PERGAMON

Computers & Fluids 28 (1999) 481–510

computers & fluids

Multi-dimensional asymptotically stable finite difference schemes for the advection–diffusion equation

Saul Abarbanel *, Adi Ditkowski

School of Mathematical Sciences, Department of Applied Mathematics, Tel-Aviv University, Tel-Aviv, Israel

Received 20 November 1997; accepted 18 May 1998

Abstract

An algorithm is presented which solves the multi-dimensional advection–diffusion equation on complex shapes to second-order accuracy and is asymptotically stable in time. This bounded-error result is achieved by constructing, on a rectangular grid, a differentiation matrix whose symmetric part is negative definite. The differentiation matrix accounts for the Dirichlet boundary condition by imposing penalty-like terms. Numerical examples in two dimensions show that the method is effective even where standard schemes, stable by traditional definitions, fail. It gives accurate, non oscillatory results even when boundary layers are not resolved.

1. Introduction

Currently there is a growing interest in long time integration for solving problems in areas such as fluid-mechanics, aero-acoustics, electro-magnetics, material-science, and others. Clearly, it will be very advantageous if one could formulate the spatial discretization in a way which guarantees that, for the semi-discrete formulation, the solution-error norm is bounded by the norm of the truncation error. Most, if not all, existing algorithms rely on stability for convergence. However, even stable schemes, which at a given time converge with mesh refinement may have a temporally growing error [1]. This is particularly true for hyperbolic operators.

* Corresponding author.Fax: +972-3-6409-357; e-mail: saul@math.tau.ac.il.

0045-7930/99/$ - see front matter
PII: S0045-7930(98)00042-5

This paper considers second-order accurate approximations to model linear advection–diffusion equations in one or more dimensions, on domains which may be irregular. By an irregular domain, we mean a body whose boundary points do not necessarily coincide with nodes of a rectangular mesh.

In Section 2 we treat a model "shock-layer" equation (linearized Burger's equation),

$$u_t + au_x = \frac{1}{R}u_{xx}; \quad t \geq 0, 0 < x < 1; \quad R \gg 1$$

We develop there the theory for the one dimensional semi-discrete system resulting from the spatial differentiation used in the finite difference algorithm. Energy methods are used in conjunction with "SAT" type terms [1, 2], in order to find boundary treatment and "artificial-viscosity-like terms", that preserve the accuracy of the scheme while constraining an energy norm of the error to be temporally bounded for all $t > 0$ by a "constant" proportional to the norm of the truncation error.

In Section 3 it is shown how the methodology developed in Section 2 is used as a building block for the multi-dimensional algorithm, even for irregular shapes.

Section 4 presents numerical results. Section 4.1 deals with the steady-state solution to the "shock-layer" equation for a large range of the "Reynolds number", R. Oscillations that appear in the numerical solution when using a standard central finite-differencing, are eliminated (or dramatically reduced) when the bounded-error algorithm is used.

Section 4.2 considers steady-state solution to a two-dimensional scalar model to the boundary layer equations,

$$u_t + au_x + bu_y = \frac{1}{R}u_{yy}; \quad R \gg 1, \quad b < 0$$

both for rectangular and trapezoidal domains. Again, the bounded-error algorithm outperforms the standard scheme in ways described therein.

Section 4.3 presents a time dependent example, modeling a boundary-layer being excited sinosoidially,

$$u_t + au_x + bu_y = \frac{1}{R}u_{yy} + \sigma b \sin[k(x - at)]$$

Here, aside from the usual performance criteria, such as error-norms and quality of the velocity profiles, we see that the error-bounded algorithm also has a significantly smaller phase error.

2. The Scalar One-Dimensional Case

Consider the scalar advection-diffusion problem

$$\frac{\partial u}{\partial t} = a\frac{\partial u}{\partial x} + \frac{1}{R}\frac{\partial^2 u}{\partial x^2} + f(x,t); \quad \Gamma_L \le x \le \Gamma_R, \quad t \ge 0, \quad a > 0^1 \tag{1a}$$

$$u(x,0) = u_0(x) \tag{1b}$$

$$u(\Gamma_L, t) = g_L(t) \tag{1c}$$

$$u(\Gamma_R, t) = g_R(t) \tag{1d}$$

[1] and $f(x, t) \in \mathcal{C}^2$.

Let us discritize Eqs. (1a)–(1 d) spatially on the following uniform grid:

Note that the boundary points, $x = \Gamma_L$ and $x = \Gamma_R$, do not necessarily coincide with x_1 and x_N. Set $x_{j+1} - x_j = h$, $1 \le j \le N-1$; $x_1 - \Gamma_L = \gamma_L h$, $0 \le \gamma_L < 1$; $\Gamma_R - x_N = \gamma_R h$, $0 \le \gamma_R \le 1$.

The projection into the above grid of the exact solution $u(x, t)$ to Eqs. (1a)–(1 d), is $u_j(t) = u(x_j, t) \triangleq \mathbf{u}(t)$. Let $\tilde{D}$ be a matrix representing $au_x + 1/Ru_{xx}$, at internal points without specifying yet how it is being constructed. Then we may write

$$\frac{\mathrm{d}}{dt}\mathbf{u}(t) = [\tilde{D}\mathbf{u}(t) + \mathbf{B} + \mathbf{T}] + \mathbf{f}(t) \tag{2}$$

where $\mathbf{T}$ is the truncation error due to the numerical differentiation, and $\mathbf{f}(t) = f(x_j, t)$, $1 \le j \le N$. The boundary vector $\mathbf{B}$ has entries whose values depend on g_L, g_R, γ_L, γ_R in such a way that $\tilde{D}\mathbf{u} + \mathbf{B}$ represents $au_x + 1/Ru_{xx}$ everywhere to the desired accuracy. The standard way of finding a numerical approximate solution to Eq. (1) is to omit $\mathbf{T}$ from Eq. (2) and solve

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v}(t) = \tilde{D}\mathbf{v}(t) + \mathbf{B} + \mathbf{f}(t) \tag{3}$$

where $\mathbf{v}(t)$ is the numerical approximation to the projection $\mathbf{u}(t)$. Subtracting Eq. (3) from Eq. (2) one gets an equation for the solution error, $\vec{\epsilon}(t) = \mathbf{u}(t) - \mathbf{v}(t)$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\vec{\epsilon} = \tilde{D}\vec{\epsilon} + \mathbf{T} \tag{4}$$

Our requirement for temporal stability is that $\|\vec{\epsilon}\|$, the L_2 norm of $\vec{\epsilon}$, be bounded by a "constant" proportional to h^m (m being the spatial order of accuracy). Note that this definition is more severe than either the G.K.S. stability criterion [3], or the definition in [1].

It can be shown that if $\tilde{D}$ is constructed in a standard manner, i.e. away from the boundaries the numerical second derivative is symmetric and the numerical first derivative is antisymmetric (and near the boundaries one uses "non-symmetric" differentiation), then there

[1] The results for the case $a < 0$ are found by an analysis anologus to the one presented in this section, and are presented in Appendix I.

are ranges of γ_R and γ_L for which $\tilde{D}$ is not negative definite. Since in the multi-dimensional case one may encounter all values of $0 \leq \gamma_L, \gamma_R \leq 1$, this is unacceptable.

The rest of this section is devoted to the construction of a scheme of second-order spatial accuracy, which is temporally stable for any γ_L, γ_R. The basic idea is to follow the procedure used in [2]. The present case is more complicated due to the difficulty in treating the advection term.

Note first that the solution projection $u_j(t)$ satisfies, besides Eq. (2), the following differential equation:

$$\frac{d\mathbf{u}}{dt} = D\mathbf{u} + \mathbf{T}_e + \mathbf{f}(t) \tag{5}$$

where now D is indeed a differentiation matrix, that does not use the boundary values and therefore $\mathbf{T}_e \neq \mathbf{T}$, but it too is a truncation error due to differentiation.

Next let the semi-discrete problem for $\mathbf{v}(t)$ be, instead of Eq. (3),

$$\frac{d\mathbf{v}}{dt} = [D\mathbf{v} - \tau_L(A_L\mathbf{v} - \mathbf{g}_L) - \tau_R(A_R\mathbf{v} - \mathbf{g}_R)] + \mathbf{f}(t) \tag{6}$$

where $\mathbf{g}_L = (1,\ldots,1)^T g_L(t)$; $\mathbf{g}_R = (1,\ldots,1)^T g_R(t)$, are vectors created from the left and right boundary values as shown. The matrices A_L and A_R are defined by the relations:

$$A_L\mathbf{u} = \mathbf{g}_L - \mathbf{T}_L, \quad A_R\mathbf{u} = \mathbf{g}_R - \mathrm{T}_\mathbf{R} \tag{7}$$

i.e. each row in A_L (A_R) is composed of the coefficients extrapolating $\mathbf{u}$ to its boundary value $\mathbf{g}_L(\mathbf{g}_R)$, at Γ_L (Γ_R) to within the desired order of accuracy (the error is then $\mathbf{T}_L(\mathbf{T}_R)$). The diagonal matrices τ_L and τ_R are given by

$$\tau_L = \mathrm{diag}(\tau_{L_1}, \tau_{L_2}, \ldots, \tau_{L_N}); \quad \tau_R = \mathrm{diag}(\tau_{R_1}, \tau_{R_2}, \ldots, \tau_{R_N}) \tag{8}$$

Subtracting Eq. (6) from Eq. (5) we get

$$\frac{d\vec{\epsilon}}{dt} = [D\vec{\epsilon} - \tau_L A_L\vec{\epsilon} - \tau_R A_R\vec{\epsilon} + \mathbf{T}_1] \tag{9}$$

where

$$\mathbf{T}_1 = \mathbf{T}_e + \tau_L\mathbf{T}_L + \tau_R\mathbf{T}_R$$

Taking the scalar product of $\vec{\epsilon}$ with Eq. (9) one gets:

$$\begin{aligned} \frac{1}{2}\frac{d}{dt}||\vec{\epsilon}||^2 &= (\vec{\epsilon}, (D - \tau_L A_L - \tau_R A_R)\vec{\epsilon}) + (\vec{\epsilon}, \mathbf{T}_1) \\ &= (\vec{\epsilon}, M\vec{\epsilon}) + (\vec{\epsilon}, \mathbf{T}_1) \end{aligned} \tag{10}$$

We notice that $(\vec{\epsilon}, M\vec{\epsilon})$ is $(\vec{\epsilon}, (M + M^T)\vec{\epsilon})/2$, where

$$M = D - \tau_L A_L - \tau_R A_R \tag{11}$$

If $(M + M^T)$ can be made negative definite then

$$(\vec{\epsilon}, (M + M^T)\vec{\epsilon})/2 \leq -c_0||\vec{\epsilon}||^2 \quad (c_0 > 0) \tag{12}$$

Eq. (10) then becomes

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}||\vec{\epsilon}||^2 \leq -c_0||\vec{\epsilon}||^2 + (\vec{\epsilon}, \mathbf{T}_1)$$

and using Schwartz's inequality we get after dividing by $\|\vec{\epsilon}\|$

$$\frac{\mathrm{d}}{\mathrm{d}t}||\vec{\epsilon}|| \leq -c_0||\vec{\epsilon}|| + ||\mathbf{T}_1||$$

and therefore (using the fact that $\mathbf{v}(0) = \mathbf{u}(0)$)

$$||\vec{\epsilon}|| \leq \frac{||\mathbf{T}_1||_M}{c_0}(1 - \mathrm{e}^{c_0 t}) \tag{13}$$

where the "*constant*" $\|\mathbf{T}_1\|_M = \max_{0 \leq \tau \leq t}\|\mathbf{T}_1(\tau)\|$.

