SIAM J. Numer. Anal.
Vol. 58, No. 2, pp. 1117–1137

CONVERGENCE IN ℓ_2 AND ℓ_∞ NORM OF ONE-STAGE AMF-W-METHODS FOR PARABOLIC PROBLEMS[*]

S. GONZÁLEZ-PINTO[†], E. HAIRER[‡], and D. HERNANDEZ-ABREU[†]

Abstract. For the numerical solution of parabolic problems with a linear diffusion term, linearly implicit time integrators are considered. To reduce the cost on the linear algebra level, an alternating direction implicit approach is applied (so-called AMF-W-methods). The present work proves optimal bounds of the global error for two classes of 1-stage methods in the Euclidean ℓ_2 norm as well as in the maximum norm ℓ_∞. The bounds are valid under a very weak step size restriction that covers PDE convergence, where the time step size is of the same order as the spatial grid size.

Key words. parabolic PDEs, time integration, W-methods, approximate matrix factorization, alternating direction implicit schemes, convergence

AMS subject classifications. 65M12, 65M20

DOI. 10.1137/19M1246262

1. Introduction.

The present article is concerned with the numerical solution of parabolic partial differential equations (PDEs) on a rectangular domain in arbitrary dimension m. We assume the space derivatives to be discretized by standard finite differences, and we solve the resulting large-dimensional ordinary differential equation by so-called AMF-W-methods. These are linearly implicit time integrators, where dimensional splitting (alternating direction implicit (ADI)) is used on the linear algebra level to increase efficiency of the integration. We are interested in getting rigorous, optimal bounds of the global error in the Euclidean ℓ_2 norm as well as in the maximum norm ℓ_∞.

There are a few convergence results for linearly implicit time integrators applied to discretized parabolic differential equations. Let us mention [21, Chap. 7] and the convergence analysis of [15] and [14]. Since they are based on estimates on the difference between the Jacobian of the vector field and its approximation, they cannot be directly applied to the approximate matrix factorization (AMF) approach.

On the other hand there exist convergence results for ADI-type time integrators, which are related (but not identical) to W-methods. Convergence of the Peaceman–Rachford integrator is proved in [11], and for the Crank–Nicolson (with locally one-dimensional splitting) in [8]. A convergence analysis (also for problems with mixed derivatives) of a modified Craig–Sneyd scheme is presented in [13]. All these results give error bounds for the Euclidean ℓ_2 norm and are related to two-dimensional PDE problems. For a modified Douglas scheme, the authors in [1] consider an arbitrary number of splitting terms and prove second order of convergence in both the Euclidean and the ℓ_∞ norm under an assumption that is satisfied for time-independent Dirichlet boundary conditions, but not for the time-dependent case.

[*] Received by the editors February 25, 2019; accepted for publication (in revised form) January 13, 2020; published electronically April 9, 2020.
https://doi.org/10.1137/19M1246262

Funding: This work was partially supported by Ministerio de Economia, Industria y Competitividad Spanish Project grant MTM2016-77735-C3-3-P and by Fonds National Suisse grant 200020_159856.

[†] Departamento de Análisis Matemático, Universidad de La Laguna, 38200-La Laguna, Spain (spinto@ull.es, dhabreu@ull.edu.es).

[‡] Section de mathématiques, Université de Genève, CH-1211, Switzerland (Ernst.Hairer@unige.ch).

Our aim is to elaborate a convergence analysis for AMF-W-methods. To emphasize the essential ideas, we only consider 1-stage W-methods, and we restrict our analysis to the linear diffusion problem

$$\partial_t u(t,\vec{x}) = \sum_{j=1}^{m} \alpha_j \, \partial_{x_j x_j} u(t,\vec{x}) + c(t,\vec{x}), \qquad t \ge 0, \tag{1.1}$$

for $\vec{x} = (x_1,\dots,x_m)^\top \in [0,1]^m$, with constants $\alpha_j > 0$, $1 \le j \le m$, and Dirichlet boundary conditions, but we admit an arbitrarily large space dimension m and consider time-dependent boundary conditions. A standard second order central finite difference discretization on a uniform grid,

$$x_j^{(i_j)} = i_j \cdot \Delta x_j, \qquad 0 \le i_j \le n_{x_j}+1, \quad 1 \le j \le m, \tag{1.2}$$

with $\Delta x_j = 1/(n_{x_j}+1)$, yields the ordinary differential equation

$$\dot U = D\,U + g(t), \qquad D = D_1 + \cdots + D_m, \quad g(t) = g_1(t) + \cdots + g_m(t), \tag{1.3}$$

where $D_j = \alpha_j \,(I_{n_{x_m}} \otimes \cdots \otimes D_{x_j x_j} \otimes \cdots \otimes I_{n_{x_1}})$. Here, the differentiation matrices $D_{x_j x_j}$ are tridiagonal with entries $(1,-2,1)/\Delta x_j^2$, respectively, and $\otimes$ stands for the Kronecker product of matrices. Observe that the dimension of system (1.3) is $n_x := n_{x_1} \cdot \ldots \cdot n_{x_m}$.

The splitting of $g(t)$ is not unique. Throughout this article we assume that the discretization of the reaction term $c(t,\vec{x})$ is entirely included in $g_1(t)$, so that $g_2(t),\dots,g_m(t)$ consist only of contributions from the boundary conditions. For homogeneous Dirichlet boundary conditions we thus have $g_j(t) = 0$ for $j = 2,\dots,m$, and for time-independent Dirichlet boundary conditions the vectors $g_2(t),\dots,g_m(t)$ are constant.

Outline of the paper. Section 2 recalls W-methods, when they are applied with the approximate matrix factorization (AMF) technique, and it presents basic formulas for the local and global errors. The power-boundedness of the stability matrix is discussed in section 3. This is standard for the Euclidean ℓ_2 norm, but a challenging problem for the maximum norm ℓ_∞. The main convergence results are given in section 4 for 1-stage AMF-W-methods (with parameter θ). For the ℓ_2 norm and $\theta = 1/2$, convergence of order 2 is proved under a step size restriction that includes PDE convergence. An error bound $\mathcal{O}(\tau^2 |\log h|)$ (where τ is the time step size and h is the mesh-width) is obtained under a still weaker restriction. Convergence order 1 is obtained in the ℓ_∞ norm, and essentially order 2 for time-independent Dirichlet boundary conditions. Section 5 extends the convergence results to a modified method. The convergence order is improved in the ℓ_∞ norm for general Dirichlet boundary conditions and $m = 2$ spatial dimensions. Some technical results are collected in the final section A.

2. Time integration – AMF-W-methods.

The space discretized problem (1.3) is a stiff differential equation. Explicit methods are not suitable for its time integration. An interesting class of methods is linearly implicit integrators, where nonlinear equations are avoided. Methods, requiring only an approximate Jacobian of the vector field, were introduced in [19] and are nowadays called W-methods [7, sect. IV.7].

The idea of using a splitting at a linear algebra level is already present in [2]. Its use in connection with W-methods is proposed in [22]; see also the monograph

[12, sect. IV.5]. Such methods are called AMF-W-methods. Recently much effort has been devoted to their construction and analysis; see, e.g., [5, 6, 16].

2.1. Formulation of AMF-W-methods. For the integration of (1.3) we consider s-stage 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 (see [5, sect. 4])

$$\begin{aligned} K_i^{(0)} &= \tau D\Big(U_n + \sum_{j=1}^{i-1} a_{ij} K_j\Big) + \tau\, g(t_n + c_i\tau) + \sum_{j=1}^{i-1} \ell_{ij} K_j, \\ (I - \theta\tau D_j) K_i^{(j)} &= K_i^{(j-1)} + \theta\rho_i \tau^2 \dot{g}_j(t_n), \quad j = 1,\dots,m, \quad K_i = K_i^{(m)}, \\ U_{n+1} &= U_n + \sum_{i=1}^{s} b_i K_i. \end{aligned} \tag{2.1}$$

It is characterized by (A, L, b, θ) with matrices $A = (a_{ij})_{j<i}$, $L = (\ell_{ij})_{j<i}$, vector $b = (b_i)_i$, and scalar θ. The coefficients ρ_i and c_i are recursively defined by $\rho_i = 1 + \sum_{j=1}^{i-1} \ell_{ij}\rho_j$, and $c_i = \sum_{j=1}^{i-1} a_{ij}\rho_j$. Later, the notation $\mathbb{1} = (1,\dots,1)^\top$ will also be used.

Among this family of methods, the simplest ones are the 1-stage methods with $b_1 = 1$, $c_1 = 0$, $\rho_1 = 1$ and free parameter $\theta > 0$:

$$\begin{aligned} K_1^{(0)} &= \tau D\, U_n + \tau\, g(t_n), \\ (I - \theta\tau D_j) K_1^{(j)} &= K_1^{(j-1)} + \theta\tau^2 \dot{g}_j(t_n), \quad j = 1,\dots,m, \\ U_{n+1} &= U_n + K_1^{(m)}. \end{aligned} \tag{2.2}$$

2.2. Local and global error. For the study of convergence we denote the global error by

$$E_n = U_n - U(t_n). \tag{2.3}$$

Using $DU(t_n) = \dot{U}(t_n) - g(t_n)$, the internal stages of the method (2.1) can be written as

$$\begin{aligned} K_i^{(0)} &= \tau D E_n + \tau D \sum_{j=1}^{i-1} a_{ij} K_j + \sum_{j=1}^{i-1} \ell_{ij} K_j + \tau \dot{U}(t_n) - \tau g(t_n) + \tau g(t_n + c_i\tau), \\ (I - \theta\tau D_1)\cdots(I - \theta\tau D_m) K_i &= K_i^{(0)} + \theta\rho_i\tau^2 \dot{\mathcal{G}}(t_n), \end{aligned}$$

where the vector $\mathcal{G}(t)$ is given by[1]

$$\mathcal{G}(t) = \sum_{i=1}^{m} \Big(\prod_{j=1}^{i-1} \big(I - \theta\tau D_j\big) \Big)\, g_i(t). \tag{2.4}$$

With the notation

$$\Pi(\theta) = (I - \theta\tau D_1)\cdots(I - \theta\tau D_m) \tag{2.5}$$

[1] By convention, the empty product is the identity matrix.

this shows that

$$\begin{aligned}\Pi(\theta)K_i - \tau D\sum_{j=1}^{i-1} a_{ij}K_j - \sum_{j=1}^{i-1}\ell_{ij}K_j &= \tau D E_n + \delta_i,\\ \delta_i = \tau\dot U(t_n) - \tau g(t_n) + \tau g(t_n+c_i\tau) &+ \theta\rho_i\tau^2\dot{\mathcal{G}}(t_n).\end{aligned} \tag{2.6}$$

The update formula of (2.1) becomes

$$E_{n+1} = E_n + \sum_{i=1}^s b_iK_i - \Big(U(t_n+\tau) - U(t_n)\Big). \tag{2.7}$$

Inserting K_i from (2.6) into this formula yields the following result.

