Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

W.H. Hundsdorfer, J.G. Verwer

Stability and convergence of the Peaceman-Rachford ADI method for initial-boundary value problems

Department of Numerical Mathematics Report NM-R8724 November

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

Stability and Convergence of the Peaceman-Rachford ADI Method for Initial-Boundary Value Problems

W.H. Hundsdorfer and J.G. Verwer
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

In this paper an analysis will be presented for the ADI (alternating direction implicit) method of Peaceman and Rachford applied to initial-boundary value problems for partial differential equations in two space dimensions. We shall use the method of lines approach. Motivated by developments in the field of stiff nonlinear ordinary differential equations, our analysis will focus on problems where the semi-discrete system, obtained after discretization in space, satisfies a one sided Lipschitz condition with a constant independent of the grid spacing. For such problems unconditional stability and convergence results will be derived.

AMS(MOS) 1980 subject classifications: 65M10, 65M15, 65M20

1982 CR Categories: 5.17

Keywords: Numerical analysis, time dependent PDEs, alternating direction implicit methods, Peaceman-Rachford method, method of lines, stability, error bounds.

Note: This report will be submitted for publication elsewhere.

1. Introduction

For many years splitting methods have proved valuable in the numerical solution of time dependent, multispace dimensional partial differential equations (PDEs). The general idea of splitting is to attack a multi-space dimensional problem in such a way that only one-space dimensional computations are required. This idea has led to the development of a great variety of so-called alternating direction (ADI) methods, locally one-dimensional (LOD) or fractional step methods, and hopscotch type methods [6]. ADI methods were first introduced by Peaceman, Douglas and Rachford for the solution of parabolic (and elliptic) equations in two [13] and three [3] space variables. The present paper is devoted to a study of the stability and convergence properties of the original Peaceman-Rachford ADI method when applied to initial-boundary value problems.

Report NM-R8724
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The idea of splitting has to do with the time integration, rather than with the space discretization. This suggests to adopt the method of lines approach [9], which has the advantage that it enables us to formulate the Peaceman-Rachford (PR) method in a very compact way for a wide class of (two-space dimensional) initial-boundary value problems, including nonlinear ones (method (2.2)). Another advantage is that it enables us to directly use ideas and results from the field of stiff ordinary differential equations (ODEs), which in the last years has witnessed interesting developments on nonlinear stability and convergence [1].

Let us give a brief outline of the paper. Section 2 is devoted to some preliminary results which are relevant for the remainder. There we link the PR method with an LOD type splitting method, based on the implicit midpoint rule. Loosely speaking, this ADI/LOD link (made before in [8]) reveals that with respect to step-by-step stability the ADI method will behave very much the same as the fully implicit midpoint rule. With the notion of step-by-step stability we mean the stability of the ODE integration formula for evolving time (A-stability is such a property). This observation is rather interesting, since the implicit midpoint rule is known to possess unconditional stability properties for nonlinear, stiff ODE problems satisfying a one-sided Lipschitz condition. In the remainder of the paper, we therefore assume such a condition to hold for the semi-discrete system under consideration.

The stability analysis for the PR method is carried out in detail in Section 3. Our analysis concentrates on unconditional stability, by which we mean that no relation is assumed between the stepsize in time and the space grid refinement. We present a result valid for nonlinear, noncommuting splitting operators, which refutes to some extent the often expressed view that for step-by-step stability commutativity is determining. In this section we also point out, through a numerical illustration, that when implementing the PR method on the computer for nonlinear problems, care must be exercised in solving the arising systems of nonlinear algebraic equations. If this is not done with sufficient accuracy, then the stability may deteriorate severely. This observation is of practical significance, because in applications one often linearizes the problem which always can be interpreted as carrying out, in a certain way, one step of the iterative Newton process. The point of view we take here is that in many cases instability is an artifact of the linearization, and not of the ADI scheme itself.

Sections 4 and 5 are devoted to full convergence properties of the PR scheme. Here we distinguish between nonlinear (Section 4) and linear (Section 5) problems. The prefix full means that we compare the numerical solution directly with the exact PDE solution. More specifically, the main objective of our convergence analysis is the order in time p featuring in a global error bound of the type

$$\|U_n - u_h(t_n)\| \leq C_1 \tau^p + C_2 \max_t \|\alpha_h(t)\|,$$

where U_n is the numerical solution at time $t = t_n$, $u_h(t_n)$ the PDE solution at $t = t_n$ restricted to the

imposed space grid, α_h the spatial truncation error, and C_1 and C_2 are constants completely independent of the stepsize τ and the space grid refinement. This independency means we examine unconditional convergence. In the nonlinear case we prove such convergence with order $p=1$, which is one less than the order on a fixed space grid. The discrepancy is caused by influence of the boundary conditions, not by lack of smoothness. Here the notions of local and global order reduction come into play, which is elucidated in an extensive discussion devoted to the linear case. There we present also convergence results with $p=2$ and briefly outline how the so-called Fairweather-Mitchell correction fits in the convergence theory.

2. Preliminaries

As mentioned already in the introduction, we follow in this paper the method of lines approach. This enables us to formulate the Peaceman-Rachford method in a compact way and, in addition, allows for the general treatment we aim at. In Section 2 we have collected some preliminary material. Section 2.1 deals with the time integration formula, while Section 2.2 contains information on the semi-discrete problems.

2.1 The Peaceman-Rachford integration formula

Consider the real Cauchy problem for the nonlinear ODE system

$$\dot{U} = F(t,U),\ 0\leqslant t\leqslant T,\ U(0) = U_0, \tag{2.1a}$$

where $U_0\in\mathbb{R}^M$ and $F:[0,T]\times\mathbb{R}^M\to\mathbb{R}^M$ are given. This system is supposed to originate from spatial discretization of an initial-boundary value PDE problem. For the moment there is no need to be more specific about (2.1a). We suppose that F can be decomposed into two more simple functions F_1 and F_2, with

$$F_1(t,v) + F_2(t,v) = F(t,v) \text{ for all } (t,v)\in[0,T]\times\mathbb{R}^M. \tag{2.1b}$$

The meaning of this linear splitting will become clear later.

The Peaceman-Rachford integration formula we examine in this paper is then given by

$$\begin{aligned} U_{n+1/2} &= U_n + \tfrac{1}{2}\tau F_1(t_{n+1/2},U_{n+1/2}) + \tfrac{1}{2}\tau F_2(t_n,U_n),\\ U_{n+1} &= U_{n+1/2} + \tfrac{1}{2}\tau F_1(t_{n+1/2},U_{n+1/2}) + \tfrac{1}{2}\tau F_2(t_{n+1},U_{n+1}). \end{aligned} \tag{2.2}$$

Here $t_{n+1/2}=t_n+\tau/2$, $t_{n+1}=t_n+\tau$ for $n\geqslant 0$, and $U_{n+1/2}$, U_{n+1} are the approximations to the exact solutions $U(t)$ of (2.1) at time $t=t_{n+1/2}$, t_{n+1}, respectively. In this one-step integration formula $(U_n\to U_{n+1})$, $U_{n+1/2}$ is always considered to be an intermediate, auxiliary vector like in Runge-Kutta

methods. Due to the one-step nature, it is easy to use variable stepsizes τ. However, in the remainder we restrict ourselves to constant values for τ.

From (2.2) two well known integration methods can be recovered. If we put $F_1=0$, $F_2=F$, the trapezoidal rule is obtained, while for $F_1=F$, $F_2=0$, (2.2) reduces to the implicit midpoint rule

$$U_{n+1} = U_n + \tau F(t_{n+1/2}, \tfrac{1}{2}U_n + \tfrac{1}{2}U_{n+1}). \tag{2.3}$$

Inspection of (2.2) and (2.3) immediately reveals the characteristic features of (2.2). This method is alternately implicit in F_1 and F_2, whereas (2.3) is (fully) implicit in F. In our application of (2.2), following Peaceman and Rachford [13], F stands for a discretized PDE operator in two space dimensions, and F_1 and F_2 are both assumed to be "one-dimensional". This implies that, per step, the costs involved in solving the implicit relations in (2.2) will be substantially lower than in a fully implicit method like (2.3).

Like the implicit midpoint and trapezoidal rule, the PR formula (2.2) has the (usual) order of consistency two for any given ODE system (2.1). This follows readily from a straightforward Taylor expansion. We mean here consistency with respect to the ODE solution U, not with respect to the underlying PDE solution. In the Sections 4, 5 we will be more specific about consistency and convergence. There we will compare the approximations U_n directly with the PDE solution.

One of the points we wish to emphasize in this paper is that the stability of the PR method is in a sense governed by the stability of an implicit midpoint LOD method. To see this, we rewrite (2.2) in the Euler fashion

$$\begin{aligned} Y_{n+1/2} &= U_n + \tfrac{1}{2}\tau F_2(t_n, U_n), \\ U_{n+1/2} &= Y_{n+1/2} + \tfrac{1}{2}\tau F_1(t_{n+1/2}, U_{n+1/2}), \\ Y_{n+1} &= U_{n+1/2} + \tfrac{1}{2}\tau F_1(t_{n+1/2}, U_{n+1/2}), \\ U_{n+1} &= Y_{n+1} + \tfrac{1}{2}\tau F_2(t_{n+1}, U_{n+1}) \end{aligned} \tag{2.4}$$

for $n \geqslant 0$. This can be rearranged as a first step

$$Y_{1/2} = U_0 + \tfrac{1}{2}\tau F_2(t_0, U_0) \tag{2.5a}$$

followed by

$$\begin{cases} U_{n+1/2} = Y_{n+1/2} + \tfrac{1}{2}\tau F_1(t_{n+1/2}, U_{n+1/2}), \\ Y_{n+1} = U_{n+1/2} + \tfrac{1}{2}\tau F_1(t_{n+1/2}, U_{n+1/2}), \end{cases} \tag{2.5b}$$

$$\begin{cases} U_{n+1} = Y_{n+1} + \frac{1}{2}\tau F_2(t_{n+1}, U_{n+1}), \\ Y_{n+3/2} = U_{n+1} + \frac{1}{2}\tau F_2(t_{n+1}, U_{n+1}) \end{cases} \tag{2.5c}$$

for $n \geqslant 0$. Note that now (2.5b) constitutes the implicit midpoint rule

$$Y_{n+1} = Y_{n+1/2} + \tau F_1(t_{n+1/2}, \tfrac{1}{2}Y_{n+1/2} + \tfrac{1}{2}Y_{n+1}). \tag{2.6}$$

Likewise, (2.5c) gives

$$Y_{n+3/2} = Y_{n+1} + \tau F_2(t_{n+1}, \tfrac{1}{2}Y_{n+1} + \tfrac{1}{2}Y_{n+3/2}). \tag{2.7}$$

Consequently, apart form start and completion, the PR scheme is equivalent to an alternate application of the implicit midpoint schemes (2.6) and (2.7). The combination of these two is just a locally one-dimensional (LOD) method. It thus follows that this implicit midpoint LOD method governs the step-by-step stability of the PR method.

