stichting
mathematisch
centrum

MC

AFDELING NUMERIEKE WISKUNDE NW 4/73 MAY

E. SLAGT

SOME APPLICATIONS OF STABILIZED RUNGE-KUTTA METHODS FOR HYPERBOLIC DIFFERENTIAL EQUATIONS

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

AMS (MOS) subject classification scheme (1970): 35F10, 35L45,65M20

Some applications of stabilized Runge-Kutta methods for hyperbolic differential equations

by

E. Slagt

Abstract

In this report two applications are given of stabilized Runge-Kutta methods for the solution of initial boundary value problems of hyperbolic type. The results of the first and second problem are compared, respectively with those of A. and F. Solomon [8] and with an approximation to the analytical solution.

Contents. page

1. Introduction

This paper contains two applications of stabilized Runge-Kutta methods for numerical integration of initial boundary value problems. These methods are described and analysed in van der Houwen [3]. Both problems considered here are chosen in the field of hyperbolic differential equations. The first problem originates from a paper by A. and F. Solomon [8]. They solved the problem numerically by means of a rather specific scheme.
Our purpose was to compare these results with those which may be obtained by the more general, stabilized Runge-Kutta methods.
The second problem is a mathematical model of dispersion of gas in a tube. This problem initially was subject of investigation by a university laboratory on chemical engineering.
The ALGOL 60 procedure "modified runge kutta" presented in Beentjes [1] was used. This procedure chooses its step sizes automatically, depending on the required accuracy and on the stability properties of the formula which is used to solve the system of differential equations.
The numerical experiments were carried out on the EL X8 computer of the Mathematical Centre.

2. The equation $(tU_t)_t = U_{xx}$

In this section we describe some methods for solving the initial boundary value problem

$$(2.1)\quad \begin{cases} (tU_t)_t = U_{xx}, & 0 \le x \le 2\pi \quad , \quad t > 0 , \\ U \quad = f(x), & 0 \le x \le 2\pi \quad , \quad t = 0 , \\ U \quad = g(t), & x = 0, x = 2\pi, \quad t \ge 0 , \end{cases}$$

for given, sufficiently differentiable functions f(x) and g(t).

In 1969 A. and F. Solomon [8] published a paper in which a numerical scheme for this problem was proposed.
We will compare our methods and results with those published in [8].
Equation (2.1) arose from considerations of heat conduction with delay and also governs the motion of a homogeneous rope with one free end when the variables t and x are interchanged.
It is hyperbolic for $t > 0$ and parabolic for $t = 0$. In the latter case the equation reduces to the heat equation

$$(2.2)\qquad U_t = U_{xx}.$$

In the following subsections we examine some methods for solving the particular initial boundary value problem

$$(2.3)\quad \begin{cases} t\, U_{tt} + U_t = U_{xx} , & 0 \le x \le 2\pi \quad , \quad t > 0 , \\ U = \cos x , & 0 \le x \le 2\pi \quad , \quad t = 0 , \\ U = J_0(2\sqrt{t}) , & x = 0, x = 2\pi \ , \quad t \ge 0 . \end{cases}$$

The analytical solution reads $U(x,t) = \cos x \; J_0(2\sqrt{t})$,
where J_0 denotes the Besselfunction of order zero.

2.1 The discretization of the space derivatives

Calling $U_t = v$ and $U_x = w$ we rewrite (2.3) as a system of first order partial differential equations

$$(2.4) \begin{cases} \begin{pmatrix} v \\ w \end{pmatrix}_t = \begin{pmatrix} -{}^1/t & {}^1/t{}^\partial/\partial x \\ {}^\partial/\partial x & 0 \end{pmatrix} \begin{pmatrix} v \\ w \end{pmatrix} , \\ v(x,0) = -\cos x \; , \\ w(x,0) = -\sin x \; , \\ v(0,t) = v(2\pi,t) = -J_1(2\sqrt{t})/\sqrt{t} \\ w(0,t) = w(2\pi,t) = 0 \; , \end{cases}$$

or more compactly,

$$(2.5) \begin{cases} \frac{\partial\vec{r}}{\partial t} = A(t,{}^\partial/\partial x)\;\vec{r} \; , \\ \vec{r}(x,0) = \vec{r}_o \; , \\ \vec{r}(0,t) = \vec{r}(2\pi,t) = \vec{r}_b . \end{cases}$$

If we discretize the operator ${}^\partial/\partial x$ in (2.5) we are led to a set of first order ordinary differential equations, i.e

$$(2.6) \quad \frac{d\vec{r}}{dt} = \begin{pmatrix} -{}^1/t & {}^1/t\frac{E_+-E_-}{2h} \\ \frac{E_+-E_-}{2h} & 0 \end{pmatrix} \vec{r} \; ,$$

where E± are the usual shift operators and h denotes the mesh size on the x-axis.
The vectorfunction $\vec{r}(t)$ has the components

$$(v_1,\ldots,v_{N-1},w_1,\ldots,w_{N-1})^T ,$$

where N is the number of points used on the x-axis.

Problem (2.5) now obtains the form

$$(2.7)\quad \begin{cases} \dot{\vec{r}} = D\,\vec{r} + \vec{g}(t), \\[2ex] v(jh) = -\cos(jh), \qquad j = 0,1,\ldots N \\ w(jh) = -\sin(jh), \end{cases}$$

with

$$D = \left[\begin{array}{cccc|cccc}
-{}^{1}/t & & & & 0 & \frac{1}{2ht} & 0\cdots & 0 \\
 & \ddots & & & -\frac{1}{2ht} & 0 & \frac{1}{2ht} & \vdots \\
 & & \ddots & & 0 & \ddots & \ddots & \ddots \\
 & & & & \vdots & -\frac{1}{2ht} & 0 & \frac{1}{2ht} \\
 & & & -{}^{1}/t & 0\cdots & 0 & -\frac{1}{2ht} & 0 \\
\hline
0 & \frac{1}{2h} & 0\cdots & 0 & & & & \\
-\frac{1}{2h} & 0 & \ddots & \vdots & & & & \\
0 & \ddots & \ddots & \frac{1}{2h} & & 0 & & \\
\vdots & & & & & & & \\
0\cdots & 0 & -\frac{1}{2h} & 0 & & & &
\end{array}\right]$$

and

$$\vec{r} = \left[\begin{array}{c} v_1 \\ \vdots \\ v_{N-1} \\ \hline w_1 \\ \vdots \\ w_{N-1} \end{array}\right] , \qquad \vec{g} = \left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ \hline \frac{1}{2h}\; J_1(2\sqrt{t})/\sqrt{t} \\ 0 \\ \vdots \\ 0 \\ -\frac{1}{2h}\; J_1(2\sqrt{t})/\sqrt{t} \end{array}\right] .$$

