Available online at www.sciencedirect.com
ScienceDirect
Applied Numerical Mathematics 57 (2007) 19–35
APPLIED NUMERICAL MATHEMATICS
www.elsevier.com/locate/apnum

Stability of ADI schemes applied to convection–diffusion equations with mixed derivative terms

K.J. in 't Hout, B.D. Welfert *

Department of Mathematics and Statistics, Arizona State University, Tempe, Arizona 85287, USA

Available online 10 January 2006

Abstract

We consider Alternating Direction Implicit (ADI) schemes for the numerical solution of initial-boundary value problems for convection–diffusion equations with cross derivative terms. We derive new linear stability results for three ADI schemes that have previously been studied in the literature. These results are subsequently used to show that the ADI schemes under consideration are unconditionally stable when applied to finite difference discretizations of general parabolic two-dimensional convection–diffusion equations. Supporting numerical evidence is included.

Keywords: Initial-boundary value problems; Convection–diffusion equations; Numerical solution; ADI splitting schemes; Von Neumann stability analysis; Fourier transformation

1. Introduction

1.1. Three alternating direction implicit schemes

We consider the large system of ordinary differential equations (ODEs)

$$U'(t) = F\big(t, U(t)\big) \quad (t \geqslant 0), \tag{1.1}$$

with given function F and initial value $U(0) = U_0$, arising from the semi-discretization of an initial-boundary value problem for a multi-dimensional convection–diffusion equation. In splitting methods for the numerical solution of (1.1), the function F is decomposed into a sum

$$F(t, v) = F_0(t, v) + F_1(t, v) + \cdots + F_k(t, v), \tag{1.2}$$

where the terms F_j are simpler to handle than F itself. We assume in this paper that F_0 is treated explicitly in time-integration schemes, whereas $F_1, F_2, \ldots, F_k$ represent stiff, unidirectional contributions in F that are treated implicitly.

The analysis to be presented in this paper is mostly relevant to ODE systems (1.1) originating from convection–diffusion problems in two spatial variables. For the formulation of the splitting schemes below, however, the particular spatial dimension of the problem is not yet important.

* Corresponding author.
E-mail addresses: khout@math.asu.edu (K.J. in 't Hout), welfert@asu.edu (B.D. Welfert).

0168-9274/$30.00
doi:10.1016/j.apnum.2005.11.011

Let time step $\Delta t > 0$. The *Douglas scheme* defines an approximation $U_n \approx U(t_n)$, with $t_n = n\Delta t$, successively for $n = 1, 2, 3, \ldots$ by

$$\begin{cases} Y_0 = U_{n-1} + \Delta t\, F(t_{n-1}, U_{n-1}), \\ Y_j = Y_{j-1} + \theta \Delta t\,(F_j(t_n, Y_j) - F_j(t_{n-1}, U_{n-1})), \quad j = 1, 2, \ldots, k, \\ U_n = Y_k. \end{cases} \tag{1.3}$$

Here $\theta > 0$ denotes a real parameter, which specifies the scheme. In (1.3), the forward Euler predictor step is followed by k implicit but unidirectional corrector steps, whose purpose is to stabilize the predictor step.

In the general form above, the scheme (1.3) has been considered for example in [8,10,11]. Special instances of (1.3) include the well-known Alternating Direction Implicit (ADI) methods of Douglas and Rachford [4], with $F_0 = 0$ and $\theta = 1$, and of Brian [1] and Douglas [3], with $F_0 = 0$ and $\theta = \frac{1}{2}$. These special ADI methods were initially developed for application in the case of the two- and three-dimensional heat equations. For a survey of their development, see e.g. Peaceman [17].

It can be verified that the classical order[1] of the scheme (1.3) is equal to 2 whenever $F_0 = 0$ and $\theta = \frac{1}{2}$, and it is of order 1 otherwise. In our application we will always have $F_0 \neq 0$ and, consequently, the order of the Douglas scheme (1.3) reduces to just 1, for any given θ. In the following we formulate two ADI schemes that attain order 2 also if $F_0 \neq 0$.

The subsequent scheme can be regarded as an extension of (1.3):

$$\begin{cases} Y_0 = U_{n-1} + \Delta t\, F(t_{n-1}, U_{n-1}), \\ Y_j = Y_{j-1} + \theta \Delta t\,(F_j(t_n, Y_j) - F_j(t_{n-1}, U_{n-1})), \quad j = 1, 2, \ldots, k, \\ \widetilde{Y}_0 = Y_0 + \sigma \Delta t\,(F_0(t_n, Y_k) - F_0(t_{n-1}, U_{n-1})), \\ \widetilde{Y}_j = \widetilde{Y}_{j-1} + \theta \Delta t\,(F_j(t_n, \widetilde{Y}_j) - F_j(t_{n-1}, U_{n-1})), \quad j = 1, 2, \ldots, k, \\ U_n = \widetilde{Y}_k. \end{cases} \tag{1.4}$$

Here $\sigma > 0$ denotes a second real parameter. If $F_0 = 0$, then (1.4) reduces to the Douglas scheme (1.3). Clearly, (1.4) uses the (stable) approximation Y_k to $U(t_n)$ obtained from the Douglas method so as to introduce a correction with respect to the F_0 part as well. The scheme (1.4) is of classical order 2 whenever $\{F_0 = 0$ and $\theta = \frac{1}{2}\}$ or $\{\theta = \sigma = \frac{1}{2}\}$, and it has order 1 otherwise. Hence, contrary to (1.3), with the scheme (1.4) order 2 can be attained independently of F_0—upon taking $\theta = \sigma = \frac{1}{2}$.

The so-called "iterated scheme" that was proposed by Craig and Sneyd [2] can be reformulated as (1.4). Correspondingly, we shall refer in this paper to (1.4) as the *Craig & Sneyd scheme*. These authors studied the scheme (1.4) in the application to pure diffusion problems where mixed spatial derivative terms are present. We note that a version of the Douglas scheme (1.3) was considered in [2] as well, where it was called the "simple scheme". It is interesting to remark that a variant of (1.4) is used in financial option pricing, cf. [18]. When applied to linear autonomous problems (1.1), this variant is readily seen to reduce to (1.4) with $k = 2$ and $\theta = \sigma = \frac{1}{2}$.

The following scheme can be regarded as a second, different extension of the Douglas scheme:

$$\begin{cases} Y_0 = U_{n-1} + \Delta t\, F(t_{n-1}, U_{n-1}), \\ Y_j = Y_{j-1} + \theta \Delta t\,(F_j(t_n, Y_j) - F_j(t_{n-1}, U_{n-1})), \quad j = 1, 2, \ldots, k, \\ \widetilde{Y}_0 = Y_0 + \sigma \Delta t\,(F(t_n, Y_k) - F(t_{n-1}, U_{n-1})), \\ \widetilde{Y}_j = \widetilde{Y}_{j-1} + \theta \Delta t\,(F_j(t_n, \widetilde{Y}_j) - F_j(t_n, Y_k)), \quad j = 1, 2, \ldots, k, \\ U_n = \widetilde{Y}_k. \end{cases} \tag{1.5}$$

If $\sigma = \frac{1}{2}$ the scheme (1.5) is identical to a scheme formulated by Hundsdorfer [10, p. 222]. Furthermore, in the case of linear autonomous problems (1.1) it is equivalent to a Rosenbrock type method discussed e.g. in Hundsdorfer and Verwer [11, p. 400]. In view of this, we shall refer in this paper to (1.5) as the *Hundsdorfer & Verwer scheme*. This scheme has been considered in [10,11] for the application to convection–diffusion–reaction problems without mixed derivative terms.

[1] I.e., the order for fixed non-stiff ODEs.

For any given θ, the scheme (1.5) is of classical order 2 if $\sigma = \frac{1}{2}$ and of order 1 otherwise, independently of F_0. Compared to (1.4), note e.g. that (1.5) uses the full right-hand side function F from (1.1) for the update $\widetilde{Y}_0$, instead of just F_0 alone.

We mention that all three schemes (1.3)–(1.5) above are closely related to so-called Approximate Matrix Factorization methods, cf. e.g. [11].