We note that the link between ADI and LOD has been made before by Gourlay and Mitchell [8]. We shall use it in our stability analysis presented in Section 3, although in a slightly different manner than above.

2.2. The semi-discrete problem

The stability and convergence analysis presented in the remainder of this paper is centered around the semi-discrete problem (2.1). This means that in a large part of our analysis there is no need to be specific about the underlying 2-dimensional PDE and its spatial discretization. The formulation (2.1) indicates that we have finite difference discretizations in mind, but finite element methods (continuous time Galerkin) could also be considered.

Let Ω_h be a space grid covering the spatial domain $\Omega \subset \mathbb{R}^2$ of the PDE. The vectors $U, F \in \mathbb{R}^M$ in (2.1) can be viewed as gridfunctions, each component (or set of components for nonscalar PDEs) corresponds to a value on a gridpoint of Ω_h. The positive parameter h refers to the grid distance, which may vary over the grid. In what follows, the limit $h \to 0$ means that the space grid is refined arbitrarily far in a suitable manner. Hereby it is emphasized that the dimension M of U and F depends on h. This dependence is suppressed in our notation. We assume that the boundary conditions on Γ, the boundary of Ω, are incorporated in the function F.

Let $u(x,t)$ $(x \in \Omega \cup \Gamma,\ t \in [0,T])$ be the exact PDE solution. The (pointwise) restriction of u to Ω_h will be denoted by u_h. In our convergence analysis we will compare the fully discrete numerical solutions U_n to $u_h(t_n)$. For this analysis we need the space truncation error $\alpha_h(t)$, defined by

$$\alpha_h(t) = \dot{u}_h(t) - F(t,u_h(t)) \quad \text{for} \quad 0 \leqslant t \leqslant T. \tag{2.8}$$

It will be assumed that (2.1) is consistent with the underlying initial-boundary value problem, in the sense that

$$\max_{0 \leqslant t \leqslant T} \|\alpha_h(t)\| \to 0 \ \text{ as } \ h \to 0. \tag{2.9}$$

Throughout this paper, $\|w\|$ denotes a chosen norm for M-dimensional vectors w, generated by an inner product $<v,w>$ on $\mathbb{R}^M$. Likewise, we denote the induced matrix norm

$$\|A\| = \sup_{w \neq 0} \|Aw\| / \|w\|$$

for $A \in L(\mathbb{R}^M)$, the space of real $M \times M$ matrices.

Our stability and convergence analysis will be restricted to semi-discrete problems satisfying the one-sided Lipschitz condition

$$<F_i(t,\tilde{w}) - F_i(t,w), \tilde{w} - w> \ \leqslant \nu \|\tilde{w} - w\|^2 \quad (i = 1,2) \tag{2.10}$$

for arbitrary $w, \tilde{w}$ in $\mathbb{R}^M$ and $0 \leqslant t \leqslant T$. Essential hereby is that the one-sided Lipschitz constant ν is independent of h, that is, of the grid spacing. We shall assume, for convenience, that $\nu \leqslant 0$, but this is not essential for what follows.

Condition (2.10) implies the exponential stability result

$$\|\tilde{U}(t) - U(t)\| \leqslant e^{2\nu t} \|\tilde{U}(0) - U(0)\|, \ \ 0 \leqslant t \leqslant T, \tag{2.11}$$

valid uniformly in h, for any pair of semidiscrete solutions $\tilde{U}, U$ of (2.1) [1]. It also implies that the spatial error $u_h(t) - U(t)$ satisfies the bound (cf. [19])

$$\|u_h(t) - U(t)\| \leqslant \frac{e^{2\nu t} - 1}{2\nu} \max_{0 \leqslant s \leqslant t} \|\alpha_h(s)\|, \ \ 0 \leqslant t \leqslant T, \tag{2.12}$$

provided $U(0) = u_h(0)$. Here one should read t for $(e^{2\nu t} - 1)/2\nu$ in case $\nu = 0$.

The well-posedness inequality (2.11) indicates that condition (2.10) is a fairly natural one. When combined with a suitable space discretization, interesting classes of PDE problems can be shown to satisfy (2.10). As an example we mention the semi-linear heat equation

$$u_t = (a_1(x,y,t)u_x)_x + (a_2(x,y,t)u_y)_y + s(x,y,t,u), \tag{2.13}$$

with a_i strictly positive and $\partial s / \partial u \leqslant \nu$. It is this type of equation for which ADI methods were originally developed.

As a word of warning, we should also note that many semi-discrete problems exist for which it may be very cumbersome to verify (2.10) for a certain norm, of course assuming such a norm exists. For example, solution dependent coefficients a_i in (2.13) cause difficulties here. Finally, the restriction that we let $w,\tilde{w}$ lie in the whole of $\mathbb{R}^M$ is not essential and is made only for convenience of presentation. In actual, nonlinear applications, it suffices to verify (2.10) with $\tilde{w}=u_h(t)$ and w lying in a tube around $u_h(t)$, $0\leqslant t\leqslant T$.

To conclude this preliminary section, we recall that in the method of lines literature, semi-discrete PDEs are often treated as stiff ODEs [1,16]. In fact, many results in the nonlinear stability theory for stiff ODEs relate to problems satisfying a one-sided Lipschitz condition like (2.10). Important parts in our stability and convergence analysis presented in the remainder of this paper, do originate from the field of stiff ODEs.

3. Stability

The entire Section 3 is devoted to stability. We will present a stability result for the PR method (2.2) which is valid for any ODE system (2.1) satisfying the one-sided Lipschitz condition (2.10) (so, F_1 and F_2 may be nonlinear and noncommuting). For simplicity of presentation it will be assumed that (2.10) holds with $\nu\leqslant 0$. The results can be easily extended to the case $\nu>0$.

We will frequently use the following norm inequality for rational functions of matrices, basically due to von Neumann (cf. [1], Theorem 2.3.1.).

Lemma 3.1. *Let $r(z)$ be a rational function, and A an $M\times M$ matrix satisfying $<Aw,w>\leqslant\nu\|w\|^2$ for all $w\in\mathbb{R}^M$. Then we have, for all $\tau>0$,*

$$\|r(\tau A)\| \leqslant \sup\{|r(z)|: z\in\mathbb{C}, \mathrm{Re}(z)\leqslant\tau\nu\}. \quad \square$$

3.1. A general stability inequality

Beside (2.2) we consider the perturbed PR scheme

$$\begin{aligned}\tilde{U}_{n+1/2} &= \tilde{U}_n+\tfrac{1}{2}\tau F_1(t_{n+1/2},\ \tilde{U}_{n+1/2})+\tfrac{1}{2}\tau F_2(t_n,\ \tilde{U}_n)+\tau\delta_{n+1/2},\\ \tilde{U}_{n+1} &= \tilde{U}_{n+1/2}+\tfrac{1}{2}\tau F_1(t_{n+1/2},\ \tilde{U}_{n+1/2})+\tfrac{1}{2}\tau F_2(t_{n+1},\ \tilde{U}_{n+1})+\tau\delta_{n+1}.\end{aligned} \tag{3.1}$$

The perturbations δ_j may stand for round-off errors, errors due to nonexactly solving the implicit relations, or for discretization errors. Let

$$\epsilon_j = \tilde{U}_j - U_j \ \text{ for } \ j = n, n+1/2 \ \text{ and } \ n\geqslant 0. \tag{3.2}$$

By subtracting (2.2) from (3.1) and using the mean value theorem, we obtain the following recursion for the errors,

$$\begin{aligned}\epsilon_{n+1/2} &= \epsilon_n + \tfrac{1}{2}\tau A_{1,n+1/2}\,\epsilon_{n+1/2} + \tfrac{1}{2}\tau A_{2,n}\,\epsilon_n + \tau\delta_{n+1/2},\\ \epsilon_{n+1} &= \epsilon_{n+1/2} + \tfrac{1}{2}\tau A_{1,n+1/2}\,\epsilon_{n+1/2} + \tfrac{1}{2}\tau A_{2,n+1}\,\epsilon_{n+1} + \tau\delta_{n+1}\end{aligned} \tag{3.3}$$

where

$$A_{ij} = \int_0^1 F_i'(t_j, \theta\tilde{U}_j + (1-\theta)U_j)\,d\theta, \quad F_i'(t,w) = \partial F_i(t,w)/\partial w \tag{3.4}$$

for $i = 1,2$ and $j = n, n+1/2$. We now eliminate $\epsilon_{n+1/2}$ from (3.3) to obtain

$$\epsilon_{n+1} = R_n\epsilon_n + \tau\rho_{n+1} \tag{3.5}$$

where

$$R_n = (I - \tfrac{1}{2}\tau A_{2,n+1})^{-1} r(\tau A_{1,n+1/2})(I + \tfrac{1}{2}\tau A_{2,n}), \tag{3.6}$$

$$\rho_{n+1} = (I - \tfrac{1}{2}\tau A_{2,n+1})^{-1}[r(\tau A_{1,n+1/2})\delta_{n+1/2} + \delta_{n+1}] \tag{3.7}$$

and $r(z) = (1-z/2)^{-1}(1+z/2)$ is the familiar stability function of the implicit midpoint rule.