THEOREM 2.1. *The global error* (2.3) *satisfies the recursion*

$$E_{n+1} = R\,E_n + S_n, \quad n\ge 0, \tag{2.8}$$

where the stability matrix $R = R(\tau D_1,\dots,\tau D_m)$ *is*

$$R(Z_1,\dots,Z_m) = I + (b^\top\otimes I)\Big(I\otimes\Pi(\theta) - A\otimes Z - L\otimes I\Big)^{-1}(\mathbb{1}\otimes Z)$$

with $\Pi(\theta) = (I-\theta Z_1)\cdots(I-\theta Z_m)$, $Z_i = \tau D_i$, $Z = Z_1+\cdots+Z_m$, *and the local error* $S_n = S_n(\tau D_1,\dots,\tau D_m)$ *is*

$$S_n(\tau D_1,\dots,\tau D_m) = \sum_{i=1}^s b_i\Delta_i - \Big(U(t_n+\tau)-U(t_n)\Big),$$

where Δ_i *is recursively defined by* $\Pi(\theta)\Delta_i = \delta_i + \sum_{j=1}^{i-1}\big(\ell_{ij}I + a_{ij}\tau D\big)\Delta_j$. □

Example 2.2. For 1-stage methods the stability matrix is given by

$$R(\tau D_1,\dots,\tau D_m) = I + \Pi(\theta)^{-1}\tau D, \qquad D = D_1+\cdots+D_m, \tag{2.9}$$

with $\Pi(\theta)$ from (2.5). The local error $S_n = S_n(\tau D_1,\dots,\tau D_m)$ is

$$S_n = \Pi(\theta)^{-1}\Big(\tau\dot U(t_n) + \theta\tau^2\dot{\mathcal{G}}(t_n)\Big) - \Big(U(t_n+\tau)-U(t_n)\Big), \tag{2.10}$$

with the vector $\mathcal{G}(t)$ given by (2.4).

If, for a given norm, the estimates $\|R\|\le 1+C_0\tau$ and $\|S_n\|\le C_1\tau^{p+1}$ hold, a standard argument yields convergence of order p, i.e., $\|E_n\|\le C\tau^p$ on a bounded time interval $0\le n\tau\le T$. If such estimates are not available with suitable constants C_0 and C_1, it is advised to solve the recursion (2.8):

$$E_n = R^nE_0 + \sum_{j=0}^{n-1} R^{n-1-j}S_j. \tag{2.11}$$

The convergence analysis, based on this relation, requires the power-boundedness of the stability matrix and a careful analysis of the local error. This is the content of the following sections.

3. Power-boundedness of the stability function. We consider the Euclidean space $\mathbb{R}^{n_x}$ with $n_x = n_{x_1} \cdot \ldots \cdot n_{x_m}$, and vectors $U = (U_{i_1,\ldots,i_m})$ and $V = (V_{i_1,\ldots,i_m})$, where $i_j = 1, \ldots, n_{x_j}$. We are mainly interested in the Euclidean inner product norm and in the maximum norm.

3.1. Euclidean $\boldsymbol{\ell_2}$ norm. The weighted inner product

$$\langle U, V\rangle = \Delta x_1 \cdot \ldots \cdot \Delta x_m \sum_{i_1=1}^{n_{x_1}} \cdots \sum_{i_m=1}^{n_{x_m}} U_{i_1,\ldots,i_m} V_{i_1,\ldots,i_m} \qquad \text{for} \quad U, V \in \mathbb{R}^{n_x},$$

with induced ℓ_2 norm

$$\|U\|_2 = \sqrt{\langle U, U\rangle} \qquad \text{for} \quad U \in \mathbb{R}^{n_x}, \tag{3.1}$$

has the advantage that, considering the diagonalization of the stability matrix, its power-boundedness can be reduced to that of a scalar. In fact, the eigenvectors of $D_1, \ldots, D_m$, and $D = D_1 + \cdots + D_m$ are all the same (see [12, p. 297]). With

$$\phi_j^{(x_i)} = \sqrt{2}\,\big(\sin(j\Delta x_i \pi), \sin(2j\Delta x_i\pi), \ldots, \sin(n_{x_i} j \Delta x_i \pi)\big)^\top \tag{3.2}$$

they are given by $\phi_{j_m}^{(x_m)} \otimes \cdots \otimes \phi_{j_1}^{(x_1)}$ with eigenvalue

$$\lambda_j^{(x_i)} = -\frac{4}{{\Delta x_i}^2} \sin^2\!\big(j \Delta x_i \tfrac{\pi}{2}\big)\alpha_i, \qquad j = 1, \ldots, n_{x_i}. \tag{3.3}$$

We have

$$D_i\, \phi_{j_m}^{(x_m)} \otimes \cdots \otimes \phi_{j_1}^{(x_1)} = \lambda_{j_i}^{(x_i)} \phi_{j_m}^{(x_m)} \otimes \cdots \otimes \phi_{j_1}^{(x_1)},$$

$$D\, \phi_{j_m}^{(x_m)} \otimes \cdots \otimes \phi_{j_1}^{(x_1)} = \big(\lambda_{j_1}^{(x_1)} + \cdots + \lambda_{j_m}^{(x_m)}\big)\, \phi_{j_m}^{(x_m)} \otimes \cdots \otimes \phi_{j_1}^{(x_1)}.$$

These eigenvectors are orthonormal with respect to the inner product. When expanding a vector U in the basis of eigenvectors

$$U = \sum_{j_1=1}^{n_{x_1}} \cdots \sum_{j_m=1}^{n_{x_m}} \widehat{U}_{j_1,\ldots,j_m}\, \phi_{j_m}^{(x_m)} \otimes \cdots \otimes \phi_{j_1}^{(x_1)},$$

we denote the Fourier coefficients by $\widehat{U}_{j_1,\ldots,j_m}$. This implies that

$$\|R(\tau D_1, \ldots, \tau D_m) U\|_2^2 = \sum_{j_1=1}^{n_{x_1}} \cdots \sum_{j_m=1}^{n_{x_m}} \Big| R\big(\tau\lambda_{j_1}^{(x_1)}, \ldots, \tau\lambda_{j_m}^{(x_m)}\big)\, \widehat{U}_{j_1,\ldots,j_m}\Big|^2,$$

which is bounded by $\|U\|_2^2$, provided that

$$\big|R\big(\tau\lambda_{j_1}^{(x_1)}, \ldots, \tau\lambda_{j_m}^{(x_m)}\big)\big| \le 1 \qquad \text{for all} \;\; \lambda_{j_i}^{(x_i)} \le 0. \tag{3.4}$$

Consequently, we have contractivity $\|R(\tau D_1, \ldots, \tau D_m)\|_2 \le 1$, and hence also power-boundedness of $R = R(\tau D_1, \ldots, \tau D_m)$ in the Euclidean ℓ_2 norm.

Example 3.1. The stability matrix (2.9) of the 1-stage AMF-W-method (2.2) satisfies $\|R\|_2 \le 1$ for $\theta \ge 1/2$. This is a consequence of (3.4), which follows from $(1-\theta z_1) \cdot \ldots \cdot (1 - \theta z_m) \ge 1 - \theta(z_1 + \cdots + z_m)$ for real $z_j \le 0$. Note that (3.4) also follows from [9, Theorem 2.1], where the more general situation of complex arguments z_j is considered.

3.2. Maximum norm in dimension $m = 2$. We next consider the maximum norm

$$\|U\|_\infty = \max_{i_1,\dots,i_m} |U_{i_1,\dots,i_m}| \qquad \text{for} \quad U \in \mathbb{R}^{n_x}. \tag{3.5}$$

From the inequality $\Delta x_1 \cdot \ldots \cdot \Delta x_m \, \|U\|_\infty^2 \le \|U\|_2^2 \le \|U\|_\infty^2$ it follows that $\|U\|_\infty \le h^{-m/2}\|U\|_2$ (with $h = \min\{\Delta x_1, \dots, \Delta x_m\}$) for all vectors U. This implies

$$\|R^n\|_\infty \le h^{-m/2}\|R^n\|_2$$

in the corresponding operator norms. Improving this bound for powers of the stability matrix in the maximum norm is more difficult.

For 1-stage methods the stability matrix is given by (2.9). For $\theta = 1/2$ and for $m = 2$ the stability matrix is a product

$$R(\tau D_1, \tau D_2) = \mathrm{Tr}(\tau\alpha_2 D_{x_2x_2}) \otimes \mathrm{Tr}(\tau\alpha_1 D_{x_1x_1}),$$

where $\mathrm{Tr}(A) = \bigl(I - \frac{1}{2}A\bigr)^{-1}\bigl(I + \frac{1}{2}A\bigr)$ is the stability matrix of the trapezoidal rule (Crank–Nicolson) applied to the one-dimensional heat equation (see [1, sect. 4.2, p. 276]). Taking the nth power of this relation and using $\|A \otimes B\|_\infty = \|A\|_\infty \|B\|_\infty$ for two matrices A, B, we obtain the relation

$$\bigl\|R(\tau D_1, \tau D_2)^k\bigr\|_\infty = \bigl\|\mathrm{Tr}(\tau\alpha_1 D_{x_1x_1})^k\bigr\|_\infty \bigl\|\mathrm{Tr}(\tau\alpha_2 D_{x_2x_2})^k\bigr\|_\infty.$$

It is proved in [4] that each of the two factors is bounded by $C_\infty < 4.325$. Larger bounds for C_∞ were stated in [3, p. 52], [17]. Other references about the power-boundedness of rational functions of matrices $D_{x_jx_j}$ can be found, e.g., in [18, Thm. 6.4.2] and [20]. Hence

$$\bigl\|R(\tau D_1, \tau D_2)^k\bigr\|_\infty \le C_\infty^2, \tag{3.6}$$

which is independent of k, τ, and the spatial step size.

For $\theta > 1/2$ the stability matrix $R = R(\tau D_1, \tau D_2)$ is no longer the tensor product of two matrices of lower dimension. To get some insight into the power-boundedness we consider the situation $\alpha_1 = \alpha_2 = 1$, $n_{x_1} = n_{x_2} = n$, and we compute numerically $\|R^k\|_\infty$ for $\tau = 1/(n+1)$, for different values of θ, and for $n = 8, 16, 32$. The result is shown in Figure 1, where $\|R^k\|_\infty$ is plotted as a function of $k\tau \le 1$. Thick curves correspond to $\theta = 1/2$, middle curves to $\theta = 3/4$, and thin curves to $\theta = 1$. The largest values of $\|R^k\|_\infty$ turn out to be for small values of k, and for large $n = 32$. These values are 5.96 for $\theta = 1/2$, 4.35 for $\theta = 3/4$, and 3.61 for $\theta = 1$. For large values of k, $\|R^k\|_\infty$ is more or less independent of the number of grid points n.