When applied to the linear scalar test equation

$$U'(t) = (\lambda_0 + \lambda_1 + \cdots + \lambda_k)U(t), \tag{1.6}$$

with complex constants λ_j $(0 \leqslant j \leqslant k)$, the Douglas scheme (1.3) reduces to

$$U_n = R(z_0, z_1, \ldots, z_k)\, U_{n-1} \tag{1.7}$$

with $z_j = \lambda_j \Delta t$ $(0 \leqslant j \leqslant k)$ and

$$R(z_0, z_1, \ldots, z_k) = 1 + \frac{z_0 + z}{p}. \tag{1.8}$$

Here, and throughout this paper, we adopt the notation

$$z = z_1 + z_2 + \cdots + z_k \quad \text{and} \quad p = (1 - \theta z_1)(1 - \theta z_2) \cdots (1 - \theta z_k). \tag{1.9}$$

The Craig & Sneyd scheme (1.4) reduces in the case of (1.6) to

$$U_n = S(z_0, z_1, \ldots, z_k)\, U_{n-1} \tag{1.10}$$

with

$$S(z_0, z_1, \ldots, z_k) = 1 + \frac{z_0 + z}{p} + \sigma \frac{z_0(z_0 + z)}{p^2}. \tag{1.11}$$

Finally, the Hundsdorfer & Verwer scheme (1.5) reduces, in the case of (1.6), to

$$U_n = T(z_0, z_1, \ldots, z_k)U_{n-1} \tag{1.12}$$

with

$$T(z_0, z_1, \ldots, z_k) = 1 + 2\frac{z_0 + z}{p} - \frac{z_0 + z}{p^2} + \sigma \frac{(z_0 + z)^2}{p^2}. \tag{1.13}$$

The iterations (1.7), (1.10), (1.12) are stable if

$$\left|R(z_0, z_1, \ldots, z_k)\right| \leqslant 1, \tag{1.14}$$

$$\left|S(z_0, z_1, \ldots, z_k)\right| \leqslant 1, \tag{1.15}$$

and

$$\left|T(z_0, z_1, \ldots, z_k)\right| \leqslant 1, \tag{1.16}$$

respectively.

1.2. Stability of the three ADI schemes in the application to multi-dimensional convection–diffusion equations

We are interested in the stability of the three schemes (1.3)–(1.5) when applied to the semi-discretized k-dimensional convection–diffusion equation

$$\frac{\partial u}{\partial t} = \mathbf{c} \cdot \nabla u + \nabla \cdot (D \nabla u) \tag{1.17}$$

on a rectangular domain, supplemented with initial and boundary conditions. Here $\mathbf{c}$ denotes a given convection vector and D is a given diffusion matrix, which is always assumed to be positive semi-definite. If D is positive definite, then (1.17) is parabolic. Our interest is in the general equation (1.17), where D is a full (non-diagonal) matrix and $\mathbf{c}$ is non-zero. Thus, both mixed derivative and convection terms are present in (1.17). Multi-dimensional convection–diffusion

equations of this kind arise in many applied areas, for example mathematical biology (cf. e.g. [5]) and financial mathematics (cf. e.g. [19]). We also note that full matrices D arise when converting, by coordinate transformation, from convection–diffusion equations with diagonal matrices D on non-rectangular domains to equations on rectangular domains (cf. e.g. [15]).

The stability results derived in the present paper for the three schemes (1.3)–(1.5) are primarily relevant to the two-dimensional case of (1.17), i.e., $k = 2$. These results for $k = 2$ however do represent a substantial improvement over known results existing in literature. Moreover, our analysis covers the above three ADI schemes together with a large variety of spatial discretizations under one single umbrella. We mention that results for arbitrary spatial dimensions $k > 2$ will be given in a forthcoming paper [7], where pure diffusion problems with mixed derivative terms are considered.

We shall analyze the stability of (1.3)–(1.5) in the application to semi-discrete versions of (1.17) using the well-known von Neumann method (Fourier transformation). Here, we adopt the usual assumption that $\mathbf{c}$ and D are constant and that the boundary condition for (1.17) is periodic. This leads to the conditions (1.14), (1.15) and (1.16), respectively, where each $z_j = \lambda_j \Delta t$ with λ_j an eigenvalue of the linear operator F_j that is obtained after semi-discretization.

We are interested in this paper in *unconditional* stability, i.e., stability without any restriction on the time step $\Delta t > 0$. In line with the von Neumann method, stability is always understood here to be with respect to the l_2-norm.

At present only few, partial results appear to be known in the literature concerning the application and the stability of the schemes (1.3)–(1.5) relevant to general convection–diffusion problems with mixed derivative terms. In the remainder of this section, we will review results that are currently available.

Craig and Sneyd [2] performed a von Neumann stability analysis of the two schemes (1.3), (1.4) relevant to Eqs. (1.17) with $\mathbf{c} = \mathbf{0}$ and D a full matrix, i.e., no convection, only diffusion. The semi-discretization was done using standard, centered finite differencing and a splitting (1.2) was considered where F_0 contains all the discretized cross derivative terms and F_j, for $1 \leqslant j \leqslant k$, represents the discretized diffusion operator in the jth spatial direction. With these choices, it was proved in [2] that the Douglas scheme (1.3) is unconditionally stable whenever the parameter θ is sufficiently large, with a lower bound on θ that only depends on k. In particular, for $k = 2$, unconditional stability of (1.3) was obtained whenever $\theta \geqslant \frac{1}{2}$ (see also McKee and Mitchell [14] if $\theta = \frac{1}{2}$). For the scheme (1.4), Craig and Sneyd [2] arrived at a similar result, under an additional condition on the parameter σ. In particular, for $k = 2$, they showed that (1.4) is unconditionally stable whenever $\theta \geqslant \sigma \geqslant \frac{1}{2}$.

The above, favorable, stability results are surprising in view of the fact that the F_0 part is integrated explicitly, whereas its scaled eigenvalues z_0, which lie on the real axis, range from large negative to large positive values (cf. also Section 3).

McKee et al. [15] examined the unconditional stability of an equivalent version of the Douglas scheme (1.3) when applied to a standard finite difference discretization of (1.17) in the two-dimensional case ($k = 2$) with convection ($\mathbf{c} \neq \mathbf{0}$) and a cross derivative term (D non-diagonal). The F_0 part in their case represents again the discretized cross derivative term. Next, F_j, for $j = 1, 2$, contains the discretized first- and second-order derivatives in the jth spatial direction. The positive result was proved in [15] that, in the presence of convection, unconditional stability of the Douglas scheme (1.3) is maintained whenever $k = 2$ and $\theta = \frac{1}{2}$.

Hundsdorfer [9] considered general, necessary and sufficient, conditions on the z_j ($0 \leqslant j \leqslant k$) with regard to the stability requirement (1.14). It was proved in [9] that if $k \geqslant 2$, $\theta \geqslant \frac{1}{2}$ and $z_0 = 0$, then (1.14) holds for all z_j ($1 \leqslant j \leqslant k$) lying in a wedge $\{\zeta \in \mathbb{C}\colon\ |\arg(-\zeta)| \leqslant \alpha\}$ if and only if the angle $\alpha \leqslant \frac{1}{k-1}\frac{\pi}{2}$. In particular, if $k = 2$ and $\theta \geqslant \frac{1}{2}$, the condition (1.14) is thus fulfilled whenever $\Re z_1, \Re z_2 \leqslant 0$ and $z_0 = 0$. Note that if $z_0 = 0$, all conclusions about (1.14) are directly valid for the requirement (1.15) as well. Additional results on (1.14) under the assumption that z_0 belongs to the stability region of the forward Euler method, i.e. $|1 + z_0| \leqslant 1$, were also derived in [9].

The above stability bound on the angle α was proved by Hundsdorfer [10] for a very general type of rational functions, including also (1.13) with $z_0 = 0$. Correspondingly, it is not clear at present whether useful ADI splitting schemes exist that are unconditionally stable when applied to general (semi-discrete) convection–diffusion problems in dimensions $k \geqslant 3$. We remark that in the case of pure diffusion problems, without convection, there is no such limitation on the dimension, cf. [2,7]. In the latter case all eigenvalues $z_0, z_1, \ldots, z_k$ are real.

The requirement (1.16) with $\sigma = \frac{1}{2}$ was investigated numerically in Hundsdorfer [10]. The numerical experiments in [10] suggested that if $k = 2$, $z_0 = 0$ and θ is larger than a certain threshold value, then (1.16) is fulfilled whenever $\Re z_1, \Re z_2 \leqslant 0$. This conjecture was proved by Lanser et al. [13], with threshold value $\theta = \frac{1}{2} + \frac{1}{6}\sqrt{3}$. Additional numerical results regarding (1.16) were presented in [10] under the assumption that z_0 belongs to the stability region of

the explicit improved Euler method, i.e., $|1+z_0+\frac{1}{2}z_0^2|\leqslant 1$. In our application, however, z_0 represents the eigenvalues of the discretized cross derivative terms, which may lie anywhere along the real axis, and we shall need a different kind of condition on z_0 than $|1+z_0+\frac{1}{2}z_0^2|\leqslant 1$, or $|1+z_0|\leqslant 1$, to assess the stability of the three ADI schemes.

