Modified Douglas Splitting Methods for Reaction-Diffusion Equations

A. Arrarás[*], K.J. in 't Hout[†], W. Hundsdorfer[‡], L. Portero[§]

Abstract

We present modifications of the second-order Douglas stabilizing corrections method, which is a splitting method based on the implicit trapezoidal rule. Inclusion of an explicit term in a forward Euler way is straightforward, but this will lower the order of convergence. In the modifications considered here, explicit terms are included in a second-order fashion. For these modified methods, results on linear stability and convergence are derived. Stability holds for important classes of reaction-diffusion equations, and for such problems the modified Douglas methods are seen to be often more efficient than related methods from the literature.

1 Introduction

In this paper we consider numerical methods for solving initial value problems for ordinary differential equations (ODEs) on a time interval $[0, T]$,

$$u'(t) = F(t, u(t)), \qquad u(0) = u_0, \tag{1.1}$$

with given $F : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ and $u_0 \in \mathbb{R}^m$. In our applications, this system will be obtained from spatial discretization of a partial differential equation (PDE) by finite differences or finite elements (with mass lumping). Then (1.1) is called a semi-discrete system and m will be proportional to the number of grid points in space. In this formulation, inhomogeneous boundary values for the PDE are directly incorporated inside F as source terms.

For many problems occurring in practice there will be a natural decomposition

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

in which the separate component functions F_j are more simple than the whole F. It will be assumed that F_0 is a non-stiff or mildly stiff term that can be treated explicitly in a time stepping method. The other terms F_j will be treated in an implicit fashion, in such a way that in each internal stage at most one of these F_j appears implicitly.

[*]Departamento de Ingeniería Matemática e Informática, Universidad Pública de Navarra, Campus de Arrosadía, 31006 Pamplona, Spain. E-mail: andres.arraras@unavarra.es

[†]Department of Mathematics and Computer Science, University of Antwerp, Middelheimlaan 1, B-2020 Antwerp, Belgium. Email: karel.inthout@uantwerp.be.

[‡]CWI, Science Park 123, Amsterdam, The Netherlands. E-mail: willem.hundsdorfer@cwi.nl

[§]Departamento de Ingeniería Matemática e Informática, Universidad Pública de Navarra, Campus de Arrosadía, 31006 Pamplona, Spain. Email: laura.portero@unavarra.es

1.1 Modified Douglas methods

In the following we denote by $\Delta t > 0$ the step-size and $u_n \approx u(t_n)$ stands for the numerical approximation to the exact solution at time level $t_n = n\Delta t$, $n \geq 0$.

A simple method that employs the splitting (1.2) is the Douglas method, which has been presented in [17, 18] as

$$
\begin{cases}
v_0 &= u_{n-1} + \Delta t\, F(t_{n-1}, u_{n-1})\,, \\
v_j &= v_{j-1} + \theta\Delta t\big(F_j(t_n, v_j) - F_j(t_{n-1}, u_{n-1})\big) \qquad (j = 1, 2, \ldots, s)\,, \\
u_n &= v_s\,,
\end{cases}
\tag{1.3}
$$

with parameter $\theta \geq \frac{1}{2}$ and with $n = 1, 2, \ldots$ indicating the time level. The internal vectors v_j, which depend of course also on n, are all consistent with the exact solution at time t_n, and consequently steady state solutions $F(u) = 0$ for autonomous equations are stationary points of the method. This is an advantage over many other splitting methods, based on Lie or Strang splitting, where such consistency is often absent.

Originally, for $F = F_1 + \cdots + F_s$ without explicit terms, such methods were introduced by J. Douglas Jr and co-workers for multi-dimensional parabolic problems with dimensional splittings; see for instance [6, 7], and also [18, p. 373] and [19] for additional references. In the terminology of [19], methods like (1.3) are also known as stabilizing correction methods.

If $\theta = \frac{1}{2}$ and $F_0 = 0$ the method (1.3) is convergent of order two in the classical ODE sense, that is, for fixed (non-stiff) ODE systems. However, for non-zero F_0 the order will only be one, because the explicit term is treated in a forward Euler fashion.

To improve this treatment of the explicit terms, we consider the following modification

$$
\begin{cases}
v_* &= u_{n-1} + \Delta t\, F(t_{n-1}, u_{n-1})\,, \\
v_0 &= v_* + \frac{1}{2}\Delta t\big(F_0(t_n, v_*) - F_0(t_{n-1}, u_{n-1})\big)\,, \\
v_j &= v_{j-1} + \frac{1}{2}\Delta t\big(F_j(t_n, v_j) - F_j(t_{n-1}, u_{n-1})\big) \qquad (j = 1, 2, \ldots, s)\,, \\
u_n &= v_s\,.
\end{cases}
\tag{1.4}
$$

Now the explicit term is treated as in the explicit trapezoidal rule (also known as the modified Euler method). As we will see, this method is indeed of order two in the classical ODE sense, and convergence with order two will also be valid for interesting classes of PDEs of reaction-diffusion type.

A closely related variant is obtained if the extra stage with F_0 is performed last, leading to

$$
\begin{cases}
v_0 &= u_{n-1} + \Delta t\, F(t_{n-1}, u_{n-1})\,, \\
v_j &= v_{j-1} + \frac{1}{2}\Delta t\big(F_j(t_n, v_j) - F_j(t_{n-1}, u_{n-1})\big) \qquad (j = 1, 2, \ldots, s)\,, \\
u_n &= v_s + \frac{1}{2}\Delta t\big(F_0(t_n, v_s) - F_0(t_{n-1}, u_{n-1})\big)\,.
\end{cases}
\tag{1.5}
$$

Again, the explicit term is treated as in the explicit trapezoidal rule, and the method is of order two in the classical ODE sense. It will be seen, however, that for semi-discrete systems obtained from PDEs proving second-order convergence is more difficult than for (1.4) if we have a splitting with $s \geq 2$.

These modifications (1.4) and (1.5) were briefly considered in [17], but the methods were deemed not to be sufficiently stable if F_0 contains discretized convection terms.

In this paper we will consider applications to reaction-diffusion equations and it will be seen that for such applications the modified Douglas methods give good results compared to some other stabilizing correction methods.

For $s = 1$, both (1.4) and (1.5) are implicit-explicit (IMEX) methods where the implicit trapezoidal rule is combined with its explicit counterpart. These two IMEX methods have been considered and tested in [18, pp. 391–395]. Additional splittings of the implicit terms, giving $s \geq 2$, are important in many practical situations.

1.2 Outline

In this paper we will study stability and convergence properties of the modified Douglas methods (1.4) and (1.5). The analysis will be performed for linear problems with $F_j(t, u) = A_j u + g_j(t)$. Numerical tests will show the relevance for nonlinear problems.

Stability is considered in Section 2 in a von Neumann-type analysis. The results indicate that the modified schemes will be mostly suited for classes of problems where all implicit terms are known a priori to have eigenvalues close to the negative real axis, such as reaction-diffusion equations.

In Section 3 error bounds for the methods are derived. These error bounds are then discussed in some detail in Section 4 for linear 2D model problems $u_t = \Delta u + f$ with dimension splitting.

In Section 5 numerical test results are presented for nonlinear reaction-diffusion problems. The modified Douglas methods will be compared to methods from the literature, in particular from [5, 12, 17]. Apart from dimension splitting, which requires Cartesian grids, we will also consider domain decomposition splitting with smooth overlappings [20], which can be used on arbitrary grids in combination with spatial discretization by finite volume or finite element methods.

2 Linear Stability

Stability will first be studied for the simple scalar test equation

$$u'(t) = \big(\lambda_0 + \lambda_1 + \cdots + \lambda_s\big)u(t)\,, \qquad \lambda_j \in \mathbb{C}\,. \tag{2.1}$$

The relevance for PDEs is discussed afterwards.

2.1 Stability for the scalar test equation

Consider equation (2.1), and let $z_j = \Delta t\,\lambda_j$. Applied to this test equation, both (1.4) and (1.5) give the recursion $u_n = R\,u_{n-1}$ where $R = r(z_0, \ldots, z_s)$ is given by the stability function

$$r(z_0, z_1, \ldots, z_s) = 1 + \frac{\big(1 + \frac{1}{2}z_0\big)\sum_{j=0}^{s} z_j}{\prod_{j=1}^{s}\big(1 - \frac{1}{2}z_j\big)}\,. \tag{2.2}$$

In this section we will consider the condition

$$|r(z_0, z_1, \ldots, z_s)| \leq 1\,, \tag{2.3}$$

which provides stability with respect to initial perturbations on arbitrary time intervals $[0, T]$. In the results, we will use

$$\mathcal{W}_\alpha = \{\zeta \in \mathbb{C} : \ \zeta = 0 \ \text{ or } \ |\arg(-\zeta)| \leq \alpha\}$$

to denote closed wedges in the left half-plane, with angle $\alpha \in [0, \frac{1}{2}\pi]$.

3

For the case that $z_0 = 0$, $s \geq 2$, the following results were derived in [16]:

$$|r(0, z_1, \ldots, z_s)| \leq 1 \text{ for all } z_i \in \mathcal{W}_\alpha, 1 \leq i \leq s \iff \alpha \leq \frac{1}{s-1}\frac{\pi}{2}, \quad (2.4)$$

and if $1 \leq k \leq s - 1$ then

$$\left.\begin{array}{c} |r(0, z_1, \ldots, z_s)| \leq 1 \text{ for all } z_1, \ldots, z_{s-k} \in \mathcal{W}_\alpha \\ \text{and } z_{s-k+1}, \ldots, z_s \leq 0 \end{array}\right\} \iff \alpha \leq \frac{1}{s-k}\frac{\pi}{2}. \quad (2.5)$$

If $z_0 = 0$ and $s = 1$, then we just have the implicit trapezoidal rule and stability will hold precisely for $z_1 \in \mathcal{W}_{\pi/2}$. Further we note that these results for $z_0 = 0$ are also relevant if $|\lambda_0| \leq L$ with a fixed $L > 0$, corresponding to the case where F_0 is genuinely non-stiff. Then $|z_0| \leq \Delta t\, L$, and with either condition (2.4) or (2.5) imposed on $z_1, \ldots, z_s$ we then get $|r(z_0, z_1, \ldots, z_s)| \leq 1 + \Delta t\, K$ with a $K > 0$, giving stability on finite time intervals $[0, T]$.

