Analysis of Global Expansion Methods: Weakly Asymptotically Diagonal Systems

L. M. DELVES

and

T. L. FREEMAN

Analysis of Global Expansion Methods: Weakly Asymptotically Diagonal Systems

L. M. DELVES
Department of Computational and Statistical Science
University of Liverpool
Liverpool, UK

and

T. L. FREEMAN
Department of Mathematics
University of Manchester
Manchester, UK

1981

ACADEMIC PRESS

A Subsidiary of Harcourt Brace Jovanovich, Publishers

London New York Toronto Sydney San Francisco

ACADEMIC PRESS INC. (LONDON) LTD
24–28 Oval Road
London NW1

US edition published by
ACADEMIC PRESS INC.
111 Fifth Avenue
New York, New York 10003

British Library Cataloguing in Publication Data

Delves, L. M.
Analysis of global expansion methods. – (Computational mathematics and applications).
1. Differential equations – Asymptotic theory
2. Integral equations
3. Matrices
I. Title II. Freeman, T. L. III. Series
515.3'5 QA371

ISBN 0-12-208880-8
LCCCN 80-42084

Preface

The study of *Weakly Asymptotically Diagonal* systems originated in an attempt to analyse the convergence of expansion methods for differential and integral equations of "global" type; that is, methods which employ expansions of the form

$$f(x) = \sum_{i=1}^{N} a_i^{(N)} h_i(x),$$

where the $h_i(x)$ have global rather than local support and are typically chosen to be orthogonal polynomials in an appropriate set of variables. For a linear problem, such methods lead to a set of linear equations for the coefficients $a_i^{(N)}$, $i = 1, \ldots, N$, and it is possible to treat both the convergence and the stability properties of the method by analysing the structure of the matrix and right-hand side of these equations. Such an approach has the advantage that these properties are then characterised directly in terms of quantities which are available without additional cost during the course of the calculations: the error estimates which result are cheap to compute.

For the analysis to be of general use, it is necessary to abstract the essential structure of the equations, and to analyse the class of matrices having this structure. The definition of a *Weakly Asymptotically Diagonal* (WAD) matrix arises directly from this necessity. Historically, two subclasses of WAD matrices were introduced first: matrices of type A, which closely model the equations that arise from Fredholm integral equations, and type B, which model differential equations.

These subclasses yield a rather straightforward analysis of the convergence and stability properties of global expansion methods for one-dimensional problems, together with cheaply computable and very effective error estimates for such methods. Systems of type A in particular yield a rather simple theory, which is discussed in detail in Chapter 2.

The more abstract general WAD definitions arise from a natural wish to broaden the class of problems covered by the analysis as much as possible, and to formalise the essential results contained in the rather specific theorems of type A and type B systems. It appears that this can be achieved in quite a natural manner, and a considerable body of analysis has grown up for WAD matrices. Although it is by no means complete, it seems worthwhile to bring together the available results together with some of their applications to expansion methods; the result is this book, which splits naturally into three sections.

The theory and numerical analysis of WAD systems is discussed in Chapters 1–8; we hope that the ordering chosen, and the examples given, illustrate sufficiently the analysis to motivate this detailed study. The third section, Chapters 12–14, discusses applications of the theory to the solution of integral and differential equations. Space considerations limit the detail that can be given in these chapters. However, it is our hope that they, and also Chapter 8 on the numerical analysis of WAD systems, form a practical justification for Chapters 1–7: namely, that the study of WAD systems, and the insight which these systems give into the structure of global expansion methods, has led directly to the development of improved algorithms and to cheap and effective error estimates for a wide class of problems. The error estimates depend on the parameters of the WAD systems, and these parameters can themselves be directly estimated during the calculations. They can, however, also be related to the analytic properties (the smoothness) of the coefficients in the equation being solved. Section 2, Chapters 9–11, is devoted to the analysis of the convergence properties of orthogonal expansions, and discusses this relationship in detail. These chapters thus form a bridge between the abstract WAD theory and its applications in Chapters 12–14. They also form a bridge between the basis—dependent analysis of this book, and the more common basis—independent analysis of the convergence of variational and Galerkin methods in which the assumed smoothness properties of the equations appear directly.