1.3. Outline of this paper

In Section 2 we consider conditions on general complex numbers $z_0, z_1, \ldots, z_k$ that imply stability of the Douglas scheme (1.3), the Craig & Sneyd scheme (1.4) and the Hundsdorfer & Verwer scheme (1.5) in the sense that (1.14), (1.15) and (1.16) hold, respectively. Each of these conditions consists of an upper bound on the modulus of z_0 in terms of $z_1, \ldots, z_k$. A particularly useful sufficient condition is introduced for the case $k=2$, see (2.4) and Theorem 2.8.

In Section 3 it is shown that the condition (2.4) is fulfilled for the eigenvalues z_0, z_1, z_2 which correspond to standard finite difference discretizations of (1.17) in two spatial dimensions, with general (positive semi-definite) diffusion matrix D and general convection vector $\mathbf{c}$. By applying Theorem 2.8, this directly leads to the main results of our paper, Theorems 3.1 and 3.2. These two theorems give positive conclusions concerning the unconditional stability of the ADI schemes (1.3), (1.4), (1.5) in the application to general two-dimensional convection–diffusion equations (1.17) with a wide range of spatial discretizations.

Numerical experiments are conducted in Section 4. The experiments agree with our two stability theorems from Section 3. In addition, they provide insight into the actual convergence behavior of the three ADI schemes.

Final comments about extensions and future research directions are given in Section 5.

2. General sufficient conditions for (1.14)–(1.16)

2.1. General sufficient conditions for (1.14) and (1.15)

For given θ, consider the following condition on $z_0, z_1, \ldots, z_k \in \mathbb{C}$:

$$p\neq 0 \quad \text{and} \quad |z_0|\leqslant \left|\frac{p}{2\theta}\right| - \left|\frac{p}{2\theta}+z\right|. \tag{2.1}$$

Recall that z, p are given by (1.9). Our first lemma concerns the requirement (1.14). Note that if $z_0=0$ and $\theta=\frac{1}{2}$, then (2.1) is equivalent to (1.14).

Lemma 2.1. *Assume* (2.1) *holds and* $\theta\geqslant\frac{1}{2}$. *Then*

$$\left|R(z_0,z_1,\ldots,z_k)\right|\leqslant 1.$$

Proof. Define

$$\widetilde{R}=\widetilde{R}(z_0,z_1,\ldots,z_k)=\frac{1}{2\theta}+\frac{z+z_0}{p}. \tag{2.2}$$

We have

$$|\widetilde{R}|=\left|\frac{1}{2\theta}+\frac{z}{p}+\frac{z_0}{p}\right|\leqslant\left|\frac{1}{2\theta}+\frac{z}{p}\right|+\left|\frac{z_0}{p}\right|\leqslant\frac{1}{2\theta}. \tag{2.3}$$

Subsequently,

$$\left|R(z_0,z_1,\ldots,z_k)\right|=\left|1-\frac{1}{2\theta}+\widetilde{R}\right|\leqslant 1-\frac{1}{2\theta}+|\widetilde{R}|\leqslant 1-\frac{1}{2\theta}+\frac{1}{2\theta}=1. \qquad \square$$

The next lemma deals with the requirement (1.15).

Lemma 2.2. *Assume* (2.1) *holds and* $\theta\geqslant\frac{1}{2}$, $\sigma\leqslant 2\theta$. *Then*

$$\left|S(z_0,z_1,\ldots,z_k)\right|\leqslant 1.$$

Proof. Consider $\widetilde{R}$ given by (2.2) and let

$$\widetilde{S} = \widetilde{S}(z_0, z_1, \ldots, z_k) = \frac{1}{2\theta} + \frac{z + z_0}{p} + \sigma \frac{z_0(z_0 + z)}{p^2}.$$

Using (2.3), we obtain

$$\begin{aligned}
|\widetilde{S}| &= \left| \frac{1}{2\theta} + \frac{z}{p} + \frac{z_0}{p} + \sigma \frac{z_0}{p} \left(\widetilde{R} - \frac{1}{2\theta} \right) \right| \\
&\leqslant \left| \frac{1}{2\theta} + \frac{z}{p} \right| + \left(1 - \frac{\sigma}{2\theta} \right) \left| \frac{z_0}{p} \right| + \sigma \left| \frac{z_0}{p} \right| |\widetilde{R}| \\
&\leqslant \left| \frac{1}{2\theta} + \frac{z}{p} \right| + \left| \frac{z_0}{p} \right| \\
&\leqslant \frac{1}{2\theta}.
\end{aligned}$$

Then we have

$$\left| S(z_0, z_1, \ldots, z_k) \right| = \left| 1 - \frac{1}{2\theta} + \widetilde{S} \right| \leqslant 1 - \frac{1}{2\theta} + |\widetilde{S}| \leqslant 1 - \frac{1}{2\theta} + \frac{1}{2\theta} = 1. \qquad \square$$

The direct verification of condition (2.1) is not straightforward in general, in view of the non-trivial formulas one has for the eigenvalues z_j (cf. (3.4)). We introduce next a condition which is much easier to verify when $k = 2$:

$$\Re z_1 \leqslant 0, \qquad \Re z_2 \leqslant 0 \quad \text{and} \quad |z_0| \leqslant 2\sqrt{\Re z_1 \Re z_2}. \tag{2.4}$$

Observe that, contrary to (2.1), condition (2.4) is independent of the parameter θ. Also, note that (2.4) reduces to $\Re z_1, \Re z_2 \leqslant 0$ whenever $z_0 = 0$. We have:

Lemma 2.3. *Assume $k = 2$. Then* (2.4) $\Rightarrow$ (2.1).

Proof. Define the vectors

$$\mathbf{v}_j = \begin{bmatrix} \sqrt{-2\Re z_j} \\ |\frac{1+\theta z_j}{\sqrt{2\theta}}| \end{bmatrix}, \quad j = 1, 2.$$

Then

$$\|\mathbf{v}_j\|_2 = \sqrt{-2\Re z_j + \frac{|1 + \theta z_j|^2}{2\theta}} = \frac{|1 - \theta z_j|}{\sqrt{2\theta}}.$$

Condition (2.4) implies $p \neq 0$. Next, we obtain

$$\begin{aligned}
|z_0| + \left| \frac{p}{2\theta} + z \right| &= |z_0| + \left| \frac{(1 - \theta z_1)(1 - \theta z_2)}{2\theta} + z_1 + z_2 \right| \\
&= |z_0| + \left| \frac{(1 + \theta z_1)(1 + \theta z_2)}{2\theta} \right| \\
&\leqslant 2\sqrt{\Re z_1 \Re z_2} + \left| \frac{(1 + \theta z_1)(1 + \theta z_2)}{2\theta} \right| \\
&= \mathbf{v}_1 \cdot \mathbf{v}_2 \leqslant \|\mathbf{v}_1\|_2 \|\mathbf{v}_2\|_2 \\
&= \frac{|1 - \theta z_1||1 - \theta z_2|}{2\theta} \\
&= \left| \frac{p}{2\theta} \right|,
\end{aligned}$$

which shows that (2.1) holds. $\square$

We note that the result of Lemma 2.3 is sharp in the sense that if $z_1 = z_2 = \zeta$ with ζ any given negative real number, then conditions (2.4) and (2.1) are (both) equivalent to $|z_0| \leqslant -2\zeta$. Further, it is easily seen in this case that if z_0 is any real number with $z_0 > -2\zeta$, then $R(z_0, z_1, z_2) > 1$ and $S(z_0, z_1, z_2) > 1$, for any given θ, σ. This shows that Lemmas 2.1 and 2.2 are sharp, in a certain sense, as well.

In Section 3 we shall give a simple proof that condition (2.4) holds in the case of 2D convection–diffusion equations (1.17) and a wide range of spatial discretizations.

2.2. General sufficient conditions for (1.16)

We first consider whether the stability requirement (1.16) is fulfilled under the condition (2.1). The following lemma gives a partial result, namely for the case of real, non-positive z_j with $j \geqslant 1$.

