SECOND ORDER IMPLICIT FOR LOCAL EFFECTS AND EXPLICIT FOR NONLOCAL EFFECTS IS UNCONDITIONALLY STABLE

CATALIN TRENCHEA†

Abstract. A family of implicit-explicit second order time-stepping methods is analyzed for a system of ODEs motivated by ones arising from spatial discretizations of evolutionary partial differential equations. The methods we consider are implicit in local and stabilizing terms in the underlying PDE and explicit in nonlocal and unstabilizing terms. Unconditional stability and convergence of the numerical scheme are proved by the energy method and by algebraic techniques. This is the first solution to the problem of finding a scheme for (1.1) that is (provably) unconditionally stable and treats the Cu term explicitly. First order schemes were known in [2, 10] and [10] gives a second order scheme stable provided all operators commute.

Key words. Unconditional stability, IMEX methods, Crank-Nicolson, BDF2

AMS subject classifications. 76D05, 65L20, 65M12

1. Introduction.

In this report we consider the system of ordinary differential equations of the form

$$u'(t) + Au(t) - Cu(t) + B(u)u(t) = f(t), \tag{1.1}$$

in which $A, B(u)$ and C are $n \times n$ matrices, $u(t)$ and $f(t)$ are n-vectors, and

$$A = A^T \succ 0, B(u) = -B(u)^T, C = C^T \succcurlyeq 0 \text{ and } \quad A - C \succ 0. \tag{1.2}$$

Here $\succ$ and $\succcurlyeq$ denote the positive definite and positive semidefinite, respectively. The key properties motivating the analysis are that A is sparse and that although C is not sparse, the action of C on a vector is inexpensive to calculate. This structure is motivated by multiscale discretizations of turbulence, but can also arise from closed-loop control problems and ensemble calculations. Given this structure of (1.1), the simplest scheme that is computationally feasible is *explicit* in the global, unstable part of (1.1), that is, Cu. Our study extends the work in [2], where the implicit-explicit [3, 1, 4, 16, 17] first order method

$$\frac{u_{n+1} - u_n}{\Delta t} + Au_{n+1} - Cu_n + B(u_n)u_{n+1} = f_{n+1} \tag{1.3}$$

was proved to be *unconditionally* stable. Here we analyze the stability of a family of three level, second order time stepping schemes:

$$\begin{aligned} &\frac{(\theta + \frac{1}{2})u_{n+1} - 2\theta u_n + (\theta - \frac{1}{2})u_{n-1}}{\Delta t} \\ &+ (A-C)^{\frac{1}{2}} A (A-C)^{-\frac{1}{2}} (\theta u_{n+1} + (1-\theta)u_n) - (A-C)^{\frac{1}{2}} C (A-C)^{-\frac{1}{2}} ((\theta+1)u_n - \theta u_{n-1}) \\ &+ B(\mathcal{E}_{n+\theta})(A-C)^{-\frac{1}{2}} \Big(\theta A (A-C)^{-\frac{1}{2}} u_{n+1} + \big((1-\theta)A - (\theta+1)C\big)(A-C)^{-\frac{1}{2}} u_n + C(A-C)^{-\frac{1}{2}} \theta u_{n-1} \Big) \\ &\quad = f_{n+\theta}, \end{aligned} \tag{1.4}$$

where $\mathcal{E}_{n+\theta} = (\theta+1)u_n - \theta u_{n-1}$ is an explicit second order approximation of $u_{n+\theta}$. The parameter $\theta \in [\frac{1}{2}, 1]$, yielding for $\theta = \frac{1}{2}$ the IMEX Crank-Nicolson with linear extrapolation

$$\begin{aligned} &\frac{u_{n+1} - u_n}{\Delta t} + (A-C)^{\frac{1}{2}} A (A-C)^{-\frac{1}{2}} \frac{u_{n+1} + u_n}{2} - (A-C)^{\frac{1}{2}} C (A-C)^{-\frac{1}{2}} \big(\tfrac{3}{2}u_n - \tfrac{1}{2}u_{n-1}\big) \\ &+ B(\mathcal{E}_{n+\theta})(A-C)^{-\frac{1}{2}} \Big(\tfrac{1}{2} A (A-C)^{-\frac{1}{2}} u_{n+1} + \big(\tfrac{1}{2}A - \tfrac{3}{2}C\big)(A-C)^{-\frac{1}{2}} u_n + \tfrac{1}{2} C (A-C)^{-\frac{1}{2}} u_{n-1} \Big) = f_{n+1}, \end{aligned} \tag{1.5}$$

†Department of Mathematics, 301 Thackeray Hall, University of Pittsburgh, Pittsburgh, PA 15260, Email: trenchea@pitt.edu. Partially supported by Air Force grant FA 9550-09-1-0058.

while $\theta = 1$ corresponds to BDF2 with linear extrapolation

$$\frac{3u_{n+1}-4u_n+u_{n-1}}{2\Delta t} + (A-C)^{\frac{1}{2}}A(A-C)^{-\frac{1}{2}}u_{n+1} - (A-C)^{\frac{1}{2}}C(A-C)^{-\frac{1}{2}}(2u_n-u_{n-1}) \tag{1.6}$$
$$+ B(\mathcal{E}_{n+\theta})(A-C)^{-\frac{1}{2}}\big(A(A-C)^{-\frac{1}{2}}u_{n+1}-2C(A-C)^{-\frac{1}{2}}u_n+C(A-C)^{-\frac{1}{2}}u_{n-1}\big) = f_{n+1}.$$

In Theorem 2.3 we prove by energy methods that, under assumptions (1.2), the methods (1.4) are also *unconditionally* stable.

