COMPUTABLE ERROR BOUNDS FOR APPROXIMATE SOLUTIONS OF ORDINARY DIFFERENTIAL EQUATIONS

by

C. GERRARD

Ph.D. Thesis

April, 1979

UNIVERSITY OF NEWCASTLE UPON TYNE

E R R A T A

Chapter 3 onwards ..... Replace occurrences of W_{ni} by w_{ni}

ACKNOWLEDGEMENTS

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

I am also greatly indebted to the typist, Mrs. S.J. Boyd, who undertook this formidable task.

Thanks to my parents for their encouragement and support throughout.

Throughout the period of research for this thesis the author was supported by the Science Research Council.

ABSTRACT

This thesis is concerned with an error analysis of approximate methods for second order linear two point boundary value problems, in particular for the method of collocation using piecewise polynomial approximations.

As in previous related work on strict error bounds an operator theoretic approach is taken. We consider operators acting between two spaces X_1 and X_2 with uniformly equivalent metrics. The concept of a "collectively compact sequence of operators" is examined in relation to "pointwise convergence" - relevant to many approximate numerical methods. The introduction of a finite dimensional projection operator permits considerable theoretical development which enables us to relate various inverse approximate operators directly to a certain inverse matrix.

The application of this theory to the approximate solution of linear two point boundary value problems is then considered. It is demonstrated how the method of collocation can be expressed in terms of a projection method applied to a certain 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. Various estimates of bounds on the inverse differential operator are presented and it is demonstrated that the "residual" can be a very useful error estimate. The use of a "weighted infinity norm" is shown to improve the applicability of the theory for "stiff" problems. Some real problems are then examined and a selection of numerical results illustrating the theory and application are presented.

The thesis concludes with a brief review, outlining some of the deficiencies in the work and possible improvements and extensions of the analysis.

CONTENTS

INDEX OF TABLES

CHAPTER 1

Introduction

Introduction

§1.1 An application

An operator approximation theory is developed in Chapter 2 and 3 in an attempt to unify and extend other work arising mainly from studies of approximate solutions to integral and differential equations. An abstract theoretical setting is maintained until the application considered in Chapter 4. This generalised approach is taken in order to permit the application of the theory to as wide a range of problems as possible.

We now examine briefly an application of the work in this thesis, the rest of the introductory chapter will consist of a survey of related work followed by a summary of the fundamentals and main results of each chapter.

We are primarily concerned with finding strict error bounds for approximate solutions of linear two point boundary value problems in ordinary differential equations. These solutions will be the result of applying a piecewise polynomial collocation method. The theory developed in Chapters 2 and 3 does, however, have much wider applications. Interesting error estimates arise as a by product of the work in Chapter 4.

We deal with linear equations of the form

$$\frac{d^2x}{dt^2}(t) + p(t)\frac{dx}{dt}(t) + q(t)\,x(t) = y(t) \qquad (1.1)$$

subject to the boundary conditions.

$$x(-1) = x(1) = 0 \qquad (1.2)$$

The theory also applies to any order linear equation provided that the L.H.S. of (1.1) can be expressed as the sum of two differential operators applied to x, one of which is invertible with the given boundary conditions. This simple second order case suffices as an example without clouding the arguments with too much detail. The approximate solution consists of a piecewise polynomial; conditions of continuity are not imposed on the second derivative.

More details of this application are given in Sections 4.1, 4.2, 4.3 and 4.4 preceding Lemma 4.1 which can be read now as part of the introduction.

§1.2 Related work

Chapter 2 consists of an operator approximation theory essentially developed by Anselone [1] which in turn was based on the work of the Russian school of functional analysts including Kantorovich [21] Akilov [22] and Krylov [23]. Other work using similar theory includes Gilbert and Colton [16], Phillips [35,36] Rall [38] and Vainikko [48]. Coldrick [10] and Cruickshank [11] also use Anselones theory. The presentation in Chapter 2 is devoid of any mention of the later application in order to retain a sufficiently wide base for the development of other applications.

Chapter 3 introduces the concept of projections in a very general manner and proceeds with the development of the theory in Chapter 2.

Chapter 4 is an illustration of the previous theory applied to approximate solutions of (1.1) obtained by a particular projection method - collocation. Projection methods for differential equations are discussed by de Boor [3] and Lucas and Reddien [31].

Collocation methods in particular are discussed widely and references include de Boor 4 , Cruickshank 11 , Diaz 14 , Hangelbrook, Koper and Leaf 17 , de Hoog and Weiss 19 , Karpilovskaja 25 , Lucas and Reddien 30 , Phillips 35 , Reddien and Schumaker 39 , Russell and Shampine 41, Russell 42 , Sincovec 45 , Vainikko 46,47 and Wright 50 . Section 4.3 of Chapter 4 describes an application of this method.

Sets of points arising as zeros of certain orthogonal polynomials are widely used in such collocation methods, for example the Chebychev zeros by Cruickshank and Wright 11 and Gauss points by de Boor 4 .

Results from approximation theory of functions of a single real variable will play an important part in Chapter 4. References include Davis 13 , Natanson 33 and Powell 37 .

Other work directly concerning piecewise polynomial approximations for differential equations includes de Boor 3,4,5 Diaz 14 , Wittenbrink 49 , Schmidt and Lancaster 44 . A series of papers in Numerische Mathematik by Ciarlet, Shulz and Varga 6,8 are also of interest.

Although not discussed in this thesis it is possible to apply the theory to other numerical methods and to a wide range of equations including partial differential equations in several variables. Finite difference methods are discussed by Pereyra and Sewell 34 and a comparison with collocation is given by Schmidt and Lancaster 44 . Integral equations and integro-differential equations are discussed by Anselone 1 , Coldrick 10 , Hanson and Phillips 18,35,36 , Hangelbrook, Koper and Leaf 17 , Mikhlin and Smolitskiy 32 .

It should be possible to apply the theory to singular boundary value problems, see de Hoog and Weiss 19 and Reddien and Schumaker 39 , by a careful choice of the "principal part" operator M. (see §2.1). Stiff boundary value problems are examined by Flaherty and O'Malley 15 and the work of §4.10 is relevant here. Linear partial differential equations can be treated by the theory in Chapters 2 and 3 but it is far more difficult to derive certain quantities required for strict error bounds than it is in the ordinary differential equation case, however, see Gilbert and Colton 16 and Kantorovich 20 . The work of de Boor 5 , Lentini and Peryra 29 ,Peryra and Sewell 34 and Russell and Christiansen 43 are also relevant to the final discussions in §4.10.

Non linear equations cannot be treated directly by the theory described here, but error bounds for each linear equation of an iterative sequence could be found and, hopefully, combined with further convergence results to produce a final error bound. Non linear problems are considered by Bellman and Kalaba 2 , Clenshaw and Norton 9 , Lucas and Reddien 30,31 , Rall 38 , Vainikko 47 . Non linear boundary conditions are discussed by Ciarlet, Shultz and Varga 7 and Reddien 40 .

§1.3 Summary

Sections 2.1, 2.2, 2.3 form an introduction to Chapter 2, they contain the basic definitions of spaces, operators and norms used throughout. The aim is to generalise and express concisely various relations between inverse operators. Most of the theorems in this chapter are well known in the setting of a Banach space X and the associated space of bounded linear operators $[X]$.

In this chapter we consider the form of these theorems when the operators map elements of a space X_1 to another space X_2. It turns out that these spaces must have essentially the same structure, but useful results are achieved later using the extended theorems. Anselones 1 concept of "collective compactness" is a convenient way of expressing the convergence properties of certain numerical procedures - which often appear in other guises.

In Chapter 3 a projection operator is introduced and various approximations are described in terms of it. A generalised approximation is developed from these ideas which includes for example the "collocation method" for differential equations and many quadrature formulae used for solving integral equations. Sections 3.1 to 3.4 can be read as an introduction to this chapter. The aim of this chapter is to produce bounds on an inverse operator (expressing x in terms of y) in terms of quantities which are computable. A projection onto a finite dimensional space is shown to permit considerable development of the theory in Chapter 2, without sacrificing generality. The bounds developed here, of course, will not be suitable for all applications and there is much room for more detailed investigations.

An important convergence theorem (3.7) relates the norm of the inverse operator directly to the norm of an inverse matrix provided certain fairly general conditions are satisfied. This theorem is extremely valuable in justifying certain error estimates for the approximate solutions.

In Chapter 4 a two point boundary value problem in ordinary differential equations is defined and expressed as an operator equation in the space of Riemann integrable functions with sup. norm.

An approximate solution generated by the method of collocation using piecewise polynomials is studied. These sections 4.1 - 4.4 form the basis of an example to which the theory in Chapters 2 and 3 can be applied. The latter half of Section 4.4 verifies that certain conditions of the theory are satisfied and concludes with a "plug in" statement of a-priori bounds on the inverse approximate operator and a-posteriori bounds on the actual differential operator. Section 4.5 verifies certain extra conditions hold in order to apply Theorem 3.7, stated here as Theorem 4.8.

Having shown that the theory is applicable to this problem Sections 4.6 - 4.10 proceed to develop concrete numerical bounds on various operators and from these show how it is possible to obtain computable error bounds for an approximate solution. Particularly of note are the improved error bounds possible using Legendre zeros for the collocation points. The use of a weighted sup. norm is demonstrated which enables realistic error bounds to be produced for "stiff" problems.

It would be interesting to apply the theory to other approximate methods such as finite difference schemes and to higher order problems but time does not permit this.

We examine in Chapter 5 some numerical evidence which prompted the theoretical investigations in this thesis. In particular the behaviour of the inverse collocation matrix and the close relationship of the error to the 'residual' are studied for a series of problems. The chapter closes with some examples of error bounds for a small group of problems.

CHAPTER 2

Theory of Approximation Methods

Theory of Approximation Methods

§2.1 Introduction

This chapter introduces the theoretical background for certain operator equations and their approximate solution. Most of the results are well known but are included for completeness. Kantorovich and Akilov [22], Anselone [1], Coldrick [10] and Cruickshank [11] cover much of the same work.

The theorems are placed in a general setting so as not to restrict their application. Later chapters will be concerned more specifically with collocation as a projection method for the approximate solution of boundary value problems in ordinary differential equations.

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

$$Mx = y \qquad y \in Y \qquad M \in [X,Y] \tag{2.1}$$

for $x \in X$.

It is not always possible to solve (2.1) analytically and often a numerical method is used to approximate (2.1), e.g.

$$\tilde{M}\tilde{x} = \tilde{y} \qquad \tilde{y} \in \tilde{Y} \qquad \tilde{M} \in [\tilde{X}, \tilde{Y}] \tag{2.2}$$

Solving for $\tilde{x} \in \tilde{X}$. This equation is usually set in a space of finite dimension and corresponds to a finite set of linear algebraic equations. Now provided $\tilde{X} \subseteq X$ and M is invertible it follows that

$$x - \tilde{x} = M^{-1}(M(x - \tilde{x})) = M^{-1}(y - M\tilde{x}) \qquad (2.3)$$

$$\text{or } ||x - \tilde{x}||_x \leq ||M^{-1}|| \cdot ||y - M\tilde{x}||_y$$

which is a strict error bound on the approximate solution.

The rest of this chapter will be concerned with the term M^{-1}.

In all of the following theory we shall be concerned with an operator M which may be split into two parts

$$M = M_1 + M_2 \qquad M_1, M_2 \in [X,Y] \qquad (2.4)$$

where M_1 is invertible. Under certain circumstances we may deduce that M is invertible. Note that equation (2.1) may now be written

$$(M_1 + M_2)x = y \qquad (2.5)$$

where we may apply $M_1^{-1} \in [Y,X]$, giving

$$(I_x + M_1^{-1}M_2)x = M_1^{-1}y \qquad (2.6)$$

Or we may replace x by $M_1^{-1}Z$ where $z = M_1 x$, giving

$$(I_y + M_2 M_1^{-1})z = y \qquad (2.7)$$

The identity operator $I_x \in [X,X]$, denoted $[X]$ and $I_y \in [Y]$.

Since M is invertible, error bounds of the form (2.3) may be recovered from (2.6) and (2.7). For example if it is known that

$M_1^{-1}(I_y + M_2M_1^{-1})^{-1} \in [Y,X]$ then

$$x - \tilde{x} = M_1^{-1}(I_y + M_2M_1^{-1})^{-1}(y-(M_1 + M_2)\tilde{x}) \qquad (2.8)$$

so that $||x - \tilde{x}|| \leq ||M_1^{-1}|| \cdot ||(I_y + M_2M_1^{-1})^{-1}|| \cdot ||(y-(M_1 + M_2)\tilde{x})||$

Because $M_1 \in [X,Y]$ is invertible there is a close relationship between the spaces X, Y and it is often the case that error bounds derived independently from (2.6) or (2.7) turn out to be equivalent when suitable practical norms are used.

§2.2 Setting for Theory

Since M_1 is invertible it permits a 1-1 correspondence between the elements of X and the elements of Y. It is neater to work in one space consisting of the elements of X or Y alone, using the equations (2.6), (2.7). It is not necessary, however, for Y and the space M_1X, for example, to have the same norm - it is only required that the norms are uniformly equivalent. With such norms the metric properties of Y and M_1X are the same, allowing much of the present published theory in this field to be generalised.

Let X_1, X_2 be normed linear spaces consisting of the same elements but with (possibly) different norms $||\cdot||_{X_1}$, $||\cdot||_{X_2}$ respectively, and let $[X_1, X_2]$ denote the space of bounded linear operators mapping $X_1 \to X_2$ with the subordinate norm. Both (2.6) and (2.7) may then be expressed, with appropriate choice of spaces,

$$\text{as} \quad (I_{12} - K)x = y \qquad y \in X_2 \qquad (2.9)$$

$$I_{12},\ K \in [X_1, X_2]$$

It will be seen later how advantage may be taken of allowing X_1 and X_2 to have different norms. We are not entirely free to choose these norms as we like because the two (distinct) identity operators $I_{12} : X_1 \to X_2$ and $I_{21} : X_2 \to X_1$ must be bounded. This also means that the metrics ρ_1, ρ_2 defined by

$$\rho_1(x,y) = ||x - y||_{x_1}$$

$$\rho_2(x,y) = ||x - y||_{x_2}$$

are uniformly equivalent. That is $\exists\, \alpha, \beta > 0$ such that

$$\alpha\rho_1(x,y) \leq \rho_2(x,y) \leq \beta\, \rho_1(x,y) \qquad \forall\, x,y \in X_1$$

This implies, letting $u = x - y$

$$\alpha||u||_{x_1} \leq ||u||_{x_2} \leq \beta\, ||u||_{x_1} \qquad \forall\, u \in X_1 \qquad (2.10)$$

Further, as the metrics ρ_1, ρ_2 are uniformly equivalent the open sets of X_1 are the same as the open sets of X_2. The definitions of closed set, closure, continunity, completeness and compactness can all be expressed in terms of open sets. This means that theorems using these properties in the setting of a normed linear space X and the space of bounded linear operators $[X]$ may be immediately generalised to the setting described at the beginning of this section. For this reason proofs of some theorems are given as references to the literature.

The distinction between the spaces X_1 and X_2 occurs only in the definition of the norms in these spaces. Any element $x \in X_1$ is also a member of X_2, and vice-versa. Further any operator $K \in [X_1, X_2]$ is a member of $[X_2, X_1]$, $[X_1]$ and $[X_2]$. An element x or operator K will, however, have different norms in each case. For example (2.10) gives

$$||x||_{x_2} \leq \beta\, ||x||_{x_1}$$

$$||x||_{x_1} \leq \frac{1}{\alpha}\, ||x||_{x_2}$$

Now consider the four identity operators

$$I_{11} \in [X_1] \ , \ I_{12} \in [X_1, X_2] \ , \ I_{21} \in [X_2, X_1] \text{ and } I_{22} \in [X_2].$$

Using the subordinate norm gives

$$\left.\begin{aligned} ||\ I_{11}\ || &= \sup_{||x||_{x_1} = 1} ||x||_{x_2} = 1 \\ ||\ I_{12}\ || &= \sup_{||x||_{x_1} = 1} ||x||_{x_2} \leq \beta \\ ||\ I_{21}\ || &= \sup_{||x||_{x_2} = 1} ||x||_{x_1} \leq \frac{1}{\alpha} \\ ||\ I_{22}\ || &= \sup_{||x||_{x_2} = 1} ||x||_{x_2} = 1 \end{aligned}\right\} \qquad (2.11)$$

Note also that

$$||\ I_{12}\ || \geq \alpha \text{ and } ||\ I_{21}\ || \geq \frac{1}{\beta}$$

Using these values and the following theorem the norm of an operator K may be bounded in any of the spaces $[X_1]$, $[X_1, X_2]$ $[X_2, X_1]$ $[X_2]$ given a bound in one.

Let A, B, C be normed linear spaces, not necessarily all different, and let K_{AB}, K_{BC} and K_{AC} be three linear operators in $[A, B]$, $[B, C]$ and $[A, C]$ respectively such that $K_{BC}\, K_{AB} = K_{AC}$. Then

$$||K_{AC}|| \leq ||K_{BC}|| \cdot ||K_{AB}|| \qquad (2.12)$$

For example $K_{12} = I_{12} K_{11}$ therefore

$$|| K_{12}|| \leq \beta || K_{11}||$$

It will usually be clear from the context in which space an operator is considered to be. If it is necessary to make a distinction for the purpose of calculating norms subscripts will be used as in (2.11) and the above example.

Most of the theory to be developed will require that certain operators are compact, although for some of the earlier theorems it is sufficient that X_1 and X_2 are complete.

§2.3 Definitions of Compactness

Let S be a subset of a normed linear space X, then S is compact iff every open cover of S has a finite subcover. S is said to be relatively compact iff $\overline{S}$ is compact. S is sequentially compact iff each sequence in S has a convergent subsequence (with limit in X). With these definitions relative and sequential compactness are equivalent.

When the space X is complete a useful result is that S is totally bounded iff S is relatively compact. (Anselone | p4)

An operator $K \in [X]$ is compact iff the set KB is relatively compact, where B is the unit ball in X : $B = \{x \in X : ||x|| \leq 1\}$. By the comments in the previous section if an operator K is compact in $[X]$ it is compact when considered as an operator between two spaces X_1 and X_2 consisting of the same elements and with uniformly equivalent norms.

§2.4 Theorems on operator inverses

In this section several theorems on operator inverses will be stated and proved. These theorems form the basis from which practical error bounds will be developed, as well as several weaker convergence results. Much use will be made of norms and the ideas of §2.2.

For simplicity we will assume throughout X_1 and X_2 are complete; for those parts of the theory where this condition is sufficient it will be restated for emphasis.

Theorem 2.1

If K is a compact operator in $[X_1, X_2]$ then the following three statements are equivalent.

(i) $(I-K)^{-1}$ exists and is bounded; $(I-K)^{-1} \in [X_2\ X_1]$

(ii) $(I-K)x = 0 \Rightarrow x = 0$

(iii) $\inf_{||x||_{X_1} = 1} ||(I-K)x||_{X_2} = M$, for some $M>0$

Proof

In the case $X_1 = X_2 = X$ is a standard result, see Appendix I of Anselone [1]. The proof generalises by the remarks of §2.2. Further we may deduce the following bound on $||(I-K)^{-1}||$.

Consider $(I-K)x \in X_2$ with $||(I-K)x||_{X_2} = 1$

Now $$||(I-K)x||_{X_2} = ||x||_{X_1} \cdot \left|\left| (I-K)\frac{x}{||x||_{X_1}} \right|\right|_{X_2}$$

But $$\left|\left|\frac{x}{||x||_{X_1}}\right|\right|_{X_1} = 1 \text{ and so} \left|\left| (I-K)\frac{x}{||x||_{X_1}}\right|\right|_{X_2} \geq M \text{ from (iii)}$$

$$||x||_{X_1} \leq \frac{1}{M}$$

Hence $\sup ||(I-K)^{-1}y||_{X_1} \leq \frac{1}{M}$

$||y||_{X_2} \leq 1$

The point in showing the above working in detail is to demonstrate that the factors α and β associated with the norms do not affect the bound on $||(I-K)^{-1}||$.

Theorem 2.2

If $K \in [X_1, X_2]$ is such that $||K_{12}|| < \alpha$ and either

(a) K is compact and X_1, X_2 are complete

or (b) X_1, X_2 are complete

Then $(I-K)$ has a unique inverse $(I-K)^{-1} \in [X_2, X_1]$ and

$$||(I-K)^{-1}|| \leq \frac{1}{\alpha - ||K_{12}||} \qquad (2.13)$$

Proof

In case (a) simply use Theorem 2.1 noting that for $x \in X_1$, $||x||_{X_1} = 1$ we have

$$||(I-K)x||_{X_2} \geq ||x||_{X_2} - ||Kx||_{X_2} \geq \alpha - ||K_{12}|| > 0$$

Note the appearance of the factor α associated with the operator I_{12}.

In case (b) we can show the convergence of the Neumann series $\sum_{m=0}^{\infty} K^m$ to $(I-K)^{-1}$.

Now $\sum_{m=0}^{\infty} K^m$ here is a mapping from X_2 to X_1 for this series to be absolutely convergent it is necessary that

$$\sum_{m=0}^{\infty} ||K_{11}^m I_{21}|| < \infty$$

Now $||I_{21}|| \leq \frac{1}{\alpha}$ and $||K_{11}|| = ||I_{21} K_{12}|| \leq \frac{||K_{12}||}{\alpha} < 1$

So $\sum_{m=o}^{\infty} ||K_{11}^m \ I_{21}|| \leq \frac{1}{\alpha} \frac{\alpha}{\alpha-||K_{12}||} = \frac{1}{\alpha-||K_{12}||} < \infty$ (2.14)

Since $\sum_{m=o}^{\infty} K_{11}^m \ I_{21}$ is absolutely convergent we can re-arrange the terms in the series

$$\left(\sum_{m=o}^{M} K_{11}^m \ I_{21} \right) \ (I_{12}-K_{12}) = I_{11} - K_{11}^{M+1} \to I_{11} \text{ as } M \to \infty$$

Similarly $\sum_{m=o}^{\infty} K_{11}^m \ I_{21}$ is a right inverse for $(I_{12}-K_{12})$.

Further from (2.14) we have

$$||(I_{12}-K_{12})^{-1}|| \leq \frac{1}{\alpha-||K_{12}||}$$

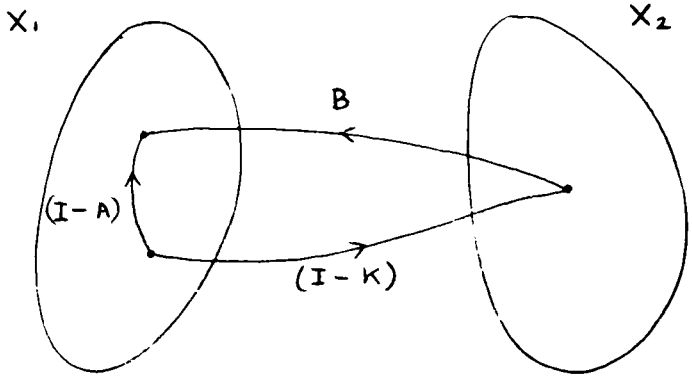
Theorem 2.3

Let $K \in [X_1, X_2]$, $B \in [X_2 \ X_1]$, $A \in [X_1]$. If $B\,(I-K) = I-A$ with $||A|| < 1$ and K and A are compact operators then

$$(I-K)^{-1} \in [X_2, X_1] \text{ and further}$$

$$||(I-K)^{-1}|| \leq \frac{||B||}{1-||A||} \qquad ||(I-K)^{-1}-B|| \leq \frac{||A|| \cdot ||B||}{1-||A||} \quad (2.15)$$

Proof



Note that $(I-A)^{-1} \in [X_1]$ and $||(I-A)^{-1}|| \leq \frac{1}{1-||A||}$

Thus $(I-K)$ has a left inverse $(I-A)^{-1} B \in [X_2, X_1]$. Th[illegible] unique inverse provided that the image of X_1 under the operator [illegible] is the whole space X_2. Anselone 1 for example show[illegible].

whenever K is a compact operator. X_1 and X_2 have the same topology and that result holds here also.

$$(I-K)^{-1} = (I-A)^{-1}B \tag{2.16}$$

So $$||(I-K)^{-1}|| \leq \frac{||B||}{1 - ||A||}$$

Also $$(I-K)^{-1}-B = ((I-A)^{-1}-I)B$$

$$= (I-A)^{-1}(I-(I-A))B$$

$$= (I-A)^{-1}AB$$

Hence $$||(I-K)^{-1}-B|| \leq \frac{||A|| \cdot ||B||}{1-||A||}$$

The conditions for the existence of $(I-K)^{-1}$ in this theorem do not depend on α or β because the operator I-A is in $[X_1]$. It must be remembered, however, that $B \in [X_2, X_1]$ and care taken that the correct norm is used for B. Note that we have maintained the assumption that X_1, X_2 are complete.

