COLLOCATION ALGORITHMS AND ERROR ANALYSIS

FOR

APPROXIMATE SOLUTIONS

OF

ORDINARY DIFFERENTIAL EQUATIONS

A.H. Ahmed

Ph.D. Thesis

June 1981

University of Newcastle upon Tyne

ACKNOWLEDGEMENTS

I should like to thank my supervisor, Dr. Kenneth Wright, whose advice and enthusiasm throughout was very much appreciated.

I am also indebted to the typist, Anne Codling, who bravely tackled this difficult task.

Throughout the period of research for this thesis the author was supported by the University of Khartoum.

ABSTRACT

This thesis is mainly concerned with an error analysis of numerical methods for two point boundary value problems, in particular for the method of collocation using polynomial and certain piecewise polynomial bases.

As in previous work on strict error bounds an operator theoretical approach is taken. The setting for the theory and the principal results for later use are firstly considered. Then two types of 'a posteriori' error bounds are developed. These bounds are made computable by relating the inverse of the approximating operator to the inverse of certain matrices formed in the actual application of the approximation method.

The application of this theory to the numerical solution of linear two point boundary value problems is then considered. It is demonstrated how the differential equation can be split to fit into the setting required by the theory. It is also demonstrated how the global and the piecewise collocation method can be expressed in terms of a projection method applied to the operator equation. The conditions required by the theory are expressed in terms of continuity requirements on the coefficients of the differential equation and in terms of the distribution of the collocation points. In examining these bounds on a variety of problems, it is noticed that with some problems the conditions for applicability may not hold except for more points than one actually required to obtain a satisfactory solution. To improve the applicability, the theory is reconsidered with a different splitting of the differential equation. The method of collocation is expressed accordingly in terms of a new projection operator which is proved to have some nice properties in practice. This new approach is then compared

with the original one and it is shown to be superior on various problems.

By examining the inverse differential operator and the residual improved error bounds and estimates are shown to be obtainable. These estimates are tested in a large variety of examples and some graphs are presented to describe their behaviour in more detail. Finally these estimates are used to develop various adaptive mesh selection algorithms for solving two point boundary value problems. These strategies are tested and compared in several representative examples and some conclusions are drawn.

The thesis concludes with a brief review of the work with an indication of possible improvements and extensions.

CONTENTS

		Page No.
CHAPTER 0	INTRODUCTION	
	0.1 Aim	1
	0.2 Summary	1
CHAPTER 1	THEORY OF APPROXIMATION METHODS AND BOUNDS ON CERTAIN INVERSE OPERATOR	
	1.1. Introduction	4
	1.2. Setting for the theory	5
	1.3 Relevant results in compact operators	7
	1.4 The theory of projection method	8
	1.4.1 Introduction	8
	1.4.2 Projection method bounds	10
	1.5 The extended projection method	12
	1.5.1 Anselone-type theory	12
	1.5.2 Application to Anselone-type theory (Cruickshank extension)	14
	1.5.3 The extended projection method bounds	15
	1.6 Bounds on the approximating inverses	16
	1.6.1 The behaviour of $\|\|Q_n\|\|$ and $\|\|W_n\|\|$	19
	1.6.2. Bounds for the approximate inverses in terms of $\|\|W_n\|\|$ and $\|\|Q_n\|\|$	23
	1.7 Summary of bounds	24
CHAPTER 2	APPLICATION OF THEORY TO THE METHOD OF COLLOCATION	
	2.1 Introduction	26
	2.2 Form of problem	27
	2.3 The global collocation method	29
	2.3.1 Introduction to the method	29
	2.3.2 Satisfaction of the criteria for the application of the theory	29
	2.3.3 The behaviour of $\|\|W_n\|\|$ and $\|\|Q_n\|\|$	31
	2.4 The piecewise collocation method	35
	2.4.1 Introduction to the method	35
	2.4.2 Satisfaction of the criteria for the application of the theory	36
	2.4.3 The behaviour of $\|\|W_n\|\|$ and $\|\|Q_n\|\|$	37
	2.5 Bounds on the inverse differential operator	39
	2.5.1 Bounds for the constant terms $\|\|G^{-1}\|\|$, $\|\|K\|\|$, $\|\|D^dK^d\|\|$ and $\|\|G^{-1}T\|\|$	39
	2.5.2 Bounds for the projected terms	41
	2.5.3 Formulation of bounds	43
	2.6 Examples and Results	45
	2.6.1 Test problems	45
	2.6.2 $\|\|W_n\|\|$ and $\|\|Q_n\|\|$	46
	2.6.3 The applicability	50
	2.6.4 Bounds on $\|\|(G-T)^{-1}\|\|$	53
	2.6.5 Conclusions	53

CHAPTER 3 PRINCIPLE PART EXTENSION FOR BETTER APPLICABILITY

3.1.1 Introduction 64
3.1.2 New splitting of the differential operator 65
3.1.3 Study of the inverse of G* 66
3.2 The global collocation and the new projection 73
3.2.1 The projection norm 73
3.2.2 Relation between ϕ_n and ϕ_n^* 77
3.2.3 The satisfaction of the conditions of Chapter 1 79
3.3 Piecewise collocation 84
3.4 The choice of the parameters λ_i 93
3.5 Numerical Application 94
3.5.1 Derivation of bounds for $||k^*||$, $||D^2k^{*2}||$, $||k^*\phi_n^*||$, $||k^*P_{np}^*||$, $||(I-\phi_n^*)k^{*d}||$ and $||(I-P_{np}^*)k^{*d}||$ 95
3.5.3. Numerical results 98

CHAPTER 4 ALGORITHMS FOR ERROR BOUNDS AND ESTIMATES

4.1 Introduction 108
4.2 The behaviour of the residual 109
4.3 Improved error bounds using the polynomial approximation of the residual 111
4.4 Error estimate and estimates of the bounds 114
4.4.1 Estimates using $||Q_n||$ 114
4.4.2 Estimates using approximation of the residual 114
4.4.3 Estimates using the LU decomposition of the solution 115
4.4.4 Estimate using the principal part of the residual and the principal part of the equation only 116
4.5 Numerical examples 116
4.5.1 General comments on the computer program 116
4.5.2 The behaviour of $||Q_n||$ 122
4.5.3 Results for the global case 125
4.5.4 The piecewise case 136

CHAPTER 5 ADAPTIVE MESH SELECTION ALGORITHMS FOR BOUNDARY VALUE PROBLEMS

5.1 Introduction 156
5.2 Introduction to the algorithms 157
5.3 Comparisons 160
5.4 Improvements in the adaptive technique 170

CHAPTER 6 CONCLUSIONS

6.1 Summary 177
6.2 Improvements and extensions 178
6.2.1 The applicability 178
6.2.2 Error bounds and estimation 179

BIBLIOGRAPHY 180

TABLES INDEX

Table		Page No.
2.1	Problem Constants	46
2.2	$\|\|W\|\|$ and $\|\|Q\|\|$ values (global polynomials)	47
2.3	$\|\|W\|\|$ and $\|\|Q\|\|$ values (piecewise)	48
2.4	$\|\|W_n\|\|$ and $\|\|Q_n\|\|$ against α	49
2.5	Applicability (global polynomials)	51
2.6	Applicability (piecewise)	52
2.7	Bounds on $\|\|(G-T)^{-1}\|\|$ Problem (1) (Global)	55
2.8	Bounds on $\|\|(G-T)^{-1}\|\|$ Problem (2) (Global)	56
2.9	Bounds on $\|\|(G-T)^{-1}\|\|$ Problem (3) (Global)	57
2.10	Bounds on $\|\|(G-T)^{-1}\|\|$ Problem (4) (Global)	58
2.11	Bounds on $\|\|(G-T)^{-1}\|\|$ Problem (1) (Piecewise)	59
2.12	Bounds on $\|\|(G-T)^{-1}\|\|$ Problem (2) (Piecewise)	60
2.13	Bounds on $\|\|(G-T)^{-1}\|\|$ Problem (3) (Piecewise)	61
2.14	Bounds on $\|\|(G-T)^{-1}\|\|$ Problem (4) (Piecewise)	62
3.1	Experiments on the new principal operator (G*) with negative values of λ	72
3.2	Experiments with positive values of λ	72
3.3	Values of $\|\|\phi_n^*\|\|$ (negative λ)	76
3.4	Values of $\|\|\phi_n^*\|\|$ (positive λ)	76
3.5	Values of $\|\|P_{np}^*\|\|$ (p = 2)	89
3.6	Values of the constants	99
3.7	$\|\|W^*\|\|$ values	101
3.8	Applicability	103
3.9	Applicability with different values of λ for Problem 1, α = 2	106
3.10	Applicability with different values of λ for Problem 1, α = 0.5,1	107
4.1	The behaviour of $\|\|Q_n\|\|$ with the simple operator $x^m = y$, m=1,2,3,4 (global)	123
4.2	The behaviour of $\|\|Q_n\|\|$, problem 5,6,7,8 (global)	123
4.3	The behaviour of $\|\|Q_n\|\|$ with the simple operator $x^m = y$, m = 1,2,3,4(piecewise)	124

4.4 The behaviour of $||Q_n||$, problem 5,6,7,8 (piecewise) 124

4.5 The behaviour of the residual, problem 1, $y = 1$ 127

4.6 The behaviour of the residual, problem 1, $y = \sqrt{|t-0.9|}$ 127

4.7 The behaviour of the residual, problem 2, $y = \cosh(1)$ 128

4.8 The behaviour of the residual, problem 2, $y = \frac{1}{t^2+0.1}$ 128

4.9 The behaviour of the residual, problem 3, $y = \frac{1}{2(t+5)}$ 129

4.10 The behaviour of the residual, problem 3, $y = \begin{cases} t^2 + \sin t + 2 & t < 0 \\ (2-t)\,e^t & t > 0 \end{cases}$ 129

4.11 The behaviour of the residual, problem 4, $y = \frac{1}{t+3}$ 130

4.12 The behaviour of the residual, problem 5, $y = 1$ 130

4.13 The error estimates, problem 1, $y = 1$ 132

4.14 The error estimates, problem 1, $y = \sqrt{|t-0.9|}$ 132

4.15 The error estimates, problem 2, $y = \frac{1}{t^2+0.1}$ 133

4.16 The error estimates, problem 3, $y = \begin{cases} t^3+\sin t + 2 & t < 0 \\ (2-t)e^t & t \geq 0 \end{cases}$ 133

4.17 The error estimates, problem 5, $y = 1$ 134

4.18 The piecewise case, problem 2, $y = \frac{1}{t^2+0.1}$ 137

4.19 The piecewise case, problem 4, $y = \frac{1}{t+3}$ 137

4.20 The piecewise case $10^{-4}\, x'' - (2 - t^2)\, x = -1$ 139

4.21 The piecewise case. $x'' + 2t\, x' + 2x = 0$ 140

5.1 The simple adaptive scheme (problem 12) 162

5.2 The simple adaptive scheme (problem 13) 163

5.3 The simple adaptive scheme (problem 14) 166

5.4 The simple adaptive scheme (problem 15) 169

5.5 The improved adaptive scheme (problem 12) 173

5.6 The improved adaptive scheme (problem 13) 174

5.7 The improved adaptive scheme (problem 14) 175

5.8 The improved adaptive scheme (problem 15) 176

FIGURES INDEX

		Page No.
3.1	Illustrates how the basis of the solution (splines) looks	86
3.2	The matrix of solution using the spline basis	90
4.1	The matrix A with 3 partitions	119
4.2	The matrix P with 3 partitions	120
4.3	The matrix GF with 3 partitions	121
4.4	Problem (10) The residual	144
4.5	Problem (10) The residual (whole range)	145
4.6	Problem (10) The GR estimate	146
4.7	Problem (10) The GR estimate (whole range)	147
4.8	Problem (10) The E2 estimate	148
4.9	Problem (10), The E2 estimate (whole range)	149
4.10	Problem (10) The residual	150
4.11	Problem (10) The residual (whole range)	151
4.12	Problem (11) The GR estimate	152
4.13	Problem (11) The GR estimate (whole range)	153
4.14	Problem (11) The E2 estimate	154
4.15	Problem (11) The E2 estimate (whole range)	155
5.1	Problem (12), (The Solution)	161
5.2	Problem (13), (The Solution)	161
5.3	Problem (14), (The Solution)	165
5.4	Problem (14), (The Residual)	165
5.5	Problem (15), (The Solution)	168

Chapter 0

Introduction

0.1 Aim

An operator approximation theory is described in chapter 1 in an attempt to unify and extend other work arising mainly from studies of approximate solutions to integral and differential equations. The theory is placed in a general setting so as to permit as wide range of application as possible.

We are primarily concerned with finding strict error bounds for the approximate solution of linear two point boundary value problems in ordinary differential equations. These solutions will be the result of applying the method of collocation using polynomial and piecewise polynomial bases. The theory developed in chapter 1, however has much wider application.

Interesting error estimates arise as a by product of the work and are used in mesh selection algorithms for collocation codes for solving boundary value problems.

0.2 Summary

In sections (1.1 - 1.5) the theoretical background to the approximation method and the main results are presented. The theoretical results employed there are of a general nature and are derived primarily by Kantorovich & Akilov (1964) and Anselone (1971). Similar investigations have been pursued by Philips (1972), Coldrick (1972), Cruickshank (1974) and Gerrard (1979). The results are only based on certain operators being compact.

In section (1.6) two important convergence theorems which relate the inverse operator to the inverse of some matrices are presented. These theorems are important for the suitability of the bounds derived at the end of the section for the approximate inverse operator and are extremely valuable in justifying certain error estimates for the approximate solutions.

In chapter 2 a two point boundary value problem in ordinary differential equations is defined and expressed in the operator form required by the theory. In section (2.3) an approximate solution generated by the global collocation method is considered. It is shown using some results of Cruickshank (1974) and Wright (1979) that all the conditions required by the theory are satisfied by the global collocation method. In a similar way the piecewise collocation method is considered in section (2.4) and the conditions of the theory are verified using certain results of Gerrard (1979).

Having shown the theory is applicable section (2.5) proceeds to develop concrete numerical bounds on various operators and from these show how it is possible to obtain computable bounds for the differential operator. In section (2.6) test problems are used to illustrate the techniques and it is noted that in some cases the number of points or partitions needed to find the bounds is larger than one would like.

In chapter 3 that problem of applicability is considered. The differential equation is expressed in a parameteric operator equation. These parameters are constants to be chosen to allow the maximum possible applicability. In section (3.1) it is shown that this new operator equation satisfies the conditions of the theory under certain restrictions on the parameters. The method of collocation defines

a new projection operator which is proved to tend in norm to the usual projection operator with both the polynomial and piecewise polynomial bases. All conditions of the theory are shown to be satisfied by both methods and numerical methods are developed to calculate the norm of these projections and bounds on all other operators. At the end the applicability is discussed and compared on the test problems.

In chapter 4 improved error bounds and estimates are developed. In section (4.2) the behaviour of the residual is studied and useful properties are obtained with both the global and piecewise method. In section (4.3) and (4.4) it is shown that by examining the inverse differential operator and the residual one can obtain closer bounds and estimates with less work. In the last sections these estimates are compared with the actual error on a large selection of problems and some graphs are presented.

In chapter 5 various adaptive mesh selection algorithms based on the error analysis developed in this thesis are presented. These algorithms are tested and compared in a selection of badly behaved problems and some conclusions are drawn.

In all the computation with global methods we use zeros of Tchebychev polynomials as collocation points. Gauss points are compared with Tchebychev points for the piecewise case in chapter 4 and accordingly are chosen for the collocation codes in chapter 5. These two sets of points are widely used in such collocation methods, for example Tchebychev zeros by Wright (1964) and Gauss points by De Boor (1973).

All the calculations were performed in double precision arithmetic on IBM 360/370 computer.

Chapter 1

Theory of approximation methods and bounds on certain inverse operator

1.1 Introduction

In this chapter we introduce the theoretical background for certain operator equation and their approximate solution. We will be interested on the main 'a posteriori' theorems based on the work of Kantorovich & Akilov (1964) and Anselone (1971). Similar work was covered by Phillips (1972), Coldrick (1972), Cruickshank (1974) and Gerrard (1979).

The theorems are placed in a general setting so as to permit several possible areas of application. In later chapters we will be concerned with collocation as a projection method for the approximation solution of boundary value problems.

We now briefly define our problem and outline the general approach which we are going to follow in dealing with it.

Let X, Y be normed linear spaces and let $||\cdot||_X$, $||\cdot||_Y$ denote the norm in X and Y respectively. Let [X, Y] denotes the space of bounded linear operators mapping $X \to Y$ with the subordinate norm. We will be concerned with solving equations of the form

$$Dx = y \; ; \; y \in Y, \; D \in [X,Y] \tag{1.1}$$

for $x \in X$.

It is not always possible to solve (1.1) analytically and often a numerical method is used to approximate (1.1), e.g.

$$\tilde{D}\,\tilde{x} = \tilde{y} \; ; \; \tilde{y} \in \tilde{Y}, \; \tilde{D} \in [\tilde{X},\tilde{Y}] \tag{1.2}$$

solving for $\tilde{x} \in \tilde{X}$. This equation is usually set in a space of finite dimensions and corresponds to a finite set of linear algebraic equations. Now provided $\tilde{X} \subset X$ and D is invertable it follows that

$$x - \tilde{x} = D^{-1}(D(x - \tilde{x})) = D^{-1}(y - D\tilde{x})$$

$$\text{or} \quad ||x-\tilde{x}|| \leq ||D^{-1}||\ ||y-D\tilde{x}||_Y \quad , \tag{1.3}$$

which is a strict error bound on the approximate solution. 'A posteriori' bounds on $||D^{-1}||$ in terms of $||\tilde{D}^{-1}||$ may sometimes be computed and everything on the right hand side of the above inequality may be bounded once the approximate solution has been calculated.

Bounds on $||\tilde{D}^{-1}||$ can be calculated by relating them to the norms of certain matrices formed in the actual application of the numerical method. The norms of these matrices will be shown to have some nice properties making this approach more suitable in practice.

In the next section we introduce the general setting for the theory and put our problem in more specific form. In section (1.4) the main results for the projection method (Kantorovich & Akilov-type) are presented. The Anselone-type theory which will be called the extended projection method is introduced in section (1.5). Section (1.6) is dealing with the norm of the inverse of the approximate operator $\tilde{D}$ and some related matrices. At the end (section (1.7)) all these results are summed up to describe different ways of calculating computable bounds for the inverse operator D^{-1}.

1.2 Setting for the theory

In all of the following theory we shall be concerned with operators D which may be split into two parts

$$D = G - T$$

where G is invertable, $G^{-1} \in [Y,X]$. In certain circumstances we may deduce that D is invertable. Note that equation (1.1) may now be written

$$(G - T)\, x = y \tag{1.4}$$

and (1.3) becomes

$$||x-\tilde{x}||_X \leq \quad\quad || (G-T)^{-1}|| \; || (G-T)\tilde{x} - y||_Y \,. \tag{1.5}$$

We may apply G^{-1} to (1.4) giving

$$(I_X - G^{-1}T)\, x = G^{-1}y, \tag{1.6}$$

or we may replace x by $G^{-1}u$ where $u = G\,x$, giving

$$(I_Y - TG^{-1})u = y\,. \tag{1.7}$$

The identity operator $I_X \in [X,X]$, denoted $[X]$, and $I_Y \in [Y]$.

Since D is invertable, error bounds of the form (1.5) may be recovered from (1.6) and (1.7). For example if it is known that $G^{-1}(I - TG^{-1})^{-1} \in [Y,X]$ then

$$x-\tilde{x} = G^{-1}\,(I - TG^{-1})^{-1}\,(y - (G - T)\tilde{x})$$

so that $||x-\tilde{x}||_X \leq ||G^{-1}||\; ||I_Y - TG^{-1})^{-1}||\; ||(y-(G-T)\tilde{x}\,||_Y$. (1.8)

Similarly from (1.6)

$$||x-\tilde{x}||_X \leq ||\,(I_X - G^{-1}T)^{-1}||\; ||G^{-1}||\; ||(G-T)\tilde{x}-y||_Y \,. \tag{1.9}$$

Because G is invertable there is a close relation between the space X, Y and it is often the case that error bounds derived independently from (1.6) and (1.7) turn out to be equivalent when suitable practical norms are used. For example if we define the norm in X by

$||x||_X = ||Gx||_Y$ as in Kantorovich & Akilov, then

$$||G^{-1}|| = \sup_{||y||=1} ||G^{-1}y||_X = ||y||_Y = 1$$

and

$$||(I_X - G^{-1}T)^{-1}|| = \sup_{||x||=1} ||(I_X - G^{-1}T)^{-1}x||_X$$

$$= \sup_{||x||_X=1} ||G(I_X - G^{-1}T)^{-1}x||_Y$$

$$= \sup_{||Gx||_Y=1} ||GI_X G^{-1} - GG^{-1}TG^{-1})^{-1}Gx||_Y$$

$$= \sup_{||y||=1} ||(I_Y - TG^{-1})^{-1}y||_Y = ||(I-TG^{-1})^{-1}||,$$

which proves that bounds (1.8) and (1.9) are equal with this choice of norms.

The theory to be developed in this chapter will deal directly with equation (1.4) working in the two spaces X,Y. The infinity norm is going to be used all through unless otherwise stated. Other work in this field deals with the more conventional setting of a single space and works with the problem in its transformed form (1.8) or (1.9). Obviously dealing with (1.8) or (1.9) means loss of accuracy before hand due to multiplication of norms, e.g.

$$||(G-T)^{-1}|| \leq ||G^{-1}||\,||(I_y - TG^{-1})^{-1}||.$$

1.3 Relevant results in compact operators

In Anselone (1971), it is generally assumed that the normed space Y is complete. The results developed here will required only certain operators being compact. The reason for introduction of the attribute compact is exhibited in the following theorems quoted from Coldrick (1972) which show that if an operator K is compact then I-K enjoys some of the pleasant theoretical properties:

Theorem (1.1) (Coldrick (1972, page 12)

If K is a compact operator on Y then the following three statements are equivalent:

(i) $(I-K)^{-1} \in [Y]$

(ii) $(I-K)\,y=0$ implies $y = 0$,

(iii) $\inf_{||y||=1} ||(I-K)y|| = M$ for some $M>0$.

This theorem is a standard result, and can be found, for example in Appendix 1 of Anselone (1971).

Theorem (1.2) (Coldrick (1972, page 13)

If K is in [Y] such that $||K|| < 1$ and EITHER

(a) K is compact

or (b) Y is complete

then $||(I-K)^{-1}|| \leq \frac{1}{1-||K||}$.

This result is given by Anelsone (1971, proposition 1) when Y is a Banach space.

The importance of these two theorems is that they play the role of propositions (1.1) and (1.2) of Anselone (1971) when Y is not a Banach space.

We introduce next the background for the theory based on the work of torovich & Akilov (1964, page 541-601).

1.4 The theory of projection methods

1.4.1 Introduction

Let X_n, Y_n be subspaces of X and Y respectively with ϕ_n a linear projection $Y \rightarrow Y_n$. (The subscript n will have significance later, denoting the dimension of the subspace, but no restriction of dimensionality is made here).

An approximate solution $x_n \in X_n \subset X$ is found by requiring that the projection ϕ_n of equation (1.4) with x_n substituted for x

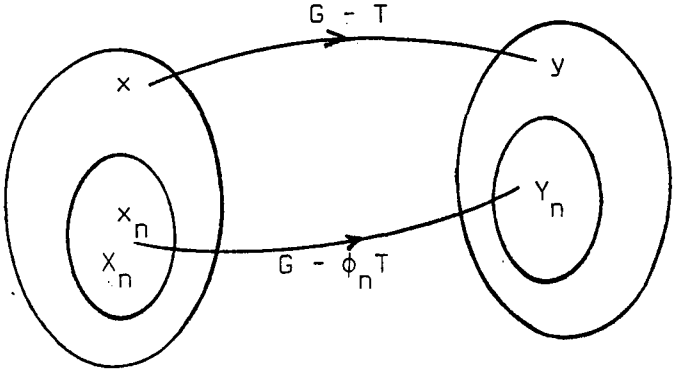
shall be zero, that is

$$\phi_n(Gx_n - Tx_n - y) = 0 \,. \tag{1.10}$$

It is assumed that $\phi_n G x_n = G x_n$, i.e. G restricted to X_n establishes a bijection between X_n and $Y_n = \phi_n Y$. Hence x_n satisfies

$$(G - \phi_n T)\, x_n = \phi_n y \,. \tag{1.11}$$

An intuitive concept of the situation described is illustrated below.



$G-T : X \rightarrow Y$

$G - \phi_n T : X_n \rightarrow Y_n$.

Note that $G-\phi_n T$ is regarded as being restricted to domain X_n.

We first state this result which is of 'a priori' nature.

Theorem (1.3)

Let Y_n be a subspace of normed linear space Y and let ϕ_n be a linear projection mapping $Y \longrightarrow Y_n$. Suppose that $K \in [Y]$ is compact and $(I-K)^{-1} \in [Y]$. Then if $\delta_n = ||(I-K)^{-1}||\, ||(I-\phi_n)K|| < 1$,

$(I-\phi_n K)^{-1}$ exists in $[Y]$ and

$$||(I-\phi_n K)^{-1}|| \leq \frac{||(I-K)^{-1}||}{1-\delta_n} \,.$$

This theorem is a reformulation of Kantorovich & Akilov theory quoted from Coldrick (1972, page 14). Its importance is it ensures that if $\delta_n \to 0$, then $(I-\phi_n K)^{-1}$ exists for sufficiently large n and its norm is uniformly bounded. This theorem can be extended as follows.

Cor (1.1) under the conditions of the above theorem with $K \equiv TG^{-1}$, $(G-\phi_n T)^{-1}$ exists in $[Y,X]$ and

$$||(G-\phi_n T)^{-1}|| \leq \frac{||G^{-1}||\ ||(I-K)^{-1}||}{1-\delta_n},$$

Proof: it follows easily from the theorem since it is assumed $G^{-1} \in [Y,X]$.

1.4.2 Projection method bounds

Theorem (1.4) with $K(\equiv TG^{-1})$ being compact, whenever $(I-\phi_n K^{-1})$ exists define $\delta_n^m = ||(\phi_n - I)K(I-\phi_n K)^{-1} K^m ||$.

Then if $\delta_n^m < 1$, $(G-T)^{-1}$ exists and

$$||(G-T)^{-1}|| \leq \sum_{i=0}^{m-1} \frac{||G^{-1}||\ ||K^i|| + ||(G-\phi_n T)^{-1}K^m||}{(1-\delta_n^m)}$$

$m = 0,1,2\ldots$

Proof: whenever $(I-\phi_n K)^{-1}$ exists, let

$$H = G^{-1}(I + K + \ldots K^{m-1}\ (I-\phi_n K)^{-1})K^m) \qquad m = 0,1,2,\ldots$$

be an approximate inverse of $(G-T)$.

$$(G-T)H = I + (\phi_n - I)\ K\ (I - \phi_n K)^{-1}\ K^m.$$

Since $(I-\phi_n K)^{-1}$ is bounded and $(\phi_n - I)K, K^m$ are compact, by Anselone (1971, page 59) $(\phi_n - I)K(I-\phi_n K)^{-1} K^m$ is compact. By theorem (1.2) if $\delta_n^m < 1$, $(I+(\phi_n - I)K(I-\phi_n K)^{-1}K^m)^{-1}$ exists and

$$||(I + (\phi_n - I)K(I-\phi_n K)^{-1}K^m)^{-1}|| < \frac{1}{1-\delta_n^m} ,$$

Which implies $G-T$ and $I-K$ have right inverses.

By theorem (1.1) since K is compact, $(I-K)^{-1}$ is unique and hence $(G-T)^{-1}$ is unique, that is the inverses are also left inverses.

Hence $$(G-T)^{-1} = H(I + (\phi_n - I)\, K\, (I - \phi_n K)^{-1}K^m)^{-1}$$

$$||(G-T)^{-1}|| \le ||H||/(1-\delta_n^m)$$

$$\le \frac{\sum_{i=0}^{m-1} ||G^{-1}||\,||K^i|| + ||(G-\phi_n T)^{-1}K^m||}{1 - \delta_n^m} . \quad (1.12)$$

The applicability of this theorem is guaranteed if $\delta_n^m \to 0$, since the existence and boundness of the approximate inverses $(I-\phi_n K)^{-1}$ and $(G-\phi_n T)^{-1}$ for sufficient, large n was ensured by theorem (1.3).

It was noted in the previous section that the approximate operator $(G-\phi_n T)^{-1}$ will be in $[Y,X]$ and $(I-\phi_n K)^{-1}$ in $[Y]$. If we denote these inverses when restricted to the subspaces by $(G-\phi_n T)^{-1}_{Y_n}$ and $(I-\phi_n K)^{-1}_{Y_n}$, then the following relations can be seen between them.

$$(I-\phi_n K)^{-1} = I + (I-\phi_n K)^{-1}_{Y_n}\, \phi_n K \quad (1.13)$$

$$(G-\phi_n T)^{-1} = G^{-1} + (G-\phi_n T)^{-1}_{Y_n}\, \phi_n K , \quad (1.14)$$

which implies

$$||(I-\phi_n K)^{-1}|| \le 1 + ||(I-\phi_n K)^{-1}_{Y_n}||\; ||\phi_n K||$$

$$||(G-\phi_n T)^{-1}|| \le ||G^{-1}|| + ||(G-\phi_n T)^{-1}_{Y_n}||\; ||\phi_n K|| .$$

Substituting these results in (1.12) we get

$$||(G-T)^{-1}|| \le \frac{||G^{-1}||\sum_{i=0}^{m} ||K^i|| + ||(G-\phi_n T)^{-1}_{Y_n}||\; ||\phi_n K^{m+1}||}{1 - \delta_n^m} \quad (1.15)$$

$$\text{if } \delta_n^m \leq \|(I-\phi_n K)\|\|K^m\|\|(1 + \|(I-\phi_n K)_Y^{-1}\| \quad \|\phi_n K\|) < 1.$$

In the next section we present a different type of projection method (the extended projection method) and derive similar bounds for the norm of the inverse operator $(G-T)^{-1}$.

1.5 The extended projection method

This method is based on the theory of Anselone (1972) and developed by Cruickshank (1974). Cruickshank in his thesis showed how this method could yield improved error bounds for approximate solution of boundary value problems obtained by global polynomial collocation methods. Here we follow that idea and develop similar bounds for the original operator $(G-T)^{-1}$. But first we introduce the theoretical background of the method.

1.5.1 Anselone-type theory

In this section we give a brief introduction to Anselone theory and state relevant results. We are not going to assume completeness as in Anselone. The theory uses the weaker pointwise(strong) convergence but requires additional compactness conditions.

Here the equation is of the form

$$(I-L)u = y \quad u, y \in Y \text{ and } L \in [Y].$$

An approximate solution $u_n \in Y$ to u is sought satisfying the equation

$$(I-L_n)u_n = y \text{ with } L_n \in [Y] \quad .$$

The following conditions are used,

With $L, L_n \in [Y]$ $(n = 1,2.....)$

(i) $L_n \to L$ i.e $\|L_n u - L u\| \longrightarrow 0 \ \forall u \in Y.$

(ii) L is compact

(iii) $[L_n]$ is collectively compact.

The theoretical results due to Anselone (1971) may now be stated.

Theorem (1.5) ('a priori') with conditions (i), (ii) and (iii) suppose $(I-L)^{-1}$ exists and define $\Delta_n = ||(I-L)^{-1}||\ ||(L_n-L)L_n||$. Then $\Delta_n \to 0$ as $n \to \infty$ and for $\Delta_n < 1$, $(I-L_n)^{-1}$ exists in $[Y]$ with

$$||(I-L_n)^{-1}|| \leq \frac{1 + ||(I-L)^{-1}||\ ||L_n||}{1 - \Delta_n} .$$

Proof: Anselone theorem (1.11) with Anselone lemma (1) replaced by theorem (1.2).

Theorem (1.6) ('a posteriori') with conditions (i), (ii) and (iii), whenver $(I-L_n)^{-1}$ exists define $\bar{\Delta}_n^m = ||(I-L_n)^{-1}||\ ||(L_n-L)L^m||$ (m integer > 1). Then, if for a particular value of n, $(I-L_n)^{-1}$ exists and $\bar{\Delta}_n^m < 1$, $(I-L)^{-1}$ exists and

$$||(I-L)^{-1}|| < \frac{1 + ||L|| + \ldots + ||L^{m-1}|| + ||(I-L_n)^{-1}||\ ||L^m||}{1 - \bar{\Delta}_n^m} .$$

Proof :

(Anelsone Theorem (1.12) with (Anselone lemma (1) replaced by theorem (1.2).

Nothing has so far been said concerning the uniform boundness of the $(I-L_n)^{-1}$, or the possibility of convergence as $n \to \infty$. However having obtained by the above result that $(I-L)^{-1}$ exists, theorem 1.5 can now be applied to show that $(I-L_n)^{-1}$ exists for all n sufficiently large and that its norm is uniformly bounded. Further the properties of collectively compactness give $\bar{\Delta}_n^m \to 0$. These deductions ensure that the estimates from the above theorem are uniformly bounded with respect to n.

1.5.2 Application to Anselone-type theory (Cruickshank extension)

We notice that Anselone-type theory is not immediately relevant to the usual projection method, since the right hand side of the approximate equation is $y \in X$ whereas in the projection method it is the projection of that term. Cruickshank (1974) working with equation (1.7) has suggested a variation analogous to the Nystrom extension for integral equations. That extension relates the two approximate methods and enables us to apply Anselone-type theory. Now we introduce that extension of Cruickshank modified for the original equation.

Suppose the operator $(G-\phi_n T) \in [X_n, Y_n]$ has a unique inverse $(G-\phi_n T)^{-1} \in [Y_n, X_n]$.

Define for each $n, w_n \in X$ by

$$w_n = G^{-1} y + G^{-1} T x_n . \tag{1.16}$$

The solution x_n can be related to w_n by using (1.11) and (1.16)

$$(G^{-1} \phi_n G) w_n = x_n \tag{1.17}$$

Define $T_n : X \to Y$

by $$T_n = T G^{-1} \phi_n G .$$

Then $$(G-T_n)w_n = (G-TG^{-1}\phi_n G)w_n$$

$$= y + TG^{-1}(G x_n - \phi_n y - \phi_n T x_n) .$$

Thus by (1.11) $$(G-T_n)w_n = y \tag{1.18}$$

It is easy to express the inverse of $(G-T_n) \in [X,Y]$ in terms of the inverse projection operator $(G-\phi_n T)^{-1}$ as follows:

$$(G-T_n)w_n = y$$
$$(G-T_n)\,(G^{-1}y + G^{-1}Tx_n) = y$$
$$(G-T_n)\,(G^{-1} + G^{-1}T\,(G-\phi_n T)^{-1}\phi_n y) = y$$
$$(G-T_n)\,(G^{-1} + G^{-1}T(G-\phi_n T)^{-1}\phi_n)y = y \quad .$$

So that $G^{-1} + G^{-1}T(G-\phi_n T)^{-1}\phi_n$ is a right inverse for $(G-T_n)$. Similarly we can show that it is also a left inverse for $(G-T_n)$, hence

$$(G-T_n)^{-1} = G^{-1} + G^{-1}T(G-\phi_n T)^{-1}_{Y_n}\,\phi_n \qquad (1.19)$$

and also $$(I-K\phi_n)^{-1} = I + K\,(I-\phi_n K)^{-1}_{Y_n}\,\phi_n \quad . \qquad (1.20)$$

1.5.3. The extended projection method bounds

We present now the following theorem which is a modification of theorem (1.6) to deal with $(G-T)^{-1}_{Y\to X}$.

Theorem (1.7) with (i), (ii) and (iii) whenever $(I-K\phi_n)^{-1}$ exists define $\Delta_n^m = ||K(\phi_n-I)\,(I-K\,\phi_n)^{-1}\,K^m\,||$. Then if $\Delta_n^m < 1$, $(G-T)^{-1}$ exists and

$$||(G-T)^{-1}|| \le ||G^{-1}||\,\sum_{i=o}^{m-1} ||K^i|| + ||(G-T_n)^{-1}||\,||K^m|| \qquad m=1,2,..$$
$$\frac{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}{1-\Delta_n^m}$$

Proof

Let $H = (I+K+\ldots+K^{m-1})\,(I-K\phi_n)^{-1}K^m$ be an approximate inverse of $(I-K)$.

$$(I-K)\,H = I + K(\phi-I)\,(I-K\,\phi_n)^{-1}\,K^m.$$

Since $(I-K\phi_n)^{-1}$ is bounded and $(K\phi_n-K)$, K^m are compact, by Anselone (1971, page 59) $K(\phi_n - I)\,(I-K\,\phi_n)^{-1}\,K^m$ is compact. Hence by theorem (1.2)

$(I+ K(\phi_n - I)(I-K\phi_n)^{-1})^{-1}K^m)^{-1}$ exists in Y and

$$||(I+K(\phi_n - I)(I-K\phi_n)^{-1}K^m)^{-1}|| \leq \frac{1}{1-\Delta_n^m} .$$

Which implies $G-T$ and $I-K$ have right inverses.

But since K is compact, by theorem (1.1) $(I-K)^{-1}$ is unique and hence $G^{-1}(I-K)^{-1} = G^{-1}H(I + K(\phi_n - I)(I - K\phi_n)^{-1}K^m)^{-1}$

and

$$||(G-T)^{-1}|| < \frac{||G^{-1}|| \sum_{i=0}^{m-1} ||K^i|| + ||(G-T_n)^{-1}||\,||K^m||}{1-\Delta_n^m} . \qquad (1.21)$$

The existence and uniform boundedness of $(I-K\phi_n)^{-1}$ and $(G-T_n)^{-1}$ follows from theorem (1.5) and collectively compactness gives $\Delta_n^m \to 0$ as mentioned before. This will ensure the applicability and the convergence of these bounds.

Finally, in order to apply bounds (1.15) and (1.21) we need to calculate bounds for the following quantities. (i) $||G^{-1}||$, (ii) $||K(I-\phi_n)||$, (iii) $||\phi_n K^m||$, (iv) $||(I-\phi_n)K^m||$, (v) $||(I-\phi_n K)_{Y_n}^{-1}||$, (vi) $||(I-K\phi_n)^{-1}||$, (vii) $||(G-\phi_n T)_{Y_n}^{-1}||$, (viii) $||(G-T_n)^{-1}||$. These bounds will be considered when applying the actual approximation method. However a general approach in bounding the norm of the approximating inverse operators is introduced in the next section.

1.6 Bounds on the approximating inverses

The idea here is to relate the approximate inverses on the finite dimensional spaces to some matrices formed in the actual application of the projection method. By putting these relations in operator form, we may express the norms of these approximate inverses in terms of the norms of these matrices. The suitability of this method depends on the behaviour of the norms of these matrices. Firstly,

we show how the equation in the X_n, Y_n subspaces can be related to the solution of a finite system of algebraic equations.

We use the vector space of dimension n, R^n, and we denote its identity by I_n (for any integer $n > o$). The projection operator ϕ_n will be taken for the purpose of illustrations to be the interpolation projection based on the n points $[\xi_j^n]_{j=1}^n$. We take as a basis for $Y_n = \phi_n Y$, the functions $[L_{n_j}]_{j=1}^n$, where

$$L_{n_j}(\xi_i^n) = \delta_{ji} \quad \text{(assuming such functions exist)}$$

In the sequal the subscript on ξ_i^n will be omitted.

Then $\phi_n y = y_n \in Y_n$ can be expressed as a linear combination of these elements L_{n_j},

$$y_n = \sum_{j=1}^{n} y_j L_{nj} .$$

By definition the basis for X_n will be $[L_{n_j}^*]_{j=1}^n$, where $L_{n_j}^* = \bar{G} L_{n_j}$, $j=1,\ldots\ldots, n$ and $x_n \in X_n$ can be expressed by

$$x_n = \sum_{j=1}^{n} \bar{x}_j L_{n_j}^* .$$

Define $H_n \in [Y, R^n]$ such that

$$(H_n u)_i = u(\xi_i) \text{ for every } u \in Y.$$

(Note that the evaluation is restricted to the interpolation points which implies $H_n \phi_n = H_n$).

$$||H_n|| = \sup_{||u||=1} ||H_n u|| = \sup_i |u(\xi_i)| = 1.$$

If we introduce a linear projection $\bar{\phi}_p : X \to X_n$ based on p points $[S_i]_{i=1}^p$ then in a similar way we define $\Phi_p \in [X, R^p]$ by

$$(\Phi_p x)_i = x(S_i) \text{ for every } x \in X.$$

Then clearly $\Phi_p \bar{\phi}_p = \Phi_p$ and $||\Phi_p|| < 1$.

Note $p = n + m$ for any integer $m \geq 0$.

We now recall equation (1.11),

$$(G-\phi_n T)x_n = y_n$$

$$\bar{\phi}_p x = x_n = (G-\phi_n T)^{-1}_{Y_n} y_n$$

$$(\Phi_p x)_j = [\Phi_p (G-\phi_n T)^{-1}_{Y_n} y_n]_j \qquad (1.22)$$

$$= [\Phi_p (G-\phi_n T)^{-1}_{Y_n} \sum_{i=1}^{n} \bar{y}_i L_{n_i}]_j \qquad j=1,\ldots,p$$

$$= \sum_{i=1}^{n} \bar{y}_i [\Phi_p (G-\phi_n T)^{-1}_{Y_n} L_{n_i}]_j \qquad j=1,\ldots,p$$

$$= \sum_{i=1}^{n} \bar{y}_i Q_{nij} \qquad j=1,\ldots,p$$

where $Q_{nij} = [\Phi_p (G-\phi_n T)^{-1}_{Y_n} L_{n_i}]_j$.

Then $\Phi_p x = Q_n H_n y$, (1.23)

where Q_n is the pxn matrix with ijth element Q_{nij}. We can go in a similar way and define an nxn matrix W_n by

$H_n u = W_n H_n y$ where the elements of W_n are $W_{nij} = (H_n (I-\phi_n K)^{-1}_{Y_n} L_{n_i})_j$.

These matrices W_n and Q_n can be interpreted in practice as the left inverse of the approximation matrix when the parameters defining the solution are taken as the values of u at the interpolation points $[\xi_j]_{j=1}^{n}$ or the values of x at the points $[S_j]_{j=1}^{p}$ respectively.

1.6.1 The behaviour of $||Q_n||$ and $||W_n||$

It is observed in later chapters that $||Q_n||$ and $||W_n||$ converge to certain values. That behaviour for $||W_n||$ has been observed previously by Cruickshank & Wright (1978) when they were considering computable error bounds for polynomial collocation methods. Later in 1979 Wright considered a theoretical justification of that behaviour and proved that under certain conditions, in fact $||W_n|| \to ||(I-K)^{-1}||$. Gerrard (1979), working with the piecewise collocation method, observed that $||W_n||$ approach the same constant irrespective of the interpolation scheme used. He also proved that $||W_n|| \to ||(I-K)^{-1}||$ if certain other conditions were satisfied.

In this section we are going to state this result in a general form and prove a similar result for $||Q_n||$.

The following assumption on the projection ϕ_n are used.

For any compact operator L,

(a) $[L\phi_n]$ is collectively compact and $L\phi_n \to L$.

(b) There is an extension operator $J_n \in [R^n, Y]$ such that $||J_n|| = 1$,

$$\phi_n J_n H_n = \phi_n \text{ and } H_n J_n = I_n \text{ and}$$

$$||L(\phi_n - I)J_n|| \to 0 .$$

Theorem (1.8)

Let K be a compact operator satisfying conditions (a) and (b). Then if (c) $||H_n(I-K)^{-1}J_n|| \to ||(I-K)^{-1}||$,

$$||W_n|| \to ||(I-K)^{-1}|| \text{ as } n \to \infty$$

The proof of this theorem was considered by Wright (1979) and by Gerrard (1979).

Theorem (1.9)

Let $X \subseteq Y$ and let K be a compact operator satisfying (a) and (b) and (d) $\| \Phi_p (G-T)^{-1} J_n \| \to \| (G-T)^{-1} \|$. Then

$$\| Q_n \| \to \| (G-T)^{-1} \| \text{ as } n \to \infty$$

where $p = n + m$ for some integer $m \geq 0$.

Proof. The proof goes through two stages.

(1) express Q_n in some equivalent operator form

(2) show that $\| Q_n - \Phi_p (G-T)^{-1} J_n \|$.

Then the result follows from (c). Also we note that since $X \subseteq Y$ and the infinity norm is used in both X and Y by taking $T = I_{X \to Y}$ in the original equation G^{-1} will be in $[Y]$ (a special form of K). This argument implies that all conditions satisfied by a general K are trivially satisfied by G^{-1}. That is G^{-1} is compact, $[G^{-1} \phi_n]$ is collectively compact and $G^{-1} \phi_n \to G^{-1}$.

(1) Q_n can be expressed in the operator form

$$Q_n = \Phi_p (G - \phi_n T)^{-1}_{Y_n} \phi_n J_n .$$

For, if we call (1.23)

$$\Phi_p x = Q_n H_n y \quad \text{for every } y \in Y \text{ and } x \in X,$$

and (1.22) in its vector form,

$$\Phi_p x = \Phi_n (G - \phi_n T)^{-1}_{Y_n} \phi_n y \text{ for every } y \in Y \text{ and } x \in X .$$

then $\Phi_p(G-\phi_n T)_{Y_n}^{-1}\phi_n y = Q_n H_n y$ for every $y \in Y$.

Hence $\Phi_p(G-\phi_n T)_{Y_n}^{-1}\phi_n = Q_n H_n$.

Multiply from the right by J_n to get

$$\Phi_p(G-\phi_n T)_{Y_n}^{-1}\phi_n J_n = Q_n \quad . \tag{1.24}$$

(2) $$||Q_n|| \to ||\Phi_p(G-T)^{-1}J_n|| = ||Q_n^*|| \quad .$$

We note here that Q_n^* depends on the particular choice of J_n unlike Q_n, because $\phi_n J_n$ is independent of the particular form of J_n.

$$Q_n - Q_n^* = \Phi_p G^{-1}(I-\phi_n K)^{-1}\phi_n J_n - \Phi_p G^{-1}(I-K)^{-1}J_n$$

$$= \Phi_p G^{-1}[(I-\phi_n K)^{-1}\phi_n - (I-K)^{-1}]J_n$$

$$= \Phi_p G^{-1}(I-\phi_n K)^{-1}[\phi_n(I-K)-(I-\phi_n K)](I-K)^{-1}J_n$$

$$= \Phi_p G^{-1}(I-\phi_n K)^{-1}(\phi_n - I)(I-K)^{-1}J_n$$

$$= \Phi_p G^{-1}[I + (I-\phi_n K)^{-1}\phi_n K\ ](\phi_n - I)(I-K)^{-1}J_n$$

$$= \Phi_p G^{-1}(\phi_n - I)(I + K(I-K)^{-1})J_n +$$

$$\Phi_p G^{-1}(I-\phi_n K)^{-1}\phi_n K(\phi_n - I)(I+K(I-K)^{-1})J_n$$

$$= \Phi_p G^{-1}(\phi_n - I)J_n + \Phi_p G^{-1}(\phi_n - I)K(I-K)^{-1}J_n$$

$$+ \Phi_p G^{-1}(I-\phi_n K^{-1})\phi_n J_n H_n K(\phi_n - I)(I+K(I-K)^{-1})J_n$$

$$= \Phi_p G^{-1}(\phi_n - I)J_n + \Phi_p G^{-1}(\phi_n - I)K(I-K)^{-1}J_n$$

$$+ Q_n H_n K(\phi_n - I)J_n + Q_n H_n K(\phi_p - I)K(I-K)^{-1}J_n \quad .$$

Hence

$$||Q_n - Q_n^*|| \leq ||\Phi_p||\ ||G^{-1}(\phi_n - I)J_n|| + ||\Phi_p||\ ||G^{-1}(\phi_n - I)K||\ ||(I-K)^{-1}||\ ||J_n||$$

$$+ ||Q_n||\ ||\Phi_p||\ ||K(\phi_n - I)J_n|| + ||Q_n||\ ||H_n||\ ||K(\phi_n - I)K||$$

$$||(I-K)^{-1}||\ ||J_n||$$

$||Q_n - Q_n^*|| \to 0$, since

(i) $||K(\phi_n - I)J_n||$, $||G^{-1}(\phi_n - I)J_n|| \to 0$ by (b)

(ii) $||K(\phi_n - I)K||$, $||G^{-1}(\phi_n - I)K|| \to 0$ by (a) and Anselone (1.8)

(iii) $||Q_n||$ is bounded as shown below.