3.3. Maximum norm in dimension $m = 3$. In dimension $m \ge 3$ (even for $\theta = 1/2$) we are no longer in the lucky situation where the stability matrix can be written as the tensor product of lower-dimensional matrices. For $m = 3$ and $\theta = 1/2$ we have

$$\begin{aligned} R(\tau D_1, \tau D_2, \tau D_3) &= \mathrm{Tr}(\tau\alpha_3 D_{x_3x_3}) \otimes \mathrm{Tr}(\tau\alpha_2 D_{x_2x_2}) \otimes \mathrm{Tr}(\tau\alpha_1 D_{x_1x_1}) \\ &\qquad - 2\, s(\tau\alpha_3 D_{x_3x_3}) \otimes s(\tau\alpha_2 D_{x_2x_2}) \otimes s(\tau\alpha_1 D_{x_1x_1}), \end{aligned}$$

where $s(A) = \frac{1}{2}A\bigl(I - \frac{1}{2}A\bigr)^{-1}$.

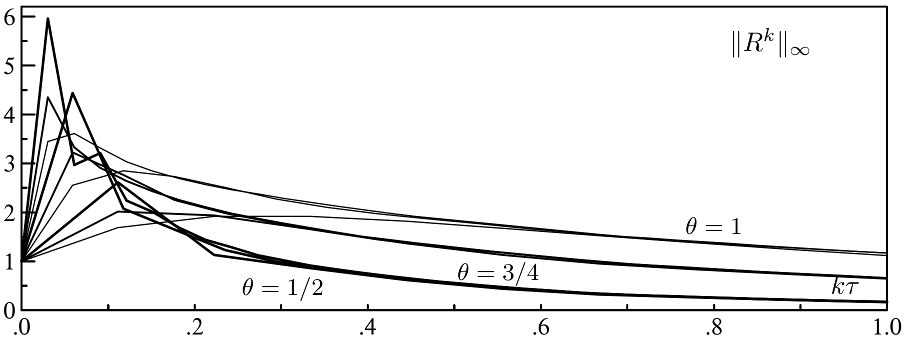

FIG. 1. $\|R^k\|_\infty$ *for* (2.9) *as a function of* $k\tau$ *with* $m = 2$, $\tau = 1/(n+1)$, *and* $n = 8, 16, 32$.

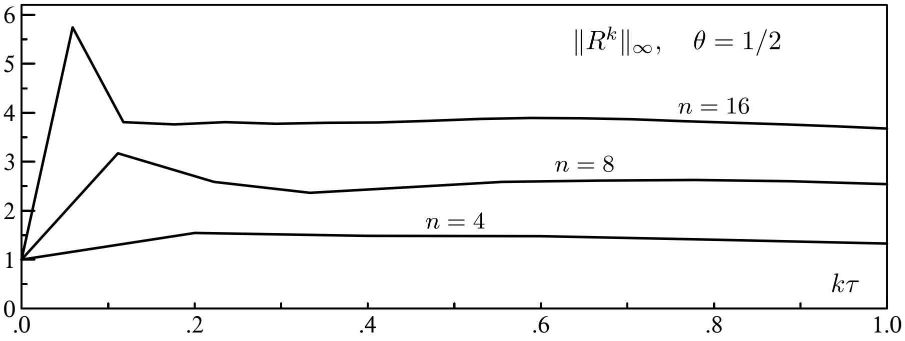

FIG. 2. $\|R^k\|_\infty$ *for* (2.9) *as a function of* $k\tau$ *with* $m = 3$, $\tau = 1/(n+1)$, *and* $n = 4, 8, 16$.

To get a feeling of the power-boundedness of $R = R(\tau D_1, \tau D_2, \tau D_3)$ we rely on numerical experiments. We consider the diffusion problem with $\alpha_1 = \alpha_2 = \alpha_3 = 1$ and $n_{x_1} = n_{x_2} = n_{x_3} = n$, and we compute numerically $\|R^k\|_\infty$ for $\theta = 1/2$, $\tau = 1/(n+1)$, and for different values of n. The result is shown in Figure 2. The largest values of $\|R^k\|_\infty$ are again for small values of k. In contrast to the results for dimension $m = 2$ (Figure 1), we observe that $\|R^k\|_\infty$ has significantly different values for large k.

4. Convergence of 1-stage methods.

We consider the 1-stage method (2.2) with parameter θ. It has classical order $p = 1$ in general, and classical order $p = 2$ for $\theta = 1/2$. The stability matrix and the local error are given in Example 2.2. The presentation (2.10) of the local error is not practical for an analysis, because for nonhomogeneous boundary conditions the vector functions $g_i(t)$ contain elements that are proportional to Δx_i^{-2}. The same is true for the vectors $D_iU(t)$. It is much better to work with the vector functions

$$\varphi_i(t) = D_iU(t) + g_i(t), \tag{4.1}$$

which are bounded (together with their time derivatives) if the spatial mesh size tends to zero. We remark that $\varphi_1(t) + \cdots + \varphi_m(t) = DU(t) + g(t) = \dot{U}(t)$.

4.1. Convergence of order 1.

We write the local error (2.10) in a form that is more appropriate for estimates. In the expression $\mathcal{G}(t)$ of (2.4) we insert $g_i(t) =$

$\varphi_i(t) - D_i U(t)$. Expanding $U(t_n+\tau)$ into a Taylor series (with remainder term), and using the identity (which is seen by induction on m)

$$I - \theta\tau \sum_{i=1}^{m} \Pi_i(\theta) D_i = \Pi(\theta) \qquad \text{with} \qquad \Pi_i(\theta) = \prod_{j=1}^{i-1} \bigl(I - \theta\tau D_j\bigr) \tag{4.2}$$

(note that $\Pi_1(\theta) = I$), the local error (2.10) can be written as

$$S_n = \theta\tau^2 \Pi(\theta)^{-1} \sum_{i=1}^{m} \Pi_i(\theta)\dot\varphi_i(t_n) - \tau^2 \int_0^1 (1-s)\,\ddot U(t_n+s\tau)\,\mathrm{d}s. \tag{4.3}$$

We require that the solution $U(t)$ of the differential equation (1.3) is such that in a given norm and on a bounded time interval

$$\begin{aligned} \|\ddot U(t)\| &\le C, \\ \|\dot\varphi_j(t)\| &\le C \quad \text{for} \;\; j = 1,\dots,m. \end{aligned} \tag{4.4}$$

This requirement is fulfilled in any norm if the solution $u(t,\vec x)$ of the linear diffusion problem is sufficiently differentiable. Moreover, we assume for the method that

$$\|R^n\| \le C \quad \text{for} \;\; n \ge 1, \tag{4.5}$$

$$\|(I - \theta\tau D_j)^{-1}\| \le C \quad \text{for} \;\; j = 1,\dots,m. \tag{4.6}$$

Note that for the weighted ℓ_2 norm (3.1) the conditions (4.5) and (4.6) are satisfied in any dimension m for $\theta \ge 1/2$. It is satisfied in dimension $m=2$ for the ℓ_∞ norm and $\theta = 1/2$.

THEOREM 4.1. *Assume that* (4.4), (4.5), *and* (4.6) *hold for a given norm, and consider the* 1*-stage method* (2.2) *with* $\theta \ge 1/2$. *Then, for an initial value satisfying* $\|E_0\| = \mathcal{O}(\tau)$, *the global error is bounded by*

$$\|E_n\| = \mathcal{O}(\tau) \qquad \textit{for} \;\; n\tau \le T,$$

where the constant symbolized by $\mathcal{O}(\cdot)$ *is independent of* τ *and the spatial discretization, but depends on the constant* C *in* (4.4), (4.5), *and* (4.6), *and on* T.

Proof. The assumptions of the theorem imply that the local error (4.3) satisfies in the given norm $S_n = \mathcal{O}(\tau^2)$. The power-boundedness of the stability matrix then proves convergence of order 1. □

The rest of this section considers the case $\theta = 1/2$, for which the method has classical order 2. A more refined analysis of the local error is necessary to get improved error estimates.

4.2. Higher order convergence in the ℓ_2 norm for $\theta = 1/2$. Our aim is to prove second order convergence for the 1-stage method (2.2) with $\theta = 1/2$. Since $\tau^2 \int_0^1 (1-s)\,\ddot U(t_n+s\tau)\,\mathrm{d}s = \frac{\tau^2}{2}\ddot U(t_n) + \frac{\tau^3}{2}\int_0^1 (1-s)^2\,\dddot U(t_n+s\tau)\,\mathrm{d}s$ and $\sum_{i=1}^{m} \dot\varphi_i(t_n) = \ddot U(t_n)$, the local error (4.3) can be written as

$$S_n = \frac{\tau^2}{2}\Pi\bigl(\tfrac12\bigr)^{-1} \sum_{i=1}^{m} \Bigl(\Pi_i\bigl(\tfrac12\bigr) - \Pi\bigl(\tfrac12\bigr)\Bigr)\dot\varphi_i(t_n) - \frac{\tau^3}{2}\int_0^1 (1-s)^2\,\dddot U(t_n+s\tau)\,\mathrm{d}s. \tag{4.7}$$

In the following we require that

$$\|\dddot{U}(t)\|_2 \le C \tag{4.8}$$

and that the functions $\dot{\varphi}_j(t) = \big(\dot{\varphi}_{i_1,\dots,i_m,j}(t)\big) \in \mathbb{R}^{n_x}$ satisfy, for $j = 1,\dots,m$,

$$\dot{\varphi}_{i_1,\dots,i_m,j}(t) = v_j\big(t, x_1^{(i_1)},\dots,x_m^{(i_m)}\big), \tag{4.9}$$

where $v_i(t,\vec{x})$ is continuously differentiable in all variables. This assumption permits us to get improved error estimates with the help of Lemma A.2. Observe that each vector $\varphi_j(t)$ defined in (4.1) is a second order approximation on the grid to the partial derivative $\partial_{x_j x_j} u$ at every time $t \in [0,T]$ if the PDE solution u is at least four times continuously differentiable in each spatial variable.

The step size restriction (4.10) in the following theorem covers the situation $\tau \approx \Delta x_l$, $\tau \to 0$, which is often called *PDE convergence.* Note that under a step size restriction $\tau \le c_0 \Delta x_l^2$ we are in the nonstiff situation and classical convergence results can be applied.