Of course, for $z_0 = 0$ there is no difference between the original Douglas method and the modifications (1.4) and (1.5). For these modifications one would hope to get stability under the restriction $|1 + z_0 + \frac{1}{2}z_0^2| \leq 1$, but we will see that in general the stronger condition $|1 + z_0| \leq 1$ is needed.

The following results provide stability under rather strict conditions. It will be shown afterwards that these strict conditions are necessary, that is, the results are sharp.

Theorem 2.1 *The stability condition (2.3) is valid if either*

$$s = 1, \quad |1 + z_0| \leq 1, \quad z_1 \in \mathcal{W}_{\pi/2}, \quad (2.6)$$

or

$$s \geq 2, \quad |1 + z_0| \leq 1, \quad z_j \in \mathcal{W}_0 \quad (j = 1, 2, \ldots, s). \quad (2.7)$$

Proof. For $s = 1$ we have

$$r(z_0, z_1) = \frac{\left(1 + z_0 + \frac{1}{2}z_0^2\right) + (1 + z_0)\frac{1}{2}z_1}{1 - \frac{1}{2}z_1}.$$

According to the maximum modulus theorem it is sufficient to consider z_0 on the boundary of the set where $|1 + z_0| \leq 1$. Therefore, put $1 + z_0 = e^{i\phi}$ with $i = \sqrt{-1}$. Then $1 + z_0 + \frac{1}{2}z_0^2 = \frac{1}{2}(1 + e^{2i\phi})$. Further we can take z_1 on the boundary of $\mathcal{W}_{\pi/2}$, that is, $z_1 = it$ for some $t \in \mathbb{R}$. For such z_0, z_1 we have

$$|r(z_0, z_1)|^2 = \frac{\frac{1}{4}|1 + e^{2i\phi}|^2 + \frac{1}{2}\mathrm{Re}\left(i(1 + e^{-2i\phi})e^{i\phi}\right)t + \frac{1}{4}|e^{i\phi}|^2 t^2}{1 + \frac{1}{4}t^2}.$$

Since the middle term in the numerator vanishes, this is seen to be bounded by one.

Next, consider $s \geq 2$. If $z_s \to -\infty$, then $r(z_0, z_1, \ldots, z_{s-1}, z_s)$ converges to

$$r_*(z_0, z_1, \ldots, z_{s-1}) = 1 - 2\frac{\left(1 + \frac{1}{2}z_0\right)}{\prod_{j=1}^{s-1}\left(1 - \frac{1}{2}z_j\right)}. \quad (2.8)$$

Furthermore, with $\theta_s = (1 - \frac{1}{2}z_s)^{-1}$, we have

$$r(z_0, \ldots, z_{s-1}, z_s) = \theta_s\, r(z_0, \ldots, z_{s-1}) + (1 - \theta_s)\, r_*(z_0, \ldots, z_{s-1}), \quad (2.9)$$

4

and if $z_s \leq 0$ (i.e. $z_s \in \mathcal{W}_0$) then $\theta_s \in [0, 1]$. Hence the maximum of $|r|$ for $z_s \leq 0$ is attained in either $z_s = 0$ or $z_s = -\infty$:

$$\max_{z_s \leq 0} |r(z_0, \ldots, z_{s-1}, z_s)| = \max\left(|r(z_0, \ldots, z_{s-1})|, |r_*(z_0, \ldots, z_{s-1})|\right). \qquad (2.10)$$

If $|1 + z_0| \leq 1$ and $z_1, \ldots, z_{s-1} \leq 0$, then $\theta = \prod_{j=1}^{s-1}(1 - \frac{1}{2}z_j)^{-1} \in [0, 1]$, and therefore

$$|r_*(z_0, \ldots, z_{s-1})| = |(1 - \theta) - \theta(1 + z_0)| \leq (1 - \theta) + \theta|1 + z_0| \leq 1.$$

Sufficiency of (2.7) now follows easily by induction with respect to s. $\qquad\square$

With respect to sharpness of the above results: first note that if $s = 1$, we get, as limit value for $z_1 \to -\infty$,

$$r_*(z_0) = -(1 + z_0).$$

This shows that even if $s = 1$, then the condition $|1 + z_0| \leq 1$ is already necessary for stability with arbitrary $z_1 \in \mathcal{W}_0$. For $s = 2$ we get

$$r_*(z_0, z_1) = -\frac{(1 + z_0) + \frac{1}{2}z_1}{1 - \frac{1}{2}z_1},$$

from which it is easily seen that in order to have $|r(z_0, z_1, z_2)| \leq 1$ for arbitrary values $|1 + z_0| \leq 1$, $z_1 \in \mathcal{W}_\alpha$ and $z_2 \in \mathcal{W}_0$, we need $\alpha = 0$.

Remark 2.2 If $s = 1$, the above results are also relevant for the Douglas method (1.3) with $\theta = \frac{1}{2}$ because the stability function of that method equals $-r_*(z_0, z_1)$. The stability result (2.6) in Theorem 2.1 for the modified Douglas methods (1.4), (1.5) is more favourable than for the method (1.3), because $|r_*(z_0, z_1)| \leq 1$ for all $|1 + z_0| \leq 1$ will require that $z_1 \in \mathcal{W}_0$, instead of $z_1 \in \mathcal{W}_{\pi/2}$. $\qquad\diamond$

2.2 Stability for PDEs

The above results for the scalar problems are directly applicable to linear systems with $F_j(u) = A_j u$ and scaled Euclidean norms, provided the matrices A_j are normal and commuting. This seems a very strict assumption, but it is common in the analysis of numerical methods for nonlinear PDEs to use linearization, neglect boundary conditions and freeze the coefficients in a so-called von Neumann analysis [17]. The λ_j in the scalar test equation then stand for eigenvalues of the linearized operators $A_j = \partial F_j(t, u)/\partial u$, and due to the freezing of coefficients and absence of boundary conditions these matrices A_j can be assumed to be normal and commuting in the von Neumann analysis.

Further we note that the above stability conditions have been derived for the case that the $z_j = \Delta t \lambda_j$ can be independent from each other. For convection-diffusion-reaction equations $u_t + \nabla f(u) = \Delta u + g(u)$, with stiff reaction term $g(u)$, it might seem attractive to treat the convective term explicitly, say $F_0(u) \approx \nabla f(u)$, $F_1(u) \approx \Delta u$ and $F_2(u) \approx g(u)$, and after linearization z_0 and z_1 are then related as being eigenvalues for the same Fourier modes. However, also for this case with related z_0, z_1, very strict stability conditions may be needed.

As an example, consider the 1D model problem $u_t + au_x = du_{xx} + cu$, where c is arbitrary negative, representing a stiff reaction term, and the advection and diffusion terms are discretized with second-order central differences on a uniform grid with mesh-width $\Delta x = h$. Then with Courant number $\nu = \Delta t\, a/h$ and cell Péclet number

$\mu = ah/d$, the eigenvalues are $z_0 = i\nu\sin(2\phi)$, $z_1 = -4(\nu/\mu)\sin^2(\phi)$ with $i = \sqrt{-1}$, $\phi \in [0, 2\pi]$, and $z_2 \le 0$ is arbitrary. It follows by some straightforward (but somewhat tedious) calculations that $\max_\phi |r_*(z_0, z_1)| \le 1$ iff $\nu \le 2/\mu$ while the condition $\max_\phi |r(z_0, z_1)| \le 1$ leads to a severe restriction on the Courant number ν for small Péclet numbers μ (asymptotically $\nu^3 \le 4\mu(1 + o(\mu))$ for $\mu \to 0$). The restrictions for stability are displayed in Figure 1, found by direct numerical computations.
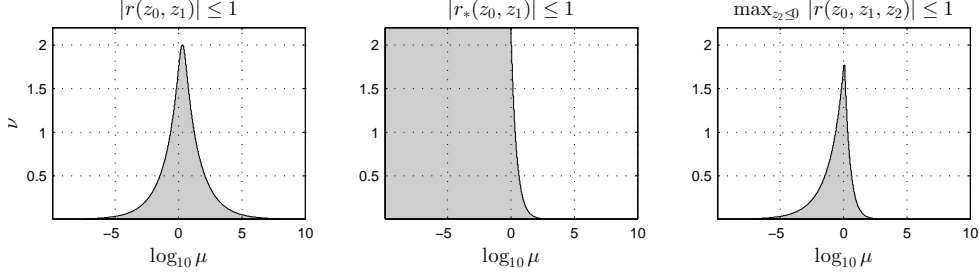

Figure 1: Domains for stability (gray) in the (μ, ν)-plane, with cell Péclet number $\mu = ah/d$ and Courant number $\nu = \Delta ta/h$, for the modified Douglas schemes with second-order central discretizations.

If the advection term is discretized by first-order upwind, then we get eigenvalues $z_0 = -\nu\big(1 - \cos(2\phi)\big) + i\nu\sin(2\phi)$ which lie on a circle in $\mathbb{C}^-$, with radius ν, touching the imaginary axis, and therefore $|1 + z_0| \le 1$ if $\nu \le 1$. According to Theorem 2.1 this gives stability under the CFL restriction $\nu \le 1$. However, since the goal of the modified Douglas method is to achieve temporal order two, having only order one in space is not that interesting.

