IMPACT OF COMPUTING IN SCIENCE AND ENGINEERING 3, 93–122 (1991)

An Adaptive Multilevel Approach to Parabolic Equations

II. Variable-Order Time Discretization Based on a Multiplicative Error Correction

FOLKMAR A. BORNEMANN

Konrad-Zuse-Zentrum für Informationstechnik Berlin, Heilbronner Strasse 10, D-1000 Berlin 31, Federal Republic of Germany

Received January 15, 1991

Folkmar A. Bornemann, An Adaptive Multilevel Approach to Parabolic Equations, II. Variable-Order Time Discretization Based on a Multiplicative Error Correction. *IMPACT of Computing in Science and Engineering* **3**, 93–122 (1991).

In continuation of part I this paper develops a variable-order time discretization in Hilbert space based on a multiplicative error correction. Matching of time and space errors as explained in part I allows to construct an adaptive multilevel discretization of the parabolic problem. In contrast to the extrapolation method in time, which has been used in part I, the new time discretization allows us to separate space and time errors and further to solve fewer elliptic subproblems with less effort—a feature which is essential in view of the application to space dimensions greater than one. Numerical examples for space dimension one are included which clearly indicate the improvement.

INTRODUCTION

In part I of this paper [2] the author developed an adaptive approach to parabolic equations by constructing a variable-order and time-step control mechanism in function space, viewing the parabolic initial–boundary-value problem as an abstract Cauchy problem in that function space. Discretization of the arising spatial partial differential equations is viewed as a *perturbation* of the discrete orbit in the function space, which has to be pushed below a level not touching the accuracy of the parabolic problem. The author developed to some extent a theory of single-step methods in Hilbert space applicable to abstract Cauchy problems which involve an m-sectorial operator. Within the scope of that theory the use of extrapolation techniques has been justified.

0899-8248/91 $3.00

For *one space dimension* the author implemented an extrapolated implicit Euler scheme together with an adaptive multilevel FEM solver, which handles the arising singularly perturbed elliptic subproblems. Very promising results were obtained [2], which led the author to try the same approach in *two space dimensions* since the theory developed is independent of space dimension.

However, *several structural drawbacks of extrapolation methods showed up:*

1. In order to keep the perturbations of the function space orbit small, the more accurately the elliptic subproblems have to be solved, the higher the suggested order of the time discretization is. This yields increasingly high computational effort for the spatial part if the order of the time discretization is increased. That increase is so high in the framework of extrapolation methods that for the 2D case an unreasonable amount of work occurs, which in turn restricts the order to the lowest possible one—thus killing the *variable* order device as a whole.

The reason: Extrapolation creates higher order approximations through higher order differences. But iterated higher differences *amplify* the spatial perturbations.

2. Extrapolation imposes elliptic subproblems for different time steps. In order to make the algebraic operation of extrapolation possible we have to use an algorithm as explained in [2, Sect. 4.3]. This yields a final triangulation which has to be good for all the time steps which have been used. In the 2D case this final triangulation contains far too many nodal points, which gives rise to far too much work.

3. Since we get the time-error estimate as a *difference* of two entities attached with a spatial error, we can only detect whether the time error is *below* the given tolerance, but we are not able to get the order of magnitude of the time error if the error is below the given tolerance. Therefore the algorithm is not able to detect stationary phases.

The first two drawbacks are *not serious* in the 1D case or for ODEs, which explains why they have first been observed in the 2D case. The third one is intimately connected to the first one.

The present paper tries to avoid these drawbacks by constructing a variable-order time discretization with the features

- Avoidance of differences for higher order approximations.
- Only one kind of elliptic problem at each time step—at the most with different right hand sides.

In Section 1 we give a formulation of the continuous problem which allows us to consider arbitrary domains Ω. This is important for the 2D case. Furthermore we give a short draft of the algorithm together with its requirements.

Section 2 is the *core* of the paper, where the variable-order time discretization is derived, which corrects error approximations by multiplication in order to avoid differences, thus leading to the title of this paper. The application to the abstract Cauchy problem is explained and the structure of the arising elliptic subproblems is analyzed.

Section 3 is devoted to the control of the perturbations arising from spatial discretization. Some details for the implementation are given.

In Section 4 we give numerical results for one space dimension—in spite of the fact that the author has already computed successfully examples in *two* space dimensions; since these results need a careful explanation of error-estimation and preconditioning of the singularly perturbed elliptic subproblems, they are subjects of a forthcoming part III of the paper. In Section 4 special attention is paid to enlightening the improvement which has been achieved in comparison with extrapolation methods. It turns out that we end up with *multigrid complexity,* i.e., the computing time is proportional to the number of introduced unknowns.

1. Preliminary Considerations

1.1. *The Problem*

Throughout this paper we are concerned with *temporally homogeneous* parabolic initial–boundary-value problems:

Given a domain $\Omega \subset \mathbb{R}^d$, a time $T > 0$, and functions f, $u_0 \in L^2(\Omega)$, solve

$$\begin{aligned} &\text{(i)}\ \frac{\partial u(t,x)}{\partial t} + A(x,\partial)u(t,x) = f(x), && x \in \Omega,\ t \in]0,T]; \\ &\text{(ii)}\ u(t,\cdot)|_{\partial\Omega} = 0, && t \in]0,T]; \\ &\text{(iii)}\ u(0,\cdot) = u_0; && \end{aligned} \tag{1.1}$$

where $A(x,\partial)$ denotes a strongly elliptic *formally selfadjoint* operator of second order,

$$A(x,\partial) = \sum_{0 \le |\rho|,|\sigma| \le 1} (-1)^{|\rho|} \partial^\rho (a^{\rho\sigma}(x)\partial^\sigma),$$

where $a^{\rho\sigma} \in L^\infty(\Omega)$, $a^{\rho\sigma} = a^{\sigma\rho}$ in the usual multiindex notation. Thus the induced continuous bilinear form $a(\cdot,\cdot)$ on $H_0^1(\Omega) \times H_0^1(\Omega)$ given by

$$a(u,v) = \sum_{0 \le |\rho|,|\sigma| \le 1} \int_\Omega a^{\rho\sigma} \partial^\rho u \partial^\sigma v dx, \qquad u, v \in H_0^1(\Omega),$$

is *symmetric.*

We will further assume the $H_0^1(\Omega)$-*ellipticity* of the form $a(\cdot, \cdot)$: There is a constant $c_1 > 0$ such that

$$a(u, u) \geq c_1 \|u\|_1^2 \qquad \text{for all } u \in H_0^1(\Omega). \tag{A1}$$

Notation. The norms of the Sobolev spaces $H^s(\Omega)$ will be denoted by $\|\cdot\|_s$ and the inner product of $L^2(\Omega)$ will be denoted by $(\cdot, \cdot)$.

The following considerations mainly serve the purpose of developing a concept of *solution* of the parabolic problem, which justifies our approach without additional regularity assumptions.

THEOREM 1.1. *Suppose that assumption* (A1) *is satisfied, then the following holds*

(a) *There is exactly one positive selfadjoint operator*

$$A: D_A \subset L^2(\Omega) \to L^2(\Omega)$$

satisfying

(i) $D_A \subset H_0^1(\Omega)$,

(ii) $a(u, v) = (Au, v)$ *for all* $u \in D_A$, $v \in H_0^1(\Omega)$. (1.2)

Furthermore we have:

(b) *The domain of definition* D_A *is dense in* $H_0^1(\Omega)$ *with respect to the Hilbert space topology of* $H_0^1(\Omega)$.

(c) *For every* $f \in L^2(\Omega)$ *the solution* $u \in H_0^1(\Omega)$ *of the variational problem*

$$a(u, v) = (f, v) \qquad \text{for all } v \in H_0^1(\Omega)$$

exists and satisfies in addition

$$u \in D_A, \qquad Au = f.$$

(d) *The square root* $A^{1/2}$ *of* A *exists with* $D_{A^{1/2}} = H_0^1(\Omega)$ *and satisfies*

$$a(u, v) = (A^{1/2}u, A^{1/2}v) \qquad \text{for all } u, v \in H_0^1(\Omega).$$

Proof. The assertions (a) and (b) are essentially the Friedrichs representation theorem of semibounded symmetric bilinear forms in Hilbert space; consult, e.g., Kato [5, pp. 322f.]. The solution $u \in H_0^1(\Omega)$ of the variational problem

exists by the Lax—Milgram Lemma and the rest of assertion (c) holds again by the Friedrichs representation theorem. For assertion (d) consult, e.g., Kato [5, pp. 331f.]. ■

Remark 1.2. Let $f \in L^2(\Omega)$. By means of the above theorem we observe that the *weak* solution u of the elliptic boundary-value problem

$$\begin{aligned} &\text{(i)}\ A(x, \partial)u(x) = f(x), \qquad x \in \Omega, \\ &\text{(ii)}\ u|_{\partial\Omega} = 0, \end{aligned} \tag{1.3}$$

exists and is given as

$$u = A^{-1} f \in D_A \subset H_0^1(\Omega).$$

Therefore we call A the *weak representation* of the differential operator $A(x, \partial)$ imposed with homogeneous Dirichlet boundary conditions.