THEOREM 4.2. *Assume that* (4.8) *and* (4.9) *hold, and consider the* 1*-stage method* (2.2) *with* $\theta = 1/2$*. Then, for an initial value satisfying* $\|E_0\|_2 = \mathcal{O}(\tau^2)$*, and under the step size restriction*

$$c_0 \Delta x_l^2 \le \tau \le c_1 \Delta x_l, \qquad l = 1,\dots,m \tag{4.10}$$

(with positive c_0, c_1*), the global error is bounded by*

$$\|E_n\|_2 = \mathcal{O}(\tau^2) \qquad \textit{for} \ \ n\tau \le T,$$

where the constant symbolized by $\mathcal{O}(\cdot)$ *is independent of* τ *and the spatial discretization, but depends on the constant* C *in* (4.8)*, on* c_0, c_1 *in* (4.10)*, on* T*, and on bounds of the spatial derivatives of* $v_i(t,\vec{x})$ *in* (4.9)*.*

Proof. The local error S_n of (4.7) is a linear combination of terms that fall into one of the following three categories:

$$\begin{aligned} &\text{(A)} \quad \tfrac{\tau^3}{2}\textstyle\int_0^1 (1-s)^2\, \dddot{U}(t_n + s\tau)\,\mathrm{d}s,\\ &\text{(B)} \quad \tau^3 \Pi\big(\tfrac{1}{2}\big)^{-1} D_l \dot{\varphi}_i(t_n),\\ &\text{(C)} \quad \tau^{2+k} \Pi\big(\tfrac{1}{2}\big)^{-1} D_{l_1} D_{l_2} \cdots D_{l_k} \dot{\varphi}_i(t_n), \qquad k \ge 2, \end{aligned}$$

where $1 \le l_1 < l_2 < \cdots < l_k \le m$. By assumption (4.8) the expression (A) is of size $\mathcal{O}(\tau^3)$. Therefore, a standard convergence argument shows that this term leads to a $\mathcal{O}(\tau^2)$ contribution in the global error.

The expressions in (B) and (C) require a refined analysis. To exploit the smooth dependence of the local error on time, we apply partial summation in (2.11) and write the global error as (assuming $E_0 = 0$)

$$E_n = \Big(\sum_{j=0}^{n-1} R^{n-1-j}\Big) S_0 + \sum_{j=0}^{n-2}\Big(\sum_{i=j+1}^{n-1} R^{n-1-i}\Big)\big(S_{j+1} - S_j\big),$$

which can also be written as

$$E_n = (I - R^n)(I - R)^{-1} S_0 + \sum_{j=0}^{n-2} (I - R^{n-1-j})(I - R)^{-1}\big(S_{j+1} - S_j\big). \tag{4.11}$$

Let us denote the expression in (B) by $\widehat{S}_n$ and its contribution to the global error by $\widehat{E}_n$. Since $R-I=\Pi\big(\frac{1}{2}\big)^{-1}\tau D$ by (2.9), we have

$$\widehat{S}_n=\tau^3\Pi\big(\tfrac{1}{2}\big)^{-1}D_l\dot{\varphi}_i(t_n)=\tau^2(R-I)D^{-1}D_l\dot{\varphi}_i(t_n)$$

and

$$\widehat{S}_{j+1}-\widehat{S}_j=\tau^3(R-I)D^{-1}D_l\int_0^1\ddot{\varphi}_i(t_j+s\tau)\,\mathrm{d}s.$$

Inserted into (4.11), the factor $(R-I)$ cancels with $(I-R)^{-1}$. Since $\|R^n\|_2\le 1$ and $\|D^{-1}D_l\|_2\le 1$, which can be seen by diagonalization of the matrices, the boundedness of the time derivatives of $\varphi_i(t)$ implies that $\|\widehat{E}_n\|_2=\mathcal{O}(\tau^2)$.

We next consider terms of the form (C). Without loss of generality we assume $l_j=j$, so that this term of the local error is

$$\widehat{S}_n=\tau^{2+k}\Pi\big(\tfrac{1}{2}\big)^{-1}D_1D_2\cdots D_k\dot{\varphi}_i(t_n) \tag{4.12}$$

with $2\le k\le m$. We expand the vector $\dot{\varphi}_i(t_n)$ in the basis of eigenvectors of D_l (see section 3.1),

$$\dot{\varphi}_i(t_n)=\sum_{i_1=1}^{n_{x_1}}\cdots\sum_{i_m=1}^{n_{x_m}}\widehat{\dot{\Phi}}_{i_1,\dots,i_m,i}(t_n)\,\phi_{i_m}^{(x_m)}\otimes\cdots\otimes\phi_{i_1}^{(x_1)}, \tag{4.13}$$

so that the norm of $(I-R^n)(I-R)^{-1}\widehat{S}_0$ (see (4.11)) becomes

$$\tau^{2+k}\left\{\sum_{i_1=1}^{n_{x_1}}\cdots\sum_{i_m=1}^{n_{x_m}}\left(\frac{1-r_{i_1,\dots,i_m}^n}{1-r_{i_1,\dots,i_m}}\right)^2\frac{\prod_{l=1}^k|\lambda_{i_l}^{(x_l)}|^2\,|\widehat{\dot{\Phi}}_{i_1,\dots,i_m,i}(t_0)|^2}{\prod_{l=1}^m(1+\frac{1}{2}\tau|\lambda_{i_l}^{(x_l)}|)^2}\right\}^{1/2}, \tag{4.14}$$

where $r_{i_1,\dots,i_m}=R\big(\tau\lambda_{i_1}^{(x_1)},\dots,\tau\lambda_{i_m}^{(x_m)}\big)$. An application of Lemma A.2, which is justified by assumption (4.9), shows that $\big(\prod_{l=1}^m|\lambda_{i_l}^{(x_l)}|\big)|\widehat{\dot{\Phi}}_{i_1,\dots,i_m,i}(t_0)|^2\le C^2$ for all $i_1,\dots,i_m$. Consequently, the expression (4.14) is bounded by $C\tau^2a(n)$, where

$$a(n)=\tau^k\left\{\sum_{i_1=1}^{n_{x_1}}\cdots\sum_{i_m=1}^{n_{x_m}}\left(\frac{1-r_{i_1,\dots,i_m}^n}{1-r_{i_1,\dots,i_m}}\right)^2\frac{\prod_{l=1}^k|\lambda_{i_l}^{(x_l)}|}{\prod_{l=1}^m(1+\frac{1}{2}\tau|\lambda_{i_l}^{(x_l)}|)^2}\prod_{l=k+1}^m\frac{1}{|\lambda_{i_l}^{(x_l)}|}\right\}^{1/2}. \tag{4.15}$$

The same computation can be done for the second summand in (4.11). The only difference is that n has to be replaced by $n-1-j$, and $\widehat{S}_0$ by $\widehat{S}_{j+1}-\widehat{S}_j$, which has the form (4.12) with $\tau\int_0^1\ddot{\varphi}_i(t_j+s\tau)\,\mathrm{d}s$ instead of $\dot{\varphi}_i(t_0)$. Applying the triangle inequality in (4.11), we get the following for all $n\ge 1$, $n\tau\le T$:

$$\|\widehat{E}_n\|_2\le C\tau^2a(n)+C\tau^3\sum_{j=0}^{n-2}a(n-1-j). \tag{4.16}$$

From Lemma 4.3 below we have under the assumption (4.10) that $|a(n)|\le M$ for $n\tau\le T$. This implies $\|\widehat{E}_n\|_2=\mathcal{O}(\tau^2)$ and completes the proof of the theorem. □

LEMMA 4.3. *Let $a(n)$ be defined by (4.15) for $n\tau\le T$ and $k\ge 2$. Then there exists a constant $M=M(T)$ such that for all τ satisfying (4.10) it holds that*

$$a(n)\le M.$$

Proof. Using the fact that

$$0 \le 1 - r_{i_1,\dots,i_m} = \frac{\tau \sum_{l=1}^{m} |\lambda_{i_l}^{(x_l)}|}{\prod_{l=1}^{m}(1+\frac{1}{2}\tau|\lambda_{i_l}^{(x_l)}|)} \le 2,$$

it follows from Lemma A.6 that

$$a(n)^2 \le (2^{2-\gamma}T^{\gamma})\tau^{2k-2} \sum_{i_1=1}^{n_{x_1}} \cdots \sum_{i_m=1}^{n_{x_m}} \frac{\prod_{l=1}^{k} |\lambda_{i_l}^{(x_l)}| \, \prod_{l=k+1}^{m} |\lambda_{i_l}^{(x_l)}|^{-1}}{\prod_{l=1}^{m}\big(1+\frac{1}{2}\tau|\lambda_{i_l}^{(x_l)}|\big)^{\gamma}\big(\sum_{l=1}^{m}|\lambda_{i_l}^{(x_l)}|\big)^{2-\gamma}},$$

where $\gamma \in [0,2]$ is for the moment a free parameter. With the help of the arithmetic-geometric mean inequality

$$\sum_{l=1}^{m} |\lambda_{i_l}^{(x_l)}| \ge \sum_{l=1}^{k} |\lambda_{i_l}^{(x_l)}| \ge k \cdot \sqrt[k]{|\lambda_{i_1}^{(x_1)}| \cdot \ldots \cdot |\lambda_{i_k}^{(x_k)}|},$$

this sum of products turns into a product of sums and yields

$$(4.17)\quad a(n)^2 \le \left(\frac{2^{2-\gamma}T^{\gamma}}{k^{2-\gamma}}\right)\tau^{2k-2}\prod_{l=1}^{k}\left(\sum_{i_l=1}^{n_{x_l}} \frac{|\lambda_{i_l}^{(x_l)}|^{1-(2-\gamma)/k}}{\big(1+\frac{1}{2}\tau|\lambda_{i_l}^{(x_l)}|\big)^{\gamma}}\right)\prod_{l=k+1}^{m}\left(\sum_{i_l=1}^{n_{x_l}}\frac{1}{|\lambda_{i_l}^{(x_l)}|}\right).$$

As a consequence of [12, Lem. 6.2, p. 298] the second product is bounded, because

$$\sum_{i_l=1}^{n_{x_l}}\frac{1}{|\lambda_{i_l}^{(x_l)}|} = \frac{(\Delta x_l)^2}{4\alpha_l}\sum_{i_l=1}^{n_{x_l}}\frac{1}{\sin^2\big(i_l\Delta x_l\frac{\pi}{2}\big)} = \mathcal{O}(1).$$

The first product can be bounded by applying Lemma A.5 with $\alpha = 2 - \frac{2(2-\gamma)}{k}$. With the choice $\gamma = \alpha < 1$ for $k = 2$, and $\gamma = 1$ for $k \ge 3$, we have $\alpha + 1 - 2\gamma > 0$, so that

$$(4.18)\qquad a(n)^2 \le C\tau^{2k-2}\prod_{l=1}^{k}\Big(\tau^{-\gamma}\Delta x_l^{2\gamma-\alpha-1}\Big) = C\prod_{l=1}^{k}\Big(\frac{\tau}{\Delta x_l}\Big)^{\alpha+1-2\gamma},$$

because $2 - \frac{2}{k} - \gamma = \alpha + 1 - 2\gamma$ for our choices of γ. The boundedness of $a(n)$ thus follows from assumption (4.10). □

Although Theorem 4.2 covers the important situation of PDE convergence, we present a further convergence result under the weaker step size restriction

$$(4.19)\qquad c_0\Delta x_l^2 \le \tau \le c_2\Delta x_l^{\beta}, \qquad \beta \in (0,1)$$

(with positive c_0, c_2), for $l = 1, \dots, m$.

