NUMERICAL ASPECTS OF THE ONE-DIMENSIONAL DIFFUSION EQUATION

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE TECHNISCHE WETENSCHAPPEN AAN DE TECHNISCHE HOGESCHOOL TE DELFT OP GEZAG VAN DE RECTOR MAGNIFICUS IR. H. J. DE WIJS, HOOGLERAAR IN DE AFDELING DER MIJNBOUWKUNDE, VOOR EEN COMMISSIE UIT DE SENAAT TE VERDEDIGEN OP WOENSDAG 24 JUNI 1964 DES NAMIDDAGS TE 4 UUR

DOOR

LEENDERT DEKKER

WISKUNDIG EN ELEKTROTECHNISCH INGENIEUR

geboren te Rotterdam

DIT PROEFSCHRIFT IS GOEDGEKEURD DOOR DE PROMOTOR

PROF. DR. E. VAN SPIEGEL

C O N T E N T S

INTRODUCTION

The investigation from which this thesis has arisen relates to the design and application of a special analogue computer for solving such problems as

$$\frac{\partial z_i}{\partial t} - a_i \frac{\partial^2 z_i}{\partial x^2} = f_i(x,t,z_1,\ldots,z_n), \qquad a_i > 0,$$

$$z_i(0,t) = \varphi_i(t),\ z_i(1,t) = \eta_i(t),\ z_i(x,0) = \gamma_i(x), \qquad i = 1,\ldots,n.$$

The realized special analogue computer does not solve a given problem itself but a numerical approximation of it obtained by replacing the differential equations by difference equations. As a consequence x and t can only have discrete values.

As for the choice of the type of the difference equation and the values of the step widthes one has to take care that the exact solution of the difference problem is a good representation of the solution of the differential problem.

The step widthes ought not to be chosen smaller than strictly needed because smaller values of the step widthes result in higher cost for obtaining a solution. In connection with this sometimes it seems sensible to fit the boundary and initial values of a given differential problem before using them with the corresponding difference problem.

Studying the criteria which the step widthes have to satisfy it has proven to be useful to apply two concepts: the external solution and the internal solution.

The special analogue computer does not solve a difference problem directly but by means of an iteration process at each time t. The rate of convergence of this iteration process proves to be high in comparison with a digital iteration process.

The thesis can be divided into two parts. In the first one, being the chapters I and II, the concepts external and internal solution are discussed from analytic point of view. The second part deals with the numerical aspects of solving problems of the above mentioned type. We remark that for a good understanding of the second part no detailed knowledge of the first two chapters is needed.

Chapter 1

EXTERNAL AND INTERNAL SOLUTION FOR ORDINARY DIFFERENTIAL EQUATIONS

1.1. Definitions

In this thesis a number of notions are used for boundary and/or initial value problems of different kinds. All these problems have boundaries on which one of the independent variables is constant. Because most problems have only one dependent variable and one or two independent variables, in this paragraph we will define the mentioned notions only for the last problems. Moreover we will restrict ourselves to linear problems. Further existence and uniqueness of the notions will be left out of consideration.

If we conceive differentiation, matrixtransformation, etc. as operations, then all considered problems can be written in the same form. Denoting the operators of a problem by O_i, $i = 1,\dots,n$, then for each problem the solution f has to satisfy a number of equations $Lf = g$, where $L = \sum_{i=1}^{n} k_i O_i$ with k_i = constant and g is a given quantity. Only one of these equations is valid in the whole region for which the problem is formulated. This equation can be a partial differential equation, an integral-difference equation, etc. and will, in general, be called system equation. The other equations are the boundary and/or initial conditions.

Denoting the independent variables by x_i, $i = 1,2$, a boundary will be called a x_i-boundary, if x_i = constant on this boundary.

First we consider a problem without boundary and initial conditions, thus only a system equation $Lf = g$ is given. We state that the operators of the system equation can be applied infinitely often to g and its transforms, already performed. Then we define the <u>external solution</u> as that linear composition of g and its transforms which satisfies the system equation in such a way that the coefficients of this linear composition are independent on the choice of g.

For a boundary and/or initial value problem we state that the operators of the system equation can be applied infinitely often to the right hand members g of the system equation and of the equations valid on the x_i-boundaries as well as to the transforms of these right hand members g. Then we define the <u>x_i-external solution</u> as that linear composition of these right hand members g and their transforms, which satisfies the system equation and the equations, valid on the x_i-boundaries, where the coefficients of the linear composition are independent on the choice of the right hand members g.

In general the external (or x_i-external) solution is not the solution of the given problem. The difference between the solution of the given problem and the external solution is defined as the <u>internal</u> (or x_i-internal) <u>solution</u>.

In this thesis we call f an elementary solution of an equation $Lf = 0$ if f satisfies this equation and is also an eigenelement of all operators O_i of L (this means $O_i f = c_i f$ with c_i = constant).

We will call f a x_1-elementary solution if

1) it is an eigenelement of all operators of the system equation with exception of the operators representing an operation with respect to x_1,
2) it satisfies the homogeneous system equation as well as the homogeneous conditions on the x_1-boundaries.

We can conceive a boundary and initial value problem as describing a physical system. If the solution of a given problem is identically zero, then we say that the corresponding physical system is in absolute rest. If only the internal solution is identically zero, then we say that the physical system is in relative rest. We call a linear physical system stable if for each deviation out of (absolute or relative) rest the physical system keeps this deviation or comes back again into the state of rest in the long run. In the contrary case we call the physical system unstable.

1.2. <u>External and internal solution for linear ordinary differential and difference equations with constant coefficients</u>

Up to now the notions external and internal solution are defined generally in a rather abstract way. Now we will illustrate these notions for simple initial value problems.

First we consider the problem

$$a_n y^{(n)}(t) + \dots + a_o y(t) = g(t),\ a_o,\dots,a_n = \text{constants, where}$$

$$y(0),\dots,y^{(n-1)}(0) \text{ are prescribed.} \tag{1}$$

The system equation of this problem is an ordinary differential equation, the operators $0_1,\dots,0_n$ are $\frac{d}{dt},\dots,\frac{d^n}{dt^n}$. We assume that g(t) can be differentiated infinitely often.
The eigenfunctions of $\frac{d}{dt}$ are

$$e^{\lambda t};\ \lambda = \text{constant.} \tag{2}$$

By definition an elementary solution of the homogeneous differential equation has to be an eigenfunction of all operators $\frac{d}{dt},\dots,\frac{d^n}{dt^n}$. So an elementary solution is $e^{\lambda_j t}$, provided λ_j satisfies

$$a_n\lambda^n + \dots + a_o = 0. \tag{3}$$

We restrict ourselves to the case that all roots of this equation are different. Then there are n elementary solutions $e^{\lambda_j t}$ and the general solution of the homogeneous differential equation is

$$\sum_{j=1}^{n} b_j e^{\lambda_j t};\ b_j = \text{constant.} \tag{4}$$

According to the definition of paragraph 1 the external solution of differential equation (1) is a linear composition of $g(t)$ and its derivatives in such a way that the coefficients of this composition are independent on the choice of $g(t)$. We write the external solution $y_e(t)$ in the form

$$y_e(t) = \sum_{k=o}^{\infty} c_k g^{(k)}(t). \tag{5}$$

Supposing that it is allowed to differentiate (5) term by term when (5) is substituted in (1) gives

$$\sum_{k=o}^{\infty} \Big(\sum_{j=o}^{\mathrm{Min}(n,k)} a_j c_{k-j} \Big) \, g^{(k)}(t) = g(t). \tag{6}$$

As the coefficients c_k must be independent on the choice of $g(t)$, it follows from (6) that these coefficients satisfy

$$\begin{array}{ll}
a_o c_o & = 1 \\
a_1 c_o + a_o c_1 & = 0 \\
\vdots & \\
a_n c_o + \cdots + a_o c_n & = 0 \\
\vdots & \\
a_n c_r + \cdots + a_o c_{n+r} & = 0. \\
\vdots &
\end{array} \tag{7}$$

Obviously the solution for the constants c_k is unique, provided $a_o \neq 0$.

The infinite system of equations, that can be obtained from (7) by omitting the first n equations, represents a recurrent relation between $(n+1)$ succeeding c's:

$$a_n c_r + \ldots\ldots\ldots + a_o c_{n+r} = 0; \; r \geqq 0. \tag{8}$$

The initial conditions of (8) are formed by the first n equations of (7). The solution of (8) can be found in the following way. In relation to (8) we consider the functions $h(x)$ which satisfy the following difference equation

$$a_n h(x) + \ldots + a_1 h(x+n-1) + a_o h(x+n) = 0. \tag{9}$$

An elementary solution of this equation is $\mu_j^{\,x}$, provided μ_j satisfies

$$a_o \mu^n + a_1 \mu^{n-1} + \ldots + a_n = 0. \tag{10}$$

Comparing (10) with (3), we deduce

$$\mu_j = \frac{1}{\lambda_j} \; ; \; j = 1,\ldots,n. \tag{11}$$

Hence the general solution of (9) is

$$\sum_{j=1}^{n} d_j \left(\frac{1}{\lambda_j}\right)^x ; \; d_j = \text{arbitrary constant.} \tag{12}$$

Substituting x = k in (12), we obtain for the general solution of (8)

$$c_k = \sum_{j=1}^{n} d_j \left(\frac{1}{\lambda_j}\right)^k . \tag{13}$$

Finally the solution of (7) can be found by choosing the constants d_j such that (13) also satisfies the first n equations of (7). Since the constants c_k are uniquely fixed by (7), we can conceive (13) as a system of n linear equations with unknowns $d_1,\ldots,d_n$. The coefficient determinant of this system is a Vandermonde determinant, which implies, that the solution of (13) is unique.

The above considerations have been based on the assumption that series (5) can be differentiated term by term. Now the condition will be determined, which g(t) has to satisfy for this. From (13) it follows

$$|c_k| = |d_1|\left|\frac{1}{\lambda_1}\right|^k \left| 1 + \sum_{j=2}^{n} \frac{d_j}{d_1}\left(\frac{\lambda_1}{\lambda_j}\right)^k \right| \leq |d_1|\left|\frac{1}{\lambda_1}\right|^k \left\{ 1 + \left| \sum_{j=2}^{n} \frac{d_j}{d_1}\left(\frac{\lambda_1}{\lambda_j}\right)^k \right| \right\}. \tag{14}$$

Suppose that the constants λ_j have been arranged in increasing absolute value. Then $|\lambda_1/\lambda_j| < 1$ for each j. This means that for each $\varepsilon > 0$, a number M_1 can be found such that

$$\left|\sum_{j=2}^{n} \frac{d_j}{d_1}\left(\frac{\lambda_1}{\lambda_j}\right)^k\right| < \varepsilon \qquad \text{for } k > M_1 . \tag{15}$$

Substitution of this in (14) gives:

$$|c_k| < (1 + \varepsilon)|d_1|\left|\frac{1}{\lambda_1}\right|^k \qquad \text{for } k > M_1 . \tag{16}$$

Suppose that in the t-interval $[t_1,t_2]$ the function g(t) has the property that two positive numbers λ and M_2 can be found for which

$$|g^{(k)}(t)| \leq \text{constant} \, . \, \lambda^k \text{ with } \lambda < |\lambda_1| \quad \text{for } k > M_2 . \tag{17}$$

If g(t) satisfies (17), it holds by virtue of (16):

$$|c_k g^{(k)}(t)| < \text{constant}.\ (1+\varepsilon)|d_1||\frac{\lambda}{\lambda_1}|^k \text{ for } k > \text{Max}(M_1,M_2). \qquad (18)$$

So for sufficiently large k the terms of series (5) in absolute value are smaller than the terms of a geometric series with ratio $|\lambda/\lambda_1|$ smaller than 1. Hence series (5) is uniformly convergent in $[t_1,t_2]$, if $g(t)$ satisfies (17). Likewise it is simple to see, if $g(t)$ satisfies (17) that the series which arise from (5) by differentiating term by term, are uniformly convergent. So the sum of series (5) is a solution of the inhomogeneous differential equation.

Hence the sum of series (5) can be considered as the definition of the external solution.

The internal solution of an inhomogeneous problem satisfies the homogeneous differential equation. Its initial conditions are equal to the differences between the given initial conditions and the corresponding values of the external solution. So the internal solution $y_i(t)$ of (1) satisfies

$$a_n y_i^{(n)}(t) + \ldots + a_o y_i(t) = 0$$

$$y_i(0) = y(0) - y_e(0);\ldots;y_i^{(n-1)}(0) = y^{(n-1)}(0) - y_e^{(n-1)}(0). \qquad (19)$$

Apparently the internal solution can be found from the general solution (4) by choosing the constants b_j such, that the initial conditions of the internal solution are satisfied.

Before illustrating the notions external- and internal solution for ordinary difference equations with constant coefficients, first the results of the preceding will be generalised. In the system equation of (1), replacing $\frac{d}{dt}$ by O_1, being some linear operator, gives

$$a_n O_1^n y + a_{n-1} O_1^{n-1} y + \ldots + a_o y = g. \qquad (20)$$

Because from $O_1 f = \lambda f$ it follows $O_1^m f = \lambda^m f$, the elementary solutions of the homogeneous operation equation, belonging to (20), are those eigenfunctions f_j of O_1, for which the eigenvalues λ_j satisfy the equation (3). If again all λ_j are different and if only one f_j corresponds to each λ_j, then the general solution of the homogeneous equation, belonging to (20), is

$$\sum_{j=1}^{n} b_j f_j \quad \text{with } b_j = \text{arbitrary constant}. \qquad (21)$$

We can deduce from (5) and (17), that the external solution y_e of (20) is

$$y_e = \sum_{k=o}^{\infty} c_k O_1^k g, \qquad (22)$$

provided g satisfies, for sufficiently large k,

$$|O_1^k g| \leqq \text{constant} \,.\, \lambda^k \text{ with } \lambda < |\lambda_1|. \tag{23}$$

Again the internal solution can be found from the general solution (21) by choosing the constants b_j such that the auxiliary conditions of the internal solution are satisfied.

Now we consider the case that O_1 is a difference operator *). An expression of the form $\Delta y/\Delta t = \{y(t+\Delta_1 t+\Delta t) - y(t+\Delta_1 t)\}/\Delta t$ is called a difference quotient of y in point t, if $\Delta_1 t, \Delta t$ are constants. Under $\Delta^n y/\Delta t^n$ will be understood the transform which is obtained from y(t) by applying the difference operator $\frac{\Delta}{\Delta t}$ successively n times. In the case $O = \frac{\Delta}{\Delta t}$ the elementary solutions f_j can be determined in the following way. The eigenfunctions f of the operator $\frac{\Delta}{\Delta t}$ are

$$f = \rho^{\frac{t}{\Delta t}};\ \rho = \text{constant}. \tag{24}$$

An eigenfunction f is an elementary solution f_j of the homogeneous difference equation, belonging to (20), if its eigenvalue satisfies equation (3), so if it holds

$$(\rho_j - 1)\rho_j^{\Delta_1 t/\Delta t} = \lambda_j \Delta t\,. \tag{25}$$

*) In this case the external solution is called "Hauptlösung" by N.E. Nörlund, page 40 of [1].

Chapter 2

EXTERNAL AND INTERNAL SOLUTION FOR THE PARABOLIC DIFFERENTIAL EQUATION

2.1. The x-external solution for a diffusion problem

We consider a physical system described by the equation $\partial z/\partial t - \partial^2 z/\partial x^2 = 0$, where the boundaries are $x = 0$ and $x = 1$. Supposing z is given on the x-boundaries and the initial state of the physical system is known, the problem can be formulated by

$$\frac{\partial z}{\partial t} - \frac{\partial^2 z}{\partial x^2} = 0$$

$$z(0,t) = \varphi(t), \quad z(1,t) = \eta(t)$$

$$z(x,0) = \gamma(x). \tag{1}$$

We assume that φ and η can be differentiated infinitely often.

First the notions x-external- and x-internal solution will be considered physically. The x-external solution $z_e(x,t)$ is the solution which is forced upon the physical system from the outside via the boundary conditions. So the external solution does satisfy

$$\frac{\partial z_e}{\partial t} - \frac{\partial^2 z_e}{\partial x^2} = 0$$

$$z_e(0,t) = \varphi(t), \quad z_e(1,t) = \eta(t). \tag{2}$$

Now we consider two known special cases of the notion "x-external solution", namely "steady state solution" for φ,η constant and "periodic steady state solution" for φ,η periodic. The stability of a problem (1) guarantees that for large t the solution $z(x,t)$ becomes also constant respectively periodic. A common property of both cases is that being constant or periodic of $\varphi,\varphi^{(1)},\ldots;\eta,\eta^{(1)},\ldots$ corresponds with a constant or periodic solution for large t and that we can build up this asymptotic solution linearly from $\varphi,\varphi^{(1)},\ldots;\eta,\eta^{(1)},\ldots$

For a problem with constant or periodic boundary conditions the difference between the solution of the given problem and the steady state or periodic steady state solution is often called "transient". Physically the x-external- and x-internal solution are generalizations of the concepts "(periodic)" steady state solution" respectively "transient" in the case that the boundary conditions depend arbitrarily on t. The x-external solution has to be fixed uniquely by φ and η. The mentioned common property of the above special cases of φ and η suggests, that it must be possible in general to build up the x-external solution linearly from φ,η and their derivatives. This agrees completely with the definition of the x-external solution, as given in paragraph 1.1.

There is also a t-external solution. However, this solution makes no sense for a physical system described by (1). In the next of this chapter "external solution" will always mean x-external solution (only in paragraph 2.8 and 2.9 some attention is paid to the t-external solution).

Now we will consider the following mathematical formulation of the external solution

$$z_e(x,t) = \sum_{k=0}^{\infty} \left\{g_k(x)\varphi^{(k)}(t) + g_k(1-x)\eta^{(k)}(t)\right\} . \tag{3}$$

We remark that the k^{th} term of this series contains only the k^{th} derivatives of φ and η.

First we will determine the functions $g_k(x)$. Series (3) satisfies the boundary conditions of (2) independent on φ and η, only if

$$g_0(0) = 1;\ g_k(0) = 0,\ k \geqq 1$$

$$g_k(1) = 0,\ k \geqq 0. \tag{4}$$

For the sake of convenience we will take $\eta(t) \equiv 0$. Substituting formally (3) in (2) (supposing (3) can be differentiated term by term) we obtain

$$g_0^{(2)}(x)\varphi(t) + \sum_{k=1}^{\infty} \left\{g_{k-1}(x) - g_k^{(2)}(x)\right\} \varphi^{(k)}(t) = 0. \tag{5}$$

This holds for arbitrary φ, only if the functions $g_k(x)$ satisfy the following infinite system of homogeneous ordinary differential equations

$$g_0^{(2)}(x) = 0$$

$$g_k^{(2)}(x) - g_{k-1}(x) = 0,\ k \geqq 1. \tag{6}$$

From (4) and (6) we have

$$g_0(x) = 1 - x. \tag{7}$$

Obviously the functions $g_k(x)$, $k \geqq 1$, satisfy the recurrent relation

$$g_k^{(2)}(x) - g_{k-1}(x) = 0;\ g_k(0) = g_k(1) = 0,\ k \geqq 1, \tag{8}$$

which has the form of an ordinary differential-difference equation with vanishing boundary conditions in $x = 0$ and $x = 1$. From (7) the functions $g_k(x)$ can be found successively by repeated integration. In this way we find for $g_1(x)$

$$g_1(x) = -\frac{1-x}{6} + \frac{(1-x)^3}{6} . \tag{9}$$

It can easily be shown that each function $g_k(x)$ is an odd polynomial in $(1-x)$ of degree $(2k+1)$.

Another formulation of the functions $g_k(x)$ can be found in the following way. The polynomials $g_k(x)$, $k \geqq 1$ form the solution of a boundary- and initial value problem

$$g_k^{(2)}(x) - g_{k-1}(x) = 0,$$

$$g_o(x) = 1 - x$$

$$g_k(0) = g_k(1) = 0, \ k \geqq 1. \tag{10}$$

Because the system equation of (10) as well as the conditions on the x-boundaries are homogeneous, the x-external solution of (10) vanishes. This means that the solution of (10) is identically equal to the x-internal solution. The last one can be written as an infinite series of x-elementary solutions. Therefore first the x-elementary solutions of (10) will be determined. By definition a solution of the system equation is a x-elementary solution, if it is an eigenfunction of $\frac{\Delta}{\Delta k}$ and if it vanishes for $x = 0$ and $x = 1$. So the x-elementary solutions must be of the form

$$\rho^k . (\text{function of } x), \ \rho = \text{constant}. \tag{11}$$

Elaboration gives the x-elementary solutions as

$$\frac{(-1)^k}{(j\pi)^{2k}} \sin j\pi x, \ j = 1,2,\ldots . \tag{12}$$

Expression (12) makes attractive the writing of $g_1(x)$ as a Fourier-series, containing only sinusterms. From (9) we derive

$$g_1(x) = -2 \sum_{j=1}^{\infty} \frac{1}{(j\pi)^2} \frac{\sin j\pi x}{j\pi}, \ 0 \leqq x \leqq 1. \tag{13}$$

Regarding (13) term by term, we obtain from (12) and (13)

$$g_k(x) = (-1)^k 2 \sum_{j=1}^{\infty} \frac{1}{(j\pi)^{2k}} \frac{\sin j\pi x}{j\pi}, \ 0 \leqq x \leqq 1. \tag{14}$$

Series (13) may be treated term by term, because the obtained series (14) is absolutely and uniformly convergent in $0 \leqq x \leqq 1$, if $k \geqq 1$.

From (14) an important limit property of the functions $g_k(x)$ can be derived in the following way. We have

$$\frac{\left|\sum\limits_{j=2}^{\infty} \frac{1}{(j\pi)^{2k}} \frac{\sin j\pi x}{j\pi}\right|}{\left|\frac{1}{\pi^{2k}} \frac{\sin \pi x}{\pi}\right|} \leqq \sum_{j=2}^{\infty} \frac{1}{j^{2k}} \left|\frac{\sin j\pi x}{j \sin \pi x}\right|. \tag{15}$$

Applying the unequality $\left|\frac{\sin j\pi x}{j \sin \pi x}\right| \leqq 1$ to (15) we can derive that for each $\varepsilon > 0$ a number K_1 can be found such that

$$\frac{\left|\sum_{j=2}^{\infty} \frac{1}{(j\pi)^{2k}} \frac{\sin j\pi x}{j\pi}\right|}{\left|\frac{1}{\pi^{2k}} \frac{\sin \pi x}{\pi}\right|} < \varepsilon, \text{ if } k > K_1. \tag{16}$$

From (14) and (16) we obtain

$$(1 - \varepsilon)\left|\sin \pi x\right| \frac{2}{\pi^{2k+1}} < \left|g_k(x)\right| < (1 + \varepsilon)\left|\sin \pi x\right| \frac{2}{\pi^{2k+1}}, \text{ if } k > K_1. \tag{17}$$

From (14) and (17) it follows

$$\lim_{k \to \infty} \frac{(-1)^k g_k(x)\pi^{2k+1}}{2 \sin \pi x} = 1, \; 0 \leqq x \leqq 1. \tag{18}$$

Now we will determine, when series (3) may be differentiated term by term. Suppose that $\eta(t) \equiv 0$ and that $\varphi(t)$ satisfies in the interval $[t_1, t_2]$

$$\left|\varphi^{(k)}(t)\right| \leqq c(\text{onstant}) \cdot a^k;\; a < \pi^2, \text{ if } k > K_2. \tag{19}$$

From (17) and (19) we have

$$\left|g_k(x)\varphi^{(k)}(t)\right| \leqq c(1 + \varepsilon)\, \frac{2}{\pi} \left(\frac{a}{\pi^2}\right)^k, \text{ if } k > \operatorname{Max}(K_1, K_2). \tag{20}$$

The right hand member of (20) is the k^{th} term of a geometric series with ratio $\frac{a}{\pi^2}$.

Hence, series (3) is absolutely and uniformly convergent in $0 \leqq x \leqq 1$, $t_1 \leqq t \leqq t_2$, if for large k $\left|\varphi^{(k)}(t)\right|$ as well as $\left|\eta^{(k)}(t)\right|$ (since we may interchange φ and η) is smaller than a constant times a^k, $a < \pi^2$.

In nearly the same way as in the case of convergence, it can be proved that series (3) diverges in general if $\left|\varphi^{(k)}(t)\right|$ and/or $\left|\eta^{(k)}(t)\right|$ are larger than a constant times π^{2k}.

So we have proved that the series $\sum_{k=0}^{\infty} \{g_k(x)\varphi^{(k)}(t) + g_k(1 - x)\eta^{(k)}(t)\}$ is a solution of problem (2), if φ and η satisfy (19). This means that it is in fact the x-external solution of problem (1), as defined in paragraph 1.1.

Finally we will show that the x-external solution (3) confirms our assertion that it is the generalization of the concepts "steady state" and "periodic steady state solution". Again we take $\eta(t) \equiv 0$.