2.2 Numerical integration methods and stability criteria.

In order to integrate system (2.7) numerically we use polynomial methods described and analysed in [2].
We have to select suitable polynomials to integrate our system.
This selection depends completely on the location of the eigenvalues of the operator D of equation (2.7) in the complex plane.
Moreover, the spectral radius σ of D plays an important role in stability considerations.
We determine these eigenvalues by substituting the vector

$$(2.8) \qquad \vec{e}(t) = \vec{r}\,'(t) \exp(i\ \omega\ j\ h)$$

into the equations

$$(2.9) \qquad \begin{pmatrix} -{}^{1}/t & {}^{1}/t\dfrac{E_{+}-E_{-}}{2h} \\ \dfrac{E_{+}-E_{-}}{2h} & 0 \end{pmatrix} \vec{e} = \delta\ \vec{e}\ .$$

It then follows that the eigenvalues δ of D are

$$(2.10) \qquad \delta_{\omega} = \frac{1}{2t}\left(-1 \pm \sqrt{1-\frac{4t\ \sin^{2}(\omega h)}{h^{2}}}\right)$$

We distinguish the following cases

1) $t \leq h^{2}/4$.

From (2.10) it follows that all eigenvalues are real and negative.
The spectral radius σ is given by ${}^{1}/t$ (see figure 2.1)

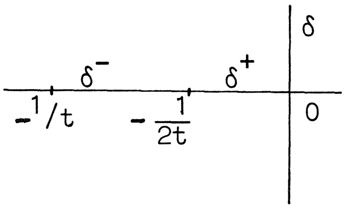

fig. 2.1 Eigenvalues of D in case 1.

2) $h^2/4 \le t \le h^2$.

In this case we see that the eigenvalues δ are situated on a cross in the left half plane (see figure 2.2).

It is easily verified that the spectral radius is still the same as in case 1.

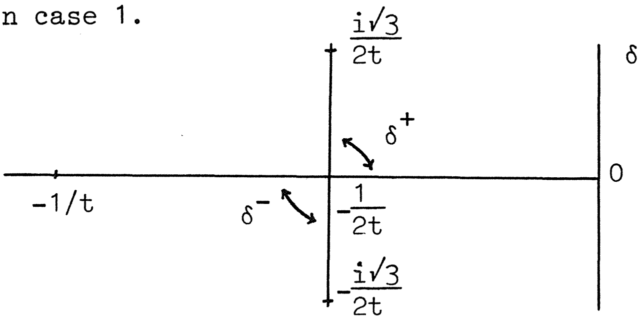

fig.2.2. Eigenvalues of D in the case $h^2/4 \le t \le h^2$.

3) $t \ge h^2$.

Again the eigenvalues δ are situated on a cross as in figure 2.2. However, the spectral radius becomes $^1/h\sqrt{t}$, because of the fact that in this case $\max_\omega|\delta| \ge \max_\omega|\text{Re}(\delta)|$.

We may conclude that in all cases the eigenvalues are situated in a region consisting of the negative x-axis and the disk $(x+h^{-2})^2 + y^2 \le h^{-4}$, where $\delta = x+iy$ (see figure 2.3).

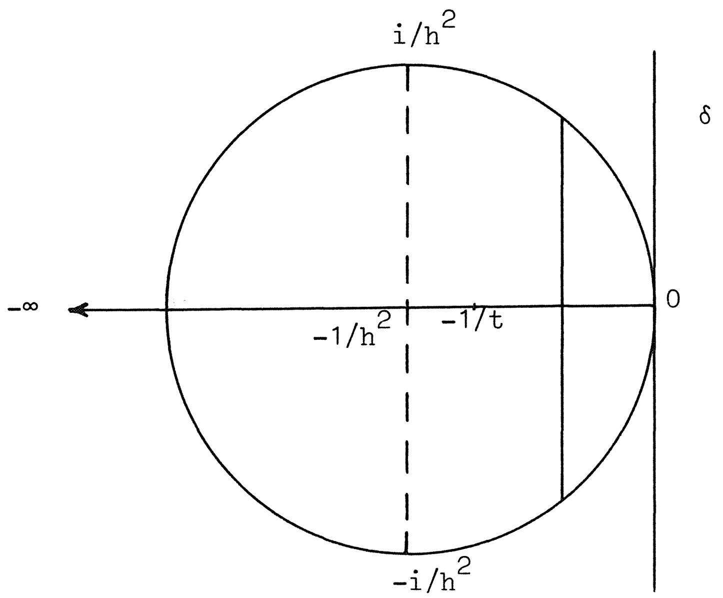

fig.2.3. Location of the eigenvalues of the matrix D.

The spectral radius is given by

$$(2.11)\qquad \sigma(t) = \begin{cases} 1/t & \text{for} \quad t \le h^2 \\ 1/h\sqrt{t} & \text{for} \quad t \ge h^2 . \end{cases}$$

Having determined the eigenvalues and the spectral radius of the Jacobian matrix D we can select the polynomials which generate our integration methods.

We also have to consider the accuracy required.

If only first order accuracy is wanted and the eigenvalues are real or "almost" real, then the Chebyshev polynomials $T_n(1+z/n^2) = \cos[n \arccos(1+z/n^2)]$ are most efficient.

For higher order accuracy we refer to [5].

If the eigenvalues are purely imaginary the polynomial

$$(2.12)\qquad P_4(z) = 1 + z + \frac{1}{2} z^2 + \frac{1}{6} z^3 + \frac{1}{24} z^4$$

is to be preferred. (see [4]).

Figure 2.3 suggests to start with one of the polynomials for real eigenvalues. However, the eigenvalues of D soon run out of the stability regions of these polynomials so that we have to switch to $P_4(z)$. Therefore we preferred to start directly with $P_4(z)$.

The stability condition becomes (see [2])

$$(2.13)\qquad \tau_k \le 2\sqrt{2}\ h \sqrt{t_k} \qquad \text{for } t \ge h^2 .$$

Polynomial (2.12) is fourth order exact. Hence, the analytical solution of (2.3) will locally satisfy the difference scheme generated by (2.12) apart from a term

$$(2.14)\qquad O(\tau^5) + O(\tau h^2).$$

Since (2.13) allows time steps of order h we have an approximation error of $O(h^3)$.