(1.24) gives $Q_n = \Phi_p G^{-1}(I - \phi_n K)_{Y_n}^{-1} \phi_n J_n$

$$= \Phi_p G^{-1}[I + \phi_n K(I - \phi_n K)_{Y_n}^{-1}]\phi_n J_n$$

$$= \Phi_p G^{-1}[1 + K(I - \phi_n K)_{Y_n}^{-1}\phi_n]J_n$$

$$= \Phi_p G^{-1}(I - K\phi_n)^{-1} J_n \text{ by (1.20)}$$

$$||Q_n|| \leq ||G^{-1}\phi_n||\ ||(I - K\phi_n)^{-1}||$$

which implies that $||Q_n||$ is uniformly bounded if $||G^{-1}\phi_n||$ and $||(I-K\phi_n)^{-1}||$ are uniformly bounded. The first condition follows from the collective compactness of $[G^{-1}\phi_n]$ and the second one follows from theorem (1.6).

This theorem as well as theorem (1.8) is a strong evidence that $||W_n||$ or $||Q_n||$ is a good choice for expressing the norm of the approximate inverse operator, for in the limit at least, it is independent of the form of approximation as long as its conditions are satisfied.

1.6.2 Bounds for the approximate inverses in terms of $||W_n||$ and $||Q_n||$

In this section we derive bounds for the norms of the approximate inverses $(I-\phi_n K)_{Y_n}^{-1}$, $(G-\phi_n T)_{Y_n}^{-1}$, $(I-K\phi_n)^{-1}$ and $(G-T_n)^{-1}$ in terms of $||Q_n||$ or $||W_n||$. These forms of bounds are justified by the nice asymptotic properties of $||Q_n||$ and $||W_n||$, mentioned at the end of the previous section, and their easy calculation in practice.

Now define an extension operator $\psi_p \in [R^p, X]$ with the following properties : $||\psi_p|| = 1$, $\phi_p\psi_p\Phi_p = \phi_p$ and $\Phi_p\psi_p = I_p$. If we recall equation (1.24)

$$Q_n = \Phi_p(G-Q_nT)_{Y_n}^{-1}\phi_n J_n,$$

and use $\phi_n J_n H_n = \phi_n$, $\phi_p\psi_p\Phi_p = \phi_p$ and $\phi_p x_n = x_n$ for every $x_n \in X_n$, then we get

$$\phi_p\psi_p\Phi_p(G-\phi_nT)_{Y_n}^{-1}\phi_n J_n H_n = (G-\phi_nT)_{Y_n}^{-1} = \phi_p\psi_p Q_n H_n,$$

and

$$||(G-\phi_nT)_{Y_n}^{-1}|| \leq ||\phi_p||\ ||Q_n|| \quad (1.25).$$

If $||\phi_p||$ is increasing with n this bound may be unsatisfactory. An alternative may be derived as follows

$$(G-\phi_nT)_{Y_n}^{-1} = G^{-1}(I-\phi_nK)_{Y_n}^{-1}$$

$$= G^{-1}\phi_n J_n H_n (I-\phi_n K)^{-1}\phi_n J_n H_n,$$

but similar to Q_n, W_n can be expressed as

$$W_n = H_n(I-\phi_nK)_{Y_n}^{-1}\phi_n J_n \qquad (1.26)$$

Therefore $(G - \phi_n T)_{Y_n}^{-1} = G^{-1}\phi_n J_n W_n H_n$, and

$$||(G - \phi_n T)_{Y_n}^{-1}|| \leq ||G^{-1}\phi_n|| \; ||W_n|| \; . \tag{1.27}$$

If we recall (1.19) $(G - T_n)^{-1} = G^{-1} + G^{-1}T(G - \phi_n T)_{Y_n}^{-1}\phi_n$ and use the idea of (1.25) we get $||(G - T_n)^{-1}|| \leq ||G^{-1}T\phi_n|| \; ||Q_n|| + ||G^{-1}||$, (1.28)

In a similar way we deal with $(I - \phi_n K)_{Y}^{-1}$ and $(I - K\phi_n)^{-1}$ and get

$$||(I - \phi_n K)_{Y_n}^{-1}|| \leq ||\phi_n|| \; ||W_n|| \tag{1.29}$$

and $$||(I - K\phi_n)^{-1}|| \leq 1 + ||K\phi_n|| \; ||W_n|| \; , \tag{1.30}$$

Since (1.29) contains the factor $||\phi_n||$ which may grow, an alternative bound which may be of better asymptotic behaviour is given by Cruickshan k and Wright (1978)

$$||(I - \phi_n K)_{Y_n}^{-1}|| \leq \frac{1 + ||K\phi_n|| \; ||W_n||}{1 - ||(I - \phi_n)K||} \; , \text{ if } ||(I - \phi_n)K|| < 1.$$

1.7 Summary of bounds

Finally we use the results obtained in the previous section and substitute, the bounds of $||(G - \phi_n T)_{Y_n}^{-1}||$, $||(I - \phi_n K)_{Y_n}^{-1}||$ in (1.15) and the bounds of $||(G - T_n)^{-1}||$, $||(I - K\phi_n)^{-1}||$ in (1.21) and get:

Results (1)

From (1.15) (the projection method),

$$||(G - T)^{-1}|| \leq \frac{||G^{-1}|| \sum_{i=o}^{m} ||K^i|| + ||\phi_p|| \; ||Q_n|| \; ||\phi_n K^{m+1}||}{1 - \delta_n^m} \tag{1.31}$$

$$\text{or } \|(G-T)^{-1}\| \leq \frac{\|G^{-1}\| \sum_{i=0}^{m} \|K^i\| + \|G^{-1}\phi_n\| \, \|W_n\| \, \|\phi_n K^{m+1}\|}{1 - \delta_n^m}$$

$$m=0,1,\ldots$$

if $\delta_n^m < 1$ where

$$\delta_n^m = \min \begin{cases} \|(I-\phi_n)K\| \, \|K^m\| (1 + \|\phi_n\| \, \|W_n\| \, \|\phi_n K\|) , \\ \|(I-\phi_n)K\| \, \|K^m\| (1 + \dfrac{(1 + \|K\phi_n\| \, \|W_n\|)}{1-\|(I-\phi_n)K\|} \|\phi_n K\|) \end{cases}$$

(1.32)

if $\|(I-\phi)K\| < 1$.

N.B. δ_n^m is a bound on a previously defined δ_n^m.

Result (2)

From (1.21) (the extended projection method)

$$\|(G-T)^{-1}\| \leq \frac{\|G^{-1}\| \sum_{i=0}^{m} \|K^i\| + \|G^{-1}T\phi_p\| \, \|Q_n\| \, \|K^m\|}{1 - \Delta_n^m} \qquad (1.33)$$

$$m=1,2,\ldots$$

If $\Delta_n^m < 1$, where

$$\Delta_n^m = \|K(\phi_n - I)K^m\| + \|K(\phi_n - I)K\phi_n\| \, \|W_n\| \, \|K^m\|$$

N.B. Δ_n^m is a bound on the previously defined Δ_n^m.

The corresponding bounds derived by Cruickshank & Wright (1978) are:

$$\|(G-T)^{-1}\| < \frac{\|G^{-1}\| \sum_{i=0}^{m} \|K^i\| + \dfrac{1 + \|K\phi_n\| \, \|W_n\| \, \|\phi_n K^{m+1}\|}{1 - \|(I-\phi_n K)\|}}{1 - \delta_{mn}} \qquad (1.34)$$

If $\delta_{mn} < 1$, where δ_{mn} is the same as δ_n^m with $\|(I-\phi_n)K\| \, \|K^m\|$ replaced by $\|(I-\phi_n)K^{m+1}\|$; and

$$\|(G-T)^{-1}\| \leq \|G^{-1}\| \sum_{i=0}^{m-1} \|K^i\| + (1+\|K\phi_n\| \, \|W_n\|)\|K^m\| / 1-\Delta_{mn} \qquad (1.35)$$

if $\Delta_{mn} = (1 + \|K\phi_n\| \, \|W_n\|) \, \|K\| \, \|(I-\phi_n)K^m\| < 1$.

We notice here that the conditions necessary for the application of results to (1.32), (1.33) seems to be of no improvement over those required by (1.34) and (1.35). However if these conditions are satisfied we expect the bounds to give closer results than the corresponding ones in (1.34) and (1.35). This expectation will be tested in the next Chapter.

CHAPTER 2

Application of theory to the methods of Collocation

2.1. Introduction

Chapter 1 has developed an abstract theory concerning bounds on the inverse operator using bounds on its approximating inverse. This Chapter is an illustration of this theory applied to approximate solution of linear differential boundary value problems obtained by a particular projection method - collocation. Projection methods for differential equations are discussed by de Boor (1966) and Lucas and Reddien (1973). Collocation methods in particular are discussed widely and references include Karpilovskaja (1963), Wright (1964), Vainikko (1965, 1966), Phillips (1969), Lucas & Reddien (1972), Russell& Shampine (1972), de Boor & Swartz (1973), Cruickshan k (1974), Russell (1977), McKeown (1977) and Gerrard (1979). In all of the early work on the collocation method, polynomials were taken to be the basic functions. Following the investigation of piecewise polynomial interpolation, collocation methods based on piecewise polynomials have been widely used.

In this Chapter we will define the problem precisely, define the approximations that will be studied and verify the conditions of the theorems in the previous chapter . At the end of the chapter we include a selection of numerical results for illustration.

The theory could be applied to linear partial differential equations with no changes. However the derivation of certain constants required for strict bounds can be extremely lengthy and time consuming. Non-linear equations cannot be treated directly by the theory, but

bounds for each linear differential operator of an iterative sequence could be found and, hopefully, combined with further convergence results to produce a final bound.

2.2. Form of problem

We shall consider in this thesis problems of the form

$$L\,x(s) = x^{(m)}(s) + \sum_{k=o}^{m-1} p_k(s)\, x^{(k)}(s) = y(s) \qquad (2.1)$$

where $p_k(s)$ are continuous, over, say, $(-1,1)$, $x \in X$ and $y \in Y$; subject to the linearly independent boundary conditions

$$\sum_{k=o}^{m-1} \{\alpha_{ik}\, x^{(k)}(-1) + \beta_{ik}\, x^{(k)}(1)\} = 0 \qquad (2.2)$$

where α_{ik}, β_{ik} are constants. Y will be taken as the space of continuous functions $C(-1,1)$ and X may be taken as a subspace of $C^{(m)}\{-1,1\}$ satisfying the boundary conditions (2.2).

We define the operator G by $G = D^m$ and T by

$T = -\sum_{j=o}^{m-1} p_j D^j$ where D denotes the differentiation with respect to s.

Thus the differential equation (2.1) plus the boundary conditions (2.2) is equivalent to the operator equation

$$Gx - Tx = y \quad .$$

N.B. We note here that G can be chosen differently and that will be

considered in detail in the next chapter.

The only condition required now is that $x^{(m)}(s) = y(s)$ with the boundary conditions (2.2) should always have a solution. This corresponds to the requirement that the operator G in (1.4) must be invertable, which in turn equivalent to the existence of Green's function for this part of the operator. The operator K is defined from T and G by

$$(Ku)(s) = \int_{-1}^{1} K(s,t)\, u(t)\, dt$$

where

$$K(s,t) = -\sum_{j=0}^{m-1} P_j(s) \frac{\partial^j}{\partial s^j} g(s,t) \qquad \text{and where}$$

$g(s,t)$ is the Green's function corresponding to the operator $Gx = u$ with the given boundary conditions. The compactness of integral operators of the form K is proved for example by Kolmogorov and Fomin (1957).

We now state the conditions required, by the theorems in the previous chapter, on the projection operator ϕ_n. The projection method, Theorem (1.4), requires

(a) $\|(I - \phi_n) K^d\| \to 0$ as $n \to \infty$ $\qquad d = 1,2\ldots$

to ensure its applicability. For the applicability of the extended projection method, (Theorem 1.7), we use the weaker pointwise convergence,

(b) $K\phi_n \to K$, but require

(c) $\{K\phi_n\}$ to be collectively compact.

To prove the asymptotic properties of $\|W_n\|$ and $\|Q_n\|$, we require with (a), (b) and (c)

(d) $||K(\phi_n - I)\, J_n|| \to 0$ as $n \to \infty$

(e) $||H_n (I - K)^{-1} J_n|| \to ||(I - K)^{-1}||$

(f) $||\Phi_{n+m}(G - T)^{-1} J_n|| \to ||(G - T)^{-1}||$.

2.3. The Global Collocation Method

2.3.1. Introduction to the method

Let Y_n be taken as the space of polynomials of degree n-1 and X_n as polynomials of degree n+m-1 satisfying the boundary conditions. The approximate solution $x_n \varepsilon X_n$ is sought by requiring it to satisfy the equation to be solved exactly at n distinct points $\{\xi_j\}_{j=1}^n$ called the colloction points, i.e.

$$(G - T)\, x_n)(\xi_j) = y(\xi_j) \qquad j=1, \ldots, n \ . \tag{2.3}$$

An approximate equation of the form (1,11) is satisfied by the collocation solution x_n, where ϕ_n can be taken to be the projection mapping each continuous function to its interpolating polynomial of degree n-1 at the collocation points. For most of the results we need to assume these points are zeros of polynomial $Q_n(t)$ orthogonal with respect to $\rho(t) \geq m^* > 0$ for which

$$\omega^2 = \int_{-1}^{1} \rho(t)\, dt \quad \text{and} \quad \omega^{*2} = \int_{-1}^{1} \frac{1}{\rho(t)}\, dt \quad \text{are bounded.}$$

In some cases the collocation points are further restricted to be zeros of Jacobi polynomials $P_n^{\alpha,\beta}$ with $-\frac{1}{2} \leq \alpha, \beta \leq \frac{1}{2}$.

2.3.2. Satisfaction of the criteria for the application of the theory

It is shown in Cruickshank ((1974), section 4.3) if the collocation points are chosen as zeros of orthogonal polynomials then,

Lemma (2.1)

(i) $K\phi_n \longrightarrow K$

(ii) $\{K\phi_n\}$ is collectively compact.

By this lemma we satisfy conditions (b) and (c). For condition (a) we may require the boundedness of some higher derivatives of the coefficient of the differential equation as shown in the following lemma.

Lemma (2.2)

$\|(I-\phi_n)K^d\| \to 0$ as $n \to \infty$ provided the dth derivatives of the coefficients of the differential equation are bounded, where d is a positive integer depends on the weight function $\rho(t)$.

Proof

$$\|(I-\phi_n)K^d\| = \sup_{\|y\|=1} \|(I-\phi_n)K^d y\| .$$

If $y_n \in Y_n$ then

$$(I-\phi_n)K^d y = (I-\phi_n)(K^d y + y_n - y_n)$$

$$= (I-\phi_n)(K^d y - y_n) ,$$

then $\|(I-\phi_n)K^d y\| \leq \|(I-\phi_n)\| \; \|K^d y - y_n\|$.

It follows from Jackson's theorem (Cheney (1966 p.147)) that provided $K^d y \in C^d(-1,1)$, there is a polynomial y_n such that

$$\|K^d y - y_n\| \leq \left(\frac{\pi}{2}\right)^d \frac{\|D^d K^d y\|}{n(n-1)\cdots(n-d+1)} .$$

Hence $\|(I-\phi_n)K^d\| \leq \left(\frac{\pi}{2}\right)^d \dfrac{\|I-\phi_n\| \; \|D^d K^d\|}{n(n-1)\cdots(n-d+1)} \to 0$ as $n \to \infty$,

for the suitable value of d. The value of d which is necessary for the convergence depends on the bound of $||\phi_n||$ which depends on the orthogonal polynomial used. For example if the weight function $\rho(t) \geq m^*$ where $m^* > 0$, it is shown in Natanson (1965, page 52) that $||\phi_n|| < O(n)$ and therefore taking d=2 will be sufficient. For Tchebychev and Legendre polynomials where $||\phi_n|| \sim O(\ln n)$ and $||\phi_n|| \sim O(\sqrt{n})$ respectively, d = 1 is sufficient.

It can be seen from the definition of K^d that $K^d y \in C^d(-1,1)$ if the dth derivatives of the coefficients of the differential equation are bounded.

In the next section we show how the global collocation method can be put into the setting of theorem (1.8) and (1.9) and satisfy the conditions required for their application.

2.3.3. The behaviour of $||W_n||$ and $||Q_n||$

Let $\bar{\Phi}_{n+m}$ be a linear projection on X mapping each element into its interpolating polynomial of degree n+m-1, constructed using the points $\{s_i\}_{i=1}^{n+m}$. It is required later that the distance between these points should tend to zero as n tends to ∞.

Define $\Psi_{n+m} \in (R^{n+m}, X)$ so that $\Psi_{n+m} \underline{x}$ is a continuous linear function, as follows

$$\Psi_{n+m}\underline{x}(s) = \begin{cases} x(s_1) & s \leq s_1 \\ \dfrac{(s - s_{j-1}) x(s_j) - (s - s_j) x(s_{j-1})}{s_j - s_{j-1}} & s_{j-1} < s < s_j \quad j = 2, \ldots, n+m \\ x(s_{n+m}) & s \geq s_{n+m} \end{cases}$$

where $s_1 < s_2 < \ldots < s_{n+m}$,

and similarly $J_n \in (R^n, Y)$ by

$$J_n \underline{u}(s) = \begin{cases} u(\xi_1) & s \leq \xi_1 \\ \dfrac{(s - \xi_{j-1})\, u(\xi_j) - (s - \xi_j)\, u(\xi_{j-1})}{\xi_j - \xi_{j-1}} & \xi_{j-1} \leq s \leq \xi_j \\ & j = 2, \ldots, n \\ u(\xi_n) & s \geq \xi_n \end{cases}$$

if $\xi_1 < \xi_2 < \xi_3 \cdots < \xi_n$

(Wright 1979) .

Then if $\Phi_{n+m} \in (X, R^{n+m})$ and $H_n \in (Y, R^n)$ are defined as in section (1.6), we can state the following Lemma:

Lemma (2.3)

(i) $\Phi_{n+m} \Psi_{n+m} = I_{n+m}$, $\bar{\Phi}_{n+m} \Psi_{n+m} \Phi_{n+m} = \bar{\Phi}_{n+m}$ and $||\Psi_{n+m}|| = 1$

(ii) $H_n J_n = I_n$, $\phi_n J_n H_n = \phi_n$ and $||J_n|| = 1$.

Proof can easily be verified from the definitions .

Lemma (2.4)

(i) $||K(\phi_n - I) J_n|| \to 0$ as $n \to \infty$

(ii) $||H_n (I - K)^{-1} J_n|| \to ||(I - K)^{-1}||$

and (iii) $||\Phi_{n+m} (G - T)^{-1} J_n|| \to ||(G - T)^{-1}||$.

Proof The first and second results are given by Wright (1979). For the proof of the first one the collocation points are chosen to be zeros of Jacobi polynomial $P_n^{\alpha,\beta}(t)$ with $-\frac{1}{2} \leq \alpha,\beta \leq \frac{1}{2}$. We consider here the third result.

Since $||\Phi_{n+m}|| = ||J_n|| = 1$ then $||\Phi_{n+m}(G-T)^{-1} J_n|| \leq ||(G-T)^{-1}||$.

Now it is sufficient to show that $||\Phi_{n+m}(G-T)^{-1}J_n|| \geq ||(G-T)^{-1}||$.

Consider a function $y_o \in Y$ with $||y_o|| = 1$ such that $x_o = (G-T)^{-1} y_o$ satisfies $||x_o|| \geq ||(G-T)^{-1}|| - \varepsilon$, for small $\varepsilon > o$. Such a function must clearly exist from the definition of $||(G - T)^{-1}||$.

Let $\underline{\omega} = \Phi_{n+m} (G - T)^{-1} J_n H_n y_o$.

Now with the piecewise linear form of J_n

$$J_n H_n y \to y \text{ for any fixed } y \in Y,$$

for n sufficiently large

$$|| \underline{\omega} - \Phi_{n+m} x_o|| < \varepsilon .$$

If the maximum distance between the $\{s_j\}$ tends to zero as $n \to \infty$, then $||\Phi_{n+m} x || \to || x||$ for any fixed $x \in X$.

Hence for n sufficiently large

$$| (|| \omega || - ||x_o || | < 2 \varepsilon$$

and so $||\omega|| \geq ||(G-T)^{-1}|| - 3\varepsilon$.

But $||\underline{\omega}|| \leq ||\Phi_{n+m}(G-T)^{-1} J_n|| \; ||H_n|| \; ||y_o|| \leq ||\Phi_{n+m}(G-T)^{-1}J_n||$.

Therefore

$$||(G-T)^{-1}|| \leq || \Phi_{n+m} (G-T)^{-1} J_n || + 3\varepsilon \quad \text{if } n \text{ is sufficiently large.}$$

Now we can state the following theorem.

Theorem (2.1) with the assumption that the collocation points are zeros of Jacobi polynomial $P_n^{\alpha,\beta}(t)$ $-\frac{1}{2} \leq \alpha,\beta \leq \frac{1}{2}$ (i) $||W_n|| \to ||(I-K)^{-1}||$ and (ii) $||Q_n|| \to ||(G-T)^{-1}||$ where W_n and Q_n are defined by

$$W_n H_n y = H_n u \text{ and } Q_n H_n y = \Phi_{n+m} x.$$

Proof

By Lemma 2.1, 2.2, 2.3 and 2.4 we can verify all the conditions required by theorem (1.8) and (1.9).

If D^k denotes the kth differential operator, $D^k G^{-1}$ can be defined by $(D^k G^{-1})\, u(s) = \int_{-1}^{1} \frac{\partial^{(k)} g(s,t)}{\partial s^k} u(t)\, dt \qquad k < m.$

If the matrix Q_n^k denotes the inverse of the collocation matrix where the parameters defining the solution are taken as the values of $x^{(k)}$ at any $(n+m-k)$ points with the distance between them tends to zero as n tends to ∞, then in a similar way we can prove the following corollary.

Cor (2.1) with the assumption of the above theorem, if Q_n^k is defined by

$$\Phi_{n+m-k} x^{(k)} = Q_n^k H_n y,$$

then $||Q_n^k|| \to ||D^k G^{-1} (I - K)^{-1}|| \qquad k = 0,1,2\ldots\ldots, m-1,$

The set of points $\{\xi_j\}$ on which W depends must be the collocation points which are restricted to being zeros of orthogonal polynomials, while the set of points $\{s_j\}$ on which Q^k depend do not have this restriction. This is related to the theorem of convergence of collocation which states:

" The approximate solution with its derivatives up to m-1 inclusive tends to the true solution uniformly as $n \to \infty$ while the highest derivative tends on the mean square norm with weight $\rho(s)$ ".

Vainikko (1966).

2.4. The piecewise collocation method

2.4.1. Introduction to the method

Define the partition Π_n of $(-1,1)$ by the points $\{t_i\}_{i=0,..,n}$ $-1=t_o<t_1<\ldots<t_n=1$. We use the space of Riemann integrable functions $R(-1,1)$ for Y here to allow the use of piecewise polynomial approximation which may have discontinuities at the mesh points. The space X is taken as a subspace of $R^m(-1,1)$ satisfying the boundary conditions.

If $\{\xi_j\}$ is a set of p distinct points on $(-1,1)$, (including possibly, the end points), then the collocation points will be defined in each interval (t_{i-1}, t_i) by

$$\xi_{ji} = t_{i-1} + \frac{1+\xi_j}{2}(t_i - t_{i-1}) \qquad \begin{matrix} i=1,\ldots,n \\ j=1,\ldots,p \end{matrix}.$$

Y_{np} is taken as the space of piecewise polynomials with degree p in each subinterval (t_{i-1}, t_i). X_{np} will be the space of piecewise polynomials of degree m+p in each subinterval (t_{i-1}, t_i) and satisfy the boundary conditions (2.2) and the continuity conditions $x_{np}^{(k)}(t_i - o) = x_{np}^{(k)}(t_i+o)$ $\begin{matrix} k=o,\ldots m-1 \\ i=1,\ldots n \end{matrix}$ $\forall\, x_{np} \in X_{np}$. The piecewise collocation method requires the approximate solution x_{np} to satisfy the equation at the np collocation points $\{\xi_{ji}\}$. This is equivalent to the approximate equation (1.11) where ϕ_n is the projection P_{np} mapping each function in R to Lagrange interpolating polynomial of degree p

on each subinterval (t_{i-1},t_i) at the collocation points $\{\xi_{ji}\}_{j=1}^{p}$ $i=1,\ldots,n$

i.e. $(P_{np}\, y)(t) = \sum_{j=1}^{p} y(\xi_{ji})\ L_{ji}(t)$ $t \in (t_{i-1},t_i)$, $i=1,\ldots,n$

where L_{ji} is the unique polynomial such that $L_{ji}(\xi_{ki}) = \delta_{kj}$ for each $i = 1, \ldots, n$. We note here that the norm of P_{np} is given by the usual polynomial projection norm ϕ_p corresponding to interpolation at the points $\{\xi_j\}_{j=1}^{p}$ and is independent of n.

2.4.2. Satisfaction of the criteria for the application of the theory.

Under the following restrictions:

(i) Each of the polynomial interpolation points ξ_j is in the interval $(I_{j-1} - 1, I_j - 1)$, where

$$I_j = \sum_{i=1}^{j} \int_{-1}^{1} L_i(t)\ dt \qquad j = 1,\ldots p,$$

$I_o = 0$ and

L_i is the unique polynomial such that $L_i(\xi_j) = \delta_{ij}$.

(ii) $||\pi_n|| = \max_i |t_i - t_{i-1}| \to 0$ as $n \to \infty$,

Gerrard (1979) in his thesis (page 51) shows that

$\int_{-1}^{1} (P_{np}\, y)(t)\ dt \to \int_{-1}^{1} y(t)\ dt$ and hence proves (theorem 4.4 page 53) that,

Lemma (2.5)

(i) $K,\ K P_{np} \in (R)$ (ii) $KR \subset C$

(iii) K is compact (iv) $\{K P_{np}\}$ is collectively compact

(v) $K P_{np} \to K$ (vi) $||(I - P_{np})K|| \to 0$.

By this Lemma we satisfy condition (a), (b) and (c).

2.4.3. The behaviour of $||W_n||$ and $||Q_n||$

Let $H_{np} \in (R, R^{np})$, $J_{np} \in (R^{np}, R)$ be defined as follows.

$$(H_{np}\, u)_{ji} = u(\xi_{ji}) \qquad j = 1,\ldots,p, \quad i = 1,\ldots,n,$$

$$(J_{np}\, u_{ji})(t) = \sum_{u=1}^{n} \sum_{j=1}^{p} u_{ji}\, \chi(\tau_{i,j-1}, \tau_{i,j})(t),$$

where $u_{ji} = u(\xi_{ji})$,

$\tau_{ij} = t_{i-1} + I_j \dfrac{(t_i - t_{i-1})}{2}$ and

$$\chi(\tau_{i,j-1}, \tau_{i,j})(t) = \begin{cases} 1 & \tau_{i,j-1} < t < \tau_{ij} \\ 0 & \text{elsewhere} \end{cases}$$

Then from Gerrard (page 56,57)

$$H_{np} P_{np} = H_{np}, \quad H_{np} J_{np} = I_{np}, \quad P_{np} I_{np} H_{np} = P_{np},$$

$$||H_{np}|| = ||J_{np}|| = 1 \quad \text{and} \quad ||K(P_{np} - I) J_{np}|| \to 0 \quad \text{as } n \to \infty.$$

In a similar way let $\{s_j\}$ be any set of p+m distinct points in (-1,1) and define $\{s_{ji}\}$ by

$$s_{ji} = t_{i-1} + \frac{(1+s_j)}{2}(t_i - t_{i-1}) \qquad \begin{matrix} i = 1,\ldots,n \\ j = 1,\ldots,p+m. \end{matrix}$$

Define the linear projection $\bar{P}_{n(p+m)} : X \to X_n$ which maps each function in X to its Lagrange interpolating polynomial of degree p+m based on the p+m nodes $\{s_{ji}\}_{j=1}^{p+m}$ on each subinterval

$(t_{i-1} - t_i)_{i=1,\ldots,n}$ i.e.

$$(\bar{P}_{n(p+m)} x)(t) = \sum_{j=1}^{p+m} x(s_{ji})\, L^{x}_{ji}(t) \; ; \quad t \in (t_{i-1}, t_i),$$

where L^{x}_{ji} is the unique polynomial such that $L^{x}_{ji}(s_{ki}) = \delta_{kj}$.

Let $\Phi_{n(p+m)}$ be the evaluation operator in $(X^{n(p+m)}, R^{n(p+m)})$

defined by $(\Phi_{n(p+m)} x)_{ji} = x(s_{ji})_{j=1,\ldots,p+m}$.

Then clearly $||\Phi_{n(p+m)}|| = 1$ and $\Phi_{n(p+m)} \bar{P}_{n(p+m)} = \Phi_{n(p+m)}$.

Define the extension operator $\Psi_{n(p+m)} \in (R^{n(p+m)}, X)$ (similar to J_{np}) such that the following conditions are satisfied.

(i) $\Phi_{n(p+m)} \Psi_{n(p+m)} = I_{n(p+m)}$ (ii) $\bar{P}_{n(p+m)} \Psi_{n(p+m)} \Phi_{n(p+m)} = \bar{P}_{n(p+m)}$

(iii) $||\Psi_{n(p+m)}|| = 1$.

Lemma (2.6) Gerrard (1979, page 58)

Let $(I - K)^{-1} \in R$ then

$$||(I - K)^{-1}|| = \sup_{y \in C} ||(I - K)^{-1} y||.$$

This lemma makes the proof of conditions (g) and (h) in R equivalent to that in C which has been considered by Lemma (2.4).

Now we are in the position to state the following theorem.

Theorem (2.2)

$$\text{If} \quad \text{(i)} \; I_{j-1} - 1 < \xi_j < I_j - 1 \qquad j = 1,\ldots,p$$

$$\text{(ii)} \; ||T_n|| \to 0 \quad \text{as } n \to \infty \,,$$

then $$||W_{np}|| \to ||(I - K)^{-1}|| \quad n \to \infty \,,$$

$$||Q_{np}|| \to ||(G - T)^{-1}|| \text{ as } n \to \infty \,,$$

where W_{np} is the npxnp matrix defined by

$$H_{np}\, u = W_{np}\, H_{np}\, y$$

and Q_{np} is the $n(p+m) \times np$ matrix defined by

$$\Phi_{n(p+m)}\, x = Q_{np} \quad H_{np}\, y \,.$$

Proof: Verify the conditions of theorem (1.7) and (1.8) using the above results.

Corollary (2.2) With the assumptions of the above theorem

$||Q^k_{np}|| \to ||D^k G^{-1}(I - K)^{-1}||$, where Q^k_{np} is defined by

$$\Phi_{(n+m-k)p}\, x^{(k)} = Q_{np}\, H_{np}\, y \,.$$

2.5. Bounds on the inverse differential operator

To use the results at the end of the previous chapter it is necessary to achieve a computable bound for each item occurring in the various expressions. On examination of these expressions it can be seen that the following are required:

2.5.1. Bounds for the constant term $||G^{-1}||$, $||K||$, $||D^d K^d||$ and $||G^{-1}T||$

(i) $$||G^{-1}|| = \sup_{||u||=1} ||G^{-1}u|| = \sup_{||u||=1} \max_s \left| \int_{-1}^{1} g(s,t)\, u(t)\, dt \right|$$

$$\leq \max_s \int_{-1}^{1} |\, g(s,t)\, |\, dt = g_o$$

(ii) $$||K|| = \sup_{||u||=1} ||Ku|| = \sup_{||u||=1} \max_s \left| \int_{-1}^{1} K(s,t)\, u(t)\, dt \right|$$

$$< \max_s \int_{-1}^{1} |K(s,t)|\, dt = K_0$$

(iii) $$||DK|| = \sup_{||u||=1} ||DKu||$$

$$(DK)\, u(s) = (DT)\, x(s) = -D\left(\sum_{j=0}^{m-1} P_j\, x^{(j)} \right)(s)$$

$$= -\sum_{j=0}^{m-1} \left(P_j'\, x^{(j)} + P_j\, x^{(j+1)}\right)(s) = \sum_{j=0}^{m-1} -P_j'(s) \int_{-1}^{1} \frac{\partial^j g(s,t)}{\partial s^j}\, u(t)\, dt$$

$$+ \sum_{j=0}^{m-2} -P_j(s) \int_{-1}^{1} \frac{\partial^{j+1} g(s,t)}{\partial s^{j+1}}\, u(t)\, dt + P_{m-1}(s)\, u(s).$$ Now if

$$\bar{K} = \max_s \left[\sum_{j=0}^{m-1} |P_j'(s)| \int_{-1}^{1} \left| \frac{\partial^j g(s,t)}{\partial s^j} \right| dt + \sum_{j=0}^{m-2} |P_j(s)| \int_{-1}^{1} \left| \frac{\partial^{j+1} g(s,t)}{\partial s^{j+1}} \right| dt \right],$$

then

$||DK|| \le \max_s |P_{m-1}(s)| + \bar{K} = K_1$ which is bounded if $\{P_j'(s)\}$ are bounded.

Similarly $||D^2K^2|| \le K_2$, where

$$K_2 = \max_S \sum_{j=0}^{m-1} |P_j''(s)| \int_{-1}^{1} \left| \frac{\partial^j g}{\partial s^j}(s,t) \right| dt + \sum_{j=0}^{m-2} |2\, P_j'(s)| \int_{-1}^{1} \left| \frac{\partial^{j+1} g}{\partial s}(s,t) \right| dt$$

$$+ \sum_{j=0}^{m-3} |P_j(s)| \int_{-1}^{1} \left| \frac{\partial^{j+2} g}{\partial s^{j+1}}(s,t) \right| dt + |P_{m-2}(s) + 2P_{m-1}'(s)|\, ||K|| + |P_{m-1}(s)|\, ||DK||$$

which is bounded if $\{P_j''(s)\}$ are bounded.

In a similar way we can derive bounds Kd for $||D^dK^d||$ which

require bounds on the dth derivatives of the coefficients of the differential equation.

$$\text{(iv)}\quad ||G^{-1}T|| = \sup_{||x||=1} ||G^{-1}Tx|| = \sup_{||x||=1} \max_s \left| \int_{-1}^{1} g(s,t) \sum_{j=0}^{m=1} P_j(t) x^{(j)}(t)\, dt \right|$$

$$= KK .$$

For the simple second order case,

$$\int_{-1}^{1} g(s,t)\,(P_1(t)\,x'(t)+P_0(t)x(t))dt = g(s,t)\,P_1(t)\,x(t)\,\Big|_{-1}^{1} \quad \backslash$$

$$- \int_{-1}^{1} \left(g(s,t)\,P_1'(t) + \frac{\partial g}{\partial t}(s,t)\,P_1(t)\right) x(t)\,dt + \int_{-1}^{1} g(s,t)P_0(t)x(t)\,dt$$

$$||G^{-1}T|| = \sup_{||x||=1} ||G^{-1}Tx|| = \sup_s \Big(\int_{-1}^{1} \Big| g(s,t)P_1'(t) + \frac{\partial g}{\partial t}(s,t)\,P_1(t)\Big|\,dt$$

$$+ \int_{-1}^{1} |g(s,t)\,P_0(t)|dt).$$

2.5.2. Bounds for the projected terms

(A) The Global Case

$$||G^{-1}\phi_n|| = \sup_{||u||=1} ||G^{-1}\phi_n u||$$

$$(G^{-1}\phi_n u)(s) = \int_{-1}^{1} g(s,t)\,\phi_n u(t)\,dt .$$

Using Cauchys inequality we have

$$((G^{-1}\phi_n u)(s))^2 \le \int \frac{1}{\rho(t)} (g(s,t))^2 dt - \int_{-1}^{1} \rho(t)\,(\phi_n u(t))^2\,dt$$

$$\le \Omega^2 \quad \Omega^{*2} \quad \max_{s,t} |g(s,t)|^2 ,$$

therefore

$$||G^{-1}\phi_n|| \leq \bar{g} = \Omega^* \Omega \; g_{max} \quad \text{here } g_{max} = \max_{s,t} |g(s,t)| .$$

Similarly $||K\phi_n|| \leq \bar{K} = \Omega^* \Omega \; K_{max}$ where $K_{max} = \max_{s,t} |k(s,t)|$.

For $||G^{-1}T\phi_{n+m}||$,

$$(G^{-1}T\phi_{n+m})x(s) = \sum_{j=0}^{m-1} \int_{-1}^{1} g(s,t)\, p_j(t)\, D^j \phi_{n+m} x(t)\, dt.$$

A simple bound can be $g_t = ||G^{-1}T||\, ||\phi_{n+m}|| = KK\, ||\phi_{n+m}||$,

where KK is the bound given for $||G^{-1}T||$. Alternatively, we may use the same arguments used in bounding $||G^{-1}T||, ||G^{-1}\phi_n||$ and get for the simple second order case

$$||G^{-1}T\phi_{n+m}|| \leq \Omega \Omega^* \max_s \{ |p_1'(s)g(s,t) + p_1(s) \frac{\partial g}{\partial t}(s,t)| + |p_o(s)g(s,t)| \}.$$

$||(I - \phi_n) K^d||$ is essentially an interpolation error and can be bound EITHER using Jackson's theorem as in lemma (2.2) and get

$$||(I - \phi_n) K^d|| \leq B_d = J_{dn}\, \nu_n\, ||D^d K^d|| ,$$

where $\nu_n = ||I - \phi_n||$ and $J_{dn} = (\frac{\pi}{2})^d \frac{1}{n(n-1)\ldots(n-d+1)}$.

OR using Peano Kernel theorem(Davis (1971, page 70))and get directly

$$||(I - \phi_n) K^d|| \leq P_{dn}\, ||D^d K^d||$$

where P_{dn} are the Peano constants for interpolation. Unfortunately the computed Peano constants are not strict bounds as the processes of numerical integration and maximisation are not exact. Moreover these constants need to be computed for every n which is time consuming. For these reasons we prefer to use Jackson's theorem.

We note also $||(I - \phi_n) K \phi_n|| < J_{1n}\nu_n\, ||D K \phi_n||$ and $||D K \phi_n||$ can be bounded for example by $\Omega \Omega^* K_{1max}$, where

$K_{1max} = \max_s | P_{m-1}^{(s)} |\, ||\varphi_n|| / \Omega \Omega^* + \bar{K}$

A bound C_d of $||\phi_n K^d||$ can be $||(I - \phi_n) K^d|| + ||K^d||$.

(B) The piecewise case

Since $||P_{np}|| = ||P_p||$ is independent of n, bounds on $||G^{-1}P_{np}||$, $||K P_{np}||$ and $||G^{-1}T P_{n(p+m)}||$ may be $||G^{-1}||\ ||P_{np}||$, $||K||\ ||P_{np}||$ and $||G^{-1}T||\ ||P_{n(p+m)}||$ respectively or may be similar to those derived for the global case.

A bound B_d for $||(I - P_{np}) K^d||$ can similarly be found by Jackson's theorem

$$B_d = \left\{\frac{\pi||\pi_n||}{4}\right\}^d \frac{||D^d K^d||\ ||(I-P_{np})||}{P(p-1).....(p-d+1)} .$$

For $||P_{np} K^d||$, its bound C_d can be either $||P_{np}||\ ||K^d||$ or $||(I - P_{np}) K^d|| + ||K^d||$.

2.5.3. Formulation of bounds

Collecting these results together bounds on $||(G - T)^{-1}||$ may now be expressed in terms of computable quantities.

Result 1 of section (1.6) gives,

$$||(G-T)^{-1}|| \leq QP_d = g_o \frac{\sum_{i=0}^{d} K_o^i + ||\phi_{n+m}||\ ||Q_n||\ C_{d+1}}{1 - \delta_n^d},$$

$$\text{or } ||(G - T)^{-1}|| \leq WP_d = g_o \frac{\sum_{i=0}^{d} K_o^i + \bar{g}\ ||W_n||\ C_{d+1}}{1 - \delta_n^d},$$

if $\delta_n^d < 1$, where

$$\delta_n^d = \min\left(B_1 K_o^d (1 + ||\phi_n||\ ||W_n||C_1),\ B_1 K_o^d \left(1+\frac{(1+\bar{K}\ ||W_n||)C_1}{(1 - B_1)}\right) \text{if } B_1 < 1\right)$$

$$d=0,1,2\ldots$$

Result 2 gives

$$||(G-T)^{-1}|| \leq QA_d = g_o \frac{\sum_{i=0}^{d} K_o^i + gt\ ||Q_n||\ K_o^d}{1 - \Delta_n^d},$$

if $\Delta_n^d < 1$

where $\Delta_n^d = (1 + \bar{K}_o ||W_n||)\ K_o^d B_1$, $\bar{K}_o = \Omega\ \Omega^*\ K_{1max}$ $d = 1,2\ldots$

For comparisons we present also bounds derived by Cruickshank and Wright (1978) and Gerrard (1979); From the projection method,

$$||(G - T)^{-1}|| \leq P_d, \text{where } P_d = g_o \frac{\left(\sum_{i=0}^{d} K_o^i + \frac{(1 + \bar{K}\ ||W_n||)C_{d+1}}{1-B_1}\right)}{1 - \delta_{dn}},$$

or
$$P_d = g_o \frac{\left(\sum_{i=0}^{d} K_o^i + ||\phi_n||\ ||W_n||\ C_{d+1}\right)}{1 - \delta_{dn}},$$

if $\delta_{dn} < 1$. $\delta_{dn} = \min\left(B_{d+1}\ (I+||\phi_n||\ ||W_n||C_1), B_{d+1}\frac{(1+(1+\bar{K}||W_n||C_1)}{(1 - B_1)} \text{if } B_1\right.$

From the extended projection method,

$$||(G-T)^{-1}|| \leq A_d, \text{where } A_d = g_o \frac{\left(\sum_{i=0}^{d-1} K_o^i + (1 + \bar{K}\ ||W_n||)\ K_o^d\right)}{1 - \Delta_{dn}},$$

if $\Delta_{dn} < 1$. $\Delta_{dn} = (1 + \bar{K}\ ||W_n||)\ K_o\ B_d$ $d = 1,2\ldots.$

We note that the bounds WP_d, QP_d and QA_d involve B_1 unlike P_d and A_d which involve B_d. So the bounds QP_d and WP_d with d>o and QA_d with d>1 are not considered further as they are not making use of the higher differentiability of the coefficients.

We note here bounds which use the norm of the projection operator may not be satisfactory for large n in the global case. However they may still give better results with small value of n and in the following the best of these bounds will be used.

2.6. Examples and Results

In this section we present a selection of numerical examples illustrating how the ideas can be applied in practice and what sort of results can be obtained. We use zeros of Tchebychev polynomials as collocation points and for piecewise case we work with two points only.

2.6.1. Test problems

For ease of comparison, we use the same four problems used by Cruickshank and Wright (1978) as basic test problems. Other problems have been considered, including higher order equations.

Problem 1

$$x'' + \alpha (1 + t^2) x = y \qquad x(\pm 1) = 0,$$

Problem 2

$$x'' - \alpha x = y \qquad x(\pm 1) = 0,$$

Problem 3

$$x'' - \frac{2\alpha}{(t+5)^2} x = y \qquad x(\pm 1) = 0,$$

Problem 4

$$x'' + \frac{2 \alpha x'}{(t + 3)} - \frac{2 \alpha x}{(t + 3)^2} = y \qquad x(\pm 1) = 0,$$

The parameter α is included to vary the stiffness of the problem.

The calculation of K K_{max}, K_{max}, K_o, K_1, KK, K_2 is straightforward and their values are given for the four problems in table (2.1).