Lemma 2.4. *Assume $z_1, z_2, \ldots, z_k$ are non-positive real numbers, condition* (2.1) *holds and $4\theta(1-\theta) \leqslant \sigma \leqslant \theta$. Then*

$$\left|T(z_0, z_1, \ldots, z_k)\right| \leqslant 1.$$

Proof. The assumptions imply

$$p \geqslant 1, \quad 1 - \frac{1}{\theta} + \frac{\sigma}{4\theta^2} + \frac{1}{2\theta p} \geqslant 0 \quad \text{and} \quad 2 - \frac{\sigma}{\theta} - \frac{1}{p} \geqslant 0.$$

Consider $\widetilde{R}$ given by (2.2) and write $T = T(z_0, z_1, \ldots, z_k)$. Using (2.3), we obtain

$$\begin{aligned}
|T| &= \left|1 + 2\left(\widetilde{R} - \frac{1}{2\theta}\right) - \frac{1}{p}\left(\widetilde{R} - \frac{1}{2\theta}\right) + \sigma\left(\widetilde{R} - \frac{1}{2\theta}\right)^2\right| \\
&\leqslant \left|1 - \frac{1}{\theta} + \frac{\sigma}{4\theta^2} + \frac{1}{2\theta p}\right| + \left|2 - \frac{\sigma}{\theta} - \frac{1}{p}\right||\widetilde{R}| + \sigma|\widetilde{R}|^2 \\
&\leqslant \left(1 - \frac{1}{\theta} + \frac{\sigma}{4\theta^2} + \frac{1}{2\theta p}\right) + \left(2 - \frac{\sigma}{\theta} - \frac{1}{p}\right)\frac{1}{2\theta} + \frac{\sigma}{4\theta^2} \\
&= 1. \qquad \square
\end{aligned}$$

In the most interesting case of $\sigma = \frac{1}{2}$, the assumption on θ in Lemma 2.4 becomes $\theta \geqslant \frac{1}{2} + \frac{1}{4}\sqrt{2} \approx 0.8536$.

For general complex $z_1, z_2, \ldots, z_k$ however, the condition (2.1) does not guarantee (1.16). To see this, consider $k = 2$ and $\sigma = \frac{1}{2}$, and take

$$z_0 = \frac{6}{5\theta} + \frac{1}{\theta} I, \quad z_1 = -\frac{1}{\theta}, \quad z_2 = -\frac{6}{5\theta} I \tag{2.5}$$

with $I = \sqrt{-1}$. It is readily verified that the inequality in (2.1) becomes an equality, and a calculation yields

$$\left|T(z_0, z_1, z_2)\right|^2 = 1 + \frac{888\theta^3 + 1041\theta^2 + (9\theta - 1)^2}{(244\theta^2)^2} > 1.$$

A counterexample for $k > 2$ is obtained using (2.5) and $z_3 = \cdots = z_k = 0$.

We next examine whether (1.16) is fulfilled for $k = 2$ under the condition (2.4). Note that the above counterexample does not apply in this case. In view of Lemma 2.3, we can (directly) replace condition (2.1) by (2.4) in Lemma 2.4. This yields a partial result on the question of whether the condition (2.4) implies (1.16) for $k = 2$. In the following we shall improve upon this result by taking a different approach than above to derive (1.16) from (2.4). To this end we introduce the quantities

$$q = p^2 + 2pz - z + \sigma z^2, \qquad w = 2p - 1 + 2\sigma z$$

and consider the following condition on $z_0, z_1, \ldots, z_k \in \mathbb{C}$:

$$p \neq 0 \quad \text{and} \quad |q| + |w||z_0| + \sigma|z_0|^2 \leqslant |p|^2. \tag{2.6}$$

Lemma 2.5. *Assume* (2.6) *holds. Then*

$$\left|T(z_0, z_1, \ldots, z_k)\right| \leqslant 1.$$

Proof. Write $T = T(z_0, z_1, \ldots, z_k)$. Then

$$\begin{aligned}
|T| &= \left|\frac{p^2 + 2p(z+z_0) - (z+z_0) + \sigma(z+z_0)^2}{p^2}\right| \\
&= \left|\frac{q + wz_0 + \sigma z_0^2}{p^2}\right| \\
&\leqslant \frac{|q| + |w||z_0| + \sigma|z_0|^2}{|p|^2} \\
&\leqslant 1. \qquad \square
\end{aligned}$$

Subsequently, we have

Lemma 2.6. *Let $k = 2$. Assume z_1, z_2 are real numbers, condition* (2.4) *holds and*

$$2(1-\theta) \leqslant \sigma \leqslant \left(1 + \frac{1}{2}\sqrt{2}\right)\theta.$$

Then (2.6) *holds.*

Combined, Lemmas 2.5 and 2.6 yield an improved result on (1.16) under the condition (2.4) with z_1, z_2 real, in that the set of allowable parameters θ, σ is substantially enlarged. If $\sigma = \frac{1}{2}$ the assumption on θ given by Lemma 2.6 becomes $\theta \geqslant \frac{3}{4}$. Contrary to Lemma 2.4 with $\sigma = \frac{1}{2}$, this now includes the range $\theta \geqslant \frac{1}{2} + \frac{1}{6}\sqrt{3} \approx 0.7887$ for which it is known [13] that (1.16) is fulfilled whenever $k = 2$, $z_0 = 0$ and $\Re z_1, \Re z_2 \leqslant 0$ (cf. also Section 1.2).

Proof. In order to keep the presentation concise, we will assume that $\sigma = \frac{1}{2}$. The proof for general σ is more technical, but follows the same lines. Note that with $\sigma = \frac{1}{2}$ we have $\theta \geqslant \frac{3}{4}$.

Let $z_1, z_2 \leqslant 0$ and define $y = 2\sqrt{z_1 z_2}$. Then $z = z_1 + z_2 \leqslant 0$ and $y + z = -(\sqrt{-z_1} - \sqrt{-z_2})^2 \leqslant 0$. Next, $p = 1 - \theta z + \frac{1}{4}\theta^2 y^2 \geqslant 1$, $w = 2p - 1 + z = 1 + (1-2\theta)z + \frac{1}{2}\theta^2 y^2 \geqslant 1$. By condition (2.4), we further have $|z_0| \leqslant y$. Consequently, condition (2.6) is fulfilled if

$$|q| + wy + \frac{1}{2}y^2 \leqslant p^2. \tag{2.7}$$

If $q \geqslant 0$, then (2.7) is equivalent to $p \geqslant \frac{1}{2} - \frac{1}{4}(y+z)$, which is easily seen to hold. If $q < 0$, then (2.7) is equivalent to

$$2p^2 - 2(y-z)p + \left(y - z + \frac{1}{2}z^2 - yz - \frac{1}{2}y^2\right) \geqslant 0,$$

which is true if the discriminant $8(y^2 - y + z) \leqslant 0$, or if $y^2 - y + z > 0$ and $p \geqslant \frac{1}{2}(y - z + \sqrt{2y^2 - 2y + 2z})$. From $z \leqslant -y$ it is readily seen that $y^2 - y + z > 0$ can only occur if $y > 2$ and, next, that the latter inequality for p is fulfilled if

$$1 + \left(\theta - \frac{1}{2}\right)y + \frac{1}{4}\theta^2 y^2 \geqslant \frac{1}{2}\left(y + \sqrt{2y^2 - 4y}\right).$$

It follows that (2.7) holds if $\theta \geqslant \Phi(y)$ for all $y > 2$ with

$$\Phi(y) = \frac{\sqrt{4y + 2\sqrt{2y^2 - 4y}} - 2}{y}.$$

A numerical calculation shows that $\max_{y>2} \Phi(y) \approx 0.7249$ (reached for $y \approx 4$). The bound

$$\Phi(y) \leqslant \frac{\sqrt{4y + \frac{2\sqrt{2}}{9} + (\frac{3}{4}y - \frac{2}{3}(1+2\sqrt{2}))^2 + 2\sqrt{2y^2 - 4y + 2}} - 2}{y} = \frac{3}{4} \quad (y > 2)$$

shows that $\theta \geqslant \Phi(y)$ whenever $y > 2$. Hence, condition (2.7) holds and this concludes the proof. □

Similarly as in Section 2.1 we have a sharpness result: if $z_1 = z_2 = \zeta$ with ζ any negative real number, and z_0 is any real number with $z_0 > -2\zeta$, then $T(z_0, z_1, z_2) > 1$, for any given θ, σ.

If $\sigma \geqslant \frac{1}{2}$, the condition on the parameter θ given by Lemma 2.6 can be further weakened when z_0 is assumed to be real as well and one aims at proving (1.16) directly:

Lemma 2.7. *Let $k = 2$. Assume z_0, z_1, z_2 are real numbers, condition* (2.4) *holds and*

$$\frac{1}{2} \leqslant \sigma \leqslant \left(1 + \frac{1}{2}\sqrt{2}\right)\theta.$$

Then (1.16) *holds.*

Proof. First note that (2.4) implies $p \geqslant 1$ and

$$z + z_0 \leqslant z_1 + z_2 + 2\sqrt{z_1 z_2} = -\left(\sqrt{-z_1} - \sqrt{-z_2}\right)^2 \leqslant 0. \tag{2.8}$$

The condition $T \geqslant -1$ is equivalent to

$$2p^2 + (2p-1)(z + z_0) + \sigma(z + z_0)^2 \geqslant 0,$$

which is true since the discriminant is $(2p-1)^2 - 8\sigma p^2 = 4(1-2\sigma)p^2 - 4p + 1 < 0$. In view of (2.8), the condition $T \leqslant 1$ is equivalent to

$$1 - (2\theta - \sigma)z + 2\theta^2 z_1 z_2 + \sigma z_0 \geqslant 0.$$

Write $A = (\sqrt{-z_1} - \sqrt{-z_2})^2$ and $B = (\sqrt{-z_1} + \sqrt{-z_2})^2$. Then, we have

$$\begin{aligned} &1 - (2\theta - \sigma)z + 2\theta^2 z_1 z_2 + \sigma z_0 \\ &\quad \geqslant 1 - (2\theta - \sigma)z + 2\theta^2 z_1 z_2 - 2\sigma\sqrt{z_1 z_2} \\ &\quad = (1 - \theta\sqrt{2z_1 z_2})^2 + \left(1 - \frac{1}{2}\sqrt{2}\right)\theta A + \left[\left(1 + \frac{1}{2}\sqrt{2}\right)\theta - \sigma\right]B \geqslant 0. \end{aligned}$$ □

If $\sigma = \frac{1}{2}$ the assumption on θ given by Lemma 2.7 becomes $\theta \geqslant 1/(2+\sqrt{2}) \approx 0.2929$. In comparison, the Douglas and Craig & Sneyd schemes require $\theta \geqslant \frac{1}{2}$ for (1.14), (1.15) to hold, respectively, under the same condition on z_0, z_1, z_2 (consider $z_0 = 0$).