REMARK 1.1. *Unfortunately, in its present form, the method proposed is more appropriate for spectral methods, where the evaluation of* $(A-C)^{-\frac{1}{2}}$ *is undemanding. Nonetheless, it is the first attempt to solve the open problem (see Remark 1.3 below) of proving by energy estimates the unconditional stability of second-order schemes in the general case of non-commuting matrices [2, 10].*

REMARK 1.2. *If the matrices* A *and* C *commute, then the computation of the inverse matrix* $(A-C)^{-\frac{1}{2}}$ *is replaced by an extra solve of a linear system at each time step. Indeed, the method (1.4) becomes*

$$\frac{(\theta+\frac{1}{2})u_{n+1} - 2\theta u_n + (\theta-\frac{1}{2})u_{n-1}}{\Delta t} + \theta A u_{n+1} + \big((1-\theta)Au_n - (\theta+1)C\big)u_n + \theta C u_{n-1} \tag{1.7}$$
$$+ B(\mathcal{E}_{n+\theta})\Big(\theta A(A-C)^{-1}u_{n+1} + \big((1-\theta)A-(\theta+1)C\big)(A-C)^{-1}u_n + \theta C(A-C)^{-1}u_{n-1}\big)\Big) = f_{n+\theta},$$

and therefore it can be written as:

(i) solve for v^{n+1}:

$$(A-C)\Big(\frac{(\theta+\frac{1}{2})v_{n+1} - 2\theta v_n + (\theta-\frac{1}{2})v_{n-1}}{\Delta t} + \theta A v_{n+1} + \big((1-\theta)Av_n - (\theta+1)C\big)v_n + \theta C v_{n-1}\Big)$$
$$+ B(\mathcal{E}_{n+\theta})\Big(\theta A v_{n+1} + \big((1-\theta)A-(\theta+1)C\big)v_n + \theta C v_{n-1}\big)\Big) = f_{n+\theta},$$

(ii) solve for u^{n+1}:

$$(A-C)v^{n+1} = u^{n+1}.$$

REMARK 1.3. *There is a an open question in J.-G. Liu's paper [10] on the stability proof, using energy estimates, for the second order IMEX scheme for the pressure projections equation formulation applied to the Stokes equation. The approximation uses second order Adams-Bashforth on the pressure term and Crank-Nicholson for the discretization of the viscous term, namely*

$$\frac{v^{n+1}-v^n}{\Delta t} - \nu\Delta\frac{v^{n+1}+v^n}{2} + \nu\mathscr{B}\frac{3v^n - v^{n-1}}{2} = 0.$$

Here $\mathscr{B} = \partial_y(\Delta_N\Delta - I)\partial_y$ *denotes the pressure operator, and is dominated by* $-\Delta$*, see [10, observations (3.9)-3.10], therefore falling under our assumptions. Moreover,* $\mathscr{B}$ *commutes with the Laplacian* Δ*, also allowing a proof by normal modes. Taking* $\theta = \frac{1}{2}$*,* $A := -\nu\Delta, C := -\nu\mathscr{B}, B = 0$*, with the proper adjustments of notation, Theorem 2.3 gives a positive answer to the problem in [10].*

OPEN PROBLEM 1. *Is there a second-order unconditionally stable method that does not necessitates the evaluation of the inverse matrix* $(A-C)^{-\frac{1}{2}}$*?*

We remark that the study of (1.1) can be partially motivated by the following problems.

Turbulent dispersion The basic model of the turbulent dispersion is that it is dissipative in the mean (see [15]). A more accurate formulation is that its dissipative effects are focused on the smallest resolved scales (see [9]). This physical idea has led to algorithms for numerical stabilization of transport-dominated phenomena based on eddy diffusivity acting only on the smallest resolved scales (e.g., [11, 6, 5]). The

natural realization of this idea for spatial discretizations of convection diffusion equations is, diffusive stabilization on all scales and then antidiffusing on the large scales. This leads to the system of ODEs

$$\dot{u}_{ij}(t) + b \cdot \nabla^h u_{ij} - (\epsilon_0(h) + \nu)\Delta^h u_{ij} + \epsilon_0(h) P_H(\Delta^h P_H(u_{ij})) = f_{ij}, \tag{1.8}$$

where standard notation is used: Δ^h is the discrete Laplacian, $\epsilon(h)$ is the artificial viscosity parameters and P_H denotes a projection onto a coarser mesh. The system (1.8) fits exactly the form (1.1), (1.2), where C is provided as the matrix arising from $\epsilon_0(h)$ term.

Spatial regularization Consider a regularizing operator $G_h : X \to V_h$ that satisfies $\langle G_h(v), v\rangle \leq \|v\|^2$. Such examples include the discrete differential and discrete Stokes differential filters, nonlinear filters, VMS regularization operator, and approximate deconvolution operator (see e.g. [13] and references therein for more details). Then (1.1) can be obtained by spatial discretization of

$$\begin{aligned} &u'(t) + u \cdot \nabla u - (\nu + \epsilon(h))\Delta u + \epsilon(h)\Delta G_h(u) + \nabla \pi = f, \\ &\nabla \cdot u = 0, \end{aligned} \qquad \text{in } \Omega \times (0,T).$$

Lions' hyperviscosity model [14]

$$\begin{aligned} &u'(t) + u \cdot \nabla u - \nu(\Delta + \epsilon(-\Delta)^\alpha)u + \epsilon(-\Delta)^\alpha u + \nabla \pi = f, \\ &\nabla \cdot u = 0, \end{aligned} \qquad \text{in } \Omega \times (0,T).$$

Nonlinear viscosity models: modified NSE of Ladyzhenskaya [12] and Smagorinsky (with $p = 3$):

$$\begin{aligned} &u'(t) + u \cdot \nabla u - \nabla \cdot (\nu \nabla u + \epsilon_\delta |\nabla u|^{p-2}\nabla u) + \nabla \cdot (\epsilon_\delta |\nabla u|^{p-2}\nabla u) + \nabla \pi = f, \\ &\nabla \cdot u = 0, \end{aligned} \qquad \text{in } \Omega \times (0,T).$$

Nonlinear spectral eddy-viscosity models of turbulence [7]

$$\begin{aligned} &u'(t) + u \cdot \nabla u - (\nu\Delta + \epsilon(-\Delta)^\alpha Q)u + \epsilon(-\Delta)^\alpha Q u + \nabla \pi = f, \\ &\nabla \cdot u = 0, \end{aligned} \qquad \text{in } \Omega \times (0,T).$$

and

$$\begin{aligned} &u'(t) + u \cdot \nabla u - (\nu\Delta u + \epsilon Q \nabla \cdot (|\nabla Q u|^{p-2}\nabla Q u)) + \epsilon Q \nabla \cdot ((|\nabla Q u|^{p-2}\nabla Q u)) + \nabla \pi = f, \\ &\nabla \cdot u = 0, \end{aligned}$$

in $\Omega \times (0,T)$, where Q is a high-pass filter, erasing all the low-frequency modes of the input (damping is applied only to the high frequency part of the solution).