If we indeed succeed in constructing M such that $M + M^{\mathrm{T}}$ is negative definite, with $c_0 > 0$ independent of the size of the matrix M as it increases, then it follows from Eq. (13) that the norm of the error will be bounded for all t by a constant which is $O(h^m)$ where m is the spatial accuracy of the finite difference scheme (6). The numerical solution is then temporally stable.

It can be shown that as $1/R \to 0$, so does c_0. When $c_0 = 0$, the differential inequality is

$$\frac{\mathrm{d}}{\mathrm{d}t}||\vec{\epsilon}|| \leq ||\mathbf{T}_1|| \tag{14}$$

leading to

$$||\vec{\epsilon}|| \leq ||\mathbf{T}_1||_M t \tag{15}$$

i.e. a linear growth in time, a result typical of hyperbolic systems. This result can also be obtained formally from Eq. (13) by letting $c_0 \to 0$ for *any* fixed t.

The rest of this section is devoted to the task of constructing M in the case of $m = 2$, i.e. a second-order accurate finite difference algorithm. We shall deal separately with the hyperbolic and parabolic parts of the R.H.S. of Eq. (11).

Let

$$M = \frac{1}{R}M_{\mathrm{P}} + aM_{\mathrm{H}} = \frac{1}{R}(D_{\mathrm{P}} - \tau_{\mathrm{L_P}}A_{\mathrm{L_P}} - \tau_{\mathrm{R_P}}A_{\mathrm{R_P}}) + a(D_{\mathrm{H}} - \tau_{\mathrm{L_H}}A_{\mathrm{L_H}} - \tau_{\mathrm{R_H}}A_{\mathrm{R_H}}). \tag{16}$$

The parabolic terms are given by:

$$D_{\mathrm{P}} = \frac{1}{h^2}\begin{bmatrix} 1 & -2 & 1 & 0 & & & \\ 1 & -2 & 1 & 0 & & & \\ 0 & 1 & -2 & 1 & & & \\ 0 & 0 & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & & \\ & & 1 & -2 & 1 & 0 & 0 \\ & & & 1 & -2 & 1 & 0 \\ & & & & 1 & -2 & 1 \\ & & & & 1 & -2 & 1 \end{bmatrix} \tag{17}$$

$$\tau_{\mathrm{L_P}} = \frac{1}{h^2}\,\mathrm{diag}[\tau^{(\mathrm{P})}_{\mathrm{L}_1}, 0, \ldots 0] = \frac{1}{h^2}\,\mathrm{diag}\left[\frac{4}{(2+\gamma_{\mathrm{L}})(1+\gamma_{\mathrm{L}})}, 0, \ldots, 0\right] \tag{18}$$

$$\tau_{\mathrm{R_P}} = \frac{1}{h^2}\,\mathrm{diag}[0, 0, \ldots, \tau^{(\mathrm{P})}_{\mathrm{R}_N}] = \frac{1}{h^2}\,\mathrm{diag}\left[0, 0, \ldots, \frac{4}{(2+\gamma_{\mathrm{R}})(1+\gamma_{\mathrm{R}})}\right] \tag{19}$$

$$A_{\mathrm{L_P}} = \begin{bmatrix} \frac{1}{2}(2+\gamma_{\mathrm{L}})(1+\gamma_{\mathrm{L}}) & -\gamma_{\mathrm{L}}(2+\gamma_{\mathrm{L}}) & \frac{1}{2}(\gamma_{\mathrm{L}}+\gamma_{\mathrm{L}}^2) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{2}(2+\gamma_{\mathrm{L}})(1+\gamma_{\mathrm{L}}) & -\gamma_{\mathrm{L}}(2+\gamma_{\mathrm{L}}) & \frac{1}{2}(\gamma_{\mathrm{L}}+\gamma_{\mathrm{L}}^2) & 0 & \cdots & 0 \end{bmatrix} \tag{20}$$

$$A_{\mathrm{R_P}} = \begin{bmatrix} 0 & \cdots & 0 & \frac{1}{2}(\gamma_{\mathrm{R}}+\gamma_{\mathrm{R}}^2) & -\gamma_{\mathrm{R}}(2+\gamma_{\mathrm{R}}) & \frac{1}{2}(2+\gamma_{\mathrm{R}})(1+\gamma_{\mathrm{R}}) \\ \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \frac{1}{2}(\gamma_{\mathrm{R}}+\gamma_{\mathrm{R}}^2) & -\gamma_{\mathrm{R}}(2+\gamma_{\mathrm{R}}) & \frac{1}{2}(2+\gamma_{\mathrm{R}})(1+\gamma_{\mathrm{R}}) \end{bmatrix} \tag{21}$$

The hyperbolic terms are given by:

$$D_{\mathrm{H}} = \frac{1}{2h}\left\{ \begin{bmatrix} -2 & 2 & & & & & & \\ -1 & 0 & 1 & & & & & \\ & -1 & 0 & 1 & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & -1 & 0 & 1 & 0 & \\ & & & & -1 & 0 & 1 & \\ & & & & & -2 & 2 & \end{bmatrix} + \begin{bmatrix} c_1 & & & & & & & \\ & c_2 & & & & & & \\ & & c_3 & & & & & \\ & & & \ddots & & & & \\ & & & & c_{N-2} & & & \\ & & & & & c_{N-1} & & \\ & & & & & & c_N & \end{bmatrix} \right.$$
$$\left. \times \begin{bmatrix} -1 & 2 & -1 & & & & & \\ 0 & -1 & 2 & -1 & & & & \\ 1 & -2 & 0 & 2 & -1 & & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & 1 & -2 & 0 & 2 & -1 & \\ & & & 1 & -2 & 1 & 0 & \\ & & & & 1 & -2 & 1 & \end{bmatrix} + 2h\tilde{c} \begin{bmatrix} 0 & -1 & 1 & & & & & \\ 1 & -1 & -1 & 1 & & & & \\ 1 & -1 & 0 & -1 & 1 & & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & 1 & -1 & 0 & -1 & 1 & \\ & & & & 1 & -1 & -1 & 1 \\ & & & & & 1 & -1 & 0 \end{bmatrix} \right\} \tag{22}$$

where

$$c_k = \frac{1}{N-1}[(c_N - c_1)k + (Nc_1 - c_N)] \tag{23}$$

and

$$\tilde{c} = \frac{1}{2}(c_1 - c_N) \tag{24}$$

For $a > 0$ in Eq. (1)a), the left boundary is, for the hyperbolic part, an "outflow" boundary on which we do not prescribe a "hyperbolic boundary condition", therefore, in this case $\tau_{\mathrm{L_H}} = 0$. When $a < 0$, then $\tau_{\mathrm{R_H}} = 0$—see Appendix I for details.

Here, with $a > 0$,

$$\tau_{\mathrm{R_H}} = \frac{1}{2h}\mathrm{diag}[0, 0, \ldots, \tau^{(\mathrm{H})}_{\mathrm{R}_{N-1}}, \tau^{(\mathrm{H})}_{\mathrm{R}_N}] \tag{25}$$

and

$$A_{\mathrm{R_H}} = \begin{bmatrix} 0 & & & & \\ & \ddots & & & 0 \\ & & 0 & & \\ 0 & & & -\gamma_{\mathrm{R}} & 1+\gamma_{\mathrm{R}} \\ & & & -\gamma_{\mathrm{R}} & 1+\gamma_{\mathrm{R}} \end{bmatrix} \tag{26}$$

Next we shall show that the parabolic part of M is negative definite. The symmetric part of M_{P}, $\tilde{M}_{\mathrm{P}} = 1/2(M_{\mathrm{P}} + M_{\mathrm{P}}^{\mathrm{T}})$, is found using Eqs. (17)–(21), to be

$$\tilde{M}_{\mathrm{P}} = \frac{1}{2h^2}\begin{bmatrix} -2 & \frac{3\gamma_{\mathrm{L}}-1}{\gamma_{\mathrm{L}}+1} & \frac{2-\gamma_{\mathrm{L}}}{2+\gamma_{\mathrm{L}}} & & & & & & & \\ \frac{3\gamma_{\mathrm{L}}-1}{\gamma_{\mathrm{L}}+1} & -4 & 2 & & & 0 & & & & \\ \frac{2-\gamma_{\mathrm{L}}}{2+\gamma_{\mathrm{L}}} & 2 & -4 & 2 & & & & & & \\ & & 2 & -4 & 2 & & & & & \\ & & & \ddots & \ddots & \ddots & & & & \\ & & & & 2 & -4 & 2 & & & \\ 0 & & & & & 2 & -4 & 2 & \frac{2-\gamma_{\mathrm{R}}}{2+\gamma_{\mathrm{R}}} & \\ & & & & & & 2 & -4 & \frac{3\gamma_{\mathrm{R}}-1}{\gamma_{\mathrm{R}}+1} & \\ & & & & & & \frac{2-\gamma_{\mathrm{R}}}{2+\gamma_{\mathrm{R}}} & \frac{3\gamma_{\mathrm{R}}-1}{\gamma_{\mathrm{R}}+1} & -2 & \end{bmatrix} \tag{27}$$

