1974

Functional Analysis Techniques in Numerical Analysis

Kenneth D. Schoenfeld
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/gradreports

Part of the Mathematics Commons

Recommended Citation

This Report is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Plan B and other Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact rebecca.nelson@usu.edu.
FUNCTIONAL ANALYSIS TECHNIQUES IN NUMERICAL ANALYSIS

by

Kenneth D. Schoenfeld

A report submitted in partial fulfillment of the requirements for the degree

of

MASTER OF SCIENCE

in

Mathematics

Approved:

UTAH STATE UNIVERSITY
Logan, Utah
1974
ACKNOWLEDGMENTS

I wish to thank those professors who contributed directly or indirectly, to the writing of this paper. In particular, I wish to thank Dr. Watson, whose comments and criticism taught me a great deal.

Kenneth D. Schoenfeld
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>I. AN OPTIMAL QUADRATURE FORMULA</td>
<td>6</td>
</tr>
<tr>
<td>II. AN OPTIMAL DIFFERENCE EQUATION</td>
<td>18</td>
</tr>
<tr>
<td>III. AN OPTIMAL FINITE DIFFERENCE METHOD</td>
<td>64</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>113</td>
</tr>
</tbody>
</table>


LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Optimal explicit formulae of Adams type and the values of $\zeta$ for the Adams-Bashforth formulae</td>
<td>57</td>
</tr>
<tr>
<td>2.2</td>
<td>Optimal implicit formulae of Adams type and the values of $\zeta$ for the Adams-Moulton formulae</td>
<td>58</td>
</tr>
</tbody>
</table>
INTRODUCTION

In this paper we will consider the problem of selecting the best, or optimal, numerical method of solution to a given mathematical problem. The admissible numerical methods will be a clearly defined set for each problem. Obviously, in order to find the best method in this set, we must have a clear mathematical formulation of just what "best" means; this will be the intent of Theorem 0.1. Intuitively, the best method will be understood to be the one which minimizes the maximum possible error where this error will be measured in terms of the norm of a given Hilbert space.

Each succeeding chapter in this paper will be a different example in finding an optimal technique. The actual examples and the numerical methods used to solve them may have little in common; however, the techniques used to prove that a certain method is optimal will have a unifying theme - each of the proofs will revolve around the use of functional analysis techniques, particularly techniques involving Hilbert space theory.

The examples will be based on sections from Numerical Processes In Differential Equations by I. Babuska, M. Prager, and E. Vitasek (these examples may be found in [1, pp. 42-47, 72-78, 170-181]). This book is an English translation of the Czech version which was published in 1964. It is also a book in which the word "obvious" is used liberally. The intent of this paper is to make some of these missing "obvious" parts of the proofs more obvious.
To begin, consider a linear functional $F$ over the Hilbert space $H$ and a space of linear functionals $M$. The functional $f \in M$ will be said to be an optimal approximation to $F$ in $M$ if for every $g \in M$ one has

$$
\sup_{\|x\|=1} |F(x) - f(x)| \leq \sup_{\|x\|=1} |F(x) - g(x)|.
$$

In order to put the above definition in proper perspective, the following example might be kept in mind. Let $H$ be a Hilbert space of integrable functions and let $F \in H^*$ be the integral operator

$$
F(x) = \int_a^b x \, d\sigma \quad \text{for } x \in H.
$$

As our approximating subspace, choose the set of all linear functionals of the form

$$
g(x) = \sum_{i=1}^n c_i x(\xi_i), \quad a \leq \xi_i \leq b.
$$

Hence, we are considering the problem of approximating the integral by a quadrature formula (The details of this example will be the subject of Chapter 1). The above definition states that the optimal approximation to $F$ in $M$, i.e., the "best" choice of the $c_i$'s, is the quadrature formula which minimizes the maximum possible error where this error is measured in the Hilbert space norm.

Another characterization of this definition of optimality will be given in the following theorem. Denote by $z_h$ the element of $H$ which by the Riesz Representation Theorem (see [7, p. 213]) gives the functional $h$, that is, $h(x) = (z_h, x)$ for every $x \in H$ where $(z_h, x)$ denotes the inner product in the Hilbert space $H$. 
Theorem 0.1. If $F$ is a linear functional and $M$ a closed subspace of the space $H^*$ of all linear functionals over the Hilbert space $H$, then there exists just one functional $f \in M$ which is the optimal approximation to $F$ in $M$ and $f \in M$ is the optimal approximation of $F$ in $M$ if and only if $g(z_{F-f}) = 0$ for every $g \in M$.

Proof: The Riesz Theorem states, among other things, that

$$\|z_{F-g}\| = \sup_{\|x\|=1} |(F-g)(x)|, \quad \text{for all } g \in H.$$ 

Therefore it is clear that the following two inequalities are equivalent:

$$\sup_{\|x\|=1} |(F-f)(x)| \leq \sup_{\|x\|=1} |(F-g)(x)|, \quad \text{for all } g \in M,$$

and

$$\|z_{F-f}\| \leq \|z_{F-g}\|, \quad \text{for all } g \in M. \quad (1)$$

Also from the Riesz Theorem, it follows that the next two statements are equivalent:

$$g(z_{F-f}) = 0, \quad \text{for all } g \in M,$$

and

$$(z_g, z_{F-f}) = 0, \quad \text{for all } g \in M. \quad (2)$$

Hence, if we can show that (1) implies (2), (2) implies (1), and that $z_f$ is unique, we will have completed the proof.

Part I. (1) implies (2).

Let $Z_M = \{z_h; h \in M\}$ and let $\{\phi_a; a \in A\}$ denote an orthonormal basis for $Z_M$. Extend $\{\phi_a; a \in A\}$ to a basis for $H$ and denote this basis by $\{\phi_b; b \in B\}$. Let $F \in H^*$, then $z_F = \sum_{b \in B} (z_F, \phi_b) \phi_b$. The element of

$$z_{F-f} = \sum_{b \in B} (z_{F-f}, \phi_b) \phi_b.$$
of $z_F$ onto $Z_M$. Consequently, $z_{F-f} = z_F - z_f = \sum_{b \in B} (z_F, \phi_b) \phi_b$. Let $h \in M$, then $(z_h, z_{F-f}) = (z_h, \sum_{b \in B} (z_F, \phi_b) \phi_b) = \sum_{b \in B} (z_F, \phi_b)(z_h, \phi_b)$ and since $z_h \in Z_M$, $z_h \perp \phi_b$ for all $b \notin A$, and consequently $(z_h, z_{F-f}) = 0$ for all $h \in M$.

Part II. (2) implies (1).

Let $f \in M$ such that $(z_h, z_{F-f}) = 0$ for all $h \in M$. Then $(z_h, z_{F-f}) = \sum_{b \in B} (z_{F-f}, \phi_b)(z_h, \phi_b) = 0$ for every $h \in M$, and since $h$ was arbitrary, $z_{F-f} \perp \phi_b$ for all $b \in A$ and $z_{F-f} = \sum_{b \in B} (z_{F-f}, \phi_b) \phi_b$.

Hence $z_{F-f} = \sum_{b \in B} (z_F - z_f, \phi_b) \phi_b = \sum_{b \in B} ((z_F, \phi_b) - (z_f, \phi_b)) \phi_b$, and since $z_f \in Z_M$, this gives $z_{F-f} = \sum_{b \in B} (z_F, \phi_b) \phi_b$. Now $z_F = \sum_{b \in B} (z_F, \phi_b) \phi_b$, and $z_{F-f} = z_F - z_f$, which implies that $z_f = \sum_{a \in A} (z_F, \phi_a) \phi_a$. Consequently $z_f$ is the orthogonal projection of $z_F$ onto $Z_M$ and $z_f$ clearly satisfies (1).

Part III. Uniqueness.

The uniqueness of the element $z_f$ can be proven as follows. Suppose that $f$ and $g$ are both elements of $M$ such that $(z_{F-g}, z_h) = (z_{F-f}, z_h) = 0$ for every $h \in M$. Thus $(f - g), (g - f) \in M$ which implies that
(z_{F-g}, z_{f-g}) = (z_{F-f}, z_{g-f}) = 0. By the Pythagorean Theorem

\[ \| z_F - z_f \|^2 = \| z_f - z_g \|^2 + \| z_F - z_g \|^2 \]

\[ = \| z_f - z_g \|^2 + \| z_F - z_f \|^2 + \| z_f - z_g \|^2 \]

and consequently \( \| z_f - z_g \|^2 = 0 \) so that \( z_f = z_g \).
I. AN OPTIMAL QUADRATURE FORMULA

For our first example, we will consider the problem of the numerical evaluation of the integral

\[ I_p = \frac{1}{2\pi} \int_0^{2\pi} f(x)e^{ipx} \, dx, \quad i = \sqrt{-1} \]

One possible approach to this problem would be to approximate \( f \) by some simple curve and then try to integrate exactly. Another possibility would be to apply a standard quadrature formula, such as Simpson's rule, to the integral.

It is our intention to start with a given quadrature formula and then to obtain the coefficients which make this formula optimal in some Hilbert space.

To proceed, let \( H_{\alpha}, \ell \geq 1 \) be the Hilbert space of classes of \( 2\pi \)-periodic, complex-valued functions \( f(x), -\infty < x < \infty \) which do not differ by more than a constant and the \( \ell \)-th derivatives of which are square integrable. The scalar product for \( H_{\alpha} \) will be given by

\[ (f, g) = \int_0^{2\pi} \frac{d^\ell f}{dx^\ell} \cdot \frac{d^\ell g}{dx^\ell} \]

Note that this scalar product partitions the \( 2\pi \)-periodic functions into equivalence classes of functions which differ by a constant since only the constant polynomial is periodic.

Define

\[ F_p(f) = \frac{1}{2\pi} \int_0^{2\pi} e^{ipx} f(x) \, dx, \quad \text{where } p \text{ is an integer}, \quad (1.1) \]
and let the approximating space, $M_n$, of linear functionals be the space of all functionals of the form

$$\phi(f) = \sum_{j=1}^{n} c_j f(2\pi j/n), \quad \sum_{j=1}^{n} c_j = 0. \quad (1.2)$$

Observe that $M_n$ is a subspace of $H_{\ell}^*$ since it follows from (1.2) that for $f, g \in M_n$, $\phi(af + \beta g) = a\phi(f) + \beta\phi(g)$, where $\alpha$ and $\beta$ are scalars, and if $f_1, f_2$ are in the same equivalence class, i.e., if $f_1$ and $f_2$ differ by the same constant at every point, then $\phi(f_1 - f_2) = \phi(k) = k \sum_{j=1}^{n} c_j = 0$, where $k$ is some constant.

We will now prove

Theorem 1.1. The functional

$$\psi_{p,n}(f) = \frac{c(p,n,\ell)}{n} \sum_{j=1}^{n} e^{i p(2\pi j/n)} f(2\pi j/n), \quad (1.3)$$

for $n > p$, is the optimal approximation to the functional $F_p(f)$ in $M_n$, where

$$c(p,n,\ell) = \frac{1}{\zeta(\ell,p/n) + (p/(n-p))^{2\ell} \zeta(\ell,(n-p)/n)} \quad (1.4)$$

$$\zeta(\ell,\alpha) = \sum_{t=0}^{\infty} \frac{1}{(1+t/\alpha)^{2\ell}}, \quad 0 < \alpha < \ell. \quad (1.5)$$

Proof: Part I. We will first show that $F_p - \psi_{p,n} \in H_{\ell}^*$ is the dual of the function

$$g(x) = \frac{e^{-ipx}}{2\pi p^{2\ell}} - c(p,n,\ell) \sum_{t=-\infty}^{\infty} \frac{e^{-i(tn-p)x}}{(tn-p)^{2\ell}}, \quad (1.6)$$
i.e., that \( F_p(f) - \psi_{p,n}(f) = (f,g) \) for all \( f \in H \). Observe that \( g \) does determine an element of \( H_k \) since \( g(x) \) is \( 2\pi \)-periodic and because we will show later that

\[
|g|^2_{H_k} = \frac{1}{2\pi} \left[ 1 - c(p,n) \right]^{1/2}.
\]

Also note that since \( \left\{ e^{ikx}; -\infty < k < \infty \right\} \) forms a complete orthogonal set in \( H_k \), any \( f \in H_k \) can be expressed in the form

\[
f(x) = \sum_{k=-\infty}^{\infty} a_k e^{ikx}
\]

where \( a_k \) is the generalized Fourier coefficient \( a_k = (f,e^{ikx}) \).

where \((,\,\,\,)\) denotes the inner product in \( H_k \). For \( L_2 \) we would have

\[
(f,f)_{L_2} = \sum_{k=-\infty}^{\infty} |a_k|^2 \quad \text{however, for } H_k
\]

\[
\int_0^{2\pi} \frac{df}{dx} \cdot \frac{df}{dx} \, dx = \int_0^{2\pi} \sum_{k=-\infty}^{\infty} (ik)^{\ell} a_k e^{ikx} \sum_{t=-\infty}^{\infty} (-it)^{\ell} a_t e^{-itx} \, dx
\]

\[
= \sum_{k=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} (ik)^{\ell} (-it)^{\ell} a_k a_t \int_0^{2\pi} e^{ikx} e^{-itx} \, dx,
\]

and since we get nonzero terms adding to the double summation only if \( t=k \),

\[
(f,f)_{\ell} = \sum_{k=-\infty}^{\infty} |a_k|^2 \, (k)^{2\ell} 2\pi < \infty.
\]

We will now proceed to show that \( F_p(f) - \psi_{p,n}(f) = (f,g) \).