2. Stability Analysis. In what follows, we use the same notation for the inner product in $\mathbb{R}^{n\times n}$ and $\mathbb{R}^{2n\times 2n}$, namely $\langle\cdot,\cdot\rangle$. First let note that by multiplication with

$$(A-C)^{-\frac{1}{2}}\Big(\theta A(A-C)^{-\frac{1}{2}}u_{n+1} + \big((1-\theta)A - (\theta+1)C\big)(A-C)^{-\frac{1}{2}}u_n + \theta C(A-C)^{-\frac{1}{2}}u_{n-1}\Big),$$

the diffusive term in (1.4) gives

$$\begin{aligned} &\Big\langle (A-C)^{\frac{1}{2}}\Big(\theta A(A-C)^{-\frac{1}{2}}u_{n+1} + \big((1-\theta)A - (\theta+1)C\big)(A-C)^{-\frac{1}{2}}u_n - \theta C(A-C)^{-\frac{1}{2}}u_{n-1}\Big), \\ &\quad , (A-C)^{-\frac{1}{2}}\Big(\theta A(A-C)^{-\frac{1}{2}}u_{n+1} + \big((1-\theta)A - (\theta+1)C\big)(A-C)^{-\frac{1}{2}}u_n + \theta C(A-C)^{-\frac{1}{2}}u_{n-1}\Big)\Big\rangle \\ &= \|\theta A(A-C)^{-\frac{1}{2}}u_{n+1} + \big((1-\theta)A - (\theta+1)C\big)(A-C)^{-\frac{1}{2}}u_n - \theta C(A-C)^{-\frac{1}{2}}u_{n-1}\|^2, \end{aligned} \tag{2.1}$$

while the convective term vanishes. Next we define the symmetric positive matrix $F \in \mathbb{R}^{n\times n}$ by

$$F = (A-C)^{-\frac{1}{2}}\big(\theta(2\theta-1)A+\theta(2\theta+1)C\big)(A-C)^{-\frac{1}{2}}, \tag{2.2}$$

and the symmetric matrix $G \in \mathbb{R}^{2n\times 2n}$ as follows

$$G=\begin{pmatrix} (A-C)^{-\frac{1}{2}}\Big(\frac{\theta(2\theta+3)}{4}A - \frac{\theta(2\theta+1)}{4}C\Big)(A-C)^{-\frac{1}{2}} & -(A-C)^{-\frac{1}{2}}\Big(\frac{(\theta+1)(2\theta-1)}{4}A + \frac{(1-\theta)(2\theta+1)}{4}C\Big)(A-C)^{-\frac{1}{2}} \\ -(A-C)^{-\frac{1}{2}}\Big(\frac{(\theta+1)(2\theta-1)}{4}A + \frac{(1-\theta)(2\theta+1)}{4}C\Big)(A-C)^{-\frac{1}{2}} & (A-C)^{-\frac{1}{2}}\Big(\frac{\theta(2\theta-1)}{4}A + \frac{\theta(-2\theta+3)}{4}C\Big)(A-C)^{-\frac{1}{2}} \end{pmatrix}. \tag{2.3}$$

We denote by $\left\|\begin{bmatrix}u\\v\end{bmatrix}\right\|_G^2$ the following scalar function of the $2n$ vector $\begin{bmatrix}u\\v\end{bmatrix}$:

$$\left\|\begin{bmatrix}u\\v\end{bmatrix}\right\|_G^2 = \left\langle \begin{bmatrix}u\\v\end{bmatrix}, G\begin{bmatrix}u\\v\end{bmatrix}\right\rangle, \tag{2.4}$$

whose values could be negative.

LEMMA 2.1. *Under the assumption (1.2), for any vectors $u, v \in \mathbb{R}^N$, we have*

$$\begin{aligned}\left\langle \begin{bmatrix}u\\v\end{bmatrix}, G\begin{bmatrix}u\\v\end{bmatrix}\right\rangle &= \tfrac{2\theta+1}{4}u^Tu + \tfrac{-2\theta+1}{4}v^Tv \\ &\quad + \tfrac{(\theta+1)(2\theta-1)}{2}(u-v)^T(u-v) + \tfrac{\theta}{2}(u-v)^T(A-C)^{-\frac{1}{2}}C(A-C)^{-\frac{1}{2}}(u-v) \\ &\geq \tfrac{2\theta+1}{4}\|u\|^2 - \tfrac{2\theta-1}{4}\|v\|^2.\end{aligned} \tag{2.5}$$

Proof. The identity (2.5) follows from algebraic manipulations, while the inequality yields from the positive-definiteness of $(A-C)^{-\frac{1}{2}}C(A-C)^{-\frac{1}{2}}$. ∎

We note that for $\theta = \frac{1}{2}$ the matrix G defined in (2.3) is symmetric and positive definite, and therefore the expression defined in (2.4) is a G-norm.

LEMMA 2.2. *Let u_n satisfy (1.4) for all $n \in \{2, \dots, \frac{T}{\Delta t}\}$. Then*

$$\begin{aligned}&\frac{1}{\Delta t}\Big\langle \big(\theta+\tfrac{1}{2}\big)u^{n+1} - 2\theta u^n + \big(\theta-\tfrac{1}{2}\big)u^{n-1}, \\ &\quad , (A-C)^{-\frac{1}{2}}\Big(\theta A(A-C)^{-\frac{1}{2}}u_{n+1} + \big((1-\theta)A-(\theta+1)C\big)(A-C)^{-\frac{1}{2}}u_n + \theta C(A-C)^{-\frac{1}{2}}u_{n-1}\Big)\Big\rangle \\ &= \frac{1}{\Delta t}\left\|\begin{bmatrix}u_{n+1}\\u_n\end{bmatrix}\right\|_G^2 - \frac{1}{\Delta t}\left\|\begin{bmatrix}u_n\\u_{n-1}\end{bmatrix}\right\|_G^2 + \frac{1}{4\Delta t}\|u_{n+1}-2u_n+u_{n-1}\|_F^2.\end{aligned} \tag{2.6}$$

Proof. The form of G-matrix (2.3) and the G-stability result (2.6) follows from standard calculations, see e.g. [8, Chapter V.6] and references therein. ∎

