# Direct solution of certain sparse linear systems

**Eindhoven University of Technology**
**Department of Mathematics and Computing Science**

RANA 88-02

February 1988

DIRECT SOLUTION OF
CERTAIN SPARSE LINEAR SYSTEMS

by

R.M.M. Mattheij

Reports on Applied and Numerical Analysis
Department of Mathematics and Computing Science
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands

# Direct Solution of Certain Sparse Linear Systems

R.M.M. Mattheij
Faculteit Wiskunde en Informatica
Technische Universiteit Eindhoven
Postbus 513
5600 MB Eindhoven
The Netherlands

ABSTRACT.

This paper deals with sparse linear systems which arise among other cases when solving boundary value problems. These matrices have a bidiagonal blockstructure plus potentially some nontrivial blocks as last column and or row. The analogy with BVP is employed to relate it to theory of discretized ODE. The underlying structure of the solution space, following from well-conditioning, then induces a direct solution method (based on decoupling). The strategy is shown to work for fairly general cases.

This paper is in final form and will not be published elsewhere. We gratefully acknowledge a grant from the Koninklijke Nederlandse Academie der Wetenschappen (Royal Dutch Academy of Sciences).

## 1. Introduction.

A number of problems in numerical analysis leads to the formulation of a linear system where the (sparse) matrix typically has a structure as in fig. 1.1.

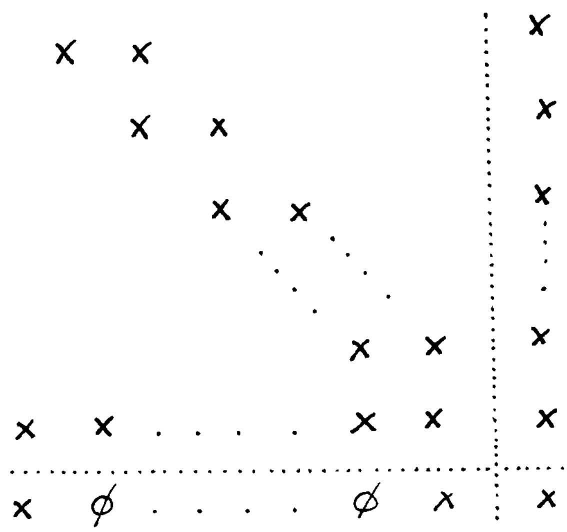

Fig. 1.1

Here $\times$ stands for a matrix which is $n \times n$ in the left upper part (as indicated by the dotted line), has $m$ rows in the last block row and has $m$ columns in the last block column.
In principle we need not require that the square blocks be nonsingular, but for our analysis it is helpful to have some simplifying assumptions. For a paper which deals with more general cases we masy e.g. refer to [7]. Matrices as in (1.1) arise, for instance, in system theory and in numerical solutions of boundary value problem (BVP) for differential equations. For a general nonlinear BVP, one usually solves via linearization, i.e. by a sequence of linear problems in which the Jacobian of some sort gives rise to a matrix as in fig. 1.1. As can be found in [1], most numerical methods actually lead (possibly after some form of condensation) to a multiple shooting system of which the left upper part in fig. 1.1 is reminiscent. Actually, in this paper we like to employ this fact by relating such matrices to special BVP. This will enable us to use a number of important results of the theory of those problems for establishing a *stable direct* method for solving linear systems with such a matrix.
To this end we first consider general *linear* ODE of the form

(1.1) $$\frac{dx}{dt} = L(t)\,x + C(t)\,\lambda + f(t)\,,\ \alpha \le t \le \beta$$

where

$L : [\alpha, \beta] \rightarrow \mathbf{R}^{nxn}$

$C : [\alpha, \beta] \rightarrow \mathbf{R}^{nxm}$

$x, f : [\alpha, \beta] \rightarrow \mathbf{R}^{n}$

$\lambda \in \mathbf{R}^{m}$.

After discretisation of (1.1) by say a difference method we obtain, possibly after condensation (e.g. a local "shooting" approach), a difference equation for the solution at certain nodes, $t_1, \cdots, t_N$ say. If we neglect discretisation errors, we can write

(1.2a) $\quad x_i = x(t_i)$

(1.2b) $\quad X^T = (x_1^T, x_N^T \lambda^T)$,

for the unknown solution values. In a multiple shooting framework, we typically get a difference equation

(1.3a) $\quad A_i x_i + B_i x_{i+1} + C_i \lambda = f_i$

By writing

(1.3b) $\quad (F^1)^T = (f_1^T, \cdots, f_{N-1}^T)$,

and

(1.4) $$\mathbf{A}^1 = \begin{bmatrix} A_1 & B_1 & & & & C_1 \\ & A_2 & B_2 & & & \vdots \\ & & \ddots & \ddots & & \vdots \\ & & & A_{N-1} & B_{N-1} & C_{N-1} \end{bmatrix},$$

the discretised ODE results in the following set of equations

(1.5) $\quad \mathbf{A}^1 X = F^1$

(To fix thoughts, often $B_i = -I$ for multiple shooting and thus $A_i$ represents the incremental matrix, going from $t_i$ to $t_{i+1}$)
As for the boundary conditions (BC), we let them be of the form