It seems from the above that advection cannot be taken explicitly with second-order central spatial discretizations. However, surprisingly, it can be done if a part of the diffusion term is added to F_0. For example, in 1D, let the discrete advection and diffusion operators be given, in stencil notation, by

$$S_a = \frac{a}{2h}\big[1,\ 0,\ -1\big], \qquad S_d = \frac{d}{h^2}\big[1,\ -2,\ 1\big]. \tag{2.11}$$

Then with $\mu = ah/d$ we can take

$$A_0 = S_a + \tfrac{1}{2}\mu S_d = \frac{a}{h}\big[1,\ -1,\ 0\big], \qquad A_1 = \big(1 - \tfrac{1}{2}\mu\big)S_d, \tag{2.12}$$

provided the cell Péclet number satisfies $\mu \le 2$. With this h-dependent splitting we now get A_0 as for first-order upwind, so stability will hold under the CFL restriction $\nu = \Delta t\, a/h \le 1$. In more spatial dimensions a similar splitting can be used.

In conclusion: even though some convection is permitted, the methods (1.4) and (1.5) with $s \ge 2$ seem best suited for classes of problems where all implicit terms are known a priori to have eigenvalues close to the real, negative axis. There are interesting problems of this type originating from systems of reaction-diffusion equations $u_t = \nabla(D(u)\nabla u) + g(u)$ with diagonal matrix D and with a nonlinear term g which is either non-stiff or such that the dominant eigenvalues of $g'(u)$ are real negative. For such problems the methods are expected to be stable.

3 Analysis of discretization errors

In this section we will analyze the discretization errors for the modified Douglas methods (1.4), (1.5) in a norm $\|\cdot\|$ on $\mathbb{R}^m$, with corresponding induced matrix norm on

6

$\mathbb{R}^{m \times m}$. The approach will be similar to [17], where the errors of the method (1.3) have been studied.

We will use the notation $\mathcal{O}(\Delta t^k)$ for a vector or matrix whose norm is bounded by $C\Delta t^k$ for small $\Delta t > 0$, with a constant C that does not depend on the mesh-width h from the spatial discretization. The discretization errors will be expressed in terms of the exact solution $u(t)$ of problem (1.1) and

$$\varphi_j(t) = F_j(t, u(t)).\tag{3.1}$$

The derivatives of $u(t)$ and $\varphi_j(t)$ appearing in the analysis are assumed to be $\mathcal{O}(1)$ for $t \in [0, T]$, that is, bounded uniformly in the mesh-width h.

The convergence properties of the methods will be studied for linear inhomogeneous systems $u'(t) = Au(t) + g(t)$ with

$$F_j(t, v) = A_j v + g_j(t) \qquad (j = 0, 1, \ldots, s).\tag{3.2}$$

If A_j is a discretized differential operator, with negative powers of the mesh-width h involved, then inhomogeneous boundary values pertinent to A_j are supposed to be contained in the source term $g_j(t)$.

In this section it will be assumed that

$$\left\|\Delta t A_0\right\| \le L < 2, \qquad \left\|(I - \tfrac{1}{2}\Delta t A_j)^{-1}\right\| \le M \quad (j = 1, 2, \ldots, s),\tag{3.3}$$

with a moderate constant $M > 0$ (independent of h and Δt). Further we will use the following notation,

$$Z_j = \Delta t A_j, \quad Z = Z_0 + Z_1 + \cdots + Z_s, \quad P_0 = I + \tfrac{1}{2}Z_0, \quad Q_j = I - \tfrac{1}{2}Z_j.$$

3.1 Internal perturbations for method (1.4)

Consider along with (1.4) the scheme with perturbations on each of the stages

$$\begin{cases} \tilde{v}_* = \tilde{u}_{n-1} + \Delta t\, F(t_{n-1}, \tilde{u}_{n-1}) + \sigma_0, \\ \tilde{v}_0 = \tilde{v}_* + \tfrac{1}{2}\Delta t\big(F_0(t_n, \tilde{v}_*) - F_0(t_{n-1}, \tilde{u}_{n-1})\big) + \rho_0, \\ \tilde{v}_j = \tilde{v}_{j-1} + \tfrac{1}{2}\Delta t\big(F_j(t_n, \tilde{v}_j) - F_j(t_{n-1}, \tilde{u}_{n-1})\big) + \rho_j \qquad (j = 1, 2, \ldots, s), \\ \tilde{u}_n = \tilde{v}_s. \end{cases}\tag{3.4}$$

Let $e_n = \tilde{u}_n - u_n$ and $w_j = \tilde{v}_j - v_j - e_{n-1}$, $w_* = \tilde{v}_* - v_* - e_{n-1}$. Subtraction of (1.4) from (3.4) then gives for the linear problems (3.2) the relations

$$w_* = Z e_{n-1} + \sigma_0, \quad w_0 = P_0 w_* + \rho_0, \quad w_j = Q_j^{-1}(w_{j-1} + \rho_j), \quad e_n = e_{n-1} + w_s.$$

Elimination of the internal quantities w_*, w_j leads to the error recursion

$$e_n = R e_{n-1} + d_n,\tag{3.5}$$

with stability matrix

$$R = I + Q_s^{-1} \cdots Q_2^{-1} Q_1^{-1} P_0 Z\tag{3.6}$$

and

$$d_n = Q_s^{-1} \cdots Q_2^{-1} Q_1^{-1}\big(P_0 \sigma_0 + \rho_0\big) + \sum_{j=1}^{s} Q_s^{-1} \cdots Q_j^{-1} \rho_j.\tag{3.7}$$

Under (3.3) the norm of d_n can be bounded in terms of the norms of σ_0 and the ρ_j. To have a favourable error propagation for the whole process it will also be assumed that

$$\|R^n\| \le K \qquad \text{(for all } n \ge 1\text{)}, \tag{3.8}$$

with some moderate constant $K > 0$ (independent of Δt and spatial mesh-width h). This will hold with $K = 1$ in the discrete L_2-norm if the matrices Z_j are normal and commuting, and the eigenvalues z_j satisfy the linear stability assumptions for the scalar problem (2.1).

3.2 Local discretization errors for method (1.4)

For the convergence analysis we will use the perturbed method (3.4) with $\tilde{u}_n = u(t_n)$. Then the e_n in (3.5) are the global discretization errors

$$e_n = u(t_n) - u_n. \tag{3.9}$$

Apparently, the term d_n in recursion (3.5) is then the local discretization error, introduced in the step from t_{n-1} to t_n. To obtain expressions for these local errors, it is convenient to take $\tilde{v}_* = \tilde{v}_j = u(t_n)$. Then the residuals are

$$\sigma_0 = \tfrac{1}{2}\Delta t^2 u''(t_n) + \mathcal{O}(\Delta t^3),$$
$$\rho_j = \tfrac{1}{2}\Delta t\big(\varphi_j(t_{n-1}) - \varphi_j(t_n)\big) = -\tfrac{1}{2}\Delta t^2 \varphi_j'(t_n) + \mathcal{O}(\Delta t^3) \tag{3.10}$$

for $j = 0, 1, \ldots, s$. Since $u''(t) = \sum_{j=0}^s \varphi_j'(t)$, this gives the following expression for the local errors:

$$d_n = \tfrac{1}{2}\Delta t^2\, Q_s^{-1}\cdots Q_1^{-1}(P_0 - I)\big(\varphi_0'(t_n) + \varphi_1'(t_n)\big)$$
$$+ \tfrac{1}{2}\Delta t^2\, Q_s^{-1}\cdots Q_1^{-1} \sum_{j=2}^s (P_0 - Q_1\cdots Q_{j-1})\varphi_j'(t_n) + \mathcal{O}(\Delta t^3). \tag{3.11}$$

It is seen from the derivation that the $\mathcal{O}(\Delta t^3)$ remainder term can be bounded with a constant determined by the maximum of $|\varphi_j''(t)|$ for $j = 0, 1, \ldots, s$ and $t \in [t_{n-1}, t_n]$.

For non-stiff problems, with $Z_j = \mathcal{O}(\Delta t)$, we will have $d_n = \mathcal{O}(\Delta t^3)$. However, for stiff systems, and in particular for semi-discrete systems derived from PDEs, we may get a lower order. In spite of this, the method will often be second-order convergent due to damping and cancellation effects, as we will see next.

3.3 Global discretization errors for method (1.4)

According to the general criterion formulated in [15], we will have second-order convergence if the local errors d_n can be decomposed as

$$d_n = (R-I)\xi_n + \eta_n \quad \text{with} \quad \xi_n = \mathcal{O}(\Delta t^2),\ \eta_n = \mathcal{O}(\Delta t^3),\ \xi_n - \xi_{n-1} = \mathcal{O}(\Delta t^3), \tag{3.12}$$

and this decomposition was also shown to be necessary in case the d_n are constant, that is, independent of n. We will discuss in detail whether (3.12) will be fulfilled. For simplicity it will be assumed that Z is invertible, although that is not strictly needed (cf. Remark 3.3 below).