Theorem 2.4

Let, K, L $\in [X_1, X_2]$ and suppose $(I-L)^{-1} \in [X_2, X_1]$ and either

(a) K, L are compact and X_1, X_2 are complete.

(b) X_1, X_2 are complete.

If $\Delta o = ||(I-L)^{-1}(K-L)|| \leq 1$ then there exists

$$(I-K)^{-1} \in [X_2, X_1] \text{ and}$$

$$||(I-K)^{-1}|| \leq \frac{||(I-L)^{-1}||}{1 - \Delta o}; \quad ||(I-K)^{-1}-(I-L)^{-1}|| \leq \frac{\Delta o||(I-L)^{-1}||}{1 - \Delta o} \tag{2.17}$$

Proof

Consider $(I-K) = (I-L)(I-(I-L)^{-1}(K-L))$

In case (a) $(I-L)^{-1}(K-L)$ is compact. Apply Theorem 2.3 with $B = (I-L)^{-1}$ and $A = (I-L)^{-1}(K-L) \in [X_1]$ to give $(I-(I-L)^{-1}(K-L))^{-1} \in [X_1]$. Hence $(I-(I-L)^{-1}(K-L))^{-1}(I-L)^{-1}$ as the inverse of $(I-K)$.

In case of (b) apply Theorem 2.2 with $X_2 = X_1$ to show that $(I-(I-L)^{-1}(K-L))^{-1} \in [X_1]$ (since $\Delta o < 1$). It is then a matter of simple algebra to show that $(I-(I-L)^{-1}(K-L))^{-1}(I-L)^{-1}$ is a left and right inverse for $(I-L)$ and hence the unique inverse.

Again the comments at the end of Theorem 2.3 apply.

Theorem 2.5

Let $K, L \in [X_1, X_2]$ be compact and suppose $(I-L)^{-1} \in [X_2, X_1]$ If $\Delta d = ||(I-L)^{-1}(K-L)K^d|| < 1$ for some $d \geq 1$ then

$$(I-K)^{-1} \in [X_2, X_1] \text{ and}$$

$$||(I-K)^{-1}|| \leq \frac{||I+K+\ldots+K^{d-1}+(I-L)^{-1}K^d||}{1-\Delta d} \qquad (2.18)$$

Proof

Let $B = I+K+\ldots+K^{d-1} + (I-L)^{-1}K^d$ in Theorem 2.3, noting that $A = (I-L)^{-1}(K-L)K^d$ is compact. Theorem 4(a) can be interpreted as a special case of Theorem 2.5 putting $d = 0$.

clear It is not whether $(I-A)^{-1}B$ is the unique inverse of $(I-K)$ in the case when K, L are not compact.

The caution at the end of Theorem 2.3 is important here. That is $(I-L)^{-1}(K-L)K^d \in [X_1]$ and if we split this operator into two parts such as $(I-L)^{-1}$ and $(K-L)K^d$ for the purpose of evaluating norms it is necessary to be consistent, e.g. we could take

$(I-L)^{-1} \in [X_2, X_1]$ and $(K-L)K^d \in [X_1, X_2]$.

Further $B = I_{21} + K_{21} + \ldots K_{11}^{d-1} I_{21} + (I_{12}-L_{12})^{-1} K_{22}^d$ and the appropriate norms must be used.

Theorem 2.4 and 2.5 can be used to bound the inverse operator $(I-K)^{-1}$ given a sufficiently close approximate inverse $(I-L)^{-1}$. Indeed these are the theorems on which the error bounds in later chapters are based.

It is now shown how these theorems, applied with certain convergence results, can be used to give a-priori bounds on a sequence $||(I-K_n)^{-1}||$ of approximate inverse operators, and give a-posteriori bounds on $||(I-K)^{-1}||$ from bounds on $||(I-K_n)^{-1}||$ for an operator K_n sufficiently "close" to K.

§2.5 Collective Compactness

It is convenient to define here the notion of "collective compactness" introduced by Anselone [1]. This idea is closely associated with the convergence of sequences of compact operators.

There are two main types of operator convergence - convergence in norm and pointwise convergence.

Let $\{K_n\}$ be a sequence of operators; then we have convergence in norm of this operator sequence to some operator K provided

$$||K_n - K|| \to 0 \text{ as } n \to \infty.$$

This is a sufficiently strong convergence criterion to allow powerful convergence theorems to be deduced from the theorems of the previous section.

Another form of convergence - pointwise convergence - often arising from numerical approximations does not admit to such easy application of these theorems. Pointwise convergence of an operator sequence $\{K_n\}$ to K means that for any given $x \in X$

$$||K_n x - Kx|| \to 0 \text{ as } n \to \infty$$

Such convergence is denoted $K_n \to K$. Convergence in norm is a uniform convergence (independent of any particular x) it implies pointwise convergence, but not vice versa.

To derive certain convergence results from the theorems of the previous section we will require the convergence in norm of certain operator sequences. The idea of collective compactness is a convenient vehicle for relating pointwise and norm convergence in the context of the application of numerical methods to operator equations.

Definition

A set $K \subset [X]$ is collectively compact provided that the set $KB = \{Kx : K \in K, x \in B\}$ is relatively compact. A sequence of operators in $[X]$ is collectively compact whenever the corresponding set is.

The important result derived by Anselone linking pointwise and norm convergence is now stated.

Anselone's Corollary 1.9

Let $K, K_n \in [X]$, $n = 1,2,\ldots$ Assume $K_n \to K$ (pointwise) and $\{K_n\}$ is collectively compact. Then

$$||(K_n - K)K|| \to 0 \qquad ||(K_n - K)K_n|| \to 0$$

Note that convergence in norm of K_n to K is achieved on a compact

subset of X, not the whole of $\mathcal{B}$.

Another useful result given by Anselone is

$\{K_n\}$ collectively compact, $K_n \to K \Rightarrow K$ compact.

Also, trivially, any $K_n \in \{K_n\}$ is compact.

It is not <u>necessary</u> to show $\{K_n\}$ is collectively compact to obtain a suitable convergence in norm. In a particular case, for example, it may be known that the elements of $K\mathcal{B}$ satisfy a Lipshitz condition in which case it may be possible to verify directly that $||K_n - K||$ or $||(K_n - K)K||$ tends to zero with increasing n. rather than attempt to show that $\{K_n\}$is collectively compact, which could involve further assumptions or restrictions.

The various possible convergence results deriving from the above discussion and theorems 2.4 and 2.5 can now be combined and stated in one pair of theorems.

<u>Theorem 2.6A</u>

Let $K \in [X_1, X_2]$ and $\{K_n\}$a sequence of approximations to K with $K_n \in [X_1, X_2]$ $n = 1,2,\ldots$ and suppose $(I-K)^{-1} \in [X_2, X_1]$ and <u>either</u>

(a) X_1 is a Banach space ; $||K_n - K|| \to 0$

(b) K, K_n $(n = 1,2,\ldots)$ compact ; $||(K_n - K)K^d|| \to 0$ for some (fixed) $d \geq 0$.

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

then there exists $N \geq 1$ such that $\forall\, n \geq N$, $||(I-K_n)^{-1}|| \leq M$ for some $M > 0$

N.B. for (b) and (c) we still require X_1 complete.

Proof

(a) Apply Theorem 2.4 (b) with L=K, $K=K_n$.

$$\Delta o_n \leq ||(I-K)^{-1}|| \cdot ||K_n-K|| \to 0.$$

so that in this case

$$||(I-K_n)^{-1}-(I-K)^{-1}|| \to 0.$$

(b) Apply Theorem 2.5.

(c) Apply Theorem 2.5 + Anselones Corollary 1.9.

Theorem 2.6B

Let $K \in [X_1, X_2]$ and $\{K_n\}$ a sequence of approximations to K with $K_n \in [X_1, X_2]$ n= 1,2,... and $(I-K_n)^{-1} \in [X_2, X_1]$ in particular $||(I-K_n)^{-1}|| \leq M$ for all $n \geq N$.

and either

(a) X_1 is a Banach space : $||K_n-K|| \to 0$

(b) K, K_n (n=1,2,...) compact ; $||(K_n-K)K^d|| \to 0$ for some (fixed) $d \geq 0$.

(c) $\{K_n\}$ collectively compact ; $K_n \to K$.

then $(I-K)^{-1}$ exists and is bounded.

Proof

(a) Apply Theorem 2.4(b) with $L=K_n$,

$$\Delta o_n \leq M \cdot ||K_n-K|| \to 0$$

so that $||(I-K)^{-1}|| \leq M$.

(b) Apply Theorem 2.5

(c) Apply Theorem 2.5 + Anselone's Corollary 1.9.

These theorems provide the basis for the more specific convergence results in Chapter 3. More detailed examination of the norms of various terms, particularly Δ, will enable us to produce strict numerical bounds on the norm of the inverse $(I-K)^{-1}$.

CHAPTER 3

The Inverse Approximate Operator

The Inverse Approximate Operator

§3.1 Introduction

The previous chapter developed theorems connecting the operator I-K and its inverse with an approximate operator I-L. The major problem in applying these theorems lies in verifying the conditions of applicability. For example in Theorem 2.5 we need to show $\Delta_d < 1$ and if we split the norm into $\Delta_d \leq ||(I-L)^{-1}|| \cdot ||(K-L)K^d||$ we need to bound the inverse approximate operator and the term $(K-L)K^d$ - which is a measure of the difference between K and L.

The second term is dealt with in quite a straightforward way by the approximation theory associated with the operator L. Bounding the inverse approximate operator, however, presents something of a problem. In this chapter it is shown that fairly general bounds on the inverse approximate operator can be obtained for a class of approximations known as "projection methods", and some related schemes.

§3.2 Projection Methods

A projection in a normed linear space X is an operator $P \in [X]$ such that

$$P(Px) = Px \qquad \forall\ x \in X$$

PX is a subspace of X with the property

$$P\tilde{u} = \tilde{u} \qquad \forall\ \tilde{u} \in PX$$

Project (2.9) onto PX, giving

$$P((I-K)u-y) = 0$$

This may be simplified by seeking an approximate solution $\tilde{u} \in PX$

$$\tilde{u} - PK\tilde{u} = Py$$

$$\text{or} \quad (I - PK)\tilde{u} = PY \qquad (3.1)$$

This is one approximate equation arising in a very natural way from projection methods; it is not the only possibility. Cruickshank [11] suggested a variation analogous to the Nyström extension for integral equations. In that thesis it was shown how this extension could yield improved error bounds for approximate solutions of boundary value problems obtained by global polynomial collocation methods.

§3.3 The extended projection method

Suppose that the operator $(I-PK) \in [PX]$ has a unique inverse $(I-PK)^{-1} \in [PX]$. For any $y \in X$ define $\tilde{u}$ by

$$\left.\begin{aligned} \tilde{u} &= (I-PK)^{-1}Py \\ \text{and } z \text{ by} \quad z &= y + K\tilde{u} \end{aligned}\right\} \tag{3.2}$$

Now observe that

$$\begin{aligned} (I-KP)x &= (I-KP)(y+K\tilde{u}) \\ &= y+(I-KP)K(I-PK)^{-1}Py-KPy \\ &= y+K(I-PK)(I-PK)^{-1}Py-KPy \\ &= y. \end{aligned}$$

So that arising from the projection method there is another approximate operator equation.

$$(I-KP)z = y \tag{3.3}$$

Note that there is now simply a y on the right hand side as required in the approach of Anselone. The solution $\tilde{u}$ can be obtained from z by noting that from (3.1)

$$\tilde{u} = Py + PK\tilde{u}$$

$$= Pz \qquad (3.4)$$

It is not proposed to discuss in this chapter whether z is a more accurate solution than $\tilde{u}$*, the relevance of this solution lies in the structure of the approximate inverse from which it arises.

It is easy to express the inverse operator $(I-KP)^{-1} \in [X]$ in terms of the inverse projection operator $(I-PK)^{-1} \in [PX]$ as follows:

$$(I-KP)z = y$$

$$(I-KP)(y+K\tilde{u}) = y$$

$$(I-KP)(I+K(I-PK)^{-1}P)y = y$$

So that $I+K(I-PK)^{-1}P$ is a right inverse for $(I-KP)$. It is a matter of simple algebra to establish that this operator is also a left inverse of I-KP, hence the unique inverse.

$$(I-KP)^{-1} = I+K(I-PK)^{-1}P \qquad (3.5)$$

§3.4 A generalisation

Anselone, in his treatment of Integral Equations, derives an approximate operator K_n in terms of n linear - functionals on X. He obtains from common quadrature formulae, operators $K_n \to K$ (pointwise) which have certain similarities to the operator KP when P is a projection onto a finite dimensional subspace of X. A generalisation which includes Anselone's method and the "extended projection method" is now described.

*See Ch.4 §4.9

The extended projection method requires the solution of $(I-KP)z = y$ for $z \in X$. The solution obtained was

$$z = (I+K(I-PK)^{-1}P)y \tag{3.6}$$

when $(I-PK)^{-1} \in [PX]$. Since PX is in this case considered to be finite dimensional, it is complete and the existence of the inverse $(I-PK)^{-1}$ implies its uniqueness.

It is possible to obtain this solution in a more direct fashion as follows. Project (3.3) onto the subspace PX to give

$$\begin{aligned} & P(I-KP)z = Py \\ \Leftrightarrow \quad & (I-PK)Pz = Py \\ \Leftrightarrow \quad & Pz = (I-PK)^{-1}Py \quad \text{when } (I-PK)^{-1} \in [PX]. \end{aligned}$$

The solution z is retrieved from Pz by (3.2)

$$z = y + KPz \tag{3.7}$$

which is equivalent to (3.6). This latter approach is generalised by replacing K in (3.3) by an operator $K^* \in [PX,X]$. Again we solve for $z \in X$

$$(I-K^*P)z = y \tag{3.8}$$

The solution is given by

$$\left.\begin{aligned} & z = y + K^*Pz \\ \text{where} \quad & Pz = (I+PK^*)^{-1}Py \end{aligned}\right\} \tag{3.9}$$

wherever $(I-PK^*)^{-1} \in [PX]$.

Clearly the extended projection method is included simply by replacing K^* by K. We now show the connection with Anselone's work.

The action of K^* on PX can be described uniquely by its

action on a set of basis elements in PX. In practice K* will often be defined in such a manner. Let $\{\Phi_i\} = B$ be a (finite) basis for PX, and let $\tilde{u} \in PX$ so that

$$\tilde{u} = \sum_{\Phi_i \in B} a_i \Phi_i$$

Clearly $K^*\tilde{u} = K^* \sum_{\Phi_i \in B} a_i \Phi_i = \sum_{\Phi_i \in B} a_i (K^* \Phi i)$

Denote $K^*\Phi_i$ by k^*_i . $\{k^*_i\}$ is a (finite) set of elements of X which define the action of K* with respect to a basis $\{\Phi_i\}$ of PX.

Consider K* defined by $k^*_i = K\Phi i$. Then $K^*\tilde{u} = \Phi a_i K\Phi_i$ $= K\Sigma a_i \Phi_i = K\tilde{u}$. This is simply the extended projection method.

The treatment of integral equations by Anselone gives an example in which the subspace PX is not uniquely determined - many subspaces suffice to define the same approximate equation and solution. In fact, in the discussion given there, PX plays a very minor role. For the purpose of illustration we will consider one particular projection.

Let K be an integral operator defined on the space of continuous functions X by

$$Kx(s) = \int_{-1}^{1} k(s,t)x(t)\, dt$$

where $k(s,t)$ is continuous on the unit square $-1 \leq s,t \leq 1$. Define linear functionals on X

$$\Psi x = \int_{-1}^{1} x(t)\, dt$$

$$\Psi_n x = \sum_{i=1}^{n} W_{ni}\, x\,(t_{ni}) \qquad n = 1,2, \ldots$$

where $-1 \leq t_{ni} < t_{ni+1} \leq 1$ and $W_{ni} \in \mathbb{R}$. Anselone (p.18) defines

an approximation to the integral operator K

$$Kx(s) = \int_{-1}^{1} k(s,t)x(t)\,dt = \Psi_t\, k(s,t)x(t)$$

by $L_n\, x(s) = \sum_{i=1}^{n} W_{ni}\, K(s,t_{ni})x(t_{ni}) = \Psi_{nt}\, k(s,t)x(t)$

where the subscript t on Ψ indicates that it applies to the following expression considered as a function of t.

This approximation is included in the generalisation as now shown. Let $P_n x(t)$ be the piecewise linear interpolant to x(t) agreeing with x(t) at the points t_{ni}, linear in-between and constant for $t \leq t_{n1}$ and $t \geq t_{nn}$. Take as a basis for $P_n X$ the 'hat' functions $\Phi_{ni}(t) \in P_n X$ defined by

$$\Phi_{ni}(t_{nj}) = \delta_{ij}$$

Let $x \in X$, $P_n x$ is given by

$$P_n\, x\,(t) = \sum_{i=1}^{n} x(t_{ni}) \cdot \Phi_{ni}(t)$$

Now define K_n by the elements $k_i \in X \quad i = 1,2 \ldots$

$$k_i\,(S) = W_{ni}\, k\,(s,t_{ni})$$

Hence $K_n P_n x(s) = L_n x(s)$, demonstrating that this approximation does in fact belong to the generalisation of the projection method. Clearly any other interpolation projection through the same points could be used to obtain the same result. Anselone chooses this particular projection because in his spaces X, $P_n X$ it has the minimum norm (=1). This helps in maintaining a reasonably tight bound on $||(I-P_n K)^{-1}||$ in the space $[P_n X]$.

§3.5 Approximate Inverse in P_nX

The generalisation in §3.4 includes a very wide class of numerical methods for solving operator equations and it has been shown that the existence of $(I-K^*P)^{-1}$ depends crucially upon the existence of $(I-PK^*)^{-1}$ in the subspace $[PX]$. It is shown in this section that if P is a projection onto a finite dimensional subspace PX then the operator $(I-PK^*)$ has an inverse in the subspace $[PX]$ if and only if a certain matrix inverse exists. Further the norm of $(I-PK^*)^{-1}$ can be expressed in terms of a certain norm of the inverse matrix.

It is now shown how the equation in the n-dimensional subspace $[P_nX]$ is related to the solution of a finite system of linear algebraic equations. We use the vector space $\mathbb{R}^n$. The identity operator in $\mathbb{R}^n$ will be denoted by I_n and an element $\underline{v} \in \mathbb{R}^n$ in component form by a column vector

$$\underline{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

Let $\{\Phi_{ni}\}$ be a basis for P_nX. An element $\tilde{u} \in P_nX$ can be expressed as a linear combination of the basis elements Φ_{ni}

$$\tilde{u} = \sum_{i=1}^{n} u_i \Phi_{ni}$$

Define two bounded linear operators $H_n \in [P_nX, \mathbb{R}^n]$, $J_n \in [\mathbb{R}^n, P_nX]$ by

$$\left.\begin{aligned} (H_n \tilde{u})_i &= u_i \\ J_n \underline{u} &= \tilde{u} \end{aligned}\right\} \qquad (3.10)$$

The general equation in the subspace is

$$(I-P_nK^*)\tilde{u} = \tilde{y}$$

Because we have now fixed $[P_nX]$ as an n-dimensional subspace K^* becomes a finite dimensional operator in $[P_nX, X]$ and will be denoted by K_n.

$$(I-P_n K_n)\tilde{u} = \tilde{y} \tag{3.11}$$

The operator H_n and its inverse J_n give a 1-1 correspondence between the space P_nX and $\mathbb{R}^n$ so that

$$(3.11) \Leftrightarrow H_n (I-P_n K_n)\tilde{u} = H_n \tilde{y}$$

$$\Leftrightarrow (H_n - H_nP_nK_n)(J_n\underline{u}) = \underline{y}$$

$$\Leftrightarrow (H_n J_n - H_nP_nK_nJ_n)\underline{u} = \underline{y}$$

$$\Leftrightarrow (I_n - \bar{K}_n)\underline{u} = \underline{y} \tag{3.12}$$

where $$\bar{K}_n = H_nP_nK_nJ_n \tag{3.13}$$

So that $(I-P_nK_n)^{-1} \in [P_nX] \Leftrightarrow (I_n-\bar{K}_n)^{-1} \in [\mathbb{R}^n]$. Moreover when $(I_n - \bar{K}_n)^{-1}$ exists

$$\tilde{u} = J_n (I_n-\bar{K}_n)^{-1} H_n \tilde{y}$$

$$\left.\begin{aligned} \text{or} \quad & (I-P_nK_n)^{-1} = J_n(I_n-\bar{K}_n)^{-1}H_n \\ \text{and} \quad & (I_n - \bar{K}_n)^{-1} = H_n(I-P_nK_n)^{-1}J_n \end{aligned}\right\} \tag{3.14}$$

There exists an interesting alternative expression for

$(I_n - \bar{K}_n)^{-1}$ involving $(I - K_n P_n)^{-1}$. Note from (3.8), (3.9) that

$$P_n (I - K_n P_n)^{-1} = (I - P_n K_n)^{-1} P_n$$

so that

$$H_n P_n (I - K_n P_n)^{-1} J_n = H_n (I - P_n K_n)^{-1} P_n J_n$$

Extend H_n to X by $H_n = H_n P_n$ and note that $P_n J_n \equiv J_n$ so that

$$H_n (I - K_n P_n)^{-1} J_n = H_n (I - P_n K_n)^{-1} J_n$$

or
$$(I_n - \bar{K}_n)^{-1} = H_n (I - K_n P_n)^{-1} J_n \tag{3.15}$$

This argument can be extended as follows.

Let $J_n^* \in [\mathbb{R}^n, X]$ be such that

$$P_n J_n^* = J_n \tag{3.16}$$

Then $H_n P_n (I - K_n P_n)^{-1} J_n^* = H_n (I - P_n K_n)^{-1} P_n J_n^*$

$$H_n (I - K_n P_n)^{-1} J_n^* = H_n (I - P_n K_n)^{-1} J_n$$

giving

$$(I_n - \bar{K}_n)^{-1} = H_n (I - K_n P_n)^{-1} J_n^* \tag{3.17}$$

§3.6 Bounds on the approximate inverse

To proceed to investigate bounds on $||(I - P_n K_n)^{-1}||$ in detail requires explicit definition of X_1 and X_2, in particular the definition of the norms in these spaces, and is better left to a discussion of the application of the theory to a particular example.

Nevertheless there are some results that can be derived for quite a general class of norms which at least illustrate what can be achieved.