The case $\varphi(t)$ = constant is trivial, because then series (3) contains only one term which is equal to the steady state solution.

The case of periodic boundary conditions will only be considered for $\varphi(t) = e^{ibt}$, $-\pi^2 < b < \pi^2$. For large t the behaviour of z with respect to t will be like e^{ibt} for each x in $0 \leqq x < 1$, because the considered problem is stable. This means that for large t the solution $z(x,t)$ is a function of the form $f(x) \cdot e^{ibt}$ which satisfies (2) for $\varphi = e^{ibt}$, $\eta = 0$. Substitution of $z_e(x,t) = f(x)e^{ibt}$ in (2) gives

$$z_e(x,t) = \frac{\sinh\sqrt{ib}(1-x)}{\sinh\sqrt{ib}}\, e^{ibt}. \tag{21}$$

However we must obtain the same function $z_e(x,t)$ from (3) and so it must hold

$$\left\{\sum_{k=0}^{\infty} g_k(x)(ib)^k\right\} \sinh\sqrt{ib} = \sinh\left\{\sqrt{ib}(1-x)\right\}. \tag{22}$$

Expansion of both members of (22) in a power series of $\sqrt{ib}$ and equalizing coefficients of $(\sqrt{ib})^{2n+1}$ gives

$$g_n(x) + \frac{g_{n-1}(x)}{3!} + \dots + \frac{g_o(x)}{(2n+1)!} = \frac{(1-x)^{2n+1}}{(2n+1)!}. \tag{23}$$

In paragraph 2.2 it will be shown in a simple way that this recurrent relation between the polynomials $g_k(x)$ is true. So it may be concluded that the x-external solution agrees with the steady state- and periodic steady state solution in the case of constant respectively periodic boundary conditions.

2.2. Expansion of a function in its even derivatives in 2 points

For a problem (2.1.1) the even derivatives with respect to x of $z_e(x,t)$ in $x=0$ and $x=1$ can be obtained for each t from φ and η with the help of the system equation of (2.1.1) as $\partial^2 z_e(0,t)/\partial x^2 = \varphi^{(1)}(t)$, $\partial^4 z_e(0,t)/\partial x^4 = \varphi^{(2)}(t)$, etc. This means that it is allowed to replace $\varphi^{(k)}(t)$ and $\eta^{(k)}(t)$, $k=0,1,\dots$ in series (2.1.3) by $\partial^{2k} z_e(0,t)/\partial x^{2k}$ and $\partial^{2k} z_e(1,t)/\partial x^{2k}$. Then we obtain

$$z_e(x,t) = \sum_{k=0}^{\infty}\left\{g_k(x)\,\frac{\partial^{2k} z_e(0,t)}{\partial x^{2k}} + g_k(1-x)\,\frac{\partial^{2k} z_e(1,t)}{\partial x^{2k}}\right\}. \tag{1}$$

We observe that in this expression only z_e and its derivatives to x appear. This suggests that a function $f(x)$ can be expanded in its even derivatives in $x=0$ and $x=1$ as

$$f(x) = \sum_{k=o}^{\infty} \left\{ g_k(x) f^{(2k)}(0) + g_k(1 - x) f^{(2k)}(1) \right\} . \tag{2}$$

We call a series of form (2) a 2 points expansion [2], [3]. Assuming that a seriesexpansion of form (2) exists for $f(x)$ as well as for its even derivatives, it can easily be deduced formally that the functions $g_k(x)$ in (2) are indeed the same functions as we have already met before.

Substituting successively $f(x) = 1, x, x^2, x^3$, the following relations arise

$$\begin{aligned}
1 &= g_o(x) + g_o(1 - x) \\
x &= g_o(1 - x) \\
x^2 &= 2g_1(x) + g_o(1 - x) + 2g_1(1 - x) \\
x^3 &= g_o(1 - x) + 6g_1(1 - x).
\end{aligned} \tag{3}$$

The first two relations yield $g_o(1-x) = x$ and from the 3^{rd} and 4^{th} relation it follows $g_1(1-x) = (x^3 - x)/6$. Continuing in this way we can calculate all functions $g_k(x)$ one after another. The substitution of x^{2n} and x^{2n+1} in (2) results into the following recurrent relations

$$\begin{aligned}
\frac{x^{2n}}{(2n)!} &= g_n(x) + \frac{g_n(1-x)}{0!} + \frac{g_{n-1}(1-x)}{2!} + \ldots + \frac{g_o(1-x)}{(2n)!} , \\
\frac{x^{2n+1}}{(2n+1)!} &= \frac{g_n(1-x)}{1!} + \frac{g_{n-1}(1-x)}{3!} + \ldots + \frac{g_o(1-x)}{(2n+1)!} .
\end{aligned} \tag{4}$$

As for the convergence of series (2) we will refer to what has been said in the last paragraph about the convergence of series (2.1.3), because in this series we may replace $\varphi^{(k)}(t)$ and $\eta^{(k)}(t)$ by $f^{(2k)}(0)$ and $f^{(2k)}(1)$.

We remark that in the derivation of (2.1.3) we have only used in fact that

a. the boundaries $x = 0$ and $x = 1$ are independent of t
b. on these boundaries there are unique relations between the even derivatives with respect to x and the derivatives with respect to t of $z_e(x,t)$.

So it will be clear that series (2) will also be the base for the external solution of other partial differential equations, for instance $\partial^2 z/\partial t^2 - \partial^2 z/\partial x^2 = 0$ or $\partial^2 z/\partial x^2 + \partial^2 z/\partial y^2 = 0$, as is shown in [4].

Omitting the terms of series (2) for $k \geqq n$ gives an approximation of $f(x)$ in the form of a polynomial of the degree $(2n + 1)$. This polynomial $F_{2n+1}(x)$ has the same even derivatives as $f(x)$ in $x = 0$ and $x = 1$ up to and including the order 2n.

Now we will consider the remainder $R_{2n+2}(x) = f(x) - F_{2n+1}(x)$.

$$R_{2n+2}(x) = \sum_{k=n+1}^{\infty} \left\{ g_k(x) f^{(2k)}(0) + g_k(1 - x) f^{(2k)}(1) \right\} . \tag{5}$$

We shall show formally that $R_{2n+2}(x)$ can be expressed into $f^{(2n+2)}(x)$. Applying (2), we can write $f^{(2n+2)}(x)$ as

$$f^{(2n+2)}(x) = \sum_{k=0}^{\infty} \left\{ g_k(x) f^{(2n+2+2k)}(0) + g_k(1-x) f^{(2n+2+2k)}(1) \right\} . \tag{6}$$

The right hand members of (5) and (6) have the same form; only the orders of corresponding derivatives of f differ an amount 2k. We are able to diminish the order of the derivatives of f in (6) with the help of the following recurrent relation

$$g_{k+1}(x) = - \int_0^1 K(x,\mu) g_k(\mu) d\mu , \tag{7}$$

where $K(x,\mu)$ is the known kernel:

$$\begin{aligned} K(x,\mu) &= \mu(1 - x), \quad 0 \leqq \mu \leqq x \\ &= x(1 - \mu), \quad x \leqq \mu \leqq 1. \end{aligned} \tag{8}$$

Relation (7) can easily be proven by integrating its right hand member by parts, using the property ${g_{k+1}}^{(2)}(x) = g_k(x)$. Integrating both members of (6) over the x-interval $[0,1]$ after multiplication by $K(x,\mu)$ gives

$$- \int_0^1 K(x,\mu) f^{(2n+2)}(\mu) d\mu = \sum_{k=1}^{\infty} \left\{ g_k(x) f^{(2n+2k)}(0) + g_k(1-x) f^{(2n+2k)}(1) \right\} . \tag{9}$$

Comparing (9) and (6) with (5) leads to the conclusion that the right hand member of (6) passes into that of (5) if we perform the above manner of integration $(n+1)$ times. In this way we obtain

$$R_{2n+2}(x) = (-1)^{n+1} \int_0^1 K_{n+1}(x,\mu) f^{(2n+2)}(\mu) d\mu, \tag{10}$$

where $K_{n+1}(x,\mu)$ is the n times iterated kernel of $K(x,\mu)$. Because $K_{n+1}(x,\mu) \geqq 0$ in the x-interval $[0,1]$ for each n, we can apply the first mean-value theorem about integrals to (10), provided $f^{(2n+2)}(x)$ is continuous in $[0,1]$. Then we obtain

$$R_{2n+2}(x) = (-1)^{n+1}\ f^{(2n+2)}(\xi)\int_0^1 K_{n+1}(x,\mu)d\mu, \quad 0 \leqq \xi \leqq 1. \tag{11}$$

It is well-known that $K_{n+1}(x,\mu)$ can be written as (see also paragraph 9)

$$K_{n+1}(x,\mu) = 2\sum_{r=1}^{\infty}\frac{\sin r\pi x \sin r\pi\mu}{(r\pi)^{2n+2}}. \tag{12}$$

Substituting (12) into (11) we may interchange the order of integration and summation because series (12) is absolutely and uniformly convergent. Elaboration gives

$$R_{2n+2}(x) = f^{(2n+2)}(\xi).\ 4(-1)^{n+1}\sum_{r=0}^{\infty}\frac{\sin(2r+1)\pi x}{\{(2r+1)\pi\}^{2n+3}},\ 0 \leqq \xi \leqq 1. \tag{13}$$

By virtue of (2.1.14) we can also write (13) as

$$R_{2n+2}(x) = \left\{g_{n+1}(x) + g_{n+1}(1-x)\right\} f^{(2n+2)}(\xi), \qquad 0 \leqq \xi \leqq 1. \tag{14}$$

We observe that in this form the remainder $R_{2n+2}(x)$ resembles very much the Lagrange remainder of the Mac Laurin-series. The only difference is that $\frac{x^{2n+2}}{(2n+2)!}$ has been replaced by $\{g_{n+1}(x) + g_{n+1}(1-x)\}$.

For large values of n it follows from (2.1.18) that

$$R_{2n+2}(x) \approx (-1)^{n+1}\frac{4\sin\pi x}{\pi^{2n+3}} f^{(2n+2)}(\xi), \qquad 0 \leqq \xi \leqq 1. \tag{15}$$

In the last two paragraphs we have found several properties of the functions $g(x)$. However these functions can be conceived to be completely known, because they are closely related to Bernoulli polynomials:

$$g_k(x) = \frac{2^{2k+1}}{(2k+1)!} B_{2k+1}(1 - \tfrac{x}{2}). \qquad ^{*)} \tag{16}$$

*) I am indebted to Prof. S.C. van Veen for pointing out this relation.

2.3. External solution of a "parabolic" difference problem

Again we consider problems of type (2.1.1) but now, however, for the case in which the differential equation has been replaced by a difference equation. The solution of the obtained difference problem is only defined for discrete values of x, $x_m = m\Delta x$ and discrete values of t, $t_n = n\Delta t$.

Denoting the difference quotient which arises instead of $\partial z/\partial t$ by $\Delta z/\Delta t$, etc. the problems to be considered are

$$\frac{\Delta z(x_m,t_n)}{\Delta t} - \frac{\Delta^2 z(x_m,t_n)}{\Delta x^2} = 0$$

$$z(0,t_n) = \varphi(t_n),\; z(1,t_n) = \eta(t_n)$$

$$z(x_m,0) = \gamma(x_m). \tag{1}$$

We see that in (1) the x-interval $[0,1]$ is divided in M subintervals, $M = 1/\Delta x$. By definition the x-external solution has to be a linear composition of φ, η and their difference quotients with respect to t. In analogy with paragraph 2.1 we only consider the external solution written as

$$z_e(x_m,t_n) = \sum_{k=0}^{\infty} \left\{ g_k^*(x_m)\varphi^{)k(}(t_n) + g_k^*(1-x_m)\eta^{)k(}(t_n) \right\}, \tag{2}$$

where for shortness sake we have used the notation $\Delta^k f/\Delta t^k = f^{)k(}$.

The functions $g_k^*(x_m)$ appear to be the solution of the following boundary- and initial value problem

$$g_k^{*)2(}(x_m) - g_{k-1}^*(x_m) = 0$$

$$g_k^*(0) = g_k^*(1) = 0, \quad k \geqq 1$$

$$g_o^*(x_m) = 1 - x_m\,. \tag{3}$$

In order to be able to obtain a solution of (3) we restrict ourselves to the case in which

$$g_k^{*)2(}(x_m) = \frac{\delta^2 g_k^*(x_m)}{(\Delta x)^2} = \frac{g_k^*(x_{m-1})-2g_k^*(x_m)+g_k^*(x_{m+1})}{(\Delta x)^2}\,. \tag{4}$$

Substituting this in (3) gives a recurrent relation between $g_k^*(x_m)$ and $g_{k-1}^*(x_m)$, $m = 1,\ldots,M-1$, in the form of a simple set of linear equations, in which $g_k(x_m)$ occurs implicitly. Denoting the matrix of this system by A and introducing a vector $\underline{v}_k$ such that its m^{th} component is equal to $g_k^*(x_m)$ we can write the set of equations as

$$A\underline{v}_k - \underline{v}_{k-1} = \underline{0}, \qquad k \geqq 0$$

$$\underline{v}_{-1} = (-\frac{1}{(\Delta x)^2}, 0, \ldots, 0). \tag{5}$$

This is an initial value problem. First we will determine the elementary solutions of the system equation of (5). By definition these are those eigenelements of A and $\frac{\Delta}{\Delta k}$, which also satisfy the system equation. If we denote the eigenvectors and eigenvalues of A by $\underline{u}_j$ and μ_j, the elementary solutions prove to be

$$(\frac{1}{\mu_j})^k \underline{u}_j, \qquad j = 1, \ldots, M-1. \tag{6}$$

It is well-known that the eigenvectors and eigenvalues of the tri-diagonal matrix A are

$$\underline{u}_j = (u_{j1}, \ldots, u_{j,M-1}), \qquad u_{jm} = \sin j\pi x_m ;$$

$$\mu_j = -\frac{2(1-\cos j\pi\Delta x)}{(\Delta x)^2} ; \qquad j = 1, \ldots, M-1 . \tag{7}$$

Obviously the eigenvectors $\underline{u}_j$ are linearly independent, so the general solution of the system equation of (5) is

$$\sum_{j=1}^{M-1} K_j (\frac{1}{\mu_j})^k \underline{u}_j . \tag{8}$$

From this the solution of (6) arises, if the constants K_j are chosen such that

$$\underline{v}_{-1} = \sum_{j=1}^{M-1} \mu_j K_j \underline{u}_j . \tag{9}$$

Denoting these values of K_j by b_j we have the functions $g_k^*(x_m)$ as

$$g_k^*(x_m) = \sum_{j=1}^{M-1} b_j (\frac{1}{\mu_j})^k \sin j\pi x_m, \qquad m = 1, \ldots, M-1, \quad k \geqq 0. \tag{10}$$

It can easily be seen that (10) also holds in $x_m = 0$ and $x_m = 1$, excepted in $x_m = 0$ for $k = 0$.

We remark that (10) can also be obtained directly from (3) in an analogous way as has been done in paragraph 2.1 for the determination of $g_k(x)$. Then we would have found the x-elementary solutions of (3) as

$$\left\{-\frac{(\Delta x)^2}{2(1-\cos j\pi\Delta x)}\right\}^k \sin j\pi x_m, \quad j = 1,\ldots,M-1, \quad m = 0,\ldots,M. \tag{11}$$

Of course from this we would finally have obtained (10) again. We call attention to one difference between both above methods of determining $g_k^*(x_m)$. While in the first method eigenvectors and eigenvalues of A must be known, they arise in the second method as a subsidiary result from (11).

Applying the difference operator δ^2 to an arbitrary odd polynomial $Q_1(x)$ results again in an odd polynomial $Q_2(x)$, whose degree is decreased by 2. For given coefficients of $Q_2(x)$ it holds that all the coefficients of $Q_1(x)$ excepted that of x are uniquely fixed. We can choose this coefficient such that $Q_1(1) = 0$. Hence starting from an odd polynomial of the first degree, $P_0(1 - x) = x$, which equals $g_1^*(1 - x_m)$ in $x = x_m$, we can construct odd polynomials in x, $P_k(1 - x)$, of degree $(2k + 1)$, $k = 1,2,$etc., which satisfy (3) for each $0 \leqq x \leqq 1$. The first three of these polynomials are

$$P_0(1 - x) = x$$

$$P_1(1 - x) = -\frac{1}{6}x + \frac{1}{6}x^3$$

$$P_2(1 - x) = \frac{7+5(\Delta x)^2}{360}x - \frac{2+(\Delta x)^2}{72}x^3 + \frac{1}{120}x^5. \tag{12}$$

We observe that only the coefficients of the polynomials $P_k(x)$ contain the parameter Δx for $k \geqq 2$ and that $P_0(x) = g_0(x)$ and $P_1(x) = g_1(x)$. From the last equalities we have

$$g_0^*(x_m) = g_0(x_m) \tag{13}$$

$$g_1^*(x_m) = g_1(x_m). \tag{14}$$

Now we will show that from these relations a seriesexpansion of $\frac{\cos y}{\sin^3 y}$ can be derived. Substituting (10) and (2.1.14) in (13) and (14) we obtain

$$\sum_{j=1}^{M-1} b_j \sin j\pi x_m = 2\sum_{j=1}^{\infty}\frac{\sin j\pi x_m}{j\pi}, \quad m = 1,\ldots,M \tag{15}$$

$$\sum_{j=1}^{M-1} \frac{b_j}{\mu_j} \sin j\pi x_m = -2 \sum_{j=1}^{\infty} \frac{\sin j\pi x_m}{(j\pi)^3}, \quad m = 0,\dots,M. \tag{16}$$

For each x_m in [0,1] it holds

$$\sin \{(\frac{2p}{\Delta x} \pm j)\pi x_m\} = \pm \sin j\pi x_m, \quad p = 0,1,\dots . \tag{17}$$

By means of this relation we can reduce the right hand members of (15) and (16) to finite series. Then equating the coefficients of $\sin j\pi x_m$ in the left and right hand member of (15) and of (16) gives, if we write $j\Delta x/2 = y$,

$$b_j = \frac{\Delta x}{\pi}[\frac{1}{y} + \sum_{r=1}^{\infty} (\frac{1}{y-r} + \frac{1}{y+r})] = \Delta x \operatorname{cotg} \pi y \tag{18}$$

$$-\frac{b_j}{\mu_j} = \frac{(\Delta x)^3}{4\pi^3}[\frac{1}{y^3} + \sum_{r=1}^{\infty}\{\frac{1}{(y-r)^3} + \frac{1}{(y+r)^3}\}] . \tag{19}$$

From these equalities by substitution of (7) into (19) we can deduce

$$\frac{\cos \pi y}{\sin^3 \pi y} = \frac{1}{\pi^3}[\frac{1}{y^3} + \sum_{r=1}^{\infty}\{\frac{1}{(y-r)^3} + \frac{1}{(y+r)^3}\}] . \tag{20}$$

In the special case $y = \frac{1}{4}$, this series reduces to a well-known series of Euler

$$\frac{\pi^3}{32} = 1 - \frac{1}{3^3} + \frac{1}{5^3} - \frac{1}{7^3} + \frac{1}{9^3} - \dots . \tag{21}$$

With the obtained knowledge about $g_k^*(x_m)$ we are able to determine the conditions for the convergence of series (2). From (7) and (10) it can easily be seen that

$$\lim_{k\to\infty} \frac{(-1)^k}{b_1}\{\frac{2(1-\cos \pi\Delta x}{(\Delta x)^2}\}^k \frac{g_k^*(x_m)}{\sin \pi x_m} = 1, \quad m = 0,\dots,M. \tag{22}$$

From the great resemblance with (2.1.18) we can conclude immediately that series (2) is absolutely convergent in $0 \leqq x_m \leqq 1$, $\tau_1 \leqq t_n \leqq \tau_2$, if for large k $|\varphi^{(k)}(t_n)|$ as well as $|\eta^{(k)}(t_n)|$ are smaller than a constant times a^k, $a < 2(1 - \cos \pi\Delta x)/(\Delta x)^2$.

In the next we will show that the external solution (2) of the difference problem approaches the external solution (2.1.3) of the differential

problem, if Δt, $\Delta x \to 0$. First we shall prove that each term of (2) tends to the corresponding term of (2.1.3), if Δt, $\Delta x \to 0$ and x, t are constant. Because the choice of x has to be at will, we consider
$h_k(x) = \sum_{j=1}^{M-1} b_j(\frac{1}{\mu_j})^k \sin j\pi x$, in which x is continuous, instead of $g_k^*(x_m)$.

We divide $h_k(x)$ as well as the Fourierseries of $g_k(x)$ into the sum of the first N terms, denoted respectively by S_N^* and S_N, and the remaining series, denoted respectively by R_N^* and R_N. Then we have

$$|h_k(x) - g_k(x)| \leqq |S_N^* - S_N| + |R_N^* - R_N|. \tag{23}$$

From (7) and (8) it can easily be understood that S_N^* tends to S_N, if $\Delta x \to 0$. For R_N^* we have

$$|R_N^*| < \sum_{j=N+1}^{M-1} |b_j||\frac{1}{\mu_j}|^k . \tag{24}$$

Because $j\Delta x < 1$, we have from (18)

$$|b_j| < \frac{2}{j\pi} \tag{25}$$

and from (7)

$$|\frac{1}{\mu_j}| < \frac{1}{(j\pi)^2\{1-\frac{(j\pi\Delta x)^2}{12}\}} < \frac{1}{(j\pi)^2(1-\frac{\pi^2}{12})} . \tag{26}$$

From this we see that for $k \geqq 1$ the right hand member of (24) is convergent if $M \to \infty$. Hence $|R_N^*|$ can be made arbitrarily small by choosing M and N sufficiently large. The same holds for $|R_N|$, because series (2.1.14) is absolutely convergent for $k \geqq 1$. From the above we conclude that $|h_k(x) - g_k(x)|$ can be made arbitrarily small in $0 \leqq x \leqq 1$ by choosing Δx small enough. This means that $h_k(x)$ converges uniformly to $g_k(x)$ in $[0,1]$, if $\Delta x \to 0$. Because $\varphi^{)k(}(t) \to \varphi^{(k)}(t)$ if $\Delta t \to 0$, it has been shown now that each term of (2) tends to the corresponding term of (2.1.3), if Δt, $\Delta x \to 0$.

The proof that the external solution (2) approaches the external solution (2.1.3) if Δt, $\Delta x \to 0$ is nearly the same as the above one and so will not be repeated here.

We observe that again series (2) relates to a 2 points expansion of a function $f(x)$, being

$$f(x) = \sum_{k=o}^{\infty} \{P_k(x) \frac{\delta^{2k}f(0)}{(\Delta x)^{2k}} + P_k(1-x) \frac{\delta^{2k}f(1)}{(\Delta x)^{2k}}\} . \tag{27}$$

Truncating this series after the n^{th} term gives a polynomial approximation of $f(x)$ that has the same even difference quotients in $x = 0$ and $x = 1$ up to and including order $2n$.

In conclusion we will determine a generating function of the polynomials $P_k(x)$. First we remark that series (2) also satisfies the system equation and the x-boundary conditions of (1), if $g_k^*(x_m)$ and $g_k^*(1-x_m)$ are replaced by $P_k(x)$ and $P_k(1-x)$, where again x is continuous. Then taking $\eta(t_n) \equiv 0$ and $\varphi(t_n)$ such that $\varphi^{)1(}(t_n) = b\varphi(t_n)$, b = constant, we obtain from (2)

$$z_e(x,t_n) = f(x)\varphi(t_n); \quad f(x) = \sum_{k=o}^{\infty} P_k(x)\ b^k . \tag{28}$$

We can also determine $f(x)$, substituting directly $z_e(x,t_n) = f(x)\varphi(t_n)$ in the system equation and the x-boundary conditions of (1). This results into

$$f(x) = \frac{\sinh A(1-x)}{\sinh A}, \quad A = \frac{1}{\Delta x} \text{ areacosh } \{1 + \frac{b(\Delta x)^2}{2}\} . \tag{29}$$

Hence

$$\frac{\sinh A(1-x)}{\sinh A} = \sum_{k=o}^{\infty} P_k(x)\ b^k . \tag{30}$$

2.4. The x-internal solution of a diffusion problem

First again we will consider problems of type (2.1.1). By definition the x-internal solution $z_i(x,t)$ is the difference between the solution of the given problem and the x-external solution, so $z_i(x,t)$ has to satisfy:

$$\frac{\partial z_i}{\partial t} - \frac{\partial^2 z_i}{\partial x^2} = 0$$

$$z_i(0,t) = z_i(1,t) = 0$$

$$z_i(x,0) = \gamma(x) - z_e(x,0) . \tag{1}$$

We will determine $z_i(x,t)$ as a series of x-elementary solutions of (1), being $e^{\lambda_j t}\alpha_j(x)$, provided $\alpha_j(x)$ satisfies

$$\alpha_j^{(2)}(x) - \lambda_j\alpha_j(x) = 0, \quad \alpha_j(0) = \alpha_j(1) = 0. \tag{2}$$

Only if $\lambda_j = -(j\pi)^2$, $j = 1,2,\dots$, we obtain a nontrivial solution of (2) as $\alpha_j(x) = \sin j\pi x$. So the x-elementary solutions of (1) are

$$e^{-(j\pi)^2 t} \sin j\pi x, \quad j = 1,2,\dots . \tag{3}$$

Hence formally we have, writing $z_i(x,0) = \sum_{j=1}^{\infty} c_j \sin j\pi x$, that

$$z_i(x,t) = \sum_{j=1}^{\infty} c_j \sin j\pi x \, e^{-(j\pi)^2 t} . \tag{4}$$

From (4) we see that the internal solution tends to zero if $t \to \infty$. This means that the corresponding physical system is stable. This stability only includes that all x-elementary solutions approach zero if $t \to \infty$. However the elementary solutions do not tend to zero in the same extent. For relative comparison of the elementary solutions there is need of a measure of stability. For each elementary solution it holds that the ratio of the values at times t and $(t-1)$ is constant. From (3) we have the growthfactor g_j of the j^{th} elementary solution as

$$g_j = e^{-(j\pi)^2} . \tag{5}$$

The inverse s_j of the growthfactor g_j can serve as a measure of stability. We also need some measure of stability for the internal solution itself, but the ratio of its values at t and $(t-1)$ is not a constant. However we observe that for large t this ratio tends to a constant, because then the internal solution is approximately equal to $c_1 \sin \pi x \, e^{-\pi^2 t}$. So it is sensible to introduce a measure of asymptotic stability, being e^{π^2}.