\[
(f,g) = \int_0^{2\pi} \left\{ \sum_{k=-\infty}^{\infty} (ik)^{\ell} a_k e^{ikx} \right\} \left\{ \frac{(ip) e^{ipx}}{2\pi p^{\ell}} \right\} \, dx
\]
\[- \frac{c(p,n,\ell)}{2\pi} \sum_{t=-\infty}^{\infty} \left( -i \right)^{\ell} (t_n - p)^{-\ell} e^{-i(t_n-p)x} \left( -1 \right)^{\ell} \sum_{k=-\infty}^{\infty} a_k \int_{0}^{2\pi} \frac{e^{i(p+k)x}}{p^\ell} \, dx \]

\[= \int_{0}^{2\pi} \left( \frac{c(p,n,\ell)}{2\pi} \right) \sum_{k=-\infty}^{\infty} a_k \int_{0}^{2\pi} \frac{e^{i(p+k)x}}{p^\ell} \, dx \]

\[= \sum_{k=-\infty}^{\infty} \left( -1 \right)^{\ell} a_k \int_{0}^{2\pi} \frac{e^{i(p+k)x}}{p^\ell} \, dx \]

\[- \frac{c(p,n,\ell)}{2\pi} \sum_{t=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_k \frac{e^{i(k-(t_n-p))x}}{(t_n-p)^\ell} \int_{0}^{2\pi} \frac{e^{i(p+k)x}}{p^\ell} \, dx \]

and since \( \int_{0}^{2\pi} e^{i(p+k)x} \, dx = 0 \) for \( p + k \neq 0 \) and

\[\int_{0}^{2\pi} e^{i(k-(t_n-p))x} \, dx = 0 \] for \( k - (t_n - p) \neq 0 \), the only nonzero contribution of these two integrals to the above summations is when \( k = -p \) and \( k = (t_n - p) \) respectively. Therefore,

\[(f,g) = \frac{(-1)^{\ell} (-p)^{\ell} a_{-p}}{2\pi p^\ell} \int_{0}^{2\pi} 1 \, dx \]

\[- \frac{c(p,n,\ell)}{2\pi} \sum_{t=-\infty}^{\infty} a_{tn-p} (tn-p)^{\ell} (tn-p)^{-\ell} \int_{0}^{2\pi} 1 \, dx, \]
which gives

\[(f,g) = a_{p} - c(p,n,ξ) \sum_{t=-\infty}^{\infty} a_{tn-p}. \tag{1.6}\]

We will now show that \(F_{p}(f) - \psi_{p,n}(f)\) is equal to the right hand side of (1.6).

\[
F_{p}(f) = \frac{1}{2\pi} \int_{0}^{2\pi} e^{ipx} f(x) = \frac{1}{2\pi} \int_{0}^{2\pi} e^{ipx} \sum_{k=-\infty}^{\infty} a_{k} e^{ikx}
= \frac{1}{2\pi} \left[ \sum_{k=-\infty}^{\infty} a_{k} e^{i(p+k)x} \right] = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} a_{k} \int_{0}^{2\pi} e^{i(p+k)x}.
\]

Using the same orthogonality argument as above, we conclude that

\[
F_{p}(f) = \frac{1}{2\pi} (a_{-p}) \int_{0}^{2\pi} 1 = a_{-p} \tag{1.7}
\]

Now,

\[
\psi_{p,n}(f) = \frac{c(p,n,ξ)}{n} \sum_{j=1}^{n} e^{ip(2\pi j/n)} f(2\pi j/n)
= \frac{c(p,n,ξ)}{n} \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} a_{k} e^{ik(2\pi j/n)} e^{ip(2\pi j/n)}
= \frac{c(p,n,ξ)}{n} \sum_{k=-\infty}^{\infty} a_{k} \sum_{j=1}^{n} (e^{i(k+p)(2\pi j/n)}). \]
When \( k+p \neq nt \), where \( t \) is an integer, the geometric sequence

\[
\sum_{j=1}^{n} \left( e^{i(k+p)(2\pi j/n)} \right)
\]

is equal to

\[
\frac{(1 - e^{(i(k+p)2\pi/n)n}) e^{(i(k+p)2\pi/n)}}{1 - e^{(i(k+p)2\pi/n)}} = 0
\]

so that nonzero terms in the double summation occur only when \( k+p \) is a multiple of \( n \), say \( n \) times \( t \), where \( n \) and \( t \) are integers. When \( k+p = nt \),

\[
\sum_{j=1}^{n} e^{i(k+p)j(2\pi/n)} = \sum_{j=1}^{n} e^{int(2\pi j/n)} = n, \text{ so}
\]

\[
\psi_{p,n}(f) = \frac{c(p,n,\xi)}{n} \sum_{k=-\infty}^{\infty} n a_k = c(p,n,\xi) \sum_{t=-\infty}^{\infty} a_{nt-p}. \quad (1.8)
\]

(Since \( k+p = nt \) the substitution \( k = nt-p \) was used in the last equality).

From (1.6), (1.7), and (1.8), it follows that

\[
F_p(f) - \psi_{p,n}(f) = (f,g), \text{ for } f \in H,
\]

and we are finished with part I of Theorem 1.1.

Part II. In part I we have shown that \((F_p - \psi_{p,n})(f) = (f,g)\).

Therefore, if we can show that \( \phi(g) = 0 \) for every \( \phi \in M_n \), we can use

Theorem 0.1 to conclude that \( \psi_{p,n} \) is the optimal approximation to \( F_p \)

in \( M_n \). We will show that \( g(2\pi j/n) = 0 \) for \( j = 1, 2, \ldots, n \) and then

from the definition of \( \phi \), (1.2), it will follow that \( \phi(g) = 0 \)

for all \( \phi \in M_n \).
Now,

\[ g(2\pi j/n) = \frac{e^{-ip(2\pi j/n)}}{2\pi p^{2\ell}} - \frac{c(p,n,\ell)}{2\pi} \sum_{t=-\infty}^{\infty} \frac{e^{i(tn-p)(2\pi j/n)}}{(tn-p)^{2\ell}} \]

\[ = \frac{e^{-ip(2\pi j/n)}}{2\pi p^{2\ell}} - \frac{c(p,n,\ell)}{2\pi} \sum_{t=-\infty}^{\infty} \frac{e^{it2\pi j} - ip2\pi j/n}{(tn-p)^{2\ell}} \]

\[ = \frac{e^{-ip2\pi j/n}}{2\pi} \left\{ \frac{1}{p^{2\ell}} - c(p,n,\ell) \sum_{t=-\infty}^{\infty} \frac{1}{(tn-p)^{2\ell}} \right\} \]

\[ = \frac{e^{-ip2\pi j/n}}{2\pi p^{2\ell}} \left\{ 1 - c(p,n,\ell) \sum_{t=-\infty}^{\infty} \frac{1}{{(tn/p) - 1}}^{2\ell} \right\}. \]

Thus,

\[ g(2\pi j/n) = \frac{e^{-ip2\pi j/n}}{p^{2\ell}2\pi} \left\{ 1 - c(p,n,\ell) \sum_{t=-\infty}^{\infty} \frac{1}{{(1 - (tn/p)^{2\ell})}} \right\}. \tag{1.9} \]

Therefore, if we can show that

\[ \sum_{t=-\infty}^{\infty} \frac{1}{{(1 - (tn/p))^{2\ell}}} = \frac{1}{c(p,n,\ell)}, \]

then \( g(2\pi j/n) = 0 \), for \( j = 0,1,...,n \). To this end, observe that
\[
\sum_{t=-\infty}^{\infty} \frac{1}{1-(tn/p)^{2\ell}} = \sum_{t=0}^{\infty} \frac{1}{1-(tn/p)^{2\ell}} + \sum_{t=1}^{\infty} \frac{1}{1-(tn/p)^{2\ell}}
\]

\[
= \sum_{t=0}^{\infty} \frac{1}{1-(tn/p)^{2\ell}} + \sum_{k=0}^{\infty} \frac{1}{1-((k+1)n/p)^{2\ell}}
\]

\[
= \zeta(\ell, p/n) + \sum_{k=0}^{\infty} \frac{1}{n-p} \left\{ \frac{1}{n} \left( \frac{n+k+n}{n-p} \right)^{2\ell} \right\}
\]

\[
= \zeta(\ell, p/n) + \frac{1}{(n-p)/n} \left\{ 1 + \frac{k}{n-p} \right\}^{2\ell}
\]

so

\[
\sum_{t=-\infty}^{\infty} \frac{1}{1-(tn/p)^{2\ell}} = \zeta(\ell, p/n) + \frac{1}{(n-p)/n} \left\{ 1 + \frac{k}{n-p} \right\}^{2\ell}
\]

From (1.4), we have

\[
\frac{1}{c(p,n,\ell)} = \zeta(\ell, p/n) + \frac{(p/(n-p))^{2\ell}}{\zeta(\ell, \frac{n-p}{n})}
\]

so that from (1.10)

\[
\sum_{t=-\infty}^{\infty} \frac{1}{1-(tn/p)^{2\ell}} = 1/c(p,n,\ell).
\]

Using this identity in (1.9), it follows that
\[ g(2\pi j/n) = e^{\frac{-ip2\pi j/n}{2\pi p^{2}} \frac{2\ell}{c(p,n,\xi)}} \left\{ 1 - \frac{c(p,n,\xi)}{c(p,n,\xi)} \right\} = 0 \]

and we are finished with part II of Theorem 1.1.

Formula (1.3) gives the optimal approximation in the space \( M_n \) to the integral \( \frac{1}{2\pi} \int_0^{2\pi} e^{ipx} f(x) \, dx \). Obviously, if we were actually to use this formula, we must choose a value for the parameter \( \lambda \).

It would be advantageous if we could make the best choice, i.e., if we could make \( |\mathcal{P}_p(f) - \psi_{p,n}(f)| \) a minimum with respect to \( \lambda \).

In the following calculations we will assume that \( f(x) \) is analytic on the set \( E[x+iy; |y| \leq R] \). Hence, from the Cauchy Integral Formula,

\[ f(\lambda)(z) = \frac{\lambda!}{2\pi i} \int_C \frac{f(z)}{(z - z_0)^{\lambda+1}} \, dz \]

where \( C \) is a circle of radius \( R \) centered at a point on the \( x \) axis, \( z_0 = x_0 \). Thus

\[ |f(\lambda)(z)| \leq \frac{\lambda!}{2\pi i} \int_C \left| \frac{f(z)}{R^{\lambda+1}} \right| \, |dz| = \frac{\lambda!M2\pi R^{\lambda+1}}{2\pi R^{\lambda+1}} = \frac{\lambda!M}{R^{\lambda}} \]

Since our functions are \( 2\pi \)-periodic, we need only consider the set \( D[x+iy; 0 \leq x \leq 2\pi, |y| \leq R] \), and on this compact set there exists an \( M \) such that \( |f(z)| \leq M \) for all \( z \in D \). Therefore,

\[ |f(\lambda)(z)| \leq \frac{\lambda!M}{2\pi R^{\lambda+1}} \int_C |dz| = \frac{\lambda!M2\pi R^{\lambda+1}}{2\pi R^{\lambda+1}} = \frac{\lambda!M}{R^{\lambda}} \]
from which follows,

\[
(f, f) = \int_0^{2\pi} f(\xi) \overline{f(\xi)} \, dx \leq \int_0^{2\pi} \left| \frac{\ell M}{R^\ell} \right|^2 = \left( \frac{\ell M}{R^\ell} \right)^2 \quad \forall \ell \geq 2
\]

and

\[
\|f\| \leq \frac{M !}{R^\ell} \sqrt{2\pi} = \frac{\ell !}{R^\ell} \quad a
\]

where \( a = M \sqrt{2\pi} \).

Appealing once again to the Riesz Representation Theorem, we have

\[
\sup_{\|f\|=1} \left| F_p(f) - \psi_p, n(f) \right| = \|g\|
\]

where \( g \) is defined by

\[
g(x) = \frac{e^{ipx}}{2\pi p} \left( \frac{2\pi}{2^\ell} \right) - \frac{c(p, n, \ell)}{2\pi} \sum_{t=-\infty}^{\infty} \frac{e^{i(tn-p)x}}{(tn-p)^{2\ell}}
\]

Now,

\[
(g, g) = \int_0^{2\pi} \left( \frac{d^\ell g}{dx^\ell} \right)^2 \, dx
\]

\[
= \int_0^{2\pi} \left\{ \left( -ip \right)^\ell e^{-ipx} \frac{2\pi}{p^{2\ell} 2\pi} - \frac{c(p, n, \ell)}{2\pi} \sum_{t=-\infty}^{\infty} \frac{e^{i(tn-p)x}}{(tn-p)^{2\ell}} \right\}
\]

\[
\cdot \left\{ \frac{(ip)^\ell e^{ipx}}{2\pi p^{2\ell}} - \frac{c(p, n, \ell)}{2\pi} \sum_{k=-\infty}^{\infty} \frac{e^{-i(kn-p)x}}{(kn-p)^{2\ell}} \right\} \, dx
\]
\[
\begin{align*}
&= \int_0^{2\pi} \frac{(-ip)^{\ell} (ip)^{\ell}}{(p^{2\ell} 2\pi)^{2}} + \frac{c^2(p, n, \ell)}{(2\pi)^2} \sum_{t=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \frac{e^{i(t-k)nx} (i(tn-p))^{\ell} (-i(kn-p))^{\ell}}{(tn-p)^{2\ell} (kn-p)^{2\ell}} \\
&- \frac{c(p, n, \ell)(-ip)^{\ell}}{(2\pi)^2 p^{2\ell}} \sum_{k=-\infty}^{\infty} \frac{e^{-iknx} (-i(kn-p))^{\ell}}{(kn-p)^{2\ell}} \\
&- \frac{c(p, n, \ell)(ip)^{\ell}}{(2\pi)^2 p^{2\ell}} \sum_{t=-\infty}^{\infty} \frac{e^{itnx} (i(tn-p))^{\ell}}{(tn-p)^{2\ell} (kn-p)^{2\ell}} \\
&= \int_0^{2\pi} \frac{1}{(2\pi)^2 p^{2\ell}} + \frac{c^2(p, n, \ell)}{(2\pi)^2} \sum_{t=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \int_0^{2\pi} \frac{e^{i(t-k)nx} (i(tn-p))^{\ell}}{(tn-p)^{2\ell} (kn-p)^{2\ell}} dx \\
&- \frac{c(p, n, \ell)(-i)^{\ell}}{(2\pi)^2 p^{2\ell}} \sum_{k=-\infty}^{\infty} \int_0^{2\pi} \frac{e^{-iknx} (-i)^{\ell}}{(kn-p)^{2\ell}} dx \\
&- \frac{c(p, n, \ell)(i)^{\ell}}{(2\pi)^2 p^{2\ell}} \sum_{t=-\infty}^{\infty} \int_0^{2\pi} \frac{e^{itnx} (-i)^{\ell}}{(tn-p)^{2\ell}} dx.
\end{align*}
\]