Since the weak representation operator A is positive selfadjoint the fractional powers A^α, $\alpha \geqslant 0$, exist and the corresponding domains of definition

$$\dot{H}^{2\alpha} = D_{A^\alpha}$$

equipped with the inner product

$$(u, v)_{\dot{H}^{2\alpha}} = (A^\alpha u, A^\alpha v) \qquad \text{for all } u, v \in \dot{H}^{2\alpha},$$

define a scale of Hilbert spaces for which the embeddings

$$\dot{H}^\alpha \hookrightarrow \dot{H}^\beta, \qquad \alpha > \beta,$$

are continuous. Hence Theorem 1.1 states that

$$D_A = \dot{H}^2 \hookrightarrow \dot{H}^1 = D_{A^{1/2}} = H_0^1(\Omega).$$

In some sense the space $\dot{H}^2$ fully describes the *regularity* of weak solutions of the problem (1.3) since

$$\|u\|_{\dot{H}^2} = \|f\|_0.$$

The term of $H^{1+s}(\Omega)$-regularity, $s \geqslant 0$, may now be expressed as the existence of a continuous embedding

$$\dot{H}^2 \hookrightarrow H^{1+s}(\Omega) \cap H_0^1(\Omega).$$

EXAMPLE 1.3. By making the weak assumptions

$$\Omega \in C^{0,1}, \tag{A2}$$

which states that Ω has Lipschitz boundary, and

$$a^{\rho\sigma} \in C^{0,t}(\bar{\Omega}) \qquad \text{for some } 0 < t \leqslant 1 \text{ whenever } |\rho| = 1, \tag{A3}$$

we gain the following regularity result due to Nečas [6]:

$$\dot{H}^2 \hookrightarrow H^{1+s}(\Omega) \cap H_0^1(\Omega) \qquad \text{for all } 0 \leqslant s < \min(t, \tfrac{1}{2}).$$

Imposing in addition

$$\Omega \text{ is convex}, \qquad t = 1, \tag{A4}$$

yields full regularity

$$\dot{H}^2 \hookrightarrow H^2(\Omega) \cap H_0^1(\Omega),$$

a result due to Kadlec [4].

With the help of the weak representation operator A we may restate our parabolic problem (1.1) as the following *abstract Cauchy problem* in $L^2(\Omega)$:

$$\begin{aligned} &\text{(i) } u' + Au = f, \\ &\text{(ii) } u(0) = u_0. \end{aligned} \tag{1.4}$$

If we denote the *holomorphic semigroup* of contractions generated by the negative selfadjoint operator $(-A)$ as

$$\mathcal{U}(t) = \exp(-tA),$$

the solution $u \in C^\infty(]0, T], \dot{H}^2)$ of (1.4) is given by

$$\begin{aligned} &\text{(i) } u(t) = [w - \mathcal{U}(t)w] + \mathcal{U}(t)u_0, \qquad \text{where} \\ &\text{(ii) } w = A^{-1}f \in \dot{H}^2. \end{aligned} \tag{1.5}$$

Exactly *this* solution will be approximated by our algorithm.

1.2. *The Algorithm*

As mentioned in the introduction and discussed in [2], the initial-value character of the abstract Cauchy problem requires discretization in time *first*.

The principle of a variable-step, variable-order discretization in time will be explained first assuming that the spatial subproblems can be solved exactly. Thereafter discretization in the spatial variables will be viewed as a perturbation of this semidiscrete orbit.

1.2.1. *Semidiscretization in Time*

As discussed in [2] a single step method

$$u_{j+1} = \Phi(u_j, \tau), \qquad j = 0, 1, \cdots$$

is applicable to the abstract Cauchy problem (1.4) whenever the corresponding rotational approximation $r_\Phi(z)$ to $\exp(-z)$ is A_0-*acceptable,* which means that

$$\begin{aligned} &\text{(i)} \ |r_\Phi(z)| < 1 \qquad \text{for } z > 0, \\ &\text{(ii)} \ |r_\Phi(\infty)| < 1. \end{aligned} \tag{1.6}$$

The rational approximation is said to be of *order* $p \geqslant 1$ whenever

$$r_\Phi(z) = e^{-z} + \mathcal{O}(z^{p+1}) \qquad \text{for } z \to 0.$$

THEOREM 1.4. *Given an* A_0*-acceptable rational approximation* $r(z)$ *to* $\exp(-z)$ *of order* p*, the single step method*

$$\Phi_r(u, \tau) = r(\tau A)u + (I - r(\tau A))A^{-1}f \tag{1.7}$$

is well defined for $\tau \geqslant 0$ *and the sequence* $u_{j+1} = \Phi(u_j, \tau)$, $j = 0, 1, \ldots$, *approximates the solution of the abstract Cauchy problem* (1.4) *at* $t = j\tau$ *with an error of*

$$\|u_j - u(t)\|_0 \leqslant C\tau^p t^{\min(1, \alpha - p)} \|u_0\|_{\dot{H}^{2\alpha}}. \tag{1.8}$$

Proof. This is the case $N = p - 1$ of Theorem 2.7 of [2]. ■

Consider now a sequence $r_j(z)$, $j = 1, 2, \ldots$, of A_0-acceptable rational approximations to $\exp(-z)$ of increasing order j together with the corresponding single step methods $\Phi_j = \Phi_{r_j}$. A variable-step variable-order method for the abstract Cauchy problem can be described as the following device:

Given an initial approximation $u^0 = \tilde{u}(t)$ at time t, a tolerance TOL, time step τ, and a suggested order k, the method computes the sequence

$$u^j = \Phi_j(u^0, \tau), \qquad j = 1, \ldots, k + 1,$$

which approximates with successively higher order the solution $\tilde{u}(t+\tau)$ of the Cauchy problem with initial data $\tilde{u}(t)$ at time τ.

As in [3] we get the error estimates

$$\epsilon_j = \|u^{j+1} - u^j\|_0 \doteq \|\tilde{u}(t+\tau) - u^j\|_0,$$

such that further the approximation u^{k+1} is accepted if

$$\epsilon_k < \mathsf{TOL}.$$

Comparison of the ϵ_j with the a priori estimate (1.8) gives new time steps

$$\tau_j = \sqrt[j+1]{\frac{\mathsf{TOL}}{\epsilon_j}}\,\tau \tag{1.9}$$

for the orders $j = 1, \ldots, k$. As new order k^* together with $\tau^* = \tau_{k^*}$ as new time step we take that order which minimizes the amount of work per unit step; i.e.,

$$\frac{A_{k^*+1}}{\tau^*} = \min_{1 \le j \le k} \frac{A_{j+1}}{\tau_j}. \tag{1.10}$$

Here A_j measures the amount of work for computing the sequence $u^1, \ldots, u^j$.

Repeated application of this procedure yields the approximate orbit in Hilbert space.

1.2.2. *Perturbations through Spatial Discretization*

Computation of $u^j = \Phi_j(\tilde{u}(t), \tau)$ requires the *weak* solution of several elliptic problems due to the denominator of the rational functions $r_j(z)$. In general we cannot get the exact functions u^j but *perturbed* functions

$$\hat{u}^j = u^j + \delta_j \qquad j = 1, \ldots, k+1,$$

with perturbations $\delta_j \in L^2(\Omega)$. The following requirements are reasonable:

Keep the perturbations δ_j below a certain level such that

- the approximation $\hat{u}^{k+1}$ is good enough with respect to TOL,
- the generated time-step sequence is nearly the same as in the case of *no* perturbations.

These requirements ensure that the problem dependent time-stepping in Hilbert space is preserved.

They can be met if we assume that we are able to compute time-error estimates

$$\hat{\epsilon}_j = \epsilon_j + \theta_j \qquad j = 1, \ldots, k,$$

as well as estimates $[\theta_j]$, $[\delta_j]$ of the spatial perturbations $|\theta_j|$, $\|\delta_j\|_0$.

We then proceed as follows: Compute time steps with respect to $\rho\mathsf{TOL}$ instead of TOL, where $0 < \rho < 1$. The approximation $\hat{u}^{k+1}$ is accepted if

$$\begin{aligned} &\text{(i)}\ \hat{\epsilon}_k + [\delta_{k+1}] < \mathsf{TOL}, \\ &\text{(ii)}\ [\theta_j] < \tfrac{1}{4}\hat{\epsilon}_j \qquad j = 1, \ldots, k. \end{aligned} \tag{1.11}$$

Implementing this computable control criterion (1.11) yields $\hat{u}^{k+1}$ accurate enough and time steps

$$\hat{\tau}_j = \sqrt[j+1]{\frac{\mathsf{TOL}}{\hat{\epsilon}_j}}\,\tau,$$

varying in comparison to the corresponding *exact* time steps τ_j as

$$\frac{1}{1.8}\tau_j \leqslant \hat{\tau}_j \leqslant 1.3\tau_j,$$

provided that $[\theta_j] \doteq |\theta_j|$, $[\delta_j] \doteq \|\delta_j\|_0$.

In order to make a passage through the criterion (1.11) possible we have to impose accuracies

$$\mathsf{eps}_j = \chi(j, k)(1 - \rho)\mathsf{TOL} \tag{1.12}$$

on the elliptic problems arising in the computation of u^j.