Suppose that the norm of an element $\tilde{u} \in P_n X$ can be related in the following manner to a norm based on the components with respect to the basis $\{\phi_{ni}\}$

$$\mu_n ||\tilde{u}||_p \leq ||\tilde{u}|| \leq \nu_n ||\tilde{u}||_p \qquad (3.18)$$

where $||\tilde{u}||_p = \left(\sum_{i=1}^{n} |u_i|^p . ||\phi_{ni}||^p \right)^{1/p} \qquad p > 0$

It is easy to verify that $||.||_p$ satisfies the properties required of a norm. Note that μ_n, ν_n (may) depend on n. Suppose further that the norms in the spaces $P_n X_1$ and $P_n X_2$ are related as follows

$$||\phi_{ni}||_{x_1} = W_{ni} ||\phi_{ni}||_{x_2} \qquad W_{ni} > 0 \quad i=1,2,\ldots n \qquad (3.19)$$

N.B. This in no way defines the norms - it is merely a property of the norms induced on $P_n X_1$ from X_1. We are now in a position to relate the norm of an element in $P_n X_1$ to that in $P_n X_2$. Denote the $||\cdot||_p$ norm of an element in X_1 by $||\cdot||_p^1$.

In $P_n X_1$

$$||\tilde{u}||_p^1 = \left(\sum_{i=1}^{n} |u_i|^p ||\phi_{ni}||_{x_1}^p \right)^{1/p}$$

$$= \left(\sum_{i=1}^{n} |u_i|^p . W_{ni} ||\phi_{ni}||_{x_2}^p \right)^{1/p}$$

$$= \left(\sum_{i=1}^{n} |W_{ni} u_i|^p . ||\phi_{ni}||_{x_2}^p \right)^{1/p} \qquad (3.20)$$

It remains to verify for these norms that (2.10) holds.

Multiply (3.20) by $\frac{1}{\max W_{ni}}$

$$\frac{1}{\max W_{ni}} ||\tilde{u}||_p^1 = \frac{1}{\max W_{ni}} \left(\sum_{i=1}^{n} |W_{ni} u_i|^p . ||\phi_{ni}||_{x_2}^p \right)^{1/p}$$

$$= \left(\sum_{i=1}^{n} \left| \frac{W_{ni}}{\max W_{ni}} u_i \right|^p . ||\Phi \quad ||_{X_2}^p \right)^{1/p} \cdot$$

$$\leq \left(\sum_{i=1}^{n} |u_i|^p . ||\Phi_{ni}||_{X_2}^p \right)^{1/p} = ||\tilde{u}||_p^2$$

So that $||\tilde{u}||_p^1 \leq \max W_{ni} \; ||\tilde{u}||_p^2$ (3.21)

Now multiply (3.20) by $\frac{1}{\min W_{ni}}$ to give

$$\frac{1}{\min W_{ni}} ||\tilde{u}||_p^1 = \left(\sum_{i=1}^{n} \left| \frac{W_{ni}}{\min W_{ni}} u_i \right|^p . ||\Phi_{ni}||_{X_2}^p \right)^{1/p}$$

So that $||\tilde{u}||_p^1 \geq \min W_{ni} \; ||\tilde{u}||_p^2$ (3.22)

Inequalities (3.18), (3.21), (3.22) can now be combined to give

$$\frac{\mu_n}{\nu_n \max W_{ni}} ||\tilde{u}||_{X_1} \leq ||\tilde{u}||_{X_2} \leq \frac{\nu_n}{\mu_n \min W_{ni}} ||\tilde{u}||_{X_1} \qquad (3.23)$$

as required.

It is a simple procedure to relate the norm of $(I-P_nK_n)^{-1}$ to that of $(I_n-\bar{K}_n)^{-1}$ using this p-norm, as follows.

Assume that the basis $\{\Phi_{ni}\}$ has been normalised in X_2 - this is not necessary but makes the algebra tidier.

Then $||\tilde{u}||_p^2 = ||\underline{u}||_p = (\Sigma|u_i|^p)^{1/p}$

Now $||(I-P_nK_n)^{-1}|_{P_nX_2}|| = \sup_{\tilde{u} \in P_n X_2 \; ||\tilde{u}||_{X_2} \leq 1} ||(I-P_nK_n)^{-1}\tilde{u}||_{X_1}$

But $||\tilde{u}||_{X_2} \leq 1 \Rightarrow ||\tilde{u}||_p^2 \leq \frac{1}{\mu_n} \Rightarrow ||\underline{u}||_p \leq \frac{1}{\mu_n}$

i.e. $||H_n\tilde{u}||_p \leq \frac{1}{\mu_n}$

Using the p-norm for the inverse matrix $(I_n-\bar{K}_n)^{-1}$ now gives

$$||(I_n-\bar{K}_n)^{-1} H_n \tilde{u}||_p \leq ||(I_n-\bar{K}_n)^{-1}||_p \frac{1}{\mu_n}$$

from which we can proceed if the coarse bound in (3.23) is used

to relate the norm for $(I-P_nK_n)^{-1} \in [P_nX_2]$ to that in $[P_nX_2, P_nX_1]$ using the ideas in §2.2. It is far better, however, to notice that this is a convenient stage to introduce the "weighted" norm $||\cdot||_p$ which will give directly the norm of $(I-P_nK_n)^{-1}$ considered as an operator in $[P_nX_2, P_nX_1]$.

The norm $||\cdot||_p^1$ is found in the same way as the norm $||\cdot||_p^2$ modified by multiplying the elements of the corresponding vector norms by the weights W_{ni}, i.e.

$$||\tilde{u}||_p^1 = \left(\sum_{i=1}^{n} \left| W_{ni}\, u_i \right|^p \right)^{1/p} \; ; \; ||\tilde{u}||_p^2 = \left(\sum_{i=1}^{n} \left| u_i \right|^p \right)^{1/p}$$

Now $((I_n-\bar{K}_n)^{-1} H_n \tilde{u})_i = \sum_{j=1}^{n} (I_n-\bar{K}_n)^{-1}_{ij} (H_n \tilde{u})_j$

so that $W_{ni} ((I_n-\bar{K}_n)^{-1} H_n \tilde{u})_i = \sum_{j=1} W_{ni} (I_n-\bar{K}_n)^{-1}_{ij} (H_n \tilde{u})_j$

Define the matrix $W_{ij} = W_{ni} (I_n-\bar{K}_n)^{-1}_{ij}$

$$||(I_n-\bar{K}_n)^{-1} H_n \tilde{u}||_p^1 \leq ||W||_p \; \frac{1}{\mu_n}$$

so finally $||(I-P_nK_n)^{-1}|| \leq \frac{\nu_n}{\mu_n} ||W||_p$ (3.24)

Note that this bound for $||(I-P_nK_n)^{-1}||$ applies only in the subspace $[P_nX_2, P_nX_1]$. The following identity can be used to extend the result to $[X_2, X_1]$.

$$(I-P_nK_n)^{-1} = I + (I-P_nK_n)^{-1} P_nK_n \qquad (3.25)$$

Hence $||(I-P_nK_n)^{-1}|| \leq$ (norm in subspace) $\cdot\, ||P_nK_n||$ where $(I-P_nK_n)^{-1}$ and P_nK_n are now considered as operators on X_2.

The foregoing analysis can be expressed more concisely as follows.

$$(I-P_nK_n)^{-1} = J_n(I_n-\bar{K}_n)^{-1}H_n \quad \text{from} \quad (3.14)$$

Let $\mathbb{R}_2^n$ denote the space $\mathbb{R}^n$ with the p norm

$$||x||_{\mathbb{R}_2^n} = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

Let $\mathbb{R}_1^n$ denote the space $\mathbb{R}^n$ with weighted p norm

$$||x||_{\mathbb{R}_2^n} = \left(\sum_{i=1}^{n} |W_{ni}\, x_i|^p\right)^{1/p}$$

Consider $H_n \in \left[P_nX_2, \mathbb{R}_2^n\right]$ then $||H_n|| \leq \frac{1}{\mu_n}$ from (3.18)

$$||(I_n-\bar{K}_n)^{-1}|| = ||W||_p \text{ by definition.}$$

And $J_n \in \left[\mathbb{R}_1^n, P_nX\right]$ so that $||J_n|| \leq \nu_n$ from (3.18).

The fact that P_nX_1 has a weighted norm does not alter this bound on $||J_n||$ because the weighting is component wise and is removed by virtue of the range space being $\mathbb{R}_1^n$ rather than simply $\mathbb{R}^n$.

Hence

$$||(I-P_nK_n)^{-1}|| \leq ||J_n||\cdot||(I_n-\bar{K}_n)^{-1}||\cdot||H_n||$$

$$\leq \frac{\nu_n}{\mu_n}||W||_p \qquad (3.24)'$$

From (3.14) we also have

$$||(I_n-\bar{K}_n)^{-1}|| \leq ||H_n||\cdot||(I-P_nK_n)^{-1}||\cdot||J_n|| \qquad (3.26)$$

$$\left(\text{Hence} \quad ||W||_p \leq \frac{\nu_n}{\mu_n}||(I-P_nK_n)^{-1}||\right)$$

Note that here $(I-P_nK_n)^{-1} \in \left[P_nX_2, P_nX_1\right]$.

Another bound for $||W||_p$ can be obtained which involves an operator defined on X_2 rather than P_nX_2 and which offers some flexibility in the choice of operator J_n.

Recall (3.17) we can choose any $J_n^* \in \left[\mathbb{R}^n, X\right]$ which satisfies

$P_n J_n^* \equiv J_n$ and write

$$(I_n - \bar{K}_n)^{-1} = H_n (I - K_n P_n)^{-1} J_n^*$$

Now consider $J_n^* \in [\mathbb{R}_1^n, X_1]$ and $||J_n^*|| \leq \nu_n^*$, we have

$$(I_n - \bar{K}_n)^{-1} || \leq ||H_n|| \cdot ||(I - K_n P_n)^{-1}|| \cdot ||J_n^*|| \qquad (3.27)$$

$$\left(\text{Hence } ||W||_p \leq \frac{\nu_n^*}{\mu_n} \; ||(I - K_n P_n)^{-1}|| \right)$$

§3.7 The Behaviour of $||W||$

In the previous section certain bounds on $||W||$ were obtained using the weighted p norm. A striking feature of $||W||$ observed in examples in later chapters is the apparent convergence to a certain value. Further, for those examples, $||W||$ seems to approach the same constant irrespective of the interpolation scheme used (e.g. polynomial interpolation at Chebychev zeros, Legendre zeros, piecewise polynomial interpolation, etc.). Implied in this observation is the use of the same norm in each scheme and for each degree of approximation. In fact to study convergence of $||W||$ we need to fix the norms in X_1 and X_2 and we will drop the subscript p of the previous chapter to indicate this. The suspicion arose from the observed behaviour of $||W||$ that perhaps $||W|| \to ||(I - K)^{-1}||$ but this is not easy to prove in general - we prove it is the case when the conditions in Theorem 7 (to follow) are satisfied.

Certain basic assumptions will first be stated which ensure that the theorems in Chapter 2 apply. (X_1 and X_2 are fixed)

(a) K is compact

(b) $(I-K)^{-1}$ exists (and is bounded)

(c) $\{K_n P_n\}$ collectively compact

(d) $K_n P_n \to K$

$\Rightarrow$ (a)

Recall the definition of the matrix $W_n = (I_n - \tilde{K}_n)^{-1}$

$$W_n = H_n (I-K_n P_n)^{-1} J_n^* \quad \text{from (3.17)}$$

Now suppose that H_n and J_n^* satisfy the following conditions in addition to those stated in their definitions.

$$\left.\begin{array}{ll} \text{(i)} & ||H_n|| = 1 \\ \text{(ii)} & ||J_n^* u|| = ||u|| \quad \forall u \in \mathbb{R}^n \ (\Leftrightarrow ||J_n^*|| = 1) \\ \text{Let} \quad S_n = J_n^* H_n & \text{(i)+(ii)} = ||S_n|| \leq 1 \\ \text{(iii)} & ||(I-S_n)x|| \to 0 \quad \forall \underline{\text{fixed}}\ x \in X. \end{array}\right\} \qquad (3.28)$$

Choose $y_\varepsilon \in X$ with $||y_\varepsilon|| \leq 1$ such that $||(I-K)^{-1}|| - ||(I-K)^{-1} y_\varepsilon|| < \varepsilon$

y_ε will be considered <u>fixed</u> below.

Let $\underline{y}_{n\varepsilon} = H_n y_\varepsilon$

$$\underline{z}_n = W_n \underline{y}_{n\varepsilon}$$

Consider $||J_n^* \underline{z}_{n\varepsilon} - (I-K)^{-1} y_\varepsilon||$

$$= ||J_n^* \underline{z}_{n\varepsilon} - S_n((I-K)^{-1} y_\varepsilon - (I-K)^{-1} y_\varepsilon) - (I-K)^{-1} y_\varepsilon||$$

$$\leq ||J_n^* \underline{z}_{n\varepsilon} - S_n (I-K)^{-1} y_\varepsilon|| + ||(I-S_n)(I-K)^{-1} y_\varepsilon||$$

$$= ||S_n (I-K_n P_n)^{-1} S_n y_\varepsilon - S_n (I-K)^{-1} y_\varepsilon|| + ||(I-S_n)(I-K)^{-1} y_\varepsilon||$$

$$= ||S_n (I-K_n P_n)^{-1} (S_n - I) y_\varepsilon + S_n \{(I-K_n P_n)^{-1} y_\varepsilon - (I-K)^{-1} y_\varepsilon\}||$$

$$+ ||(I-S_n)(I-K)^{-1} y_\varepsilon||$$

$$\leq ||S_n (I-K_n P_n)^{-1} (I-S_n) y_\varepsilon|| \qquad \text{(A)}$$

$$+ ||S_n \{(I-K_n P_n)^{-1} y_\varepsilon - (I-K)^{-1} y_\varepsilon\}|| \qquad \text{(B)}$$

$$+ \|(I-S_n)(I-K)^{-1}y_\varepsilon\| \qquad (C)$$

Consider (C); $(I-K)^{-1}y_\varepsilon \in X$ is fixed so by (iii) (C) $\to 0$.

Consider (A); Theorem 2.6A tells us for $n \geq N, \|(I-K_nP_n)^{-1}\| < M$

Further $\|S_n\| \leq 1$ and $\|(I-S_n)y_\varepsilon\| \to 0$ hence (A) $\to 0$.

Consider (B); Theorem 1.10 (Anselone P8) (Also see §4.9) gives

$$\|(I-K_nP_n)^{-1}y_\varepsilon-(I-K)^{-1}y_\varepsilon\|$$

$$\leq \frac{\|(I-K)^{-1}\|\cdot\|K_nP_ny_\varepsilon - Ky_\varepsilon\| + \Delta\|(I-K)^{-1}y_\varepsilon\|}{1-\Delta}$$

if $\Delta = \|(I-K)^{-1}(K_nP_n-K)K_nP_n\| < 1$

From conditions (b), (c), (d) above, $\Delta \to 0$; also $(K_nP_n-K)y_\varepsilon \to 0$, giving $\|(I-K_nP_n)y_\varepsilon-(I-K)^{-1}y_\varepsilon\| \to 0$. Again $\|S_n\| \leq 1$ so (B) $\to 0$

Finally $\|J_n*\underline{z}_{n\varepsilon} - (I-K)^{-1}y_\varepsilon\| \to 0.$ (3.29)

Now $\|\underline{z}_{n\varepsilon}\| \leq \|W_n\|\cdot\|y_{n\varepsilon}\|$ but $\|y_{n\varepsilon}\| \leq \|y_\varepsilon\| \leq 1$

therefore using (ii) and (3.29)

$$\|W_n\| \geq \|\underline{z}_{n\varepsilon}\| = \|J_n*\underline{z}_{n\varepsilon}\| \to \|(I-K)^{-1}y_\varepsilon\|$$

ε is arbitrary so in the limit $\|W_n\|$ is bounded below by $\|(I-K)^{-1}\|$.

Suppose that in addition to conditions (i), (ii) and (iii) P_nK_n and J_n, J_n* satisfy the following

$$\left.\begin{array}{ll} \text{(iv)} & \|P_nK_n - K\| \to 0 \\ \text{(v)} & \|P_nK_n(J_n-J_n*)\| \to 0 \end{array}\right\} \qquad (3.30)$$

Then we can derive an asymptotic upper bound on $||W||$ as follows

Recall (3.17), $W_n = H_n (I-P_nK_n)^{-1} J_n$

Let $$W_n^* = H_n (I-P_nK_n)^{-1} J_n^* \tag{3.31}$$

Consider $W_n - W_n^* = H_n(I-P_nK_n)^{-1}(J_n-J_n^*)$

$$= H_n(I-(I-P_nK_n)^{-1}P_nK_n)(J_n-J_n^*)$$

$$= H_n(I-P_nK_n)^{-1}(P_nK_n)(J_n-J_n^*)$$

Since $H_nJ_n = H_nJ_n^* = I_n$

From condition (iv) and Theorem 4 we have

$$||(I-P_nK_n)^{-1} - (I-K)^{-1}|| \to 0 \tag{3.32}$$

So that $||(I-P_nK_n)^{-1}|| \to ||(I-K)^{-1}||$

Applying condition (v) and (i) we have

$$||W_n - W_n^*|| \to 0 \tag{3.33}$$

But $||W_n^*|| \leq ||(I-P_nK_n)^{-1}|| \to ||(I-K)^{-1}||$

So that in the limit $||W_n||$ is bounded above by $||(I-K)^{-1}||$.

Together with the asymptotic lower limit on $||W_n||$ we have

Theorem 3.7

(In the context of previous chapters) Suppose

(b) $(I-K)^{-1}$ exists

(c) $\{K_nP_n\}$ collectively compact

(d) $K_nP_n \to K$

(i) $||H_n|| = 1$

(ii) $||J_n^*|| = 1$

(iii) $||(I-J_n^*H_n)x|| \to 0$ for any fixed $x \in X$

(iv) $||P_nK_n - K|| \to 0$

(v) $||P_nK_n(J_n-J_n^*)|| \to 0$

Then $||W_n|| \to ||(I-K)^{-1}||$

We shall find that all the conditions of Theorem 3.7 are satisfied by the problem examined in Chapter 4.

We can obtain the result of Theorem 3.7 from the result in §3.6 in the particularly simple case of $||J_n|| = 1$; $||H_n|| = 1$ plus (b), (c), (d), (iv) for then (3.24)' + (3.26) gives

$$||W_n|| = ||(I-P_nK_n)^{-1}|| \to ||(I-K)^{-1}||$$

Note also from (3.32) and (3.33)

$$||W_n - H_n (I-K)^{-1} J_n^*|| \to 0 \qquad (3.34)$$

$W_n \in [\mathbb{R}^n]$ is a discrete analogue of the operator $(I-K)^{-1}$ on the space $J_n^* \mathbb{R}^n$.

CHAPTER 4

Theory of Application to Two Point Boundary Value Problems

Theory of Application to Two Point Boundary Value Problems

§4.1 Introduction

Chapters 2 and 3 have developed an abstract theory concerning approximate operator inverses and bounds. The theory can be applied to any problems for which the conditions of the theorems are satisfied. The rest of this thesis will be concerned with the application to linear two point boundary value problems of the second order in ordinary differential equations. In this chapter we will define the problem precisely, define the approximations that will be studied and verify the conditions of the theorems in previous chapters. At the end of the chapter various error estimates will be discussed which involve somewhat less work than the strict error bounds for the examples in Chapter 5.

The theory could be applied to higher order problems with various boundary conditions and 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. Error estimates, however, could be produced in a similar manner to this chapter with far less effort.

Although the theory does not apply directly to non-linear problems, the results obtained in the linear case could form a useful starting point for further investigations.

We use the space of Riemann integrable functions for X to allow the use of piecewise polynomial approximations.

§4.2 Form of Problem

We consider problems of the form

$$\left.\begin{aligned} x''(t) + p(t)\,x'(t) + q(t)\,x(t) &= y(t) \\ x(-1) = x(1) &= 0 \end{aligned}\right\} \quad (4.1)$$

with $p, q, y \in C[-1, 1] \subset R[-1, 1]$ (Riemann integrable on $[-1, 1]$)

$x \in R^{(2)}[-1, 1]$

The interval $[-1, 1]$ will be dropped from further notation. In order to develop strict bounds we shall generally require that p, q, y have a bounded modulus of continuity or possess higher order of continuity than C.

The equation (4.1) can be expressed in the form (2.5) by noting that D^2 is invertible with the boundary conditions in (4.1). The inverse can be expressed as the well known integral operator

$$(G\,x)(s) = \int_{-1}^{1} g(s,t)\,x(t)\,dt \quad (4.2)$$

$$\text{where } g(s,t) = \begin{cases} \tfrac{1}{2}(s+1)(t-1) & s < t \\ \tfrac{1}{2}(s-1)(t+1) & s \geq t \end{cases} \quad (4.3)$$

Then $D^2 G\,x \equiv x$ almost everywhere for $x \in R$

$GD^2\,x \equiv x$ for $x \in R^{(2)}$ satisfying the boundary conditions. Keller (26, p108) gives the Greens functions (4.3) for a variety of differential operators and boundary conditions.

Consider the integral operator K defined by

$$(Kx)\,s = \int_{-1}^{1} k(s,t)x(t)\,dt$$

$$\text{where } -k(s,t) = p(s)\,\frac{\partial g}{\partial s}(s,t) + q(s)\,g(s,t) \quad (4.4)$$

If x satisfies the boundary conditions of (4.1) it can be shown

that

$$-(Kx'')_s = -\int_{-1}^{1} k(s,t)\,\frac{d^2x}{dt^2}\,dt = p(s)\,x'(s) + q(s)x(s),\ x'' \in R \tag{4.5}$$

and equation (4.1) can be expressed in the integral form

$$(I-K)\,u = y \tag{4.6}$$

where $u = x'' \in R$.

An alternative expression for K which will prove useful later is now given. Define the operator H as follows

$$\left(Hx\right)(s) = \int_{-1}^{s} x(t)dt = \int_{-1}^{1} h(s,t)\,x(t)\,dt \tag{4.7}$$

$$\text{where } h(s,t) = \begin{cases} 1 & t \le s \\ 0 & ,\ t > s \end{cases}$$

In a similar manner to deriving (4.6) we have

$$\left(Kx\right)(s) = p(s)\left(Hx\right)(s) + q(s)\left(H^2x\right)(s) - \left(\frac{p(s)}{2} + q(s)\,\frac{(s+1)}{2}\right)(H^2x)(1) \qquad \text{for } x \in R \tag{4.8}$$

The usual sup norm will be used on R, so that R is a Banach space. R takes the place of the space X in chapters 2 and 3. Since most of this chapter will not be concerned with the details of using a weighted sup norm we need make no distinction between the spaces X_1 and X_2 for the present.

§4.3 The method of collocation

The method of collocation requires that an approximate solution $\tilde{x}$ satisfies the equation to be solved exactly at a finite set of points. The approximate solution will generally be an element of a finite dimensional subspace of R- yielding a finite set of linear equations to solve for $\tilde{x}$.

Suppose that the equation we wish to solve is

$$(I-K)x=y \qquad x,\ y \in R, \quad K \in [R].$$

And the approximate solution $\tilde{x} \in \tilde{R} \subset R$ is obtained by requiring

$$((I-K)\tilde{x})(s_i) = y(s_i) \qquad i=1,\ldots,m \ ; \ s_i \in [-1,\ 1].$$

This is equivalent to using an interpolation projection method to solve the equation

$\tilde{R}$ has an m dimensional basis $\{\Psi i\}$; construct a new basis $\{\Phi i\}$ such that

$$\Phi i\ (s_j) = \delta_{ij} \qquad i,j = 1,\ldots,m$$

we need m distinct s_j in order to form this basis.