TABLE (2.1) Problem Constants

Problem	$\frac{K}{\alpha}$ K_{max}	K_{max}/α	K_0/α	K_1/α	K K/α	K_2/α^2
1	1	1	0.5	2	0.583	3
2	0.5	0.5	0.5	1	0.5	0.5
3	0.0625	0.0625	0.0625	0.15625	0.0366	0.018
4	0.95	1.25	1.25	2.25	0.306	6.26

2.6.2. $||W_n||$ and $||Q_n||$

The convergence of $||W_n||$ and $||Q_n||$ is observed in table (2.2) and (2.3). When considering approximating $||(G-T)^{-1}||$ in a later chapter, the behaviour of $||Q_n||$ will be examined in more detail but the relevant points to be noted here are ;

(1) $||Q_n|| < ||G^{-1}||\ ||W_n||$ for all n and for all examples. This can be seen EITHER from the operator definition of Q_n in (1.24)

$$Q_n = \Phi_{n+m} (G-T)^{-1}_{Yn}\ \phi_n\ J_n = \Phi_{n+m}\ G^{-1}\ \phi_n\ J_n\ W_n, \text{ giving}$$

$$||Q_n|| \leq ||G^{-1}\phi_n||\ ||W_n||.$$

OR from theorems (2.1) and (2.2) for sufficiently large n

$$||Q_n|| \sim ||(G-T)^{-1}|| \leq ||G^{-1}||\ ||(I-K)^{-1}|| \sim ||G^{-1}||\ ||W_n||.$$

(2) This inequality becomes more obvious when we increase the value of α as seen in table (2.4). It will be shown later that this inequality is more pronounced if we consider higher order differential

TABLE (2.2) $||W||$ and $||Q||$ values (global polynomials)

Problem	α	0.5		1.0		2.0		100.0	
	n	$\|\|W_n\|\|$	$\|\|Q_n\|\|$	$\|\|W_n\|\|$	$\|\|Q_n\|\|$	$\|\|W_n\|\|$	$\|\|Q_n\|\|$	$\|\|W_n\|\|$	$\|\|Q_n\|\|$
1	5	1.3258	0.6515	1.9318	0.9318	13.4358	6.2179	526.8720	4.2706
	10	1.3254	0.6402	1.9306	0.9156	13.4678	6.1361	31.9845	0.1595
	15	1.3258	0.6515	1.9321	0.9321	13.5002	6.2501	19.9719	0.1273
	20	1.3257	0.6482	1.9318	0.9271	13.4935	6.2157	20.2048	0.1247
	25	1.3258	0.6515	1.9320	0.9321	13.5002	6.2501	21.9437	0.1345
	30	1.3257	0.6500	1.9320	0.9297	13.4973	6.2338	23.2598	0.1361
2	5	1.0144	0.4134	1.0234	0.3516	1.0346	0.2704	0.8169	0.0162
	10	1.0809	0.4067	1.1315	0.3464	1.1841	0.2665	1.0489	0.0126
	15	1.1210	0.4134	1.2014	0.3519	1.2956	0.2705	1.0758	0.0111
	20	1.1409	0.4114	1.2362	0.3503	1.3519	0.2693	1.1383	0.0106
	25	1.1542	0.4314	1.2954	0.3519	1.3894	0.2705	1.2387	0.0103
	30	1.1624	0.4125	1.2738	0.3512	1.4129	0.2699	1.3292	0.0102
3	5	1.0025	0.4917	1.0048	0.4836	1.0088	0.4683	1.0475	0.1102
	10	1.0090	0.4834	1.0177	0.4757	1.0339	0.4608	1.1567	0.1119
	15	1.0124	0.4918	1.0244	0.4837	1.0470	0.4683	1.3660	0.1121
	20	1.0144	0.4893	1.0283	0.4813	1.0546	0.4662	1.4831	0.1120
	25	1.0156	0.4917	1.0306	0.4836	1.0591	0.4683	1.5592	0.1119
	30	1.0164	0.4904	1.0321	0.4826	1.0621	0.4673	1.6114	0.1120
4	5	1.4935	0.4815	2.0570	0.4563	3.3172	0.3967	32.1639	0.2927
	10	1.5393	0.4799	2.1727	0.4608	3.6498	0.4091	198.9996	0.2283
	15	1.5483	0.4815	2.1956	0.4582	3.7182	0.4109	132.0303	0.0400
	20	1.5515	0.4826	2.2038	0.4606	3.7425	0.4109	148.5863	0.02188
	25	1.5529	0.4815	2.2075	0.4610	3.7539	0.4106	163.8159	0.0198
	30	1.5537	0.4828	2.2096	0.4603	3.7601	0.4102	173.4017	0.0198

TABLE (2.3) $||W||$ as $||Q||$ values (piecewise)

Problem	α	0.5		1.0		2.0		100.0	
	n	$\|\|W_n\|\|$	$\|\|Q_n\|\|$	$\|\|W_n\|\|$	$\|\|Q_n\|\|$	$\|\|W_n\|\|$	$\|\|Q_n\|\|$	$\|\|W_n\|\|$	$\|\|Q_n\|\|$
1	5	1.3244	0.6460	1.9242	0.9197	12.3286	5.6404	12.3415	0.0503
	10	1.3256	0.6486	1.9304	0.9268	13.2077	6.0791	88.8749	0.5567
	15	1.3257	0.6509	1.9312	0.9307	13.3679	6.1807	41.5038	0.2662
	20	1.3257	0.6508	1.9316	0.9307	13.4268	6.2071	29.0360	0.1851
	25	1.3257	0.6513	1.9318	0.9316	13.4525	6.2251	25.5366	0.1645
	30	1.3257	0.6512	1.9319	0.9315	13.4675	6.2309	24.9923	0.1524
2	5	1.1243	0.4117	1.2080	0.3509	1.3084	0.2701	0.8277	0.0100
	10	1.1653	0.4120	1.2790	0.3509	1.4218	0.2698	1.2500	0.0100
	15	1.1784	0.4132	1.3020	0.3518	1.4593	0.2704	1.4539	0.0100
	20	1.1856	0.4131	1.3146	0.3517	1.4797	0.2703	1.5709	0.0100
	25	1.1897	0.4133	1.3218	0.3519	1.4914	0.2704	1.6466	0.0100
	30	1.1925	0.4133	1.3269	0.3518	1.4997	0.2704	1.6997	0.0100
3	5	1.0129	0.4889	1.0253	0.4811	1.0489	0.4661	1.4113	0.1119
	10	1.0165	0.4898	1.0325	0.4817	1.0628	0.4664	1.6285	0.1083
	15	1.0178	0.4914	1.0350	0.4834	1.0677	0.4680	1.7111	0.1105
	20	1.0185	0.4912	1.0363	0.4831	1.0702	0.4677	1.7547	0.1091
	25	1.0189	0.4916	1.0371	0.4836	1.0717	0.4680	1.7813	0.1103
	30	1.0191	0.4915	1.0376	0.4834	1.0728	0.4680	1.7992	0.1094
4	5	1.4835	0.4758	2.0370	0.4486	3.990	0.4093	1.5049	0.0170
	10	1.5181	0.4829	2.1199	0.4607	3.5032	0.4103	7.7034	0.0182
	15	1.5302	0.4814	2.1500	0.4613	3.5867	0.4107	25.6737	0.01893
	20	1.5364	0.4826	2.1656	0.4609	3.6309	0.4108	75.7068	0.01932
	25	1.5402	0.4827	2.1751	0.4605	3.6583	0.4108	292.2509	0.0248
	30	1.5427	0.4823	2.1815	0.4610	3.6769	0.4109	995.6873	0.1197

TABLE (2.4) $||W_n||$ and $||Q_n||$ against α

Problem	1		2		3		4	
α	$\|\|W_n\|\|$	$\|\|Q_n\|\|$	$\|\|W_n\|\|$	$\|\|Q_n\|\|$	$\|\|W_n\|\|$	$\|\|Q_n\|\|$	$\|\|W_n\|\|$	$\|\|Q_n\|\|$
0.0001	1.0001	0.4975	1.0000	0.4974	1.0000	0.4974	1.0001	0.4975
0.1	1.0524	0.5218	1.0331	0.4776	1.0029	0.4958	1.1016	0.4952
1	1.9318	0.9271	1.2362	0.3503	1.0283	0.4813	2.2038	0.4606
2	13.4935	6.2157	1.3519	0.2693	1.0546	0.4662	3.7425	0.4109
100	20.2048	0.1247	1.1383	0.0106	1.4831	0.1120	148.5863	0.2188
1000	7.6801	0.0032	1.050	0.0019	1.1851	0.0153	226.8	0.03667
10000	2.1528	0.0002	0.9587	0.0003	1.0525	0.0025	13.9584	0.0027

n = 20

equations.

These two points indicate that bounds using $||Q_n||$ are expected to be superior to those using $||W_n||$.

(3) Finally the relatively large values of $||W_n||$ and $||Q_n||$ in problem 1, $\alpha = 2$ occur because there the problem is nearly singular: the equation

$$\lambda x'' + (1 + t^2)\, x = y \qquad x(\pm 1) = 0$$

has an eigen value near $\lambda = 0.46$.

2.6.3. The Applicability

Tables (2.5) and (2.5) give the smallest number of points or subintervals n, for which $\delta_{1n}, \delta_{2n}, \Delta_{1n}$ and Δ_{2n} are less than unity and hence for which the bounds are applicable. We note that:

(1) For many cases the number of points or partitions needed are much more than one would like. This problem will be dealt with in the next two chapters.

(2) Bounds using second derivatives are of better applicability than those using first derivatives only. It is shown in Gerrard (1979) in general the applicability is better when higher derivatives are used. That is easily justified from the type of bounds derived for $||(I - \phi_n)\, K^d||$ in (2.5.2).

(3) With the global polynomials the extended projection method (Δ_{1n}, Δ_{2n}) gives better applicability than the projection method (δ_{1n} and δ_{2n}), while with the piecewise case the projection method is superior in most problems. This is due to the poor bound on

TABLE (2.5) Applicability

Global Polynomials (number of points required)

Problem	α	$\delta_{1n} = \delta_n^1$	δ_{2n}	$\Delta_{1n} = \Delta_n^1$	Δ_{2n}
1	0.5	19	8	3	3
	1	82*	22	36	9
	2	>120*	>120*	>120*	95*
	100	>120*	>120*	>120*	>120*
2	0.5	7	3	2	2
	1	26	7	6	2
	2	104*	20	61	9
	100	>120*	>120*	>120*	>120*
3	0.5	2	2	2	2
	1	2	2	2	2
	2	3	2	2	2
	100	>120*	>120*	>120*	>120*
4	0.5	33	12	13	6
	1	>120*	38	>120*	24
	2	>120*	>120*	>120*	>120*
	100	>120*	>120*	>120*	>120*

*assumes $||W_n||$ constant

TABLE (2.6) Applicability

(Piecewise) (number of partitions required)

Problem	α	$\delta_{1n} = \delta_n^1$	δ_{2n}	$\Delta_{1n} = \Delta_n^1$	Δ_{2n}
1	0.5	5	3	2	2
	1.0	13	7	5	4
	2.0	160*	30	155*	27
	100.0	>200*	>200*	>200*	>200*
2	0.5	2	2	2	2
	1.0	5	4	2	2
	2.0	14	6	12	5
	100.0	>200*	>200*	>200*	>200*
3	0.5	2	2	2	2
	1.0	2	2	2	2
	2.0	2	2	2	2
	100.0	>200*	>100*	>200*	>200*
4	0.5	7	4	4	3
	1.0	21	11	21	11
	2.0	125*	34	200*	53
	100.0	>200*	>200*	>200*	>200*

*assumes $||W_n||$ constant

$||(I - \phi_n K)^{-1} Y_n||$ with the global case.

(4) Generally it seems the piecewise collocation method is more readily applicable than the global one.

2.6.4. Bounds on $||(G - T)^{-1}||$

We now examine the bounds on $||(G - T)^{-1}||$ using the formulae in section (2.5.3) with d = 1,2 i.e. using first and second derivatives of the coefficients only. With each problem we use the suitable values of α which allow early application of these bounds.

Tables (2.7 - 2.10) give the results for the global collocation method. If we compare P_d, WP_d, QP_d, A_d, QA_d presented in these tables we may observe the following inequalities:

Problem (1) $\alpha = 0.5$ $QA_1 < A_1 << P_1 < WP_1 < QP_1$,

$\alpha = 0.5,1$ $A_2 < P_2$,

Problem (2) $\alpha = 0.5$ $QA_1 < A_1 < P_1 < QP_1 < WP_1$,

$\alpha = 1$ $QA_1 < A_1 < P_1 < QP_1 < WP_1$,

$\alpha = 0.5,1,2$ $A_2 < P_2$

Problem (3) $\alpha = 0.5,1,2$ $QA_1 < A_1 < P_1 < QP_1 < WP_1$

$\alpha = 0.5,1.2$ $A_2 < P_2$

Problem (4) $\alpha = 0.5$ $QA_1 < A_1 << QP_1 < WP_1 < P_1$

$\alpha = 0.5$ $A_2 < P_2$

From these inequalities we notice:

(1) The extended projection method bounds ($QA_1, A_{1,2}$) are much more accurate than the projection method ones (QP_1, WP_1, $P_{1,2}$). This is confirmed from the expressions for $P_{1,2}, QP_1, A_{1,2}, QA_1$ which give when $B_1 < 1$,

$A_{1,2} < P_{1,2}$ or $QA_1 < QP_1$.

(2) If we consider each method separately we observe that with the extended projection method in all problems

For the usual projection method, in most cases QP_1 and WP_1 have no improvement over P_1.

In general if $||K|| > \frac{1}{2}$ it can be seen that $WP_1 < P_1$.

(3) If we compare the bounds using the first derivatives with those using second derivatives we observe that the latter are always better. That is obviously due to the bounds derived for $||(I - \phi_n)K^d||$.

Tables (2.11→2.14) give similar results for the piecewise collocation method. All the problems have consistently satisfied the following inequality.

$$QA_1 < QP_1 < WP_1 < A_1 < P_1, \quad A_2 < P_2$$

That confirms (i) the superiority of the bounds using the matrix Q (ii) the improvement of P_d by WP_d which is expected by the theory. Like the global case bounds using second derivatives are better. Gerrard (1973) shows further that if we use higher derivatives then the projection method will be superior to the extended projection method. That is because the first uses $||(I - P_{np})K^{d+1}||$ while the second uses $||(I - P_{np})K^d||$ and better bounds for $||(I - P_{np}K)^{-1}_{Yn}||$ are available with the piecewise method.

TABLE (2.7) Bounds on $||(G - T)^{-1}||$

Problem (1) (Global)

(i) Using first derivatives

α	n	P_1	WP_1	QP_1	A_1	QA_1
0.5	5				1.1587	0.8416
	10				0.9585	0.7704
	15				0.9006	0.7458
	20	5.6941	5.8610	6.9245	0.8734	0.7317
	25	2.8566	3.0241	3.6632	0.8578	0.7232
	30	2.1175	2.2724	2.7965	0.8479	0.7170
1	5					
	10					
	15					
	20					
	25					
	30					

(ii) Using second derivatives

α	n	P_2	A_2
0.5	5		0.7646
	10	1.2556	0.7037
	15	0.9623	0.6947
	20	0.8605	0.6923
	25	0.8163	0.6917
	30	0.7918	0.6917
1	5		
	10		2.2106
	15	201.6139	1.5651
	20	4.9507	1.424
	25	3.3429	1.3714
	30	2.8139	1.3478

TABLE (2.8) Bounds on $||(G - T)^{-1}||$

Problem (2) (Global)

(i) Using first derivatives

α	n	P_1	WP_1	QP_1	A_1	A_1
0.5	5				0.8356	0.8281
	10	2.5179	2.6470	2.4443	0.7846	0.7730
	15	1.4888	1.6561	1.5854	0.7625	0.7557
	20	1.2147	1.3820	1.3481	0.7506	0.7467
	25	1.0867	1.2511	0.2432	0.7435	0.7425
	30	1.0119	1.1727	1.1798	0.7385	0.7395
1	5				19.0553	17.9044
	10				2.7131	2.4740
	15				1.9724	1.8082
	20				1.7099	1.5744
	25	12.1448	11.0939	9.3647	1.5763	1.4610
	30	5.2210	4.9283	4.2239	1.4934	1.3913

(ii) Using second derivatives

α	n	P_2	A_2
0.5	5	0.9002	0.6838
	10	0.7869	0.6782
	15	0.7440	0.6772
	20	0.7258	0.6769
	25	0.7158	0.6769
	30	0.7093	0.6768
1	5	5.3775	1.2312
	10	1.7874	1.0899
	15	1.5987	1.0688
	20	1.5060	1.0625
	25	1.3922	1.0604
	30	1.3264	1.0593
2	5		
	10		8.8678
	15	10.7778	4.4260
	20	6.4739	3.7576
	25	5.4617	3.5176
	30	5.0276	3.3987

TABLE (2.9) Bounds on $||(G - T)^{-1}||$

Problem (3) (Global)

(i) Using first derivatives

α	n	P_1	WP_1	QP_1	A_1	QA_1
0.5	5	0.6101	0.6584	0.6682	0.5179	0.5175
	10	0.5668	0.6100	0.6183	0.5174	0.5171
	15	0.5517	0.5880	0.6029	0.5172	0.5169
	20	0.5438	0.5765	0.5938	0.5171	0.5169
	25	0.5390	0.5692	0.5889	0.5717	0.5168
	30	0.5357	0.5644	0.5852	0.5170	0.5168
1	5	0.7848	0.8723	0.8893	0.5407	0.5392
	10	0.6567	0.7394	0.7524	0.5389	0.5372
	15	0.6175	0.6874	0.7138	0.5380	0.5366
	20	0.5981	0.6612	0.6923	0.5375	0.5363
	25	0.5896	0.6451	0.6809	0.5372	0.5361
	30	0.5787	0.6342	0.6725	0.5370	0.5366
2	5	1.7790	1.8801	1.9040	0.6059	0.5994
	10	0.9625	1.1080	1.1174	0.5966	0.5892
	15	0.8139	0.9400	0.9789	0.5922	0.5858
	20	0.7508	0.8657	0.9147	0.5899	0.5840
	25	0.7156	0.8231	0.8814	0.5884	0.5830
	30	0.6930	0.7952	0.8586	0.5873	0.5824

(ii) Using second derivatives

α	n	P_2	A_2
0.5	5	0.5185	0.5162
	10	0.5171	0.5162
	15	0.5168	0.5162
	20	0.5166	0.5161
	25	0.5165	0.5161
	30	0.5165	0.5161
1	5	0.5445	0.5337
	10	0.5380	0.5335
	15	0.5364	0.5335
	20	0.5356	0.5335
	25	0.5352	0.5335
	30	0.5349	0.5335
2	5	0.6383	0.5747
	10	0.5967	0.5732
	15	0.5875	0.5729
	20	0.5835	0.5728
	25	0.5813	0.5727
	30	0.5798	0.5727

TABLE (2.10) Bounds on $||(G - T)^{-1}||$

Problem (4)(Global)

(i) Using first derivatives

α	n	P_1	WP_1	QP_1	A_1	QA_1
0.5	5					
	10					
	15				10.2359	6.4816
	20				4.5575	2.9074
	25				3.3731	2.1638
	30				2.8578	1.8431

(ii) Using second derivatives

α	n	P_2	A_2
0.5	5		
	10	12.7048	2.0017
	15	3.4532	1.6471
	20	2.7255	1.5484
	25	2.3688	1.5055
	30	2.1238	1.4827

TABLE (2.11) Bounds on $||(G - T)^{-1}||$

Problem (1) (Piecewise)

(i) Using first derivatives

α	n	P_1	WP_1	QP_1	A_1	QA_1
0.5	5	18.2766	8.1982	2.1942	1.5881	0.8271
	10	4.3168	1.7537	1.5698	1.4694	0.7656
	15	3.4491	1.3439	1.2146	1.4337	0.7473
	20	3.1357	1.1941	1.0838	1.4165	0.7383
	25	2.9741	1.1162	1.0163	1.4063	0.7331
	30	2.8756	1.0684	0.9747	1.3996	0.7296
1	5				20.7728	11.8663
	10				3.9562	2.2646
	15	17.8255	9.2949	7.8883	3.1136	1.7851
	20	8.5673	4.3670	3.7197	2.8139	1.6132
	25	6.5362	3.2824	2.8045	2.6602	1.5255
	30	5.6458	2.8055	2.4010	2.5667	1.4718

(ii) Using second derivatives

α	n	P_2	A_2
0.5	5	2.2284	1.3873
	10	1.7784	1.3529
	15	1.7053	1.3467
	20	1.6770	1.3445
	25	1.6623	1.3436
	30	1.6533	1.3430
1	5		3.6147
	10	4.5960	2.3382
	15	3.4460	2.1944
	20	3.1443	2.1482
	25	3.0096	2.1274
	30	2.9435	2.1163

TABLE (2.12) Bounds on $||(G-T)^{-1}||$

Problem (2) (Piecewise)

(i) Using first derivatives

α	n	P_1	WP_1	QP_1	A_1	QA_1
0.5	5	3.9812	1.5474	1.2309	1.4451	0.7203
	10	3.0165	1.1115	0.8957	1.3998	0.6959
	15	2.7902	1.0068	0.8178	1.3851	0.6881
	20	2.6896	0.9598	0.7826	1.3779	0.6842
	25	2.6325	0.9329	0.7629	1.3736	0.6819
	30	2.5958	0.9157	0.7500	1.3708	0.6804
1	5	68.1637	32.6952	20.8309	2.9716	1.4065
	10	5.6576	2.5781	1.6503	2.3822	1.1131
	15	4.3324	1.9316	1.2465	2.2311	1.0386
	20	3.8803	1.7099	1.1077	2.1625	1.0043
	25	3.6512	1.5969	1.0379	1.1230	0.9848
	30	3.5134	1.5288	0.9956	2.0975	0.9721
2	5					
	10					
	15	48.9174	26.9986	12.0137	18.0123	6.6479
	20	13.2619	2.2533	3.2326	9.8920	3.6246
	25	9.2276	5.0168	2.2410	7.7752	2.8380
	30	5.5123	2.7383	1.5375	5.2287	2.0418

(ii) Using second derivatives

α	n	P_2	A_2
0.5	5	1.7182	1.3443
	10	1.6517	1.3400
	15	1.6346	1.3393
	20	1.6269	1.3391
	25	1.6625	1.3391
	30	1.6197	1.3390
1	5	3.2325	2.0783
	10	2.7144	2.0045
	15	2.6139	1.9928
	20	2.5730	1.9895
	25	2.5507	1.9882
	30	2.5370	1.9877
2	5		15.1116
	10	2.4460	6.1051
	15	5.1706	5.5103
	20	5.7899	5.3382
	25	5.6121	5.2648
	30	5.5123	5.2287

TABLE (2.13) Bounds on $||(G - T)^{-1}||$

Problem (3) (Piecewise)

(i) Using first derivatives

α	n	P_1	WP_1	QP_1	A_1	QA_1
0.5	5	2.1566	0.5874	0.5735	1.0336	0.5166
	10	2.0994	0.5609	0.5504	1.0332	0.5164
	15	2.0810	0.5523	0.5430	1.0330	0.5163
	20	2.0719	0.5480	0.5393	1.0329	0.5163
	25	2.0665	0.5455	0.5371	1.0329	0.5162
	30	2.0629	0.5438	0.5356	1.0328	0.5162
1	5	2.3396	0.6865	0.6538	1.0725	0.5354
	10	2.2087	0.6267	0.6023	1.0704	0.5344
	15	2.1684	0.6080	0.5864	1.06966	0.5440
	20	2.1487	0.5989	0.5786	1.0693	0.5338
	25	2.1371	0.5934	0.5740	1.0691	0.5337
	30	2.1294	0.5899	0.5709	1.0690	0.5337
2	5	2.8152	0.9344	0.8462	1.1687	0.5806
	10	2.4632	0.7762	0.7127	1.1587	0.5806
	15	2.3650	0.7311	0.6756	1.1554	0.5738
	20	2.3188	0.7097	0.6577	1.1537	0.5729
	25	2.2919	0.6973	0.6475	1.1527	0.5724
	30	2.2743	0.6891	0.6406	1.1521	0.5721

(ii) Using second derivatives

α	n	P_2	A_2
0.5	5	1.0658	1.0323
	10	1.0647	1.0323
	15	1.0644	1.0323
	20	1.0643	1.0323
	25	1.0642	1.0323
	30	1.0641	1.0323
1	5	1.1389	1.0669
	10	1.1341	1.0668
	15	1.1328	1.0668
	20	1.1322	1.0668
	25	1.1319	1.0668
	30	1.1317	1.0668
2	5	1.3104	1.1450
	10	1.2881	1.1439
	15	1.2826	1.1437
	20	1.2801	1.1437
	25	1.2787	1.1436
	30	1.2778	1.1436

TABLE (2.14) Bounds on $||(G - T)^{-1}||$

Problem (4) (Piecewise)

(i) Using first derivatives

α	n	P_1	WP_1	QP_1	A_1	QA_1
0.5	5				5.3685	2.3359
	10	9.4302	4.7493	3.0980	3.5814	1.3053
	15	5.8865	2.9014	1.9020	3.1231	1.1348
	20	4.9590	2.4157	1.5922	2.9349	1.0652
	25	4.5317	2.1913	1.4484	2.8323	1.0271
	30	4.2859	1.0620	1.3648	2.7678	1.0031

(ii) Using second derivatives

α	n	P_2	A_2
0.5	5	8.9983	3.4573
	10	3.8801	2.7253
	15	3.4789	2.6237
	20	3.3442	2.5904
	25	2.2790	2.5756
	30	3.2412	2.5678
1	5		
	10		
	15	17.5602	19.3604
	20	12.2169	13.6171
	25	10.6674	11.9800
	30	9.9884	11.2493

2.6.5. Conclusions

The numerical results show that improved "aposteriori" bounds for the inverse differential operator can indeed be found if we consider the differential equation in the original operator form,

$$(G - T)\ x = y, \quad \text{instead of the transformed one,}$$

$$(I - K)\ x = y.$$

The introduction of the matrix Q_n which is shown to tend to the norm of the inverse differential operator is a main factor of the closeness of these bounds. The improvement is more obvious with the piecewise collocation method than the global case, and within the global case it is more obvious with the extended projection than the projection method. This is clearly due to the involvement of the projection norm (which is not uniformly bounded) in the latter case. However in that case we have got WP_d bounds which use $||G^{-1}\phi_n||\ ||W_n||$, instead of $||\phi_n||\ ||Q_n||$, as alternatives.

Unfortunately these bounds do not make use of the higher differentiability of the coefficients of the differential equation, and so do not obtain corresponding improvements in applicability in those found by Cruickshank & Wright (1978) and Gerrard (1979). Still with all these bounds the applicability is a difficulty and an alternative approach is introduced in the next Chapter to deal with it. If strict bounds are not required useful estimates of $||(G - T)^{-1}||$ are shown to be obtainable in Chapter 4.

CHAPTER THREE

Principal part extension for better applicability

3.1.1. Introduction

We have seen that the major problem with the bounds produced in the previous chapter is that for many practical problems an inordinate amount of work is necessary to produce any strict bound at all. The aim of this chapter is to consider this problem in order to improve applicability of these bounds.

As indicated in Chapter 1, the theory can be applied to general equations of the form

$Dx = (G - T)x = y$, where G is invertable. Clearly different choices of G may be treated. However there are some practical difficulties which would limit them. Firstly the inverse of G needs to be known explicitly. Secondly a procedure for calculating the projection norm or a bound on it needs to be available. Thirdly all the assumptions on section (2.2) should be satisfied.

Perhaps the simplest extension is to define the principle part of the differential operator G by

$$Gx = x^{(m)} + \lambda_{m-1} x^{(m-1)} + - - - - + \lambda_o x ,$$

where the λ's are some parameters to be chosen to give the highest possible applicability with reasonable amount of work. The inverse of this new G (will be called G*) is considered in (3.1.3.).

In section (3.2) the global collocation method is considered with the new projection (ϕ_n^*). A method of calculating $||\phi_n^*||$ is given and it is proved that $||\phi_n^*||$ and $||\phi_n||$ are asymptotically the same if the collocation points are Tchebycheu zeros and this is assumed throughout this section.

The conditions of the theory of Chapter 1, are all shown to be satisfied. Similar results are given for the piecewise collocation method in section (3.3). In (3.4) we consider the problem of choosing the λ_i values.

Experimental results are given for the simple second order case and G* is chosen in its simplest form,

$$G^*x = x'' + \lambda x, \qquad x(\pm 1) = 0.$$

Tables for the bounds of $||G^{*-1}||$, $||\phi_n^*||$ and $||\phi_{np}^*||$ are presented for different values of λ. Cubic splines are introduced to give simple calculation of $||P_{np}^*||$.

At the end of the chapter experiments on the applicability with the test problems introduced in the previous chapter are compared and discussed.

3.1.2. New splitting of the differential operator

The differential equation 2-1 is now put in the new operator form $(G^* - T^*)x = y$

where $$G^*x = x^{(m)} - \sum_{k=0}^{m-1} \lambda_k x^{(k)} \qquad (3\text{-}1)$$

$\lambda_k \neq 0$ for at least one value of k

and $$T^* x = - \sum_{k=0}^{m-1} (P_k - \lambda_k) x^{(k)}, \qquad (3\text{-}2)$$

In order to apply the theory in the previous chapters we need:

(i) To know the inverse of G^* explicitly.

(ii) $G^*X_n = Y_n$, where X_n is the subspace of the approximate solution and $Y_n = \phi_n^* Y$ for some bounded projection ϕ_n^*. i.e. $\phi_n^* G^* x_n = G^* x_n \quad \forall x_n \in X_n$.

(iii) Knowledge about this projection ϕ_n^* so that $||\phi_n^*||$ or a bound on it can be calculated.

(iv) Conditions of theorem (1-7) to be satisfied and a matrix similar to W_n, call it W_n^*, can be defined such that

$$||W_n^*|| \rightarrow ||(I - K^*)^{-1}|| \quad \text{where } K^* = T^* G^{*-1}.$$

3.1.3. Study of the inverse of G^*

$$G^*x = u^* \tag{3.3}$$

is a linear inhomogeneous differential equation with constant coefficients. The solution of this type of equation if it exists may be found analytically. The solution of this equation is equivalent to the existence of the inverse of G^* which can be expressed by

$$(G^{*-1} u^*)(s) = \int_{-1}^{1} g^*(s,t)\, u^*(t)\, dt$$

where $g^*(s,t)$ is the Green's function of (3-3) with the boundary conditions given in (2.2). The operator K^* is defined in a similar way by

$$(K^*u^*)(s) = \int_{-1}^{1} \sum_{k=m-1}^{o} (P(s)-\lambda_k)_k \frac{\partial^{(k)}}{\partial s^{(k)}} g^*(s,t)\, u^*(t)dt .$$

To study the behaviour of G^{*-1} in more detail we consider the following simple second order example,

$$G^*x = x'' + \lambda x = y \qquad x(\pm 1) = 0 .$$

If we solve this equation with the method of variation of constants (for example) we will reach the following solution for negative λ,

$$x(s) = \frac{e^{\sqrt{\lambda}(2+s)} - e^{-\sqrt{\lambda}s}}{3\sqrt{\lambda}\left(e^{\sqrt{\lambda}} - e^{-\sqrt{\lambda}}\right)} \int_{-1}^{1} \sinh\sqrt{\lambda}(t-1)\, y(t)\, dt$$

$$+ \frac{1}{\sqrt{\lambda}} \int_{-1}^{s} \frac{e^{\sqrt{\lambda}(s-t)} - e^{\sqrt{\lambda}(s-t)}}{2}\, y(t)\, dt .$$

The corresponding Green's function can be defined by

$$g^*(s,t) = \begin{cases} \frac{1}{\sqrt{\lambda}} \left[w(s) \sinh\sqrt{\lambda}(t-1) + \sinh\sqrt{\lambda}(s-t) \right] & t \leq s \\ \frac{1}{\sqrt{\lambda}} w(s) \sinh\sqrt{\lambda}(t-1) & t > s \end{cases}$$

where $w(s) = \dfrac{e^{\sqrt{\lambda}(2+s)} - e^{-\sqrt{\lambda}s}}{3\sqrt{\lambda}\left(e^{\sqrt{\lambda}} - e^{-\sqrt{\lambda}}\right)}$.

Then $||G^{*-1}|| \leq \sup_s \int_{-1}^{1} |g^*(s,t)|\, dt$

$$= \sup_s \frac{1}{\lambda} \{w(s)\{\cosh(2\sqrt{\lambda}) - 1\} + 1 - \cosh\sqrt{\lambda}(s+1)\} = \frac{1}{\lambda}\{1 - \frac{2(1 - e^{2\sqrt{\lambda}})}{e^{\sqrt{\lambda}} - e^{3\sqrt{\lambda}}}\} .$$

$$g_s^*(s,t) = \begin{cases} w^*(s)\sinh\sqrt{\lambda}(t-1) + \cosh\sqrt{\lambda}(s-t) & t \leq s \\ w^*(s)\sinh\sqrt{\lambda}(t-1) & t > s \end{cases}$$

where $w^*(s) = (e^{\sqrt{\lambda}(2+s)} + e^{-\sqrt{\lambda}s}) / (e^{3\sqrt{\lambda}} - e^{-\sqrt{\lambda}})$.

Hence $|| D\, G^{*-1} || \leq \sup_s \int_{-1}^{1} |g_s^*(s,t)|\, dt$

$$= \sup_s \frac{1}{\sqrt{\lambda}} \{1 + \{2e^{-\sqrt{\lambda}} - e^{\sqrt{\lambda}(2-s)} - e^{\sqrt{\lambda}s} + e^{\sqrt{\lambda}(2s+1)} + e^{\sqrt{\lambda}(1-2s)} - e^{\sqrt{\lambda}(2+s)} - e^{-\sqrt{\lambda}s} - e^{3\sqrt{\lambda}} - e^{-\sqrt{\lambda}}\} / (e^{3\sqrt{\lambda}} - e^{-\sqrt{\lambda}})\} = \frac{1}{\sqrt{\lambda}}\{1 + \frac{2(e^{-\sqrt{\lambda}} - e^{\sqrt{\lambda}})}{(e^{3\sqrt{\lambda}} - e^{-\sqrt{\lambda}})}\} .$$

In a similar way for positive λ the Green's function can be shown to be

$$g^*(s,t) = \begin{cases} \dfrac{1}{\sqrt{\lambda}\ \sin 2\sqrt{\lambda}} \sin\sqrt{\lambda}(s-1)\ \sin\sqrt{\lambda}\ (t+1) & t < s \\ \dfrac{1}{\sqrt{\lambda}\ \sin 2\sqrt{\lambda}} \sin\sqrt{\lambda}\ (s+1)\ \sin\sqrt{\lambda}\ (t-1) & t \geq s \end{cases}$$

which is singular for $\lambda = \dfrac{k^2\pi^2}{4}$ $\qquad k = 1,2,\ldots$

Hence $\| G^{*-1} \| = \sup_s \int_{-1}^{1} | g^*(s,t) | \, d t$

$$= \sup_s \left\{ \left| \frac{\sin\sqrt{\lambda}\ (s-1)}{\sqrt{\lambda}\ \sin 2\sqrt{\lambda}} \right| \int_{-1}^{s} \left|\sin\sqrt{\lambda}\ (t+1)\right| dt + \left| \frac{\sin\sqrt{\lambda}(s+1)}{\sqrt{\lambda}\ \sin 2\sqrt{\lambda}} \right| \int_{-s}^{1} \left|\sin\sqrt{\lambda}(t-1)\right| dt \right\}$$

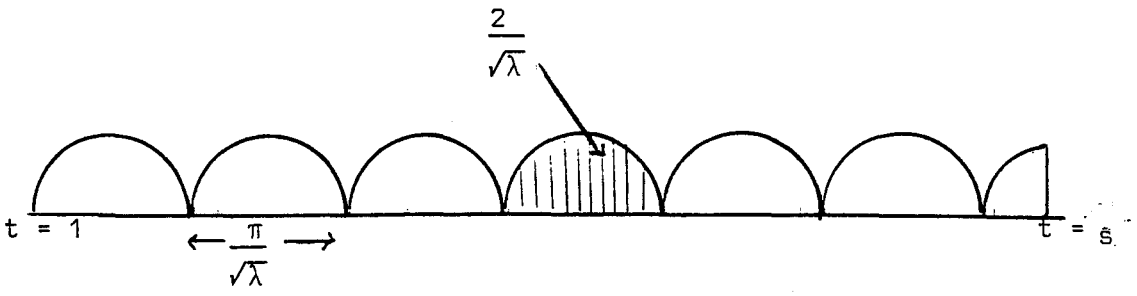
To evaluate $\int_{-1}^{s} \left|\sin\sqrt{\lambda}(t+1)\right| dt$ we count the number of half periods of the function $\sin\sqrt{\lambda}(t+1)$ in the interval $\{-1,s\}$ say n,

$$n = \text{integer part of } \frac{(1+s)\sqrt{\lambda}}{\pi} .$$

Then we multiply n by the area over a half period which is $\int_0^{\frac{\pi}{\sqrt{\lambda}}} \sin\sqrt{\lambda}\ (t+1)\, d t = \dfrac{2}{\sqrt{\lambda}}$

and add $\int_{\frac{-n\pi}{\sqrt{\lambda}}}^{s+1-\frac{n\pi}{\sqrt{\lambda}}} \left| \sin\sqrt{\lambda}\ (t+1)\right| dt = \sin\sqrt{\lambda}\ t = \dfrac{1}{\sqrt{\lambda}} \{1-\cos\sqrt{\lambda}(s+1)\}$

the area of the remaining fraction of period. This integration can be seen in the following diagram.



Similarly $\int_s^1 |\sin \sqrt{\lambda}\,(t-1)|\, dt = \frac{2m}{\sqrt{\lambda}} + \frac{1}{\sqrt{\lambda}} \{1 - \cos\sqrt{\lambda}(1-s)\}$

where m = integer part of $\frac{(1-s)\sqrt{\lambda}}{\pi}$, the number of half periods of the function $\sqrt{\lambda}\,(t-1)$ in the interval $(s,1)$.

That gives $||G^{*-1}|| \leq \sup_s \left|\frac{\sin\sqrt{\lambda}\,(s-1)}{\lambda \sin 2\sqrt{\lambda}}\right| \{2n + 1 - \cos\sqrt{\lambda}\,(s+1)\}$

$$+ \left|\frac{\sin\sqrt{\lambda}\,(s+1)}{\lambda \sin 2\sqrt{\lambda}}\right| \{2m + 1 - \cos\sqrt{\lambda}(1-s)\},$$

We can show in a similar way

$$||D\,G^{*-1}|| < \sup_s \left|\frac{\cos\sqrt{\lambda}(s-1)}{\sqrt{\lambda}\sin 2\sqrt{\lambda}}\right| \{2m + 1 - \cos\sqrt{\lambda}\,(s+1)\}$$

$$+ \left|\frac{\cos\sqrt{\lambda}\,(s+1)}{\sqrt{\lambda}\sin 2\sqrt{\lambda}}\right| \{2m + 1 - \cos\sqrt{\lambda}\,(s-1)\},$$

Tables (3.1) and (3.2) give the values of $g^*_{max} = \max_{s,t} g^*(s,t)$, BG^{*-1} (the bound obtained for $||G^{*-1}||$) and BDG^{*-1} (the bound obtained for $||DG^{*-1}||$) respectively for different values of λ. We note that:

(i) For small $|\lambda|$, the results for G^* are exactly like the usual

operator G(g_{max} = $||G^{-1}||$ = ½, $||D\ G^{-1}||$ = 1) which is what is expected by definition of G* when $|\lambda|$ small G* $\simeq$ G. When $|\lambda|$ is very small we observe different results and B G* in table 1 seems to be diverging. These odd results were found to be due to rounding errors in their evaluation which involve division by λ.

(ii) In table (3.1)(negative λ), we observe:

When $|\lambda|$ becomes larger g^*_{max}, B G^{*-2} and B D G^{*-1} become smaller, obviously due to the division by λ or $\sqrt{\lambda}$. This is a first indication for better applicability with this new operator when the chosen λ is -ve, since

$$g^*_{max} \leq g_{max},\ B\ G^{*-1} \leq B\ G^{-1} \text{ and } B\ D\ G^{*-1} \leq B\ D\ G^{-1} \quad \forall\ \lambda\ .$$

(iii) In table (3.2) (positive λ) we observe:

For λ = 2.467401 and 9.869604 we observe very large values and that is because of singularity at the points $\frac{\pi}{4}$ and π. This is the main problem with this operator i.e. to have singularity (G^{*-1} undefined). The only way to overcome this problem is to note the points of singularity and to avoid them beforehand.

Away from singularity we observe for large λ $g^*_{max} \to \frac{1}{\sqrt{\lambda}}$ but B D G^{*-1} is not affected. This obviously follows from the expressions of these terms.

Unfortunately, the bounds here are not monotically decreasing as above, but they oscillate with large limits near the eigenvalues. This behaviour should be taken into account when we consider the choice of λ later.

Experiments on G*

TABLE (3.1) (Negative λ)

λ	g^*_{max}	$B\ G^{*-1}$	$B\ D\ G^{*-1}$
-1×10^{-16}	0.5	2.08×10^{7}	1.07553
-1×10^{-12}	0.5	41.63335	1.0004
-1×10^{-7}	0.5	0.5	1.0
-0.0001	0.49998	0.49998	0.99997
-0.01	0.49834	0.4979	0.9967
-0.5	0.43053	0.41344	0.86106
-1.0	0.3808	0.35195	0.7616
-5.0	0.21856	0.15773	0.43711
-10.0	0.15755	0.09155	0.3151
-100	0.05	0.01	0.1

TABLE (3.2) (Positive λ)

λ	g^*_{max}	$B\ G^{*-1}$	$B\ D\ G^{*-1}$
1×10^{-16}	0.5	0.5829	1.0408
1×10^{-12}	0.5	0.5	1.0
1×10^{-7}	0.5	0.5	1.0
0.0001	0.5	0.5	1.0
0.01	0.5017	0.5021	1.0033
0.5	0.6042	0.6307	1.2084
1.0	0.7787	0.8508	1.5574
2.467401	1×10^{8}	1.2×10^{7}	1.99×10^{7}
9.869604	2.5×10^{7}	3.4×10^{7}	1.4×10^{7}
100	0.1092	0.1219	1.5032
10^{4}	0.0114	0.0136	1.4487
10^{12}	1.5×10^{-6}	1.9×10^{-6}	1.9327

3.2. The Global Collocation and the new projection

X_n is taken as before to be the space of polynomials of degree n+m-1 satisfying the boundary conditions. We introduce Y_n^* to be the space of polynomials generated by $\{ G^* \Psi_r \}_{r=1,--,n}$ where $\{ \Psi_r \}_{r=1,--,n}$ is a basis of X_n. Now if G^{*-1} is well defined, G^* establishes a bijection between X_n and $Y_n^* = \phi_n^* Y$. So that $\phi_n^* G^* x_n = G^* x_n$ for all x_n in X_n and ϕ_n^* defines a linear projection from Y to Y_n^*.

3.2.1. The projection norm

In the usual analysis we show that the approximate solution satisfies $\phi_n (G - T) x_n = \phi_n y$, where ϕ_n is the polynomial interpolation projection based on the n collocation points $\{\xi_j\}_{j=1}^n$ which are chosen throughout this section to be zeros of Tchebychev polynomial.

With this new operator splitting $G^* - T^*$ the subspaces $G X_n$, $G^* X_n$ will in general be different. However to analyse this new interpolation projection it is convenient to consider

$G^* = G - T_\lambda$, with corresponding projection ϕ_n^*. (T_λ is some special form of T defined by $T_\lambda x = \sum_{i=m-1}^{o} \lambda_i x^{(i)}$. In this case

$(I - \phi_n T_\lambda G^{-1}) U_n = \phi_n y$ is equivalent to

$$I U_n^* = \phi_n^* y \qquad (3.4)$$

and $$x_n = G^{*-1} \phi_n^* y, \qquad (3.5)$$

N.B. x_n is not affected by the choice of splitting.

Now we want to calculate the projection norm for ϕ_n^*.

$(\phi_n^* y)(t)$ can be expressed as

$$\sum_{j=1}^{n} \ell_j^*(t)\, y(\xi_j)$$

where $\ell_j^*(t)$ are the generalized Lagrange interpolating coefficients, then

$$\| \phi_n^* \| = \sup_t \sum_{j=1}^{n} | \ell_j^*(t) | \text{ as usual.}$$

By definition $\ell_j^*(\xi_k) = \delta_{jk}$ and may be expressed as

$\ell_j^*(t) = \phi_n^* J \underline{e}_j$ where $\underline{e}_j$ is the jth unit vector and J is the extension operator defined in Chapter 2.

Now returning to equation (3.4) we see that

$U_n^* = \phi_n^* y$, and so $\ell_j^*(t)$ can be found by solving

$$G^* x^{(j)} = J \underline{e}_j \tag{3.6}$$

by collocation using the points $\{\xi_j\}$ $j=1,\cdots,n-1$.

If the numerical solution of (3.6) is $x_n^{(j)}(t)$ then clearly

$$\ell_j^*(t) = U^*{}_n^{(j)}(t) = G^* x_n^{(j)}(t). \tag{3.7}$$

This may well be the simplest way of calculating $\ell_j^*(t)$ and hence $\|\phi_n^*\|$.

For illustration if we define G^* by

$$G^* x = x'' + \lambda x \qquad x(\pm 1) = 0 \quad ,$$

then we want to solve

$G^* x^{(j)} = J \underline{e}_j$ $j=1, \ldots, n$ by collocation. If we take Tchebychev polynomials $\{T_i(t)\}_{i=1}^{n}$ as basis for the solution,

then $x_n^{(j)}(t) = \sum_{i=1}^{n} C_i^j \, T_i(t)$.

If we take $\{\xi_j\}_{j=1}^{n}$ as collocation points then the collocation method gives the following algebraic equations

$$\underline{A}\ \underline{C}^{(j)} = \underline{e}_j \quad , \qquad j=1,---,n$$

$\underline{C}^j = \{C_1^j, C_2^j, ----, C_n^j\}^T$ and A is n x n matrix corresponds to G*. The solution for the $\underline{C}$ will be simply the jth column of the inverse of the collocation matrix A. If $\bar{a}_{ij}$ is the ijth element of A^{-1} then

$$x_n^{(j)}(t) = \sum_{i=1}^{n} \bar{a}_{ij} \, T_i(t) \ , \qquad \text{and by (3.7)}$$

$$\ell_j^*(t) = G^* x_n^{(j)}(t) = \sum_{i=1}^{n} a_{ij} \, (T_i''(t) + \lambda \, T_i(t)) .$$

That gives $||\phi_n^*|| = \sup_t \left| \sum_{j=1}^{n} \left| \sum_{i=1}^{n} a_{ij} \, (T_i''(t) + \lambda \, T_i(t)) \right| \right.$.

Tables (3.3,4) describes the behaviour of $||\phi_n^*||$ for different values of λ. The most striking feature is that for every value of λ it looks as if $||\phi_n^* \quad - \quad \phi_n|| \to 0$. The theory behind this behaviour will be considered in the next section.

We also observe that the effect of singularity (for example $\lambda = 9.869604$) is less than one would expect. This may be due to the approximations involved and the cancellations occurring in multiplying back by G* in (3.7).

TABLE (3.3) $||\phi_n^*||$

(a) Negative λ

n \ λ	-1×10^{-5}	-0.5	-10	-50	0
5	1.98885	1.97824	1.82248	1.88896	1.98885
10	2.42883	2.42593	2.37461	2.21149	2.42883
15	2.68671	2.68539	2.66094	2.57132	2.68671
20	2.86977	2.86902	2.85486	2.79987	2.86977
25	3.01179	3.01130	3.00210	2.96532	3.01179
30	3.12784	3.12750	3.12104	3.09484	3.12784

Table (3.4)

(b) Positive λ

n \ λ	1×10^{-5}	0.5	9.869604	50	0
5	1.98885	1.9998	5.42223	1.99848	1.98885
10	2.42883	2.43175	1.44×10^{3}	2.92604	2.42883
15	2.68671	2.68805	2.71379	2.84581	2.68671
20	2.86977	2.87053	2.88501	2.95305	2.86977
25	3.01179	3.07228	3.02157	3.06370	3.01179
30	3.12784	3.12818	3.13465	3.16345	3.12784

3.2.2. Relation between ϕ_n and ϕ_n^*

Recall equation (3.5)

$$x_n = G^{*-1}\, \phi_n^*\, y \ .$$

In the usual analysis

$$x_n = (G - \phi_n T_\lambda)^{-1}_{Y_n}\ \phi_n\, y \ ,$$

therefore $$G^{*-1}\, \phi_n^*\, y = (G - \phi_n T_\lambda)^{-1}_{Y_n}\, \phi_n\, y \ .$$

Multiply on the left by G^* to get

$$\phi_n^*\, y = G^* (G - \phi_n T_\lambda)^{-1}_{Y_n}\ \phi_n\, y \ .$$

Hence

$$\phi_n^* = G^* (G - \phi_n T_\lambda)^{-1}_{Y_n}\ \phi_n$$

$$= (G - T_\lambda)(G - \phi_n T_\lambda)^{-1}_{Y_n}\ \phi_n \qquad (3.8)$$

$$= (I - T_\lambda G^{-1})\, G\, G^{-1} (I - \phi_n T_\lambda G^{-1})^{-1}_{Y_n}\ \phi_n$$

$$= (I - T_\lambda G^{-1})(I - \phi_n T_\lambda G^{-1})^{-1}_{Y_n}\ \phi_n \ . \qquad (3.9)$$

The following identity may be called several times,

$$\phi_n (I - T_\lambda G^{-1} \phi_n)^{-1} = (I - \phi_n T_\lambda G^{-1})^{-1}_{Y_n}\ \phi_n \ , \qquad (3.10)$$

for

$$(I - \phi_n T_\lambda G^{-1})^{-1}_{Y_n}\ \phi_n = (I + \phi_n T_\lambda G^{-1} (I - \phi_n T_\lambda G^{-1})^{-1}_{Y_n})\, \phi_n$$

$$= \phi_n (I + T_\lambda G^{-1} (I - \phi_n T_\lambda G^{-1})^{-1}_{Y_n}\ \phi_n) \ .$$

But from (1.28) $(I - T_\lambda G^{-1} \phi_n)^{-1} = (I + T_\lambda G^{-1} (I - \phi_n T_\lambda G^{-1})^{-1}_{Y_n})$,

therefore $(I - \phi_n T_\lambda G^{-1})^{-1}_{Y_n}\ \phi_n = \phi_n (I - T_\lambda G^{-1} \phi_n)^{-1}$.