This book contains a number of new and extended results, but its main purpose is to draw together previously published results. We are grateful to all of our colleagues who have worked in this field, and particularly to Drs M. Bain, K. O. Mead and F. A. Musa, for permission to draw upon their theses and published work, as listed in the references, as well as for many discussions. Without their collaboration and cooperation, this book could not have been written. Responsibility for any errors which remain in it, however, rests with us. We would also like to acknowledge with gratitude the patience and forbearance of Miss K. Anderson and Mrs G. M. Eyres, in seeing us through successive drafts.

Liverpool & Manchester
March 1981

L. M. Delves
T. L. Freeman

Contents

Section 2: Orthogonal Expansions

9. The Convergence of Orthogonal Expansions on the Finite Interval

10. Orthogonal Expansions on Infinite Intervals

11. The Convergence of Multi-dimensional Expansions

Section 3: Applications

12. The Galerkin Solution of Integral Equations

13. Global Expansion Methods for Ordinary Differential Equations

Section 1

Theory of WAD Systems

1

Introduction and Motivation

1.1 Introduction

This book is primarily concerned with the analysis of a class of infinite matrices: the class of *Weakly Asymptotically Diagonal* (WAD) matrices of the title. This analysis is interesting in its own right; however, the major interest lies (for the authors at least) in the application of the theory to problems in the general field of approximation theory. WAD matrices arise naturally when global or global element expansion methods are used to solve numerically a wide class of problems which involve integral and differential equations in one or more dimensions; the WAD assumptions were in fact introduced in order to model the structure of the matrices that arise from these problems. In this chapter we therefore try to motivate the analysis of Chapters 2–7 by giving a brief discussion of such methods. We demonstrate by example their main source of interest: the possibility of obtaining very rapid convergence; convergence which in "suitable" cases is much more rapid than that attainable with a conventional finite difference or finite element approach. However, rapid convergence is of little use in practice unless it can be recognised: what is needed is a computable error estimate which reflects the actual error as faithfully as possible while adding as little as possible to the cost of the calculations. The provision of such error estimates is one of the main achievements of the WAD theory developed here. The estimates depend on an analysis of the structure of the defining equations for the expansion method; this structure is displayed in a simple case in the example of Section 1.5. Once the structure is understood, surprisingly simple and effective error estimates follow. In addition, the structure suggests efficient ways of both setting up and of solving the defining equations; see Chapters 8 and 12–14 for a discussion of the savings which can be made.

1.2 Galerkin methods for linear operator equations

1.2.1 The Galerkin formalism

Let R be a Hilbert space with inner product $(\cdot, \cdot)$ and $\mathscr{L}: R \to R$ a linear operator defined on R (or perhaps a subspace of R). Let g be a known element of R, and f the solution (assumed to exist and to be unique in R) of the linear operator equation

$$\mathscr{L}f = g. \tag{1.2.1}$$

An *expansion method* introduces a complete set $\{h_i\}$, $i = 1, 2, \ldots,$ of elements in R, and expansions for the exact solution of (1.2.1) and the truncated form f_N which approximates it. Thus

$$f = \sum_{i=1}^{\infty} b_i h_i \tag{1.2.2}$$

$$f_N = \sum_{i=1}^{N} a_i^{(N)} h_i. \tag{1.2.3}$$

The method also provides an algorithm for computing the coefficients $a_i^{(N)}$ in the approximate solution. Many algorithms can be constructed; we consider here only the *unsymmetric Galerkin method* or *method of moments*, which introduces a second set of elements $\{\hat{h}_i\}$ and computes $\mathbf{a}^{(N)} = (a_i^{(N)})$ as the solution of the $N \times N$ linear system

$$\mathbf{L}\mathbf{a}^{(N)} = \mathbf{g} \tag{1.2.4}$$

where the matrix $\mathbf{L}$ and vector $\mathbf{g}$ are defined by

$$L_{ij} = (\hat{h}_i, \mathscr{L}h_j), \quad i, j = 1, \ldots, N, \tag{1.2.5}$$

$$g_i = (\hat{h}_i, g), \quad i = 1, \ldots, N. \tag{1.2.6}$$

In practice, the sets $\{\hat{h}_i\}$, $\{h_i\}$ may be related; for example, we may choose $\hat{h}_i = \mathscr{L}h_i$ (method of least squares) or $\hat{h}_i = wh_i$ (weighted Galerkin). The choice $\hat{h}_i = h_i$ yields the symmetric Galerkin technique; we shall use the generic term Galerkin method, the precise choice of $\hat{h}_i$ being evident from the context.