We note that due to the one-sided Lipschitz condition (2.10) with $\nu \leqslant 0$, all operations above are justified for arbitrary $\tau > 0$. The implicit relations are uniquely solvable, and the following matrix norm inequalities follow from Lemma 3.1,

$$\|r(\tau A_{ij})\| \leqslant 1 \quad \text{for all} \quad \tau > 0, \tag{3.8}$$

$$\|(I - \tfrac{1}{2}\tau A_{ij})^{-1}\| \leqslant 1 \quad \text{for all} \quad \tau > 0. \tag{3.9}$$

This lemma also shows that when $\nu > 0$, the upper bound 1 in (3.8) is to be replaced by $(1-\tau\nu/2)^{-1}(1+\tau\nu/2)$ and the corresponding range for τ by $\tau\nu < 2$. Similarly, for $\nu < 0$ the upper bound 1 in (3.9) can be sharpened to $(1-\tau\nu/2)^{-1}$ for all $\tau > 0$, while for $\nu > 0$ this same bound holds for $\tau\nu < 2$. Essential for application to PDEs of these inequalities is their validity uniformly in the mesh width of the space grid.

Direct application of (3.8) to obtain a bound for $\|R_n\|$ is not possible in general, due to the fact that $A_{2,n}$ may vary with n, and that $A_{2,n}$ and $A_{1,n+1/2}$ need not commute. Should A_2 be independent of n and commute with the matrices $A_{1,n+1/2}$, then $R_n = r(\tau A_2)r(\tau A_{1,n+1/2})$, so that $\|R_n\| \leqslant 1$ for arbitrary $\tau > 0$. In this case we immediately derive from (3.5)-(3.9) the global error bound

$$\|\epsilon_n\| \leqslant \|\epsilon_0\| + 2D \quad \text{for all} \quad \tau>0,\ n\geqslant 1, \tag{3.10}$$

where D is an upperbound for all $\|\delta_j\|$, $j=n,n+1/2$ and $n\geqslant 0$. This error bound expresses stability of the PR scheme with respect to initial errors ϵ_0 and perturbations δ_j.

We now consider the general case (where the matrices do not commute and $A_{2,n}$ varies with n), and introduce the following transformation of the errors ϵ_n for $n\geqslant 0$,

$$\hat{\epsilon}_n = (I-\tfrac{1}{2}\tau A_{2,n})\epsilon_n. \tag{3.11}$$

These transformed errors satisfy

$$\hat{\epsilon}_{n+1} = \hat{R}_n\hat{\epsilon}_n + \tau\hat{\rho}_{n+1}, \tag{3.12}$$

with

$$\hat{R}_n = r(\tau A_{1,n+1/2})r(\tau A_{2,n}), \tag{3.13}$$

$$\hat{\rho}_{n+1} = r(\tau A_{1,n+1/2})\delta_{n+1/2} + \delta_{n+1}. \tag{3.14}$$

The effect of this transformation is that the new amplification operator $\hat{R}_n$ is factored into two operators both with norm $\leqslant 1$, like in LOD methods [18], which gives us the global bound

$$\|\hat{\epsilon}_n\|\leqslant\|\hat{\epsilon}_0\| + \max_{1\leqslant k\leqslant n}\|\hat{\rho}_k\| \quad \text{for all} \quad \tau>0,\ n\geqslant 1. \tag{3.15}$$

Since $\|\epsilon_n\|\leqslant\|\hat{\epsilon}_n\|$ and $\|\hat{\rho}_{n+1}\|\leqslant\|\delta_{n+1/2}\|+\|\delta_{n+1}\|$ (cf. (3.8), (3.9)), we obtain the following stability result.

THEOREM 3.2. *Consider (2.2) and (3.1) with perturbations $\|\delta_j\|\leqslant D$. Suppose the one-sided Lipschitz condition (2.10) holds with $\nu\leqslant 0$. Then the errors $\epsilon_n=\tilde{U}_n - U_n$ satisfy*

$$\|\epsilon_n\| \leqslant \|(I-\tfrac{1}{2}\tau A_{2,0})\epsilon_0\| + 2D \quad \text{for all} \quad \tau>0,\ n\geqslant 1, \tag{3.16}$$

where $A_{2,0}$ is given by (3.4). □

The transformation (3.11) leading to this result is inspired by the ADI-LOD link outlined in Section 2.2. For linear problems with constant coefficients a similar result was obtained by Douglas and Gunn [2].

The bound (3.16) expresses stability of the PR scheme w.r.t. the transformed initial error $\hat{\epsilon}_0$ and the original perturbations δ_j. We will comment on $\|\hat{\epsilon}_0\|$ in the next section. For the moment we note

that if $\|\hat{\epsilon}_0\| \leq C\|\epsilon_0\|$ with $C>0$ independent of h (for instance if $\epsilon_0=0$ or ϵ_0 is a smooth gridfunction, so that $\|A_{2,0}\epsilon_0\| \leq C'\|\epsilon_0\|$) then (3.16) shows unconditional stability without the common assumptions that F_1 and F_2 are linear and commuting. Note that when $A_{2,n}$ is independent of n and commutes with $A_{1,n+1/2}$, no additional smoothness of ϵ_0 is required (cf. (3.10)).

3.2. Stabilization of the first step

The transformed initial error $\hat{\epsilon}_0=(I-\frac{1}{2}\tau A_{2,0})\epsilon_0$ can be interpreted as the difference of two explicit Euler steps with negative stepsize $-\tau$,

$$\hat{\epsilon}_0 = [\tilde{U}_0-\tfrac{1}{2}\tau F_2(t_0,\tilde{U}_0)]-[U_0-\tfrac{1}{2}\tau F_2(t_0,U_0)]. \tag{3.17}$$

In general we may have $\|\hat{\epsilon}_0\| \gg \|\epsilon_0\|$, due to the explicitness in (3.17). Of course, if ϵ_0 is negligible, for instance if round-off is the only error sourse, then $\hat{\epsilon}_0$ will be small for reasonable values of h. However, if ϵ_0 is not very small, for example if $\tilde{U}_0$ is obtained from experimental data with significant errors, then $\hat{\epsilon}_0$ may become quite large and grow with spatial refinement. In such a situation we can stabilize the PR scheme by computing the first approximation U_1 by the backward Euler-LOD method, and apply (2.2) only for $n \geq 1$.

We thus consider the scheme with first step

$$\begin{aligned} U_{1/2} &= U_0 + \tau F_1(t_{1/2},U_{1/2}), \\ U_1 &= U_{1/2} + \tau F_2(t_1,U_1), \end{aligned} \tag{3.18a}$$

and for $n=1,2,3,...$

$$\begin{aligned} U_{n+1/2} &= U_n+\tfrac{1}{2}\tau F_1(t_{n+1/2},U_{n+1/2})+\tfrac{1}{2}\tau F_2(t_n,U_n), \\ U_{n+1} &= U_{n+1/2}+\tfrac{1}{2}\tau F_1(t_{n+1/2},U_{n+1/2})+\tfrac{1}{2}\tau F_2(t_{n+1},U_{n+1}). \end{aligned} \tag{3.18b}$$

On a fixed space grid the LOD-scheme has only order 1 in time, but since we only perform one LOD step, the order of the process (3.18) will still be 2 on fixed space grids. Assume as before that (2.10) holds with $\nu \leq 0$. Repeating the stability analysis of the previous section, we now obtain

$$\|\epsilon_n\| \leq \|(I-\tfrac{1}{2}\tau A_{2,1})\epsilon_1\| + 2D \quad \text{for all } \tau>0,\ n\geq 2 \tag{3.19}$$

with D an upper bound for the $\|\delta_j\|$ ($j=3/2,2,5/2,...$). The error ϵ_1 is now given by

$$\epsilon_{1/2} = (I-\tau A_{1,1/2})^{-1}[\epsilon_0+\tau\delta_{1/2}],\ \ \epsilon_1 = (I-\tau A_{2,1})^{-1}[\epsilon_{1/2}+\tau\delta_1] \tag{3.20}$$

where $\tau\delta_{1/2},\tau\delta_1$ are perturbations on the right hand side of (3.18a). By using Lemma 3.1, it follows

that

$$\|(I-\tfrac{1}{2}\tau A_{2,1})\epsilon_1\| \leqslant \|(I-\tfrac{1}{2}\tau A_{2,1})(I-\tau A_{2,1})^{-1}\|\ \|\epsilon_{1/2}+\tau\delta_1\| \leqslant$$

$$\leqslant \|\epsilon_{1/2}\|+\tau\|\delta_1\| \leqslant \|\epsilon_0\|+\tau(\|\delta_{1/2}\|+\|\delta_1\|).$$

Thus we obtain for scheme (3.18) the stability result

$$\|\epsilon_n\| \leqslant \|\epsilon_0\|+\tau(\|\delta_{1/2}\|+\|\delta_1\|)+2D \quad \text{for all} \quad \tau>0,\ n\geqslant 1. \tag{3.21}$$

Amplification of ϵ_0 through $\hat{\epsilon}_0$ is thus prevented.

We have no practical experience with scheme (3.18). In some numerical experiments with disturbed initial values no large errors were found in the original PR scheme, so that there was little need for stabilization. We think that starting with one LOD step (or a few) may be advantageous in situations where ϵ_0 contains very high frequencies. Like the trapezoidal rule and implicit midpoint rule, the ADI scheme damps such high frequent error components very slowly, whereas the LOD-scheme has strong damping properties [7,20].

3.3. A practical observation

The PR-scheme is implicit, and thus in actual application nonlinear algebraic equations have to be solved by an appropriate iterative method. Because the implicitness is "one-dimensional", it is feasible to implement a Newton-Raphson type method using a direct (band) solver for the arising linear systems, as it is customary in the field of stiff ODEs.

Let $\tilde{U}_n$ denote the numerical values generated by an implemented PR scheme in an actual application. These numerical values can be thought of as being solutions of the perturbed scheme (3.1) where the δ_j are errors due to approximately solving the implicit relations, or, in other words, for stopping the iteration process. If the conditions of Theorem 3.2 hold, one can conclude that the implementation is stable if the stopping criterium is based on the residual test $\|\delta_j\| \leqslant$ [prescribed tolerance].

