On the conditioning of multipoint and integral boundary value problems

Eindhoven University of Technology
Department of Mathematics and Computing Science

RANA 88-03

March 1988

ON THE CONDITIONING OF MULTIPOINT AND INTEGRAL BOUNDARY VALUE PROBLEMS

by

F.R. de Hoog

and

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

ON THE CONDITIONING OF MULTIPOINT AND

INTEGRAL BOUNDARY VALUE PROBLEMS

F.R. de Hoog
CSIRO
Division of Mathematics
and Statistics
P.O. Box 1965
Canberra ACT 2601
Australia.

R.M.M. Mattheij
Eindhoven University of Technology
Department of Mathematics and
Computing Science
P.O. Box 513
5600 MB Eindhoven
The Netherlands.

On the Conditioning of Multipoint and Integral Boundary Value Problems

F.R. de Hoog and R.M.M. Mattheij

Abstract

We investigate linear multipoint boundary value problems from the point of view of the condition number and properties of the fundamental solution. It is found that when the condition number is not large, the solution space is polychotomic. On the other hand if the solution space is polychotomic then there exist boundary conditions such that the associated boundary value problem is well conditioned.

§1. Introduction

Consider the first order system of ordinary differential equations

(1.1) $\quad Ly := y' - Ay = f \;, \quad 0<t<1$

where $A \in L_1^{n\times n}(0,1)$ and $f \in L_1^{n}(0,1)$.

We are interested in the solution of (1.1) that satisfies the multipoint boundary condition (BC)

(1.2) $\quad \mathcal{B}y := \sum_{i=1}^{N} B_i y(t_i) = b.$

Here,

$$0 = t_1 < \ldots < t_N = 1$$

and the matrices $B_i \in \mathbb{R}^{n\times n}$, $k = 1,\ldots,N$ have been scaled so that for instance

(1.3) $\quad \sum_{i=1}^{N} B_i B_i^T = I$

The restriction $t_1 = 0$, $t_N = 1$ has been introduced for notational convenience and is not restrictive provided we allow for the possibility that $B_0 = 0$ and $B_N = 0$.

One of the simplest examples of a multipoint boundary value problem is that of a dynamical system with n states which are observed at different times. Further examples and a description of numerical schemes for the solution of such equations may be found in [12], [1], [11].

From the theory of boundary value problems, (1.1), (1.2) has a unique solution if $\mathcal{B}Y$ is nonsingular for any fundamental solution Y of L (see for example Keller [8]). In the sequel we assume this is the case. Then, given any fundamental solution Y of (1.1) we may write the solution of (1.1), (1.2) as

(1.4) $\quad y(t) = \Phi(t)b + \int_0^1 G(t,s)f(s)ds \;, \quad 0\le t\le 1$

where

(1.5a) $\quad \Phi(t) := Y(t)\ (\mathcal{B}Y)^{-1}$

and

(1.5b)
$$G(t,s)=\begin{cases} \Phi(t) \sum\limits_{i=1}^{k} B_i\Phi(t_j)\Phi^{-1}(s) , & t_k < s < t_{k+1} , \ t > s \\ -\Phi(t) \sum\limits_{i=k+1}^{N} B_i\Phi(t_j)\Phi^{-1}(s) , & t_k < s < t_{k+1} , \ t < s \end{cases}$$

The function G is the Green's function associated with (1.1) (1.2).

We can now use (1.4) to examine the conditioning of (1.1), (1.2). Let $|\cdot|$ denote the usual Euclidean norm in $\mathbb{R}^n$ and define

$$||u||_\infty := \sup_t |u(t)| \ , \ u \in [L_\infty(0,1)]^n$$

$$||u||_1 = \int_0^1 |u(t)|dt \ , \ u \in [L_1(0,1)]^n.$$

Then it follows from (1.3) that

(1.6) $\quad ||y||_\infty \leq \beta\ |\mathcal{B}y| + \alpha\ ||\mathcal{L}y||_1$

where

(1.7a) $\quad \alpha := \sup\limits_{t,s} |G(t,s)|$

and

(1.7b) $\quad \beta := \sup\limits_{t} |\Phi(t)|$

the quantities α,β defined by (1.7) serve quite well as a condition numbersfor the boundary value problem in the sense that they give a measure for the sensitivity of (1.1), (1.2) to changes in the data. Consequently, if α or β is large, we may expect to have difficulties in obtaining an accurate numerical approximation to the solution of (1.1), (1.2).

If α is of moderate size, the solution space of (1.1) has properties that can (and should) be used in the construction of algorithms for calculating an approximate solution of (1.1), (1.2). For the two point case (i.e. N=2), de Hoog and Mattheij [5], [6] have shown that the solution space is dichotomic when α is not too large. A dichotomic solution space (see section 4 for a more detailed discussion of dichotomy) essentially means that non-increasing modes of the solution space can be controlled by boundary conditions imposed on the left while non-decreasing modes can be controlled by boundary conditions imposed on the right. This concept is the basis for algorithms using decoupling ideas (see for example [10], [11]. The aim of this paper is to generalize the results of [5], [6] to (1.1), (1.2) with $N \geq 2$. In this case the notion of dichotomy has to be generalized and it turns out that, for well conditioned problems, the solution space consists of modes that can be controlled at one of the points $t_1, \ldots t_N$ (see section 4). This has allowed us to generalize the ideas of decoupling to multipoint problems but that is discussed elsewhere [11].

In general one may say that if $N > n$ there is a redundancy in the number of conditions involved. It is therefore crucial to pick precisely n appropriate points from which modes are actually controlled by suitable conditions. It is quite natural to consider then a limit case of multipoint BC, viz an integral condition (which incidentally generalizes two and multipoint conditions in an obvious way), so

$$(1.8) \qquad \mathcal{B}y := \int_0^1 B(\tau)\; y(\tau)\; d\tau = b.$$

Such BC arise directly when L_p norms are used to scale the solution (possibly after linearization) as in eigenvalue problems.