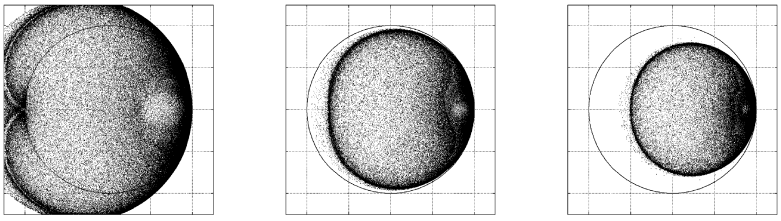
The important question remains whether (1.16) is fulfilled with $k = 2$ for general complex z_0, z_1, z_2 satisfying (2.4). Note that real z_1, z_2 are just relevant to the case of (semi-discretized) pure diffusion equations (1.17), where $\mathbf{c} = \mathbf{0}$. An analytical answer to our question appears hard to obtain. For example, it does not seem possible to apply the maximum modulus principle as is done in [13] for the (special) case of $z_0 = 0$. We therefore conducted a numerical test to gain insight into the possible result.

Let $r_{i,j}$ for $i, j = 0, 1, 2$ denote independent, uniformly distributed random numbers in the interval $[0, 1]$. In the experiment, we consider complex (random) points

$$z_j = -10^{1-5r_{1,j}} \pm I\, 10^{1-5r_{2,j}}$$

for $j = 1, 2$ and

$$z_0 = 2\sqrt{\Re z_1 \Re z_2}\; r_{1,0} e^{2\pi I r_{2,0}}$$



$\theta = \frac{1}{2}$ $\theta = \frac{1}{2} + \frac{1}{6}\sqrt{3}$ $\theta = 1$

Fig. 1. The values $T(z_0, z_1, z_2)$ for random points (z_0, z_1, z_2) satisfying (2.4).

so that (2.4) holds. The values $T(z_0, z_1, z_2)$ are then plotted in the complex plane, together with the unit circle, for the three cases $\theta = \frac{1}{2}$, $\theta = \frac{1}{2} + \frac{1}{6}\sqrt{3}$ and $\theta = 1$ with always $\sigma = \frac{1}{2}$, see Fig. 1. In each case, 10^6 random triplets (z_0, z_1, z_2) were used.[2]

The results of Fig. 1 support the interesting conjecture that, if $\sigma = \frac{1}{2}$ and $\theta \geqslant \frac{1}{2} + \frac{1}{6}\sqrt{3}$, then (1.16) is fulfilled for $k = 2$ whenever (2.4) holds (with general complex z_0, z_1, z_2). On the other hand, this conclusion does clearly not appear to be valid when $\sigma = \frac{1}{2}$ and $\theta = \frac{1}{2}$.

2.3. Main conclusions of Section 2 relevant to $k = 2$

In the following theorem, we summarize the main results from Sections 2.1, 2.2 on the stability requirements (1.14)–(1.16) relevant to $k = 2$. Unless stated otherwise, z_0, z_1, z_2 are assumed to be complex numbers here.

Theorem 2.8. *Assume $k = 2$ and* (2.4) *holds. Let $c = 1 + \frac{1}{2}\sqrt{2}$. Then*

(a) (1.14) *is fulfilled whenever $\theta \geqslant \frac{1}{2}$,*
(b) (1.15) *is fulfilled whenever $\theta \geqslant \frac{1}{2}$ and $\sigma \leqslant 2\theta$,*
(c) (1.16) *is fulfilled for real z_1, z_2 whenever $2(1-\theta) \leqslant \sigma \leqslant c\,\theta$,*
(d) (1.16) *is fulfilled for real z_0, z_1, z_2 whenever $\min\{\frac{1}{2}, 2(1-\theta)\} \leqslant \sigma \leqslant c\,\theta$.*

Next, we have

Conjecture 2.9. *If $k = 2$ and* (2.4) *holds, then* (1.16) *is fulfilled whenever $\sigma = \frac{1}{2}$ and $\theta \geqslant \frac{1}{2} + \frac{1}{6}\sqrt{3}$.*

3. Application to two-dimensional convection–diffusion equations

We consider Eq. (1.17) with $k = 2$ in the unit square $[0, 1] \times [0, 1]$ and write

$$\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}.$$

We rearrange it as

$$\frac{\partial u}{\partial t} = (d_{12} + d_{21})u_{xy} + (c_1 u_x + d_{11} u_{xx}) + (c_2 u_y + d_{22} u_{yy}). \tag{3.1}$$

We require that D be positive semi-definite. This is equivalent to

$$d_{11} \geqslant 0, \qquad d_{22} \geqslant 0 \quad \text{and} \quad (d_{12} + d_{21})^2 \leqslant 4 d_{11} d_{22}. \tag{3.2}$$

We approximate all spatial derivatives in (3.1) using second-order central discretizations on a rectangular grid with mesh widths Δx and Δy in the x and y directions, respectively:

[2] Considering values z_j of larger modulus does not affect the visual appearance of the figure.

$$(u_x)_{i,j} \approx \delta_x u_{i,j} = \frac{u_{i+1,j} - u_{i-1,j}}{2\Delta x}, \tag{3.3a}$$

$$(u_y)_{i,j} \approx \delta_y u_{i,j} = \frac{u_{i,j+1} - u_{i,j-1}}{2\Delta y}, \tag{3.3b}$$

$$(u_{xx})_{i,j} \approx \Delta_x u_{i,j} = \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{(\Delta x)^2}, \tag{3.3c}$$

$$(u_{yy})_{i,j} \approx \Delta_y u_{i,j} = \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{(\Delta y)^2}, \tag{3.3d}$$

$$\begin{aligned}(u_{xy})_{i,j} \approx{}& \frac{(1+\beta)(u_{i+1,j+1} + u_{i-1,j-1}) - (1-\beta)(u_{i-1,j+1} + u_{i+1,j-1})}{4\Delta x \Delta y} \\ &+ \frac{4\beta u_{i,j} - 2\beta(u_{i+1,j} + u_{i,j+1} + u_{i-1,j} + u_{i,j-1})}{4\Delta x \Delta y},\end{aligned} \tag{3.3e}$$

where β denotes a real parameter with $-1 \leqslant \beta \leqslant 1$ and we use the notation $u_{i,j} = u(i\Delta x, j\Delta y, t)$. The right-hand side of (3.3e) is the most general form of a second-order approximation for the cross derivative u_{xy} based on a centered 9-point stencil. When $\beta = 0$, (3.3e) reduces to the standard 4-point stencil (see e.g. [16, p. 81])

$$(u_{xy})_{i,j} \approx \delta_x \delta_y u_{i,j} = \frac{u_{i+1,j+1} + u_{i-1,j-1} - u_{i-1,j+1} - u_{i+1,j-1}}{4\Delta x \Delta y}.$$

The arrangement (3.1) naturally leads to a splitted, semi-discrete system of the form (1.1), (1.2) where $F_j(t, v) = A_j v$ for $j = 0, 1, 2$ with constant matrices A_j. The matrix A_0 represents the cross derivative term in (3.1) and A_1, A_2 represent the spatial derivatives in the x and y directions, respectively.