THEOREM 2.3. *Assuming that (1.2) holds, let u_n satisfy (1.4), with u_0, u_1 given, and $\theta \in [\frac{1}{2}, 1]$. Then*

the following equality holds

$$\begin{aligned}
&\frac{1}{\Delta t}\left\|\begin{bmatrix} u_N \\ u_{N-1}\end{bmatrix}\right\|_G^2 + \frac{1}{4\Delta t}\sum_{n=1}^{N-1}\|u_{n+1}-2u_n+u_{n-1}\|_F^2 \\
&\quad + \sum_{n=1}^{N-1}\|\theta A(A-C)^{-\frac{1}{2}}u_{n+1} + \big((1-\theta)A-(\theta+1)C\big)(A-C)^{-\frac{1}{2}}u_n + \theta C(A-C)^{-\frac{1}{2}}u_{n-1}\|^2 \\
&= \frac{1}{\Delta t}\left\|\begin{bmatrix} u_1 \\ u_0\end{bmatrix}\right\|_G^2 \\
&+ \sum_{n=1}^{N-1}\Big\langle f_{n+\theta}, (A-C)^{-\frac{1}{2}}\Big(\theta A(A-C)^{-\frac{1}{2}}u_{n+1} + \big((1-\theta)A-(\theta+1)C\big)(A-C)^{-\frac{1}{2}}u_n + \theta C(A-C)^{-\frac{1}{2}}u_{n-1}\Big)\Big\rangle,
\end{aligned} \tag{2.7}$$

and the energy estimate

$$\begin{aligned}
&\|u_N\|^2 + \tfrac{1}{2\theta+1}\sum_{n=1}^{N-1}\|u_{n+1}-2u_n+u_{n-1}\|_F^2 \\
&\quad + \tfrac{2}{2\theta+1}\Delta t\sum_{n=1}^{N-1}\|\theta A(A-C)^{-\frac{1}{2}}u_{n+1} + \big((1-\theta)A-(\theta+1)C\big)(A-C)^{-\frac{1}{2}}u_n + \theta C(A-C)^{-\frac{1}{2}}u_{n-1}\|^2 \\
&\le \big(\tfrac{2\theta-1}{2\theta+1}\big)^N\|u_0\|^2 + 2\left\|\begin{bmatrix} u_1 \\ u_0\end{bmatrix}\right\|_G^2 + \Delta t\sum_{n=1}^{N-1}\|(A-C)^{-\frac{1}{2}}f_{n+\theta}\|^2.
\end{aligned} \tag{2.8}$$

Proof. The estimate (2.7) is a straightforward consequence of (2.1) and Lemma 2.2. Using the Cauchy-Schwarz and Young inequalities, we have that the forcing term in (2.7) can be bounded as follows

$$\begin{aligned}
&\Big\langle f_{n+\theta}, (A-C)^{-\frac{1}{2}}\Big(\theta A(A-C)^{-\frac{1}{2}}u_{n+1} + \big((1-\theta)A-(\theta+1)C\big)(A-C)^{-\frac{1}{2}}u_n + \theta C(A-C)^{-\frac{1}{2}}u_{n-1}\Big)\Big\rangle \\
&\le \frac{1}{2}\|(A-C)^{-\frac{1}{2}}f_{n+\theta}\|^2 + \frac{1}{2}\|\theta A(A-C)^{-\frac{1}{2}}u_{n+1} + \big((1-\theta)A-(\theta+1)C\big)(A-C)^{-\frac{1}{2}}u_n + \theta C(A-C)^{-\frac{1}{2}}u_{n-1}\|^2,
\end{aligned}$$

which gives

$$\begin{aligned}
&\left\|\begin{bmatrix} u_N \\ u_{N-1}\end{bmatrix}\right\|_G^2 + \frac{1}{4}\sum_{n=1}^{N-1}\|u_{n+1}-2u_n+u_{n-1}\|_F^2 \\
&\quad + \frac{\Delta t}{2}\sum_{n=1}^{N-1}\|\theta A(A-C)^{-\frac{1}{2}}u_{n+1} + \big((1-\theta)A-(\theta+1)C\big)(A-C)^{-\frac{1}{2}}u_n + \theta C(A-C)^{-\frac{1}{2}}u_{n-1}\|^2 \\
&\le \left\|\begin{bmatrix} u_1 \\ u_0\end{bmatrix}\right\|_G^2 + \frac{\Delta t}{2}\sum_{n=1}^{N-1}\|(A-C)^{-\frac{1}{2}}f_{n+\theta}\|^2.
\end{aligned}$$

Using (2.5) we obtain

$$\begin{aligned}
&\|u_N\|^2 + \frac{1}{2\theta+1}\sum_{n=1}^{N-1}\|u_{n+1}-2u_n+u_{n-1}\|_F^2 \\
&\quad + \frac{2}{2\theta+1}\Delta t\sum_{n=1}^{N-1}\|\theta A(A-C)^{-\frac{1}{2}}u_{n+1} + \big((1-\theta)A-(\theta+1)C\big)(A-C)^{-\frac{1}{2}}u_n + \theta C(A-C)^{-\frac{1}{2}}u_{n-1}\|^2 \\
&\le \frac{2\theta-1}{2\theta+1}\|u_{N-1}\|^2 + \frac{4}{2\theta+1}\left\|\begin{bmatrix} u_1 \\ u_0\end{bmatrix}\right\|_G^2 + \frac{2}{2\theta+1}\Delta t\sum_{n=1}^{N-1}\|(A-C)^{-\frac{1}{2}}f_{n+\theta}\|^2,
\end{aligned}$$

which by induction completes the proof. ∎

3. Consistency and convergence. With $\mathcal{E}^t_{n+\theta} = (\theta+1)u(t_n) - \theta u(t_{n-1})$, corresponding to $\mathcal{E}_{n+\theta}$ the explicit approximation of $u(t_{n+\theta})$, the local truncation error for (1.4) is