Appealing again to the fact that \( \int_0^{2\pi} e^{ikx} dx = 0 \) for \( k \neq 0 \),

\[
(g, g) = \frac{1}{2\pi p^{2\ell} 2\pi} + \frac{c^2(p, n, \ell)}{(2\pi)^2} \int_0^{2\pi} \sum_{t=-\infty}^{\infty} \frac{1}{(tn-p)^{2\ell}} dx
\]

\[
- \frac{c(p, n, \ell)(-i)^{\ell}}{(2\pi)^2 p^{2\ell}} \left\{ \frac{(-i)^{\ell} 2\pi}{(-p)^{\ell}} \right\} - \frac{c(p, n, \ell)(i)^{\ell}}{(2\pi)^2 p^{2\ell}} \left\{ \frac{(i)^{\ell} 2\pi}{(-p)^{\ell}} \right\}
\]
and since
\[ \sum_{t=-\infty}^{\infty} \frac{1}{(tn-p)^{2\ell}} = \frac{1}{p^{2\ell} c(p,n,\ell)}, \]

\[ (g,g) = \frac{1}{2\pi p^{2\ell}} + \frac{c(p,n,\ell)}{2\pi p^{2\ell}} - \frac{2c(p,n,\ell)}{2\pi p^{2\ell}} \]

\[ = \frac{1}{2\pi p^{2\ell}} \left\{ 1 - c(p,n,\ell) \right\}, \text{ and} \]

\[ \left\| g \right\|_{H^\ell} = \left\| F_p - \psi_{p,n} \right\|_{H^\ell}^* = \frac{1}{\sqrt{2\pi p^{\ell}}} \left\{ 1 - c(p,n,\ell) \right\}^{1/2}. \]

Noting that \( |F_p(f) - \psi_{p,n}(f)| \leq \left\| F_p - \psi_{p,n} \right\|_{H^\ell}^* \left\| f \right\|_{H^\ell}, \)

it then follows from (1.11) that
\[ |F_p(f) - \psi_{p,n}(f)| \leq \frac{a\ell!}{2\pi p^{\ell}} Q(\ell,p,n) \]

where \( Q(\ell,p,n) = \frac{1}{\sqrt{2\pi}} \left\{ 1 - c(p,n,\ell) \right\}^{1/2}. \)

Consequently, inequality (1.12) gives a bound on the error between
\( F_p(f) \) and \( \psi_{p,n}(f) \) which may be used to minimize this error with respect to \( \ell. \)
II. AN OPTIMAL DIFFERENCE EQUATION

In this chapter, we will be considering difference equations and their application to the problem of finding approximate solutions to initial value problems. The differential equations will be of the form

$$y' = f(x, y), \quad y(x_0) = y_0.$$  \hspace{1cm} (2.1)

The general difference formula to be used to solve equations of the above form is given by

$$\sum_{n=0}^{k} \alpha_n y_n = h \sum_{n=0}^{k} \beta_n f(x_n, y_n),$$  \hspace{1cm} (2.2)

where $\alpha_n, \beta_n$ are given constants with $\alpha_k \neq 0$.

The initial conditions are

$$y_n = y_0, \quad n = 0, 1, \ldots, k-1,$$

and $h$ is the uniform step size over the interval $(x_0, x_0 + kh)$. Here, and in the rest of this chapter, $y_n$ will be used to denote a solution to the difference equation while $y(x_0 + nh)$ will denote the value of the exact solution to the differential equation at the same point. Thus $y_n$ predicts the value of $y(x_0 + nh)$. In many situations, $y_n$ will not be exactly equal to $y(x_0 + nh)$, but the value of $y_n$ should be close to that of $y(x_0 + nh)$.

Two types of difference formulae which frequently occur are implicit and explicit formulae. The characteristic which distinguishes between
these two types is the value of $\beta_k$. If $\beta_k \neq 0$, (2.2) is said to be an implicit formula of order $k$. If $\beta_k = 0$, (2.2) is said to be an explicit formula or order $k$. Note that if $\beta_k = 0$, then $y_k$ has the coefficient $\alpha_k \neq 0$, and $y_k$ can be solved for explicitly; in the case $\beta_k \neq 0$, $y_k$ is involved in $f(x_k, y_k)$ and is therefore implicitly defined. Thus, an explicit difference formula of order $k$ is one which can be represented in the form

$$\sum_{n=0}^{k} \alpha_n y_n = h \sum_{n=0}^{k-1} \beta_n f(x_n, y_n), \quad \alpha_k \neq 0,$$

while an implicit formula of order $k$ must be represented in the form

$$\sum_{n=0}^{k} \alpha_n y_n = h \sum_{n=0}^{k} \beta_n f(x_n, y_n), \quad \beta_k, \alpha_k \neq 0.$$

The difference formula (2.2) (both implicit and explicit variations) is said to be of degree $p \geq 0$ if it fulfills the $p+1$ conditions

$$\sum_{n=0}^{k} \alpha_n = 0, \quad \sum_{n=0}^{k} \frac{s \alpha}{s!} = \sum_{n=0}^{s-1} \frac{\beta}{(s-1)! n},$$

for $s = 1, 2, \ldots, p$. \hspace{1cm} (2.3)

This definition is motivated by the fact that the conditions in (2.3) imply (at least for functions for which the Taylor expansion is valid) that the difference
\[
\sum_{n=0}^{k} \alpha_n y(x_0 + nh) - h \sum_{n=0}^{k} \beta_n y'(x_0 + nh)
\]

is of order \(h^{p+1}\) so that the solution to the difference formula (2.3) approximates the exact solution to the differential equation \(y' = f(x,y)\) locally with an accuracy of order \(h^{p+1}\). To see this, merely expand \(y(x_0 + nh)\) and \(y'(x_0 + nh)\) about \(x_0\) to get

\[
\sum_{n=0}^{k} \alpha_n y(x_0 + nh) = \sum_{n=0}^{k} \alpha_n \sum_{s=0}^{\infty} \frac{(nh)^s y(s)(x_0)}{s!}
\]

And

\[
h \sum_{n=0}^{k} \beta_n y'(x_0 + nh) = \sum_{n=0}^{k} \beta_n \sum_{s=1}^{\infty} \frac{(nh)^{s-1} y(s)(x_0)}{(s-1)!}
\]

Hence

\[
\sum_{n=0}^{k} \alpha_n y(x_0 + nh) - h \sum_{n=0}^{k} \beta_n y'(x_0 + nh)
\]

\[= y(x_0) \sum_{n=0}^{k} \alpha_n + \sum_{n=0}^{k} \sum_{s=1}^{\infty} h^s y(s)(x_0) \left\{ \frac{\alpha_n s}{s!} - \frac{n^{s-1} \beta_n}{(s-1)!} \right\}.
\]

Thus, if the \(\alpha's\) and \(\beta's\) satisfy (2.3),

\[
\sum_{n=0}^{k} \alpha_n y(x_0 + nh) - h \sum_{n=0}^{k} \beta_n y'(x_0 + nh)
\]

\[= \sum_{n=0}^{\infty} \sum_{s=p+1}^{\infty} h^s y(s)(x_0) \left\{ \frac{\alpha_n s}{s!} - \frac{n^{s-1} \beta_n}{(s-1)!} \right\} = 0(h^{p+1}).
\]
Observe the difference in our use of the words "order" and "degree." Our usage is not universal, but in this paper, the order of a difference equation indicates the upper summation index while the degree of a difference equation indicates the ease with which the solution to the difference equation converges (as a function of $h$) to the exact solution of the differential equation.

We will now give two examples which will illustrate the application of implicit and explicit difference formulae to the solution of differential equations.

**Example 1.** Consider solving the differential equation

$$y' = f(x,y), \quad y(0) = y_0 = 2,$$

by means of the Adams-Bashforth formula

$$y_{n+2} = y_{n+1} + h \left( \frac{3}{2} y'_{n+1} - \frac{1}{2} y'_n \right).$$

This is an explicit formula of order 2 since it can be expressed in the form

$$\sum_{n=0}^{2} \alpha_n y_n = h \sum_{n=0}^{1} \beta_n y'_n$$

where $\alpha_0 = 0$, $\alpha_1 = -1$, $\alpha_2 = 1$, $\beta_0 = -1/2$, and $\beta_1 = 3/2$.

It is also a formula of degree 2 since (2.3) gives us that

$$\sum_{n=0}^{2} \alpha_n = 0.$$
(-1) + 2(1) = -1/2 + 3/2 \quad (s = 1), \text{ and} \\
\frac{(-1) + 2^2(1)}{2!} = \frac{3/2}{2!} \quad (s = 2), \text{ but} \\
\frac{(-1) + 2^3(1)}{3!} \neq \frac{3/2}{2!} \quad (s = 3).

Note that \( y_0 \) and \( y_1 \) must be determined before the Adams-Bashforth formula can be used. \( y_0 \) is given by the initial condition \( y(0) = 2 \), but in order to determine \( y_1 \), another difference formula, such as the Euler formula

\[
\frac{y_1 - y_0}{h} = y'_0
\]

must be applied to (2.4). However, once the initial conditions

\[
y_n = y_{0,n'}, \quad n = 0,1,
\]

have been determined, the Adams-Bashforth formula can be used repeatedly to generate all values \( y_n \).

**Example 2.** Consider solving the differential equation (2.4) by means of the difference formula

\[
y_{n+2} = y_{n+1} + h \{ (9/16)y'_{n+2} + (5/8)y'_{n+1} - (1/16)y'_n \}.
\]

Using the same techniques we used in the previous example, we see that this is an implicit formula of degree 2 and order 2. This formula is said to be of the Adams-Moulton type.

After obtaining the initial conditions

\[
y_n = y_{0,n'}, \quad n = 0,1,
\]

by the same methods used in example 1, we are then able to use the Adams-Moulton formula as an iterative procedure in determining the value of \( y_k \).
Notice that an implicit formula does not predict new $y_k$. The importance of an implicit formula lies in the fact that the stepping procedure used to generate the $y_n$'s generally gives a more accurate approximation to the solution of the differential equation than the explicit method of the same degree and order. In addition, implicit methods are stable under a wider range of conditions. Thus, one possible method for obtaining more accurate answers to the initial value problem (2.4) would be to use the Adams-Bashforth formula to predict solutions to the differential equation at new points and then use the Adams-Moulton formula to increase the accuracy of these predictions.

The remainder of this chapter will be devoted to constructing optimal difference formulae (optimal in some specified Hilbert space).

To begin, let $k, \ell \geq 2$ be integers and let $h > 0$. Denote by $H_{k,h}^{\ell}$ the Hilbert space of the classes of real-valued functions $f(x)$, $0 \leq x \leq kh$, which differ at most by a polynomial of degree $\ell-1$ with derivatives of order $\ell$ in the $L^2$ sense on the interval $[0, kh]$ and with the operation of scalar product

$$ (f,g)_\ell = \int_0^{kh} \frac{d^\ell f}{dx^\ell} \cdot \frac{d^\ell g}{dx^\ell} \, dx. $$

A few comments are now in order. First, it is clear that this inner product will not distinguish between polynomials of degree $\ell-1$. Next, the $\ell$'th derivatives of $f(x)$ need not be classical derivatives but rather $L^2$ derivatives. For example, the function $xH(x)$, where $H(x)$ is the heaviside function defined by
is not differentiable in the classical sense but is in the $L^2$ sense, i.e., $xH(x) = \int_0^{kh} H(x) \, dx$ so that $\frac{d}{dx} xH(x) = H(x)$ almost everywhere. Further $H(x)$ could be viewed as the distributional derivative of $xH(x)$ (just as are classical derivatives, $\frac{d}{dx} xH(x) = H(x) + x\delta(x) = H(x)$). However, for $f^{(\ell-1)}(x)$ to have a strictly distributional derivative, such as $\delta(x)$, would require that the integral of $\delta^2(x)$ exist. Therefore the $\ell$'th derivatives cannot be strictly distributional derivatives, but must also be derivatives in the $L^2$ or classical sense.

Let $F(f)$ be the linear functional on $H_{k,h}^n$ defined by

$$F(f) = f(kh) + \sum_{j=0}^{k-1} \alpha_j f(jh) - h \sum_{j=0}^{k} \beta_j f'(jh)$$

(2.5)

where

$$\sum_{j=0}^{k-1} \alpha_j = -1, \quad k^s + \sum_{j=0}^{k-1} \alpha_j j^s = \sum_{j=0}^{k} s\beta_j j^{s-1},$$

(2.6)

$s = 1, \ldots, \ell-1$. Observe that (2.5) implies that $\alpha_k = 1$.

Next, we will prove that if (2.5) defines a linear functional on $H_{k,h}^n$, then (2.6) must hold. At the conclusion of the proof, it should be evident that (2.6) also determines sufficient conditions for (2.5) to define a linear functional on $H_{k,h}^n$. 