We now decompose $\tilde{M}_{\mathrm{P}}$ as follows:

$$
\tilde{M}_{\mathrm{P}} = \frac{1}{2h^2}\left\{ \alpha \begin{bmatrix}
-4 & 2 & & & & & & \\
2 & -4 & 2 & & & & & \\
 & 2 & -4 & 2 & & & & \\
 & & \ddots & \ddots & \ddots & & & \\
 & & & 2 & -4 & 2 & & \\
 & & & & 2 & -4 & 2 & \\
 & & & & & 2 & -4 & 2 \\
 & & & & & & 2 & -4
\end{bmatrix} \right.
$$

$$
+ (1-\alpha) \begin{bmatrix}
0 & 0 & 0 & & & & & & \\
0 & 0 & 0 & & & & & & \\
0 & 0 & -2 & 2 & & & & & \\
 & & 2 & -4 & 2 & & & & \\
 & & & \ddots & \ddots & \ddots & & & \\
 & & & & 2 & -4 & 2 & 0 & \\
 & & & & 0 & 2 & -2 & 0 & 0 \\
 & & & & & & 0 & 0 & 0 \\
 & & & & & & 0 & 0 & 0
\end{bmatrix}
$$

$$
\left. + \begin{bmatrix}
-2(1-2\alpha) & \frac{3\gamma_{\mathrm{L}}-1}{\gamma_{\mathrm{L}}+1} - 2\alpha & \frac{2-\gamma_{\mathrm{L}}}{2+\gamma_{\mathrm{L}}} & & & & & & \\
\frac{3\gamma_{\mathrm{L}}-1}{\gamma_{\mathrm{L}}+1} - 2\alpha & -4(1-\alpha) & 2(1-\alpha) & & & & & & \\
\frac{1-\gamma_{\mathrm{L}}}{2+\gamma_{\mathrm{L}}} & 2(1-\alpha) & -2(1-\alpha) & & & & & & 0 \\
 & & & 0 & & & & & \\
 & & & & 0 & & & & \\
0 & & & & & -2(1-\alpha) & 2(1-\alpha) & \frac{2-\gamma_{\mathrm{R}}}{2+\gamma_{\mathrm{R}}} & \\
 & & & & & 2(1-\alpha) & -4(1-\alpha) & \frac{3\gamma_{\mathrm{R}}-1}{\gamma_{\mathrm{R}}+1} - 2\alpha & \\
 & & & & & \frac{2-\gamma_{\mathrm{R}}}{2+\gamma_{\mathrm{R}}} & \frac{3\gamma_{\mathrm{R}}-1}{\gamma_{\mathrm{R}}+1} - 2\alpha & -2(1-2\alpha)
\end{bmatrix} \right\}
\tag{28}
$$

We look for $1 > \alpha > 0$ such that the second and third matrices in Eq. (28) are non-positive definite. The first matrix in Eq. (28) is already negative definite by the argument leading to Eq. (60), in [2]. By the same argument it immediately follows that its largest eigenvalue is smaller than $-\alpha\pi^2$. For $0 < \alpha < 1$, the second matrix in Eq. (28) is non-positive definite, see Eqs. (63) and (64) in [2]. The third matrix in Eq. (28) has two square 3×3 corners which are negative for $0 < \alpha < 0.275$. This completes the proof that $\tilde{M}_{\mathrm{P}}$ is indeed negative definite.

Next we would like to show that $\tilde{M}_{\mathrm{H}} = 1/2(M_{\mathrm{H}} + M_{\mathrm{H}}^{\mathrm{T}})$ is non-positive definite. Using Eqs. (22)–(26) we have

$$\tilde{M}_{\mathrm{H}} = \frac{1}{4h}\begin{bmatrix} -4-2c_1 & 1+2c_1 & 0 & & & & & \\ 1+2c_1 & -2c_1 & 0 & & & & & \\ & & & & & & 0 & \\ 0 & 0 & 0 & & & & & \\ & & & 0 & & & & \\ & & & & \ddots & & & \\ & 0 & & & 2c_N + 2\gamma_{\mathrm{R}}\tau_{N-1}^{(\mathrm{H})} & & -1 & -2c_N - (1+\gamma_{\mathrm{R}})\tau_{N-1}^{(\mathrm{H})} + \gamma_{\mathrm{R}}\tau_N^{(\mathrm{H})} \\ & & & -1 & -2c_N - (1+\gamma_{\mathrm{R}})\tau_{N-1}^{(\mathrm{H})} + \gamma_{\mathrm{R}}\tau_N^{(\mathrm{H})} & & & 4 + 2c_N - 2(1+\gamma_{\mathrm{R}})\tau_N^{(\mathrm{H})} \end{bmatrix} \tag{29}$$

We now write $\tilde{M}_{\mathrm{H}}$ as the sum of three "corner-matrices",

$$\tilde{M}_{\mathrm{H}} = \frac{1}{4h}[m_{\mathrm{H}_1} + m_{\mathrm{H}_2} + m_{\mathrm{H}_3}] \tag{30}$$

where

$$m_{\mathrm{H}_1} = \begin{bmatrix} -4-2v_1 & 1+2c_1 & & & \\ & & 0 & & \\ 1+2c_1 & -2c_1 & & & \\ & & 0 & & \\ & 0 & & \ddots & \\ & & & & 0 \end{bmatrix}$$

$$m_{\mathrm{H}_2} = \begin{bmatrix} 0 & & & \\ & 0 & & 0 \\ & \ddots & & \\ & 2\gamma_{\mathrm{R}}\tau_{N-1}^{(\mathrm{H})} & & -1-(1+\gamma_{\mathrm{R}})\tau_{N-1}^{(\mathrm{H})} + \gamma_{\mathrm{R}}\tau_N^{(\mathrm{H})} \\ 0 & & & \\ & -1-(1+\gamma_{\mathrm{R}})\tau_{N-1}^{(\mathrm{H})} + \gamma_{\mathrm{R}}\tau_N^{(\mathrm{H})} & & 4-2(1+\gamma_{\mathrm{R}})\tau_N^{(\mathrm{H})} \end{bmatrix} \tag{31}$$

$$m_{\mathrm{H}_3} = c_N \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & \ddots & \\ & & 2 & -2 \\ & & -2 & 2 \end{bmatrix}$$

Clearly m_{H_3} is N.P.D. (non-positive definite) for $\forall c_N \leq 0$. Also, m_{H_1} is N.P.D. for $c_1 \geq 1/4$. A

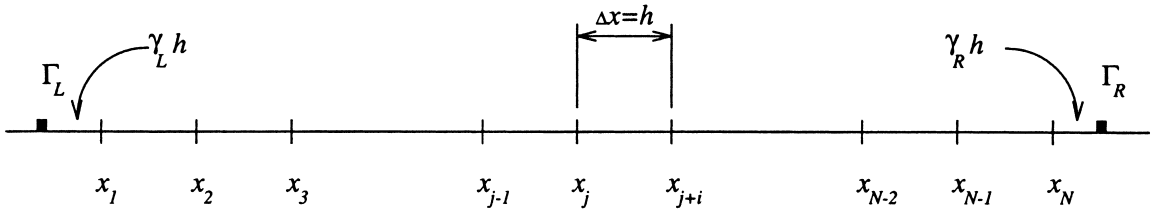

Fig. 1. One-dimensional grid.

simple computation shows that m_{H_2} is N.P.D. if τ_{N-1} and τ_N satisfy

$$\tau_N^{(\mathrm{H})} = \frac{2+\delta}{1+\gamma_{\mathrm{R}}} \quad (\delta \geq 0) \tag{32}$$

$$\tau_{N-1}^{(\mathrm{H})} = -\frac{1-\gamma_{\mathrm{R}}(1-\delta)}{(1+\gamma_{\mathrm{R}})^2} \tag{33}$$

Thus we have proved that $\tilde{M}_{\mathrm{H}}$ is indeed non-positive definite, and therefore $\tilde{M} = 1/R\tilde{M}_{\mathrm{P}} + a\tilde{M}_{\mathrm{H}}$ is negative definite for $\forall 1/R$, $a > 0$, with its eigenvalues bounded away below zero by $-\alpha\pi^2/R$, $0 < \alpha < 0.275$.

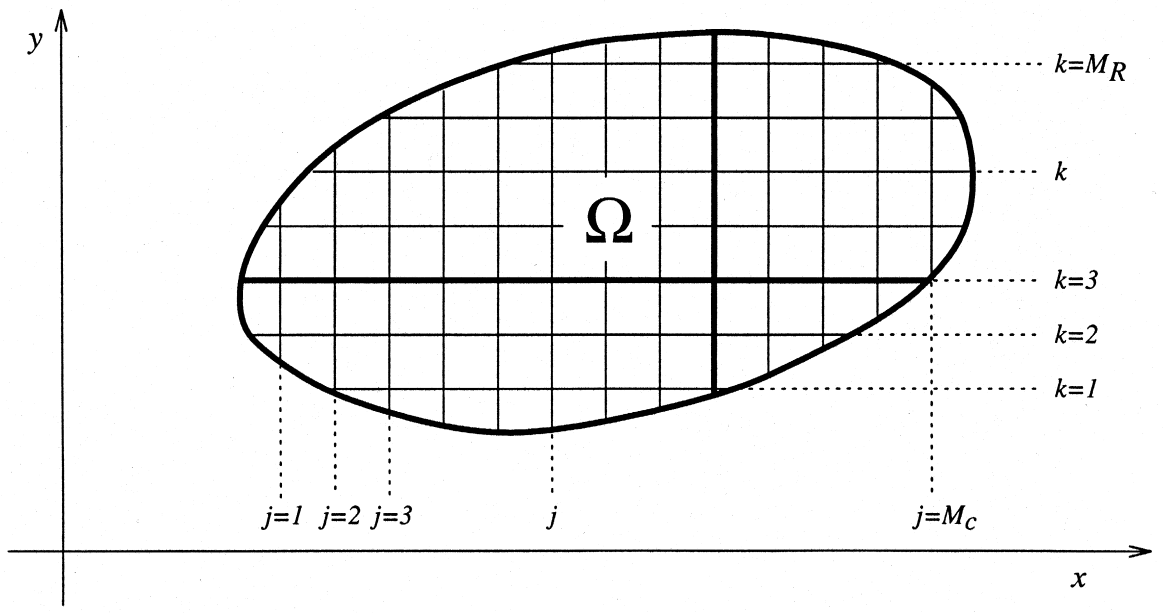

Fig. 2. Two-dimensional grid.

