High order PDE-convergence of AMF-W methods for 2D-linear parabolic problems

S. González-Pinto[a,1], E. Hairer [b,1], D. Hernández-Abreu[a,1]

[a] *Departamento de Análisis Matemático. Universidad de La Laguna. 38071, La Laguna, Spain. (spinto@ull.edu.es, dhabreu@ull.edu.es)*
[b] *Section de Mathématiques, Université de Genève, CH-1205 Genève, Switzerland. (Ernst.Hairer@unige.ch).*

Abstract

The orders of PDE-convergence in the Euclidean norm of s-stage AMF-W-methods for two-dimensional parabolic problems on rectangular domains are considered for the case of Dirichlet boundary conditions and an initial condition. The classical algebraic conditions for order p with $p \leq 3$ are shown to be sufficient for PDE-convergence of order p (independently of the spatial resolution) in the case of time-independent Dirichlet boundary conditions. Under additional conditions, PDE-convergence of order $p = 3.25 - \epsilon$ for every $\epsilon > 0$ can be obtained. In the case of time-dependent boundary conditions the order reduction is more dramatic, but order $p = 2.25 - \epsilon$ for every $\epsilon > 0$ can be achieved.

Keywords: parabolic problem, AMF-W method, PDE-convergence, order conditions, fractional order
AMS subject classifications: 65M12, 65M15, 65M20.

1. Introduction

This article is devoted to the numerical treatment of linear diffusion problems on an open rectangular domain (unit square) with Dirichlet boundary conditions,

$$\begin{aligned} \partial_t u = a\,\partial_{xx}^2 u + b\,\partial_{yy}^2 u + c(t,x,y), \quad (x,y) \in \Omega = (0,1)^2,\ t \in (0,T], \\ u(t,x,y) = \beta(t,x,y), \quad (x,y) \in \partial\Omega,\ t \in (0,T], \end{aligned} \tag{1}$$

and the initial condition $u(0,x,y) = u_0(x,y),\ (x,y) \in \Omega$ ($\partial\Omega$ denotes the boundary of Ω). The coefficients a and b are assumed to be positive constants, and the inhomogeneity $c(t,x,y)$ sufficiently smooth. The Dirichlet boundary conditions (BCs) can be either time-dependent or time-independent, which makes an important difference in the convergence order of the methods considered below. A second order central difference space discretisation on a uniform grid $\{(x_i,y_j)\,|\,i = 1,\dots,N,\ j = 1,\dots,M\}$ with spacings $\Delta x = 1/(N+1)$ and $\Delta y = 1/(M+1)$ yields an ordinary differential equation

$$\dot U = DU + g(t), \qquad D = D_1 + D_2, \qquad D_1 = a(I_M \otimes D_{xx}),\ D_2 = b(D_{yy} \otimes I_N) \tag{2}$$

$$g(t) = (g_{i,j}(t))_{i,j=1}^{N,M}, \quad g_{i,j}(t) = c(t,x_i,y_j) + \beta_1(t,x_i,y_j) + \beta_2(t,x_i,y_j), \quad \text{where}$$

$$(\Delta x)^2 \beta_1(t,x_i,y_j) = \begin{cases} a\,\beta(t,0,y_j) & \text{if } x_i = \Delta x, \\ a\,\beta(t,1,y_j) & \text{if } x_i = 1-\Delta x, \\ 0 & \text{otherwise,} \end{cases} \qquad (\Delta y)^2 \beta_2(t,x_i,y_j) = \begin{cases} b\,\beta(t,x_i,0) & \text{if } y_j = \Delta y, \\ b\,\beta(t,x_i,1) & \text{if } y_j = 1-\Delta y, \\ 0 & \text{otherwise,} \end{cases}$$

*Corresponding author: D. Hernández-Abreu (dhabreu@ull.edu.es)
[1] This work was partially supported by the Swiss National Science Foundation, grant No. 200020_192129, and by the University of La Laguna

with the tridiagonal Toeplitz matrices $D_{xx} = tridiag\,(1, -2, 1)/\Delta x^2$ and $D_{yy} = tridiag\,(1, -2, 1)/\Delta y^2$, whose dimensions are N and M, respectively. Here, $U(t) = (U_{ij}(t))_{i,j}$, where $U_{ij}(t) \approx u(t, x_i, y_j)$ approximates the solution of (1) on the grid. The vector $g(t) = (g_{ij}(t))_{i,j}$ contains contributions from the inhomogeneity and from the boundary conditions. Throughout the paper, I_d will stand for the identity matrix of a given dimension d.

For a fine grid the differential equation (2) is stiff and a numerical integration with an implicit method typically requires the solution of a large linear system with matrix $(I_{NM} - \tau\theta D)$. For the sake of efficiency, one is interested to replace this matrix by the product $(I_{NM} - \tau\theta D_1)(I_{NM} - \tau\theta D_2)$, so that only the solution of linear systems with tridiagonal matrices is needed. This kind of splitting goes back to the pioneering papers by Douglas [2] and Douglas & Rachford [3]. Since then, the terminology ADI-type methods (Alternating Direction Implicit) has been extensively used to refer to them in the literature. A good review about this approach for the time integration of partial differential equations (PDEs) can be found in van der Houwen & Sommeijer [20].

Fractional orders of convergence in the ℓ_p norms for Rosenbrock and Runge-Kutta methods applied to parabolic PDEs are proven in [17] and [16], respectively, but in both cases splitting was not considered. In [15, Section 6] convergence results for general W-methods applied to semilinear and quasilinear parabolic problems are given and it is shown that in certain Sobolev norms a PDE convergence order between one and two can be achieved. Also in this work, ADI-type methods are not treated.

PDE-convergence of order two (convergence in time, independently of the spatial resolution) for 2D problems in the weighted Euclidean norm for some ADI methods are shown in [11] for the Peaceman–Rachford method, in [1] for a modified Douglas splitting and in [13] for problems with mixed derivatives by using a modified Craig–Sneyd scheme. PDE-convergence in the ℓ_2 and maximum norms for 1-stage AMF-W methods in the case of m spatial dimensions is considered in [5]. In both norms, convergence of order one is proven. Furthermore, convergence of order two for a particular method under some mild ratio between the spatial and time resolutions is also shown. Additionally, in [6] PDE-convergence of order two in the maximum norm for the former one stage AMF-W method as well as for the Douglas method is derived.

Outline of the paper. The aim of the present article is to consider numerical methods for linear parabolic differential equations that permit to achieve a PDE-convergence of high order. Section 2 recalls AMF-W-methods, which are related to Rosenbrock methods with inexact Jacobian, but use a dimensional splitting on the linear algebra level for reasons of efficiency. Non-stiff order conditions up to order 3 are given. The main convergence results (for the Euclidean norm) are presented in Section 3. For time-independent Dirichlet boundary conditions we present conditions on the parameters of the method that imply PDE-convergence of order p with $p \leq 3$, and we discuss methods from the literature that satisfy these conditions. We show that, if in addition a subset of the order conditions for order 4 is satisfied, one can achieve order $p = 3.25 - \epsilon$ for every $\epsilon > 0$. For general boundary conditions, methods with PDE-convergence of order $p \leq 2$ are discussed, and conditions are given that permit to obtain order $p = 2.25 - \epsilon$ for every $\epsilon > 0$. The local error of the methods is analysed in Section 4. General convergence theorems are given in Section 5. A new convergence theorem is presented that permits to prove the sharp results for the order of PDE-convergence. Detailed proofs of the convergence statements are given in Sections 6 and 8 for general Dirichlet boundary conditions, and in Sections 7 and 9 for time-independent Dirichlet boundary conditions. Section 10 is devoted to numerically illustrate the sharpness of the PDE-convergence orders of Section 3.

2. AMF-W methods and their (non-stiff) order

According to the splitting in (2) we also split

$$g(t) \equiv g(t,\cdot,\cdot) = g_1(t) + g_2(t), \quad g_1(t) = c(t,\cdot,\cdot) + \beta_1(t,\cdot,\cdot), \quad g_2(t) = \beta_2(t,\cdot,\cdot), \tag{3}$$

where "$\cdot$" refers to the associated space grid-points. Observe that $g_1(t)$ contains the contribution from $a \cdot \partial_{xx}^2 u$ and from $c(t, x, y)$, whereas $g_2(t)$ only the one of $b \cdot \partial_{yy}^2 u$. When inhomogeneous Dirichlet boundary conditions are imposed $g_1(t)$ and $g_2(t)$ contain negative powers of Δx and Δy, respectively.

The discretized problem (2)-(3) can be written as

$$U'(t) = F(t,U) = F_1(t,U) + F_2(t,U), \quad U(0) = U_0, \qquad F_j(t,U) = D_jU + g_j(t), \ (j = 1,2),$$

and the classical ADI Douglas method [12, Chap. IV.3] is given by ($\theta \geq 1/2$ for stability reasons)

$$\begin{aligned} v_0 &= U_n + \tau F(t_n, U_n), \\ v_i &= v_{i-1} + \theta\tau \left(F_i(t_{n+1}, v_i) - F_i(t_n, U_n)\right), \quad i = 1,2, \\ U_{n+1} &= v_2. \end{aligned}$$

This method requires the solution of linear systems with coefficient matrices of type $(I_{NM} - \theta\tau D_1)(I_{NM} - \theta\tau D_2)$ instead of $(I_{NM} - \theta\tau D)$ as mentioned in the introduction. In this paper we consider the so-called AMF-W-methods [5, 7, 8], which are a kind of W-methods [19], [10, Chap. IV.7], modified in such a way that they can cope adequately with 2D-parabolic problems (also with mD-parabolic problems) by using some splitting in the Jacobian through the AMF (Approximate Matrix Factorization [20]) and they involve similar computational costs as ADI methods for the case of linear parabolic problems and one stage [5, 6]. For more stages, the computational cost per integration step of AMF-W methods increases but this can be compensated by the increase in the convergence order.

For the time integration of (2)-(3) we consider AMF-W methods. Given a numerical approximation $U_n \approx U(t_n)$ at t_n, the approximation $U_{n+1} \approx U(t_{n+1})$ at $t_{n+1} = t_n + \tau$ is defined by

$$\begin{aligned} &K_i^{(0)} = \tau D\Big(U_n + \sum_{j=1}^{i-1} a_{ij}K_j\Big) + \tau\, g(t_n + c_i\tau) + \sum_{j=1}^{i-1} \ell_{ij}K_j, \\ &(I_{MN} - \theta\tau D_1)K_i^{(1)} = K_i^{(0)} + \theta\rho_i\tau^2\dot{g}_1(t_n + \eta\tau), \\ &(I_{MN} - \theta\tau D_2)K_i = K_i^{(1)} + \theta\rho_i\tau^2\dot{g}_2(t_n + \eta\tau), \quad i = 1,2,\ldots,s, \\ &U_{n+1} = U_n + \sum_{i=1}^{s} b_iK_i. \end{aligned} \tag{4}$$

It is characterized by (A, L, b, θ, η), where $A = (a_{i,j})_{j<i}$, $L = (\ell_{i,j})_{j<i}$ and $b = (b_i)_i$ are matrices or vectors and $\theta > 0$ and $\eta \geq 0$ are two constants. The coefficients ρ_i and c_i are defined by $\rho_i = 1 + \sum_{j=1}^{i-1} \ell_{ij}\rho_j$ and $c_i = \sum_{j=1}^{i-1} a_{ij}\rho_j$. In vector form they are

$$\rho = (I_s - L)^{-1}\mathbf{1} \qquad \text{and} \qquad c = A\rho, \tag{5}$$

where $\mathbf{1}$ denotes the vector of dimension s with all entries equal to 1. We also use the notation $c^r = (c_i^r)_i$.

A one-step method, like (4), has non-stiff (or classical) order p, if for fixed Δx and Δy the global error satisfies

$$U_n - U(t_n) = \mathcal{O}(\tau^p) \quad \text{for} \quad 0 \leq t_n = n\tau \leq T. \tag{6}$$

In this definition the constant symbolised by $\mathcal{O}$ is allowed to depend on negative powers of Δx and Δy. For small values of p, order conditions are obtained straight-forwardly by expanding the local error into powers of τ. Using the vector notation

$$\tilde{A} = A(I_s - L)^{-1}, \quad \tilde{b}^\top = b^\top(I_s - L)^{-1}, \quad \tilde{\Gamma} = \theta(I_s - L)^{-1} \tag{7}$$

together with ρ and c from above, we obtain the following order conditions (non-stiff case) which arise when W-methods that satisfy $W - F'(t_n, U_n) = \mathcal{O}(\tau)$ (hereafter, $F'(t,U) \equiv \partial F(t,U)/\partial U$) are applied to Initial Value Problems in ODEs of the form $\dot{U} = F(t,U)$, see e.g. [19], [14], [7, sect. 2] or [10, p. 114-117],

order $p = 1 \iff \tilde{b}^\top\mathbf{1} = 1.$

order $p = 2 \iff \tilde{b}^\top\mathbf{1} = 1$ and $\tilde{b}^\top(\tilde{A} + \tilde{\Gamma})\mathbf{1} = 1/2.$

$$\text{order } p=3 \quad \Longleftrightarrow \quad \tilde b^\top \mathbf{1}=1,\quad \tilde b^\top \tilde A\,\mathbf{1}=1/2,\quad \tilde b^\top \tilde\Gamma\,\mathbf{1}=0 \quad \text{and}$$
$$\tilde b^\top c^2 = 1/3,\quad \tilde b^\top(\tilde A+\tilde\Gamma)^2\mathbf{1}=1/6.$$

The main interest of the present work is to derive conditions on the coefficients of the method, where (6) is satisfied with a constant (symbolised by $\mathcal{O}$) that is independent of Δx and Δy. In such a situation we say that the method is PDE-convergent of order p. It will be seen below that the convergence results are more favorable for the case of time-independent Dirichlet boundary conditions (which of course also apply to homogeneous BCs). In many cases and for methods of high consistency order ($p \geq 3$) the convergence order for time-independent BCs is typically one unit larger than for time-dependent BCs.