Theorem 3.1

$||\phi_n^*||$ and $||\phi_n||$ are asymptotically the same.

Proof

$$\phi_n^* - \phi_n = (I - T_\lambda G^{-1})(I - \phi_n T_\lambda G^{-1})^{-1} Y_n \phi_n - \phi_n \quad \text{by (3.9)}$$

$$= (I - T_\lambda G^{-1}) \phi_n (I - T_\lambda G^{-1} \phi_n)^{-1} - \phi_n \quad \text{by (3.10)}$$

$$= \phi_n (I - T_\lambda G^{-1} \phi_n)^{-1} - T_\lambda G^{-1} \phi_n (I - T_\lambda G^{-1} \phi_n)^{-1} - \phi_n$$

$$= \phi_n (I - T_\lambda G^{-1} \phi_n)^{-1} - (I - T_\lambda G^{-1} \phi_n)^{-1} + I - \phi_n$$

$$= (I - \phi_n)(I - (I - T_\lambda G^{-1} \phi_n)^{-1})$$

$$= (\phi_n - I)(T_\lambda G^{-1} \phi_n)(I - T_\lambda G^{-1} \phi_n)^{-1} ,$$

or $$||\phi_n^* - \phi_n|| \leq ||(\phi_n - I) T_\lambda G^{-1}|| ||\phi_n|| \, ||(I - T_\lambda G^{-1} \phi_n)^{-1}|| .$$

By lemma (2.2) $||(\phi_n - I) T_\lambda G^{-1}|| ||\phi_n|| \to 0$ for Tchebycheu zeros, and from (1.30) $||(I - T_\lambda G^{-1} \phi_n)^{-1}||$ is uniformly bounded.

Therefore $||\phi_n^* - \phi_n|| \to 0$ q.e.d.

Corollary 3.1

If $\overline{W}_n$ denotes the W_n matrix of the operator G^*, then

$$||\phi_n^*|| \leq (||\phi_n|| + ||T_\lambda G^{-1} \phi_n||)(||\overline{W}_n|| \, ||T_\lambda G^{-1} \phi|| + 1) .$$

Proof

From (3.9) and (3.10) $\phi_n^* = (I - T_\lambda G^{-1}) \phi_n (I - T_\lambda G^{-1} \phi_n)^{-1}$,

therefore

$$||\phi_n^*|| \leq (||\phi_n|| + ||T_\lambda G^{-1} \phi_n||) ||(I - T_\lambda G^{-1} \phi_n)^{-1}|| .$$

From (1.30) $||(I-T_\lambda G^{-1}\phi_n)^{-1}|| \leq (||\overline{W}_n||\,||T_\lambda G^{-1}\phi_n|| + 1)$. (3.11)

Therefore $||\phi_n^*|| \leq (||\phi_n|| + ||T_\lambda G^{-1}\phi_n||)(||\overline{W}_n||\,||T_\lambda G^{-1}\phi_n|| + 1)$.

The significance of this theorem and corollary is that they ensure $||\phi_n^*||$ will behave exactly like $||\phi_n||$ for sufficiently large n and is not going to be very much worse for small n.

3.2.3. The satisfaction of the conditions of Chapter 1.

Now it remains to show that the conditions required by the projection and the extended projection methods are satisfied. Also we introduce a new matrix W_n^*, which is the W matrix for this new operator splitting, and we show that it has got indeed the same asymptotic properties shown for W_n in the previous chapter. i.e.

$$||W_n^*|| \rightarrow ||(I - K^*)|| \quad K^* = T^*G^{*-1}.$$

Lemma (3.1) K^* is compact.

Proof

The compactness of integral operators of the forms K^* is given for example by Kolomogorov and Fomin (1957).

Lemma (3.2) The sequence $K^*\phi_n^*$ is uniformly bounded.

Proof $K^* = T^* G^{*-1} = T^* G^{-1} (I - T_\lambda G^{-1})^{-1}$.

(3.9) gives $\phi_n^* = (I - T_\lambda G^{-1})(I - \phi_n T_\lambda G^{-1})^{-1}_{Y_n} \phi_n$,

Therefore $K^* \phi_n^* = T^*G^{-1}(I-T_\lambda G^{-1})^{-1}\,(I-T_\lambda G^{-1})\,(I-\phi_n T_\lambda G^{-1})^{-1}_{Y_n}\,\phi_n$

$$= T^*G^{-1}\,(I - \phi_n\, T_\lambda G^{-1})^{-1}_{Y_n}\,\phi_n \qquad (3.12)$$

$$= T^*G^{-1}\,\phi_n\,(I - T_\lambda G^{-1}\phi_n)^{-1} \quad \text{by (3.10)} .$$

therefore $||K^*\,\phi_n^*|| \leq || T^*G^{-1}\phi_n ||\; ||\,(I - T_\lambda G^{-1}\phi_n)^{-1}\,||$.

Since $T^*G^{-1}\phi_n \equiv TG^{-1}\phi_n|_{T=T^*}$ i.e. some form of $K\phi_n$,

Then $||T^*G^{-1}\phi_n||$ is uniformly bounded by Lemma (2.1).

But $||(I - T_\lambda G^{-1}\,\phi_n)^{-1}||$ is uniformly bounded by (3.11) ,

therefore $||K^*\,\phi_n^*||$ is uniformly bounded .

Lemma (3.3)

The sequence $\{K^*\phi_n^*\}$ is collectively compact.

Proof

To prove $K^*\phi_n^*$ is collectively compact we require the set $\mathcal{K}^*\,U = \{K^*\phi_n^*\, y : n \in N,\ Y \in U\}$ to be relatively compact, where U is the unit ball and N is the set of positive integers. The result is achieved by means of the Arzela-Ascoli theorem, given for example by Kantorovich and Akilov (1964) by proving equicontinuity and uniform boundedness of $\mathcal{K}^*U$.

In Lemma (3.2) it was shown that $K^*\phi_n^*$ is uniformly bounded and thus it remains to show that the equicontinuity condition is satisfied. With $-1 \underline{\leq} s_1 \underline{\leq} s_2 \underline{\leq} 1$ and $u \in U$ (3.12) gives

$|(K^*\phi_n^* y)(s_1) - (K^*\phi_n^* y)(s_2)| = |(T^*G^{-1}U_n)(s_1) - (T^*G^{-1}U_n)(s_2)|$, where

$U_n = (I-\phi_n T_\lambda G^{-1})^{-1}\phi_n y$. Let $K^{**} = T^* G^{-1}$ have kernel $K^{**}(s,t)$. Then using Cauchy's inequality as in Cruickshanck (1974, page 103),

$$|(K^* \phi_n^* y)(s_1) - (K^*\phi_n^* y)(s_2)| \leq \{\int_{-1}^{1} \frac{1}{\rho(t)} \{K^{**}(s_1,t) - K^{**}(s_2,t)\}^2 dt\}^{\frac{1}{2}}$$

$$\{\int_{-1}^{1} \rho(t)\, u_n^2(t)\, dt\}^{\frac{1}{2}}.$$

First term clearly tends to zero as $s_1 \to s_2$ independently of y.

Second term is uniformly bounded since

$$\int \rho(t)\, U_n^2(t)\, dt = \Sigma\, \overline{w}_i \{U_n(\xi_i)\}^2 \leq ||U_n||_\infty \Sigma\, \overline{w}_i \quad \text{where}$$

$\underline{U}_n = \overline{W}_n \underline{y}$ and

$$||\underline{U}_n|| \leq ||\overline{W}_n|| \quad ||\underline{y}|| \leq ||\overline{W}_n|| \text{ as } ||\underline{y}|| \leq 1.$$

Lemma (3.4)

$$K^* \phi_n^* \to K^* \text{ as } n \to \infty.$$

<u>Proof</u> $(K^*\phi_n^* - K^*) = (K^*\phi_n^* + K^*\phi_n - K^*\phi_n - K^*)$.

By lemma (2.1)

$$K^* \phi_n \to K^*$$

i.e. $||K^*(\phi_n - I)\, U|| \to 0$ for every $U \in Y$. Then

$$||K^*\phi_n^* U - K^* U|| \leq ||K^*(\phi_n^* - I)\, U|| + ||K^*||\ ||\phi_n^* - \phi_n||\ ||U|| \to 0$$

since $||\phi_n - \phi_n|| \to 0$ by theorem (3.1) and $||K^*||$, $||U||$ are bounded.

Lemma (3.5)

$$\| (I - \phi_n^*) \overset{d}{K^*} \| \to 0 \quad \text{as } n \to \infty$$ provided that the dth derivatives of the coefficients of the differential equation are bounded.

Proof

$$\| (I - \phi_n^*) \overset{d}{K^*} \| = \| I - \phi_n^* \overset{d}{K^*} + \phi_n \overset{d}{K^*} - \phi_n \overset{d}{K^*} \|$$

$$\leq \| (I - \phi_n) \overset{d}{K^*} \| + \| \overset{d}{K^*} \| \| \phi_n - \phi_n^* \| \to 0,$$

since the first part tends to zero by lemma (2.2) and the second part tends to zero by theorem (3.1).

Lemma (3.1), (3.3) and (3.4) show that (1) K^* is compact (2) $K^* \phi_n^* \to K^*$ and (3) $K^*\phi^*$ is collectively compact which are all the conditions required by the extended projection method. Lemma (3.5) also ensures the applicability of the usual projection method. We show next that the result of theorem (1.8) can be in fact applied with this new splitting.

Theorem (3.2)

$$\| W_n^* \| \to \| (I - K^*)^{-1} \| \quad \text{as } n \to \infty .$$

Proof

The proof goes as follows

(1) We show $(I - K^*)^{-1} = (I - T_\lambda G^{-1}) (I - K)^{-1}$

(2) We show $(I - \phi_n^* K^*)^{-1} \phi_n^* = (I - T_\lambda G^{-1}) (I - \phi_n K)^{-1} \phi_n$

and hence $W_n^* = H_n (I - T_\lambda G^{-1}) (I - \phi_n K)^{-1} \phi_n J_n$.

Then by theorem (2.1) and its corollary

$$\left| \; ||W_n^*|| - ||(I - K^*)^{-1}|| \; \right| \leq \left| \; ||H_n(I-\phi_n K)^{-1}\phi_n|| - ||(I-K)^{-1}|| \; \right|$$

$$+ \sum_{i=o}^{m-1} |\lambda_i| \left| \; ||H_n D^i G^{-1}(I-\phi_n K)^{-1}\phi_n J_n|| - ||D^i G^{-1}(I-K)^{-1}|| \; \right|$$

$$\to \quad o \qquad \text{as} \quad n \to \infty .$$

(1) By definition $K^* = T^* G^{-1} = T^* G^{-1}(I - T_\lambda G^{-1})^{-1}$.

But $T^* = T - T_\lambda$,

therefore $K^* = (T - T_\lambda)\, G^{-1}\, (I - T_\lambda G^{-1})^{-1}$

$$= (K - T_\lambda G^{-1})\, (I - T_\lambda G^{-1})^{-1} . \qquad (3.13)$$

Hence $(I - K^*)^{-1} = (I - (K - T_\lambda G^{-1})\, (I - T_\lambda G^{-1}))^{-1}$

$$= (I - K(I - T_\lambda G^{-1})^{-1} + T_\lambda G^{-1}\, (I - T_\lambda G^{-1})^{-1} \;)^{-1}$$

$$= (I - K\,(I - T_\lambda G^{-1})^{-1} + (I - T_\lambda G^{-1})^{-1} - I)^{-1}$$

$$= ((I - K)\, (I - T_\lambda G^{-1})^{-1})^{-1}$$

$$= (I - T_\lambda G^{-1})\, (I - K)^{-1} .$$

(2) We use (3.13) and (3.9)

$$(I - \phi_n^* K^*)^{-1}\, \phi_n^* = (I - \phi_n^*\, (K - T_\lambda G^{-1})\, (I - T_\lambda G^{-1})^{-1})^{-1}\, \phi_n^*$$

$$= (I - \phi_n^*\, (K(I - T_\lambda G^{-1})^{-1} - T_\lambda G^{-1}(I - T_\lambda G^{-1})^{-1})^{-1}\, \phi_n^*$$

$$= (I - \phi_n^* (K(I - T_\lambda G^{-1})^{-1} - (I - T_\lambda G^{-1})^{-1} + I) \;)^{-1}\, \phi_n^*$$

$$= (I + \phi_n^*\, (I - K)\, (I - T_\lambda G^{-1})^{-1} - \phi_n^*)^{-1}\, \phi_n^*$$

$$= (I + \phi_n^*\, (I - K)\, (I - T_\lambda G^{-1})^{-1} - \phi_n^*)^{-1}\, (I - T_\lambda G^{-1})$$

$$(I - \phi_n\, T_\lambda G^{-1})^{-1}_{Y_n}\, \phi_n$$

$$= \{(I-\phi_n T_\lambda G^{-1})Y_n\ (I - T_\lambda G^{-1})^{-1}(I+(I-T_\lambda G^{-1})\ (I-\phi_n T_\lambda G^{-1})^{-1}\phi_n\ (I-K)$$
$$(I - T_\lambda G^{-1})^{-1} - (I - T_\lambda G^{-1})\ (I -\phi_n T_\lambda G^{-1})^{-1}_{Y_n}\ \phi_n)\}^{-1}\ \phi_n$$

$$= \{(I-\phi_n T_\lambda G^{-1})Y_n\ (I-T_\lambda G^{-1})^{-1} + (I-\phi_n T_\lambda G^{-1})^{-1}_{Y_n}\ \phi_n\ (I-K)\ (I-T_\lambda G^{-1})^{-1}$$
$$- (I - \phi_n\ T_\lambda G^{-1})^{-1}\ \phi_n\}^{-1}\ \phi_n$$

$$= \{(I-\phi_n T_\lambda G^{-1})Y_n\ (I-T_\lambda G^{-1})^{-1} +\phi_n\ (I-k)\ (I-T_\lambda G^{-1})^{-1} - \phi_n\}^{-1}\ \phi_n$$

$$= \{(I-\phi_n T_\lambda G^{-1} + \phi_n - \phi_n K)\ (I - T_\lambda G^{-1})^{-1} - \phi_n\}^{-1}\ \phi_n$$

$$= \{(I-\phi_n K)\ (I-T_\lambda G^{-1})^{-1} + \phi_n\ (I-T_\lambda G^{-1})\ (I-T_\lambda G^{-1})^{-1} - \phi_n\}^{-1}\ \phi_n$$

$$= (I - T_\lambda G^{-1})\ (I - \phi_n K)^{-1}\ \phi_n \qquad \text{q.e.d.}$$

3.3. Piecewise Collocation

The space X_{np}, the collocation points $\{\xi_{ji}\}_{\substack{i=1,---,n\\ j=1,---,p}}$ and the projection ρ_{np} are defined exactly as in (2.4). The space Y^*_{np} is defined to be the space of piecewise polynomials generated by $G^*\Psi_i$ $i = 1,---,np$ where $\{\Psi_i\}_{i=1,---,np}$ is a basis of X_{np}. If G^{*-1} exists then clearly G^* will establish a bijection between X_{np} and Y^*_{np}.

Let P^*_{np} be a linear projection from Y_{np} to Y^*_{np}, then if we follow the analysis in Section (3.2.1) we reach corresponding to (3.4) and (3.5),

$$I\ U^*_{np} = P^*_{np}\ y \quad (3.14)$$

$$X_{np} = G^{*-1}\ P^*_{np}\ y \quad (3.15)\ .$$

The projection P^*_{np} is defined by

$$(P^*_{np}\ y)\ (t) = \sum_{i=1}^{np} L^*_i(t)\ \ y(\xi_i)$$

where L^*_i is the unique polynomial such that $L^*_i(\xi_j) = \delta_{ij}$ and $\{\xi_j\}_{j=1,\ldots,np}$ are given by $\xi_{(k-1)p+j} = \xi_{jk}$ $k=1,\ldots,n$ $j=1,\ldots,p$.

We note here that P^*_{np} cannot be defined in each interval (t_{i-1}, t_i) separately as is the case for P_{np}. That can be seen easily from equation (3.14) $U^*_{np} = G\,x_{np} + \sum_{i=1}^{m-1} \lambda_i x^{(i)}_{np} = P^*_{np}\,y$ where we have assumed continuity of $x^{(i)}_{i=1,\ldots,m-1}$. But for the P_{np} projection

$$U_{np} = G\,x_{np} = P_{np}\,y$$ and continuity of $G\,x$ is not required.

To calculate $L^*_i(t)$ we solve,

$$G^*\,x^{(i)} = J_{np}\,\underline{e}_i \qquad i=1,\ldots,np\ ,$$

by the piecewise collocation method using the points ξ_{ji} $i=1,\ldots,n$ $j=1,\ldots,p$.

Then $L^*_i(t) = U^{*(i)}_{np}(t) = G^*\,x^{(i)}_{np}(t)$.

$$||P^*_{np}|| = \max_t \sum_{i=1}^{np} |L^*_i(t)| \ .$$

For illustration let G^* be defined by

$$G^*\,x = x'' + \lambda\,x \qquad x(\pm 1) = 0 \ .$$

We want to solve

$$G^*\,x^{(j)} = J_{np}\,\underline{e}_j \qquad j=1,\ldots,np$$ by piecewise collocation.

If we take $p=2$ the cubic splines U_i, V_i $l=1,\ldots,n$ defined below are suggested as a basis for the solution x_{np}. If $h_j = t_j - t_{j-1}$ then

$$U_j(t) = \begin{cases} -(t-t_{j-1})^2 \dfrac{(t-\frac{1}{2}(3t_j - t_{j-1}))}{\frac{1}{2} h^3_j} & \text{in } (t_{j-1}, t_j) , \\ (t-t_{j+1})^2 \dfrac{(t-\frac{1}{2}(3t_j - t_{j+1}))}{\frac{1}{2} h^3_{j+1}} & \text{in } (t_j, t_{j+1}) , \\ 0 \quad \text{elsewhere} . \end{cases}$$

$$V_j(t) = \begin{cases} (t_j - t) \dfrac{(t - t_{j-1})^2}{h^2_j} & \text{in } (t_{j-1}, t_j) , \\ (t_j - t) \dfrac{(t - t_{j+1})^2}{h^2_{j+1}} & \text{in } (t_j, t_{j+1}) , \\ 0 \text{ elsewhere} . \end{cases}$$

We notice that

$$U_j(t_{j-1}) = U_j(t_{j+1}) = 0 \quad \text{and} \quad U_j(t_j -) = U_j(t_j +) = 1 ,$$

$$U'_j(t_{j-1}) = U'_j(t_{j+1}) = 0 \quad \text{and} \quad U'_j(t_j -) = U'_j(t_j +) = 0 ,$$

$$V_j(t_{j-1}) = V_j(t_{j+1}) = 0 \quad \text{and} \quad V_j(t_j -) = V_j(t_j +) = 0 ,$$

$$V'_j(t_{j-1}) = V'_j(t_{j+1}) = 0 \quad \text{and} \quad V'_j(t_j -) = V'_j(t_j +) = 1 ,$$

which shows that U_i and V_i satisfy the boundary and the continuity conditions.

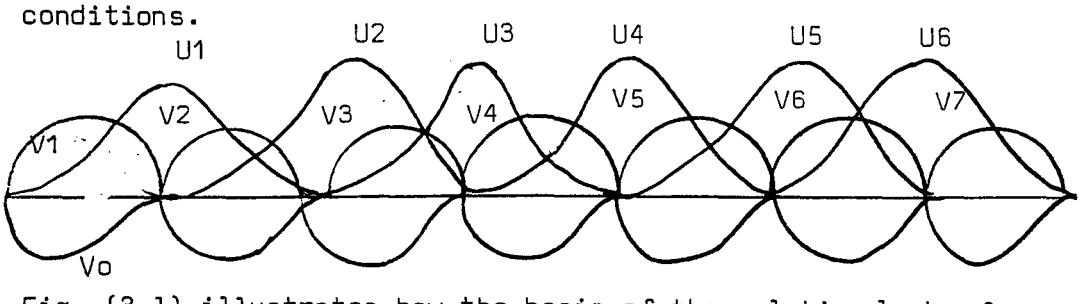


Fig. (3.1) illustrates how the basis of the solution looks for $n = 7$.

looks

If we arrange the U_i and V_i in the set $\{\Psi_i\}_{i=1}^{2n}$ such that

$$\Psi_{2i+1} = V_i \qquad i=0,\ldots, n-1$$

$$\Psi_{2i} = U_i \qquad i=1,\ldots, n-1$$

and $\Psi_{2n} = V_n$, then the basis functions for interpolation $\{\phi_i\}_{i=1}^{2n}$ are defined by

$$\phi_i = G^* \Psi_i \qquad i=1,\ldots, 2n .$$

If ξ_{1j}, ξ_{2j} are the two collocation points in the jth interval $(t_{j-1}, t_j)_{j=1,\ldots,n}$, then the set of all collocation points $\xi_{i\ i=1,\ldots,2n}$ is given by

$$\xi_{2i+1} = \xi_{1j} \qquad i=0,\ldots, n-1$$

$$\xi_{2i} = \xi_{2i} \qquad i=1, \ldots, n.$$

If L_j^* denotes the unique polynomial such that $L_j^*(\xi_i) = \delta_{ij}$ then L_j^* can be expressed uniquely by

$$L_j^*(t) \quad \sum_{k=1}^{2n} \alpha_k^{(j)} \phi_k(t) \qquad j=1,\ldots, 2n .$$

To find $L_j(t)$ we solve for $\alpha_k^{(j)}$ $\begin{matrix} k=1,\ldots,2n \\ j=1,\ldots,2n \end{matrix}$

using $L^*_j(\xi_i) = \delta_{ij}$ $\begin{matrix} i=1,\ldots,2n \\ j=1,\ldots,2n \end{matrix}$.

The interpolation norm is given by

$$\| P^*_{np} \| = \sup_t \sum_{j=1}^{2n} | L^*_j(t) | .$$

We note that by definition of $\phi_k(t)$ $k=1,\ldots,2n$

$$\phi_1(\xi_j) = \begin{cases} c \neq o & j=1,2 \\ o & \text{elsewhere} \end{cases};$$

for $i = 2, \ldots, 2n-1$,

$$\phi_i(\xi_j) = \begin{cases} c \neq o & j=2i-1,\ 2i,\ 2i+1,\ 2i+2 \\ o & \text{elsewhere} \end{cases}$$

and

$$\phi_{2n}(\xi_j) = \begin{cases} c \neq o & j=2n,\ 2n-1 \\ o & \text{elsewhere} \end{cases}$$

Where c is constant.

which makes the matrix of the solution to have the special shape given in Fig. (3.2).

Table (3.5) gives the values of $||P^*_{np}||$ for different values of λ. We notice the same property observed for $||\phi^*_n||$, that is $||P^*_{np}||$ is tending to the same value. This here is 1.414213 which is the value of $||P_{n2}|| = ||\phi_2||$ when the collocation points are Tchebychev zeros. That indicates $||P^*_{np}|| \to ||P_{np}|| = ||\phi_p||$ which can be proved following exactly the arguments of theorems (3.1) and using Lemma (2.5) instead of Lemma (2.2).

Theorem (3.3)

$$||P^*_{np}|| \to ||P_{np}|| = ||\phi_p|| \text{ as } n \to \infty.$$

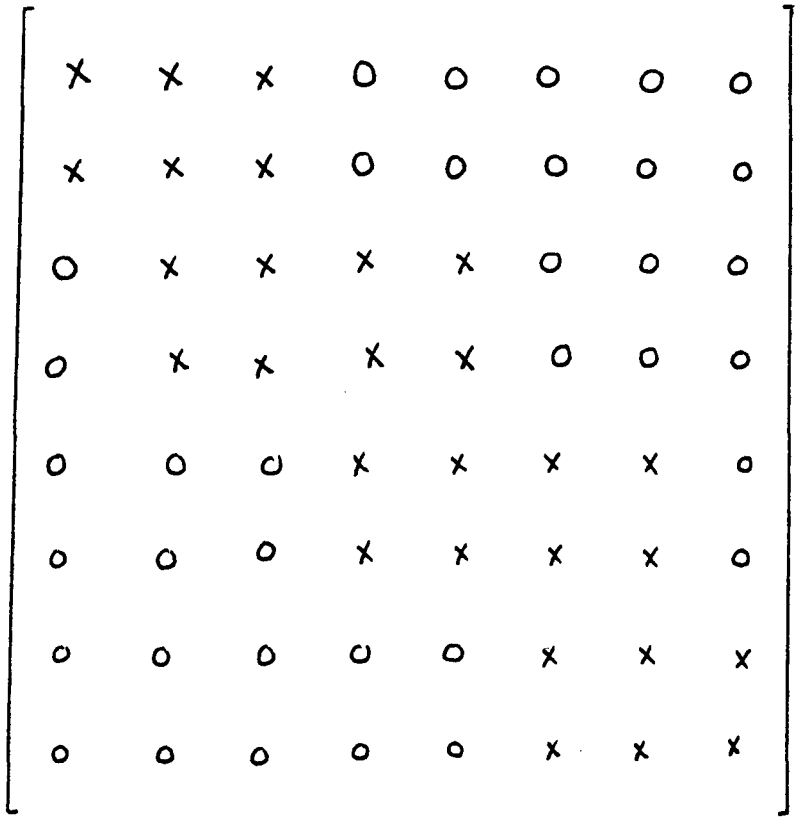
Corollary (3.2)

If $\overline{W}_{np}$ denotes the W_{np} matrix of the operator G^*, then $||P^*_{np}||$ can be bounded by

Fig. (3.2) The solution matrix (splines)

Number of intervals = 4

Number of collocation points = 2



x non zero element

o zero element

This can be considered as a band matrix of order 8 and 2 sub-diagonals and 2 super-diagonal elements in a typical row can be stored as:- (8x5) array , and can be solved for example by the NAG special procedures Fϕ1BMF and Fϕ4AVF.

TABLE (3.5) $||p^*_{np}||$ p=2

(a) Negative λ

n \ λ	-1×10^{-5}	-0.5	-10	-50	-500
5	1.414213	1.412528	1.421329	1.423347	1.858299
10	1.414213	1.413833	1.419330	1.430870	1.555555
15	1.414213	1.414051	1.417218	1.426088	1.421426
20	1.414213	1.414124	1.416134	1.422605	1.430752
25	1.414213	1.414157	1.415541	1.420364	1.438364
30	1.414213	1.414174	1.415181	1.418887	1.431551

(b) Positive λ

n \ λ	1×10^{-5}	0.5	9.869602	50	500
5	1.414213	1.417604	4.800538	3.869048	3.168807
10	1.414213	1.415125	4.798982	2.325582	65.11853
15	1.414213	1.414629	4.833981	1.859074	15.40038
20	1.414213	1.414450	4.920161	1.672735	8.694045
25	1.414213	1.414366	4.991944	1.579949	10.13541
30	1.414213	1.414320	5.029934	1.530478	19.23538

$$||\phi_p|| + ||T_\lambda G^{-1}|| \; ||\phi_p|| \; ||(\bar{W}_{np}|| + ||T_\lambda G^{-1}|| \; ||\phi_p|| + 1).$$

Proof. Follow corollary (3.1) with $||\phi_p|| = ||P_{np}||$.

This theorem and corollory like the global case ensure that $||P^*_{np}||$ is not going to be much worse than $||P_{np}||$ for any value of n.

Lemma (3.6)

The sequence $\{K^*P^*_{np}\}$ is uniformly bounded.

Proof.

The proof follows by the same arguments of Lemma (3.2) with Lemma (2.5), used instead of Lemma (2.1).

Lemma (3.7)

The sequence $\{K^*P^*_{np}\}$ is collectively compact.

Proof.

If we follow the proof of Lemma (3.3) until we get

$$|((K^*P^*_{np})\, y)(s_1) - ((K^*P^*_{np})\, y)(s_2)| = |(T^*G^{-1}U_{np})(s_1) - (T^*G^{-1}U_{np})(s_2)|,$$

where U_{np} is the approximate solution of the highest derivative when the differential operator $G + T_\lambda$ is solved by the piecewise collocation method of section (2.4). Since $T^* G^{-1}$ is some special form of K we use Cauchy inequality and uniform boundedness of $||\bar{W}_{np}||$ as in Lemma (3.3) to show

$$|(T^*G^{-1} U_{np})(s_1) - (T^*G^{-1} U_{np})(s_2)| \to 0$$

when $s_2 \to s_1$.

Lemma (3.8)

$$K^* P^*_{np} \to K^* \quad \text{as } n \to \infty .$$

Proof

ollows by Theorem (3.3) and Lemma (2.5), v), using the same arguments of the global case (Lemma 3.4).

Lemma (3.9)

$$||(I - P^*_{np}) K^* || \to 0 \quad \text{as } n \to \infty .$$

Proof

ollows from Theorem (3.3) and Lemma (2.5), vi) using the same arguments of Lemma (3.5).

Now we have shown all the conditions required for the application of the projection and the extended projection methods are indeed satisfied by the piecewise collocation methods with this new splitting.

Finally it remains to prove that the norm of the matrix W^*_{np}, which is the W_{np} matrix with the new splitting, tends to the norm of $(I-K^*)^{-1}$ as stated in the following theorem.

Theorem (3.4)

$$|| W^*_{np} || \to ||(I - K^*)^{-1} || .$$

Proof

follow the same arguments of theorem (3.2) and use theorem (2.2) and its corollary instead of theorem (2.1) and its corollary.

By this theorem we are in a position to apply the results of Chapter 1 to both the global and piecewise collocation method with this

new splitting. But before that we need to consider some practical ways of choosing the parameters λ_i $i=0,\ldots,m-1$.

3.4. The choice of the parameters λ_i

The main question now is what are the values of these parameters $\{\lambda_i\}_{i=0}^{m-1}$ which will maximize the applicability? If we go back to section (1.6) we see that factors involved in the applicability conditions are $||\phi_n^*||$, $||W^*||$, $||K^*||$ and $||D^m K^{*m}||$.

We have shown that $||\phi_n^*||$ behaves almost like $||\phi_n||$ and the effect of the parameters is very small except when G^* is nearly singular. In such exceptional cases the effect on K^* will be even more and no improvement is expected. If we also note that

$$||D^m K^{*m}|| \leq ||D^m K^*|| \; ||K^{*m-1}|| \text{ and}$$

$$||W_n^*|| \to ||(I - K^*)^{-1}|| \leq 1 + ||K^*|| + \ldots + ||K^{*m}|| \; ||(I-K^*)^{-1}||,$$

it may be the easiest way to consider minimising

$$||K^*|| \leq \sup_s \sum_{i=0}^{m-1} |p_i(s) - \lambda_i| \int_{-1}^{1} \left| \frac{\partial^{(i)}}{\partial s^i} g^*(s,t) \right| dt .$$

Here we have two independent parts inside the summation depending on the parameters; the coefficient part $|p_i(s) - \lambda_i|$ and the Green's function part $\int_{-1}^{1} \left| \frac{\partial^{(i)}}{\partial s^i} g^*(s,t) \right|$. The first part is minimised in infinity norm by $\lambda_i^* = \frac{1}{2}\{\max_s |p_i(s)| + \min_s |p_i(s)|\}$, the best approximation of $P_i(s)$. The second part is expected to behave in the same order as the norm of the inverse of the original operator G except when G^* is nearly singular. The odd case now is when G^* is nearly singular at the λ's giving the best approximations of $p_i(s)$. Such cases can be overcome by testing different values of λ_i in the region of λ_i

until $||K^*||$ takes its minimum value.

3.5. Numerical Application.

3.5.1. Introduction.

In our numerical application we will continue to work with the simple second order case

$$x''(s) + p(s)\,x'(s) + q(s)\,x(s) = y(s) \quad , \qquad x(\pm 1) = 0$$

and G^* is taking the simplest form,

$$G^*\,x = x'' + \lambda x \quad ,$$

i.e. the parameter of x' is taken zero. Obviously if $p(s) \neq 0$, then the inclusion of that parameter will give better results. Generalization to higher order equations and more complicated G^* is straightforward but a bit tedious.

The test problems will be the same ones considered with the original method in the previous chapter. Problem 2 will be neglected because it is trivial with the above G^*.

To start with λ will be taken as the one point best approximation of $q(s)$. If no good improvement is achieved, it may be an indication that G^* is nearly singular with that choice λ (especially when $p(s) \equiv 0$), and hence $||k^*||$ is tested with other values near that λ.

Before we go into numerical results we derive bounds for $||K^*||$, $||DK^*||$, $||D^2K^{*2}||$ and $||K^*\phi^*||$ following a similar way to the original method.

3.5.2. Derivation of bounds for $||K^*||$, $||DK^*||$ and $||D^2K^{*2}||$, $||K^*\phi_n^*||$, $||K^*P_{np}^*||$ $||(I - \phi_n^*)K^{*d}||$ and $||(I - P_{np}^*)K^{*d}||$.

By definition

$$(K^*u^*)(s) = \int_{-1}^{1} \left[-(q(s) - \lambda)\, g^*(s,t) - p(s) \frac{\partial g^*(s,t)}{\partial s}\right] u^*(t)\, dt$$

$$= \int_{-1}^{1} K^*(s,t)\, u^*(t)\, dt \quad ,$$

where $k^*(s,t) = -(q(s) - \lambda)\, g^*(s,t) - p(s) \frac{\partial g^*}{\partial s}(s,t)$.

(i) $$||K^*|| = \sup_{||u^*|| = 1} \sup_s \left| \int_{-1}^{1} K^*(s,t)\, u^*(t)\, dt \right|$$

$$\leq \sup_s |(q(s) - \lambda)| \int_{-1}^{1} |g^*(s,t)|\, dt + |p(s)| \int_{-1}^{1} \left|\frac{\partial g^*(s,t)}{\partial s}\right| dt$$

$$= K_o^* \quad .$$

(ii) To bound $||DK^*||$ consider DK^*U expressed in terms of x so that $x'' + \lambda x = u^*$. Then

$$DK^*u^* = DT^*x = -D(px' + (q-\lambda)x)$$

$$= -px'' - p'x' + \lambda x' - q\,x' - q'x$$

$$= -px'' - \lambda px + \lambda px - p'x' + \lambda x' - qx' - q'x$$

$$= -p(x'' + \lambda x) + (\lambda p - q')\,x - (p' + q - \lambda)\,x' \quad .$$

Returning to the variable u^* we have

$$(DK^*u^*)(s) = -p(s)\,u^*(s^*) - \int_{-1}^{1}\left((p'(s) + q(s) + q(s) - \lambda)\frac{\partial g^*(s,t)}{\partial s} + (q(s) - \lambda p(s))g^*(s,t)\right) u^*(t)\,dt$$

$$||DK^*|| \leq \sup_s |p(s)| + |p'(s) + q(s) - \lambda| \int_{-1}^{1} \left|\frac{\partial g^*(s,t)}{\partial s}\right| dt$$

$$+ |q'(s) - \lambda p(s)| \int_{-1}^{1} |g^*(s,t)|\,dt = K_1^* .$$

Therefore $||DK^*||$ can be bounded by K_1^* if bounds for p' and q' are available.

(iii) Similarly for $||D^2K^{*2}||$, consider

$D^2K^{*2}u^* = D^2K^*\nu^*$ where $\nu^* = K^*u^*$ and put

$\nu^* = G^*w^* = w^{*\prime\prime} + \lambda w^*$ then

$$D^2K^*\nu^* = D^2T^*w^* = D^2\{-Pw^* - (q-\lambda)w^*\}$$

$$= -p(w^{*\prime\prime\prime} + \lambda w^{*\prime}) - \{2p' + (q-\lambda)\}(w^{*\prime\prime} + \lambda w^*)$$

$$- \{p'' + 2q' - \lambda p\}w^{*\prime} - \{q'' - 2\lambda p' + (q-\lambda)\lambda\}w^* .$$

$$(D^2K^*)\nu^*(s) = -p(s)\,\nu^{*\prime}(s) - \{2p'(s) + (q(s) - \lambda)\}\nu^*(s)$$

$$-\{p''(s) + 2q'(s) - \lambda p(s)\}\int_{-1}^{1} \frac{\partial g^*(s,t)}{\partial s}\,\nu^*(t)\,dt - \{q''(s) -$$

$$2\lambda p'(s) + (q(s) - \lambda)\lambda\}\int_{-1}^{1} g^*(s,t)\,\nu^*(t)\,dt .$$

$$||D^2K^{*2}|| \leq \max \{|p(s)| \; ||DK^*|| + |2p'(s) + q(s) - \lambda| \; ||K^*||$$

$$+ \{|p''(s) + 2q'(s) - \lambda\, p(s)| \int_{-1}^{1} \left|\frac{\partial g^*}{\partial s}(s,t)\right| dt$$

$$+ |q''(s) - 2\lambda\, p'(s) + (q(s) - \lambda)\lambda| \int_{-1}^{1} |g^*(s,t)|dt\}||K^*|\} = K_2^* .$$

Hence a bound K_2^* for $||D^2K^{*2}||$ can be obtained if bounds for the second derivatives are available.

(iv) $K^*\phi_n^*$. To bound $K^*\phi_n^*$ we may need to use (3.9) .

$$\phi_n^* = (I - T_\lambda G^{-1})(I - \phi_n T_\lambda G^{-1} \phi_n)^{-1} \phi_n$$

$$= (I - T_\lambda G^{-1} + \phi_n T_\lambda G^{-1} - \phi_n T_\lambda G^{-1})(I - \phi_n T_\lambda G^{-1} \phi_n)^{-1} \phi_n$$

$$= (\phi_n - I) T_\lambda G^{-1} (I - \phi_n T_\lambda G^{-1} \phi_n)^{-1} \phi_n + \phi_n .$$

$$K^*\phi_n^* = K^*(\phi_n - I) T_\lambda G^{-1} (I - \phi_n T_\lambda G^{-1} \phi_n)^{-1} \phi_n + K^* \phi_n$$

$$= K^*(\phi_n - I) T_\lambda G^{-1} \phi_n J_n H_n (I - \phi_n T_\lambda G^{-1} \phi_n)^{-1} \phi_n J_n H_n + K^* \phi_n$$

If $\overline{W}_n$ is the W matrix of the operator G^*, then

$$||K^*\phi_n^*|| \leq ||K^*|| \; ||(\phi_n - I) T_\lambda G^{-1}|| \; ||\phi_n|| \; ||\overline{W}_n|| + ||K^*\phi_n|| .$$

$||K^*||$ is bounded by K_o^* and $||K^*\phi_n||$ is bounded by $K_{max}^* \; \Omega \; \Omega^*$ as it was shown in section (2.5.2). $||(I - \phi_n)T_\lambda G^{-1}||$ can be bounded using Jackson's theorem as in section (2.5.2).

$$||(I-\phi_n)T_\lambda G^{-1}|| \leq (\frac{\pi}{2})^2 \frac{||I-\phi_n||\ ||D\ T_\lambda G^{-1}||}{\ell n\ n}$$

If for the simple case, $T_\lambda G^{-1} = \lambda G^{-1}$,

then $$||(I-\phi_n)\lambda G^{-1}|| \leq (\frac{\pi}{2})^2 \frac{(1+||\phi_n||)|\lambda|}{n} \sim \frac{\ell n(n)}{n}$$

for Tchebychev zeros. This implies that for a sufficiently large n

$$||K^*\phi_n^*|| \sim ||K^*\phi_n|| + (\frac{\ell n\ n}{n})^2 \to ||K^*\phi_n||.$$

(v) For piecewise case $||K^*P_{np}^*||$ can be bounded by $||K^*||\ ||P_{np}^*||$.

(vi) $||(I - \phi_n^*)\ K^{*d}||$

As shown above, ϕ_n^* can be written as

$$\phi_n^* = (\phi_n - I)\ T_\lambda G^{-1}\ (I - \phi_n T_\lambda G^{-1}\phi_n)^{-1}\ \phi_n + \phi_n .$$

$$(I-\phi_n^*)\ K^{*d} = (I-\phi_n)T_\lambda G^{-1}(I-\phi_n T_\lambda G^{-1}\phi_n)^{-1}\phi_n\ K^{*d} + (I-\phi_n)K^{*d},$$

$$||(I-\phi_n^*)K^{*d}|| \leq ||(I-\phi_n)T_\lambda G^{-1}||\ ||\phi_n||\ ||\bar{w}_n||\ ||K^{*d}| + ||(I-\phi_n)K^{*d}||,$$

and $||(I - \phi_n)K^{*d}||$ is bounded as before, (2.5.2).

3.5.3. Numerical results

(1) The Constants

Table (3.6) gives the values of K^*_{max}, K_o^*, K_1^*, K_2^* and their ratio over their corresponding values with the original splitting

Problem	α	λ	BG^{*-1}	BDG^{*-1}	K^*_{max}	K^*_o	K^*_1	K^*_2	$\frac{K^*_{max}}{K_{max}}$	$\frac{K^*_o}{K_o}$	$\frac{K^*_1}{K_1}$	$\frac{K^*_2}{K_2}$
1	0.5	0.75	0.7247	1.3577	0.1697	0.1812	0.3694	0.5373	0.3394	0.7248	0.3694	0.7164
	1	1.5	1.2988	2.2645	0.5661	0.6494	1.2876	6.2071	0.5661	1.2988*	0.6438	2.0690*
	2	3	2.4094	7.2154	1.8073	2.4094*	7.1925*	141.0485	0.9037	2.4094*	1.7981*	11.7542*
	100	150	0.1792	2.1251	6.7139	8.0191	97.3315	10191.23	0.0671	0.1604	0.4867	0.3397
	1000	1500	0.0352	1.4482	14.0512	15.8731	724.0929	38460.2	0.0014	0.0317	0.3620	0.0128
3	0.5	-0.04514	0.4908	0.9852	0.0045	0.0045	0.01711	0.00035	0.1440	0.1440	0.219	0.0778
	1	-0.09028	0.4819	0.9790	0.0089	0.0088	0.0337	0.0014	0.1408	0.1408	0.2157	6.0778
	2	-0.18056	0.4650	0.9439	0.0174	0.0171	0.0656	0.0052	0.1369	0.1368	0.2099	0.0722
	100	-9.028	0.0998	0.3312	0.4036	0.2058	1.1505	1.5655	0.0645	0.0329	0.0736	0.0087
	1000	-90.28	0.01107	0.1052	1.5859	0.2998	3.6546	18.2770	0.0254	0.0048	0.0234	0.0015
4	0.5	-0.15625	0.4694	0.9510	0.4851	0.4755	0.8269	1.0702	0.7762	0.7608	0.7350	0.6838
	1	-0.3125	0.4422	0.9074	0.9657	0.9074	1.6238	4.1937	0.7726	0.7259	0.7217	0.6699
	2	-0.625	0.3962	0.8332	1.9155	1.6665	3.1457	16.1507	0.8513	0.7407	0.6291	0.6450
	100	-31.25	0.0318	0.1789	75.6156	17.8880	165.4939	29153.01	0.6049	0.1430	0.7356	0.4657
	1000	-312.5	0.0032	0.0751	499.9936	56.5685	1825.271	2897246	0.4000	0.0453	0.8112	0.4628

Table (3.6) Value of the constants

(table (2.1)). If we look to these ratios we can easily observe

(i) Some odd values in problem (1) (values with *) where the new splitting gives worse results. This is obviously due to the large values of BG^{*-1} and BDG^{*-1} at $\lambda = 1.5,3$. The easiest practical way for avoiding these nearly singular cases is to consider other values of λ nearby and to choose the best of them, as shown later.

(ii) Huge reductions were achieved in problem 3. That is because (1) λ is negative and hence BG^{*-1} and BDG^{*-1} are well behaved. (2) The function q does not vary very much and can be well approximated by a constant. (3) p(s) = 0. (p is not accounted for with this simple G*).

(iii) For problem (4) all values are reduced but the reductions are not as in (2) since here $p \neq 0$. Obviously if G* includes an approximation of p then similar reductions are expected.

(iv) For most of the cases the reductions increase with α and the maximum reductions occur in K_2.

(2) $||W_n^*||$

Table (3.7) gives the values of $||W_n^*||$ for both global and piecewise case. The values in each row are tending to a constant which confirms theorems (3.2) and (3.4). In comparing these values of $||W_n^*||$ with values of $||W_n||$ in tables (2.2) and (2.3) we notice the following:

(i) For problem 1 $||W_n^*|| < ||W_n||$ for every n and for every value of α. From section (3.4) that is what we expect for $\alpha = 0.5$, 100 where K was shown to be reduced.

Table (3.7) $\|w^*\|$ values

	Problem	α	λ	n = 5	n = 10	n = 15	n = 20	n = 25	n = 30
(a) Global	1	0.5	0.75	1.0122	1.0671	1.100	1.1136	1.1248	1.1300
		1	1.5	1.3572	1.3402	1.3573	1.3543	1.3573	1.3565
		2	3.0	5.6781	5.4391	5.6996	5.6359	5.6996	5.6733
		100	150	112.6927	6.0762	7.3665	5.7066	7.7255	7.0296
	3	0.5	-0.04514	1.0042	1.0045	1.0045	1.0045	1.0045	1.0045
		1	-0.09078	1.0027	1.0094	1.0127	1.0144	1.0155	1.0162
		2	-0.18056	1.0060	1.0181	1.0244	1.0278	1.0300	1.0313
		100	-90.78	1.2439	1.2402	1.2436	1.2450	1.2436	1.2462
	4	0.5	-0.15625	1.5017	1.5415	1.5492	1.5520	1.5533	1.5540
		1	-0.3125	2.0659	2.1750	2.1967	2.2043	2.2079	2.2099
		2	-0.6250	3.3371	3.6551	3.7205	3.7439	3.7547	3.7607
		100	-31.25	34.1241	200.3131	132.3251	148.7529	163.9271	173.4809
(b) Piece-wise	1	0.5	0.75	1.0978	1.1324	1.1416	1.1475	1.1503	1.1526
		1	1.5	1.3501	1.3558	1.3568	1.3572	1.3573	1.3574
		2	3.0	5.2240	5.5806	5.6477	5.6710	5.6816	5.6872
		100	150	3.2398	32.2530	13.3198	10.46120	8.9978	8.7398
	3	0.5	-0.04514	1.0045	1.0046	1.0046	1.0046	1.0046	1.0046
		1.0	-0.09078	1.0124	1.0161	1.0174	1.0181	1.0184	1.0187
		2.0	-0.18056	1.0240	1.0311	1.03360	1.0348	1.0356	1.0361
		100	-90.78	1.2517	1.2497	1.2489	1.2499	1.2502	1.2502
	4	0.5	-0.15625	1.4933	1.5230	1.5336	1.5389	1.5422	1.5444
		1	-0.3125	2.058	2.1308	2.1573	2.1711	2.1795	2.1852
		2	-0.625	3.3382	3.5278	3.6032	3.6434	3.6683	3.6853
		100	-31.25	1.9127	8.1019	26.3493	76.9849	295.8121	1005.084

But for $\alpha = 1$ and 2 where $||K||$ is larger the only explanation may be due to the singularity at $\alpha = 2.19$ which will affect the numerical calculation of $||W_n||$ more than $||W_n^*||$.

(ii) For problem (3) $||W_n^*|| < ||W_n||$ as expected.

(iii) For problem (4) although $||K^*|| < ||K||$, $||W_n^*||$ is almost similar to $||W_n||$ or slightly larger. This is probably due to the domination of $p(s)$, ($|p(s)| >> |q(s)|$), we notice in the piecewise case (b) that for $\alpha = 100$, $||W_n^*||$ looks as if it is diverging, but if we look to more values of $||W_n^*||$ we will notice that was just a matter of accident, for example $||W_{40}^*|| = 224.777$.