To apply criterion (3.12), the local errors will be written in the form

$$d_n = (R-I)\zeta_n + \mathcal{O}(\Delta t^3). \tag{3.13}$$

8

These ζ_n are then natural candidates for ξ_n in (3.12). Furthermore, to express the errors in a compact way, we introduce

$$S_j = 2\left(I - Q_1 Q_2 \cdots Q_j\right) \qquad (j = 1, 2, \ldots, s). \qquad (3.14)$$

Then $S_1 = Z_1$, $S_2 = Z_1 + Z_2 - \frac{1}{2}Z_1 Z_2$ and for $2 < j \leq s$ we have

$$S_j = \sum_{1 \leq i \leq j} Z_i - \sum_{1 \leq i_1 < i_2 \leq j} \tfrac{1}{2} Z_{i_1} Z_{i_2} + \sum_{1 \leq i_1 < i_2 < i_3 \leq j} \tfrac{1}{4} Z_{i_1} Z_{i_2} Z_{i_3} + \cdots + \left(\tfrac{-1}{2}\right)^{j-1} Z_1 Z_2 \cdots Z_j.$$

For the modified Douglas method (1.4) we obtain from (3.11)

$$\begin{aligned} d_n = {} & \tfrac{1}{2}\Delta t^2 (R - I)Z^{-1}P_0^{-1}\Big[(P_0 - I)\big(\varphi_0'(t_n) + \varphi_1'(t_n)\big) \\ & + \sum_{j=2}^{s}(P_0 - Q_1 Q_2 \cdots Q_{j-1})\varphi_j'(t_n)\Big] + \mathcal{O}(\Delta t^3). \end{aligned}$$

Using $P_0 - Q_1 \cdots Q_j = \frac{1}{2}(Z_0 + S_j)$, this gives the expression (3.13) with

$$\zeta_n = \tfrac{1}{4}\Delta t^2 Z^{-1}P_0^{-1}\Big[Z_0 u''(t_n) + \sum_{j=2}^{s} S_{j-1}\varphi_j'(t_n)\Big]. \qquad (3.15)$$

To find conditions so that $\zeta_n = \mathcal{O}(\Delta t^2)$, we assume for convenience that Z_0 commutes with the other Z_j. Then the factor P_0^{-1} can be moved to the front, and according to (3.3) we have $\|P_0^{-1}\| \leq (1 - \frac{1}{2}L)^{-1}$. Furthermore, we will have $Z^{-1}S_{j-1}v(t) = \mathcal{O}(1)$ if all products $Z^{-1}Z_{i_1} \cdots Z_{i_k}v(t)$ are $\mathcal{O}(1)$ for $1 \leq i_1 < \cdots < i_k < j$. Consequently, criterion (3.12) can be applied with $\xi_n = \zeta_n$ under the following conditions:

$$A^{-1}A_0 v(t) = \mathcal{O}(1) \qquad \text{for } v = u'', u''' \text{ and } t \in [0, T], \qquad (3.16a)$$

$$\Delta t^{k-1}\, A^{-1}A_{i_1}A_{i_2} \cdots A_{i_k}v(t) = \mathcal{O}(1) \qquad \begin{array}{l} \text{for } v = \varphi_j', \varphi_j'',\ t \in [0, T] \text{ and} \\ 1 \leq i_1 < i_2 < \cdots < i_k < j \leq s. \end{array} \qquad (3.16b)$$

In summary, we have obtained the following convergence result.

Theorem 3.1 *Consider linear problems (3.2) with $\varphi_j'(t), \varphi_j''(t) = \mathcal{O}(1)$ for $t \in [0, T]$, $j = 0, 1, \ldots, s$. Assume (3.3) and (3.8) hold, $A_0 A_j = A_j A_0$ for $j = 1, 2, \ldots, s$, and the conditions (3.16) are satisfied. Then method (1.4) will be convergent of order two on the interval $[0, T]$.*

If $s = 2$, this shows convergence with order two under the condition (3.16a) and $A^{-1}A_1 v(t) = \mathcal{O}(1)$ for $v = \varphi_2', \varphi_2''$, $t \in [0, T]$. For $s = 3$ we get the additional conditions $A^{-1}A_1 v(t) = \mathcal{O}(1)$, $A^{-1}A_2 v(t) = \mathcal{O}(1)$ and $\Delta t\, A^{-1}A_1 A_2 v(t) = \mathcal{O}(1)$ for $v = \varphi_3', \varphi_3''$, $t \in [0, T]$. These results for $s \leq 3$ are the same as in [17, Thm. 2.1, 2.2] for the case $F_0 = 0$; see also [18, pp. 380, 381]. So, in these theoretical results, inclusion of an explicit term as in the modified Douglas method (1.4) does not lead to conditions for second-order convergence that are more strict than for $F_0 = 0$.

The assumption that A_0 commutes with the other A_j has been made only to get more simple conditions. Note that if $A_0 = \mathcal{O}(1)$, then $P_0 = I + \mathcal{O}(\Delta t)$, in which case the conditions (3.16) again ensure second-order convergence.

Example 3.2 The validity of these convergence conditions has been discussed in [17, 18] for the 3D heat equation in a cube with inhomogeneous Dirichlet boundary conditions, using dimension splitting for standard finite differences with uniform mesh-width h in all directions, and the discrete L_2-norm $\|v\|_2 = (\frac{1}{m} \sum_{i=1}^{m} |v_i|^2)^{1/2}$. Then $\|A^{-1}\|_2 \leq 1$, $\|A^{-1}A_j\|_2 \leq 1$. Furthermore, in this norm, a condition $\Delta t A^{-1} A_1 A_2 v = \mathcal{O}(1)$ will hold if v is the restriction to the grid of a smooth function $w(x)$ which is zero at the boundaries pertinent to either A_1 or A_2. If w does not satisfy such a homogeneous boundary condition, we will still have $\Delta t A^{-1} A_1 A_2 v = \mathcal{O}(|\log(h)|)$ if $\Delta t \sim h$ and $\Delta t A^{-1} A_1 A_2 v = \mathcal{O}(1)$ if $\Delta t \sim h^{1+\varepsilon}$ for some $\varepsilon > 0$. So in this case convergence with order two is ensured under a mild time step restriction.

For larger s additional conditions will appear. For example, for $s = 4$ the essential new condition will be $\Delta t^2 A^{-1} A_1 A_2 A_3 v(t) = \mathcal{O}(1)$ for $v = \varphi_4', \varphi_4''$, which may lead to additional time step constraints, depending on the boundary values assumed by the grid function φ_4. $\diamond$

Remark 3.3 In the above it has been assumed that $Z = \Delta t A$ is invertible, but this is not really needed. The above conditions (3.16) are all of the form $A^{-1}w = \mathcal{O}(1)$. Instead, criterion (3.12) can already be applied if there is a $v \in \mathbb{R}^m$ such that $Av = w$ and $v = \mathcal{O}(1)$. This may hold even if A is singular, as it will be for example for parabolic problems with Neumann conditions on the whole boundary. $\diamond$

3.4 Discretization errors for method (1.5)

In the same way, we can derive error recursions for method (1.5). As perturbed scheme we now consider

$$
\begin{cases}
\tilde{v}_0 = \tilde{u}_{n-1} + \Delta t F(t_{n-1}, \tilde{u}_{n-1}) + \sigma_0, \\
\tilde{v}_j = \tilde{v}_{j-1} + \frac{1}{2}\Delta t \big(F_j(t_n, \tilde{v}_j) - F_j(t_{n-1}, \tilde{u}_{n-1})\big) + \rho_j \qquad (j = 1, 2, \ldots, s), \quad (3.17) \\
\tilde{u}_n = \tilde{v}_s + \frac{1}{2}\Delta t \big(F_0(t_n, \tilde{v}_s) - F_0(t_{n-1}, \tilde{u}_{n-1})\big) + \rho_0.
\end{cases}
$$

With $e_n = \tilde{u}_n - u_n$ and $w_j = \tilde{v}_j - v_j - e_{n-1}$, subtraction of (1.5) from (3.17) gives for the linear problems (3.2) the relations

$$
w_0 = Ze_{n-1} + \sigma_0, \qquad w_j = Q_j^{-1}(w_{j-1} + \rho_j), \qquad e_n = e_{n-1} + P_0 w_s + \rho_0.
$$

After elimination of the internal quantities w_j we obtain the error recursion (3.5) with

$$
R = I + P_0 Q_s^{-1} \cdots Q_2^{-1} Q_1^{-1} Z \tag{3.18}
$$

and

$$
d_n = P_0 Q_s^{-1} \cdots Q_1^{-1}(\sigma_0 + \rho_1) + P_0 Q_s^{-1} \cdots Q_2^{-1}\rho_2 + \cdots + P_0 Q_s^{-1}\rho_s + \rho_0. \tag{3.19}
$$

If $\tilde{u}_n = \tilde{v}_j = u(t_n)$, the residuals are the same as in (3.10). Using the fact that $u''(t) = \sum_{j=0}^{s} \varphi_j'(t)$, this gives the following expression for the local errors:

$$
d_n = \frac{1}{2}\Delta t^2 (P_0 Q_s^{-1} \cdots Q_1^{-1} - I)\varphi_0'(t_n)
$$
$$
+ \frac{1}{2}\Delta t^2 P_0 Q_s^{-1} \cdots Q_1^{-1} \sum_{j=2}^{s}(I - Q_1 \ldots Q_{j-1})\varphi_j'(t_n) + \mathcal{O}(\Delta t^3). \tag{3.20}
$$

By some calculations, it is seen that this local error fits in the form (3.13) with

$$
\zeta_n = \frac{1}{4}\Delta t^2 Z^{-1} \Big[(Z_0 + S_s)P_0^{-1}\varphi_0'(t_n) + \sum_{j=2}^{s} S_{j-1}\varphi_j'(t_n) \Big]. \tag{3.21}
$$

10

So the main difference between these local errors and the ones for the method (1.4) is the factor multiplying φ_0' in formula (3.21). As we will see, this factor will lead to complications in proving convergence with order two if $s \geq 2$.

First consider $s = 1$. Then we have the simple expression $\zeta_n = \frac{1}{4}\Delta t^2 P_0^{-1} \varphi_0'(t_n)$. Taking $\xi_n = \zeta_n$ thus shows convergence of order two under the sole condition

$$\varphi_0'(t), \varphi_0''(t) = \mathcal{O}(1) \qquad \text{for } t \in [0, T]. \tag{3.22}$$

However, theoretical problems arise for $s = 2$. Then we will have a local error with $\zeta_n = -\frac{1}{8}\Delta t^2 Z^{-1} Z_1 Z_2 P_0^{-1} \varphi_0'(t_n) + \mathcal{O}(\Delta t^2)$, but in some experiments convergence with order two in the L_2-norm was observed even if this $\zeta_n \neq \mathcal{O}(\Delta t^2)$. To prove this convergence, a decomposition (3.12) with $\xi_n \neq \zeta_n$ will be needed. As mentioned already, criterion (3.12) was shown to be necessary for constant d_n in [15, Lemma 5.2]. For that proof the following decomposition was used:

$$\xi_n = (\lambda R - I)^{-1}(R - I)\zeta_n, \qquad \eta_n = (\lambda - 1)R\xi_n + \mathcal{O}(\Delta t^3), \tag{3.23}$$

with $\lambda = 1/(1+\Delta t)$. In the experiments, this ξ_n was indeed seen to be $\mathcal{O}(\Delta t^2)$. However, because of the inverse of $\lambda R - I$, this term ξ_n has a rather complicated structure, and it is not clear how to prove that it is $\mathcal{O}(\Delta t^2)$ under transparent, realistic conditions.