With respect to later on we will also regard here the case (1), in which the equation $\partial z/\partial t - \partial^2 z/\partial x^2 = 0$ is replaced by $\partial z/\partial t - \partial^2 z/\partial x^2 - pz = 0$. Then we find the elementary solutions as

$$\sin j\pi x \, e^{\{p-(j\pi)^2\}t}, \quad j = 1,2,\dots . \tag{6}$$

Hence this physical system is only stable if $p < \pi^2$.

2.5. Internal solution of a "parabolic" difference problem

In the next we will determine the internal solution $z_i(x_m,t_n)$ of problem (2.3.1), which has to satisfy

$$\frac{\Delta z_i(x_m,t_n)}{\Delta t} - \frac{\Delta^2 z_i(x_m,t_n)}{(\Delta x)^2} = 0$$

$$z_i(0,t_n) = z_i(0,t_n) = 0$$

$$z_i(x_m,0) = \gamma(x_m) - z_e(x_m,0). \tag{1}$$

Again we consider (1) only in the case that $\frac{\Delta^2}{(\Delta x)^2}$ is taken equal to $\frac{\delta^2}{(\Delta x)^2}$. Then the x-elementary solutions of (1) prove to be

$$\sin j\pi x_m \, g^n_{j,\Delta t} \, , \qquad j = 1,2,\dots , \tag{2}$$

where $t_n = n\Delta t$ and $g_{j,\Delta t}$ represents the growthfactor per step Δt of the j^{th} elementary solution. Writing the initial condition $z_i(x_m,0)$ as $\sum_{j=1}^{M-1} d_j \sin j\pi x_m$, we have

$$z_i(x_m,t_n) = \sum_{j=1}^{M-1} d_j \sin j\pi x_m \, g^n_{j,\Delta t} \, . \tag{3}$$

The value of $g_{j,\Delta t}$ depends amply on the choice of the difference operator $\frac{\Delta}{\Delta t}$. As examples we will consider the difference equations of Milne and of Laasonen

$$\frac{\delta_+ z(x_m,t_n)}{\Delta t} - \frac{\delta^2 z(x_m,t_n)}{(\Delta x)^2} = 0 \qquad \text{Milne,} \tag{4}$$

$$\frac{\delta_- z(x_m,t_n)}{\Delta t} - \frac{\delta^2 z(x_m,t_n)}{(\Delta x)^2} = 0 \qquad \text{Laasonen,} \tag{5}$$

where δ, δ_+ and δ_- are the central, forward and backward difference. In this order the growthfactor proves to be

$$g_{j,\Delta t} = 1 - \frac{2\Delta t}{(\Delta x)^2}(1 - \cos j\pi\Delta x) \tag{6}$$

$$g_{j,\Delta t} = \frac{1}{1+\frac{2\Delta t}{(\Delta x)^2}(1-\cos j\pi \Delta x)} \, . \tag{7}$$

For the difference equation of Milne the following remarks can be made. All growthfactors are smaller than +1 and by choosing Δt large enough each growthfactor can be made smaller than -1. For small values of Δt the 1st elementary solution has the smallest stability, but enlarging Δt changes this and $g_{M-1,\Delta t}$ is the growthfactor which becomes first smaller than -1. It holds $|g_{j,\Delta t}| \leqq 1$ for each j or in other words the internal solution is stable if

$$\frac{\Delta t}{(\Delta x)^2} \leqq \frac{1}{1+\cos \pi \Delta x} . \tag{8}$$

For the difference equation of Laasonen it holds $0 < g_{j,\Delta t} < 1$ for each j. So the internal solution is unconditionally stable. For each Δt and Δx the 1^{st} elementary solution has the smallest stability.

Finally we remark without proving it here that the internal solution of the considered difference problems approaches the internal solution of the corresponding differential problem, if $\Delta t, \Delta x \to 0$.

2.6. A more general seriesexpansion of the external solution

In paragraph 2.1 we have discussed one special seriesexpansion of the external solution, having a rather strong condition of convergence. We remark that this condition of convergence does not give any indication about the existence of the external solution. This can be shown for a special case of problem (2.1.1), namely $\varphi(t) = e^{at}$, $\eta(t) = 0$. Then all derivatives of $\varphi(t)$ are proportional to $\varphi(t)$ itself and so by definition the external solution can be determined directly by substituting $z_e(x,t) = r(x)\varphi(t)$ in (2.1.2). Regardless the value of a, the external solution proves to be

$$z_e(x,t) = \frac{\sinh\sqrt{a}(1-x)}{\sinh\sqrt{a}} e^{at}. \tag{1}$$

The same result can be obtained from (2.1.3), but only if $|a| < \pi^2$, because otherwise series (2.1.3) is not convergent.

In this paragraph we will consider a more general seriesexpansion of the external solution having a parameter such that the region of convergence can be shifted at will by means of this parameter.

Introducing a new dependent variable

$$u(x,t) = z(x,t)e^{-at}, \tag{2}$$

we obtain from (2.1.1), if we take $\eta(t) = 0$ and write $\alpha(t) = \varphi(t)e^{-at}$, that $u(x,t)$ has to satisfy

$$\frac{\partial u}{\partial t} \overset{?}{+} \frac{\partial^2 u}{\partial x^2} + au = 0$$

$$u(0,t) = \alpha(t), \ u(1,t) = 0$$

$$u(x,0) = \gamma(x). \tag{3}$$

In analogy with the choice of the seriesexpansion of $z_e(x,t)$ in paragraph 2.1, we will try to find a seriesexpansion of $u_e(x,t)$ of the form

$$u_e(x,t) = \sum_{k=0}^{\infty} G_k(x,a)\, \alpha^{(k)}(t). \tag{4}$$

Obviously $G_k(x,a)$ equals $g_k(x)$ for $a = 0$. Formal substitution of series (4) into (3) gives as recurrent relation between the functions $G_k(x,a)$, $k \geqq 1$

$$G_k^{(2)}(x,a) - a\, G_k(x,a) = G_{k-1}(x,a), \quad G_k(0) = G_k(1) = 0, \tag{5}$$

while $G_o(x,a)$ has to satisfy

$$G_o^{(2)}(x,a) - a\, G_o(x,a) = 0, \quad G_o(0) = 1, \quad G_o(1) = 0. \tag{6}$$

Assuming that (2) also holds for the external solutions, we obtain from (4), if $\alpha^{(k)}(t)$ is expressed in $\varphi(t)$ and its derivatives

$$z_e(x,t) = \sum_{k=0}^{\infty} G_k(x,a) \left\{ \sum_{j=0}^{k} \binom{k}{j} (-a)^{k-j} \varphi^{(j)}(t) \right\}. \tag{7}$$

Of course we have to verify that this seriesexpansion is true. However first we will determine the functions $G_k(x,a)$. The first two can easily be obtained from (5) and (6) as

$$G_o(x,a) = \frac{\sinh\sqrt{a}(1-x)}{\sinh\sqrt{a}} \tag{8}$$

$$G_1(x,a) = \frac{1}{2\sqrt{a}} \frac{\sinh\sqrt{a}x}{(\sinh\sqrt{a})^2} - \frac{1}{2\sqrt{a}} \frac{\cosh\sqrt{a}(1-x)}{\sinh\sqrt{a}}. \tag{9}$$

Again it will prove sensible to write the functions $G_k(x,a)$ as Fourierseries with only sin terms. This can most easily be done by writing first $G_1(x,a)$ as Fourierseries and after that by formally substituting one after another the functions $G_k(x,a)$ into (5). In this way we find

$$G_k(x,a) = (-1)^k\, 2 \sum_{j=1}^{\infty} \frac{j\pi}{\{(j\pi)^2 + a\}^{k+1}} \sin j\pi x. \tag{10}$$

We observe that treating these series formally is indeed allowed because they are all absolutely and uniformly convergent in $[0,1]$ for each value of a. For large values of k one of the terms of series (10) will become dominating each of the others. We abbreviate $(-1)^k 2j\pi/\{(j\pi)^2 + a\}^{k+1}$ as t_j and denote t_j of the dominating term by t_J. It holds that $|t_j/t_J| \to 0$ for $k \to \infty$ if

$$\left|\frac{(j\pi)^2+a}{(J\pi)^2+a}\right| > 1; \quad j \neq J, \quad j = 1,2,\ldots . \tag{11}$$

It can easily be derived that (11) is satisfied if

$$a_1 = -\frac{J^2+(J+1)^2}{2} < \operatorname{Re}\frac{a}{\pi^2} < -\frac{(J-1)^2+J^2}{2} = a_2, \quad J \neq 1 \text{ and}$$

$$-\frac{5}{2} < \operatorname{Re}\frac{a}{\pi^2}, \quad J = 1. \tag{12}$$

Now we will outline how to prove that the J^{th} term also dominates all other terms of (10) together, if a satisfies (12). We divide series (10) into three parts:

1) $p_1 = \sum_{j=1}^{J-1} t_j \sin j\pi x$

2) $p_2 = t_J \sin J\pi x$

3) $p_3 = \sum_{j=J+1}^{\infty} t_j \sin j\pi x$. (13)

Because the first part is a finite sum, we can conclude immediately that $\left|\frac{p_1}{t_J}\right| \to 0$, if $k \to \infty$. It can also be proven that $\left|\frac{p_3}{t_J}\right| \to 0$ if $k \to \infty$, but we omit this proof, because it is similar to the corresponding proof for the polynomials $g_k(x)$ in paragraph 2.1. Hence we have

$$\lim_{k\to\infty} \frac{(-1)^k G_k(x,a)}{t_J \sin J\pi x} = 1. \tag{14}$$

From this the condition of convergence of series (7) can be derived; again we suffice to refer to paragraph 2.1. It proves if a satisfies (12) that series (7) is absolutely and uniformly convergent in the x,t-region $0 \leqq x \leqq 1$, $t_1 \leqq t \leqq t_2$, if in this t-interval for large k $|\{\varphi(t)e^{-at}\}^{(k)}|$ as well as $|\{\eta(t)e^{-at}\}^{(k)}|$ are smaller than a constant.c^k, $c < |(J\pi)^2 + a|$. If this condition of convergence is satisfied, series (7) is a solution of (2.1.2).

In the following we shall show that series (7) represents indeed the external solution by proving that series (2.1.3) and (7) in their own region of convergence equal one and the same function. So it will be proven that

$$e^{at}\left\{\sum_{k=0}^{\infty} G_k(x,a)\alpha^{(k)}(t)\right\} \equiv \sum_{k=0}^{\infty} g_k(x)\varphi^{(k)}(t). \tag{15}$$

For this we need more knowledge about the relationship between the functions $G_k(x,a)$ and the polynomials $g_k(x)$. Differentiating both members of (10) with respect to a gives

$$G_{k+1}(x,a) = \frac{1}{k+1}\,\frac{\partial G_k(x,a)}{\partial a}\,. \tag{16}$$

With the help of this recurrent relation we can express $G_k(x,a)$ into $G_o(x,a)$. Hence, the following identity is true

$$\sum_{k=o}^{\infty} G_k(x,a)(b-a)^k = \sum_{k=o}^{\infty} \frac{(b-a)^k}{k!}\,\frac{\partial^k G_o(x,a)}{\partial a^k} = G_o(x,b), \tag{17}$$

provided $|b-a| < |(J\pi)^2 + a|$. This restriction can easily be understood from the condition of convergence of series (4) by choosing $\alpha(t) = e^{(b-a)t}$. From (17) and (8) we conclude

$$\frac{\sinh\sqrt{b}(1-x)}{\sinh\sqrt{b}} = \sum_{k=o}^{\infty} G_k(x,a)(b-a)^k = \sum_{k=o}^{\infty} g_k(x)b^k\;, \tag{18}$$

provided $|b-a| < |(J\pi)^2 + a|$ and $|b| < \pi^2$. From the first condition it follows that b must be inside the circle with centre a and that passes through $-(J\pi)^2$. The second condition of (18) requires that b is inside the circle with centre 0 and radius π^2. So b must be inside the common part of both circles. It depends on the values of a if both circles do intersect each other. The choice of Re a is already limited by (12) to the interval (a_1,a_2); however for the choice of Im a there are no limitations. So for each value of J we find that both circles cut each other by choosing $|\mathrm{Im}\,a|$ sufficiently large. In the common part of both circles the power series in (18) are absolutely and uniformly convergent. Arranging both series in powers of b, we obtain

$$g_k(x) = \sum_{j=o}^{\infty} \binom{k+j}{k} G_{k+j}(x,a)(-a)^j, \tag{19}$$

while arranging in powers of $(b-a)$ results in

$$G_k(x,a) = \sum_{j=o}^{\infty} \binom{k+j}{k} g_{k+j}(x)a^j. \tag{20}$$

Now we are able to prove (15). Arranging both series of (15) as $\varphi^{(k)}(t)$, $k = 0,1,\dots,$ it follows from (19) that the coefficients of $\varphi^{(k)}(t)$ in both series are the same. Hence, indeed relation (15) is valid inside the common region of convergence or in other words both series in (15) represent one and the same function of x and t. Moreover from (18) it follows that series (4) equals the external solution in the case $\varphi(t) = e^{bt}$, $\eta(t) = 0$. Hence we can define the external solution for the above restriction of

Im a by seriesexpansion (4) inside its region of convergence. It is not difficult to show that this is true for each choice of a, excepted the particular values $a = -(j\pi)^2$.

Now we will consider the special case, $\varphi(t) = e^{-(j\pi)^2 t}$, $j = 1,2,\ldots$. We observe that for this $\varphi(t)$ series (4) does not converge for any value of a and that $z_e(x,t)$ in (1) becomes infinite. From (1) we see that if $a \to -(j\pi)^2$, the behaviour of $z_e(x,t)$ as function of x and t becomes like that of the j^{th} elementary solution. Together with the external solution also the amplitude of the j^{th} elementary solution in the internal solution becomes infinite. However the sum of both stays finite. For the physical system it means that it is forced via the boundary $x = 0$ in its j^{th} eigenvibration and so there occurs some kind of resonance. This resonance will be considered in more detail, if moreover the initial condition of problem (2.1.1) is taken zero. Writing the initial value of the external solution (1) as a Fourierseries, we obtain for the internal solution

$$z_i(x,t) = -2 \sum_{r=1}^{\infty} \frac{r}{(r\pi)^2 + a} \sin r\pi x \; e^{-(r\pi)^2 t}. \tag{21}$$

From this we find as the limit of the sum of the external solution and the j^{th} term of (21), if $a \to -(j\pi)^2$:

$$(1 - x)\cos j\pi x \; e^{-(j\pi)^2 t} + 2j\pi t \sin j\pi x \; e^{-(j\pi)^2 t}.$$

Finally we remark that it can be verified, just as has been done in paragraph 2.1, that seriesexpansions (4) and (7) are related to a 2 points expansion of a function f(x).

$$f(x) = \sum_{k=0}^{\infty} \{G_k(x,a)O_1^k f(0) + G_k(1 - x,a)O_1^k f(1)\}, \text{ where}$$

$$O_1 f(x) = f^{(2)}(x) - a\, f(x). \tag{23}$$

2.7. External solution for boundary conditions of mixed type

In this paragraph we will discuss briefly the problem

$$\frac{\partial z}{\partial t} - \frac{\partial^2 z}{\partial x^2} = 0$$

$$z(0,t) + \mu_1 \frac{\partial z(0,t)}{\partial x} = \varphi(t), \quad z(1,t) + \mu_2 \frac{\partial z(1,t)}{\partial x} = \eta(t)$$

$$z(x,0) = \gamma(x), \tag{1}$$

which is a modification of problem (2.1.1), having boundary conditions of the mixed type.

By definition the external solution of (1) can be written in its simplest form as

$$z_e(x,t) = \sum_{k=0}^{\infty} \{h_{1,k}(x)\varphi^{(k)}(t) + h_{2,k}(x)\eta^{(k)}(t)\}. \tag{2}$$

From the formal substitution of this into (1) we find that both functions $h_{1,k}(x)$, $h_{2,k}(x)$ satisfy the same recurrent relation

$$h_k^{(2)}(x) - h_{k-1}(x) = 0,\ h_k(0) + \mu_1 h_k^{(1)}(0) = h_k(1) + \mu_2 h_k^{(1)}(1) = 0,$$

$$k \geqq 1, \tag{3}$$

while $h_{1,0}(x)$ and $h_{2,0}(x)$ prove to be

$$h_{1,0}(x) = \frac{1+\mu_2 - x}{1+\mu_2-\mu_1},\quad h_{2,0} = \frac{x-\mu_1}{1+\mu_2-\mu_1}. \tag{4}$$

Hence the functions $h_{1,k}(x)$ and $h_{2,k}(x)$ are polynomials in x, which can be determined one after another from (3) and (4). We can also write them as infinite series of the x-elementary solutions of (3), which are found to be

$$(-1)^k \frac{\sin(a_j x + b_j)}{a_j^{2k}},\ \text{where } \mathrm{tg}\, b_j = -\mu_1 a_j \text{ and } \mathrm{tg}(a_j + b_j) = -\mu_2 a_j. \tag{5}$$

From this we see, that in general the constants a_j are no multiples of $j\pi$ and so the functions $h_{1,k}(x)$ and $h_{2,k}(x)$, written as series of x-elementary solutions, are no Fourierseries.

Again series (2) is based on a 2 points expansion of a function f(x),

$$f(x) = \sum_{k=0}^{\infty} \{h_{1,k}(x,\mu_1,\mu_2)O_1^k(\mu_1)f(0) + h_{2,k}(x,\mu_1,\mu_2)O_1^k(\mu_2)f(1)\},\ \text{where}$$

$$O_1^k(\mu)f(x) = f^{(2k)}(x) + \mu f^{(2k+1)}(x). \tag{6}$$

2.8. External solution for the inhomogeneous parabolic differential equation

In the preceding we have studied the x-external solution in the case that the forcing of the considered physical system happened via the boundaries. Now we will consider that the forcing only occurs via the right hand member of the differential equation, thus that z(x,t) satisfies

$$\frac{\partial z}{\partial t} - \frac{\partial^2 z}{\partial x^2} = f(x,t)$$

$$z(0,t) = z(1,t) = 0$$

$$z(x,0) = \gamma(x). \tag{1}$$

We assume that f can be differentiated infinitely often. By means of $f(x,t)$ the physical system is forced in each point x as function of time t. We observe that if forcing via the boundaries also occurs, the x-external solution can be found by superposing the x-external solutions of (1) and (2.1.1).

By definition $z_e(x,t)$ should be a linear combination of $f(x,t)$ and its derivatives with respect to x and t. However in this way substitution in the differential equation and the conditions on the x-boundaries of (1) cannot give an identity, because in the differential equation of (1) a term $z(x,t)$ is absent. Obviously we can get out of this complication by permiting derivatives to x or t of negative order in the linear combination.

Now we will consider that $f(x,t)$ can be written as

$$f(x,t) = \sum_{k=o}^{\infty} \{g_k(x) \frac{\partial^{2k} f(0,t)}{\partial x^{2k}} + g_k(1-x) \frac{\partial^{2k} f(1,t)}{\partial x^{2k}}\} . \tag{2}$$

First we will determine the contribution to the x-external solution of the general term of (2). We denote this contribution by $z_{ek}(x,t)$. If we take $\partial^{2k} f(1,t)/\partial x^{2k} = 0$ and abbreviate $\partial^{2k} f(0,t)/\partial x^{2k}$ by $u_k(t)$, then $z_{ek}(x,t)$ has to satisfy

$$\frac{\partial z_{ek}}{\partial t} - \frac{\partial^2 z_{ek}}{\partial x^2} = g_k(x) u_k(t), \quad z_{ek}(0,t) = z_{ek}(1,t) = 0. \tag{3}$$

Writing $z_{ek}(x,t)$ in the form

$$z_{ek}(x,t) = \sum_{j=o}^{\infty} Q_j(x) u_k^{(j)}(t), \tag{4}$$

we obtain that the functions $Q_j(x)$, $j \geqq 1$ have to satisfy

$$Q_j^{(2)}(x) - Q_{j-1}(x) = 0, \; Q_j(0) = Q_j(1) = 0, \text{ while}$$

$$Q_o^{(2)}(x) = - g_k(x). \tag{5}$$

Comparing this with (2.1.8) we see immediately that

$$Q_j(x) = - g_{k+j+1}(x). \tag{6}$$

Substituting this into (4) gives

$$z_{ek}(x,t) = - \sum_{j=o}^{\infty} g_{k+j+1}(x) u_k^{(j)}(t). \tag{7}$$

So formally the x-external solution of (1) can be written in the form

$$z_e(x,t) = - \sum_{k=o}^{\infty} \sum_{j=o}^{\infty} \{g_{k+j+1}(x) \frac{\partial^{2k+j} f(0,t)}{\partial x^{2k} \partial t^j} + g_{k+j+1}(1-x) \frac{\partial^{2k+j} f(1,t)}{\partial x^{2k} \partial t^j}\} . \tag{8}$$

In the above we have considered the x-external solution. However, although (1) is a two boundary value problem, the superposition principle also allows us to consider the external solution as descended from $f(x,t)$ in a most convenient way. For instance it can be simplest for a given $f(x,t)$ to consider the t-external solution, satisfying

$$\frac{\partial z_e}{\partial t} - \frac{\partial^2 z_e}{\partial x^2} = f(x,t), \quad z_e(x,0) = 0. \tag{9}$$

Then we can obtain in a similar way as above, writing $f(x,t)$ as a Mac Laurin series, that the external solution descended from $f(x,t)$ conceived as t-external solution is

$$z_e(x,t) = \sum_{k=o}^{\infty} \sum_{j=o}^{\infty} \frac{t^{k+j+1}}{(k+j+1)!} \frac{\partial^{k+2j} f(x,0)}{\partial x^{2j} \partial t^k} . \tag{10}$$

2.9. External and internal solution in relation to the function of Green

Again we consider problems of type (2.1.1). Solving these problems is often done with the help of Laplace-transformation. The advantage of this transformation is that the homogeneous partial differential equation together with the initial condition are transformed into an inhomogeneous ordinary differential equation. Denoting the Laplace-transform $\int_0^\infty z(x,t)e^{-pt}dt$ by $z(x,p)$, we can write the transformed problem as

$$\begin{aligned} &z^{(2)}(x,p) - pz(x,p) = - \gamma(x) \\ &z(0,p) = \varphi(p), \ z(1,p) = \eta(p). \end{aligned} \tag{1}$$

Utilizing the function of Green, belonging to boundary problem (1),

$$\begin{aligned} G(x,\mu,p) &= \frac{\sinh\sqrt{p}\mu \sinh\sqrt{p}(1-x)}{\sqrt{p} \sinh\sqrt{p}}, \quad 0 \leqq \mu \leqq x \\ &= \frac{\sinh\sqrt{p}x \sinh\sqrt{p}(1-\mu)}{\sqrt{p} \sinh\sqrt{p}}, \quad x \leqq \mu \leqq 1, \end{aligned} \tag{2}$$

the solution of (1) can be written in the form

$$z(x,p) = [\int_0^1 G(x,\mu,p)\gamma(\mu)d\mu] + [\frac{\sinh\sqrt{p}(1-x)}{\sinh\sqrt{p}}\varphi(p)] + [\frac{\sinh\sqrt{p}x}{\sinh\sqrt{p}}\eta(p)]. \tag{3}$$

We see that in (3) the influence of $\gamma(x)$, $\varphi(p)$ and $\eta(p)$ arises separately. For instance the first term of the right hand member of (3) is the solution of $z(x,p)$ if the boundary conditions are zero. From (3) we can obtain the solution of (2.1.1) by transforming (3) to the t-region.