(3) The applicability

Tables (3.8) a) and b) give the number n required for the applicability for each of the methods in section (2.5.3) with this new splitting for global and piecewise methods respectively. In general we see how much the improvement achieved with this new approach. If we go into detailed comparisons with Tables (2.5) and (2.6) concentrating on cases where the number required is more than one actually needs, we may notice

Problem (1) : $\alpha = 0.5$ $\delta_1^* << \delta_1$ $\delta_2^* << \delta_2$ $\Delta_1^* < \Delta_1$ and $\Delta_2^* < \Delta_2$

$\alpha = 1$ $\delta_1^* << \delta_1$ $\delta_2^* < \delta_2$ $\Delta_1^* << \Delta_1$ and $\Delta_2^* > \Delta_2$

$\alpha = 2$ No improvement. This case and case Δ_2 with $\alpha = 1$ are expected due to the bad results we have got for the constants there. These cases will be reconsidered.

TABLE (3.8) Applicability

(a) Global

Problem	α	λ	δ_1^*	δ_2^*	Δ_1^*	Δ_2^*
1	0.5	0.75	4	4	2	2
	1	1.5	35	19	18	12
	2	3	>120*	>120*	>120*	>120*
	100	150	>120*	>120*	>120*	>120*
3	0.5	-0.04514	2	2	2	2
	1	-0.09028	2	2	2	2
	2	-0.18056	2	2	2	2
	100	-9.02778	22	10	4	3
4	0.5	-0.15625	17	6	5	4
	1	-0.3125	97*	22	65	16
	2	-0.6250	>120*	74	>120*	77
	100	-31.25	>120*	>120*	>120*	>120*

* assuming $||W_n^*||$ constant and $||\phi_n^*|| = ||\phi_n^*||$

TABLE (3.8) (cont'd) Applicability

(b) Piecewise

Problem	α	λ	δ_1^*	δ_2^*	Δ_1^*	Δ_2^*
1	0.5	0.75	3	2	2	2
	1	1.5	11	10	7	7
	2	3	>200*	>200*	>200*	>200*
	100	150	>200*	>200*	>200*	>200*
3	0.5	-0.04514	2	2	2	2
	1	-0.09028	2	2	2	2
	2	-0.18056	2	2	2	2
	100	-9.02778	9	5	2	2
4	0.5	-0.15625	5	4	2	2
	1	-0.3125	14	8	11	7
	2	-0.625	48	23	63	28
	100	-31.25	>200*	>200*	>200*	>200*

* assuming $||W_n^*||$ constant and $||\phi_n^*|| = 1.414$

Problem (3) : Huge improvements are acheived with $\alpha = 100$ and $\alpha = 1000$

Problem (4) : $\alpha = 0.5$ $\delta_1^* << \delta_1$, $\delta_2^* << \delta_2$, $\Delta_1^* << \Delta_1$, $\Delta_2^* < \Delta_2$

$\alpha = 1$ $\delta_1^* << \delta_1$, $\delta_2^* << \delta_2$, $\Delta_1^* << \Delta_1$, $\delta_2^* << \delta_2$

$\alpha = 2$ $\delta_2^* << \delta_2$ and $\Delta_2^* << \Delta_2$. δ_1 and Δ_1 are improved with the piecewise method but nothing appeared to happen with the global method.

(4) The re-consideration of problem (1)

Table (3.9) (a) gives the values of the constants for problem $(1,\alpha=2)$ with different values of λ near 3 the best approximation of $2(1+t^2)$. We look for a λ with the smallest $||K||$. That appears to be $\lambda = 1$. Unfortunately the other constants don't take their minimum there,but they are not that much worse. This situation suggests putting into consideration other values of λ for testing.

Tables (3.9)(b) &(c) describe the applicability for $\lambda = 0.5$, 0.75, 1, 1.25 and 1.5 with the global and piecewise methods respectively. We observe that with all these λ's a good improvement has been achieved. Further we observe that the best applicability occurred with $\lambda = 1$ where $||K||$ is minimum which supports our method in dealing with the choice of the λ's.

Table (3.10) gives new results for $\alpha = 0.5$, 1 with $\lambda = 0.3125$ and 0.625 respectively. These λ's are chosen in respect of the above results for case $\alpha = 2$. We observe that better results are obtained and the previous odd case ($\alpha = 1, \Delta_2 < \Delta_2^*$) now is eliminated

TABLE (3.9) Applicability

(Problem 1, $\alpha = 2$)

(a) The constants

λ	BG^{*-1}	BDG^{*-1}	K^*_{max}	K^*_0	K^*_1	K^*_2
0.25	0.5580	1.0926	0.9567	0.9795	4.0973	12.2351
0.5	0.6307	1.2085	0.9099	0.9608	4.2296	12.6509
0.75	0.7247	1.3577	0.8604	0.9459	4.4126	13.3480
1	0.8508	1.5574	0.8085	0.9365	4.6722	14.4779
1.25	1.0288	1.8386	0.7575	0.9401	5.0562	16.4127
1.5	1.2988	2.2645	0.7402	0.9657	5.6613	19.9100
1.75	1.7572	2.9863	0.7779	1.0382	6.7193	27.1389
2	2.706	4.4789	0.9258	1.2342	8.9578	46.6914
2.25	5.8386	9.4009	1.5140	1.9816	16.4517	152.4970
2.5	39.0761	122.3044	15.2872	19.5380	183.4566	19146.00
2.75	4.5240	13.8011	2.5872	3.3930	17.2482	378.7914

(b) Applicability (Global)

λ	δ^*_1	δ^*_2	Δ^*_1	Δ^*_2
0.5	>120*	95*	>120*	61
0.75	>120*	92*	>120*	55
1	>120*	85*	>120*	48
1.25	>120*	85*	>120*	49
1.5	>120*	87*	>120*	49

(c) Applicability (Piecewise)

λ	δ^*_1	δ^*_2	Δ^*_1	Δ^*_2
0.5	130*	30	120*	26
0.75	120*	29	105*	25
1	110*	28	95*	24
1.25	120*	28	105*	24
1.5	135*	31	120*	26

Applicability

TABLE (3.10) (Problem 1, $\alpha = 0.5, 1$)

(a) The Constants

α	λ	BG^{*-1}	BDG^{*-1}	K^*_{max}	K^*_0	K^*_1	K^*_2
0.5	0.3125	0.5746	1.1191	0.1278	0.1347	0.7694	0.2251
1	0.625	0.6745	1.2781	0.2822	0.3136	1.7573	1.4722

(b) Applicability (Global)

α	λ	δ^*_1	δ^*_2	Δ^*_1	Δ^*_2
0.5	0.3125	8	3	2	2
1	0.625	35	9	5	3

(c) Applicability (Piecewise)

α	λ	δ^*_1	δ^*_2	Δ^*_1	Δ^*_2
0.5	0.3125	4	2	2	2
1	0.625	9	5	2'	2'

CHAPTER FOUR

Algorithms for Error Bounds and Estimates

4.1. Introduction

If we define the residual of any approximate solution x_n of

$(G - T)\ x = y$ by

$$r_n = y - (G - T)\, x_n \tag{4.1}$$

then the error $e_n = x_n - x$ in x_n is related to r_n by

$$(G - T)\ e_n = r_n, \quad \text{or if } (G - T)^{-1} \text{ exists,}$$

$$\text{then} \quad e_n = (G - T)^{-1} r_n \,. \tag{4.2}$$

(4.2) gives a straightforward bound on e_n,

$$\| x - x_n \| = \| e_n \| \leq \| (G-T)^{-1} \| \ \| r_n \| . \tag{4.3}$$

r_n can be calculated by substituting x_n in the differential equation and bounds on $\| (G - T)^{-1} \|$ are available from the previous chapters. We will show in this chapter that by examining the inverse operator $(G - T)^{-1}$ and the residual we can obtain closer bounds with less work.

It is seen by experiments at a large variety of problems that $\| Q_n \|$ settles very early and gives a good estimation of the norm of the inverse differential operator $(G - T)^{-1}$. This property is used to justify a simple estimate of the bound on x which is shown to be of high quality.

It is also shown that by examining the inverse operator and the residual as before we can derive different error bound estimates

and error estimates which vary in their closeness according to the amount of work involved.

The practical implementation of these ideas are discussed and tested on different examples for both the global and piecewise collocation method.

4.2. The behaviour of the residual

(a) Global case

from (4-1) the residual

$$r_n = y - (G - T)\, x_n$$

$$= y - (G - T - \phi_n T + \phi_n T)\, x_n$$

$$= y - (G - \phi_n T) x_n + T x_n - \phi_n T x_n$$

$$= y - \phi_n\, y + T x_n - \phi_n\, T x_n$$

$$= (I - \phi_n)\,(y + T x_n)\,. \qquad (4.4)$$

It can be seen from (4.4) that r_n is the remainder of the interpolation of the function $y + Tx_n$. This useful property of the residual will be used to obtain effective algorithms for error bounds and estimates. It will be shown also that it is essential for the efficiency and the reliability of these algorithms to have the residual well approximated by a polynomial. That can be shown to depend on the smoothness of the coefficients and the right hand side of the differential equation. For, let $R_n = \phi_N r_n$ be a polynomial of degree $N - 1$, $N > n$ which agrees with the residual at the collocation points and any other $N - n$ distinct points. Then

$$r_n - R_n = (I - \phi_N)\, r_n = (I - \phi_N)\,(I - \phi_n)\,(y + T x_n).$$

But by Jackson's theorem (Chen ey, p.147), if $(y + T x_n) \in C^d(-1,1)$ then there is a polynomial $\tilde{u}$ of degree $N - 1$ such that

$$||(y + T x_n) - \tilde{u}|| \leq J_{dN} ||D^d (y + Tx_n)|| .$$

Therefore $||r_n - R_n|| = ||(I - \phi_N)(I - \phi_n)(y + T x_n - \tilde{u} + \tilde{u})||$

$$= ||(I - \phi_N)((I - \phi_n)\tilde{u} + (I - \phi_N)(I - \phi_n)(y + Tx_n - \tilde{u}))||$$

$$= ||(I - \phi_N)(I - \phi_n)(y + T x_n - \tilde{u})|| \text{ ?}$$

$$\leq ||I - \phi_N|| \; ||(I - \phi_n|| \; J_{dN} \; ||D^d (y + Tx_n)||.$$

If we write $r_n = R_n + \Psi_n$ then

$$||r_n|| \leq ||R_n|| + ||\Psi_n||.$$

Now if the extra Points are chosen so that $||I - \phi_N||$ becomes small for large n and N e.g. if the collocation points are zeros of T_n and the extra Points maxima of T_{2n} then provided the right hand side y and the coefficients of the differential equation $\{P_i\}$ are sufficiently differentiable r_n can be expressed as a polynomial R_n plus small quantity Ψ_n if n is sufficiently large. This polynomial R_n will be called a principal part of the residual and Ψ_n the modified residual.

(b) Piecewise case

In a similar way the residual for x_{np} can be expressed as $r_{np} = (I - P_{np})(y + T x_{np})$

i.e. a reminder of the piecewise interpolation of the function $y + T x_{np}$.

If we suppose that $D^d (y + T x_{np})$ is continuous on the ith interval (t_{i-1}, t_i), then using Jackson's theorem as above there is

a piecewise polynomial u_i of degree $k-1$ such that

$$||y(t) + T x_{np}(t) - u_i(t)|| \leq J_{dk} \left(\frac{\pi}{4} (t_i - t_{i-1}) \right)^d \sup_{t_{i-1}<t<t_i} D^d (y+Tx_{np})(t) = Li .$$

If $R_i = P_k r_i$ is the polynomial which agrees with the residual in ith partition, r_i, at the collocation points and other $(k-p)$ points then as in the global case,

$$r_i(t) - R_i(t) = (I-P_{nK}) (I-P_{np}) (y(t)+Tx_{np}(t) + u_i(t) - u_i(t)) \quad t\varepsilon (t_{i-1}, t_i)$$

$$= (I - P_{nk}) (I - P_{np}) (y(t) + Tx_{np}(t) - u_i(t))$$

$$||r_i - R_i|| \leq ||I - P_{nk}|| \; ||I - P_{np}|| \; L_i .$$

If $\Psi_i = r_i - R_i$ then

$$||r_i|| \leq ||R_i|| + ||\Psi_i||, \text{where } ||\Psi_i|| < L_i.$$

This shows that the residual r_{np} can be expressed as a piecewise polynomial plus a remainder and that for sufficiently large n

$$||\Psi_i|| << ||R_i||. \quad i = 1,\ldots,n.$$

4.3. Improved error bounds using the polynomial approximation of the residual

(A) Global Case

If we recall (4.2),

$$e_n = (G - T)^{-1} r_n$$

then by (4.5)

$$e_n = (G - T)^{-1} (R_n + \Psi_n) = (G - T)^{-1} R_n + (G - T)^{-1} \Psi_n$$

$$= G^{-1}(I - K)^{-1} R_n + (G - T)^{-1} \Psi_n$$

$$= G^{-1} (I + (I - K)^{-1}K) R_n + (G - T)^{-1} \Psi_n$$

$$= G^{-1} R_n + (G - T)^{-1} K R_n + (G - T)^{-1} \Psi_n \quad . \qquad (4.10)$$

Now since R_n is a continuous function, by definitions of G^{-1} and K

$$(G^{-1}R_n)(s) = \int_{-1}^{1} g(s,t) R_n(t)\, dt \qquad (4.11)$$

and $$(K R_n)(s) = \sum_{k=0}^{m-1} P_k(s) \int_{-1}^{1} g^k(s,t) R_n(t)\, dt \, . \qquad (4.12)$$

That gives,

$$||G^{-1} R_n|| = \sup_s |(G^{-1} R_n)(s)| \quad \text{and} \quad ||K R_n|| = \sup_s |(K R_n)(s)| .$$

Therefore from (4.10)

$$||e_n|| \leq ||G^{-1}R_n|| + ||(G - T)^{-1}|| (||K R_n|| + ||\Psi_n|| \qquad (4.13)$$

Since R_n is a polynomial and Green's function $g(s,t)$ is a piecewise polynomial, the integrals (4.11) and (4.12) can be found exactly. This bound (4.13) is expected to be very accurate since the principal part of the error $||G^{-1}R_n||$ is exact. The accuracy obviously depends on how much R_n is taking from the residual r_n which can be checked by the size of $||\Psi_n||$.

We note here also that because R_n is a highly oscillatory function and K is an integral operator, we expect many cancellations in the integration (4.12) which makes $||K R_n||$ much smaller than $||r_n||$

(assuming $||\Psi_n||$ is sufficiently small). That is guaranteed for n sufficiently large, depending on the smoothness of y and $\{p_i\}$ as shown in the previous section. Now having $||KR_n||$ and $||\Psi_n||$ relatively small compared with $||G^{-1}R_n||$ in (4.13) justifies using a crude bound in $||(G - T)^{-1}||$. For example one could use a smaller n value for finding a bound on $||(G - T)^{-1}||$ than the one being used for the calculation of r_n or (R_n).

(b) Piecewise case

If we consider any subinterval (t_{i-1}, t_i) then the error e_i in that interval can be expressed as in (4.10) by

$$e_i = (G^{-1} R)_i + ((G - T)^{-1} K R)_i + ((G - T)^{-1}) \Psi_i \qquad (4.14)$$

where $(G^{-1}R)_i$ and $(K R)_i$ are defined here by

$$(G^{-1}R)_i (s) = \sum_{j=1}^{n} \int_{-1}^{1} g_{ij}(s,t) R_i(t)\, dt \qquad (4.15)$$

$$(K R)_i (s) = -\sum_{k=0}^{m-1} P_k(\tfrac{1}{2}(t_i - t_{i-1})\, s + t_i + t_{i-1}) \sum_{j=1}^{n} \int_{-1}^{1} g_{ij}^{k}(s,t) R_i(t)\, dt \qquad (4.16)$$

and $g_{ij}(s,t) = g(\tfrac{1}{2}(t_i - t_{i-1})\, s + t_i + t_{i-1},\ \tfrac{1}{2}(t_i - t_{i-1})t + t_i + t_{i-1})$. (4.17)

The integrals (4.15) and (4.16) can be found exactly and we get

$$||e_i|| \leq ||(G^{-1}R)_i|| + ||(G - T)^{-1}|| \,(||K R|| + ||\Psi||)\,. \qquad (4.18)$$

For the same reasons mentioned for the global case $||e_i||$ can give a close bound for the error in the ith interval (t_{i-1}, t_i) and for sufficiently large n $||KR_i||$ and $||\Psi_i||$ will be small enough to accept a crude bound on $||(G - T)^{-1}||$.

4.4. Error Estimates and Estimates of the bounds

We have shown in Chapter 2 that $||Q_n|| \to ||(G - T)^{-1}||$ as $n \to \infty$. We have also shown that the type of bounds derived for $||(G - T)^{-1}||$ involve matrix inversion and may not be applicable for small values of n. That is the main criticism of these bounds. One of the suggestions here is to take $||Q_n||$ as an estimate for $||(G - T)^{-1}||$. This estimate clearly has no formal restriction on the value of n, but one would of course check that $||Q_n||$ really has settled down when using them.

4.4.1. Estimates using $||Q_n||$ only

$||e_n|| \leq ||(G - T)^{-1}|| \; ||r_n||$, then the above result justifies the following simple bound estimation.

$$||e_n|| \sim e_Q = ||Q_n|| \; ||r_n|| \quad . \tag{4.19}$$

If $||Q_n||$ has already settled down when $n = n^*$ then it will be unreasonable to recalculate $||Q_n||$ with larger value of n. Hence a cheaper estimate can be taken as

$$e_Q = ||Q_n^*|| \; ||r_n|| \quad . \tag{4.20}$$

When $||r_n||$ is sufficiently small, $|[||(G - T)^{-1}|| - ||Q_n^*||]| \; ||r_n||$ will be very small (compared to the actual bound) which makes this estimate very close to the actual bound. Results show that this estimate is closer than the estimates given by Cruickshanck (1974) and McKeown (1978).

4.4.2. Estimates using approximation of the residual

One may recall (4.13)

$$||e_n|| \leq ||G^{-1}R_n|| + ||(G - T)^{-1}|| \; (||K\,R_n|| + ||\Psi_n|| \;)$$

and take as an estimate

$$e_1^* = ||G^{-1}R_n|| + ||Q_n^*|| \, (||K\,R_n|| + ||\Psi_n||) \, . \qquad (4.21)$$

We note here that the calculation of Ψ_n involves r_n which requires the evaluation of the differential equation with the approximate solutions $\{x_n^{(k)}\}_{k=0}^{m}$ substituted for $(x^{(k)})_{k=0}^{m}$. That means to have an accurate bound of $||\Psi_n||$ may need to evaluate Ψ_n at a large sample of points which will be very expensive. However by the theory in section (4.2), we expect $||\Psi_n||$ to be very small for sufficiently large n. That can be checked with a small selected number of points. If it is true one may neglect $||\Psi_n||$ and use the cheaper estimate

$$e_2^* = ||G^{-1}R_n|| + ||Q_n^*|| \; ||K\,R_n|| \, . \qquad (4.22)$$

4.4.3. Estimates using the LU decomposition of the solution matrix

Recall (4.10)

$$e_n = G^{-1} R_n + (G - T)^{-1} K R_n + (G - T)^{-1} \Psi_n$$

Now since $K R_n$ can be found exactly by (4.12) and it does not vanish at the collocation points (unlike the residual) one can apply the LU decomposition of the solution matrix and find $(G - T)^{-1} K R_n$. Then the following estimate can be obtained

$$E_1 = ||G^{-1} R_n + (G - \phi_n T)^{-1} \phi_n K R_n|| + ||Q_n^*|| \; ||\Psi_n|| . \qquad (4.23)$$

As above one may neglect $||\Psi_n||$ and take as cheaper estimate

$$||E_2|| = ||G^{-1} R_n + (G - \phi_n T)^{-1} \phi_n K R_n|| \, . \qquad (4.24)$$

This estimate is very cheap to calculate since it doesn't require the calculation of $||Q_n^*||$ or $||\Psi_n||$. Later it will be shown that with many problems this estimate is very close to the actual error.

4.4.4. Estimate using the principal part of the residual and the principal part of the equation only

Finally one may assume further that $(G - T)^{-1} K R_n$ is sufficiently small compared with $G^{-1}R_n$ and take $||G^{-1}R_n||$ as a cheap simple estimate of the error. This estimate may not be reliable if the problem is nearly singular ($||(G - T)^{-1}||$ very large), or it has got large coefficients, i.e. $||K R_n|| > ||G^{-1}R_n||$. However apart from these exceptional cases this estimate is very close to the actual error as shown in results later.

4.5. Numerical examples

4.5.1. General comments on the computer program

The program basically sets up and solves the linear algebraic system for the collocation method producing the approximate solution x_n. The bases elements of the solution are chosen as Tchebychev polynomials $\{T_i\}$ or Legendre polynomials $\{P_i\}$. The collocation points $\{\xi_j\}$ are chosen as Tchebychev or Legendre zeros depending on the base. A sequence of procedures are written to compute all the different quantities required. It is of practical interest to mention the following points.

(1) The collocation matrix for the piecewise case (the solution matrix) takes the special-block form diagrammed in Fig. (4.1). This block matrix is solved by special library procedure.

(2) To calculate the matrix Q the LU decomposition of the solution matrix A is used to compute its inverse. Then $Q = PA^{-1}$ where P is the evaluation matrix of the basis elements at the collocation points. In principle the evaluation of P could be at any arbitrary points $\{s_i\}$ as shown in section (2.4.3). The only reason for this choice is to make use of the elements $\{T_i(\xi_j)\}$ or $\{P_i(\xi_j)\}$ already calculated in constructing A.

(3) In calculating R_n the choice of its number of coefficients N is arbitrary in principle. It could be n + 1, n+ 2, or larger. Since (n+1) maxima of r_n are known for the Tchebychev case it is of practical convenience to take N = 2n and express R_n by

$$R_n = T_n \sum_{i=0}^{n}{}' a_i T_i .$$

Then using the orthogonality properties of Tchebychev series we can express the coefficients by

$$a_i = \frac{2}{(n+1)} \sum_{j=0}^{n}{}'' r_n\left(\cos\left(\frac{\pi j}{n+1}\right)\right) \cos\left(\frac{ij\pi}{n+1}\right) .$$

For the Legrende case R_n is expressed by

$$R_n = P_n \sum_{i=0}^{n}{}' a_i T_i$$

and the coefficients $\{a_i\}$ are expressed by

$$a_i = \frac{2}{(n+1)} \sum_{j=0}^{n} \frac{r_n\left(\cos\frac{\pi j}{n+1}\right)}{P_n\left(\cos\frac{\pi j}{n+1}\right)} \cos\left(\frac{ij\pi}{n+1}\right).$$

Since the local maxima of Tchebychev and Legrende polynomials are asymptotically the same , (Szegő (1939) , theorem (6.21.2)), $P_n\left(\cos\frac{\pi j}{n+1}\right)$ will not be small and so the a_i can be calculated accurately.

We note here that the tail coefficients of R_n could be taken as a measure for the size of $||\Psi_n||$. For example when these coefficients are sufficiently small it is an indication that $||\Psi_n||$ is small and it may be neglected beforehand. We may also note that when these coefficients are sufficiently small they could be missed in calculating R_n and save some effort.

(4) If we assume that Green's function is not known explicitly then

$$(G^{-1}R_n)(s) = \int_{-1}^{s} \frac{(s-t)^{m-1}}{(m-1)!} R_n(t)\,dt + a_o T_o(s) + a_1 T_1(s) + \ldots + a_{m-1} T_{m-1}(s)$$

where the parameters $a_o, a_1, \ldots, a_{m-1}$ are found from the boundary conditions,

$$\sum_{k=0}^{m-1} \alpha_{ik} (G^{-1}R_n)^{(k)}(-1) + \sum_{k=0}^{m-1} \beta_{ik} (G^{-1}R_n)^{(k)}(1) = \gamma_i \qquad i=1,2,\ldots,m .$$

We note that $\int_{-1}^{s} \frac{(s-t)}{(m-1)!} R_n(t)\,dt$ is a multiple integral of R_n and could be expressed as a summation of Tchebychev series, using the properties of Tchebychev series, (Fox & Parker (1968)).

If $(G^{-1}R_n)(s)$ is known then $(K R_n)(s)$ is simply

$$(K R_n)(s) = - \sum_{k=0}^{m-1} P_k(s) (G^{-1}R_n)^{(k)}(s).$$

In a similar way for the piecewise case

$$(G^{-1}R_i)(s) = (\tfrac{1}{2}(t_i - t_{i-1})^{m-1}) \int_{-1}^{s} \frac{(s-t)^{m-1}}{(m-1)!} R_i(t)\,dt + \sum_{j=0}^{m-1} a_{ij} T_j(\tfrac{1}{2}(t_i - t_{i-1})\, s + t_i + t_{i-1})$$

and

FIGURE (4.1)

The matrix A with 3 partitions

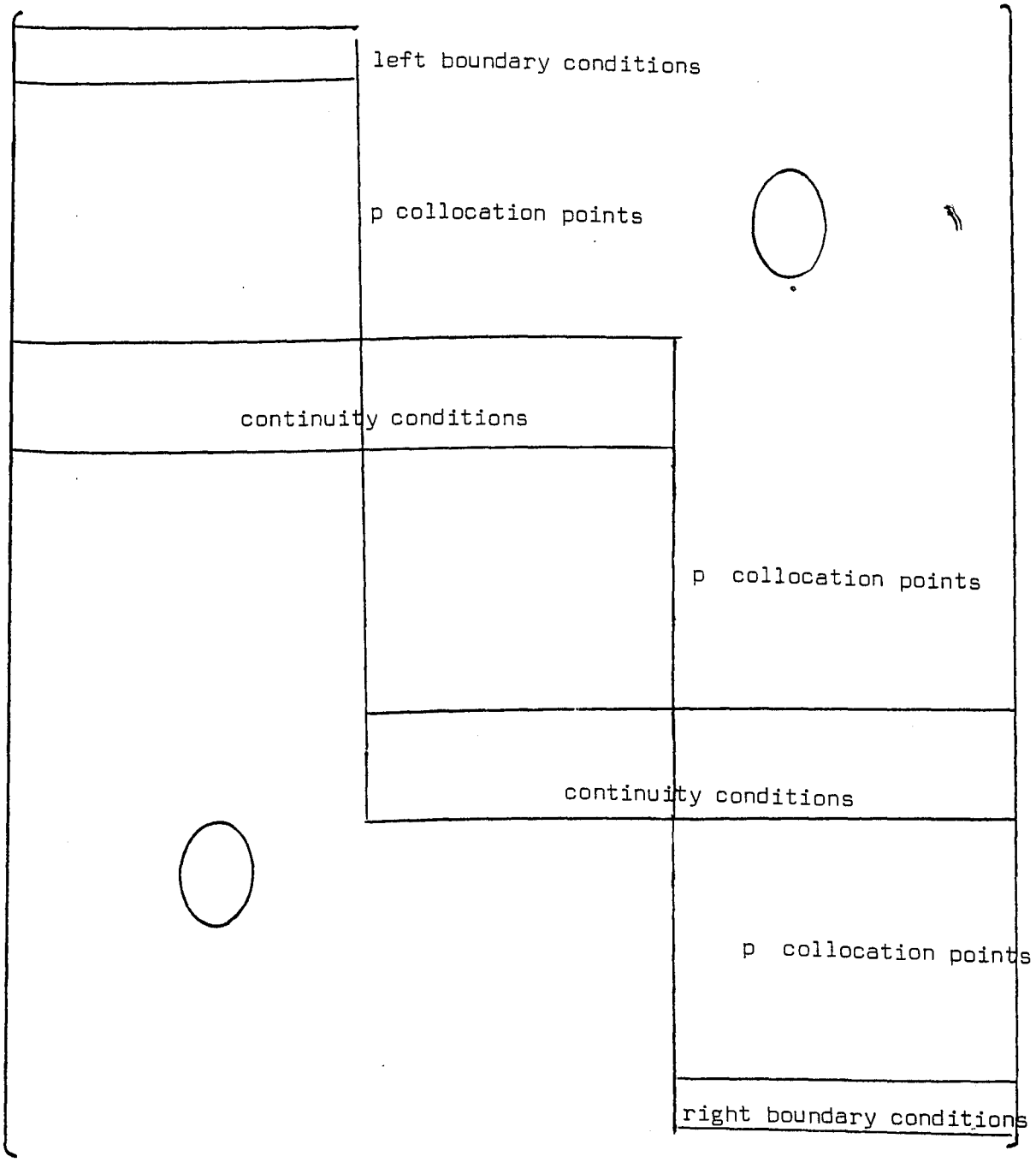
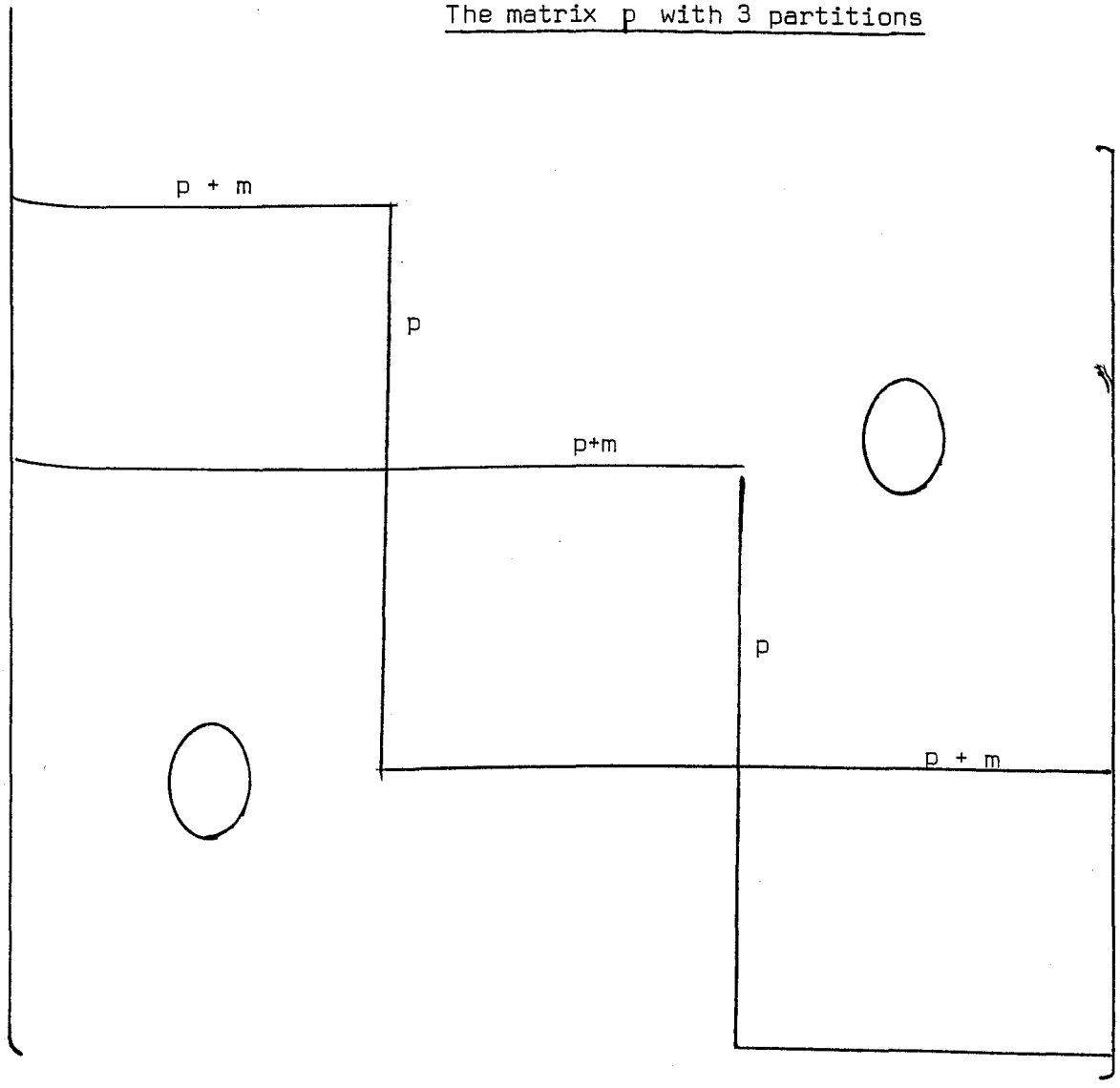


FIGURE (4.2)

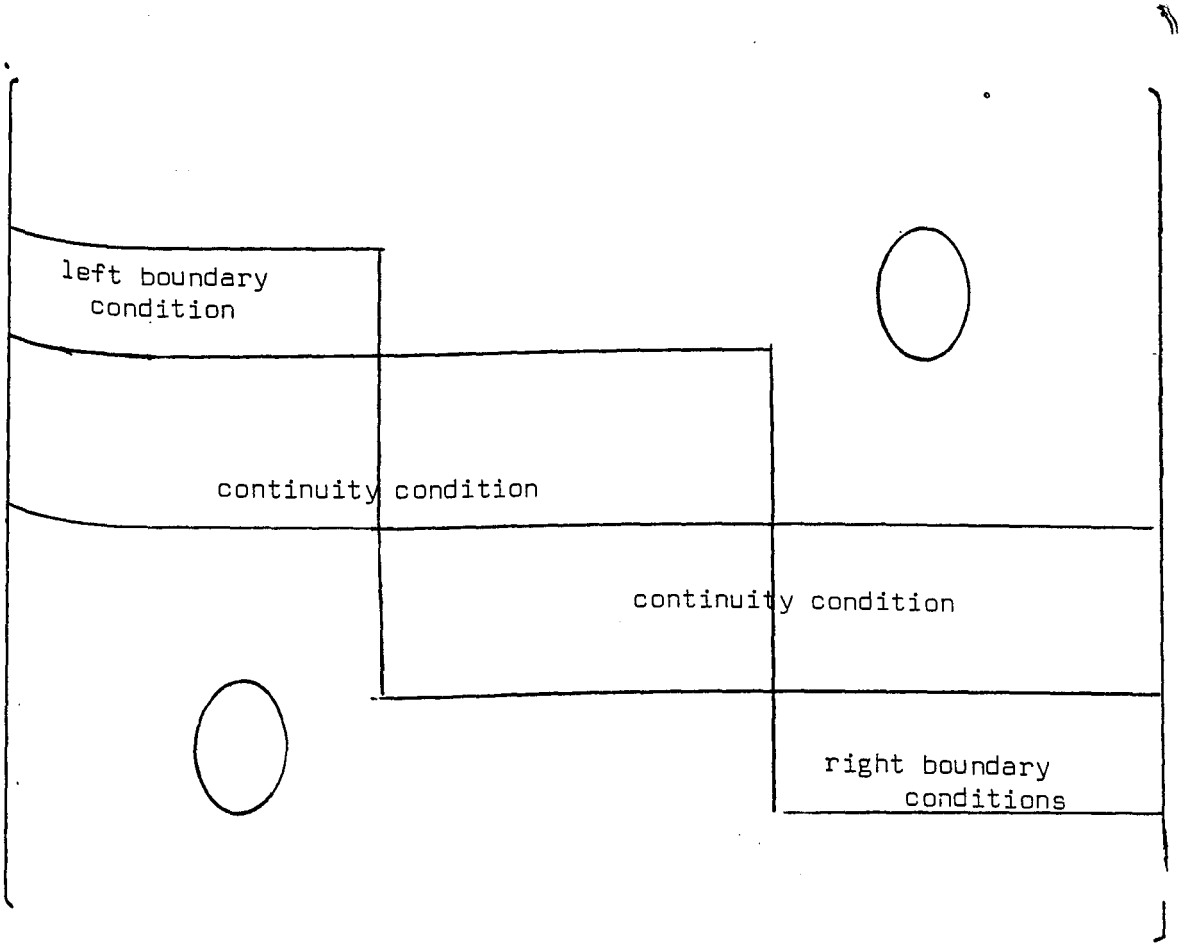
The matrix p with 3 partitions



Typical element $P_{ij} = T_i(\xi_j)$ or $=P_i(\xi_j)$.

FIGURE (4.3)

The matrix GF with 3 partitions



$$(K R_i)^{(k)}(s) = -\sum_{k=0}^{m-1} P_k\left(\tfrac{1}{2}(t_i - t_{i-1})\, s + t_i + t_{i-1}\right) (G^{-1}R_i)^{(k)}(s).$$

We note here that in solving for the constants $\{a_{ij}\}_{j=0}^{m-1}$ $i=i,\ldots,n$ using continuity and boundary conditions we obtain a matrix FG taking the block form described in Fig. (4.3). To solve for these constants the special library procedure is used with no collocation points.

4.5.2. The behaviour of $||Q_n||$

Table (4.1) gives the values of $||Q_n||$ for the simple differential operators (i) $x' = y, x(-1) = 0$; (ii) $x'' = y, x(\pm 1) = 0$; (iii) $x^{(3)} = y; x(\pm 1)$, $x'(-1) = 0$ and (iv) $x^{(4)} = y$ $x(\pm 1) = x'(\pm 1) = 0$. In each column $||Q_n||$ is tending to a constant which in this special case equals $||G^{-1}||$. That confirms $||Q_n||$ indeed tends to $||(G - T)^{-1}||$ as proved previously. Table (4.3) considers the same problems for the piecewise case with two and four collocation points and shows that $||Q_n||$ is similarly tending to $||G^{-1}||$.

As further illustration to the problems considered in Chapter 2, we consider the following problems which vary in order and smoothness.

Problem (5)	$x'' + (t)\, x = y$	$x(\pm 1) = 0$
Problem (6)	$x''' + x'' + \sin(t)\, x = y$	$x(\pm 1) = x'(-1) = 0$
Problem (7)	$x^{(iv)} + 200\, x = y$	$x(\pm 1) = x'(\pm 1) = 0$
Problem (8)	$x^{(iv)} + x''' + e^{t} x = y$	$x(\pm 1) = x'(\pm 1) = 0$

The values of $||Q_n||$ of these problems are given in table (4.2) for the global case and in table (4.4) for the piecewise case.

TABLE (4.1) The behaviour of $||Q_n||$ (Global)

n \ m	1	2	3	4
1	1.7071	0.5000	0.1967	0.04167
2	1.8660	0.4268	0.1732	0.0363
3	1.9239	0.5000	0.1957	0.04167
4	1.9511	0.4665	0.1983	0.0397
5	1.9659	0.5000	0.1916	0.0422
7	1.9808	0.5000	0.1924	0.4167
10	1.9898	0.4915	0.1947	0.0406
14	1.9945	0.4952	0.1973	0.0410
19	1.9969	0.5000	0.1974	0.04167
25	1.9982	0.5000	0.1967	0.04167

TABLE (4.2) The behaviour of $||Q_n||$ (Global)

n \ Problem	5	6	7	8
1	0.5000	0.1475	0.0045	0.0400
3	0.5126	0.1649	0.0053	0.0389
5	0.5473	0.1629	0.0052	0.0395
8	0.5581	0.1641	0.0053	0.0387
12	0.5621	0.1631	0.0053	0.0391
17	0.5651	0.1637	0.0053	0.0391
22	0.5651	0.1639	0.0053	0.0391
27	0.5651	0.1640	0.0053	0.0391

TABLE (4.3) The behaviour of $||Q_n||$ (piecewise)

m	1		2		3		4	
n	2 points	4 points	2 points	4 points	2 points	4 points	2points	4 points
1	1.70711	1.92388	0.25000	0.42678	0.14226	0.20152	0.01042	0.03107
2	1.85355	1.96194	0.48928	0.49928	0.18698	0.19733	0.03990	0.04155
4	1.92678	1.98097	0.59732	0.49982	0.19448	0.19748	0.04122	0.04164
5	1.94142	1.98478	0.49000	0.49707	0.19574	0.19750	0.04002	0.04179
10	1.97071	1.99239	0.49957	0.49997	0.19706	0.19752	0.04160	0.04166
20	1.98536	1.99619	0.49989	0.49999	0.19742	0.19753	0.04167	0.04167

TABLE (4.4) The behaviour of $||Q_n||$ (piecewise)

Problem	5		6		7		8	
n	2 points	4 points	2 points	4 points	2 points	4 points	2 points	4 points
1	0.3037	0.5064	0.1245	0.1634	0.0034	0.0046	0.0138	0.0315
2	0.5341	0.5661	0.1641	0.1637	0.0054	0.0053	0.0394	0.0392
4	0.5602	0.5666	0.1583	0.1626	0.0053	0.0053	0.0394	0.0392
5	0.5588	0.5645	0.1645	0.1634	0.0052	0.0053	0.0387	0.0392
10	0.5658	0.5668	0.1637	0.1640	0.0053	0.0053	0.0393	0.0393
20	0.5666	0.5668	0.1640	0.1640	0.0053	0.0053	0.0392	0.0392

We notice from these results and by testing the behaviour of $||Q_n||$ with other different problems that $||Q_n||$ settles down very quickly and gives a sufficient approximation for $||(G - T)^{-1}||$ with very small values of n. In comparing the rate of convergence of $||Q_n||$ with the residual $||r_n||$ and the last coefficient of the solution, $|C_n|$, we observed, that of $||Q_n||$ was always higher especially when the right hand side is less smooth. That behaviour is expected since the rate of convergence of the residual and the coefficients depend on the right hand side as well as the left hand side while $||Q_n||$ depends on the left hand side only.

This quick convergence of $||Q_n||$, even when the coefficients on the left handside are not smooth as in problem 5, relative to r_n supports the idea of taking $||Q_n^*||$ (n* small) as an estimate of $||(G - T)^{-1}||$ and avoid the expensive calculation of $||Q_n||$ for larger values of n.

4.5.3. Results for the global case

1. The behaviour of the residual

We mentioned in section (4.1) that the residual,

$$r_n = y - (G - T)\, x_n$$

is calculated by substituting x_n in the differential equation. In (4.4.1(3)) we described how to calculate the principal part of the residual

$$R_n = T_n \sum_{k=0}^{n-1} a_k T_k \quad .$$

Then the modified residual Ψ_n is simply calculated by subtracting R_n from r_n. The behaviour of the residual which was mentioned to be like the remainder of the interpolation of the function $(y + T x_n)$ is considered by examining the

values of $||r_n||$ and $||\Psi_n||$ with different problems.

In table (4.5) problem 1 is considered with the smooth right hand side y = 1. We observe that $||\Psi_n|| = 0$ for all n> 4. That can be seen since $y + T x_n = 1 + (1 + t^2) x_n$ is a polynomial of degree n + 3 and R_n is a polynomial of degree 2n - 1.

In table (4.6) problem 1 is considered with the singular right hand side $y = \sqrt{|(t - 0.9)|}$. Here we notice the slow convergence of $||r_n||$ and that $||\Psi_n||$ is not much smaller than $||r_n||$ with the all given values of n.

Problem 2 is considered first with y = cosh(1) in Table (4.7) and then with $y = \frac{1}{t^2 + .1}$, which is nearly singular at zero, in table (4.8). The notable thing here is that in the second case when n becomes sufficiently large (n>8) then r_n starts to settle down and $||\Psi_n||$ tends very quickly to zero. This behaviour is because y is an analytic function.

In tables (4.9) and (4.10) the function $y + T x_n$ is smooth and $||R_n||$ gives a good approximation of the residual very early. In table (4.10) where y is less smooth (has discontinuities in the third derivatives) we observe that the convergence is slower but quicker than that shown in tables(4.6) and (4.8).

In table (4.12) problem 5 which has got a discontinuous coefficient on the left hand side is considered with smooth right hand side y = 1. We notice here that the residual is behaving better than that in table (4.6) which confirms that the right hand side is dominant in the residual behaviour.

N.B. $||r_n||$, $||R_n||$, $||G^{-1}R_n||$, $||K R_n||$ and $||\Psi_n||$ are evaluated using 200 equally spaced points.

TABLE (4.5) The behaviour of the residual

Problem 1 $y = 1$

n	$\|\|r_n\|\|$	$\|\|\Psi_n\|\|$	$\|\|G^{-1}R_n\|\|$	$\|\|K\ R_n\|\|$
4	0.01086	0	0.0119	0.0119
6	0.0030	0	7.5×10^{-5}	9.4×10^{-5}
7	0.0004	0	5.3×10^{-6}	7.4×10^{-6}
8	0.0002	0	3.8×10^{-6}	3.8×10^{-6}
11	2.4×10^{-7}	0	1.3×10^{-9}	1.3×10^{-9}
13	4.0×10^{-8}	0	1.2×10^{-11}	1.2×10^{-11}
18	5.6×10^{-13}	0	1.7×10^{-15}	1.7×10^{-15}
23	2.3×10^{-15}	0	0	0

TABLE (4.6)

Problem 1 $y = \sqrt{|t - 0.9|}$

n	$\|\|r_n\|\|$	$\|\|\Psi_n\|\|$	$\|\|G^{-1}R_n\|\|$	$\|\|K\ R_n\|\|$
4	0.1979	0.1539	0.03196	0.03145
6	0.2958	0.1879	0.0057	0.0007
7	0.2712	0.0919	0.0143	0.0174
8	0.2267	0.1591	0.0039	0.0046
11	0.1056	0.0984	0.004	0.0048
13	0.1909	0.1212	0.0018	0.0022
18	0.0835	0.0797	0.0018	0.0022
23	0.1059	0.1004	0.0003	0.0004

TABLE (4.7) The behaviour of the residual

Problem (2) $y = \cosh(1)$

n	$\|\|r_n\|\|$	$\|\|\Psi_n\|\|$	$\|\|G^{-1} R_n\|\|$	$\|\|K R_n\|\|$
4	5.7×10^{-3}	0	5.7×10^{-4}	5.7×10^{-4}
6	4.5×10^{-5}	0	1.3×10^{-6}	1.3×10^{-6}
7	3.9×10^{-7}	0	4.92×10^{-9}	4.92×10^{-9}
8	2×10^{-7}	0	3.5×10^{-9}	3.5×10^{-9}
11	2×10^{-12}	0	7.3×10^{-15}	7.3×10^{-15}
13	3.9×10^{-15}	0	0	0
23	0	0	0	0

TABLE (4.8) Problem (2) $y = \frac{1}{t^2 + 0.1}$

n	$\|\|r_n\|\|$	$\|\|\Psi_n\|\|$	$\|\|G^{-1}R_n\|\|$	$\|\|K R_n\|\|$
4	5.289	1.21	0.926	0.926
6	3.009	0.3873	0.296	0.296
7	1.093	0.2015	0.1387	0.1387
8	1.645	0.116	0.0947	0.0947
11	0.3126	0.0175	0.0137	0.0137
13	0.173	0.00517	0.0043	0.0043
18	0.0738	2.2×10^{-4}	4.2×10^{-4}	4.2×10^{-4}
23	7.7×10^{-3}	1×10^{-5}	1.77×10^{-5}	1.77×10^{-5}

TABLE (4.9) The behaviour of the residual

Problem (3) $y = \frac{1}{2(t+5)}$

n	$\|\|r_n\|\|$	$\|\|\Psi_n\|\|$	$\|\|G^{-1}R_n\|\|$	$\|\|K R_n\|\|$
4	1.3×10^{-4}	5.6×10^{-8}	9×10^{-6}	7.2×10^{-7}
6	2.5×10^{-7}	1.1×10^{-11}	5.8×10^{-8}	5.7×10^{-9}
7	3.2×10^{-7}	1.5×10^{-13}	5.2×10^{-9}	4.6×10^{-10}
8	4×10^{-8}	2×10^{-15}	5.4×10^{-10}	4.2×10^{-11}
11	7.1×10^{-11}	0	4.7×10^{-13}	4.2×10^{-14}
13	9.7×10^{-13}	0	4.7×10^{-15}	4×10^{-16}
18	0	0	0	0
23	0	0	0	0

TABLE (4.10)

Problem (3) $y = \begin{cases} t^3 + \sin t + 2 & t < 0 \\ (2-t)e^t & t \geq 0 \end{cases}$

n	$\|\|r_n\|\|$	$\|\|\Psi_n\|\|$	$\|\|G^{-1}R_n\|\|$	$\|\|K R_n\|\|$
4	0.0556	0.0224	0.007	0.00057
6	7.4×10^{-3}	5.3×10^{-4}	5.5×10^{-4}	4.4×10^{-5}
7	2×10^{-3}	3.4×10^{-4}	2×10^{-4}	1.7×10^{-5}
8	2.8×10^{-3}	1.9×10^{-4}	1.4×10^{-4}	1.2×10^{-5}
11	4.4×10^{-4}	7.1×10^{-5}	2.9×10^{-5}	2.3×10^{-6}
13	2.6×10^{-4}	4.1×10^{-5}	1.5×10^{-5}	1.2×10^{-6}
18	2.2×10^{-4}	1.3×10^{-5}	4.3×10^{-6}	3.4×10^{-7}
23	4.5×10^{-5}	6.6×10^{-6}	1.4×10^{-6}	1.1×10^{-7}

TABLE (4.11) The behaviour of the residual

Problem (4) $y = \frac{-1}{t+3}$

n	$\|\|r_n\|\|$	$\|\|\Psi_n\|\|$	$\|\|G^{-1}R_n\|\|$	$\|\|K\ R_n\|\|$
4	0.0243	4.2×10^{-5}	0.001	0.0026
6	0.00135	8×10^{-8}	2.9×10^{-5}	1.2×10^{-4}
7	5.2×10^{-4}	3.7×10^{-9}	4.5×10^{-6}	2.5×10^{-5}
8	7.6×10^{-5}	1.5×10^{-10}	8.4×10^{-7}	5.2×10^{-6}
11	8.2×10^{-7}	1.5×10^{-15}	4.6×10^{-9}	4.1×10^{-8}
13	3.6×10^{-8}	0	1.5×10^{-10}	1.5×10^{-9}
18	1.3×10^{-11}	0	2.7×10^{-14}	3.9×10^{-13}
23	0.7×10^{-15}	0	0	0

TABLE (4.12) Problem (5) $y = 1$

n	$\|\|r_n\|\|$	$\|\|\Psi_n\|\|$	$\|\|G^{-1}\ R_n\|\|$	$\|\|K\ R_n\|\|$
4	0.2175	0.0396	0.0352	0.0088
6	0.1081	0.0209	0.0114	0.0026
7	0.0497	0.0274	0.0068	0.002
8	0.0757	0.0167	0.006	0.0014
11	0.031	0.0167	0.0028	0.0007
13	0.0262	0.0137	0.002	0.0005
18	0.0319	0.0084	0.0011	0.0003
23	0.0147	0.007	0.00064	0.00017

2. The behaviour of $||G^{-1}R_n||$ and $||K R_n||$

With the values of $||r_n||$ and $||\Psi_n||$ the corresponding values of $||G^{-1}R_n||$ and $||K R_n||$ are also given. In comparing $||G^{-1}R_n||$ and $||K R_n||$ with $||R_n||$ observe thefollowing:

(i) with all values of n even small ones $||G^{-1}R_n|| < ||R_n||$ and when n is sufficiently large (depending on $||K||$) $||K R_n|| < ||R_n||$.