In applications one frequently circumvents the difficulties connected with the solution of nonlinear algebraic equations by applying linearization, which corresponds to using just one iteration, in some way or another, of a Newton-Raphson type iterative process. In the above setting this means that the δ_j are not controlled, and that, consequently, the stability may deteriorate. The following numerical example, quoted from van der Houwen and Sommeijer [10], serves to illustrate this phenomenon.

Consider the nonlinear parabolic equation

$$u_t = 3(u^2u_x)_x + 3(u^2u_y)_y + xyu - 9t^2(x^2+y^2)u^3, \tag{3.22}$$

with exact solution $u(x,y,t)=\exp(xyt)$, on the unit square $0\leqslant x,y\leqslant 1$ and $0\leqslant t\leqslant 1$. We assume Dirichlet boundary conditions. On a uniform grid, with mesh width h in both directions, we apply the difference formula (similar in y-direction)

$$(u^2u_x)_x|_{(x_i,y_j)} \approx h^{-2}\{u^2_{i+1/2,j}\ u_{i+1,j}-(u^2_{i+1/2,j}+u^2_{i-1/2,j})u_{ij}+u^2_{i-1/2,j}\ u_{i-1,j}\}$$

where $u_{i\pm 1/2,j}=(u_{i\pm 1,j}+u_{i,j})/2$. In the standard way, including equal distribution of the term $xyu-9t^2(x^2+y^2)u^3$ over F_1 and F_2 and natural ordering of gridpoints, one can now set up the semi-discrete system (2.1) and apply scheme (2.2) for the time integration.

Table 3.1 shows the errors at $t=1$, measured in the discrete L_2-norm. In the left part the errors are given for the case where only one iteration step of the Newton-Raphson process is used for solving the nonlinear algebraic equations, and the entries in the right part correspond to 2 iterations (which is sufficient for these τ,h values; more iterations do not alter accuracy). The deterioration of stability is clearly visible.

h^{-1} \ τ^{-1}	10	20	40	80	160	10	20	40	80	160
10	-3.23	2.37	2.89	3.27	3.37	1.88	2.35	2.87	3.26	3.37
20	*	*	-0.12	3.41	3.83	1.81	2.27	2.82	3.38	3.82
40	*	*	*	*	3.89	1.77	2.24	2.78	3.36	3.92
80	*	*	*	*	*	1.76	2.22	2.76	3.34	3.93
160	*	*	*	*	*	1.75	2.21	2.75	3.33	3.92
	1 Newton iteration					2 Newton iterations				

TABLE 3.1. The entries are $-^{10}\log$ ||error at $t=1$||. The symbol * denotes instability (overflow or near overflow).

We emphasize once more that this deterioration of stability is an artifact of the chosen implementation. The experiment shows that the PR method itself is stable for the chosen values of τ and h. Unfortunately, we do not know whether the one sided Lipschitz condition (2.10) is valid, in some suitable norm, for this problem. Finally it should be stressed that the present experiment does not stand on its own. One easily conceives more examples, see for instance [1], Sect. 9.4, for a related discussion.

4. A General Convergence Result

The remainder of this paper is devoted to an investigation of the full convergence properties of the PR scheme (2.2). The prefix full indicates that we shall compare the fully discrete numerical solution U_n directly to the exact PDE solution $u_h(t_n)$, without using the intermediate ODE solution $U(t_n)$. The main objective of our investigation is the order p in time featuring in the general error bound

$$\|u_h(t_n)-U_n\| \leq C_1\tau^p + C_2 \max_{0\leq t\leq T}\|\alpha_h(t)\| \quad (\text{for all } \tau,h>0,\ 0\leq t_n\leq T) \tag{4.1}$$

where C_1, C_2 are constants independent of τ and h, and α_h is the spatial truncation error (2.8). Note that τ and h are allowed to tend to zero simultaneously and independent of each other (unconditional convergence). We assume in the following $U_0 = u_h(0)$.

Convergence will be proved here by using the stability estimates of Section 3 for perturbations δ_j. Let $\tilde{U}_j = u_h(t_j)$ for $j = n, n+1/2$ and $n \geq 0$. The δ_j then stand for residual discretization errors, and the $\epsilon_n = u_h(t) - U_n$ are global discretization errors. By a Taylor expansion of $u_h(t)$ around $t = t_{n+1/2}$, we obtain from the first equation in (3.1)

$$\delta_{n+1/2} = \tfrac{1}{2}\dot{u}_h(t_{n+1/2}) - \tfrac{1}{8}\tau\ddot{u}_h(s_{n+1/2}) - \tfrac{1}{2}F_1(t_{n+1/2}, u_h(t_{n+1/2})) - \tfrac{1}{2}F_2(t_n, u_h(t_n))$$

for some intermediate point $s_{n+1/2} \in (t_n, t_{n+1/2})$. Since $\dot{u}_h(t) = F_1(t, u_h(t)) + F_2(t, u_h(t)) + \alpha_h(t)$, it follows that

$$\delta_{n+1/2} = -\tfrac{1}{8}\tau\ddot{u}_h(s_{n+1/2}) + \tfrac{1}{2}[F_2(t_{n+1/2}, u_h(t_{n+1/2})) - F_2(t_n, u_h(t_n))] + \tfrac{1}{2}\alpha_h(t_{n+1/2}). \tag{4.2}$$

In a similar way we get

$$\delta_{n+1} = \tfrac{1}{8}\tau\ddot{u}_h(s_{n+1}) - \tfrac{1}{2}[F_2(t_{n+1}, u_h(t_{n+1})) - F_2(t_{n+1/2}, u_h(t_{n+1/2}))] + \tfrac{1}{2}\alpha_h(t_{n+1/2}) \tag{4.3}$$

where $s_{n+1} \in (t_{n+1/2}, t_{n+1})$.

Since u_h is the restriction to Ω_h of the exact PDE solution the terms $\ddot{u}_h(s)$ are bounded uniformly in h. Assuming

$$\|F_2(t+\tau, u_h(t+\tau)) - F_2(t, u_h(t))\| \leq C\tau \quad (\text{for } \tau,h>0,\ 0\leq t\leq T-\tau) \tag{4.4}$$

with constant $C>0$, the stability estimate (3.16) directly leads to the following convergence result.

THEOREM 4.1. *Let F_1, F_2 satisfy the one-sided Lipschitz condition (2.10) with $\nu \leq 0$. Assume $u_h \in C^2[0,T]$ and (4.4) holds. Then there are $C_1, C_2 > 0$ such that*

$$\|u_h(t_n) - U_n\| \leq C_1\tau + C_2 \max_{0\leq t\leq T}\|\alpha_h(t)\| \quad (\text{for } \tau,h>0,\ 0\leq t_n\leq T). \quad \square \tag{4.5}$$

The assumption (4.4) is a natural one; since $F(t,u_h(t))=\dot{u}_h(t)-\alpha_h(t)$, a bound as in (4.4) will hold in general for the whole gridfunction $F(t,u_h(t))$, provided only that u is smooth. So, what we assume in fact here, is that this smoothness property is maintained in the splitting of F.

The bound (4.5) only shows order $p=1$ in time, whereas the order on fixed space grids is known to be 2. This discrepancy is caused by the fact that we have avoided bounds on partial derivatives of F_1 and F_2. These contain negative powers of the mesh width in space, so that an error bound based on these quantities becomes useless when $h\to 0$. The material presented in the next section, where we examine linear problems, elucidates this point. There we shall also derive bounds (4.1) with $p=2$.

5. Convergence for Linear Problems with Constant Coefficients

In the following we restrict our attention to initial-boundary value problems where the differential operators in space are linear and constant in time. The semi-discrete system then becomes

$$\dot{U} = f(t,U) = AU + g(t) \tag{5.1}$$

where A is constant. We assume that A can be split, in a natural way, into A_1+A_2, and (cf. (2.10))

$$<A_i v,v>\leqslant 0 \quad \text{for all} \quad v\in\mathbb{R}^M \text{ and } i=1,2. \tag{5.2}$$

The inhomogenous term $g(t)$ will contain two contributions,

$$g(t) = b(t)+f(t). \tag{5.3}$$

Here $b(t)=b_1(t)+b_2(t)$ is supposed to emanate from the boundary conditions, and $f(t)$ represents a source term. For f we shall consider splittings $f_1(t)=\theta f(t)$, $f_2(t)=(1-\theta)f(t)$ with $\theta\in[0,1]$, and $F_i(t,v)=A_i v+b_i(t)+f_i(t)$ for $v\in\mathbb{R}^M, t\in[0,T]$. Note that A_1, A_2 need not commute.

In the remaining sections $O(\tau^p h^k)$ will be used to denote a scalar, or vector, whose absolute value, or norm, is bounded by $C\tau^p h^k$ for all possible τ and h, with $C>0$ a constant independent of τ,h. This notation will also be used for $k=0$; $O(\tau^p)$ thus stands for a term which can be bounded by $C\tau^p$ uniformly for $h>0$.