The approximation errors due to discretization of time and space derivatives are of order h^5 and h^3, respectively.
In general, the best results are obtained if both errors are of the same order, since in that case the possibility of partial cancellation exists. Therefore we can either lower the order of the method or raise the order of the operator which approximates the space derivative.
However, since we are interested in a rather accurate result, we decided for the latter. In the following subsection we derive an operator which approximates the operator $\partial/\partial x$ with third order accuracy as $h \to 0$.

2.3 Higher order discretization of the space derivative.

Taking into consideration arguments given at the end of section 2.2 we need an operator that approximates the derivative $\partial/\partial x$ with third order accuracy. For that purpose we define the operator

$$(2.15) \quad \Delta \equiv aE_-^2 + bE_- + c + dE_+ + eE_+^2 ,$$

where a,b,c,d and e are weight parameters to be determined in such a way that Δ approximates the operator $\partial/\partial x$ with third order accuracy. After expansion of the operator Δ, it turns out that conditions for a third order approximation are

$$(2.16) \quad \begin{cases} a + b + c + d + e = 0 , \\ -2a - b \quad + d + 2e = h^{-1}, \\ 4a + b \quad + d + 4e = 0 , \\ -8a - b \quad + d + 8e = 0 , \\ 16a + b \quad + d + 16e = 0 . \end{cases}$$

A simple calculation reveals that (2.16) is solved by the set of values

$$(2.17) \quad a = \frac{1}{12h} , \; b = -\frac{2}{3h} , \; c = 0 , \; d = \frac{2}{3h} \text{ and } e = -\frac{1}{12h} .$$

The operator Δ can be applied at the internal gridpoints (jh,t_k) $j = 2,3,\ldots,N-2$.
At the point (h,t_k) we define the operator

$$(2.18) \qquad \Delta_l' \equiv a'E_- + b' + c'E_+ + d'E_+^2 + e'E_+^3$$

and at the point $((N-1)h,t_k)$ the similar operator

$$(2.19) \qquad \Delta_r' \equiv a''E_+ + b'' + c''E_- + d''E_-^2 + e''E_-^3 \;.$$

It turns out that a third order approximation to ${}^{\partial}/\partial x$ at the boundary points is obtained by

$$(2.20) \qquad \begin{array}{l} a' = -a'' = -{}^{1}/4h \;,\; b' = -b'' = -{}^{5}/6h \;,\; c' = -c'' = {}^{3}/2h, \\ d' = -d'' = -{}^{1}/2h \;,\; e' = -e'' = {}^{1}/12h \;. \end{array}$$

Problem (2.5) can now be approximated by an initial value problem for the system of ordinary differential equations (2.7),

$$(2.21) \qquad D = \left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array}\right] \;, \qquad A = \begin{bmatrix} -{}^{1}/t & & 0 \\ & \ddots & \\ 0 & & -{}^{1}/t \end{bmatrix},$$

$$(2.22) \qquad B = \begin{bmatrix} {}^{b'}/t & {}^{c'}/t & {}^{d'}/t & {}^{e'}/t & 0 & \cdots & \cdots & 0 \\ {}^{b}/t & {}^{c}/t & {}^{d}/t & {}^{e}/t & 0 & & & \vdots \\ {}^{a}/t & {}^{b}/t & {}^{c}/t & {}^{d}/t & {}^{e}/t & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & {}^{a}/t & {}^{b}/t & {}^{c}/t & {}^{d}/t & & {}^{e}/t \\ \vdots & & 0 & {}^{a}/t & {}^{b}/t & {}^{c}/t & & {}^{d}/t \\ 0 & \cdots & 0 & {}^{e''}/t & {}^{d''}/t & {}^{c''}/t & & {}^{b''}/t \end{bmatrix}$$

(2.27) $C = t\,B$

and

$$(2.24)\qquad \vec{g} = -\frac{J_1(2\sqrt{t})}{\sqrt{t}}\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hline a' \\ a \\ 0 \\ \vdots \\ 0 \\ e \\ a'' \end{bmatrix}$$

From (2.17) and (2.20) it follows that

$$(2.25)\qquad B = \frac{1}{12ht}\begin{bmatrix} -10 & 18 & -6 & 1 & 0 & \cdots & 0 \\ -8 & 0 & 8 & -1 & 0 & & \vdots \\ 1 & -8 & 0 & 8 & -1 & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & 1 & -8 & 0 & 8 & & -1 \\ \vdots & & 0 & 1 & -8 & 0 & 8 \\ 0 & \cdots & 0 & -1 & 6 & -18 & 10 \end{bmatrix}$$

It is easily verified that the eigenvalues λ of the operator Δ are

(2.26) $\lambda = i \sin \omega h(4-\cos \omega h)\;/3h$.

The eigenvalues of the matrix D become

$$(2.27) \qquad \delta_\omega = \frac{1}{2t}\left(-1 \pm \sqrt{1-\frac{4t\sin^2\alpha(4-\cos\alpha)^2}{9h^2}}\right) ,$$

where $\alpha = \omega h$ (cf.(2.10)).

An analysis of formula (2.27) similar to section 2.2 reveals that the spectral radius

$$(2.28) \qquad \sigma(t) = \begin{cases} {}^1/t & \text{for} \quad t \leq {}^{12h^2}/(3+8\sqrt{6}) \simeq {}^{4h^2}/7.53 \\ 1.{}^4/h\sqrt{t} & \text{for} \quad t \geq {}^{4h^2}/7.53 . \end{cases}$$

The stability condition becomes

$$(2.29) \qquad \tau_k \leq 2h\sqrt{2t_k}\ /1.4 .$$

Now the analytical solution of (2.3) locally satisfies scheme (2.7) apart from a residual term

$$(2.30) \qquad O(\tau^5) + O(\tau h^4) = O(h^5) .$$

2.4 Actual computation scheme and numerical results.

Since the spectral radius of the matrix D is infinite for $t = 0$, we have to start our calculations in a way different from the one we described in the preceding sections.