EXAMPLE 1.5. The extrapolated implicit Euler scheme yields, as shown in [2],

$$\chi(j, k) = \frac{1}{j}\,\alpha_j^{k+1},$$

with coefficients α_j^k quite *small* for higher k, for instance $\alpha_5^5 = 6.5_{10} - 3$. Thus extrapolation *amplifies* spatial perturbations. This amplification is due to the fact that we build higher and higher order *differences* whose perturbations stay in the order of magnitude of the initial perturbation—but do not decrease like the differences.

2. Variable-Order Time Discretization Based on a Multiplicative Error Correction

In this section we will derive the new time discretization and apply it to the abstract Cauchy problem.

2.1. *A Family of Rational Approximations to* $\exp(-z)$

The drawbacks, which have been mentioned in the introduction as well as in Example 1.5, of extrapolation methods or related methods like deferred corrections are a result of the fact that the error estimation is built as a *difference* of two approximations of different order,

$$\eta_j = u^{j+1} - u^j$$

—a fact which is very similar to the "cancellation effect."

In contrast, we are searching for a method which computes η_j *directly* in such a way that the higher order approximation is given as

$$u^{j+1} = u^j + \eta_j,$$

in order to avoid any cancellation. Thus the corresponding rational approximation $r_j(z)$ to e^{-z} can be written as

$$r_{j+1}(z) = r_j(z) + \rho_j(z).$$

We require several features for the rational functions $r_j(z)$ and $\rho_j(z)$:

(R1) r_j should be an L_0-acceptable approximation to e^{-z} of order $j > 0$.

(R2) The corrections ρ_{j+1} should be obtained *multiplicatively*,

$$\rho_{j+1}(z) = \gamma_{j+1}\rho(z)\rho_j(z), \qquad j = 1, 2, \ldots,$$

with a rational function ρ and coefficients γ_j.

L_0-acceptability denotes in addition to A_0-acceptability that

$$r_j(\infty) = 0.$$

Discussion of the Requirements. Requirement (R1) yields the existence and continuity of the mappings

$$r_j(\tau A)\colon \dot{H}^\alpha \to \dot{H}^{\alpha+2} \qquad \text{for } \alpha \geq 0.$$

Thus we have modeled the effect of *parabolic smoothing*. The *multiplicative error correction* (R2) is motivated by the aims of the least possible need of memory and the least possible effort of work, since it means that

1. We need only memorize the actual approximation $r_j(\tau A)$ and the last correction $\rho_{j-1}(\tau A)$ in order to get a new correction $\rho_j(\tau A)$ and thereafter a new approximation $r_{j+1}(\tau A)$;

2. We always have to perform the *same* type of elliptic problem, that is, the evaluation of $\rho(\tau A)$, in contrast to extrapolation methods which have to compute the *different* elliptic problems $(I + \tau/jA)^{-1}$ for varying j.

Derivation of the Approximations. Up to now we have constructed our rational approximations as

$$r_{j+1}(z) = r_1(z) + \sum_{k=0}^{j-1} \bar{\alpha}_k \rho^k(z)\rho_1(z), \qquad j = 1, 2, \ldots,$$

with $\bar{\alpha}_k = \prod_{l=1}^{k} \gamma_{l+1}$ for $k \geqslant 1$, $\bar{\alpha}_0 = 1$. By choosing

$$\rho_1(z) = \gamma_1 \rho^{\nu}(z) r_1(z),$$

where γ_1 and the integer $\nu \geqslant 1$ will be specified later on, and

$$\begin{aligned} &\text{(i)} \quad \alpha_{k+\nu} = \gamma_1 \bar{\alpha}_k, && k \geqslant 0, \\ &\text{(ii)} \quad \alpha_0 = 1, \\ &\text{(iii)} \quad \alpha_i = 0, && i = 1, \ldots, \nu - 1, \end{aligned} \tag{2.1}$$

we gain the relation

$$r_j(z) = r_1(z) \sum_{k=0}^{j-2+\nu} \alpha_k \rho^k(z), \qquad j = 1, 2, \ldots. \tag{2.2}$$

Thus the approximation r_j is a refinement of r_1. Since we want to refine the implicit Euler, which belongs to parabolic problems [1], we choose

$$r_1(z) = \frac{1}{1 + z}.$$

Now the rational function ρ should have the same denominator as r_1, which means that we have to solve the *same* elliptic problem as in $r_1(\tau A)$ in order to evaluate $\rho(\tau A)$. Thus we get

$$\rho(z) = \frac{\pi(z)}{1+z},$$

with $\pi(z)$ a polynomial in z. Because of $\alpha_0 = 1$, $r_j(0) = r_1(0) = 1$ we have $\pi(0) = 0$. Moreover the L_0-acceptability of the r_j yields

$$|\rho(z)| \leqslant M \qquad \text{for } z \to +\infty,$$

with a certain $M > 0$. Hence we get deg $\pi = 1$. Since we have not yet specified the coefficients $\{\alpha_k\}_{k\geqslant 0}$, we get

$$\rho(z) = \frac{z}{1+z}.$$

Our considerations have led so far to the following problem: Find coefficients $\{\alpha_k\}_{k\geqslant 0}$ such that

$$e^{-z} = \frac{1}{1+z} \sum_{k=0}^{\infty} \alpha_k \left(\frac{z}{1+z}\right)^k.$$

Upon introducing

$$w = \frac{z}{1+z} \tag{2.3}$$

we observe that the $\{\alpha_k\}_{k\geqslant 0}$ should be generated by the function

$$\frac{1}{1-w} \exp\left(\frac{w}{w-1}\right).$$

This function is intimately connected with the *Laguerre polynomials*, since

$$\frac{1}{1-w} \exp\left(\frac{xw}{w-1}\right) = \sum_{k=0}^{\infty} L_k(x) w^k, \qquad |w| < 1, \tag{2.4}$$

where $L_k(\cdot)$ denotes the Laguerre polynomial of degree k. Thus we have

$$e^{-z} = \frac{1}{1+z} \sum_{k=0}^{\infty} L_k(1) \left(\frac{z}{1+z}\right)^k, \qquad \Re z > -\tfrac{1}{2}, \tag{2.5}$$

and get

$$\alpha_k = L_k(1), \qquad k = 0, 1, \ldots .$$

Since $\alpha_1 = L_1(1) = 0$ and $\alpha_2 = L_2(1) = -\frac{1}{2}$ we obtain by (2.1) that $\nu = 2$ and $\gamma_1 = -\frac{1}{2}$. By (2.2) our rational approximation $r_j(z)$ is given as

$$r_j(z) = \frac{1}{1+z} \sum_{k=0}^{j} L_k(1) \left(\frac{z}{1+z} \right)^k . \tag{2.6}$$

To end up with a recurrence formula for the rational functions r_j in connection with requirement (R2), we trace the derivation backward:

$$\begin{aligned}
&\text{(i)} \quad r_1(z) = \frac{1}{1+z} \\
&\text{(ii)} \quad \rho_1(z) = -\frac{1}{2} \frac{z^2}{(1+z)^2} r_1(z) \\
&\text{(iii)} \quad r_{j+1}(z) = r_j(z) + \rho_j(z), \qquad j = 1, 2, \cdots \\
&\text{(iv)} \quad \rho_{j+1}(z) = \gamma_{j+1} \frac{z}{1+z} \rho_j(z), \qquad j = 1, 2, \cdots \\
&\text{(v)} \quad \gamma_{j+1} = \frac{L_{j+2}(1)}{L_{j+1}(1)}, \qquad j = 1, 2, \ldots .
\end{aligned} \tag{2.7}$$

LEMMA 2.1. *The rational approximation $r_j(z)$ to e^{-z} defined by (2.6) is of order at least j. Furthermore we have*

(a) *$r_j(z)$ is L_0-acceptable whenever $|L_k(1)| \leq 1$ for $0 \leq k \leq j$.*

(b) *$r_j(z)$ can be computed by means of the recurrence formula (2.7) whenever $L_k(1) \neq 0$ for $2 \leq k \leq j-1$.*

Proof. The first assertion can be seen by comparison of (2.6) with (2.5). Part (b) follows from (2.7(v)). The remaining part (a) can be proved as follows: By using the transformation (2.3) we have

$$\begin{aligned}
r_j(z) &= \frac{1}{1+z} \sum_{k=0}^{j} L_k(1) \left(\frac{z}{1+z} \right)^k \\
&= (1-w) \sum_{k=0}^{j} L_k(1) w^k .
\end{aligned}$$

The interval $z \in]0, \infty[$ is mapped to $w \in]0, 1[$ which yields

$$\begin{aligned} |r_j(z)| &\leq (1-w) \sum_{k=0}^{j} |L_k(1)| w^k \\ &\leq (1-w) \sum_{k=0}^{j} w^k \\ &< 1, \end{aligned}$$

whenever $z > 0$ and $|L_k(1)| \leq 1$ for $0 \leq k \leq j$. ■

Next we show that our approximations are in fact special cases of the so called *RD–Padé approximations* to e^{-z} introduced by Nørsett [7]. (RD = restricted denominator).