(1.6a) $$\sum_{i=1}^{N} M_i \, x(t_i) + E_N \, \lambda = b$$

(1.6b) $\quad P_1 \, x(t_i) + P_N \, x(t_N) + E_{N+1} \, \lambda = c$

where

$M_i \in \mathbf{R}^{nxn}$

$E_N \in \boldsymbol{R}^{nxm}$

$E_{N+1} \in \boldsymbol{R}^{mxm}$

$P_1, P_2 \in \boldsymbol{R}^{mxn}$

$b \in \boldsymbol{R}^n, c \in \boldsymbol{R}^m$

(More general BC are also possible, but (1.6) is sufficiently general, given $t_1, \cdots, t_N$). Clearly (1.6) gives rise to $n + m$ additional equations

(1.7) $\quad \mathbf{A}^2 X = F^2,$

where

(1.8a) $\quad \mathbf{A}^2 = \begin{bmatrix} M_1 & M_2 & \cdots & M_N & E_N \\ P_1 & 0 & \cdots & P_N & E_{N+1} \end{bmatrix}$

(1.8b) $\quad F^2 = \begin{bmatrix} b \\ c \end{bmatrix}$

There are several special cases which deserve a special consideration:

I <u>Two Point BVP:</u>

$$\left.\begin{matrix} C_1, & \cdots, & C_N \\ P_1, & P_N & \\ E_N, & E_{N+1} & \end{matrix}\right\} \text{are absent}$$

$$M_2 = \cdots = M_{N-1} = 0$$

II <u>Multipoint BVP:</u>

$$\left.\begin{matrix} C_1, & \cdots, & C_N \\ P_1, & P_N & \\ E_N, & E_{N+1} & \end{matrix}\right\} \text{are absent}$$

III <u>Two point parameter BVP:</u>

$$M_2 = \cdots = M_{N-1} = 0$$

In section 2 we discuss the properties of the underlying subspaces in each of the cases I, II, III. The more general structure of the matrix A, defined by

(1.9) $$\mathbf{A} = \left[ \frac{\mathbf{A}^1}{\mathbf{A}^2} \right]$$

will then be dealt with in section 3. The basic algorithm on which our method is built is given in section 4. Finally, we show how this is used to solve the general problem.

(1.10) $$\mathbf{A}\, X = F ,$$

where

$$F = \begin{bmatrix} F^1 \\ F^2 \end{bmatrix}$$

## 2. On dichotomy and (skew) polychotomy

The various kinds of BVP have solution spaces with a potentially different character. This is a very important aspect as our direct methodes for solving (1.10) will be based on finding a stable way for actually determining this solution space.

Let us consider the homogeneous part of (1.3a)

$$A_i\ x_i + B_i\ x_{i+1} = 0, \tag{2.1}$$

and let us denote by $\{\Phi_i\}_{i=1}^N$ a *fundamental system* of (2.1). In order to derive some properties for $\{\Phi_i\}$, we shall assume throughout this paper that $\| \mathbf{A}^{-1} \|$ is not large:

**Assumption 2.2** Let $\kappa$ be a not large positive constant and let $\| \mathbf{A}^{-1} \| \le \kappa$. □

For the simplest case, with two point BC, a simple splitting of the solution space can be indicated in which the Green's function (burried in $\mathbf{A}^{-1}$) is used, cf [2,3]. We summarize the discrete analogue of [3] in the following

**Property 2.3** (case I, Two Point BVP).
If $C_1, \cdots, C_N, P_1, P_N, E_N, E_{N+1}$ are absent and $M_2 = \cdots = M_{N-1} = 0$, then it is not restrictive to assume that $\{\Phi_i\}_{i=1}^N$ is *dichotomic* ,i.e. there exists an orthogonal projection $P$ with

$$\| \Phi_i\ P\ \Phi_{j+1}^{-1} \| \le \kappa,\ i > j$$

$$\| \Phi_i\ (I-P)\ \Phi_{j+1}^{-1} \| \le \kappa,\ i \le j$$

□

From this property if follows e.g. for a solution $\{\phi_i\}$, with $\phi_i = \Phi_i\, c$ , $c \in \mathbb{R}^n$,

$$\frac{\| \phi_i \|}{\| \phi_{j+1} \|} = \frac{\| \Phi_i\ Pc \|}{\| \Phi_{j+1}\ Pc \|} \le \max_{y \ne 0} \frac{\| \Phi_i\ P\ \Phi_{j+1}^{-1} y \|}{\| y \|} \le \kappa,\ i > j$$

In other words, $\{\phi_i\}$ is "non-increasing". Similarly $\{\psi_i\}$, with $\psi_i = \Phi_i\ (I-P)d$ , $d \in \mathbb{R}^n$ is a "non-decreasing" solution. Moreover, dichotomy implies that the two thus indicated solution subspaces cannot be directionally close (cf [2]).

For multipoint BC there exists the following generalisation (see [4]).

**Property 2.4** (case II, Multipoint BVP)
If $C_1, \cdots, C_N, P_1, P_N, E_N, E_{N+1}$ are absent, then it is not restrictive to assume that $\{\Phi_i\}_{i=1}^N$ in *polychotomic* , i.e. there exist orthogonal projections $P_1, .., P_l$ , $l \le n$ with

$$\sum_{j=1}^{l} P_j = I \, , \; \sum_{j=1}^{l} \operatorname{rank}(P_j) = n$$

and corresponding indices $1 = i_1 , \ \cdots \ , i_l = N$ such that

$$\| \Phi_i \, (\sum_{j=1}^{k} P_j) \, \Phi_{j+1}^{-1} \| \le \kappa , \; i_k \le j < i_{k+1} , \; i > j$$

$$\| \Phi_i \, ( \sum_{j=k+1}^{n} P_j) \, \Phi_{j+1}^{-1} \| \le \kappa , \; i_k \le j < i_{k+1} , \; i \le j$$

□

Hence for multipoint BVP we see that on each interval $[ t_{i_k} , t_{i_{n+1}} ]$ the solution space is dichotomic, but that the dimension of the "non-increasing" mode subspace may *increase* from $[ t_{i_k} , t_{i_{k+1}} ]$ to $[ t_{i_{k+1}} , t_{i_{k+2}} ]$.
As before the subspaces do not make small angles. So a typical solution may be either "non-decreasing" or "non-increasing" everywhere, or may be "non-decreasing" till some point $t_{i_k}$, after which it becomes "non-increasing". In fig. 2.1 we have sketched these three possibilities.