3. Main PDE-convergence results

Removing $g_1(t)$ and $g_2(t)$ from (4) we obtain $U_{n+1} = R(\tau D_1, \tau D_2)U_n$ with a stability function given by

$$\begin{aligned} R(z_1,z_2) &= 1 + b^\top P(z_1,z_2)^{-1}\mathbf{1}\cdot(z_1+z_2), \qquad \text{where} \\ P(z_1,z_2) &= (1-\theta z_1)(1-\theta z_2)I_s - (z_1+z_2)A - L \end{aligned} \tag{8}$$

is a triangular matrix of dimension s. Notice that the matrices D_1 and D_2 commute and that $R(\tau D_1, \tau D_2)$ and $P(\tau D_1, \tau D_2)$ are defined by replacing z_j by τD_j, $j = 1, 2$, in the convenient way by using the Kronecker product. Here, the identity matrices have the adequate dimensions, so that in the left side of the Kronecker product they have dimension s and the dimension of D in the right side of it,

$$\begin{aligned} R(\tau D_1,\tau D_2) &= I_{NM} + (b^\top \otimes I)P(\tau D_1,\tau D_2)^{-1}(\mathbf{1}\otimes \tau D), \\ P(\tau D_1,\tau D_2) &= I_s \otimes \big((I_{MN}-\theta\tau D_1)(I_{MN}-\theta\tau D_2)\big) - A\otimes \tau D - L\otimes I_{MN}. \end{aligned} \tag{9}$$

Besides the standard assumption on the stability function our convergence analysis requires also an assumption on the real-valued function

$$q(X) = \sum_{k=0}^{s-1} b^\top X^k\,\mathbf{1}, \tag{10}$$

defined for $s \times s$ matrices X. Consequently, $q(\zeta A)$ is a scalar real polynomial of degree $s-1$ of the real variable ζ. We note that the derivative of $q(X)$, applied to a $s \times s$ matrix H, is given by

$$q'(X)H = b^\top\big(H + (XH+HX) + (X^2H + XHX + HX^2) + \ldots\big)\mathbf{1}. \tag{11}$$

The importance of this function will become clear from Lemma 7 below.

All convergence results and order statements of the present work are with respect to the Euclidean norm

$$\|U\| = \sqrt{\langle U,U\rangle}, \qquad \langle U,V\rangle = \Delta x\,\Delta y \sum_{i=1}^{N}\sum_{j=1}^{M} U_{ij}V_{ij} \tag{12}$$

for vectors $U, V \in \mathbb{R}^N \times \mathbb{R}^M$. The convergence results of this section are proved in Sections 6, 7, 8, and 9.

3.1. Time-independent boundary conditions

We first consider the case of time-independent boundary conditions. This implies that the derivatives $\dot g_1(t)$ and $\dot g_2(t)$ in (4) do not contain any negative powers of Δx and Δy.

Theorem 1. *Let* (2) *be the space discretisation of* (1) *with time-independent Dirichlet boundary conditions. If, for an s-stage AMF-W method* (4),

(a) the stability function satisfies $-1 \leq R(z_1,z_2) < 1$ *for* $z_1 \leq 0$, $z_2 \leq 0$ *and* $(z_1,z_2) \neq (0,0)$,
(b) we have $q(\zeta A) > 0$ *for* $\zeta \in (-\theta^{-1}, 0]$, *and* $q'(-\theta^{-1}A)(\theta^{-1}A + L) > 0$ *if* $q(-\theta^{-1}A) = 0$,

(c) the (non-stiff) order conditions of Section 2 hold for some integer $p \leq 3$,

then the AMF-W method is PDE-convergent of order p.

We conjecture that AMF-W methods (4) cannot be PDE-convergent of order $p \geq 4$.

Corollary 2. *The following AMF-W methods reach PDE-order $p = 3$, for the time integration of the spatial discretization of the problem (1) when time-independent boundary conditions are imposed:*

1. *the Hundsdorfer–Verwer method in [12, p. 155, p. 400-405]*

$$A = \begin{pmatrix} 0 & 0 \\ 2/3 & 0 \end{pmatrix}, \quad L = \begin{pmatrix} 0 & 0 \\ -4/3 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 5/4 \\ 3/4 \end{pmatrix}, \tag{13}$$

   and $\theta = (3 + \sqrt{3})/6$,
2. *the 3-stage AMF-W methods in [8, Theorem 1] for $\theta \geq 1/3$,*
3. *the 4-stage AMF-W methods in [7, Corollary 1 and Theorem 3] with $\theta \geq \theta_1 = 0.36367...$,*
4. *the 3− and 4−stage AMF-W methods by Rang & Angermann in [18].*

Proof. Assumption (a) of Theorem 1 on the stability function and Assumption (c) on the non-stiff order are satisfied by construction of the methods.

For all the 2-stage methods in [12, p. 400-405] we have $q(\zeta A) = (\zeta + 4)/2$. If $\theta > \frac{1}{4}$, $q(\zeta A) > 0$ for $\zeta \in [-\theta^{-1}, 0]$. For $\theta = 1/4$ we have $q'(-\theta^{-1}A)(\theta^{-1}A + L) = b^\top H\mathbf{1} = 1 > 0$, with $H = \theta^{-1}A + L$ by (11).

For the 3-stage methods of item 2 it holds $q(\zeta A) = (\zeta + 3)(\zeta + 6)/6 > 0$ for $\zeta \in [-\theta^{-1}, 0]$, whenever $\theta > \frac{1}{3}$. For $\theta = 1/3$ we have from (11) that $q'(-\theta^{-1}A)(\theta^{-1}A + L) = \frac{3}{2} > 0$.

For the 4-stage methods of item 3, $q(\zeta A) = 1 + (\frac{1}{2}\zeta + 1) + (\frac{1}{6}\zeta^2 + \zeta + 1) + \frac{1}{24}(\zeta - q_1)^3$, with $q_1 := -3 + \sqrt{3}$. For this polynomial it holds $q(\zeta A) > 0$, $\zeta \in [-\theta^{-1}, 0]$, whenever $\theta > \theta_1 = 0.36367\ldots$, where θ_1 is the real root of $1 + \left(-13 + 3\sqrt{3}\right)\theta - 18\left(-4 + \sqrt{3}\right)\theta^2 + 6\left(-21 + 5\sqrt{3}\right)\theta^3 = 0$. For $\theta = \theta_1 = 0.36367\ldots$ we have from (11) that $q'(-\theta^{-1}A)(\theta^{-1}A + L) = 1.76069\ldots > 0$.

Regarding the methods of item 4 in [18], for the 3−stage methods ROS3w and ROS3Dw it holds $q(\zeta A) = 1.852859819860479 + 0.6176199399534931\zeta > 0$, for $\zeta \in [-\theta^{-1}, 0]$ with $\theta = 0.4358665215084590$. For the 3−stage method ROS3Pw, one has $q(\zeta A) = 2.366025403784439 + 0.7886751345948129\zeta > 0$, for $\zeta \in [-\theta^{-1}, 0]$ with $\theta = 0.78867513459481287$. For the 4−stage methods ROS34PW1a and ROS34PW1b, we have $q(\zeta A) = 2.942190516733562 + 1.521014142404432\zeta + 0.1868086517466218\zeta^2 > 0$, for $\zeta \in [-\theta^{-1}, 0]$ with $\theta = 0.4358665215084590$. For the 4−stage methods ROS34PW2, it holds $q(\zeta A) = 3.135557863716253 + 1.573942402757592\zeta + 0.07464477599087557\zeta^2 - 0.04293183963321737\zeta^3 > 0$, for $\zeta \in [-\theta^{-1}, 0]$ with $\theta = 0.4358665215084590$. For the 4−stage methods ROS34PW3, with $\theta = 1.0685790213016289$, it holds $q(\zeta A) = 5.065434545246438 + 6.766542550555094\zeta + 4.550637949792242\zeta^2 + 1.191138422438117\zeta^3 > 0$, for $\zeta \in [-\theta^{-1}, 0]$. □

3.2. Time-dependent boundary conditions

We consider here the general case of time-dependent Dirichlet boundary conditions, so that the derivatives $\dot{g}_1(t)$ and $\dot{g}_2(t)$ typically contain terms with negative powers of Δx and Δy, respectively. In the following we assume that the time step size satisfies $\tau \geq c_0 \max(\Delta x^2, \Delta y^2)$ for some $c_0 > 0$, which is a natural assumption when applying linearly implicit integration methods. Note that we shall replace Assumption (b) of Theorem 1 by a slightly stronger assumption.

Theorem 3. *Let* (2) *be the space discretisation of* (1) *with time-dependent Dirichlet boundary conditions. An s-stage AMF-W method satisfying the Assumption* (a) *of Theorem 1 and*

(b^*) *all zeros of the polynomial* $q(\zeta A)$ *are outside the closed interval* $[-\theta^{-1}, 0]$,

is PDE-convergent of

order $p = 1$ *if* $\tilde{b}^\top \mathbf{1} = 1$, *order* $p^* = 2$ *if* $\tilde{b}^\top \mathbf{1} = 1, \quad \tilde{b}^\top (\tilde{A} + \tilde{\Gamma})\, \mathbf{1} = 1/2.$

Here, $p^* = 2$ *means that we have order* $2 - \epsilon$ *for every* $\epsilon > 0$.

Remark 1. Under the assumptions of Theorem 3 we have order $p = 2$, if in addition to the two order conditions either $\eta = 1/2$ is imposed or the step size satisfies $\tau \le c_1 \min(\Delta x, \Delta y)$ with some $c_1 > 0$.

Remark 2. If we relax Assumption (b^*) to Assumption (b) of Theorem 1, the statement for order $p = 1$ is still true, but we only get order $p^* = 1.5$ (order $p = 1.5$ under the restriction $\tau \le c_1 \min(\Delta x, \Delta y)$) if the order conditions for order 2 are satisfied. This is confirmed by numerical experiments with the methods in Corollary 2 in the following cases, item 1 with $p = s = 2$ and $\theta = 1/4$, item 2 with $p = s = 3$ and $\theta = 1/3$, and the one in item 3 with $\theta = \theta_1$.

We conjecture that for time-dependent Dirichlet boundary conditions AMF-W methods (4) cannot be PDE-convergent of order $p \ge 3$.

Corollary 4. *The following AMF-W methods reach PDE-order* $p^* = 2$ *(and* $p = 2$ *when* $\eta = 1/2$*), for the time integration of the spatial discretization of the problem (1) when time-dependent boundary conditions are imposed:*

1. *the* 2*-stage AMF-W-methods collected in [12, p. 400–405]) for* $\theta > 1/4$,
2. *the methods of Corollary 2, except those of item* 2 *with* $s = p = 3$ *and* $\theta = 1/3$, *and those of item* 3 *with* $s = 4, p = 3$, *and* $\theta = \theta_1$,
3. *the 3-stage method given in [4].*

Proof. The proof for the methods of items 1 and 2 follow as in the proof of Corollary 2.

For the method of item 3, Assumption (a) of Theorem 1 on the stability function and the non-stiff order conditions for order $p \le 2$ in Section 2 are satisfied by construction of the method. It also holds that $q(\zeta A) = 1.613684936166224 + 0.7045616453887415\zeta + 0.08333333333333333\zeta^2 > 0$ for $\zeta \in [-\theta^{-1}, 0]$ with $\theta = 0.4358665215084590$. ☐

3.3. Fractional order

We are not able to find methods of order $p = 3$ for time-independent boundary conditions. However, with additional order conditions we can improve the order up to 2.25.