Define a mapping P_m as follows

$$\left(P_m\ x\right)(t) = \sum_{i=1}^{m} \Phi_i(t)\ .\ x\ (s_i)$$

P_m is obviously a projection and gives rise to the same approximate solution as the collocation method - albeit from a different set of linear algebraic equations because of a change of basis perhaps.

§4.4 Piecewise polynomial method

We define here a projection of R onto a space of piecewise polynomials on $[-1,1]$. Define the partition T_n of $[-1,1]$ by the points t_i $\quad i=0,\ldots,n$

$$-1 = t_0 < t_1 < \ldots < t_n = 1 \tag{4.10}$$

On each of the intervals $[t_{i-1},t_i)$ $\quad i=1,\ldots,n-1$, $[t_{n-1},t_n]$ the approximation will consist of a polynomial in t. The order of each of these polynomials will be the same and each polynomial will be defined by interpolation within the corresponding interval on a set of points, the relative distribution of which will remain the same for all intervals. No conditions of continuity will be imposed from one interval to the next, i.e. we do not assume that

$$\tilde{x}\ (t_i - 0) = \tilde{x}\ (t_i + 0)$$

at any partition point t_i .

Let $\{\xi_j\}$ be a set of p distinct points in $[-1,1]$, (including, possibly, the end points). The collocation, or interpolation points, will be defined in each interval $[t_{i-1}, t_i]$ by

$$\xi_{ji} = t_{i-1} + \frac{(1+ \xi_j)}{2} (t_i - t_{i-1}) \qquad (4.11)$$

$$i=1,\ldots,n \ ; \ j=1,\ldots,p$$

The projection P_{np} is defined on each interval $[t_{i-1}, t_i]$ by

$$\left(P_{np}\ x\right)(t) = \sum_{j=1}^{p} x\ (\xi_{ji})\ L_{ji} \qquad (4.12)$$

for i=1,...,n.

L_{ji} is the (unique) polynomial such that $L_{ji}\ (\xi_{ki})=\delta_{kj}$ for each i=1,...,n. L_j will denote the (unique) polynomial such that $L_j\ (\xi_i) = \delta_{ij}$.

Note that the norm of P_{np} is given by the usual polynomial projection norm corresponding to interpolation at the points ξ_j and is independent of n.

$$||P_{np}|| = \max_i \sup_{t\in[t_{i-1},t_i]} \sum_{j=1}^{p} \left|L_{ji}(t)\right|$$

$$= \sup_{t\in[-1,1]} \sum_{j=1}^{p} \left|L_j(t)\right| = ||P_p||$$

Define

$$l_j = \sum_{i=1}^{j} \int_{-1}^{1} L_i(t)\ dt$$

for j=1,...,p and l_o=0

In order to satisfy conditions of theorems in chapters 2 and 3 we shall restrict the projection P_p and P_{np} to those having the following properties

$$\text{(i)} \quad l_{j-1} \leq \xi_j + 1 \leq l_j \qquad j = 1,\ldots,p$$

$$\text{(ii)} \quad ||T_n|| = \max_{i=1,\ldots,n} |t_i - t_{i-1}| \to 0 \text{ as } n \to \infty \tag{4.13}$$

The geometric significance of (i) and (ii) may not be immediately obvious. The second condition simply ensures that as n increases the width of the polynomial pieces decreases. There is no restriction at this stage on the manner in which this happens. Condition (i) means that each of the polynomial interpolation points ξ_j of P_p lies in the interval $\left[l_{j-1}-1, l_j-1\right]$ where the l_j are the sum of the integration weights (positive) for the points ξ_1 to ξ_{j-1}. This ensures that we can determine a direct correspondence (Lemma 4.1) between P_{np} and a certain Riemman sum (S_{np}).

Define a projection S_p on R as follows

$$\left(S_p\, x\right)(t) = \sum_{j=1}^{p} x\left(\xi_j\right) X\left[l_{j-1}-1,\, l_j-1\right](t) \tag{4.14}$$

where $X\left[a,b\right](t) = \begin{cases} 1 & a \leq t < b \\ 0 & \text{elsewhere} \end{cases}$

$S_p\, x$ is a piecewise constant function.

Lemma 4.1

$$\int_{-1}^{1} \left(P_p\, x\right)(t)\, dt = \int_{-1}^{1} \left(S_p\, x\right)(t)\, dt \tag{4.15}$$

Proof

$$\int_{-1}^{1} \left(P_p\, x\right)(t)\, dt = \int_{-1}^{1} \sum_{j=1}^{p} x\left(\xi_j\right) L_j(t)\, dt$$

$$= \sum_{j=1}^{p} x\left(\xi_j\right) \int_{-1}^{1} L_j(t)\, dt$$

Similarly $\int_{-1}^{1} \left(S_p\, x\right)(t)\, dt \quad \sum_{j=1}^{p} x\left(\xi_j\right) \int_{-1}^{1} \chi\ \left[l_{j-1}-1,\ l_j-1\right](t)\, dt$

but $\int_{-1}^{1} \chi\left[l_{j-1}-1,\ l_{j-1}\right](t)\, dt = l_j - l_{j-1} = \int_{-1}^{1} l_j(t)\, dt$ giving (4.15)

Lemma 4.2

$$\int_{-1}^{1} \left(P_{np}\, x\right)(t)\, dt \to \int_{-1}^{1} x(t)\, dt \quad \text{for any } x \in R \qquad (4.16)$$

Proof

Extend S_p in the obvious manner to S_{np} on the partition T_n, then

$$\int_{-1}^{1} \left(P_{np}\, x\right)(t)\, dt = \int_{-1}^{1} \left(S_{np}\, x\right)(t)\, dt$$

Then note that since $||T_n|| \to 0$, $\left(S_{np}\, x\right)(t)$ is a Riemann sum so that

$$\int_{-1}^{1} \left(S_{np}\, x\right)(t)\, dt \to \int_{-1}^{1} x(t)\, dt \quad \text{for } x \in R$$

We now introduce some notation in order to make use of §2.10 in Anselone 1 .

For any kernel $k(s,t)$ define the functions $k_s(t)$ and $k^t(s)$ as follows (cf Anselone §2.8)

$$k_s(t) = k^t(s) = k(s,t)$$

A set F of functions $x(t)$, $-1 \le t \le 1$ is <u>regular</u> if for each $x \in F$ and each $m = 1,2,\ldots$ there exist x_m, $x^m \in C$ such that

$$x_m(t) \le x(t) \le x^m(t) \qquad t \in [-1,1]$$

$\int_{1}^{1} \left(x^m(t) - x_m(t)\right) dt \to 0$ uniformly for $x \in F$ as $m \to \infty$ and for each fixed m, the sets

$$F_m = \{x_m : x \in F\},\ F^m = \{x^m : x \in F\}$$

are totally bounded.

Regular sets have the following properties. Subsets and closures of regular sets are regular, the convex hull of a regular set is regular. If F and G are regular then $F \cup G$,

$$F+G = \{x+y : x \in F, y \in G\}$$

$$F.G = \{x.y : x \in F, y \in G\}$$

are regular.

Lemma 4.3

(a) $\{k_s : 1 \leq s \leq 1\}$ is regular.

(b) $||k_s - k_{s'}||_1 \to 0$ as $s' \to s$ $\quad -1 \leq s, s' \leq 1$

(c) $k^t \in R$, $-1 \leq t \leq 1$

Proof

(a) The sets $\{p_s\} = \{p(s) : -1 \leq s \leq 1\}$

and $\{q_s\} = \{q(s) : -1 \leq s \leq 1\}$

are regular because $p(s)$ and $q(s)$ are uniformly continuous on the interval $[-1, 1]$ and each function in $\{p_s\}$ or $\{q_s\}$ is constant.

The set $\{g_s(t)\}$ is regular because $g(s,t)$ is continuous.

The set $\{t-1\}$ is trivially regular.

The set $\{h_s(t)\}$ is regular because each $h_s(t)$ is constant with a single jump discontinuity at $t=s$.

Recall the definition of $h(s,t)$ in (4.7) and note that

$$\frac{\partial g}{\partial s}(s,t) = (t-1).h(s,t)$$

Hence $\{k_s\} = \{p_s\}\{t-1\}\{h_s\} + \{q_s\}\{g_s\}$ is regular

$$(b)\quad \int_{-1}^{1} |k(s,t) - k(s',t)|\, dt$$

$$\leq \int_{-1}^{\min(s,s')} |k(s,t)-k(s',t)|\,dt + \int_{\max(s,s')}^{1} |k(s,t)-k(s',t)|\, dt$$

$$+2\,|s-s'|\,.\sup_{-1\leq s,t\leq 1} |k(s,t)|$$

For t in the intervals $[-1, \min(s,s'))$ and $(\max(s,s'), 1]$, $k(s,t)$ is a continuous function of s for $s>\min(s,s')$ and $s<\max(s,s')$ respectively. Also $k(s,t)$ is bounded since $\{k_s\}$ is regular, so that finally

$$||k_s - k_{s'}||_1 \to 0 \quad \text{as} \quad |s-s'| \to 0$$

$$(c)\quad k^t(s) = \begin{cases} \tfrac{1}{2}(t-1)[-p(s)-q(s)(s+1)] & \text{for } s < t \\ \tfrac{1}{2}(t+1)[-p(s)-q(s)(s-1)] & \text{for } s \geq t \end{cases}$$

For each $t \in [-1,1]$, $k^t \in R$.

Now with Lemma 4.3, 4.2 and property 4.13 we have satisfied conditions (a)-(e) required for Theorem 2.13 of Anselone which we now state as Theorem 4.4. The proof follows that given in Anselone with very little modification.

Theorem 4.4

K, $KP_{np} \in [R]$, $KRcC$, K is compact

$\{KP_{np}\}$ is collectively compact (p fixed)

$KP_{np} \to K$ as $n \to \infty$

$||(I-P_{np})K|| \to 0$ as $\to \infty$

Proof

Since $\{k_s\}$ is regular, $k_s \in R_1$ (R with the seminorm $||x||_1$).

Define $f:[-1,1] \to R_1$ by $f(s)=k_s$. Then f is a continuous function on a compact set so that the convergence in Lemma 4.3(b) is uniform.

$$|(Kx)(s)| \leq ||k_s||_1 \cdot ||x|| = \{Kx: ||x|| \leq 1\} \text{ is bounded}$$

$$|(Kx)(s)-(Kx)(s')| \leq ||k_s - k_{s'}||_1 \cdot ||x|| \to \{Kx: ||x|| \leq 1\} \text{ is equicontinuous.}$$

Hence $K \in [R]$, $KR \subset C$, K is compact.

$P_{np} \in [R]$ therefore $KP_{np} \in [R]$ and each KP_{np} is compact.

Consider the linear functionals φ and φ_n, defined on R as follows

$$\varphi x = \int_1^1 x(t)\, dt \ , \varphi_n x = \int_1^1 \left(Pnp\ x\right)(t)\, dt.$$

Lemma 4.2 says that $\varphi n \to \varphi$.

Let u be (fixed) in R. Then $F = \{k_s u\}$ is regular. Choose x_m and x^m as in the definition of a regular set. Since $\varphi n \to \varphi$ uniformly on the totally bounded sets F_m and F^m and since $(x^m - x_m) \to 0$ uniformly for $x \in F$ we have $\varphi n \to \varphi$ uniformly for $x \in F$. Since

$$\left(Ku\right)(s) = \varphi\left(k_s\ u\right) \text{ and } \left(KP_{np}\ u\right)(s) = \varphi n\ \left(k_s\ u\right)$$

we have $||KP_{np}\ u - Ku|| \to 0$.

Note that $$\left|\left(KP_{np}\ x\right)(s)-\left(KP_{np}\ x\right)(s')\right| \leq \int_1^1 |k_s(t)-k_{s'}(t)| \cdot \left|\left(P_{np}\ x\right)(t)\right| dt$$
$$\leq ||k_s - k_{s'}||_1 \cdot ||P_{np}\ x||$$

By an argument similar to that proving K compact we have

$$\left(KP_{np}\ x\right)(s)-\left(KP_{np}\ x\right)(s') \to 0 \text{ as } n \to \infty \text{ and } s' \to s$$

uniformly for $||x|| \leq 1$ and $-1 \leq s \leq 1$

This implies that the sequence of sets $\{KP_{np}\ x : ||x|| \leq 1\}$ is asymptotically equicontinuous. These sets are bounded uniformly because $||KP_{np}|| \leq ||K|| \cdot ||P_{np}|| \leq M < \infty$. Also each of these sets

is totally bounded since each KP_{np} is compact. A construction similar to that used in Arzela's theorem then gives

$\{KP_{np}\, x : n \geq 1, \|x\| \leq 1\}$ is totally bounded, therefore relatively compact in R. Thus $\{KP_{np}\}$ is collectively compact. Note that $P_{np} \to I$ on C so that we have

$$\|(I-P_{np})K\| \to 0 \text{ as } n \to \infty \qquad p \text{ fixed.}$$

The conclusions of theorems 2.3 - 2.6 in Chapter 2 now apply, for example

Theorem 4.5

Suppose $(I-K)^{-1}$ exists, then there exists $N_1 > 1$ such that

$\forall n \geq N_1 \qquad \|(I-P_{np}K)^{-1}\| \leq M_1$

Use theorem 2.6A(a)

Also there exists $N_2 \geq 1$ such that $\forall n \geq N$

$$\|(I-KP_{np})^{-1}\| \leq M_2$$

Use theorem 2.6A(c)

Theorem 4.6

Suppose there exists n such that either

$$\|(I-P_{np}K)^{-1}(K-P_{np}K)\| < 1$$

or $\|(I-KP_{np})^{-1}(K-KP_{np})K^d\| < 1$ for some $d \geq 1$

Then $(I-K)^{-1}$ exists. Use theorem 2.4 or 2.5.

§4.5 The Behaviour of $||W_n||$

We remain in the space $R = X_1 = X_2$. §3.5 - §3.7 in Chapter 3 make use of the operators H_n, J_n, J_n^* sometimes called restriction (H_n) and extension (for J_n and J_n^*) operators, they must possess the following properties in order to apply the theorems in those sections.

$$H_{np} \in [R, \mathbb{R}^{np}],\ J_{np} \in [\mathbb{R}^{np}, P_{np} R],\ J_{np}^* \in [\mathbb{R}^{np}, R]$$

(a) $H_{np} P_{np} = H_{np}$ (e) $||H_{np}|| = 1$

(b) $H_{np} J_{np}^* = I_{np}$ (f) $||J_{np}^*|| = 1$

(c) $P_{np} J_{np}^* H_{np} = P_{np}$ (g) $J_{np}^* H_{np} \to I$ on R

(d) $J_{np} H_{np} = P_{np}$ ($\Leftrightarrow J_{np} = P_{np} J_{np}^*$) (4.17)

These conditions imply $||J_{np}|| = ||P_{np}||$

Define H_{np}, J_{np}, J_{np}^* as follows

$$(H_{np} x)_{ji} = x(\xi_{ji}) \qquad j=1,\dots,p\ ;\ i=1,\dots,n \tag{4.18}$$

$$(J_{np} u_{ji})(t) = \sum_{i=1}^{n} \chi_{[t_{i-1},t_i)}(t) \sum_{j=1}^{p} u_{ji} L_{ji}(t) \tag{4.19}$$

$$(J_{np}^* u_{ji})(t) = \sum_{i=1}^{n} \sum_{j=1}^{p} u_{ji} \chi_{[\tau_{i,j-1},\tau_{i,j})}(t) \tag{4.20}$$

where $\tau_{ij} = t_{i-1} + l_j (t_i - t_{i-1})/2$ (NB $J_{np}^* H_{np} \equiv S_{np}$ in Lemma 4.2)

We use the sup norm on R and the maximum norm on $\mathbb{R}^{np}$

Conditions (a) - (f) above are satisfied trivially by these definitions. Condition (g) does not hold on R, but only on C, and we must make a slight addition to the proof of Theorem 3.7.

However we now have, directly from (3.26), and (3.27) since

$$||(I-P_{np}K)|_{P_{np}K}|| \le ||(I-P_{np}K)||$$

$$\left.\begin{aligned} &||W_n||_\infty \le ||P_{np}|| \cdot ||(I-P_{np}K)^{-1}|| \le ||P_{1p}|| \cdot M_1 \text{ for } n \ge N_1 \\ &\text{and } ||W_n||_\infty \le ||(I-KP_{np})^{-1}|| \le M_2 \text{ for } n \ge N_2 \end{aligned}\right\} \tag{4.21}$$

For the stronger result of theorem 3.7 we need two further conditions, namely that $||P_{np}K(J_n - J_n^*)|| \to 0$ as $n \to \infty$ and that $(I-K)^{-1}$ has the same norm in the spaces $[C]$ and $[R]$.

Since P_{np} is uniformly bounded, or from the fact that $||K-P_{np}K|| \to 0$, it is sufficient to show that

$$||K(J_{np} - J_{np}^*)|| \to 0.$$

Recall the integral operator H defined by 4.7 and consider

$$||H(J_{np} - J_{np}^*)||$$

$$= \max_{|a_{ji}| \leq 1} \sup_{s \in [-1,1]} \left| \sum_{i=1}^{n} \sum_{j=1}^{p} a_{ji} \int_{-1}^{1} \left(\chi_{[t_{i-1}, t_i)}(t) \, L_{ji}(t) - \chi_{[\tau_{i,j-1}, \tau_{ij})}(t) \right) dt \right| \tag{4.21}$$

Now since $\int_{t_{i-1}, t_i)} L_{ji}(t)\, dt = \tau_{ij} - \tau_{i,j-1}$ $\quad i=1,\ldots,n$ $\quad j=1,\ldots,p$

we can replace the lower bound of the integral in (4.21) by the largest element of T_n less than s call it t_k, we have $|s-t_k| < ||T_n||$ and we can simplify the bound in (4.21) to

$$||H(J_n - J_n^*)|| \leq \max_{|a_{ji}| \leq 1} \sup_{s \in [-1,1]} \int_{t_k}^{s} \sum_{j=1}^{p} \left| L_{ji}(t) - \chi_{[\tau_{i,j-1}, \tau_{ij})} \right| dt$$

$$\leq ||T_n|| \cdot \left(||P_{np}|| + 1 \right) \tag{4.22}$$

Since $||Tn|| \to 0$ we have

$$||H(J_{np} - J_{np}^*)|| \to 0 \text{ as } n \to \infty \tag{4.23}$$

In view of (4.23) and (4.8) we have

$$||K(J_{np} - J_{np}^*)|| \to 0 \text{ as } n \to \infty \tag{4.24}$$

Lemma 4.7

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

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

Proof

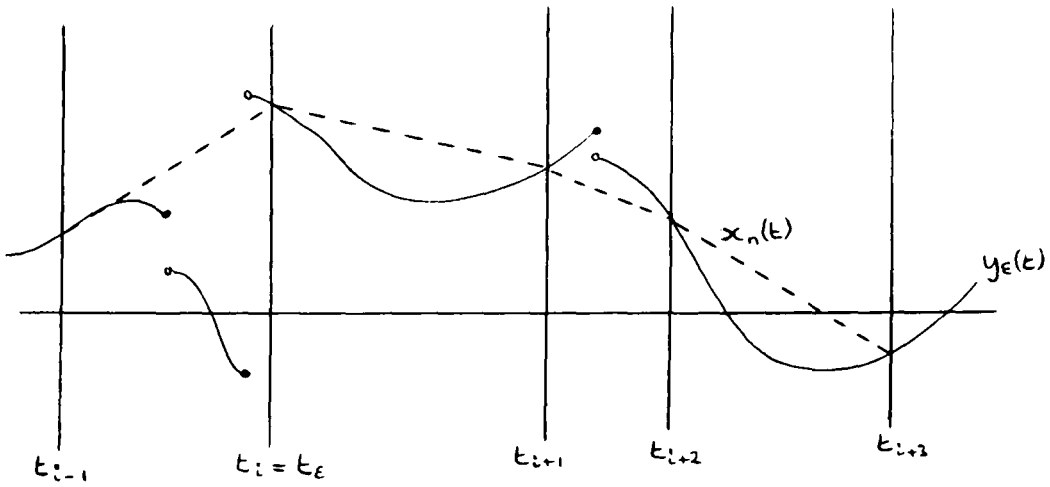
Let $y_\varepsilon \in R$ be such that $||y_\varepsilon|| \leq |$

$$||(I-K)^{-1}|| - ||(I-K)^{-1} y_\varepsilon|| < \varepsilon \qquad \text{(A)}$$

Let $t_\varepsilon \in [-1,1]$ such that

$$||(I-K)^{-1} y_\varepsilon|| - ||((I-K)^{-1} y_\varepsilon)(t_\varepsilon)|| < \varepsilon \qquad \text{(B)}$$

Let x_n be a sequence of continuous piecewise linear function defined by $x_n(t) = (P_{n2}\, y_\varepsilon)\ (t)$ with $\xi_1 = -1$, $\xi_2 = +1$, $t_\varepsilon = t_i$ for some i=0...n. ($||P_{n2}||=1$)



By the results leading up to theorem 4.4 we have

$$||Kx_n - Ky_\varepsilon|| \to 0 \text{ as } n \to \infty$$

Since each $x_n \in C$ (forced by $\xi_1 = -1$, $\xi_2 = +1$) we can find an x_N

say such that

$$||Kx_N - Ky_\varepsilon|| \leq \varepsilon \Big/ ||(I-K)^{-1}||$$

The operator $(I-K)^{-1}$ can be expressed as

$$(I-K)^{-1} = I + (I-K)^{-1}K$$

Therefore $(I-K)^{-1}(x_N - y_\varepsilon)(t_\varepsilon) = (x_N(t_\varepsilon) - y_\varepsilon(t_\varepsilon)) + (I-K)^{-1}(Kx_N - Ky_\varepsilon)(t_\varepsilon)$

But by construction we have $x_N(t_\varepsilon) = y_\varepsilon(t_\varepsilon)$

Hence $||((I-K)^{-1}y_\varepsilon)(t_\varepsilon)|| - ||((I-K)^{-1}x_N)(t_\varepsilon)|| \leq \varepsilon$ (C)

Adding A, B, C

$$||(I-K)^{-1}|| - ||((I-K)^{-1}x_N)(t_\varepsilon)|| < 3\varepsilon$$

But $||((I-K)^{-1}x_N)(t_\varepsilon)|| \leq \sup_{y \in C} ||(I-K)^{-1}y||$ since $||x_N|| \leq 1$

As ε is arbitrary we have

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

$C \subset R$ so that, finally

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

Theorem 4.8

$$||W_n||_\infty \to ||(I-K)^{-1}||$$

Proof

Verify conditions of theorem 3.7 hold, apart from condition (iii). Replace y in proof of the theorem by an $x_\varepsilon \in C$

Lemma 4.7 guarantees that we can do this, and condition (iii) is satisfied for $x \in C$.

Wright [51] has demonstrated the conclusion of theorem 4.8 for $(I-K)^{-1} \in C$ and approximations consisting of global polynomial collocation at the zeros of polynomials orthogonal with respect to a weight function of the form

$$(1-t)^{\alpha}\,(1-t)^{\beta} \; ; \; -\tfrac{1}{2} \leq \alpha,\ \beta \leq \tfrac{1}{2}$$

The result of theorem 4.8 is strong evidence that with the infinity norm the matrix W_n is a good choice for expressing the inverse approximate operator, for, in the limit at least, it is not dependent on the form of the approximation. We can express the properties that an approximation must have to satisfy the conditions required to prove theorem 4.8 concisely as:-

for any $x \in R$, $S_{np}\,x$ is a Riemann sum.

Theorem 4.8 together with numerical corroboration, presented later, provides the justification for the form of bounds on the inverse approximate operators in the next section, which are there expressed in terms of computable quantities.