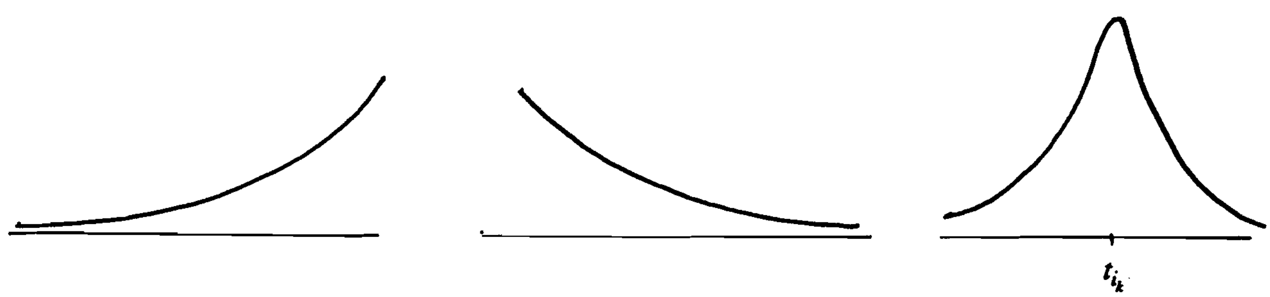


Fig. 2.1

Finally, we have

**Property 2.5** (case III, Two Point BVP with parameters)
If $M_2 = \ \cdots \ = M_N = 0$, then it is not restrictive to assume that $\{ \Phi_i \}_{i=1}^{N}$ is

*skew polychotomic*, i.e. there exist orthogonal projections $P_1, \cdots, P_l$, $l \le n$, with $\sum_{j=1}^{l} P_j = I$, $\sum_{j=1}^{l} \mathrm{rank}(P_j) = n$ and corresponding indices $1 = i, \cdots, i_l = N$ such that

$$\| \Phi_{j+1} (\sum_{j=1}^{k} = P_j) \Phi_i^{-1} \| \le \kappa, \; i_k \le j < i_{k+1}, \; i > j$$

$$\| \Phi_{j+1} ( \sum_{j=k+1}^{n} P_j) \Phi_i^{-1} \| \le \kappa, \; i_k \le j \le i_{k+1}, \; i \le j$$

Hence for parameter BVP we see that again the solution space is dichotomic interval wise, but that the dimension of the *non–decreasing* mode subspace may increase from $[t_{i_k}, t_{i_{k+1}}]$ to $[t_{i_{k+1}}, t_{i_{k+2}}]$. So a typical solution may either be "non-increasing" or "non-decreasing" or be "non-increasing" till some point $t_{i_k}$ after which it berowes "non-decreasing". In fig 2.1 we have sketched these three possibilities.

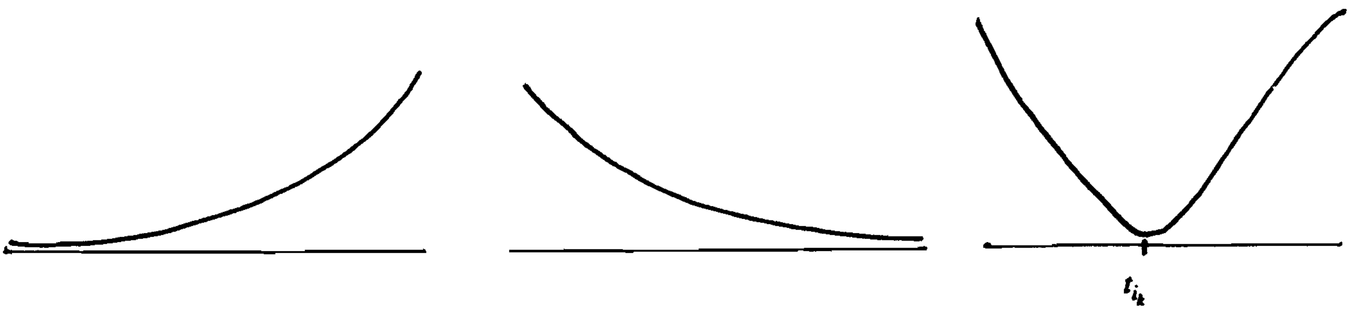


Fig. 2.2

It should be noted that case II is related to case III in that after some manipulation a matrix arising in one case can be viewed as the transpose of the other. To see these, consider e.g. case II.