Let $F(f)$ be the linear functional on $H_{k,h}^n$ defined by
Since \( f-g \) is a polynomial of degree at most \( \ell-1 \),

\[
f(jh) - g(jh) = \sum_{s=0}^{m} c_s(jh)^s; \quad c_m \neq 0, \quad 0 \leq m \leq \ell-1,
\]

and

\[
f'(jh) - g'(jh) = \sum_{s=1}^{m} sc_s(jh)^{s-1}.
\]

The case \( m = 0 \), where \( f - g \) is a constant polynomial, will be taken care of at the end of this proof. Therefore, suppose that \( n \) is an integer such that \( 1 \leq n \leq \ell-1 \), then

\[
F(f-g) = \sum_{s=0}^{n-1} c_s(kh)^s + \sum_{j=0}^{k-1} \alpha_j \sum_{s=0}^{n} c_s(jh)^s
\]

\[
- h \sum_{j=0}^{k} \beta_j \sum_{s=1}^{n} sc_s(jh)^{s-1}
\]

\[
= \sum_{s=0}^{n} c_s(kh)^s + \sum_{s=0}^{n} c_s h^s \sum_{j=0}^{k-1} \alpha_j j^s
\]

\[
- h \sum_{s=1}^{n} c_s h^{s-1} \sum_{j=0}^{k} s \beta_j j^{s-1}
\]

\[
= c_0 + c_0 \sum_{j=0}^{k-1} \alpha_j +
\]
\[ + \sum_{s=1}^{n} h \sum_{c_s} \left\{ k^s + \sum_{j=0}^{k-1} \alpha_j j^s - \sum_{j=0}^{k} s \alpha_j j^{s-1} \right\}, \]

and if \( h \neq 0 \), then both conditions in (2.6) are necessary to insure that \( F(f-g) = 0 \).

If \( f-g = c_0 \), where \( c_0 \) is some constant, we can use the argument above to conclude that

\[ F(f-g) = c_0 + c_0 \sum_{j=0}^{k-1} \alpha_j. \]

Thus, for \( c_0 \neq 0 \), the first condition in (2.6) must be satisfied to insure that \( F(f-g) = 0 \).

Also note that (2.6) is equivalent to (2.3) so

\[ \sum_{n=0}^{k} \alpha_n y_n = h \sum_{n=0}^{k} \beta_n f(x_n, y_n), \quad \alpha_k = 1, \]

is a difference formula of degree \( \ell-1 \).

To define approximating subspaces in \( \{ H_{k,h} \}^* \), we select all linear functionals of the form

\[ \phi(f) = h \sum_{j=0}^{k} a_j f'(jh), \text{ for the implicit case,} \quad (2.7) \]

where \( \sum_{j=0}^{k} s \sum_{j} a_j j^{s-1} = 0, \quad s = 1, 2, \ldots, \ell-1, \)

and \( \phi(f) = \sum_{j=0}^{k-1} b_j f'(jh), \text{ for the explicit case,} \quad (2.8) \)
where \[ \sum_{j=0}^{k-1} s b_j j^{s-1} = 0, \quad s = 1, 2, \ldots, \ell-1. \]

Let \( M_1^k \) and \( M_2^k \) respectively represent the two subspaces defined by (2.7) and (2.8). \( M_1^k \) and \( M_2^k \) can be shown to actually be subspaces of \( \{H_{\ell}^k, h\}^* \) by using the same argument which we used previously to show that \( F(f) \) is a linear functional on \( H_{\ell}^k, h \).

We shall also require that \( k \geq \ell-1 \) and \( k \geq \ell \) for \( M_1^k \) and \( M_2^k \) respectively. This condition guarantees that \( M_1^k \) and \( M_2^k \) will consist of linear functionals other than the identically zero functional. To see this, note that if the \( a_j \)'s and the \( b_j \)'s in the formulas

\[
\sum_{j=0}^{k} s a_j j^{s-1} = 0, \quad s = 1, 2, \ldots, \ell-1,
\]

and

\[
\sum_{j=0}^{k-1} s b_j j^{s-1} = 0, \quad s = 1, 2, \ldots, \ell-1,
\]

are thought of as unknowns, we have \( k+1 \) and \( k \) unknowns respectively in \( \ell-1 \) equations. Thus the conditions that \( k \geq \ell-1 \) for \( M_1^k \) and \( k \geq \ell \) for \( M_2^k \) guarantee that in each case there will be more unknowns than equations. Consequently there will exist nonzero sets of \( a_j \)'s and \( b_j \)'s which will satisfy the above equations and will then determine nonzero functionals in \( M_1^k \) and \( M_2^k \).

By Theorem (0.1), this optimal approximation to (2.5) in \( M_1^k \) or \( M_2^k \) is unique. We will denote these approximations respectively by
\[ \phi_0(f) = h \sum_{j=0}^{k} a_j^0 f'(jh), \quad \text{and} \]
\[ \phi_0(f) = h \sum_{j=0}^{k-1} b_j^0 f'(jh). \]

By Theorem (0.1), we also have
\[
\sup_{\|f\|_g=1} | F(f) - \phi_0(f) | \leq \sup_{\|f\|_g=1} | F(f) - \phi(f) |
\]
for any \( \phi \in M_1^k \), so
\[
\sup_{\|f\|_g=1} | f(kh) + \sum_{j=0}^{k-1} \alpha_j f(jh) - h \sum_{j=0}^{k} (\beta_j + a_j^0) f'(jh) | \leq \sup_{\|f\|_g=1} | f(kh) + \sum_{j=0}^{k} \alpha_j f(jh) - h \sum_{j=0}^{k} \beta_j + a_j^0 f'(jh) |
\]
where equality occurs only if \( a_j^0 = a_j \) for all \( j \). We obtain a similar result for \( \phi \in M_2^k \).

The formula
\[
\sum_{j=0}^{k} \alpha_j y_j = h \sum_{j=0}^{k} (\beta_j + a_j^0) f(x_j, y_j),
\]
where \( \alpha_k = 1 \), will be called the optimal implicit formula of order \( k \) and degree \( \ell - 1 \). Similarly,
\[
\sum_{j=0}^{k} \alpha_j y_j = h \sum_{j=0}^{k-1} (\beta_j + b_j^0) f(x_j, y_j),
\]

(2.12)

where \( \alpha_k = 1 \), will be called the optimal explicit formula of order \( k \) and degree \( k-1 \).

The condition \( \alpha_k = 1 \) in the above formulae is necessary to insure uniqueness, for if we did not require that \( \alpha_k = 1 \), we could then get an infinite number of equivalent formulae simply by multiplying through by different constants. By requiring that \( \alpha_k = 1 \), we insure that there is just one optimal implicit or explicit formula.

From here on, we will consider only the case \( h = 1 \). (The abbreviation \( H^k_\ell \) for \( H^{k,1}_\ell \) will sometimes be used). The same results will hold for arbitrary \( h \), but this restriction will simplify the notation.

An optimal implicit formula of degree \( \ell-1 \) will be called \( \zeta \) percent correct (with respect to \( F(f) \)) where

\[
\zeta = 100 \frac{\left( \sup_{\| f \| _\ell = 1} \left| f(k) + \sum_{j=0}^{k-1} \alpha_j f(j) - \sum_{j=0}^{k} \beta_j + b_j^0 f'(j) \right| \right)}{\| F(f) \| _\ell}.
\]

Percent correctness for explicit formulae is similarly defined.

We will now prove two theorems which will allow us to construct optimal implicit and explicit formulae.
**Theorem 2.1.** Let \( g(x) \), \( 0 \leq x \leq k \), satisfy the conditions

\[
g^{(2\ell)}(x) = 0, \quad i < x < i+1, \quad i = 0, 1, \ldots, k-1; \tag{2.13}
\]
\[
g^{(2\ell-1)}(j+0) - g^{(2\ell-1)}(j-0) = \alpha_j, \quad j = 1, \ldots, k-1; \tag{2.14}
\]
\[
g^{(2\ell-2)}(j+0) - g^{(2\ell-2)}(j-0) = \beta_j, \quad j = 1, \ldots, k-1; \tag{2.15}
\]
\[
g^{(2\ell-1)}(0) = \alpha_0, \quad g^{(2\ell-1)}(k) = -1,
\]
\[
g^{(2\ell-2)}(0) = \beta_0, \quad g^{(2\ell-2)}(k) = -\beta_k,
\]
\[
g^{(j)}(0) = 0, \quad j = \ell, \ldots, 2\ell-3,
\]
\[
g^{(j)}(k) = 0, \quad j = \ell, \ldots, 2\ell-3. \tag{2.16}
\]

Then,

\[
\sup_{\|f\|_x = 1} |f(k) + \sum_{j=0}^{k-1} \alpha_j f(j) - \sum_{j=0}^{k} \beta_j f(j) | = \left\{ \int_0^k \left( g^{(\ell)} \right)^2 dx \right\}^{1/2}.
\]

Note the lack of jump conditions in all derivatives less than the \((2\ell-2)\)'nd derivative. By this lack, we shall mean that there are no jumps in these derivatives, i.e., that the derivatives \(2\ell-3, 2\ell-4, \ldots, 1\) are continuous.

As a preliminary result we will prove that

\[
f(k) + \sum_{j=0}^{k-1} \alpha_j f(j) = \sum_{j=0}^{k} \beta_j f(j) = (-1)^{\ell} \int_0^k g^{(\ell)} f(\ell) dx.
\]

The conclusion to Theorem 2.1 will then follow immediately by appealingly to the Riesz Theorem.
Since $F(f)$ is a linear functional on the Hilbert space $H^k$, the Riesz Representation Theorem guarantees that there exists a function $h(x)$ such that

$$(h,f) = F(f), \quad \text{for all } f \in H^k.$$

This $h(x)$ is a unique element of $H^k$. Note, however, that $h(x)$ is not a unique point function since the inner product involves $l'$th derivatives, which means that $h(x)$ will be determined only up to a polynomial of degree $l-1$.

Before proving the theorem, let us consider two examples of explicit formulae.

**Example 1.** Consider the functional

$$F(f) = f(2) - f(1) - (3/2)f'(1) + (1/2)f'(0).$$

This is the Adams-Bashforth formula of a previous example. Suppose for the moment that we assume that the function $f$ has sufficient differentiability and any other conditions needed to guarantee that it belongs to a suitable class of test functions defined on the interval $[0,k]$. Recall that for this formula $l = 3$ and $k = 2$. Viewed as a distribution

$$F(f) = f(2) - f(1) - (3/2)f'(1) + (1/2)f'(0)$$

$$= \int_0^2 \{ \delta(x-2) - \delta(x-1) + (3/2)\delta'(x-1) - (1/2)\delta'(x) \} f \, dx$$
Further, suppose we look at $F(f)$ as $(g^{(2\ell)}, f)$ in the distributional sense, that is

$$F(f) = (g^{(2\ell)}, f) = -(g^{(2\ell-1)}, f') = \ldots = (-1)^{\ell}(g^{(\ell)}, f^{(\ell)})$$

where $g^{(2\ell)} = \delta(x-2) - \delta(x-1) + 3/2 \delta'(x-1) - 1/26(x)$.

In this example, this means that

$$F(f) = -\int_0^2 [H(x-2) - H(x-1) + 3/2\delta(x-1) - 1/26(x)] f'(x) \, dx$$

$$= (-1)^2 \int_0^2 [(x-2)H(x-2) - (x-1)H(x-1) + 3/2H(x-1) - 1/26(x)] f''(x) \, dx$$

$$= (-1)^3 \int_0^2 [(x-2)^2H(x-2) - (x-1)^2H(x-1) + 3/2(x-1)H(x-1) - 1/2xH(x)] f'''(x) \, dx,$$

so

$$g^{(\ell)}(x) = g^{(3)}(x) = \left(\frac{x-2}{2}\right)^2 - \left(\frac{x-1}{2}\right)^2 H(x-1) + 3/2(x-1)H(x-1) - 1/2xH(x),$$

and

$$g(x) = \left(\frac{x-2}{5!}\right)^5 H(x-2) - \left(\frac{x-1}{5!}\right)^5 H(x-1) + 3/2(x-1)^4 H(x-1) - 1/2x^4 H(x) + p(x),$$

where $p(x)$ is an arbitrary polynomial of degree at most 2.

We now claim that $(-1)^3g$ is the $h$ given by the Riesz Theorem. This will be verified by showing that $((-1)^3g, f)_\ell = (-1)^3(g, f)_\ell$ will
yield $F(f)$ by formally integrating $(-1)^3 (g, f)_3$ by parts. Note that this process will be carried out simply by appealing to classical calculus—distribution theory will be used nowhere. This proceeds as follows: Let $f \in H_3^2$, then

$$(-1)^3 (g, f)_3 = -\int_0^2 \left[ \left( \frac{(x-2)^2}{2} H(x-2) - \frac{(x-1)^2}{2} H(x-1) + \frac{3}{2} (x-1) H(x-1)^2 \right. \right.$$

$$- \frac{1}{2} x H(x) \big] f^{(3)}(x) \big] \, dx$$

$$= - \left[ \left\{ (x-2)^2 H(x-2) - (x-1)^2 H(x-1) + \frac{3}{2} (x-1) H(x-1)^2 \right. \right.$$

$$- \frac{1}{2} x H(x) \big] f^{(2)}(x) \big] \big] \big]_0^2 + \int_0^2 \left[ (x-2) H(x-2) - (x-1) H(x-1) + \frac{3}{2} H(x-1)^2 \right.$$  

$$- \frac{1}{2} x H(x) \big] f^{(2)}(x) \big] \, dx.$$

The first expression in the last equality is equivalent to

$$1/2 f^{(2)}(2) - 3/2 f^{(2)}(2) + f^{(2)}(2) = 0.$$

The last two terms in the integral can be integrated. Thus

$$\int_0^2 \left[ \frac{3}{2} H(x-1) - \frac{1}{2} H(x) \right] f^{(2)}(x) \, dx = \int_0^2 f^{(2)}(x) \, dx - \int_0^2 \frac{1}{2} f^{(2)}(x) \, dx$$

$$= f'(2) - f'(1) + 1/2 f'(0).$$

Integrating the remaining terms by parts, we obtain
\[ \int_0^2 \left[ (x-2)H(x-2) - (x-1)H(x-1) \right] f''(x) \, dx \]
\[ = \left[ \left( (x-2)H(x-2) - (x-1)H(x-1) \right) f'(x) \right]^2_0 \]
\[ - \int_0^2 \left[ H(x-2) - H(x-1) \right] f'(x) \, dx \]
\[ = -f'(2) + \int_1^2 f'(x) \, dx \]
\[ = -f'(2) + f(2) - f(1). \]

Putting everything together,
\[ (-1)^3 (g,f)_3 = f'(2) - 3/2 f'(1) + 1/2 f'(0) - f'(2) + f(2) - f(1) \]
\[ = f(2) - f(1) - 3/2 f'(1) + 1/2 f'(0). \]

Therefore, we have shown that
\[ (-1)^3 (g,f)_3 = F(f) \]

without appealing to distribution theory. Nevertheless, the correct results are obtained using the formal manipulations on distributions.