We consider the Taylor series

$$v(x,\tau_\circ) = v(x,0) + \tau_\circ\, v_t(x,0) + {}^{\tau_\circ^2}/2!\ v_{tt}(x,0) + \ldots$$

and determine the following limits (cf.2.4)

$$\lim_{t\to 0} v_t = \lim_{t\to 0} \left(\frac{-v+w_x}{t}\right) = {}^1/2!\ v_{xx}(x,0) \ ,$$

$$\lim_{t\to 0} v_{tt} = \lim_{t\to 0} \frac{\partial}{\partial t} \left(\frac{-v+w_x}{t}\right) = {}^1/3!\ v_{xxxx}(x,0) \ ,$$

$$\lim_{t\to 0} v_{ttt} = \lim_{t\to 0} \frac{\partial^2}{\partial t^2} \left(\frac{-v+w_x}{t}\right) = {}^1/4!\ v_{xxxxxx}(x,0) \ .$$

Hence, the Taylor expansion becomes

$$(2.31) \qquad v(x,\tau_\circ) = \cos x\left(-1+\frac{\tau_\circ}{1!2!}-\frac{\tau_\circ^{\,2}}{2!3!}+\frac{\tau_\circ^{\,3}}{3!4!}-\ldots\right) = -\cos x\ J_1(2\sqrt{\tau_\circ})/\sqrt{\tau_\circ}$$

In a similar way we find that

$$(2.32) \qquad w(x,\tau_\circ) = \sin x\left(-1+\frac{\tau_\circ}{(1!)^2}-\frac{\tau_\circ^{\,2}}{(2!)^2}+\frac{\tau_\circ^{\,3}}{(3!)^2}-\ldots\right) = -\sin x\ J_0(2\sqrt{\tau_\circ})$$

After initializing the vectors $\vec{v}$ and $\vec{w}$ at $t = \tau_\circ$ by (2.31) and (2.32) respectively, we calculate the values of $\vec{v}(\tau_\circ + k\tau_k)$ and $\vec{w}(\tau_\circ + k\tau_k)$ by a fourth order Runge Kutta method which reads, represented by the array form introduced by Butcher, as follows

$$(2.33) \qquad \begin{array}{c|c} M & \Lambda \\ \hline & \Theta \end{array} \quad = \quad \begin{array}{c|cccc} 0 & & & & \\ {}^1/2 & {}^1/2 & & & \\ {}^1/2 & 0 & {}^1/2 & & \\ 1 & 0 & 0 & & \\ \hline & {}^1/6 & {}^1/3 & {}^1/3 & {}^1/6 \end{array}$$

Here M is the column vector $(\mu_0, \ldots, \mu_{m-1})$, Λ the lower triangular matrix containing the parameters $\lambda_{j,l}$ and Θ the row vector $(\theta_0, \ldots, \theta_{m-1})$.

Let $\tau_\circ = {}^h/2$ then, by virtue of (2.11), we may take for the next step size

$$\tau_1 = 2\sqrt{2}\ \tau_\circ .$$

After this step $t > h^2$ holds. Hence, the step size is determined by

$$\tau_k = 2\sqrt{2t_k}\, h \text{ and } \tau_k = 2\sqrt{2t_k}\, h/1.4$$

for first and third order approximations of the space derivative, respectively.

At every t-level we have to determine the values of u_k by a quadrature formula, e.g. Simpsons rule,

$$(2.34)\quad \begin{cases} u_k((j+2)h) = u_k(jh) + \frac{h}{3}\left[w_k(jh) + 4w_k((j+1)h) + w_k((j+2)h)\right] + O(h^5) \\ u_k(0) = J_0(2\sqrt{t_k}). \end{cases}$$

If higher order accuracy is required we can use one of Bode's rules [6].

This completes the description of the method used.

In table I and II the results of the experiments are listed for discretization I and II, respectively.

Table I First order exact approximation of $\partial/\partial x$ (discr. I).

t	$k \diagdown x$	0	$\pi/5$	$2\pi/5$	$3\pi/5$	$4\pi/5$	π	ε_{rel}	ε_{abs}
1.99958	19	0	$.37_{10}{}^{0}$	$.15_{10}{}^{+1}$	$-.19_{10}{}^{+1}$	$-.89_{10}{}^{0}$	$-.79_{10}{}^{0}$	$.71_{10}{}^{0}$	$.34_{10}{}^{-2}$
10.40000	40	0	$.50_{10}{}^{0}$	$.17_{10}{}^{+1}$	$-.12_{10}{}^{+1}$	$-.21_{10}{}^{0}$	$-.70_{10}{}^{0}$	$.44_{10}{}^{0}$	$.27_{10}{}^{-2}$
30.13000	66	0	$.29_{10}{}^{0}$	$.22_{10}{}^{+1}$	$-.38_{10}{}^{+1}$	$-.20_{10}{}^{+1}$	$-.18_{10}{}^{+1}$	$.14_{10}{}^{+1}$	$.61_{10}{}^{-2}$
60.23000	92	0	$.66_{10}{}^{-1}$	$-.13_{10}{}^{+1}$	$.35_{10}{}^{+1}$	$.20_{10}{}^{+1}$	$.18_{10}{}^{+1}$	$.14_{10}{}^{+1}$	$.38_{10}{}^{-2}$
100.02800	118	0	$.67_{10}{}^{0}$	$.35_{10}{}^{+1}$	$-.48_{10}{}^{+1}$	$-.21_{10}{}^{+1}$	$-.18_{10}{}^{+1}$	$.16_{10}{}^{+1}$	$.67_{10}{}^{-2}$

In these tables the relative errors in percents, i.e

$$(2.35)\quad R_j = 100\ \varepsilon_k^{(j)}\ /\ \tilde{U}(x_j,t_k).$$

are shown. The error $\varepsilon_k^{(j)}$ is defined by the difference between the analytical solution of (2.3)

(2.36) $\tilde{U}(x,t) = \cos x \; J_0(2\sqrt{t})$

and the numerical solution $u_k[j]$ at the point (jh,t_k)

$$\varepsilon_k^{(j)} = \tilde{U}(x_j,t_k) - u_k[j].$$

The errors ε_{rel} and ε_{abs} in the last two columns of table I and II are defined by

(2.37) $\varepsilon_{rel} = 100 \; \| \varepsilon_k \| / \| \tilde{U}(x,t_k) \|$

and

(2.38) $\varepsilon_{abs} = \| \varepsilon_k \|$,

where $\| \;\; \|$ denotes the Euclidean norm over the grid points $x_j = jh$ $(j = 0,1,\ldots,50)$.
We obtain far more accurate results if we apply the third order exact approximation of $\partial/\partial x$ as given in section 2.3.