DEFINITION 2.2. A rational approximation to e^{-z} of the form

$$R_p^j(z) = \frac{\sum_{k=0}^{j} a_k z^k}{(1+\sigma z)^p}$$

of order at least $j \leq p$ is called a *(j, p)-RD-Padé approximation.*

LEMMA 2.3, (Nørsett [7]). *The (j, p)-RD-Padé approximation is uniquely given by*

$$R_p^j(z) = \frac{\sum_{k=0}^{j} (-1)^{p+k} L_p^{(p-k)}(1/\sigma)(\sigma z)^k}{(1+\sigma z)^p}.$$

Proof. Corollary 2.1 of [7]. ■

COROLLARY 2.4. *The approximation $r_j(z)$ is the $(j, j+1)$-RD-Padé approximation $R_{j+1}^j(z)$ with $\sigma = 1$ and given as*

$$r_j(z) = \frac{\sum_{k=0}^{j} (-1)^{j+1-k} L_{j+1}^{(j+1-k)}(1) z^k}{(1+z)^{j+1}}. \tag{2.8}$$

Proof. Lemma 2.1 states that the rational function $r_j(z)$ is of order j at least. Furthermore it is clearly of the same form as $R_{j+1}^j(z)$ with $\sigma = 1$, which is by Lemma 2.3 uniquely given as (2.8). ■

Remark 2.5. A direct proof of Eq. (2.8) is possible by using (2.6) and the relation

$$\sum_{k=0}^{j} \binom{p-k-1}{j-k} L_k(x) = (-1)^{p+j} L_p^{(p-j)}(x),$$

which can be obtained by differentiating $p-j$ times with respect to x in the generating function formula (2.4).

The characterization of $r_j(z)$ as an RD-Padé approximation yields some consequences which are listed below.

COROLLARY 2.6. *The error of $r_j(z)$ as an approximation to e^{-z} is explicitly given as*

$$r_j(z) - e^{-z} = \frac{z^{j+1}}{(1+z)^{j+1}} e^{-z} \left(z \int_0^1 e^{zt} L_{j+1}(t) dt - e^z L_{j+1}(1) \right). \quad (2.9)$$

Proof. The assertion follows from Corollary 2.4 together with Theorem 4.2 of [7]. ■

COROLLARY 2.7. *The order of $r_j(z)$ is j if $L_{j+1}(1) \neq 0$, otherwise the order is $j+1$.*

Proof. Follows directly from Corollary 2.6. ■

We have seen in Lemma 2.1 and Corollary 2.7 that many of the properties of the approximations $r_j(z)$ depend on the values of the Laguerre polynomials $L_k(x)$ at $x = 1$. Investigation of these values yields our main result:

THEOREM 2.8. *The rational functions $r_j(z)$ given by (2.6) are L_0-acceptable approximations to e^{-z} of order j* at least *for $1 \leqslant j \leqslant 100$. For these j the functions $r_j(z)$ can be computed by means of the recurrence (2.7).*

Proof. Examination of the values of $L_j(1)$ for $1 \leqslant j \leqslant 101$ shows that the assumptions of Lemma 2.1 and Corollary 2.7 are valid for those j. This can be done for instance by means of the formula-manipulating language REDUCE. ■

Remark 2.9. The author conjectures that Theorem 2.8 is valid for all $j \geqslant 1$, but has found no better estimate than $|L_j(1)| < \sqrt{e}$ for all $j \geqslant 1$ in the literature.

Remark 2.10. In his paper [7], Nørsett is mainly interested in optimizing the order of RD-Padé approximations by choosing special values of σ. However, these values of σ *depend* on the chosen order, thus making a recurrence like (2.7) impossible. Because (2.7) is *essential* for a cheap realization of $r_j(\tau A)$ with the possibility of *varying* order, we have chosen $\sigma = 1$.

We close by listing the first coefficients γ_j of the recurrence (2.7) in Table I.

TABLE I
THE COEFFICIENTS γ_j FOR $j = 2, \ldots, 9$

j	Numerator of γ_j	Denominator of γ_j
2	4	3
3	15	16
4	56	75
5	185	336
6	204	1295
7	−6209	1632
8	112400	55881
9	1520271	1124000

Note added in proof. The approximations $r_j(z)$, $j = 1, \ldots, 10$ are $A(\vartheta)$-stable in the sense of [2, Sec. 2.2] with an angle $\vartheta \geqslant 88.9°$, thus making the approximations applicable to a wide range of problems.

2.2. *The Variable-Order Single Step Method in Hilbert Space*

Here we will explain the single step methods corresponding to the just derived family of rational approximations. Given

$$u^0 = \tilde{u}(t)$$

and a time step $\tau \geqslant 0$ the recurrence (2.7) yields

$$\begin{aligned}
&\text{(i)}\ u^1 = r_1(\tau A)u^0 + (I - r_1(\tau A))A^{-1}f \\
&\text{(ii)}\ \eta_1 = -\tfrac{1}{2}(\tau A(I + \tau A)^{-1})^2(u^1 - A^{-1}f) \\
&\text{(iii)}\ u^{j+1} = u^j + \eta_j \qquad\qquad j = 1, 2, \cdots \\
&\text{(iv)}\ \eta_{j+1} = \gamma_{j+1}\tau A(I + \tau A)^{-1}\eta_j \qquad j = 1, 2, \cdots
\end{aligned} \tag{2.10}$$

if we remember the construction (1.7) of single step methods from the rational function. The *update relation* (iv) specifies the meaning of what we called a direct computation of the error corrections η_j.

If we make use of the relation

$$I - (I + \tau A)^{-1} = \tau A(I + \tau A)^{-1},$$

we are able to find a simpler expression for the terms u^1, η_1:

$$\begin{aligned}
&\text{(i)}\ u^1 = (I + \tau A)^{-1}(u^0 + \tau f) \\
&\text{(ii)}\ \eta_0 = u^1 - u^0 \\
&\text{(iii)}\ \eta_1 = \tfrac{1}{2}\tau A(I + \tau A)^{-2}\eta_0.
\end{aligned} \tag{2.11}$$

Remark 2.11. Another version of representing u^1, η_0 would be

$$\text{(i)}\ \eta_0 = \tau(I + \tau A)^{-1}(f - Au^0)$$
$$\text{(ii)}\ u^1 = u^0 + \eta_0,$$

which puts the difference at a more desirable place. However, this is *only* possible if $u^0 \in \dot{H}^2$.

By means of the representation (2.10) we observe that for

$$u^0, f \in L^2(\Omega)$$

the approximation and corrections possess the necessary regularity:

$$u^j, \eta_j \in \dot{H}^2 \qquad \text{for } j \geqslant 1.$$

Since A is the weak representation of the elliptic operator $A(x, \partial)$, problems of the kind

$$u = (I + \tau A)^{-1} w, \qquad w \in L^2(\Omega),$$

are *equivalent* to the variational problem

$$(u, v) + \tau a(u, v) = (w, v) \qquad \text{for all } v \in H_0^1(\Omega);$$

whereas problems of the kind

$$\eta = \tau A(I + \tau A)^{-1}\zeta, \qquad \zeta \in \dot{H}^2,$$

are equivalent to the variational problem

$$(\eta, v) + \tau a(\eta, v) = \tau a(\zeta, v) \qquad \text{for all } v \in H_0^1(\Omega).$$

The equivalence is backed by Theorem 1.1.

Finally the time-error estimators are

$$\epsilon_j = \|\eta_j\|_0 \qquad \text{for } j \geqslant 1. \tag{2.12}$$

3. The Matching of Spatial Errors and Algorithmic Details

In this section we will specify the perturbation concept introduced in the preliminary draft of Section 1. For that purpose we derive expressions for the

estimators $[\theta_j]$, $[\delta_j]$ as well as for the accuracy function χ. Finally we discuss some algorithmic details.

3.1. *The Perturbation Estimators* $[\theta_j]$, $[\delta_j]$

In order to realize (2.10) we have to approximate the arising (weak) elliptic problems. Since we have seen that they are equivalent to the variational forms, an adaptive FEM method is ideally suited for our purposes. However, it has to fulfill certain requirements already discussed in [2, Sect. 4]. For the following the main point of interest is that the elliptic solver may solve within a given accuracy eps and delivers an error estimate. Then we can proceed as follows:

By using the elliptic solver within the given accuracy eps we first get an approximation of u^1

$$\hat{u}^1 = u^1 + \delta_1 \in L^2(\Omega),$$

together with an estimate $[\delta_1] \leqslant \text{eps}$ of $\|\delta_1\|_0$.