**Example 2.** Consider the functional

\[ F(f) = f(2) - f(1) - f'(1). \]

Here \( \ell = 2 \) and \( k = 2 \). Using distributional techniques again,

\[ F(f) = \int_0^2 \left[ \delta(x-2) - \delta(x-1) + \delta'(x-1) \right] f(x) \, dx \]
\[
\begin{align*}
&= - \int_0^2 [H(x-2) - H(x-1) + \delta(x-1)] f'(x) \, dx \\
&= \int_0^2 [(x-2)H(x-2) - (x-1)H(x-1) + H(x-1)] f^{(2)}(x) \, dx.
\end{align*}
\]

Therefore,
\[
g^{(2)}(x) = g^{(1)}(x) = (x-1)H(x-2) - (x-1)H(x-1) + H(x-1)
\]

and
\[
g(x) = \left(\frac{x-2}{3!}\right)^3 H(x-2) - \left(\frac{x-1}{3!}\right)^3 H(x-1) + \left(\frac{x-1}{2!}\right)^2 H(x-1) + p(x)
\]

where \( p(x) \) is an arbitrary polynomial of degree at most one.

Again without using distribution theory, we will show that this \( g(x) \) is the function which satisfies the relation
\[
(-1)^2 (g,f) = F(f), \quad f \in H^1_{\mathbb{R}}.
\]

\[
\begin{align*}
(-1)^2 (g,f)_2 &= \int_0^2 [(x-2)H(x-2) - (x-1)H(x-1) + H(x-1)] f^{(2)}(x) \, dx \\
&= \int_0^2 [(x-2)H(x-2) - (x-1)H(x-1)] f^{(2)}(x) \, dx + \int_0^2 H(x-1) f^{(2)}(x) \, dx \\
&= \int_0^2 [H(x-2) - H(x-1)] f'(x) \, dx - \int_0^2 [H(x-2) - H(x-1)] f'(x) \, dx \\
&\quad + \int_1^2 f^{(2)}(x) \, dx.
\end{align*}
\]
\begin{align*}
&= -f'(x) + f'(x) \bigg|_1^2 - \int_0^2 H(x-1)f'(x) \, dx \\
&= -f'(2) + f'(2) - f'(1) + f(2) - f(1) \\
&= f(2) - f(1) - f'(1) \\
&= F(f).
\end{align*}

Therefore, \((-1)^2(f,g) - F(f), f \in H^2\), and this \(g(x)\), which we constructed by appealing to distribution theory, once again has proven to be the correct function.

**Proof of Theorem 2.1**

**Part I**

We will first show that given any linear functional on \(H^k\) of the form

\[ F(f) = \sum_{j=0}^{k} \alpha_j f(j) - \sum_{j=0}^{k} \beta_j f'(j) \]

where the \(\alpha_j\)'s and \(\beta_j\)'s satisfy (2.6), we can find a function \(g(x)\) such that

\[ F(f) = \langle g, f \rangle_{H^k} \text{ for all } f \in H^k. \]

We claim that

\begin{align*}
\sum_{j=0}^{k} \alpha_j \frac{(x-1)^{2j-1}}{(2j-1)!} H(x-j) + \sum_{j=0}^{k} \beta_j \frac{(x-1)^{2j-2}}{(2j-2)!} H(x-j) \\
+ \sum_{j=0}^{k} c_j x^j, & \quad (2.17)
\end{align*}
where \( H(x-j) \) is the Heaviside function and the \( c_j \)'s are arbitrary constants, is the correct function. This formula was derived by using the same distributional techniques we used in the examples.

Note that for \( m = 2\ell-(s+1), m = \ell, \ldots, 2\ell-2 \), or, equivalently, \( s = \ell-1, \ldots, 3, 2, 1 \),

\[
g^{(m)}(x) = \sum_{j=0}^{k} a_j \frac{x-j}{s!} H(x-j) + \sum_{j=0}^{k} b_j \frac{(x-j)^{s-1}}{(s-1)!} H(x-j), \quad (2.18)
\]

so that \( g^{(\ell)}(x), \ldots, g^{(2\ell-3)}(x) \) are continuous functions and \( g^{(2\ell-2)}(x) \) is discontinuous. Here we are assuming that \( \ell \leq 2\ell-3 \) which implies that \( \ell \geq 3 \). In the special case \( \ell = 2, \ell = 2\ell-2 \), so that the \( \ell \)'th derivative is also discontinuous.

In order to show that

\[
F(f) = (-1)^{\ell} (g,f)_{\ell}, \text{ for all } f \in H^k_{\ell},
\]

we will want to claim that repeated integration by parts gives us

\[
\int_{0}^{k} g^{(\ell)} f^{(\ell)} \, dx = g^{(\ell)} f^{(\ell-1)} \bigg|_{0}^{k} - \int_{0}^{k} g^{(\ell-1)} f^{(\ell+1)} \, dx
\]

\[
= 0 - \int_{0}^{k} g^{(\ell-1)} f^{(\ell+1)} \, dx
\]

\[
= \ldots
\]

\[
= (-1)^{\ell-1} f^{(2\ell-3)} f^{(2)} \bigg|_{0}^{k} - \int_{0}^{k} g^{(2\ell-2)} f^{(2)} \, dx
\]

\[
= (-1)^{\ell-2} \int_{0}^{k} g^{(2\ell-2)} f^{(2)} \, dx.
\]
After we have verified that

\[
(g,f)^\ell = (-1)^{\ell-2} \int_0^k g^{(2\ell-2)}(x) f'(x) \, dx,
\]

(2.19)

We will merely integrate the right hand side by classical calculus techniques to conclude that

\[
(-1)^{\ell}(g,f)^\ell = F(f).
\]

Note that for \( \ell = 2 \), \((g,f)^\ell\) is equal to the right hand side of (2.19) by definition.

Clearly if (2.19) is to be correct,

\[
g^{(m)}f(2\ell-m) \bigg|_0^k
\]

must be equal to zero for \( m = \ell, \ldots, 2\ell-3 \). As we stated before, \( g^{(m)}(x), m = \ell, \ldots, 2\ell-3 \), is continuous. Since \( f^{(1)}(x) \) is continuous for \( i = \ell, \ldots, 2 \), a sufficient condition for (2.20) to be zero is that \( g^{(m)}(0) = g^{(m)}(k) = 0 \) for \( m = \ell, \ldots, 2\ell-3 \).

Now for \( m = 2\ell-(s+1) = \ell, \ldots, 2\ell-3 \),

\[
g^{(m)}(0) = \sum_{j=0}^{k-1} \alpha_j \frac{(-1)^s}{s!} H(-j) + \sum_{j=0}^{k-1} \beta_j \frac{(-1)^{s-1}}{(s-1)!} H(-j).
\]

Therefore, since \( H(-j) = 0 \) except possibly for \( j = 0 \), and then \((0)^s = (0)^{s-1} = 0 \), it follows that

\[
g^{(m)}(0) = 0, \text{ } m = \ell, \ldots, 2\ell-3.
\]

To show that \( g^{(m)}(k) = 0 \), however, is not so simple. The proof relies heavily on (2.6) and on properties of the coefficients of a
binomial expansion. Let $\binom{s}{n} = \frac{s!}{(s-n)! \ n!}$, then

\[
\binom{s}{n} = \binom{s-1}{n} \frac{s}{(s-n)} ,
\]

\[
(1-1)^n = \sum_{n=0}^{s} (-1)^n \binom{s}{n} = 0 ,
\]

and

\[
\sum_{n=0}^{s-1} (-1)^n \binom{s}{n} = (-1)^{s-1} .
\]

Now for $m = (2\ell-(s+1)) = \ell, \ldots, 2\ell-3$ (or $s = 2, \ldots, \ell-1$),

\[
g^{(m)}(k) = \sum_{j=0}^{k-1} \alpha_j \frac{(k-j)^s}{s!} + \sum_{j=0}^{k-1} \beta_j \frac{(k-j)^{s-1}}{(s-1)!}
\]

so that

\[
s! g^{(m)}(k) = \sum_{j=0}^{k-1} \alpha_j \sum_{n=0}^{s} \binom{s}{n} k^n (-j)^{s-n}
\]

\[
+ \sum_{j=0}^{k-1} s \beta_j \sum_{n=0}^{s-1} \binom{s-1}{n} k^n (-j)^{s-1-n}
\]

\[
= k^s \sum_{j=0}^{k-1} \alpha_j + \sum_{n=0}^{k-1} \sum_{n=0}^{s-1} \alpha_j \binom{s}{n} k^n (-j)^{s-n}
\]

\[
+ \sum_{j=0}^{k-1} \sum_{n=0}^{s-1} \beta_j \binom{s-1}{n} k^n (-j)^{s-n-1}
\]

\[
= -k^s + \sum_{n=0}^{s-1} k^n \binom{s}{n} \left\{ \sum_{j=0}^{k-1} \alpha_j (-j)^{s-n} + (s-n) \beta_j (-j)^{s-n-1} \right\}
\]
From (2.6) it follows that

\[ s!g^{(m)}(k) = -k^s + \sum_{n=0}^{s-1} \binom{s}{n} (-1)^{s-n} \left\{ \sum_{j=0}^{k-1} \alpha_j i^{s-n} - (s-n) \beta_j i^{s-n-1} \right\}. \]

Therefore \( s!g^{(m)}(k) = 0 \) for \( m = \ell, \ldots, 2\ell-3 \), which implies that \( g^{(m)}(k) = 0 \) for \( m = \ell, \ldots, 2\ell-3 \).

Thus \( g^{(m)}(0) = g^{(m)}(k) = 0, m = \ell, \ldots, 2\ell-3 \), and as a resulting

\[ g^{(m)}_f(2\ell-m) \bigg|_0^k = 0. \]

Consequently, integrating \( (g,f) \) repeatedly by parts, we obtain the formula

\[ (g,f)_\ell = (-1)^{\ell-2} \int_0^k (2\ell-2)(x)f^{(2)}(x) \, dx. \]
Recalling that
\[ g(2k-2)(x) = \sum_{j=0}^{k} \alpha_j(x-j)H(x-j) + \sum_{j=0}^{k} \beta_j H(x-j), \]
we continue the process of integration:

\[
(g,f)_l = (-1)^{l-2} \int_0^k \left\{ \sum_{j=0}^{k} \alpha_j(x-j)H(x-j) + \sum_{j=0}^{k} \beta_j H(x-j) \right\} f^{(2)}(x) \, dx
\]

\[
= (-1)^{l-2} \sum_{j=0}^{k} \alpha_j(x-j)H(x-j)f^{(2)}(x) \, dx
\]

\[
+ (-1)^{l-2} \int_0^k \sum_{j=0}^{k} \beta_j H(x-j)f^{(2)}(x) \, dx
\]

\[
= (-1)^{l-2} \sum_{j=0}^{k} \alpha_j(x-j)H(x-j)f'(x) \bigg|_0^k
\]

\[
- (-1)^{l-2} \sum_{j=0}^{k} \alpha_j H(x-j)f'(x)
\]

\[
+ (-1)^{l-2} \sum_{j=0}^{k} \beta_j H(x-j)f^{(2)}(x) \, dx
\]

\[
= (-1)^{l-2} \sum_{j=0}^{k} \alpha_j(k-j)H(k-j)f'(k)
\]

\[
+ (-1)^{l-1} \sum_{j=0}^{k-1} \alpha_j f(x) \, dx
\]
\[-\frac{1}{(k)_{-2}} \sum_{j=0}^{k-1} \alpha_j (k-j) f'(k) + (-1)^{k-1} \sum_{j=0}^{k-1} \alpha_j \{ f(k) - f(j) \} \]
\[+ (-1)^{k-2} \sum_{j=0}^{k-1} \beta_j \{ f'(k) - f'(j) \} \]
\[= (-1)^{k-2} k f'(k) \sum_{j=0}^{k-1} \alpha_j + (-1)^{k-1} f'(k) \sum_{j=0}^{k-1} \alpha_j \]
\[+ (-1)^{k-1} f(k) \sum_{j=0}^{k-1} \alpha_j + (-1)^k \sum_{j=0}^{k-1} \alpha_j f(j) \]
\[+ (-1)^{k-2} f'(k) \sum_{j=0}^{k-1} \beta_j + (-1)^{k-1} \sum_{j=0}^{k-1} \beta_j f'(j). \]