Theorem 5. *Let* (2) *be the space discretisation of* (1) *with time-dependent Dirichlet boundary conditions. An s-stage AMF-W method satisfying the Assumption* (a) *of Theorem 1 and Assumption* (b^*) *of Theorem 3 is PDE-convergent of*

order $p^* = 2.25$ *if* $\eta = 1/2$ *and the order conditions of Section 2 for order* 3 *hold.*

The improvement of the order for time-independent boundary conditions requires a subset of the order conditions for order 4. They are

$$\begin{aligned}
&\tilde{b}^\top (\tilde{A} + \tilde{\Gamma})\, \tilde{\Gamma} \mathbf{1} = 0 \\
&\tilde{b}^\top (\tilde{A} + \tilde{\Gamma})^3 \mathbf{1} = 1/24 \\
&\tilde{b}^\top \tilde{\Gamma}\, (\tilde{A} + \tilde{\Gamma}) \mathbf{1} = 0 \\
&\tilde{b}^\top (\tilde{A} + \tilde{\Gamma})\, c^2 = 1/12 \\
&\tilde{b}^\top c^3 = 1/4.
\end{aligned} \tag{14}$$

Theorem 6. *Let* (2) *be the space discretisation of* (1) *with time-independent Dirichlet boundary conditions. An s-stage AMF-W method satisfying the Assumptions (a) and (b) of Theorem 1 is PDE-convergent of*

order $p^* = 3.25$ *if* *the* 5 *order conditions of Section 2 for order* 3 *and the conditions* (14) *hold.*

The proof of the statements in Theorems 1, 3, 5, 6 and the justification of the Remarks 1 and 2 are presented in Sections 6, 7, 8, and 9, below.

4. Local error

For the study of the local error of (4) we consider an initial value $\tilde{U}_n = U(t_n)$ on the exact solution, and we denote the internal approximations and the numerical solution after one step by $\tilde{K}_i^{(l)}, \tilde{K}_i$ and $\tilde{U}_{n+1}$. Our aim is to expand the local error

$$\nu_n = U(t_n+\tau) - \tilde{U}_{n+1} = U(t_n+\tau) - U(t_n) - \sum_{i=1}^{s} b_i \tilde{K}_i \tag{15}$$

in terms of smooth and bounded coefficients. In addition to $U(t)$ we assume that the functions $(m=1,2)$

$$\varphi_m(t) = D_m U(t) + g_m(t) \tag{16}$$

are smooth and have derivatives with bounds in the Euclidean norm that are independent of the space discretisation. Observe that from (2)-(3) we have that $\varphi_1(t) = D_1 U + \beta_1(t) + c(t,x,y) = (a\partial_{xx}^2 + \mathcal{O}(\Delta x^2))u(t,x,y) + c(t,x,y)$ and $\varphi_2(t) = D_2 U + \beta_2(t) = (b\partial_{yy}^2 + \mathcal{O}(\Delta y^2))u(t,x,y)$.

For the study of the local error we first have to express the vectors $\tilde{K}_i$ in terms of smooth quantities. With the abbreviation

$$\pi(\tau D_1, \tau D_2) = (I_{MN} - \theta\tau D_1)(I_{MN} - \theta\tau D_2)$$

we obtain from a multiplication of the equation for $\tilde{K}_i$ with $(I_{MN} - \theta\tau D_1)$

$$\pi(\tau D_1, \tau D_2)\tilde{K}_i = \tilde{K}_i^{(0)} + \theta\rho_i\tau^2\Big(\dot{g}_1(t_n+\eta\tau) + (I_{MN} - \theta\tau D_1)\dot{g}_2(t_n+\eta\tau)\Big) \tag{17}$$

where

$$\tilde{K}_i^{(0)} = \tau D\Big(U(t_n) + \sum_{j=1}^{i-1} a_{ij}\tilde{K}_j\Big) + \tau g(t_n + c_i\tau) + \sum_{j=1}^{i-1} \ell_{ij}\tilde{K}_j.$$

Inserting $\tilde{K}_i^{(0)}$ into the formula for $\tilde{K}_i$ and collecting suitable terms, we obtain

$$\begin{aligned} &\pi(\tau D_1, \tau D_2)\tilde{K}_i - \tau D\sum_{j=1}^{i-1} a_{ij}\tilde{K}_j - \sum_{j=1}^{i-1} \ell_{ij}\tilde{K}_j \\ &= \tau D U(t_n) + \tau g(t_n + c_i\tau) + \theta\rho_i\tau^2\Big(\dot{g}_1(t_n+\eta\tau) + (I_{MN} - \theta\tau D_1)\dot{g}_2(t_n+\eta\tau)\Big). \end{aligned}$$

Using (16) we express all appearances of $g_m(t)$ by derivatives of the smooth functions $U(t)$ and $\varphi_l(t)$. This yields

$$\begin{aligned} &\pi(\tau D_1, \tau D_2)\tilde{K}_i - \tau D\sum_{j=1}^{i-1} a_{ij}\tilde{K}_j - \sum_{j=1}^{i-1} \ell_{ij}\tilde{K}_j \\ &= \tau D U(t_n) + \tau\big(\dot{U}(t_n + c_i\tau) - DU(t_n + c_i\tau)\big) \\ &+ \theta\rho_i\tau^2\Big(\ddot{U}(t_n+\eta\tau) - D\dot{U}(t_n+\eta\tau) - \theta\tau D_1\big(\dot{\varphi}_2(t_n+\eta\tau) - D_2\dot{U}(t_n+\eta\tau)\big)\Big). \end{aligned} \tag{18}$$

By an expansion into a Taylor series around $\tau = 0$, the right-hand side of this expression becomes

$$\sum_{m\geq 1} \tau^m \Big(\alpha_i^{(m)} U^{(m)}(t_n) + \beta_i^{(m)} \tau D U^{(m)}(t_n) + \gamma_i^{(m)}\tau^2 D_1 D_2 U^{(m)}(t_n) - \gamma_i^{(m)}\tau^2 D_1 \varphi_2^{(m)}(t_n)\Big) \tag{19}$$

with coefficients given by $\alpha_i^{(1)} = 1$, and

$$\begin{aligned} \alpha_i^{(m)} &= \tfrac{1}{(m-1)!}c_i^{m-1} + \theta\rho_i\tfrac{1}{(m-2)!}\eta^{m-2}, \quad m\geq 2, \\ \beta_i^{(m)} &= -\tfrac{1}{m!}c_i^{m} - \theta\rho_i\tfrac{1}{(m-1)!}\eta^{m-1}, \quad m\geq 1, \\ \gamma_i^{(m)} &= \theta^2\rho_i\tfrac{1}{(m-1)!}\eta^{m-1}, \quad m\geq 1. \end{aligned} \tag{20}$$

Since we are interested in convergence estimates in the Euclidean norm we use a Fourier-type analysis and diagonalise the appearing matrices. We let $\{\phi_k^{(x)}\}_k$ be a basis such that $D_{xx}\phi_k^{(x)} = \lambda_k^{(x)}\phi_k^{(x)}$ and, similarly, $\{\phi_l^{(y)}\}_l$ a basis such that $D_{yy}\phi_l^{(y)} = \lambda_l^{(y)}\phi_l^{(y)}$ (see Section 3 of [5]). The set $\{\phi_l^{(y)} \otimes \phi_k^{(x)}\}_{k,l}$ is then an orthonormal basis with respect to the inner product (12), and we have

$$\tau D_1(\phi_l^{(y)} \otimes \phi_k^{(x)}) = z_1(\phi_l^{(y)} \otimes \phi_k^{(x)}), \qquad \tau D_2(\phi_l^{(y)} \otimes \phi_k^{(x)}) = z_2(\phi_l^{(y)} \otimes \phi_k^{(x)})$$

with negative real numbers $z_1 = \tau a\, \lambda_k^{(x)}$, $z_2 = \tau b\, \lambda_l^{(y)}$, where we suppress the dependence on k and l (recall that $a > 0$ and $b > 0$ are the diffusion coefficients). Furthermore, we use the notation

$$U(t) = \sum_{k,l} \widehat{U}_{kl}(t)(\phi_l^{(y)} \otimes \phi_k^{(x)}), \qquad \varphi_2(t) = \sum_{k,l} \widehat{\varphi}_{2,kl}(t)(\phi_l^{(y)} \otimes \phi_k^{(x)}).$$

Recall that by Parseval's identity we have $\|U\| = \|\widehat{U}\|_2$, where $\widehat{U} = (\widehat{U}_{kl})_{kl}$.

In the basis $\{\phi_l^{(y)} \otimes \phi_k^{(x)}\}_{k,l}$ the equation (18) becomes

$$\begin{aligned} &\Big(\pi(z_1,z_2)\tilde{K}_i - (z_1+z_2)\sum_{j=1}^{i-1} a_{ij}\tilde{K}_j - \sum_{j=1}^{i-1} \ell_{ij}\tilde{K}_j\Big)_{k,l} \\ &= \sum_{m\ge 1} \tau^m \Big(\Big(\alpha_i^{(m)} + \beta_i^{(m)}(z_1+z_2) + \gamma_i^{(m)} z_1 z_2\Big)\widehat{U}_{kl}^{(m)}(t_n) - \gamma_i^{(m)} z_1 \tau \widehat{\varphi}_{2,kl}^{(m)}(t_n)\Big), \end{aligned}$$

where $\pi(z_1,z_2) = (1-\theta z_1)(1-\theta z_2)$. With the triangular matrix, see (8),

$$S(z_1,z_2) := P(z_1,z_2)^{-1} = \big(\pi(z_1,z_2)I_s - (z_1+z_2)A - L\big)^{-1}, \tag{21}$$

where we suppress the dependence on θ, this yields for the Fourier coefficients of $\tilde{K}_i = \sum_{k,l} \tilde{K}_{i,kl}\phi_l^{(y)} \otimes \phi_k^{(x)}$

$$\tilde{K}_{i,kl} = \sum_{m\ge 1} \tau^m \sum_{j=1}^{i} S_{ij}(z_1,z_2)\Big(\big(\alpha_j^{(m)} + \beta_j^{(m)}(z_1+z_2) + \gamma_j^{(m)} z_1 z_2\big)\widehat{U}_{kl}^{(m)}(t_n) - \gamma_j^{(m)} z_1 \tau \widehat{\varphi}_{2,kl}^{(m)}(t_n)\Big).$$

Inserted into the relation (15), the (k,l)-component of the local error $\nu_n = \sum_{k,l} \nu_{n,kl}\phi_l^{(y)} \otimes \phi_k^{(x)}$ becomes

$$\begin{aligned} \nu_{n,kl} = \sum_{m\ge 1} \tau^m \bigg(\frac{1}{m!} - \sum_{i=1}^{s} b_i \sum_{j=1}^{i} S_{ij}(z_1,z_2)\Big(\alpha_j^{(m)} + \beta_j^{(m)}(z_1+z_2) + \gamma_j^{(m)} z_1 z_2\Big)\bigg)\widehat{U}_{kl}^{(m)}(t_n) \\ + \sum_{m\ge 1} \tau^{m+1} \sum_{i=1}^{s} b_i \sum_{j=1}^{i} S_{ij}(z_1,z_2)\gamma_j^{(m)} z_1 \widehat{\varphi}_{2,kl}^{(m)}(t_n). \end{aligned} \tag{22}$$

By using vector notation

$$\alpha^{(m)} = (\alpha_i^{(m)})_i, \quad \beta^{(m)} = (\beta_i^{(m)})_i, \quad \gamma^{(m)} = (\gamma_i^{(m)})_i,$$

we can rewrite (22) in the form

$$\begin{aligned} \nu_{n,kl} = \sum_{m\ge 1} \tau^m \bigg(\frac{1}{m!} - b^\top S(z_1,z_2)\Big(\alpha^{(m)} + \beta^{(m)}(z_1+z_2) + \gamma^{(m)} z_1 z_2\Big)\bigg)\widehat{U}_{kl}^{(m)}(t_n) \\ + \sum_{m\ge 1} \tau^{m+1} b^\top S(z_1,z_2)\gamma^{(m)} z_1 \widehat{\varphi}_{2,kl}^{(m)}(t_n). \end{aligned} \tag{23}$$

5. PDE-convergence

After some preliminary estimates we formulate assumptions on the local error in theorems that permit to prove the convergence statements of Section 3.