$$\begin{aligned}
\tau_{n+1}(\Delta t) = {} & \frac{(\theta+\frac{1}{2})u(t_{n+1}) - 2\theta u(t_n) + (\theta-\frac{1}{2})u(t_{n-1})}{\Delta t} - u'(t_{n+\theta}) \\
& + (A-C)^{\frac{1}{2}} A\big(\theta(A-C)^{-\frac{1}{2}}u(t_{n+1}) + (1-\theta)(A-C)^{-\frac{1}{2}}u(t_n) - (A-C)^{-\frac{1}{2}}u(t_{n+\theta})\big) \\
& - (A-C)^{\frac{1}{2}} C\big((\theta+1)(A-C)^{-\frac{1}{2}}u(t_n) - \theta(A-C)^{-\frac{1}{2}}u(t_{n-1}) - (A-C)^{-\frac{1}{2}}u(t_{n+\theta})\big) \\
& + \Big(B(\mathcal{E}^t_{n+\theta}) - B(u(t_{n+\theta}))\Big)(A-C)^{-\frac{1}{2}}\Big(A(A-C)^{-\frac{1}{2}}(\theta(A-C)^{-\frac{1}{2}}u(t_{n+1}) + (1-\theta)(A-C)^{-\frac{1}{2}}u(t_n)\big) \\
& \qquad - C\big((\theta+1)(A-C)^{-\frac{1}{2}}u(t_n) - \theta(A-C)^{-\frac{1}{2}}u(t_{n-1})\big)\Big) \\
& + B(u(t_{n+\theta}))\Big[(A-C)^{-\frac{1}{2}}\Big(A(\theta(A-C)^{-\frac{1}{2}}u(t_{n+1}) + (1-\theta)(A-C)^{-\frac{1}{2}}u(t_n)\big) \\
& \qquad - C\big((\theta+1)(A-C)^{-\frac{1}{2}}u(t_n) - \theta(A-C)^{-\frac{1}{2}}u(t_{n-1})\big)\Big) - (A-C)^{-\frac{1}{2}}u(t_{n+\theta})\Big].
\end{aligned}$$

THEOREM 3.1. *Assume that (1.2) holds and $f \in C^1([0,T]), u \in C^2([0,T])$. Then the local truncation error is $\mathcal{O}(\Delta t^2)$, the methods (1.4) are convergent, and if $e_0 = e_1 = 0$ the global error satisfies*

$$\|e_N\|^2 \le \frac{4}{2\theta+1}\exp\big(\tfrac{2\theta-1}{2\theta+1} + \tfrac{8}{2\theta+1}T\kappa\|(A-C)^{-\frac{1}{2}}\|^2\big)\|(A-C)^{-\frac{1}{2}}\|^2 U^2 \Delta t^4,$$

where

$$\begin{aligned}
U = {} & \max_{[t_{n-1},t_{n+1}]}\|u''(t)\|_2 \frac{\theta\big(\theta^3+(1+\theta)^3\big)}{3} \\
& + \max_{[t_{n-1},t_{n+1}]}\|(A-C)^{-\frac{1}{2}}A(A-C)^{-\frac{1}{2}}u'(t)\|_2 \frac{\theta(1-\theta)}{2} \\
& + \max_{[t_{n-1},t_{n+1}]}\|(A-C)^{-\frac{1}{2}}C(A-C)^{-\frac{1}{2}}u'(t)\|_2 \frac{2\theta^3+\theta^2+1}{2} \\
& + \max_{[t_{n-1},t_{n+1}]}\Big\|\tfrac{d}{dt}B(u(\cdot))\Big\|_2 \max_{[t_{n-1},t_{n+1}]}\|u(\cdot)\|_2 \frac{\max\{\theta(1-\theta), 2\theta^3+2\theta^2+1\}}{2} \\
& + \max_{[t_{n-1},t_{n+1}]}\Big\|\tfrac{d}{dt}B(u(\cdot))\Big\|_2 \max_{[t_{n-1},t_{n+1}]}\|(A-C)^{-\frac{1}{2}}A(A-C)^{-\frac{1}{2}}u'(\cdot)\|_2 \frac{\theta(1-\theta)}{2}\Delta t^2 \\
& + \max_{[t_{n-1},t_{n+1}]}\Big\|\tfrac{d}{dt}B(u(\cdot))\Big\|_2 \max_{[t_{n-1},t_{n+1}]}\|(A-C)^{-\frac{1}{2}}C(A-C)^{-\frac{1}{2}}u'(\cdot)\|_2 \frac{2\theta^3+\theta^2+1}{2}\Delta t^2 \\
& + \max_{[t_n,t_{n+1}]}\|B(u(\cdot))\|_2 \max_{[t_n,t_{n+1}]}\|(A-C)^{-\frac{1}{2}}A(A-C)^{-\frac{1}{2}}u'(\cdot)\|_2 \frac{\theta(1-\theta)}{2} \\
& + \max_{[t_n,t_{n+1}]}\|B(u(\cdot))\|_2 \max_{[t_{n-1},t_{n+1}]}\|(A-C)^{-\frac{1}{2}}C(A-C)^{-\frac{1}{2}}u'(\cdot)\|_2 \frac{2\theta^3+\theta^2+1}{2}.
\end{aligned}$$

Proof. Using the Taylor expansion around $t_{n+\theta} := t_n + \theta\Delta t$ we obtain

$$\|\tau_{n+1}(\Delta t)\|_2 \le U\Delta t^2 \tag{3.1}$$

which proves the consistency of methods (1.4). The error $e_n = u(t_n) - u_n$ satisfies

$$\begin{aligned}
&\frac{(\theta+\frac{1}{2})e_{n+1} - 2\theta e_n + (\theta-\frac{1}{2})e_{n-1}}{\Delta t} \\
&\quad + (A-C)^{\frac{1}{2}}A(A-C)^{-\frac{1}{2}}(\theta e_{n+1} + (1-\theta)e_n) - (A-C)^{\frac{1}{2}}C(A-C)^{-\frac{1}{2}}((\theta+1)e_n - \theta e_{n-1}) \\
&\quad + B(\mathcal{E}_{n+\theta})(A-C)^{\frac{1}{2}}\Big(A(A-C)^{-\frac{1}{2}}\big(\theta e_{n+1} + (1-\theta)e_n\big) - C(A-C)^{-\frac{1}{2}}\big((\theta+1)e_n - \theta e_{n-1}\big)\Big) \\
&= \tau_{n+1}(\Delta t) \\
&\quad - \big(B(\mathcal{E}^t_{n+\theta}) - B(\mathcal{E}_{n+\theta})\big)(A-C)^{-\frac{1}{2}}\Big(A(A-C)^{-\frac{1}{2}}\big(\theta u(t_{n+1}) + (1-\theta)u(t_n)\big) \\
&\qquad\qquad - C(A-C)^{-\frac{1}{2}}\big((\theta+1)u(t_n) - \theta u(t_{n-1})\big)\Big).
\end{aligned} \tag{3.2}$$