In conclusion, the variant Douglas method with $s \geq 2$ was observed to be convergent in the L_2-norm of order two for many problems, but simple conditions for such convergence have not been found. For that, instead of (3.23) a more tractable decomposition might be needed.

Remark 3.4 To prove the necessity of (3.12), it was assumed in [15] that $\|R\| \leq 1$. With our stability assumption (3.8), where $K \geq 1$, this is achieved with a new norm $\|v\|_* = \sup_{n \geq 0} \|R^n v\|$ for $v \in \mathbb{R}^m$. This new norm is equivalent to the old one, $\|v\| \leq \|v\|_* \leq K\|v\|$ for all $v \in \mathbb{R}^m$, and we have $\|R\|_* \leq 1$. $\diamond$

4 Results for 2D model problems with dimension splitting

4.1 Convergence in the discrete L_2-norm

Numerical comparisons of the modified Douglas methods will be presented in Section 5 for some interesting reaction-diffusion problems. To illustrate the error build-up, we first show some results for the Douglas methods applied to a model problem consisting of the inhomogeneous heat equation

$$u_t = u_{xx} + u_{yy} + f(x, y, t) \tag{4.1}$$

with prescribed exact solution

$$u(x, y, t) = \sin(t)\left((1 + 2x^2)(1 + y^2) - 1\right) \tag{4.2}$$

on the unit square $(x, y) \in \Omega = [0, 1]^2$ with time $t \in [0, 1]$ and Dirichlet boundary conditions. For this model problem, the source term f as well as the initial and boundary values are fitted to the exact solution. Furthermore we use dimension splitting with standard second-order differences in space and mesh-width $\Delta x = \Delta y = h$. In this example, the source term is put into F_0 and F_1, F_2 contain the discretized difference operators in the x- and y-direction, respectively, together with the relevant boundary values.

Because the solution is chosen to be a polynomial of degree less than four in both x and y, there will be no spatial errors. So the errors given in the following tables are just temporal errors.

In Table 1 the local errors $\|e_1\|_2 = \|d_1\|_2$ of the first time step are given in the discrete L_2-norm for various step-sizes Δt with the ratio $\Delta t/h$ held fixed. For these local errors we see an $\mathcal{O}(\Delta t^2)$ behaviour with all three schemes. For the original Douglas method (1.3) this is in line with it being a first-order method, with an error proportional to Δt^2 after one step. However, for the two modified methods (1.4) and (1.5) we see a similar behaviour. This is due to order reduction; on a fixed spatial grid the local errors are proportional to Δt^3, even though very small time steps will be needed to see that asymptotic behaviour.

Table 1: Local errors $\|e_1\|_2 = \|d_1\|_2$, at time $t_1 = \Delta t$, for the heat equation (4.1), (4.2), with the Douglas schemes, $\Delta t/h = 1$.

$1/\Delta t$	50	100	200	400
Meth. (1.3)	$1.31 \cdot 10^{-3}$	$3.58 \cdot 10^{-4}$	$9.54 \cdot 10^{-5}$	$2.49 \cdot 10^{-5}$
Meth. (1.4)	$2.14 \cdot 10^{-4}$	$4.91 \cdot 10^{-5}$	$1.10 \cdot 10^{-5}$	$2.40 \cdot 10^{-6}$
Meth. (1.5)	$6.70 \cdot 10^{-4}$	$1.42 \cdot 10^{-4}$	$3.02 \cdot 10^{-5}$	$6.40 \cdot 10^{-6}$

The global errors $\|e_N\|_2$ in the discrete L_2-norm at the output time $t_N = 1$ are shown in Table 2. Here the Douglas method (1.3) is seen to be convergent with order one, whereas for the modified methods (1.4) and (1.5) we get convergence with order two. For method (1.4) this is in agreement with the result of Theorem 3.1.

For method (1.5) simple convergence criteria are lacking, but in the table an order two convergence behaviour is observed. In the present test, the ratio $\Delta t/h$ is held fixed, and then it can be shown (cf. Example 3.2) that $\zeta_n = \mathcal{O}(|\log(h)|\Delta t^2)$, which shows convergence with 'practically' order two. If Δt would tend to zero much slower than h, say $\Delta t \sim \sqrt{h}$, a more complicated error decomposition, such as (3.23), would be needed to demonstrate the order two convergence.

Finally we mention that in this test the error constants are smaller for the modification (1.4) than for the variant (1.5). It will be seen in Section 5, with more interesting test problems, that this is not always the case.

4.2 Convergence in the maximum norm

The global errors for the model problem (4.1), (4.2) in the maximum norm $\|\cdot\|_\infty$ are given in Table 3. As seen from this table, there is now a clear distinction between the two methods (1.4) and (1.5). Method (1.4) retains its second-order convergence, but the order of convergence for (1.5) drops to one. In fact, the errors for the latter method are now even larger than for the original Douglas method (1.3).

To understand these results, let us first note that the local errors are $\mathcal{O}(\Delta t^2)$ for all the methods, just as in Table 1. Furthermore, the methods are stable in the maximum norm. For this model problem with $A_0 = 0$ and commuting A_1 and A_2, we have $R = R_1 R_2$ with matrices $R_j = (I - \frac{1}{2}Z_j)^{-1}(I + \frac{1}{2}Z_j)$ for $j = 1, 2$. It is known, for the one-dimensional heat equation, that the estimate $\sup_{n \geq 0} \|R_j^n\|_\infty \leq \kappa$ is valid with a modest value $\kappa > 0$, cf. [4, 8, 23]. Since R_1 and R_2 commute, it follows that the stability condition (3.8) will hold with constant $K = \kappa^2$.

The local errors for method (1.4) are given by (3.13) with $\zeta_n = \frac{1}{4}\Delta t^2 A^{-1} A_1 \varphi_2'(t_n)$, and $\varphi_2(t)$ is the restriction to the grid of $u_{yy}(x, y, t)$. In this test we observed that $\|A^{-1}A_1\|_\infty \sim |\log(h)|$. This leads directly to the global error bound $\|e_N\|_\infty = \mathcal{O}(|\log(h)|\Delta t^2)$, which is 'practically' order two for fixed ratios $\Delta t/h$.

Table 2: Global errors $\|e_N\|_2$ at final time $t_N = 1$ for the heat equation (4.1), (4.2), with the Douglas schemes, $\Delta t/h = 1$.

$1/\Delta t$	50	100	200	400
Meth. (1.3)	$2.52 \cdot 10^{-3}$	$1.22 \cdot 10^{-3}$	$6.04 \cdot 10^{-4}$	$3.00 \cdot 10^{-4}$
Meth. (1.4)	$1.21 \cdot 10^{-4}$	$3.04 \cdot 10^{-5}$	$7.64 \cdot 10^{-6}$	$1.91 \cdot 10^{-6}$
Meth. (1.5)	$6.63 \cdot 10^{-4}$	$1.60 \cdot 10^{-4}$	$3.90 \cdot 10^{-5}$	$9.60 \cdot 10^{-6}$

Table 3: Global errors $\|e_N\|_\infty$ at final time $t_N = 1$ for the heat equation (4.1), (4.2), with the Douglas schemes, $\Delta t/h = 1$.

$1/\Delta t$	50	100	200	400
Meth. (1.3)	$4.37 \cdot 10^{-3}$	$2.16 \cdot 10^{-3}$	$1.07 \cdot 10^{-3}$	$5.36 \cdot 10^{-4}$
Meth. (1.4)	$3.11 \cdot 10^{-4}$	$7.93 \cdot 10^{-5}$	$2.00 \cdot 10^{-5}$	$5.04 \cdot 10^{-6}$
Meth. (1.5)	$1.05 \cdot 10^{-2}$	$5.04 \cdot 10^{-3}$	$2.46 \cdot 10^{-3}$	$1.21 \cdot 10^{-3}$

To show that method (1.5) is only convergent with order one in the maximum norm, it is convenient to consider the Dirichlet problem for the heat equation (4.1) on the unit square with solution

$$u(x,y,t) = 1 - \tfrac{1}{2}tx^2 \,. \tag{4.3}$$

This leads to very simple expressions for the discretization errors, due to the fact that we now have $u_{tt} = u_{yy} = 0$ and $u_{xx} = -t$ is constant in space. In the following we consider again a uniform grid with $\Delta x = \Delta y = h$, with k points per direction, $h = 1/(k+1)$, giving a total number of grid points $m = k^2$.