THEOREM 4.4. *In the situation of Theorem* 4.2, *where the step size restriction* (4.10) *is relaxed to* (4.19), *the global error is bounded by*

$$\|E_n\|_2 = \mathcal{O}(\tau^2|\log h|) \qquad \textit{for} \quad n\tau \le T,$$

where $h = \min_{j=1,\dots,m}\Delta x_j$.

Proof. The proof is the same as that for Theorem 4.2. The only difference is that a different bound for $a(n)$ in (4.15) will be obtained.

To get an estimate for $a(n)$ under the assumption (4.19) we follow the proof of Lemma 4.3. For a parameter $\gamma \in [0,2]$ we again have $\alpha = 2 - \frac{2(2-\gamma)}{k}$. As long as $\alpha + 1 - 2\gamma > 0$ the left inequality in (4.18) holds and we have

$$a(n)^2 \le C\tau^{2k-2} \prod_{l=1}^{k} \Big(\tau^{-\gamma} \Delta x_l^{2\gamma-\alpha-1} \Big) = C \prod_{l=1}^{k} \Big(\frac{\tau}{\Delta x_l^{\beta}} \Big)^{(\alpha+1-2\gamma)/\beta}, \tag{4.20}$$

provided that

$$\alpha + 1 - 2\gamma = \beta \Big(2 - \frac{2}{k} - \gamma \Big). \tag{4.21}$$

Inserting $\alpha = 2 - \frac{2(2-\gamma)}{k}$, we can compute γ from this relation. We obtain

$$\gamma = 1 + \frac{(1-\beta)(k-2)}{(2-\beta)k - 2} \qquad \text{and} \qquad \alpha + 1 - 2\gamma = \frac{\beta\,(k-2)^2}{k\big((2-\beta)k-2\big)}.$$

One can check that $\gamma \in [1,2]$ for $k \ge 2$, and one sees that $\alpha + 1 - 2\gamma > 0$ for $k \ge 3$. Consequently, (4.20) implies $a(n) \le M$ for all $k \ge 3$.

For $k = 2$, we have $\gamma = 1$ and $\alpha = 1$, so that $\alpha + 1 - 2\gamma = 0$. An application of Lemma A.5 in (4.17) for the sums

$$\sum_{i_l=1}^{n_{x_l}} \frac{|\lambda_{i_l}^{(x_l)}|^{1-(2-\gamma)/k}}{\big(1 + \frac{1}{2}\tau|\lambda_{i_l}^{(x_l)}|\big)^{\gamma}}$$

yields

$$a(n)^2 \le C\tau^2 \prod_{l=1}^{2} \bigg(\tau^{-1} \Big(1 + \Big| \log\Big(\frac{\tau}{\Delta x_l^2} \Big) \Big| \Big) \bigg). \tag{4.22}$$

From the step size restriction (4.19) we have $c_0 \le \tau/\Delta x_l^2 \le c_2 \Delta x_l^{\beta-2}$, and consequently

$$\Big| \log\Big(\frac{\tau}{\Delta x_l^2} \Big) \Big| \le C + (2-\beta)|\log \Delta x_l|,$$

so that $a(n)^2 \le M^2 |\log h|^2$ follows from (4.22).

The estimates $a(n) \le M$ for $k \ge 3$, and $a(n) \le M|\log h|$ for $k = 2$, together with (4.16), complete the proof of the theorem. □

Remark 4.5. The 1-stage method (2.2) is a simple variant of the Douglas method, where the difference $\theta\tau(g_j(t_{n+1}) - g_j(t_n))$ is replaced by $\theta\tau^2\dot{g}_j(t_n)$. For the Douglas method with $\theta = \frac{1}{2}$, second order of convergence in both the Euclidean and maximum norms is proved in [1, Thm. 3.1] under the assumption (see [1, (3.16b), p. 271])

$$\tau^{k-1} D^{-1} D_{l_1} D_{l_2} \cdots D_{l_k} v(t_n) = \mathcal{O}(1), \quad 1 \le l_1 < \cdots < l_k < i \le m \quad (v = \dot{\varphi}_i, \ddot{\varphi}_i). \tag{4.23}$$

The validity of this condition is discussed for $m = 3$ in [1, Ex. 3.2, p. 272]. The same condition is considered in [10, Thm. 3.2] in order to prove convergence for linear multistep methods with stabilizing corrections applied to split ordinary differential equations. For general $m \geq 3$ and $k < m$ one cannot expect better than

$$\tau^{k-1} \| D^{-1} D_{l_1} D_{l_2} \cdots D_{l_k} v(t_n) \|_2 \leq c\, \tau^{k-1} h^{(4-3k)/2}$$

(see Lemma A.3 below). This shows that for $\tau \approx h$ the expression (4.23) is unbounded in general for $k \geq 3$, i.e., for $m \geq 4$.

The proof of Theorem 4.2 avoids the assumption (4.23). Instead, the closely related terms of type (C) are estimated directly without splitting them into a product of $\tau^2(R - I) = \tau^3 \Pi\big(\frac{1}{2}\big)^{-1} D$ and the expression in (4.23).

4.3. Higher order convergence in the ℓ_∞ norm for $\theta = 1/2$.

Convergence of the 1-stage AMF-W-method (2.2) in the ℓ_∞ norm is less favorable. Numerical experiments show that, in general, the order of PDE convergence is not more than 1, and this is already covered by Theorem 4.1. However, second order convergence can be proved for linear diffusion problems with time-independent boundary conditions.

Throughout this subsection we assume, in addition to (4.8), that the solution $U(t)$ and the functions $\varphi_i(t)$ of (4.1) satisfy, for $k \geq 0$,

$$\begin{aligned} &\| D_{l_1} D_{l_2} \cdots D_{l_k} \ddot{U}(t) \| \leq C && \text{for} \quad l_1 < l_2 < \cdots < l_k, \\ &\| D_{l_1} D_{l_2} \cdots D_{l_k} \dot{\varphi}_i(t) \| \leq C && \text{for} \quad l_1 < l_2 < \cdots < l_k < i. \end{aligned} \tag{4.24}$$

This condition is motivated by the following example.

Example 4.6 (time-independent boundary conditions). For the linear diffusion problem (1.1) we consider time-independent Dirichlet boundary conditions

$$u(t, \vec{x}) = b(\vec{x}), \qquad \vec{x} \in \partial\Omega, \quad \Omega = (0,1)^m.$$

The standard second order space discretization yields the ordinary differential equation (1.3), where the inhomogeneity $g(t) = g_1(t) + \cdots + g_m(t)$ consists of the discretization of the reaction term $c(t, \vec{x})$ and of the boundary conditions. We assume the splitting to be such that $c(t, \vec{x})$ only contributes to $g_1(t)$.

We further assume that the components of the solution vector $U(t)$ equal the values on the grid (1.2) of a smooth function $v(t, \vec{x})$ that satisfies the boundary condition $v(t, \vec{x}) = b(\vec{x})$ for $\vec{x} \in \partial\Omega$. The time derivative, which we denote by a dot, therefore yields $\dot{v}(t, \vec{x}) = \vec{0}$ for $\vec{x} \in \partial\Omega$. This implies that $D_{l_k} \ddot{U}(t)$ is an approximation to $\partial_{x_{l_k} x_{l_k}} \ddot{v}(t, \vec{x})$ on the grid. This function is bounded and vanishes on $\partial\Omega$ with the exception of the faces where $x_{l_k} \in \{0, 1\}$. In a next step, we notice that $D_{l_{k-1}} D_{l_k} \ddot{U}(t)$ is an approximation to $\partial_{x_{l_{k-1}} x_{l_{k-1}}} \partial_{x_{l_k} x_{l_k}} \ddot{v}(t, \vec{x})$ on the grid, which is bounded and vanishes on $\partial\Omega$, with the exception of the faces where either $x_{l_k} \in \{0, 1\}$ or $x_{l_{k-1}} \in \{0, 1\}$. An induction argument proves the first bound of (4.24). To prove the second bound of (4.24) we just note that $\dot{\varphi}_i(t) = D_i \dot{U}(t)$ for $i \geq 2$, which is a consequence of the choice of the splitting.

THEOREM 4.7. *Assume that* (4.8) *and* (4.24) *hold, and that the stability matrix* (2.9) *is power-bounded, and consider the* 1*-stage method* (2.2) *with* $\theta = 1/2$. *Then, for an initial value satisfying* $\|E_0\|_\infty = \mathcal{O}(\tau^2)$, *the global error is bounded by*

$$\|E_n\|_\infty = \mathcal{O}(\tau^2) \qquad \textit{for} \;\; n\tau \leq T,$$

where the constant symbolized by $\mathcal{O}(\cdot)$ *is independent of* τ *and the spatial discretization, but depends on the constant* C *in* (4.8) *and* (4.24) *and on* T.

Proof. The local error (4.3) can be written as

$$S_n = \frac{\tau^2}{2}\Pi\big(\tfrac12\big)^{-1}\bigg(\sum_{i=1}^m \Pi_i\big(\tfrac12\big)\dot\varphi_i(t_n) - \Pi\big(\tfrac12\big)\ddot U(t_n)\bigg) - \frac{\tau^3}{2}\int_0^1 (1-s)^2\,\dddot U(t_n+s\tau)\,\mathrm{d}s.$$

Since $\ddot U(t) = \dot\varphi_1(t) + \cdots + \dot\varphi_m(t)$, the local error S_n is a linear combination of expressions of the form

$$\begin{aligned}
&\text{(A)} \qquad \tfrac{\tau^3}{2}\textstyle\int_0^1 (1-s)^2\,\dddot U(t_n+s\tau)\,\mathrm{d}s,\\
&\text{(B)} \qquad \tau^{2+k}\Pi\big(\tfrac12\big)^{-1} D_{l_1}D_{l_2}\cdots D_{l_k}\dot\varphi_i(t_n), \quad 1\le l_1<\cdots<l_k<i\le m,\\
&\text{(C)} \qquad \tau^{2+k}\Pi\big(\tfrac12\big)^{-1} D_{l_1}D_{l_2}\cdots D_{l_k}\ddot U(t_n), \quad 1\le l_1<\cdots<l_k\le m,
\end{aligned}$$

where $k\ge 1$. By assumption (4.8) the expression (A) is of size $\mathcal{O}(\tau^3)$. For the maximum norm we have $\|(I-\frac{\tau}{2}D_j)^{-1}\|_\infty \le 1$ for $\tau\ge 0$, so that also $\|\Pi(\frac12)^{-1}\|_\infty\le 1$. By assumption (4.24) the expressions (B) and (C) are thus bounded by $\mathcal{O}(\tau^{2+k})$, with $k\ge 1$. A standard convergence argument then shows that the global error is bounded by $\mathcal{O}(\tau^2)$. □

Remark 4.8. Second order convergence in the maximum norm is also a consequence of [1, Thm. 3.1] under the assumption (4.23). Let us comment on the validity of this assumption. Example 4.6 shows that, assuming time-independent boundary conditions, the entries of the vector $W(t) = D_{l_1}D_{l_2}\cdots D_{l_k}V(t)$ (for $V(t)=\ddot U(t)$ or $V(t)=\dot\varphi_i(t)$) can be considered as the values on the grid of a function $w(t,\vec x)$ that is smooth in the spatial variables. This implies that condition (4.24) is satisfied. From Lemma A.4 (case $k=0$) it follows that $\|D^{-1}W(t)\|_\infty = \mathcal{O}(1)$, so that also condition (4.23) is fulfilled (even without the factor τ^{k-1}).