From (2.8) we have

$$\|u_n\|_2 \le \Lambda_1 := \Big(\|u_0\|^2 + 2\Big\|\begin{bmatrix}u_1\\u_0\end{bmatrix}\Big\|^2_G + T\|(A-C)^{-\frac{1}{2}}\|^2 \max_{t\in[0,T]}\|f(t)\|^2\Big)^{\frac{1}{2}} \quad \forall n = 1,\dots,N,$$

also from (1.1) we obtain

$$\|u(t)\|_2 \le \Lambda_2 := \Big(\|u(0)\|^2 + \int_0^T \|(A-C)^{-1}f(s)\|^2 ds\Big)^{\frac{1}{2}},$$

and let define

$$\Lambda_3 = \max_{n=1,\dots,N}\|(A-C)^{-\frac{1}{2}}\Big(A(A-C)^{-\frac{1}{2}}\big(\theta u(t_{n+1}) + (1-\theta)u(t_n)\big) - C(A-C)^{-\frac{1}{2}}\big((\theta+1)u(t_n) - \theta u(t_{n-1})\big)\Big)\|_2.$$

The last term in the RHS of (3.2) writes

$$\begin{aligned}
&\big(B(\mathcal{E}^t_{n+\theta}) - B(\mathcal{E}_{n+\theta})\big)(A-C)^{-\frac{1}{2}}\Big(A(A-C)^{-\frac{1}{2}}\big(\theta u(t_{n+1}) + (1-\theta)u(t_n)\big) \\
&\qquad\qquad - C(A-C)^{-\frac{1}{2}}\big((\theta+1)u(t_n) - \theta u(t_{n-1})\big)\Big) \\
&= \int_0^1 \frac{d}{ds}\big[B(s\mathcal{E}^t_{n+\theta} + (1-s)\mathcal{E}_{n+\theta})\big]ds(A-C)^{-\frac{1}{2}}\Big(A(A-C)^{-\frac{1}{2}}\big(\theta u(t_{n+1}) + (1-\theta)u(t_n)\big) \\
&\qquad\qquad - C(A-C)^{-\frac{1}{2}}\big((\theta+1)u(t_n) - \theta u(t_{n-1})\big)\Big) \\
&= \int_0^1 \nabla_u\big[B(u)(A-C)^{-\frac{1}{2}}\Big(A(A-C)^{-\frac{1}{2}}\big(\theta u(t_{n+1}) + (1-\theta)u(t_n)\big) \\
&\qquad\qquad - C(A-C)^{-\frac{1}{2}}\big((\theta+1)u(t_n) - \theta u(t_{n-1})\big)\Big)\big]|_{u=s\mathcal{E}^t_{n+\theta}+(1-s)\mathcal{E}_{n+\theta}} ds\,\big(\mathcal{E}^t_{n+\theta} - \mathcal{E}_{n+\theta}\big),
\end{aligned}$$

which since $B(\cdot)$ is C^1 implies

$$\begin{aligned}
&\|\big(B(\mathcal{E}^t_{n+\theta}) - B(\mathcal{E}_{n+\theta})\big)(A-C)^{-\frac{1}{2}}\Big(A(A-C)^{-\frac{1}{2}}\big(\theta u(t_{n+1}) + (1-\theta)u(t_n)\big) \\
&\qquad\qquad - C(A-C)^{-\frac{1}{2}}\big((\theta+1)u(t_n) - \theta u(t_{n-1})\big)\Big)\|_2 \\
&\le 2\kappa(\|e_n\|_2 + \|e_{n-1}\|), \forall n,
\end{aligned}$$

where

$$2\kappa = \max_{s\in[0,1],\|U_1\|_2\le 2\Lambda_1,\|U_2\|_2\le\Lambda_3,\|V_2\|_2\le 2\Lambda_2}\|\nabla_u\big[B(sV_2 + (1-s)U_1)U_2\big]\|_2.$$

Multiplying (3.2) by $(A-C)^{-\frac{1}{2}}\Big(A(A-C)^{-\frac{1}{2}}\big(\theta e_{n+1}+(1-\theta)e_n\big)-C(A-C)^{-\frac{1}{2}}\big((\theta+1)e_n-\theta e_{n-1}\big)\Big)$ and taking the sum from $n=1$ to $N-1$ we obtain

$$\begin{aligned}
&\left\|\begin{bmatrix} e_N \\ e_{N-1}\end{bmatrix}\right\|_G^2+\frac{1}{4}\sum_{n=1}^{N-1}\|e_{n+1}-2e_n+e_{n-1}\|_F^2\\
&\quad+\frac{\Delta t}{2}\sum_{n=1}^{N-1}\|\theta A(A-C)^{-\frac{1}{2}}e_{n+1}+\big((1-\theta)A-(\theta+1)C\big)(A-C)^{-\frac{1}{2}}e_n+\theta C(A-C)^{-\frac{1}{2}}e_{n-1}\|^2 \qquad (3.3)\\
&\le\left\|\begin{bmatrix} e_1 \\ e_0\end{bmatrix}\right\|_G^2+\Delta t\sum_{n=1}^{N-1}\|(A-C)^{-\frac{1}{2}}\tau_{n+1}(\Delta t)\|^2+\kappa\|(A-C)^{-\frac{1}{2}}\|^2\Delta t\sum_{n=1}^{N-1}(\|e_n\|^2+\|e_{n-1}\|^2).
\end{aligned}$$

Using again (2.5), after some calculation we get

$$\begin{aligned}
&\|e_N\|^2+\frac{1}{2\theta+1}\sum_{n=1}^{N-1}\|e_{n+1}-2e_n+e_{n-1}\|_F^2\\
&\quad+\frac{2}{2\theta+1}\Delta t\sum_{n=1}^{N-1}\|\theta A(A-C)^{-\frac{1}{2}}e_{n+1}+\big((1-\theta)A-(\theta+1)C\big)(A-C)^{-\frac{1}{2}}e_n+\theta C(A-C)^{-\frac{1}{2}}e_{n-1}\|^2\\
&\le\frac{4}{2\theta+1}\Big(\left\|\begin{bmatrix} e_1 \\ e_0\end{bmatrix}\right\|_G^2+\Delta t\sum_{n=1}^{N-1}\|(A-C)^{-\frac{1}{2}}\tau_{n+1}(\Delta t)\|^2\Big)+\Theta\Delta t\sum_{n=0}^{N-1}\|e_n\|^2,
\end{aligned}$$

where

$$\Theta=\max\Big\{\frac{2\theta-1}{(2\theta+1)\Delta t}+\frac{4}{2\theta+1}\kappa\|(A-C)^{-\frac{1}{2}}\|^2,\frac{8}{2\theta+1}\kappa\|(A-C)^{-\frac{1}{2}}\|^2\Big\}.$$