Since \[\sum_{j=0}^{k-1} \alpha_j = -1,\]

\[-\frac{1}{(k)_{-2}} \sum_{j=0}^{k-1} \alpha_j = (-1)^k f(k), \]

and \[-\frac{1}{(k)_{-2}} k f'(k) \sum_{j=0}^{k-1} \alpha_j = (-1)^{k-1} k f'(k), \]

it follows that \[(g,f)_k = (-1)^k \sum_{j=0}^{k} \alpha_j f(j) + (-1)^{k-1} \sum_{j=0}^{k-1} \beta_j f'(j) \]
\[+ (-1)^{k-1} k f'(k) + (-1)^{k-2} f'(k) \left\{ \sum_{j=0}^{k-1} \beta_j - \sum_{j=0}^{k-1} \alpha_j j \right\}. \]
From (2.6), \[ \sum_{j=0}^{k-1} \beta_j - \sum_{j=0}^{k-1} \alpha_j = k - \beta_j. \]

As a result \[ (f,g)_\ell = (-1)^\ell \sum_{j=0}^{k} \alpha_j f(j) + (-1)^{\ell-1} \sum_{j=0}^{k-1} \beta_j f(j) \]

\[ + (-1)^{\ell-1} k f'(k) + (-1)^{\ell-2} f'(k) k + (-1)^{\ell-1} f'(k) \beta_k \]

\[ = (-1)^\ell \sum_{j=0}^{k} \alpha_j f(j) + (-1)^{\ell-1} \sum_{j=0}^{k} \beta_j f'(j), \]

and \[ (-1)^\ell (g,f)_\ell = \sum_{j=0}^{k} \alpha_j f(j) - \sum_{j=0}^{k} \beta_j f'(j) = F(f). \]

Therefore, the function \( g(x) \) defined by (2.17) is such that \((-1)^\ell g(x) \) is the function \( h(x) \) guaranteed by the Riesz Theorem, i.e.,

\[ (h,f)_\ell = (-1)^\ell (g,f)_\ell = F(f). \]

**Part II**

We will now show that the conditions (2.13), (2.14), (2.15), and (2.16) in the hypothesis of Theorem 2.1 determine a unique function in \( H^k_\ell \), i.e., unique up to a polynomial of degree \( \ell-1 \), and that the function determined by these conditions is actually the function we have previously defined in (2.16). We will then have completed the proof of Theorem 2.1.
Condition (2.13) stated that

\[ g^{(2\ell)}(x) = 0, \text{ } 1 \leq x < i+1, \text{ } i = 0, \ldots, k-1. \]

Integrating \( g^{(2\ell)} \) over each of the subintervals, we obtain the formula

\[ g^{(2\ell-1)}(x) = c_i, \text{ } i \leq x < i+1, \text{ } i = 0, \ldots, k-1 \]

(2.22)

where the \( c_i \)'s must satisfy

\[ c_{i+1} - c_i = \alpha_{i+1}, \text{ } c_0 = \alpha_0. \]

(2.23)

Here we are appealing to (2.14) and (2.16) to get the conditions on the \( c_i \)'s. From (2.23) it is clear that we can represent each of the constants, \( c_i \), in the form

\[ c_i = \sum_{j=0}^{i} \alpha_j \]

so that

\[ g^{(2\ell-1)}(x) = \sum_{j=0}^{i} \alpha_j, \text{ } i \leq x < i+1, \text{ } i = 0, \ldots, k-1. \]

(2.24)

Note that (2.6) implies that \( g^{(2\ell-1)}(k) = -1. \)

On integrating \( g^{(2\ell-1)}(x) \), we find that \( g^{(2\ell-2)}(x) \) must be expressed in the form

\[ g^{(2\ell-2)}(x) = xc_i + d_i, \text{ } i \leq x < i+1, \text{ } i = 0, \ldots, k-1 \]

(2.25)

where \( d_0 = \beta_0 \), and

\[ (i+1)c_{i+1} + d_{i+1} - ((i+1)c_i + d_i) = \beta_{i+1}. \]

(2.26)
Here our jump conditions come from (2.15), and (2.16). Using (2.26), we find that the $d_i$'s can be represented in the form

$$d_i = \sum_{j=0}^{i} \beta_j - \sum_{j=0}^{i} \alpha_j.$$  

It then follows from (2.25) that

$$g^{(2\ell-2)}(x) = \sum_{j=0}^{i} (x-j)\alpha_j + \sum_{j=0}^{i} \beta_j, \quad i<x<i+1, \quad i = 0, \ldots, k-1 \quad (2.27)$$

Remembering that (2.6) states that

$$\sum_{j=0}^{k} \alpha_j = -1, \quad \text{and} \quad k + \sum_{j=0}^{k-1} \alpha_j = \sum_{j=0}^{k} \beta_j,$$

it is clear that

$$g^{(2\ell-2)}(k) = -\beta_k.$$  

Integrating once again, we obtain the relation

$$g^{(2\ell-3)}(x) = \sum_{j=0}^{i} \frac{(x-j)^2}{2} \alpha_j + \sum_{j=0}^{i} x\beta_j + d_i,$$

$$i<x<i+1, \quad i = 0, \ldots, k-1.$$  

Since $g^{(2\ell-3)}(x)$ must be continuous, it follows that

$$\sum_{j=0}^{i} \frac{(i+1-j)^2}{2} \alpha_j + \sum_{j=0}^{i} (i+1)\beta_j + d_i.$$
\[- \left\{ \sum_{j=0}^{i+1} \frac{(i+1) - 1}{2} a_j + \sum_{j=0}^{i+1} (i+1) b_j + d_{i+1} \right\} = 0.\]

Simplifying this equation, we get the recursion relation

\[d_{i+1} = d_1 - (i+1) b_{i+1}\]

from which we obtain the representation

\[d_1 = - \sum_{j=0}^{i} j b_j, \quad j = 0, \ldots, k-1.\]

Hence,

\[g^{(2\ell-3)}(x) = \sum_{j=0}^{1} \frac{(x-1)^2}{2} a_j + \sum_{j=0}^{1} (x-j) b_j. \quad (2.28)\]

Here we have dropped the piecewise definition of \(g^{(2\ell-3)}(x)\) since \(g^{(2\ell-3)}(x)\) is continuous and its values at the net points are uniquely determined.

The same argument (following (2.21)) we used to show that \(g^{(2\ell-3)}(2) = 0\), where \(g\) was defined by (2.17), works here to verify that \(g^{(2\ell-3)}(k) = 0\), where \(g^{(2\ell-3)}(x)\) is defined by (2.28). Consequently, (2.28) determines what the \(2\ell-3'rd\) derivative of the \(g\) function given by Theorem 2.1 must look like.

Using the same techniques we used above for \(g^{(2\ell-3)}(x)\) we can show, for \(m = \ell, \ldots, 2\ell-3\), that
\[ g^{(m)}(x) = \sum_{j=0}^{i} \frac{(x-1)^{s} \alpha_j}{s!} + \sum_{j=0}^{i} \frac{(x-1)^{s-1} \beta_j}{(s-1)!}, \quad m = (2-(s+1)), \]

determines the unique \( m \)'th derivatives of \( g \). We have now satisfied the hypothesis to Theorem 2.1 and have obtained a representation for the \( \ell \)'th through \( 2\ell \)'th derivatives of \( g(x) \) (in the open intervals). From this point on, the process of integration gives arbitrary constants. Hence,

\[ g(x) = \sum_{j=0}^{i} \frac{\alpha_j}{(2\ell-1)!} (x-1)^{2\ell-1} + \sum_{j=0}^{i} \frac{\beta_j}{(2\ell-2)!} (x-1)^{2\ell-2} + p(x) \quad (2.29) \]

where \( p(x) \) is an arbitrary polynomial of degree \( \ell-1 \). Thus, \( g(x) \) is uniquely determined up to a polynomial of degree \( \ell-1 \).

On comparing (2.29) to (2.17), we see that, aside from the arbitrary polynomials, (2.29) and (2.17) give identical point values. Thus (2.29) and (2.17) are merely different representations of the same function (uniquely determined in \( H^k_{\ell} \)). Consequently, the hypothesis to Theorem 2.1 determines a function which we have already shown gives the desired conclusion, and we have completed the proof of Theorem 2.1.

**Theorem 2.2.** The formulae (2.11) and (2.12) will be optimal if, and only if, the function \( g(x) \), determined in Theorem 2.1, can be chosen such that for the explicit and implicit formulae

\[ g'(j) = 0 \text{ for } j = 0, \ldots, k-1 \text{ and } \]
\[ g'(j) = 0 \text{ for } j = 0, \ldots, k, \]

respectively.
Proof. Part I. Since the proofs for both the explicit and the implicit cases are similar, we will establish that (2.30) is a necessary and sufficient condition only for implicit formulae. To prove sufficiency, assume that (2.30) holds and let \( \phi \in \mathcal{M}_k \), then

\[
F(f) - \phi_o(f) = f(k) + \sum_{j=0}^{k-1} \alpha_j f(j) - \sum_{j=0}^{k} (\beta_j + a_j^0) f(j).
\]

From Theorem 2.1, there exists a \( g(x) \) such that

\[
F(f) - \phi_o(f) = (f, (-l)^l g).
\]

Let \( G(x) = (-l)^l g(x) \), then \( G(x) \) is the \( z_{F-\phi} \) mentioned in Theorem 0.1. Since \( g'(j) = 0 \) for \( j = 0, \ldots, k \) implies that \( G'(j) = 0 \) for \( j = 0, \ldots, k \), it follows from the definition of \( \phi(f) \), (2.7), that for \( \phi \in \mathcal{M}_k \)

\[
\phi (z_{F-\phi}) = \sum_{j=0}^{k} a_j G'(j) = 0.
\]

Therefore, from Theorem 0.1, \( \phi_o \) is the optimal approximation to (2.9) in \( \mathcal{M}_k \) and (2.11) is the optimal implicit formula of order \( k \) and degree \( l-1 \).

Part II. We will now prove that (2.30) is a necessary condition for (2.9) to be optimal. Therefore assume that \( g(x) \) is the optimal approximation to (2.5) in \( \mathcal{M}_k \). By the preceding theorem, if \( g(x) \) is the function corresponding to the optimal implicit formula, then one has for all \( a_j \) satisfying the conditions

\[
\sum_{j=0}^{k} s a_j j^{s-1} = 0, \quad s = 1, \ldots, l-1.
\]  

(2.31)
that
\[ \sum_{j=0}^{k} a_j g'(j) = 0. \] (2.32)

Considering the \( a_j \) as unknowns, (2.31) determines \( \ell-1 \) equations in \( k+1 \) unknowns. Consequently, \( k+1-(\ell-1) \) of the \( a_j \) may be chosen arbitrarily and the other \( a_j \) then solved for in terms of these choices. Theorem 2.1 determines \( g(x) \) up to a polynomial of degree \( \ell-1 \). This implies that we can modify \( g(x) \) by a polynomial of degree \( \ell-1 \), or, equivalently, \( g'(x) \) by a polynomial of degree \( \ell-2 \), and still keep \( g(x) \) in the same equivalence class. In a polynomial of degree \( \ell-2 \), we can control its value at \( \ell-1 \) points. Thus, by modifying \( g'(x) \) by a polynomial of degree \( \ell-2 \), we can insures that \( g'(j) = 0 \) for \( j = 0, \ldots, \ell-2 \) so that \( g'(j) \) may be nonzero for at most \( k-(\ell-2) = k+1-(\ell-1) \) values of \( j \). But then from the arbitrary nature of \( k+1-(\ell-1) \) of the \( a_j \), it follows that in order to satisfy (2.32) for all possible values of the \( a_j \), \( g'(j) \) must be zero for \( j = 0, \ldots, k \).

Examples

The following examples will illustrate the way in which the previous theorems can be used to construct optimal formulae.

Example 1. For our first example, we will construct optimal explicit formulae of the form

\[ y_{n+3} = y_{n+2} = h \left\{ \beta_2 y'_{n+2} + \beta_1 y'_{n+1} + \beta_0 y'_{n} \right\} \] (2.33)
Part I. We will first show that the optimal explicit formula of order 3 and degree 2 \((\ell = 3)\) is

\[
y_{n+3} = y_{n+2} + h \{17/16 y_{n+2} = 3/4 y_{n+1} + 1/8 y_n\}.
\]  

(2.34)

Note that the function \(g(x)\) mentioned in theorem 2.1 is given, in general, by

\[
g(x) = \sum_{j=0}^{k-1} \alpha_j (x-j)^{2\ell-1} H(x-j) + \sum_{j=0}^{k-1} \beta_j (x-j)^{2\ell-2} H(x-j)
\]

\[
+ \sum_{j=0}^{\ell-1} \frac{c_j x^j}{j!}
\]

so that

\[
g'(x) = \sum_{j=0}^{k-1} \frac{\alpha_j (x-j)^{2\ell-2} H(x-j)}{(2\ell-2)!} + \sum_{j=0}^{k-1} \frac{\beta_j (x-j)^{2\ell-3} H(x-j)}{(2\ell-3)!}
\]

\[
+ \sum_{j=1}^{\ell-1} \frac{c_j x^{j-1}}{(j-1)!}
\]

In our case, since \(\ell = 3\) and \(k = 3\), we have

\[
g'(x) = -\frac{(x-2)^4 H(x-2)}{4!} + \sum_{j=1}^{2} \frac{\beta_j (x-j)^3 H(x-j)}{3!} + \sum_{j=1}^{2} \frac{c_j x^{j-1}}{(j-1)!}
\]

Appealing to theorem 2.2, \(g(x)\) must satisfy the conditions
\[ g'(0) = c_1 = 0, \]
\[ g'(1) = (\beta_0/3!) + c_2 = 0, \]
and
\[ g'(x) = (\beta_0(2)^3/3!) + (\beta_1/3!) + 2c_2 = 0. \]

(2.6) gives us the conditions
\[ 5 = 2\beta_1 + 4\beta_2, \]
and
\[ 1 = \beta_0 + \beta_1 + \beta_2. \]