We will use Kronecker products for vectors and matrices. Then the restriction to the grid of a product function $\phi(x)\psi(y)$ can be written as $b \otimes a \in \mathbb{R}^m$ with $a, b \in \mathbb{R}^k$, $a_i = \phi(ih)$ and $b_i = \psi(ih)$. Let $e = (1, 1, \ldots, 1)^T \in \mathbb{R}^k$. Since there is no spatial error, $\varphi_1(t)$ is the restriction to the grid of $u_{xx} = -t$ and $\varphi_2(t) = 0$, and it readily follows that the local errors d_n for method (1.5) are constant in n, with

$$d_n \equiv d = \tfrac{1}{2}\Delta t^2 \big(Q_2^{-1}Q_1^{-1} - I\big)v \,, \qquad v = e \otimes e \,. \tag{4.4}$$

Akin to the decomposition (3.23), let $\lambda = 1/(1 + \Delta t)$ and consider the vector

$$\xi = (\lambda R - I)^{-1}d \,, \tag{4.5}$$

where $R = I + Q_2^{-1}Q_1^{-1}Z$ according to formula (3.18) with $Z_0 = 0$. It is seen by some calculations that

$$\xi = \tfrac{1}{2}\Delta t^2(1 + \Delta t)B^{-1}w \quad \text{with} \quad B = -\Delta t\, Q_1 Q_2 + Z \,, \;\; w = (I - Q_1 Q_2)v \,. \tag{4.6}$$

If $D = \mathrm{tridiag}(1, -2, 1) \in \mathbb{R}^{k \times k}$ and $\mu = \tfrac{1}{2}\Delta t/h^2$, then $A_1 = h^{-2}I \otimes D$ and $A_2 = h^{-2}D \otimes I$ are the finite difference approximations to $\partial^2/\partial x^2$ and $\partial^2/\partial y^2$, respectively, and

$$Q_1 = I \otimes (I - \mu D) \,, \qquad Q_2 = (I - \mu D) \otimes I \,.$$

13

Let $\hat{e} = De$. Then $\hat{e} = -(e_1 + e_k)$, with e_1, e_k the first and last unit vectors in $\mathbb{R}^k$, and

$$w = \mu\, e \otimes \hat{e} + \mu\, \hat{e} \otimes e - \mu^2 \hat{e} \otimes \hat{e}\,.$$

The first component of w equals $w_1 = -2\mu - \mu^2$. Hence

$$\|w\|_\infty \geq \mu^2 + 2\mu\,.$$

In fact, a little inspection shows that equality holds. Further we have

$$\|B\|_\infty \leq \Delta t\|Q_1 Q_2\|_\infty + \|Z\|_\infty \leq \Delta t\|I - \mu D\|_\infty^2 + 4\mu\,\|D\|_\infty = \Delta t(1 + 4\mu)^2 + 16\mu\,.$$

Since $\|B^{-1}w\|_\infty \geq \|w\|_\infty / \|B\|_\infty$ it now follows that

$$\|\xi\|_\infty \;\geq\; \tfrac{1}{2}\Delta t^2 \frac{\mu^2 + 2\mu}{\Delta t(1 + 4\mu)^2 + 16\mu}\,. \tag{4.7}$$

For the case that $\Delta t = h$, we have $\mu = 1/(2\Delta t)$, which gives

$$\|\xi\|_\infty \;\geq\; \frac{\Delta t}{8((2 + T)^2 + 8)} \qquad \text{whenever} \quad \Delta t \leq T\,. \tag{4.8}$$

From the proof of [15, Lemma 5.2] it is seen that convergence of order p in the maximum norm implies $\|\xi\|_\infty = \mathcal{O}(\Delta t^p)$, and therefore we have $p \leq 1$. By stability and the local error bound $\|d_n\|_\infty = \mathcal{O}(\Delta t^2)$ we know the order is at least one, showing that the order of convergence equals $p = 1$ for $\Delta t = \Delta x = \Delta y = h$, $h \to 0$.

This result with order one convergence for solution (4.3) is in accordance with the errors in Table 3 for solution (4.2). For the more simple solution (4.3) the local error of method (1.4) is $d_n = \frac{1}{2}\Delta t^2 Q_2^{-1}(Q_1^{-1} - I)\varphi_2'(t_n) = 0$, so this method is now exact. This reveals a striking difference between the two methods (1.4) and (1.5).

5 Numerical comparisons for reaction-diffusion problems

5.1 Extended stabilizing correction methods

For some nonlinear reaction-diffusion problems, we will compare the modified Douglas methods with two extended methods that require per step (approximately) twice the computational work of (1.4) and (1.5).

The following method was derived by Hundsdorfer [17], based on a Rosenbrock-type method with matrix factorization from Verwer et al. [24]. With internal vectors v_j^* and v_j, the approximations u_n are computed from

$$\begin{cases} v_0^* &= u_{n-1} + \Delta t\, F(t_{n-1}, u_{n-1})\,, \\ v_j^* &= v_{j-1}^* + \theta\Delta t\big(F_j(t_n, v_j^*) - F_j(t_{n-1}, u_{n-1})\big) \qquad (j = 1, 2, \ldots, s)\,, \\ v_0 &= u_{n-1} + \tfrac{1}{2}\Delta t\big(F(t_{n-1}, u_{n-1}) + F(t_n, v_s^*)\big)\,, \\ v_j &= v_{j-1} + \theta\Delta t\big(F_j(t_n, v_j) - F_j(t_n, v_s^*)\big) \qquad (j = 1, 2, \ldots, s)\,, \\ u_n &= v_s\,. \end{cases} \tag{5.1}$$

The underlying implicit method is A-stable for any $\theta \geq \frac{1}{4}$, and for $\theta = 1 \pm \frac{1}{2}\sqrt{2}$ this implicit method has optimal damping at infinity (L-stability). Furthermore, method (5.1) is of order two in the ODE sense for any value of θ. Some stability and convergence results, pertinent to PDEs, can be found in [10, 11, 12, 13, 17].

A related method, due to in 't Hout & Welfert [12], is given by

$$
\begin{cases}
v_0^* & = \ u_{n-1} + \Delta t\, F(t_{n-1}, u_{n-1})\,, \\
v_j^* & = \ v_{j-1}^* + \theta \Delta t \big(F_j(t_n, v_j^*) - F_j(t_{n-1}, u_{n-1})\big) \qquad (j = 1, 2, \ldots, s)\,, \\
v_0 & = \ v_0^* + \tfrac{1}{2}\Delta t\big(F_0(t_n, v_s^*) - F_0(t_{n-1}, u_{n-1})\big) \\
& \qquad + \big(\tfrac{1}{2} - \theta\big)\Delta t \sum_{j=1}^{s} \big(F_j(t_n, v_s^*) - F_j(t_{n-1}, u_{n-1})\big)\,, \\
v_j & = \ v_{j-1} + \theta \Delta t\big(F_j(t_n, v_j) - F_j(t_{n-1}, u_{n-1})\big) \qquad (j = 1, 2, \ldots, s)\,, \\
u_n & = \ v_s\,.
\end{cases}
\tag{5.2}
$$

Here setting $\theta = \tfrac{1}{2}$ gives the method of Craig & Sneyd [5]. For any choice of θ, method (5.2) is of order two in the ODE sense, and the underlying implicit method is the same as for (5.1). Results on convergence for PDEs can be found in [14]. Stability results for parabolic equations with mixed derivatives were obtained in [9, 10, 12].

In the numerical tests the modified Douglas methods (1.4) and (1.5) will be compared with the methods (5.1) and (5.2) with parameter value $\theta = 1 - \tfrac{1}{2}\sqrt{2}$, and for (5.2) we will consider also $\theta = \tfrac{1}{2}$, the Craig-Sneyd method. In the comparisons the results for the Douglas method (1.3) with $\theta = \tfrac{1}{2}$ are also included, even though we know already that this method is only of order one if $F_0 \neq 0$. The methods (5.1) and (5.2) were also tested with parameter value $\theta = \tfrac{1}{4}$, but this produced errors very similar to $\theta = 1 - \tfrac{1}{2}\sqrt{2}$, with lines in the error plots practically on top of each other; for clarity those results are not included in the figures.

In the plots we will indicate the methods (5.1) and (5.2) with $\theta = 1 - \tfrac{1}{2}\sqrt{2}$ as HV and HW, respectively. The Craig-Sneyd method, (5.2) with $\theta = \tfrac{1}{2}$, is indicated as CS.

5.2 A traveling wave problem with dimension splitting

As a first test we consider a traveling wave problem in two spatial dimensions on the unit square $(x, y) \in \Omega = [0, 1]^2$. For this problem we use dimensional splitting with $A_1 \approx \partial^2/\partial x^2$, $A_2 \approx \partial^2/\partial y^2$, where the approximations are obtained with standard second-order finite differences on a uniform Cartesian grid with $\Delta x = \Delta y = h$. The nonlinear reaction term is treated explicitly. An exact solution is known, and time-dependent Dirichlet boundary conditions are fitted to this exact solution.

The equation is given by

$$
u_t = \epsilon(u_{xx} + u_{yy}) + \gamma u^2(1 - u)\,,
\tag{5.3a}
$$

with $\gamma, \epsilon > 0$, for which we have the traveling wave solution

$$
u(x, y, t) = \frac{1}{1 + \exp(\beta(r - r_0))}\,, \qquad r = \cos(\alpha)x + \sin(\alpha)y - c\,t\,,
\tag{5.3b}
$$

with constant $\beta = \tfrac{1}{2}\sqrt{2\gamma/\epsilon}$ and wave velocity $c = \sqrt{\gamma\epsilon/2}$. The angle α and shift parameter r_0 are arbitrary; in the test we will take $\alpha = \tfrac{1}{6}\pi$ and $r_0 = 1 - c$. Further we take $\gamma = 50$, and for ϵ two values are considered: $\epsilon = 1$ and $\epsilon = \tfrac{1}{50}$. For the larger value of ϵ the solution is very smooth, and the time-dependent boundary conditions then influence the errors of the schemes. For $\epsilon = \tfrac{1}{50}$ the largest errors are found, for all schemes, in the interior of the domain, showing that the boundary conditions are then no longer a dominating error source.

A similar problem was used in [17] for tests with the Douglas method (1.3) where the nonlinear reaction term was treated implicitly. Since this reaction term is not very
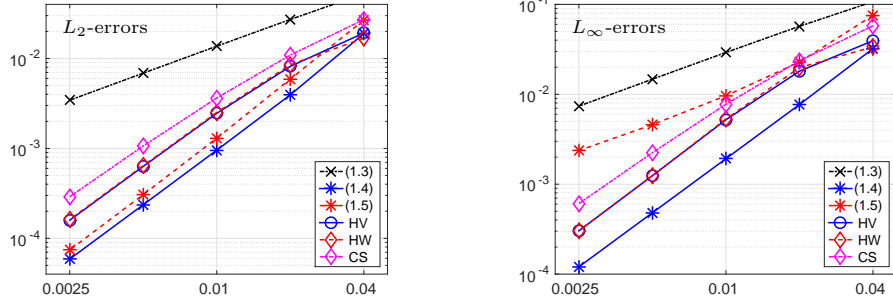
Figure 2: L_2-errors [left] and L_∞-errors [right] versus mesh-width $h = \frac{1}{25}, \frac{1}{50}, \ldots, \frac{1}{400}$, for problem (5.3) with $\epsilon = 1$, output time $T = 1$. Time step $\Delta t = h$ for the Douglas methods (1.3), (1.4), (1.5), and $\Delta t = 2h$ for the extended methods HV [i.e. (5.1) with $\theta = 1 - \frac{1}{2}\sqrt{2}$], HW [i.e. (5.2) with $\theta = 1 - \frac{1}{2}\sqrt{2}$] and CS [i.e. (5.2) with $\theta = \frac{1}{2}$].

stiff, it is more natural to treat it in an explicit fashion, as will be done here. Method (1.3) will then be first-order convergent only.