Next we *fix* the triangulation chosen by the elliptic solver in order to compute $\hat{u}^1$ and compute

$$\begin{aligned}\hat{\eta}_1 &= \tfrac{1}{2}\tau A(I+\tau A)^{-2}(\hat{u}^1 - u^0) + \hat{\omega}_1 + \tfrac{1}{2}\tau A(I+\tau A)^{-1}\hat{\omega}_0 \\ &= \eta_1 + \omega_1\end{aligned}$$

on that triangulation. Here $\hat{\omega}_0$ is the error of the approximation $\tilde{\eta}_1$ of

$$(I+\tau A)^{-1}(\hat{u}^1 - u^0),$$

and $\hat{\omega}_1$ denotes the error made while solving the second elliptic problem

$$\tfrac{1}{2}\tau A(I+\tau A)^{-1}\tilde{\eta}_1.$$

Hence the elliptic solver provides estimates $[\hat{\omega}_0]$, $[\hat{\omega}_1]$ of the norms of the corresponding errors. We gain the representation

$$\omega_1 = \tfrac{1}{2}\tau A(I+\tau A)^{-2}\delta_1 + \tfrac{1}{2}\tau A(I+\tau A)^{-1}\hat{\omega}_0 + \hat{\omega}_1$$

and can derive, by using the important estimates

$$\|\tau A(I+\tau A)^{-1}\|, \|(I+\tau A)^{-1}\| \leqslant 1,$$

the estimator

$$[\theta_1] = \tfrac{1}{2}([\delta_1] + [\hat{\omega}_0]) + [\hat{\omega}_1]. \tag{3.1}$$

By successively computing

$$\begin{aligned}\hat{\eta}_{j+1} &= \gamma_{j+1}\tau A(I + \tau A)^{-1}\hat{\eta}_j + \hat{\omega}_{j+1} \\ &= \eta_{j+1} + \omega_{j+1},\end{aligned}$$

where $\hat{\omega}_{j+1}$ denotes the error made by the elliptic solver, we get

$$\omega_{j+1} = \gamma_{j+1}\tau A(I + \tau A)^{-1}\omega_j + \hat{\omega}_{j+1}.$$

This yields the estimator

$$[\theta_{j+1}] = |\gamma_{j+1}|[\theta_j] + [\hat{\omega}_{j+1}]. \tag{3.2}$$

Herein $[\hat{\omega}_{j+1}]$ denotes the error estimate given by the elliptic solver. The spatial perturbations of the correction η_j give rise to perturbations of the approximations u^{j+1} since we compute

$$\begin{aligned}\hat{u}^{j+1} &= \hat{u}^j + \hat{\eta}_j \\ &= u^{j+1} + \delta_{j+1}.\end{aligned}$$

Thus we end up with the successive estimates

$$[\delta_{j+1}] = [\delta_j] + [\theta_j]. \tag{3.3}$$

Together with the computationally available time-error estimates

$$\hat{\epsilon}_j = \|\hat{\eta}_j\|_0, \tag{3.4}$$

we have described the whole family of required estimators.

3.2. *The Accuracy Function* χ

We have to determine the accuracy eps, in order to make a passage through the criterion (1.11) possible. As shown in Section 1 this determination will result in a relation between the accuracy eps and the parabolic accuracy TOL of the form (1.12).

To this end we observe that the effect of the perturbation δ_1 will dominate the effects due to the perturbations $\hat{\omega}_j$ in general. *This observation can be*

backed by a careful frequency analysis of some model problems and is the reason we fix the triangulation after the computation of $\hat{u}^1$. Hence it is reasonable to determine eps by the following procedure:

Set $[\delta_1] = \mathsf{eps}_{k+1}$ and $[\hat{\omega}_j] = 0$ for $j = 1, 2, \ldots$. Compute eps_{k+1} in such a way that

$$(1 - \rho)\mathsf{TOL} = [\delta_{k+1}].$$

By using the recurrences (3.2) and (3.3) and the explicit formula for the coefficients γ_j (2.7(v)) we get

$$\begin{aligned}[\delta_{k+1}] &= (1 + \sum_{j=1}^{k} \prod_{i=1}^{j} |\gamma_i|)\mathsf{eps}_{k+1} \\ &= \sum_{j=0}^{k+1} |L_k(1)| \cdot \mathsf{eps}_{k+1}.\end{aligned}$$

We obtain the same result by using the representation (2.6) of $r_{k+1}(z)$, if we assume that the error δ_1 is due to the term ahead of the sum.

Hence in order to compute the sequence $\hat{u}^1, \ldots, \hat{u}^{k+1}$ for some $k \geq 1$ we have to impose the elliptic accuracy given by

$$\mathsf{eps} = (1 - \rho)\chi(k)\mathsf{TOL} \tag{3.5}$$

with

$$\chi(k) = (\sum_{j=0}^{k+1} |L_k(1)|)^{-1}. \tag{3.6}$$

We can estimate this factor from below a priori if we remember that $L_0(1) = 1$, $L_1(1) = 0$, and $|L_j(1)| \leq 1$ *at least* for $2 \leq j \leq 101$ as stated in the proof of Theorem 2.8:

$$\chi(k) \geq \frac{1}{k+1}, \qquad \text{at least for } 1 \leq k \leq 100.$$

This result is highly satisfactory if we compare it with the function χ obtained for extrapolation methods. The relevant first values of $\chi(k)$ are even more satisfying, as shown in Table II.

TABLE II
THE COEFFICIENTS $\chi(k)^{-1}$ FOR $k = 1, \ldots, 9$

k	1	2	3	4	5	6	7	8	9
$\chi(k)^{-1}$	1.5	2.2	2.8	3.3	3.5	3.6	3.7	4.0	4.4

Remark 3.1. The question of the "correct" value for the elliptic accuracy eps is *not* a question of *guaranteeing* the pass through the control criterion (1.11) for *all* possible situations, which yields far too *pessimistic* values and in turn much more effort than needed. However, it is a question of making the pass through *possible* for a large class of *realistic*, i.e. quite probable, situations. This yields more *optimistic* accuracies, and it is intended by some heuristic considerations as well as experience that it is not *too optimistic*. Such unreasonably optimistic accuracies would cause too many time-step and order suggestions to be withdrawn, which in turn leads to more work than needed. Looking at the elliptic accuracy (3.5) and the assumptions leading to it, we should bear that balance in mind.

3.3. *Algorithmic Details*

3.3.1. *Information Theoretic Standard Model*

As discussed in Deuflhard [3] for ODEs and by the author in [1] for parabolic equations, time-step and order control along the draft of Section 1 becomes a reliable device if we supplement it by an *information theoretic standard model* as introduced in [3]. By comparing the computed time-error estimates $\hat{\epsilon}_j$ with the standard behavior predicted by that model we can implement three devices:

- convergence monitor
- order window
- device for possible increase of order greater than the computed k.

For the meaning of these terms we refer to [3].

In [1, 3] the information theoretic standard model is derived for extrapolation methods, but it needs only little change for our new time-discretization: Just replace the coefficients $\alpha(k, q)$ of formula (3.8) in [3] by

$$\alpha(k, q) = \sqrt[k+1]{\frac{(\rho \mathrm{TOL})^{(k+2)/(q+2)}}{\rho \mathrm{TOL}}}. \tag{3.7}$$

3.3.2. *Consistency Estimator*

To avoid step-size reductions in transient phases we can use a consistency estimator as introduced by the author in [2]. We estimate the maximal value of α, such that

$$u^0 \in \dot{H}^{2\alpha}.$$

In view of Theorem 1.4 we can use exactly the same consistency estimator as in [2].

3.3.3. *Optimal Choice of the Factor* ρ

We want to optimize the factor ρ with respect to the expected work. This can be done at least locally in the time direction as follows: The local amount of work to realize our algorithm depends on ρ roughly by the factor

$$\frac{1}{(1-\rho)^{d/2}\sqrt{\rho}},$$

where d denotes the dimension of the spatial part. This factor may be obtained if we assume the lowest order in time and if we model the work, that the elliptic solver needs to achieve a certain accuracy in the L^2-norm, as in the case of quasi-uniform triangulations. Minimizing that factor gives the optimal value

$$\rho_d = \frac{1}{d+1}. \tag{3.8}$$

3.3.4. *Details for the* 1D *Case*

We use the same elliptic solver as explained in [2, Sect. 4]. The measures for the amount of work as introduced in Section 1.2 should be chosen as

$$A_j = \frac{j+3}{\sqrt{\chi(j-1)}} \qquad j = 2, 3, \ldots. \tag{3.9}$$

Herein we assumed that creating the final mesh and solving for $\hat{u}^0$ is twice as expensive as the computation of one correction $\hat{\eta}_j$. Furthermore the amount of work principle stated in [2] has been used.

Knowing the amount of work in advance we are able to study the order control *qualitatively* in dependence on the imposed accuracy TOL—by using the information theoretic standard model. This study shows that the minimal

value of (1.10), that determines the optimal order, lies between neighbors which are nearly of the same size. In order to avoid a nasty oscillation between such neighboring orders we require that

$$\frac{A_{k+2}}{\tau_{k+1}} \leqslant \sigma \frac{A_{k+1}}{\tau_k}$$

before taking the order $k + 1$ into account. The value

$$\sigma = 0.9$$

has turned out to be a good choice. Making this choice we gain the following nice result: The maximal possible order suggestion which we expect then is

$$k_{\max} = \lfloor 1 - \log_{10} \mathsf{TOL} \rfloor, \tag{3.10}$$

at least for tolerances $\mathsf{TOL} \in [10^{-6}, 10^{-1}]$. The numerical examples of the next section confirm this a priori result.

4. Numerical Examples in One Space Dimension

In this section we will demonstrate the efficiency of the new time discretization by means of two 1D examples. The new algorithm is implemented in the program KASTIO1, where the number 1 indicates the space dimension. It uses the same elliptic solver as the program KASTIX1 of the author [2], which is a realization of the extrapolated implicit Euler scheme as discretization in time. The advantages of the new discretization are enlightened by comparison of the behavior of KASTIO1 and KASTIX1.