When the boundary condition is periodic, the A_j are Kronecker products of circulant (thus normal) matrices and commute with each other. In this case stability can be investigated using the model scalar equation (1.6) with $k = 2$ and with λ_j an eigenvalue of A_j for $j = 0, 1, 2$; this is equivalent to a von Neumann stability analysis. Write $I = \sqrt{-1}$. It is readily verified, upon inserting discrete Fourier modes, that the scaled eigenvalues $z_j = \lambda_j \Delta t$ are given by

$$\begin{aligned}z_j &= c_j q_j \left(\frac{1}{2} e^{I\phi_j} - \frac{1}{2} e^{-I\phi_j}\right) + d_{jj} a_j \left(e^{I\phi_j} - 2 + e^{-I\phi_j}\right) \\ &= -2d_{jj} a_j (1 - \cos\phi_j) + I c_j q_j \sin\phi_j\end{aligned} \tag{3.4a}$$

for $j = 1, 2$, and

$$z_0 = (d_{12} + d_{21}) b\big[-\sin\phi_1 \sin\phi_2 + \beta(1 - \cos\phi_1)(1 - \cos\phi_2)\big], \tag{3.4b}$$

where

$$q_1 = \frac{\Delta t}{\Delta x}, \qquad q_2 = \frac{\Delta t}{\Delta y}, \qquad a_1 = \frac{\Delta t}{(\Delta x)^2}, \qquad a_2 = \frac{\Delta t}{(\Delta y)^2} \quad \text{and} \quad b = \frac{\Delta t}{\Delta x \Delta y}.$$

The angles ϕ_j are integer multiples of $2\pi/m_j$ ($j = 1, 2$) where m_1, m_2 are the dimensions of the grid in the x and y directions, respectively.

By (3.2) we have $d_{jj} \geqslant 0$ and consequently $\Re z_j \leqslant 0$ ($j = 1, 2$). Using the condition (3.2), the Cauchy–Schwarz inequality applied to the vectors

$$\mathbf{v}_1 = \begin{bmatrix} -\sin\phi_1 \\ \beta(1 - \cos\phi_1) \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} \sin\phi_2 \\ 1 - \cos\phi_2 \end{bmatrix},$$

and $\beta^2 \leqslant 1$, we obtain

$$\begin{aligned}|z_0|^2 &\leqslant 4d_{11}d_{22} \cdot b^2 \cdot \|\mathbf{v}_1\|_2^2 \, \|\mathbf{v}_2\|_2^2 \\ &\leqslant 4d_{11}d_{22} \cdot a_1 a_2 \cdot 2(1 - \cos\phi_1) \cdot 2(1 - \cos\phi_2) \\ &= 4 \cdot 2d_{11}a_1(1 - \cos\phi_1) \cdot 2d_{22}a_2(1 - \cos\phi_2) \\ &= 4 \cdot \Re z_1 \cdot \Re z_2.\end{aligned} \tag{3.5}$$

Thus we have shown that the condition (2.4) is fulfilled, independently of Δt, Δx, $\Delta y > 0$. By invoking Theorem 2.8, parts a and b, we arrive at the following result for the Douglas and Craig & Sneyd schemes:

Theorem 3.1. *Consider equation* (1.17) *for* $k = 2$ *with* (3.2) *and periodic boundary condition. Assume* (1.1), (1.2) *is obtained after semi-discretization and splitting of* (1.17) *as described in this section. Let the parameters* θ, σ *be such that* $\theta \geqslant \frac{1}{2}$ *and* $\sigma \leqslant 2\theta$. *Then the two schemes* (1.3), (1.4) *are both unconditionally stable when applied to* (1.1), (1.2). *Moreover, this conclusion remains valid when any other, stable finite difference discretizations for* u_x, u_y *are used in place of* (3.3a), (3.3b).

The last part of Theorem 3.1 follows easily from the fact that with any other given stable[3] finite difference discretizations of u_x, u_y one obtains the bound

$$\Re z_j \leqslant -2d_{jj}a_j(1-\cos\phi_j) \leqslant 0, \quad j = 1, 2.$$

As a consequence, the above proof of (2.4) is still valid, except that in (3.5) one needs to replace the last equality "=" by an inequality "$\leqslant$".

Theorem 3.1 substantially extends and improves the results from the literature for $k = 2$ reviewed in Section 1.2. In particular, for the Craig & Sneyd scheme (1.4) we have unconditional stability in the case of general 2D convection–diffusion equations, instead of pure diffusion equations [2], and simultaneously, for a larger set of parameters θ, σ than in [2]. Furthermore, our result for the two schemes (1.3), (1.4) is valid for a much wider range of spatial discretizations of the mixed derivative and convection terms than considered in [2,15].

By application of part d of Theorem 2.8, we arrive at the following result for the Hundsdorfer & Verwer scheme:

Theorem 3.2. *Consider Eq.* (1.17) *for* $k = 2$ *with* $\mathbf{c} = \mathbf{0}$ *and* (3.2) *and with periodic boundary condition. Assume* (1.1), (1.2) *is obtained after semi-discretization and splitting of* (1.17) *as described in this section. Assume the parameters* θ, σ *are such that* $\min\{\frac{1}{2}, 2(1-\theta)\} \leqslant \sigma \leqslant (1+\frac{1}{2}\sqrt{2})\theta$. *Then the scheme* (1.5) *is unconditionally stable when applied to* (1.1), (1.2).

Theorem 3.2 appears to be the first result in the literature on the (unconditional) stability of the scheme (1.5) in the case of diffusion equations with mixed derivative terms. We currently do not have a proof of whether the unconditional stability is maintained in the presence of convection, but, in view of Conjecture 2.9, we suspect that a result completely similar to Theorem 3.1 will hold for the scheme (1.5), at least when $\sigma = \frac{1}{2}$ and $\theta \geqslant \frac{1}{2} + \frac{1}{6}\sqrt{3}$.

4. Numerical experiments

We discuss numerical experiments in the case of the 2D convection–diffusion equation (1.17) on the unit square $\Omega = [0, 1] \times [0, 1]$ with

$$\mathbf{c} = -\begin{bmatrix} 2 \\ 3 \end{bmatrix}, \qquad D = 0.025\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \tag{4.1}$$

and with initial condition

$$u(x, y, 0) = e^{-4(\sin^2\pi x + \cos^2\pi y)} \quad (0 \leqslant x, y \leqslant 1). \tag{4.2}$$

Note that the requirement (3.2) is fulfilled for the matrix D above. In fact, the last inequality in (3.2) becomes an equality for D given by (4.1). Hence, for the given diagonal entries of the diffusion matrix, the coefficient corresponding to the mixed derivative term is the largest possible.

In our experiments we consider a periodic boundary condition, i.e., $u(x \pm 1, y \pm 1, t) = u(x, y, t)$ for all x, y, t. The semi-discretization of the initial-boundary value problem is performed as in Section 3, with the parameter β in (3.3e) taken equal to zero. The exact solution U to the obtained initial value problem for (1.1), with $F(t, v) = Av$ and constant matrix A, is given by $U(t) = e^{tA}U_0$. Fig. 2 shows the solution values $U(0)$ and $U(2)$ displayed on the grid in Ω, so that they represent the exact solution to the initial-boundary value problem for (1.17) at $t = 0$ and $t = 2$ (we chose $\Delta x = \Delta y = 1/40$).

We subsequently employ a splitting (1.2) of the semi-discrete system (1.1) as given in Section 3. To the resulting splitted semi-discrete problem, with $k = 2$, we have applied each of the three ADI schemes from Section 1.1:

[3] I.e., the corresponding semi-discretizations of the 1D scalar equations $u_t = cu_x$ for $c = c_1$, c_2, with periodic boundary condition, are stable.

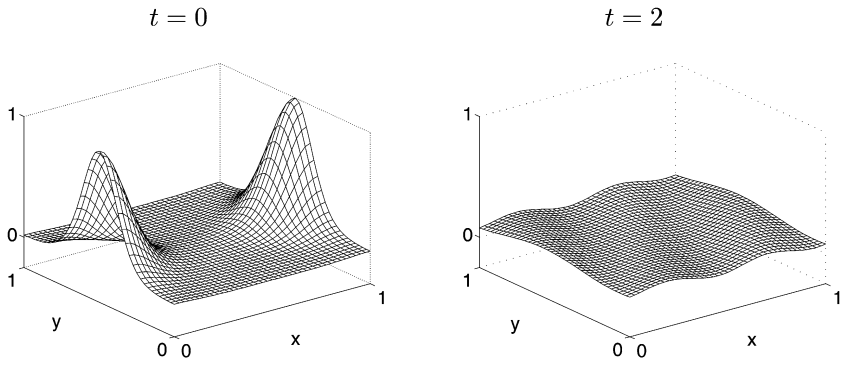


Fig. 2. Exact solution $U(t)$ at $t = 0, 2$.

the Douglas scheme (1.3) with $\theta = \frac{1}{2}$, the Craig & Sneyd scheme (1.4) with $\theta = \sigma = \frac{1}{2}$, and the Hundsdorfer & Verwer scheme (1.5) with $\sigma = \frac{1}{2}$ and the two values $\theta = \frac{1}{2}$, $\theta = \frac{1}{2} + \frac{1}{6}\sqrt{3}$, respectively.