3. The Scalar Two-Dimensional Case

We consider an inhomogeneous advection–diffusion equation, with constant coefficients, in a domain Ω. To begin with we shall assume that Ω is convex and has a boundary $\partial\Omega \in C^2$. The convexity restriction is for the sake of simplicity in presenting the basic idea; it will be removed later. The problem statement is:

$$\frac{\partial u}{\partial t} = a\frac{\partial u}{\partial x} + b\frac{\partial u}{\partial y} + v_1\frac{\partial^2 u}{\partial x^2} + v_2\frac{\partial^2 u}{\partial y^2} + f(x, y; t); \quad t > 0, \quad v_1, v_2 > 0 \tag{34a}$$

$$u(x, y, 0) = u_0(x, y) \tag{34b}$$

$$u(x, y, t)|_{\partial\Omega} = u_B(t) \tag{34C}$$

We shall refer to the grid representation in Fig. 2

We have M_R rows and M_C columns inside Ω. Each row and each column has a discretized structure as in the one-dimensional case, see Fig. 1. Let the number of grid points in the kth row be denoted by R_k and similarly let the number of points in the jth column be C_j. Let the solution projection be designated by $u_{j,k}(t)$. By $\mathbf{U}(t)$ we mean, by analogy to the one-dimensional case,

$$\begin{aligned}\mathbf{U}(t) &= (u_{1,1}, u_{2,1}, \ldots, u_{R_1,1}; u_{1,2}, u_{2,2}, \ldots, u_{R_2,2}; \ldots; u_1, M_R, u_2, M_R, \ldots, u_{R_{MR}}, M_R) \\ &\equiv (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{M_R})\end{aligned} \tag{35}$$

Thus, we have arranged the solution projection in vectors according to rows, starting from the bottom of Ω.

If we arrange this array by columns (instead of rows) we will have the following structure,

$$\begin{aligned}\mathbf{U}^{(C)}(t) &= (u_{1,1}u_{1,2}, \ldots, u_1, C_1; u_{2,1}, u_{2,2} \ldots, u_2, C_2; \ldots; U_{M_C,1}, u_{M_C,2}, \ldots, u_{M_C}, C_{M_C}) \\ &\equiv (\mathbf{u}_1^{(C)}, \mathbf{u}_2^{(C)}, \ldots, \mathbf{u}_{M_C}^{(C)}).\end{aligned} \tag{36}$$

Clearly

$$\mathbf{U}^{(C)}(t) = P\mathbf{U} \tag{37}$$

where P is an orthogonal permutation matrix, of order $l \times l$, l being the number of grid points within Ω.

The operator $v_1\ \partial^2/\partial x^2 + a\ \partial/\partial x$ in Eq. (34), including the boundary terms, is represented on the kth row by $M_k^{(x)}$, whose structure is given by Eq. (16) and the definition following it Eqs. (see (17)–(26)). Similarly let $M_j^{(y)}$ represent $v_2\ \partial^2/\partial y^2 + b\ \partial/\partial y$ on the jth column. With this notation, by analogy to Eq. (6), the two dimensional semi-discrete problem becomes

$$\frac{d\mathbf{V}}{dt} = (\mathcal{M}^{(x)} + P^T\mathcal{M}^{(y)}P)\mathbf{V} + \mathbf{G}^{(x)} + P^T\mathbf{G}^{(y)} + \mathbf{f}(t) \tag{38}$$

where $\mathbf{V}$ is the numerical approximation of $\mathbf{U}$

$$\mathcal{M}^{(x)} = \begin{bmatrix} M_1^{(x)} & & & & \\ & \ddots & & & \\ & & M_k^{(x)} & & \\ & & & \ddots & \\ & & & & M_{M_R}^{(x)} \end{bmatrix}; \mathcal{M}^{(y)} = \begin{bmatrix} M_1^{(y)} & & & & \\ & \ddots & & & \\ & & M_j^{(y)} & & \\ & & & \ddots & \\ & & & & M_{M_C}^{(y)} \end{bmatrix} \tag{39}$$

and

$$\begin{aligned} \mathbf{G}^{(x)} &= G_P^{(x)} + G_H^{(x)} \\ &= [(\tau_{L_1}^{(P)}\mathbf{g}_{L_1} + \tau_{R_1}^{(P)}\mathbf{g}_{R_1}), \ldots, (\tau_{L_k}^{(P)}\mathbf{g}_{L_k} + \tau_{R_k}^{(P)}\mathbf{g}_{R_k}), \ldots, (\tau_{L_{M_R}}^{(P)}\mathbf{g}_{L_{M_R}} + \tau_{R_{M_R}}^{(P)}\mathbf{g}_{R_{M_R}})] + [(\tau_{L_1}^{(H)}\mathbf{g}_{L_1} \\ &\quad + \tau_{R_1}^{(H)}\mathbf{g}_{R_1}), \ldots, (\tau_{L_k}^{(H)}\mathbf{g}_{L_k} + \tau_{R_k}^{(H)}\mathbf{g}_{R_k}), \ldots, (\tau_{L_{M_R}}^{(H)}\mathbf{g}_{L_{M_R}} + \tau_{R_{M_R}}^{(H)}\mathbf{g}_{R_{M_R}})] \end{aligned}$$

$$\begin{aligned} \mathbf{G}^{(y)} &= G_P^{(y)} + G_H^{(y)} \\ &= [(\tau_{B_1}^{(P)}\mathbf{g}_{B_1} + \tau_{T_1}^{(P)}\mathbf{g}_{T_1}), \ldots, (\tau_{B_j}^{(P)}\mathbf{g}_{B_j} + \tau_{T_j}^{(P)}\mathbf{g}_{T_j}), \ldots, (\tau_{B_{M_C}}^{(P)}\mathbf{g}_{B_{M_C}} + \tau_{T_{M_C}}^{(P)}\mathbf{g}_{T_{M_C}})] \\ &\quad + [(\tau_{B_1}^{(H)}\mathbf{g}_{B_1}\tau_{T_1}^{(H)}\mathbf{g}_{T_1}), \ldots, (\tau_{B_j}^{(H)}\mathbf{g}_{B_j} + \tau_{T_j}^{(H)}\mathbf{g}_{T_j}), \ldots, (\tau_{B_{M_C}}^{(H)}\mathbf{g}_{B_{M_C}} + \tau_{T_{M_C}}^{(H)}\mathbf{g}_{T_{M_C}})] \end{aligned} \tag{40}$$

The subscripts B_j (“B“ for bottom) play the same role as L_k (“L“ for left). The same remark applied to subscripts T_j (“T“ for top) and R_k (“R“ for right).

Note that $\tau_{L_k}^{H}(\tau_{R_k}^{H}) = 0$ when $a > 0 (a < 0)$. Similarly $\tau_{B_j}^{H}(\tau_{T_j}^{H}) = 0$ when $b > 0 (b < 0)$.

Designating the two dimensional array of errors, ϵ_{ij}, by $\mathbf{E} = \mathbf{U} - \mathbf{V}$, the equation for $\mathbf{E}$ becomes

$$\frac{d\mathbf{E}}{dt} = [\mathcal{M}^{(x)} + P^T\mathcal{M}^{(y)}P]\mathbf{E} + \mathbf{T} \tag{41}$$

where $\mathbf{T}$ represents the sum of the various truncation errors.

The time rate of change of $\|\mathbf{E}\|^2$ is given by

$$\frac{1}{2}\frac{d}{dt}||\mathbf{E}||^2 = (\mathbf{E}, (\mathcal{M}^{(x)} + P^T\mathcal{M}^{(y)}P)\mathbf{E}) + (\mathbf{E}, \mathbf{T}) \tag{42}$$

By the same arguments that follow Eq. (3).15) in [2] it is clear that the norm of the error, $\|\mathbf{E}\|$, is bounded by a constant, where the “constant” $\|\mathbf{T}\|_M = \max_{0 \le \tau \le t}\|\mathbf{T}(\tau)\|$.

In [2], it was shown that if the domain Ω is not convex or simply connected, the above results still hold. This is also true here.

Note that if $1/R = \nu = 0$ (or $\nu_1 = \nu_2 = 0$ in the two-dimensional case) then the differentiation operator, M, becomes non-positive definite. In that case, it follows immediately from Eq. (42) that the bound on the error-norm is not a “constant” but grows *linearly* in time.

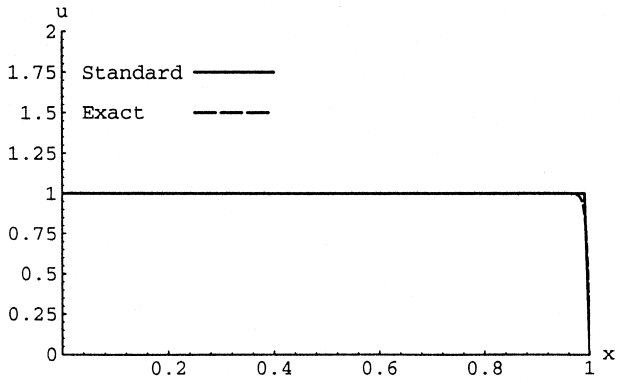

Fig. 3. Standard scheme, $R_C = 2$.