Table II Third order exact approximation of $\partial/\partial x$ (discr. II).

t	k \ x	0	$\pi/5$	$2\pi/5$	$3\pi/5$	$4\pi/5$	π	ε_{rel}	ε_{abs}
1.99958	26	0	$-.34_{10}{-2}$	$-.96_{10}{-2}$	$.11_{10}{-1}$	$.43_{10}{-2}$	$.36_{10}{-2}$	$.39_{10}{-2}$	$.19_{10}{-4}$
10.31000	55	0	$.96_{10}{-3}$	$.23_{10}{-2}$	$-.13_{10}{-2}$	$-.15_{10}{-2}$	$.22_{10}{-4}$	$.66_{10}{-3}$	$.40_{10}{-5}$
30.45000	92	0	$-.35_{10}{-2}$	$-.92_{10}{-2}$	$.11_{10}{-1}$	$.46_{10}{-2}$	$.39_{10}{-2}$	$.40_{10}{-2}$	$.16_{10}{-4}$
60.36000	128	0	$.50_{10}{-2}$	$.13_{10}{-1}$	$-.11_{10}{-1}$	$-.32_{10}{-2}$	$-.22_{10}{-2}$	$.41_{10}{-2}$	$.12_{10}{-4}$
100.02800	164	0	$.27_{10}{-3}$	$.32_{10}{-3}$	$.74_{10}{-3}$	$.68_{10}{-3}$	$.66_{10}{-3}$	$.46_{10}{-3}$	$.19_{10}{-5}$

In order to compare our results with those of Solomon we introduce a measure for the computational labour by

(2.39) $C = \dfrac{K\,n\,c}{100h}$,

where K denotes the number of integration steps and n the order of a method with respect to t. (c=1 for a three-point formula and c=2 for a five-point formula).

Two methods are comparable if for both schemes the computational labour is nearly equal. Therefore we choose in discr. II the value of h twice as large as we used in computing table II, i.e. $h={}^{4\pi}/100$.

In table III and IV we show the results for both Solomons method and discr. II.

Table III. Comparison of Solomons scheme and discr. II.

	$k \backslash x$	0	$^{\pi}/5$	$^{2\pi}/5$	$^{3\pi}/5$	$^{4\pi}/5$	π	ε_{rel}	ε_{abs}
t=1.99958									
Solomon	46	0	$.71_{10}{}^{-2}$	$-.20_{10}{}^{-1}$	$.11_{10}{}^{0}$	$.79_{10}{}^{-1}$	$.78_{10}{}^{-1}$	—	—
II	15	0	$.66_{10}{}^{-3}$	$-.11_{10}{}^{-1}$	$.24_{10}{}^{-1}$	$.13_{10}{}^{-1}$	$.11_{10}{}^{-1}$	$.88_{10}{}^{-2}$	$.43_{10}{}^{-4}$
t=100.0280									
Solomon	319	0	$.32_{10}{}^{-2}$	$.18_{10}{}^{-1}$	$.34_{10}{}^{-1}$	$-.20_{10}{}^{-1}$	$.18_{10}{}^{-1}$	—	—
II	85	0	$.22_{10}{}^{-2}$	$.52_{10}{}^{-2}$	$.24_{10}{}^{-1}$	$.16_{10}{}^{-1}$	$.14_{10}{}^{-1}$	$.11_{10}{}^{-1}$	$.43_{10}{}^{-3}$

Introducing the discrete Euclidean norm $\|\ \|_R$ over the grid points $x_j = jh$ $(j = 0,5,\ldots,25)$ we can produce table IV.

Table IV. Overall comparison of Solomons scheme and discr. II.

K	t=1.99958	c	$\|\varepsilon_k\|_R / \|\tilde{U}\|_R$	h
46	Solomon	7	$15.73_{10}{}^{-2}$	$^{2\pi}/100$
15	II	10	$3.12_{10}{}^{-2}$	$^{4\pi}/100$
	t=100.0280			
319	Solomon	51	$4.69_{10}{}^{-2}$	$^{2\pi}/100$
85	II	54	$2.87_{10}{}^{-2}$	$^{4\pi}/100$

We may conclude that the method described in this paper and which is applicable to a large class of partial differential equations, is at least as good as the scheme of A. and F. Solomon, which, as it seems, was constructed for just one particular equation. With our method we can cover all physical problems which give rise to a hyperbolic or parabolic system of differential equations with the only restriction that the eigenvalues of the Jacobian matrix D should have no positive real parts.

3. A dispersion problem.

The next initial boundary value problem arose during chemical technological investigations [7].

$$(3.1)\quad \begin{cases} U_t \quad = -(vU)_x + dU_{xx}, & 0 \le x < \infty, \\ U(x,0) = 0, & \\ U(0,t) = g(t) & \end{cases}$$

with $d = .434_{10}{}^{-1}$ and

$$v(x) = (c_1 + c_2 x)^{-\frac{1}{2}}\ , \quad c_1 = .4272_{10}{}^{-1}\ , \quad c_2 = -.597_{10}{}^{-4}\ .$$

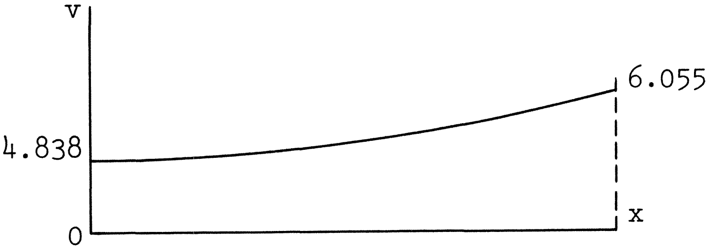

fig.3.1. The function v(x)

The function g(t) is not given in an analytical form, but as a table of values $g(t_k)$, $t_k = 2.56(.08)7.96$.

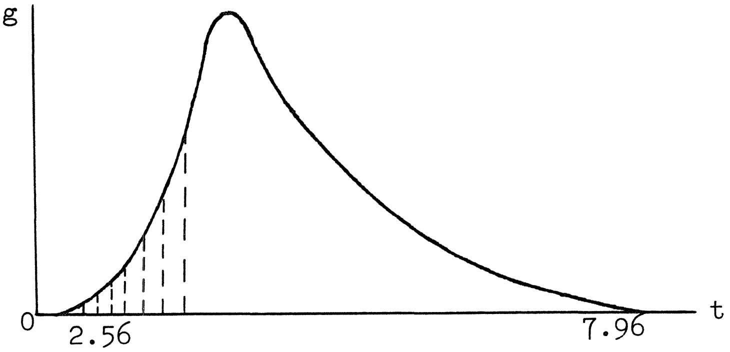

fig.3.2. The function g(t)

Problem (3.1) describes the density of a certain gas in a tube with length L = 258.8. The function g(t) represents a gas injection started at t = 2.56 untill t = 7.96.