Notation. In the tables of this section we make use of the following—besides the notation introduced earlier in this paper:

Max. order k: During a run the program has computed a sequence $u_1, \ldots, u_{k+1}$ of approximations *at least for one time layer.* Thus the maximal given order of approximation is $k + 1$ whereas the maximal order, for which an error estimation has been performed, is k.

$[N] = (\sum_{j \geqslant 1}$ No. of nodal points of time step $j)$/No. of time steps,

CPU = computing time in seconds on a SPARC-station1+,

N_{tot} = total No. of points over all time layers divided by 1000,

$$\kappa = \text{CPU}/N_{\text{tot}}.$$

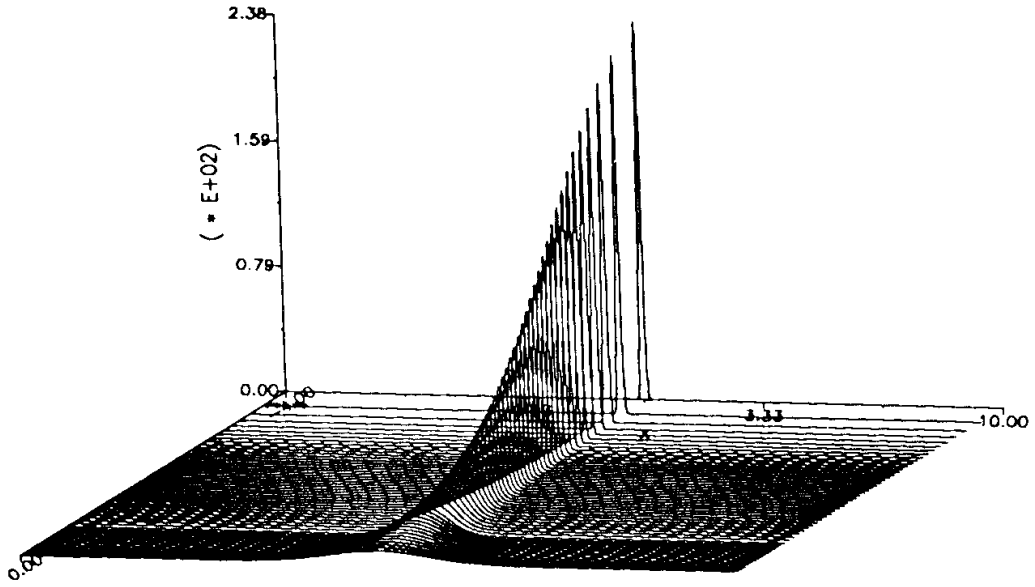

FIG. 1. Evolution of point-source, time in log-scale (Example 1).

For the meaning of the mean value $[N]$ see [2]. Since it indicates the effort for every nodal point, κ is something like a *complexity index*.

EXAMPLE 4.1. *Point-source.* This model problem has been proposed by the author in [1] to test the time-stepping procedure. We solve the homogeneous heat equation on the spatial interval $I = [-10, 10]$ with the following approximate δ-function as initial data:

$$u_0(x) = 250 \exp(-250x^2).$$

The Dirichlet boundary conditions can be considered as zero for $t \leqslant 1$ to model the evolution of u_0 on the whole real axis; the solution computed by KASTIO1 can be seen in Fig. 1.

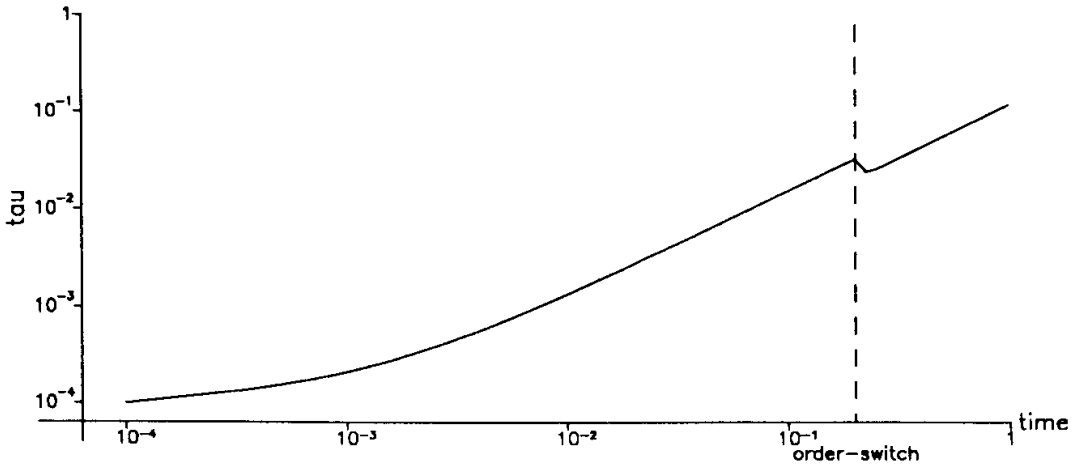

FIG. 2. Automatic increase of the time step (Example 1).

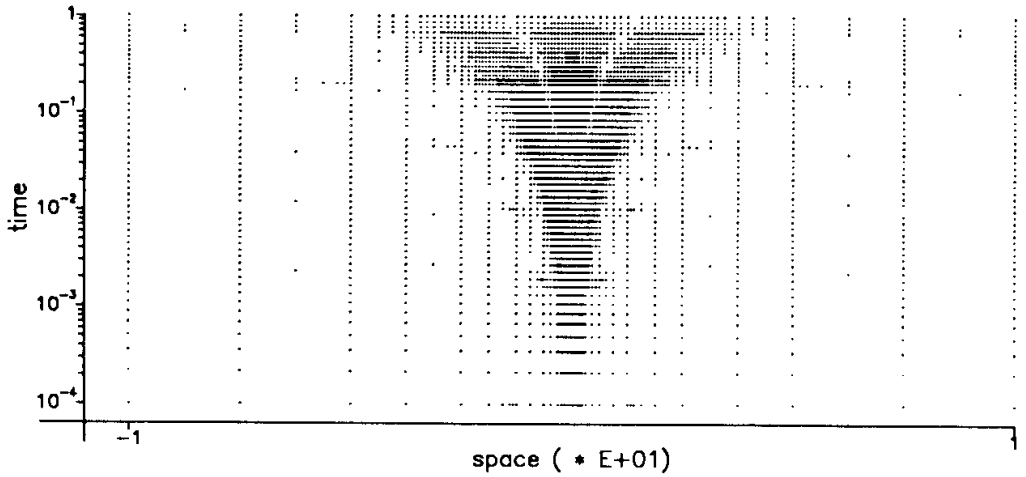

FIG. 3. Mesh development for the point-source (Example 1).

Because of the exponential decay of the solution as shown in Fig. 1 we expect an increase of the time step according to a power law, which really occurs automatically in the performance of KASTIO1 as shown in Fig. 2; the corresponding development of the space mesh is shown in Fig. 3.

Comparison of Table III with Table IV clearly shows the first drawback of extrapolation, mentioned in the introduction: Increasing cost with increasing order. KASTIX1 needs more accurate TOL to increase the order, and drops that order more quickly than KASTIO1. The slow increase of time steps as shown in Table III for KASTIO1 means that the new time discretization is able to use the higher orders quite long—a feature which had to be expected in view of the low cost of the higher orders. The maximal orders occurring in the runs of KASTIO1 nicely confirm the theoretical prediction (3.10) of Section 3. Moreover the complexity index κ is nearly constant for KASTIO1; thus we can speak of *multigrid complexity* of that program.

Fig. 4 shows that the error estimation of KASTIO1 is very reliable. In the run the chosen order is 2 for $t < 0.2$ and 1 for $t \geq 0.2$. The jump of the error

TABLE III
NEW (KASTIO1): PERFORMANCE FOR VARIABLE ORDER (EXAMPLE 1)

TOL	Time steps	Max. order	[N]	$L^\infty([0, T], L^2(I))$ norm of true error	CPU	N_{tot}	κ
10^{-1} [a]	55	2	147	$4.90_{10} - 2$	13	8	1.6
10^{-2}	66	3	513	$2.73_{10} - 3$	61	34	1.8
10^{-3}	79	4	1748	$1.67_{10} - 4$	289	138	2.1
10^{-4}	87	5	5632	$8.87_{10} - 6$	1145	490	2.3

[a] Run represented in Figs. 1–3.

TABLE IV
OLD (KASTIX1): PERFORMANCE FOR VARIABLE ORDER (EXAMPLE 1)

TOL	Time steps	Max. order	[N]	$L^\infty([0, T], L^2(I))$ norm of true error	CPU	N_{tot}	κ
10^{-1}	55	1	186	$3.96_{10} - 2$	28	10	2.8
10^{-2}	118	1	634	$4.76_{10} - 3$	286	75	3.8
10^{-3}	99	2	3758	$4.36_{10} - 4$	2115	372	5.7
10^{-4} [a]	—	—	—	—	—	—	—

[a] Run exceeds storage capabilities of the workstation used.

at this switching time nicely reflects the whole behavior of time-step and order control. Moreover it shows the quality of the error prediction for the next step, since the estimated error is just below the given tolerance.