Solving this system of linear equations we find that
\[ \beta_0 = -6c_2, \]
\[ \beta_1 = 36c_2, \]
\[ \beta_2 = -18c_2 + 5/4, \]
and
\[ c_2 = -1/48. \]

Consequently \( \beta_0 = 1/8, \beta_1 = -3/4, \) and \( \beta_2 = 17/16. \)

Part II. We will now show that the optimal explicit formula of order 3 \((\ell = 4)\) is
\[
Y_{n+3} = Y_{n+2} = h\{5/12y'_n - 4/3y'_{n+1} + 23/12y''_{n+2}\}. \quad (2.35)
\]

Note that in this case \(k = \ell + 1\) so that \(\beta_0, \beta_1,\) and \(\beta_2\) will be completely determined by (2.6), i.e., the \(\beta\)'s must satisfy the conditions in (2.6) for \(s = 1, 2, 3,\) so
\[ \beta_0 + \beta_1 + \beta_2 = 1, \]
\[ 2\beta_1 + 4\beta_2 = 5, \]
and
\[ 3\beta_1 + 12\beta_2 = 19. \]

Solving this system, we find that \(\beta_0 = 5/12, \beta_1 = -4/3,\) and \(\beta_2 = 23/12.\)
Part III. In order to find the percent correctness of (2.34) with respect to (2.35) we must find

\[
\sup_{\|f\|_{L^1}} \left| f(3) - f(2) - \frac{1}{8}f'(0) - \frac{3}{4}f'(1) + \frac{17}{16}f'(2) \right|
\]

where \( \ell = 3 \). From Theorem 2.1,

\[
\zeta = 100 \frac{\int_0^3 (g_1^{(3)})^2 \, dx}{\int_0^3 (g_2^{(3)})^2 \, dx}^{1/2}
\]

where

\[
g_1^{(3)}(x) = -\frac{(x-2)^2}{2}H(x-2) + \frac{1}{8}xH(x) - \frac{3}{4}(x-1)H(x-1) + \frac{17}{16}(x-2)H(x-2)
\]

and

\[
g_2^{(3)}(x) = -\frac{(x-2)^2}{2}H(x-2) + \frac{5}{12}xH(x) - \frac{4}{3}(x-1)H(x-1) + \frac{23}{12}(x-2)H(x-2).
\]

On computing the quotient for integrals above, we see that \( \zeta = 82.7211 \).

Example 2. For our second example, we will consider implicit formulae of the form

\[
y_{n+2} = y_{n+1} + h \{ \beta_2 y'_n + \beta_1 y'_n + \beta_0 y'_n \}.
\]
Part I. We will first construct the optimal implicit formula of the above form of order 2 and degree 1 ($\ell = 2$). The $g'(x)$ given in Theorem 2.1 is

$$g'(x) = -\frac{(x-1)^2}{2}H(x-1) + \sum_{j=0}^{1} \beta_j (x-j)H(x-j) + c_1.$$  

Thus, Theorem 2.2 implies that

$$g'(0) - c_1 = 0,$$

$$g'(1) = \beta_0 = 0,$$

and $$g'(2) = -1/2 + \beta_1 = 0.$$  

Therefore $\beta_0 = 0$ and $\beta_1 = 1/2$. From (2.6) it follows that

$$1 = \beta_0 + \beta_1 + \beta_2.$$  

Thus $\beta_2 = 1/2$. Consequently,

$$y_{n+2} = y_{n+1} + h \{1/2y'_{n+2} - 1/2y'_{n+1}\}$$  

(2.37)

is the optimal implicit formula of degree 1 and order 2.

Part II. We will now show that the optimal implicit formula of degree 2 ($\ell = 3$) is

$$y_{n+2} = y_{n+1} + h \{7/16y'_{n+2} + 5/8y'_{n+1} - 1/16y'_{n}\}.$$  

(2.38)

Using the same methods as before, we find that

$$g'(x) = -\frac{(x-1)^4}{4!}H(x-1) + \sum_{j=0}^{1} \beta_j \frac{(x-1)^3}{3!}H(x-j) + c_2x + c_1.$$
so \[ g'(0) = c_1 = 0, \]
\[ g'(1) = \beta_0/6 + c_2 = 0, \]
\[ g'(2) = -1/24 + \beta_0(8/6) + \beta_1(1/6) + 2c_2 = 0, \]

and from (2.6),
\[ 1 = \beta_0 + \beta_1 + \beta_2, \]

and \[ 3 = 2\beta_1 + 4\beta_2. \]

Consequently
\[ \beta_0 = -6c_2, \]
\[ \beta_1 = 36c_2 + 1/4, \]
\[ \beta_2 = -18c_2 + 5/8, \]

and \[ c_2 = 1/8(12), \]

so that \[ \beta_0 = -1/16, \beta_1 = 5/8, \text{ and } \beta_2 = 7/16. \]

Part III. The implicit formula of maximum possible degree
(degree 3, \( l = 4 \)) is given by

\[
y_{n+2} = y_{n+1} + h \left\{ 5/12y_{n+2} + 2/3y_{n+1} - 1/2y_n \right\}. \tag{2.39}
\]

This formula is derived by appealing to (2.6) from which we get the equations
\[
1 = \beta_0 + \beta_1 + \beta_2 \]
\[
3 = 2\beta_1 + 4\beta_2 \]
\[
7 = 3\beta_1 + 24\beta_2. \]

Solving this system for \( \beta_0, \beta_1, \text{ and } \beta_2, \) we arrive at (2.39).
To find the percent correctness of (2.37) with respect to (2.39) we appeal to Theorem 2.1 from which it follows that

\[
\zeta = 100 \frac{\sup_{\|f\|_2 = 1} \left| f(2) - f(1) - \left\{ \frac{1}{2}f'(2) + \frac{1}{2}f'(1) \right\} \right|}{\sup_{\|f\|_2 = 1} \left| f(2) - f(1) - \left\{ \frac{5}{12}f'(2) + \frac{2}{3}f'(1) - \frac{1}{12}f'(0) \right\} \right|}
\]

\[
100 \frac{\left\{ \int_0^2 \left( g_1^{(2)}(x) \right)^2 \right\}^{1/2}}{\left\{ \int_0^2 \left( g_2^{(2)}(x) \right)^2 \right\}^{1/2}}
\]

where

\[
g_1^{(2)}(x) = -(x-1)H(x-1) + \frac{1}{2}H(x-1)
\]

and

\[
g_2^{(2)}(x) = -(x-1)h(x-1) - \frac{1}{12}H(x) + \frac{2}{3}H(x-1).
\]

Now

\[
\int_0^2 (g_1^{(2)})^2 dx = \int_0^2 (x^2 - 3x + \frac{9}{4})H(x-1) dx
\]

\[
\int_0^2 (g_2^{(2)})^2 dx = \int_0^2 \left\{ (x^2 - \frac{19}{6}x + \frac{45}{18})H(x-1) + \frac{1}{144}H(x) \right\} dx
\]

\[
\int_0^1 x^2 - 3x + \frac{9}{4} dx = \frac{1}{12}
\]

\[
\int_0^1 x^2 - \frac{19}{6}x + \frac{45}{18} dx + \int_0^2 \left( \frac{1}{144} \right) dx = \frac{7}{72}
\]

Consequently, \( \zeta = 100 \frac{\sqrt{\frac{1}{12}}}{\sqrt{\frac{7}{72}}} = 92.5820. \)
The percent correctness of (2.38) with respect to (2.39) is given by

\[
\zeta = 100 \frac{\int_0^2 (g_1(3))^2 \, \text{d}x}{\int_0^2 (g_2(3))^2 \, \text{d}x}^{1/2}
\]

where, in this case,

\[
g_1(3)(x) = \frac{-(x-1)^2}{2} \text{H}(x-1) - \frac{1}{16x}\text{H}(x) + \frac{5}{8}(x-1)\frac{\text{H}(x)}{2}
\]

and

\[
g_2(3)(x) = \frac{-(x-1)^2}{2} \text{H}(x-1) - \frac{1}{12x}\text{H}(x) + \frac{2}{3}(x-1)\text{H}(x-1).
\]

It turns out that \(\zeta = 97.5665\).

We will now list a few optimal difference formulae. Some of these formulae have already been seen as previous examples; others are new. Then, some comments will be made regarding the use of optimal formulae.

The information in the following two tables was taken from [1, pp. 76-77].

A few comments are now in order. First, for each optimal formula we derived in the examples, the \(\alpha_j, j = 0, 1, \ldots, k\), were given and then the \(\beta_j\) computed, i.e., we optimized only with respect to changes in the \(\beta_j\). The reason for this is that certain kinds of stability depend only on the values of the \(\alpha_j\). Thus, by fixing the \(\alpha_j\), we can insure the stability of the difference equation and then determine the optimal set of \(\beta_j\).
Table 2.1. Optimal explicit formulae of Adams type and the values of $\zeta$ for the Adams-Bashforth formulae

$k=2$

$$y_{n+2} = y_{n+1} + h(\beta_1 y'_{n+1} + \beta_0 y'n)$$

<table>
<thead>
<tr>
<th>$l$</th>
<th>$\beta$</th>
<th>$\zeta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>75.5928</td>
</tr>
<tr>
<td>3</td>
<td>1.5</td>
<td>100.000</td>
</tr>
</tbody>
</table>

$k=3$

$$y_{n+3} = y_{n+2} + h(\beta_2 y'_{n+2} + \beta_1 y'_{n+1} + \beta_0 y'n)$$

<table>
<thead>
<tr>
<th>$l$</th>
<th>$\beta_2$</th>
<th>$\beta_1$</th>
<th>$\beta_0$</th>
<th>$\zeta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>49.7416</td>
</tr>
<tr>
<td>3</td>
<td>1.625</td>
<td>-0.75</td>
<td>0.125</td>
<td>82.7211</td>
</tr>
<tr>
<td>4</td>
<td>1.9166</td>
<td>-1.33</td>
<td>0.4166</td>
<td>100.0000</td>
</tr>
</tbody>
</table>

$k=4$

$$y_{n+4} = y_{n+3} + h(\beta_3 y'_{n+3} + \beta_2 y'_{n+2} + \beta_1 y'_{n+1} + \beta_0 y'n)$$

<table>
<thead>
<tr>
<th>$l$</th>
<th>$\beta_3$</th>
<th>$\beta_2$</th>
<th>$\beta_1$</th>
<th>$\beta_0$</th>
<th>$\zeta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30.845</td>
</tr>
<tr>
<td>3</td>
<td>1.6333</td>
<td>-0.8</td>
<td>0.2</td>
<td>-0.0333</td>
<td>54.3341</td>
</tr>
<tr>
<td>4</td>
<td>2.07954</td>
<td>-1.821</td>
<td>0.9053</td>
<td>-0.1628</td>
<td>85.7927</td>
</tr>
<tr>
<td>5</td>
<td>2.291666</td>
<td>-2.45833</td>
<td>1.54166</td>
<td>-0.37500</td>
<td>100.0000</td>
</tr>
</tbody>
</table>
Table 2.2. Optimal implicit formulae of Adams type and the values of $\zeta$ for the Adams-Moulton formulae

<table>
<thead>
<tr>
<th>$k=1$</th>
<th>$y_{n+1} = y_n = h(\beta_1 y_{n+1} + \beta_0 y'_n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k=2$</td>
<td>$y_{n+2} = y_{n+1} + h(\beta_2 y'<em>{n+2} + \beta_1 y'</em>{n+1} + \beta_0 y'_n)$</td>
</tr>
<tr>
<td>$k=3$</td>
<td>$y_{n+3} = y_{n+2} + h(\beta_3 y'<em>{n+3} + \beta_2 y'</em>{n+2} + \beta_1 y'_{n+1} + \beta_0 y'_n)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$k=1$</th>
<th>$\ell$</th>
<th>$\beta_1$</th>
<th>$\beta_0$</th>
<th>$\zeta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.5</td>
<td>0.5</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>0.5</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$k=2$</th>
<th>$\ell$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_0$</th>
<th>$\zeta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>92.5820</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.4375</td>
<td>0.625</td>
<td>-0.0625</td>
<td>97.5665</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.4166</td>
<td>0.666</td>
<td>-0.0833</td>
<td>100.000</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$k=3$</th>
<th>$\ell$</th>
<th>$\beta_3$</th>
<th>$\beta_2$</th>
<th>$\beta_1$</th>
<th>$\beta_0$</th>
<th>$\zeta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>80.5387</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.4333</td>
<td>0.65</td>
<td>-0.1</td>
<td>0.0166</td>
<td>84.6415</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.38762</td>
<td>0.75378</td>
<td>-0.1704</td>
<td>0.02904</td>
<td>95.7169</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.375</td>
<td>0.79166</td>
<td>-0.20833</td>
<td>0.04166</td>
<td>100.000</td>
<td></td>
</tr>
</tbody>
</table>
The rest of this section will be devoted to an explanation of how to use Table 2.1 and Table 2.2. We will not discuss which table or which value of \( k \) should be used. These choices depend on facts outside the scope of this paper. What we will devote our time to is discussing the choice of a value for \( \lambda \) given a particular \( k \).

Observe that if

\[
F(g) = 0 \text{ for all } g \in H^k_{\lambda},
\]

i.e., if

\[
g(kh) + \sum_{j=0}^{k-1} \alpha_j g(jh) - h \sum_{j=0}^{k} \beta_j g'(jh) = 0 \text{ for all } g \in H^k_{\lambda},
\]

then the difference equation

\[
\sum_{j=0}^{k} \alpha_j y_j = h \sum_{j=0}^{k} \beta_j f(x_j, y_j)
\]

would solve the differential equation

\[
y' = f(x, y)
\]

exactly, as long as \( y \in H^k_{\lambda} \). Unfortunately, we cannot guarantee that

\[
F(g) = 0 \text{ for all } g \in H^k_{\lambda}.
\]

What can be done is to use the optimal approximation, \( \phi_0 \), to \( F \). Then by definition

\[
\sup_{\| f \|_{\lambda} = 1} |F(f) - \phi_0(f)| \leq \sup_{\| f \|_{\lambda} = 1} |F(f) - \phi(f)| \quad \text{(2.40)}
\]
where $\phi \in (H^k_\lambda)^*$ is any element of $M^k_1$ of $M^k_2$, depending on whether we are dealing with implicit or explicit formulae. Thus $F - \phi_0$ minimizes the error in the worst case.