Here we have

$$\mathbf{A} = \begin{bmatrix} A_1 & B_1 & & & & C_1 \\ & A_2 & B_2 & & & C_2 \\ & & \ddots & \ddots & & \cdots \\ & & & A_{N-1} & B_{N-1} & C_{N-1} \\ M_1 & \emptyset & \cdots & \emptyset & M_N & E_N \\ P_1 & \emptyset & \cdots & \emptyset & P_N & E_{N+1} \end{bmatrix} \tag{2.6}$$

After writing the first block row of $\mathbf{A}^T$ last we obtain

$$\tilde{\mathbf{A}}^T := \begin{bmatrix} B_1^T & A_2^T & & & & \\ & B_2^T & \ddots & & & \\ & & \ddots & A_{N-1}^T & & \\ & & & B_{N-1}^T & M_N^T & P_N^T \\ C_1^T & C_2^T & \cdots & C_{N-1}^T & E_N^T & E_{N+1}^T \\ A^{1^T} & \emptyset & \cdots & \emptyset & M_1^T & P_1^T \end{bmatrix} \tag{2.7}$$

Next we realize that the last two block columns (with n and m columns respectively) constitute a full column rank (= n+m) system. Hence it is not restrictive to assume $P_N^T = 0$. Since now $\begin{bmatrix} E_{N+1}^T \\ P_1^T \end{bmatrix}$ must have rank m we may permute the last m + n rows of $\tilde{\mathbf{A}}^T$ in order to have a non-singular block at the place of $E_{N+1}^T$. For simplicity let us assume that $E_{N+1}^T$ is already non-singular; then we can use this to eliminate $C_1^T, \cdots, C_{N-1}^T, E_N^T$, thereby producing potential fill-in in the last block row. Summarizing, there exists a well-conditioned equivalence transformation for $\tilde{\mathbf{A}}^T$ giving a matrix $\tilde{\mathbf{A}}^T$ say with

$$\tilde{\mathbf{A}}^T := \begin{bmatrix} B_1^T & A_2^T & & & & & \\ & B_2^T & A_3^T & & & & \\ & & \ddots & \ddots & & & \\ & & & \ddots & A_{N-1}^T & & \\ & & & & B_{N-1}^T & M_N^T & \emptyset \\ \emptyset & & & \cdots & \emptyset & \emptyset & E_{N+1}^T \\ * & * & \cdots & \cdots & \cdots & * & * \end{bmatrix} \tag{2.8}$$

obviously, we can eliminate the last block variable, which effectively means deleting the N-th block row and (N + 1)-st block column (both of rank m). Clearly the resulting matrix is of type II.

## 3. The structure of the solution space, general case .

Now consider the general case where **A** has the form as in fig. 1.1, i.e. we have a multipoint parameter problem. The situation here is more complicated. Rather than analyzing it in detail, we shall consider an example to illustrate this. Since the solution methods we shall propose later only work for any of the cases I, II or III, mentioned before, this will be sufficient to motivate our approach for the more general case.

Consider the following matrix **A**

(3.1) $$\mathbf{A} = \begin{bmatrix} \varepsilon\ 0 & -1\ 0 & 0\ 0 & 0 \\ 0\ 1 & 0\ -\varepsilon & 0\ 0 & 0 \\ 0\ 0 & 1\ 0 & -\varepsilon\ 0 & 1 \\ 0\ 0 & 0\ \varepsilon & 0\ -1 & 0 \\ 1\ 0 & 0\ 0 & 0\ 0 & 0 \\ 0\ 0 & 0\ 1 & 0\ 0 & 0 \\ 0\ 0 & 0\ 0 & 1\ 0 & 1 \end{bmatrix}, \varepsilon > 0$$

From (3.1) we see

(3.2a) $$\begin{bmatrix} \varepsilon & 0 \\ 0 & 1 \end{bmatrix} \Phi_1 = \begin{bmatrix} 1 & 0 \\ 0 & \varepsilon \end{bmatrix} \Phi_2$$

(3.2b) $$\begin{bmatrix} 1 & 0 \\ 0 & \varepsilon \end{bmatrix} \Phi_2 = \begin{bmatrix} \varepsilon & 0 \\ 0 & 1 \end{bmatrix} \Phi_3$$

Writing $\Phi_i = \left[ \phi_i^1 \mid \phi_i^2 \right]$, we may identify $\{ \phi_i^1 \}$, $\{ \phi_i^2 \}$ as follows

(3.3a) $$\phi_1^1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \phi_2^1 = \begin{bmatrix} \varepsilon \\ 0 \end{bmatrix}, \phi_3^1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

(3.3b) $$\phi_1^2 = \begin{bmatrix} 0 \\ \varepsilon \end{bmatrix}, (\phi_2^2) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \phi_3^2 = \begin{bmatrix} 0 \\ \varepsilon \end{bmatrix}$$

see fig. 3.1

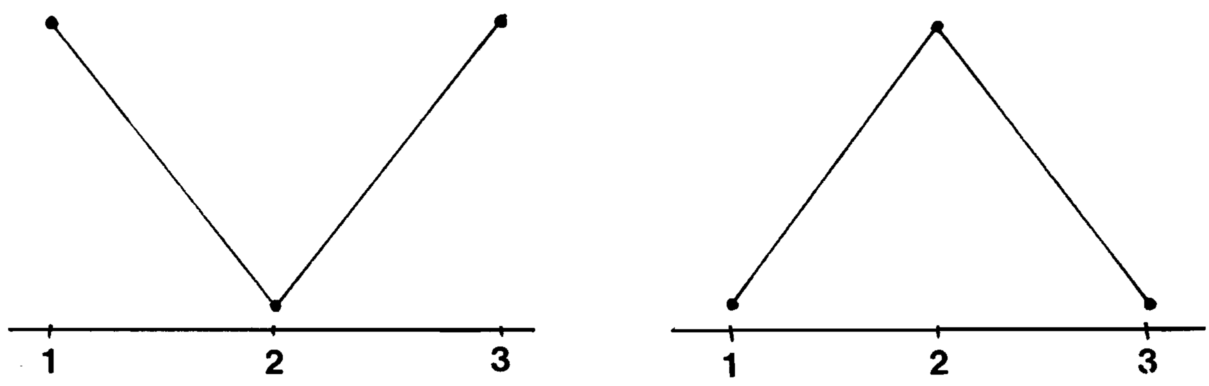


Fig. 3.1

In other words, since these basis modes are clearly directionally well separated, we may conclude that there is no polychotomic, no skew polychotomic fundamental solution . Yet, the matrix **A** is well conditioned. For, if we premultiply **A** by