Figure 5 shows the error estimators in more detail:

$$\hat{\epsilon}_j = \text{TIME},$$

$$[\delta_{j+1}] = \text{STAT-TOT},$$

$$[\delta_1] = \text{STAT1}.$$

Here j denotes the actually chosen order. As expected we observe that the jump of the estimated error at $t = 0.2$ is due to $\hat{\epsilon}_j$. As long as the order remains constant the time-stepping procedure leaves the time-error component of the whole error nearly constant. The equal shape of the behavior of $[\delta_1]$ and $[\delta_{j+1}]$ backs the assertion that δ_1 dominates all other spatial perturbations—a feature detailly discussed in Section 3.2. This feature is shown more quantitatively in Fig. 6. Herein the quotient $[\delta_{j+1}]/[\delta_1]$ is shown together with the corre-

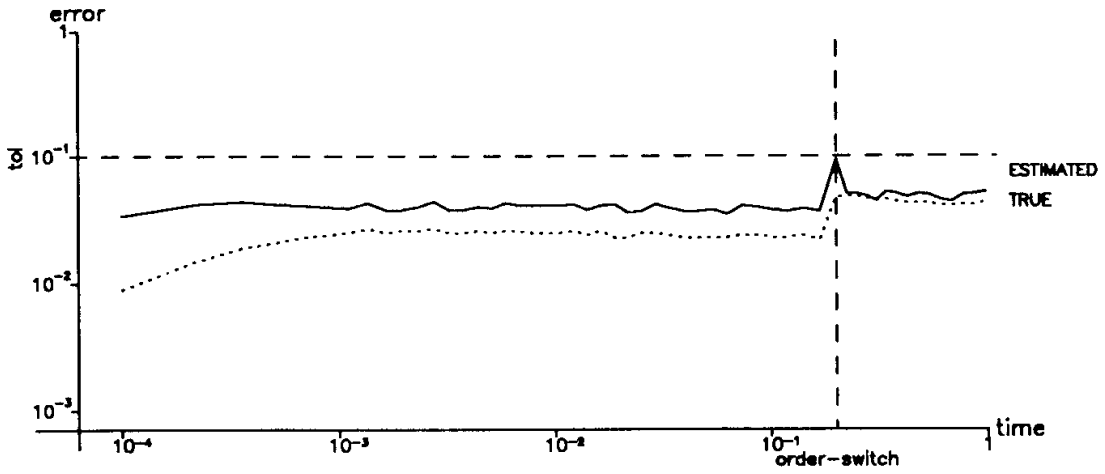

FIG. 4. Estimated vs. true error; KASTIO1 for TOL = 10^{-1} (Example 1).

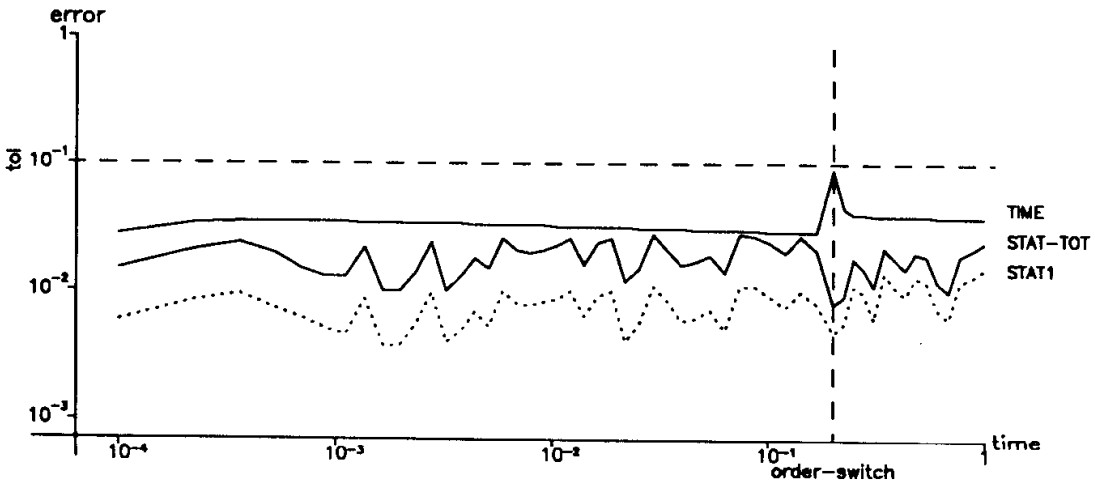

FIG. 5. The estimators $\hat{\epsilon}_j$, $[\delta_1]$ and $[\delta_{j+1}]$; KASTIO1 and TOL = 10^{-1} (Ex. 1).

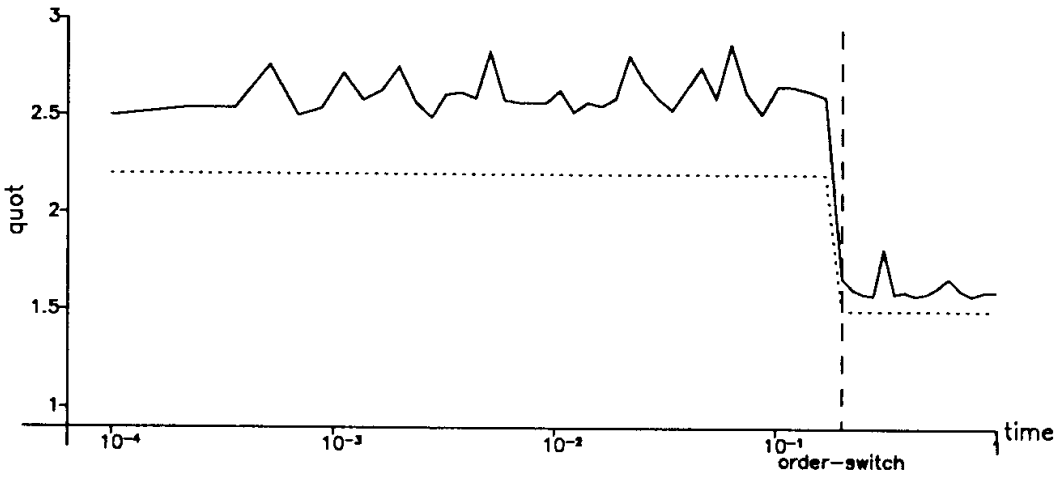

FIG. 6. The quotient $[\delta_{j+1}]/[\delta_1]$; KASTIO1 for TOL = 10^{-1} (Example 1).

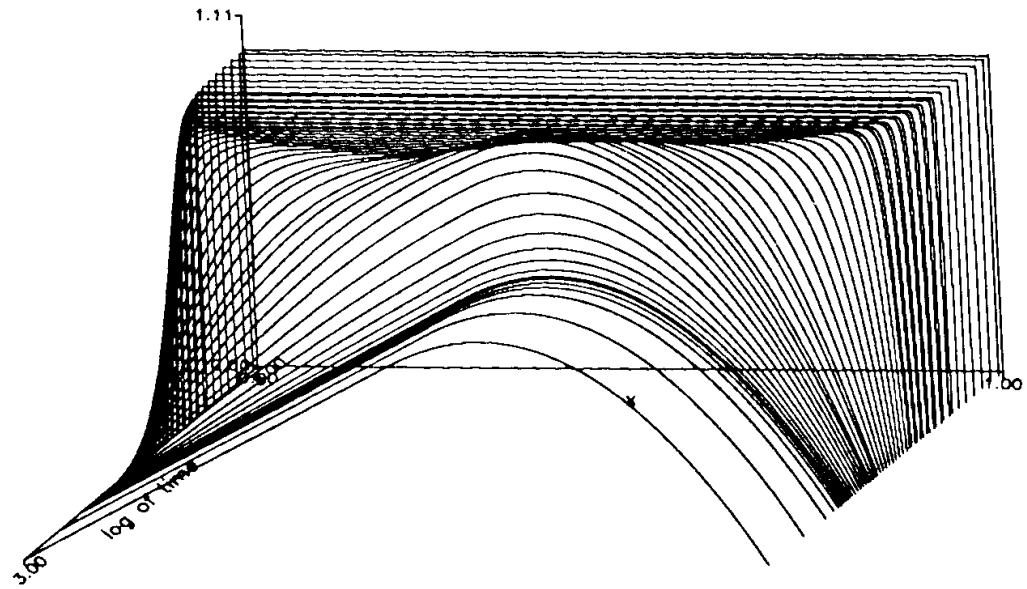

FIG. 7. Evolution of a boundary-inconsistency into a stationary solution, time in log-scale (Example 2).