Therefore, from the discrete Grönwall lemma, we deduce the following error estimate

$$\begin{aligned}
&\|e_N\|^2+\frac{1}{2\theta+1}\sum_{n=1}^{N-1}\|e_{n+1}-2e_n+e_{n-1}\|_F^2\\
&\quad+\frac{2}{2\theta+1}\Delta t\sum_{n=1}^{N-1}\|\theta A(A-C)^{-\frac{1}{2}}e_{n+1}+\big((1-\theta)A-(\theta+1)C\big)(A-C)^{-\frac{1}{2}}e_n+\theta C(A-C)^{-\frac{1}{2}}e_{n-1}\|^2\\
&\le\frac{4}{2\theta+1}\exp\big(\tfrac{2\theta-1}{2\theta+1}+\tfrac{8}{2\theta+1}T\kappa\|(A-C)^{-\frac{1}{2}}\|^2\big)\Big(\left\|\begin{bmatrix} e_1 \\ e_0\end{bmatrix}\right\|_G^2+\Delta t\sum_{n=1}^{N-1}\|(A-C)^{-\frac{1}{2}}\tau_{n+1}(\Delta t)\|^2\Big).
\end{aligned}$$

Finally, the convergence result follows from the consistency bound (3.1). □

4. Numerical verification of Theorem 2.3. We give two numerical tests that confirm the theory. In all test cases, the initial conditions are

$$u_0=(1,1)^T \quad \text{and} \quad u_1=(1,1)^T,$$

and the matrices A and C are

$$A=(\nu+\varepsilon)\begin{pmatrix}1 & 0\\ 0 & 100\end{pmatrix},\quad C=\varepsilon\begin{pmatrix}1 & 0\\ 0 & 100\end{pmatrix},$$

where

$$\nu=0.001,\quad \varepsilon=0.01.$$

The time interval is $[0, 50]$, and we take $f = (0, 0)^T$.

Test 1. In the first case the matrix B is

$$B(\mathcal{E}_{n+\theta}) = \|\mathcal{E}_{n+\theta}\| \begin{pmatrix} 0 & 10 \\ -10 & 0 \end{pmatrix},$$

With the time steps

$$\Delta t = 0.25 \quad \text{and} \quad \Delta t = 0.125,$$

both methods CN-AB2 ($\theta = \frac{1}{2}$) and BDF2-AB2 ($\theta = 1$) are observed to be stable, see Figures 4.1 and 4.2, respectively.

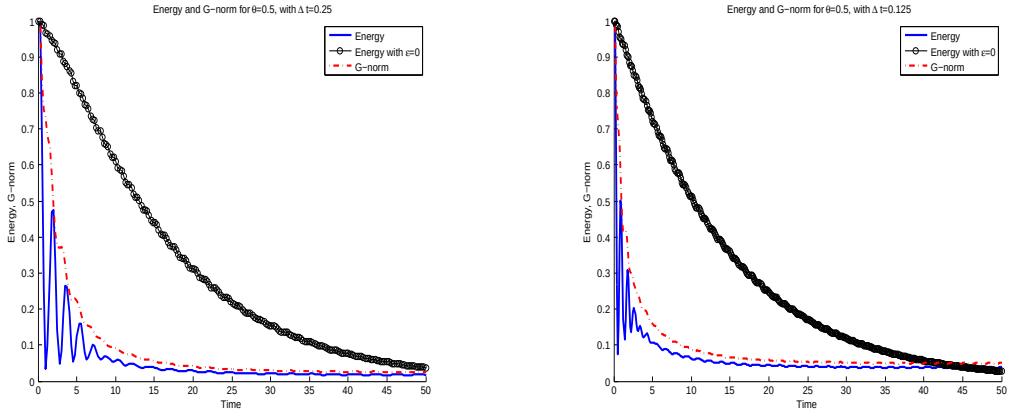

Fig. 4.1: Crank-Nicolson-AB2 method (1.5).

Test 2. In the second case the matrix B is

$$B(\mathcal{E}_{n+\theta}) = \|\mathcal{E}_{n+\theta}\| \begin{pmatrix} 0 & 100 \\ -100 & 0 \end{pmatrix},$$

where

$$\nu = 0.001, \quad \varepsilon = 0.01.$$

With the time steps

$$\Delta t = 0.25 \quad \text{and} \quad \Delta t = 0.125,$$

both methods CN-AB2 ($\theta = \frac{1}{2}$) and BDF2-AB2 ($\theta = 1$) are observed to be stable, see Figures 4.3 and 4.4, respectively.

REFERENCES

[1] G. Akrivis, M. Crouzeix, and C. Makridakis, *Implicit-explicit multistep methods for quasilinear parabolic equations*, Numer. Math., 82 (1999), pp. 521–541. 1

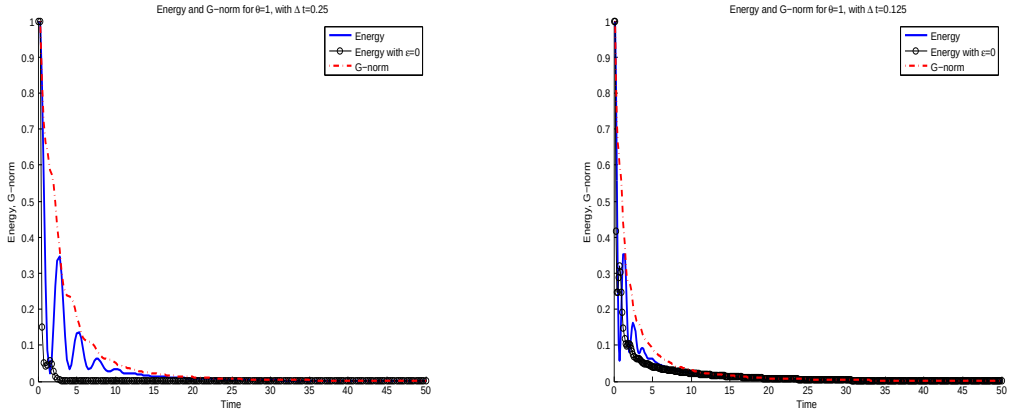

Fig. 4.2: BDF2-AB2 method (1.6).