We remark that transforming the three terms of (3) separately to the t-region can be inefficient, because each term has obtained one or two discontinuities, which of course altogether cancel each other. We will demonstrate this for the simple problem

$$\frac{\partial z}{\partial t} - \frac{\partial^2 z}{\partial x^2} = 0, \quad z(0,t) = z(1,t) = z(x,0) = 1, \tag{4}$$

having the solution $z(x,t) = 1$. Then expression (3) becomes

$$z(x,p) = [\int_0^1 G(x,\mu,p)d\mu] + [\frac{\sinh\sqrt{p}(1-x)}{p\sinh\sqrt{p}}] + [\frac{\sinh\sqrt{p}x}{p\sinh\sqrt{p}}]. \tag{5}$$

Transformation to the t-region yields

$$z(x,t) = [2\sum_{j=1}^{\infty}\frac{1-(-1)^j}{j\pi}\sin j\pi x\, e^{-(j\pi)^2 t}] + [(1-x) - 2\sum_{j=1}^{\infty}\frac{\sin j\pi x}{j\pi}e^{-(j\pi)^2 t}] +$$

$$+ [x + 2\sum_{j=1}^{\infty}\frac{(-1)^j}{j\pi}\sin j\pi x\, e^{-(j\pi)^2 t}] = 1. \tag{6}$$

Indeed we see that transforming the three terms of (5) separately is much more complicated than transforming all terms together.

If the concepts external and internal solution are used, the solution of (2.1.1) is also divided into parts, but now separate determination of these parts is always efficient. For problem (4) we obtain

$$z_e(x,t) = (1-x) + x = 1 \rightarrow z_i(x,0) = 0 \rightarrow z_i(x,t) = 0 \rightarrow z(x,t) = 1. \tag{7}$$

The solution of problem (1) can also be determined with the help of the concepts external and internal solution instead of utilizing the function of Green. Again we know a priori that no discontinuities are introduced then. Because the external and internal solutions of (1) are not the Laplace-transforms of the external and internal solutions of (2.1.1), we will denote the first ones by $z_e^*(x,p)$ and $z_i^*(x,p)$. Applying the results of paragraph 1.2 we have

$$z_e^*(x,p) = \sum_{k=o}^{\infty} \frac{1}{p^{k+1}} \gamma^{(2k)}(x). \tag{8}$$

Hence $z_i^*(x,p)$ is fixed by

$$z_i^{*(2)}(x,p) - p\, z_i^*(x,p) = 0,$$

$$z_i\,(0,p) = \varphi(p) - \sum_{k=o}^{\infty} \frac{\gamma^{(2k)}(0)}{p^{k+1}};\quad z_i^*(1,p) = \eta(p) - \sum_{k=o}^{\infty} \frac{\gamma^{(2k)}(1)}{p^{k+1}}. \tag{9}$$

The solution of this problem can be found in the usual way, but we will do it applying the 2 points expansion (2.6.23). From (9) we have $(\frac{d^2}{dx^2} - p)^r\, z_i^*(x,p) = 0,\ r \geqq 1$ and so (2.6.23) yields

$$z_i^*(x,p) = \{\varphi(p) - \sum_{k=o}^{\infty} \frac{\gamma^{(2k)}(0)}{p^{k+1}}\}\frac{\sinh\sqrt{p}(1-x)}{\sinh\sqrt{p}} + \{\eta(p) - \sum_{k=o}^{\infty} \frac{\gamma^{(2k)}(1)}{p^{k+1}}\}\frac{\sinh\sqrt{p}x}{\sinh\sqrt{p}}. \tag{10}$$

From (8) and (10) again the solution of (2.1.1) can be found by transformation to the t-region. Immediately we see, that the transform of (8) is

$$z_e^*(x,t) = \sum_{k=o}^{\infty} \frac{t^k}{k!} \gamma^{(2k)}(x), \tag{11}$$

which is the t-external solution of problem (2.1.1). Mostly the transformation of (10) is not simple and the result, $z_i^*(x,t)$, is the solution of a problem (2.1.1) which has still external forcing.

In figure 1 the division into parts of the solution of (4) is represented for the three considered ways.

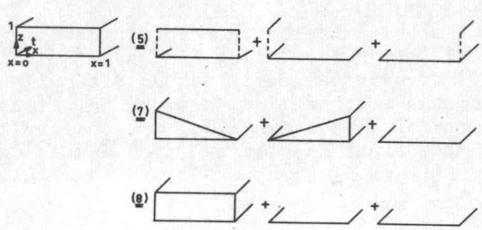

Figure 1. Some ways of dividing the solution of (4) into parts.

If $\varphi(p)$ and $\eta(p)$ can be written as powerseries of $\frac{1}{p}$, then in (3) as well as in (10) terms like $\sinh\sqrt{p}(1-x)/p^{r+1}\sinh\sqrt{p}$ occur. We shall determine the inverse Laplace-transform, L^{-1}, of this term with the help of the polynomials $g_k(x)$. From (2.6.18) we have

$$\frac{\sinh\sqrt{p}(1-x)}{p^{r+1}\sinh\sqrt{p}} = \sum_{k=0}^{r} g_k(x)p^{k-r-1} + \sum_{k=r+1}^{\infty} g_k(x)p^{k-r-1} . \tag{12}$$

Denoting $\sum_{k=r+1}^{\infty} g_k(x)p^{k-r-1}$ by $f(x,p)$, it can easily be seen (look also at paragraph 2.8) that

$$f^{(2)}(x,p) - p\, f(x,p) = g_r(x); \; f(0,p) = f(1,p) = 0. \tag{13}$$

The corresponding problem in the t-region is

$$\frac{\partial f(x,t)}{\partial t} - \frac{\partial^2 f(x,t)}{\partial x^2} = 0; \; f(0,t) = f(1,t) = 0; \; f(x,0) = - g_r(x). \tag{14}$$

From (2.1.14) it follows

$$f(x,t) = (-1)^{r+1}\, 2 \sum_{j=1}^{\infty} \frac{\sin j\pi x}{(j\pi)^{2r+1}} e^{-(j\pi)^2 t} . \tag{15}$$

Hence

$$L^{-1}\left\{\frac{\sinh\sqrt{p}(1-x)}{p^{r+1}\sinh\sqrt{p}}\right\} = \sum_{k=0}^{r} \frac{g_k(x)t^{r-k}}{(r-k)!} + (-1)^{r+1}\, 2 \sum_{j=1}^{\infty} \frac{\sin j\pi x}{(j\pi)^{2r+1}} e^{-(j\pi)^2 t}. \tag{16}$$

Finally we will determine $G(x,\mu,t)$ formally from $G(x,\mu,p)$ with the help of the polynomials $g_k(x)$. Applying (3) to (13) yields

$$\int_0^1 G(x,\mu,p)g_k(\mu)d\mu = - \sum_{j=k+1}^{\infty} g_j(x)p^{j-k-1}, \quad k \geqq 0. \tag{17}$$

We observe that this recurrent relation between the polynomials $g_k(x)$ simplifies much for $p = 0$; then we obtain again the recurrent relation (2.2.7). If we write $G(x,\mu,p)$ in (17) as power series in p, then interchanging the order of summation and integration yields

$$g_{k+m+1}(x) = - \int_0^1 \frac{\partial^m G(x,\mu,0)}{m!\,\partial p^m} g_k(\mu)d\mu . \tag{18}$$

We see that this is a recurrent relation between $g_{k+m+1}(x)$ and $g_k(x)$. We remark that the same relation would be obtained by applying transformation (2.2.7) $(m+1)$ times and so the kernel of the integral equation (18) is equal to $(-1)^m K_{m+1}(x,\mu)$. Substituting (2.1.14) into (18) and after that interchanging the order of summation and integration yields

$$\int_0^1 \frac{\partial^m G(x,\mu,0)}{m!\,\partial p^m} \sin n\pi\mu \, d\mu = \frac{(-1)^m}{(n\pi)^{2m+2}} \sin n\pi x, \quad n = 1,2,\dots . \tag{19}$$

This means that the eigenvalues and normalized eigenfunctions of the kernel $\partial^m G(x,\mu,0)/m!\,\partial p^m$ occurring in (19) are

$$\frac{(-1)^m}{(n\pi)^{2m+2}}, \quad \sqrt{2} \sin n\pi x, \quad n = 1,2,\dots . \tag{20}$$

Hence by virtue of a well-known theorem

$$\frac{\partial^m G(x,\mu,0)}{m!\,\partial p^m} = 2 \sum_{r=1}^{\infty} \frac{(-1)^m \sin r\pi x \sin r\pi\mu}{(r\pi)^{2m+2}} . \tag{21}$$

From this we have

$$G(x,\mu,p) = 2 \sum_{r=1}^{\infty} \frac{\sin r\pi x \sin r\pi\mu}{p+(r\pi)^2} . \tag{22}$$

Hence

$$G(x,\mu,t) = 2 \sum_{r=1}^{\infty} \sin r\pi x \sin r\pi\mu \, e^{-(r\pi)^2 t} . \tag{23}$$

2.10. External and internal solution for two simultaneous partial differential equations

In the preceding paragraphs of this chapter we have considered problems concerning only one partial differential equation. Now we will treat briefly the concepts external and internal solution for the following problem with two simultaneous partial differential equations

$$\frac{\partial z_1}{\partial t} - a_1 \frac{\partial^2 z_1}{\partial x^2} - d_{11} z_1 - d_{12} z_2 = 0, \quad a_1 > 0$$

$$\frac{\partial z_2}{\partial t} - a_2 \frac{\partial^2 z_2}{\partial x^2} - d_{21} z_1 - d_{22} z_2 = 0, \quad a_2 > 0$$

$$z_{1,2}(0,t) = \varphi_{1,2}(t); \; z_{1,2}(1,t) = \eta_{1,2}(t); \; z_{1,2}(x,0) = \gamma_{1,2}(x). \tag{1}$$

For convenience sake we will take $\eta_1(t) = \varphi_2(t) = \eta_2(t) = 0$ and denote $\varphi_1(t)$ by $\varphi(t)$.

By definition we have the x-external solution $\{z_{1e}(x,t),\ z_{2e}(x,t)\}$ as

$$z_{1e}(x,t) = \sum_{k=o}^{\infty} G_k(x)\varphi^{(k)}(t),\ z_{2e}(x,t) = \sum_{k=o}^{\infty} H_k(x)\varphi^{(k)}(t). \tag{2}$$

It follows from substitution of (2) into (1) that $G_k(x)$ and $H_k(x)$ have to satisfy

$$a_1 G_k^{(2)} - G_{k-1} + d_{11}G_k + d_{12}H_k = 0,\ G_k(0) = G_k(1) = 0$$

$$a_2 H_k^{(2)} - H_{k-1} + d_{21}G_k + d_{22}H_k = 0,\ H_k(0) = H_k(1) = 0, \tag{3}$$

while $G_o(x)$ and $H_o(x)$ are fixed by

$$a_1 G_o^{(2)} + d_{11}G_o + d_{12}H_o = 0,\ G_o(0) = 1;\ G_o(1) = 0$$

$$a_2 H_o^{(2)} + d_{21}G_o + d_{22}H_o = 0,\ H_o(0) = H\ (0) = 0. \tag{4}$$

We shall write $H_k(x)$ and $G_k(x)$ as series of x-elementary solutions of (3). It can easily be seen that

$$\{g_j^{\ k} \sin j\pi x,\ \ K_j g_j^{\ k} \sin j\pi x\} \tag{5}$$

is a x-elementary solution of (3) provided

$$[\{d_{11} - a_1(j\pi)^2\}g_j - 1] + \qquad d_{12}\ K_j = 0$$

$$d_{21} + [\{d_{22} - a_2(j\pi)^2\}g_j - 1]K_j = 0. \tag{6}$$

Denoting the values of g_j, for which (6) has a solution by g_{j1} and g_{j2} and the corresponding values of K_j by K_{j1}, K_{j2}, we have the general solution $\{G_k(x),\ H_k(x)\}$ of (3) as

$$\{\sum_{j=1}^{\infty} (A_j g_{j1}^{\ k} + B_j g_{j2}^{\ k})\sin j\pi x,\ \sum_{j=1}^{\infty} (A_j K_{j1} g_{j1}^{\ k} + B_j K_{j2} g_{j2}^{\ k})\sin j\pi x\}. \tag{7}$$

Writing the solution of (4) as $\{G_o(x),\ H_o(x)\} = \{\sum_{j=1}^{\infty} G_{oj} \sin j\pi x$, $\sum_{j=1}^{\infty} H_{oj} \sin j\pi x\}$, from (7) the solution of (3) and (4) arises if

$$A_j + B_j = G_{oj}$$

$$A_j K_{j1} + B_j K_{j2} = H_{oj}. \tag{8}$$

By definition the internal solution of (1) is the difference between the solution itself and the external solution of (1) and so $\{z_{1i}(x,t), z_{2i}(x,t)\}$ can be written as a series of x-elementary solutions of (1). Trying an elementary solution of the form

$$\{\sin j\pi x\, e^{\mu_j t},\ L_j \sin j\pi x\, e^{\mu_j t}\}, \tag{9}$$

then substitution into (1) shows that this is true provided that the substitution of g_j and K_j in (6) by $\frac{1}{\mu_j}$ and L_j gives an identity. This means

$$\mu_{j1,2} = \frac{1}{g_{j1,2}}\ ;\quad L_{j1,2} = K_{j1,2}. \tag{10}$$

Hence the internal solution $\{z_{1i}(x,t),\ z_{2i}(x,t)\}$ can be written as

$$\{\sum_{j=1}^{\infty} (C_j\, e^{\mu_{j1}t} + D_j\, e^{\mu_{j2}t})\sin j\pi x,\ \sum_{j=1}^{\infty} (C_j K_{j1}\, e^{\mu_{j1}t} + D_j K_{j2}\, e^{\mu_{j2}t})\sin j\pi x\}, \tag{11}$$

where the constants C_j and D_j are such that (11) equals

$$[\,\{\gamma_1(x) - z_{1e}(x,0)\},\ \{\gamma_2(x),\ z_{2e}(x,0)\}\,] \quad \text{for } t = 0.$$

2.11. External solution for the damped vibrating string

In this paragraph we will consider the concept x-external solution for the damped vibrating string as an example of a nondiffusion problem. The displacement of the string satisfies

$$a_1 \frac{\partial^2 z}{\partial t^2} + a_2 \frac{\partial z}{\partial t} - \frac{\partial^2 z}{\partial x^2} = 0,$$

$$z(0,t) = \varphi_1(t),\ z(1,t) = \varphi_2(t),\ z(x,0) = \eta_1(x),\ \frac{\partial z(x,0)}{\partial x} = \eta_2(x). \tag{1}$$

Denoting the operator $(a_1 \frac{\partial^2}{\partial t^2} + a_2 \frac{\partial}{\partial t})$ by O_1, we see comparing (1) with (2.1.1) that O_1 occurs instead of $\frac{\partial}{\partial t}$. Thus the x-external solution of (1) can be obtained from (2.1.3) by replacing the operator $\frac{d}{dt}$ by O_1. This results into

$$z_e(x,t) = \sum_{k=o}^{\infty} \{g_k(x)O_1^k\varphi_1(t) + g_k(1 - x)O_1^k\varphi_2(t)\}. \tag{2}$$

We see that in the special case $(a_1,a_2) = (0,1)$ and $(a_1,a_2) = (1,0)$ series (2) is the x-external solution of, respectively, the parabolic equation in elementary form and the hyperbolic equation in elementary form (undamped vibrating string).

If we take for instance $\varphi_1(t) = e^{bt}$, $\varphi_2(t) = 0$, then (2) becomes

$$z_e(x,t) = \{\sum_{k=o}^{\infty} g_k(x)(a_1 b^2 + a_2 b)^k\} e^{bt}$$

$$= \frac{\sinh\{\sqrt{a_1 b^2 + a_2 b}(1-x)\}}{\sinh\sqrt{a_1 b^2 + a_2 b}} e^{bt} . \tag{3}$$

Chapter 3

INTERPRETATION OF A DIFFUSION PROBLEM AS A DIFFERENCE PROBLEM

3.1. Introduction

Assuming that the determination of the solution of a problem

$$\frac{\partial z}{\partial t} - \frac{\partial^2 z}{\partial x^2} - pz = 0$$

$$z(0,t) = \varphi(t), \; z(1,t) = \eta(t)$$

$$z(x,0) = \gamma(x) \tag{1}$$

is done approximately by solving numerically a corresponding difference problem with the help of a computer, we will consider in this chapter some aspects of the accuracy of the (analytic) solution of the difference problem with regard to the solution of (1).

We will limit ourselves to one class of difference equations, which can be written as

$$\frac{\delta_- z(x_m,t_n)}{\Delta t} - a\frac{\delta^2 z(x_m,t_n)}{(\Delta x)^2} - (1-a)\frac{\delta^2 z(x_m,t_{n-1})}{(\Delta x)^2} - apz(x_m,t_n) - (1-a)pz(x_m,t_{n-1}) = 0, \tag{2}$$

where again δ, δ_- mean the central and backward difference. We observe that the difference equations of Milne and of Laasonen, considered in paragraph 2.5, are special cases of (2).

For difference equation (2) the elementary solutions of problem (1) in difference form prove to be

$$\sin j\pi x_m \, g_j^{t_n}, \quad j = 1,\dots,M-1, \tag{3}$$

where

$$g_j = \left\{\frac{1-(1-a)q_j\Delta t}{1+aq_j\Delta t}\right\}^{\frac{1}{\Delta t}}, \quad q_j = \frac{2(1-\cos j\pi\Delta x)}{(\Delta x)^2} - p. \tag{4}$$

The accuracy of the interpretation of the solution of (1) using difference equation (2) depends among other things on the choice of Δt and Δx. The accuracy can be increased by reducing Δt and Δx. However, then in general the cost of determining the solution increases. So it is important to choose Δt and Δx not too small. We will see that regarding this choice it may be favourable to fit the given initial condition $\gamma(x)$ in a special way before using it for the difference problem.

3.2. Accuracy

Assuming $p < \pi^2$, we know from (2.4.6) that the internal solution of (3.1.1) tends to zero if $t \to \infty$. So for large values of t the accuracy of the representation of (1) by a difference problem only depends on the difference between the external solutions. We restrict ourselves to problems of type (3.1.1), for which the external solution does not change much as a function of x and t. This means that for large t also Δx and Δt can be chosen large.

For small t in many cases the internal solution of (3.1.1) is not small in comparison with the external solution and does change much as function of x and t. So for small t the choice of Δt and Δx has often to be done with regard to a good representation of the internal solution.

The accuracy of the interpretation of a differential problem by a difference problem will be better if the deviation in behaviour of the solutions of both problems is less. This means that it is not primarily important to reduce the difference between both solutions itself as much as possible, but that also the significant derivatives with respect to x and t must be interpreted sufficiently accurate by the difference problem.

We remark that if the solution of a problem (3.1.1) has to be determined after a specified time T, it would be attractive to be able to choose Δx and Δt such that the external solution as well as the elementary solutions, being present noticeably at $t = T$, are interpreted just sufficiently accurate by the difference problem. We will denote these optimum values of Δx and Δt by Δx_{opt} and Δt_{opt}.

3.3. Fitting the initial condition

From formula (3.1.3) we see that the elementary solutions of the differential problem with $j \geqq M$, so the high frequency elementary solutions, are not recognized in the difference problem. However, if the same initial condition is used, the presence of high frequency elementary solutions in the differential problem does influence the solution of the difference problem by a change of the amplitudes of the low frequency elementary solutions. Even in the case of equal external solutions or in other words of equal initial conditions of the internal solutions - writing these initial conditions as Fourier series - the coefficients c_j and d_j of the obtained infinite and finite Fourier series differ from each other. From (2.3.17) we have

$$d_j - c_j = \sum_{r=1}^{\infty} \{c_{2rM+j} - c_{2rM-j}\}, \quad j = 1,\ldots,M-1. \tag{1}$$

Often this phenomenon of changing the amplitudes of the low frequency elementary solutions is called "folding".

In many cases for $\Delta t = \Delta t_{opt}$ and $\Delta x = \Delta x_{opt}$ the error in the amplitudes of the low frequency elementary solutions will be too large. This error can be dininished in two ways.

One way is by reducing Δx. Then, however, a number of the original high frequency elementary solution become recognizable in the difference problem. Of course these elementary solutions also have to be interpreted by the difference problem. Because a right representation of higher frequency elementary solutions demands a smaller Δt, reducing Δx often results inevitably in reducing Δt. The values of Δt and Δx for which the amplitude error caused by folding is small enough and the recognizable elementary solutions are interpreted sufficiently well are dependent on the course of the Fourier coefficients c_j as function of j.

Another possibility of reducing the folding-error is removing the high frequency components in the initial condition of the internal solution of the differential problem before using it in the difference problem. We will call this transformation of the initial condition "fitting the initial condition". If the initial condition is fitted, then we can take $\Delta x = \Delta x_{opt}$ and $\Delta t = \Delta t_{opt}$. This means that in this way solving the difference problem takes no more computing time and memory capacity than strictly necessary. Of course fitting the initial condition takes also computing time. However roughly we can say that fitting the initial condition will be the best way of reducing the folding-error, if the initial condition of the internal solution contains many high frequency components.

Sometimes it may also be sensible to fit the boundary conditions $\varphi(t)$ and $\eta(t)$ by means of smoothing, for instance if φ and η contain high frequency noise.

The effect of fitting the initial condition can only be predicted precisely for linear problems, because only then do the elementary solutions not influence each other. It depends on the measure of nonlinearity as to whether or not fitting the initial condition can be applied. If a problem is very nonlinear then small values of Δt and Δx must be taken because of the interaction between the elementary solutions and so there is no need of fitting the initial condition.

In many cases the solution of a problem (3.1.1) has to be determined a number of successive times. For each next time more high frequency components can be omitted out of the initial condition. Gradually Δt and Δx can be enlarged to the values necessary for determining the external solution.

3.4. Behaviour of the elementary solutions as function of time

In the last paragraph we have already discussed the effect of the discretization of x with regard to the interpretation of the amplitudes of the elementary solutions. Now we will consider the effect of discretizing x and t on the behaviour of the elementary solutions as function of t.

The discretization of x does change the growthfactor of the j^{th} elementary solution of the differential problem $e^{p-(j\pi)^2}$ by the replacement of $y_1 = -(j\pi)^2$ in the exponent by $y_2 = 2(\cos j\pi\Delta x - 1)/(\Delta x)^2$. In figure 1 both

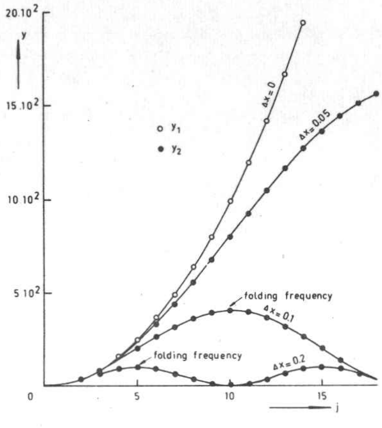

Figure 1. Influence of the discretization of x.

of these expressions are given as function of j. The folding reflects itself in the symmetry of the second one with regard to the values $j = k/\Delta x$, $k = 1,2$, etc.. For smaller values of Δx both expressions are approximately equal over a larger j-range.

The discretization of t does change the e-power itself of $e^{p-(j\pi)^2}$. The effect of discretizing both x and t is already given in paragraph 3.1. In order to obtain a quantity containing Δx as well as Δt implicitly we replace the growthfactor $g_{j,\Delta t}$ of the difference problem by

$$g_u = \frac{1+(1-a)u}{1-au}, \quad u = \{p - \frac{2(1-\cos j\pi\Delta x)}{(\Delta x)^2}\}\Delta t. \tag{1}$$

From this we have

$$g_u = 1 + u + au^2 + \ldots + a^{n-1}u^n + \ldots, \quad |au| < 1. \tag{2}$$

In the figures 2 and 3 the graph of g_u as function of u is given. This has been done for the difference equation of Milne $(a = 0, b = 0)$, Laasonen $(a = 1, b = 0)$ and of Crank-Nicolson $(a = \frac{1}{2}, b = 0)$ and also for a difference equation $(a = \frac{2}{3}, b = -\frac{1}{6})$ which will be discussed in the next paragraph. Figure 2 has linear scales while in figure 3 the u-scale is a logarithmic

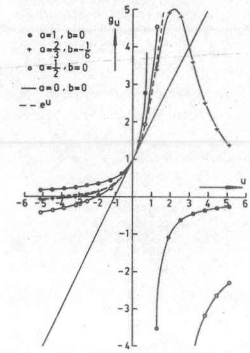

Figure 2. Curve of g_u as a function of u for a number of difference equations.