For the Douglas and Craig & Sneyd schemes under consideration, the assumption on the parameters θ, σ from Theorem 3.1 is fulfilled. For the two Hundsdorfer & Verwer schemes, the requirement on θ, σ from Conjecture 2.9 and the related discussion following Theorem 3.2 is fulfilled if $\theta = \frac{1}{2} + \frac{1}{6}\sqrt{3}$, but *not* if $\theta = \frac{1}{2}$.

Fig. 3 displays, for two choices of the spatial mesh width, the normalized global error

$$e = \frac{1}{\sqrt{m_1 m_2}} \left\| U(2) - U_{2N} \right\|_2 \tag{4.3}$$

as a function of $\Delta t = 1/N$, where m_1 and m_2 denote the grid dimensions in the x and y directions, respectively. For the (only) purpose of indicating the stiffness of the initial value problems (1.1), we have added in Fig. 3 the global errors obtained with the explicit Euler method, RK1, and the second-order explicit improved Euler method, RK2.

The results of Fig. 3 are clearly consistent with an unconditionally stable behavior of the Douglas and Craig & Sneyd schemes, cf. Theorem 3.1. Furthermore, Fig. 3 provides additional support for the conjecture following Theorem 3.2 on the unconditional stability of the Hundsdorfer & Verwer scheme when applied to general parabolic 2D convection–diffusion equations: for $\theta = \frac{1}{2} + \frac{1}{6}\sqrt{3}$ the error curve is consistent with an unconditionally stable behavior of the scheme, whereas for $\theta = \frac{1}{2}$ it is not.

To gain more insight in the latter case of $\theta = \frac{1}{2}$, we have displayed in Fig. 4 the spectral radii (norms) of the numerical time-step iteration matrices, arising in the experiments, versus the time step Δt. Clearly, the regions of time steps in Fig. 3 where the scheme (1.5) with $\theta = \sigma = \frac{1}{2}$ has large global errors—relative to the other three ADI schemes—correspond exactly to the regions in Fig. 4 where the spectral radius of the numerical iteration matrix is greater than 1. Thus this strongly indicates that the large errors in Fig. 3 for the scheme (1.5) with $\theta = \sigma = \frac{1}{2}$ are indeed caused by a lack of unconditional stability.[4] For the other three ADI schemes, the spectral radii are always equal to 1, and this agrees with a presence of unconditional stability.

It is interesting to consider in Fig. 3 also the convergence behavior of the ADI schemes. When the time step is sufficiently small, the Douglas scheme is clearly outperformed by the Craig & Sneyd and the Hundsdorfer & Verwer schemes. Fig. 3 suggests that the global errors for the Douglas scheme behave like $C\Delta t$, whereas for the Craig & Sneyd and Hundsdorfer & Verwer schemes they behave like $C(\Delta t)^2$, with certain constants C that only depend weakly on the spatial mesh widths.[5]

We remark that we have also performed numerical experiments in the case where the periodic boundary condition is replaced by the Dirichlet boundary condition $u(x, y, t) = u(x, y, 0)$ for $(x, y) \in \partial\Omega$ and $t \geqslant 0$, with $u(x, y, 0)$ given by (4.2). The semi-discretization was again done as in Section 3, except that, in order to avoid unwanted oscillations, the convection terms were discretized using upwind differencing near the outflow part of the boundary. In the Dirichlet case, we obtained exactly the same conclusions concerning unconditional stability as in the periodic case above. Note,

[4] Clearly, there does appear to be *conditional* stability.

[5] Even though this agrees with the "non-stiff" classical orders of the schemes, cf. Section 1.1, it is a non-trivial result to prove and we leave this issue for future research.

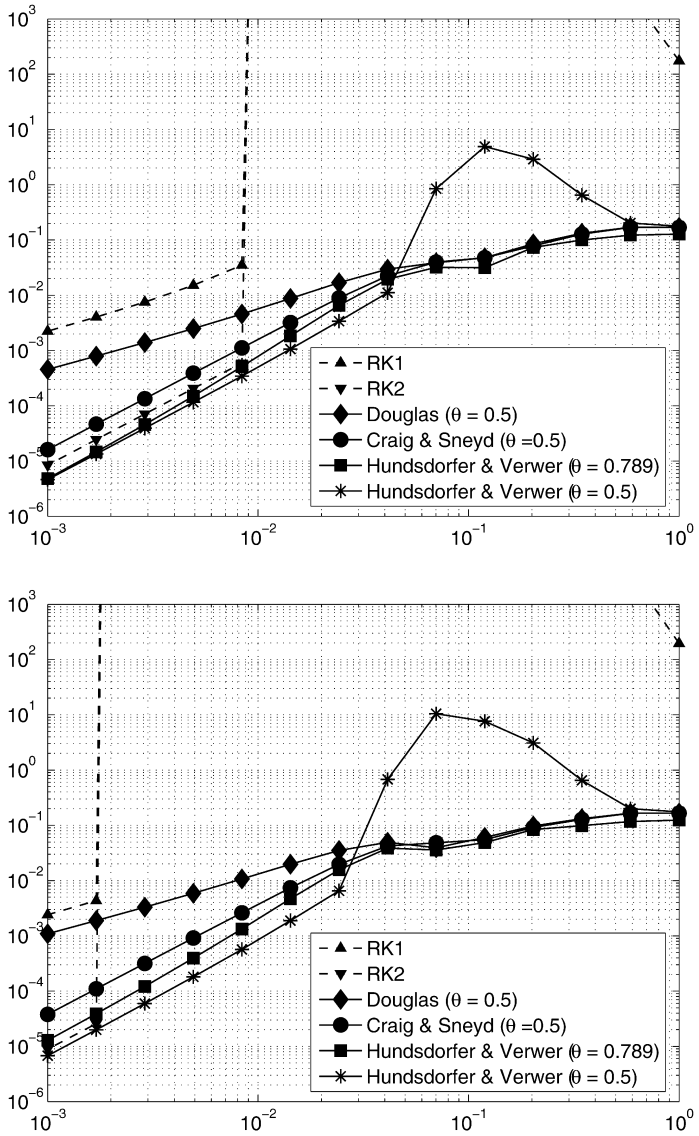


Fig. 3. Convergence curves e vs. Δt in the case of the semi-discretized problem (1.17) with (4.1), (4.2) and periodic boundary condition for $\Delta x = \Delta y = \frac{1}{20}$ (top) and $\Delta x = \Delta y = \frac{1}{40}$ (bottom).

however, that the stability analysis in this paper is not directly relevant to the Dirichlet case, since the present matrix A is not normal. Hence, we have some evidence that the positive stability results for the ADI schemes (1.3)–(1.5) derived in this paper extend to cases beyond the von Neumann framework.

5. Concluding remarks and extensions

We have presented in this paper several linear stability results for the three ADI schemes (1.3)–(1.5) formulated in Section 1.1. In particular:

- Our proof of the unconditional stability of the Douglas scheme (1.3) when applied to general parabolic two-dimensional convection–diffusion equations (1.17) substantially simplifies the proof given in [15], while also holding for a wider range of spatial discretizations.

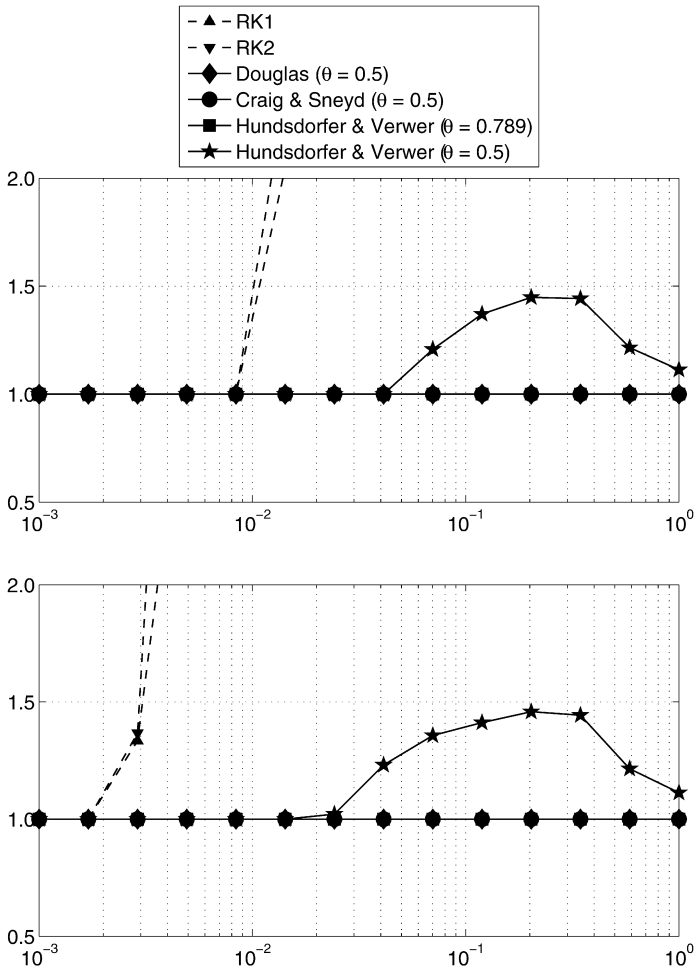


Fig. 4. Spectral radii of the numerical time-step iteration matrices vs. Δt in the case of the semi-discretized problem (1.17) with (4.1), (4.2) and periodic boundary condition for $\Delta x = \Delta y = \frac{1}{20}$ (top) and $\Delta x = \Delta y = \frac{1}{40}$ (bottom).