The errors at the output time $T = 1$ in the discrete L_2-norm and maximum norm are presented in the plots as function of the mesh-width h. The time step is taken as $\Delta t = h$ for the Douglas methods (1.3), (1.4), (1.5), and $\Delta t = 2h$ for the extended methods (5.1) and (5.2). The computational work to cover the time interval $[0, T]$ is then very similar for all the schemes. Further it should be noted that the errors in the plots are the total errors in space and time, but it was verified that the temporal errors are the dominating ones.

The errors for the smooth case, with $\epsilon = 1$, are given in Figure 2. As for the linear problem of the previous section, method (1.5) suffers from order reduction in the maximum norm, due to the boundary conditions. In the L_2-norm the order of convergence is two with this method, showing that the largest errors are confined to small regions near the boundaries. The errors for the methods (5.1) and (5.2) with $\theta = 1 - \frac{1}{2}\sqrt{2}$ are very similar; the errors for (5.2) are slightly smaller but this is not really visible. For (5.2) with $\theta = \frac{1}{2}$ (the Craig-Sneyd method) the errors are a little larger.

The results for $\epsilon = \frac{1}{50}$ are shown in Figure 3. Here method (1.5) is convergent with order two in the maximum norm. In fact, the errors for this method are now slightly
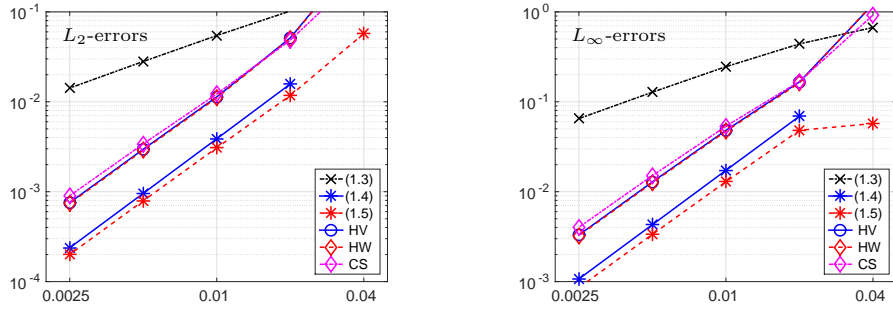

Figure 3: L_2-errors [left] and L_∞-errors [right] versus mesh-width $h = \frac{1}{25}, \frac{1}{50}, \ldots, \frac{1}{400}$, for problem (5.3) with $\epsilon = \frac{1}{50}$, output time $T = 1$. Time step $\Delta t = h$ for the Douglas methods, and $\Delta t = 2h$ for the extended methods (legend names as in Figure 2).

16

smaller than for (1.4). Both modifications of the Douglas method give smaller errors than the extended stabilizing correction methods (5.1) and (5.2). These extended methods produced very similar results in this test.

For $h = \frac{1}{25}$ method (1.4) is unstable, and the same was observed for the methods (5.1) and (5.2) with $\theta = \frac{1}{4}$. Surprisingly, method (1.5) is more stable here. It should be noted, however, that for $h = \frac{1}{30}$ all the methods in the test are stable whereas for $h = \frac{1}{20}$ all the methods become unstable, so the differences between the methods is not very large in this respect.

5.3 A pattern formation problem on a hexagon with domain decomposition splitting

Dimensional splitting requires Cartesian grids. This can be a serious drawback in applications. Non-Cartesian grids, in combination with finite elements (with mass lumping) or finite volume discretizations, can be used with splittings based on domain decomposition, as introduced by Mathew, Polyakov, Russo & Wang [20].

To illustrate this we consider the following system of reaction-diffusion equations, describing the interaction of two chemical species,

$$
\begin{aligned}
u_t &= D_1(u_{xx} + u_{yy}) + \kappa(a - u + u^2 v), \\
v_t &= D_2(v_{xx} + v_{yy}) + \kappa(b - u^2 v),
\end{aligned}
\tag{5.4}
$$

on a hexagonal spatial domain Ω. The vertices of this hexagon are located at the points $(\pm 1, 0)$, $(\frac{1}{2} \pm 1, \frac{1}{2}\sqrt{3})$ and $(\frac{1}{2} \pm 1, -\frac{1}{2}\sqrt{3})$. The initial condition is

$$
u(x, y, 0) = a + b + 10^{-3} e^{-100\left(\left(x - \frac{1}{4}\right)^2 + \left(y - \frac{1}{6}\right)^2\right)}, \qquad v(x, y, 0) = b/(a + b)^2,
$$

and at the boundaries homogeneous Neumann conditions are imposed. The parameter values are $D_1 = 0.05$, $D_2 = 1$, $\kappa = 100$, $a = 0.1305$ and $b = 0.7695$. The initial condition consists of a small Gaussian perturbation added to the chemical steady state $u \equiv a + b = 0.90$, $v \equiv b/(a + b)^2 = 0.95$. Due to the reaction and diffusion processes, this small perturbation is amplified and spread, leading to the formation of patterns with spots. Figure 4 shows snapshots of the time evolution of the u-component at times $t = 0.5$ and $t = 1$. The values for u vary between 0.2 and 2.8, whereas those for v range from 0.4 to 1.2, with smaller values in the spots where u is maximal.
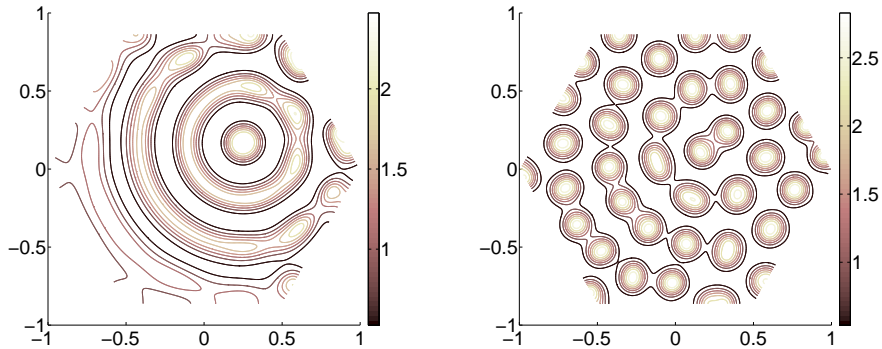

Figure 4: Contour plots of the u-component at times $t = 0.5$ [left] and $t = 1$ [right].

This model is due to Schnakenberg [22]. It was considered on a rectangular domain in [18] as a test for IMEX schemes with $s = 1$, treating the 2D diffusion terms implicitly and the nonlinear reaction terms explicitly. It should be noted that the IMEX method (4.12) of [18, p. 391], which gave favourable results in that test, corresponds to the modified Douglas method (1.5) in case $s = 1$.

Here, we consider the model to test the behaviour of the stabilizing correction methods with $s > 2$, using domain decomposition splitting of the diffusion terms, discretized by finite elements on a triangular grid. We will primarily consider domain decomposition splitting with subdomains formed by squares restricted to Ω, giving $s = 4$, as in Figure 5. In that figure also the subdomains formed by triangles with $s = 6$ are shown; comments for this are given at the end of this section. The domain decomposition splitting technique is briefly outlined here for the squares. A more general and detailed description of this technique can be found in [2, 20]; for further applications we refer to [1, 3].

Consider a two-dimensional problem posed on a region Ω which is embedded in a square $[a, b]^2$. Suppose K is a positive integer and $z_k = a + (b - a)(k - \frac{1}{2})/K$ for $k = 1, 2, \ldots, K$. Let $\psi(z) : [a, b] \to [0, 1]$ be a smooth function (e.g. a spline) such that

$$\psi(z) = \begin{cases} 1 & \text{if } |z - z_k| < r \text{ with } k \text{ odd}, \\ 0 & \text{if } |z - z_k| < r \text{ with } k \text{ even}, \end{cases}$$

where $r < 1/(2K)$. We now define $\psi_1(x, y) = \psi(x)\psi(y)$, $\psi_2(x, y) = (1 - \psi(x))\psi(y)$, $\psi_3(x, y) = \psi(x)(1 - \psi(y))$ and $\psi_4(x, y) = (1 - \psi(x))(1 - \psi(y))$. Then, the two-dimensional Laplace operator $\Delta = \nabla^2$ can be split in a natural way into four parts, $\Delta = \sum_{j=1}^4 \Delta_j$ with

$$\Delta_j u = \nabla\big(\psi_j(x, y)\nabla u\big) \quad \text{for } 1 \leq j \leq 4.$$

With Ω_j being the support of ψ_j in Ω, it is seen that $\Delta_j u$ will be zero outside Ω_j. Moreover, Ω_j consists of a number of disconnected component sets. Consequently, if A_j is the discretization of Δ_j, then a linear system with matrix $I - \theta\Delta t A_j$ will break up into a number of parallel linear systems, each corresponding to one of the component sets of Ω_j.