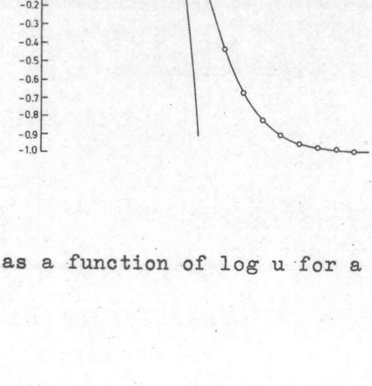

Figure 3. Curve of g_u as a function of log u for a number of difference equations.

one and so values of u differing an octave have equal distances along the u-axis.

In the next we will only consider problems of type (3.1.1) for $p = 0$. First we will compare the growthfactors per step of the j^{th} elementary solutions for small values of Δt and Δx. Then the difference between these growthfactors can be written as

$$g_u - e^{-(j\pi)^2\Delta t} = \{u + (j\pi)^2\Delta t\} + \{au^2 - \frac{(j\pi)^4(\Delta t)^2}{2!}\} + \{a^2u^3 - \frac{(j\pi)^6(\Delta t)^3}{3!}\} + \ldots =$$

$$= \{\frac{(j\pi)^4(\Delta x)^2\Delta t}{12} \ldots\} + \{(a - \frac{1}{2!})(j\pi)^4(\Delta t)^2 \ldots\} + \{-(a^2 - \frac{1}{3!})(j\pi)^6(\Delta t)^3 \ldots\} + \ldots . \quad (3)$$

So the difference between the growthfactors per step is, for small Δt and Δx, linearly dependent on $(\Delta x)^2\Delta t$ and in general also on $(\Delta t)^2$. We see that for small values of Δt and Δx the best choice of a that can be made is $a = \frac{1}{2}$ or in other words the best results will be obtained for the difference equation of Crank-Nicolson. For $a = \frac{1}{2}$ and arbitrary choice of $\Delta t/\Delta x$ the difference between the growthfactors per step is in first order approximation proportional to $(j\pi)^4\{(\Delta x)^2 - (j\pi)^2(\Delta t)^2\}\Delta t$. So if we want to interprete the j^{th} elementary solution as good as possible then Δt and Δx have to satisfy:

$$\frac{\Delta x}{\Delta t} = j\pi . \quad (4)$$

In this case (3) becomes

$$g_u - e^{-(j\pi)^2\Delta t} = \frac{1}{180}(j\pi\Delta x)^5 \ldots . \quad (5)$$

The difference in amplitudes of the j^{th} elementary solutions $g_u^{\frac{t}{\Delta t}} - e^{-(j\pi)^2 t}$ is a function of time. If we take t continuously it can easily be shown that for small Δt and Δx this difference in amplitudes has an extremum occurring at

$$t \approx \frac{1}{(j\pi)^2} . \quad (6)$$

For this value of t the difference in amplitudes equals

$$e^{-1}\{1 + \frac{\ln g_u}{(j\pi)^2\Delta t}\} = \frac{e^{-1}}{(j\pi)^2\Delta t}\left[\{u + (j\pi)^2\Delta t\} + (a - \frac{1}{2})u^2 + \ldots\right] =$$

$$= e^{-1}\left[\{\frac{(j\pi)^2(\Delta x)^2}{12} \ldots\} + (a - \frac{1}{2})\{(j\pi)^2\Delta t \ldots\} + \ldots\right] . \quad (7)$$

So, in general, the extremum is

$$\approx e^{-1}(j\pi)^2\{\frac{(\Delta x)^2}{12} + (a - \frac{1}{2})\Delta t\} ; \qquad (8)$$

but in the special case $a = \frac{1}{2}$ it becomes

$$\approx \frac{e^{-1}(j\pi)^2}{12} \{(\Delta x)^2 - (j\pi\Delta t)^2\} \qquad (9)$$

and if, moreover, Δt and Δx satisfy (4) then the extremum is

$$\approx \frac{e^{-1}}{180} (j\pi\Delta x)^4 . \qquad (10)$$

Above we have seen that the difference equation of Crank-Nicolson is the best one for small values of u. Now, however, we will show that it is not the best one for large values of $|u|$. From (1) we have

$$\lim_{|u|\to\infty} g_u = - \frac{1-a}{a} . \qquad (11)$$

Large values of -u correspond with large values of j and Δt and thus also with small values of $e^{-(j\pi)^2\Delta t}$. So we see that for large values of -u the best choice is a = 1.

In the case $a = \frac{1}{2}$ the growthfactor $g_u \approx -1$ for large negative values of u. This means if Δt is not small enough that for the difference equation of Crank-Nicolson the high frequency elementary solutions alternate nearly undamped while they do not alternate and are very damped in the differential problem. As already said before without fitting the initial condition Δx has to be chosen such that the folding error is sufficiently small and Δt such that all recognizable frequencies are interpreted sufficiently well. From the above it will be clear that without fitting it can happen that Δt and Δx have to be chosen because of a good interpretation of the high frequency elementary solutions and so much larger than Δt_{opt} and Δx_{opt}. Then the advantage has been lost that regarding the low frequency elementary solutions the largest values of Δt and Δx can be applied if $a = \frac{1}{2}$.

3.5. A difference equation interpreting good low frequency as well as high frequency elementary solutions

We have seen that the difference equation of Crank-Nicolson has a growthfactor which tends to -1 for the high frequency elementary solutions, if Δt is not small enough. The reason for this is that numerator and denumerator of g_u as given in (3.4.1) are polynomials in u of the same degree.

It can easily be seen that for p = 0 we can obtain also a second power of u in the denumerator of g_u by adding a term $b\Delta t\{\delta^4 z(x_m,t_n)/(\Delta x)^4\}$, b = constant, to the difference equation (3.1.2) which passes then into

$$\frac{\delta_- z(x_m, t_n)}{\Delta t} - a \frac{\delta^2 z(x_m, t_n)}{(\Delta x)^2} - (1-a) \frac{\delta^2 z(x_m, t_n)}{(\Delta x)^2} - b\Delta t \frac{\delta^4 z(x_m, t_n)}{(\Delta x)^4} = 0; \tag{1}$$

while the formula for g_u becomes

$$g_u = \frac{1+(1-a)u}{1-au-bu^2}. \tag{2}$$

From this we have

$$g_u = 1 + u + (a + b)u^2 + (a^2 + ab + b)u^3 + \ldots, \text{ if } |au + bu^2| < 1. \tag{3}$$

Comparing this with (3.4.2) we see that for small values of u the best choice of $(a+b)$ is done if $(a+b) = \frac{1}{2}$. Then we can still choose freely one of the constants a and b. For instance this choice can be made in order to make the third term of (3.4.3) vanishes in first order approximation. Then $a = \frac{2}{3}$, $b = -\frac{1}{6}$. However in this case g_u becomes negative for $u < -3$ and reaches a minimum of about $-0,1$ for $u \approx -8$. Sometimes it may be better if we take a and b such that g_u cannot become negative. From (2) we see that for that purpose $a \geqq 1$. An attractive choice is $a = 1$, $b = -\frac{1}{2}$, because then the numerator does no longer contain u. For this choice of a and b difference equation (1) becomes

$$\frac{\delta_- z(x_m, t_n)}{\Delta t} - \frac{\delta^2 z(x_m, t_n)}{(\Delta x)^2} + \frac{\Delta t}{2} \frac{\delta^4 z(x_m, t_n)}{(\Delta x)^4} = 0, \tag{4}$$

and g_u is

$$g_u = \frac{1}{1-u+\frac{1}{2}u^2}. \tag{5}$$

In fig. 1 for some values of a and b the course of g_u as function of u is given.

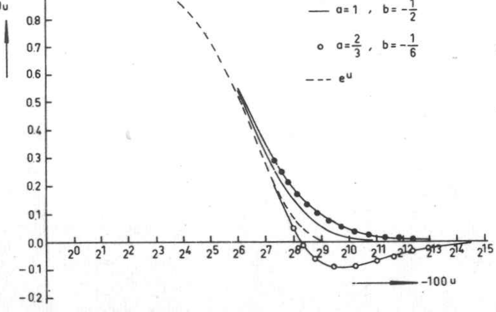

Figure 1. Curve g_u as a function of log u for difference equation (1) for some values of a and b.

Chapter 4

A SPECIAL ANALOGUE COMPUTER

4.1. Design of the special analogue computer

In this chapter a special analogue computer that can solve problems of the following type (1) will be briefly discussed

$$\frac{\partial z}{\partial t} - a\,\frac{\partial^2 z}{\partial x^2} = f(x,t,z), \qquad a > 0$$

$$z(0,t) = \varphi(t),\ z(1,t) = \eta(t);\ z(x,0) = \gamma(x), \tag{1}$$

where a does not need to be a constant [5], [6], [7].

The discussion will be limited to the principal and technical design of this computer and so here we will not enter into the general question when it will be sensible to design and apply a special analogue computer.

In the special analogue computer a problem (1) is interpreted by a difference problem. Figure 1 represents the blockdiagram of the computer in the case that the difference equation of Laasonen is used.

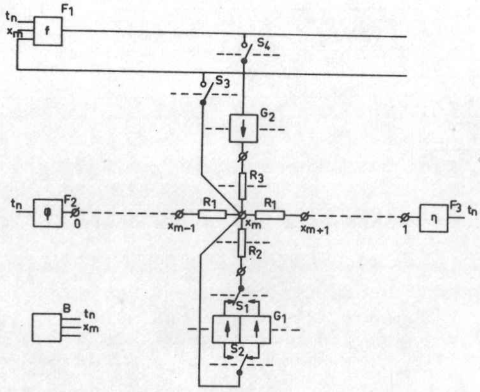

Figure 1. Blockdiagram of the special analogue computer.

The part of the difference equation, $\{\delta_- z(x_m,t_n)/\Delta t - \delta^2 z(x_m,t_n)/(\Delta x)^2\}$ is realized in the computer by means of a resistance network, which is given in figure 2 in more detail.

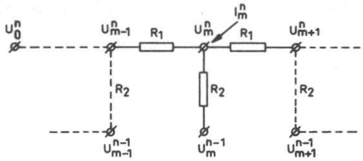

Figure 2. Resistance network for the difference equation of Laasonen.

We see that the balance of the currents in the resistance network itself agrees with $\{\delta_- z(x_m,t_n)/\Delta t - \delta^2 z(x_m,t_n)/(\Delta x)^2\} = 0$. In order that the voltages $U(x_m,t_n)$ in the internal nodes satisfy the difference equation as a consequence of the balance of currents the term $f(x,t,z)$ is realized by means of injection currents I_m^n flowing to the internal nodes.

From the figures 1 and 2 it will be clear that we must still have available $U(x_m,t_{n-1})$ during the computation of $U(x_m,t_n)$. Therefore, untill this computation has been finished the voltages obtained from the preceding computation are kept with the help of a set of analogue memories G_1. Each of these memories has two locations. During the computation of the solution at t_n the memories G_1 deliver the solution at t_{n-1} to the resistance network from the one half of their locations, while the other half is set to the voltages present in the network. So at the end of the computation these locations are already adjusted to the obtained solution. Passing from t_n to t_{n+1} in each memory G_1 the function of its locations are interchanged by means of the set of switches S_1 and S_2.

Often the injection currents I_m^n cannot be obtained from the voltages $U(x_m,t_n)$ in a simple way because of the nonlinear relation between f and z. From figure 2 we see that

$$\frac{1}{R_1}(U_m^n - U_{m-1}^n) + \frac{1}{R_1}(U_m^n - U_{m+1}^n) + \frac{1}{R_2}(U_m^n - U_{m-1}^{n-1}) = I_m^n . \qquad (2)$$

So U_m^n satisfies problem (1) in difference form, if we take care that the injection voltage V_m^n satisfies

$$V_m^n = U(x_m,t_n) + f\{x_m,t_n,U(x_m,t_n)\} \; ; \quad \frac{R_2}{R_1} = \frac{\Delta t}{(\Delta x)^2} \, , \quad \frac{R_3}{R_1} = \frac{1}{(\Delta x)^2} . \quad (3)$$

The boundary conditions $U(0,t_n) = z(0,t_n)$ and $U(1,t_n) = z(1,t_n)$ are obtained from the analogue function generators F_2 and F_3.

The solution at t_n can only be calculated straightforward without an iteration process if as many function generators F_1 are used as there are internal nodes [8]. However then the computer will often be too expensive. Therefore only one function generator F_1 is used to create the injection voltages for all nodes. As a consequence the solution at each t_n can only be obtained by means of an iteration process. During this iteration process the function generator F_1 is utilized in each iteration cycle one after another for all internal nodes. If the iteration process is convergent, then the injection voltages V_m^n will gradually become equal to $U(x_m,t_n) + f\{x_m,t_n,U(x_m,t_n)\}$. We still remark that if $f(x,t,z) = g(x)h(x,t,z)$, the function $g(x)$ can be interpreted by chosing $R_3 = R_1/(\Delta x)^2 g(x)$.

During the iteration process the last calculated injection voltage of each node has to be maintained untill the next calculation of this injection voltage. Keeping of the injection voltages is realized with the help of a set of analogue memories G_2. The switching of the function generator F_1 is performed by means of the set of switches S_3 and S_4.

In an analogue computer the calculations are often performed with a level of accuracy where the limit of the physical discrimination of the computer is nearly reached, because then the computer is used most efficiently. This means that the above iteration process has to be continued untill physically the voltages in the internal nodes of the network no longer change noticeably. So it is not possible to mark very accurately by observing when the iteration process is finished. Moreover an accurate check about the state of the iteration process by means of measuring is not attractive technically. Therefore, the iteration process is finished after a specified number of iteration cycles by a logical decision.

The computer is controlled by means of a central control unit B. This unit realizes all logic manipulations, e.g., the switching of the switches $S_1,\ldots,S_4$. The programming of the control unit is mainly a fixed one, but for instance the number of cycles per iteration process is adjustable.

We remark that if the differential equation in problem (1) is changed by adding a term $\frac{\partial z}{\partial x}$ then in the special analogue computer only the resistance network has to be changed. Also other types of boundary conditions can easily be realized. So we see that the described special analogue computer can also be used for other types of boundary and intial value problems with only few modifications.

4.2. Some technical details

The presence of many "parallel" arithmetic units in an analogue computer is no guarantee that the computing speed will be larger than that of a digital computer. This depends very much on the dynamic properties of the applied components.

In the special analogue computer a large number of analogue switches is used. These switches are closed and opened many times during the computation of the solution of a problem (4.1.1). The switching times contribute to the computing time. This means that the computing time can only be small if the switching times are also small. Therefore in the special analogue computer only analogue switches of electronic type have been applied [9].

An important demand for the technical design of the above special analogue computer has been to make the different kinds of analogue electronic circuits as simple as possible. The reason for this is that each of these circuits appears in a large number and so they form an important part of the cost of the special analogue computer. This demand included that no high accuracy could be the aim. In order to obtain a total accuracy of the computer of about one percent, all types of electronic circuits are designed for this accuracy. Because of the large number of electronic circuits, we can expect the sum of the individual errors not to differ much from the individual error itself if there are no important systematic errors.

All analogue and digital electronic circuits of the computer are transistorized.

In principle each of the analogue memory locations is a capacitor, which is read into and read out from the resistance network via buffer amplifiers.

Each analogue switch consists of one transistor which is used in voltage saturation or in current saturation. The time for opening or closing the applied switches is about 2 μsec.

Chapter 5

DIGITAL ITERATIVE METHODS FOR SOLVING DIFFUSION PROBLEMS

5.1. Introduction

Again we consider problems of type (4.1.1) and it will be assumed that such a problem will be interpreted by a difference problem in the same way as in chapter 3 before solving numerically.

At each time t_n the solution of a difference problem is fixed by $z(0,t_n)$, $z(1,t_n)$ and $z(x_m,t_{n-1})$, $m = 1,\dots,M-1$, as a set of equations. The special analogue computer described before determines the solutions at succeeding times t_n by solving these sets of equations one after another starting with the set of equations at $t = \Delta t$. Generally the equations of these sets are nonlinear as a consequence of the presence of the terms $f\{x_m,t_n,z(x_m,t_n)\}$, $m = 1,\dots,M-1$. As exposed in the last chapter the special analogue computer determines the solution of each set of equations by means of an iteration process.

This iteration process differs much with that one that occurs if the above mentioned sets of equations are solved iteratively with the help of a digital computer. For the special analogue computer we have that during the time in which the function generator is connected with the m^{th} node of the resistance network a set of equations is solved in parallel. However, for a digital computer we have that during the time of an iteration cycle in which computations are executed for the point x_m only one equation is solved.

We remark that before starting the iteration process concerning the time t_n it can be sensible to predict the solution as good as possible by means of the already determined solution at preceding times. In this chapter we assume that the prediction is such that only small changes of z occur during an iteration process so that it is allowed to linearize the iteration problem as for the behaviour of the iteration error as function of the number of iterations s.

In this chapter we will examine some digital iterative methods. We will see that the behaviour of the iteration processes is dependent on $\frac{\partial f}{\partial z}$. We assume that $\frac{\partial f}{\partial z} = p$ is independent of x in order that the x-elementary solutions of the iteration problems are simple functions of x.

First the general properties of the iteration processes will be examined. Only in paragraph 4 will we enter into more details for one special iteration process where we will also refer to the demand that Δt and Δx have to be small enough in order to obtain a sufficiently accurate interpretation of the differential problem.

5.2. Some iteration processes

We will restrict ourselves to those iteration processes for which in each iteration cycle only those values of z are used obtained in this cycle and in the preceding one. The order in each iteration cycle will be $x_1, x_2, \ldots, x_{M-1}$. We will denote the value of z in the point (x_m, t_n) obtained in the s^{th} iteration cycle by $z(x_m, t_n, s)$, while the iteration error $\{z(x_m, t_n, s) - z(x_m, t_n)\}$ will be denoted by $\varepsilon(x_m, s)$.

We observe that in advance it is not fixed uniquely which relations $\varepsilon(x_m, s)$ has to satisfy because of the free choice of the way in which the results of the $(s-1)^{th}$ iteration cycle are utilized in the s^{th} cycle. As already said, during an iteration cycle the digital computer solves one after another $(M-1)$ equations. In the m^{th} one of these equations the values of z in the points x_{m-1}, x_m and x_{m+1} appear. As for z in x_{m+1} we can only choose $z(x_{m+1}, t_n, s-1)$, but in x_{m-1} we have choice between $z(x_{m-1}, t_n, s-1)$ and $z(x_{m-1}, t_n, s)$. We will not limit ourselves to one special case now. In order to be able to consider the general case we introduce the symbol s^* having the meaning that for each term of the iteration equation containing s^*, s^* can be equal to s or $(s-1)$ dependent on the final choice of the iteration equation.

In this way it follows from (3.1.2) that the iteration problem that is solved during an iteration process can be written as

$$\frac{z(x_m, t_n, s_1{}^*) - z(x_m, t_{n-1})}{\Delta t} - a\, \frac{z(x_{m-1}, t_n, s_2{}^*) - 2z(x_m, t_n, s_3{}^*) + z(x_{m+1}, t_n, s-1)}{(\Delta x)^2} -$$

$$- (1-a)\, \frac{\delta^2 z(x_m, t_{n-1})}{(\Delta x)^2} = af\{x_m, t_n, z(x_m, t_n, s_4{}^*)\} +$$

$$+ (1-a) f\{x_m, t_{n-1}, z(x_m, t_{n-1})\},$$

$$z(0, t_n) = \varphi(t_n), \; z(1, t_n) = \eta(t_n)$$

$$z(x_m, t_n, 0) = \gamma(x_m, t_n), \tag{1}$$

where $\gamma(x_m, t_n)$ is the predicted solution at t_n.

Because at least one of the quantities s^* occurring in (1) has to be equal to s, there are 15 different kinds of iteration processes.

Linearizing iteration problem (1) gives

$$\frac{\varepsilon(x_m, s_1{}^*)}{\Delta t} - a\, \frac{\varepsilon(x_{m-1}, s_2{}^*) - 2\varepsilon(x_m, s_3{}^*) + \varepsilon(x_{m+1}, s-1)}{(\Delta x)^2} - ap\; \varepsilon(x_m, s_4{}^*) = 0$$

$$\varepsilon(0, s) = \varepsilon(1, s) = 0, \; \varepsilon(x_m, 0) = \gamma(x_m, t_n) - z(x_m, t_n). \tag{2}$$

It will be clear that we do not know $\varepsilon(x_m,0)$ in advance.

By definition the x-elementary solutions of (2) are $\alpha(x_m)g^s$, $g =$ constant $\neq 0$, provided $\alpha(x_m)$ satisfies the following set of linear equations

$$b\alpha(x_{m-1}) + c\alpha(x_m) + \alpha(x_{m+1}) = 0, \quad \alpha(0) = \alpha(1) = 0$$

$$b = g_2^*; \; c = -\frac{(\Delta x)^2}{a\Delta t} g_1^* - 2g_3^* + p(\Delta x)^2 g_4^*, \tag{3}$$

where g_i^*, $i = 1,\ldots,4$, is equal to g if $s_i^* = s$ and equal to 1 if $s_i^* = s-1$.

Denoting the matrix of the set of equations (3) by A, we can also write (3) as

$$A\underline{u} = \underline{0}, \; \underline{u} = \{\alpha(x_1),\ldots,\alpha(x_{M-1})\}. \tag{4}$$

There is only a nontrivial solution of (3) and (4), if the determinant of A, denoted by A_d, vanishes. In general $A_d = 0$ is equivalent whith a $(M-1)^{th}$ power equation in g. If all vectors $\underline{u}$ belonging to the $(M-1)$ roots g of this power equation are independent, then we can write $\varepsilon(x_m,s)$ as

$$\varepsilon(x_m,s) = \sum_{j=1}^{M} K_j \; \alpha_j(x_m)g_j^{\,s}. \tag{5}$$

In an elementary way we can easily prove that the x-elementary solutions of (2) are

$$b_j^{\frac{m}{2}} \sin j\pi x_m \; g_j^{\,s}, \tag{6}$$

provided g_j satisfies

$$c_j + 2\sqrt{b_j} \cos j\pi\Delta x = 0. \tag{7}$$

From this we can conclude immediately that for arbitrary constants b and c the eigenvalues μ_r and eigenvectors $\underline{w}_r$ of A are

$$\mu_r = c + 2\sqrt{b} \cos \frac{j\pi}{M}$$

$$\underline{w}_r = (w_{r1},\ldots,w_{r,M-1}), \quad w_{rm} = b^{\frac{m}{2}} \sin \frac{j\pi m}{M}. \tag{8}$$

5.3. Components of the iteration error

In this paragraph we will discuss briefly the possibilities for the components of the iteration error.

As already said in general $A_d = 0$ is equivalent with a $(M-1)^{th}$ power equation in g. The only exception occurs if s_1^*, s_3^* and s_4^* are equal to $(s-1)$, However, from (5.2.2) we see that in this case ε in x_{M-1} does not change at all during the iteration process and so we can not obtain the solution of (4.1.1) in difference form by means of this type of iteration process. Therefore we will leave this case out of consideration.

Concerning b there are two possibilities; b can be equal to 1 or equal to g. In the first case (5.2.7) yields one root g for each j. This means that the solution of (5.2.2) can be written as

$$\varepsilon(x_m, s) = \sum_{j=1}^{M-1} K_j g_s^{\,s} \sin j\pi x_m . \qquad (1)$$

If $b = g$, then (5.2.7) is a square equation in $\sqrt{g}$, which for any choice of c can be written as

$$(\sqrt{g_j})^2 + 2d_o \cos j\pi\Delta x\, (\sqrt{g_j}) + e = 0, \quad d_o, e = \text{real constants}. \qquad (2)$$

In the usual way we denote the roots of (2) by $(\sqrt{g_j})_1$ and $(\sqrt{g_j})_2$, while the squares of $(\sqrt{g_j})_1$ and $(\sqrt{g_j})_2$ will be denoted by g_{j1} and g_{j2}. From (5.2.8) we have that the contribution to the iteration error belonging to j can be written as

$$\{K_{j1}(\sqrt{g_j})_1^{m+2s} + K_{j2}(\sqrt{g_j})_2^{m+2s}\} \sin j\pi x_m . \qquad (3)$$

It can easily be shown that in the case $d_o < 0$ the sum of the contributions belonging to j and $M-j, j < \frac{M}{2}$ can be written as

$$K_1 g_{j1}^{\frac{m}{2}+s} \sin j\pi x_m + K_2 g_{j2}^{\frac{m}{2}+s} \sin(M-j)\pi x_m, \quad \text{if } e < 0; \qquad (4)$$

$$(K_1 g_{j1}^{\frac{m}{2}+s} + K_2 g_{j2}^{\frac{m}{2}+s}) \sin j\pi x_m, \quad \text{if } 0 < e < d_o^{\,2}\cos^2 j\pi\Delta x \qquad (5)$$

$$\{K_1 + K_2(\tfrac{m}{2}+s)\}\, g_{j1}^{\frac{m}{2}+s} \sin j\pi x_m, \quad \text{if } e = d_o^{\,2}\cos^2 j\pi\Delta x; \qquad (6)$$

$$K_1 |g_1|^{\frac{m}{2}+s} \sin\{(\tfrac{m}{2} + s)\arg g_{j1} + K_2\}\sin j\pi x_m, \text{ if } e > d_o^{\,2}\cos^2 j\pi\Delta x, \tag{7}$$

where K_1 and K_2 are arbitrary constants. If $d_o > 0$ then j and M - j have to be interchanged in the above expressions.