(ii) When n becomes larger $||G^{-1}R_n|| << ||R_n||$ and $||K R_n|| << ||R_n||$ because in such case we will have higher oscillation on R_n and hence more cancellations will occur in the evaluation of $G^{-1}R_n$ and $K R_n$.

We also observe that $G^{-1}R_n$ and $K R_n$ have a similar rate of convergence. That can be seen by definition of $G^{-1}R_n$ and $K R_n$ since the coefficients of the equation and Green's function and its derivations are bounded.

3. Estimates of the Error bound

The estimates derived in section (4.3) are

(1) $e_Q = ||Q_n^*||\ ||r_n||$

(2) $e_1^* = ||G^{-1}R_n|| + ||Q_n^*|| (||K R_n|| + ||\Psi_n||)$

(3) $e_2^* = ||G^{-1}R_n|| + ||Q_n^*||\ ||K R_n||$

(4) $E_1 = ||G^{-1}R_n + (G - \phi_n T)^{-1}\phi_n K R_n|| + ||Q_n^*||\ ||\Psi_n||$

(5) $||E_2|| = ||G^{-1}R_n + (G - \phi_n T)^{-1}\ \phi_n\ K R_n||$.

These estimates are compared with the actual error e_n in tables (4.13) to (4.17). We note here that when the actual solution is unknown a solution with n = 40 is used in estimating e_n. $||e_n||$and $||E_2||$ are evaluated using 200 equally spaced points.

The error estimates

TABLE (4.13) Problem 1 $y = 1$

n	e_Q	$e_1^* = e_2^*$	$E_1 = \|\|E_2\|\|$	$\|\|e_n\|\|$
4	0.1012	0.023	0.0173	0.0168
6	2.8×10^{-3}	1.6×10^{-4}	8.9×10^{-5}	9.1×10^{-5}
7	3.7×10^{-4}	1.2×10^{-5}	5.9×10^{-6}	6×10^{-6}
8	1.8×10^{-4}	7.3×10^{-6}	4×10^{-6}	4×10^{-6}
11	2.6×10^{-7}	2.2×10^{-9}	1.1×10^{-9}	1.1×10^{-9}
13	3.7×10^{-8}	2×10^{-11}	1×10^{-11}	1×10^{-11}
18	5.2×10^{-13}	3.1×10^{-15}	1.6×10^{-15}	2.4×10^{-15}

TABLE (4.14) Problem 1 $y = \sqrt{|(t-0.9)|}$

n	e_Q	e_1^*	e_2^*	E_1	$\|\|E_2\|\|$	$\|\|e_n\|\|$
4	0.1845	0.2047	0.0602	0.1965	0.0531	0.0177
6	0.2782	0.18145	0.0122	0.1839	0.0088	0.0022
7	0.2528	0.1162	0.0305	0.1094	0.0237	0.012
8	0.2113	0.1525	0.0081	0.1549	0.0067	0.0005
11	0.0984	0.0961	0.0082	0.0984	0.0067	0.00025
13	0.1779	0.1150	0.0039	0.1159	0.0029	0.0007
18	0.0778	0.0763	0.0039	0.0773	0.0030	7.78×10^{-5}
23	0.0988	0.0943	0.00067	0.0940	0.0005	1.3×10^{-4}

The error estimates

TABLE (4.15) Problem 2 $y = \frac{1}{t^2+0.1}$

n	e_0	e_1^*	e_2^*	E_1	$\|\|E_2\|\|$	$\|\|e_n\|\|$
4	1.86	1.678	1.2422	1.125	0.707	0.606
6	1.0557	0.536	0.3997	0.366	0.2297	0.212
7	0.382	0.259	0.1881	0.174	0.103	0.0984
8	0.5789	0.1688	0.1280	0.1159	0.0751	0.0713
11	0.1065	0.0246	0.0184	0.0124	0.0106	0.0101
13	0.0609	7.6x10-3	5.8×10^{-3}	0.0042	0.0034	0.0032
18	0.0260	6.45×10^{-4}	5.6×10^{-4}	4.6×10^{-4}	3.9×10^{-4}	3.8×10^{-4}
23	2.7×10^{-3}	2.72×10^{-5}	2.32×10^{-5}	1.98×10^{-5}	1.63×10^{-5}	1.6×10^{-5}

TABLE (4.16) Problem 3 $y = \begin{cases} t^3 + \sin t + 2 & \text{if } t < 0 \\ (2-t)\, e^t & \text{if } t \geq 0 \end{cases}$

n	e_0	e_1^*	e_2^*	E_1	$\|\|E_2\|\|$	$\|\|e_n\|\|$
4	0.0269	0.0181	0.0073	0.0177	0.0069	0.0063
6	3.6×10^{-3}	8.2×10^{-4}	5.6×10^{-4}	7.9×10^{-4}	5.4×10^{-4}	4.9×10^{-4}
7	9.7×10^{-4}	3.7×10^{-4}	2.1×10^{-4}	3.6×10^{-4}	2×10^{-4}	2×10^{-4}
8	1.3×10^{-3}	2.4×10^{-4}	1.5×10^{-4}	2.3×10^{-4}	1.4×10^{-4}	1.3×10^{-4}
11	2.13×10^{-4}	5×10^{-5}	1.5×10^{-5}	5.9×10^{-5}	2.9×10^{-5}	3×10^{-5}
13	1.26×10^{-4}	1.75×10^{-5}	1.55×10^{-5}	1.6×10^{-5}	1.4×10^{-5}	1.5×10^{-5}
18	1.06×10^{-4}	1.06×10^{-5}	4.6×10^{-6}	1×10^{-5}	4.2×10^{-6}	3.9×10^{-6}
23	2.2×10^{-5}	4.6×10^{-6}	1.3×10^{-6}	3.7×10^{-6}	1.4×10^{-6}	1.5×10^{-6}

The error estimates

TABLE (4.17) Problem 5 y = 1

n	e_Q	e_1^*	e_2^*	E_1	$\|\|E_2\|\|$	$\|\|e_n\|\|$
4	0.1232	0.0625	0.0401	0.0627	0.0403	0.0284
6	0.0611	0.0244	0.0126	0.0244	0.0126	0.0086
7	0.028	0.0234	0.0079	0.0228	0.0073	0.0099
8	0.0428	0.0162	0.0078	0.0161	0.0066	0.0044
11	0.0175	0.0126	0.0032	0.0124	0.003	0.0041
13	0.0148	0.0100	0.0023	0.0100	0.0022	0.003
18	0.0181	0.0061	0.0013	0.0060	0.0012	0.0007
28	0.0115	0.0037	0.0005	0.0037	0.0005	0.000_2

(1) In table (4.13) (problem 1, y = 1) we observe

(i) $e_1^* \equiv e_2^*$ and $E_1 \equiv ||E_2||$. That is obvious because $||\Psi_n|| \equiv 0$ as shown in Table (4.5).

(ii) e_1^* becomes closer to the actual error when n increases while E_1 gives very close results even with small value of n. This is justified since E_1 in this case gives estimation of the error itself while e_1^* is an estimate of the bound which becomes closer to the actual error when $K R_n$ becomes very small.

(iii) e_Q is not as close as e_1^* but is not that far away.

(iv) We note that at n = 13 E_1 is exactly equal to $||e_n||$ while for n = 18 becomes a bit different. This may be due to round off error.

(2) In table (4.14) (problem 1, $y = \sqrt{|(x+0.9)|}$, we observe

(i) e_Q is dominated by the residual and e_1^* and E_1 are dominated by the modified residual $||\Psi_n||$ which is like the residual in this case (refer to table (4.6)).

(ii) e^*_2 and E_2 are not that close but they are acceptable. These poor results are expected with this example since the residual is not sufficiently small and R_n is not dominant. However that is guaranteed when n is sufficiently large as shown in section (4.2) and then we expect these estimates to be much closer to the actual error.

(3) Problems2,3 and 5 are considered with different right hand sides in tables (4.15), (4.16) and (4.17). We can see that when the residual becomes sufficiently small e_Q becomes closer to the actual

error while the estimates e_2^* and E_2 becomes closer only when R_n gives a good approximation of r_n.

(4) In comparing the amount of work it is very difficult to form firm conclusions, because many factors are involved. However in general one expects e_1^* and E_1 to be the most expensive since they require the evaluation of the residual or the modified residual. In comparing e_2^* with E_2, E_2 is the cheapest since e_2^* still requires the evaluation of the coefficients of the differential equation in evaluating $K R_n$.

Finally these estimates were compared with the estimates given by Cruickshanck (1974) and McKeown (1978) and we found that even e_Q is superior to them. However with more analysis these estimates might be improved even more.

4.5.4. The piecewise case

1. Numerical results

Here we expect better results than for the global case when $y + Tx_n$ is less smooth due to the superiority of piecewise interpolation in such cases. That may be seen if we consider the less smooth case (problem 2, $y = \frac{1}{t^2 + 0.1}$) and the smooth one (problem 4, $y = \frac{1}{t+3}$). In the first case we may notice that with the piecewise case (table 4.18) $||\Psi_n||$ takes relatively small values with small n compared with the global case (table 4.8), while in the second case $||\Psi_n|| \to 0$ more quickly with the global case (see table (4.11) and table (4.19)).

If we consider the estimates in table (4.18) and (4.19) we see that they behave in general like the global case. We observe that e_Q gives close results when the residual is sufficiently small,

The piecewise case

TABLE (4.18) Problem 2, $y = \frac{1}{t^2+0.1}$

Interp.	n	partition	$\|\|r_n\|\|$	$\|\|\Psi_n\|\|$	$\|\|G^{-1}R_n\|\|$	$\|\|K R_n\|\|$	e_Q	e_1^*	e_2^*	E_1	$\|\|E_2\|\|$	$\|\|e_n\|\|$
	3	1	0.1457	0.0011	0.0489	0.0489	0.0513	0.0665	0.0661	0.0336	0.0332	0.0479
3	4	2	0.5221	0.0161	0.0244	0.0244	0.1836	0.0387	0.0331	0.0242	0.0186	0.0197
Tcheby.	6	3	0.2209	0.0044	0.0047	0.0047	0.0777	0.0079	0.0064	0.0053	0.0038	0.0041
points	8	4	0.1066	0.00092	0.00095	0.00095	0.0375	0.0016	0.0013	0.0011	0.00082	0.00087

TABLE (4.19) Problem 4, $y = -\frac{1}{(t+3)}$

Interp.	n	partition	$\|\|r_n\|\|$	$\|\|\Psi_n\|\|$	$\|\|G^{-1}R_n\|\|$	$\|\|KR_n\|\|$	e_Q	e_1^*	e_2^*	E_1	$\|\|E_2\|\|$	$\|\|e_n\|\|$
	3	1	0.0082	8.7×10^{-6}	0.00015	0.00074	0.0037	0.00039	0.00039	0.00017	0.00017	0.00017
3 Gauss	4	1	0.0042	2.2×10^{-6}	5.2×10^{-5}	3.2×10^{-4}	0.0019	1.9×10^{-4}	1.9×10^{-4}	5.8×10^{-5}	5.8×10^{-5}	5.5×10^{-5}
points	6	1	0.0015	2.6×10^{-7}	9.7×10^{-6}	8.5×10^{-5}	6.9×10^{-4}	4.8×10^{-5}	4.8×10^{-5}	1.07×10^{-5}	1.07×10^{-5}	1.02×10^{-5}
	8	1	7×10^{-4}	5.7×10^{-8}	2.7×10^{-6}	3.2×10^{-5}	3.2×10^{-4}	1.7×10^{-5}	1.7×10^{-5}	3.04×10^{-6}	3.04×10^{-6}	2.9×10^{-6}

while e_2^* and E_2 become closer when the residual is well approximated by its principal part. It is also notable that E_2 is the closest in most cases.

We should note here that in the calculations of these estimates a lot of work is needed with the piecewise case since R_n needs to be calculated at every partition.

For further investigation two more problems are examined.

Problem 9

$$\varepsilon\, x''(t) - (2 - t^2)\, x(t) = -1, \quad x(\pm 1) = 0 \qquad \varepsilon = 10^{-4}$$

This problem is taken from Russell and Shampine (1972). The solution is symmetric about zero and it has a boundary layer of width $\sqrt{\varepsilon}$ at 1.

Problem 10

$$x''(t) + 2\gamma t\, x'(t) + 2\gamma(t) = 0 \;, \quad x(0) = 0 \quad x(1) = e^{-\gamma} \quad \gamma = 1$$

This problem is taken from Russell and Christiansen (1976). The solution is $e^{-\gamma x^2}$ which is well behaved when $\gamma = 1$. Large values of γ will be considered later.

These problems are considered with both Tchebychev and Gauss points. We notice in table (4.20) (Problem 9) that the modified residual is very small relative to the residual which takes very large values. With small n(n=4) the estimates with both Tchebychev and Gauss points are very poor, but when n becomes larger these estimates start to settle down and E_2 becomes very close to the actual error. The results for problem (10) which is a smooth one, are given in table (4.21). We observe there the estimates give very close results even with small values

The piecewise case

TABLE (4.20) $10^{-4}\, x'' - (2-t^2)x = -1$

Interpol.	n	partition	$\|\|r_n\|\|$	$\|\|\Psi_n\|\|$	$\|\|G^{-1}R_n\|\|$	$\|\|K\,R_n\|\|$	e_Q	e_1^*	E_1	$\|\|e_n\|\|$
	4	4	8262.89	15.34	18.36	209879.9	0.7154	36.5	4.64	0.5914
	8	8	5395.14	2.3342	2.7420	29378.56	0.5023	5.47	0.9967	0.3276
3 Tcheby-	16	16	2184.431	0.1980	0.2414	2706.16	0.2130	0.5053	0.15	0.0906
chev points	32	32	569.9	0.0093	0.0256	271.48	0.0520	0.0503	0.0127	0.0134
3	4	4	8825.58	16.27	12.49	14517.5	0.7888	25.46	2.1457	0.2685
Gauss	8	8	6483.5	2.8202	2.16	23469.4	0.5922	4.307	0.5029	0.1607
points	16	16	3054.53	0.2806	0.2270	2364.5	0.2831	0.44	0.0921	0.0512
	32	32	871.88	0.0144	0.0135	137.74	0.0792	0.0260	0.0091	0.0098

The piecewise case

TABLE (4.21) $x'' + 2t\,x' + 2x = 0$

3 points	n	partit-ion	$\|\|r_n\|\|$	$\|\|\Psi_n\|\|$	$\|\|G^{-1}R_n\|\|$	$\|\|K\,R_n\|\|$	e_Q	e_1^*	e_2^*	e_1^{**}	e_2^{**}	$\|\|e_n\|\|$
Tcheby-chev	4	3	0.0022	0	8.56×10^{-6}	1.2×10^{-4}	3×10^{-4}	2.6×10^{-5}	2.6×10^{-5}	1.05×10^{-5}	1.05×10^{-5}	$9.4\times10^{-$
	8	5	3×10^{-4}	0	4.4×10^{-7}	8.05×10^{-6}	4.2×10^{-5}	1.56×10^{-6}	1.56×10^{-6}	1.05×10^{-7}	1.05×10^{-7}	$5.2\times10^{-$
	16	11	3.13×10^{-5}	0	2.6×10^{-8}	4.5×10^{-7}	4.37×10^{-6}	9.9×10^{-8}	9.9×10^{-8}	3.5×10^{-8}	3.5×10^{-8}	3×10^{-8}
	32	21	4.05×10^{-6}	0	1.5×10^{-9}	2.9×10^{-8}	5.7×10^{-7}	5.3×10^{-9}	5.3×10^{-9}	2.1×10^{-9}	2.1×10^{-9}	$1.7\times10^{-$
Gauss	4	2	0.0043	0	2.7×10^{-6}	5×10^{-5}	6×10^{-4}	7.7×10^{-6}	7.7×10^{-6}	2.8×10^{-6}	2.8×10^{-6}	$2.4\times10^{-$
	8	4	5.3×10^{-4}	0	7.8×10^{-8}	3.7×10^{-6}	7.5×10^{-5}	5.8×10^{-7}	5.8×10^{-7}	7.9×10^{-8}	7.9×10^{-8}	$7.6\times10^{-$
	16	7	6.7×10^{-5}	0	2.4×10^{-9}	2.1×10^{-7}	9.3×10^{-6}	3.1×10^{-8}	3.1×10^{-8}	2.3×10^{-9}	2.3×10^{-9}	$2.3\times10^{-$
	32	14	8.3×10^{-6}	0	7.4×10^{-11}	1.4×10^{-8}	1.16×10^{-6}	1.9×10^{-9}	1.9×10^{-9}	7.3×10^{-11}	7.3×10^{-11}	$7.2\times10^{-$

of n.

In comparing Gauss points with Tchebychev ones we may notice that the error is smaller and the estimates are closer in the Gauss case. This superiority of Gauss points confirms the result of De Boor and Swartz (1973).

Finally we may notice that with these examples as well as others when using the equally spaced partitions scheme, the maximum error always occurs in the same place. That suggests looking for a new dividing scheme which may use these estimates to make partitions with larger error having small sizes. That will be considered in the next Chapter.

2. Graphs for the residual, the estimates and the actual errors.

For close comparisons some graphs are plotted for the residual, the estimates and the actual errors in a sample of subintervals and in the whole interval with both Tchebychev and Gauss points.

The problems considered are problem (10) and the following problem.

Problem (11) $x''(t) + (3\cot(t) + 2\tan(t))x'(t) + 2x(t) = 0$

$x(30^o) = 4 \qquad x(60^o) = 4/3$

This problem from Russell and Shampine (1972) and it has the solution $\frac{1}{\sin^2(t)}$.

In studying these graphs we may notice the following.

FIRSTLY THE RESIDUAL

(i) The choice of the collocation points is well reflected in the behaviour of the residual since it is like the interpolation error of $y + Tx_n$, with the collocation points taken as interpolation points. With Tchebychev points we observe the minimax property while the minimum over squares is observed with Gauss points.

(ii) The residual gives a very close estimate to the error in the highest derivative of the solution. That is expected when $||K\, r_{np}||$ is sufficiently small which is often the case.

(iii) We may also notice the discontinuity of the residual at the joining points. That is expected since we haven't assumed continuity on the highest derivative of the solution.

SECONDLY THE ERROR

(i) The error takes very small values at the end points of the partition with Gauss points. We will show that the error there exactly

equals $((G - T)^{-1}\Psi)_i$ as follows.

Recall (4.14)

$$e_i = (G^{-1}R)_i + ((G - T)^{-1} K R)_i + ((G - T)^{-1} \Psi)_i \cdot$$

At the end points the discontinuity of Green's function will be removed and (4.15) becomes

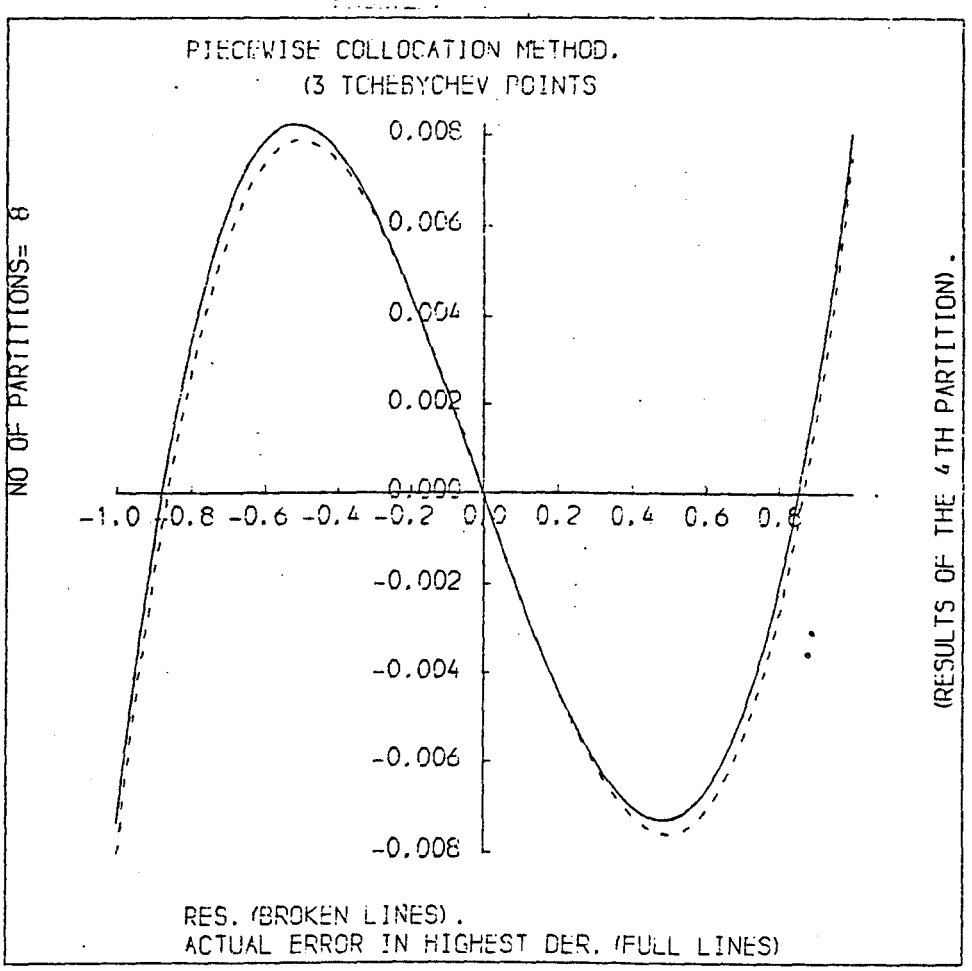
$$(G^{-1}R)_i (1) = \sum_{j=1}^{n} \int_{-1}^{1} g_{ij}(1,t) R_i(t)\, dt .$$

But if $R_i(t)$ is a polynomial and of degree $2p-m-1$ and $g(1,t)$ is a polynomial of degree $< m$ by Gaussian quadrature $G^{-1}R_i(1)$ will become zero. In a similar way it can be shown that $(K R_i)(s)$ defined in (4.16) is zero at 1. This makes the error at end points $\leq ||(G - T)^{-1}||\ \ ||\Psi_i||$.

This property is not observed with the Tchebychev case since the weight function there is $(1 - t^2)^{-\frac{1}{2}}$ but we may notice that the error is still smaller at the end points than in the middle of the sub-range. That follows since the integral of a Tchebychev polynomial is of order $\frac{1}{r}$ at the middle points while is of order $\frac{1}{r^2}$ at the end points (where r is the degree of the polynomial).

(ii) The estimates E_2 and $GR = (G^{-1}R)_i$ are very close to the actual error and are closer with Gauss points as expected.

Fig. (4.4) Problem (10) The residual



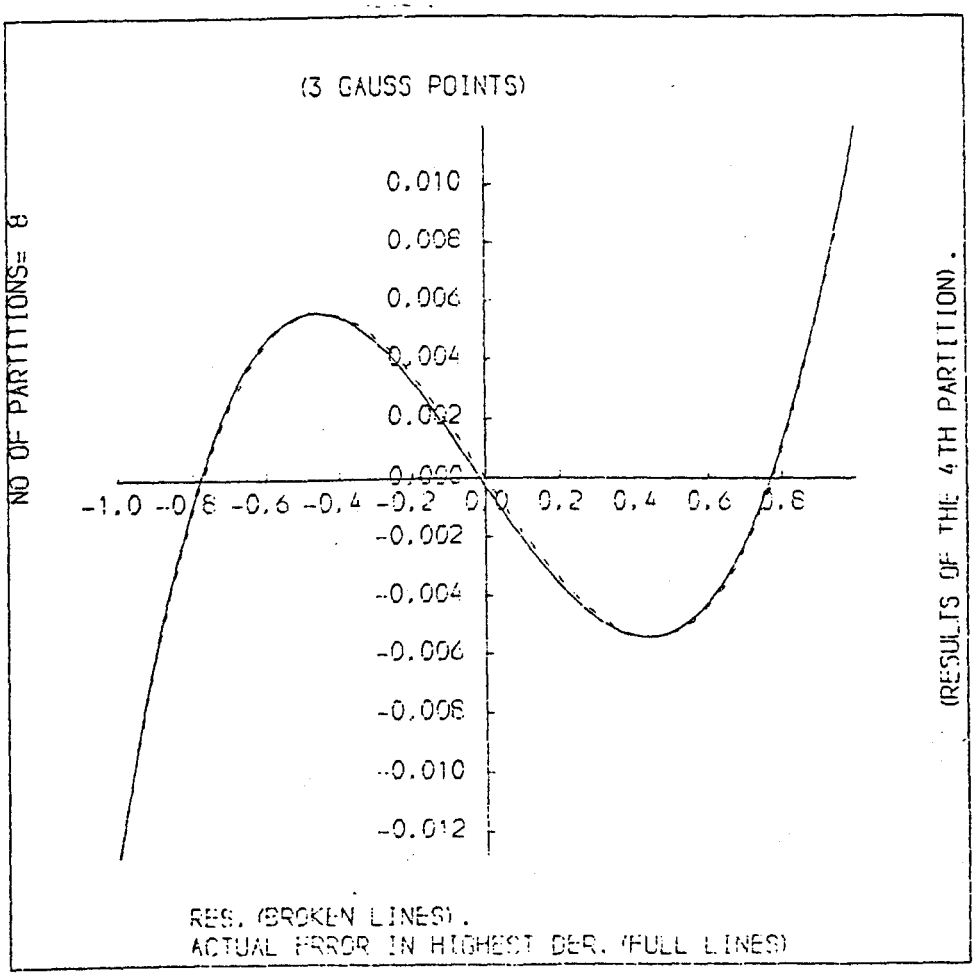
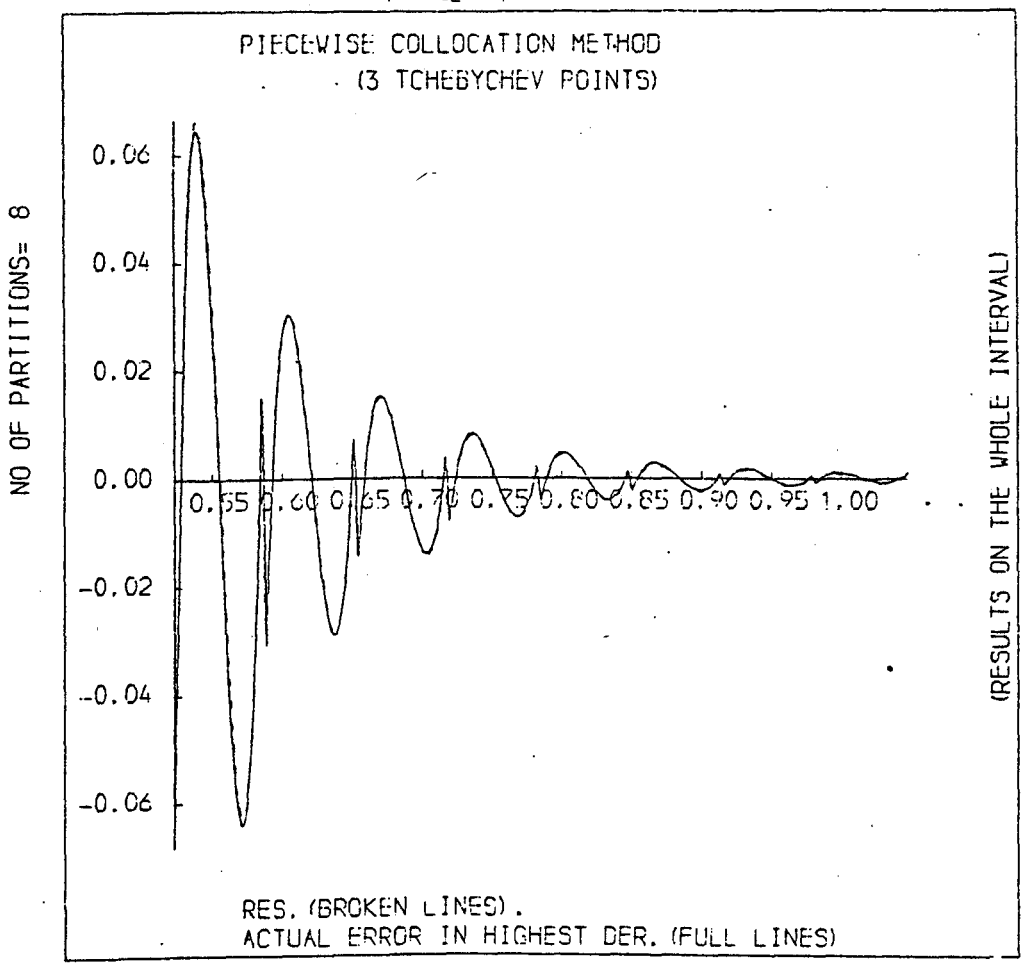


Fig. (4.5) Problem (10) The residual



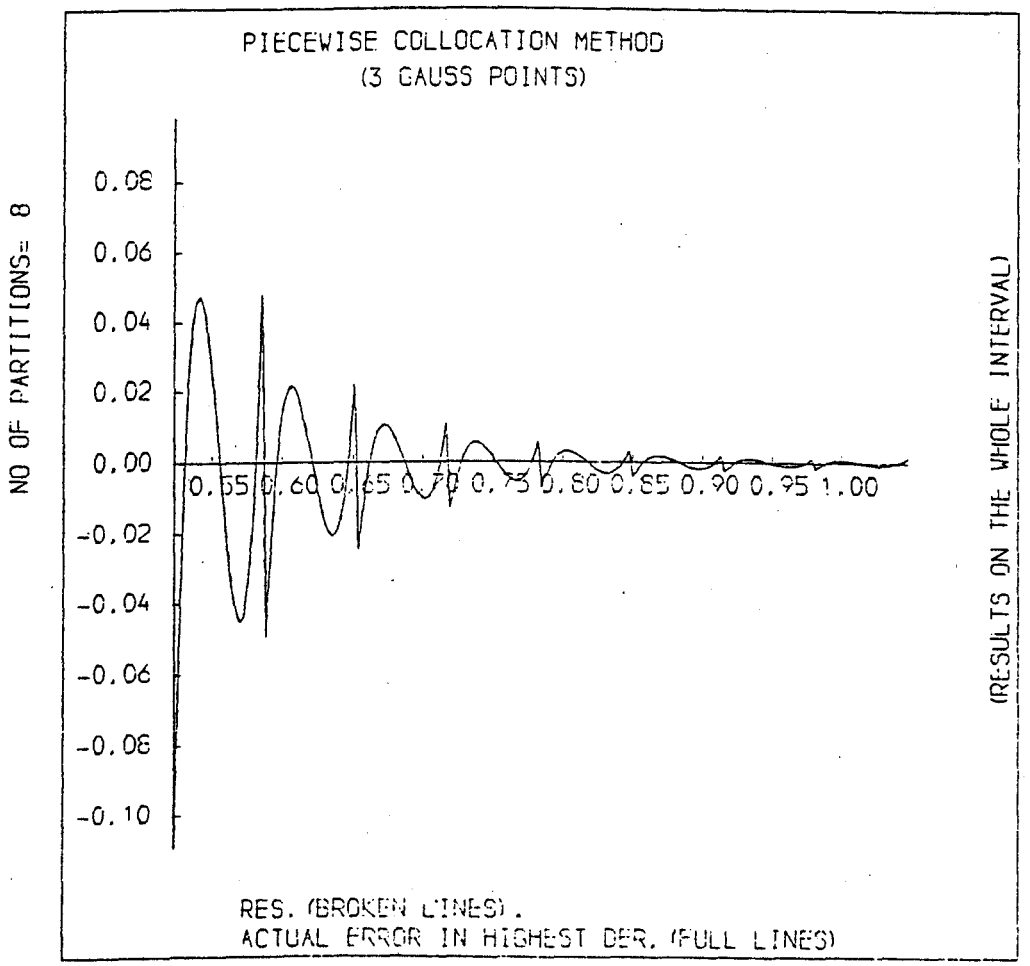
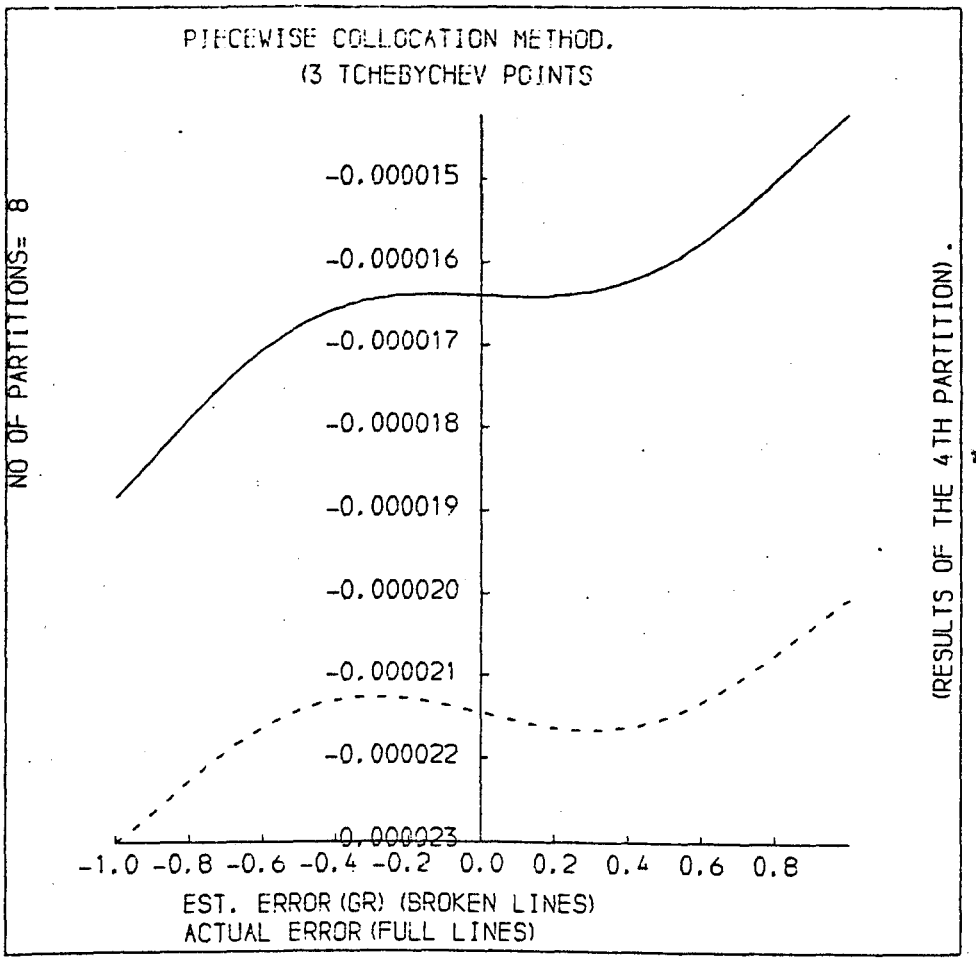


Fig. (4.6) Problem (10) The GR estimate



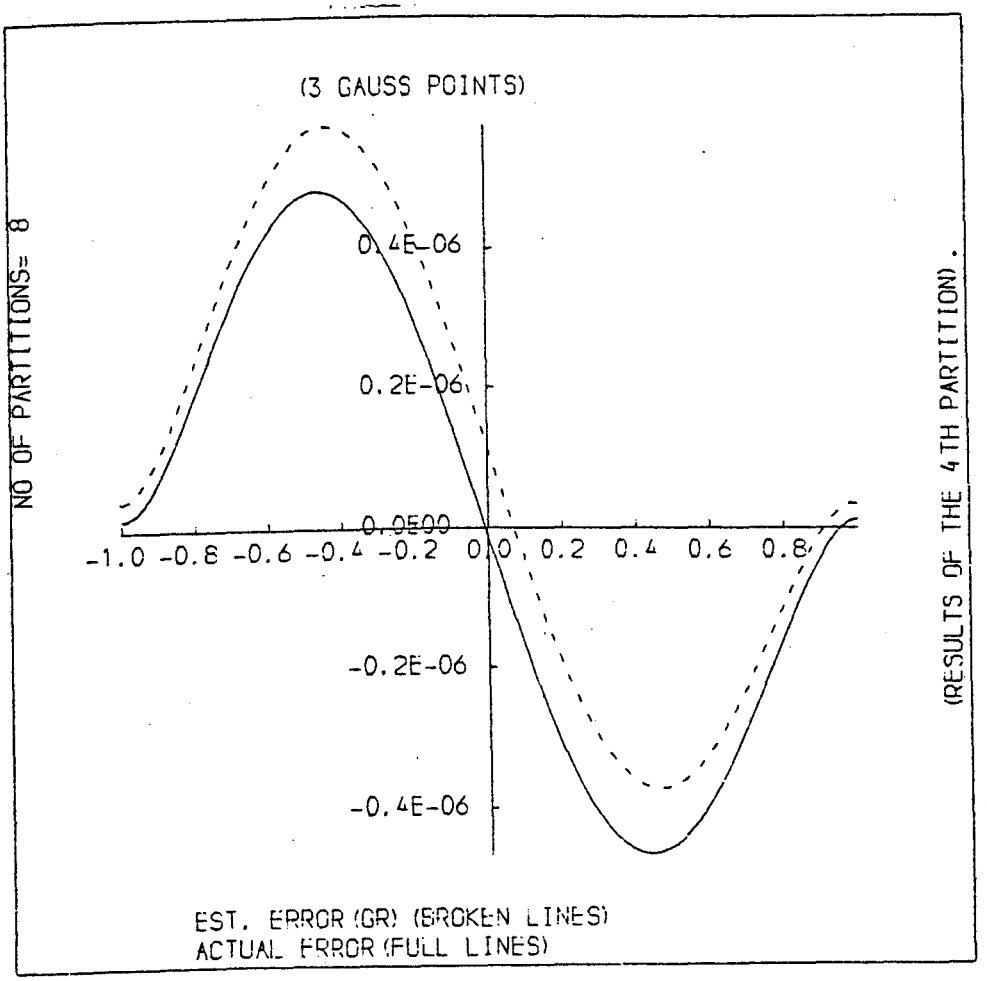
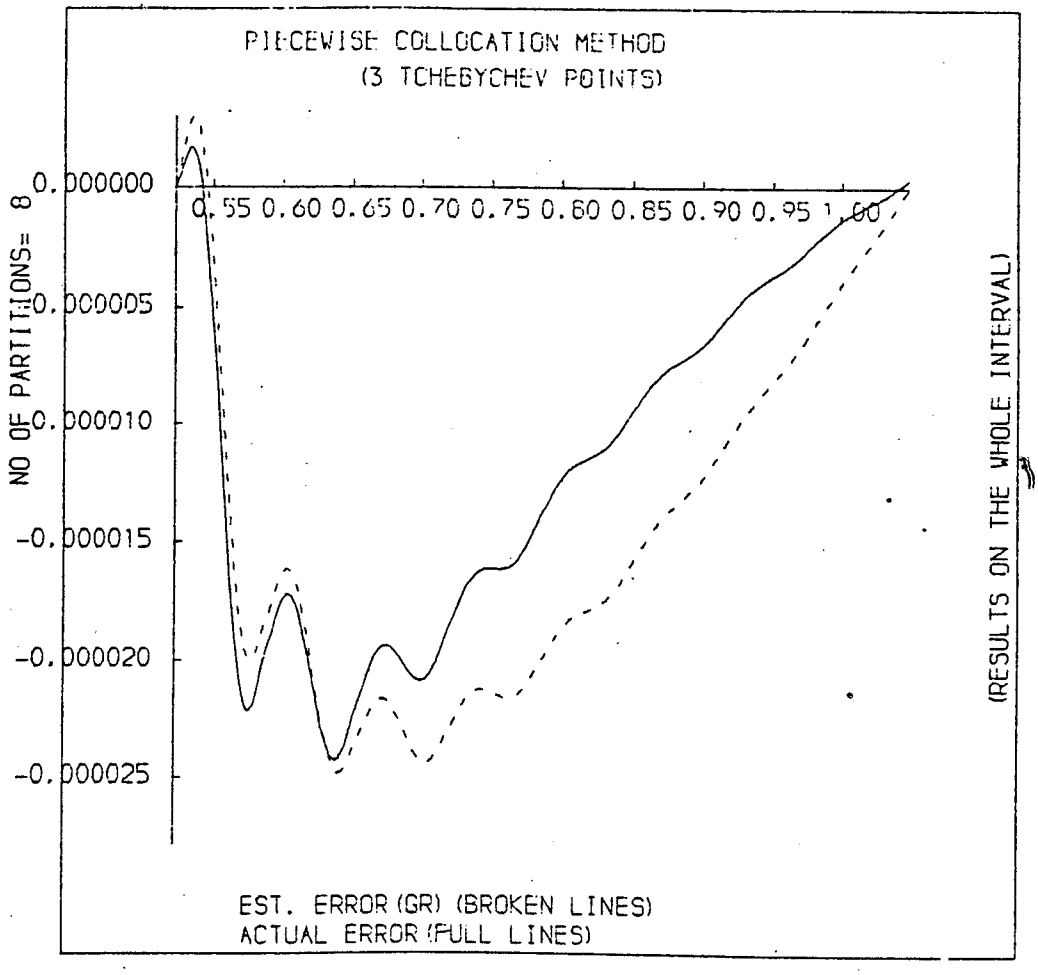


Fig (4.7) Problem (10) The GR estimate



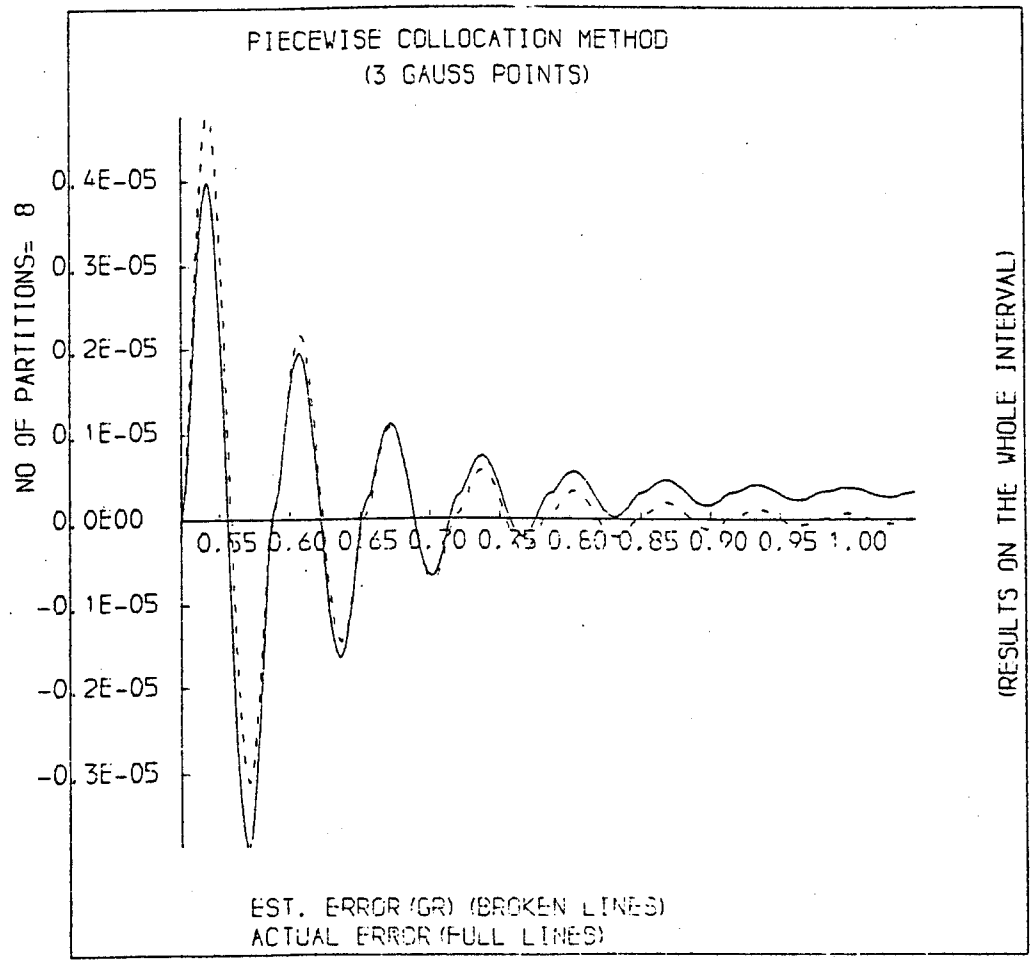
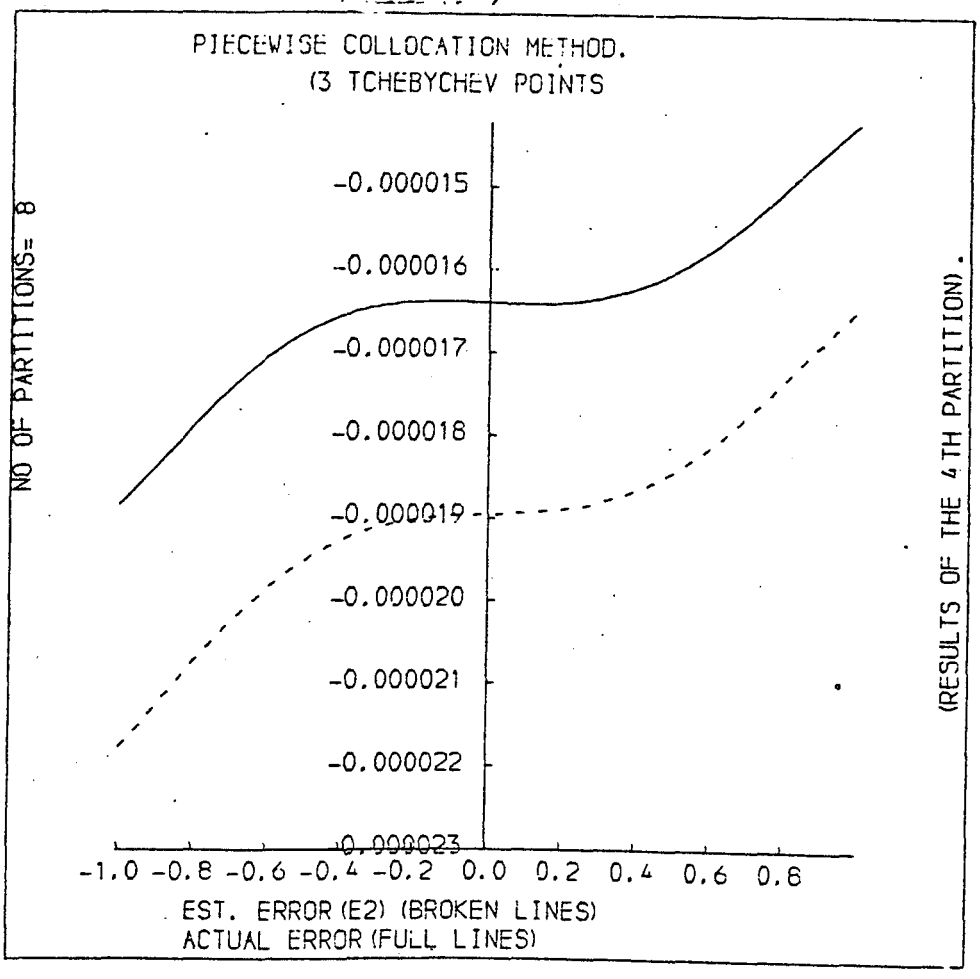