4. Numerical Examples

4.1. One-dimensional case

Here we consider the problem

$$\frac{\partial u}{\partial t} + u_x = \frac{1}{R} u_{xx}, \quad t \geq 0, \quad 0 \leq x \leq 1 \tag{43}$$

$$\begin{aligned} u(0, t) &= 1 \\ u(1, t) &= 0 \\ u(x, 0) &= u_0(x) \end{aligned}$$

The steady state solution to (43) is:

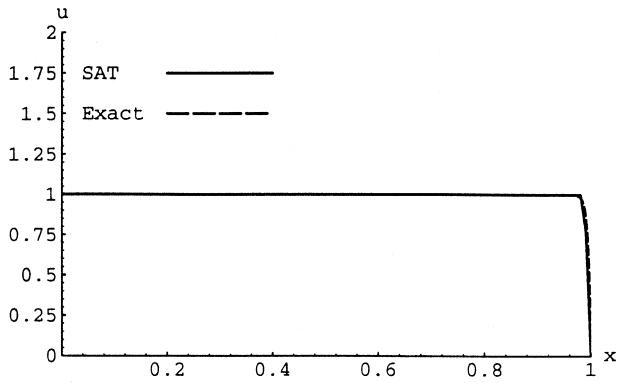

Fig. 4. SAT, $R_C = 2$.

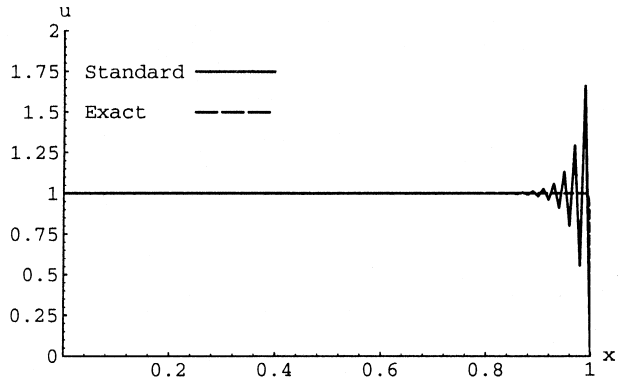

Fig. 5. Standard scheme, $R_C = 10$.

$$u(x) = \frac{1 - e^{-R(1-x)}}{1 - e^{-R}} \tag{44}$$

Note that R $(=1/\nu)$ plays the role of Reynolds number in this model for a "linear shock layer".

Eq. (43) was solved numerically by two methods. In one (referred to as "standard") we use central differencing for the spatial differentiation, and fourth-order Runge–Kutta in time. In this "standard" case, there is no need for special treatment at the boundaries.

The numerical approximation $\mathbf{v}$, in this "standard" case, satisfies the following finite difference equation:

$$\frac{1}{2h}(v_{j+1} - v_{j-1}) - \frac{1}{Rh^2}(v_{j+1} - 2v_j + v_{j-1}) = 0 \quad (0 \le j \le n) \tag{45}$$

with $v_0 = 1$ and $v_N = 0$. The solution to Eq. (45) is:

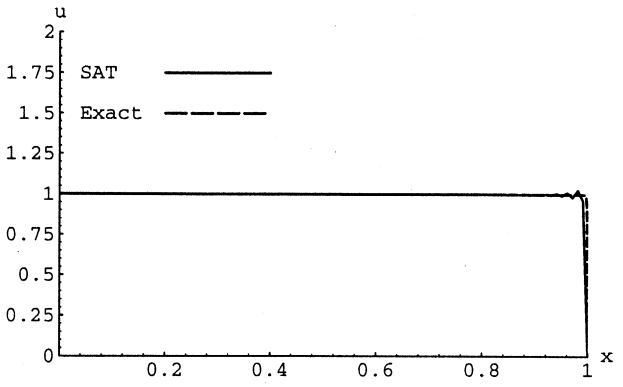

Fig. 6. SAT, $R_C = 10$.

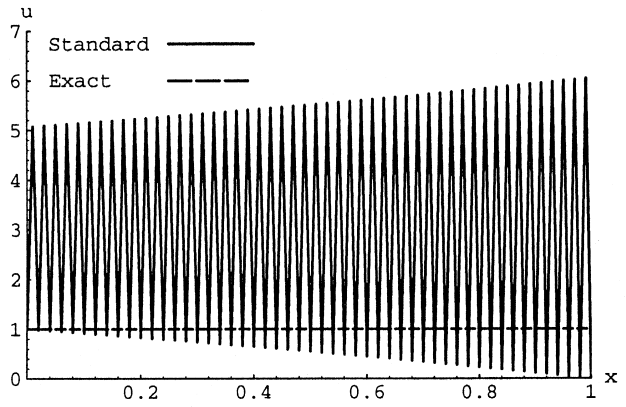

Fig. 7. Standard scheme, $R_C = 1000$.

$$v_j = \frac{\kappa^j - \kappa^{2N-j}}{1 - \kappa^{2N}}, \quad \kappa = \frac{2 + hR}{2 - hR} \tag{46}$$

Notice, that if the "cell Reynolds number," $R_C = hR > 2$, then $\kappa < 0$ and the numerical solution, v_j, will be oscillatory. If $R_C < 2$ then we resolve the "shock layer" (or "boundary layer") and the solution will be smooth.

Numerical steady-state solutions of Eq. (43) using the "standard scheme", and using the "bounded-error" algorithm, Eq. (6), described above are shown in Figs. 3–8 for $\Delta x = 1/100$ and various values of R. Both schemes were advanced to steady state using fourth-order Runge–Kutta. It is clear that when $R_C < 2$, both schemes give good results. For $R_C = 10$ ($R = 1000$) both show oscillations, but the new algorithm approximates the exact solution much better. When $R_C = 10^3$ ($R = 10^5$), the "standard" numerical solution is useless while the "bounded-error" scheme gives excellent results; in fact far better than for $R_C = 10$.

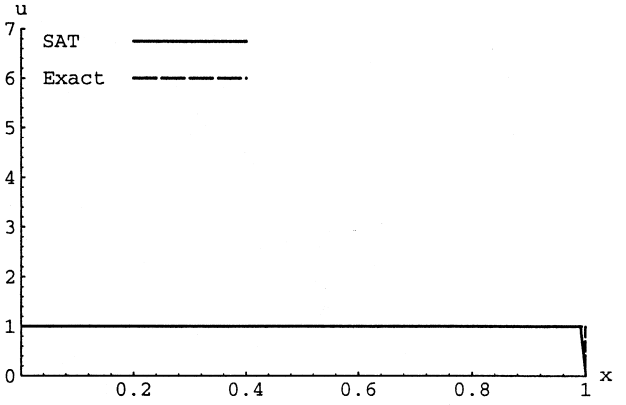

Fig. 8. SAT, $R_C = 1000$.

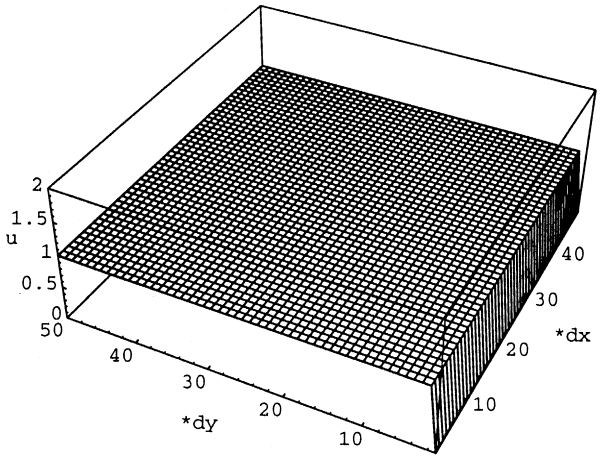

Fig. 9. Exact solution.

4.2. A steady-state two-dimensional case

Here we shall consider a steady-state problem, which models, in a way, the two-dimensional boundary layer equations. The formulation is as follows (the time derivative is left in the equation, since the approach to steady state will be via temporal advance):

$$u_t + au_x + bu_y = \frac{1}{R}u_{yy}; \quad t \geq 0;\ 0 \leq x < 1;\ 0 \leq y \leq 1 \tag{47}$$

$$u(0, y, t) = \frac{1 - e^{bRy}}{1 - e^{bR}} + \frac{1}{10} bR e^{bRy/2} \sin \pi y \tag{47a}$$

$$u(x, 0, t) = 0 \tag{47b}$$

$$u(x, 1, t) = 1 \tag{47c}$$

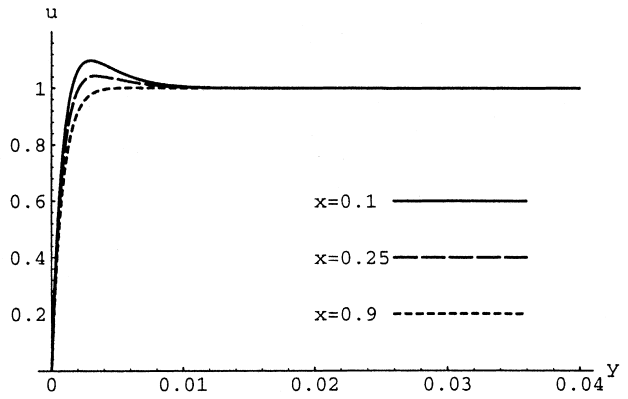

Fig. 10. Exact solution near the boundary.

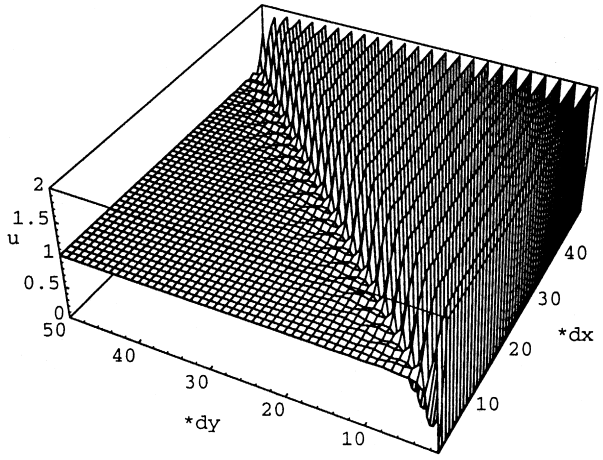

Fig. 11. Standard scheme, $b = -1$.

We also take $a = 1$, and in order to have a growing "boundary layer" on $y = 0$, we must set $b < 0$.