5.1. Preliminary estimates and useful formulas

It is helpful to express the matrix $S(z_1,z_2)$ of (21) as

$$\begin{aligned} S(z_1,z_2) = \frac{1}{\pi(z_1,z_2)}(I_s - X)^{-1} &= \frac{1}{\pi(z_1,z_2)}\Big(I_s + X + \ldots + X^{s-1}\Big), \\ X &= \frac{z_1+z_2}{\pi(z_1,z_2)}A + \frac{1}{\pi(z_1,z_2)}L. \end{aligned} \tag{24}$$

Since we always assume $\theta > 0$, the powers of X are bounded for $z_1, z_2 \le 0$. Consequently, there exists a positive constant C, such that

$$\|S(z_1,z_2)\|_\infty \le C/\pi(z_1,z_2). \tag{25}$$

The following lemma is essential for obtaining the convergence results of the present work. [2]

Lemma 7. *Let the stability function $R(z_1,z_2)$ be of order $p \ge 1$ and satisfy Assumption (a) of Theorem 1.*

- *Under the Assumption (b^*) of Theorem 3 we have (with a positive constant C_1)*

$$C_1\frac{|z_1+z_2|}{\pi(z_1,z_2)} \le \big|R(z_1,z_2)-1\big| \qquad \textit{for}\ \ z_1,z_2 \le 0. \tag{26}$$

- *Under the weaker Assumption (b) of Theorem 1 we still have (with a positive constant C_1)*

$$C_1\frac{|z_1+z_2|}{\pi(z_1,z_2)^2} \le \big|R(z_1,z_2)-1\big| \qquad \textit{for}\ \ z_1,z_2 \le 0. \tag{27}$$

Proof. (i) We consider the function

$$G_1(z_1,z_2) := \Big(R(z_1,z_2)-1\Big)\frac{\pi(z_1,z_2)}{z_1+z_2} = b^\top\big(I_s + X + \ldots + X^{s-1}\big)\mathbf{1} = q(X),$$

which follows from (8) and (24), and from the definition (10). At the origin we have $X = L$ which implies $G_1(0,0) = b^\top(I_s - L)^{-1}\mathbf{1} = \tilde b^\top\mathbf{1} = 1$, because $p \ge 1$. By Assumption (a) we therefore have $G_1(z_1,z_2) > 0$ for all $z_1,z_2 \le 0$. In the limit $z_2 \to -\infty$ we obtain

$$G_1(z_1,-\infty) = b^\top\big(I_s + \zeta A + \ldots + \zeta^{s-1}A^{s-1}\big)\mathbf{1} = q(\zeta A), \qquad \zeta = \frac{1}{(-\theta)(1-\theta z_1)}.$$

The interval $z_1 \in [-\infty,0]$ corresponds to $\zeta \in [-\theta^{-1},0]$. As a consequence of Assumption (b^*) of Theorem 3 we therefore have $G_1(z_1,-\infty) \ge \kappa > 0$ for $z_1 \in [-\infty,0]$, where κ is a positive constant. By symmetry, also $G_1(-\infty,z_2) \ge \kappa > 0$ holds for $z_2 \in [-\infty,0]$. By continuity of $G_1(z_1,z_2)$ it follows that there exists $K > 0$ (typically very large), such that $G_1(z_1,z_2) > \kappa/2$ if either $z_1 \le -K$ or $z_2 \le -K$. The positivity of $G_1(z_1,z_2)$ on the compact set $[-K,0]\times[-K,0]$ proves the existence of a constant $C_1 > 0$ such that $G_1(z_1,z_2) \ge C_1$ for all $z_1 \le 0$ and $z_2 \le 0$. This proves the inequality (26).

(ii) The Assumption (b) is weaker than (b^*), because it admits the situation, where $q(\zeta A)$ vanishes for $\zeta = -\theta^{-1}$, i.e., for $z_1 = 0$. Therefore, $G_1(0,-\infty) = 0$, so that (26) does not hold for (z_1,z_2) close to $(0,-\infty)$. In this situation we consider the function

$$G_2(z_1,z_2) := \Big(R(z_1,z_2)-1\Big)\frac{\pi(z_1,z_2)^2}{z_1+z_2} = \pi(z_1,z_2)\,q(X).$$

[2] Throughout the paper, $C, C_0, C_1, \ldots$ will stand for positive constants independent of τ, Δx and Δy, which may take different values at each appearance.

We put $w_2 = z_2^{-1}$ and we expand this expression around $w_2 = 0$. Since $X = \zeta A + w_2\,\zeta\,(\theta^{-1}A + L) + \mathcal{O}(z_1 w_2) + \mathcal{O}(w_2^2)$, the Taylor expansion of $q(X)$ yields

$$q(X) = q(\zeta A) + w_2\,\zeta\, q'(\zeta A)(\theta^{-1}A + L) + \mathcal{O}(z_1 w_2) + \mathcal{O}(w_2^2).$$

For $z_1 = 0$ (i.e., for $\zeta = -\theta^{-1}$) and $q(-\theta^{-1}A) = 0$ we thus have, using $\pi(0, z_2) = (w_2 - \theta)/w_2$,

$$G_2(0, z_2) = q'(-\theta^{-1}A)(\theta^{-1}A + L) + \mathcal{O}(w_2),$$

so that $G_2(0, -\infty) > 0$ by Assumption (b) of Theorem 3. As in the proof of part (i) we now conclude that $G_2(z_1, z_2) \geq C_1 > 0$ for $z_1 \leq 0$ and $z_2 \leq 0$. This proves the estimate (27). □

Let us mention three formulas for $S(z_1, z_2)$ that will be useful in the convergence analysis. First, using the identity $V^{-1} - W^{-1} = W^{-1}(W - V)V^{-1}$ with $W = S(z_1, z_2)^{-1}$ and $V = S(0, 0)^{-1} = (I_s - L)$ we get

$$S(z_1, z_2) - S(0, 0) = S(z_1, z_2)\Big((z_1 + z_2)(\theta I_s + A) - z_1 z_2 \theta^2 I_s\Big)(I_s - L)^{-1}. \tag{28}$$

Second, for $W = \big(\pi(z_1, z_2) S(z_1, z_2)\big)^{-1} = I_s - X$ and $V = S(0, 0)^{-1}$ we obtain

$$S(z_1, z_2) - \frac{S(0, 0)}{\pi(z_1, z_2)} = \frac{S(z_1, z_2)}{\pi(z_1, z_2)}\Big((z_1 + z_2)(A + \theta L) - z_1 z_2 \theta^2 L\Big)(I_s - L)^{-1}. \tag{29}$$

The third relation follows from the trivial identity $S(z_1, z_2) = \pi(z_1, z_2)^{-1} I_s + S(z_1, z_2) X$ and it is

$$S(z_1, z_2) - \frac{I_s}{\pi(z_1, z_2)} = \frac{S(z_1, z_2)}{\pi(z_1, z_2)}\Big((z_1 + z_2)A + L\Big). \tag{30}$$

5.2. Conventional convergence theorem

The global error $E_n = U_n - U(t_n)$ of method (4) satisfies a recursion

$$E_{n+1} = R(\tau D_1, \tau D_2) E_n - \nu_n,$$

where $R(\cdot, \cdot)$ is the stability function in (8)-(9) and ν_n is the local error (15). In the following we denote by $\nu_{n,kl}$ the Fourier coefficients of ν_n, and correspondingly for parts of the local error. To emphasize the dependence on k and l, we write μ_k and μ_l instead of z_1 and z_2, respectively.

Theorem 8. *Consider an AMF-W method, and assume that the local error can be split as* $\nu_n = \nu_n^a + \nu_n^b$, *such that in Fourier space*

$$\begin{array}{lrcl} (a) & |\nu_{n,kl}^a| & \leq & C\,\tau^{p+1} |\widehat{\chi}_{kl}(t_n)| \\ (b) & |\nu_{n,kl}^b| & \leq & C\,\tau^{p} \big|R(\mu_k, \mu_l) - 1\big| |\widehat{\chi}_{kl}(t_n)| \end{array}$$

where $\mu_k = \tau a \lambda_k^{(x)}$, $\mu_l = \tau b \lambda_l^{(y)}$, *and* $\widehat{\chi}_{kl}(t_n)$ *represent the Fourier coefficients of either* $U^{(m)}(t_n)$ *or* $\varphi_2^{(m)}(t_n)$. *Moreover, we assume that the difference* $\nu_{n+1,kl}^b - \nu_{n,kl}^b$ *satisfies the same estimate as* $\nu_{n,kl}^b$ *with one additional factor* τ. *Then, under the Assumptions (a) and (c) of Theorem 1, we have convergence of order* p *in the Euclidean norm, i.e., the global error satisfies*

$$\|E_n\| \leq C_1\,\tau^p \qquad \textit{for} \qquad n\tau \leq T.$$

Proof. (a) Since $\sum_{k,l} |\widehat{\chi}_{kl}(t_n)|^2$ is bounded, a standard application of Lady Windermere's fan [9, p. 38-39], shows that we have order p (also for non-constant time step sizes).

(b) A more refined analysis (see e.g., [12, Section II.2.3]) shows that, for a constant step size application, we have order p if the local error can be written as $\nu_n^b = \big(R(\tau D_1, \tau D_2) - I_{MN}\big)\kappa_n$ or $\nu_{n,kl}^b = (R(\mu_k, \mu_l) - 1)\kappa_{n,kl}$ with $\|\kappa_n\|^2 = \sum_{k,l} \kappa_{n,kl}^2 \leq C_2 \tau^{2p} \sum_{k,l} |\widehat{\chi}_{kl}(t_n)|^2 \leq C_3 \tau^{2p}$ and $\|\kappa_{n+1} - \kappa_n\| \leq C_4 \tau^{p+1}$. But this is just what we require. □

5.3. Novel convergence theorem

The previous convergence theorem is not sufficient to prove the results stated in Section 3. We use the techniques developed in [5] to get sharp estimates for the global error. We recall that the weak step size restriction $\tau \geq c_0 \max(\Delta x^2, \Delta y^2)$ will be assumed to hold.

Theorem 9. *Consider an AMF-W method satisfying Assumption (a) of Theorem 1. Suppose that the Fourier coefficients of the local error ν_n are bounded as*

$$|\nu_{n,kl}| \leq C\,\tau^r \frac{|\mu_k|^{\alpha_1}|\mu_l|^{\alpha_2}}{\pi(\mu_k,\mu_l)^{\gamma_1}}|\widehat{\chi}_{kl}(t_n)|, \tag{31}$$

where μ_k, μ_l, $\widehat{\chi}_{kl}(t_n)$ are as in Theorem 8, and $\alpha_1, \alpha_2, \gamma_1$ are non-negative constants. Moreover, we assume that the difference $\nu_{n+1,kl} - \nu_{n,kl}$ satisfies the same estimate as $\nu_{n,kl}$ with one additional factor τ.

- *Under the Assumption (b^*) of Theorem 3 we have, for $n\tau \leq T$,*

$$\|E_n\| \leq C_1 \tau^{p^*} \qquad \textit{with} \qquad p^* = r \quad \textit{if} \quad \alpha_1 = \alpha_2 = \gamma_1 = 1. \tag{32}$$

  Recall that $p^* = r$ means that we have order $r - \epsilon$ for every $\epsilon > 0$.

- *Under the weaker Assumption (b) of Theorem 1 we still have, for $n\tau \leq T$,*

$$\|E_n\| \leq C_1 \tau^{p^*} \qquad \textit{with} \qquad p^* = \begin{cases} r - 1/3 & \textit{if} \ \ \alpha_1 = \alpha_2 = \gamma_1 = 1 \\ r - 1/2 & \textit{if} \ \ \alpha_2 = 0,\ \alpha_1 = \gamma_1 = 1 \\ r + 1/3 & \textit{if} \ \ \alpha_1 = \alpha_2 = 1,\ \gamma_1 = 2. \end{cases} \tag{33}$$

Proof. The proof follows the reasoning of [5, Section 4.2]. Using the identity

$$E_n = -(I_{MN} - R^n)(I_{MN} - R)^{-1}\nu_0 - \sum_{j=0}^{n-2}(I_{MN} - R^{n-1-j})(I_{MN} - R)^{-1}(\nu_{j+1} - \nu_j)$$

we have

$$\|E_n\| \leq C\,\tau^r a(n) + C\,\tau^{r+1}\sum_{j=0}^{n-2} a(n-1-j), \tag{34}$$

where

$$a(n) = \tau^{\alpha_1+\alpha_2}\left\{\sum_{k=1}^{N}\sum_{l=1}^{M}\Big(\frac{1-r_{kl}^n}{1-r_{kl}}\Big)^2 \frac{|\lambda_k^{(x)}|^{2\alpha_1}|\lambda_l^{(y)}|^{2\alpha_2}|\widehat{\chi}_{kl}(t_n)|^2}{(1+\theta\tau|\lambda_k^{(x)}|)^{2\gamma_1}(1+\theta\tau|\lambda_l^{(y)}|)^{2\gamma_1}}\right\}^{1/2}.$$