The use of a Galerkin technique for numerical calculations raises a number of interesting and interrelated questions:

(i) *Posing the problem*

How do we pose a given type of problem in the form (1.2.1)? In particular, for differential equations, this question usually reduces to: How do we treat the boundary conditions?

(ii) *Choosing the expansion set*
How do we choose the basis set $\{h_i\}$ in (1.2.2), (1.2.3), and the companion set $\{\hat{h}_i\}$ in (1.2.5), (1.2.6)? How does this choice affect the accuracy of the calculation? The cost of the calculation? The stability against numerical or round-off errors? The convenience?

(iii) *Computing the solution*
Setting up the matrix problem (1.2.4) usually involves providing numerical approximations to the inner products involved; the choice of approximation can crucially affect both the accuracy obtained and the time taken. For the large systems which result from multi-dimensional problems, the methods used to solve (1.2.4) are also important since the solution time can be as great as or greater than the time taken to form the equations.

(iv) *Analysing the errors*
Three classes of numerical error can be distinguished in a Galerkin calculation:

(a) *Truncation errors*
These stem from the truncation of (1.2.2) after N terms.
(b) *Discretisation errors*
These stem from the differences $b_i - a_i^{(N)}$, $i = 1, 2, \ldots, N$.
(c) *Quadrature errors*
Given that we approximate the inner products involved (using quadrature rules) we in fact solve not (1.2.4) but the perturbed system

$$(\mathbf{L} + \delta\mathbf{L})(\mathbf{a}^{(N)} + \delta\mathbf{a}^{(N)}) = \mathbf{g} + \delta\mathbf{g}, \tag{1.2.7}$$

these perturbations yielding an additional source of error, $\delta\mathbf{a}^{(N)}$.

1.2.2 An example

We provide a partial answer to questions (i) and (ii) above by means of an example. Consider the real, linear Fredholm integral equation of the second kind

$$f(x) = g(x) + \lambda \int_a^b K(x, y) f(y)\, dy. \tag{1.2.8}$$

Under suitable assumptions (on the kernel λK, and driving term g), this equation has a unique solution†. If we choose an inner product on the

† For example: if $g(x)$ is continuous in $[a, b]$, $K(x, y)$ is continuous in $[a, b] \times [a, b]$, and λ is not a characteristic value of (1.2.8), then there exists a unique continuous solution $f(x)$.

interval $[a, b]$:

$$(f_1, f_2) \equiv \int_a^b f_1(x) f_2(x) w(x)\,dx \tag{1.2.9}$$

and set

$$\hat{h}_i(x) = h_i(x), \quad i = 1, 2, \ldots \tag{1.2.10}$$

then (1.2.5), (1.2.6) takes the form

$$L_{ij} = \int_a^b w(x) h_i(x) h_j(x)\,dx - \lambda \int_a^b w(x) h_i(x) \int_a^b K(x, y) h_j(y)\,dy\,dx \tag{1.2.11}$$

$$g_i = \int_a^b w(x) h_i(x) g(x)\,dx. \tag{1.2.12}$$

Equations (1.2.4), (1.2.11), (1.2.12) are the formal defining equations for the Galerkin method applied to the problem (1.2.8). The same equations result if we choose the *unweighted inner product*

$$(f_1, f_2) = \int_a^b f_1(x) f_2(x)\,dx \tag{1.2.12a}$$

and set

$$\hat{h}_i(x) = w(x) h_i(x), \quad i = 1, 2, \ldots. \tag{1.2.12b}$$

A *numerical method* results from these defining equations once we:

(a) specify the basis set $\{h_i\}$;
(b) specify how the integrals are to be performed.

We are not concerned here with (b) (but see Chapter 12). The choice (a) lies primarily between *local* and *global* bases; we can informally describe these as follows:

Local basis
We split the region $a \leqslant x \leqslant b$ into intervals of width h. On each subinterval $[a + mh, a + (m + 1)h]$ the approximate solution $f_N(x)$ is taken to be a polynomial of fixed degree p. Continuity conditions may be imposed across subinterval boundaries (for example, a *spline basis* results if $f_N(x)$ is constrained to have $p - 1$ continuous derivatives on $[a, b]$). Convergence is obtained by letting $h \to 0$ for fixed p.