$$(3.4) \qquad \boldsymbol{L} := \begin{bmatrix} 1 & & & & & & \\ 0 & 1 & & & & & \\ 1 & 0 & 1 & & & & \\ & & & 1 & & & \\ & & & & 1 & & \\ & & & & & 1 & \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 \end{bmatrix},$$

we see that $\boldsymbol{L}$ **A** is a permutation matrix perturbed by a matrix of norm $\approx \varepsilon$. Hence we conclude that $\| \mathrm{A}^{-1} \| \le 2(1+\varepsilon) =: \kappa$.

However, we may reduce, at the expense of increasing the dimension of the recursion, the general problem to a multipoint BVP, giving the possibility to identify the resulting subspace as polychotomic. We proceed as follows:
Write

$$(3.5) \qquad z_i = \begin{bmatrix} x_i \\ \lambda \end{bmatrix}$$

For the sake of simplicity we take $B_i = -I$ for $i = 1, \cdots, N-1$ from now on. Then we obtain the augmented system.

(3.6) $$z_{i+1} = \begin{bmatrix} A_i & C_i \\ 0 & I \end{bmatrix} z_i \text{ (+ inh. term)}$$

and from (1.6) the *multipoint* BC

(3.7) $$\sum_{i=1}^{N} \mathbf{M}_i \; z_i = \mathbf{b} := \begin{bmatrix} b \\ c \end{bmatrix}$$

where

(3.8a) $$\mathbf{M}_1 = \begin{bmatrix} M_1 & \varnothing \\ P_1 & \varnothing \end{bmatrix}$$

(3.8b) $$\mathbf{M}_i = \begin{bmatrix} M_i & \varnothing \\ \varnothing & \varnothing \end{bmatrix}, \quad 2 \le i \le N-1$$

(3.8c) $$\mathbf{M}_N = \begin{bmatrix} M_N & E_N \\ P_N & E_{n+1} \end{bmatrix}$$

From (3.6) and (3.7) we find the an augmented system

(3.9) $$\hat{\mathbf{A}} = \begin{bmatrix} A_1 & C_1 & -I & \varnothing & & & & & & \\ \varnothing & I & \varnothing & -I & & & & & & \\ & & A_2 & C_2 & -I & \varnothing & & & & \\ & & \varnothing & I & \varnothing & -I & & & & \\ & & & & \ddots & & & \ddots & & \\ & & & & & & A_{N-1} & C_{N-1} & -I & 0 \\ & & & & & & \varnothing & I & \varnothing & -I \\ & \mathbf{M}_1 & \vdots & \vdots & \vdots & & \mathbf{M}_{N-1} & & & \mathbf{M}_N \end{bmatrix}$$

By obvious permutations $\mathbf{P}_1$, $\mathbf{P}_2$ we find

(3.10) $$\hat{\hat{\mathbf{A}}} := \mathbf{P}_1 \, \hat{\mathbf{A}} \, \mathbf{P}_2 = \left[\begin{array}{ccccc:ccccc} A_1 & -I & & & & C_1 & & & & \\ & A_2 & -I & & & & \ddots & & & \\ & & \ddots & \ddots & & & & \ddots & & \\ & & & A_{N-1} & -I & & & & C_{N-1} & \\ M_1 & M_2 & \cdots & \cdots & M_N & & & & & E_N \\ \hdashline P_1 & \varnothing & \cdots & \varnothing & P_N & \varnothing & \varnothing & \varnothing & & E_{N+1} \\ & & & & & \varnothing & I & -I & & \\ & & & & & & & \ddots & \ddots & \\ & & & & & & & & I & -I \\ & & & & & I & -I & & \varnothing & \varnothing \end{array}\right]$$

Finally, after postmultiplication by the matrix $\mathbf{B}$, defined as

$$(3.11) \qquad \mathbf{B} := \left[\begin{array}{ccc:ccc} I & & & & & \\ & \ddots & & & \varnothing & \\ & & I & & & \\ \hdashline & & & I & & \\ \varnothing & & & \vdots & \ddots & \\ & & & I & & I \end{array}\right]$$

($\mathbf{B}$ is partitioned in the same way as $\hat{\mathbf{A}}$),
we obtain

$$(3.12) \qquad \hat{\mathbf{A}}\,\mathbf{B} = \begin{bmatrix} A_1 & -I & & & & C_1 & & & & & \\ & A_2 & -I & & & C_2 & C_2 & & & & \\ & & \ddots & \ddots & & \vdots & & \ddots & & & \\ & & & A_{N-1} & -I & C_{N-1} & & & \ddots & & \\ M_1 & M_2 & & & M_N & E_N & & & & C_{N-1} & \\ P_1 & & & & P_N & E_{N+1} & & & & & \\ & & & & & & I & -I & & & \\ & & & & & & & I & -I & & \\ & & & \varnothing & & & & & \ddots & \ddots & \\ & & & & & & & & & I & -I \\ & & & & & & -I & & & 0 & 0 \end{bmatrix}$$

From (3.12) we then immediately find

**Property 3.13** $\| \mathbf{A}^{-1} \|_\infty \leq \kappa_\infty (1 + N(1 + \max_i \| C_i \|_\infty)$ $\square$

We therefore conclude that $\hat{\mathbf{A}}$ has a similar conditioning as $\mathbf{A}$. Now we can use Property 2.4 and deduce that the augmented BVP (3.6), (3.7) has a polychotomic fundamental solution.