§4.6 Bounds on the Inverse Approximate Operator

In order to develop practical error bounds for approximate solutions arising from equations such as (3.1) or (3.3) it is necessary to derive bounds on the inverse approximate operators $(I-P_{np}\,K)^{-1}$ or $(I-K\,P_{np})^{-1}$. It will not be sufficient to know that such bounds will exist, we shall want to be able to compute a numerical value for the bound for any given n. The theorems in Chapter 2 will be of no use since we have no bound on $(I-K)^{-1}$, instead we will use the results in §3.6 to relate these inverse approximate operators to the matrix W_n.

Recall (3.24)

$$||(I-P_{np}\,K)^{-1}|| \leq ||P_{np}||\cdot||W_n||$$

The inverse approximate operator here has range space $P_{np}\,R$ and to bound $(I-P_{np}\,K)$ on the whole of R we can use the identity (3.25).

$$||(I-P_{np}\,K)^{-1}|| \leq 1 + ||(I-P_{np}\,K)|_{P_{np}\,R}||\cdot||P_{np}\,K||$$

Giving

$$||(I-P_{np}\,K)^{-1}|| \leq 1 + ||P_{np}||\cdot||W||\cdot||P_{np}\,K|| \qquad (4.25)$$

Since $||P_{np}||$ is independent of n, $||W_n|| \to ||(I-K)^{-1}||$ and $||P_{np}\,K|| \leq ||(P_{np}-I)K||+||K|| \to ||K||$, (4.25) is a fairly satisfactory bound.

Recall that $(I-K\,P_{np})$ can be expressed in terms of $(I-P_{np}\,K)^{-1}|_{P_{np}\,R}$ from (3.5).

$$(I-K\,P_{np})^{-1} = I + K(I-P_{np}\,K)^{-1}\,P_{np}$$

Giving

$$||(I-K\,P_{np})^{-1}|| \leq 1+||K||\cdot||P_{np}||\cdot||W_n|| \qquad (4.26)$$

Cruickshank, deriving bounds similar to (4.25) and (4.26) notes that for the global polynomial case $||P_{lp}||$ increases with p when Chebychev or Legendre collocation points are used $\left(||P_{np}||\right.$ remains constant with $\left.n\right)$. Bounds which do not increase with $||P_{lp}||$ are derived by considering $||K(I-P_{lp}\,K)^{-1}P_{lp}||$ in more detail and making use of the fact that the collocation points are zeros of orthogonal polynomials. The result achieved for Chebychev polynomials is

$$\left.\begin{aligned} ||K(I-P_{lp}\,K)^{-1}P_{lp}|| &\le \sup_{-1\le s,t\le 1}|k(s,t)|\frac{\pi}{\sqrt{2}}\,||W_n||_\infty \\ \text{For Legendre polynomials} & \\ ||K(I-P_{lp}\,K)^{-1}P_{lp}|| &\le \sup_{-1\le s,t\le 1}|k(s,t)|\cdot 2||W_n||_\infty \end{aligned}\right\} \quad (4.27)$$

To express a bound for $(I-P_{lp}\,K)^{-1}$ in terms of $K(I-P_{lp}\,K)^{-1}P_{lp}$ we use the identify

$$(I-P_{lp}\,K)^{-1} = I + P_{lp}\,K\,(I-P_{lp}\,K)^{-1}$$

$$= I + (P_{lp}-I)K(I-P_{lp}\,K)^{-1} + K(I-P_{lp}\,K)^{-1}$$

Then, provided $||(P_{lp}-I)K|| = \delta\rho < 1$ note that

$$||(I-P_{lp}\,K)^{-1}|_{P_{lp}\,C}|| \le \frac{1 + ||K(I-P_{lp}K)^{-1}P_{lp}||}{1-\delta\rho} \qquad (4.28)$$

Then use (3.25)

All these expressions (4.25) - (4.28) depend on the integral operator K as well as $||W_n||_\infty$, which is not very satisfactory when K is large. It is easy to show that

$$||(I-P_{np}\,K)^{-1}-(I-K)^{-1}|| \to 0 \text{ as } n \to \infty$$

In view of the fact that $||W_n||_\infty \to ||(I-K)^{-1}||$ we must have asymptotically $||(I-P_{np}\,K)^{-1}|| \sim ||W_n||_\infty$ but for the present we must rely on (4.25) and (4.26). A small improvement in the bound (4.26) might be achieved for the piecewise polynomial case by similar analysis to that for (4.27) but for small values of p it is probably not worthwhile.

§4.7 Bounds on K^d, $(I-P_{np})K^d$

The bounds in the previous section, together with bounds on K^d, $(I-P_{np})\,K^d$ will enable us to use theorems 2.4 and 2.5 of Chapter 2 to obtain strict bounds on $||(I-K)^{-1}||$.

Recall the definition of K as an integral operator

$$\left(K\,x\right)(s) = \int_{-1}^{1} k(s,t)x(t)\,dt$$

$$||\,K\,x\,|| = \sup_{s\in[-1,1]} \left| \int_{-1}^{1} k(s,t)x(t)\,dt \right|$$

$$\leq \sup_{s\in[-1,1]} \int_{-1}^{1} \left|k(s,t)\right| dt \;.\; ||x|| \qquad (4.29)$$

we can use (4.29) to bound $||K||$. From 4.4 we have

$$||K|| \leq \sup_{s\in[-1,1]} \left(|p(s)| \int_{-1}^{1} \left| \frac{\partial g}{\partial s}(s,t) \right| dt + |q(s)| \int_{-1}^{1} \left|g(s,t)\right| dt \right) \qquad (4.30)$$

It is easy to show that

$$\int_{-1}^{1} \left|g(s,t)\right| dt = \tfrac{1}{2}(1-s^2) \;;\; \int_{-1}^{1} \left| \frac{\partial g}{\partial s}(s,t) \right| dt = \tfrac{1}{2}(1+s^2) \qquad (4.31)$$

for g(s,t) defined by (4.3).

To bound K^d, note that

$$||K^d|| \leq ||K||^d \qquad (4.32)$$

P_{np} is an interpolation projection so that $(I-P_{np})x$ is the 'error' in interpolating a given function x. Most of the error bounds for interpolation are given in terms of higher derivatives of x. For that reason we now consider the operator $D^r K^d$ $1 \leq r \leq d$

defined by

$$\left(D^r K^d x\right)(s) = \frac{d^r}{ds^r} K^d x\ (s) \tag{4.33}$$

We will develop bounds k_d^r on $\left|\left|D^r K^d\right|\right|$

Define $p_j(s) = \left|p^{(j)}(s)\right|$, $q_j(s) = \left|q^{(j)}(s)\right|$ when $p,q \in C^{(j)}$

Define $k_1^o(s) = p_o(s)\tfrac{1}{2}(1+s) + q_o(s)\tfrac{1}{2}(1-s^2)$ (4.34)

Then $||K|| \leq \sup_{s\in[-1,1]} k_1^o(s)$

Consider $-\left(DKx''\right)(s)$, $x'' \in R$; $p,q \in C^{(1)}$

$$= \frac{d}{ds}\int_{-1}^{1}\left(p(s)\ \frac{\partial g}{\partial s}\ (s,t) + q(s)g(s,t)\right)x''(t)\ dt$$

$$= \frac{d}{ds}\left[p(s)\ x'(s) + q(s)\ x\ (s)\right]$$

$$= p(s)\ x''\ (s) + \left[p'(s) + q(s)\right]\ x'(s) + q'(s)\ x\ (s)$$

$$= p(s)\ x''\ (s) + \int_{-1}^{1}\left[(p'(s)+q(s)\frac{\partial g}{\partial s}(s,t)+q'(s)g(s,t)\right]x''\ (t)\ dt$$

(The fact that $x(-1) = x(1) = 0$ in no way restricts x'')

Define

$$k_1^1(s) = p_1(s) + (p_1(s)+q_o(s))\ \frac{(1+s^2)}{2} + q_1(s)\ \frac{(1-s^2)}{2} \tag{4.35}$$

Then $||DK|| \leq \sup_{s\in[-1,1]} k_1^1(s)$

Consider $-(DK^2x)$, make the substitution $Kx = y$

then

$$(DK^2x)\ (s) = (DK\ y)\ (s)$$

Define $k_2^1(s) = k_1^1(s) \cdot ||k_1^0||$

Then $||DK^2|| \leq \sup_{s \in [-1,1]} k_1^1(s) \cdot ||k_1^0||$ (4.36)

Similarly define $k_n^0(s) = k_1^0(s) \cdot ||k_1^0||^{n-1}$

Then $||K^n|| \leq \sup_{s \in [-1,1]} k_1^0(s) \cdot ||k_1^0||^{n-1}$ (4.37)

We lose something in the bounds (4.36) and (4.37) by failing to take into account the form of $(Kx)(s)$, but the extra algebra involved for higher order cases (e.g. D^3K^5) might outweigh the reductions in bounds.

Consider $(D^rK^rx)(s)$, $r \geq 2$

Make the substitution $K^{r-1}x = y''$, $y \in C^{(r+1)}$ satisfies BC.

$$(D^rK^rx)(s) = (D^rKy'')(s)$$

$$= D^r(p(s)y''(s) + q(s)y(s))$$

This can be expanded in a binomial series involving derivatives of p and q and y. Note that

$$D^{r+1}y = D^{r-1}y'' = D^{r-1}K^{r-1}x$$

$$D^ry = D^{r-2}K^{r-1}x = D^{r-2}K^{r-2}(Kx)$$

$$\vdots$$

$$y'' = K(K^{r-2}x)$$

$$y'(s) = \int_{-1}^{1} \frac{\partial g}{\partial s}(s,t) \left(K^{r-1}x\right)(t)\, dt$$

$$y(s) = \int_{-1}^{1} g(s,t) \left(K^{r-1}x\right)(t)\, dt$$

So that from the expressions $k(s)$ up to $k_{r-1}^{r-1}(s)$ and $p_o(s)$, $q_o(s)$ up to $p_r(s)$ $q_o(s)$ we can derive an expression for $k_r^r(s)$ with

$$||D^r K^r|| \leq \sup_{s\in[-1,1]} k_r^r(s)$$

Trivially $$||D^r K^d|| \leq \sup_{s\in[-1,1]} k_r^r(s)\cdot||k_1^o||^{d-r} \qquad 1\leq r\leq d$$

This recursive definition of $k_r^r(s)$ is readily implemented as a recursive procedure call in a computer program. The s-dependence of the terms of $k_r^r(s)$ could be dropped, giving somewhat poorer bounds.

As examples consider $k_2^2(s)$, $k_3^3(s)$

$$k_2^2(s) = p_o(s)\,y_3(s) + 2p_1(s)y_2(s) + p_2(s)y_1(s)$$
$$+ q_o(s)y_2(s) + 2q_1(s)y_1(s) + q_2(s)y(s)$$

where $y_3(s) = k_1^1(s)$

$y_2(s) = k_1^o(s)$ (4.38)

$y_1(s) = \tfrac{1}{2}(1+s^2)\cdot||k_1^o||$

$y_o(s) = \tfrac{1}{2}(1-s^2)\cdot||k_1^o||$

$$k_3^3(s) = p_o(s)y_4(s) + 3p_1(s)y_3(s) + 3p_2(s)y_2(s) + p_3(s)y_1(s)$$
$$+ q_o(s)y_3(s) + 3q_1(s)y_2(s) + 3q_2(s)y_1(s) + q_3(s)y_o(s)$$

where $y_4(s) = k_2^2(s)$

$y_3(s) = k_1^1(s)\cdot||k_1^o||$

$y_2(s) = k_1^o(s)\cdot||k_1^o||$ (4.39)

$y_1(s) = \tfrac{1}{2}(1+s)\cdot||k_1^o||^2$

$y_o(s) = \tfrac{1}{2}(1-s)\cdot||k_1^o||^2$

Error bounds for the interpolant $P_{np}K^d x$ can be found from Jacksons Theorem, which states that if $u \in C^{(d)}$ there exists a polynomial $\tilde{u}$ of degree n-1 with

$$||u-\tilde{u}|| \leq \left(\frac{\pi}{2}\right)^d \frac{||u^{(d)}||}{n(n-1)\cdots(n-d+1)}, \quad n\geq d\geq 1$$

In the notation of this chapter this gives, provided $p,q \in C^{(r)}$

$$||(I-P_{np})\ K^d|| \leq \left(\frac{\pi \cdot ||T_n||}{4}\right) \frac{||D^r K^d||(1+||P_{np}||)}{p(p-1)\ldots(p-r+1)}, \quad \begin{matrix} p \geq r \geq 1 \\ d \geq r \end{matrix} \qquad (4.40)$$

where we can take

$$||D^r K^d|| \leq \sup_{s \in [-1,1]} k_d^r(s)$$

The Peano kernel expression of the remainder for polynomial interpolation has smaller constants of proportionality than does Jacksons theorem, particularly for larger p and r as can be seen in Tables 38 and 39 . Unfortunately the computed Peano constants are not strict theoretical bounds as the processes of numerical integration and maximisation are not exact. Strict bounds on the errors in computing the Peano constants could, in principle, be found but this would involve considerably more effort.

§4.8 Bounds on $(I-K)^{-1}$

We now have all the terms required for bounding $(I-K)^{-1}$ from theorems 2.4 or 2.5 applied to approximations of the form (3.1) or (3.3) in a computable form. Call approximations of the form (3.1) "projection" and (3.3) "extended". We can now compare the usefulness of the "projection" and "extended" approximations.

"projection" approximation

Theorem 4.9

Suppose that for some n, p, $d \geq 1$, W_n exists and

$$\Delta = \{1+||P_{np}|| \cdot ||W_n|| \cdot (||K|| + ||(I-P_{np})K||)\} \cdot ||(I-P_{np})K^d|| < 1$$

Then

$(I-K)^{-1}$ exists and

$$||(I-K)^{-1}|| < \frac{\sum_{i=o}^{d-1}||K||^i+\{1+||P_{np}||\cdot||W_n||\cdot(||K||+||(I-P_{np})K||)\}||K||^{d-1}}{1-\Delta}$$

where $||K||^o = 1$ and the summation is empty if $d < 2$

"extended" approximation

Theorem 4.10

Suppose that for some n, p, $d \geq 1$, W_n exists and

$$\Delta=\{1+||K||\cdot||P_{np}||\cdot||W_n||\}\cdot||K||\cdot||(I-P_{np})K^d|| < 1$$

Then

$(I-K)^{-1}$ exists and

$$||(I-K)^{-1}|| \leq \frac{\sum_{i=o}^{d-1}||K||^i+\{1+||K||\cdot||P_{np}||\cdot||W_n||\}||K||^d}{1-\Delta}$$

Proof of the above theorems, substitute (4.25), (4.26) into theorem (2.4) and (2.5). $||K||^i$ and $||(I-P_{np})K^d||$ are computed from (4.36) and (4.40). The expression $(||K|| + ||(I-P_{np})K||)$ is used instead of $||P_{np}||\cdot||K||$ for $||P_{np}\ K||$ in theorem 4.9 because in most practical applications where $\Delta < 1$ it is generally true that $||(I-P_{np})K|| < (||P_{np}||-1)\cdot||K||$. (A notable exception is when $||Pnp|| = 1$, e.g. a continuous piecewise linear interpolant).

If $||K|| \leq \frac{1}{2}$ it will generally be better to use theorem 2.2, and certainly much easier. The above theorems will be most useful when $K >> \frac{1}{2}$. The "projection" approximation bound will generally be applicable with less work (smaller n) and give smaller bounds on $(I-K)^{-1}$ than the "extended" approximation. The situation for global polynomial collocation (P_{lp}) is somewhat more complicated by poorer bounds on $||(I-P_{lp}K)^{-1}||$.

Since $||W_n|| \to ||(I-K)^{-1}||$ these bounds leave a lot to be

desired, but alternative bounds for the errors in the approximate solutions mitigate this to some extent.

§4.9 Error Bounds

Define the residual of any approximate solution x_n of

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

by

$$r_n = y - (I-K)\, x_n$$

If $(I-K)^{-1}$ exists then

$$r_n = (I-K)(x-x_n) = (I-K)\, e_n$$

and we have the following error bound on x_n.

$$||x - x_n|| = ||e_n|| \leq ||(I-K)^{-1}|| \cdot ||r_n|| \qquad (4.41)$$

We can calculate r_n from x_n, y, p and q and bounds on $||(I-K)^{-1}||$ are available from theorems 4.9 and 4.10.

Unfortunately we are not yet in a position to say that $||r_n||$ tends to zero for the approximations described in this chapter.

Note the identity

$$T^{-1} - S^{-1} = T^{-1}\,(S-T)\,S^{-1}$$

Applied to an approximation such as $(I-KP_{np})z_n = y$ (4.42)

gives

$$\left((I-KP_{np})^{-1} - (I-K)^{-1}\right)y = \left((I-KP_{np})^{-1}(K-KP_{np})(I-K)^{-1}\right)y$$

$$z_n - x = (I-KP_{np})^{-1}\,(K-KP_{np})x$$

since x is fixed $||Kx-KP_{np}x|| \to 0$ (from Theorem 4.4).

Also $||(I-KP_{np})^{-1}||$ is bounded uniformly for n sufficiently large (Theorem 4.5), therefore

$$||z_n - x|| \to 0 \quad \text{as } n \to \infty$$

But $z_n = y + Kx_n$ where x_n is given by

$$(I-P_{np}K)x_n = P_{np}y \tag{4.43}$$

giving $$||Kx_n - Kx|| \to 0 \quad \text{as } n \to \infty$$

Note that the residual for x_n can be expressed as follows

$$\begin{aligned} r_n &= y - (I-K)x_n \\ &= y - (I-P_{np}K)x_n + (I-P_{np})Kx_n \\ &= (I-P_{np})(y + Kx_n) \end{aligned} \tag{4.44}$$

So

$$||r_n|| \leq ||(I-P_{np})y|| + (1+||P_{np}||)||Kx-Kx_n|| + ||(I-P_{np})K||\cdot||x||$$

and provided y is sufficiently "smooth", $||(I-P_{np})y||$ tends to zero, giving $$||r_n|| \to 0 \text{ and } ||x_n-x|| \to 0 \quad \text{as } n \to \infty$$

Note that we have convergence of the "extended" solution z_n without this restriction on y.

It can be seen from (4.44) that r_n will behave like the remainder for interpolation of the function $y+Kx_n$ and useful properties of r_n can be obtained from this fact.

It was discovered during the course of numerical experiments that the error term e_n for piecewise polynomial solutions was approximately equal to the residual. This "property" was remarkably consistent over a wide range of problems and held to a high degree of accuracy for large n. In view of the error bound (4.41) which depends significantly on a bound for $||(I-K)^{-1}||$ an attempt at a theoretical justification of this "property" was prompted. We start with an identity which expresses $(I-K)^{-1}$ in terms of the resolvent operator $(I-K)^{-1}K$.

$$(I-K)^{-1} = I+(I-K)^{-1}K$$

Hence
$$e_n = (I+(I-K)^{-1}K)r_n$$
$$= r_n + (I-K)^{-1}Kr_n \qquad (4.45)$$

If we can show that $||Kr_n||$ tends to zero faster than $||r_n||$ we have $e_n \doteq r_n$ for n sufficiently large. We now examine the behaviour of r_n and Kr_n in terms of the derivatives of y and Kx_n.

Suppose that $f^{(p+r)}$ is uniformly continuous on the intervals (t_{i-1}, t_i) $i = 1,\ldots n$. Then by Jacksons theorem there exists polynomials $u_i(t)$ of degree p+r-1 with

$$||f(t) - u_i(t)|| \leq \left(\frac{\pi(t_i-t_{i-1})}{4}\right)^{p+r} \frac{t_{i-1}^{\sup<s<t_i} \left|f^{(p+r)}(s)\right|}{(p+r)!} = \varepsilon_i \qquad (4.46)$$

for $t \in [t_{i-1}, t_i)$ $i=1,\ldots,n$.

Denote $(I-P_{np})u_i(t) = v_i(t)$ for $t \in [t_{i-1},t_i)$.

Then

$$\left.\begin{aligned} &||(I-P_{np})\ f(t) - v_i(t)|| \leq \varepsilon_i\ (1+||P_{np}||) \\ \text{and}\quad &||v_i(t)|| \leq ||(I-P_{np})f(t)|| + \varepsilon_i(1+||P_{np}||) \end{aligned}\right\} \qquad (4.47)$$

Now, for f substitute $y+Kx_n$ and recall (4.44).

Equation (4.47) tells us that provided y and Kx_n are sufficiently differentiable, r_n can be expressed as a piecewise polynomial "principal part" plus a remainder. For example if $||f^{(2)}(t)|| \leq 1$, $||f^{(4)}(t)|| \leq 30$.p=2, n=50 and each $(t_i-t_{i-1})=0.04$ then (4.47) tells us that the interpolation error $(I-P_{np})f(t)$ is within approximately 0.25% of a piecewise polynomial $v_i(t)$ of magnitude $\frac{(1+||P_{np}||)}{2000}$.

This effect becomes more pronounced as r and n increase. In a computer program one could use Jacksons theorem and evaluate f(t) at various points in order to obtain a tighter bound on $v_i(t)$.

The piecewise polynomial $v_i(t)$ is zero at each point ξ_{ji} $j=1,\ldots,p$; $i=1,\ldots,n$ and is of degree $p+r-1$ in each interval $[t_{i-1}, t_i)$, also an overall bound is given by (4.47). We hope to use this information to obtain better bounds on $Kv(t)$ than we can achieve on $K(I-P_{np})f(t)$. A particularly interesting example is given by using the Gauss-Legendre zeros for ξ_j. For such points it is known that the integral of any polynomial of degree up to $2p-1$ is given by a weighted sum of the values at the points ξ_j, therefore

$$\left(Hv\right)(t_i) = \int_1^t v(t)\,dt = \sum_{j=1}^{i} \int_{[t_{j-1},t_j)} v_j(t)dt = 0 \qquad i=o,\ldots,n$$

since $v_j(t) = 0$ at every ξ_{ji}. Further, since we have a bound on $||v_i||$ we can say that

$$||H\,v|| \leq \frac{||T_n||}{2}\left[||(I-P_{np})f(t)|| + \varepsilon\,(1+||P_{np}||)\right] \qquad (4.48)$$

where $\varepsilon = \max\,(\varepsilon_1,\ldots,\varepsilon_n)$

we then have the following bound on $K(I-P_{np})f(t)$

$$||K(I-P_{np})f|| \leq (p+2q_o)\frac{||T_n||}{2}\left[||(I-P_{np})f||+\varepsilon(1+||P_{np}||)\right] + ||K||\cdot\varepsilon\cdot(1+||P_{np}||) \qquad (4.49)$$

Since the polynomials forming $v_i(t)$ can have degree at most $2p-1$ it follows that we must take $r \leq p$. Equation (4.49) shows that if we use Legendre zeros ($p>1$) $||K(I-P_{np})f||$ tends to zero faster than does $||(I-P_{np})f||$, provided f is sufficiently differentiable (piecewise).

In order to apply the above argument to Kr_n we need bounds on high order derivatives of y and Kx_n. Bounds on high order derivatives of Kx_n exist in the open intervals (t_{i-1},t_i) and since

x_n is known we can compute them by a similar process to that described for bounding $||D^r K^d||$ in §4.7.