5.1. The structure of the local discretization error

In this section we shall derive, by using the residual errors δ_j, an expression for the discretization error which is introduced in the PR-process (2.2) in one single step. By expanding the formula (4.2) for $\delta_{n+1/2}$ somewhat further, we get

$$\delta_{n+1/2} = -\tfrac{1}{8}\tau\ddot{u}_h(t_{n+1/2})+\tfrac{1}{4}\tau\dot{F}_2(t_{n+1/2},u_h(t_{n+1/2}))+\tfrac{1}{2}\alpha_h(t_{n+1/2})+$$

$$+\frac{1}{48}\tau^2\dddot{u}_h(s'_{n+1/2})-\frac{1}{16}\tau^2\ddot{F}_2(s''_{n+1/2},u_h(s''_{n+1/2}))$$

for certain intermediate points $s'_{n+1/2},s''_{n+1/2}\in(t_n,t_{n+1/2})$. Note that now total derivatives w.r.t. t of $F_2(t,u_h(t))$ come into play. Instead of (4.4) we shall impose in the following the slightly stronger condition $\ddot{F}_2(t,u_h(t))=O(1)$, which also holds for reasonable splittings provided u is smooth. Define for $t\in[0,T]$

$$w_h(t) = -\tfrac{1}{8}\ddot{u}_h(t)+\tfrac{1}{4}\dot{F}_2(t,u_h(t)). \tag{5.4}$$

We then obtain, assuming $u_h(t)$ to be three times continuously differentiable,

$$\delta_{n+1/2} = \tau w_h(t_{n+1/2})+\tfrac{1}{2}\alpha_h(t_{n+1/2})+O(\tau^2). \tag{5.5a}$$

Similarly

$$\delta_{n+1} = -\tau w_h(t_{n+1/2})+\tfrac{1}{2}\alpha_h(t_{n+1/2}) + O(\tau^2). \tag{5.5b}$$

With our choice $\tilde{U}_j=u_h(t_j)$, the error $\tau\rho_{n+1}$ defined in (3.5) represents the discretization error introduced in one single step of the process (2.2) (local discretization error). Since A_1 and A_2 are constant, we have

$$\begin{aligned}\tau\rho_{n+1} &= (I-\tfrac{1}{2}\tau A_2)^{-1}[r(\tau A_1)\tau\delta_{n+1/2} + \tau\delta_{n+1}] = \\ &= (I-\tfrac{1}{2}\tau A_2)^{-1}(I-\tfrac{1}{2}\tau A_1)^{-1}[(I+\tfrac{1}{2}\tau A_1)\tau\delta_{n+1/2}+(I-\tfrac{1}{2}\tau A_1)\tau\delta_{n+1}].\end{aligned}$$

From (5.5) and (3.9) we get the following result.

LEMMA 5.1. *Consider the semi-discrete system (5.1) with A_1,A_2 satisfying (5.2). Suppose $u_h\in C^3[0,T]$ and $\|\ddot{F}_2(t,u_h(t))\|=O(1)$ $(0\leq t\leq T)$. Then we have for the local discretization error*

$$\tau\rho_{n+1} = (I-\tfrac{1}{2}\tau A_2)^{-1}(I-\tfrac{1}{2}\tau A_1)^{-1}[\tau^3 A_1 w_h(t_{n+1/2})+\tau\alpha_h(t_{n+1/2})] + O(\tau^3). \qquad \square \tag{5.6}$$

It should be noted that (5.6) does not yield a bound $\|\tau\rho_{n+1}\|=O(\tau^3)+O(\tau)\|\alpha_h(t_{n+1/2})\|$ in general, due to the fact that $A_1w_h(t)$ need not be $O(1)$ for $h\to 0$, unless the gridfunction $w_h(t)$ satisfies certain homogeneous boundary conditions imposed by A_1 (these conditions are unnatural; see for example Section 5.2 and [15,21]). The eventual unboundedness of $A_1w_h(t)$ thus originates from the boundaries, and will not show up if one considers pure Cauchy problems with $\Omega=\mathbb{R}^2$.

We do have, in view of Lemma 3.1, $\|(I-\frac{1}{2}\tau A_2)^{-1}\|\leq 1$ and $\|(I-\frac{1}{2}\tau A_1)^{-1}\tau A_1\|\leq 2$, which implies

$\|\tau\rho_{n+1}\| = O(\tau^2) + O(\tau)\|\alpha_h(t_{n+1/2})\|$. As we shall see in the next section such a bound is nearly optimal. At first sight this only leads to a global result $\|\epsilon_n\| = O(\tau) + O(1)\max\|\alpha_h(t)\|$, and this was already established for nonlinear problems. In Section 5.3 it will be shown, however, that cancellation of errors may occur, which then leads to a second order result $\|\epsilon_n\| = O(\tau^2) + O(1)\max\|\alpha_h(t)\|$.

5.2. Local error analysis for a simple heat equation

In this section we consider the inhomogeneous model problem

$$\begin{aligned} u_t &= u_{xx} + u_{yy} + s(x,y,t) \quad \text{on } \Omega, \\ u(x,y,t) &= u_\Gamma(x,y,t) \quad \text{on } \Gamma = \partial\Omega, \\ u(x,y,0) &= u_0(x,y) \quad \text{on } \Omega \cup \Gamma \end{aligned} \tag{5.7}$$

where $0 \leqslant t \leqslant T$ and Ω is the unit square $(0,1)\times(0,1)$. Using standard space discretization on a uniform mesh, we obtain a system (5.1) for which precise bounds on the local errors can be given. First we describe the matrices A_1, A_2 and vectors b_1, b_2, f appearing in the semi-discrete system.

Let $\Omega_h = \{(x_i,y_j) : x_i = ih,\ y_j = jh,\ 1 \leqslant i,j \leqslant m\}$ with $h = 1/(m+1)$. We identify gridfunctions on Ω_h and vectors in $\mathbb{R}^M (M = m^2)$ in a natural way, assuming row-wise ordering on Ω_h. Thus $w : \Omega_h \to \mathbb{R}^M$ will also be written as $w = (w_1^T, \ldots, w_m^T)^T$ with $w_j = (w_{1j}, \ldots, w_{mj})^T \in \mathbb{R}^m$, $w_{ij} = w(x_i,y_j)$, and $w(x_i,y_j)$ will be called a component of the gridfunction w. Further we shall use, for matrices and vectors, the direct (Kronecker) product $\otimes$ (see [12], Sect. 12.1, 12.2 for standard properties). If $v = (v_i)$, $\tilde{v} = (\tilde{v}_i) \in \mathbb{R}^m$, then $v \otimes \tilde{v} = (v_1\tilde{v}^T, \ldots, v_m\tilde{v}^T)^T \in \mathbb{R}^M$ corresponds to the gridfunction with values $v_j\tilde{v}_i$ at (x_i,y_j).

The matrices A_1 and A_2 can be written as

$$A_1 = -I \otimes Q, \quad A_2 = -Q \otimes I \tag{5.8}$$

where I is the $m \times m$ identity matrix, and $Q = h^{-2}\text{tridiag}(-1,2,-1) \in L(\mathbb{R}^m)$ is the usual finite difference operator approximating $-\partial^2/\partial x^2$ in one dimension with Dirichlet boundary conditions (the first and last row of Q contain nonzero entries $h^{-2}(2,-1)$ and $h^{-2}(-1,2)$ respectively). The boundary values are incorporated in $b(t) = b_1(t) + b_2(t) \in \mathbb{R}^M$, $b_1(t)$ having nonzero components $h^{-2}u_\Gamma(x \pm h,y,t)$ on $(x,y) \in \Omega_h$ adjacent to the vertical boundaries, and $b_2(t)$ with nonzero components $h^{-2}u_\Gamma(x,y \pm h,t)$ near the horizontal boundaries. Further $f(t)$ is the restriction of $s(x,y,t)$ to Ω_h.

The matrices A_1 and A_2 are symmetric and negative definite. Thus they satisfy (5.2) w.r.t. the standard inner product $\langle w,\tilde{w}\rangle = h^2 w^T\tilde{w}$. Besides the norm $\|w\| = \langle w,w\rangle^{1/2}$ on $\mathbb{R}^M$, we also will use $|v| = (hv^Tv)^{1/2}$ for $v \in \mathbb{R}^m$.

For the local errors we have (cf. (5.6))

$$\tau\rho_{n+1} = \tau^3(I-\tfrac{1}{2}\tau A_2)^{-1}(I-\tfrac{1}{2}\tau A_1)^{-1}A_1 w_h(t_{n+1/2}) + O(\tau)\alpha_h(t_{n+1/2}) + O(\tau^3), \tag{5.9}$$

where $w_h(t)$, defined in (5.4), is a smooth gridfunction.

LEMMA 5.2. *For any* $\gamma\in[0,1/4)$ *there is constant* $C_\gamma>0$ *such that*

$$\|\tau^3(I-\tfrac{1}{2}\tau A_2)^{-1}(I-\tfrac{1}{2}\tau A_1)^{-1}A_1 w_h(t)\| \leqslant C_\gamma \tau^{2+\gamma} \quad (\text{for } \tau,h>0,\ 0\leqslant t\leqslant T).$$

PROOF. Since $\|(I-\frac{1}{2}\tau A_2)^{-1}\|\leqslant 1$, it is sufficient to prove the above bound for $\tau^3(I-\frac{1}{2}\tau A_1)^{-1}A_1 w$ with $w=w_h(t)$. Let $\tilde{A}_1=-A_1$. This matrix is positive definite, and we can write for arbitrary $\gamma\in[0,1/4)$

$$\|\tau^3(I-\tfrac{1}{2}\tau A_1)^{-1}A_1 w\| = \tau^{2+\gamma}\|(I+\tfrac{1}{2}\tau\tilde{A}_1)^{-1}(\tau\tilde{A}_1)^{1-\gamma}\tilde{A}_1^{\gamma} w\|.$$

The matrix $(I+\frac{1}{2}\tau\tilde{A}_1)^{-1}(\tau\tilde{A}_1)^{1-\gamma}$ is symmetric with eigenvalues contained in

$$\{(1+\tfrac{1}{2}\tau\lambda)^{-1}(\tau\lambda)^{1-\gamma} : \lambda>0\} \subset (0,2),$$

and thus its norm is bounded by 2. Further we have $\tilde{A}_1^{\gamma}=I\otimes Q^{\gamma}$. Hence $\tilde{A}_1^{\gamma}w=((Q^{\gamma}w_1)^T,\ldots,(Q^{\gamma}w_m)^T)^T$ and

$$\|\tilde{A}_1^{\gamma}w\|^2 = h^2(\tilde{A}_1^{\gamma}w)^T(\tilde{A}_1^{\gamma}w) = h\sum_{j=1}^{m}|Q^{\gamma}w_j|^2.$$

It will be shown in the appendix that $|Q^{\gamma}w_j|$ is bounded uniformly for $h>0$ (with a bound only depending on smoothness properties of w_j, which are determined by smoothness of u, cf. (5.4)). Therefore we also have $\|\tilde{A}_1^{\gamma}w\|=O(1)$, which completes the proof. □