Here we used $\mu_k = \tau\lambda_k^{(x)}$ and $\mu_l = \tau\lambda_l^{(y)}$, ignoring the positive diffusion coefficients a and b, and $r_{kl} = R(\mu_k, \mu_l)$. It follows from [5, Lemma A.6 and Lemma A.2] that for an arbitrarily chosen $\gamma \in [0, 2]$ we have

$$\Big(\frac{1-r_{kl}^n}{1-r_{kl}}\Big)^2 \leq \tau^{-\gamma}2^{2-\gamma}T^\gamma \frac{1}{(1-r_{kl})^{2-\gamma}} \qquad \text{and} \qquad \sqrt{\lambda_k^{(x)}\lambda_l^{(y)}}\,|\widehat{\chi}_{kl}(t_j)| \leq C_2,$$

where $n\tau \leq T$. This implies

$$a(n) \leq C_3\tau^{\alpha_1+\alpha_2}\left\{\sum_{k=1}^{N}\sum_{l=1}^{M}\frac{\tau^{-\gamma}}{(1-r_{kl})^{2-\gamma}}\frac{|\lambda_k^{(x)}|^{2\alpha_1-1}|\lambda_l^{(y)}|^{2\alpha_2-1}}{(1+\theta\tau|\lambda_k^{(x)}|)^{2\gamma_1}(1+\theta\tau|\lambda_l^{(y)}|)^{2\gamma_1}}\right\}^{1/2}. \tag{35}$$

Now we use Assumption (b^*) and the inequality (26) of Lemma 7. We thus get the estimate

$$a(n) \leq C_4\tau^{\alpha_1+\alpha_2}\left\{\tau^{-2}\sum_{k=1}^{N}\sum_{l=1}^{M}\frac{|\lambda_k^{(x)}|^{2\alpha_1-1}|\lambda_l^{(y)}|^{2\alpha_2-1}}{(1+\theta\tau|\lambda_k^{(x)}|)^{2\gamma_1-2+\gamma}(1+\theta\tau|\lambda_l^{(y)}|)^{2\gamma_1-2+\gamma}(|\lambda_k^{(x)}|+|\lambda_l^{(y)}|)^{2-\gamma}}\right\}^{1/2}. \tag{36}$$

With the help of the arithmetic-geometric mean inequality $|a| + |b| \geq 2|ab|^{1/2}$ we get

$$a(n) \leq C_5 \tau^{\alpha_1+\alpha_2} \left\{ \tau^{-2} \sum_{k=1}^{N} \sum_{l=1}^{M} \frac{|\lambda_k^{(x)}|^{2\alpha_1-2+\gamma/2} |\lambda_l^{(y)}|^{2\alpha_2-2+\gamma/2}}{(1+\theta\tau|\lambda_k^{(x)}|)^{2\gamma_1-2+\gamma}(1+\theta\tau|\lambda_l^{(y)}|)^{2\gamma_1-2+\gamma}} \right\}^{1/2}. \tag{37}$$

Since the double sum can be written as a product of two single sums, we are in the position to apply Lemma A.5 from [5], which states: for all $\alpha \geq 0$ and $\beta \geq 0$ there exists a constant $C > 0$, such that

$$\sum_k \frac{|\lambda_k|^{\alpha/2}}{(1+\theta\tau|\lambda_k|)^\beta} \leq C\tau^{-(\alpha+1)/2} \qquad \text{if} \qquad \alpha + 1 - 2\beta < 0, \tag{38}$$

where λ_k stands for $\lambda_k^{(x)}$ or $\lambda_k^{(y)}$. For the situation $\alpha_1 = \alpha_2 = \gamma_1 = 1$, the double sum in (37) is the product of two identical single sums of the form (38) with $\alpha = \gamma$ and $\beta = \gamma$. With the choice $\gamma = 1 + 2\epsilon$, where $\epsilon > 0$ can be arbitrarily small, each single sum is bounded by $C\tau^{-1-\epsilon}$. Consequently, we have that $a(n) \leq C_6\tau^{-\epsilon}$. This proves the statement (32) under the Assumption (b^*).

We next consider Assumption (b) of Theorem 1, so that we have to work with (27) of Lemma 7. Starting with (35) and using (27) the same computation as above leads to

$$a(n) \leq C_6 \tau^{\alpha_1+\alpha_2} \left\{ \tau^{-2} \sum_{k=1}^{N} \sum_{l=1}^{M} \frac{|\lambda_k^{(x)}|^{2\alpha_1-2+\gamma/2} |\lambda_l^{(y)}|^{2\alpha_2-2+\gamma/2}}{(1+\theta\tau|\lambda_k^{(x)}|)^{2\gamma_1-4+2\gamma}(1+\theta\tau|\lambda_l^{(y)}|)^{2\gamma_1-4+2\gamma}} \right\}^{1/2}. \tag{39}$$

This is again a product of two factors of the form (38).

For $\alpha_1 = \alpha_2 = \gamma_1 = 1$ we have two identical factors with $\alpha = \gamma$ and $\beta = 2\gamma - 2$. Inserting the values of α and β into (38) we get $\gamma > 5/3$ and $(\alpha + 1)/2 = 4/3 + \epsilon$. This yields the statement of the theorem.

For $\alpha_1 = \alpha_2 = 1$, $\gamma_1 = 2$ we also have two identical factors, this time with $\alpha = \gamma$ and $\beta = 2\gamma$. Inserted into (38) we get $\gamma > 1/3$ and $(\alpha + 1)/2 = 2/3 + \epsilon$. This yields the statement of the theorem.

We finally consider the case $\alpha_1 = \gamma_1 = 1$, $\alpha_2 = 0$. Instead of the arithmetic-geometric mean inequality we use the trivial inequality $|\lambda_k^{(x)}| + |\lambda_l^{(y)}| \geq |\lambda_k^{(x)}|$ and thus obtain

$$a(n) \leq C_5 \tau^{\alpha_1+\alpha_2} \left\{ \tau^{-2} \sum_{k=1}^{N} \sum_{l=1}^{M} \frac{|\lambda_k^{(x)}|^{2\alpha_1-3+\gamma} |\lambda_l^{(y)}|^{-1}}{(1+\theta\tau|\lambda_k^{(x)}|)^{2\gamma_1-4+2\gamma}(1+\theta\tau|\lambda_l^{(y)}|)^{2\gamma_1-4+2\gamma}} \right\}^{1/2}. \tag{40}$$

Now we have one factor with $\alpha = 2(\gamma - 1)$ and $\beta = 2\gamma - 2$. The other factor is bounded by $\sum_l |\lambda_l^{(y)}|^{-1}$, which is $\mathcal{O}(1)$ by [12, Lemma 6.2, p. 298]. Inserting the values of α and β into (38) we get $\gamma > 3/2$ and $(\alpha + 1)/2 = 1 + \epsilon$. This shows that $a(n) = \mathcal{O}(\tau^{-1/2-\epsilon})$ and proves the statement of the theorem. □

Remark 3. We note that under the step size restriction $\tau \leq c_1 h$ with $h = \min(\Delta x, \Delta y)$ the order statements (32) and (33) hold with p (and not only with p^*). Instead of (38) we have to use the bound

$$\sum_k \frac{|\lambda_k|^{\alpha/2}}{(1+\theta\tau|\lambda_k|)^\beta} \leq C\tau^{-\beta} h^{2\beta-\alpha-1} \qquad \text{if} \qquad \alpha + 1 - 2\beta > 0, \tag{41}$$

which also follows from [5, Lemma A.5]. For example, in the situation of (32) we have $\alpha = \gamma$ and $\beta = \gamma$, so that the expression (41) is bounded by $C\tau^{-1}(\tau/h)^{1-\gamma}$ for every $\gamma < 1$, and we have $a(n) \leq C_7(\tau/h)^{1-\gamma}$. The same argument can be applied to the cases of (33).

6. Proof of Theorem 3

This section is devoted to the proof of Theorem 3 and to an explanation of the remarks related to this theorem. We recall that Theorem 3 is concerned with general boundary conditions.

6.1. Order $p = 1$

In the local error (23) we consider the first term ($m = 1$)

$$\delta_1 := \tau\Big(1 - b^\top S(z_1, z_2)\big(\alpha^{(1)} + \beta^{(1)}(z_1 + z_2) + \gamma^{(1)} z_1 z_2\big)\Big)\widehat{U}_{kl}^{(1)}(t_n). \tag{42}$$

All other terms are of size $\mathcal{O}(\tau^2)$, more precisely, of type (a) in Theorem 8. Using the identity

$$S(z_1, z_2)^{-1}\rho = \mathbf{1} - (z_1 + z_2)(\theta\rho + c) + z_1 z_2 \theta^2 \rho = \alpha^{(1)} + \beta^{(1)}(z_1 + z_2) + \gamma^{(1)} z_1 z_2,$$

which follows from the definition of the vectors ρ and c (Section 2) and from (20), we obtain

$$\delta_1 = \tau\big(1 - b^\top \rho\big)\widehat{U}_{kl}^{(1)}(t_n). \tag{43}$$

Since $b^\top \rho = \tilde{b}^\top \mathbf{1} = 1$, this proves the first statement of Theorem 3.

6.2. Order $p = 2$

We have to consider the coefficient of τ^2 in (23). Because of $\|S(z_1, z_2)z_1\| \le C|z_1 + z_2|/\pi(z_1, z_2)$ and (26) the coefficient of $\widehat{\varphi}_{2,kl}^{(1)}(t_n)$ is of type (b) in Theorem 8 with $p = 2$.

We next consider the term with $\tau^2 \widehat{U}_{kl}^{(2)}(t_n)$ in (23), which is

$$\delta_2 := \tau^2\Big(\frac{1}{2} - b^\top S(z_1, z_2)\big(\alpha^{(2)} + \beta^{(2)}(z_1 + z_2) + \gamma^{(2)} z_1 z_2\big)\Big)\widehat{U}_{kl}^{(2)}(t_n).$$

We note that for $z_1 = z_2 = 0$ the expression in the large brackets becomes

$$\frac{1}{2} - b^\top S(0,0)\alpha^{(2)} = \frac{1}{2} - b^\top (I_s - L)^{-1}(c + \theta\rho) = \frac{1}{2} - \tilde{b}^\top(\tilde{A} + \tilde{\Gamma})\mathbf{1} = 0, \tag{44}$$

which vanishes by the condition of order two. Now, subtracting (44), multiplied by $\tau^2\,\widehat{U}_{kl}^{(2)}(t_n)$, from δ_2 and using (28) thus yields $\delta_2 = \delta_2^a + \delta_2^b$ with

$$\begin{aligned} \delta_2^a &= -\tau^2 b^\top S(z_1, z_2)(z_1 + z_2)\big((\theta I_s + A)(I_s - L)^{-1}\alpha^{(2)} + \beta^{(2)}\big)\widehat{U}_{kl}^{(2)}(t_n) \\ \delta_2^b &= -\tau^2 b^\top S(z_1, z_2) z_1 z_2 \big(\gamma^{(2)} - \theta^2 (I_s - L)^{-1}\alpha^{(2)}\big)\widehat{U}_{kl}^{(2)}(t_n). \end{aligned} \tag{45}$$

Because of (25) and (26) the term δ_2^a is of type (b) in Theorem 8 and gives a $\mathcal{O}(\tau^2)$ contribution to the global error. Concerning the term δ_2^b we note that it is of the form (31), so that Theorem 9 can be applied. Under the assumption (b^*) we thus have order $p = 2^*$. This completes the proof of Theorem 3. □

6.3. Comments on Remark 1

To prove order $p = 2$ and not only $p = 2^*$ we have two possibilities. Either we assume the step size restriction $\tau \le c_1 \min(\Delta x, \Delta y)$ with some $c_1 > 0$ (see Remark 3) or we assume the additional order condition $\eta = 1/2$. This can be seen as follows: with the aim of applying (30) in δ_2^b we consider the expression

$$b^\top \big(\gamma^{(2)} - \theta^2 (I_s - L)^{-1}\alpha^{(2)}\big) = \theta^2 \eta\, \tilde{b}^\top \mathbf{1} - \theta^2\, \tilde{b}^\top (\tilde{A} + \tilde{\Gamma})\mathbf{1} = \theta^2(\eta - 1/2),$$

which vanishes for $\eta = 1/2$. Using (30) we get

$$\delta_2^b = -\tau^2 b^\top \frac{S(z_1, z_2)}{\pi(z_1, z_2)}\Big((z_1 + z_2)A + L\Big) z_1 z_2 \big(\gamma^{(2)} - \theta^2 (I_s - L)^{-1}\alpha^{(2)}\big)\widehat{U}_{kl}^{(2)}(t_n). \tag{46}$$

The estimate (25) implies that the coefficient of $\widehat{U}_{kl}^{(2)}(t_n)$ is bounded by $C\tau^2|z_1 + z_2|/\pi(z_1, z_2)$, so that item (b) of Theorem 8 can be applied. This completes the proof of order $p = 2$. □

6.4. Comments on Remark 2

Instead of Assumption (b^*) we consider the weaker Assumption (b). We have $\delta_1 = 0$ so that order $p = 1$ is still obtained. The term δ_2^b is of the form (31) with $\alpha_1 = \alpha_2 = \gamma_1 = 1$ and $r = 2$. This contributes to the global error with a term of order $p^* = r - 1/3$. We split the remaining term δ_2^a into one that is of the form (31) with $\alpha_1 = \gamma_1 = 1$, $\alpha_2 = 0$ and $r = 2$, and another with $\alpha_2 = \gamma_1 = 1$, $\alpha_1 = 0$ and $r = 2$. By symmetry, both terms are equivalent and contribute with an error term of order $p^* = r - 1/2$. Furthermore, since $\|S(z_1, z_2)z_1\| \le C|z_1 + z_2|/\pi(z_1, z_2)$, the coefficient of $\widehat{\varphi}^{(1)}_{2,kl}(t_n)$ in (23) produces a contribution to the global error of size $\mathcal{O}(\tau^{p^*})$, with $p^* = 1.5$, according to (33). We thus have order $p^* = 1.5$.