5. Convergence of the modified 1-stage method. We consider the modification of the 1-stage AMF-W-method (2.2) given by

$$\begin{aligned}
K_1^{(0)} &= \tau D\,U_n + \tau\, g(t_n),\\
(I-\theta\tau D_j)K_1^{(j)} &= K_1^{(j-1)} + \theta\tau^2\dot g_j(t_n+\tau/2), \quad j=1,\ldots,m, \qquad (5.1)\\
U_{n+1} &= U_n + K_1^{(m)},
\end{aligned}$$

where the derivatives $\dot g_j$ are evaluated at $t_n+\tau/2$ and not at t_n. We are mainly interested in the case $\theta=1/2$. The modification (5.1) does not change the computational work, but it provides improved convergence in ℓ_∞ for $m=2$ when time-dependent boundary conditions are imposed. Observe that the stability function is again (2.9). The local error $S_n = S_n(\tau D_1,\ldots,\tau D_m)$ is

$$S_n \;=\; \Pi(\theta)^{-1}\Big(\tau\dot U(t_n) + \theta\tau^2\dot{\mathcal{G}}(t_n+\tfrac{\tau}{2})\Big) - \Big(U(t_n+\tau)-U(t_n)\Big), \qquad (5.2)$$

where the vector $\mathcal{G}(t)$ is given by (2.4). Using the identity (4.2) it can be written as

$$\begin{aligned}
S_n = \Pi(\theta)^{-1}\bigg(\tau\big(\dot U(t_n)-\dot U(t_n+\tfrac{\tau}{2})\big) + \theta\tau^2\sum_{i=1}^m \Pi_i(\theta)\dot\varphi_i(t_n+\tfrac{\tau}{2})\bigg)&\\
- \Big(U(t_n+\tau)-U(t_n)-\tau\dot U(t_n+\tfrac{\tau}{2})\Big).& \qquad (5.3)
\end{aligned}$$

Under the assumption (4.4) this representation of the local error shows that $S_n = \mathcal{O}(\tau^2)$, which proves convergence of order 1, provided that the norm satisfies (4.5) and (4.6).

With the aim of improving the convergence order for $\theta = 1/2$, we expand the functions in (5.3) into a Taylor series around $t_n + \frac{\tau}{2}$. This gives

$$\begin{aligned} S_n &= \frac{\tau^2}{2}\Pi(\tfrac12)^{-1}\bigg(\sum_{i=1}^{m}\Pi_i(\tfrac12)\,\dot\varphi_i(t_n+\tfrac{\tau}{2}) - \ddot U(t_n+\tfrac{\tau}{2})\bigg) + S_n^2 \\ &= \frac{\tau^2}{2}\Pi(\tfrac12)^{-1}\bigg(\sum_{i=2}^{m}\Big(\Pi_i(\tfrac12) - I\Big)\,\dot\varphi_i(t_n+\tfrac{\tau}{2})\bigg) + S_n^2, \qquad (5.4)\\ S_n^2 &= -\frac{\tau^3}{2}\int_0^1 k(s)\,\dddot U(t_n+s\tau)\,\mathrm{d}s + \tau^3\Pi(\tfrac12)^{-1}\int_0^{1/2} s\,\dddot U(t_n+s\tau)\,\mathrm{d}s, \end{aligned}$$

where the kernel $k(s)$ is given by $k(s) = \min\big(s^2, (1-s)^2\big)$. All convergence results of section 4 can be directly transferred to the modified method.

Theorem 5.1. *The statements of Theorems* 4.1, 4.2, 4.4, *and* 4.7 *remain true for the modification* (5.1) *of the* 1*-stage W-method.* □

The advantage of the modified method is that, compared to the representation of S_n in the proof of Theorem 4.7, the expression containing $\ddot U(t)$ is no longer present. This permits us to get an improved convergence result for $m = 2$ space dimensions in the maximum norm and for time-dependent boundary conditions.

Theorem 5.2. *Let* $m = 2$, *assume that* (4.5), (4.6), (4.8), (4.9) *hold, and consider the modified* 1*-stage method* (5.1) *with* $\theta = 1/2$. *Then, for an initial value satisfying* $\|E_0\|_\infty = \mathcal{O}(\tau^2)$, *and for* $h = \min(\Delta x_1, \Delta x_2)$, *the global error is bounded by*

$$\|E_n\|_\infty = \mathcal{O}(\tau^2|\log h|^2) \qquad \text{for} \ \ n\tau \le T,$$

where the constant symbolized by $\mathcal{O}(\cdot)$ *is independent of* τ *and the spatial discretization, but depends on the constant* C *in* (4.5), (4.6), *and* (4.8) *and on* T.

Proof. We consider the representation (5.4) of the local error. By assumption (4.8) we have $\|S_n^2\|_\infty = \mathcal{O}(\tau^3)$, so that its contribution to the global error is of size $\mathcal{O}(\tau^2)$. For $m = 2$, the remaining term of the local error is

$$\widehat{S}_n = -\frac{\tau^3}{4}\Pi(\tfrac12)^{-1}D_1\dot\varphi_2(t_n+\tfrac{\tau}{2}).$$

From Lemma A.4 below (with $k = 1$ and $m = 2$) we get that

$$\|D^{-1}D_1\varphi_2^{(l)}(t)\|_\infty \le C|\log h|^2 \qquad \text{for} \quad l = 1, 2. \tag{5.5}$$

Using $R - I = \Pi(\frac12)^{-1}\tau D$, this implies that

$$\|(I-R)^{-1}\widehat{S}_0\|_\infty \le C_1\tau^2|\log h|^2, \qquad \|(I-R)^{-1}(\widehat{S}_{j+1} - \widehat{S}_j)\|_\infty \le C_2\tau^3|\log h|^2.$$

The power-boundedness of the stability matrix together with (4.11) then yields the desired estimate for the global error. □

Corollary 5.3. *Under the step size restriction* (4.19), *the estimate of Theorem* 5.2 *is equivalent to*

$$\|E_n\|_\infty = \mathcal{O}(\tau^2|\log \tau|^2) \qquad \text{for} \ \ n\tau \le T,$$

where the constant symbolized by $\mathcal{O}(\cdot)$ *is independent of* τ *and the spatial discretization, but depends on the constants* c_0*,* c_2*, and* β *in* (4.19)*, on the constant* C *in* (4.5)*,* (4.6)*, and* (4.8)*, and on* T*.*

Proof. It follows from (4.19) that, for $c_2 h^\beta \le 1$,

$$\widetilde{c}_0 |\log\ h| \ge |\log\ t| \ge \widetilde{c}_1 |\log\ h|,$$

with positive constants $\widetilde{c}_0, \widetilde{c}_1$ depending on c_0, c_2, and β. This implies the stated equivalence. ∎

6. Conclusion.

In this article we have rigorously proved PDE convergence for 1-stage AMF-W-methods in the Euclidean ℓ_2 norm and in the maximum norm ℓ_∞, and we have obtained optimal rates of convergence. We expect the developed techniques to be useful for obtaining

- optimal convergence rates in ℓ_∞ of ADI-type integrators of low order, such as the Peaceman–Rachford, Craig–Sneid, and Douglas schemes;
- optimal convergence rates in ℓ_2 and in ℓ_∞ for s-stage AMF-W-methods with $s \ge 2$.

Further interesting problems are the study of the power-boundedness of the stability matrix in the ℓ_∞ norm, and an extension of the convergence estimates to linear diffusion problems with mixed derivative terms.

Appendix A.

This section collects some technical results that have been used to prove optimal convergence of the 1-stage methods. The estimates for the discrete sine transform of a grid vector U are an essential ingredient and are related to the results of [12, pp. 296–300].

A.1. Properties of the discrete sine transform.

On the interval $[0,1]$ we consider a smooth function $u(x)$, which can have nonzero values at the endpoints. For $n \ge 1$ and $\Delta x = 1/(n+1)$ we put $x^{(i)} = i\Delta x$ for $i = 1, \dots, n$ and $U_i = u(x^{(i)})$, and we write $U = \sum_{j=1}^n \widehat{U}_j \phi_j^{(x)}$. By the orthonormality of the eigenvectors $\phi_j^{(x)}$ of (3.2), the coefficient $\widehat{U}_j$ is given by the discrete sine transform

$$\widehat{U}_j = \sqrt{2}\,\Delta x \sum_{i=1}^n U_i\ \sin(ij\Delta x\pi). \tag{A.1}$$

Interpreted as a Riemann sum, the values $\widehat{U}_k$ are seen to be uniformly bounded when $\Delta x \to 0$. The following lemma yields a sharper bound.