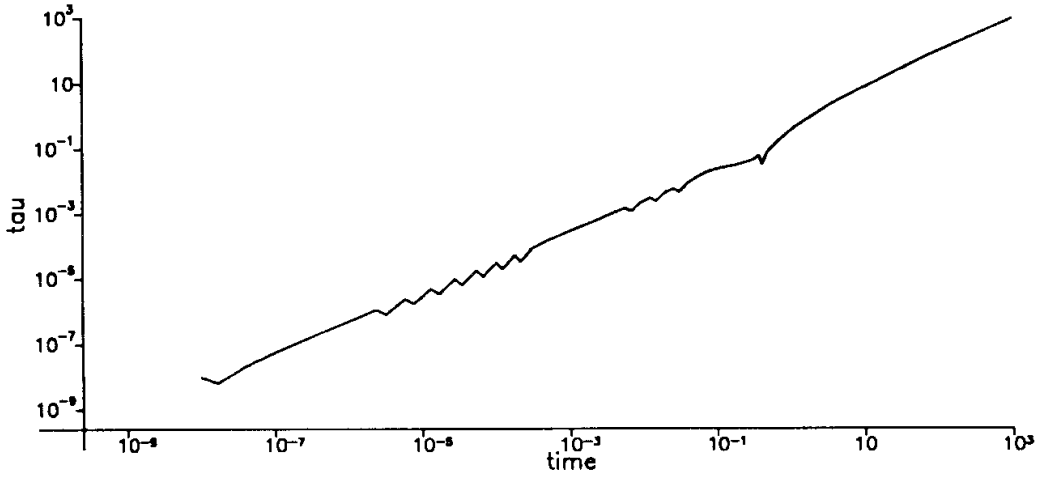

FIG. 8. Automatic increase of the time step (Example 2).

sponding propagation function $\chi(j)^{-1}$ (dotted line). It shows that our model of Section 3.2 slightly underestimates the error propagation.

EXAMPLE 4.2. *Inconsistent initial data.* This example is very challenging for the order and time-step control mechanism because of its transient phase. Moreover the solution runs into a stationary one. Thus we are able to study the third drawback of the extrapolation method KASTIX1 as mentioned in the introduction: KASTIX1 is not able to detect stationary phases.

The problem consists of the simple heat equation on the spatial interval $I = [0, 1]$ with a simple time independent source term. We impose homogeneous Dirichlet boundary conditions and choose

$$u_0 \equiv 1.$$

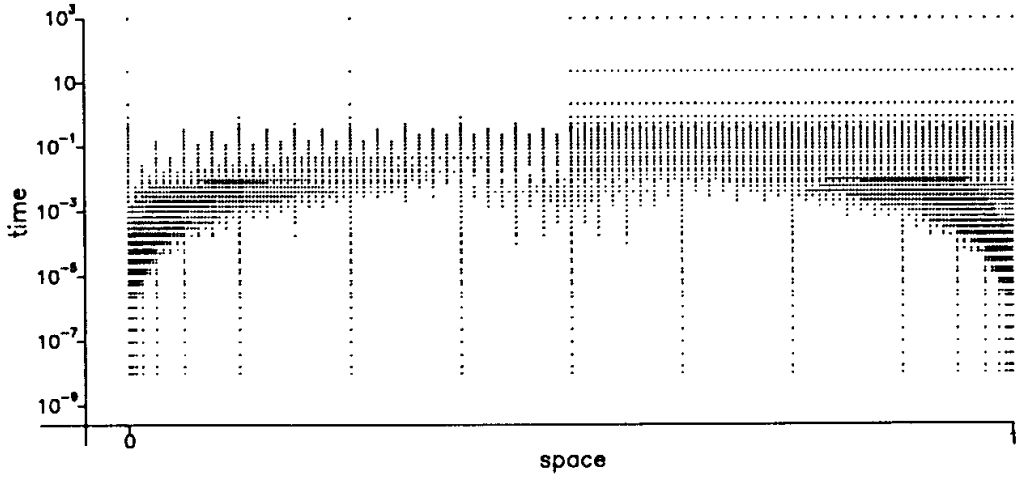

FIG. 9. Mesh development for the boundary-inconsistency (Example 2).

TABLE V
NEW (KASTIO1): PERFORMANCE FOR VARIABLE ORDER (EXAMPLE 2)

TOL	Time steps	Max. order	[N]	$L^\infty([0, T], L^2(I))$ norm of est. error	CPU	N_{tot}	κ
10^{-1}	18	1	13	$2.49_{10}-2$	0.4	0.2	2.1
10^{-2}	25	2	31	$9.21_{10}-3$	1.1	0.8	1.4
10^{-3} [a]	52	3	89	$9.88_{10}-4$	5.9	4.6	1.3
10^{-4}	83	4	282	$9.93_{10}-5$	35.6	23.4	1.5
10^{-5}	79	6	1002	$9.81_{10}-6$	147.3	79.1	1.9

[a] Run represented in Figs. 7–10.

Because u_0 does not satisfy the Dirichlet condition, the initial data are inconsistent. The source is chosen in order to get a stationary solution which is linear in [0, 0.5] and a parabola in [0.5, 1]. The solution computed by KASTIO1 can be seen in Fig. 7.

Again we expect an increase of the time step according to a power law, which really occurs automatically in the performance of KASTIO1 as shown in Fig. 8. The corresponding development of the space mesh is shown in Fig. 9.

Comparison of Tables V and VI shows that KASTIX1 chooses lower orders than KASTIO1 as in Example 1 and needs far more time steps. The latter observation can be explained by the above mentioned third drawback since the solution becomes stationary roughly at $t = 1$. For all tolerances KASTIO1 needs only 3 time steps to come from $t = 1$ to $t = 1000$, whereas KASTIX1 spends about 35 time steps for the same task (TOL = 10^{-1}, 10^{-2}). Moreover the need of computing time and of storage is much higher in the case of the extrapolation method than in the case of the new time discretization. Finally the complexity index κ shows for KASTIO1 *multigrid complexity* in contrast to KASTIX1.

TABLE VI
OLD (KASTIX1): PERFORMANCE FOR VARIABLE ORDER (EXAMPLE 2)

TOL	Time steps	Max. order	[N]	$L^\infty([0, T], L^2(I))$ norm of est. error	CPU	N_{tot}	κ
10^{-1}	141	1	15	$8.03_{10}-2$	4.2	2.1	2.0
10^{-2}	131	1	33	$6.18_{10}-3$	8.9	4.2	2.1
10^{-3}	55	2	169	$9.85_{10}-4$	33.6	9.3	3.6
10^{-4}	164	2	483	$9.71_{10}-5$	524.1	79.0	6.6
10^{-5} [a]	—	—	—	—	—	—	—

[a] Run exceeds storage capabilities of the workstation used.

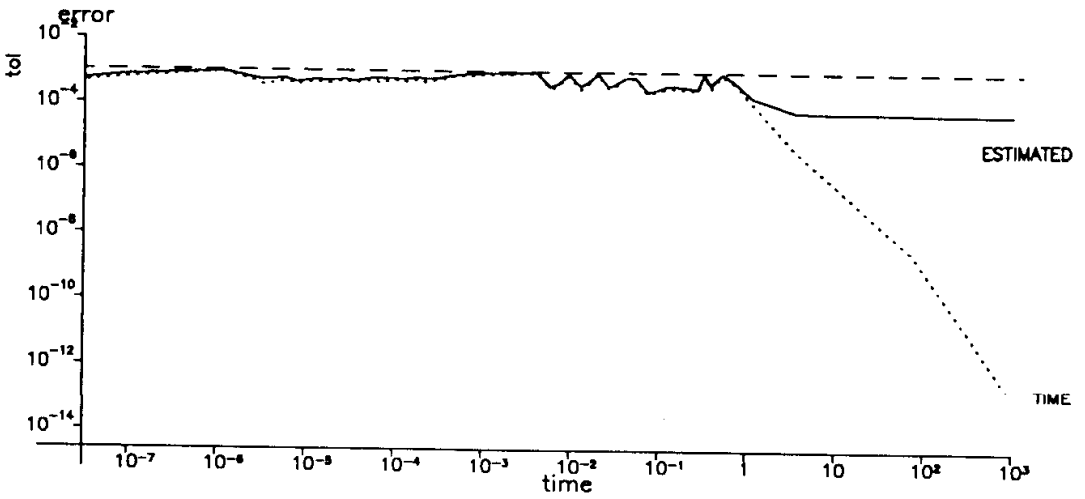

FIG. 10. Error behavior of KASTIO1 for TOL = 10^{-3} (Example 2).

Figure 10, where the estimated time-error component is plotted in addition to the estimated total error, nicely shows how KASTIO1 is able to detect stationary phases.

ACKNOWLEDGMENTS

The author thanks P. Deuflhard and M. Wulkow for helpful discussions.

REFERENCES

1. F. A. Bornemann, "Adaptive Multilevel Discretization in Time and Space for Parabolic Partial Differential Equations," Technical Report TR 89-7, Konrad-Zuse-Zentrum, Berlin (1989).
2. F. A. Bornemann, An adaptive multilevel approach to parabolic equations. I. General theory and 1D-implementation. *IMPACT Comput. Sci. Engrg.* **2**, 279–317 (1990).
3. P. Deuflhard, Order and stepsize control in extrapolation methods, *Numer. Math.* **41**, 399–422 (1983).
4. J. Kadlec, O reguljarnosti rešenija sadači Puassona na oblasti s granicej, lokol'no podobnoj granice vypukloj oblasti. *Czechoslovak Math. J.* **14**, 386–393 (1964).
5. T. Kato, "Perturbation Theory for Linear Operators." Second corrected printing of the second edition. Springer-Verlag, Berlin/Heidelberg/New York (1984).
6. J. Nečas, Sur la coercivité des formes sesqui-linéaires elliptiques, *Rev. Roumaine Math. Pures Appl.* **9**, 47–69 (1964).
7. S. P. Nørsett, Restricted Padé approximations to the exponential function. *SIAM J. Numer. Anal.* **15**, 1008–1029 (1978).