One may treat the (discrete) multipoint case separately from (1.8). However, as it turns out, it is possible to construct a general mechanism which handles the integral BC as well. The price to be paid for this is that our proofs will be based on functional analytic arguments and thus are less constructive as could be given for the discrete case.
The reward though is that we have been able to get sharp bounds in our estimates, sharpening even the bounds given for the two point case in [6].

§2 Notation and Assumptions

In this section we review some basic results which we need later on in our analysis. For some general references regarding Green's functions one may consult e.g. [2] and [9].

2.1 Boundary conditions and their normalisation

Consider the general boundary condition (BC)

$$(2.1) \qquad \mathcal{B}y = b$$

$\mathcal{B}$ is a bounded linear operator from $L^n_{1,1}(0,1)$ to $\mathbb{R}^n$. Note that this includes the BC of type (1.2) and (1.8) as well. By $L^n_{1,1}(0,1)$ we mean those functions whose first derivative is in $L^n_1(0,1)$. We introduce the norm

$$||u||_\infty = \max_{0\le t\le 1} |u(t)| \ , \ u \in L^n_{1,1}(0,1)$$

where

$$|a| = \Big(\sum_{i=1}^n a_i^2\Big)^{\frac{1}{2}}, a \in \mathbb{R}^n.$$

For any $a \in \mathbb{R}^n$, $a^T\mathcal{B}$ is a linear functional from $L^n_{1,1}[0,1]$ to $\mathbb{R}$. We define

$$||a^T\mathcal{B}||_\infty = \sup_{u\in L^n_{1,1}(0,1)} \frac{|a^T \mathcal{B}u|}{||u||_\infty}$$

$$\rho_1(\mathcal{B}) = \max_{a\in\mathbb{R}^n} \frac{||a^T\mathcal{B}||_\infty}{|a|} =: ||\mathcal{B}||_\infty$$

and

$$\rho_n(\mathcal{B}) = \min_{a\in\mathbb{R}^n} \frac{||a^T\mathcal{B}||_\infty}{|a|}$$

<u>Lemma 2.1</u> Let $0 < \rho_1(\mathcal{B}) < \infty$. Then, there exists a matrix $C \in \mathbb{R}^{n\times n}$ such that

$$||C\mathcal{B}||_\infty = \rho_1(C\mathcal{B}) = 1$$

and

$$\rho_n(C\mathcal{B}) \geq \rho_n(E\mathcal{B})/\rho_1(E\mathcal{B}) \quad , \forall E \in \mathbb{R}^{n\times n}$$

<u>Proof</u> If $\rho_n(\mathcal{B}) = 0$, the result is trivial. We therefore assume $\rho_n(\mathcal{B}) > 0$ and let

$$\mathcal{D} = \{E \in \mathbb{R}^{n\times n} \mid \rho_1(E\mathcal{B}) = 1\}$$

Since $\rho_n(E\mathcal{B})$ is continuous in E and $\mathcal{D}$ is closed and bounded, it follows that there is a matrix $C \in \mathcal{D}$ such that

$$\rho_n(C\mathcal{B}) \geq \rho_n(E\mathcal{B}) \qquad \forall E \in \mathcal{D}.$$

This is equivalent to the statement of the lemma. ⊠

This now gives us a possibility to scale the BC, cf. (1.3) in a meaningful way:

<u>Assumption 2.1</u> In the sequel, we shall assume that the BC (2.1) has been scaled so that

$$\text{(2.2a)} \qquad \rho_1(\mathcal{B}) = ||\mathcal{B}||_\infty = 1$$

and

$$\text{(2.2b)} \qquad \rho_n(\mathcal{B}) \geq \rho_n(E\mathcal{B})/\rho_1(E\mathcal{B}) \quad , \forall E \in \mathbb{R}^{n\times n}$$

⊠

In addition to assumption 2.1 we have

<u>Assumption 2.2</u> Let (1.1), (2.1) have a solution for every $f \in L_1^n(0,1)$

and $b \in \mathbb{R}^n$.

⊠

Then, $\mathcal{B}\, Y \in \mathbb{R}^{n\times n}$ is nonsingular where $Y \in L_{1,1}^{n\times n}(0,1)$ is the solution of

(2.3a) $\quad \mathcal{L}Y = 0,\ Y(0) = F$

and $F \in \mathbb{R}^{n\times n}$ is nonsingular.

On defining

(2.3b) $\quad \Phi(t) := Y(t)\ (\mathcal{B}Y)^{-1}$

we can write any function $y \in L_{1,1}^{n}(0,1)$ as

$$y = \mathcal{P}y + (I-\mathcal{P})y$$

(2.4) $$\quad = \mathcal{P}y + \mathcal{G}(\mathcal{L}y)$$

where

(2.5a) $\quad \mathcal{P}y := \Phi\ (\mathcal{B}y)$

(2.5b) $\quad \mathcal{G}f := \int_0^1 G(t,s)f(s)ds\ ,\ f \in L_1^n(0,1)$

and G is the <u>Green's function</u> defined by

(2.6a) $\quad G(t,s) = \Phi(t)\ \{H(t,s) - \mathcal{B}(\Phi\ H(\cdot,s))\}\ \Phi^{-1}(s)$

with

(2.6b) $$\quad H(t,s) = \begin{cases} I, & t > s \\ 0, & t < s \end{cases}$$

(cf. the special case (1.4), where $\mathcal{B}$ is given by (1.2)).

Remark 2.1 The operator $\mathcal{B}$ in the term $\mathcal{B}(\Phi\ H(\cdot,s))$ above should be interpreted as an extension of $\mathcal{B}$ to an operator from $L_\infty^n(0,1)$ to $\mathbb{R}^n$. Note however that a sensible extension of $\mathcal{B}$ to $L_\infty^n(0,1)$ is assured by the Hahn-Banach Theorem.