LEMMA A.1. *Let* $U_i = u(x^{(i)})$ *(for* $i = 1, \dots, n$*) with a continuously differentiable function* $u(x)$*. Then the coefficients* (A.1) *fulfill*

$$\widehat{U}_j\ \sin\bigl(j\Delta x \tfrac{\pi}{2}\bigr) = \mathcal{O}(\Delta x), \qquad j = 1, \dots, n,$$

where the constant symbolized by $\mathcal{O}(\cdot)$ *is independent of* n*.*

Proof. Using the identity $2\sin\alpha \sin\beta = \cos(\alpha - \beta) - \cos(\alpha + \beta)$, we get

$$\frac{1}{\Delta x}\widehat{U}_j\ \sin\bigl(j\Delta x \tfrac{\pi}{2}\bigr) = -\frac{\sqrt{2}}{2} \sum_{i=1}^n U_i\ \bigl(V_{i+1} - V_i\bigr),$$

where $V_i = \cos\bigl((i-\frac{1}{2})j\Delta x\pi\bigr)$, $i = 1,\ldots,n+1$. Summation by parts of this relation yields

$$\frac{\sqrt{2}}{\Delta x}\widehat{U}_j\,\sin\bigl(j\Delta x\tfrac{\pi}{2}\bigr) = -\Bigl(U_nV_{n+1} - U_1V_1\Bigr) + \sum_{i=1}^{n-1} V_{i+1}\,\bigl(U_{i+1}-U_i\bigr),$$

which is seen to be $\mathcal{O}(1)$, because $U_{i+1}-U_i = \Delta x\int_0^1 u'(x^{(i)}+\tau\Delta x)\,\mathrm{d}\tau$ contains an additional factor Δx. ◻

Improved bounds can be obtained for the situation, where $u(0) = u(1) = 0$. The property stated in Lemma A.1 can be extended to an arbitrary number of spatial dimensions m. On $[0,1]^m$ we consider a smooth function $u(x_1,\ldots,x_m)$ and a grid vector $U = (U_{i_1,\ldots,i_m}) \in \mathbb{R}^{n_x}$ with $U_{i_1,\ldots,i_m} = u(x_1^{(i_1)},\ldots,x_m^{(i_m)})$ on the interior of the grid (1.2). We write it as

$$U = \sum_{j_1=1}^{n_{x_1}}\cdots\sum_{j_m=1}^{n_{x_m}} \widehat{U}_{j_1,\ldots,j_m}\phi_{j_m}^{(x_m)}\otimes\cdots\otimes\phi_{j_1}^{(x_1)}, \tag{A.2}$$

where, from the orthonormality of the eigenvectors $\phi_{j_l}^{(x_l)}$ of (3.2), the coefficients $\widehat{U}_{j_1,\ldots,j_m}$ are given by

$$\widehat{U}_{j_1,\ldots,j_m} = (\sqrt{2})^m\Delta x_1\cdot\ldots\cdot\Delta x_m\sum_{i_1=1}^{n_{x_1}}\cdots\sum_{i_m=1}^{n_{x_m}} U_{i_1,\ldots,i_m}\prod_{l=1}^{m}\sin\bigl(i_lj_l\Delta x_l\pi\bigr).$$

Lemma A.2. *Let $U_{i_1,\ldots,i_m} = u(x_1^{(i_1)},\ldots,x_m^{(i_m)})$ (for $i_j = 1,\ldots,n_{x_j}$) with a continuously differentiable function $u(x_1,\ldots,x_m)$. Then we have*

$$\widehat{U}_{j_1,\ldots,j_m}\prod_{l=1}^{m}\sin\bigl(j_l\Delta x_l\tfrac{\pi}{2}\bigr) = \mathcal{O}(\Delta x_1\cdot\ldots\cdot\Delta x_m), \qquad j_l = 1,\ldots,n_{x_l},$$

where the constant symbolized by $\mathcal{O}(\cdot)$ is independent of n_{x_l}, $1 \le l \le m$.

Proof. The proof follows along the lines of the proof of Lemma A.1 by using induction on m. Given $m \ge 2$, assume that the statement holds for $m-1$. Then, for

$$\mathcal{P} := \frac{1}{\Delta x_1\cdot\ldots\cdot\Delta x_m}\,\widehat{U}_{j_1,\ldots,j_m}\prod_{l=1}^{m}\sin\bigl(j_l\Delta x_l\tfrac{\pi}{2}\bigr),$$

we have that

$$\begin{aligned}\mathcal{P} &= \Bigl(-\frac{\sqrt{2}}{2}\Bigr)^m\sum_{i_1=1}^{n_{x_1}}\cdots\sum_{i_m=1}^{n_{x_m}} U_{i_1,\ldots,i_m}\prod_{l=1}^{m}\Bigl(V_{i_l+1}^{(x_l)} - V_{i_l}^{(x_l)}\Bigr)\\ &= \Bigl(-\frac{\sqrt{2}}{2}\Bigr)^m\sum_{i_m=1}^{n_{x_m}}\widetilde{U}_{i_m}\Bigl(V_{i_m+1}^{(x_m)} - V_{i_m}^{(x_m)}\Bigr),\end{aligned}$$

where $V_{i_l}^{(x_l)} = \cos\bigl((i_l-\frac{1}{2})j_l\Delta x_l\pi\bigr)$, and

$$\widetilde{U}_{i_m} = \sum_{i_1=1}^{n_{x_1}}\cdots\sum_{i_{m-1}=1}^{n_{x_{m-1}}} U_{i_1,\ldots,i_m}\prod_{l=1}^{m-1}\Bigl(V_{i_l+1}^{(x_l)} - V_{i_l}^{(x_l)}\Bigr).$$

Summation by parts in the relation for $\mathcal{P}$ yields

$$\mathcal{P} = \Big(-\frac{\sqrt{2}}{2}\Big)^m \Big(\widetilde{U}_{n_{x_m}} V^{(x_m)}_{n_{x_m}+1} - \widetilde{U}_1 V^{(x_m)}_1 - \sum_{i_m=1}^{n_{x_m}-1} V^{(x_m)}_{i_m+1}\Big(\widetilde{U}_{i_m+1} - \widetilde{U}_{i_m}\Big)\Big).$$

The induction hypothesis implies that $\widetilde{U}_{i_m} = \mathcal{O}(1)$, $i_m = 1,\dots,n_{x_m}$. Furthermore, $\widetilde{U}_{i_m+1} - \widetilde{U}_{i_m} = \mathcal{O}(\Delta x_m)$, because $u(x_1,\dots,x_m)$ is continuously differentiable in the variable x_m. This implies $\mathcal{P} = \mathcal{O}(1)$ and proves the statement of the lemma. □

A.2. Operator estimates in the ℓ_2 norm and in the ℓ_∞ norm. Here, we prove estimates for $D^{-1}D_{l_1}D_{l_2}\cdots D_{l_k}U$ that were used in Remarks 4.5 and 4.8.

We first consider the ℓ_2 norm. In this case, negative powers of h may arise when $k \ge 2$. Observe that for $k = 1$ it holds that $\|D^{-1}D_{l_1}\|_2 \le 1$. This follows by diagonalization and (3.3), because $|\lambda^{(x_1)}_{i_1}/(\lambda^{(x_1)}_{i_1} + \cdots + \lambda^{(x_m)}_{i_m})| \le 1$ for all $i_1,\dots,i_m$.

Lemma A.3. *Let $U \in \mathbb{R}^{n_x}$ be the restriction of a continuously differentiable function $u(x_1,\dots,x_m)$ to the interior points of the grid* (1.2). *For the ℓ_2 norm, we then have for distinct indices $l_1,\dots,l_k$ and $2 \le k \le m$ that*

$$\|D^{-1}D_{l_1}D_{l_2}\cdots D_{l_k}U\|_2 = \mathcal{O}\big(h^{(4-3k)/2}\big).$$

Proof. Without loss of generality and for ease of notation, we assume $l_j = j$ for $j = 1,\dots,k$. Writing U as in (A.2), in the basis of eigenvectors we obtain

$$E_2 := \|D^{-1}D_1D_2\cdots D_kU\|_2^2 = \sum_{i_1=1}^{n_{x_1}}\cdots\sum_{i_m=1}^{n_{x_m}} |\widehat{U}_{i_1,\dots,i_m}|^2 \frac{|\lambda^{(x_1)}_{i_1}|^2\cdot\ldots\cdot|\lambda^{(x_k)}_{i_k}|^2}{(|\lambda^{(x_1)}_{i_1}| + \cdots + |\lambda^{(x_m)}_{i_m}|)^2}.$$

From Lemma A.2 and (3.3) we get $|\widehat{U}_{i_1,\dots,i_m}|^2|\lambda^{(x_1)}_{i_1}|\cdot\ldots\cdot|\lambda^{(x_m)}_{i_m}| \le C$ for all $i_1,\dots,i_m$. Neglecting the term $|\lambda^{(x_{k+1})}_{i_{k+1}}| + \cdots + |\lambda^{(x_m)}_{i_m}|$ in the denominator, and using the arithmetic mean–geometric mean (AM-GM) inequality

$$|\lambda^{(x_1)}_{i_1}| + \cdots + |\lambda^{(x_k)}_{i_k}| \ge k\cdot\sqrt[k]{|\lambda^{(x_1)}_{i_1}|\cdot\ldots\cdot|\lambda^{(x_k)}_{i_k}|}, \tag{A.3}$$

we have that

$$E_2 \le C\sum_{i_1=1}^{n_{x_1}}\cdots\sum_{i_m=1}^{n_{x_m}} |\lambda^{(x_1)}_{i_1}|^{1-\frac{2}{k}}\cdot\ldots\cdot|\lambda^{(x_k)}_{i_k}|^{1-\frac{2}{k}}\cdot|\lambda^{(x_{k+1})}_{i_{k+1}}|^{-1}\cdot\ldots\cdot|\lambda^{(x_m)}_{i_m}|^{-1}. \tag{A.4}$$

Now, from (3.3) and [12, Lem. 6.2], we get

$$\sum_{i_l=1}^{n_{x_l}} |\lambda^{(x_l)}_{i_l}|^{-\beta} = \begin{cases} \mathcal{O}(h^{2\beta-1}) & \text{if } \beta < 1/2,\\ \mathcal{O}\big(|\log h|\big) & \text{if } \beta = 1/2,\\ \mathcal{O}(1) & \text{if } \beta > 1/2. \end{cases} \tag{A.5}$$

Inserted into (A.4), these estimates give $E_2 = \mathcal{O}\big((h^{-3+\frac{4}{k}})^k\big)$, which completes the proof of the lemma. □

For the ℓ_∞ norm we get the following bounds.

LEMMA A.4. *Let $U \in \mathbb{R}^{n_x}$ be the restriction of a continuously differentiable function $u(x_1, \dots, x_m)$ to the interior points of the grid (1.2). For the ℓ_∞ norm, we then have for distinct indices $l_1, \dots, l_k$ that*

$$\|D^{-1}D_{l_1}D_{l_2}\cdots D_{l_k}U\|_\infty = \begin{cases} \mathcal{O}(1) & \text{if } k = 0, \\ \mathcal{O}\big(|\log h|^m\big) & \text{if } k = 1, \\ \mathcal{O}\big(h^{2-2k}|\log h|^{m-k}\big) & \text{if } 2 \le k \le m. \end{cases}$$

Proof. As in the proof of Lemma A.3 we assume that $l_j = j$ for $j = 1, \dots, k$, and we write U in the basis of eigenvectors. This yields

$$E_\infty := \|D^{-1}D_1D_2\cdots D_kU\|_\infty \le C\sum_{i_1=1}^{n_{x_1}}\cdots\sum_{i_m=1}^{n_{x_m}} |\widehat{U}_{i_1,\dots,i_m}| \frac{|\lambda_{i_1}^{(x_1)}|\cdot\ldots\cdot|\lambda_{i_k}^{(x_k)}|}{|\lambda_{i_1}^{(x_1)}|+\cdots+|\lambda_{i_m}^{(x_m)}|}.$$

As a consequence of $|\widehat{U}_{i_1,\dots,i_m}||\lambda_{i_1}^{(x_1)}|^{\frac12}\cdot\ldots\cdot|\lambda_{i_m}^{(x_m)}|^{\frac12} \le C$ (for all $i_1, \dots, i_m$) we have

$$E_\infty \le C'\sum_{i_1=1}^{n_{x_1}}\cdots\sum_{i_m=1}^{n_{x_m}} \frac{|\lambda_{i_1}^{(x_1)}|^{\frac12}\cdot\ldots\cdot|\lambda_{i_k}^{(x_k)}|^{\frac12}\cdot|\lambda_{i_{k+1}}^{(x_{k+1})}|^{-\frac12}\cdot\ldots\cdot|\lambda_{i_m}^{(x_m)}|^{-\frac12}}{|\lambda_{i_1}^{(x_1)}|+\cdots+|\lambda_{i_m}^{(x_m)}|}.$$

For the case $k = 0$ we use the AM-GM inequality (A.3) for $|\lambda_{i_1}^{(x_1)}|+\cdots+|\lambda_{i_m}^{(x_m)}|$ and the bound (A.5) with $\beta = \frac12 + \frac1m$. This yields $E_\infty = \mathcal{O}(1)$.