The process proceeds as follows:
A density wave is built up and after some time, say t_1, it travels along the positive x-axis by the convection term $-vU_x$. The wave is slightly damped by the term $-v'U$ and spread out by diffusion. The maximal density or concentration (U_{max}) moves on untill it reaches the end of the tube at $x = L$, where measurements are done.
There are two difficulties in integrating problem (3.1) numerically

1. Untill the time $t = t_1$ the step size τ_k has exactly to be .08 or, if this is not possible in view of accuracy or stability requirements, we have to interpolate between the given values of the function $g(t)$.

2. The function $g(t)$ reaches in a very short time a large value and then decreases to 0 in a slightly longer time. This implies the presence of very large derivatives.
   Hence, if we do not take the value of Δx very small the numerical scheme will introduce spurious oscillations, since the coefficients of the differential equation depend on x.
   Thus it will take a very long time to integrate problem (3.1) untill U_{max} reaches $x = L$. Therefore we will use a smoothing technique to suppress the incorrect oscillations.

3.1 The discretization of the space derivatives.

If we discretize the operators $\partial/\partial x$ and $\partial^2/\partial x^2$ in problem (3.1) by central differences we are led to a set of ordinary differential equation of the form

$$(3.3) \qquad \frac{d\vec{U}}{dt} = D\vec{U} + \vec{F}(t) \ ,$$

where

$$(3.4)\quad D=\begin{bmatrix} -2d/h^2 & -v(2h)/2h+d/h^2 & 0 & \cdots & \cdots & 0 \\ v(h)/2h+d/h^2 & -2d/h^2 & -v(3h)/2h+d/h^2 & & & \vdots \\ 0 & \ddots & \ddots & \ddots & & 0 \\ \vdots & & v((N-3)h)/2h+d/h^2 & -2d/h^2 & -v((N-1)h)/2h+d/h^2 \\ 0 & \cdots & 0 & v((N-2)h/2h+d/h^2 & -2d/h^2 \end{bmatrix}$$

and

$$\vec{U}=\begin{bmatrix} U(h) \\ \vdots \\ U((N-1)h \end{bmatrix} \qquad \vec{F}(t)=g(t)\begin{bmatrix} v(0)/2h+d/h^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

At t=2.56 we have $U_j=0$ j=0,...,N.

The initial boundary value problem (3.1) has now become a Cauchy problem, since the left boundary function g(t) is included in the vector $\vec{F}$. For reasons of simplicity we suppose that the tube is longer than L = 258.8, such that the differential equation still holds, when the wave passes L.

3.2 The spectral radius of the Jacobian matrix and the numerical scheme.

As mentioned in section 2.2 the eigenvalues δ and the spectral radius σ of the Jacobian matrix D play an important role in stability considerations.

After substituting the vector

$$(3.5)\qquad \vec{e}(t)=\vec{r}\,'(t)\exp(i\omega jh)$$

into equation

$$(3.6) \qquad D\vec{e} = \delta\vec{e} \ ,$$

it follows that

$$(3.7) \qquad \delta_j = [{}^{-4d}/h^2 + (v_{j+1}-v_{j-1})/h] \sin^{2\omega h}/2 - (v_{j+1}-v_{j-1})/2h$$

$$- {}^{i}/2h(v_{j+1}+v_{j-1}) \sin \omega h,$$

where v_j denotes $v(jh)$.

It is easily verified that the spectral radius becomes

$$(3.8) \qquad \sigma = \max_{0 \le j \le N} \{{}^{16d^2}/h^4 - {}^{4d(v_{j+1}-v_{j-1})}/h^3 + {}^{(v_{j+1}^2+v_{j-1}^2)}/2h^2\}^{\frac{1}{2}}$$

From (3.7) it follows that the eigenvalues are complex and "almost" purely imaginary.

In section 2.2 and in [4] it was pointed that for this case the polynomial

$$(3.9) \qquad P_4(z) = 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \frac{1}{24}z^4$$

is an appropriate choice.

The 4-point Runge-Kutta method with third order accuracy generated by (3.9) is characterized by the array form

$$\begin{array}{c|cccc} \mu_1 & \lambda_{10} & & & \\ \mu_2 & 0 & \lambda_{21} & & \\ \mu_3 & 0 & 0 & \lambda_{32} & \\ \hline & \theta_1 & \theta_2 & \theta_3 & \theta_4 \end{array} = \begin{array}{c|cccc} {}^{8}/17 & {}^{8}/17 & & & \\ {}^{8}/15 & 0 & {}^{17}/60 & & \\ {}^{2}/3 & 0 & 0 & {}^{5}/12 & \\ \hline & {}^{1}/4 & 0 & 0 & {}^{3}/4 \end{array}$$

(see section 2.4 and [3]).

Runge-Kutta methods make use of intermediate time levels, hence for $t<t_1$ we have to interpolate the given values $g(t_k)$ at the points

$$\zeta_k{}^j = t_k + \mu_j\tau_k \quad (j=1,2,3).$$

We determined these intermediate values $g(\zeta_k{}^j)$ by means of the 4-point Lagrange interpolation formula

$$(3.11) \quad g(\zeta_k{}^j) = \sum_{i=k-1}^{k+2} \prod_{\substack{l=k-1\\ l\neq i}}^{k+2} \left[\frac{\zeta_k{}^j - t_l}{t_i - t_l}\right] g(t_i) \; (j=1,2,3).$$

Method (3.10) is third order exact, hence we have a discretization error

$$(3.12) \quad O(\tau^4) + O(\tau h^2).$$

From the stability condition

$$(3.13) \quad \tau \le 2\sqrt{2}/\sigma(D)$$

and (3.8) it follows that $\tau = O(h)$.
We may conclude that the approximation error of the difference scheme generated by (3.10) is of order h^3.