Remark 2.2 P is a projection from $L_{1,1}^n(0,1)$ onto the solution space $\{Ya \mid a \in \mathbb{R}^n\}$. Given such a projection P, we can define a linear operator

$$\mathcal{B} = C\ Y^{-1}P$$

where $C \in \mathbb{R}^{nxn}$ is a scaling matrix chosen so that (1.1), (2.2 a,b) holds. Lemma 2.1 ensures the existence of such a matrix.

Remark 2.3 It is easy to verify that the Green's function has the form

$$(2.7) \qquad G(t,s) = \begin{cases} Y(t)\ \ (I-E(s))\ Y^{-1}(s) & , t > s \\ \\ -Y(t)\ \ (E(s)\ Y^{-1}(s) & , t < s \end{cases}$$

where $E \in L_\infty^{nxn}(0,1)$. Conversely, given a function of the form (2.7), we have

$$L\{\int_0^1 G(\cdot,s)f(s)ds\} = f\ ,\ f \in L_1^n(0,1).$$

In addition, if we define

$$(Py)(t) := y(t) - \int_0^1 G(t,s)(Ly)(s)ds,$$

then

$$(Py)(t) = y(t) - \int_0^t Y(t)\ Y^{-1}(s)(Ly)(s)ds$$

$$+ Y(t)\int_0^1 E(s)Y^{-1}(s)(Ly)(s)ds$$

$$= Y(t)\ \{Y^{-1}(0)y(0) + \int_0^1 E(s)Y^{-1}(s)(Ly)(s)ds\}$$

It is easily verified that P is a projection. Thus, B defined by

$$By = C\ \{Y^{-1}(0)y(0)\ + \int_0^1 E(s)Y^{-1}(s)(Ly(s))ds\}\ ,$$

where $C \in \mathbb{R}^{nxn}$ is a scaling matrix choosen so that (2.2 a,b) holds, gives a bounded linear operator for which G is the associated Green's function.

2.2 Auerbach's lemma

Let V be a normed linear space of dimension k with norm denoted by $||\cdot||$ and let V^* be the space of all linear functionals from $V \to \mathbb{R}$.

Define a norm on V^* by

$$(2.8) \qquad ||y^*||^* = \sup_{x \in V} \frac{y^*(x)}{||x||}, \quad y^* \in V^*.$$

Definition 2.1 A boundary of V is any set

$$\mathcal{D} \subseteq \{y^* \in V^* \mid ||y^*||^* \leq 1\}$$

such that

$$||x|| = \sup_{y^* \in \mathcal{D}} y^*(x) \quad , \forall x \in V.$$

Lemma 2.2. (Auerbach see [4, lemma 4]).

If $\mathcal{D}$ is a closed boundary of V then there exist $y_i^* \in \mathcal{D}$, $y_j \in V$; $i,j = 1,\ldots,k$ such that

$$y_i^*(y_j) = \delta_{ij} \;, \; ||y_i^*||^* = 1, \; ||y_j|| = 1 \;; \; i,j = 1,\ldots,k.$$

Since $\{y^* \in V^* | \; ||y^*||^* \leq 1\}$ is a closed boundary, it follows immediately that

Corollary 2.1 There exist $y_i^* \in V^*$, $y_j \in V$; $i,j = 1,\ldots,k$ such that

$$y_i^*(y_j) = \delta_{ij} \;, \; ||y_i^*||^* = 1, \; ||y_j|| = 1 \;; \; i,j = 1,\ldots,k.$$

§3 Conditioning of Differential Equations

In this section we consider the relation between α and β and the effect of the normalisation of the BC as in assumption 2.1.

Recall that for $y \in L^n_{1,1}(0,1)$ (cf.(2.4))

$$y(t) = \Phi(t)\ \mathcal{B}y + \int_0^1 G(t,s)\,(\mathcal{L}y)(s)ds$$

Hence, on taking norms

$$||y||_\infty \le \beta|\mathcal{B}y| + \alpha||\mathcal{L}y||_1$$

where

$$\beta = ||\Phi||_\infty = \max_{a\in\mathbb{R}^n} \frac{||\Phi a||_\infty}{|a|}$$

$$\alpha = \sup_{t,s} |G(t,s)|$$

In addition to α and β, it is useful to also consider

$$P.: = Y\ (\mathcal{B}Y)^{-1}\ \mathcal{B}.$$

Lemma 3.1

$$\rho_n\ (\mathcal{B})\beta \le \| P \|_\infty \le \rho_1\ (\mathcal{B})\beta$$

Proof the result follows immediately from the definition of $\rho_1\ (\mathcal{B})$ and $\rho_n\ (\mathcal{B})$. #

Lemma 3.2 Let $\hat{B}$ be a linear operator from $L^{n}_{1,1}$ (0,1) to $\mathbb{R}^{n}$ and $\hat{\alpha}$ be the constant associated with $\hat{B}$ and the differential equation (1.1). Then,

$$\hat{\alpha} \leqq (1 + \|\hat{P}\|_{\infty})\ \alpha, \text{ where } \hat{P}y = y\ (\hat{B}y)^{-1}\ \hat{B}y$$

Proof Let

$$\hat{\Phi} := y\ (\hat{B}y)^{-1} \text{ and } \hat{G}f := \int_0^1 \hat{G}\ (.,s)\ f(s)ds,$$

where $\hat{G}$ is defined similarly to G in (2.6a), i.e. B replaced by $\hat{B}$.

Clearly, $\hat{\Phi} = y\ (\hat{B}y)^{-1}$ and consequently $P = P\ \hat{P}$

That is, $\hat{G}f = (I-\hat{P})\ Gf$ and hence

$$||\hat{G}f||_{\infty} \leqslant (1 + ||\hat{P}||_{\infty})||Gf||_{\infty}$$

Thus, $\hat{\alpha} \leqslant (1 + \|\hat{P}\|_{\infty})\ \alpha$

□

It is clear that the result of Lemmas 3.1 and 3.2 can be combined to give

$$\hat{\alpha} \leqq (1 + \rho_1\ (\hat{B})\ \hat{\beta}\ )\ \alpha$$

Since it has been assumed that (2.2 a,b) holds we obtain the estimate

(3.1) $\hat{\alpha} \leqq (1 + \hat{\beta})\ \alpha$

Note however that α and $\|P\|_{\infty}$ are independent of the scaling (2.2 a,b) but that $\rho_1\ (B)$, $\rho_n\ (B)$ and β are not. We therefore examine some of the ramifications of assumption 2.1

Lemma 3.3

$$\rho_n\ (B) \geqq n^{-1}$$

Proof let

$$V = \{a^T\ B\ |\ a \in \mathbb{R}^n\}.$$

That is, V are the linear functionals of the form $a^T\mathcal{B}$. Since $\mathcal{B}\Phi = I$, $\dim(V) = n$.