Instead of forming (3.5) we may alternatively define

$$(3.14) \qquad y_i := \begin{bmatrix} x_i \\ \sum_{j=1}^{i-1} M_j \, x_j \end{bmatrix}$$

Then we obtain the following augmented system

$$(3.15) \qquad y_{i+1} = \begin{bmatrix} A_i & \varnothing \\ M_i & I \end{bmatrix} y_i + 1 \begin{bmatrix} C_i \\ 0 \end{bmatrix} \lambda,$$

for which we have the two point BC

(3.16a) $$\begin{bmatrix} \varnothing & I \\ \varnothing & \varnothing \end{bmatrix} y_1 + \begin{bmatrix} \varnothing & \varnothing \\ M_N & I \end{bmatrix} y_N + \begin{bmatrix} 0 \\ E_N \end{bmatrix} \lambda = \begin{bmatrix} 0 \\ b \end{bmatrix}$$

(3.16b) $$[P_1 \ \varnothing]\, y_1 + [P_N \ \varnothing]\, y_N + E_{N+1}\, \lambda = C$$

By similar arguments as used above for the reduction to a multipoint case, it is not hard to show now that (3.15), (3.16) constitutes a well conditioned system if **A** is. Hence we find from Property 2.5 that the fundamental solution of (3.15) may be thought skew polychotomic.

Whether we prefer the augmented system (3.6) or (3.15) depends mainly on the sizes of the blocks. Clearly (3.6) involves order 2n matrices. For N not too small the multipoint system therefore seems to be preferable if $m \leq n$.

## 4. Solution methods

From the preceding sections we deduce that we may restrict ourselves to constructing direct solvers for cases I, II and III. As has been shown e.g. in [6], a crucial concept in solving BVP numerically is *decoupling*. We merely repeat some important aspects here.
Assume we have a dichotomy (on intervals $[t_{i_k}, t_{i_{k+1}}]$ this is always the case)
The method consists of the following major steps

(i) Transform the recursion (2.1) such that the matrices are in upper triangular form. This is done recursively and an appropiate initialization should take place.

(ii) Compute the various modes of the two ("dichotomic") subspaces as well as an appropiate particular solution by employing forward and backward recursion of the decoupled parts.

(iii) Use superposition to single out the required solution from the BC.

We demonstrate this for a general inhomogeneous recursion (cf [8]). Let

(4.1) $$A_i x_i + B_i x_{i+1} = f_i$$

Choose a suitable orthogonal matrix $Q_1$ ; by "suitable" we mean such that the first, say k, columns of $Q_1$ (if rank(P) = k) span a subspace of initial values of modes that do not belong to $\{\Phi_i P\}_{i=1}^{N}$ (cf Property 2.3), i.e. no "non decreasing" ones. Then compute

(4.2a) $$A_1 Q_1 = R_1 U_1,$$

where $R_1$ is orthogonal and $U_1$ upper triangular.
Next compute

(4.2b) $$R^{-1} B_1 = V_1 Q_2^{-1},$$

where $V_1$ is upper triangular and $Q_2$ orthogonal.
The general step reads:
Perform a "QU-decomposition":

(4.3a) $$A_i Q_i = R_i U_i$$

and a "UQ-decomposition"

(4.3b) $$R_i^{-1} B_i = V_i Q_{i+1}^{-1}$$

If we set

(4.4a) $$\tilde{M}_1 := M_1 Q_1$$

(4.4b) $$\tilde{M}_N := M_N Q_N,$$

and define

(4.5) $$\mathbf{R} := \begin{bmatrix} R_1^{-1} & & & & \\ & R_2^{-1} & & & \\ & & \ddots & & \\ & & & R_N^{-1} & \\ & & & & I \end{bmatrix}$$

(4.6) $$\mathbf{Q} := \begin{bmatrix} Q_1 & & & \\ & Q_2 & & \\ & & \ddots & \\ & & & Q_N \end{bmatrix}$$

Then the transformation (4.2) - (4.4) result in the following system

(4.7) $$\tilde{\mathrm{A}} := \mathbf{R}\ \mathrm{A}\ \mathbf{Q} = \begin{bmatrix} U_1 & V_1 & & & \\ & U_2 & V_2 & & \\ & & \ddots & & \\ & & & U_{N-1} & V_{N-1} \\ \tilde{M}_1 & \varnothing & & \varnothing & \tilde{M}_N \end{bmatrix}$$

By introducing new variables

(4.8) $$\tilde{x}_i := Q_i^{-1}\, x_i\,,\ \tilde{f}_i := R_i\, f_i\,,$$

we see that (4.7) represents a *decoupled* recursion for $\{\tilde{x}_i\}$,

(4.9) $$U_i\, \tilde{x}_i + V_i\, \tilde{x}_{i+1} = \tilde{f}_i$$

Because of our well conditioning, we have dichotomy, also for the transformed problem: Assume rank(P) = k (see Property 2.3). Then partition matrices and vectors like:

(4.10) $$U_i = \begin{bmatrix} U_i^{11} & U_i^{12} \\ \varnothing & U_i^{22} \end{bmatrix},\ U_i^{11} \in \mathbf{R}^{k\times k}$$

(4.11) $$\tilde{x}_i = \begin{bmatrix} \tilde{x}_i^1 \\ \tilde{x}_i^2 \end{bmatrix}$$

We now have

**Property 4.12** The partitioning, induced by the dichotomy, implies a decoupling of non increasing and non decreasing modes, which can therefore be computed separately and in a stable way.