- To our knowledge the unconditional stability result proved in this paper for the Craig & Sneyd scheme (1.4) when applied to general parabolic two-dimensional convection–diffusion equations is new in the literature.
- To our knowledge the unconditional stability result proved in this paper for the Hundsdorfer & Verwer scheme (1.5) when applied to general parabolic two-dimensional diffusion equations (1.17) (with $\mathbf{c} = \mathbf{0}$) is new in the literature.

The strength of condition (2.4) in proving (1.14)–(1.16) extends to certain higher-order multi-dimensional partial differential equations, for which the ADI approach becomes even more appealing in view of the increasingly more stringent stability restrictions on the time step for explicit integrators. For example, the two-dimensional, fourth-order problem

$$\frac{\partial u}{\partial t} = -\Delta^2 u = -2u_{xxyy} - u_{xxxx} - u_{yyyy}, \tag{5.1}$$

naturally involves a cross derivative term. The biharmonic operator Δ^2 arises in many important linear and nonlinear problems (e.g. the Kuramoto–Sivashinsky equation modelling thermal diffusive instabilities). Approximating the spatial derivatives using second-order central discretizations

$$(u_{xxxx})_{i,j} \approx \delta_x^2 u_{i,j}, \qquad (u_{yyyy})_{i,j} \approx \delta_y^2 u_{i,j}, \qquad (u_{xxyy})_{i,j} \approx \delta_x \delta_y u_{i,j}$$

on a uniform rectangular grid (cf. (3.3c,d)) yields, with the cross derivative term in F_0 and the derivatives in the x and y directions in F_1 and F_2, respectively, the following formulas for the scaled eigenvalues z_0, z_1, z_2:

$$z_0 = -8\tilde{b}(1-\cos\phi_1)(1-\cos\phi_2)$$

and

$$z_j = -4\tilde{a}_j(1-\cos\phi_j)^2, \quad j=1,2,$$

where

$$\tilde{a}_1 = \frac{\Delta t}{(\Delta x)^4}, \qquad \tilde{a}_2 = \frac{\Delta t}{(\Delta y)^4}, \qquad \tilde{b} = \frac{\Delta t}{(\Delta x)^2(\Delta y)^2}.$$

Note that these eigenvalues are real. Clearly, $z_1 \leqslant 0$, $z_2 \leqslant 0$ and $|z_0| = 2\sqrt{z_1 z_2}$, so that condition (2.4) holds. In particular, all the ADI schemes discussed in this paper can be applied to solve (5.1) with suitable values θ, σ.

Finally, we mention some important issues which are currently under investigation:

- Similar to the analysis in [2] for the schemes (1.3) and (1.4), we are considering [7] the unconditional stability of the Hundsdorfer & Verwer scheme (1.5) when applied to semi-discrete pure diffusion equations with full matrix D in arbitrary spatial dimensions $k \geqslant 2$. It appears that the results for the scheme (1.5) are more favorable compared to (1.3) and (1.4).
- A proof of Conjecture 2.9 concerning the Hundsdorfer & Verwer scheme. Numerical experiments indicate that the condition (2.4) implies condition (2.6), for the relevant parameter values θ, σ. Current efforts are aimed at establishing this implication, which would yield a proof of the conjecture using Lemma 2.5.
- An analysis of the convergence behavior of the ADI schemes when applied to semi-discrete convection–diffusion equations with full diffusion matrix D, cf. Section 4. Here the interest is in bounds on the global error of the type $C(\Delta t)^r$ with $r \geqslant 1$ and a constant C that is independent of the mesh widths in the semi-discretization. For various time-integration schemes, including (1.3) and (1.5), global error bounds of this type have been derived in the literature for the case of equations with diagonal matrix D, see e.g. [10,11].
- The numerical experiments mentioned at the end of Section 4 on a two-dimensional convection–diffusion equation with a Dirichlet boundary condition suggest that the stability results derived in this paper extend to cases beyond the von Neumann framework. A rigorous proof of stability for $k=2$ and $F_j(t,v) = A_j v$, $j=0,1,2$, with non-normal and/or non-commuting matrices A_j, requires bounding matrix norms such as, for Craig & Sneyd's scheme,

  $$\left\| I + \left(I + \sigma\Delta t P^{-1} A_0\right)\Delta t P^{-1} A \right\|$$

  where $A = A_0 + A_1 + A_2$, $P = (I - \theta\Delta t A_1)(I - \theta\Delta t A_2)$, and I is the identity matrix. Contractivity results relevant to non-normal, commuting matrices may be obtained along the lines of [6,12].

Acknowledgements

The authors would like to thank two anonymous referees for their constructive remarks, which have enhanced the presentation of this paper.

References

[1] P.L.T. Brian, A finite-difference method of high-order accuracy for the solution of three-dimensional transient heat conduction problems, AIChE J. 7 (1961) 367–370.
[2] I.J.D. Craig, A.D. Sneyd, An alternating-direction implicit scheme for parabolic equations with mixed derivatives, Comput. Math. Appl. 16 (1988) 341–350.
[3] J. Douglas, Alternating direction methods for three space variables, Numer. Math. 4 (1962) 41–63.
[4] J. Douglas, H.H. Rachford, On the numerical solution of heat conduction problems in two and three space variables, Trans. Amer. Math. Soc. 82 (1956) 421–439.
[5] J. Elf, P. Lötstedt, P. Sjöberg, Problems of high dimension in molecular biology, in: W. Hackbusch (Ed.), Proceedings of the 19th GAMM-Seminar, Max-Planck-Institute for Mathematic in the Sciences, Leipzig, 2003, pp. 21–30, http://www.mis.mpg.de/conferences/gamma/2003/.
[6] K.J. in 't Hout, On the contractivity of implicit–explicit linear multistep methods, Appl. Numer. Math. 42 (2002) 201–212.

[7] K.J. in 't Hout, B.D. Welfert, Unconditional stability of ADI schemes applied to multi-dimensional diffusion equations with mixed derivative terms, in preparation.
[8] P.J. van der Houwen, J.G. Verwer, One-step splitting methods for semi-discrete parabolic equations, Computing 22 (1979) 291–309.
[9] W. Hundsdorfer, Stability of approximate factorization with θ-methods, BIT 39 (1999) 473–483.
[10] W. Hundsdorfer, Accuracy and stability of splitting with Stabilizing Corrections, Appl. Numer. Math. 42 (2002) 213–233.
[11] W. Hundsdorfer, J.G. Verwer, Numerical Solution of Time-Dependent Advection–Diffusion–Reaction Equations, Springer Ser. Comput. Math., vol. 33, Springer, Berlin, 2003.
[12] J.C. Jorge, F. Lisbona, Contractivity results for alternating direction schemes in Hilbert spaces, Appl. Numer. Math. 15 (1994) 65–75.
[13] D. Lanser, J.G. Blom, J.G. Verwer, Time integration of the shallow water equations in spherical geometry, J. Comput. Phys. 171 (2001) 373–393.
[14] S. McKee, A.R. Mitchell, Alternating direction methods for parabolic equations in two space dimensions with a mixed derivative, Comput. J. 13 (1970) 81–86.
[15] S. McKee, D.P. Wall, S.K. Wilson, An alternating direction implicit scheme for parabolic equations with mixed derivative and convective terms, J. Comput. Phys. 126 (1996) 64–76.
[16] A.R. Mitchell, D.F. Griffiths, The Finite Difference Method in Partial Differential Equations, Wiley, New York, 1980.
[17] D.W. Peaceman, Fundamentals of Numerical Reservoir Simulation, Developments in Petroleum Science, vol. 6, Elsevier, Amsterdam, 1977.
[18] C. Randall, PDE Techniques for Pricing Derivatives with Exotic Path Dependencies or Exotic Processes, Lecture Notes, Workshop CANdiensten, Amsterdam, May 2002.
[19] D. Tavella, C. Randall, Pricing Financial Instruments: The Finite Difference Method, Wiley, New York, 2000.