For $\ell \in V$, define

$$||\ell|| = \sup_{y \in L^n_{1,1}(0,1)} \frac{(\ell y)}{||y||_\infty} = ||\ell||_\infty$$

V equipped with the norm $||\cdot||$ is an n dimensional normed space. From Auerbach's Theorem (Corollary 2.1), there exist

$\ell_j^* \in V^*$, $\ell_i \in V$; $i,j = 1,\ldots,n$ such that

$$\ell_j^*(\ell_i) = \delta_{ij}, \quad ||\ell_j^*||^* = ||\ell_i|| = 1; \quad i,j = 1,\ldots,n.$$

Clearly, for some $E \in \mathbb{R}^{n\times n}$,

$$a^T E \mathcal{B} = \sum_{i=1}^n a_i \ell_i, \quad \forall a = (a_1,\ldots,a_n)^T \in \mathbb{R}^n.$$

Furthermore,

$$||a^T E \mathcal{B}||_\infty = ||\sum_{i=1}^n a_i \ell_i||_\infty$$

$$\geqslant \frac{|\sum_{j=1}^n a_j \ell_j^* (\sum_{i=1}^n a_i \ell_i)|}{||\sum_{j=1}^n a_j \ell_j^*||^*} \geqslant \frac{|a|}{\sqrt{n}}$$

Thus, $\rho_n(E\mathcal{B}) \geqslant \frac{1}{\sqrt{n}}$.

In addition, $$||a^T E\mathcal{B}||_\infty = ||\sum_{i=1}^n a_i \ell_i||_\infty$$

$$\leqslant \sum_{i=1}^n |a_i| \, ||\ell_i||_\infty \leqslant n^{\frac{1}{2}} |a|$$

Thus, $\rho_1(E\mathcal{B}) \leqslant n^{\frac{1}{2}}$

and hence from (2.2b)

$$\rho_n(\mathcal{B}) \geqslant \frac{\rho_n(E\mathcal{B})}{\rho_1(E\mathcal{B})} \geqslant n^{-1}.$$

□

For boundary conditions of the form (1,2) we can obtain somewhat sharper estimates.

Lemma 3.4 For $\mathcal{B}$ given by (1.2) and satisfying (1.1), (2.1),

$$\rho_n(\mathcal{B}) \geq N_1^{-\frac{1}{2}}$$

where N_1 is the number of nontrivial matrices B_i in (1.2).

Proof Without loss of generality, we take $N_1 = N$

$$\| a^T E \mathcal{B} \|_\infty = \sum_{i=1}^{N} | B_i^T E^T a|$$

$$\leq N^{\frac{1}{2}} (a^T E \sum_{i=1}^{N} B_i B_i^T E^T a)^{\frac{1}{2}}$$

$$\leq N^{\frac{1}{2}} | E \sum_{i=1}^{N} B_i B_i^T E^T |^{\frac{1}{2}} | a|$$

Thus, $\rho_1 (E \mathcal{B}) \leq N^{\frac{1}{2}} | E \sum_{i=1}^{N} B_i B_i^T E^T |^{\frac{1}{2}}$

On the other hand,

$$\| a^T E \mathcal{B} \|_\infty = \sum_{i=1}^{N} |B_i^T E^T a|$$

$$\geq (a^T E \sum_{i=1}^{N} B_i B_i^T E^T a)^{\frac{1}{2}} \geq |a| / | (E \sum_{i=1}^{N} B_i B_i^T E^T)^{-1} |^{\frac{1}{2}}$$

Thus, $\rho_n (E \mathcal{B}) \geq 1 / (E \sum_{i=1}^{N} B_i B_i^T E^T)^{-1} |^{\frac{1}{2}}$

If we now take $E = (\sum_{i=1}^{N} B_i B_i^T)^{-\frac{1}{2}}$

then, from (2.2b), $\rho_n(\mathcal{B}) \geq \dfrac{\rho_n (E\mathcal{B})}{\rho_1 (E\mathcal{B})} \geq N^{-\frac{1}{2}}$

For an important class of boundary conditions, the bound in Lemma 3.4 is attained.

<u>Lemma 3.5</u> Let $\mathcal{B}$ be given by (1.2),

$$\sum_{i=1}^{N} \text{rank}\,(B_i) = n$$

and N_1 be the number of nontrivial matrices B_i in (1.2). Then,

$$\frac{\rho_n(\mathcal{B})}{\rho_1(\mathcal{B})} \leqslant N_1^{-\frac{1}{2}}$$

In addition, (2.2a,b) holds if and only if

$$\sum_{i=1}^{N} B_i B_i^T = N_1^{-1} I$$

<u>Proof</u>: Let us assume without loss of generality that $N_1=N$,

$$B_i^T B_i \eta_i = \sigma_i^2 \eta_i \quad , \quad |\eta_i| = 1 \quad , i = 1,\ldots,N$$

and

$$w_1 = 1 \quad , w_k = \text{sign}\left\{ \eta_k^T B_k^T \sum_{i=1}^{k-1} w_i B_i \eta_i \right\}, \; k=2,\ldots,N$$

Now,

$$\rho_1(\mathcal{B}) = \max_a \left\{ \frac{\sum_{i=1}^N |a^T B_i|}{|a|} \right\} \geqslant \max_a \left\{ \frac{\left|\sum_{i=1}^N w_i a^T B_i \eta_i\right|}{|a|} \right\} =$$

$$= \left|\sum_{i=1}^N w_i B_i \eta_i\right| \geqslant \left(\sum_{i=1}^N \eta_i^T B_i^T B_i \eta_i\right)^{\frac{1}{2}} = \left(\sum_{i=1}^N \sigma_1^2\right)^{\frac{1}{2}} .$$

This result holds for all singular values σ_i, and we may therefore take $\sigma_i = |B_i|$. Then $\rho_1(\mathcal{B}) \geqslant \left(\sum_{i=1}^N |B_i|^2\right)^{\frac{1}{2}}$.