With the above lemma we obtain $\|\tau\rho_{n+1}\|=O(\tau^q)+O(\tau)\|\alpha_h(t_{n+1/2})\|$ with $q\approx 2.25$. Note that this is only slightly better than the bound with $q=2$ which we derived directly from (5.6) for arbitrary problems (5.1) satisfying (5.2). In order to demonstrate the sharpness of these bounds, we consider the model problem (5.7) with boundary conditions $u_\Gamma=0$, initial value $u_0=0$, and source term

$$s(x,y,t) = \phi(x)\phi(y) - t(\phi(x)+\phi(y)),\quad \phi(z)=\tfrac{1}{2}z(z-1)\ \ (0\leqslant z\leqslant 1).$$

The exact solution is $u(x,y,t)=t\phi(x)\phi(y)$. Since $\ddot{u}_h=0$ and $b=0$, we have

$$A_1 w_h(t) = \tfrac{1}{4}A_1\dot{F}_2(t,u_h(t)) = \tfrac{1}{4}A_1[A_2\dot{u}_h(t) + \dot{f}_2(t)]$$

with $f_2(t)=(1-\theta)f(t)$. Viewed as gridfunctions, $\dot{u}_h(t)$ and $\dot{f}(t)$ have values $\phi(x)\phi(y)$, $-\phi(x)-\phi(y)$, respectively, for $(x,y)\in\Omega_h$. As vectors in $\mathbb{R}^M$ they can be written as

$$\dot{u}_h(t) = v\otimes v,\quad \dot{f}(t) = -e\otimes v - v\otimes e$$

where $e=(1,\ldots,1)^T, v=(v_1,\ldots,v_m)^T\in\mathbb{R}^m$ with $v_i=\frac{1}{2}ih(ih-1)$ $(1\leqslant i\leqslant m)$. We have $Qv=-e$. By using standard properties of direct products it follows from (5.8) that

$$A_1A_2\dot{u}_h(t) = e\otimes e,\quad A_1\dot{f}(t)=-e\otimes e+v\otimes Qe.$$

Since no space errors are present here (the solution is quadratic), relation (5.9) gives

$$\tau\rho_{n+1} = \tfrac{1}{4}(1-\theta)\tau^3(I-\tfrac{1}{2}\tau A_2)^{-1}(I-\tfrac{1}{2}\tau A_1)^{-1}[v\otimes Qe] + O(\tau^3).$$

It follows that

$$\tau\rho_{n+1} = \tfrac{1}{4}(1-\theta)\tau^3[(I+\tfrac{1}{2}\tau Q)^{-1}v\otimes (I+\tfrac{1}{2}\tau Q)^{-1}Qe] + O(\tau^3),$$

and finally

$$\|\tau\rho_{n+1}\|=\tfrac{1}{4}(1-\theta)\tau^3|(I+\tfrac{1}{2}\tau Q)^{-1}v|\,|(I+\tfrac{1}{2}\tau Q)^{-1}Qe| + O(\tau^3).$$

From $(I+\frac{1}{2}\tau Q)^{-1}v = v+\frac{1}{2}\tau(I+\frac{1}{2}\tau Q)^{-1}e = v+O(\tau)$, we see that there is a constant $C'>0$ such that $|(I+\frac{1}{2}\tau Q)^{-1}v|>C'$ whenever $\tau>0$ is sufficiently small. In the appendix it will be shown that there exist a $C''>0$ such that $|(I+\frac{1}{2}\tau Q)^{-1}Qe|\geqslant C''\tau^{-3/4}$ for all $h>0$ sufficiently small and τ/h^2 bounded away from 0. For such τ and h, and $\theta\neq 1$, we thus have

$$\|\tau\rho_{n+1}\|\geqslant C\tau^{2.25} \tag{5.10}$$

with $C>0$ independent of τ and h.

In case $\theta=1$ no order reduction occurs for this specific example with $u_\Gamma=0$; we get $\|\tau\rho_{n+1}\|=O(\tau^3)$ as on fixed space grids, since then $w_h(t)$ vanishes near the vertical boundaries, so that $A_1w_h(t)=O(1)$. For more complicated examples with time dependent boundary conditions the order will reduce also if $\theta=1$.

The observation that inhomogeneous terms cause local order reduction seems to originate with D'Yakonov [4]. Fairweather and Mitchell [5] introduced a correction term which restores the local order; they considered (5.7) with $s=0$ and time dependent boundaries, but as we saw above such a correction is also necessary if $u_\Gamma=0$, $s\neq 0$. The correction consists of replacing in the PR scheme

$b_1(t)$ by $\bar{b}_1(t)=b_1(t)-c(t)$, $c(t)$ being the correction term still to be specified. This corresponds to a change in the boundary values for the intermediate solution $U_{n+1/2}$. With this correction we derive, in the same way as before, the expression for local errors

$$\tau\rho_{n+1} = (I-\tfrac{1}{2}\tau A_2)^{-1}(I-\tfrac{1}{2}\tau A_1)^{-1}\tau^3 \nu_h(t_{n+1/2})+O(\tau)\alpha_h(t_{n+1/2})+ O(\tau^3) \tag{5.11}$$

where now

$$\nu_h(t) = A_1 w_h(t) + \tau^{-2}c(t). \tag{5.12}$$

The gridfunction $w_h(t)$ is smooth, and thus all components of $A_1 w_h(t)$ are $O(1)$, except those corresponding to a gridpoint adjacent to the vertical boundaries. Consider the gridpoint (x_1,y_j) near the left boundary. There we have

$$[A_1 w_h(t)]_{1j} = h^{-2}(-2w_{1j}(t) + w_{2j}(t)).$$

The correction $c(t)$ can now be used to compensate for the missing value $w_{0j}(t)$. Thus we define

$$\begin{aligned} c_{1j}(t) &= \tau^2 h^{-2} w_{0j}(t), \\ w_{0j}(t) &= -\tfrac{1}{8}\ddot{u}_{0j}(t)+\tfrac{1}{4}h^{-2}(\dot{u}_{0j+1}(t)-2\dot{u}_{0j}(t)+\dot{u}_{0j-1}(t))+\tfrac{1}{4}(1-\theta)\dot{s}_{0j}(t) \end{aligned} \tag{5.13}$$

where $u_{0j}(t)=u(0,y_j,t)$, $s_{0j}(t)=s(0,y_j,t)$. In a similar way we define $c_{mj}(t)$ to compensate for the missing values near the right boundary, and we take $c_{ij}(t)=0$ for gridpoints with $1<i<m$. This causes $\nu_h(t)=O(1)$, so that $\tau\rho_{n+1}=O(\tau^3)+O(\tau)\alpha_h(t_{n+1/2})$.

The correction (5.13) slightly differs from the one in [5]. The reason is that we started from the particular form for the local error with w_h given by (5.4). Since $\|(I-\frac{1}{2}\tau A_1)^{-1}\tau A_1\|=O(1)$, it follows that (5.11) also holds with $\nu_h(t)=A_1\tilde{w}(t)+\tau^{-2}c(t)$,

$$\begin{aligned} \tilde{w}(t)=&-\tfrac{1}{2}\tau^{-2}[u_h(t-\tfrac{1}{2}\tau)-2u_h(t)+u_h(t+\tfrac{1}{2}\tau)] \\ &+\tfrac{1}{4}\tau^{-1}[F_2(t+\tfrac{1}{2}\tau,u_h(t+\tfrac{1}{2}t))-F_2(t-\tfrac{1}{2}\tau,u_h(t-\tfrac{1}{2}\tau))]. \end{aligned}$$

This leads to (5.13) with $\ddot{u}_{0j}(t)$ and $\dot{u}_{0j}(t)$ replaced by standard differences, which is then the same as the original correction of Fairweather and Mitchell [5]. Generalization of this boundary value correction to a large class of initial-boundary value problems can be found in [17].

5.3. Cancellation of local errors

Let q be the order in time of the local discretization errors. One then naively expects order $p = q - 1$ for the global errors, as a result of addition of all the local errors. For the example of Section 5.2 this would give $p = 1.25$ only, instead of second order in time as on a fixed space grid. In this section it will be shown that under suitable assumptions we still have $p = 2$, due to cancellation of local errors. In other words, the local errors suffer from a reduction in order, but in the transition from local to global this reduction is annihilated. A similar behaviour can be observed with the implicit midpoint rule, and to some extent, with other Runge-Kutta methods [15,21]. A comprehensive analysis for the implicit midpoint rule can be found in [11]. The proof of Theorem 5.3 below was inspired by this analysis.

We consider again the general linear problem (5.1), but it will be assumed now that there exists a constant $C > 0$ (independent of τ, h), such that

$$\|A^{-1}A_1\| \leqslant C, \tag{5.14}$$

$$\|R^n\| \leqslant C \quad \text{for all} \quad n \geqslant 0 \quad \text{with} \quad t_n \in [0,T]. \tag{5.15}$$

Here $R = (I - \frac{1}{2}\tau A_2)^{-1} r(\tau A_1)(I + \frac{1}{2}\tau A_2)$ is the matrix governing stability of the PR scheme (cf. formula (3.6)). Thus (5.15) is a Lax-Richtmeyer type condition for stability [14]. If A_1, A_2 are commuting, negative definite matrices, then both (5.14) and (5.15) hold with $C = 1$. The results below are thus applicable to the heat equation with standard space discretization on a uniform mesh, which we considered in the previous section.

THEOREM 5.3. *Suppose the conditions of Lemma 5.1 are satisfied, and (5.14), (5.15) hold. Then there are constants $C_1, C_2 > 0$ such that*

$$\|u_h(t_n) - U_n\| \leqslant C_1\tau^2 + C_2 \max_{0 \leqslant t \leqslant T} \|\alpha_h(t)\| \quad (\text{for all} \quad \tau, h > 0, \ 0 \leqslant t_n \leqslant T). \tag{5.16}$$

PROOF. Consider the recursion for the global errors $\epsilon_n = u_h(t_n) - U_n$,

$$\epsilon_{n+1} = R\epsilon_n + \tau\rho_{n+1} \quad (n \geqslant 0)$$

(cf. (3.5)). The local errors, given by (5.6), can be written as

$$\tau\rho_{n+1} = \tau^3(I - \tfrac{1}{2}\tau A_2)^{-1}(I - \tfrac{1}{2}\tau A_1)^{-1} A_1 w_h(t_{n+1}) + O(\tau)\alpha_h(t_{n+1/2}) + O(\tau^3).$$

Note that w_h is evaluated here at $t = t_{n+1}$. For the proof of this theorem we may omit the terms

$O(\tau)\alpha_h(t_{n+1/2})+O(\tau^3)$, since these will give only a contribution $O(1)\max\|\alpha_h(t)\|+O(\tau^2)$ to the global bound.