Under the step size restriction $\tau \le c_1 \min(\Delta x, \Delta y)$ with some $c_1 > 0$ we get order $p = 1.5$ instead of $p = 1.5 - \epsilon$ for every $\epsilon > 0$ (see Remark 3).

7. Proof of Theorem 1

In the situation of time-independent Dirichlet boundary conditions the function $g_2(t)$ is constant, so that $\varphi_2^{(m)}(t) = D_2 U^{(m)}(t)$ for $m \ge 1$. This implies that the term involving $\dot{g}_2(t)$ vanishes in (17), so that the two terms with factor $\gamma_j^{(m)}$ in (19) cancel. For the Fourier coefficients of the local error we thus have

$$\nu_{n,kl} = \sum_{m\ge 1} \tau^m \Big(\frac{1}{m!} - b^\top S(z_1, z_2)\big(\alpha^{(m)} + \beta^{(m)}(z_1 + z_2)\big)\Big) \widehat{U}^{(m)}_{kl}(t_n). \tag{47}$$

It should be noted that (47) is just obtained from (23) by setting $\gamma^{(m)} = 0$, $m \ge 1$. Moreover, we have that in the Euclidean norm (12), for $m \ge 1$,

$$\begin{aligned} DU^{(m)}(t) &= U^{(m+1)}(t) - g^{(m)}(t) = \mathcal{O}(1), \\ D_1 D_2 U^{(m)}(t) &= \big(a\, \partial^2_{xx} + \mathcal{O}(\Delta x^2)\big)\big(b\, \partial^2_{yy} + \mathcal{O}(\Delta y^2)\big) \frac{\partial^m u(t,x,y)}{\partial t^m} = \mathcal{O}(1). \end{aligned} \tag{48}$$

In Fourier coefficients this reads

$$\begin{aligned} (z_1 + z_2)\, \widehat{U}^{(m)}_{kl}(t) &= \tau \langle DU^{(m)}(t), \phi^{(y)}_l \otimes \phi^{(x)}_k \rangle = \tau \widehat{\chi}_{1,kl}(t), \qquad \text{for } m \ge 1, \\ z_1 z_2\, \widehat{U}^{(m)}_{kl}(t) &= \tau^2 \langle D_1 D_2 U^{(m)}(t), \phi^{(y)}_l \otimes \phi^{(x)}_k \rangle = \tau^2 \widehat{\chi}_{2,kl}(t), \qquad \text{for } m \ge 1, \end{aligned} \tag{49}$$

where $\chi_1(t) = DU^{(m)}(t)$ and $\chi_2(t) = D_1 D_2 U^{(m)}(t)$.

7.1. Order $p = 1$

The term with $m = 1$ in (47), which we denote by δ_1, is equal to (42) with $\gamma^{(1)}$ replaced by zero. The computation of Section 6.1 shows that under the assumption $\tilde{b}^\top \mathbf{1} = 1$ it is equal to

$$\delta_1 = \tau b^\top S(z_1, z_2) \theta^2 \rho\, z_1 z_2 \widehat{U}^{(1)}_{kl}(t), \tag{50}$$

which is of size $\mathcal{O}(\tau^3)$ by (49). Theorem 8, item (a), therefore proves that δ_1 leads to a $\mathcal{O}(\tau^2)$ contribution in the global error.

7.2. Order $p = 2$

We just showed that δ_1 leads to an order 2 term in the global error. The term with $m = 2$ is given by (45) with $\gamma^{(2)}$ replaced by zero. It is $\delta_2 = \delta_2^a + \delta_2^b$, where

$$\begin{aligned} \delta_2^a &= -\tau^2 b^\top S(z_1, z_2)\big((\theta I + A)(I_s - L)^{-1}\alpha^{(2)} + \beta^{(2)}\big)(z_1 + z_2)\widehat{U}^{(2)}_{kl}(t_n) \\ \delta_2^b &= \tau^2 b^\top S(z_1, z_2) \theta^2 (I_s - L)^{-1} \alpha^{(2)} z_1 z_2\, \widehat{U}^{(2)}_{kl}(t_n). \end{aligned} \tag{51}$$

By (49), δ_2^a and δ_2^b are of size $\mathcal{O}(\tau^3)$ and $\mathcal{O}(\tau^4)$, respectively, and Theorem 8, item (a), can be applied.

7.3. Order $p = 3$

We consider the terms $m = 1, 2, 3$ in (47) separately.

The term with $m = 1$ is given by (50) and it was shown to be of size $\mathcal{O}(\tau^3)$. To eliminate the part that cannot be put into type (b) of Theorem 8, we require that

$$0 = b^\top S(0,0)\rho = b^\top (I_s - L)^{-1}\rho = \tilde{b}^\top \rho = \theta^{-1}\tilde{b}^\top \tilde{\Gamma}\mathbf{1}. \tag{52}$$

This relation together with (29) yields

$$\delta_1 = \tau b^\top \Big(S(z_1, z_2) - \frac{S(0,0)}{\pi(z_1,z_2)} \Big) \theta^2 \rho z_1 z_2 \widehat{U}_{kl}^{(1)}(t_n) \; = \; \delta_1^a + \delta_1^b$$

with

$$\begin{aligned} \delta_1^a &= \tau b^\top \frac{S(z_1,z_2)}{\pi(z_1,z_2)} (z_1+z_2)(A+\theta L)(I_s - L)^{-1}\theta^2 \rho z_1 z_2 \widehat{U}_{kl}^{(1)}(t_n), \\ \delta_1^b &= -\tau b^\top \frac{S(z_1,z_2)}{\pi(z_1,z_2)} z_1 z_2 \theta^2 L (I_s - L)^{-1}\theta^2 \rho z_1 z_2 \widehat{U}_{kl}^{(1)}(t_n). \end{aligned} \tag{53}$$

Using (49) and (27) the term δ_1^a is seen to be of type (b) in Theorem 8 with $p = 3$. To the term δ_1^b we can apply Theorem 9 with $\alpha_1 = \alpha_2 = 1$ and $\gamma_1 = 2$, which gives an order $3 + 1/3 - \epsilon$ for every $\epsilon > 0$.

The term with $m = 2$ is $\delta_2 = \delta_2^a + \delta_2^b$ with summands given in (51). It has been shown above that δ_2^b leads to a $\mathcal{O}(\tau^3)$ contribution of the global error. Regarding δ_2^a we note that the order conditions imply

$$\begin{aligned} & b^\top S(0,0)\big((\theta I_s + A)(I_s - L)^{-1}\alpha^{(2)} + \beta^{(2)}\big) = \tilde{b}^\top \big((\tilde{A} + \tilde{\Gamma})\alpha^{(2)} + \beta^{(2)}\big) \\ & \quad = \tilde{b}^\top \Big((\tilde{A} + \tilde{\Gamma})^2 \mathbf{1} - \frac{1}{2}c^2 - \theta\eta\rho \Big) = \frac{1}{6} - \frac{1}{2}\cdot\frac{1}{3} - 0 = 0. \end{aligned} \tag{54}$$

Hence, we have

$$\delta_2^a = -\tau^2 b^\top \Big(S(z_1,z_2) - \frac{S(0,0)}{\pi(z_1,z_2)} \Big) \mu^{(2)} (z_1 + z_2)\, \widehat{U}_{kl}^{(2)}(t_n),$$

where $\mu^{(2)} = (\tilde{A} + \tilde{\Gamma})\alpha^{(2)} + \beta^{(2)}$. With the identity (29) we thus obtain $\delta_2^a = \delta_2^{a,1} + \delta_2^{a,2}$, where

$$\begin{aligned} \delta_2^{a,1} &= -\tau^2 b^\top \frac{S(z_1,z_2)}{\pi(z_1,z_2)} (A+\theta L)(I_s - L)^{-1}\mu^{(2)} (z_1+z_2)^2\, \widehat{U}_{kl}^{(2)}(t_n), \\ \delta_2^{a,2} &= \tau^2 b^\top \frac{S(z_1,z_2)}{\pi(z_1,z_2)} \theta^2 L (I_s - L)^{-1}\mu^{(2)} (z_1+z_2) z_1 z_2\, \widehat{U}_{kl}^{(2)}(t_n). \end{aligned} \tag{55}$$

Using (49) both terms are seen to be of type (b) in Theorem 8, $\delta_2^{a,1}$ with $p = 3$ and $\delta_2^{a,2}$ with $p = 4$. They give $\mathcal{O}(\tau^3)$ and $\mathcal{O}(\tau^4)$ contributions to the global error.

For $m = 3$ we have that the error is given by $\delta_3 = \delta_3^a + \delta_3^b$ with

$$\delta_3^a = \tau^3 \Big(\frac{1}{3!} - b^\top S(z_1,z_2)\alpha^{(3)} \Big) \widehat{U}_{kl}^{(3)}(t_n), \qquad \delta_3^b = -\tau^3 b^\top S(z_1,z_2)\beta^{(3)}(z_1+z_2)\, \widehat{U}_{kl}^{(3)}(t_n). \tag{56}$$

The term δ_3^b is of size $\mathcal{O}(\tau^4)$ and Theorem 8, item (a), can be applied. From the order conditions we deduce that

$$\frac{1}{3!} - b^\top S(0,0)\alpha^{(3)} = \frac{1}{3!} - \tilde{b}^\top \Big(\frac{c^2}{2} + \theta\eta\rho \Big) = 0,$$

because $\tilde{b}^\top \rho = 0$ by (52). This relation simplifies δ_3^a to

$$\begin{aligned} \delta_3^a &= -\tau^3 b^\top \big(S(z_1,z_2) - S(0,0) \big)\alpha^{(3)}\, \widehat{U}_{kl}^{(3)}(t_n) \\ &= -\tau^3 b^\top S(z_1,z_2)\big((z_1+z_2)(\theta I_s + A) - z_1 z_2 \theta^2 I \big)(I_s - L)^{-1}\alpha^{(3)}\, \widehat{U}_{kl}^{(3)}(t_n) \end{aligned} \tag{57}$$

by (28). The estimates (49) together with the boundedness of $S(z_1, z_2)$ imply that all expressions are of type (a) with $p = 3$. This completes the proof of Theorem 1. □

8. Proof of Theorem 5 - fractional order $p = 2.25$

This section treats general boundary conditions and uses Assumption (b^*) of Theorem 3. To get convergence of order $p = 2.25$ (in fact $p = 2 + \alpha$ with $\alpha < 1/4$) we need the conditions of order for $p = 3$ (see Section 2) and also $\eta = 1/2$.

The analysis is based on [12, Lemma III.6.5] which states that for a $\mathcal{C}^2$ function $\chi(x, y)$ and using the notation $\mu_k = \tau a \lambda_k^{(x)}$ and $\mu_l = \tau b \lambda_l^{(y)}$ as in Theorem 8 we have for every α, $0 < \alpha < 1/4$,

$$(|\mu_k|^\alpha + |\mu_l|^\alpha)|\widehat{\chi}_{kl}(t)| \le \tau^\alpha u_{kl} \qquad \text{with} \qquad \sum_{k=1}^{N}\sum_{l=1}^{M} |u_{kl}|^2 \le C, \tag{58}$$

where the constant C is independent of N and M. In the following we again use the notation z_1 and z_2 for μ_k and μ_l, respectively. The local error (23) is a sum $\sum_{m\ge 1}(\delta_m + \omega_m)$, where

$$\begin{aligned} \delta_m &= \tau^m \Big(\frac{1}{m!} - b^\top S(z_1, z_2)\Big(\alpha^{(m)} + \beta^{(m)}(z_1 + z_2) + \gamma^{(m)} z_1 z_2\Big)\Big) \widehat{U}_{kl}^{(m)}(t_n), \\ \omega_m &= \tau^{m+1} b^\top S(z_1, z_2)\gamma^{(m)} z_1 \widehat{\varphi}_{2,kl}^{(m)}(t_n). \end{aligned} \tag{59}$$

By (43) the order condition $\tilde{b}^\top \mathbf{1} = 1$ implies $\delta_1 = 0$. The terms δ_m for $m \ge 4$ and ω_m for $m \ge 2$ have a sufficiently high power of τ, so that they give a $\mathcal{O}(\tau^3)$ contribution to the global error (by item (a) of Theorem 8 for δ_m and by item (b) for ω_m). The terms δ_2, δ_3, and ω_1 remain to be considered.