A special case arises if e = 0. Now we will consider this case in more detail. If e = 0, then all quantities s_i^* are equal to s and so we can write the system equation of (5.2.2) as

$$\varepsilon(x_m, s-1) = -\{1 + q(1 + \delta_+)\}\varepsilon(x_{m-2}, s)$$

$$q = \{-\frac{(\Delta x)^2}{a\Delta t} - 2 + p(\Delta x)^2\}, \tag{8}$$

where again δ_+ denotes the forward difference.

From (2) we see that $g_{j1} = 0$ and so there are left only approximately $\frac{M}{2}$ x-elementary solutions. This means that $\varepsilon(x_m,s)$ cannot be written as a linear combination of only x-elementary solutions. We shall show in the following that besides the x-elementary solutions there are still other functions of x_m and s satisfying the difference equation (8) and the boundary conditions $\varepsilon(0,s) = \varepsilon(1,s) = 0$.

We observe that for given $\varepsilon(x_m,s)$, m = 1,...,M - 1 the value of $\varepsilon(x_1,s-1)$ is not fixed uniquely by (8). This means that we can choose $\varepsilon(x_1,s-1)$ arbitrarily without changing $\varepsilon(x_m,s)$, m = 1,...,M - 1. In figure 1 a table shows the influence of the choice $\varepsilon(x_1,s-1) = 1$, $\varepsilon(x_1,s-i) = 0$, $i \geqq 2$ on the behaviour of $\varepsilon(x_m,s-i)$ as function of x_m and i, if, moreover, we have taken $\varepsilon(x_m,s) = 0$, m = 1,...,M - 1. If $\varepsilon(x_m,0)$ is taken equal to row i of this table, then the behaviour of $\varepsilon(x_m,s)$ as a function of x_m for s = 1,2,... is represented by the rows (i - 1), (i - 2),... of the table. We see that for this choice of $\varepsilon(x_m,0)$ the iteration error vanishes if $s \geqq i$. From (8) it can easily be derived that if $s \leqq i - 1$ this iteration error can be written as

$$(-1)^{i-s-1}\{1 + q(1 + \delta_+)\}^{i-s-1}\{U(x_{m-2i+2s+1}) - U(x_{m-2i+2s})\}, \tag{9}$$

where U(x) = 0, if x < 0 and U(x) = 1, if $x \geqq 0$.

row	iteration	x = 0	x	$2\Delta x$	$3\Delta x$	$4\Delta x$	$5\Delta x$	$6\Delta x$	$7\Delta x$	$8\Delta x$	$9\Delta x$
0	s	0	0	0	0	0	0	0	0	0	0
1	s-1	.	1	0	0	.	.	.	.	.	.
2	s-2	.	0	$-q$	-1	0	0	.	.	.	.
3	s-3	.	.	0	q^2	$2q$	1	0	0	.	.
4	s-4	0	0	0	0	$-q^3$	$-3q^2$	$-3q$	-1	0	0

Figure 1. Table of iteration errors becoming identically zero after a finite number of iterations.

All functions of type (9) are independent of each other. Because the total number of x-elementary solutions and functions of type (9) is equal to $(M-1)$, we conclude that in the case $e = 0$ the iteration error $\varepsilon(x_m, s)$ can be written as a linear combination of x-elementary solutions and functions of type (9).

5.4. A special iteration process

In this paragraph we will consider iteration problem (5.2.1) in the case that the most recent data are used in the iteration process. Because of the nonlinearity of $f(x,t,z)$ in general the only possibility is to take the value of z in this function equal to $z(x_m, t_n, s-1)$ unless solving of each iteration equation is done by means of a second iteration process. Excluding this possibility we have $s_1^* = s_2^* = s_3^* = s$ and $s_4^* = s-1$. For shortness sake we denote $z(x_m,t_n)$, $z(x_m,t_n,s)$, $f\{x_m,t_n,z(x_m,t_n)\}$ and $f\{x_m,t_n,z(x_m,t_n,s)\}$ respectively, by $z(m,n)$, $z_s(m,n)$, $f(m,n,z)$ and $f(m,n,z_s)$. Then from (5.2.1) the iteration formula can be obtained as

$$z_s(m,n) = C_1\{z_s(m-1,n) + z_{s-1}(m+1,n)\} + C_2 f(m,n,z_{s-1}) + F(m,n), \quad (1)$$

where

$$F(m,n) = C_3 z(m,n-1) + C_4\{z(m-1,n-1) + z(m+1,n-1)\} + C_5 f(m,n-1,z),$$

$$C_1 = \frac{a\gamma}{1+2a\gamma}, \quad C_2 = \frac{a\Delta t}{1+2a\gamma}, \quad C_3 = \frac{1-2(1-a)\gamma}{1+2a\gamma}, \quad C_4 = \frac{(1-a)\gamma}{1+2a\gamma},$$

$$C_5 = \frac{(1-a)\Delta t}{1+2a\gamma}, \quad \gamma = \frac{\Delta t}{(\Delta x)^2}. \quad (2)$$

The corresponding iteration formula for the iteration error is

$$\varepsilon(x_m,s) = C_1\{\varepsilon(x_{m-1},s) + \varepsilon(x_{m+1},s-1)\} + C_2 p\varepsilon(x_m,s-1). \quad (3)$$

From this it follows that the constants d_o and e in (5.3.2) are

$$d_o = -C_1, \quad e = -C_2 p. \tag{4}$$

We see that e is proportional to p, so the behaviour of the iteration process depends much on the value of p.

$p > 0$

If $p > 0$ it follows from (5.3.4), that $\varepsilon(x_m,s)$ can be written as

$$\varepsilon(x_m,s) = \sum_{j=1}^{M-1} D_j \, g_{j,1}^{\frac{m}{2}+s} \sin j\pi x_m, \tag{5}$$

where the constants D_j are such that (5) also holds for $s = 0$. From (5.3.2) it follows that $g_{1,1}$ is the largest growthfactor and so the iteration process is convergent if $g_{1,1} < 1$. This means by virtue of (5.3.2), (2) and (4) that for $p > 0$ the iteration process is convergent if

$$p < \frac{1}{a\Delta t} + 2\,\frac{(1-\cos\pi\Delta x)}{(\Delta x)^2} = p_1 . \tag{6}$$

$p(\Delta x)^2 < -C_1\cos^2\pi\Delta x$

The discriminant of the square equation (5.3.2) is negative for each j, if $p(\Delta x)^2 < -C_1\cos^2\pi\Delta x$. Then all growthfactors are complex, except $g_{\frac{M}{2}} = -e$ if M = even. Considering only the case M = odd from (5.3.7) we have that in this p-region $\varepsilon(x_m,s)$ can be written as

$$\varepsilon(x_m,s) = \sum_{j=1}^{\frac{M-1}{2}} D_j |g_{j,1}|^{\frac{m}{2}+s} \sin\{(\tfrac{m}{2}+s)\arg g_{j1} + \phi_j\}\sin j\pi x_m, \tag{7}$$

where again the constants D_j and ϕ_j are fixed by $\varepsilon(x_m,0)$. Because $|g_j| = e$ for each j, we have that the iteration process is convergent in this p-region, if

$$p > -\frac{1+2a\gamma}{a\Delta t} = p_2 . \tag{8}$$

$p = 0$

In the case $p = 0$ for each j one of the roots vanishes and so $\varepsilon(x_m,s)$

is a linear combination of x-elementary solutions $g_{j1}^{\frac{m}{2}+s} \sin j\pi x_m$, $j = 1, \ldots, \frac{M-1}{2}$ and of functions of type (5.3.9).

$- C_1 \cos^2 \pi\Delta x < p(\Delta x)^2 < 0$

In the p-interval $- C_1 \cos^2 \pi\Delta x < p(\Delta x)^2 < 0$ there can arise three possibilities for the component of the iteration error belonging to some value of j. Firstly, if the discriminant of (5.3.2) is positive then the contribution belonging to j is given by (5.3.5). Secondly, if the discriminant is negative then this contribution is given by (5.3.7). The third possibility is that the discriminant vanishes for one value of j. Then for this value of j the contribution belonging to j is given by (5.3.6). Leaving out of consideration the third possibility, $\varepsilon(x_m, s)$ can be written in the p-interval $- C_1 \cos^2 \pi\Delta x < p(\Delta x)^2 < 0$ as

$$\varepsilon(x_m, s) = \sum_{j=1}^{J} \left(D_{j1} g_{j1}^{\frac{m}{2}+s} + D_{j2} g_{j2}^{\frac{m}{2}+s}\right) \sin j\pi x_m + $$

$$+ \sum_{j=J+1}^{\frac{M-1}{2}} D_j |g_{j1}|^{\frac{m}{2}+s} \sin\left\{\left(\frac{m}{2}+s\right)\arg g_{j1} + \psi_j\right\} \sin j\pi x_m . \qquad (9)$$

Again $g_{1,1}$ is the in absolute value largest growthfactor and so from (6) it follows that in the last considered p-interval the iteration process is convergent for each p.

From the above we can conclude that the considered iteration process is convergent if p satisfies (6) as well as (8), so if

$$- \frac{1+2a\gamma}{a\Delta t} < p < \frac{1}{a\Delta t} + 2\,\frac{(1-\cos \pi\Delta x)}{(\Delta x)^2} . \qquad (10)$$

We observe that the iteration process can always be made convergent by choosing Δt small enough.

In the next chapter it will prove that p_1 as given in (6) is also a bound for the convergence of the iteration process occurring in the special analogue computer described in chapter 4.

In the preceding we have examined the iteration processes without paying much attention to a good interpretation of the given differential problem by the corresponding difference problem. Now we will consider this aspect in relation to the choice of the type of the iteration process. For a good interpretation at least the growthfactors of the 1st elementary solutions of the differential and corresponding difference problem must be nearly equal. From (2.4.6) and (3.4.1) it follows that for that purpose at least $p\Delta t$ has to satisfy if $a = \frac{1}{2}$

$$-1 + \gamma(1 - \cos \pi\Delta x) \ll \frac{p\Delta t}{2} \ll 1 + \gamma(1 - \cos \pi\Delta x). \qquad (11)$$

This condition can be satisfied for each value of p by choosing Δt sufficiently small.

Regardless of the choice of γ, it can be shown that in the p-interval given by (11) only the iteration processes are convergent for which $(s_1^*, s_2^*, s_3^*, s_4^*)$ equals (s,s,s,s), (s,s,s,s-1), (s-1,s,s,s) or (s-1,s,s,s-1). For p = 0 in the same order the in absolute value largest growthfactor of these iteration processes prove to be if $a = \frac{1}{2}$

$$\frac{\gamma^2\cos^2\pi\Delta x}{(1+\gamma)^2}, \quad \frac{\gamma^2\cos^2\pi\Delta x}{(1+\gamma)^2}, \quad \frac{\gamma\cos\pi\Delta x}{1+\gamma}, \quad \frac{\gamma\cos\pi\Delta x}{1+\gamma}. \qquad (12)$$

From this we see that the iteration processes (s,s,s,s) and (s,s,s,s-1) are much more favourable than the other two. This is also true if $p \neq 0$. As already said the iteration process (s,s,s,s) cannot be applied without introducing a second iteration process. So the iteration process (s,s,s,s-1) is the best one which can be chosen for a digital computer if a difference problem of the considered type has to interpret a given differential problem. For this iteration process the graph $|g|_{max}$ as function of p is given in figure 1 for Δt = 0,1 and $\Delta x = \frac{1}{4}$.

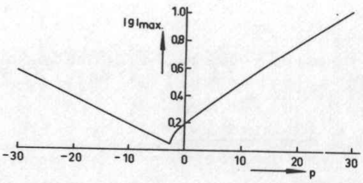

Figure 1. Curve $|g|_{max}$ as function of p for Δt = 0,1 and $\Delta x = \frac{1}{4}$.

Chapter 6

ITERATION PROCESS APPLIED IN THE SPECIAL ANALOGUE COMPUTER

6.1. Introduction

In this chapter we will discuss the iteration process for the special analogue computer described in chapter 4.

We will use $z_r(m,s)$ to denote the value of z in the point (x_m,t_n) obtained during the s^{th} iteration cycle in the time that the analogue function generator is connected with the node x_r of the resistance network. The difference $z_r(m,s) - z(x_m,t_n)$ will be called the iteration error $\varepsilon_r(m,s)$. We will restrict ourselves to the iteration process in which the most recent data are used. In that case during the time that the function generator is used in node x_r the special analogue computer solves simultaneously the following set of equations.

$$- C_1 z_r(m-1,s) + z_r(m,s) - C_1 z_r(m+1,s) - C_2 f\{x_m, z_m(m,s)\} = F(m,n), \quad 0 < m \leqq r$$

$$- C_1 z_r(m-1,s) + z_r(m,s) - C_1 z_r(m+1,s) - C_2 f\{x_m, z_m(m,s-1)\} = F(m,n), \quad r < m < M$$

$$z_r(x_o,s) = \varphi(t_n), \quad z_r(x_M,s) = \eta(t_n), \qquad (1)$$

where $F(m,n)$ and the constants C_1, C_2 have the same meaning as in (5.4.2).

From this we can derive by linearizing that the iteration error satisfies

$$A\, \underline{w}_{rs} = - \frac{C_2 p}{C_1} \underline{u}_{rs}; \quad \underline{w}_{rs} = \{\varepsilon_r(1,s),\ldots,\varepsilon_r(M-1,s)\},$$

$$\underline{u}_{rs} = \{\varepsilon_1(1,s),\ldots,\varepsilon_r(r,s),\ \varepsilon_{r+1}(r+1,s-1),\ldots,\varepsilon_{M-1}(M-1,s-1)\}, \qquad (2)$$

where A is a tridiagonal matrix, having coefficients $a_{i+1,i} = a_{i,i+1} = 1$ and $a_{ii} = -1/C_1 = c$.

In one iteration cycle the set of equation (1) is solved one after another for $r = 1,\ldots,M-1$. These $(M-1)$ sets of equations form together the following set of $(M-1)^2$ equations:

$$\begin{pmatrix} A+E_1 & 0 & \cdots & 0 \\ E_1 & A+E_2 & \ddots & \vdots \\ \vdots & E_2 & \ddots & 0 \\ E_1 & E_2 & \cdots & A+E_{M-1} \end{pmatrix} \begin{pmatrix} \underline{w}_{1s} \\ \vdots \\ \underline{w}_{M-1,s} \end{pmatrix} = - \begin{pmatrix} 0 & E_2 & E_3 & \cdots & E_{M-1} \\ \vdots & \ddots & E_3 & & \vdots \\ \vdots & & \ddots & \ddots & E_{M-1} \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix} \begin{pmatrix} \underline{w}_{1s-1} \\ \vdots \\ \underline{w}_{M-1,s-1} \end{pmatrix} \tag{3}$$

where all coefficients e_{ij} of E_k, $k = 1,\dots,n$, vanish except e_{kk} which is equal to $C_2 p/C_1$. From this we see that the above iteration process is equivalent with a successive block iteration process for the determination of the solution of the following set of equations

$$\begin{pmatrix} A+E_1 & E_2 & \cdots & E_{M-1} \\ E_1 & A+E_2 & & \vdots \\ \vdots & E_2 & \ddots & E_{M-1} \\ E_1 & E_2 & \cdots & A+E_{M-1} \end{pmatrix} \begin{pmatrix} \underline{w}_1 \\ \vdots \\ \underline{w}_{M-1} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} . \tag{4}$$

We observe that the solution itself of (1) is not utilized at all in a direct way. When the analogue function generator is switched to node $(r + 1)$ of the resistance network, only the value of $f\{x,t,z_r(r,s)\}$ is hold. This means that indirectly from the solution of (1) only $z_r(r,s)$ is utilized. Therefore we will restrict ourselves to the determination of $\varepsilon_m(m,s)$, $m = 1,\dots,M-1$. Denoting the coefficients of A^{-1}, the inverse of A, by d_{ij} and writing (m,s) instead of $\varepsilon_m(m,s)$ we find from (2) that $\varepsilon(m,s)$, $m = 1,\dots,M-1$ satisfies

$$d_{m1}\varepsilon(1,s) + \dots + (d_{mm} + d)\varepsilon(m,s) + d_{m,m+1}\varepsilon(m+1,s-1) + \dots + d_{m,M-1}\varepsilon(M-1,s-1) = 0, \tag{5}$$

where $d = \dfrac{C_1}{C_2 p}$. Denoting the vector $\{\varepsilon(1,s),\dots,\varepsilon(M-1,s)\}$ by $\underline{\varepsilon}_s$, the set of equations (5) can also be written as

$$B\,\underline{\varepsilon}_s + C\,\underline{\varepsilon}_{s-1} = \underline{0}\ , \tag{6}$$

where the coefficients of the matrices B and C are

$$b_{ij} = 0, \quad j > i; \quad b_{ij} = d_{ij}, \quad j < i; \quad b_{ii} = d_{ii} + d$$

$$c_{ij} = 0, \quad j \leqq i; \quad c_{ij} = d_{ij}, \quad j > i. \tag{7}$$

By definition the elementary solutions of (6) are of the form $g^s \underline{v}$, provided g and $\underline{v}$ satisfy

$$C \, \underline{v} = - g B \, \underline{v} \, . \tag{8}$$

This means that $-g$ and $\underline{v}$ have to be an eigenvalue and an eigenvector of a general eigenvalue problem. From (7) and (8) we have that g is a root of

$$\begin{vmatrix} (d_{11}+d)g & d_{12} & \cdots & d_{1,M-1} \\ d_{21}\,g & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & d_{M-2,M-1} \\ d_{M-1,1}\,g & \cdots & d_{M-1,M-2}\,g & (d_{M-1,M-1}+d)g \end{vmatrix} = 0 \, . \tag{9}$$

We see that always one of the roots of (9) vanishes, which root will be denoted by g_{M-1}. For some special cases it is possible to determine the roots of (9) analyticly. However, we will not discuss these special cases and limit ourselves in the next paragraphs to find the region of convergence of the iteration process and the approximated behaviour of g as function of p.

6.2. Coefficients of matrix A^{-1}

It will be clear that before the roots of (6.1.9) can be determined, first we have to find the coefficients d_{ij} of matrix A^{-1}.

By definition it must hold

$$A \, A^{-1} = I, \quad I = \text{identical matrix}. \tag{1}$$

From the coefficients of A, look at (6.1.2), we can easily derive by multiplying the i^{th} row of A and the j^{th} column of A^{-1} that the coefficients d_{ij} of A^{-1} satisfy

$$\begin{aligned} d_{i-1,j} + c d_{ij} + d_{i+1,j} &= 0 \quad \text{if } i \neq j \\ &= 1 \quad \text{if } i = j, \end{aligned} \tag{2}$$

provided $d_{o,j}$ and $d_{M,j}$ are assumed to be zero.

The elementary solutions of (2) are $G_{1,2}^{i}$, where G_1 and G_2 are the roots of the square equation

$$G^2 + cG + 1 = 0. \tag{3}$$

For problem (2) the homogeneous system equation is valid in two different i-intervals, namely $0 < i < j$ and $j < i < M$. So problem (2) can be conceived to be the mathematical description of the state of two bounded physical systems of equal construction but of different length, having interaction in the point $i = j$. The interaction condition is $d_{j-1,j} + cd_{jj} + d_{j+1,j+1} = 1$. The conditions $d_{oj} = d_{Mj} = 0$ are boundary conditions of the first respectively the second system. The general solutions of both systems are linear combinations of $G_1^{\,i}$ and $G_2^{\,i}$. So the solution of (2) can be written as

$$\begin{aligned} d_{ij} &= k_{j1}G_1^{i-j} + k_{j2}G_2^{i-j}, \quad 0 \leqq i \leqq j \\ &= K_{j1}G_1^{i-j} + K_{j2}G_2^{i-j}, \quad j \leqq i \leqq M , \end{aligned} \tag{4}$$

provided

$$\begin{aligned} G_1^{-j}k_{j1} + G_2^{-j}k_{j2} &= 0 \\ k_{j1} + k_{j2} - K_{j1} - K_{j2} &= 0 \\ (c + G_1^{-1})k_{j1} + (c + G_2^{-1})k_{j2} + G_1K_{j1} + G_2K_{j2} &= 1 \\ G_1^{M-j}K_{j1} + G_2^{M-j}K_{j2} &= 0 . \end{aligned} \tag{5}$$

Introducing as abbreviation

$$V_i = G_1^{\,i} - G_1^{\,-i}, \tag{6}$$

it can easily be derived that the solution of (4) and (5) is

$$d_{ij} = -\frac{V_i V_{M-j}}{V_1 V_M}, \quad i \leqq j; \quad d_{ij} = -\frac{V_j V_{M-i}}{V_1 V_M}, \quad i \geqq j. \tag{7}$$

From (5.4.2) because of the relation $cC_1 = -1$ we see that always $c < -2$ and that c becomes more negative the smaller $a\gamma$ is. In the case $c \ll -2$, we have from (3) that $G_1^{-1} \approx 0$ and so it follows from (6) and (7)

$$d_{ij} \approx \frac{-1}{G_1^{\,|i-j|+1}} \ . \tag{8}$$

This means that d_{ij} is approximately constant on each of the diagonals $i - j$ = constant. The coefficients of a diagonal decrease by choosing the diagonal more far away from the principal diagonal $i = j$. This reduction is about a factor $\frac{1}{G_1}$ per diagonal.

As illustration in figure 1 the behaviour of V^i is given as function of i for $a\gamma = 5$. We observe that even for this large value of $a\gamma$ the approximation (8) can already be applied for small values of i and j.

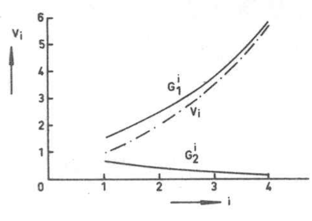

Figure 1. Curve of V_i as function of i for $\Delta t = 0{,}1$ and $a = \frac{1}{2}$.

6.3. Convergence of the iteration process

Now we will determine the region of convergence of the iteration process applied in the special analogue computer.

For one part of this region of convergence it can the most easily be done by regarding the iteration process from a physical point of view. Let us consider a physical system with unknown z of which the steady state is described by the analogue circuitry of figure 1.

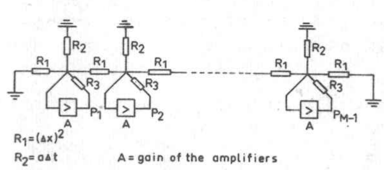

Figure 1. A fundamental system belonging to the analogue iteration process.

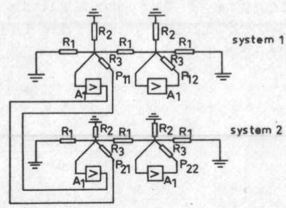

Figure 2. A way of coupling two fundamental systems.

The condition of stability of this circuitry can easily be obtained by conceiving the physical system to be time dependent in the way of (3.1.2) for a = 1. Then it follows from (3.1.4) that the physical system of figure 1 is stable if

$$A < 1 + \frac{1}{a\Delta t} + \frac{2(1-\cos \pi\Delta x)}{(\Delta x)^2} = A_1 . \tag{1}$$

In the special case $A = A_1$ the open loop gain in each of the points P_i, $i = 1,\dots,M-1$, is equal to 1. This means that a system consisting of some systems of the type of figure 1 coupled by replacing in some way outputs of amplifiers by the output of a corresponding one is also stable. Taking for instance the coupled system of figure 2, the open loop gain in P_{11} as well as P_{21} is equal to 1, because both are the squares of the open loop gain in P_1 of the fundamental system.

If in the above mentioned type of a coupled system the amplifiers have input-memories and the interconnections between the individual systems are only realized for a while periodicly one after another as closed loops by means of switches at the inputs of the amplifiers, then only the transient will change, but the new physical system remains stable for $A = A_1$. We observe that the iteration process applied in the special analogue computer is such a physical system, what is illustrated in figure 3 for M = 3.