In addition, for $\sigma_k \neq 0$,

$$\rho_n(\mathcal{B}) = \min_a \left\{ \frac{\sum_{i=1}^N |a^T B_i|}{|a|} \right\} = \sigma_k \min_a \left\{ \frac{\sum_{i=1}^N |a^T B_i|}{|a| \quad |B_k \eta_k|} \right\} \leqslant$$

$$\leqslant \sigma_k \min_a \left\{ \frac{\sum_{i=1}^N |a^T B_i|}{|a^T B_k \eta_k|} \right\} = \sigma_k .$$

Note that the last equality is not valid if Σ rank $(B_i) > n$. Nor is it valid for an arbitrary vector η_k.

Thus $$\frac{\rho_n(\mathcal{B})}{\rho_1(\mathcal{B})} \leqslant \min_k \frac{\sigma_k}{\left(\sum_{i=1}^N |B_i|^2\right)^{\frac{1}{2}}} \leqslant N^{-\frac{1}{2}}$$

which proves the first part of the lemma.

Now let (2.2a,b) hold. From lemma 3.4 and the result above

$$\sigma_k = N^{-\frac{1}{2}} \left(\sum_{i=1}^N |B_i|^2\right)^{\frac{1}{2}} .$$

Since, σ_k is an arbitrary singular value, all the singular values are equal, and using (2.2a) it follows that $\sum_{i=1}^{N} B_i B_i^T = N^{-1} I$.

Finally, let $\sum_{i=1}^{N} B_i B_i = N^{-1} I$.

Then, as previously, $\rho_1(\mathcal{B}) \geqslant \left(\sum_{i=1}^{N} |B_i|^2 \right)^{\frac{1}{2}} = 1$ and $\rho_1(\mathcal{B}) \leqslant N^{\frac{1}{2}} \ \left| \sum_{i=1}^{N} B_i B_i^T \right|^{\frac{1}{2}} = 1$.

Thus, $\rho_1(\mathcal{B}) = 1$.

In addition, as in lemma 3.4

$$\rho_n(\mathcal{B}) \geqslant 1 \Big/ \left| \left(\sum_{i=1}^{N} B_i B_i^T \right)^{-1} \right|^{\frac{1}{2}} = N^{-\frac{1}{2}}$$

and as this is the best possible, (2.21,b) holds. #

We now have the tools to assess the condition numbers α, β. Let us consider in particular (1.1) and the multipoint BC (1.2),

$$\mathcal{B}\, y = \sum_{i=1}^{N} B_i \, y(t_i),$$

for which we have the following useful properties

(3.2) $\Phi(t)\, B_i = G^+(t,t_i) - G^-(t,t_i)$, $i=1,\dots,N$,

where

(3.3a) $G^+(t,t_i) = \lim_{s \to t_i^+} G(t,s)$, $i=1,\dots,N-1$,

(3.3b) $G^-(t,t_i) = \lim_{s \to t_i^-} G(t,s)$, $i=2,\dots,N$,

(3.3c) $G^+(t,1) = G^-(t,0) = 0$.

Theorem 3.1 For $\mathcal{B}$ given by (2.1) and satisfying (2.2 a,b)

$$\beta \leqslant \frac{2 N_1 \alpha}{\rho_n(\mathcal{B})} \leqslant 2 N_1 \alpha \min (n, N^{\frac{1}{2}})$$

where N_1 is the number of nontrivial matrices B_1 in (3.2). If, in addition $\sum_{i=1}^{N} \operatorname{rank}(B_i) = n$,

then $\beta \leqslant 2 N_1 \alpha$.

Proof. Without loss of generality, we take $N_1 = N$. From (3.2), (3.3)

$$|\Phi(t) B_i| \leqslant 2\alpha$$

and hence $|\Phi(t)| \leqslant \left(\sum_{i=1}^{N} |\Phi(t) B_i|^2\right)^{\frac{1}{2}} \left|\left(\sum_{i=1}^{N} B_i B_i^T\right)^{-1}\right|^{\frac{1}{2}} \leqslant$

$$\leqslant 2 \alpha N^{\frac{1}{2}} \left| \left(\sum_{i=1}^{N} B_i B_i^T\right)^{-1} \right|^{\frac{1}{2}} .$$

The first result now follows from the inequality

$$\rho_n(\mathcal{B}) \leqslant N^{\frac{1}{2}} / \left| \left(\sum_{i=1}^{N} B_i B_i^T\right)^{-1} \right|^{\frac{1}{2}}$$

and lemmas 3.3 and 3.4.

However, if $\sum_{i=1}^{N} \operatorname{rank}(B_i) = n$

it follows from lemma 3.5 that $\left|\left(\sum_{i=1}^{N} B_i B_i^T\right)^{-1}\right|^{\frac{1}{2}} = N^{\frac{1}{2}}$

and this establishes the second part of the theorem. #

Thus, when $\mathcal{B}$ is given by (2.1) and N is not too large, the single parameter α is a suitable measure of the conditioning of the problem. However, as $N \to \infty$ we cannot bound β in terms of α using the results of

Theorem 3.1 which suggests that in general it is not possible to obtain such bounds. This is confirmed by the following example.