8.1. *Local error term ω_1*

Recall from (52) that $b^\top S(0,0)\gamma^{(1)} = \theta^2 \tilde{b}^\top \rho = 0$. With the aim of using (29) we subtract the expression $\tau^2 b^\top \big(S(0,0)/\pi(z_1, z_2)\big)\gamma^{(1)} z_1 \widehat{\varphi}_{2,kl}^{(1)}(t_n)$ from ω_1 and obtain

$$\omega_1 = \tau^2 b^\top \frac{S(z_1, z_2)}{\pi(z_1, z_2)} \Big((z_1 + z_2)(A + \theta L) - z_1 z_2 \theta^2 L\Big)(I_s - L)^{-1}\gamma^{(1)} z_1 \widehat{\varphi}_{2,kl}^{(1)}(t_n).$$

Here, we use the bound (58) for $|z_1|^\alpha |\varphi_{2,kl}^{(1)}(t_n)|$. Because of (25) the remaining factor is bounded by $|z_1 + z_2|/\pi(z_1, z_2)$. By Theorem 8 this term gives a $\mathcal{O}(\tau^{2+\alpha})$ contribution in the global error (with $\alpha < 1/4$).

8.2. *Local error term δ_2*

We have $\delta_2 = \delta_2^a + \delta_2^b$ with δ_2^a and δ_2^b from (45). Using the order conditions for $p = 3$ we obtain $\delta_2^a = \delta_2^{a,1} + \delta_2^{a,2}$ with $\delta_2^{a,1}$ and $\delta_2^{a,2}$ from (55). We now apply

$$|z_1 + z_2| \le \big(|z_1|^{1-\alpha} + |z_2|^{1-\alpha}\big)\big(|z_1|^\alpha + |z_2|^\alpha\big), \tag{60}$$

use the bound (58), and deduce that both $\delta_2^{a,1}$ and $\delta_2^{a,2}$ are bounded by $C\tau^{2+\alpha}|z_1 + z_2|/\pi(z_1, z_2)$. This permits us to apply Theorem 8 and to conclude.

Under the assumption $\eta = 1/2$ the term δ_2^b of (45) can be written in the form (46). As before, the bound (58) permits us to prove $|\delta_2^b| \le C\tau^{2+\alpha}|z_1 + z_2|/\pi(z_1, z_2)$, so that Theorem 8, item (b), yields the statement.

8.3. *Local error term δ_3*

We write the error term as $\delta_3 = \delta_3^a + \delta_3^b$, where

$$\delta_3^a = \tau^3\Big(\frac{1}{3!} - b^\top S(z_1, z_2)\alpha^{(3)}\Big)\widehat{U}_{kl}^{(3)}(t_n), \qquad \delta_3^b = -\tau^3 b^\top S(z_1, z_2)\Big(\beta^{(3)}(z_1 + z_2) + \gamma^{(3)} z_1 z_2\Big)\widehat{U}_{kl}^{(3)}(t_n).$$

The term with factor $\beta^{(3)}$ satisfies the Assumption (b) of Theorem 8 and leads to a $\mathcal{O}(\tau^3)$ contribution of the global error. For the term with factor $\gamma^{(3)}$ we split $-z_2 = |z_2|$ as $|z_2|^{1-\alpha}|z_2|^\alpha$, we use (58) and the fact that $\|S(z_1,z_2)\||z_1||z_2|^{1-\alpha}$ is bounded by (25). Hence, item (a) of Theorem 8 is satisfied with $p = 2+\alpha$, giving a $\mathcal{O}(\tau^{2+\alpha})$ contribution to the global error.

The computation of Section 7.3 right after formula (56) shows that

$$\delta_3^a = -\tau^3 b^\top S(z_1,z_2)\big((z_1+z_2)(\theta I_s + A) - z_1 z_2 \theta^2 I\big)(I_s - L)^{-1}\alpha^{(3)}\,\widehat{U}_{kl}^{(3)}(t_n).$$

We are now in precisely the same situation as before. There is one term with factor $\tau^3 S(z_1,z_2)(z_1+z_2)$, and another with factor $\tau^3 S(z_1,z_2)z_1z_2$. Therefore, also δ_3^a leads to $\mathcal{O}(\tau^3)$ and $\mathcal{O}(\tau^{2+\alpha})$ contributions of the global error. □

9. Proof of Theorem 6 - fractional order $p = 3.25$

This section considers time-independent boundary conditions and uses Assumption (b) of Theorem 1. If, in addition to the order conditions for $p = 3$ (see Section 2) the conditions (14) are satisfied, we prove convergence of order $p = 3+\alpha$ for every $\alpha < 1/4$. Combining the estimates (49) with those of [12, Lemma III.6.5] we obtain for every α, $0 < \alpha < 1/4$, that

$$(|\mu_k|^\alpha + |\mu_l|^\alpha)|\mu_k + \mu_l||\widehat{U}_{kl}^{(m)}(t)| \le \tau^{1+\alpha} v_{kl} \tag{61}$$

$$(|\mu_k|^\alpha + |\mu_l|^\alpha)|\mu_k\mu_l||\widehat{U}_{kl}^{(m)}(t)| \le \tau^{2+\alpha} w_{kl} \tag{62}$$

with bounded $\sum_{kl}|v_{kl}|^2$ and $\sum_{kl}|w_{kl}|^2$. As before, we write z_1 and z_2 for μ_k and μ_l, respectively. For time-independent boundary conditions the local error (47) is $\sum_{m\ge 1}\delta_m$, where δ_m is given by (59) with $\gamma^{(m)}$ replaced by zero.

9.1. Local error term δ_1.

From the computation in Section 7.3 we have $\delta_1 = \delta_1^a + \delta_1^b$ with δ_1^a and δ_1^b from (53). Assuming

$$0 = b^\top S(0,0)(A+\theta L)(I_s - L)^{-1}\theta\rho = \tilde{b}^\top(\tilde{A}+\tilde{\Gamma})\,\tilde{\Gamma}\mathbf{1},$$

(the second equality follows from (7) and (52)) which is the first condition of (14), we get

$$\begin{aligned}\delta_1^a &= -\tau b^\top \frac{\big(S(z_1,z_2) - S(0,0)/\pi(z_1,z_2)\big)}{\pi(z_1,z_2)}\mu\,(z_1+z_2)z_1z_2\widehat{U}_{kl}^{(1)}(t)\\ &= -\tau b^\top \frac{S(z_1,z_2)}{\pi(z_1,z_2)^2}\Big((z_1+z_2)(A+\theta L) - z_1z_2\theta^2 L\Big)(I_s-L)^{-1}\mu\,(z_1+z_2)z_1z_2\widehat{U}_{kl}^{(1)}(t)\end{aligned}$$

where we apply the notation $\mu = (A+\theta L)(I_s-L)^{-1}\theta^2\rho$. Using (62) and the estimate (25), this expression can be bounded by $C\tau^{3+\alpha}|z_1+z_2|/\pi(z_1,z_2)^2$. Therefore, by (27) and Theorem 8, δ_1^a leads to a $\mathcal{O}(\tau^{3+\alpha})$ term in the global error.

The term δ_1^b of the local error has been shown in Section 7.3 to be of order $3+1/3-\epsilon$ (for every $\epsilon > 0$). This is better than order 3.25.

9.2. Local error term δ_2.

We use the splitting $\delta_2 = \delta_2^{a,1} + \delta_2^{a,2} + \delta_2^b$ from (51) and (55), and we start by considering $\delta_2^{a,2}$. It has been shown in Section 7.3 to contribute as $\mathcal{O}(\tau^4)$ to the global error.

With the aim of applying (29) to the term $\delta_2^{a,1}$ we consider the expression

$$\begin{aligned}&b^\top S(0,0)(A+\theta L)(I_s-L)^{-1}\mu^{(2)} = \tilde{b}^\top(\tilde{A}+\tilde{\Gamma}-\theta I)\mu^{(2)}\\ &= \tilde{b}^\top(\tilde{A}+\tilde{\Gamma})^3\mathbf{1} - \tilde{b}^\top(\tilde{A}+\tilde{\Gamma})\frac{c^2}{2} - \eta\,\tilde{b}^\top(\tilde{A}+\tilde{\Gamma})\tilde{\Gamma}\mathbf{1} = \frac{1}{24} - \frac{1}{2}\cdot\frac{1}{12} - 0 = 0,\end{aligned}$$

which vanishes because of $\tilde{b}^\top \mu^{(2)} = 0$ (see (54)) and the first, second, and fourth conditions of (14). The application of (29) now gives

$$\delta_2^{a,1} = -\tau^2 b^\top \frac{S(z_1,z_2)}{\pi(z_1,z_2)^2}\Big((z_1+z_2)(A+\theta L) - z_1 z_2 \theta^2 L\Big)(I_s - L)^{-1}(A+\theta L)(I_s-L)^{-1}\mu^{(2)}(z_1+z_2)^2\, \widehat{U}_{kl}^{(2)}(t_n),$$

which by (60) and (61) can be bounded by $C\tau^{3+\alpha}|z_1+z_2|/\pi(z_1,z_2)^2$, so that item (b) of Theorem 8 can be applied.

To get a desired estimate for δ_2^b from (51) we consider

$$b^\top S(0,0)\theta^2(I_s-L)^{-1}\alpha^{(2)} = \theta\, \tilde{b}^\top \tilde{\Gamma}(c+\theta\rho) = \theta \tilde{b}^\top \tilde{\Gamma}(\tilde{A}+\tilde{\Gamma})\mathbf{1} = 0$$

which vanishes by the third condition of (14). An application of (29) thus yields

$$\delta_2^b = \tau^2 b^\top \frac{S(z_1,z_2)}{\pi(z_1,z_2)}\Big((z_1+z_2)(A+\theta L) - z_1 z_2\theta^2 L\Big)(I_s-L)^{-1}\theta^2(I_s-L)^{-1}\alpha^{(2)} z_1 z_2\, \widehat{U}_{kl}^{(2)}(t_n)$$

which, by (49), can be bounded by $C\tau^{4+\alpha}$, so that item (a) of Theorem 8 can be applied.

9.3. Local error term δ_3.

By (56) and (57) the local error term δ_3 can be written as $\delta_3 = \delta_3^c + \delta_3^d$, where

$$\begin{aligned}
\delta_3^c &= -\tau^3 b^\top S(z_1,z_2)\big((\theta I_s + A)(I_s-L)^{-1}\alpha^{(3)} + \beta^{(3)}\big)(z_1+z_2)\, \widehat{U}_{kl}^{(3)}(t_n),\\
\delta_3^d &= \tau^3 b^\top S(z_1,z_2)\theta^2(I_s-L)^{-1}\alpha^{(3)} z_1 z_2\, \widehat{U}_{kl}^{(3)}(t_n).
\end{aligned}$$

Using (49), the term δ_3^d is bounded by $\mathcal{O}(\tau^5)$ and gives a $\mathcal{O}(\tau^4)$ contribution to the global error by Theorem 8, item (a). For the term δ_3^c we consider the expression

$$b^\top S(0,0)\big((\theta I_s + A)(I_s-L)^{-1}\alpha^{(3)} + \beta^{(3)}\big) = \tilde{b}^\top\Big((\tilde{A}+\tilde{\Gamma})\Big(\frac{c^2}{2}+\theta\eta\rho\Big) - \Big(\frac{c^3}{6}+\frac{\theta\eta^2\rho}{2}\Big)\Big) = \frac{1}{2\cdot 12} - \frac{1}{6\cdot 4} = 0,$$

which vanishes as a consequence of the first, fourth, and fifth conditions of (14), and of (52). With the help of (29) we thus get

$$\delta_3^c = -\tau^3 b^\top \frac{S(z_1,z_2)}{\pi(z_1,z_2)}\Big((z_1+z_2)(A+\theta L) - z_1z_2\theta^2 L\Big)(I_s-L)^{-1}\Big((\theta I_s+A)(I_s-L)^{-1}\alpha^{(3)}+\beta^{(3)}\Big)(z_1+z_2)\, \widehat{U}_{kl}^{(3)}(t_n).$$

Using the upper relation of (49) yields the bound $C\tau^{4+\alpha}$ for δ_3^c, so that this gives a $\mathcal{O}(\tau^{3+\alpha})$ contribution of the global error by Theorem 8, item (a).