For the case $k \ge 1$ we first bound the denominator $|\lambda_{i_1}^{(x_1)}|+\cdots+|\lambda_{i_m}^{(x_m)}|$ from below by $|\lambda_{i_1}^{(x_1)}|+\cdots+|\lambda_{i_k}^{(x_k)}|$ and then apply the AM-GM inequality (A.3). The bounds (A.5), with $\beta = -\frac12 + \frac1k$ for $1 \le l \le k$ and with $\beta = \frac12$ for $k+1 \le l \le m$, then yield the statement of the lemma. ∎

A.3. Some auxiliary lemmas.

LEMMA A.5. *Let $\theta > 0$, let $n \ge 1$ be a positive integer, and let $h = 1/(n+1)$. If $\tau \ge ch^2$ for some constant $c > 0$, then for all $\alpha \ge 0$ and $\gamma \ge 0$ there exists a constant C independent of τ and h such that*

(A.6)
$$\sum_{i=1}^{n} \frac{h^{-\alpha}\sin^\alpha(ih\frac{\pi}{2})}{\big(1+\theta\tau h^{-2}\sin^2(ih\frac{\pi}{2})\big)^\gamma} \le \begin{cases} C\tau^{-\gamma}h^{2\gamma-\alpha-1} & \text{if } \alpha+1-2\gamma > 0, \\ C\tau^{-\gamma}\big(1+\big|\log\big(\frac{\tau}{h^2}\big)\big|\big) & \text{if } \alpha+1-2\gamma = 0, \\ C\tau^{-(\alpha+1)/2} & \text{if } \alpha+1-2\gamma < 0. \end{cases}$$

Proof. In the denominator of (A.6) the term $\theta\tau h^{-2}\sin^2(ih\frac{\pi}{2})$ becomes dominant over 1 when $\tau h^{-2}(ih)^2 > 1$, i.e., $i \gtrsim \tau^{-1/2}$. We therefore separate the sum in (A.6) into two terms, $\mathcal{S}_l + \mathcal{S}_r$, where $\mathcal{S}_l$ denotes the sum over the index set $\mathcal{I}_l = \{i \in \{1, \dots, n\} \,|\, i \le \tau^{-1/2}\}$, and $\mathcal{S}_r$ is the sum over the remaining indices, $\mathcal{I}_r = \{1, \dots, n\} \setminus \mathcal{I}_l$.

Since $\sin x \le x$, for all $x \in [0, \frac{\pi}{2}]$, the sum $\mathcal{S}_l$ can be bounded by

$$\mathcal{S}_l \le h^{-\alpha}\sum_{i\in\mathcal{I}_l}\sin^\alpha(ih\tfrac{\pi}{2}) \le \left(\frac{\pi}{2}\right)^\alpha\sum_{i\in\mathcal{I}_l} i^\alpha \le C_l\tau^{-(\alpha+1)/2}. \tag{A.7}$$

For the other sum we get

$$\mathcal{S}_r \le \sum_{i\in\mathcal{I}_r}\frac{h^{-\alpha}\sin^\alpha(ih\frac{\pi}{2})}{(\theta\tau)^\gamma h^{-2\gamma}\sin^{2\gamma}(ih\frac{\pi}{2})} = \theta^{-\gamma}\tau^{-\gamma}h^{2\gamma-\alpha}\sum_{i\in\mathcal{I}_r}\sin^{\alpha-2\gamma}\left(ih\tfrac{\pi}{2}\right).$$

Using $\frac{x}{2} \le \sin x \le x$, depending on the sign of $\alpha - 2\gamma$, we obtain

$$\mathcal{S}_r \le C_r \tau^{-\gamma} \sum_{i \in \mathcal{I}_r} i^{\alpha - 2\gamma} \le C_r' \tau^{-\gamma} \int_{\tau^{-1/2}}^{n+1} x^{\alpha - 2\gamma}\, dx. \tag{A.8}$$

If $\alpha+1-2\gamma > 0$, then $\mathcal{S}_r \le C_r'' \tau^{-\gamma} \left((h^{-1})^{\alpha-2\gamma+1} - (\tau^{-1/2})^{\alpha-2\gamma+1} \right)$. As a consequence of $\tau \ge ch^2$, the term with $\tau^{-\gamma} h^{2\gamma-\alpha-1}$ dominates that with $\tau^{-(\alpha+1)/2}$, so that $\mathcal{S}_l + \mathcal{S}_r \le C\tau^{-\gamma} h^{2\gamma-\alpha-1}$.

If $\alpha + 1 - 2\gamma = 0$, it follows from (A.8) that $\mathcal{S}_r \le C_r' \tau^{-\gamma} \log \left(h^{-1}/\tau^{-1/2} \right)$. This proves the statement for this case.

If $\alpha + 1 - 2\gamma < 0$, we get $\mathcal{S}_r \le C_r'' \tau^{-\gamma} \left((\tau^{-1/2})^{\alpha-2\gamma+1} - h^{2\gamma-\alpha-1} \right)$ from (A.8). In this case $\tau^{-(\alpha+1)/2}$ dominates $\tau^{-\gamma} h^{2\gamma-\alpha-1}$ because of $\tau \ge ch^2$. Therefore, we obtain $\mathcal{S}_l + \mathcal{S}_r \le C\tau^{-(\alpha+1)/2}$, which completes the proof. □

LEMMA A.6. *Let* $\gamma \in [0, 2]$. *For all* $n \ge 1$ *and* $x \in [0, 2]$*, it holds that*

$$(1 - (1 - x)^n)^2 \le 2^{2-\gamma} (nx)^\gamma.$$

Proof. Consider the function

$$f(x) := (1 - (1 - x)^n)^2 - 2^{2-\gamma} (nx)^\gamma.$$

For $x \in [0, \frac{2}{n}]$ it follows from Bernoulli's inequality $(1 - x)^n \ge 1 - nx$ that

$$f(x) \le (nx)^2 - 2^{2-\gamma}(nx)^\gamma = (nx)^\gamma \left((nx)^{2-\gamma} - 2^{2-\gamma} \right) \le 0.$$

For $x \in [\frac{2}{n}, 2]$ we have

$$2^{2-\gamma}(nx)^\gamma \ge 2^{2-\gamma} 2^\gamma = 4 \quad \text{and} \quad f(x) \le (1 - (1 - x)^n)^2 - 4 \le 0.$$

This concludes the proof. □

REFERENCES

[1] A. ARRARÁS, K. J. IN 'T HOUT, W. HUNDSDORFER, AND L. PORTERO, *Modified Douglas splitting methods for reaction-diffusion equations*, BIT, 57 (2017), pp. 261–285.

[2] R. M. BEAM AND R. F. WARMING, *An implicit finite-difference algorithm for hyperbolic systems in conservation-law form*, J. Comput. Phys., 22 (1976), pp. 87–110.

[3] N. BOROVYKH, D. DRISSI, AND M. N. SPIJKER, *A bound on powers of linear operators, with relevance to numerical stability*, Appl. Math. Lett., 15 (2002), pp. 47–53.

[4] I. FARAGÓ AND 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), pp. 133–140.

[5] S. GONZÁLEZ-PINTO, E. HAIRER, D. HERNÁNDEZ-ABREU, AND S. PÉREZ-RODRÍGUEZ, *AMF-type W-methods for parabolic problems with mixed derivatives*, SIAM J. Sci. Comput., 40 (2018), pp. A2905–A2929, https://doi.org/10.1137/17M1163050.

[6] S. GONZÁLEZ-PINTO, D. HERNÁNDEZ-ABREU, AND S. PÉREZ-RODRÍGUEZ, *Rosenbrock-type methods with inexact AMF for the time integration of advection diffusion reaction PDEs*, J. Comput. Appl. Math., 262 (2014), pp. 304–321.

[7] E. HAIRER AND G. WANNER, *Solving Ordinary Differential Equations* II. *Stiff and Differential-Algebraic Problems*, 2nd ed.,Springer Ser. Comput. Math. 14, Springer-Verlag, Berlin, 1996.

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

[9] W. HUNDSDORFER, *Stability of approximate factorization with θ-methods*, BIT, 39 (1999), pp. 473–483.

[10] W. Hundsdorfer and K. J. in 't Hout, *On multistep stabilizing correction splitting methods with applications to the Heston model*, SIAM J. Sci. Comput., 40 (2018), pp. A1408–A1429, https://doi.org/10.1137/17M1146026.

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

[12] W. Hundsdorfer and J. G. Verwer, *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*, Springer Ser. Comput. Math. 33, Springer-Verlag, Berlin, 2003.

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

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

[15] A. Ostermann and M. Roche, *Rosenbrock methods for partial differential equations and fractional orders of convergence*, SIAM J. Numer. Anal., 30 (1993), pp. 1084–1098, https://doi.org/10.1137/0730056.

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

[17] S. J. Serdyukova, *The uniform stability with respect to initial data of a six-point symmetrical scheme for the heat conduction equation*, in Numerical Methods for the Solution of Differential and Integral Equations and Quadrature Formulae, Nauka, Moscow, 1964, pp. 212–216 (in Russian).

[18] M. N. Spijker, *Lecture Notes on Stability Estimates and Resolvent Conditions in the Numerical Solution of Initial Value Problems*, preprint, University of Leiden, 1998.

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

[20] F. A. J. Straetemans, *Resolvent conditions for discretizations of diffusion-convection-reaction equations in several space dimensions*, Appl. Numer. Math., 28 (1998), pp. 45–67.

[21] K. Strehmel and R. Weiner, *Linear-implizite Runge-Kutta-Methoden und ihre Anwendung*, Teubner-Texte Math. 127, B. G. Teubner Verlagsgesellschaft mbH, Stuttgart, 1992, with English, French, and Russian summaries.

[22] J. G. Verwer, E. J. Spee, J. G. Blom, and W. Hundsdorfer, *A second-order Rosenbrock method applied to photochemical dispersion problems*, SIAM J. Sci. Comput., 20 (1999), pp. 1456–1480, https://doi.org/10.1137/S1064827597326651.