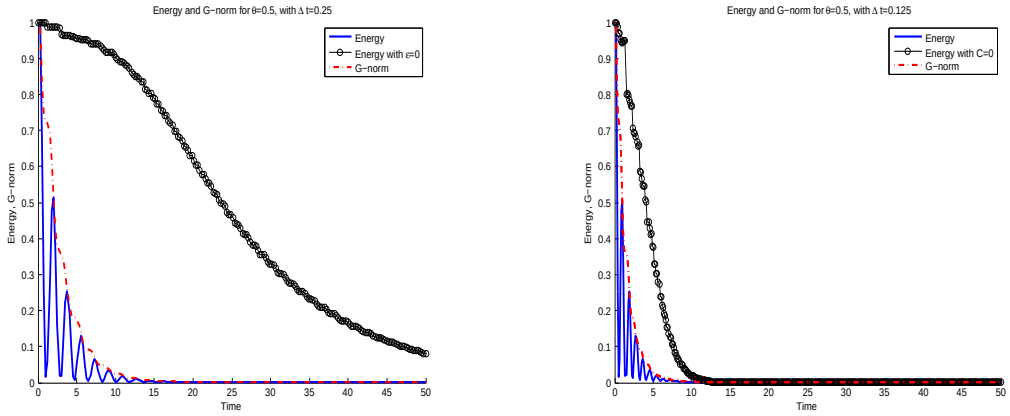

Fig. 4.3: Crank-Nicolson-AB2 method (1.5).

[2] M. Anitescu, F. Pahlevani, and W. J. Layton, *Implicit for local effects and explicit for nonlocal effects is unconditionally stable*, Electron. Trans. Numer. Anal., 18 (2004), pp. 174–187 (electronic). 1, 2

[3] U. M. Ascher, S. J. Ruuth, and B. T. R. Wetton, *Implicit-explicit methods for time-dependent partial differential equations*, SIAM J. Numer. Anal., 32 (1995), pp. 797–823. 1

[4] J. Frank, W. Hundsdorfer, and J. G. Verwer, *On the stability of implicit-explicit linear multistep methods*, Appl. Numer. Math., 25 (1997), pp. 193–205. Special issue on time integration (Amsterdam, 1996). 1

[5] J.-L. Guermond, *Stabilization of Galerkin approximations of transport equations by subgrid modeling*, M2AN Math. Model. Numer. Anal., 33 (1999), pp. 1293–1316. 2

[6] J.-L. Guermond, *Subgrid stabilization of Galerkin approximations of monotone operators*, C. R. Acad. Sci. Paris, Série I, 328 (1999), pp. 617–622. 2

[7] M. Gunzburger, E. Lee, Y. Saka, C. Trenchea, and X. Wang, *Analysis of nonlinear spectral eddy-viscosity models of turbulence*, J. Sci. Comput., 45 (2010), pp. 294–332. 3

[8] E. Hairer and G. Wanner, *Solving ordinary differential equations. II*, vol. 14 of Springer Series in Computational

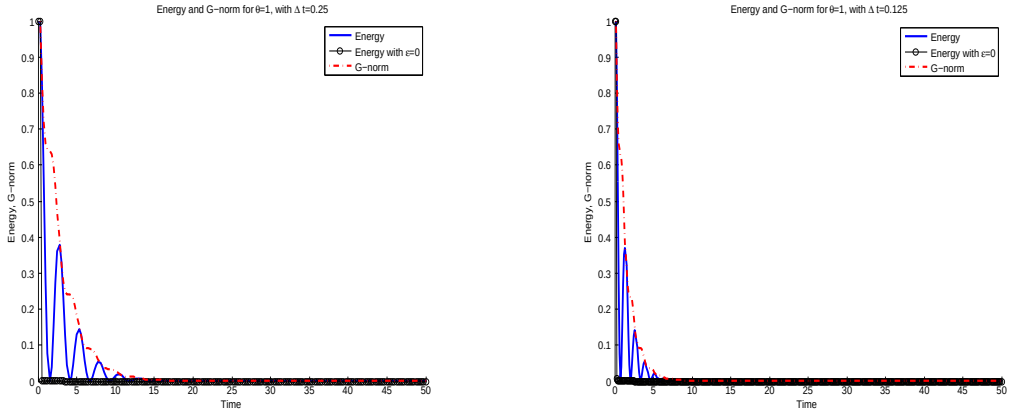

Fig. 4.4: BDF2-AB2 method (1.6).

Mathematics, Springer-Verlag, Berlin, 2010. Stiff and differential-algebraic problems, Second revised edition. 4

[9] T. J. Hughes, L. Mazzei, and K. E. Jansen, *Large eddy simulation and the variational multiscale method.*, Comput. Vis. Sci., 3 (2000), pp. 47–59. 2

[10] H. Johnston and J.-G. Liu, *Accurate, stable and efficient Navier-Stokes solvers based on explicit treatment of the pressure term*, J. Comput. Phys., 199 (2004), pp. 221–259. 1, 2

[11] S. Kaya, *Numerical analysis of a subgrid scale eddy viscosity method for higher Reynolds number flow problem*, tech. rep., University of Pittsburgh, 2002. 2

[12] O. A. Ladyženskaja, *Modifications of the Navier-Stokes equations for large gradients of the velocities*, Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI), 7 (1968), pp. 126–154. 3

[13] W. Layton, N. Mays, M. Neda, and C. Trenchea, *Numerical analysis of modular regularization methods for the BDF2 time discretization of the Navier-Stokes equations*, ESAIM: Mathematical Modelling and Numerical Analysis, (submitted 2011). 3

[14] J.-L. Lions, *Quelques méthodes de résolution des problèmes aux limites non linéaires*, Dunod, 1969. 3

[15] B. Mohammadi and O. Pironneau, *Analysis of the k-epsilon turbulence model*, RAM: Research in Applied Mathematics, Masson, Paris, 1994. 2

[16] S. J. Ruuth, *Implicit-explicit methods for reaction-diffusion problems in pattern formation*, J. Math. Biol., 34 (1995), pp. 148–176. 1

[17] J. Varah, *Stability restrictions on second order, three level finite difference schemes for parabolic equations*, SIAM J. Numer. Anal., 17 (1980), pp. 300–309. 1