3.3. An approximation to the analytical solution.

In order to get an impression of the analytical solution of problem (3.1) we assume the coefficients of the differential equation to be constant instead of slowly varying:

$$(3.14) \quad \begin{cases} U_t = aU + bU_x + dU_{xx}, \\ U(x,0) = 0, \\ U(0,t) = g(t), \\ U(\infty,t) < \infty. \end{cases}$$

Let $\bar{u}(x,s) = \mathcal{L}U(x,t) = \int_0^\infty e^{-st}\, U(x,t)dt$;

then problem (3.14) becomes

$$(3.15)\quad \begin{cases} d\bar{u}_{xx} + b\bar{u}_x + (a-s)\bar{u} = 0 \\ \bar{u}(0,s) = \bar{g}(s) \\ \bar{u}(\infty,s) < \infty \end{cases}$$

The general solution of (3.15) is given by

$$(3.16)\quad \bar{u}(s,x) = A(s)\, e^{(p+q)x} + B(s)\, e^{(p-q)x},$$

where $p = {}^{-b}/2d$ and $q = (b^2-4d(a-s))^{\frac{1}{2}}/2d$.
Taking for the values of a and b the average values of $-v'(x)$ and $-v(x)$ on $(0,L)$ respectively, we may state

$$(3.17)\quad \begin{cases} b < 0, \text{ hence } p > 0, \\ a < 0 \text{ and } d > 0, \text{ hence } q > 0,\ q \in \mathbb{R}. \end{cases}$$

Furthermore, (3.17) yields

$$(3.18)\quad \begin{cases} A(s) \equiv 0, \\ B(s) \equiv \bar{g}(s) \end{cases}$$

so that the solution of the transformed problem becomes

$$\bar{u}(x,s) = \bar{g}(s)\, \exp(p-q)x).$$

Hence,

$$U(x,t) = \exp(px)\, g(t) * \mathcal{L}^{-1}\{\exp(-q(s)x)\} =$$

$$(3.19)\quad g(t) * \frac{x}{2\sqrt{\pi d t^3}}\, \exp(-\frac{(x+bt)^2}{4dt} + at).$$

The function g(t) is given at the discrete points t_k=2.56 + k*.08, k=0,1,...,80, as follows

$g(t_k)$= 0, 6, 150, 998, 2796, 4238, 4320, 3480, 2480, 1690, 1152, 804, 578, 426, 322, 250, 198, 158, 130, 108, 90, 76, 66, 56, 48, 42, 38, 34, 30, 26, 24, 22, 20, 18, 16, 14, 14, 14, 12, 10, 10, 10, 10, 8, 8, 8, 8, 8, 6, 6, 6, 6, 6, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 0.

In order to make the convolution (3.19) not too complicated we represent g(t) by

$$(3.20) \qquad g(t) = C\ \delta(t-\alpha).$$

Then, the analytical solution of problem (3.14) reduces to

$$(3.21) \qquad U(x,t) = \frac{C\ x}{2\sqrt{\pi d(t-\alpha)^3}} \exp\{-\frac{(x+b(t-\alpha))^2}{4d(t-\alpha)} + a(t-\alpha)\}\ .$$

3.4 Numerical results and smoothing.

If we apply method (3.10) to problem (3.3) with the given boundary function U(0,t) = g(t) we get results which start to oscillate after 36 steps at t=5.52. These small oscillations do not vanish but, on the contrary, become more serious when t increases (see fig. 3.3).

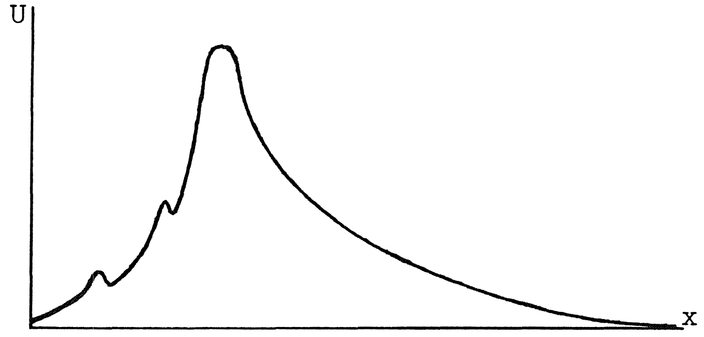

fig.3.3. Spurious waves are superposed on the solution.

By convection these oscillations move on along the x-axis and their wavelengths increase since the coefficients of the differential equation depend on x. At t=5.52 only three netpoints are involved but at t=6.00 the oscillations already cover five netpoints. Hence, we need a smoothing technique that keeps the oscillations under control during the whole integration process. It turned out that a smoothing operator of the form

$$(3.22)\quad S \equiv \beta + \alpha\, E_{2+} + \gamma E_{2-}\ ,$$

where $E_{2\pm}\ U_j = U_{j\pm 2}$, sufficiently damped the oscillations. For reasons of consistency the coefficients α, β and γ have to be determined in such way that

$$(3.23)\quad U_j^s \equiv S\ U_j = U_j + O(h^2).$$

Expansion of the shifted functions U_{j+n} $(n=\pm 2)$ reveals that (3.23) is satisfied if

$$(3.24)\quad \begin{cases} \alpha + \beta + \gamma = 1, \\ -2\alpha \qquad + 2\gamma = 0. \end{cases}$$

Hence,

$$(3.25)\quad U_j^{s^n} = \beta U_j^n + \frac{1-\beta}{2}(u_{j+2}^n + u_{j-2}^n)\ .$$

Substitution of the Fourier component $\hat{u}_\omega\ e^{i\omega jh}$ into (3.25) yields

$$(3.26)\quad \hat{u}_\omega^s = (\beta + (1-\beta)\cos 2\omega h)\ \hat{u}_\omega\ .$$

The multiplicative factor of (3.26) is real, so that the phases of the components are unaffected.

From (3.26) it follows that the high-frequency oscillations with $\omega = \frac{\pi}{2h}$ will vanish if $\beta = \frac{1}{2}$ (see fig. 3.4).

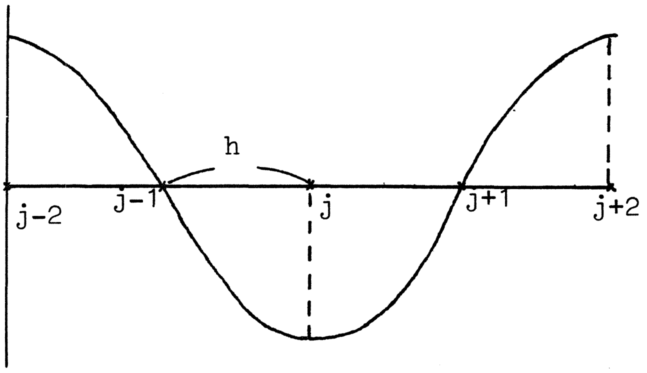

fig.3.4. Oscillations damped by the operator S.