The property (Hv) $(t_i) = 0$ for various degrees of piecewise polynomial holds for many sets of points other than Legendre zeros - for example any symmetrical arrangement of the ξ_j about 0 with p odd will suffice for v of degree p. What is particularly interesting about the Legendre polynomials $P_i(t)$ is that all the integrals from the first up to the i th over $[-1,1]$ are zero. This is easily seen from the orthogonality relationship.

$$\int_{-1}^{1} P_i(t)\ P_j(t)\ dt = 0 \qquad i \neq j$$

Then since $\left(H^{j-1}1\right)(t)$ is a polynomial of degree j-1,

$$0 = \int_{1}^{1} P_i(t)\ \left(H^{j-1}1\right)(t)\ dt \qquad j=1,\ldots,i$$

$$= \left(HP_i\right)(t)\left(H^{j-1}1\right)(t)\Big|_{-1}^{1} - \int_{-1}^{1}\left(HP_i\right)(t)\left(H^{j-2}1\right)(t)\ dt$$

$$\vdots$$

$$\therefore \qquad \int_{-1}^{1}\left(H^{j-1}P_i\right)(t) \quad = \quad \left(H^{j}P_i\right)(1) \quad = \quad 0,\ j=1,\ldots,i \qquad (4.50)$$

Using (4.49) we have $e_n \triangleq r_n$. Then apply (4.49) again to (4.45) taking p=1, q=0 and p=0, q=1 to give improved bounds for He_n and He_n^2; these are related to, respectively, the errors in first derivative and actual approximate solution of the original differential equation (4.1). From the above observations we can produce error bounds much more in keeping with actual measured errors than the bounds given by (4.41). Indeed we can obtain a higher order of

accuracy for first derivative and actual error bounds from this analysis (see also De Boor & Swartz for a proof based on "order" arguments [4].)

Perhaps the most important conclusion of this section is that provided p, q, and y are sufficiently differentiable we may take the norm of the residual r_n as an error estimate for the approximate solution x_n. This estimate is likely to be in better keeping with the actual error than the bound (4.41).

Unfortunately it is not possible to determine how large n must be in order that this estimate is valid without entering again into a discussion on strict error bounds. Further investigation might reveal other criteria upon which some measure of confidence in this estimate could be based.

See the results in Chapter 5 for a comparison of r_n and e_n for some sample problems.

§4.10 Use of weighted norm

We keep the definitions of the previous sections in this chapter but extend the examination of the results to use a weighted norm as introduced in §2.2. By the comments there most of the theory extends trivially but particular care must be exercised in extending any convergence proofs and where norms are calculated. We will make the following assumptions about the weighted norm and the approximations used in conjunction with it.

The norm in X_2 will be the usual sup norm

The norm in X_1 will be defined by

$$\left.\begin{aligned} ||x||_{x_1} &= \max_{i=1,\ldots m} \; w_i \cdot \sup_{t\varepsilon[s_{i-1},s_i)} |x(t)| \\ \text{with } 0 \leq \varepsilon \leq w_i &\leq 1 \qquad i=1,\ldots,m \\ -1 = s_o < s_1 &< \ldots < s_m = 1 \end{aligned}\right\} \qquad (4.51)$$

The points s_i and weights w_i will be considered as fixed in any convergence results (increasing n). Further any approximation using this weighted norm will be such that $\{s_i\} \subseteq \{t_i\}$. That is the set of partition points of the piecewise polynomial approximations contains the set of points defining the weighted norm (hence $m \leq n$)

We may note that 2.10 becomes

$$\frac{1}{\max w_i} ||x||_{x_1} \leq ||x||_{x_2} \leq \frac{1}{\min w_i} ||x||_{x_1} \qquad (4.52)$$

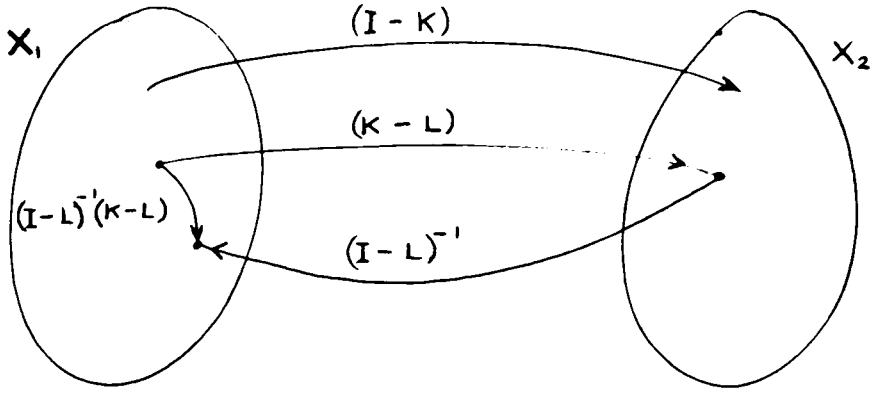
We have seen in §3.6 how the norms of $(I-P_nK_n)^{-1}$ and are related when using weighted norms but first we return to the basic equations in Chapter 2 in order to see how advantage may be taken of this weighted norm.

Recall Δo and Δd used in Theorems 2.4 and 2.5

$$\Delta o = ||(I-L)^{-1}(K-L)||$$

$$\Delta d = ||(I-L)^{-1}(K-L)K^d||$$

Our ability to make use of Theorems 2.4 and 2.5 for producing error bounds depends critically on finding an approximate operator L sufficiently "close" to K to make Δo or Δd less than 1. Consider the operator $(I-K) \; \varepsilon \; [X_1, X_2]$ so that $(I-L)^{-1} \; \varepsilon \; [X_2, X_1]$ and $(K-L) \; \varepsilon \; [X_1, X_2]$.



Instead of computing the values of Δo and Δd in the usual manner of

$$\Delta o \leq ||(I-L)^{-1}|| \cdot ||K-L||$$

with $(I-L)^{-1}$ and $(K-L)$ considered both in X (X_1) we are at liberty to use instead the norm of $(I-L)^{-1}$ considered as an operator in $[X_2, X_1]$ and the norm of $(K-L)$ considered as an operator in $[X_1, X_2]$. (Note we may consider $K^d \in X_1$ for the purposes of calculating $||(K-L)K^d||$). In particular we consider this technique applied to Theorem 4.6.

Bounds on $(I-P_{np}K)^{-1}\big|_{P_{np}R}$ are given in §4.6 in terms of W_n and other factors, also

$$||(I-P_{np}K)^{-1}|| \leq 1 + ||(I-P_{np}K)^{-1}\big|_{P_{np}R}|| \cdot ||P_{np}K|| \text{ and}$$

$$||(I-KP_{np})^{-1}|| \leq 1 + ||K|| \cdot ||(I-P_{np}K)^{-1}\big|_{P_{np}R}||$$

To express this in $[X_2, X_1]$ terms we evaluate $||P_{np}K||$ in $[X_2]$, $||K||$ in $[X_1]$ and use the now weighted form of $||W_n||$ developed in §3.6 (Note also $||I_{21}|| \leq 1$).

Summarising, we find bounds in $[X_2, X_1]$ for 4.25 and 4.26 simply by using the weighted row'norm described in §3.6 - the

appropriate weight is determined simply by examining which interval $[s_{i-1}, s_i)$ the collocation point corresponding to that row falls into.

To develop bounds on $(I-P_{np})K$ in the space $[X_1, X_2]$ we must re-examine the analysis of §4.7. Let w(t) be the piecewise constant function defined by

$$w(t) = w_i \quad t \in [s_{i-1}, s_i) \tag{4.53}$$

Then $||x||_{X_1} = ||wx||_{X_2}$

Recall the expression for DKx :

$$-\Big[DKx\Big](s)=p(s)x(s)+\int_{-1}^{1}\left((p'(s)+q(s))\frac{\partial g}{\partial s}(s,t)+q'(s)g(s,t)\right)x(t)\,dt$$

Substitute $z(t) = x(t)\cdot w(t)$ with $||x||_{X_1} \leq 1$ (so that $||z||_{X_2} \leq 1$)

$$-\Big[DKx\Big](s)=p(s)\frac{z(s)}{w(s)}+\int_{-1}^{1}\left((p'(s)+q(s))\frac{\partial g}{\partial s}(s,t)+q'(s)g(s,t)\right)\frac{z(t)}{w(t)}\,dt$$

Taking norms in the space $[X_1]$.

$$||DK|| \leq \sup_{s\,[-1,1]}\left(p_o(s)+(p_1(s)+q(s))\int_{-1}^{1}\left|\frac{\partial g}{\partial s}(s,t)\frac{w(s)}{w(t)}\right|dt+q_1(s)\int_{-1}^{1}\left|g(s,t)\frac{w(s)}{w(t)}\right|dt\right.$$

Consider $$\int_{-1}^{1}\left|g(s,t)\frac{w(s)}{w(t)}\right|dt \tag{4.54}$$

$$= \int_{-1}^{s} \tfrac{1}{2}(1-s)(1+t)\frac{w(s)}{w(t)}\,dt + \int_{s}^{1} \tfrac{1}{2}(1+s)(1-t)\frac{w(s)}{w(t)}\,dt$$

And $$\int_{-1}^{1} \left|\frac{\partial g}{\partial s}(s,t)\frac{w(s)}{w(t)}\right| dt \qquad (4.55)$$

$$= \int_{-1}^{s} \tfrac{1}{2}(1+t)\frac{w(s)}{w(t)}\, dt + \int_{s}^{1} \tfrac{1}{2}(1-t)\frac{w(s)}{w(t)}\, dt$$

Now provided

$$\left.\begin{aligned} &\text{(i)}\ \frac{1}{w(t)} \leq \max\left(\frac{1}{1+t}, \frac{1}{1-t}\right) \\ &\text{and (ii)}\ w(s) \leq w(t) \text{ for } |s| \geq |t| \end{aligned}\right\} \qquad (4.56)$$

we have

$$\frac{1+t}{w(t)} \leq 1 \quad \text{for} \quad t \leq 0$$

$$\frac{1-t}{w(t)} \leq 1 \quad \text{for} \quad t \geq 0$$

$$\frac{w(s)}{w(t)} \leq 1 \quad \text{for} \quad |s| \geq |t|$$

Giving the following bounds for (4.54) and (4.55)

$$\int_{-1}^{1} \left|\frac{\partial g}{\partial s}(s,t)\, \frac{w(s)}{w(t)}\right| dt \leq 1 + \frac{|s|}{2} \qquad (4.57)$$

$$\int_{-1}^{1} \left|g(s,t)\, \frac{w(s)}{w(t)}\right| dt \leq \frac{1+|s|}{2} - s^2 \qquad (4.58)$$

If (i) and (ii) are not satisfied bounds still exist but may be much larger.

Finally, in $[X_1]$ terms,

$$||DK|| \leq \sup_{s\in[-1,1]} \left(p_o(s) + \left(p_1(s)+q_o(s)\right)\left(1+\frac{|s|}{2}\right) + q_1(s)\left(\frac{1+|s|}{2} - s^2\right)\right) \qquad (4.59)$$

The fact that the bound on (DKx) (s) is larger in an interval with small weight is offset by the fact that due to (4.56) this interval must be narrower and the interpolation error bounds correspondingly

smaller. The inequalities (4.57) (4.58) give the following bound on K in $[X_1]$.

$$||K|| \leq \sup_{s\in[-1,1]} \left(p_o(s) \left(1+\frac{|s|}{2}\right) + q_o(s)\left(\frac{1+|s|}{2} - s^2\right)\right) \quad (4.60)$$

Using these results the analysis of §4.7 extends quite simply to give bounds on $(I-P_{np})K^d$ in $[X_1, X_2]$. Unfortunately $||K||$ in $[X_1, X_2]$ is not bounded with m and this places a further restriction on the weights if we wish to use this analysis in higher order cases (d > 0). To bound K in $[X_1, X_2]$ note that if we take

$$W_i = \tfrac{1}{2}(s_i - s_{i-1}) \quad (4.61)$$

then noting that $\int_{-1}^{1} 1/w(t)\, dt \leq 2m$, we have

$$\int_{-1}^{1} \left| g(s,t)\frac{1}{w(t)} \right| dt \leq m \quad (4.62)$$

$$\int_{-1}^{1} \left| \frac{\partial g}{\partial s}(s,t)\frac{1}{w(t)} \right| dt \leq 2m \quad (4.63)$$

We are now in a position to select the points s_i in such a manner that the weights w_i reduce the effect of large modulus row sums of the matrix W_n. The elements in rows of Wn corresponding to points near ±1 can grow alarmingly as K increases in size due to the boundary layer effect. The above analysis shows that we can improve the applicability of error bounds, that is obtain strict error bounds for fewer collocation points, at the expense of larger error bounds on the second derivative near the points ±1. Error bounds on the actual approximate solution values are not greatly increased since from (4.62) we have, in the space $[X_1, X_2]$

$$||G|| \leq 2 \quad (4.64)$$

If these procedures for producing error bounds for approximate solutions were part of a computer program for solving the differential equation it would be a simple matter to arrange that should large modulus row sums in W_n occur, then small intervals with small weights (subject to (4.56)) are introduced to improve the conditions of applicability. Also, since $\{s_i\} \subset \{ti\}$, this would suggest using smaller intervals for the piecewise polynomial approximation near -1 or +1 or both if the inverse matrix W_n has large row sums in these "positions". In view of (4.56) it would seem logical to distribute these intervals in inverse proportion to the modulus row sums of W_n. Since m may be much smaller than n it is possible that one weight would apply to several rows of W_n.

The effect of W_n on the error bounds in the weighted norm is particularly interesting in view of the close relationship between W_n and $(I-K)^{-1}$ established in (3.34), which would again suggest that poorer error bounds on the second derivative might be expected in positions corresponding to large modulus row sums of W_n.

CHAPTER 5

Examples

§5.1 Introduction

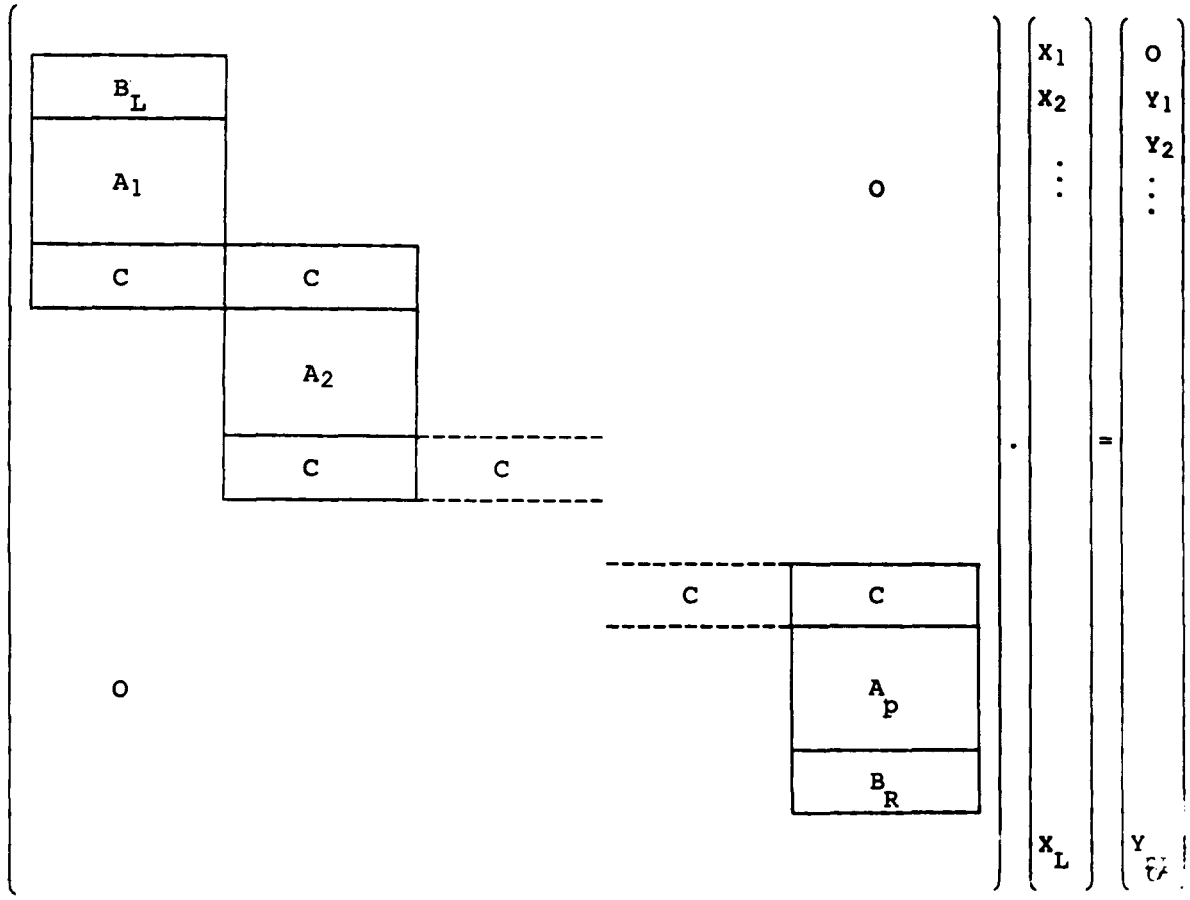
In this chapter we present a selection of numerical results illustrating both the motivation for the investigations in previous chapters and their consequences. All calculations were performed in double precision arithmetic on an IBM 360/370 computer.

A brief description of the program, written in ALGOLW, is now given. The program basically sets up and solves the linear algebraic system for the piecewise polynomial collocation method, producing the approximate solution x_n and the residual r_n for any given p, q and y. This program represents a considerable fraction of the work involved in this thesis and was developed over an extended period. It consists of a sequence of procedures, or subroutines, each of which can perform one well defined operation. These are followed by the program proper which calls into action the procedures required to solve a particular problem. An assortment of short procedures generates various arrangements of points in the interval $[-1,1]$, another procedure translates these to any given interval $[a, b]$. These are the "collocation" points used in generating the approximate solution. Much of the computation involving the polynomial forms of the piecewise sections is performed in terms of Chebychev series for reasons of numerical stability. Two further procedures define the coefficients of the problem, p and q in (4.1), and the boundary conditions.

The algebraic problem, corresponding to the approximation arising from the collocation projection (§4.3), is expressed in terms of a block matrix. Each of the blocks of this matrix relates the point values of the second derivative of the approximation x_n to the point values of the right hand side y. There are cross conditions coupling these blocks which represent the continuity

requirements of x_n and x'_n. These extra equations could be eliminated by constructing basis polynomi als for x_n which automatically satisfy the continuity requirements.

Block matrix structure:

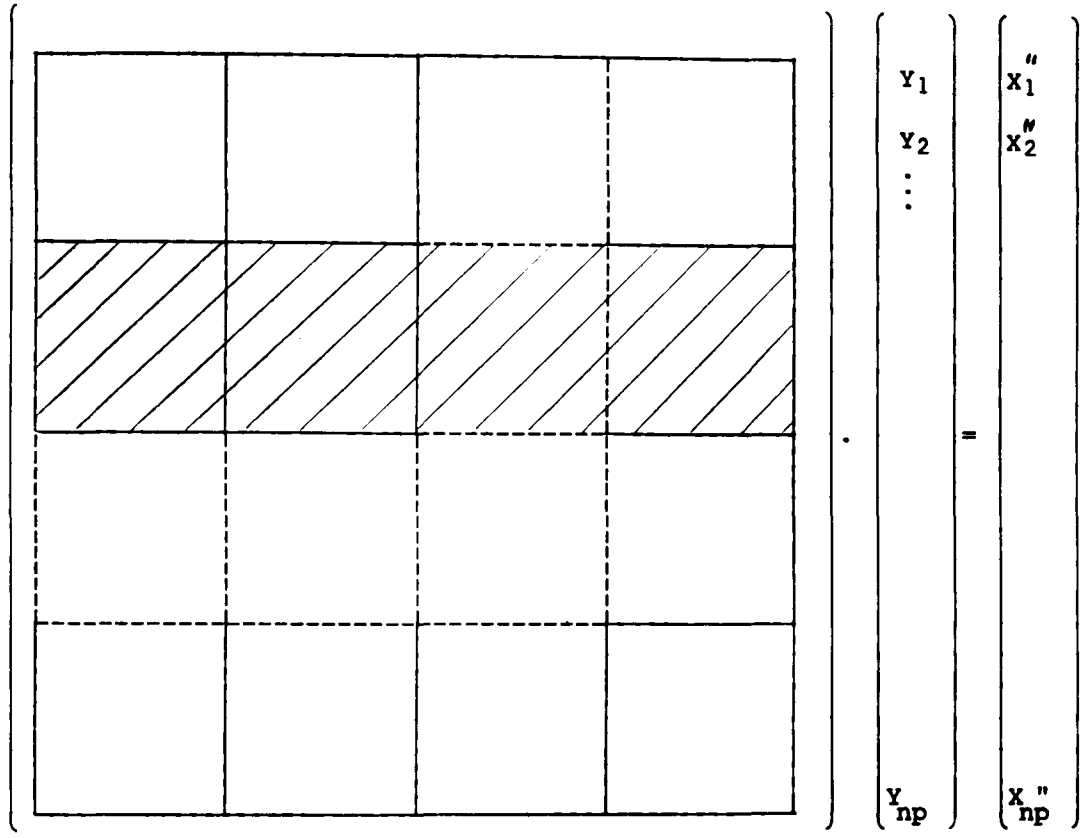


The values in regions A represent the collocation conditions within each sub interval. The values in regions B represent the boundary conditions x ($\pm$1) = O. The values in regions C represent the cross conditions required to ensure continuity of x_n and x'_n. Since there are an equal number of collocation points within each sub interval each block A is the same size. Regions B and C have similar structures each expressing the values of x_n and x'_n at the end points of the sub intervals in terms of the values of x_n at the collocation points. The vector corresponding to the right hand side y has zeros at the ends and zeros in between each sub interval to account for the

boundary and continuity conditions. The vector of x_n'' values has NP entries ($L=(P+2)\cdot N$).

The <u>inverse</u> point collocation matrix W which expresses x_n'' in terms of y is a full matrix.

W matrix structure :



The maximum modulus row sums are computed for the rectangular sub-matrix (shaded) for the sub intervals contained within a weighted interval $[s_{i-1}, s_i)$. The block matrix equation is represented in terms of a three dimensional array, one dimension each for the partitions, collocation points and coefficients of x_n. Much of the setting up of these equations is achieved by generating simpler three dimensional arrays representing the coefficients and boundary conditions then performing a three dimensional matrix multiplication with another, fairly simple, array representing, in effect, the differential operators $d^r/(dt)^r$. The resulting block matrix is

then passed to a forward elimination procedure. This procedure was carefully written to take full advantage of the block structure of the matrix. A back substitution procedure then finally yields the approximate solution for any given right hand side y. Repeated calls of this procedure are used to generate the inverse of the original block diagonal matrix.

A small procedure at the end computes the residual, and "error" by comparison with a high order approximate solution.

§5.2 The behaviour of ||W||

The point inverse matrix W exhibits some remarkable properties. For any given (soluble) problem the matrix norm is largely unaffected by the particular form of the piecewise polynomial collocation scheme, provided this scheme satisfies the requirements of Theorem 4.8. Four second order problems are examined in detail to illustrate this behaviour. A comparison of unweighted infinity norms is made using equispaced polynomial sections in $[-1,1]$. Chebychev points are usually used as the collocation points in each sub interval. Further results show the behaviour for one problem using other partitions and other sets of collocation points, in particular the breakdown of the conclusion of Theorem 4.8 when certain of its conditions are not satisfied.

Problem 1

$$x'' + x' + x = y$$
$$x(-1) = x(1) = 0$$

Problem 2

$$x'' + 2(1+t^2)x = y$$
$$x(-1) = x(1) = 0$$

Problem 3

$$x'' - 2x = y$$

$$x(-1) = x(1) = 0$$

Problem 4

$$x'' + \frac{2x'}{t+3} - \frac{2x}{(t+3)^2} = y$$

$$x(-1) = x(1) = 0$$

N represents the number of piecewise polynomial sections used for the approximation.