□

So, from (4.9), (4.10) and (4.11) we find

(4.13a) $\quad U_i^{22}\,\tilde{x}_i^{\,2} + V_i^{22}\,\tilde{x}_{i+1}^{\,2} = \tilde{f}_i^{\,2}$

In Property 4.12 it is meant that (4.13a) is a *stable* recursion in *forward* direction, so for $i = 1, \cdots, N-1$. We also obtain

(4.13b) $\quad U_i^{11}\,\tilde{x}_i^{\,1} + V_i^{11}\,\tilde{x}_{i+1}^{\,1} = f_i - U_i^{12}\,\tilde{x}_i^{\,2} - V_i^{12}\,\tilde{x}_{i+1}^{\,2}$

Considering the right hand side in (4.13b) as a source term (known if (4.13a) has been carried out earlier), it is implied that (4.13b) is a *stable* recursion in *backward* direction, so for $i = N-1, \cdots, 1$ .

These decoupled recursions are used now to complete both a fundamental solution and some particular solution. First the computation of a fundamental solution:
Choose

(4.14a) $\quad \tilde{\Phi}_1^2 = [\varnothing \mid I]\ (I = I_k),$

and compute $\tilde{\Phi}_2^2, \cdots, \tilde{\Phi}_N^2$ via the homogeneous part of (4.13a). Then use this to compute $\tilde{\Phi}_N^1, \cdots, \tilde{\Phi}_1^{1L}$ from (4.13b), satisfying the "terminal condition"

(4.14b) $\quad \tilde{\Phi}_N^1 = [I \mid \varnothing],\ (I = I_{n-k})$

Next, compute a particular solution:
Choose

(4.15a) $\quad \tilde{p}_1^{\,2} = 0,$

and compute $\tilde{p}_2^{\,2}, \cdots, \tilde{p}_N^{\,2}$ via (4.13a). Then choose

(4.15b) $\quad \tilde{p}_N^{\,1} = 0,$

and compute $\tilde{p}_{N-1}^{\,1}, \cdots, \tilde{p}_1^{\,1}$, via (4.13b).
Linearity implies that there exists a vector $c$ such that

(4.16) $\quad \tilde{x}_i = \tilde{\Phi}_i\, c + \tilde{p}_i,\ i = 1, \cdots, N,$

which can be found from the BC (cf 4.4)

(4.17) $\quad \tilde{M}_1\,\tilde{x}_1 + \tilde{M}_N\,\tilde{x}_N = b,$

Hence we have to solve the *th* order linear system

(4.18) $\quad [\tilde{M}_1\,\tilde{\Phi}_1 + \tilde{M}_N\,\tilde{\Phi}_N]\,c = b - \tilde{M}_1\,\tilde{p}_1 - \tilde{M}_N\,\tilde{p}_N.$

Once we have found $\{\tilde{x}_i\}$ via (4.16) our desired solution $\{x_i\}$ follows from (cf (4.8) )

(4.19) $\quad x_i := Q_i\,\tilde{x}_i$

For multipoint problems this uniform partitioning may not be stable anymore, as the dichotomy

may change from interval to interval. However the partitioning of the blocks at the various intervals can be done in a monotomic way (cf Property 2.4).
This was outlined more precisely in [5]. We shall not go into detail here but remark that the essential adaptation is rather obvious: One should prescribe that part of the identity matrix as an "initial-middle-terminal" value for $\{\tilde{\Phi}_i\}$ where a corresponding "multipoint" condition gives rise to this. This can be done in practice by monitoring the change in dichotomy pattern (as actually shown by incremental values found from comparing $U_i$, $V_i$). At a point where some unit vectors serve as "starting values" for certain components of $\tilde{\Phi}_i$ we also choose the corresponding coordinates of $\tilde{p}_i$ to be zero.
We illustrate this by the following three point BC (cf (1.6a) )

(4.20) $$\tilde{M}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tilde{M}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tilde{M}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Denoting by superscripts the index of the rows and coordinates in $\{\tilde{\Phi}_i\}$ and $\{\tilde{p}_i\}$ respectively, we choose

(4.21a) $$\tilde{\Phi}_1^3 = [0\ 0\ 1]$$

Then $\tilde{\Phi}_2^3$ is found from the (decoupled) recursion By choosing

(4.21b) $$\tilde{\Phi}_2^2 = [0\ 1\ 0]$$

we can use $\begin{bmatrix} 010 \\ 00* \end{bmatrix}$ as an initial value on the last interval. For backward recursion we finally set

(4.21c) $$\tilde{\Phi}_3^1 = [1\ 0\ 0],$$

and compute $\tilde{\Phi}_2^1$; at this point we augment this part of the fundamental matrix with $\tilde{\Phi}_2^2$, so use $\begin{bmatrix} *** \\ 010 \end{bmatrix}$ as "starting value" for backward recursion on the first interval. Schematically, this yields a fundamental solution $\{\tilde{\Phi}_i\}$ as in fig. 4.1

$$\tilde{\Phi}_1 = \begin{bmatrix} \Delta & \Delta & \Delta \\ 0 & \Delta & \Delta \\ 0 & 0 & 1 \end{bmatrix}, \tilde{\Phi}_2 = \begin{bmatrix} \Delta & \Delta & \Delta \\ 0 & 1 & 0 \\ 0 & 0 & * \end{bmatrix}, \tilde{\Phi}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix}$$

* = computed forward

Δ = computed backward

fig. 4.1

In a similar way we find $\{\tilde{p}_i\}$ viz. using the BC

(4.22) $\tilde{p}_1^3 = 0 \, , \tilde{p}_2^2 = 0 \, , \tilde{p}_3^1 = 0$

Superposition and backtransformation go essentially the same as for the two point case.