9.4. Local error term δ_4.

We have

$$\delta_4 = \tau^4\Big(\frac{1}{4!} - b^\top S(z_1,z_2)\Big(\alpha^{(4)} + \beta^{(4)}(z_1+z_2)\Big)\Big)\widehat{U}_{kl}^{(4)}(t_n).$$

As a consequence of (49) the term with factor $\beta^{(4)}$ is bounded by $\mathcal{O}(\tau^5)$. It gives rise to a contribution of size $\mathcal{O}(\tau^4)$ in the global error. The remaining expression, which we denote by δ_4^a, can be treated with (28) and yields

$$\delta_4^a = \tau^4\Big(\frac{1}{4!} - b^\top S(z_1,z_2)\alpha^{(4)}\Big)\widehat{U}_{kl}^{(4)}(t_n) = \tau^4\Big(\frac{1}{4!} - b^\top S(0,0)\alpha^{(4)} - b^\top\big(S(z_1,z_2)-S(0,0)\big)\alpha^{(4)}\Big)\widehat{U}_{kl}^{(4)}(t_n).$$

The fifth condition of (14) and (52) imply $b^\top S(0,0)\alpha^{(4)} = 1/4!$, so that by (28) this term becomes

$$\delta_4^a = -\tau^4 b^\top\Big(S(z_1,z_2)\big((z_1+z_2)(\theta I_s + A) - z_1z_2\theta^2 I_s\big)(I_s-L)^{-1}\alpha^{(4)}\Big)\widehat{U}_{kl}^{(4)}(t_n).$$

The Assumption (49) now leads to global error terms of size $\mathcal{O}(\tau^4)$ and $\mathcal{O}(\tau^5)$ by applying Theorem 8, item (a). □

10. Numerical illustration

In order to numerically illustrate the sharpness of the orders of PDE-convergence of Section 3, we consider the linear diffusion partial differential equation with constant coefficients (1) for $(x, y) \in (0, 1)^2$, $t \in [0, 1]$, where $g(t, x, y)$ is selected in such way that

$$u(t,x,y) = u_e(t,x,y) := \mathrm{e}^t\Big(4^2 x(1-x)y(1-y) + \kappa\Big(\big(x+\frac{1}{3}\big)^2 + \big(y+\frac{1}{4}\big)^2\Big)\Big) \tag{63}$$

is the exact solution. We impose the initial condition $u(0, x, y) = u_e(0, x, y)$ and Dirichlet boundary conditions. If $\kappa = 0$ we have homogeneous boundary conditions, but when $\kappa = 1$ we get inhomogeneous time-dependent Dirichlet conditions. Furthermore, we take $a = b = 1$ in (1).

We apply the MOL approach on a uniform grid with meshwidth $\Delta x = \Delta y = 1/(N + 1)$ for a given integer N. Hence, a semi-discretized system of the form (2) with dimension N^2 is obtained. Observe that the exact solution (63) is a polynomial of degree 2 in each spatial variable so that the global errors come only from the time discretization. AMF-W-methods (4) with either $\eta = 0$ or $\eta = \frac{1}{2}$ will be applied to (2) with fixed step size $\tau = \Delta x = \Delta y = 2^{-j}$, $2 \leq j \leq 10$.

In Tables 1-4 below, the global errors in the Euclidean norm (PDE-GE2) and the corresponding estimates of the PDE-order of convergence (PDE-ORD2) are presented for several methods in the literature when they are applied to the ODE (2) with either time-independent or time-dependent BCs. Special attention is paid to the cases $\eta = 1/2$ and $\eta \neq 1/2$.

Table 1 shows that order 3.25^* can be attained for arbitrary η for time-independent BCs, e.g., with the 4−stage AMF-W-method based on Kutta's 3/8−rule introduced in [7, p. 154]. This method is order three as W-method, order 4 as Rosenbrock method and fulfils the conditions of (14). For the same method, Table 2 (right) shows that order 2.25^* is attained for time-dependent BCs when choosing $\eta = 1/2$. For any other choice $\eta \neq 1/2$, only order 2 is obtained for such kind of boundary conditions, see Table 2 (left).

Table 3 (right) shows that order 2.25^* can be attained by third-order methods for time-dependent BCs when $\eta = \frac{1}{2}$. This is illustrated by means of the 3−stage `W3b` method in [8, p. 573] with $\theta = 0.5$ (for $\eta \neq \frac{1}{2}$ the observed order is just 2, although, for the sake of brevity, we do not include the corresponding results). This method is order two as W-method and order 3 as Rosenbrock method

Table 3 (left), for the same `W3b` method with $\theta = 1/3$, and Table 4 (left), for 2-stage second-order W-method in [12, p. 400–405] with $\theta = 1/4$, show that only order 1.5 is attained in general by second- or higher-order methods for time-dependent BCs if (26) is not fulfilled (see Remark 2). Table 4 (right) shows that order 2 is recovered by the 2-stage second-order W-method in [12, p. 400–405] whenever $\theta > 1/4$.

$N+1$	PDE-GE2	PDE-ORD2	$N+1$	PDE-GE2	PDE-ORD2
4	0.3647D-02	—	4	0.3304D-02	—
8	0.5555D-03	2.715	8	0.4963D-03	2.735
16	0.8154D-04	2.768	16	0.7092D-04	2.807
32	0.9349D-05	3.125	32	0.8045D-05	3.140
64	0.9267D-06	3.335	64	0.8044D-06	3.322
128	0.9154D-07	3.340	128	0.8105D-07	3.311
256	0.9274D-08	3.303	256	0.8356D-08	3.278
512	0.9547D-09	3.280	512	0.8702D-09	3.263
1024	0.9926D-10	3.266	1024	0.9108D-10	3.256

Table 1: Statistics on (2) for $\kappa = 0$ with the 4−stage AMF-W-method ($p = 4$) based on Kutta's 3/8−rule [7, p. 154] with $\theta = 0.5$ ($\eta = 0$ for the table on the left and $\eta = 1/2$ for the table on the right).

References

[1] Arrarás, A., in 't Hout, K. J., Hundsdorfer, W., and Portero, L. (2017). Modified Douglas splitting methods for reaction-diffusion equations. *BIT*, 57(2):261–285.

$N+1$	PDE-GE2	PDE-ORD2	$N+1$	PDE-GE2	PDE-ORD2
4	0.2084D-01	—	4	0.1503D-01	—
8	0.4261D-02	2.290	8	0.2907D-02	2.370
16	0.9023D-03	2.240	16	0.5729D-03	2.343
32	0.1985D-03	2.185	32	0.1167D-03	2.296
64	0.4468D-04	2.151	64	0.2423D-04	2.267
128	0.1024D-04	2.126	128	0.5078D-05	2.255
256	0.2386D-05	2.101	256	0.1068D-05	2.250
512	0.5646D-06	2.079	512	0.2248D-06	2.248
1024	0.1354D-06	2.060	1024	0.4731D-07	2.248

Table 2: Statistics on (2) for $\kappa = 1$ with the 4−stage AMF-W-method ($p = 4$) based on Kutta's 3/8−rule [7, p. 154] with $\theta = 0.5$ ($\eta = 0$ for the table on the left and $\eta = 1/2$ for the table on the right).

$N+1$	PDE-GE2	PDE-ORD2	$N+1$	PDE-GE2	PDE-ORD2
4	0.1999D-01	—	4	0.1471D-01	—
8	0.4451D-02	2.167	8	0.3653D-02	2.009
16	0.1096D-02	2.021	16	0.7945D-03	2.201
32	0.3186D-03	1.783	32	0.1683D-03	2.239
64	0.1034D-03	1.624	64	0.3649D-04	2.205
128	0.3529D-04	1.551	128	0.8018D-05	2.186
256	0.1230D-04	1.521	256	0.1757D-05	2.190
512	0.4322D-05	1.509	512	0.3822D-06	2.201
1024	0.1524D-05	1.504	1024	0.8243D-07	2.213

Table 3: Statistics on (2) for $\kappa = 1$ with the 3−stage AMF-W-method ($p = 3$) based on the `W3b` method in [8, p. 573] with $\theta = 1/3$ (left table) and $\theta = 0.5$ (right table). In both tables we have taken $\eta = 1/2$.

$N+1$	PDE-GE2	PDE-ORD2	$N+1$	PDE-GE2	PDE-ORD2
4	0.4157D-01	—	4	0.4008D-01	—
8	0.1098D-01	1.920	8	0.1060D-01	1.919
16	0.3046D-02	1.850	16	0.2845D-02	1.897
32	0.9029D-03	1.754	32	0.7826D-03	1.862
64	0.2848D-03	1.665	64	0.2146D-03	1.866
128	0.9421D-04	1.596	128	0.5700D-04	1.913
256	0.3213D-04	1.552	256	0.1442D-04	1.982
512	0.1115D-04	1.527	512	0.3485D-05	2.049
1024	0.3904D-05	1.514	1024	0.8130D-06	2.100

Table 4: Statistics on (2) for $\kappa = 1$ with the AMF-W-method based on the 2-stage W-methods ($p = 2$) in [12, p. 400–405] with $\theta = 1/4$ (left table) and $\theta = 0.26$ (right table). Here, $\eta = 1/2$ (for $\eta \neq 1/2$ the same orders of convergence were observed).

[2] Douglas, Jr., J. (1955). On the numerical integration of $\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 = \partial u/\partial t$ by implicit methods. *J. Soc. Indust. Appl. Math.*, 3:42–65.

[3] Douglas, Jr., J. and Rachford, Jr., H. H. (1956). On the numerical solution of heat conduction problems in two and three space variables. *Trans. Amer. Math. Soc.*, 82:421–439.

[4] Gerisch, A. and Verwer, J. G. (2002). Operator splitting and approximate factorization for taxis-diffusion-reaction models. *Appl. Numer. Math.*, 42(1-3):159–176. Ninth Seminar on Numerical Solution of Differential and Differential-Algebraic Equations (Halle, 2000).

[5] González-Pinto, S., Hairer, E., and Hernández-Abreu, D. (2020). Convergence in ℓ_2 and in ℓ_∞ norm of one-stage AMF-W-methods for parabolic problems. *SIAM J. Numer. Anal.*, 58(2):1117–1137.

[6] González-Pinto, S. and Hernández-Abreu, D. (2022). Convergence in the maximum norm of ADI-type methods for parabolic

problems. *Appl. Numer. Math.*, 171:269–280.

[7] González-Pinto, S., Hernández-Abreu, D., and Pérez-Rodríguez, S. (2017). W-methods to stabilize standard explicit Runge-Kutta methods in the time integration of advection-diffusion-reaction PDEs. *J. Comput. Appl. Math.*, 316:143–160.

[8] González-Pinto, S., Hernández-Abreu, D., Pérez-Rodríguez, S., and Weiner, R. (2016). A family of three-stage third order AMF-W-methods for the time integration of advection diffusion reaction PDEs. *Applied Mathematics and Computation*, 274:565–584.

[9] Hairer, E., Nørsett, S. P., and Wanner, G. (1993). *Solving Ordinary Differential Equations I. Nonstiff Problems.* Springer Series in Computational Mathematics 8. Springer, Berlin, 2nd edition.

[10] Hairer, E. and Wanner, G. (1996). *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems.* Springer Series in Computational Mathematics 14. Springer-Verlag, Berlin, 2nd edition.

[11] Hundsdorfer, W. and Verwer, J. (1989). Stability and convergence of the Peaceman-Rachford ADI method for initial-boundary value problems. *Math. Comp.*, 53(187):81–101.

[12] Hundsdorfer, W. and Verwer, J. (2003). *Numerical solution of time-dependent advection-diffusion-reaction equations*, volume 33 of *Springer Series in Computational Mathematics.* Springer-Verlag, Berlin.

[13] in 't Hout, K. J. and Wyns, M. (2016). Convergence of the modified Craig-Sneyd scheme for two-dimensional convection-diffusion equations with mixed derivative term. *J. Comput. Appl. Math.*, 296:170–180.

[14] Kaps, P. and Ostermann, A. (1989). Rosenbrock methods using few LU-decompositions. *IMA J. Numer. Anal.*, 9(1):15–27.

[15] Lubich, C. and Ostermann, A. (1995). Linearly implicit time discretization of non-linear parabolic equations. *IMA J. Numer. Anal.*, 15(4):555–583.

[16] Ostermann, A. and Roche, M. (1992). Runge-Kutta methods for partial differential equations and fractional orders of convergence. *Math. Comp.*, 59(200):403–420.

[17] Ostermann, A. and Roche, M. (1993). Rosenbrock methods for partial differential equations and fractional orders of convergence. *SIAM J. Numer. Anal.*, 30(4):1084–1098.

[18] Rang, J. and Angermann, L. (2005). New Rosenbrock W-methods of order 3 for partial differential algebraic equations of index 1. *BIT*, 45(4):761–787.

[19] Steihaug, T. and Wolfbrandt, A. (1979). An attempt to avoid exact Jacobian and nonlinear equations in the numerical solution of stiff differential equations. *Math. Comp.*, 33(146):521–534.

[20] van der Houwen, P. J. and Sommeijer, B. P. (2001). Approximate factorization for time-dependent partial differential equations. *J. Comput. Appl. Math.*, 128(1-2):447–466. Numerical analysis 2000, Vol. VII, Partial differential equations.