The analytic solution of this problem is:

$$u(x, y) = \frac{1 - e^{bRy}}{1 - e^{bR}} + \frac{1}{10} bR e^{bRy/2} \exp\left[\left(-\frac{b^2R^2}{4} - \pi^2\right)\frac{x}{Ra}\right] \sin \pi y \tag{48}$$

Fig. 9 is a three-dimensional rendition of $u(x, y)$ for $R = 90{,}000$. (This three-dimensional plot looks the same to the eye for various $-1 < b < -4/\sqrt{R} = -4/300$.) Fig. 10 is a plot of the "velocity profile" inside the "boundary-layer" ($0 < y < 0.04$) at $x = 0.1$, 0.25, 0.9 and $b = -4/\sqrt{R}$. The "bumps" at $x = 0.1$ and $x = 0.25$ may be considered as "emulating" results of fluid mechanics computation for an incompressible flow near the entrance to a channel, see e.g. [4].

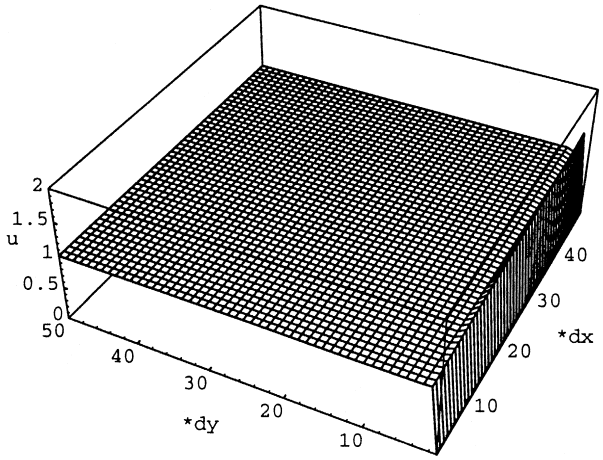

Fig. 12. Standard scheme, $b = -4/300$.

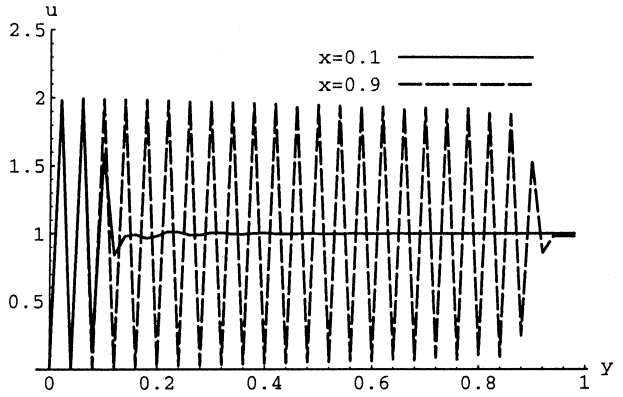

Fig. 13. Standard scheme, $b = -1$.

The numerical solution of Eq. (47) using a standard central differencing scheme depends strongly on the value of b (at a given R). Figs. 11 and 12 show the three-dimensional plot of $v_{j,k}$ with $b = -1$ and $b = -4/\sqrt{R} = -4/300$. Figs. 13 and 14 show the profiles at $x = 0.1$ and $x = 0.9$ for $b = -1$ and $-4/300$, respectively. It should be emphasized that the "peak" in Fig. 11 has nothing to do with the "bumps" in the exact solution (see Fig. 10). The "peak" occurs way outside the boundary layer, and also the amplitude behavior with the x-coordinate is counter to that of Fig. 10. The "peak" is due to a purely numerical oscillation.

The same series of plots, but as computed by the new algorithm, is shown in Figs. 15–18.

It should be noted (see Table 1) that the "bounded-error" algorithm converges to steady state (residual L_2 norm $< 10^{-13}$) an order of magnitude faster than the standard scheme when using the same Δt, while CPU-time/iteration is about the same. The standard scheme may be run at bigger Δt (by about a factor of 2) while the SAT algorithm was already at its maximum CFL number. If we let each scheme run at its own maximum Δt then the run times are about equal, but the difference in errors remains.

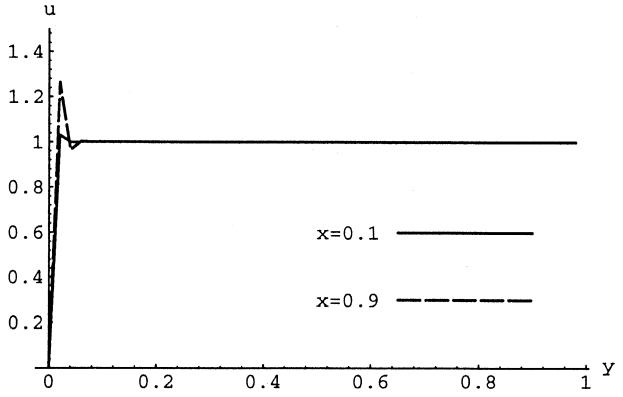

Fig. 14. Standard scheme, $b = -4/300$.

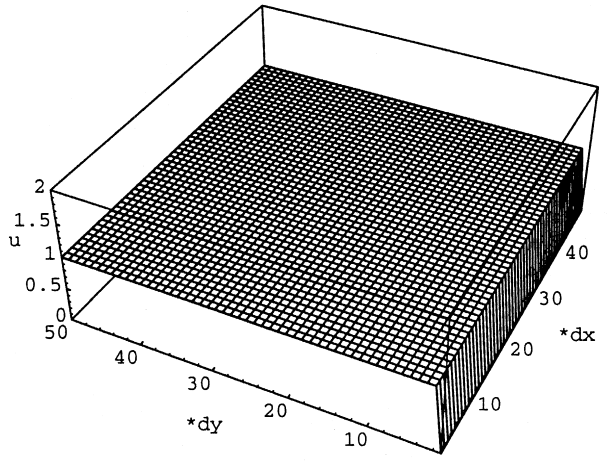

Fig. 15. SAT, $b = -1$.

We also ran the same equations for a non-strictly rectangular geometry, where the upper boundary instead of being $y = 1$ is $y = 1-(\tan\ \theta)x$, where θ is the angle which the upper boundary makes with the x-axis, see Fig. 19.

For many θs the results of the performance of the two schemes are unaffected by the change. However, there are some θs for which the standard scheme converges to steady state much slower than before at its own maximum allowed Δt, while the performance of the bounded-error algorithm remains the same as before. For example, see Table 2, for the case of $\theta = 3.9°$. As in [2], the point is that for non-rectangular geometry the distance that a boundary is away from a computational mode, γh, might become extremely small and this causes the deterioration in the performance of the standard scheme. Here it is reflected in the fact that the standard scheme cannot "support" the larger allowed Δt that can be achieved for the case

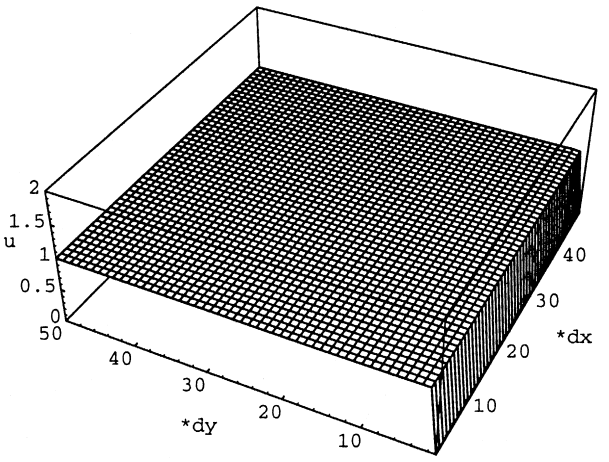

Fig. 16. SAT, $b = -4/300$.

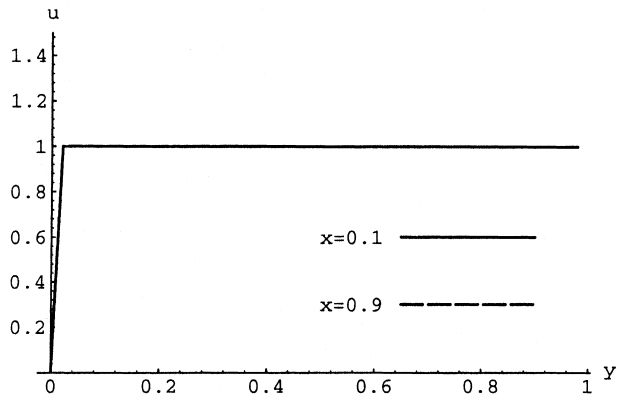

Fig. 17. SAT, $b=-1$.

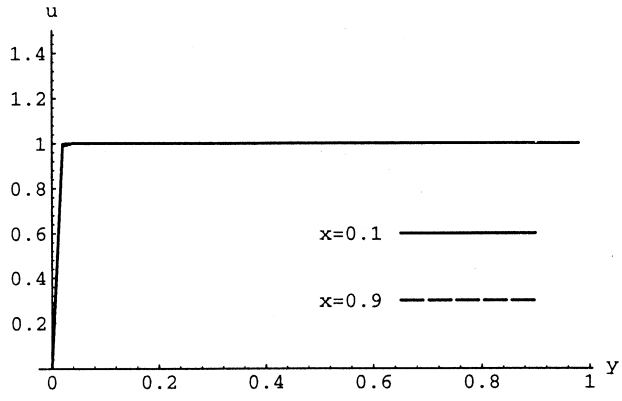

Fig. 18. SAT, $b=-4/300$.

Table 1
Rectangular geometry results

	Time to steady-state	L_2 residual	L_1 norm of the error	L_2 norm of the error	L_∞ norm of the error	Max error location
$b=-1$						
SAT	21.09	9.911×10^{-14}	8.805×10^{-05}	1.076×10^{-04}	3.108×10^{-04}	45, 46
Standard	417	9.987×10^{-14}	0.485139	0.674233	−1.00423	10, 4
$b=-4/300$						
SAT	52.64	9.943×10^{-14}	1.665×10^{-04}	1.142×10^{-03}	0.01220	50, 2
Standard	416	9.967×10^{-14}	3.362×10^{-03}	2.447×10^{-02}	−0.2864	50, 2

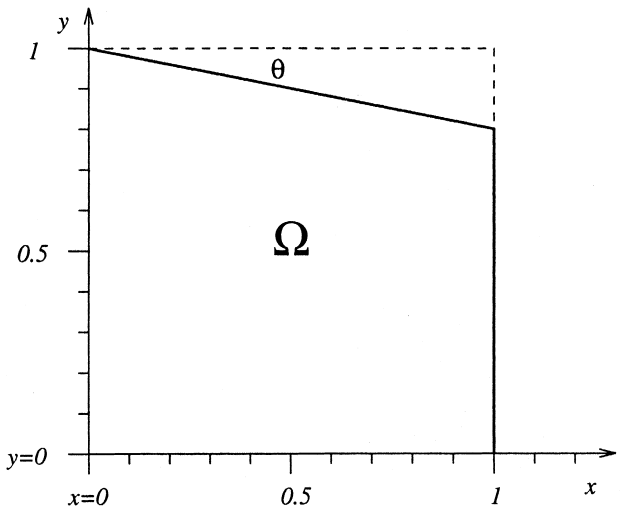

Fig. 19. The trapezoid geometry.