Fig. (4.8) Problem (10) The E1 estimate



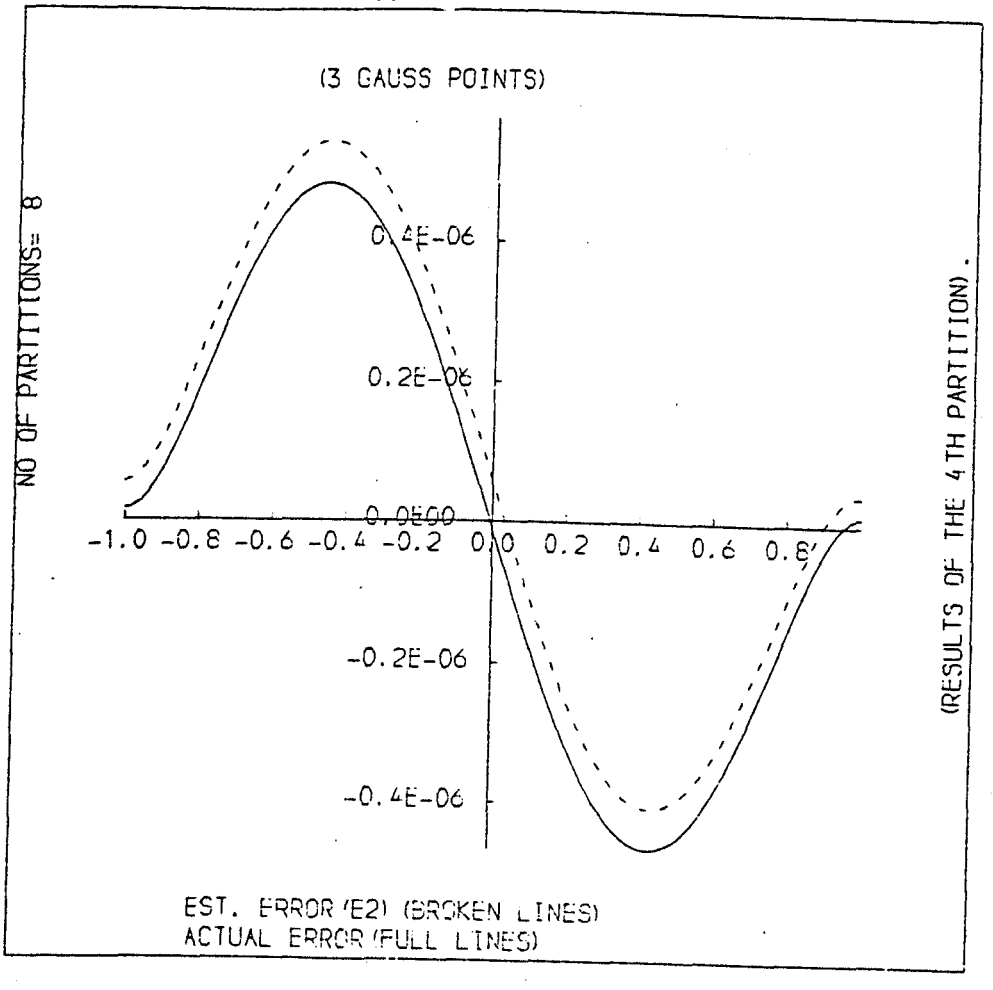
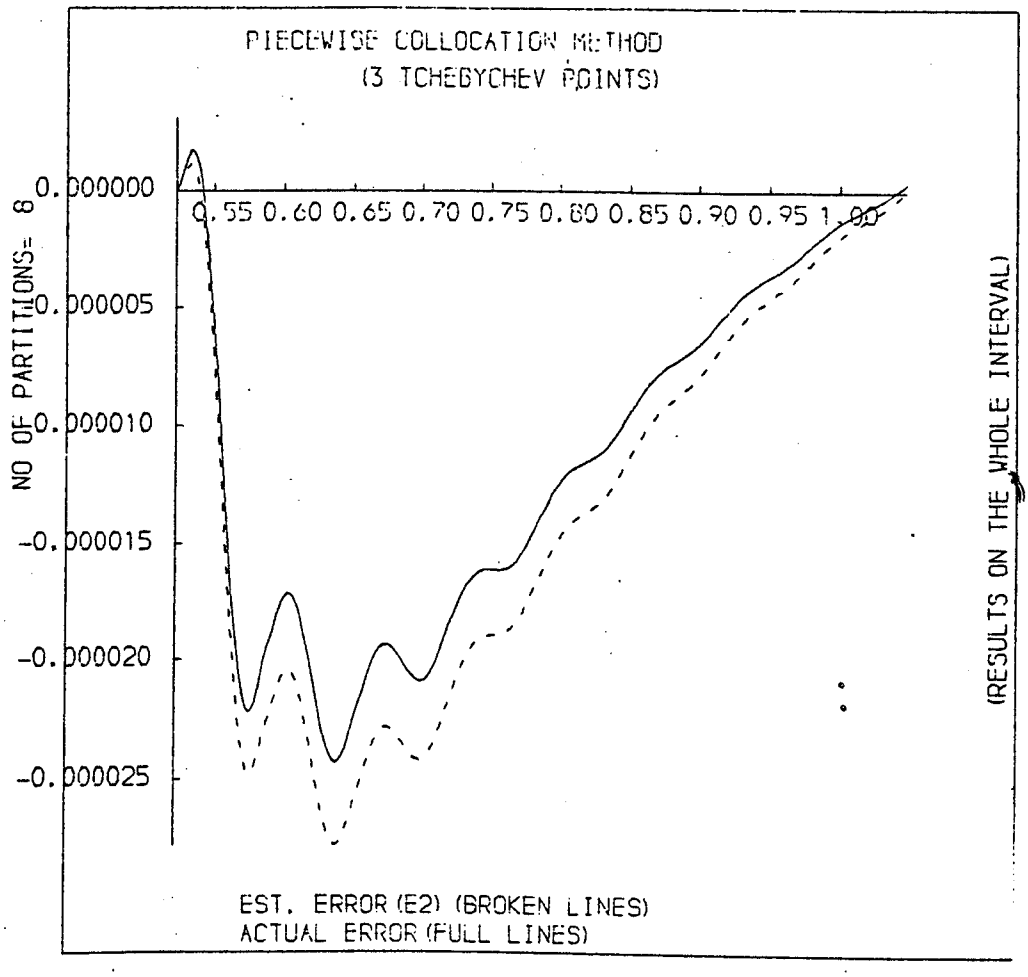


Fig. (4.9) Problem (10), The E2 estimate



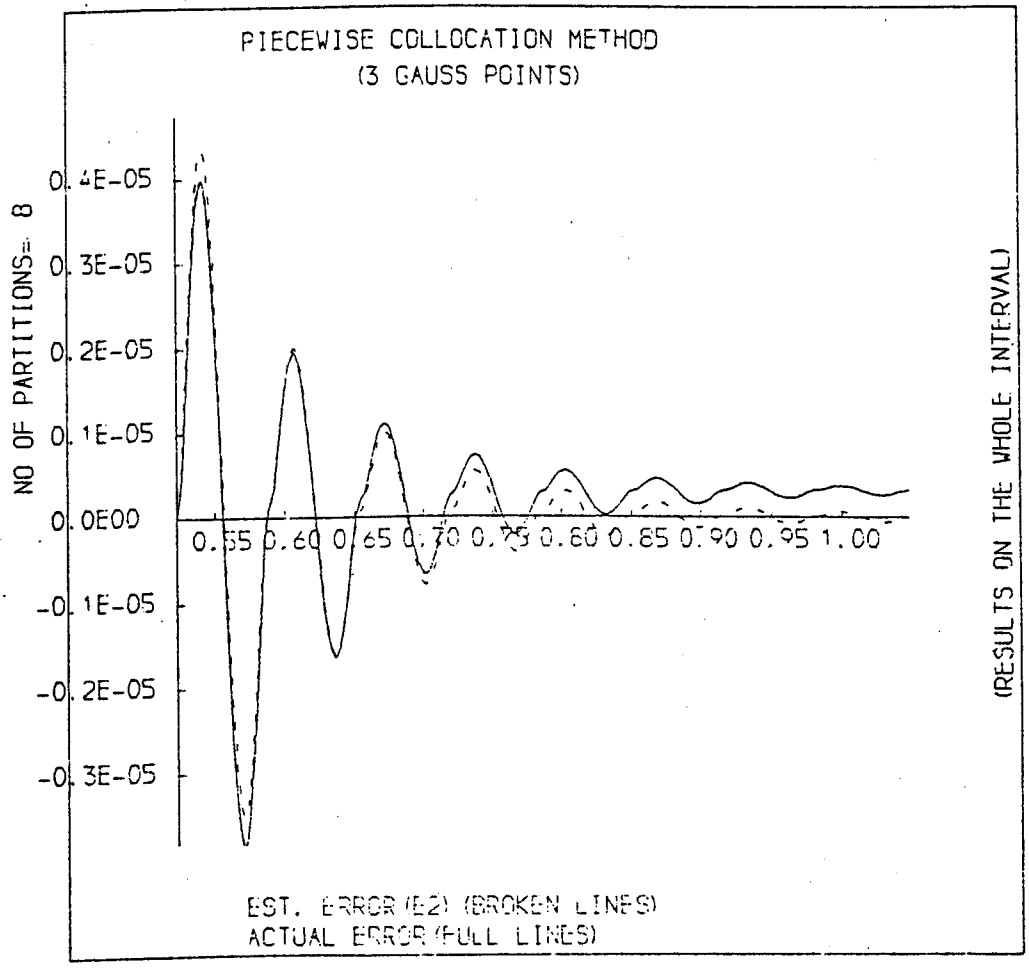
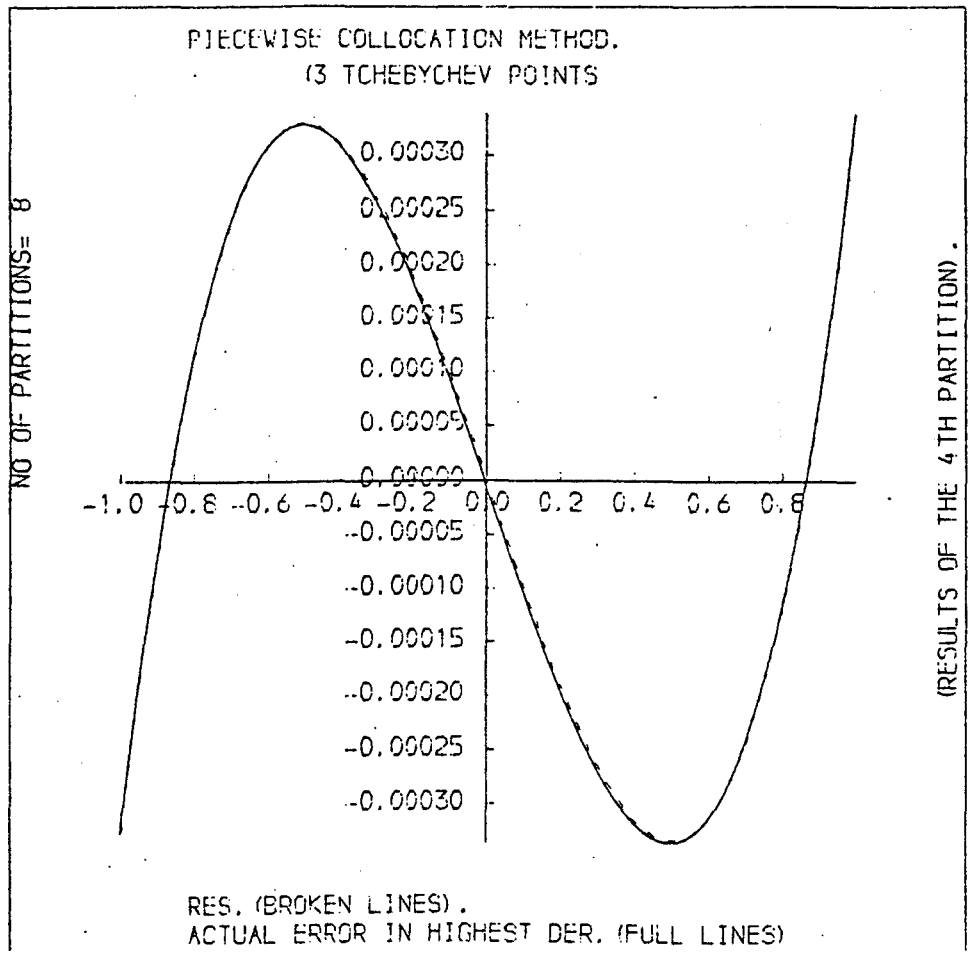


Fig. (4.10) Problem (11) The residual



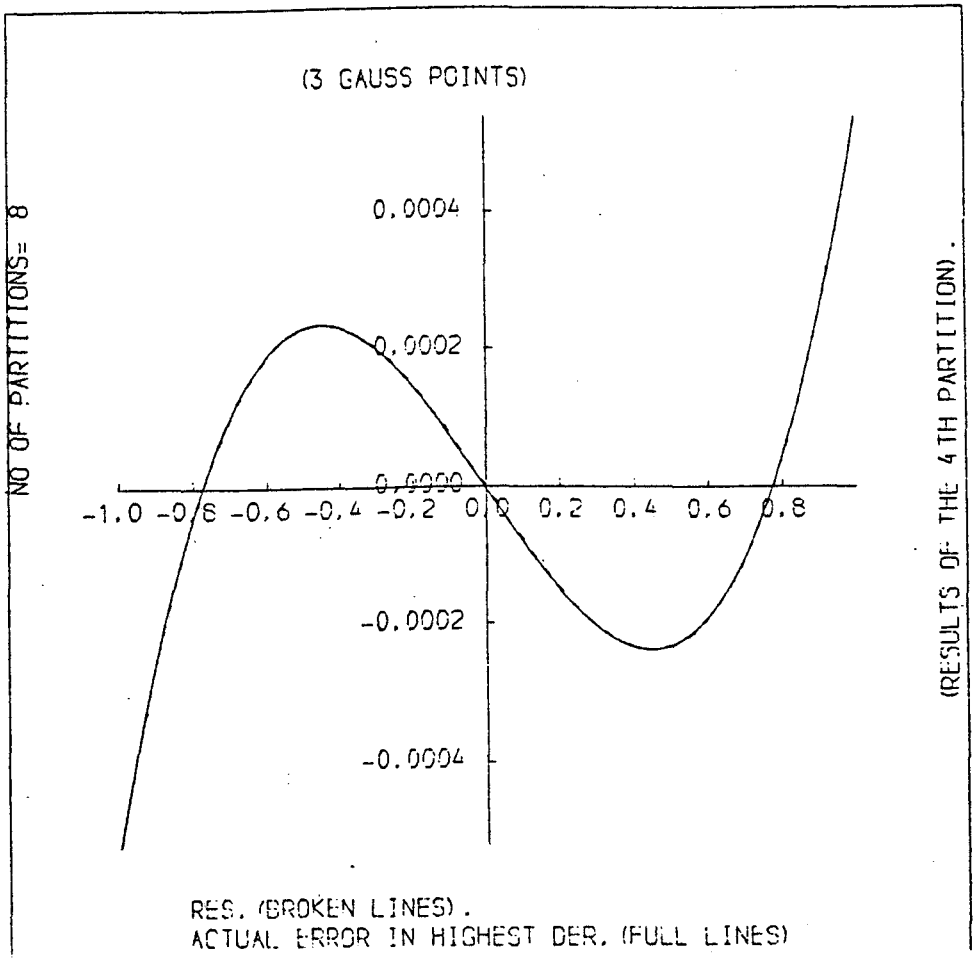
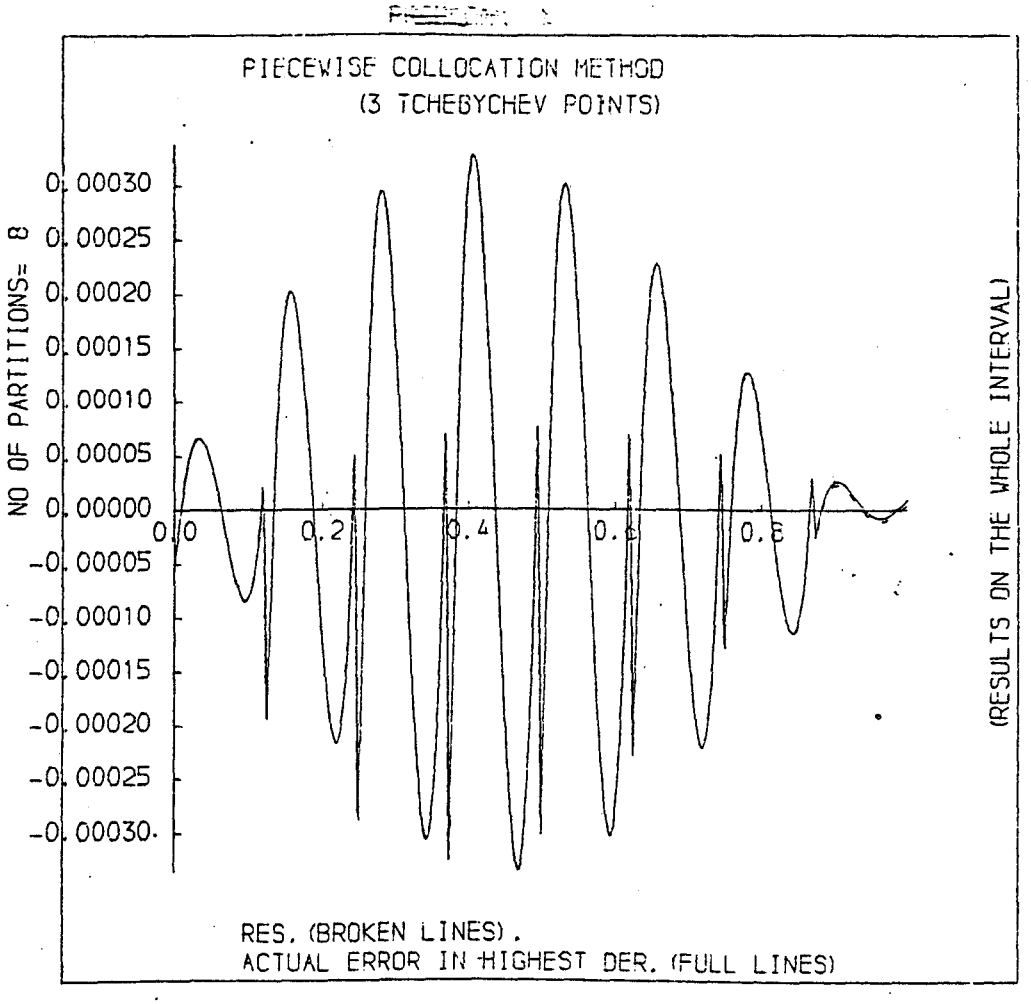


Fig. (4.11) Problem (10) The residual



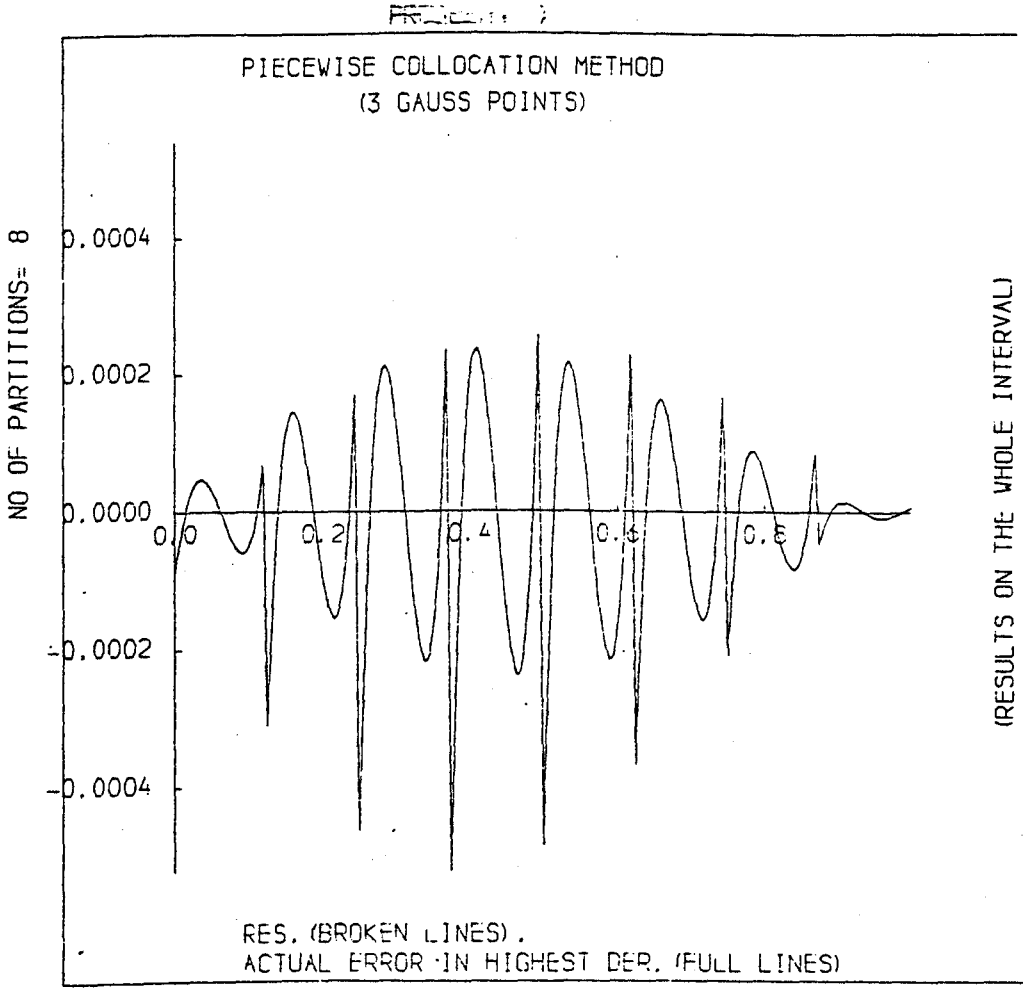
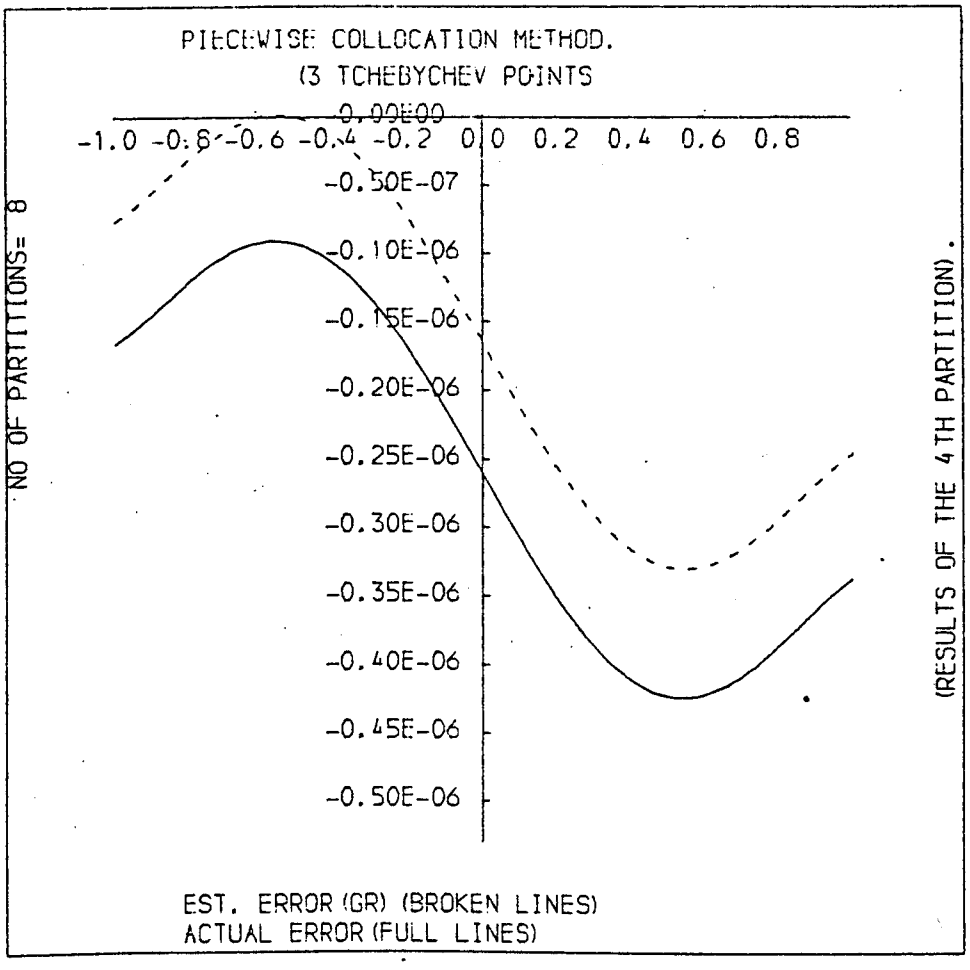


Fig. (4.12) Problem (11) The GR estimate



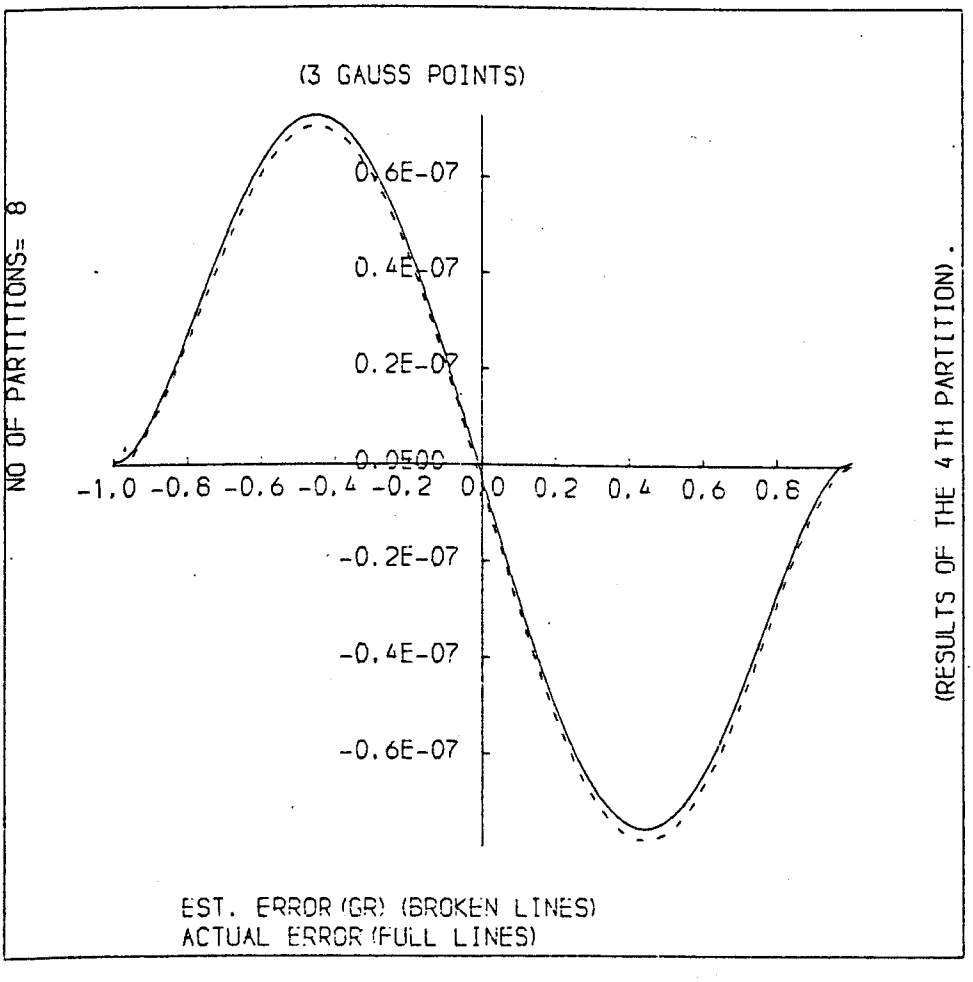
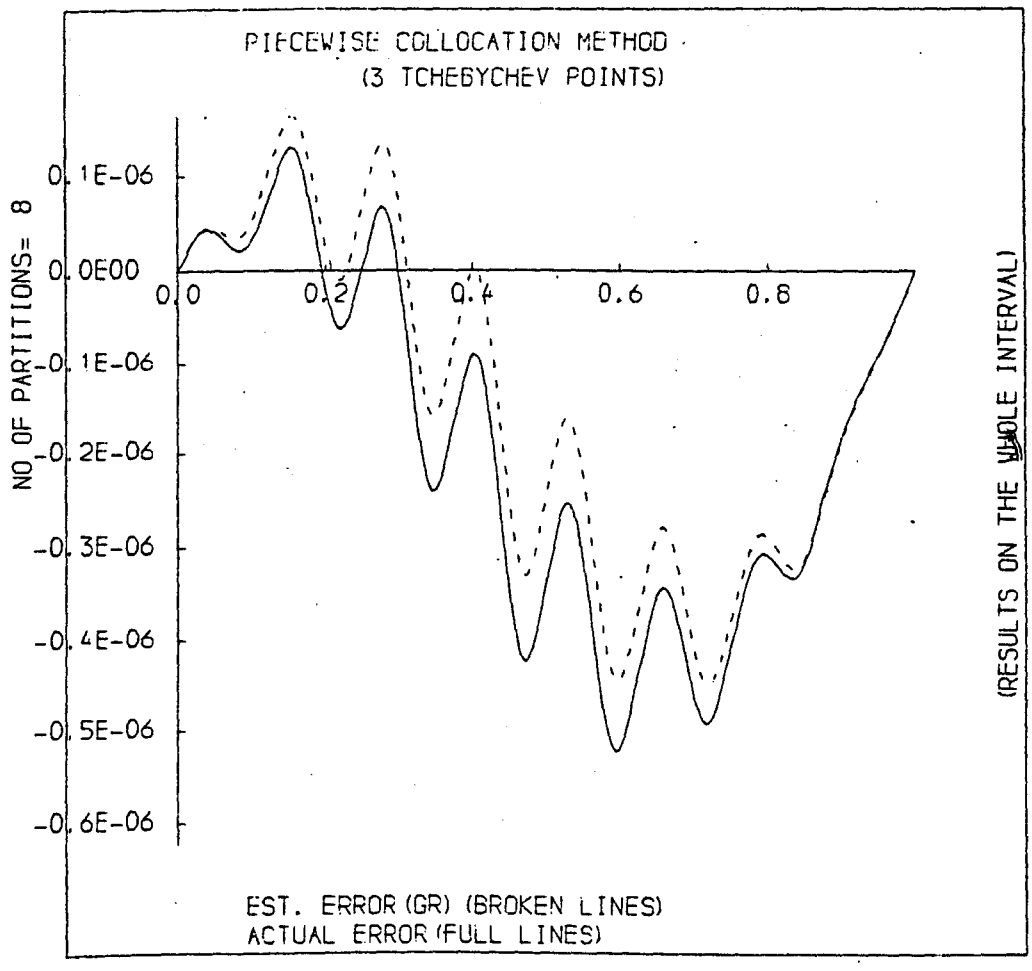


Fig.(4.13) Problem (11) The GR estimate



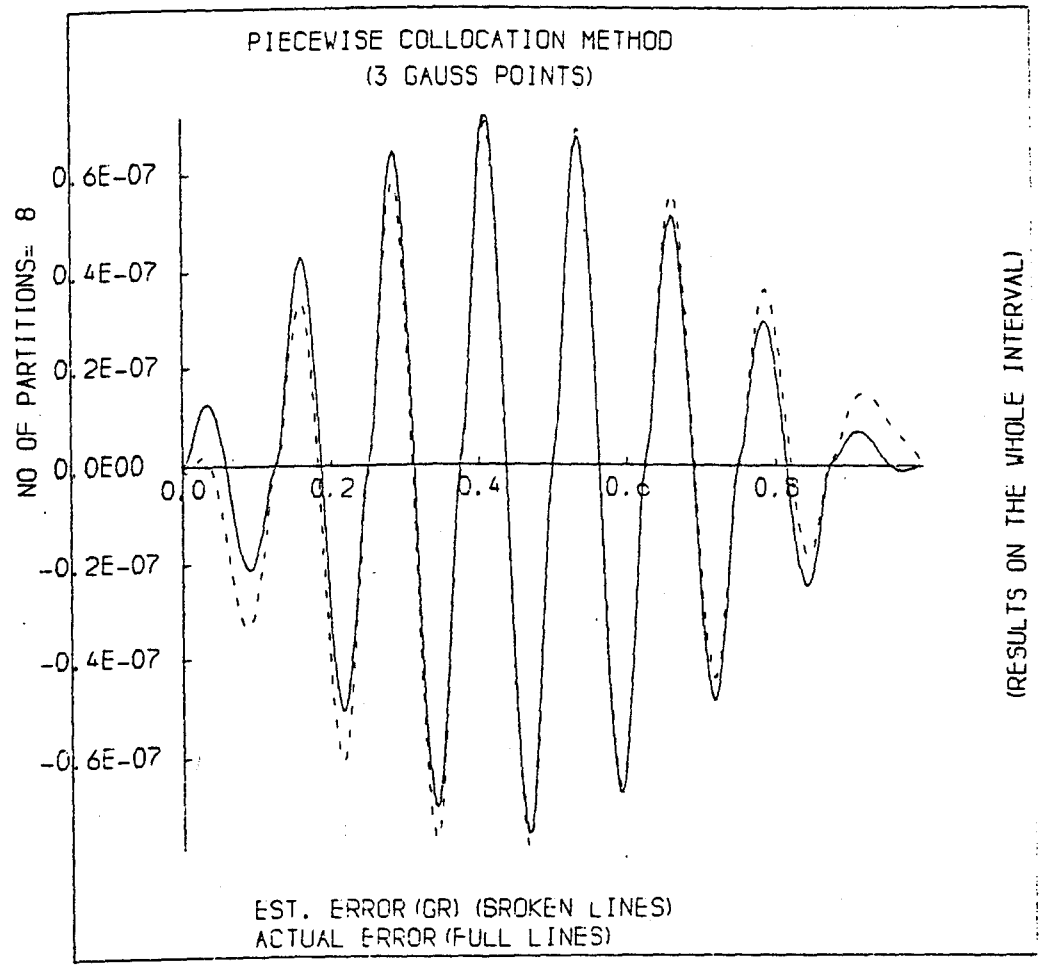
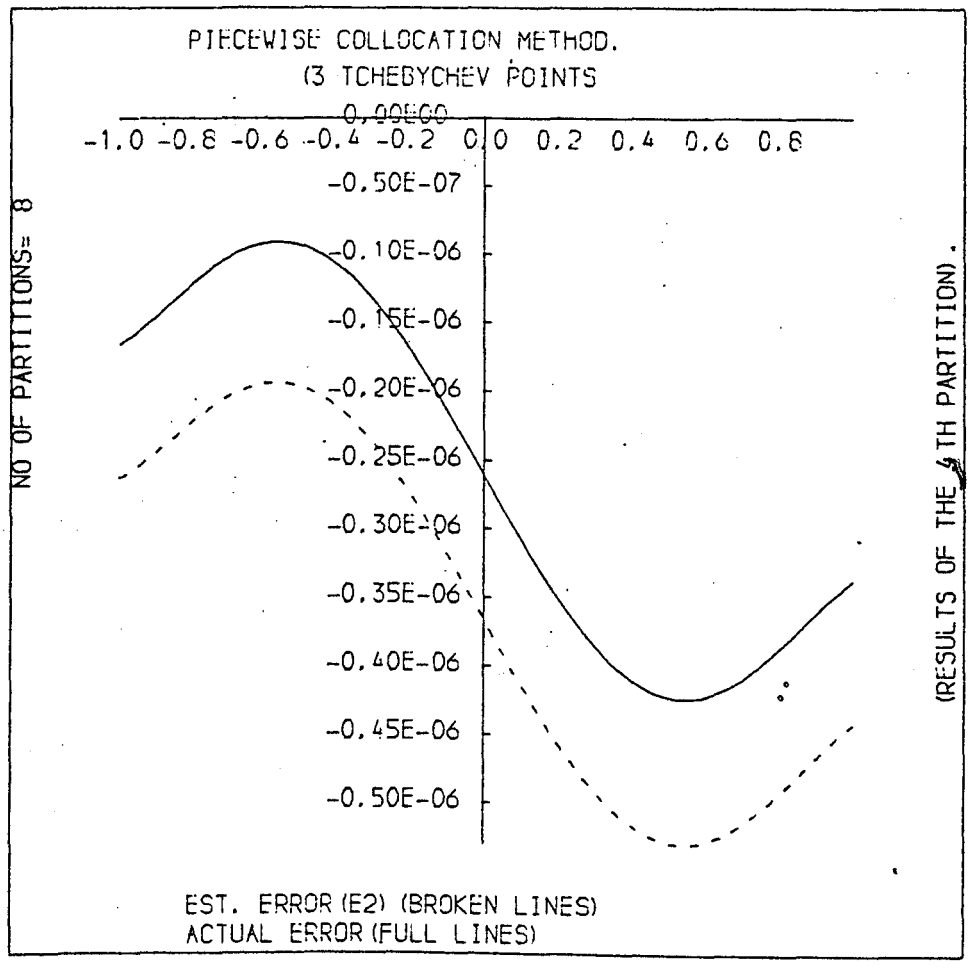


Fig. (4.14) Problem (11) The E2 estimate



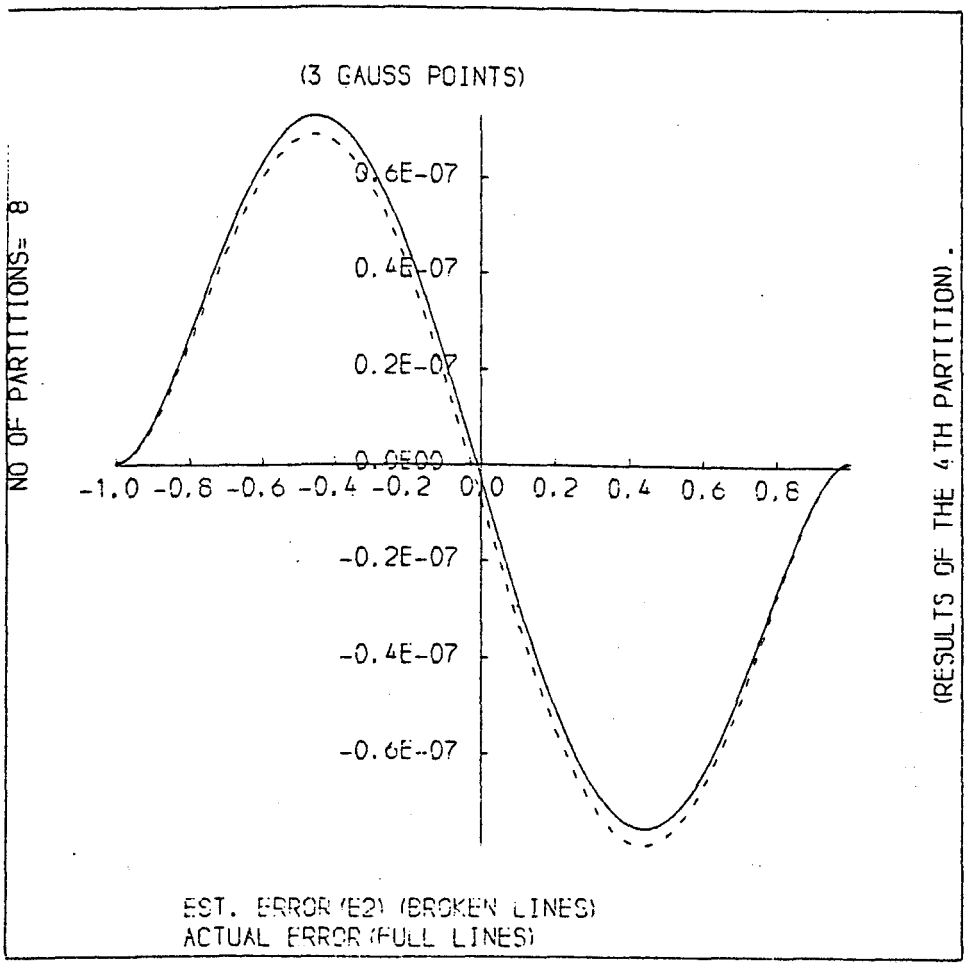
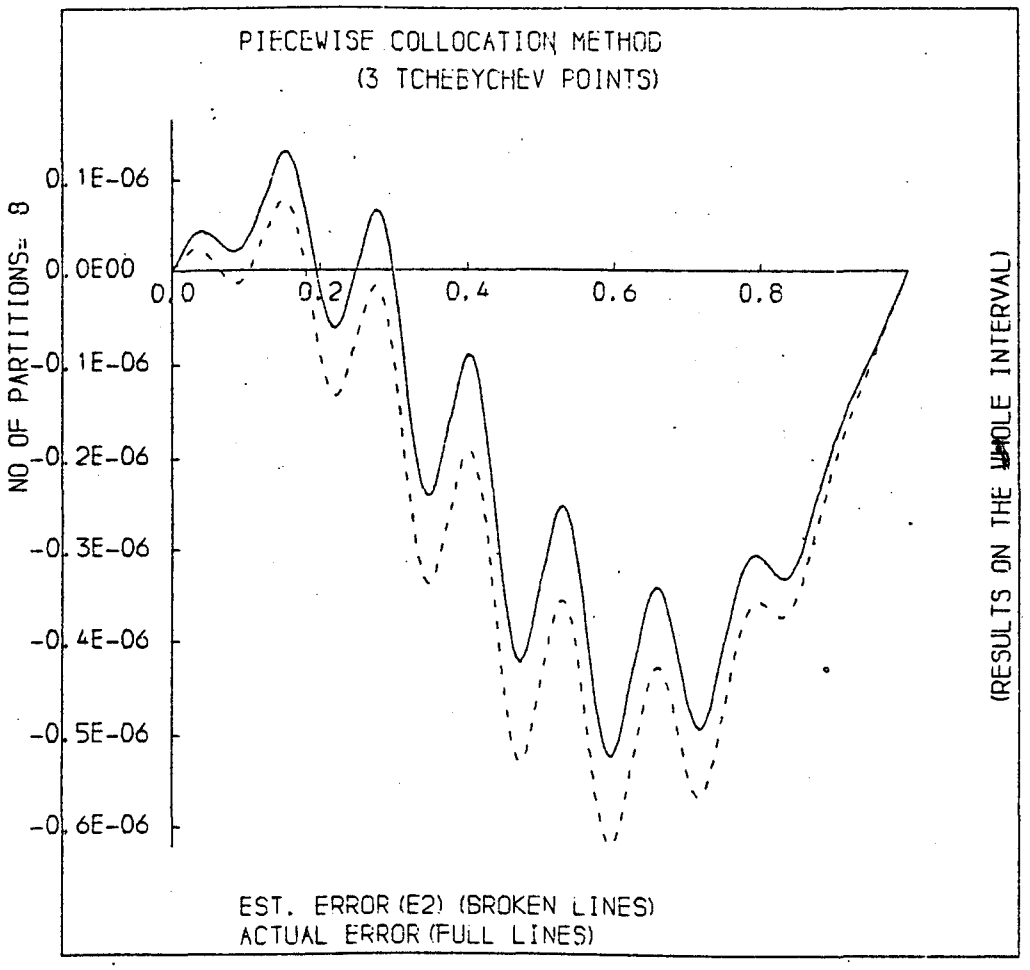
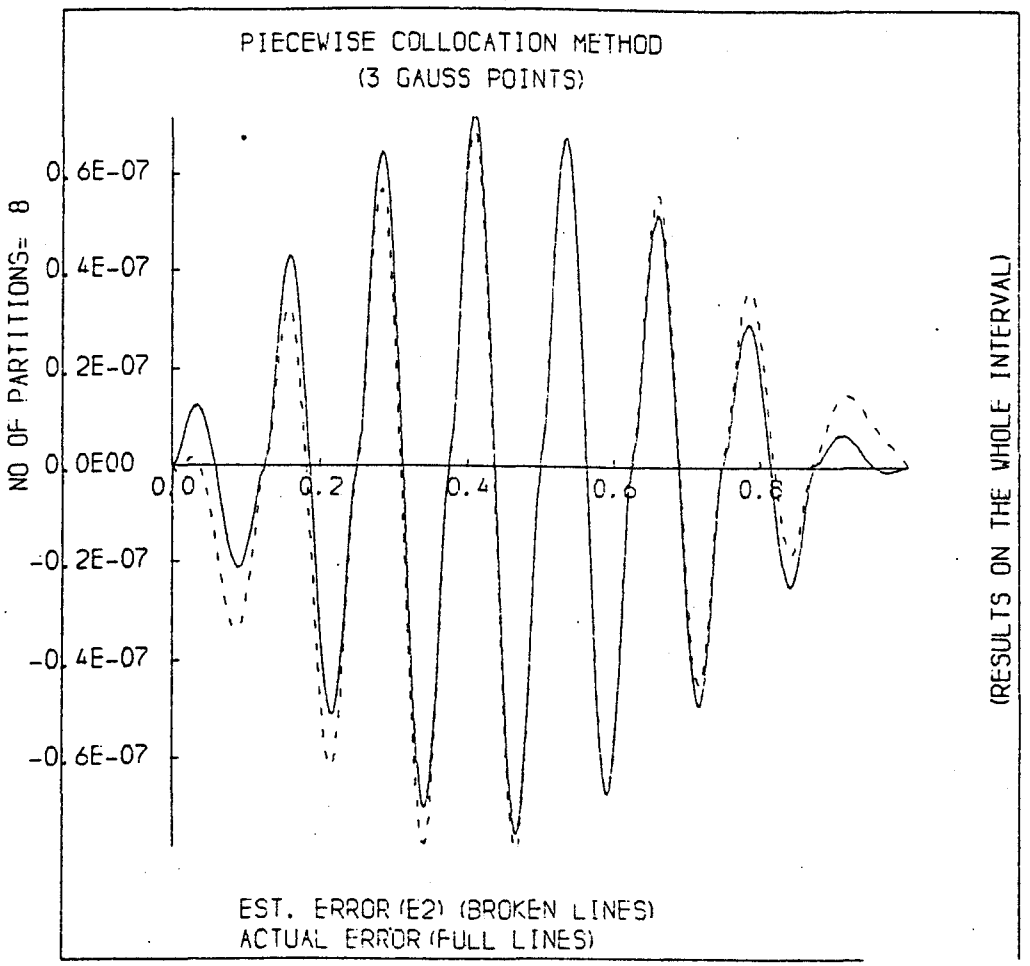


Fig. (4.15) Problem (11) The E2 estimate





CHAPTER FIVE

Adaptive mesh selection algorithms for boundary value problems

5.1. Introduction

We have shown in the previous chapter that using some nice properties of the residual and the collocation matrix, close and simple error estimates could be developed. In this chapter we will examine the use of these ideas as selection criteria for adaptive mesh selection algorithms. Despite the tremendous importance of adaptive procedure to general computer codes for solving boundary value problems viable methods using collocation have only recently been investigated. Algorithms, based on the local behaviour of the error shown in de Boor and Swartz (1973), have been theoretically investigated by de Boor (1973) and later in 1978, improved by Russell and Christiansen (1978). Their most effective algorithm is implemented in a computer code "COLSYS" for solving systems of boundary value problems (Ascher, Christiansen and Russell (1978)).