The energy spectrum of the operator S is given by

(3.27) $A_S^2(\xi) = [\frac{1}{2}(1+\cos 2\xi)]^2 = \cos^4(\xi)$ (see fig. 3.5).

where $\xi = \omega h$.

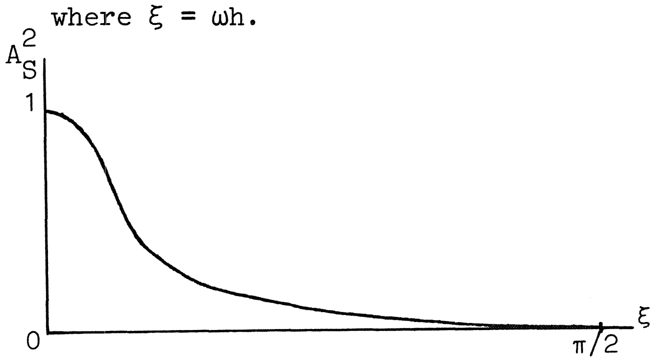

fig.3.5. Energy spectrum of S.

We applied smoothing after every integration step only in those points where oscillations started.

It turned out that this partially applied, first order exact smoothing operator removed all oscillations during the whole interval of integration. (see figure 3.6).

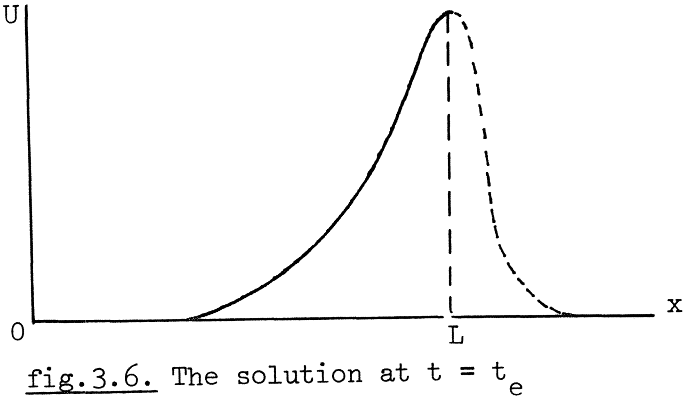

fig.3.6. The solution at $t = t_e$

We compared the numerical solution and the approximated analytical solution (3.21) at t=25.57 and t=51.17.

The most significant quantities of the solutions are (see figure 3.7)

1. The value U_{max},
2. The coordinate x_{max} for which $U(x_{max}) = U_{max}$
3. The width $\varepsilon = x_2 - x_1$, where x_2 and x_1 are the coordinates for which $U = \frac{1}{2}U_{max}$.

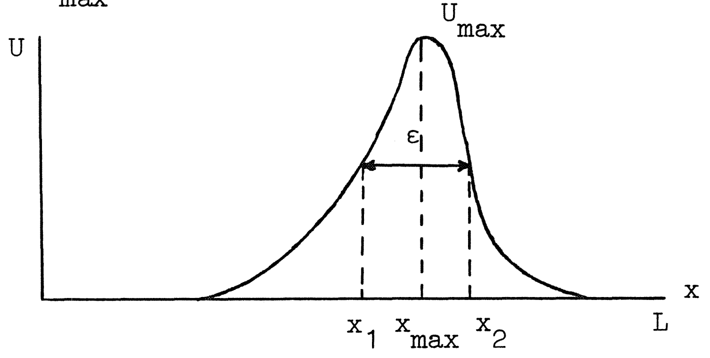

fig. 3.7 Significant quantities of the solution

At t=51.17 only the value of $\varepsilon' = x_{max} - x_1$ could be determined, because at that time $x_{max} \simeq L$ and the values of U for x > L were not computed.

We also evaluated the convolution integral (3.19) by means of Simpsons rule.

In table V some numerical results are shown.

Table V Comparison of the numerical and analytical solution.

	ε	x_{max}	U_{max}	
formula (3.21)	3.2	114.6	2635.5	
formula (3.19)	4.0	113.4	2035.0	t = 25.57
numerical solution	4.4	112.6	1963.5	
	1	x_{max}	U_{max}	
formula (3.21)	2.4	258.6	1686.9	
formula (3.19)	2.9	258.3	1402.2	t = 51.17
numerical solution	3.1	256.2	1379.5	

Although the chemist injects a δ-like boundary function we may conclude from table V that formula (3.21) gives a less accurate approximation of the solution of problem (3,1).
The convolution integral (3.19) well agrees with the numerical solution in all quantities.

Of course the numerical treatment is far more laborious than the evaluation of the convolution integral, however, with the numerical method we can also treat differential equations of which the coefficients vary considerably.

References.

[1] Beentjes, P.A., *Een ALGOL 60 versie van gestabiliseerde Runge-Kutta methoden*, NR rapport 23/72, Mathematisch Centrum, Amsterdam (1972).

[2] Houwen, P.J. van der, *One step methods for linear initial-value problems I, Polynomial methods*, TW report 119, Mathematisch Centrum, Amsterdam (1970).

[3] Houwen, P.J. van der, *Stabilized Runge-Kutta methods with limited storage requirements*, TW report 124/71, Mathematisch Centrum, Amsterdam (1971).

[4] Houwen, P.J. van der, Beentjes, P., Dekker, K. and Slagt, E., *One step methods for linear initial-value problems III, Numerical examples*, TW report 130/71, Mathematisch Centrum, Amsterdam (1971).

[5] Houwen, P.J. van der, Kok, J., *Numerical solution of a minimax problem*, TW report 123/71, Mathematisch Centrum, Amsterdam (1971).

[6] *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series 55, eds: Abramowitz, M. and Stegun, I.A.

[7] *De Ingenieur*, Orgaan van het Koninklijk Instituut van Ingenieurs, Jaargang 84, no. 26, 30 juni 1972.

[8] Solomon, A. and Solomon, F., *The initial-value problem for the equation* $(tU_t)_t = U_{xx}$, Mathematics of Computation, volume 24, number 111, July 1970.