P represents the number of collocation points in each section.

In tables 1 - 4 we use Chevbchev collocation points

Problem 1 : $||W||$ values

TABLE 1

P \ N	1	5	10	15	20	25
1	2.000	2.818	2.923	2.958	2.975	2.985
2	2.688	2.963	2.995	3.006	3.011	3.014
5	2.975	3.016	3.021	3.023	3.023	
10	3.013	3.023	3.025			
15	3.020	3.023				
25	3.024					

Problem 2 : $||W||$ values

TABLE 2

P \ N	1	5	10	15	20	25
1	singular	11.814	12.950	13.249	13.354	13.407
2	4.000	12.329	13.208	13.368	13.427	13.452
5	13.436	13.500	13.500	13.500	13.500	
10	13.468	13.500	13.500			
15	13.500	13.500				
25	13.500					

These relatively large values for $||W||$ occur because problem 2 is nearly singular : the equation

$$\lambda x'' + (1+t^2)x = y$$

$$x(-1) = x(1) = 0$$

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

Problem 3 : $||W||$ values

TABLE 3

P \ N	1	5	10	15	20	25
1	0.500	1.113	1.306	1.382	1.419	1.444
2	0.857	1.308	1.428	1.459	1.480	1.491
5	1.035	1.483	1.520	1.525	1.530	
10	1.184	1.512	1.536			
15	1.296	1.518				
25	1.389					

Problem 4 : ||W|| values

TABLE 4

P \ N	1	5	10	15	20	25
1	0.900	1.694	1.919	2.008	2.056	2.085
2	1.607	2.037	2.180	2.150	2.166	2.175
5	2.057	2.181	2.198	2.203	2.206	
10	2.173	2.206	2.210			
15	2.196	2.211				
25	2.208					

As further illustration consider Problem 1 using Legendre zeros as the collocation points in each of the equispaced sub intervals.

TABLE 5

P \ N	1	5	10	15	20	25
1	2.000	2.818	2.923	2.958	2.975	2.985
2	2.474	2.934	2.982	2.997	3.004	3.009
5	2.924	3.007	3.016	3.020		
10	3.000	3.021	3.023			
15	3.014	3.024				
25	3.021					

Also consider Problem 1 using Chebychev zeros for both the collocation points and defining the location of the sub intervals.

TABLE 6

P \ N	1	5	10	15	20	25
1	2.000	3.004	3.022	3.024	3.025	3.025
2	2.688	3.016	3.024	3.025	3.025	3.026
5	2.975	3.024	3.026	3.026	3.026	
10	3.013	3.025	3.026			
15	3.020	3.026				
25	3.024					

As a further illustration consider again Problem 1 with equispaced sub intervals and with equispaced collocation points within each sub interval (end points not included). For greater than a few points in each sub interval this approximation scheme does not satisfy condition (i) in (4.13) and the conclusion of Theorem 4.8 does not apply.

TABLE 7

P \ N	1	5	10	15	20	25
1	2.000	2.818	2.923	2.958	2.975	2.985
2	2.336	2.916	2.973	2.991	3.000	3.006
5	2.789	2.984	3.006	3.012	3.016	
10	3.130	6.100	5.177			
15	21.06	173.3				
25	562.6					

Now consider the case P=2 with $\xi_1 = -0.8$, $\xi_2 = -0.7$ for N=1 to 30.

TABLE 8

1	5	10	15	20	25	30
71.06	34.08	32.64	32.21	32.00	31.88	31.79

Also consider P=3 with $\xi_1 = -0.1$, $\xi_2 = 0.0$, $\xi_3 = 0.1$

TABLE 9

1	5	10	15	20	25	30
27.29	118.28	126.78	129.35	130.58	131.30	131.77

Although convergence to 3.026 is not observed in the last two examples, $||W||$ does appear to be approaching some limiting value. No further studies of this phenomenon have been made.

It is possible that the conclusion of Theorem 4.8 does hold for slightly weaker conditions than those given but these have not been determined. Theorem 4.8 is applicable to many of the commonly used approximation schemes.

In order to study the behaviour of a "stiff" problem and the later application of the weighted norm we now introduce a parameter α into Problem 1 as follows.

Problem 1A

$$x'' + \alpha x' + \alpha x = y$$

$$x(-1) = x(1) = 0$$

We study the behaviour of the matrix W for an approximation consisting of 5 equispaced sub intervals in $[-1,1]$ and a variable number P of Chebychev collocation points within each sub interval. The maximum modulus row sums are shown separately for each sub interval.

α = 10

TABLE 10

P	1	2	5	10	15
[-1,-.6]	29.440	67.873	60.342	64.468	65.268
[-.6,-.2]	12.371	1.717	2.800	3.007	3.052
[-.2, .2]	4.267	1.529	2.217	2.379	2.410
[.2, 0.6]	2.026	1.362	2.089	2.253	2.285
[0.6 1.0]	1.250	1.262	1.978	2.141	2.173

TABLE 11

α = 100

P	1	2	5	10	15
[-1,-.6]	2.602	1.851	250.552	512.407	575.036
[-.6,-.2]	2.851	1.172	14.827	6.273	2.016
[-.2, .2]	3.163	0.713	1.769	2.029	2.000
[.2, .6]	3.483	0.515	1.576	2.02o	1.988
[.6, 1]	3.852	0.573	1.563	2.015	1.981

Note that as α increases the maximum modulus row sums take longer to settle down to a steady value. Also there is a marked tendency for the largest values to occur at one end - this is in contrast to the near eigenvalue Problem 2 where the value of 13.5 is maintained over the whole range.

In order to make use of the weighted norm to reduce the effect of such large modulus row sums on the applicability of the bounds on $(I-K)^{-1}$ we are forced to place more collocation points in the region of largest modulus row sums, accordingly we now consider the following partition of $[-1, 1]$; T_5 = (-1, -0.95, -0.90, -0.3, 0.3, 1).

$\alpha = 10$

TABLE 12

P	1	2	5	10	15
[-1, -.95]	47.539	118.862	65.154	65.733	65.834
[-.95, -.9]	30.029	75.718	41.591	41.693	42.028
[-.9, -.3]	6.544	27.078	23.104	26.119	26.616
[-.3, .3]	5.019	1.605	2.104	2.335	2.379
[.3, 1]	2.767	1.129	1.935	2.177	2.233

$\alpha = 100$

TABLE 13

P	1	2	5	10	15
[-1, -.95]	16.619	1969.922	567.940	621.682	632.284
[-.95, -.9]	6.698	23.857	6.189	6.209	6.316
[-.9, -.3]	0.300	0.401	1.320	2.219	1.937
[-.3, .3]	0.365	0.598	1.525	2.229	2.039
[.3, 1]	0.406	0.700	1.471	2.261	2.057

Note that once over the initial "hurdle" and the approximation is becoming reasonably good the modulus row sums now appear to approach their limiting values more quickly. For $\alpha = 10$ this partition is a bit severe and is beginning to fill up parts of W in an undesirable fashion. For $\alpha = 100$, however, this partition is still not fine enough so that finally we study the behaviour of W using the following partitions :

$$\alpha = 10 : T_5 = (-1, -0.9, -0.8, -0.6, 0, 1)$$

$$\alpha = 100: T_5 = (-1, -0.99, -0.96, -0.9, 0, 1)$$

(See §5.4 for further discussion of these choices).

$\alpha = 10$

TABLE 14

P	1	2	5	10	15
[-1, -.9]	43.046	72.133	64.474	65.551	65.753
[-.9, -.8]	16.190	28.942	26.419	26.868	26.953
[-.8, -.6]	4.809	10.536	11.018	11.410	11.485
[-.6, 0]	0.865	1.890	2.631	2.967	3.034
[0, 1]	0.524	1.066	1.900	2.239	2.317

$\alpha = 100$

TABLE 15

P	1	2	5	14	15
[-1, -.99]	110.359	1454.910	626.147	637.031	639.179
[-.99, -.96]	21.698	412.062	222.125	234.328	236.714
[-.96, -.9]	2.360	7.427	12.112	13.473	13.750
[-.9, 0]	0.093	0.288	1.083	2.238	2.290
[0, 1]	0.133	0.410	1.379	2.242	2.238

There is not meant to be any implication that these choices of partitions for these problems will necessarily give a more accurate approximate solution. We only use such partitions in order to demonstrate the better applicability, using the weighted norm, of the bounds on $(I-K)^{-1}$. Further research may show that the modulus row sums of the inverse point matrix W provide an aid to the selection of an optimum partition, in terms of the accuracy of the approximate solution (see results in §5.6).

§5.3 Problem constants (see §4.7)

TABLE 16

	$\|\|K\|\|$	$\|\|DK\|\|$	$\|\|D^2K^2\|\|$	$\|\|D^3K^3\|\|$	$\|\|D^4K^4\|\|$	$\|\|D^5K^5\|\|$	$\|\|D^6K^6\|\|$
1	1	1	2	3	5	8	13
2	1	4	12	40	136	560	2224
3	1	2	2	4	4	8	8
4	1.25	2.25	6.47	22.65	93.40	443.62	2384.43

Some cancellation has been used in obtaining bounds on $||K||$, $||DK||$ $||D^2K^2||$ but not for the higher order bounds. Further algebraic manipulation would give smaller values. Note the relatively slow rate of increase of the bounds for problems 1 and 3 with increasing order - this is because p and q are simply constants (with zero derivatives). We also have the following numerical bounds on G and DG

$$||G|| \leq 0.5 \; ; \; ||DG|| \leq 1.0$$

Note that since we shall work with the spaces X_2 and X_1 (which has a weighted norm) as described in §4.10 we have the following bounds on G and DG in $[X_1]$

$$||G|| \leq .5625 \; : \; ||DG|| \leq 1.5$$

The various bounds on D^rK^r must also be computed in $[X_1]$. This is done in a similar manner to the procedure described in §4.7. For problem 1A we have:

TABLE 17

$\|\|K\|\|$	$\|\|DK\|\|$	$\|\|D^2K^2\|\|$	$\|\|D^3K^3\|\|$	$\|\|D^4K^4\|\|$	$\|\|D^5K^5\|\|$	$\|\|D^6K^6\|\|$
$1.75\|\alpha\|$	$2.5\|\alpha\|$	$4.25\|\alpha\|^2$	$8.63\|\alpha\|^3$	$16.07\|\alpha\|^4$	$31.16\|\alpha\|^5$	$59.27\|\alpha\|^6$

All of these constants for problems 1, 1A, 2, 3, 4 were computed by a program based on the procedure described in §4.7. Since constants

of all orders up to 6 were required the program built up a table of these constants in an iterative fashion rather than computing each constant recursively which would have duplicated much of the work involved. There is great scope for improving these bounds using algebraic manipulation and numerical maximisation procedures.

Better bounds could be developed for the weighted norm. The advantage of the bounds in §4.10 is that they are independent of the partition and weights, provided that these satisfy the conditions outlined in that section.

§5.4 Applicability

We are now in a position to compare the applicability of the bounds in Theorems 4.9 and 4.10 on the inverse operator $(I-K)^{-1}$. For applicability we must have $\Delta < 1$. The number of equispaced sub intervals required using Chebychev collocation points is now compared for problems 2, 1A for the "projection" and the "extended" approximation.

Problem 2 : Applicability (Projection)

TABLE 18

P \ d	1	2	3	4	5	6
1	58*	-	-	-	-	-
2	80*	10	-	-	-	-
5	104*	6	4	4	4	-
10	94*	5	3	2	2	2
15	87*	4	2	2	2	2
25	75*	3*	2*	1	1	1

NB *estimate based on constancy of $||W||$

Problem 2 : Applicability (Extended)

TABLE 19

P \ d	1	2	3	4	5	6
1	58*	-	-	-	-	-
2	57*	8	-	-	-	-
5	50*	5	3	3	3	-
10	40*	3	2	2	2	2
15	33*	2	2	1	1	1
25	25*	2*	1	1	1	1

Note the slightly better applicability of the "extended" method here. This is because of the bound on $||K||$ being as small as 1.0. For most practical applications we would expect $||K|| >> 1$ and then the projection method would be superior.

Problem 1A : $\alpha = 10$, Applicability (Projection) Estimate

TABLE 20

P \ d	1	2	3	4	5	6
1	6610	-	-	-	-	-
2	9340	257	-	-	-	-
5	12431	170	52	35	32	-
10	11584	118	32	20	15	14
15	10525	93	24	14	11	10
25	9079	68	17	10	7	6

Problem 1A : α = 10, Applicability (Extended) Estimates

TABLE 21

P \ d	1	2	3	4	5	6
1	66100	-	-	-	-	-
2	66051	684	-	-	-	-
5	59060	370	87	52	43	-
10	46518	236	51	27	20	17
15	38552	177	37	20	14	12
25	29881	122	25	13	9	8

It can now be seen that in order to produce strict bounds on $(I-K)^{-1}$ we could be required to find bounds on the inverse of matrices far larger than that required to produce a "good" approximate solution. For this reason it is important to keep the bounds on the inverse approximate operator as small as possible. It would be particularly useful to find an extension from the subspace $P_n X$ to X which does not involve the operator K since the bound on $||K||$ (10) and the large row sums of N are the main cause of the poor applicability in problem 1A. Note that it is particularly important to make use of higher derivative bounds when these are available.

The effect of the large modulus row sums in problem 1A ($\alpha >> 1$) can be reduced by using the weighted norm. We compare the result of using an unweighted equispaced partition and the non uniform partition P : (-1, -0.99, -0.96, -0.9, 0, 1) for α = 100.

TABLE 22

P	min. w_i	w_i	width	W w_i
-1, -0.99	.01	0.01	0.01	6.392
-0.99, -0.96	.04	0.04	0.04	9.469
-0.96, -0.9	.1	0.1	0.06	1.375
-0.9, 0	.1	1	0.9	2.290
0, 1	1	1	1	2.238

The column min w_i indicates the minimum value we may take for w_i in accordance with condition (i). We use the w_i indicated in order not to make the bounds on DK, D^2K^2 in $[X_1]$ excessive.

TABLE 23

Term	Unweighted	Weighted
$\|\|(DK)_{11}\|\|$	100	250
$\|\|(D^2K^2)_{11}\|\|$	20000	42500
$\|\|((I-P_{np})K)_{12}\|\|$	4.28	26.75
$\|\|((I-P_{np})K^2)_{12}\|\|$	6.90	91.59
$\|\|W\|\|$	575	9.47
$\|\|K_{22}\|\|$	100	100
$\|\|(I-P_{np}K)^{-1}{}_{21}\|\|$	427801	7047
Δ_1	1.83'6	1.89'5
Δ_2	2.95'6	6.45'5

for α=10 with P : (-1, -.9, -.8, -.6, 0, 1) and weights 0.1, 0.1, 0.2, 1, 1 the reduction in Δ is not so apparent since the modulus row sum value near -1 is not as pronounced, however we obtain

Term	Unweighted	Weighted
Δ_1	2082	1313
Δ_2	335	450

The reduction in the value of Δ by using a non-uniform partition and a weighted norm in the first example is considerable. It is difficult to demonstrate actual applicability in these cases because K and $||W||$ are so large and a very finely divided partition would be required. We would have to compute the inverse of a very large matrix and decide how to distribute the partition. Using Δ_2 for α=100 demonstrates that should we further subdivide the partitions indicated the weighted norm would require inversion of a matrix something less than a half the size originally required.

It would be useful to determine a procedure for making best use of the partition and weights for improving the applicability. The examples above are, perhaps, a little unfair in leaving the partition so coarse away from -1.

In conclusion, it must be stated that although we can in principle determine strict bounds on $(I-K)^{-1}$ from Theorems 4.9, 4.10 these bounds will only be applicable with a reasonable amount of work when K and W are not too large. We have achieved applicability for problems 1-4 with PxN < 100 but not for problem 1A (α=10, 100). It is most important, if we are to apply this theory to a useful range of problems, to determine better bounds on the "approximation" error $||(I-P_{np})K^d||$ and on the inverse approximate operators $||(I-P_{np}K)^{-1}||$ or $||(I-K\,P_{np})^{-1}||$.

§5.5 Bounds on $(I-K)^{-1}$

We now examine the "bounds" on $(I-K)^{-1}$ for problems 2 and 1A as given by the "projection" and "extended" approximations. These "bounds" are really estimates since we have assumed $\Delta \to 0$ as $N \to \infty$ in using Theorems 4.9 and 4.10.

Problem 2 : Bounds on $(I-K)^{-1}$ (Projection)

TABLE 24

P \ d	1	2	3	4	5	6
1	14.5	-	-	-	-	-
2	28	29	-	-	-	-
5	60.8	61.8	62.8	63.8	64.8	-
10	84.7	85.7	86.7	87.7	88.7	89.7
15	101.4	102.4	103.4	104.4	105.4	106.4
25	125.5	126.5	127.5	128.5	129.5	130.5

Problem 2 : Bounds on $(I-K)^{-1}$ (Extended)

TABLE 25

P \ d	1	2	3	4	5	6
1	15.5	-	-	-	-	-
2	21.1	22.1	-	-	-	-
5	30.4	31.4	32.4	33.4	34.4	-
10	35.6	36.6	37.6	38.6	39.6	40.6
15	38.8	39.8	40.8	41.8	42.8	43.8
25	43	44	45	46	47	48

Problem 1A : α=10 Bounds on $(I-K)^{-1}$ (Projection)

TABLE 26

P \ d	1	2	3	4	5	6
1	661	-	-	-	-	-
2	1321	13211	-	-	-	-
5	2925	29251	292511	2.93'6	2.93'7	-
10	4093	40931	409311	4.09'6	4.09'7	4.09'8
15	4912	49121	491211	4.91'6	4.91'7	4.91'8
25	6086	60861	608611	6.09'6	6.09'7	6.09'8

Problem 1A : α=10 Bounds on $(I-K)^{-1}$ Extended

TABLE 27

P \ d	1	2	3	4	5	6
1	6611	-	-	-	-	-
2	9341	93411	-	-	-	-
5	13901	139011	1.39'6	1.39'7	1.39'8	-
10	16441	164411	1.64'6	1.64'7	1.64'8	1.64'9
15	18021	180211	1.80'6	1.80'7	1.80'8	1.80'9
25	20051	200511	2.01'6	2.01'7	2.01'8	2.01'9

It is an unfortunate fact that the bounds on $||(I-K)^{-1}||$ are largest under precisely the conditions we need for best applicability. Since we have not defined N in these tables we have assumed $N \to \infty$ and $\Delta \to 0$ and the bounds given above are bounds on the numerator in Theorem 4.9 and 4.10. To give a full description of the bounds and estimates for each N, P, d and problem with projection and extended methods would simply generate too much data.

Theorem 4.8 gives the following estimates for $||(I-K)^{-1}||$, based on the convergence observed in tables 1 - 4.

TABLE 28

Problem	$||(I-K)^{-1}||$
1	3.03
2	13.50
3	1.5
4	2.2

Bounds on $(I-K)^{-1}$ using the weighted norm are obtained in a similar manner but it must be remembered that in order to recover unweighted error bounds we must multiply the bound on $||(I-K)_{21}^{-1}||$

by $1/w_i$ in each corresponding interval $[s_{i-1}, s_i)$.

§5.6 Residual and Error

Finally we compare the residual and "error" in the second derivative of the approximate solution x_n (r_n and e_n in the terminology of §4.9). This "error" is obtained by comparison with a high order (75) global polynomial collocation solution using Chebychev zeros. For each problem we take the right hand side $y \equiv 1$. For each combination of N (the number of equispaced intervals in $[-1, 1]$) and P (the number of Chebychev collocation points within each interval) the upper represents a bound on e_n, the lower a bound on r_n. The numbers given are the maximum of the values obtained by comparisons at 100 equispaced points in each interval and are not strict bounds.

Problem 1 : Error and residual

TABLE 29

P \ N	1	5	10	15	20	25
2	0.168	1.46'-2	4.44'-3	2.09'-3	1.21'-3	7.91'-4
	0.164	1.47'-2	4.32'-3	2.02'-3	1.17'-3	7.59'-4
5	2.31'-4	2.46'-7	9.09'-9	1.23'-9		
	2.26'-4	2.47'-7	9.09'-9	1.23'-9		
10	9.97'-10					
	9.87'-10					

Problem 2 : Error and residual

TABLE 30

P \ N	1	5	10	15	20	25
2	9.56	1.39	0.382	0.172	9.67'-2	6.19'-2
	2.26	0.726	0.214	9.80'-2	5.56'-2	3.58'-2
5	0.105	4.70'-4	1.66'-5	2.20'-6	5.29'-7	1.74'-7
	0.121	4.63'-4	1.66'-5	2.19'-6	5.27'-7	1.74'-7
10	2.10'-4	4.82'-11				
	2.10'-4	4.47'-11				
15	7.21'-9					
	7.18'-9					

Problem 3 : Error and residual

TABLE 31

P \ N	1	5	10	15	20	25
2	0.273	1.49'-2	4.29'-3	2.00'-3	1.15'-3	7.50'-4
	0.332	1.60'-2	4.43'-3	2.04'-3	1.17'-3	7.58'-4
5	2.68'-4	5.99'-7	2.19'-8	3.04'-9	7.40'-10	
	2.75'-4	6.02'-7	2.20'-8	3.04'-9	7.41'-10	
10	8.35'-9					
	8.36'-9					

Problem 4 : Error and residual

TABLE 32

P \ N	1	5	10	15	20	25
2	0.296	5.40'-2	1.77'-2	8.63'-3	5.10'-3	3.36'-3
	0.401	5.02'-2	1.61'-2	7.84'-3	4.62'-3	3.04'-3
5	9.97'-3	2.67'-5	1.22'-6	1.83'-7	4.65'-8	
	9.89'-3	2.68'-5	1.21'-6	1.83'-7	4.64'-8	
10	7.96'-6	2.89'-11				
	7.99'-6	2.88'-11				
15	3.19'-9					
	3.22'-9					

Smaller entries are omitted in order that the effect of rounding errors does not become apparent. Note the close relationship between e_n and r_n (see §4.9). As further evidence of this property we consider problems 1 and 2 using Legendre zeros for the collocation points in each interval and problem 4 with a modified right hand side y= -1/(t+3).

The accuracy with which $||e_n||$ follows $||r_n||$ for the Legendre zeros is in accordance with the theory in §4.9. The correspondence for Chebychev zeros is also quite good but not as striking.

Problem 1L : Error and residual

TABLE 33

P \ N	1	5	10	15	20	25
2	0.217	1.99'-2	5.72'-3	2.67'-3	1.53'-3	9.93'-4
	0.241	1.98'-2	5.72'-3	2.67'-3	1.53'-3	9.93'-4
5	4.52'-4	4.30'-7	1.58'-8			
	4.47'-4	4.31'-7	1.58'-8			
10	1.48'-9					
	1.46'-9					

Problem 2L : Error and residual

TABLE 34

P \ N	1	5	10	15	20	25
2	7.57	1.08	0.288	0.129	6.42'-2	4.11'-2
	7.65	1.08	0.288	0.129	6.42'-2	4.11'-2
5	0.278	8.15'-4	2.90'-5	1.91'-6	4.58'-7	1.51'-7
	0.278	8.15'-4	2.90'-5	1.91'-6	4.58'-7	1.51'-7
10	3.10'-4	6.57'-11				
	3.10'-4	6.61'-11				
15	1.09'-8					
	1.09'-8					

Problem 4 : y ≡ -1/(t+3) Error and residual

TABLE 35

P \ N	1	5	10	15	20	25
2	0.145	2.57'-2	8.37'-3	4.09'-3	2.41'-3	1.59'-3
	0.193	2.39'-2	7.63'-3	3.71'-3	2.18'-3	1.44'-3
5	4.65'-3	1.24'-5	5.66'-7	8.50'-8	2.16'-8	
	4.61'-3	1.24'-5	5.63'-7	8.47'-8	2.15'-8	
10	3.66'-6	1.33'-11				
	3.70'-6	1.33'-11				
15	1.48'-9					
	1.49'-9					

In order to demonstrate how useful the matrix W might be in studying stiff problems we turn now to problem 1A with α=10, first we consider an equispaced partition E and Chebychev zero as for problems 1 - 4.