With an adaption of argument one can also use such a technique in the skew polychotomic case, which we shall not dwell upon here. (more details are given in [10])

## 5. Conclusions

Given a system **A** as in fig. 5.1,

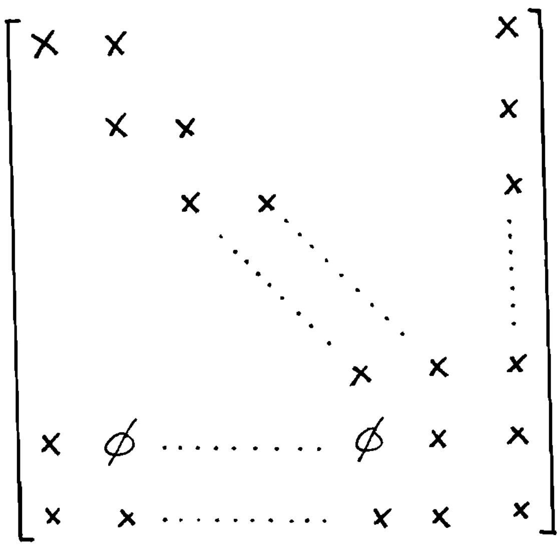

fig. 5.1

We decide from the last block row/columns, whether we effectively have either of the following cases

I two-point BVP

II multipoint BVP

III two point BVP with parameters

IV multipoint BVP with parameters

In case I we use the "standard" decoupling technique, outlined in section 4. In cases II and III we use an adapted decoupling technique (in which a possible change of the dichotomy partitioning is accounted for).

Finally, in case IV we rewrite the BVP as either a case II or a case III BVP as shown in section 2, after which we can use the method for case II or III repspectively. An augmented multipoint system is to be preferred over an augmented parameter system as long as $m \leq n$.

**References**

[1] U.M. Ascher, R.M.M. Mattheij, R.D. Russell, Numerical Solution of Boundary Value ODEs, Prentice Hall, 1988

[2] F. de Hoog, R.M.M. Mattheij, On Dichotomy and Well-conditioning in BVP, SIAM J. Numer. Anal. 24 (1987), 89-105

[3] F. de Hoog, R.M.M. Mattheij, The Role of Conditioning in Shooting Techniques, in : Proceedings of a Workshop on Numerical Boundary Value ODEs (U. Ascher, R.D. Russell, eds), Birkhauser (1985), 21-54.

[4] F. de Hoog, R.M.M. Mattheij, An Algorithm for Solving Multipoint Boundary Value Problems, Computing 38 (1987), 219-234.

[5] F. de Hoog, R.M.M. Mattheij, On the conditioning of Multipoint Boundary Value Problems, to appear in SIAM J. Math. Anal.

[6] R.M.M. Mattheij, Decoupling and Stability of BVP Algorithms, SIAM Review 27 (1985), 1-44.

[7] R.M.M. Mattheij, Decoupling and Stability of General Block Bidiagonal Systems, report TUE 1988.

[8] R.M.M. Mattheij, G.W.M. Staarink, An efficient Method for Solving General Linear Two Point BVP, SIAM J. Sci. Stat. Comp. 5 (1984), 745-763.

[9] R.M.M. Mattheij, G.W.M. Staarink, Numerical Software for Solving BVP of ODE, manuscript, 1988.

PREVIOUS PUBLICATIONS IN THIS SERIES:

| Number | Author(s) | Title | Month |
|---|---|---|---|
| 87-13 | P.H. van Lieshout/<br>P.M.J. Rongen/<br>A.A.F. van de Ven | A variational approach to magneto-elastic buckling problems for systems of ferro-magnetic or superconducting beams. | Nov. '87 |
| 87-14 | R.M.M. Mattheij | Implementing multiple shooting for non-linear BVP | Dec. '87 |
| 87-15 | R.M.M. Mattheij | A sequential algorithm for solving boundary value problems with large Lipschitz constants | Dec. '87 |
| 87-16 | A.A.F. van de Ven/<br>Tani/Otomo/Shindo | Magnetoelastic buckling of two nearby ferromagnetic rods in a magnetic field | Dec. '87 |
| 87-17 | A.J.E.M. Janssen/<br>S.J.L. van Eijndhoven | Spaces of type *W*, growth of Hermite coefficients, Wigner distribution and Bargmann transform | April '87 |
| 87-18 | L. Dortmans/<br>A. Sauren/<br>A.A.F. van de Ven | A note on the reduced creep function corresponding to the quasi-linear visco-elastic model proposed by Fung | April '87 |
| 88-01 | G.A. Kluitenberg/<br>L. Restuccia | On some generalizations of the Debye equation for dielectric relaxation | Feb. '88 |
| 88-02 | R.M.M. Mattheij | Direct Solution of Certain Sparse Linear Systems | Feb. '88 |