$\theta = 0$. For complex geometries it is very difficult to predict a priori what range the values of γ will take. The SAT methods (the bounded error algorithm) are insensitive to the variations in γ caused by the geometry of the domain.

4.3. A two-dimensional time dependent example

To check on the temporal "performance" of the bounded-error scheme, we considered the following problem:

$$u_t + au_x + bu_y = \frac{1}{R}u_{yy} + \sigma b \sin[k(x-at)];\ t \geq 0, \quad 0 \leq x < 1, \quad 0 \leq y \leq 1 \tag{49a}$$

$$u(x, y, 0) = \frac{1 - e^{bRy}}{1 - e^{bR}} + \frac{bR}{10} e^{bRy/2} e^{-(b^2R^2/4+\pi^2)x/Ra} \sin \pi y + y\sigma \sin kx \tag{49b}$$

$$u(0, y, t) = \frac{1 - e^{bRy}}{1 - e^{br}} + \frac{bR}{10} e^{bRy/2} \sin \pi y - y\sigma \sin kat \tag{49c}$$

$$u(x, 0, t) = 0 \tag{49d}$$

$$u(x, 1, t) = 1 + \sigma \sin[k(x-at)] \tag{49e}$$

The exact solution of Eq. (49) is:

$$u(x, y, t) = \frac{1 - e^{bRy}}{1 - e^{bR}} + \frac{bR}{10} e^{bRy/2} e^{-(b^2R^2/4+\pi^2)x/Ra} \sin \pi y + y\sigma \sin[k(x-at)] \tag{50}$$

Again we take $a = 1$, $R = 90{,}000$, $b = -1$,and $-4/\sqrt{R}$. The parameters σ and k have certain constraints. If we want $u > 0$, we must take $\sigma < 1$. The number of computational nodes, N, puts a lower bound of $2\pi N$ on the wavelength, $1/k$, i.e. $1 < k < 2\pi N$. In the actual computations we used $\sigma = 1/2$ and $k = 30$. All the plots for this time dependent case are shown

Table 2
Trapezoid geometry results

	Time to steady-state	L_2 residual	L_1 norm of the error	L_2 norm of the error	L_∞ norm of the error	Max error location
$b=-4/300$						
SAT	52.56	9.984×10^{-14}	1.707×10^{-04}	1.156×10^{-03}	0.01220	50, 2
Standard	401.11	9.995×10^{-14}	3.448×10^{-03}	2.479×10^{-02}	−0.2864	50, 2

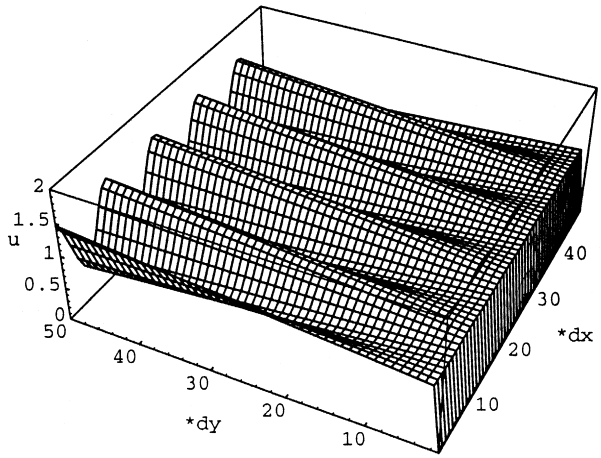

Fig. 20. Exact solution.

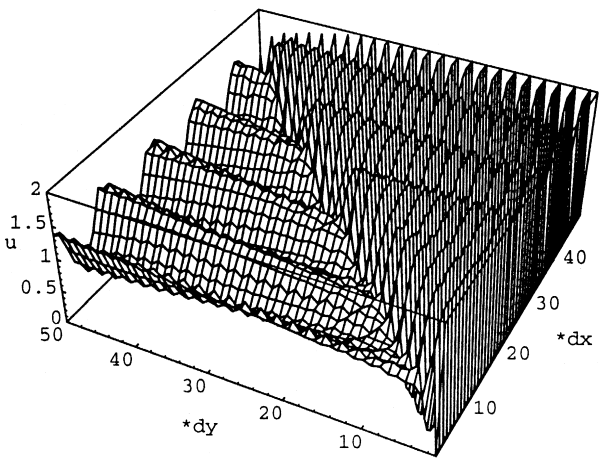

Fig. 21. Standard scheme, $b=-1$.

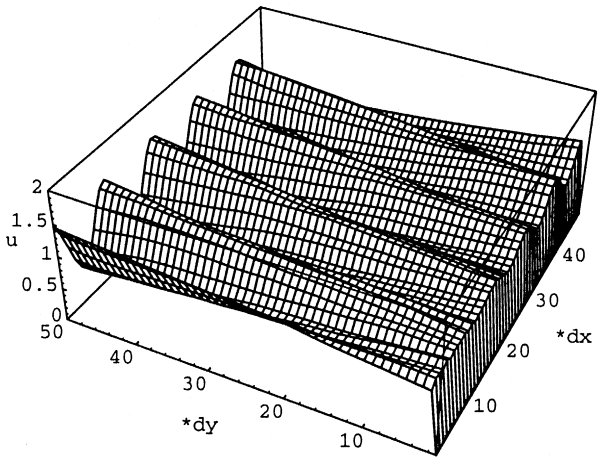

Fig. 22. SAT, $b = -1$.

for $t = 10$. Fig. 20 shows a three-dimensional plot of $u(x, y,10)$. As in the steady-state case, the plot looks the same to the eye for various $-1 < b < -4/\sqrt{R} = -4/300$. Figs. 21 and 22 show the three-dimensional plots of $v_{j,k}$ for the standard and bounded-error schemes respectively. Fig. 23 shows an x-profile of v at $y = 0.2$, for both schemes and the exact profile, for $b = -1$. Fig. 24 gives the same profiles at $y = 0.8$. These plots bring out the differences in the phase errors of the numerical algorithms. Figs. 25–28 repeat the same information as given in Figs. 21–24, but for $b = -4\sqrt{R} = -4/300$. The efficacy of the bounded-eror algorithm is quite evident—even when $b = -4/\sqrt{R}$, where the norm-errors away from the bondary layer are not dissimilar, and the phase error of the right running waves is quite a bit smaller in the case of the proposed present scheme.

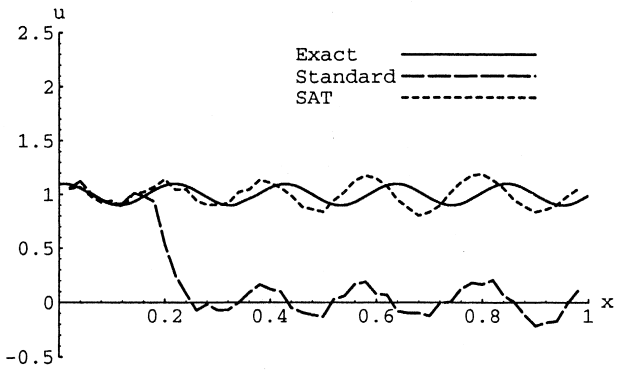

Fig. 23. $b = -1$, $y = 0.2$ profiles.

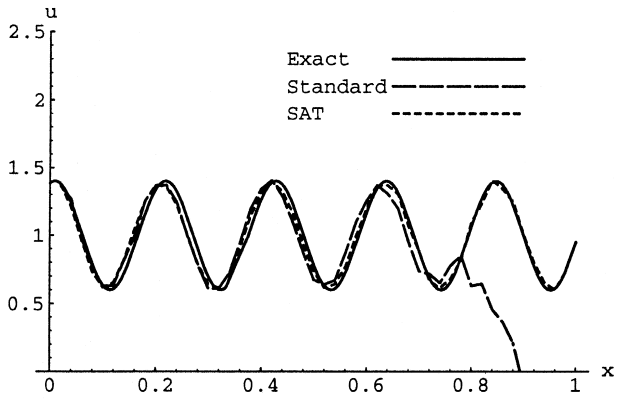

Fig. 24. $b = -1$, $y = 0.8$ profiles.

5. Conclusions

(i) A second-order method has been developed which renders spatial second derivative finite difference operators negative definite. This is not surprising, since negative definiteness was achieved for fourth-order parabolic operators in [2].
(ii) A second order method has been developed which renders spatial first derivative finite difference operators non-positive definite. For the case when boundary points do not coincide with grid nodes ($\gamma \neq 1$), this is a new result.
(iii) The results (i) and (ii) allow us to construct a solution operator for the advection diffusion problem (and, of course, the diffusion equation) which is negative definite, thereby ensuring asymptotic temporal stability.

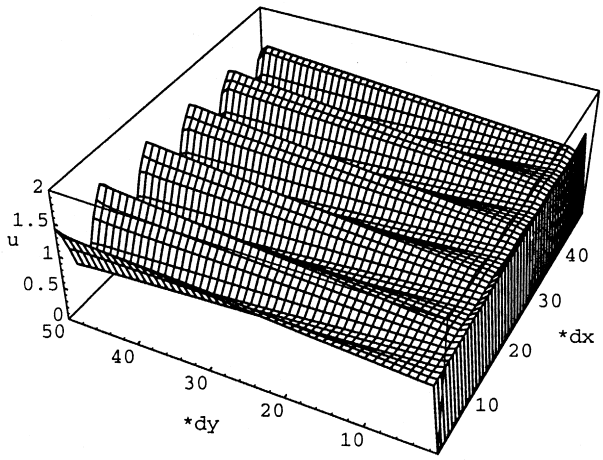

Fig. 25. Standard scheme, $b = -4/300$.