Example 3.1 Consider the problem

$$\mathcal{L}y = y' + ay \ , \ a > 0,$$

$$\mathcal{B}y = \int_0^1 y(s)ds,$$

for which $\alpha = 1, \beta = a\ (1-e^{-a})$ and $\rho_1\ (\mathcal{B}) = 1$. Clearly, β becomes unbounded as $a \to \infty$. #

Thus, in general both α and β need to be addressed in a discussion of stability.

§4. Polychotomy

For two point boundary value problems (i.e. N=2) it has become almost traditional to assume that the solution space

$$S(t) = \{\Phi(t)c \mid c \in \mathbb{R}^n\}$$

can be separated into a space

$$I(t) = \{\Phi(t)Pc \mid c \in \mathbb{R}^n\}\ ,\ P^2 = P$$

of 'non-decreasing' solutions and a space

$$D(t) = \{\Phi(t)(I-P)c \mid c \in \mathbb{R}^n\}$$

of 'non-increasing' solutions. In addition, if neither $I(t)$ nor $D(t)$ is trivial, (i.e. $P \neq 0,I$) it is usually assumed that the angle $0 < \eta(t) < \pi/2$ between $I(t)$ and $D(t)$, defined by

$$\cos \eta(t) = \max_{y_1 \in I(t),\ y_2 \in D(t)} \frac{|y_1^T y_2|}{|y_1|\ |y_2|}$$

is not too small. This has led to the notion of

<u>Definition 4.1</u> The solution space is <u>dichotomic</u> if there exists a projector P and a constant κ such that

(4.1a) $\quad |\Phi(t)P\Phi^{-1}(s)| < \kappa\ ;\ t > s$

(4.1b) $\quad |\Phi(t)(I-P)\Phi^{-1}(s)| < \kappa\ ;\ t < s$

κ is called the <u>dichotomy constant</u>

Although a projector always exists such that (4.1) is valid for some constant κ, we are primarily interested in the case when κ is of moderate size. In fact a more precise definition would involve the size of κ as well; we do not dwell on this however. It turns out that dichotomy is intimately connected with the conditioning of two point boundary value problems. Specifically, de Hoog and Mattheij [4],[5] have shown that:

<u>Theorem 4.1</u> When $N = 2$, there exists a projector P such that (4.1) holds with $\kappa = \alpha + 4\alpha^2$. Altematively, if (4.1) holds, then there exist matrices $B_1, B_2 \in \mathbb{R}^{n\times n}$ such that $\alpha \le \kappa$.

Thus, if $N = 2$ and α is of moderate size, the solution space is dichotomic (i.e. κ is also of moderate size). Conversely, if the solution space is dichotomic, there is a two point boundary value problem for which the condition number is not too large.

However, a well conditioned multipoint problem does not necessarily have a dichotomic solution space as can be seen from

<u>Example 4.1</u> Consider the problem

$$y' + 2\lambda(t-\tfrac{1}{2})y = f \quad , \quad \lambda > 0$$

$$y(\tfrac{1}{2}) = 1.$$

For this example,

$$\Phi(t) = \exp(-\lambda(t-\tfrac{1}{2})^2)$$

$$y(t) = \Phi(t) + \int_{\frac{1}{2}}^{t} \Phi(t)\Phi^{-1}(s)f(s)ds$$

and hence

$$\alpha = 1 \text{ (for all } \lambda).$$

Thus the problem is well conditioned but the fundamental solution now increases on the interval $0<t<\frac{1}{2}$ and decreases on $\frac{1}{2}<t<1$. Such behaviour is quite common in multipoint problems. Indeed, the results of de Hoog and Mattheij [4],[5] can be used to show that there exist projectors $\hat{P}_i$, $i=1,\dots,N-1$ such that

$$|\Phi(t)\hat{P}_i\Phi^{-1}(s)| < \kappa \ , \ t_i<s<t<t_{i+1}$$

$$|\Phi(t)(I-\hat{P}_i)\ \Phi^{-1}(s)| < \kappa \ , \ \ t_i<t<s<t_{i+1}$$

where κ is of moderate size if α is not large. Thus, on each interval $t_i<t<t_{i+1}$, $i=1,\dots,N-1$ the solution space is dichotomic.

However the examination of a number of well conditioned multipoint problems has suggested that additional structure is present in the solution space. This leads to the following generalization of dichotomy.

Definition 4.2 The solution space $S(t)$ is polychotomic if, for some $M \in \mathbb{N}$, and $0= x_1 \leqslant x_2 \leqslant \dots \leqslant x_M = 1$ there exist projectors P_k, $k = 1,\dots,M$ and a constant κ such that

$$\sum_{k=1}^{M} P_k = I \ , \ P_i P_j = P_j P_i \ = \delta_{ij} P_j$$

$$(4.2a) \qquad |\Phi(t) \sum_{j=1}^{k} P_j \ \Phi^{-1}(s)| < \kappa \ ; \ x_k < s < x_{k+1} \ , \ t > s$$

$$(4.2b) \qquad |\Phi(t) \sum_{j=k+1}^{M} P_j \ \Phi^{-1}(s)| < \kappa \ ; \ x_k < s < x_{k+1} \ , \ t < s$$

In section 5 we show that the concept of polychotomy is closely related to the conditioning of multipoint boundary value problems in the sense that κ will be of moderate size when α is not too large. It turns out that this relationship can be exploited in the construction of efficient numerical schemes for the solution of (1.1),(1.2) and this is discussed in detail in [5].

§5. Bounds for Polychotomy

In this section we show how the condition number α can be used to obtain bounds for κ. Initially we consider separable boundary conditions.

5.1 Separable BC

Definition 5.1 The boundary condition (1.2) is called separable if

$$\sum_{i=1}^{N} \text{rank } (B_i) = n.$$

Thus for separable boundary conditions, the solution space consists of a number of modes each of which is controlled by a condition at one of the points when rank $(B_i) \neq 0$.

We shall see that when the boundary condition (1.2) is separable, the solution space is polychotomic with constant $\kappa = \alpha$. However before we can show this some preliminary results are required.