Problem 1A : α=10 Error and residual

TABLE 36

P \ N	1	5	10	15	20	25
2	61.4	30.0	8.49	4.58	2.93	2.03
	3.67	25.9	7.04	3.65	2.26	1.54
5	22.9	0.134	8.75'-3	1.43'-3	3.89'-4	1.39'-4
	7.6	0.131	8.46'-3	1.41'-3	3.83'-4	1.36'-4
10	0.124	2.15'-6	4.70'-9			
	0.126	2.11'-6	4.64'-9			
15	3.01'-4					
	3.01'-4					
25	4.43'-11					
	4.27'-11					

Now consider the partition P:(-1, -0.9, -0.8, -0.6, 0, 1). We compare the solution error, second derivative error, residual using Chebychev, Legendre collocation points.

Problem 1A : α=10 Errors and residual

TABLE 37

P	2	5	10	15	Partition	Points
Error	0.328	2.97'-4	1.43'-9	7.10'-14	Equispaced	Chebychev
in	2.86'-2	7.19'-5	2.44'-9	7.21'-14	P	Chebychev
x_n	5.48'-3	2.50'-5	8.52'-10	7.35'-14	P	Legendre
Error	30.0	0.134	2.15'-6	4.90'-12	Equispaced	Chebychev
in	5.06	1.48'-2	1.67'-6	3.86'-11	P	Chebychev
x_n''	2.44	1.40'-2	1.60'-6	3.02'-11	P	Legendre
Residual	25.9	0.131	2.11'-6	3.11'-11	Equispaced	Chebychev
	2.36	1.44'-2	1.63'-6	3.56'-11	P	Chebychev
	2.56	1.22'-2	1.62'-6	3.05'-11	P	Legendre

The errors have been considerably reduced by using the non-uniform partition and the error in the actual approximate solution reduced still further by using Legendre zeros as the collocation points. These results suggest that the most efficient way of solving stiff problems by collocation would be to use Legendre zeros and a finely spaced non-uniform partition.

These numerical investigations are by no means exhaustive and further studies might point the way towards more theoretical investigations. It is hoped that at least they are sufficient to illustrate the practical consequences of the theory in Chapters 2 - 4.

CHAPTER 6

Conclusions

§6.1 Theory

An operator approximation theory has been developed in Chapter 2 which, using the concept of two spaces with different norms, has enabled an improvement to be made in the applicability of computable a-posteriori error bounds. Other work in this field has used similar theory in the more conventional setting of a single space X. We cannot allow these spaces to be different in a topological sense but there is sufficient latitude in the choice of metrics (norms) to permit us to take advantage of the form of certain operators especially for "stiff" problems which turn out to be most difficult to subject to error analysis. With further work it might be possible to produce bounds on the inverse operator $(I-K)^{-1}$ of a more suitable form. The major problem with the bounds produced in this thesis is that for many practical problems an inordinate amount of work is necessary to produce any strict bound at all.

Approximations of the "projection" and "extended projection" type are related in a general manner. This leads to a generalisation which includes a very wide class of approximation methods. Chapter 3 continues to study the particular form of the inverse approximate operator for this generalised approximation and finally relates this operator directly to a matrix inverse in the finite dimensional case. So called "transfinite" methods are not considered here but perhaps the theory could be developed in this direction if desired. It is most difficult to develop satisfactory bounds on the inverse approximate operator. The bounds developed here might not be suitable for all applications and it is in this area that perhaps the most profitable improvements could be made.

This chapter concludes with a convergence theorem of particular importance to the later discussion of error estimates. It is also of interest in its own right since its conclusion is both simple and highly significant : $||W_n|| \to ||(I-K)^{-1}||$. Some of the conditions of this theorem raise certain problems, namely in the selection of a J_n^* to satisfy (ii), (iii) and (v). These conditions might be better phrased or even eliminated in favour of some more simple criteria. Perhaps the most important feature of this chapter is that the theory is applicable to a wide range of problems from initial value to partial differential equations and for approximations including collocation, quadrature methods, finite elements, etc.

§6.2 Application

Some simple boundary value problems in ordinary differential equations are used to illustrate an application of the theory. A piecewise polynomial approximation is used and a collocation projection generates the approximate equation. The conditions required by the theory in chapters 2 and 3 are expressed in simple algebraic terms which helps to throw some light on the effectiveness of different approximations. A slight modification is required in order to apply Theorem 3.7, again concerning the conditions relating to J_n^*. Computable bounds are derived on $||(I-K)^{-1}||$ which can be used to bound the error in terms of the residual. In particular the advantage of using Legendre zeros for the piecewise polynomial collocation method is demonstrated. It would be most useful to determine criteria which govern the close relationship between the second derivative error and the residual. If this could be expressed in the general setting of the previous chapters we might

then have a widely applicable and highly effective error estimate. The conditions relating to the use of a "weighted" norm are described in simple terms and various quantities are computed which enable a straightforward use of the previous procedures. The weighted norm allows use to be made of the location of the partition points of the piecewise approximation in order to reduce the effect of large modulus row sums in the matrix W. These large rows in W can occur near $\pm$ 1 due to the "stiff" behaviour of $(I-K)^{-1}$ for large K. The improvement in applicability is considerable but this is at the expense of poorer error bounds for the second derivative of the approximate solution in the vicinity of $\pm$ 1. Error bounds for the approximate solution itself are not so much affected.

Chapter 5 completes the discussion of this application with a selection of problems and a summary of the numerical properties of various piecewise polynomial approximations. A study of the relative effectiveness of these different schemes in terms of applicability of Theorems 4.9 and 4.10, the bounds given by them and the relation between the residual and second derivative error is given. The conclusions are that practical bounds on $(I-K)^{-1}$ may be computed provided that K is not too large and provided that the problem is not too near singular.

Considerable improvements could be made in the numerical estimates of certain quantities such as $||D^r K^r||$ and it would be possible to arrange for the automatic selection of partition and weights with the objective of minimising Δ. The most important conclusion of the examination of residuals and errors for these problems is that a very good estimate for $||e_n||$ is given by $||r_n||$, in fact we have $||e_n - r_n|| << ||r_n||$ for n sufficiently large by using Legendre zeros as the collocation points in each sub interval.

Appendix 1

Interpolation Constants

The interpolation error bounds in Table 38 are those derived from Jacksons theorem (see (4.40)) with n=1. Entries in column d give bounds on the infinity norm of the interpolation error of a function f with $||f^{(d)}|| \leq 1$. Bounds on $||P_{1p}||$ for interpolation at Chebychev zeros are given by Powell 37 . These numbers are all easily computed and are included for comparison with the interpolation error bounds given by Peano's Theorem (Davis 13 , p.70):

Let $L(p) = 0$ for all $p \in P_n$. Then, for all $f \in C^{d+1}[a, b]$ for $d \leq n$

$$L(f) = \int_a^b f^{(d+1)}(t)\, K(t)\, dt$$

where $K(t) = \frac{1}{d!} L_x (x-t)_+^d$

and $$(x-t)_+^d = \begin{cases} (x-t)^d & x \geq t \\ 0 & x < t \end{cases}$$

In our notation we have $P_{1p}(q) = 0$ for all $q \in P_{p-1}$ consequently we take $f \in C^d[-1, 1]$ with $d \leq p$. The values in the tables were computed using exactintegration formulae and approximate maximisation and thus are not strict bounds but these could be determined with more effort. Each entry is computed from the expression

$$\max_{x \in [-1,1]} \int_{-1}^{1} \left| \frac{1}{(d-1)!} P_{1p} (x-t)_+^{d-1} \right| dt$$

where P_{1p} is applied to $(x-t)_+^{d-1}$ considered as a function of x.

TABLE 38

JACKSON CONSTANTS

P	1	2	3	4	5
1	3.140	-	-	-	-
2	1.900	2.9800	-	-	-
3	1.400	1.1000	1.720000	-	-
4	1.120	0.5860	0.460000	0.722000	-
5	0.939	0.3690	0.193000	0.152000	0.238000
6	0.813	0.2550	0.100000	0.052500	0.041200
7	0.719	0.1880	0.059100	0.023200	0.012200
8	0.645	0.1450	0.037900	0.011900	0.004680
9	0.587	0.1150	0.025900	0.006770	0.002130
10	0.539	0.0940	0.018500	0.004140	0.001080
11	0.498	0.0783	0.013700	0.002680	0.000602
12	0.464	0.0663	0.010400	0.001820	0.000357
13	0.434	0.0569	0.008120	0.001280	0.000223
14	0.409	0.0494	0.006460	0.000923	0.000145
15	0.386	0.0433	0.005230	0.000685	9.78'-05
16	0.366	0.0383	0.004300	0.000520	6.80'-05
17	0.348	0.0342	0.003580	0.000401	4.85'-05
18	0.332	0.0307	0.003010	0.000315	3.54'-05
19	0.317	0.0277	0.002560	0.000251	2.63'-05
20	0.304	0.0251	0.002190	0.000203	1.99'-05
21	0.292	0.0229	0.001890	0.000165	1.53'-05
22	0.281	0.0210	0.001650	0.000136	1.19'-05
23	0.270	0.0193	0.001440	0.000113	9.38'-06
24	0.261	0.0178	0.001270	9.52'-05	7.47'-06
25	0.252	0.0165	0.001130	8.04'-05	6.02'-06
26	0.244	0.0153	0.001000	6.85'-05	4.89'-06
27	0.236	0.0143	0.000897	5.87'-05	4.01'-06
28	0.229	0.0133	0.000805	5.06'-05	3.31'-06
29	0.222	0.0125	0.000726	4.39'-05	2.76'-06
30	0.216	0.0117	0.000657	3.82'-05	2.31'-06

TABLE 39

CHEBYCHEV ZEROS (INTEGRAL |U|)

P	1	2	3	4	5	6	7	8	9	10
1	1.000	-	-	-	-	-	-	-	-	-
2	0.707	0.25000	-	-	-	-	-	-	-	-
3	0.547	0.11300	0.083300	-	-	-	-	-	-	-
4	0.495	0.07320	0.031900	0.031300	-	-	-	-	-	-
5	0.425	0.04900	0.015600	0.009800	0.012500	-	-	-	-	-
6	0.391	0.03790	0.009700	0.004470	0.003520	0.005210	-	-	-	-
7	0.351	0.02900	0.006280	0.002370	0.001390	0.001300	0.002230	-	-	-
8	0.327	0.02400	0.004540	0.001450	0.000691	0.000482	0.000517	0.000977	-	-
9	0.301	0.01950	0.003290	0.000923	0.000375	0.000215	0.000174	0.000208	0.000434	-
10	0.283	0.01680	0.002560	0.000640	0.000228	0.000111	7.35'-05	6.63'-05	8.71'-05	0.000195
11	0.264	0.01420	0.001980	0.000448	0.000143	6.16'-05	3.49'-05	2.60'-05	2.58'-05	3.69'-05
12	0.251	0.01250	0.001610	0.000332	9.60'-05	3.17'-05	1.85'-05	1.18'-05	9.66'-06	1.04'-05
13	0.237	0.01090	0.001300	0.000242	6.55'-05	2.31'-05	1.04'-05	5.84'-06	4.12'-06	3.67'-06
14	0.226	0.00973	0.001080	0.000192	4.70'-05	1.52'-05	6.21'-06	3.15'-06	1.96'-06	1.51'-06
15	0.214	0.00862	0.000900	0.000149	3.39'-05	1.02'-05	3.82'-06	1.77'-06	9.94'-07	6.77'-07
16	0.206	0.00782	0.000770	0.000120	2.55'-05	7.12'-06	2.48'-06	1.06'-06	5.43'-07	3.33'-07
17	0.197	0.00704	0.000655	9.56'-05	1.92'-05	5.02'-06	1.64'-06	6.47'-07	3.07'-07	1.72'-07
18	0.189	0.00644	0.000569	7.87'-05	1.49'-05	3.68'-06	1.12'-06	4.15'-07	1.83'-07	9.47'-08
19	0.182	0.00586	0.000492	6.45'-05	1.16'-05	2.70'-06	7.77'-07	2.70'-07	1.11'-07	5.37'-08
20	0.176	0.00541	0.000434	5.41'-05	9.25'-06	2.04'-06	5.56'-07	1.82'-07	7.06'-08	3.19'-08
21	0.169	0.00495	0.000379	4.51'-05	7.34'-06	1.54'-06	3.99'-07	1.24'-07	4.53'-08	1.93'-08
22	0.164	0.00462	0.000339	3.86'-05	6.00'-06	1.20'-06	2.96'-07	8.74'-08	3.03'-08	1.22'-08
23	0.158	0.00427	0.000301	3.27'-05	4.87'-06	9.34'-07	2.20'-07	6.18'-08	2.04'-08	7.77'-09
24	0.154	0.00399	0.000271	2.83'-05	4.04'-06	7.42'-07	1.67'-07	4.49'-08	1.41'-08	5.12'-09
25	0.149	0.00371	0.000242	2.43'-05	3.34'-06	5.88'-07	1.27'-07	3.26'-08	9.82'-09	3.40'-09
26	0.145	0.00349	0.000220	2.13'-05	2.81'-06	4.76'-07	9.88'-08	2.44'-08	7.03'-09	2.33'-09
27	0.141	0.00327	0.000198	1.85'-05	2.36'-06	3.84'-07	7.67'-08	1.82'-08	5.04'-09	1.60'-09
28	0.137	0.00308	0.000181	1.63'-05	2.01'-06	3.16'-07	6.08'-08	1.39'-08	3.70'-09	1.13'-09
29	0.133	0.00288	0.000164	1.42'-05	1.69'-06	2.58'-07	4.78'-08	1.05'-08	2.70'-09	7.94'-10
30	0.131	0.00274	0.000151	1.27'-05	1.47'-06	2.16'-07	3.87'-08	8.24'-09	2.04'-09	5.78'-10

TABLE 40

LEGENDRE ZEROS (INTEGRAL |U|)

P	1	2	3	4	5	6	7	8	9	10
1	1.000	-	-	-	-	-	-	-	-	-
2	0.845	0.3330	-	-	-	-	-	-	-	-
3	0.742	0.1570	0.13300	-	-	-	-	-	-	-
4	0.668	0.0991	0.04930	0.057100	-	-	-	-	-	-
5	0.612	0.0707	0.02580	0.017800	0.025400	-	-	-	-	-
6	0.568	0.0540	0.01580	0.008010	0.006930	0.011500	-	-	-	-
7	0.532	0.0430	0.01060	0.004330	0.002790	0.002840	0.005330	-	-	-
8	0.502	0.0354	0.00757	0.002600	0.001360	0.001040	0.001200	0.002490	-	-
9	0.476	0.0298	0.00564	0.001690	0.000749	0.000465	0.000405	0.000521	0.001170	-
10	0.454	0.0256	0.00434	0.001150	0.000447	0.000237	0.000169	0.000164	0.000230	0.000554
11	0.435	0.0223	0.00343	0.000817	0.000283	0.000132	8.04'-05	6.40'-05	6.79'-05	0.000103
12	0.418	0.0196	0.00277	0.000600	0.000188	7.82'-05	4.20'-05	2.87'-05	2.50'-05	2.88'-05
13	0.403	0.0175	0.00227	0.000452	0.000129	4.88'-05	2.35'-05	1.42'-05	1.06'-05	1.01'-05
14	0.389	0.0157	0.00189	0.000348	9.19'-05	3.18'-05	1.39'-05	7.52'-06	5.00'-06	4.07'-06
15	0.377	0.0142	0.00160	0.000274	6.70'-05	2.14'-05	8.58'-06	4.23'-06	2.53'-06	1.83'-06
16	0.366	0.0129	0.00136	0.000218	4.99'-05	1.48'-05	5.50'-06	2.50'-06	1.36'-06	8.86'-07

BIBLIOGRAPHY

(1) P.M. Anselone (1971), "Collectively Compact Operator Approximation Theory". Prentice Hall.

(2) R. Bellman and R.E. Kalaba (1965), "Quazilinearisation and Non-linear Boundary Value Problems", Elsevier.

(3) C.R. de Boor, "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 Arbour.

(4) C.R. de Boor and B. Swartz (1973), "Collocation at Gaussian points", S.I.A.M. Journal of Numerical Analysis 10, p. 582.

(5) C.R. de Boor (1973), "Good approximation by splines with variable knots", Conference on Numerical Solution of Differential equations. Lecture notes in Mathematics 363, 12-20.

(6) P.G. Ciarlet, M.H. Shultz and R.S. Varga (1967), "Numerical methods of high order accuracy for the solution of boundary value problems. I. One dimensional problem", Numer. Math. 9, p. 394.

(7) P.G. Ciarlet, M.H. Shultz and R.S. Varga (1968), "Numerical methods of high order accuracy for non-linear boundary value problems II Non-linear boundary conditions", Numer. Math. 11, p. 331.

(8) P.G. Ciarlet, M.H. Shultz and R.S. Varga (1969), "Numerical methods of high order accuracy for the solution of boundary value problems. V. Monotone operator theory." Numer. Math. 13, p. 51.

(9) C.W. Clenshaw and H.J. Norton (1963), "The solution of non-linear ordinary differential equations in Chebychev series", Computer Journal 6, p. 88.

(10) D.B. Coldrick (1972), "Methods for the numerical solution of integral equations of the second kind", Doctoral thesis, University of Toronto.

(11) D.M. Cruickshank (1974), "Error Analysis of Collocation Methods for the Numerical Solution of Ordinary Differential Equations", Doctoral thesis, University of Newcastle upon Tyne.

(12) D.M. Cruickshank and K. Wright, "Computable error bounds for polynomial collocation methods", S.I.A.M. J.N.A.15, p.134-151.

(13) P.J. Davis (1963), "Interpolation and Approximation", Blaisdell.

(14) J.C. Diaz (1977), "A collocation galerkin method for two point boundary value problems using continuous piecewise polynomial spaces", S.I.A.M. J.N.A. 14, Pps. 844-858.

(15) J.E. Flaherty and R.E. O'Malley, "The numerical solution of Boundary Value Problems for Stiff boundary value problems", Mathematics of Computation (1977) 31, Pps. 66-93.

(16) R.P. Gilbert and D.L. Colton (1971), "On the numerical treatment of partial differential equations by function theoretic methods", Proceedings of Synspade 1970, Numerical Solution of Partial Differential Equations - II, May 1970, Maryland (Ed. Hubbard), Academic Press.

(17) R.J. Hangelbrook, H.G. Koper and G.K. Leaf (1977), "Collocation methods for integro-differential equations", S.I.A.M. J.N.A.14, pps. 377-390.

(18) J. Hanson and J.L. Phillips, "An adaptive numerical method for solving Linear Fedholm Integral Equations of the First Kind". Numer, Math. 24, pps. 291-307.

(19) F.R. de Hoog and R. Weiss (1978), "Collocation methods for singular boundary value problems", S.I.A.M. J.N.A.15, pps.198-217.

(20) L.V. Kantorovich (1934), "A method of approximate solution of partial differential equations". Doklady Akademii Nauk SSSR II, p.532.

(21) L.V. Kantorovich (1948), "Functional analysis and applied mathematics", Uspekhi Matem. Nank. 3, No. 6, P. 89.

(22) L.V. Kantorovich and G.P. Akilov (1964), "Functional Analysis in Normed Spaces", Pergamon.

(23) L.V. Kantorovich and V.I. Krylov (1958), "Approximate Methods of Higher Analysis", Noordhoff.

(24) E.B. Karpilovskaja (1953), "On the convergence of an interpolation method for ordinary differential equations", Uspekhi Matam. Nank. 8, No. 3, p. 111.

(25) E.B. Karpilovskaja (1963), "Convergence of the collocation method", Sov. Math. 4, p. 1070.

(26) H.B. Keller (1968), "Numerical methods for two-point boundary value problems", Blaisdell.

(27) H.B. Keller (1969), "Accurate difference mthods for linear ordinary differential systems subject to linear constraints", S.I.A.M. 6, p. 8.

(28) W.J. Kammerer, G.W. Reddien and R.S. Varga, (1974), "Quadratic Interpolation Splines", S.I.A.M. 22, p. 241.

(29) M. Lentini and V. Pereyra(1977), "An adaptive finite difference solver for non-linear two point boundary value problems with mild boundary layers", S.I.A.M. 14, p. 91.

(30) T.R. Lucas and G.W. Reddien Jr. (1972), "Some collocation methods for non-linear boundary value problems", S.I.A.M. 9, no. 2, p. 341.

(31) T.R. Lucas and G.W. Reddien (1973), "A high order projection method for non-linear two point boundary value problems", Numer. Math. 20, p. 257.

(32) S.G. Mikhlin and K.L. Smolitskiy (1967), "Approximate methods for the Solution of Differential and Integral Equations", Elsevier.

(33) I.P. Natanson (1965), "Constructive Function Theory, Vol. III", Ungar.

(34) V. Pereyra and E.G. Sewell (1975), "Mesh selection for discrete (finite difference piecewise polynomial) solutions of boundary value problems in ordinary differential equations", Numer. Math. 23, p. 261.

(35) J.L. Phillips (1969), "Collocation as a projection method for solving integral and other operator equations", Doctoral thesis, Purdue University, Lafayette, Ind.

(36) J.L. Phillips (1972), title as above. S.I.A.M. 9, No. 1, p. 14.

(37) M.J.D. Powell (1967), "On the maximum errors of polynomial approximations defined by interpolation and by least squares criteria". Computer Journal 9, p. 404.

(38) L.B. Rall (1969), "Computational solution of Non-linear Operator Equations", Wiley.

(39) G.W. Reddien and L.L. Schumaker, (1976), "On a collocation method for Singular two point boundary value problems", Numer. Math. 25, p. 427.

(40) G.W. Reddien (1976), "Approximate methods for two point boundary value problems with non-linear boundary conditions", S.I.A.M. 13, p. 405.

(41) R.D. Russell and L.F. Shampine (1972), "A collocation method for boundary value problems", Numer. Math. 19, p.1.

(42) R.D. Russell (1977), "A comparison of Collocation and Finite Differences for two point boundary value problems". S.I.A.M. 14, p. 1.

(43) R.D. Russell and J. Christiansen (1978), "Adaptive mesh selection for solving boundary value problems". SIAM 15, p. 59.

(44) Eckard Schmidt and Peter Lancaster (1975), "Bases of splines assoçiated with constant coefficient differential operators." S.I.A.M. 12, p630.

(45) R.F. Sincovec (1977), "On the relative efficiency of higher order collocation methods for solving two point boundary value problems." S.I.A.M. 14, p112.

(46) G.M. Vainikko (1965), "On the stability and convergence of the collocation method." Differentsial'nye Uravneniya 1, p244.

(47) G.M. Vainikko (1966), "The convergence of the collocation method for nonlinear differential equations." U.S.S.R. Comp. Math. and Math. Phys. 6 No.1, p47.

(48) G.M. Vainikko (1969), "The compact approximation principle in the theory of approximation methods." U.S.S.R. Comp. Math. and Math. Phys. 9, No.4, p1.

(49) K.A. Wittenbrink (1973), "High order projection methods of moment and collocation type for Nonlinear boundary value problems." Computing II, p255-274.

(50) K. Wright (1964). "Chebychev collocation methods for ordinary differential equations." Computer Journal 6, p358.

(51) K. Wright (1979). Technical report No.135. Computing Laboratory University of Newcastle upon Tyne.