This description of $f_N(x)$ is not in terms of an expansion of the form (1.2.3), but for any choice of p and of continuity constraints, it can be put in that form. For example, if $p = 1$ (piecewise linear approximation) and $f_N(x)$ is constrained to be continuous on $[a, b]$, the description can be rephrased as follows:

Define the *hat function* $\text{Hat}_i(x)$:

$$\begin{aligned}
\text{Hat}_i(x) &= 0, & x &\leqslant a + (i-1)h \\
&= \frac{x - a - (i-1)h}{h}, & a + (i-1)h &< x \leqslant a + ih \\
&= \frac{a + (i+1)h - x}{h}, & a + ih &< x \leqslant a + (i+1)h \\
&= 0, & x &> a + (i+1)h
\end{aligned}$$

$$h = (b - a)/N. \qquad (1.2.13)$$

Thus, $\text{Hat}_i(x)$ is zero except on the interval $[a + (i-1)h, a + (i+1)h]$ and is linear over the two halves of this interval, with discontinuous first derivative at $x = a + ih$, and at $a + (i-1)h$, $a + (i+1)h$. Now choose as basis

$$h_i(x) = \text{Hat}_i(x), \quad i = 1, \ldots, N. \qquad (1.2.14)$$

Global basis

The local basis described above has the feature that each of the *basis functions* $\text{Hat}_i(x)$ depends explicitly on N, and has *local support*; that is, is zero everywhere except over a small subinterval of $[a, b]$. A *global basis* is one in which the basis functions have support over the whole region; in practice, they are also chosen to be independent of N. Within these restrictions, many choices are available. The most common choice is to take $f_N(x)$ to be a *polynomial of degree* $N - 1$ *in* x.

We can clearly achieve this in a number of ways; for example, the choice

$$h_i(x) = x^{i-1}, \quad i = 1, 2, \ldots, \qquad (1.2.15)$$

and

$$h_i(x) = T_{i-1}\left(\frac{2x - (a+b)}{b - a}\right), \quad i = 1, 2, \ldots, \qquad (1.2.16)$$

where $T_k(z)$, $-1 \leqslant z \leqslant 1$, is a Chebyshev polynomial of degree k, and each defines a polynomial approximating function $f_N(x)$; we show in Section 1.3 that these choices are in fact *equivalent* in the sense that, if quadrature and round-off errors are ignored, they will yield the *same solution* $f_N(x)$.

The choice between *local* and *global* basis is crucial; the two types lead to quite different numerical techniques both for setting up and for solving the Galerkin equations, and to quite different types of error analysis. They also perform quite differently in practice, as the following example demonstrates.

In (1.2.8), we take

$$\lambda K(x, y) = e^{xy}$$
$$g(x) = e^x - (e^{x+1} - 1)/(x + 1) \quad (1.2.17)$$
$$a = 0, \quad b = 1.$$

Then the exact solution of (1.2.8) is

$$f(x) = e^x. \quad (1.2.18)$$

Table 1.1 shows the maximum errors, $\|f - f_N\|_\infty$, obtained using expansion methods with the expansion sets (1.2.14) and (1.2.16).

Looking first at the results using a local expansion, we remark that the errors reduce only relatively slowly as N increases. The results given are in fact well fitted by the form

$$\|f - f_N\| \sim CN^{-p}, \quad p = 2,$$

and it is a simple matter to predict this value for the exponent p in advance: it is related to the continuity of the expansion set $\{\text{Hat}_i(x)\}$ used, and not at all to the problem being solved, provided that this is "smooth enough". It is more difficult to predict the amplitude C; still, it is certainly an advantage of local methods that the behaviour of the error can be predicted, and hence checked.

Looking next at the global expansion results, we see that very rapid convergence is obtained; for this problem, there is no doubt that the global method is preferable. This rapid convergence is typical of that achieved with a global polynomial basis for problems which have "smooth" solutions. Of course, not all problems are of this type. When the solution is "non-smooth", a global basis may converge no faster than a local basis, and may well be more expensive overall. Further, techniques exist for improving the performance of both types of expansion. For example, extrapolation procedures such as the "deferred approach to the limit" may be applied to increase the convergence rate of a local calculation, while we may choose to subdivide the region $[a, b]$ into two or more subintervals and apply a global expansion over each subinterval, to improve the performance of a global method. It is not the purpose of this book to argue for or against the use of global as opposed to local bases. Instead, we note the interestingly fast convergence which can be achieved, and ask the question: can this error be predicted in practice for a global method, as it can for a local method? In particular, we would like to provide error estimates for a global expansion method which are *cheaply computable* and *realistic*. It transpires that such an analysis of some generality can be given, provided that we limit attention to orthogonal bases as typified by (1.2.16); fortunately, the use of such bases,