Lemma 5.1 If $C_k \in \mathbb{R}^{n\times n}$, $k = 1,\ldots,N$

$$\sum_{k=1}^{N} C_k = I$$

and

$$\sum_{k=1}^{N} \text{rank}(C_k) = n$$

then C_k, $k=1,\ldots,N$ are projectors (i.e. $C_i C_j = C_j C_i = \delta_{ij} C_j$).

Proof The result follows from the arguments used in [6, Theorem 3.2]. ⊠

Lemma 5.2 For $E_k \in \mathbb{R}^{n\times n}$, $k = 1,\ldots,N$, let

$$\sum_{k=1}^{N} E_k = I \ , \qquad \sum_{k=1}^{N} \operatorname{rank}(E_k) = n$$

and define

$$\hat{G}(t,s) = \begin{cases} Y(t) \sum_{k=1}^{i} E_k Y^{-1}(s) \ ; \ t_i < s < t_{i+1}, \ t > s \\ \\ -Y(t) \sum_{k=i+1}^{N} E_k Y^{-1}(s) \ ; \ t_i < s < t_{i+1}, \ t < s \end{cases}$$

where Y is a fundamental solution of (1.1). Then there exists a boundary condition

$$\text{(5.1)} \qquad \hat{\mathcal{B}}y := \sum_{i=1}^{N} \hat{B}_i y(t_i)$$

satisfying $\operatorname{rank}(\hat{B}_i) = \operatorname{rank}(E_i)$ and

$$\sum_{i=1}^{N} \hat{B}_i \hat{B}_i^T = N_1^{-1} I$$

such that $\hat{G}$ is the Green's function associated with (1.1),(5.1) and N_1 is the number of nontrivial matrices E_i.

Proof: Consider the LQ^T decomposition

$$[E_1 Y^{-1}(t_1) | E_2 Y^{-1}(t_2) | \dots | E_N Y^{-1}(t_N)] = LQ^T$$

where $L \in \mathbb{R}^{nxn}$ is lower triangular and nonsingular and $Q \in \mathbb{R}^{(N+1)nxn}$ is orthogonal (i.e. $Q^T Q = I$). Now define $\hat{B}_i \in \mathbb{R}^{nxn}$, k=1,...,N by

$$[\hat{B}_1 | \hat{B}_2 | \dots | \hat{B}_N] := N_1^{-1} Q^T \ .$$

If we define

$$\hat{\Phi}(t) := Y(t)(\hat{\mathcal{B}}Y)^{-1},$$

we see that $\hat{\Phi}(t) = Y(t)L$. It is easy then to verify that $\hat{G}$ is the Green's function associated with (1.1),(5.1), viz.

$$\hat{G}(t,s) = \begin{cases} \hat{\Phi}(t) \sum_{i=1}^{K} \hat{B}_i \Phi(t_i) \Phi^{-1}(s), & t > s \\ -\hat{\Phi}(t) \sum_{i=k+1}^{N} \hat{B}_i \Phi(t_i) \Phi^{-1}(s), & t < s \end{cases}$$

can be identified with $\hat{G}(t,s)$

⊠

The relationship between polychotomy and the condition number for separable boundary conditions is now straightforward. Specifically we have

Theorem 5.1 If the boundary condition (1.2) is separable then the solution space is polychotomic with $\kappa \leqslant \alpha$

Conversely, if the solution space of (1.1) is polychotomic with constant κ, then there exists a separable boundary condition (1.2), satisfying assumption 2.1, such that $\alpha \leqslant \kappa$

Proof If the boundary condition (1.2) is separable

$$\sum_{i=1}^{N} \text{rank } (B_i) = n$$

and

$$\sum_{i=1}^{N} B_i \Phi(t_i) = I \quad (\text{cf.}(2.3b))$$

Thus

$$\sum_{i=1}^{N} \text{rank } (B_i \Phi(t_i)) = n$$

and from Lemma 5.1,

$$P_i = B_i\,\Phi(t_i)\ ,\ i = 1,\dots,N$$

are projectors. On substituting for P_i in the Green's function (1.5) and comparing the resulting expression with the definition of polychotomy (see definition 5.1), we find that (4.2) holds with $\kappa = \alpha$, $M = N$ and $x_j = t_j$.

If on the other hand the solution is polychotomic, then

$$|G(t,s)| \le \kappa$$

where

$$G(t,s) = \begin{cases} Y(t) \sum_{i=1}^{k} P_i\, Y^{-1}(s)\ ; & x_k < s < x_{k+1}\ ,\ t > s \\ -Y(t) \sum_{i=k+1}^{M} P_i\, Y^{-1}(s)\ ; & x_k < s < x_{k+1}\ ,\ t < s \end{cases}$$

and

$$\sum_{i=1}^{M} P_i = I\ ,\ P_i P_j = P_j P_i = \delta_{ij} P_j\ .$$

But from Lemma's 5.2 and 3.5 there exists a separable boundary condition of the form (1.2) which satisfies assumption 2.1 and is such that G is the Green's function associated with (1.1),(1.2) when $N = M$ and $t_i = x_i$.

#

5.2 General BC

We now turn again to the general BC (2.1) and show how we can select appropriate separable BC from them; this is based on the theory given in section 2.