We define, for all $n\geqslant 0$,

$$\bar{\rho}_n = \tau A^{-1}A_1 w_h(t_n),\quad \bar{\epsilon}_n = \epsilon_n+\tau\bar{\rho}_n.$$

Using the equality $R-I=(I-\frac{1}{2}\tau A_2)^{-1}(I-\frac{1}{2}\tau A_1)^{-1}\tau A$, we get

$$\rho_{n+1} = (R-I)\bar{\rho}_{n+1}.$$

Therefore the $\bar{\epsilon}_n$ satisfy the recursion

$$\bar{\epsilon}_{n+1} = R\bar{\epsilon}_n + \tau R(\bar{\rho}_{n+1}-\bar{\rho}_n)\quad (n\geqslant 0).$$

The stability assumption (5.15) provides the global estimate

$$\|\bar{\epsilon}_n\|\leqslant C\|\bar{\epsilon}_0\|+Ct_n\max_k\|\bar{\rho}_{k+1}-\bar{\rho}_k\|.$$

Since we have, in view of (5.14), $\|\epsilon_k-\bar{\epsilon}_k\|=\|\tau\bar{\rho}_k\|=O(\tau^2)$ and $\|\bar{\rho}_{k+1}-\bar{\rho}_k\|=O(\tau^2)$ for all $k\geqslant 0$, the second order result for $\|\epsilon_n\|$ now follows. □

Lest we miss the obvious, the introduction of the new errors $\bar{\epsilon}_n,\bar{\rho}_n$ in the above proof is redundant if $A_1 w_h(t)=0(1)$ uniformly for $h\to 0$. We can prove then (5.16) (by using the stability result (3.15)) without the assumptions (5.14) and (5.15). However, the material presented in Section 5.2 shows that in general $\|A_1 w_h(t)\|\to\infty$ for $h\to 0$, unless the PDE solution u_h and the inhomogeneous term f_2 meet additional conditions near the boundary Γ. These conditions are unnatural, in the sense that they are only imposed by the PR scheme and they are unrelated to smoothness of the PDE solution.

Convergence results for general ADI methods applied to linear initial-boundary value problems were obtained by Douglas and Gunn [2]. They considered homogeneous boundary conditions, in which case it can be shown that the local error of the PR method has the same order as the local error of the Crank-Nicolson type scheme

$$(I-\tfrac{1}{2}\tau A)U_{n+1}=(I+\tfrac{1}{2}\tau A)U_n + \tau\phi_n \tag{5.17a}$$

with inhomogeneous term

$$\phi_n = f_1(t_{n+1/2})+\tfrac{1}{2}f_2(t_n)+\tfrac{1}{2}f_2(t_{n+1})-\tfrac{1}{4}\tau A_1(f_2(t_{n+1})-f_2(t_n)). \tag{5.17b}$$

Due to this special inhomogeneous term, the same additional conditions show up if one tries to prove

second order convergence for this scheme. Moreover, the approach of Douglas and Gunn is hard to generalize for inhomogeneous boundary conditions.

As an illustration of the local order reduction and cancellation of local errors, we consider the simple problem

$$u_t = (1+y)u_{xx} + u_{yy} + s(x,y,t) \tag{5.18}$$

on the unit rectangle with Dirichlet boundary conditions. The source term and initial-boundary values are chosen such that the exact solution is $u(x,y,t)=\exp(x+y+t)$. In space standard discretization was used on a uniform mesh with grid distance h in both directions. The PR scheme (2.2) was applied with $\tau=h$ and equal distribution of the source term ($f_1=f_2=f/2$; the choice $f_1=f$, $f_2=0$ leads to similar results). The following table shows the number of correct digits $-^{10}\log \|\epsilon_n\|$ for $n=1$ (local error) and $n=N, \tau N=1$ (global error).

τ^{-1}	5	10	20	40
local error	2.03	2.58	3.18	3.80

τ^{-1}	5	10	20	40
global error	1.68	2.20	2.76	3.35

TABLE 5.1. Errors for PR scheme applied to (5.18) with $\tau=h$.

One nicely sees second order, approximately, for both local and global errors (increase of $\simeq 0.6$ upon step halving). We note that since the continuous operators $(1+y)\partial^2/\partial x^2$ and $\partial^2/\partial y^2$ do not commute, the matrices A_1 and A_2 will not commute either. Therefore we do not know whether the stability condition (5.15) holds. The numerical results, however, indicate that the conclusions of Theorem 5.3 are valid here.

With a Fairweather-Mitchell type correction we would obtain third order locally, but still second order for the global errors. Although this correction technique does not increase the global order, the numerical results of Sommeijer et al. [17] indicate that in many cases the numerical approximations will become more accurate (the error constant C_1 in (5.16) may be smaller for corrected schemes).

ACKNOWLEDGEMENT: The authors are grateful to Joke Blom for carrying out the numerical experiments.

REFERENCES

[1] K. DEKKER and J.G. VERWER, *Stability of Runge-Kutta methods for stiff nonlinear differential equations,* North-Holland, Amsterdam - New York - Oxford, 1984.

[2] J. DOUGLAS, JR., and J.E. GUNN, *A general formulation of alternating direction methods, Part I: Parabolic and hyperbolic problems,* Numer. Math., 6 (1964), pp. 428-453.

[3] J. DOUGLAS, JR., and H.H. RACHFORD, *On the numerical solution of heat conduction problems in two and three space variables,* Trans. Amer. Math. Soc., 82 (1956), pp. 421-439.

[4] E.G. D'YAKONOV, *Difference schemes with a disintegrating operator for multi-dimensional problems,* Zh. Vychisl. Math. Fiz., 2 (1962), pp. 549-568.

[5] G. FAIRWEATHER and A.R. MITCHELL, *A new computational procedure for A.D.I. methods,* SIAM J. Numer. Anal., 4 (1967), pp. 163-170.

[6] A.R. GOURLAY, *Splitting methods for time dependent partial differential equations,* The State of the Art in Numerical Analysis , D. Jacobs (ed.), Academic Press, London - New York - San Francisco (1977), pp. 757-791.

[7] A.R. GOURLAY and J.LI. MORRIS, *The extrapolation of first order methods for parabolic partial differential equations* II, SIAM J. Numer. Anal., 17 (1980), pp. 641-655.

[8] A.R. GOURLAY and A.R. MITCHELL, *The equivalence of certain alternating direction and locally one dimensional difference methods,* SIAM J. Numer. Anal., 6 (1969), pp. 37-46.

[9] P.J. VAN DER HOUWEN and J.G. VERWER, *One-step splitting methods for semi-discrete parabolic equations,* Computing, 22 (1979), pp. 291-309.

[10] P.J. VAN DER HOUWEN and B.P. SOMMEIJER, *Improving the stability of predictor-corrector methods by residue smoothing,* Report NM-R8707, Centre for Mathematics and Computer Science, Amsterdam, 1987.

[11] H.F.B.M. KRAAIJEVANGER, *B-convergence of the implicit midpoint rule and the trapezoidal rule,* BIT, 25 (1985), pp. 652-666.

[12] P. LANCASTER and M. TISMENETSKY, *The theory of matrices,* Academic Press, Orlando, 1985.

[13] D.W. PEACEMAN and H.H. RACHFORD, JR., *The numerical solution of parabolic and elliptic differential equations,* J. Soc. Ind. Appl. Math., 3 (1955), pp. 28-41.

[14] R.D. RICHTMEYER and K.W. MORTON, *Difference methods for initial-value problems,* Interscience Publ., New York - London - Sydney, 1967.

[15] J.M. SANZ-SERNA, J.G. VERWER and W.H. HUNDSDORFER, *Convergence and order reduction of Runge-Kutta schemes applied to evolutionary partial differential equations,* Numer. Math., 50 (1987), pp. 405-418.

[16] J.M. SANZ-SERNA and J.G. VERWER, *Stability and convergence in the stiff ODE/PDE interface,* Applied Numerical Mathematics, to appear.

[17] B.P. SOMMEIJER, P.J. VAN DER HOUWEN and J.G. VERWER. *On the treatment of time-dependent boundary conditions in splitting methods for parabolic differential equations,* Int. J. Numer. Math. Engnr., 17 (1981), pp. 335-346.

[18]] J.G. VERWER, *Contractivity of locally one-dimensional splitting methods,* Numer. Math., 44 (1984), pp. 247-259.

[19] J.G. VERWER and J.M. SANZ-SERNA, *Convergence of method of lines approximations to partial differential equations,* Computing, 33 (1984), pp. 297-313.

[20] J.G. VERWER and H.B. DE VRIES, *Global extrapolation of a first order splitting method,* SIAM J. Sci. Stat. Comput., 6 (1985), pp. 771-780.

[21] J.G. VERWER, *Convergence and order reduction of diagonally implicit Runge-Kutta schemes in the method of lines,* Numerical Analysis, D.F. Griffiths & G.A. Watson (eds.), Pitman Research Notes in Mathematics Series 140 (1986), pp. 220-237.

APPENDIX

In this appendix some technical results will be derived concerning the finite difference operator approximating $-\partial^2/\partial x^2$, which were used in Section 5.2. Similar results for the continuous operator were obtained by Brenner, Crouzeix and Thomée (RAIRO Analyse Numérique 16 (1982), 5-26). The discrete case has some minor additional difficulties.