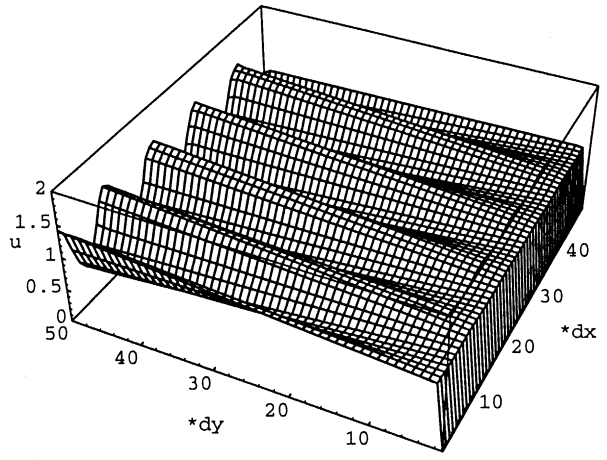

Fig. 26. SAT, $b = -4/300$.

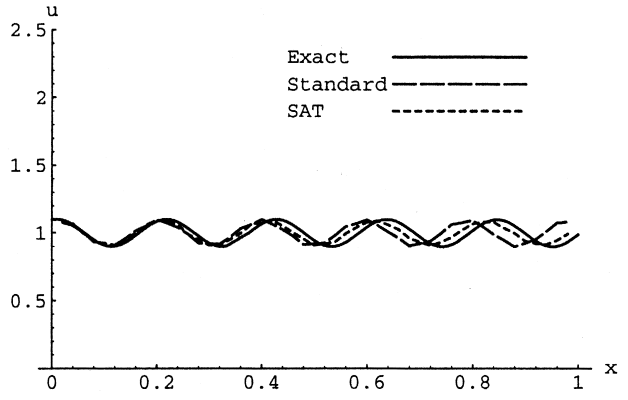

Fig. 27. $b = -4/300$, $y = 0.2$ profiles.

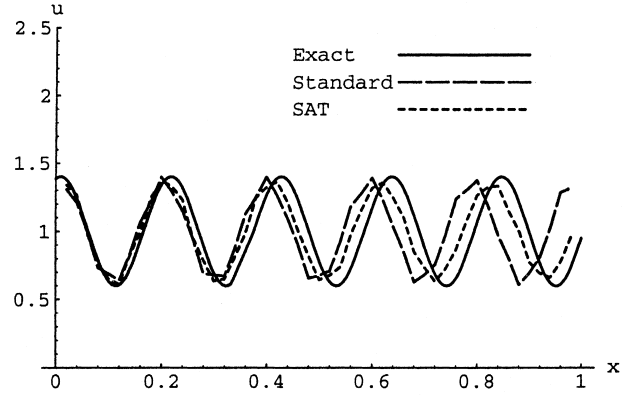

Fig. 28. $b = -4/300$, $y = 0.8$ profiles.

(iv) The construction of these operators allows an immediate simple generalization to multi-dimensional problems, on complex domains which are covered by rectangular meshes. The proofs of the boundedness of the error-norms carry over rigorously to the (linear) multi-dimensional cases.
(v) Numerous numerical examples demonstrate the efficacy of this methodology.

Appendix A

As in the $a > 0$ case the hyperbolic terms are given by:

$$D_{\mathrm{H}} = \frac{1}{2h}\left\{ \begin{bmatrix} -2 & 2 & & & & & & \\ -1 & 0 & 1 & & & & & \\ & -1 & 0 & 1 & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & -1 & 0 & 1 & 0 & \\ & & & & -1 & 0 & 1 & \\ & & & & & -2 & 2 & \end{bmatrix} + \begin{bmatrix} c_1 & & & & & & & \\ & c_2 & & & & & & \\ & & c_3 & & & & & \\ & & & \ddots & & & & \\ & & & & c_{N-2} & & & \\ & & & & & c_{N-1} & & \\ & & & & & & c_N & \end{bmatrix} \right.$$

$$\left. \times \begin{bmatrix} -1 & 2 & -1 & & & & & \\ 0 & -1 & 2 & -1 & & & & \\ 1 & -2 & 0 & 2 & -1 & & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & 1 & -2 & 0 & 2 & -1 & \\ & & & 1 & -2 & 1 & 0 & \\ & & & & 1 & -2 & 1 & \end{bmatrix} + 2h\tilde{c} \begin{bmatrix} 0 & -1 & 1 & & & & & & \\ 1 & -1 & -1 & 1 & & & & & \\ 1 & -1 & 0 & -1 & 1 & & & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & & & \\ & & 1 & -1 & 0 & -1 & 1 & & \\ & & & & 1 & -1 & -1 & 1 & \\ & & & & & 1 & -1 & 0 & \end{bmatrix} \right\} \tag{A1}$$

where

$$c_k = \frac{1}{N-1}[(c_N - c_1)k + (Nc_1 - c_N)] \tag{A2}$$

and

$$\tilde{c} = \frac{1}{2}(c_1 - c_N) \tag{A3}$$

For $a < 0$ in Eq. (1)a), the right boundary is, for the hyperbolic part, an "outflow" boundary on which we do not prescribe a "hyperbolic boundary condition", therefore, in this case $\tau_{R_H} = 0$, and

$$\tau_{L_H} = \frac{1}{2h}\,\mathrm{diag}[\tau_{L_1}^{(H)}, \tau_{L_2}^{(H)}, 0, \ldots, 0, 0] \tag{A4}$$

$$A_{L_H} = \begin{bmatrix} 1+\gamma_L & -\gamma_L & & & \\ 1+\gamma_L & -\gamma_L & & 0 & \\ & & 0 & & \\ & & & \ddots & \\ & 0 & & & 0 \end{bmatrix} \tag{A5}$$

Next we would like to show that $\tilde{M}_H = 1/2(M_H + M_H^T)$ is non-negative definite; then $a\tilde{M}_H$ is *non-positive definite*. Using Eqs. (A1)–(A5) we have

$$\tilde{M}_H = \frac{1}{4h}\begin{bmatrix} -4-2c_1-2(1+\gamma_L)\tau_1^{(H)} & 1+2c_1 & -(1+\gamma_L)\tau_2^{(H)}+\gamma_L\tau_1^{(H)} & 0 & & & \\ 1+2c_1-(1+\gamma_L)\tau_2^{(H)}+\gamma_L\tau_1^{(H)} & & -2c_1+2\gamma_L\tau_2^{(H)} & 0 & & & 0 \\ 0 & & 0 & 0 & & & \\ & & & & \ddots & & \\ 0 & & & & & 2c_N & -1-2c_N \\ & & & & & -1-2c_N & 4+2c_N \end{bmatrix} \tag{A6}$$

We now write $\tilde{M}_H$ as the sum of three "corner-matrices",

$$\tilde{M}_H = \frac{1}{4h}[m_{H_1} + m_{H_2} + m_{H_3}] \tag{A7}$$

where

$$m_{\mathrm{H}_1} = c_1 \begin{bmatrix} -2 & 2 & & & \\ 2 & -2 & & 0 & \\ & & 0 & & \\ & 0 & & \ddots & \\ & & & & 0 \end{bmatrix}$$

$$m_{\mathrm{H}_2} = \begin{bmatrix} -4-2(1+\gamma_{\mathrm{L}})\tau_1^{(\mathrm{H})} & 1-(1+\gamma_{\mathrm{L}})\tau_2^{(\mathrm{H})}+\gamma_{\mathrm{L}}\tau_1^{(\mathrm{H})} & 0 & & \\ 1-(1+\gamma_{\mathrm{L}})\tau_2^{(\mathrm{H})}+\gamma_{\mathrm{L}}\tau_1^{(\mathrm{H})} & +2\gamma\tau_2^{(\mathrm{H})} & 0 & 0 & \\ 0 & 0 & 0 & & \\ & & & 0 & \\ & 0 & & & \ddots \\ & & & & 0 \end{bmatrix} \tag{A8}$$

$$m_{\mathrm{H}_3} = c_N \begin{bmatrix} 0 & & & & \\ & 0 & & & \\ & & \ddots & & \\ & & & 2c_N & -1-2c_N \\ & & & -1-2c_N & 4+2c_N \end{bmatrix}$$

Clearly m_{H_1} is N.N.D. (non-negative definite) for $\forall c_1 \le 0$. Also, m_{H_3} is N.N.D. for $c_N \ge -1/4$. A simple computation shows that m_{H_2} is N.N.D. if τ_1 and τ_2 satisfy

$$\tau_1^{(\mathrm{H})} = -\frac{2+\delta}{1+\gamma_{\mathrm{L}}} \quad (\delta \ge 0) \tag{A9}$$

$$\tau_2^{(\mathrm{H})} = \frac{1-\gamma_{\mathrm{L}}(1-\delta)}{(1+\gamma_{\mathrm{L}})^2} \tag{A10}$$

Thus we have proved that $\tilde{M}_{\mathrm{H}}$ is indeed non-negative definite, and therefore $\tilde{M} = 1/R\tilde{M}_{\mathrm{P}} + a\tilde{M}_{\mathrm{H}}$ is negative definite for $\forall 1/R > 0$, with its eigenvalues bounded away from zero by $-\alpha\pi^2/R$, $(0 < \alpha < 0.275\}$, as in the $a > 0$ case treated in the text.

References

[1] Carpenter MH, Gottlieb D, Abarbanel S. time stable boundary conditions for finite difference schemes solving hyperbolic systems: methodology and application to high order compact schemes. NASA Contractor Report 191436, ICASE Report no. 93-9 (in press).

[2] Abarbanel S, Ditkowski A. Multi-dimensional asymptotically stable 4th-order accurate schemes for the diffusion equation. ICASE Report no. 96-8, February 1996, also J. Comp. Physics, V.133, pp. 279–288 (1997).

[3] Gustafsson B, Kreiss HO, Sundström A. Stability theory of difference approximations for mixed initial boundary value problems, II. Math Comput 1972;26:649–86.
[4] Abarbanel S, Bennet S, Brandt A, Gillis J. Velocity profiles of flow at low reynolds numbers. J Appl Mech 1970;37E(1):1–3.