Let

$$S = \{Ya \mid a \in \mathbb{R}^n\}$$

with

$$||y|| = ||y||_\infty \quad , \; y \in \mathcal{S}$$

Clearly, $\mathcal{S}$ equipped with the norm $||.||$ is a normed space of dimension n. In addition,

$$\mathcal{D} = \{y^* \in \mathcal{S}^* | y^*(y) = c^T y(t), \; |c| = 1, \; 0 \le t \le 1\}$$

is a closed boundary for $\mathcal{S}$. Hence, from Auerbach's Lemma (Lemma 2.2) there exist $y_j^* \in \mathcal{D}$, $y_i \in \mathcal{S}$; i,j = 1,...,n such that

$$y_j^*(y_i) = \delta_{ij} \; , \; ||y_j^*||^* = 1 \; , \quad ||y_i||_\infty = 1 \; ; \quad i,j = 1,\ldots,n$$

That is, there exist $c_j \in \mathbb{R}_1^n, |c_j| = 1$, points t_j with $0 \le t_j \le 1$ $j = 1,\ldots,n$ and $y_i \in \mathcal{S}$, $i = 1,\ldots,n$ such that

$$c_j^T \, y_i(t_j) = \delta_{ij} \; , \; |c_j| = ||y_i||_\infty = 1 \; ; \; i,j = 1,\ldots,n. \tag{5.2}$$

Furthermore,

$$c_j = y_j(t_j)$$

and hence

$$c_i^T \, c_j = 0 \text{ if } i \neq j \text{ and } t_i = t_j \tag{5.3}$$

Let

$$(\hat{P}y)(t) := \sum_{i=1}^{n} y_i(t) \; c_i^T \; y(t_i)$$

Thus,

$$\begin{aligned} ||\hat{P}y||_\infty &\le \sum_{i=1}^{n} ||y_i||_\infty \;\; ||y||_\infty \\ &\le n \; ||y||_\infty \end{aligned}$$

Hence

$$||\hat{P}||_\infty \le n$$

and, as in Lemma 3.2, we find that

$$\hat{\alpha} \le (1 + ||\hat{P}||_\infty)\alpha$$

$$\le (n+1)\,\alpha.$$

In addition, we have

$$\hat{\mathcal{B}}\;\hat{\Phi} = I$$

where

(5.4) $$\hat{\Phi} = N_1^{\frac{1}{2}} \left[y_1 | \ldots | y_n \right]$$

(5.5) $$\hat{\mathcal{B}}y := \sum_{i=1}^{N} {}_i\hat{\mathcal{B}}y\,(t_i)$$

$$B_k = N_1^{-\frac{1}{2}} \begin{bmatrix} o \\ C_k \\ o \end{bmatrix} \leftarrow \text{k+h position}$$

and N_1 is the number of <u>distinct points</u> in the set $\{t_k\}$. From (5.2), (5.3)

$$\sum_{k=1}^{n} \hat{B}_k \hat{B}_k^T = N_1^{-1}\, I$$

and hence from lemma 3.5, the boundary condition $\tilde{B}$ defined by (5.5), which is clearly separable, satisfies (2.2 a,b). Finally from (5.2), (5.5)

$$|\hat{\Phi}\;(t)| \leqslant N_1^{\frac{1}{2}}\, n^{\frac{1}{2}}\;.$$

Thus, we have shown

<u>Theorem 5.2</u> For a general BC (2.1) one can construct a separable BC $\hat{\mathcal{B}}$ of the form $\hat{\mathcal{B}}\,y := \sum_{i=1}^{n} \hat{B}_i\,y(t_i)$, with $t_i \in [0,1]$, such that $\hat{\mathcal{B}}$ satisfies (2.2a) and (2.2b) and for which (cf.(1.7))

$$\hat{\beta} := \sup_t |\hat{\Phi}(t)| \le n$$

$$\hat{\alpha} := \sup_{s,t} |\hat{G}(s,t)| \le (n+1)\ \alpha$$

<u>Corollary 5.1</u> If the BVP (1.1),(2.1) has a condition number α, then the solution space is polychotomic with

$$\kappa \le (n+1)\ \alpha$$

Note that the result of this corollary is somewhat different from Theorem 3.16 of [6], where bounds are derived $\sim \alpha^2$ for the two point case. For large α we may therefore say that this more general result is sharper, though not constructive.

<u>References</u>

[1] R.P. Agarwal, The numerical solution of multipoint boundary value problems. J. Comp.Appl.Math., 5 (1979), 17-24.

[2] F.V. Atkinson, Discrete and Continuous Boundary Problems, Academic Press, New York, 1964.

[3] C. de Boor, H.O. Kreiss, On the condition of the linear system associated with discretized BVPs of ODEs, SIAM J. Numer. Anal. 23 (1985), 936-939.

[4] E.W. Cheney, K.H. Price, Minimal Projections in Approximation Theory (A. Talbot ed.) Academic Press (1970), 261-289.

[5] F.R. de Hoog, R.M.M. Mattheij, The role of conditioning in shooting techniques, in 'Numerical Boundary Value ODES', edited by U. Ascher and R. Russell, Birkhäuser, 1985, 21-54.

[6] F.R. de Hoog, R.M.M. Mattheij, On dichotomy and well conditioning in BVP, to appear in SIAM J. Numer. Anal.

[7] F.R. de Hoog, R.M.M. Mattheij, An algorithm for solving multipoint boundary value problems,Computing 38 (1987), 219-234.

[8] H.B. Keller, Numerical Solution of Two-Point Boundary Value Problems, SIAM Regional Conference Series 24, Philadelphia, 1976.

[9] W.S. Loud, Generalized inverses and generalized Green's functions, J. SIAM 14 (1966), 342-369.

[10] R.M.M. Mattheij, Decoupling and stablility of algorithms for boundary value problems. SIAM Rev, 27 (1985), 1-44.

[11] R.M.M. Mattheij, G.W.M. Staarink, An efficient algorithm for solving general linear two-point BVP. SIAM J. Sci. Stat. Comp., 5 (1984), 745-763.

[12] W. Welsh, T. Ojika, Multipoint boundary value problems with discontinuities, I. Algorithms and applications. J. Comp. Appl. Math. 6 (1980), 133-143.

PREVIOUS PUBLICATIONS IN THIS SERIES:

Number	Author(s)	Title	Month
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/ Tani/Otomo/Shindo	Magnetoelastic buckling of two nearby ferromagnetic rods in a magnetic field	Dec. '87
87-17	A.J.E.M. Janssen/ S.J.L. van Eijndhoven	Spaces of type W, growth of Hermite coefficients, Wigner distribution and Bargmann transform	April '87
87-18	L. Dortmans/ A. Sauren/ 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/ 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
88-03	F.R. de Hoog R.M.M. Mattheij	On the conditioning of multipoint and integral boundary value problems	March '88