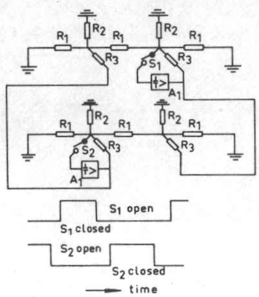

Figure 3. The analogue iteration process as a time-dependent coupling of two fundamental systems.

In the case $A = A_1$ we can say that regarded from one individual system all other ones are non-active. If $|A| < A_1$ then surely all individual systems are passive and so each of the above class of coupled systems is stable. Thus we know in advance that the iteration process applied in the special analogue computer is convergent if

$$- 2 - \left\{ \frac{1}{a\Delta t} + \frac{2(1-\cos \pi\Delta x)}{(\Delta x)^2} \right\} < p < \frac{1}{a\Delta t} + \frac{2(1-\cos \pi\Delta x)}{(\Delta x)^2} = p_1 . \qquad (2)$$

The remaining part of the region of convergence can only be determined by utilizing the special proporties of the considered iteration process. Now we will outline how convergence can be proven for $-\infty < p < 0$. From (6.2.7) it follows that all coefficients of the first column of (6.1.9) except the first two will vanish by substracting $(V_1/V_2) \cdot 2^{nd}$ column. Likewise all coefficients of the first row of (6.1.9) can be made equal to zero except the first two. Continuing in this way we find, if D_{M-i} is used to denote the determinant arising from that of (6.1.9) by omitting the first $(i-1)$ rows and columns, as recurrent relation between three successive D's

$$D_{M-i} - P_i D_{M-i-1} + Q_i D_{M-i-2} = 0, \quad i = 1,\dots,M-3, \qquad (3)$$

where

$$P_i = \{d_{ii} + d(1+\frac{V_i^2}{V_{i+1}^2})\}g - \frac{V_i}{V_{i+1}} d_{ii+1}, \quad Q_i = -\frac{V_i}{V_{i+1}} dg\{d_{ii+1} - \frac{V_i}{V_{i+1}}(d_{i+1i+1}+d)g\}. \tag{4}$$

We see that for $d = 0$ or in other words $p = \pm\infty$ all Q_i's vanish. In this case we can easily find that the roots g of (6.1.9) are

$$g_j = g_{M-j-1} = \frac{V_j V_{M-j-1}}{V_{j+1} V_{M-j}}, \; j = 1,\ldots,\frac{M-3}{2}; \; g_{\frac{M-1}{2}} = V^2_{\frac{M-1}{2}} / V^2_{\frac{M+1}{2}} \text{ if } M = \text{odd},$$

$$g_j = g_{M-j-1} = \frac{V_j V_{M-j-1}}{V_{j+1} V_{M-j}}, \; j = 1,\ldots,\frac{M-2}{2} \text{ if } M = \text{even}. \tag{5}$$

From (6.2.3) and (6.2.6) it follows that all these roots are positive and smaller than 1. So the iteration process is convergent for $p = \pm\infty$. All these roots are approximately equal to $\frac{1}{G^2}$.

Now we will prove that the iteration process is convergent for each $p < 0$ by showing that in this p-interval the rate of convergence increases by enlarging p. From (6.1.7) we see that for given $\underline{\varepsilon}_{s-1}$ the solution $\underline{\varepsilon}_s$ of (6.1.6) depends on the value of d. Diminishing d by a small quantity $-\Delta$, $\Delta < 0$, the solution of (6.1.6) will change. Denoting this solution by $\underline{\varepsilon}_s^*$ and introducing as abbreviations

$$G_i = \frac{d_{ii}+d}{d_{ii}+d+\Delta}, \quad E_i = \frac{d+\Delta}{d_{ii}+d+\Delta}\frac{V_{M-i}}{V_{M-i+1}}, \quad E_i^* = \frac{d+\Delta}{d_{ii}+d+\Delta}$$

$$F_i = \frac{d_{i-1i}}{d_{ii}+d+\Delta}, \quad H_i = \frac{\Delta}{d_{ii}+d+\Delta}, \quad F_i^* = \frac{d_{ii}}{d_{ii}+d+\Delta} \tag{6}$$

it can easily be found from (6.1.6) that

$$\varepsilon^*(i,s) = G_i\,\varepsilon(i,s) + \sum_{j=1}^{i-1} F_{j+1}E_{j+2}\ldots E_i H_j\,\varepsilon(j,s). \tag{7}$$

By squaring (7) if we apply $2|\varepsilon(i,s)\varepsilon(j,s)| \leqq \varepsilon^2(i,s) + \varepsilon^2(j,s)$ and neglect terms proportional to Δ^2 we can obtain

$$\sum_{i=1}^{M-1} \{\varepsilon^*(i,s)\}^2 \leqq \sum_{i=1}^{M-1} \{\varepsilon(i,s)\}^2. \tag{8}$$

We see that the length of the vector solution $\underline{\varepsilon}_s$ of (6.1.6) decreases for given $\underline{\varepsilon}_{s-1}$ if d is diminished. This means because this holds for arbitrary $\underline{\varepsilon}_{s-1}$ that for negative p the rate of convergence increases if p is enlarged. From this and (2) it follows that the iteration process applied in the special analogue computer is convergent if $p < p_1$.

From the convergence for $p = +\infty$ it follows that the iteration process will also be convergent for large values of p, but this region of convergence will be left out of consideration here.

6.4. Approximation of g(p)

The coefficients P_i and Q_i of the recurrent relation (6.3.3) are functions of i. However at the end of paragraph 6.2 we have seen that in practice these coefficients are nearly constant. This means that it has sense to consider (6.3.3) in the case that P_i and Q_i are replaced by P and Q, being the averages in some meaning of P_i and Q_i over the interval $i = 1, \ldots, M-3$.

For convenience sake we divide all rows of D_{M-1} by $\sqrt{Q}$ before evaluating this determinant. Denoting the thus obtained determinant by D^*_{M-1} we obtain a normalized recurrent relation

$$D_n^* - P^* D_{n-1}^* + D_{n-2}^* = 0, \quad n = 3, \ldots, M-1, \tag{1}$$

where $P^* = -P/\sqrt{Q}$. The determinant D^*_{M-1} is a polynomial in P^* of degree $M-1$. So $D^*_{M-1} = 0$ has $M-1$ roots P^*. Each root P^* represents an equation in g. The roots g of all these equations together will be used as approximations of the roots g of (6.1.9).

The determinants D_n^* are tridiagonal ones of which the coefficients on one and the same diagonal are constant except the last coefficient of the main diagonal. So it will be clear that we can adjoin a boundary and initial value problem to $D^*_{M-1} = 0$ in the same way as is mentioned in paragraph 2.3 for the determination of the functions $g_k^*(x_m)$. However, we will not do this here because above D^*_{M-1} is already defined as the solution in $n = M - 1$ of an initial value problem given by difference equation (1) and the initial values D_1^* and D_2^*.

The elementary solutions of (1) are λ^n, provided λ is equal to one of the roots λ_1 and λ_2 of

$$\lambda^2 - P^*\lambda + 1 = 0. \tag{2}$$

From this we find the solution of (1) as

$$D_n^* = K_1 \lambda_1^n + K_2 \lambda_2^n ;$$

$$K_1 = \frac{\lambda_2 - D_1^*}{\lambda_1\lambda_2(\lambda_2 - \lambda_1)}, \quad K_2 = \frac{\lambda_1 - D_1^*}{\lambda_1\lambda_2(\lambda_1 - \lambda_2)} . \tag{3}$$

Vanishing of D_{M-1}^* is aequivalent with

$$(\frac{\lambda_1}{\lambda_2})^{M-1} = - \frac{K_2}{K_1} = \frac{\lambda_1 - D_1^*}{\lambda_2 - D_2^*} . \tag{4}$$

This equality is satisfied by $\lambda_1 = e^{i\theta}$, $0 < \theta < \pi$, provided

$$\sin(M - 2)\theta - D_1^* \sin(M - 1)\theta = 0 . \tag{5}$$

We observe that D_1^* is a function of θ. So it will be clear that the transcedental equation (5) cannot be solved in general. But it can easily be shown that the roots θ_k, $k = 1,\ldots,M-2$ of (5) satisfy the unequality

$$\frac{(k-1)\pi}{M-1} < \theta_k < \frac{(k+1)\pi}{M-1} . \tag{6}$$

From $\lambda_1 = e^{i\theta}$ it follows that $P^* = 2\cos\theta$ or in other words

$$P^2 - 4Q\cos^2\theta = 0 . \tag{7}$$

From (6.3.4) it follows that P and Q can be written in the form

$$P = (\alpha_1 d + \alpha_2)g + \alpha_3, \quad Q = (\beta_1 d^2 + \beta_2 d)g^2 + \beta_3 dg, \tag{8}$$

where $\alpha_1,\ldots,\alpha_3$ and $\beta_1,\ldots,\beta_3$ are the averages of the corresponding quantities of (6.3.4). Substituting (8) into (7) yields

$$\{(\alpha_1^2 - 4\beta_1\cos^2\theta)d^2 + (2\alpha_1\alpha_2 - 4\beta_2\cos^2\theta)d + \alpha_2^2\}g^2 +$$

$$+ \{(2\alpha_1\alpha_3 - 4\beta_3\cos^2\theta)d + 2\alpha_2\alpha_3\}g + \alpha_3^2 = 0 . \tag{9}$$

The values of θ, satisfying (5), are functions of d which contain M as parameter. But it is sure that $0 \leqq \cos^2\theta \leqq 1$ independent on M. This means that for each M the maximum in the interval $0 \leqq \cos^2\theta \leqq 1$ of the in absolute value largest root g of (9) is an upper bound of the absolute values of the roots g of (9) for the right values of $\cos^2\theta$. For large values of M this upper bound is nearly the lowest upper bound.

It will be clear that for small M this upper bound is only a rough one. In this case we can better apply the following approximation. First the coefficients d_{ij} of (6.1.9) are approximated as given in (6.2.8). Then the contribution to the left member of (6.1.9) descended from those terms of the determinant containing the same coefficients of the principal diagonal can be written as a polynomial in $1/G_1$ multiplied by these coefficients. The lowest power of $\frac{1}{G_1}$ in this polynomial occurs for those terms only containing coefficients of the three inner diagonals. Secondly we limit the above polynomials in $1/G_1$ to the lowest power of $1/G_1$. In fact this means that we equate all coefficients to zero except those of the three inner diagonals. By these approximations (6.1.9) is passed into a special case of (5.2.3). From (5.2.7) we find that about the half of the roots g of (6.1.9) can be written as

$$g_j \approx \frac{4 \cos^2 j\pi\Delta x}{G_1{}^2(M^2G_1 - p)}\, p^2 = g_j{}^*, \quad j = 1,\ldots,\approx \frac{M}{2}, \tag{10}$$

while the other roots vanish approximately. We see that the in absolute largest growthfactor occurs for j = 1. As illustration in figure 1 the curve $g_1(p)$ as well as the curve $g_1{}^*(p)$ is given for $\Delta t = 0{,}1$ and $\Delta x = \frac{1}{4}$, so for the same parameters as in figure (5.4.1). In figure 2 for the same case all roots of (6.1.9) are represented as function of p.

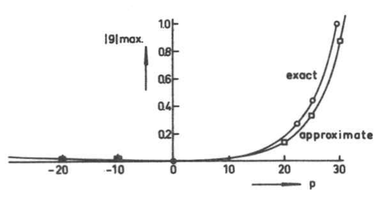

Figure 1. Exact and approximate curve of $|g|_{max}$ as a function of p for $\Delta t = 0{,}1$ and $\Delta x = \frac{1}{4}$.

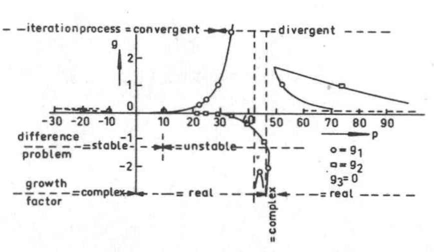

Figure 2. Growth factors g as function of p for $\Delta t = 0{,}1$ and $\Delta x = \frac{1}{4}$.

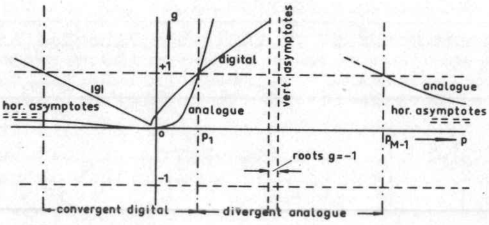

Figure 3. A sketch of $|g|_{max}$ as function of p for the analogue iteration process as well as for a digital one.

In figure 3 the behaviour of the iteration process for a digital as well as for the special analogue computer is characterized by means of sketches of the in absolute value largest growthfactor as function of p. We see that the region of convergence for the analogue computer is much larger than for a digital computer. Moreover within the common region of convergence the rate of convergence for the analogue computer is considerably larger than for a digital computer.

Chapter VII

NUMERICAL RESULTS

7.1. Introduction

In this chapter we will only discuss some numerical calculations performed to check analytic results. All these calculation have been executed on the digital computer of the Technical University of Delft.

One after another we will consider the following aspects of solving numerically a boundary and initial value problem.

1) Stability of a problem described by a pair of simultaneous "parabolic" differential equations,

2) Convergence of the iteration process occurring for this problem,

3) Fitting of a difference problem to a differential problem.

7.2. Flowdiagram for digital solving of "diffusion" problems having a set of differential equations

In figure 1 the flow diagram is represented according to which some programs have been made for digital solving of problems of the following type

$$\frac{\partial z_i}{\partial t} - a_i \frac{\partial^2 z_i}{\partial x^2} = f_i(x,t,z\ ,\dots,z_N), \qquad i = 1,\dots,N$$

$$z_i(0,t) = \varphi_i(t),\ z_i(1,t) = \eta_i(t);\ z_i(x,0) = \gamma_i(x). \qquad (1)$$

We remark that no efforts have been made to obtain an optimum efficiency of the flowdiagram and the corresponding programs because they were destined for research calculations.

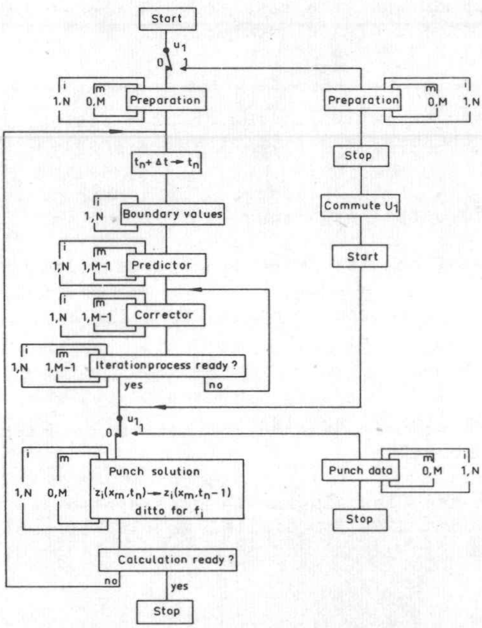

Figure 1. Flow diagram of a digital method of solving diffusion problems of type (1).

On behalf of the calculation of the solution at time t_n for each unknown z_i, i = 1,...,N and for each x_m, m = 1,...,M - 1 a quintet of quantities have to be kept. Moreover the total number of sub-programs is proportional to N. It is evident that a demand for the design of the main program had to be that the number of unknowns N and the number of x-intervals M can be chosen arbitrarily without changing the main program itself. We remark that for that purpose the holding of the mentioned set of quintets and the labeling of the sub-programs had to be arranged in a special way.

Now we will discuss what happens if switch U_1 is in position 0. After starting the calculation first a number of constants are calculated and kept. Further, some quantities relating to the initial time t_o are punched and the time is enlarged with Δt. Then successively the left and right boundary conditions, the prediction of the solution, and the terms $f(x,t,z_1,\ldots,z_N)$ corresponding to the predicted solution are calculated.

After this the iteration process starts where the predicted solution is the initial condition of the iteration process. With the help of an iteration formula (corrector) in each iteration cycle z_i and f_i, $i = 1, \ldots, N$ are corrected. Next it is checked if this corrected solution z_i for each x_m, $m = 1,\ldots,M-1$ differs less than δ from the solution z_i calculated in the preceding iteration. If the check is negative then one more iteration cycle is executed. The iteration process is stopped as soon as the check is positive. The results of the iteration process are punched and moreover they are read into the memory locations reserved for the solution of the preceding time.

Next the check "ready" is performed. If this check is negative then the main program is repeated from the instruction which enlarges the time with Δt.

7.3. Stability of a problem described by two simultaneous "diffusion" equations

In paragraph 2.10 we have discussed briefly the mathematical formulation of the external and internal solution for problems having two simultaneous "diffusion" equations. Now we will consider the analytic and numerical solution of the following example

$$\frac{\partial z_1}{\partial t} - 4\frac{\partial^2 z_1}{\partial x^2} + 19{,}9\pi^2 z_1 - 18\pi^2 z_2 = -13{,}025\pi^2 x^2 + 3{,}8\pi^2 t$$

$$\frac{\partial z_2}{\partial t} - \frac{\partial^2 z_2}{\partial x^2} + 18\pi^2 z_1 - 13{,}1\pi^2 z_2 = -8{,}6\pi^2 x^2 + 9{,}8\pi^2 t$$

$$z_1(0,t) = 2t \quad z_1(1,t) = \tfrac{1}{4} + 2t \quad z_1(x,0) = \tfrac{1}{4}x^2 + \sin 2\pi x$$

$$z_2(0,t) = 2t \quad z_2(1,t) = 1 + 2t \quad z_2(x,0) = x^2 + 2\sin 2\pi x. \qquad (1)$$

In paragraph 2.4 we have seen that for problems with one diffusion equation the 1st x-elementary solution has always the smallest stability. For problems with two simultaneous "diffusion" equations this is no longer true in general.

Denoting the largest root of (2.10.10) by μ_1 and $(j\pi)^2$ by y, it can easily be shown that the possibilities of the behaviour of μ_1 as function of y are as represented in figure 1.

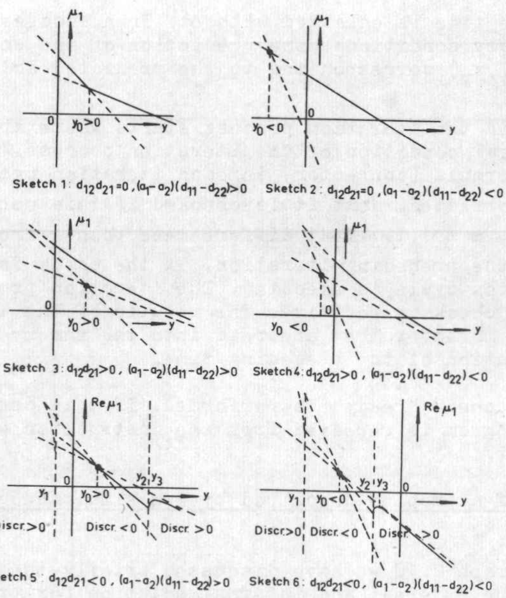

Figure 1. Different possibilities of $\mu_1(y)$.

In the sketches of figure 1 the curve $\mu_1(y)$ is only of interest for $y > 0$. In each sketch the straight lines are the asymptotes of $\mu_1(y)$ if $y \to \pm\infty$. For the sketches 1,...,4 one of the 1st elementary solutions has the largest growthfactor. This is not always true for the sketches 5 and 6, dependent on the choice of the constants a and d.

For problem (1) it holds $d_{12}d_{21} < 0$, $(a_1 - a_2)(d_{11} - d_{22}) > 0$. So the behaviour of μ_1 as function of y is as given in sketch 5. The constants a and d have been chosen such that the relative maximum is positive and occurs for $y_3 = 4\pi^2$, so for $j = 2$. In figure 2 for problem (1) we have drawn $\mathrm{Re}\mu_1$ as function of y.

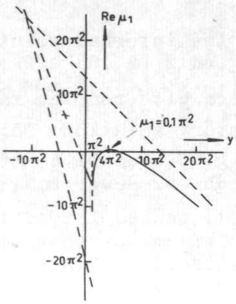

Figure 2. Reμ_1 as function of y for problem (1).

The maximum of Reμ_1 is equal to $0,1\pi^2$, so the asymptotic stability is equal to $e^{0,1\pi^2}$. Obviously problem (1) is unstable. For large values of t in general one of the 2nd elementary solutions will become dominating. From (2.10.10) this 2nd elementary solution proves to be
$\left\{\sin 2\pi x\ e^{0,1\pi^2 t}, 2 \sin 2\pi x\ e^{0,1\pi^2 t}\right\}$.

The numerical calculation of the solution of problem (1) has been done in the case that the differential equations of (1) have been replaced by difference equations of Crank-Nicolson. For this choice of difference equation it holds that $g_{\Delta t} = (1 + \frac{1}{2}c\,\Delta t)/(1 - \frac{1}{2}c\,\Delta t)$ where c satisfies (2.10.10) provided in this μ and $(j\pi)^2$ are replaced by respectively c and $2(1 - \cos j\pi\Delta x)/(\Delta x)^2$.

The right hand members of the differential equations as well as the boundary conditions of (1) are chosen such that the external solutions of differential problem (1) and the corresponding difference problem are equal. Moreover, the initial conditions are taken such that the internal solution only consists of the above mentioned 2nd elementary solution. It can easily be found that the analytic solution of (1) is

$$z_1(x,t) = (\tfrac{1}{4}x^2 + 2t) + \sin 2\pi x\ e^{0,1\pi^2 t},$$

$$z_2(x,t) = (x^2 + 2t) + 2 \sin 2\pi x\ e^{0,1\pi^2 t}. \qquad (2)$$

Problem (1) in difference form has been solved for $\Delta t = 0{,}01$, $\Delta x = 0{,}1$ and $\delta = 0{,}001$. Herewith from (2.10.10) we find that again the in absolute value largest growthfactors occurs for one of the 2^{nd} elementary solutions,

$$\left\{\sin 2\pi x_m\ 1{,}0097^{t_n/\Delta t},\quad 1{,}97 \sin 2\pi x_m\ 1{,}0097^{t_n/\Delta t}\right\}. \qquad (3)$$

The initial conditions of the internal solution of the difference problem are $z_1(x_m,0) = \sin 2\pi x_m$ and $z_2(x_m,0) = 2 \sin 2\pi x_m$. The ratio of these conditions is 2. This differs a bit from the ratio $L = 1{,}97$ of the components of the above 2^{nd} elementary solution. This means that also the other 2^{nd} elementary solution will give a small contribution to the internal solution of the difference problem. However, the growthfactor $g_{\Delta t}$ of this elementary solution is $\approx - 0{,}13$ and so its contribution to the internal solution tends rapidly to 0. This means that the ratio of the amplitudes of $\sin 2\pi x_m$ in the solution will change in a few times from 2 into 1,97.

In figure 3 the numerical solution of problem (1) is given for a number of times.

$z_1(x_m,t_n)$

$t_n \backslash x_m$	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
.00	.00	.59	.96	.97	.63	.06	-.50	-.83	-.79	-.39	.25
.01	.02	.63	1.01	1.02	.67	.08	-.50	-.84	-.80	-.38	.27
.02	.04	.65	1.04	1.05	.69	.10	-.48	-.83	-.79	-.37	.29
.03	.06	.68	1.07	1.08	.72	.12	-.47	-.82	-.78	-.35	.31
.04	.08	.71	1.10	1.11	.74	.14	-.45	-.81	-.77	-.34	.33

$z_2(x_m,t_n)$

$t_n \backslash x_m$	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
.00	.00	1.19	1.94	1.99	1.34	.25	-.82	-1.41	-1.26	-.37	1.00
.01	.02	1.22	1.99	2.04	1.37	.27	-.81	-1.42	-1.27	-.36	1.02
.02	.04	1.26	2.03	2.08	1.40	.29	-.81	-1.42	-1.27	-.35	1.04
.03	.06	1.29	2.07	2.12	1.44	.31	-.80	-1.42	-1.27	-.35	1.06
.04	.08	1.32	2.11	2.16	1.47	.33	-.79	-1.42	-1.27	-.34	1.08

Figure 3. Numerical solution of problem (1) for $\Delta t = 0{,}01$ and $\Delta x = 0{,}1$.

7.4. Iteration process for a problem described by two simultaneous "diffusion" equations

First we will consider problem (2.10.1), if in this the differential quotients are replaced by difference quotients in the way of Crank-Nicolson. If (2.10.1) is solved iteratively then during the iteration process at t_n in fact the solution of a boundary and initial value problem is determined which problem we will call iteration problem. The boundary conditions of the iteration problem are equal to those of (2.10.1) at t_n, while its initial condition is equal to the predicted solution of (2.10.1) at t_n. We remark that convergence of the iteration process has the same meaning as stability of the iteration problem. The solution of (2.10.1) at t_n is equal to the external solution or what is the same the steady state solution of the iteration problem.