Let $m \in \mathbb{N}$ and $h = 1/(m+1)$. Consider the $m \times m$ matrix

$$Q = h^{-2} \begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix}$$

This matrix has the eigenvalue-eigenvector decomposition

$$Q = V \Lambda V^{-1}$$

with

$$V = [v_1, v_2, \ldots, v_m], \quad v_j = \sqrt{2}(\sin(jh\pi), \sin(2jh\pi), \ldots, \sin(mjh\pi))^T \in \mathbb{R}^m$$

and

$$\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_m), \quad \lambda_j = 4h^{-2}\sin^2(jh\pi/2).$$

We have $hv_i^T v_j = \delta_{ij}$ (Kronecker delta). Therefore $hV^T V = I$, and

$$Q = hV \Lambda V^T.$$

On the $\mathbb{R}^m$ we consider the innerproduct $(v,w) = hv^T w$ and norm $|w| = (w,w)^{1/2}$. For any $w \in \mathbb{R}^m$ and $\psi: \mathbb{C} \to \mathbb{C}$, analytic on the positive real axis, we have

$$\psi(Q)w = \sum_{j=1}^{m} (v_j, w)\psi(\lambda_j)v_j,$$

and, since $(v_i, v_j) = \delta_{ij}$, it follows that

$$|\psi(Q)w|^2 = \sum_{j=1}^{m} |(v_j, w)\psi(\lambda_j)|^2.$$

For any two real functions f, g we will use the notation $f(x) \sim g(x)$ $(x \downarrow 0)$ if there are positive numbers C_0, C_1 and H such that $C_0 g(x) \leq f(x) \leq C_1 g(x)$ for $0 < x \leq H$. Further $x_j = jh$ for

$j=1,2,...,m.$

LEMMA 1. *Suppose* $f\in C(0,1]$, $\gamma\geqslant 0$ *and* $f(x)\sim x^{-\gamma}$ $(x\downarrow 0)$. *Then, for* $h\downarrow 0$,

$$h\sum_{j=1}^{m} f(jh)\sim \begin{cases} 1 & \text{if } \gamma<1, \\ \log(1/h) & \text{if } \gamma=1, \\ h^{1-\gamma} & \text{if } \gamma>1. \end{cases}$$

PROOF. Since we can split f into a monotonically decreasing part ($\sim x^{-\gamma}$ for $x\downarrow 0$) and a bounded remainder, it is clear that we may assume without loss of generality that f itself is monotonically decreasing. Then

$$h\sum_{j=1}^{m} f(x_j) \geqslant \int_{h}^{1} f(x)dx.$$

On the other hand

$$h\sum_{j=1}^{m} f(x_j) = hf(h) + h\sum_{j=2}^{m} f(x_j) \leqslant hf(h) + \int_{h}^{1} f(x)dx.$$

We have $hf(h)\sim h^{1-\gamma}$ $(h\downarrow 0)$. The integral $\int_h^1 f(x)dx$ is $\sim -\log h$ $(h\downarrow 0)$ if $\gamma=1$, and $\sim 1+h^{1-\gamma}(h\downarrow 0)$ if $\gamma\neq 1$. □

In the remainder e_j will stand for the vector in $\mathbb{R}^m$ with $j-th$ component equal to 1 and the other components 0. The vector $(1;1,...,1)^T\in\mathbb{R}^m$ will be denoted by e.

LEMMA 2. *Let* $\gamma\geqslant 0$. *We have*

$$\sup_{h>0} |Q^{\gamma}e|<\infty \Leftrightarrow \gamma<1/4.$$

PROOF. Since $Qe=h^{-2}(e_1+e_m)$, it follows that

$$Q^{\gamma}e = h^{-2}Q^{\gamma-1}(e_1+e_m) = h^{-2}\sum_{j=1}^{m}[h(v_{j1}+v_{jm})]\lambda_j^{\gamma-1}v_j$$

with $v_{ji}=\sqrt{2}\sin(ijh\pi)$ the i-th component of v_j. If j is even, we have $v_{j1}+v_{jm}=0$, while $v_{j1}+v_{jm}=2v_{j1}$ if j is odd. In the limit $h\downarrow 0$, we thus obtain

$$|Q^\gamma e|^2 = h^{-2}\sum_{j=1}^{m}|(\nu_{j1}+\nu_{jm})\lambda_j^{\gamma-1}|^2 \sim 2h^{-2}\sum_{j=1}^{m}|\nu_{j1}\lambda_j^{\gamma-1}|^2 =$$

$$= 4^{2\gamma-1}h^{2-4\gamma}\sum_{j=1}^{m}|\sin(jh\pi)\,\sin(jh\pi/2)^{2\gamma-2}|^2 =$$

$$= 4^{2\gamma}h^{1-4\gamma}h\sum_{j=1}^{m}\cos^2(jh\pi/2)\,\sin(jh\pi/2)^{4\gamma-2}.$$

From lemma 1 we see that $|Q^\gamma e|^2 \sim h^{1-4\gamma}(1+h^{4\gamma-1})$ $(h\downarrow 0)$ for $\gamma \neq 1/4$, whereas $|Q^\gamma e|^2 \sim -\log h$ $(h\downarrow 0)$ for $\gamma = 1/4$. □

LEMMA 3. *Let* $\phi \in C^3[0,1]$ *and* $w=(w_1,w_2,\ldots,w_m)^T$ *with* $w_j=\phi(x_j)$ *for* $1\leqslant j\leqslant m$, $m\in\mathbb{N}$. *Suppose* $\gamma<1/4$. *Then* $\sup_{h>0}|Q^\gamma w|<\infty$.

PROOF. We have

$$Qw = (a_1+h^{-2}b_1, a_2, a_3, \ldots, a_m+h^{-2}b_m)^T$$

with $a_j = h^{-2}(-\phi(x_{j-1})+2\phi(x_j)-\phi(x_{j+1})) \approx -\phi''(x_j)$ and $b_1=\phi(0)$, $b_m=\phi(1)$. Let $a=(a_1,a_2,\ldots,a_m)^T$. Then

$$Q^\gamma w = Q^{\gamma-1}a + h^{-2}Q^{\gamma-1}[b_1e_1 + b_me_m].$$

Since Q is positive definite with eigenvalues larger than 1, the vector $Q^{\gamma-1}a$ is bounded, uniformly for $h>0$, for any $\gamma<1$. In the proof of lemma 2 we saw that $h^{-2}Q^{\gamma-1}e_j$ $(j=1$ or $m)$ is also bounded uniformly in h, provided that $\gamma<1/4$. □

COROLLARY 4. Let w be as in lemma 3. For any $\gamma\in[0,1/4)$ there exsits a constant $C_\gamma>0$ such that

$$|(I+\tfrac{1}{2}\tau Q)^{-1}Qw| \leqslant C_\gamma \tau^{\gamma-1} \quad \text{(for all } \tau,h>0).$$

PROOF. The matrix $(I+\frac{1}{2}\tau Q)^{-1}(\tau Q)^{1-\gamma}$ is symmetric with eigenvalues in the interval (0,2), so that we have for the norm of this matrix $|(I+\frac{1}{2}\tau Q)^{-1}(\tau Q)^{1-\gamma}|\leqslant 2$. Since $|Q^\gamma w|$ is uniformly bounded for $h>0$, the proof follows from the inequality

$$|(I+\tfrac{1}{2}\tau Q)^{-1}Qw| \leqslant \tau^{\gamma-1}|(I+\tfrac{1}{2}\tau Q)^{-1}(\tau Q)^{1-\gamma}|\,|Q^\gamma w|.$$ □

LEMMA 5. *Let $\alpha>0$, and assume $\tau/h^2 \geq \alpha$. Then there exists a constant $C>0$ such that for all $\tau>0$*

$$|(I+\tfrac{1}{2}\tau Q)^{-1}Qe| \geq C\tau^{-3/4}.$$

PROOF. It is sufficient to consider $h>0$ sufficiently small. Let

$$\mu(\tau,h) = |(I+\tfrac{1}{2}\tau Q)^{-1}Qe|^2.$$

Similar as in the proof of lemma 2, we obtain for $h\downarrow 0$

$$\mu(\tau,h) = h^{-2}\sum_{j=1}^{m}|(v_{j1}+v_{jm})(1+\tfrac{1}{2}\tau\lambda_j)^{-1}|^2 \sim 2h^{-2}\sum_{j=1}^{m}|v_{j1}(1+\tfrac{1}{2}\tau\lambda_j)^{-1}|^2.$$

It follows that for arbitrary $\beta>0$ and $h>0$ sufficiently small

$$\mu(\tau,h) \geq 4(1+\tfrac{1}{2}\beta)^{-2}h^{-2}\sum_{j\in J_\beta}\sin^2(jh\pi)$$

where $J_\beta=\{j:\ \frac{1}{2}\beta\leq\tau\lambda_j\leq\beta\}$. We take $\beta=2\alpha$, so that $\beta h^2/4\tau\leq 1/2$. Since $\tau\lambda_j = 4\tau h^{-2}\sin^2(jh\pi/2)$, the index j belongs to J_β iff

$$\beta h^2/8\tau \leq \sin^2(jh\pi/2) \leq \beta h^2/4\tau, \quad \text{i.e.,}$$

$$2(\pi h)^{-1}\arcsin\sqrt{\beta h^2/8\tau} \leq j \leq 2(\pi h)^{-1}\arcsin\sqrt{\beta h^2/4\tau}.$$

For each $j\in J_\beta$ we thus have $\sin^2(jh\pi)\geq\beta h^2/8\tau$. Inspection of the graph of the arcsin function shows that the number of terms in J_β is $\sim h^{-1}\sqrt{h^2/\tau}=\tau^{-1/2}$ (for $h\downarrow 0$). From the above lower bound for $\mu(\tau,h)$ it now follows that there exists a constant $C>0$ such that

$$\mu(\tau,h)\geq C^2h^{-2}(h^2/\tau)\tau^{-1/2} = C^2\tau^{-3/2}$$

for $h>0$ sufficiently small. □

REMARK. The upper and lower bounds of corollary 4 and lemma 5 also hold if $(I+\frac{1}{2}\tau Q)^{-1}$ is replaced by $\psi(\tau Q)$ with ψ an arbitrary rational function satisfying $\psi(\infty)=0$ and $|\psi(z)|\leq 1$ for $z\in\mathbb{C}, \mathrm{Re}\, z\geq 0$.

If w is the restriction to $\{x_j\}$ of a smooth function $\phi:[0,1]\to\mathbb{R}$ with $\phi(0)=\phi(1)=0$, then a bound as in corollary 4 also holds for $|\psi(\tau Q)Q^2w|$ (sharpness then follows by considering $\phi(x)=x(x-1)$). □