Now if we consider the boundary value problem (2.1), (2,2), solved by the method of piecewise collocation using p collocation points, then we want to efficiently determine a partition of (-1,1),

$$\pi = -1 = t_0 < t_1 < \ldots\ldots\ldots < t_n = 1$$

such that n is small but if $x_{np}(t)$ is computed using π then the error $e_n(t) = x(t) - x_{np}(t)$ satisfies

$$||e_n|| < \text{a desired tolerance TOL} \qquad (5.1)$$

Our technique here will be as follows. Solve the problem with small n using the equally spaced scheme. Determine the partitions with maximum share in the error $||e_n||$ using some selection criterion.

Halve each of these partitions and resolve. Stop when (5.1) is satisfied ($\equiv \pi$ is determined). With this technique it is possible to make use of matrices required in the previous stage. In the collocation code "COLSYS" a new partition which may be completely different from the previous one is determined in each stage. In such cases a full construction of a solution matrix is required each time. Another criticism of that algorithm that with badly behaved problems it may need to solve the problem with large values of n several times before it takes the right direction. e.g. (Ascher, Christiansen and Russell 1978, example 1).

In section (5.2) we will introduce the selection criteria that we want to use. These criteria will be tested and compared in section (5.3), on a variety of badly behaved problems using the simple adaptive procedure - select the subinterval with the largest effect and halve it. After selecting our preferences from among them efficiencies and improvements are considered more closely in section (5.4).

5.2. Introduction to the algorithms

The algorithms chosen for examination here can be divided into three groups according to their selection criterion.

(1) Algorithm based on the first derivatives.

This is a simple minded one included only for comparison.

(2) Algorithms based on error estimations using the residual.

These error estimations are

(i) $(GR)_i = G^{-1}R_i$ where $G^{-1}R_i$ is defined in (4.15).

(ii) $E_2 = G^{-1}R_i + (G - \phi_n T)^{-1} \phi_n K R_i$ where $K R_i$ defined in (4.16).

and (iii) Algorithm based on error estimation using the collocation matrix Q_n.

Using the simple adaptive proceudre, if we assume the subinterval to be halved is the j*th subinterval then j* is determined as follows:

(1) Using the first derivative

Algorithm (1)

Find the maximum of the first derivative of the approximate solution x_{np} in the ith partition. Then

$$j^* = \max_i (h_i ||x'_{npi}||)$$ where h_i = the size of the ith partition and $||x'_{np}||_i$ is the maximum approximate derivative in the ith partition.

(2) Using the residual

Algorithm (2i)

$$j^* = \max_i (GR)_i$$

We expect here the subinterval with maximum share in the error $||e_n||$ to be the subinterval with largest estimate GR.

Algorithm (2ii)

$$j^* = \max_i ||E_2^i||$$

We expect here that the subinterval with maximum estimate $||E_2^i||$ gives the maximum effect on the error $||e_n||$.

Algorithm (2iii)

$$j^* = \max_i ||r_i||$$

Here the subinterval with maximum residual is expected to give the maximum share in the error.

(3) Using the Q matrix

If the elements of the matrix Q are

$$\{q_{ij}\}_{\substack{i=1,\ldots,np \\ j=1,\ldots,np}},$$

let Q_{ij} denote the sub-block matrix of all elements in Q correspond to the jth partition in the ith partition, i.e.

$$Q_{ij} \equiv \{q_{kl}\}_{\substack{k=i,\ i+1,\ldots i+p-1 \\ l=j,\ j+1,\ldots j+p-1}} \qquad \begin{matrix} i = 1,\ldots,n \\ j = 1\ldots,n \end{matrix}$$

An estimate of the error in the ith partition may be taken as

$$e_Q^i = \sum_{j=1}^{n} ||Q_{ij}|| \ ||r_j|| \quad \text{where} \quad ||r_j|| \quad \text{is the}$$

maximum residual in the jth partition. Here instead of looking directly to the subinterval with maximum residual, we look firstly for the sub-interval with maximum error, say the i*th partition.

$$i^* = \max_i e_Q^i.$$

Then we look for the partition where the residual gives the maximum share of that error i.e. look for j* where

$$j^* = \max_j ||Q_{i^*j}|| \ ||r_j|| .$$

These five algorithms are tested and compared on four test problems which range from mildly difficult to difficult ones in the next section. The piecewise collocation method with three points (Tchebychev and Gauss) are used. $||x_{np}||_i$, $||r_i||$, $(GR)_i$, $||\Xi_2^i||$ and $||e_n||_i$ are evaluated using 200 equally spaced points.

5.3. Comparisons

The algorithms are compared, in the size of reduction in the actual error e_n and the number of mesh points in the difficult range, using different values of n (the number of mesh points). The initial solution is found using 5 mesh points equally spaced.

Problem (12)

$$x'' - 10^8 (2-s^2) x = -10^8 \qquad x(\pm 1) = 0$$

from Russell and Shampine (1978). This problem has a unique solution symmetric about zero and having a boundary layer of width $\sim 10^{-4}$ at -1. The solution is

$$x(s) \sim \frac{1}{(2-s^2)} - e^{-10^4(1+s)} - e^{-10^4(1+s)} \quad , \quad \text{and}$$

is illustrated graphically in Fig. (5.1).

We observe in Table (5.1) that:

(i) Algorithm 2(i) is the slowest. That is expected since K is very large and in such case GR will not give a good estimate of the error as noted earlier.

(ii) Algorithm 2(ii) is not as fast as others since it involves $(G - \phi_n T)^{-1} \phi_n K R_n$ which may not be reliable with small values of n with this type of badly behaved problems.

(iii) In comparing other algorithms we find that all have done almost the same work.

Problem (13)

$$x'' + 300 s x' + 300 x = 0 \qquad x(0) = 1 \qquad x(1) = e^{-150}$$

from Russell and Christiansen (1978).

The solution $x(s) = e^{-150 s^2}$ decreases rapidly from $x(0) = 1$, $x'(0) = 0$ to near zero for $s > 0$. This behaviour is

Fig. (5.1) The solution, Problem 12

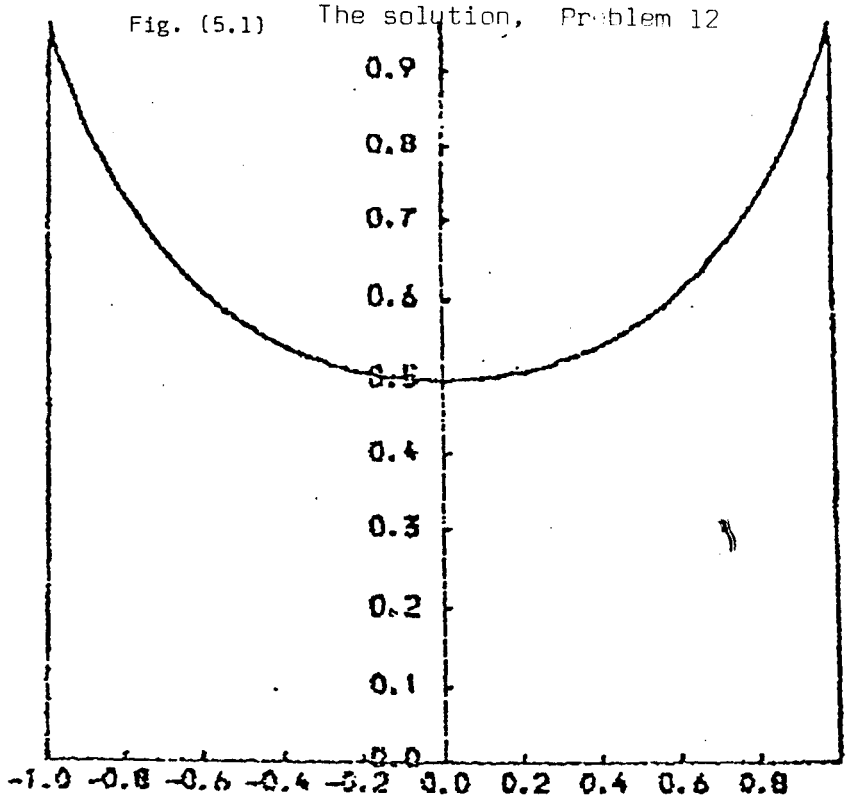
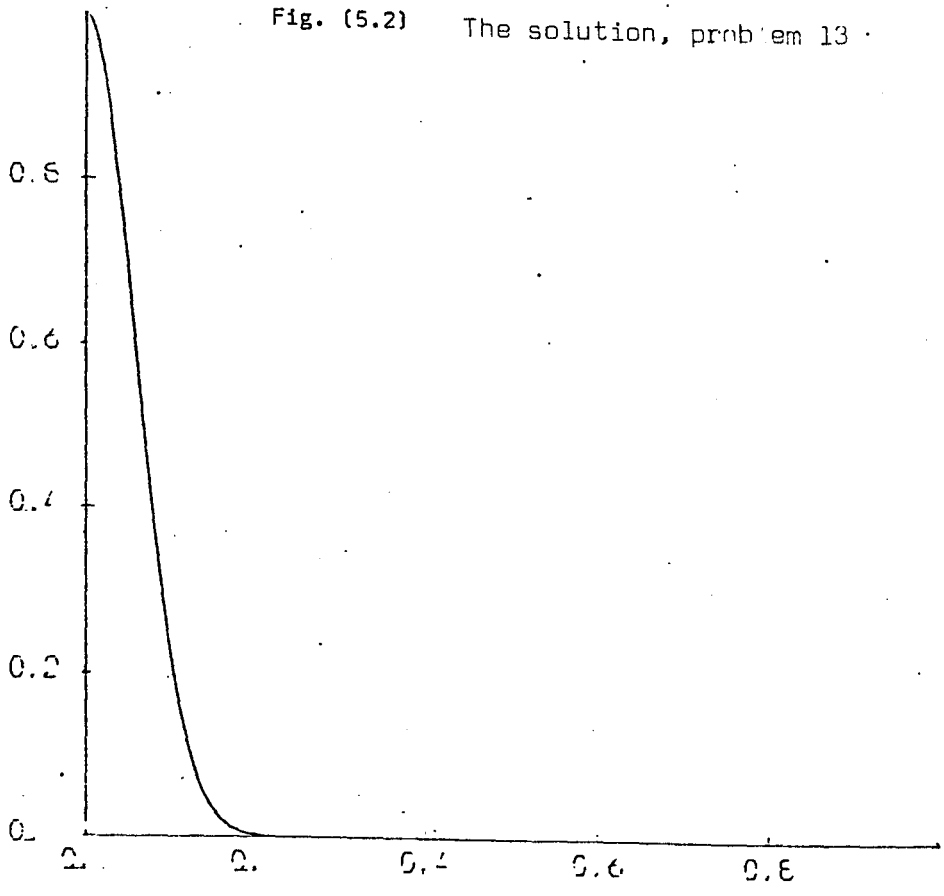


Fig. (5.2) The solution, problem 13



The simple adaptive scheme

TABLE (5.1)

Problem 12

Algorithm	n	Tchebychev points: number of mesh points (-1,-0.9)	Tchebychev points: e_n	Gauss point: number of mesh points (-1, -0.9)	Gauss point: e_n
1	5	0	0.81671	0	0.87444
	10	4	0.75970	4	0.87444
	15	6	0.64408	7	0.87445
	20	7	0.41869	7	0.28698
	25	9	0.02190	9	0.01266
	30	12	4.7×10^{-4}	12	2.4×10^{-4}
2(i)	5	0	0.81671	0	0.87444
	10	1	0.81671	2	0.87445
	15	4	0.67585	4	0.67585
	20	6	0.28698	7	0.28698
	25	8	0.14572	8	0.07570
	30	8	0.14572	8	0.07570
2(ii)	5	0	0.81671	0	0.87444
	10	2	0.81008	2	0.87445
	15	4	0.67585	4	0.67585
	20	6	0.28648	7	0.28698
	25	8	0.14572	8	0.07570
	30	9	0.02190	9	0.012660
2(iii)	5	0	0.81671	0	0.87444
	10	0	0.81008	2	0.87445
	15	4	0.75970	4	0.67585
	20	7	0.41869	7	0.28698
	25	9	0.02190	9	0.01266
	30	12	4.7×10^{-4}	12	2.4×10^{-4}
3	5	0	0.81671	0	0.87444
	10	4	0.75970	4	0.87444
	15	6	0.64408	7	0.87445
	20	7	0.41869	7	0.28695
	25	9	0.02190	9	0.01266
	30	12	4.7×10^{-4}	12	2.4×10^{-4}

The simple adaptive scheme

TABLE (5.2) Problem 13

Algorithm	n	Tchebychev points: number of mesh points in(0,0.1)	e_n	Gauss points: number of mesh points in (0,0.1)	e_n
1	5	0	0.29872	0	0.04697
	10	4	0.00273	4	6.8×10^{-4}
	15	8	0.00280	8	6.8×10^{-4}
	20	12	0.00280	13	6.8×10^{-4}
	25	15	0.00280	18	6.8×10^{-4}
2(i)	5	0	0.29872	0	0.04697
	10	2	9.5×10^{-4}	2	2×10^{-4}
	15	4	4.8×10^{-5}	5	6.5×10^{-6}
	20	7	7.5×10^{-6}	8	7.1×10^{-7}
	25	8	7×10^{-6}	8	3.4×10^{-7}
2(ii)	5	0	0.29872	0	0.04697
	10	4	0.00376	5	0.00103
	15	9	0.00370	11	0.00103
	20	14	0.00370	15	0.00103
	25	19	0.00370	20	0.00103
2(iii)	5	0	0.29872	0	0.04697
	10	3	1.9×10^{-4}	3	9.9×10^{-5}
	15	5	9.2×10^{-5}	6	9.3×10^{-6}
	20	8	1.03×10^{-5}	8	7.1×10^{-7}
	25	12	7×10^{-6}	8	3.4×10^{-7}
3	5	0	0.29872	0	0.04697
	10	2	9.5×10^{-4}	2	2×10^{-4}
	15	4	4.9×10^{-5}	4	6.5×10^{-6}
	20	7	3.2×10^{-5}	7	4.1×10^{-6}
	25	9	4.9×10^{-6}	9	2.6×10^{-7}

described graphically in Fig. (5.2). We observe in table (5.2):

(i) Here algorithm (1) moves too many points into the region of difficulty and accuracy is lost assymptotically.

(ii) Algorithm 2(ii) has failed due to the unreliable estimate $(G - \phi_n T)^{-1} \phi_n K R_n$.

Problem (14)

$$x'' + 10^6 s\, x' = -\pi^2 \cos(\pi s) - 10^6 \pi s \sin(\pi s).$$

$$x(-1) = -2 \qquad x(1) = 0.$$

The solution is

$$x(s) = \cos(\pi s) + \operatorname{erf}\left(\frac{10^3}{\sqrt{2}} s\right) / \operatorname{erf}\left(\frac{10^3}{\sqrt{2}}\right).$$

Ascher, Christiansen and Russell (1978).

The solution has a turning point at $x = 0$. The transition layer is of width $\sim 10^{-3}$. A graph for the solution is given in Fig. (5.3). We observe with this problem

(i) Algorithm (1) failed because the width of the transition layer where the derivative is expected to take its maximum is very small (10^{-3}) and may be missed in the evaluation procedure. At the same time the derivative is of similar size around the boundary layer which makes the algorithm tend to divide the interval equally.

(ii) The algorithms using the residual (2(i), 2(ii) and 2(iii) have failed because the residual in this example behaves very differently to the solution. If we look to Fig. (5.4) of the right hand side, $y = -\pi^2 \cos(\pi s) - 10^6 \pi s \sin(\pi s)$, we observe that it behaves very badly towards the end and that explains why these algorithms keep on dividing the end subintervals leaving the middle ones where

Fig.(5.3) The solution, problem 14

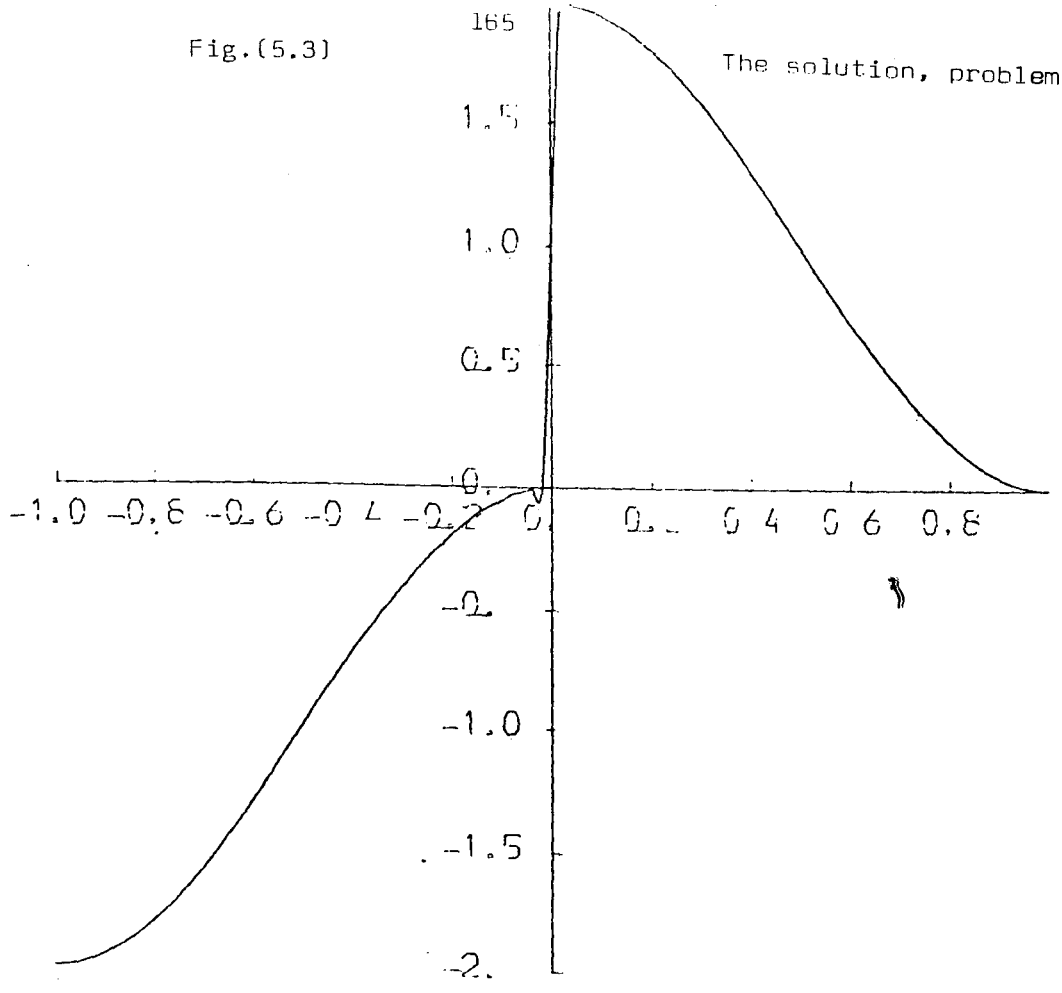
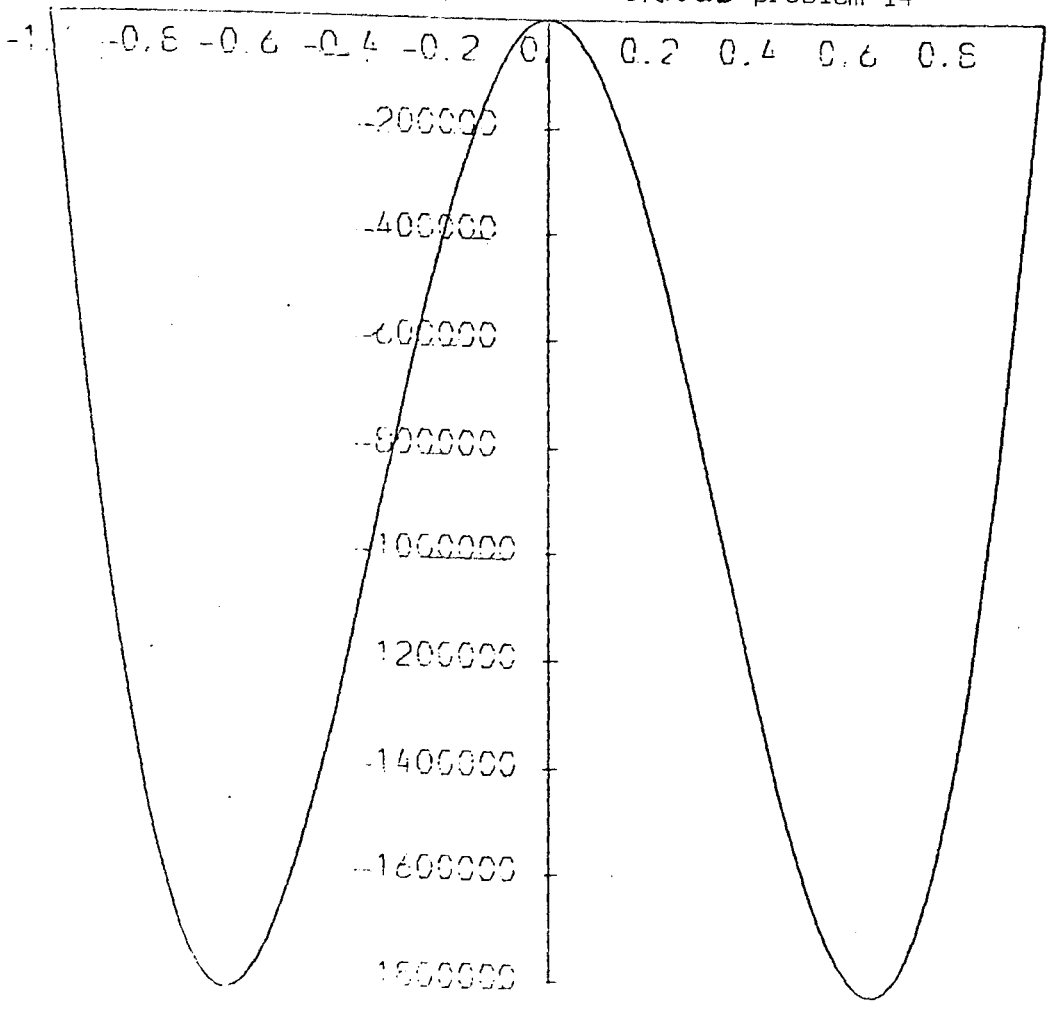


Fig. (5.4) The residual problem 14



The simple adaptive scheme

TABLE (5.3) Problem 14

Algorithm	n	Tchebychev points: number of mesh points in(-0.1,0.1)	Tchebychev points: e_n	Gauss points: number of mesh points in(-0.1,0.1)	Gauss points: e_n
1	5	0	0.96458	0	0.95495
	10	0	284.45	0	188.96
	15	2	141.89	2	95.067
	20	2	71.185	2	48.054
	25	4	33.604	4	22.996
	30	4	33.143	4	22.774
	40	4	17.239	4	22.717
2(i)	5	0	0.96458	0	0.95495
	10	0	0.96410	0	0.95496
	15	0	0.96410	0	0.95496
	20	0	0.96410	0	0.95496
	25	0	0.96410	0	0.95496
	30	0	0.96410	0	0.95496
	40	0	0.96410	0	0.95496
2(ii)	5	0	0.96458	0	0.95495
	10	0	0.96408	0	0.95496
	15	0	0.96398	0	0.95496
	20	0	0.96398	0	0.95496
	25	0	0.96398	0	0.95496
	30	0	0.96398	0	0.95496
	40	0	0.96398	0	0.95496
2(iii)	5	0	0.96458	0	0.95495
	10	0	0.96447	0	0.95496
	15	0	0.96435	0	0.95496
	20	0	0.96435	0	0.95496
	25	0	0.96435	0	0.95496
	30	0	0.96435	0	0.95496
	40	0	0.96435	0	0.95496
3	5	0	0.96458	0	0.95495
	10	2	281.48	2	188.96
	15	4	67.549	4	46.226
	20	6	28.333	4	46.284
	25	7	12.426	6	10.284
	30	12	0.0500_{-4}	6	10.209
	40	18	4.8×10^{-4}	6	10.123

the error takes its maximum.

(iii) Algorithm (3) has done very well here, especially in the Tchebychev case. The error has been reduced from 281.48 with 10 mesh points (2 points in (-0.1, 0.1) to $4.8\ 10^{-4}$ with 40 mesh points (18 points in (-0.1, 0.1) . This algorithm has shown its capability with this delicate example because it takes into its account the effect of Green's function in the error (Q matrix) as well as the residual.

(iv) We note that Tchebychev points here give better results. But one expects Gauss to do better with larger value of n, since its solution hasn't settled down yet.

Problem 15

$$x'' + \frac{2}{s}\, x' + \frac{1}{s^4}\, x = 0 \qquad x\left(\frac{1}{3\pi}\right) = 0 \qquad x(1) = \sin(1)$$

From Russell and Christiansen (1978).

The solution is $\sin\left(\frac{1}{s}\right)$ which is oscillatory. A graph for the solution is given in Fig. (5.5).

We observe in table (5.4) that

(i) Algorithm 1 is the worst for the same reason as in problem (13).

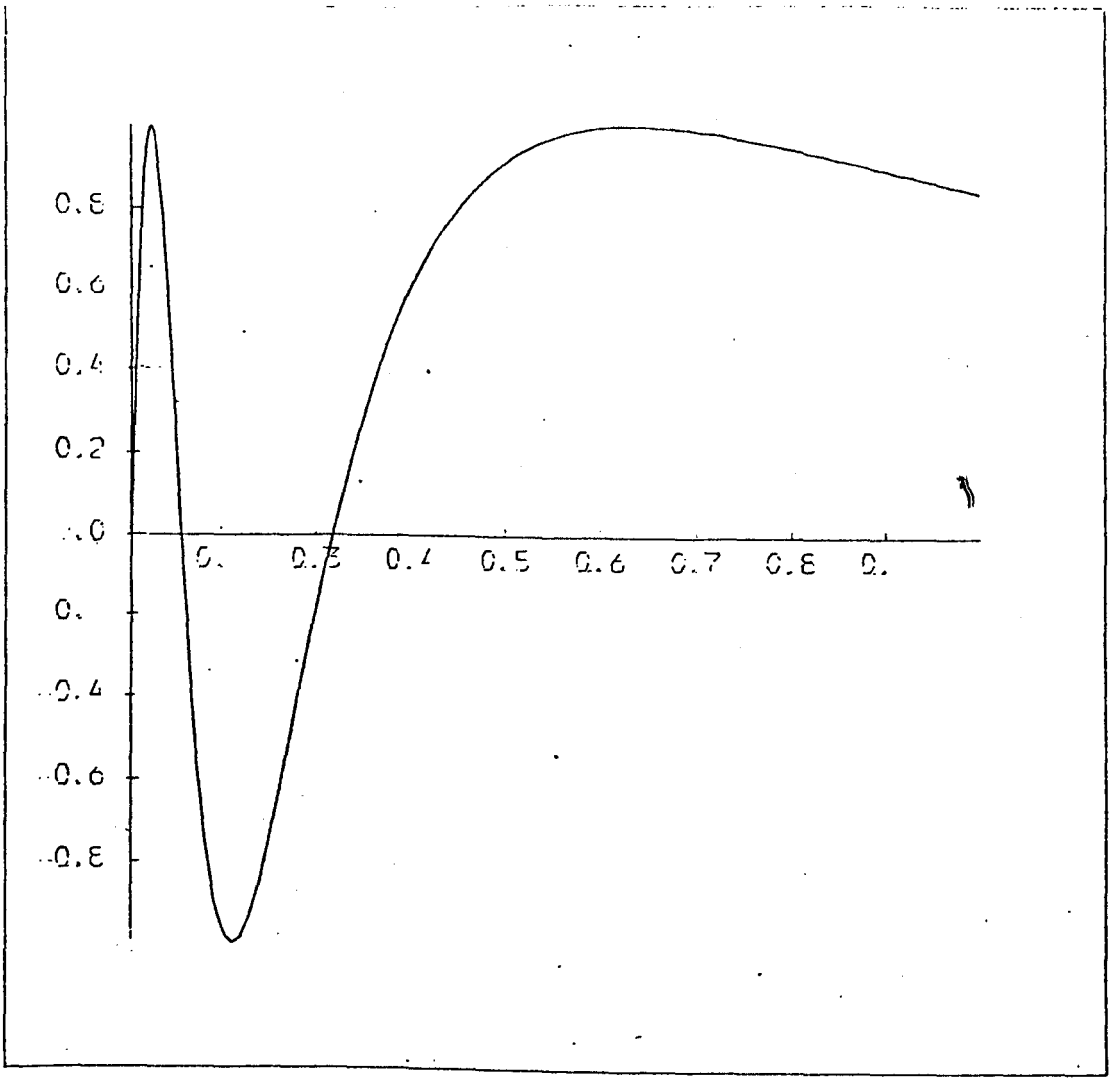
(ii) Algorithm 2(ii) has done better here since K is very small which makes the unreliable term $(G - \phi_n T)^{-1}\phi_n\ K\ R_n$ negligible and 2(ii) becomes like 2(i).

Finally we may summarise these results in the following way:

(1) Algorithm (3) shows its superiority and capability with all examples. Its drawback may be that it involves the calculation of Q_n which is very expensive.

(2) Algorithm 2(iii) comes in the second position. It fails only when the right hand side of the equation is very badly behaved on a region

Fig. 5.5 The solution, problem 15



The simple adaptive scheme

TABLE (5.4) Problem 15

Algorithm	n	Tchebychev points: number of mesh points in $(\frac{1}{3\pi}, \frac{2}{3\pi})$	Tchebychev points: e_n	Gauss points: number of mesh points in $(\frac{1}{3\pi}, \frac{1}{3\pi})$	Gauss points: e_n
1	5	0	2.1332	0	2.5264
	10	4	0.0176	4	0.00323
	15	7	0.00146	7	1.7×10^{-5}
	20	9	8.4×10^{-4}	9	1.4×10^{-4}
	25	13	7.6×10^{-4}	13	1.4×10^{-4}
2(i)	5	0	2.1332	0	2.5264
	10	4	0.01295	3	0.00166
	15	6	0.00167	6	1.4×10^{-4}
	20	10	9.2×10^{-4}	10	4.2×10^{-5}
	25	12	1.2×10^{-4}	11	3×10^{-6}
2(ii)	5	0	2.1332	0	2.5264
	10	4	0.01295	3	0.00166
	15	6	0.0167	6	1.4×10^{-3}
	20	10	9.2×10^{-4}	10	4.2×10^{-5}
	25	12	1.2×10^{-4}	11	3×10^{-6}
2(iii)	5	0	2.1332	0	2.5264
	10	5	0.01789	5	0.00323
	15	9	0.00560	9	1.9×10^{-4}
	20	13	0.00[illegible]8	12	7.1×10^{-5}
	25	16	3.9×10^{-4}	16	8.6×10^{-6}
3	5	0	2.1332	0	2.5264
	10	4	0.01295	4	0.00323
	15	7	0.00167	5	1.9×10^{-4}
	20	10	9.2×10^{-4}	9	7.1×10^{-5}
	25	12	1.2×10^{-4}	11	8.6×10^{-6}

different from the region of difficulty of the solution as in problem (14). However it is cheap and simple.

(3) Algorithm 2(i) may be recommended for problems with small K and smooth right hand side. In such problems it will do well and will be superior over others due to its simplicity.

(4) The reliability of E_2 depends on $(G - \phi_n T)\phi_n$ K R_n giving a reliable estimate of $(G - T)^{-1}$ K R - which is not expected with these types of badly behaved problems - and on the right hand side being smooth.

(5) Algorithm (1) sometimes moves too many points into the region of difficulty and the accuracy is lost assymptotically. However it often gives good initial approximation and may be improved using for example the above error estimates to overcome that problem.

5.4. Improvements in the adaptive technique

In the previous section we have examined our error estimations when used on a selection criterion. In this section we will improve the efficiency of the adaptive technique in order to develop a competitive code for solving boundary value problem.

It is shown for example in Russell and Christiansen (1978) when the number of collocation points p is greater than the order of the differential equation, m, then

$$\|r_i(t)\| \sim C_i \; h_i^p + O(h^{p+1}) \; .$$

If for example, $r_i >> r_j$, and we halve the ith interval d times such that $r_i \simeq r_j$ then $c_i \left(\frac{h_i}{2^d}\right)^p \simeq c_j h_j^p$. That gives

$$d \sim \text{round} \left(\log \frac{c_i \, h_j^p}{c_j h_i^p} \Big/ \; p \log 2\right)$$

$$= \text{round} \left(\log \left(\frac{r_i}{r_j}\right) \Big/ \; p \log 2\right).$$

This formula is basic to our new dividing scheme which can be described as follows:

(i) Look for the subinterval with maximum effect, say the j*th one.

(ii) Look for the subinterval with minimum effect, say the k*th one.

(iii) Compare r_{j*} and r_{k*} and find d using the above technique. If d = 0 then halve all subintervals. If $d \geq 1$ then look for the subinterval with maximum effect other than the j*th one, say the j**th one.

(iv) Compare r_{j*} and r_{j**} and find d as above. If d = 0 then halve the j*th and j** subinterval. If d > 0 then halve the j*th subinterval d times going in the direction of the local maximum derivative.

This scheme converges very quickly with well behaved problem and treats carefully difficult ones. It is examined using algorithm 3 with the test problem (12-15). The results are given in tables (5.5) to (5.8) respectively. It can be seen that a lot of work has been saved without any loss of accuracy in comparison with the one division/step scheme.

The timing is included in these tables to see the effect of using the solution matrix in the previous step. We can see for example, in table (5.5) that the time used for the solution with 5 mesh points is 1773 while for the solution with 21 mesh points it is 1701, and similarly in the other table.

We notice also that e_Q is not giving close estimates of the error. This is not unexpected, since e_Q gives an estimate of the bound not the actual error as discussed earlier. However one could use for error estimation E_2 which is very close when the solution is well approximated.

Finally we conclude that the selection criterion based on error estimation using the Q matrix is very powerful and if the dividing scheme is improved further to increase the number of divisions/step an efficient competitive code for boundary value problems could be developed.

The improved adaptive scheme

TABLE (5.5) Problem 12

step	n	e_n	e_Q	number of mesh points in (-1,-0.9)	number of mesh points in (0.9,1)	Time for the solution
0	5	0.87444	0.91900	0	0	1773
1	13	0.87444	0.91900	7	0	1038
2	21	0.19675	0.75843	7	7	1701
3	24	0.19675	0.75843	10	7	1934
4	27	9.9×10^{-4}	0.01626	10	10	2182
5	28	9.9×10^{-4}	0.01626	11	10	2286
6	29	2.4×10^{-4}	0.0047	11	11	2357
7	31	2.4×10^{-4}	0.0047	13	11	2510
8	33	5.9×10^{-5}	0.00197	13	13	2838
9	35	5.9×10^{-5}	0.00197	15	13	3013

The improved adaptive scheme

TABLE (5.6) Problem 13

step	n	e_n	e_Q	number of mesh points in (0, 0.1)	Time for the solution
0	5	0.04697	4.5146	0	1769
1	7	0.00448	0.22408	1	541
2	8	8.2×10^{-4}	0.08217	2	621
3	9	2×10^{-4}	0.03405	2	703
4	11	2×10^{-4}	0.01629	2	1917
5	13	9.2×10^{-5}	0.0052	4	1067
6	15	6.5×10^{-6}	0.00194	6	1525

The improved adaptive scheme

TABLE (5.7) Problem 14 $\delta = 10^{-6}$

step	n	e_n	e_Q^*	number of partitions in(-0.1,1)	Time for the solution
0	5	0.30067	4001.0	0	1790
1	9	153.54	1.13×10^{5}	3	592
2	11	132.51	44789	4	720
3	13	71.612	9942	5	852
4	15	50.054	6966	5	991
5	17	28.397	1959	6	1122
6	19	28.356	2026	6	1259
7	21	28.317	1911	6	1391
8	23	28.29	1761	6	1521
9	25	12.424	535.02	7	1668
10	27	3.224	42.314	8	1792
11	29	3.314	40.102	9	1926
12	31	0.4183	2.1754	10	2059
13	33	0.4374	2.3621	11	3076
14	35	0.05	0.2701	12	3207
15	37	0.00456	0.02467	14	3313
16	39	0.00456	0.02485	16	3535
17	41	0.00107	0.0108	17	3717
18	42	0.00157	0.0108	18	3808
19	43	4.8×10^{-4}	0.0018	19	3896

The improved adaptive scheme

TABLE (5.8) Problem 15

step	n	e_n	e_Q^*	number of partitions $(\frac{1}{3\pi},\frac{2}{3\pi})$	time for solution
1	5	2.5264	63.308	0	1840
2	7	0.04608	1.6671	2	577
3	9	0.00363	0.34389	3	745
4	11	0.00164	0.1479	5	903
5	13	1.6×10^{-4}	0.07911	5	1095
6	15	1.4×10^{-4}	0.05204	7	1277
7	17	6.7×10^{-5}	0.04002	7	1460
8	19	6.9×10^{-5}	0.02915	8	1672
9	21	7.1×10^{-5}	0.02028	9	1885
10	23	4.2×10^{-5}	0.01541	10	2054
11	25	5.8×10^{-6}	0.01123	11	2054

CHAPTER SIX Conclusions

6.1. Summary

The principal part of this thesis has been in developing algorithms for computing strict error bounds plus others in error estimates and mesh selection for the numerical solution by the methods of collocation of linear differential equations. In Chapter 1 the main 'a posteriori' theorems of Kantorovich and Akilov and of Anselone have been extended and direct bounds on the inverse operator $(G - T)^{-1}$ have been derived. The inverse approximating $(G - \phi_n T)^{-1}$ has been related to the inverse of some collocation matrices and readily practical formulae for the bounds have been presented. Despite the closeness of these bounds, unfortunately the conditions of applicability, which were the major problems with the previous analysis, have turned out to be Less satisfactory.

As a by-product of this analysis the matrix Q arose in a similar way to W when dealing with $(I - K)^{-1}$. It was proved that $||Q_n|| \to ||(G-T)^{-1}||$. This convergence theorem, is of particular importance to the later discussion of error estimation, and it has justified the choice of $||Q_n||$ for expressing the norm of the approximating inverse.

To improve the applicability of the theory the principal part of the differential equation has been defined in terms of some parameters in Chapter 3. The conditions required by the theory have been expressed as before $(G \equiv D^m)$ in terms of continuity requirements on the coefficients of the differential equation and in terms of the distribution of the collocation points with some restriction on the values of the parameters. Numerical examples were considered at the end of the chapter and considerable improvements in applicability have been achieved with a simple choice of

the parameters.

It has been proved in chapter 3 that the norm of the projection operator ϕ^*_n and the usual interpolation projection ϕ_n are asymptotically the same. For small values of n $||\phi^*_n||$ is not much worse than $||\phi_n||$. This result may be used in further investigation with ϕ^*_n as a polynomial interpolation projection.

In Chapter 4 properties of the residual and the differential operator have been used to develop various algorithms for calculating bounds and estimates of the error. It has been shown if the problem is not nearly singular then $||Q_n||$ gives a good estimate of the inverse differential operator. It has also been shown that when k is not too large and the right hand side y is sufficiently smooth, the residual can be well approximated by a polynomial. This result with the oscillatory behaviour of the residual have justified simple close estimates of the error. Although estimates using the Q matrix are not as close as these estimates, they tend to be more reliable with difficult problems.

In Chapter 5 these estimates have been used in a mesh selection algorithm for solving boundary value problems. After examination on a variety of badly behaved problems, it has been shown that algorithm using the Q matrix is the most reliable. Improvements and efficiencies in the adaptive techniques are finally considered.

6.2. Improvements and extensions

6.2.1. The applicability

(a) Theory

We observe in table (2.1) that bounds calculated for the operator $G^{-1}T$ are smaller than those of $K \equiv T\,G^{-1}$ and with further work could be much smaller. That happens because the coefficients are included in the

integration in the first case. This is an indication that deltas defined in terms of $G^{-1}T$ may be much smaller and hence one would expect better applicability of bounds derived for the inverse operator,

$$(I_x - G^{-1}T)\, x = G^{-1} y.$$

This of course requires revision of the theory in the first chapter and search for suitable matrices similar (may be the same) to W and Q which can be used in bounding the approximate inverses.

(b) Applications

The results given in Chapter 3 indicate that further improvements in applicability could be achieved if all the parameters $\{\lambda_i\}_{i=1}^{m-1}$ are included and carefully chosen. One needs to examine the optimal choice of the parameters as well as the application of the idea to higher order equations.

6.2.2. Error bounds and estimation

The error bounds and estimations described in Chapter 4 could be furtherly investigated with partial differential equations. Also examination on non-linear equations with the relevant modification needs to be considered.

Finally the adaptive algorithm using the Q matrix in the selection criterion is very promising and with further research for efficiencies and improvements a competitive collocation code could be developed.

BIBLIOGRAPHY

Anselone, P.M. (1971) "Collectively Compact Operator Approximation Theory". Prentice Hall.

Ascher, U., Christiansen, J., Russell, R.D. (1978) "COLSYS - A collocation code for boundary problems". Proc. Conf. for BVP's, Houston.

de Boor, C.R. (1966) "The method of projections as applied to the numerical solution of two point boundary value problems using cubic splines". Doctoral thesis, University of Michigan, Ann Arbor.

de Boor, C.R. (1973) "Good approximation by splines with variable knots". Conference on Numerical Solution of Differential Equations. Lecture notes in Mathematics 363, 12-20.

de Boor, C.R. and Swartz, B. (1973) "Collocation at Gaussian points". S.I.A.M. Journal of Numerical Analysis 10, p.582.

Cheney, E.W. (1966) "Introduction to Approximation Theory", McGraw-Hill.

Coldrick, D.M. (1972) "Methods for the Numerical Solution of Integral Equations of the second kind", Doctoral thesis, University of Toronto.

Cruickshank, D.M. (1974) "Error Analysis of Collocation Methods for the Numerical Solution of Ordinary Differential Equations", Doctoral thesis, University of Newcastle upon Tyne.

Cruickshank, D.M. and Wright, K. (1978) "Computable error bounds for polynomial collocation methods". S.I.A.M. J.N.A. 15, p. 134-151.

Davis, P.J. (1963) "Interpolation and Approximation", Blaisdell.

Fox, L. and Parker, I.B. (1968) "Tchebychev polynomials in Numerical Analysis", London, O.U.P.

Gerrard, C. (1979) "Computable error bounds for Approximate Solution of Ordinary Differential Equations", Doctoral thesis, University of Newcastle upon Tyne.

Kantorovich, L.V. and Akilov, G.P. (1964) "Functional Analysis in Normed Spaces", Pergamon

Karpilovskaja, E.B. (1963) "Convergence of Collocation method", Sov. Math. 4, p. 1070.

Kolmogorov, A.N. and Fomin, S.V. (1957) "Functional Analysis", Graylock.

Lucas, T.R. and Reddien, G.W. (1972), "Some collocation methods for non-linear boundary value problems", S.I.A.M. 9, No. 2, p.341.

Lucas, T.R. and Reddien, G.B. (1973), "A high order projection method for non-linear two point boundary value problems". Numer. Math. 20,p.257.

McKeown, G. (1977), "Numerical Solution of Two Point Boundary Value Problems in Differential Equations by Collocation Methods", Doctoral thesis, Victoria University of Manchester.

Natanson, I.P. (1965) "Constructive Function Theory, Volume III", Ungar.

Phillips, J.L. (1969) "Collocation as a projection method for solving integral equations", Doctoral thesis, Purdue University, Lafayette, Ind.

Phillips, J.L. (1972), title as above, S.I.A.M. J.N.A. 9 , No. 1 p. 14.

Russell, R.D. (1977) "A Comparison of Collocation and finite differences for two point boundary value problems" S.I.A.M. J.N.A. 14, p. 1.

Russell, R.D. and Christiansen, J. (1978), "Adaptive mesh selection for solving boundary value problems". S.I.A.M. J.N.A. 15 p. 59.

Russell, R.D. and Shampine, L.F. (1972) "A collocation method for boundary value problems" Numer. Math. 9, p.1.

Szego, G. (1939), "Orthogonal Polynomials" (1975) 4th Edition A.M.S. Colloquium Publications XXIII.

Vainikko, G.M. (1966) "On stability and convergence of the collocation method", Differentsial 'nye Uraveniya 1, p. 244.

Vainikko, G.M. (1966), "The Convergence of the collcoation method for non-linear differential equations", U.S.S.R. Comp. Math. and Math. Phys. 6, no. 1, p.47.

Wright, K. (1964), "Tchebychev collocation methods for ordinary differential equations", Computer Journal 6, p. 358.

Wright, K. (1979), Technical report No. 135. Computing Laboratory, University of Newcastle upon Tyne.