The x-elementary solutions of the iteration problem are $\{\alpha_1(x_m)g^s, \alpha_2(x_m)g^s\}$, provided

$$\begin{aligned} &b_1\,\alpha_1(x_{m-1}) + c_1\,\alpha_1(x_m) + \alpha_1(x_{m+1}) + d_1\,\alpha_2(x_m) = 0 \\ &b_2\alpha_2(x_{m-1}) + c_2\alpha_2(x_m) + \alpha_2(x_{m+1}) + d_2\alpha_1(x_m) = 0 \\ &\alpha_1(0) = \alpha_1(1) = \alpha_2(0) = \alpha_2(1) = 0, \end{aligned} \tag{1}$$

where

$$b_1 = g^*_{11},\quad c_1 = -\frac{2}{a_1\gamma}\,g^*_{12} - 2g^*_{13} + \frac{\Delta t}{a_1\gamma}\,g^*_{14}d_{11},\quad d_1 = \frac{\Delta t}{a_1\gamma}\,g^*_{15}d_{12}$$

$$b_2 = g^*_{21},\quad c_2 = -\frac{2}{a_2\gamma}\,g^*_{22} - 2g^*_{23} + \frac{\Delta t}{a_2\gamma}\,g^*_{24}d_{22},\quad d_2 = \frac{\Delta t}{a_2\gamma}\,g^*_{25}d_{21}. \tag{2}$$

Again g^*_{1i} and g^*_{2i}, $i = 1,\dots,5$ can be equal to g or equal to 1 dependent on the final choice of the iteration formulas.

It can easily be found that in the case $b_1 = b_2 = b$ the solution of (1) is

$$\alpha_{1j}(x_m) = b_j^{\frac{m}{2}}\sin j\pi x_m,\quad \alpha_{2j}(x_m) = K_j\, b_j^{\frac{m}{2}}\sin j\pi x_m\,, \tag{3}$$

provided

$$\begin{aligned} (c_{1j} + 2\sqrt{b_j}\cos j\pi\Delta x) + {}& d_{1j}\,K_j = 0 \\ d_{2j} + (c_{2j} + 2\sqrt{b_j}&\cos j\pi\Delta x)\,K_j = 0. \end{aligned} \tag{4}$$

Now we will discuss some numerical calculations concerning problem (7.3.1) which are performed in the case that the difference equation of Milne is used as a predictor and the iteration process is applied for which

$$b_j = g_j,\quad c_{1j} = -2\,\frac{1+a_1\gamma}{a_1\gamma}\, g_j + \frac{\Delta t}{a_1\gamma}\, d_{11},\quad d_1 = \frac{\Delta t}{a_1\gamma}\, d_{12},$$

$$c_{2j} = -2\,\frac{1+a_2\gamma}{a_2\gamma}\, g_j + \frac{\Delta t}{a_2\gamma}\, d_{22},\quad d_2 = \frac{\Delta t}{a_2\gamma}\, d_{21}\,. \tag{5}$$

First we will take $\Delta t = \Delta x = 0,1$. Then from (4) and (5) it follows that g satisfies

$$45100(\sqrt{g})^4 - 80840(\sqrt{g})^3 + 20479(\sqrt{g})^2 + 15256\sqrt{g} + 1542 = 0 \quad \text{for } j = 1$$

$$45100(\sqrt{g})^4 - 68765(\sqrt{g})^3 + 10476(\sqrt{g})^2 + 12976\sqrt{g} + 1542 = 0 \quad \text{for } j = 2. \tag{6}$$

From this we obtain that the in absolute value largest growthfactors are

$$g = 1,162\; e^{\pm\frac{2\pi i}{23,9}} \qquad \text{for } j = 1$$

$$g = 0,932\; e^{\pm\frac{16,1\pi i}{180}} \qquad \text{for } j = 2. \tag{7}$$

It can be shown that the absolute value of the stability of an elementary solution of the above iteration problem increases with j. So the iteration problem is only unstable for j = 1. We observe that the error in the predicted solution at t_1 is a constant times $\sin 2\pi x_m$ and that the contributions to the internal solution of the iteration problem for j = 2 are damped sin-functions into the x-direction of the same frequency as can be understood from (3) and (7). This includes that the internal solution of the iteration problem also contains the 1st elementary solutions and so the iteration process does not converge. It can easily be derived that for large values of s the internal solution $\{z_{1i}(x_m,s), z_{2i}(x_m,s)\}$ of the iteration problem will approach a constant times $\{f_1(x_m,s), f_2(x_m,s)\}$ where

$$f_1(x_m,s) = (1,162)^{\frac{m}{2}+s}\ \sin\{\frac{2\pi}{23,9}\,(\frac{m}{2} + s + \varphi_1)\}\,\sin\frac{m\pi}{10}\,,$$

$$f_2(x_m,s) = 1,885(1,162)^{\frac{m}{2}+s}\ \sin\{\frac{2\pi}{23,9}\,(\frac{m}{2} + s + \varphi_1 + 1,65)\}\sin\frac{m\pi}{10}\,. \tag{8}$$

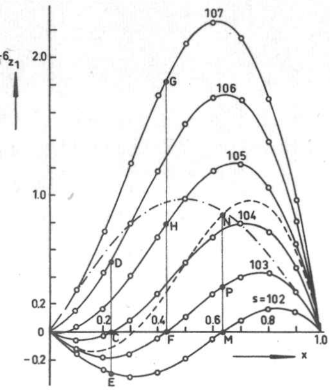

Figure 1. Numerical solution of the iteration problem for $\Delta t = \Delta x = 0{,}1$.

In figure 1 the numerical values of $z_1(x_m,s)$ are represented as function of x_m for $s = 102,\ldots,107$. Moreover for $s = 102$ in this figure the curve $z_2(x_m,s)$ is given as a dotted line. From the zero passings of $z_1(x_m,s)$ and $z_2(x_m,s)$ we have a phase shift between $z_1(x_m,s)$ and $z_2(x_m,s)$ to the value of 1,65. From CD/CE we can calculate the absolute value of the growthfactor per step into the s-direction. In this way we find 1,16. From FG/FH we can determine the period of the factor $\sin\{\frac{2\pi}{23{,}9}(\frac{m}{2}+s+\varphi_1)\}$ occurring in (8). The result is a period of 24 instead of the theoretical value 23,9. The —.— line represented in figure 1 seems to be proportional to $\sin\frac{m}{10}$ which can easily be verified theoretically. Calculating finally K_1 from the ratio $\frac{MN}{MP}$ we find $K_1 = 1{,}87$. All together we can say that the numerical results agree well with (8).

In order to obtain a convergent iteration process Δt has to be chosen sufficiently smaller than 0,1. For small values of Δt the power equation of the 4^{th} degree in $\sqrt{g}$ given by (6) can be approximated by a square equation in g. The roots of this square equation in absolute value are smaller than 1, if $\Delta t < \frac{1}{40}$. Taking $\Delta t = 0{,}01$ indeed the iteration process is convergent.

In figure 2 the solution of the thus obtained iteration problem in $x = 0,1$ is given for $t = 0,01,\ldots,0,05$. From the final results for z_2 of the iteration processes at times 0,03, 0,04 and 0,05 we obtain $g_{\Delta t} = 1,0097$. From the final results for z_1 and z_2 at time 0,05 it follows that the ratio of the components of the internal solution z_{2i}/z_{1i} is equal to 1,972. We see that these results agree well with (7.3.3).

t=0,01		
s=1	0,638.412.604	1,222.354.26
2	0,631.884.902	1,218.010.05
3	0,628.245.608	1,218.202.77
4	0,627.217.676	1,220.417.33
5	0,627.695.21	1,222.648.79
6	0,628.597.799	1,223.973.96
7	0,629.277.917	1,224.352.91
8	0,629.551.972	1,224.166.86
9	0,629.519.381	1,223.827.11

t=0,02		
s=1	0,651.660.801	1,254.314.86
2	0,652.734.419	1,255.073.44
3	0,653.346.564	1,255.068.01

t=0,03		
s=1	0,680.040.899	1,286.692.46
2	0,679.719.617	1,286.393.56

t=0,04		
s=1	0,705.467.092	1,318.725.70
2	0,705.583.648	1,318.435.06

t=0,05		
s=1	0,731.191.158	1,349.767.69
2	0,731.559.496	1,350.309.07

Figure 2. Numerical solution of the iteration problem for $\Delta t = 0,01$ and $\Delta x = 0,1$.

7.5. Fitting a difference problem to a differential problem

In paragraph 3.3 we have shown if a given differential problem is interpreted by a difference problem that it can be sensible to fit the given initial condition before using it in the difference problem. Now we will illustrate this for the following problem

$$\frac{\partial z}{\partial t} - \frac{\partial^2 z}{\partial x^2} + z = 0.$$

$$z(0,t) = z(1,t) = 1,\ z(x,0) = 10 - 1{,}1 \sin \pi x - 9\,\frac{\sinh x + \sinh(1-x)}{\sinh 1}. \tag{1}$$

Applying (2.6.4) and (2.6.8) yields the external solution as

$$z_e(x,t) = \frac{\sinh x + \sinh(1-x)}{\sinh 1}\ . \tag{2}$$

It can easily be shown that the internal solution is

$$z_i(x,t) = \sum_{j=1}^{\infty} b_j \sin j\pi x\, e^{-\{1+(j\pi)^2\}t}\ , \tag{3}$$

where

$$b = \frac{40}{\pi(1+\pi^2)} - 1{,}1,\quad b_{2j} = 0,\quad b_{2j+1} = \frac{40}{(2j+1)\pi\{1+(2j+1)^2\pi^2\}}\ ,\ j \geqq 1. \tag{4}$$

Replacing the differential equation of (1) by the difference equation of Crank-Nicolson the external solution becomes

$$z_e(x_m,t_n) = \frac{\sinh Ax_m + \sinh A(1-x_m)}{\sinh A}\ ,\quad \cosh A\Delta x = 1 + \frac{(\Delta x)^2}{2}\ . \tag{5}$$

Without fitting the initial condition the internal solution is

$$z_i(x_m,t_n) = \sum_{j=1}^{\infty} c_j \sin j\pi x_m (g_{j\Delta t})^{t_n/\Delta t}$$

$$= \sum_{j=1}^{M-1} d_j \sin j\pi x_m (g_{j\Delta t})^{t_n/\Delta t}\ , \tag{6}$$

where

$$c_1 = \frac{36}{\pi(1+\pi^2)} + \frac{4A^2}{\pi(A^2+\pi^2)} - 1{,}1,\quad c_{2j} = 0,$$

$$c_{2j+1} = \frac{1}{(2j+1)\pi}\left\{\frac{36}{1+(2j+1)^2\pi^2} + \frac{4A^2}{1+A^2(2j+1)^2\pi^2}\right\} \tag{7}$$

and the constants c and d satisfy (3.3.1).

Now we will determine the values of Δt and Δx for which at $t = 0{,}1$ the solution of the difference problem differs about 10^{-4} from that of the differential problem. From figure 1 we see that the external solution differ about 10^{-4} if $\Delta x = \frac{1}{9}$. As a consequence the initial conditions of the internal solutions are nearly the same (see figure 2). Because problem (1) is symmetric with respect to $x = \frac{1}{2}$ the internal solution only contains elementary solutions with odd values of j. In figure 3 the growth-factors per 0,1 time-unit of these elementary solutions are represented.

x	$z_e(x,t)$	$z_e(x_m,t_n)$
$0 , 1$	1,000.00	1,000.00
$\frac{1}{9} , \frac{8}{9}$	0,954.73	0,954.77
$\frac{2}{9} , \frac{7}{9}$	0,921.25	0,921.33
$\frac{3}{9} , \frac{6}{9}$	0,899.16	0,899.26
$\frac{4}{9} , \frac{5}{9}$	0,888.19	0,888.29

Figure 1. External solutions for $\Delta x = \frac{1}{9}$.

j	1	3	5	7	9	11	13	15	17	19
b_j	,0714	,0472	,0103	,0038	,0018	,0010	,0006	,0004	,0003	,0002
c_j	,0713	,0472	,0103	,0037	,0018	,0010	,0006	,0004	,0003	,0002
d_j	,0712	,0470	,0098	,0028						

Figure 2. Fourier-coefficients of the initial conditions of the internal solutions for $\Delta x = \frac{1}{9}$.

j	$e^{-0,1(1+j^2\pi^2)}$	$(g_{j,\Delta t})^{\frac{0.1}{\Delta t}}$				
		$\Delta t=\frac{1}{10}$	$\Delta t=\frac{1}{20}$	$\Delta t=\frac{1}{30}$	$\Delta t=\frac{1}{40}$	$\Delta t=\frac{1}{60}$
1	,337	,300	,331	,337	,338	0,340
3	,000	-,608	,119	-,005	,000	0,000
5	,000	-,811	,428	-,142	,028	0,000
7	,000	-,870	,571	-,280	,101	0,005

Figure 3. Growth-factors per 0,1 time-unit of the odd elementary solutions for $\Delta x = \frac{1}{9}$.

From the figures 2 and 3 it follows that at $t = 0,1$ only the 1st elementary solution of the differential problem has a noticeable value. So fitting of the initial condition means omitting the elementary solutions with $j > 1$. From figure 3 we see that $\Delta t = \frac{1}{30}$ is the best one for a good interpretation of the 1st elementary solution by the difference problem. This agrees with relation (3.4.4).

It will be clear from the figures 2 and 3 that we cannot apply $\Delta t = \frac{1}{30}$ if the initial condition is not fitted because for the difference problem the higher frequency elementary solutions are not damped enough. Therefore without fitting we have to reduce Δt, but this means that the 1st elementary solution is no longer interpreted accurately enough unless Δx is also reduced.

In figure 4 the difference between the solutions of the differential problem and the difference problem is represented for a number of choices of Δt and Δx. From this figure it can be concluded that without fitting the initial condition Δt and Δx have to be chosen equal to, respectively, $\frac{1}{120}$ and $\frac{1}{18}$ in order to obtain an error of about 10^{-4}. This means that for problem (1) the ratio of the numbers of values of $z(x_m, t_n)$ which have to be calculated for the nonfitted difference problem and for the fitted one is equal to 8.

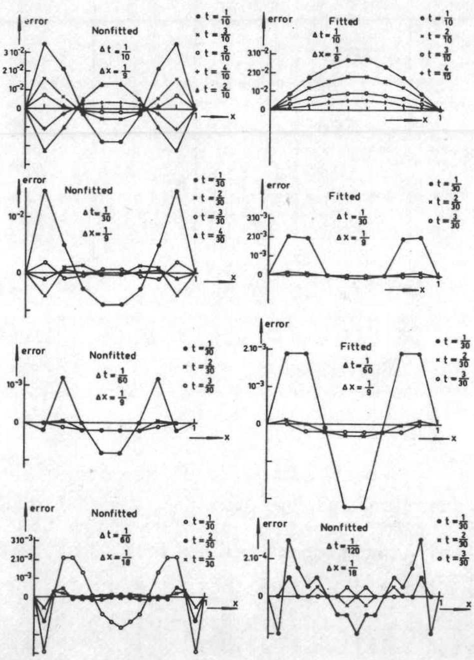

Figure 4. Error as function of x and t for a number of choices of Δx and Δt.

Finally we remark that the above example is not one for which fitting has an extravagant effect because the amplitudes of the higher frequency elementary solutions are small compared with the amplitude of the 1^{st} elementary solution.

REFERENCES

[1]. Nörlund, N.E., Vorlesungen über Differenzenrechnung, Springer Berlin, 1924.

[2]. Léauté, M.H., Développement d'une fonction à une seule variable, dans un intervalle donné, suivant les valeurs moyennes de cette fonction et de ses dérivées successives dans cet intervalle, Journal de Liouville, juin 1881, pp. 185-200.

[3]. Appell, P., Expression d'un polynome en fonction des valeurs moyennes du polynome et de ses dérivées dans un intervalle, Eléments d'Analyse Mathématique, Carré et Naud, Paris 1898, pp. 216-220.

[4]. Dekker, L., Uitwendige en inwendige oplossing bij rand- en beginwaardeproblemen, Report Techn. Un. Delft, 1962.

[5]. Dekker, L., Analogue computation of special diffusion-type problems, Proc. 3rd Int. An. Comp. Meetings, Bruxelles 1962, pp. 178-188.

[6]. Dekker, L., De toepassing van analogons bij het oplossen van warmte-overdrachtsproblemen, De Ingenieur, 36, 1961, pp. 0145-149.

[7]. Müller, H.G., Hybride rekenmachine voor het oplossen van partiële differentiaalvergelijkingen, Report Techn. Un. Delft, 1962.

[8]. Visman, A.C.F., A diode function generator made suitable for two variables. To be published in Proc. Int. Ass. An. Comp.

[9]. Müller, H.G., Onderzoek betreffende de electronische verwezenlijking van een tweetal schakelfuncties, Report Techn. Un. Delft, 1961.

[10]. Dekker, L., Optimizing a diffusion problem with the help of a high speed special analogue computer, Report Techn. Un. Delft, 1964.

[11]. Stolz, C., Convergence of a seriesexpansion of a function $f(x)$ in its even derivatives in $x = 0$ and $x = 1$, Report Techn. Un. Delft, 1964.

SUMMARY

This thesis has arisen from an investigation concerning the design and application of a special analogue computer for solving such problems as

$$\frac{\partial z_i}{\partial t} - a_i \frac{\partial^2 z_i}{\partial x^2} = f_i(x,t,z_1,\ldots,z_n), \quad a_i > 0,$$

$$z_i(0,t) = \varphi_i(t), \; z_i(1,t) = \eta_i(t), \; z_i(x,0) = \gamma_i(x), \quad i = 1,\ldots,n.$$

In the thesis some aspects of analytic and numerical solving of problems of the above type are discussed.

In chapter I some notions are defined two of which we will call "external solution" and "internal solution". The first one is a genearalization of the "steady state solution", while the internal solution is aequivalent to the "transient". These concepts are illustrated in this chapter for ordinary differential and difference equations.

Chapter II deals with these concepts for diffusion problems in differential as well as in difference form. Mathematically, the external solution is constructed as a series expansion with respect to the derivatives of the boundary conditions. It proves that these series relate to 2 points series expansions of a function f(x) like

$$f(x) = \sum_{k=o}^{\infty} \left\{ g_k(x) \frac{d^{2k}f(0)}{dx^{2k}} + g_k(1-x) \frac{d^{2k}f(1)}{dx^{2k}} \right\},$$

where the functions $g_k(x)$ are simple polynomials.

In chapter III a way is discussed in which the initial condition of a diffusion problem in differential form can be fitted in order to increase the accuracy of the solution of the corresponding difference problem.

In chapter IV a brief description is given of the designed special analogue computer.

Chapter V deals with the rate of convergence of digital iteration processes as can be applied for solving the above mentioned problems.

In chapter VI the same is done as in chapter V but for the iteration process as occurs with the special analogue computer.

In chapter VII some numerical results are discussed.

SAMENVATTING

Dit proefschrift is voortgekomen uit een onderzoek betreffende de ontwikkeling en toepassing van een speciale analoge rekenmachine bestemd voor het oplossen van problemen van het type

$$\frac{\partial z_i}{\partial t} - a_i \frac{\partial^2 z_i}{\partial x^2} = f_i(x,t,z_1,\ldots,z_n), \quad a_i > 0,$$

$$z_i(0,t) = \varphi_i(t), \; z_i(1,t) = \eta_i(t), \; z_i(x,0) = \gamma_i(x), \; i = 1,\ldots,n.$$

In het proefschrift worden enige aspecten besproken van analytisch en numeriek oplossen van deze problemen.

In hoofdstuk I worden enige begrippen gedefiniëerd waarvan we er hier twee noemen: de uitwendige oplossing en de inwendige oplossing. De eerste is een generalisatie van het begrip blijvende oplossing, terwijl de inwendige oplossing overeenstemt met het begrip inschakelverschijnsel. Deze twee begrippen worden in dit hoofdstuk toegelicht voor gewone differentiaal- en differentievergelijkingen.

Hoofdstuk II handelt over deze begrippen voor diffusieproblemen in zowel differentiaal- als differentievorm. Wiskundig wordt de uitwendige oplossing geconstrueerd als een reeksontwikkeling naar de afgeleiden van de randvoorwaarden. Het blijkt dat zulke reeksontwikkelingen verband houden met 2-puntsontwikkelingen van een functie f(x) zoals

$$f(x) = \sum_{k=0}^{\infty} \left\{ g_k(x) \frac{d^{2k}f(0)}{dx^{2k}} + g_k(1-x) \frac{d^{2k}f(1)}{dx^{2k}} \right\},$$

waarbij de functies $g_k(x)$ eenvoudige polynomen zijn.

In hoofdstuk III wordt een manier besproken waarop de beginvoorwaarde van een diffusieprobleem in differentiaalvorm kan worden aangepast teneinde de nauwkeurigheid van de oplossing van het bijbehorende differentieprobleem te verbeteren.

In hoofdstuk IV wordt een korte beschrijving gegeven van de ontwikkelde speciale analoge rekenmachine.

Hoofdstuk V heeft betrekking op de convergentiesnelheid van digitale iteratieprocessen zoals kunnen worden toegepast voor het oplossen van de voornoemde problemen.

In hoofdstuk VI gebeurt hetzelfde als in hoofdstuk V, maar nu voor het iteratieprocess zoals het optreedt bij de speciale analoge rekenmachine.

In hoofdstuk VII worden enige numerieke resultaten besproken.

STELLINGEN

1. De begrippen uitwendige en inwendige oplossing zijn ook bruikbaar bij andere dan in dit proefschrift beschreven problemen.

2. De in dit proefschrift beschouwde reeksontwikkeling van een functie $f(x)$,

$$\sum_{k=0}^{\infty} \{g_k(x) \frac{d^{2k}f(0)}{dx^{2k}} + g_k(1 - x) \frac{d^{2k}f(1)}{dx^{2k}}\},$$

heeft voor elke x als som $f(x)$, indien in het interval $0 \leqq x \leqq 1$ voor grote k geldt $|\frac{d^{2k}f(x)}{dx^{2k}}| < c\, \pi^{2k}$ met c = constante

3. Met behulp van de in dit proefschrift gedefiniëerde polynomen $g_k(x)$ kan een functie $f(x)$ formeel worden uitgedrukt in de even afgeleiden in $x = 0$ en de oneven afgeleiden in $x = 1$ door de reeksontwikkeling

$$f(x) = \sum_{k=0}^{\infty} \left[2^{2k} \{g_k(\tfrac{x}{2}) + g_k(1 - \tfrac{x}{2})\} \frac{d^{2k}f(0)}{dx^{2k}} + 2^{2k+1} \frac{d\{g_{k+1}(\frac{1+x}{2}) - g_{k+1}(\frac{1-x}{2})\}}{dx} \frac{d^{2k+1}f(1)}{dx^{2k+1}} \right].$$

4. Bij een probleem, beschreven door

$$\frac{\partial z_i}{\partial t} - a_i \frac{\partial^2 z_i}{\partial x^2} = \sum_{j=1}^{n} d_{ij}\, z_j,\ z_i(0,t) = \varphi_i(t),\ z_i(1,t) = \eta_i(t),\ z_i(x,0) = \gamma_i(x),$$

a_i = positieve constante, $i = 1,\ldots,n$,

zal bij gelijke a_i's de kleinste stabiliteit optreden voor één van de 1^e x-elementaire oplossingen.

5. Zowel bij het onderwijs als bij het speurwerk in de numerieke wiskunde is tot op heden te weinig aandacht geschonken aan rekenmethoden, die worden toegepast bij analoge rekenmachines.

6. Een duidelijk inzicht in de stabiliteit van een differentievergelijking kan vaak worden verkregen door de differentievergelijking op te vatten als de beschrijving van een evenwicht van elektrische stromen.

7. Indien een numerieke berekening betrekking heeft op het gedrag van een fysisch systeem, dient een maat voor de nauwkeurigheid van de berekening zodanig te zijn dat een grotere nauwkeurigheid overeenstemt met een be-

tere beschrijving van het fysische systeem.

8. Het is te betreuren dat de ontwikkeling van hybride rekenmachines vrijwel uitsluitend geschiedt door producenten en gebruikers van analoge rekenmachines.

9. Het bepalen van de eigenwaarden en eigenvectoren van een eenvoudige bandmatrix alsook het inverteren hiervan kan vaak zinvol worden teruggebracht tot het bepalen van de oplossingen van een randwaardeprobleem.

10. Indien een tijdsignaal $f(t)$ moet worden vastgelegd in een geheugen en hierbij vereist is dat signaalwaarden met een verschil ε nog juist van elkaar moeten kunnen worden onderscheiden, dan verdient registratie van de tijdstippen t_i waarop $|f(t_i) - f(t_{i-1})| = \varepsilon$ vaak de voorkeur boven registratie van de signaalwaarden op gelijke tijdsafstanden.

11. Een zinvolle toepassing van de in dit proefschrift beschreven speciale analoge rekenmachine is het gebruik als snelle voorspellende rekenmachine bij het optimaliseren van diffusieprocessen