Since the subdomains in which the spatial domain is decomposed do not have to be aligned with the PDE domain nor with the spatial discretization grid, this technique
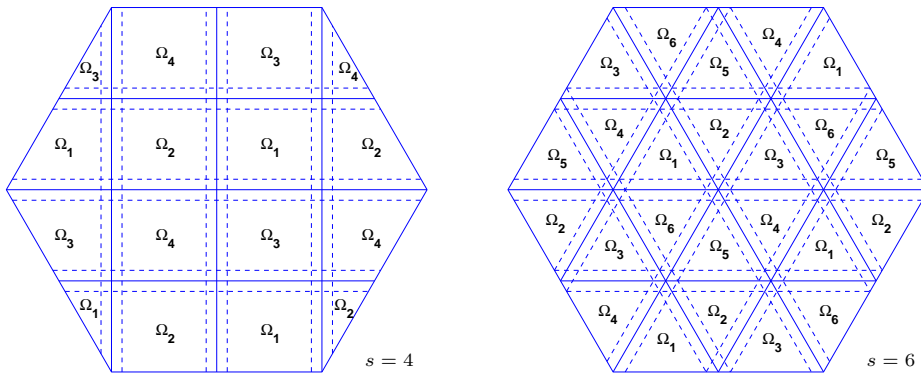

Figure 5: Decomposition of the hexagonal domain Ω into squares, $s = 4$ [left] and triangles, $s = 6$ [right]. The dashed lines indicate the overlap of the regions.
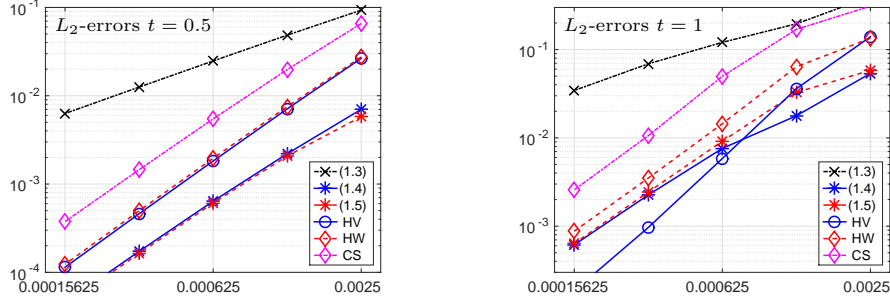
Figure 6: L_2-errors versus τ for problem (5.4) at $t = 0.5$ [left] and $t = 1$ [right] with fixed spatial grid and $\tau = \frac{1}{400}, \frac{1}{800}, \ldots, \frac{1}{6400}$. Step-sizes $\Delta t = \tau$ for the Douglas methods, and $\Delta t = 2\tau$ for the extended methods (legend names as in Figure 2).

is much more flexible in this sense than the (more classical) alternating direction approach.

Following the above procedure, we consider a decomposition of our hexagonal domain Ω into four overlapping subdomains $\{\Omega_j\}_{j=1}^4$, each of which consists of four disjoint components, as shown in Figure 5 (left).

The equations (5.4) are discretized on a triangular grid composed of 98304 equilateral triangles with length $h = \frac{1}{256}$ for the sides, using standard piecewise linear finite elements with mass lumping. On this fixed spatial grid we take time steps $\Delta t = \frac{1}{200}, \frac{1}{400}, \ldots, \frac{1}{6400}$ for the Douglas methods (1.3), (1.4) and (1.5). For the extended methods (5.1) and (5.2), the time steps are taken twice as large, again ensuring that the work to cover a time interval $[0, T]$ is the same for all methods. For these step-sizes the temporal errors for the u-component are plotted in the discrete L_2-norm. These errors have been obtained by comparison with a time-accurate reference solution on the same grid.

The left panel of Figure 6 shows the errors at time $t = 0.5$. The Douglas method is seen to be first-order convergent (since $F_0 \neq 0$), while the other methods all converge with order two. Among them, the Craig-Sneyd method gives the largest errors, followed by the extended methods (5.1) and (5.2), with $\theta = 1 - \frac{1}{2}\sqrt{2}$, which perform very similar. Finally, the modified Douglas methods (1.4) and (1.5) provide the most accurate solutions.

The right panel of Figure 6 displays the errors at time $t = 1$. Here, similar comments apply for all the methods, except for method (5.1) which is now the most accurate for small step-sizes. The reason why this method performs better for this larger output time is not clear. It seems that the errors keep increasing in time for all the schemes, but this increase is much larger for method (5.2) and the modified Douglas methods than for method (5.1). Some additional tests showed that this behaviour is irrespective of the value of the parameter θ in the extended methods.

As seen in Figure 5, the decomposition of the hexagon with squares gives component sets of different size. To balance the size of the disjoint components, we could also consider products of functions in the directions given by $y = 0$, $y = \sqrt{3}x$ and $y = -\sqrt{3}x$. This would lead to a domain decomposition into 6 subdomains, each consisting of disjoint triangular component sets, as shown in the right panel of Figure 5, giving a value $s = 6$ in the splitting formula (1.2). For this decomposition the errors of the stabilizing correction methods were very similar to the decomposition with squares, again with a somewhat different behaviour at the different output times, so these

19

results are not reproduced here.

6 Concluding remarks

As we have seen in the tests for reaction-diffusion problems, the modified Douglas methods (1.4) and (1.5) can be considerably more efficient than the extended stabilizing correction methods (5.1) and (5.2).

For more general applications, it should be noted that the modified Douglas methods may not be as stable as the extended methods with parameter $\theta > \frac{1}{4}$. For example, for parabolic problems with cross derivatives that are treated explicitly, instabilities have been observed with the modified Douglas methods while the extended methods still give good results with properly chosen parameter values θ, see e.g. [9]–[14].

The variant (1.5) is more likely to suffer from order reduction than the modification (1.4). Such an order reduction can be avoided by using boundary corrections, see e.g. [21]. However, since boundary corrections have to be derived anew for each problem, a method that does not need this has a distinct advantage over methods that do need such corrections.

References

[1] A. Arrarás, F.J. Gaspar, L. Portero, C. Rodrigo, *Domain decomposition multigrid methods for nonlinear reaction–diffusion problems.* Commun. Nonlinear Sci. Numer. Simul. 20 (2015), 699–710.

[2] A. Arrarás, L. Portero, *Improved accuracy for time-splitting methods for the numerical solution of parabolic equations.* Appl. Math. Comput. 267 (2015), 294–303.

[3] A. Arrarás, L. Portero, I. Yotov, *Error analysis of multipoint flux domain decomposition methods for evolutionary diffusion problems.* J. Comput. Phys. 257 (2014), 1321–1351.

[4] N. Borovykh, D. Drissi, M.N. Spijker, *A bound on powers of linear operators, with relevance to numerical stability.* Appl. Math. Letters 15 (2002), 47–53.

[5] 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.

[6] J. Douglas, *Alternating direction methods for three space variables.* Numer. Math. 4 (1962), 41–63

[7] J. Douglas, J.E. Gunn, *A general formulation of alternating direction methods.* Numer. Math. 6 (1964), 428–453

[8] I. Faragó, C. Palencia, *Sharpening the estimate of the stability constant in the maximum-norm of the Crank-Nicolson scheme for the one-dimensional heat equation.* Appl. Numer. Math. 42 (2002), 133–140.

[9] K.J. in 't Hout, C. Mishra, *Stability of the modified Craig-Sneyd scheme for two-dimensional convection-diffusion equations with mixed derivative term.* Math. Comp. Simul. 81 (2011), 2540–2548.

[10] K.J. in 't Hout, C. Mishra, *Stability of ADI schemes for multidimensional diffusion equations with mixed derivative terms.* Appl. Numer. Math. 74 (2013), 83–94.

[11] K.J. in 't Hout, B.D. Welfert, *Stability of ADI schemes applied to convection-diffusion equations with mixed derivative terms.* Appl. Numer. Math. 57 (2007), 19–35.

[12] K.J. in 't Hout, B.D. Welfert, *Unconditional stability of second-order ADI schemes applied to multi-dimensional diffusion equations with mixed derivative terms.* Appl. Numer. Math. 59 (2009), 677–692.

[13] K.J. in 't Hout, M. Wyns, *Convergence of the Hundsdorfer-Verwer scheme for two-dimensional convection-diffusion equations with mixed derivative term.* Proceedings Icnaam-2014, T.E. Simos, C. Tsitouras (eds.), AIP Conf. Proc. 1648, 850054 (2015).

[14] K.J. in 't Hout, M. Wyns, *Convergence of the Modified Craig-Sneyd scheme for two-dimensional convection-diffusion equations with mixed derivative term.* To appear in J. Comp. Appl. Math. (2015).

[15] W. Hundsdorfer, *Unconditional convergence of some Crank-Nicolson LOD methods for initial-boundary value problems.* Math. Comp. 58 (1992), 35–53.

[16] W. Hundsdorfer, *A note on stability of the Douglas splitting method.* Math. Comp. 67 (1998), 183–190.

[17] W. Hundsdorfer, *Accuracy and stability of splitting with stabilizing corrections.* Appl. Numer. Math. 42 (2002), 213–233.

[18] W. Hundsdorfer, J.G. Verwer, *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations.* Springer, 2003.

[19] G.I. Marchuk, *Splitting and alternating direction methods.* In: *Handbook of Numerical Analysis I.* Eds. P.G. Ciarlet, J.L. Lions, North-Holland, 1990, 197–462.

[20] T.P. Mathew, P.L. Polyakov, G. Russo, J. Wang, *Domain decomposition operator splittings for the solution of parabolic equations.* SIAM. J. Sci. Comput. 19 (1998), 912–932.

[21] L. Portero, J.C. Jorge, B. Bujanda, *Avoiding order reduction of fractional step Runge-Kutta discretizations for linear time dependent coefficient parabolic problems.* Appl. Numer. Math. 48 (2004), 409–424.

[22] J. Schnakenberg, *Simple chemical reaction systems with limit cycle behaviour.* J. Theor. Biol. 81 (1979), 389–400.

[23] S.I. Serdjukova, *Uniform stability with respect to the initial data of a six-point symmetric scheme for the heat equation.* Zh. Vychisl. Mat. i Mat. Fiz. 4 (1964), 212–216.

[24] J.G. Verwer, E.J. Spee, J.G. Blom, W. Hundsdorfer, *A second-order Rosenbrock method applied to photochemical dispersion problems.* SIAM J. Sci. Comput. 20 (1999), 1456–1480.