Consequently, the first step in using optimal difference formulae is to consider the initial value problem

$$y' = f(x, y), \ y(x_0) = y_0,$$

and decide to which space the solution belongs. If $y \in H^k_\lambda$, it may be advantageous to employ an optimal formula in $(H^k_\lambda)^*$. (The rest of this discussion will be limited to explicit formulae. The same results hold for implicit formulae but the indexing is slightly different.)

For example, if the difference formula is of order $k$ and the solution has $k+1$ or more derivatives, then $y \in H^k_{\lambda=k+1}$ and the optimal difference formula which is of degree $\lambda = k+1$ is the difference formula of maximum possible degree. We will denote this formula by $F$. (Note that this difference formula determines a linear functional on $H^k_{\lambda=k+1}$ so that $F$ may be considered either as a difference equation or as an element of $(H^k_{\lambda=k+1})^*$.) Therefore, for $y \in H^k_{\lambda=k+1}$, $F$ minimizes the worst error.

Suppose, however, that $y$ is only an element of $H^k_\lambda$, $\lambda \leq k+1$, and not of $H^k_{\lambda=k+1}$. Then $F(f)$ is not the optimal formula. In fact, if $(F - \phi_0)(f)$ is the optimal formula, (2.40) implies that

$$\sup_{\|f\|_{\lambda}=1} |(F - \phi_0)(f)| \leq \sup_{\|f\|_{\lambda}=1} |F(f)|.$$
Consequently
\[ |(F - \phi)(f)| \leq \sup |F(f)| \text{ for all } f \in H^k_{\chi}. \]
\[ \|f\|_{\chi} = 1 \]

We cannot say, however, that
\[ |(F - \phi)(f)| \leq |F(f)| \text{ for every } f \in H^k_{\chi}. \quad (2.41) \]

The opposite is true in many examples. However, for certain functions in $H^k_{\chi}$, (2.41) is true, i.e., in certain cases the optimal formula is better. Thus, if we wish to minimize the error in a worst case situation, the optimal difference formula is preferred.

The percentage correctness, $\zeta$, gives us the relation of the maximum possible error of $F - \phi$ to that of $F$, i.e.,
\[ \zeta = 100 \frac{\sup_{\|f\|_{\chi} = 1} |F(f) - \phi(f)|}{\sup_{\|f\|_{\chi} = 1} |F(f)|}. \]

Thus a formula $F - \phi$ for which $\zeta = 75\%$ has only $75\%$ the error as does $F$ (where this error is measured in a worst case situation for each formula).

Next, we will illustrate these ideas by two examples.

**Example 1.** Consider the differential equation
\[ y' = 2x. \]
Then $y = x^2$ and $y$ has infinitely many derivatives. Thus, one would expect a difference equation of high order to give the best answers.

Consider the difference equations
\[ y_{n+2} = y_{n+1} + y'_n \quad (2.42) \]
and

\[ y_{n+2} = y_{n+1} + \frac{3}{2}y'_{n+1} - \frac{1}{2}y'_n. \]  

(2.43)

These are optimal explicit formulae of order \( k = 2 \) and degree one and two respectively. Applying these equations to our problem, we see that according to (2.42)

\[ y(2) = y(1) + y'(1) \]
\[ = 1 + 2 \]
\[ = 3 \]

and according to (2.43)

\[ y(2) = y(1) + \frac{3}{2}y'(1) - \frac{1}{2}y'(0) \]
\[ = 1 + \frac{3}{2}2 - \frac{1}{2}0 \]
\[ = 4. \]

Thus, the difference formula of higher degree predicts the value of \( y(2) \) exactly while the difference formula of degree one shows an error.

**Example 2.** Consider the differential equation

\[ y'(x) = 2x - 2(x-1)H(x-1). \]

Then the solution of \( y(x) = x^2 - (x-1)^2H(x-1) \) and \( y(x) \) has a discontinuous second derivative. Using (2.42) to predict the value of \( y(2) \), we find

\[ y(2) = y(1) + y'(1) \]
\[ = 1 + 2 \]
\[ = 3. \]
Using (2.43), we find

\[ y(2) = y(1) + \frac{3}{2}y'(1) - \frac{1}{2}y'(1) \]
\[ = 1 + \frac{3}{2}2 - \frac{1}{2}0 \]
\[ = 4. \]

Since the exact value of \( y(2) \) is 3, the difference formula of degree one gives better answers than the formula of degree two.

Therefore, as these two examples point out, the differentiability of the solution determines which optimal solution can be expected to yield the minimum error. Also, since the differentiability of the right hand side of \( y' = f(x,y) \) determines the differentiability of \( y \), the optimal degree for a given differential equation can be predicted apriori.
III. AN OPTIMAL FINITE DIFFERENCE METHOD

Preliminary Notation

The notation listed here will be discussed more fully later. We list it here merely as a reference.

(1) A: \( Ay = (py')' - qy \) where \( q \in C^2[a,b], p \in C'[a,b], p \geq p_0 > 0, \) and \( q \geq 0. \)

(2) B: \( By = (py')' \).

(3) L: \( Ly = -qy. \)

(4) H: The Hilbert space in which we will be optimizing.

(5) \( H_h \): The subspace of \( H \) consisting of all functions which vanish at the points \( x_k = a + kh, \) \( k = 0,1,\ldots,n. \) Here \( h = \frac{b-a}{n} \), where \( n \) is some integer, and \( b \) and \( a \) are the end points of the interval over which we are seeking a solution to some differential equation.

(6) \( E_h \): An \( n+1 \) dimensional vector space.

(7) \( \psi_h \): The mapping from \( C \) (the space of continuous functions) into \( E_h \) defined by
\[ \psi_h f = (f(a), f(a+h), \ldots, f(b)). \]

(8) \( \psi_h^{-1} \): A mapping from \( E_h \) into \( H_h \), where, at least in the examples, \( \psi_h^{-1}(\vec{u}) \) is the piecewise linear function connecting the coordinates of the \( n+1 \) dimensional vector \( \vec{u} \).
(9) $P_h$: The projection mapping from $H$ into $H_h$.

(10) $P_h^*: The$ projection mapping from $H$ into $H_h^*$.

(11) $/\!/ /$: At different times, the symbol will denote

(a) absolute value

(b) maximum norm on $E_h$ or $C$, and

(c) the matrix norm induced by the maximum norm in $E_h$.

(12) $\|\|$: The norm on $H$.

(13) $\|\|_\ast$: The norm of a mapping from $H$ into $E_h$ defined by

$$
\|A_h f\|_\ast = \sup_{f \in H} \frac{|A_h f|}{\|f\|}.
$$

In this chapter, we will be discussing the problem of finding optimal numerical solutions to boundary-value problems. Consider the self-adjoint second-order differential equation

$$
(p(x)y'(x))' - q(x)y(x) = f(x) \quad (3.1)
$$

over the interval $[a,b]$ with the boundary conditions

$$
\alpha_1 y'(a) - \beta_1 y(a) = \gamma_1 \quad (3.2)
$$

$$
\alpha_2 y'(b) + \beta_2 y(b) = \gamma_2.
$$

Here, $p(x)$, $q(x)$, and $f(x)$ are assumed to satisfy on $[a,b]$ the conditions $p(x) \geq p_0 > 0$, $q(x) \geq 0$, $q \in C[a,b]$, $p \in C[a,b]$, $f \in C[a,b]$; $\alpha_i$ and $\beta_i$ are assumed to be non-negative numbers such that $\alpha_i + \beta_i > 0$,
i = 1, 2. (After the first example, we will be interested only in the boundary conditions \( y(a) = y(b) = 0 \).)

In what follows, we will require that solutions to (3.1) be unique. To see that the assumptions on (3.1) and (3.2) are sufficient to guarantee uniqueness of solutions, one may consult Keller [2, p. 11, 12].

We will use the method of finite differences to solve boundary-value problems. With this method, we will be concerned with the values of the solution \( y(x) \) at certain points over a given interval \([a, b]\). Let \( h = (b-a)/n \) where \( n \) is some integer. Define \( \bar{y}_h \) by

\[
\bar{y}_h = (y(a), y(a+h), y(a+2h), \ldots, y(b)).
\]

Solving by the method of finite differences will then consist of finding an \( n+1 \) dimensional vector \( \bar{y}^*_h \) whose components are close to the actual values of the solution to the differential equation at the indicated points, i.e., close to \( \bar{y}_h \). To do this, we will find matrices \( A_h, B_h \) of order \( (n+1) \), and an \( (n+1) \) dimensional vector \( g_h \) such that

\[
A_h \bar{y}_h - B_h g_h = c_h, \quad \| A_h^{-1} c_h \| = o(1) \text{ as } h \to 0, \tag{3.3}
\]

where \( A_h, B_h \) and \( g_h \) do not depend on \( \bar{y}_h \). Here, \( \| \| \) denotes some vector norm. The particular choice of this norm is of no importance to the general statement of the problem.

Let \( \bar{y}^*_h = (y_0, \ldots, y_n) \) represent the solution to the matrix equation

\[
A_h \bar{y}^*_h - B_h \bar{g}_h = \bar{0}. \tag{3.4}
\]

The values of \( \bar{y}^*_h \) will then approximate values of the solution to the given boundary-value problem, since subtracting (3.4) from (3.3) we get
\[
(y_h - \tilde{y}_h^*) = A_h^{-1}(\varepsilon_h).
\]

From (3.3) it then follows that \(y_h \rightarrow \tilde{y}_h^*\) in norm.

As an example of the method of finite differences, consider solving (3.1) by using the approximations

\[
p(x_{k+1/2})y'(x_{k+1/2}) \approx \frac{1}{h} \{y(x_{k+1}) - y(x_k)\} p(x_{k+1/2}), \text{ and}
\]

\[
(p(x)y'(x))'_{x=x_k} \approx \frac{1}{h} \{p(x_{k+1/2})y'(x_{k+1/2}) - p(x_{k-1/2})y'(x_{k-1/2})\}.
\]

Here we are using the notation \(x_k\) to mean the point \(x \in [a,b]\) such that \(x = a + kh\) where \(h = \frac{b-a}{n}\) for some integer \(n\) and \(x_{k+1/2}\) denotes the midpoint between \(x_k\) and \(x_{k+1}\). Making the above substitutions in (3.1), we obtain

\[
- p(x_{k-1/2})y(x_{k-1}) + \left( p(x_{k-1/2}) + p(x_{k+1/2}) + h^2 q(x_k) \right) y(x_k)
- p(x_{k+1/2})y(x_{k+1}) = -h^2 f(x_k).
\]

For the boundary conditions, we will use the approximations

\[
y'(x_0) \approx \frac{y(x_1) - y(x_0)}{h}
\]

and\[
y'(x_n) \approx \frac{y(x_n) - y(x_{n-1})}{h}.
\]

Making these substitutions in (3.2), we obtain
We can now represent (3.5) and (3.6) in a matrix equation of the form (3.4) as follows. Let

\[ A_h = \begin{bmatrix}
\alpha_1 + h\beta_1 & -\alpha_1 & 0 & \cdots & 0 \\
-p(\frac{x_{1/2}}{2}) & p(\frac{x_{1/2}}{2}) + p(\frac{x_3}{2}) + h^2 q(\frac{x_1}{2}) & -p(\frac{x_3}{2}) & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
& & & & -\alpha_2, \alpha_2 + h\beta_2
\end{bmatrix}, \]

\[ B_h = h^2 I, \] where \( I \) is the identity matrix, and

\[ \bar{g}_h = \begin{bmatrix}
f(x_0) \\
\vdots \\
f(x_n)
\end{bmatrix} \]

Thus the solution, \( \bar{y}_h \), to the equation

\[ A_h \bar{y}_h = B_h \bar{g}_h \]

will approximate the solution of the differential equation (3.1) with boundary conditions (3.2).
Since the purpose of this chapter is to construct optimal finite-difference methods, we will define optimality in the same way we have in previous sections, i.e., a numerical method will be said to be optimal if it minimizes the worst possible error. The mathematical framework which we will use to illustrate this concept is extremely involved. Nevertheless, this mathematical structure needs to be constructed to work with this intuitively simple definition of optimality.

In this section we will restrict (3.2) so that we will be considering the self-adjoint second-order differential equation

\[(p(x)y')' - q(x) y - f(x)\]  \hspace{1cm} (3.7)

with the boundary conditions

\[y(a) = y(b) = 0.\]  \hspace{1cm} (3.8)

To begin, let \(C\) denote the space of all continuous functions defined on the interval \([a,b]\), \(E_h\) the vector space of dimension \((n+1)\), and \(\psi_h\) the linear mapping of the space \(C\) into \(E_h\) defined by

\[\psi_h f = (f(a), f(a+h), \ldots, f(b-h), f(b)).\]

At times, we will designate \(\psi_h f\) by \(\tilde{f}_h\).

A finite difference method will be defined, in the same spirit as the definition which was used in the example, as the process of construction of matrices \(A_h\) and \(B_h\) such that the solution of the matrix equation

\[A_h \tilde{y}_h = h^2 B_h \tilde{f}_h\]  \hspace{1cm} (3.9)
will give values which will approximate the values of the exact solution, \( y(x) \), at the points \( x_k, k = 0, 1, \ldots, n \), where \( x_0 = a \) and \( x_n = b \). Since the \( h^2 \) factor will appear in all entries of \( B_h \) it is factored out of the matrix. The \( h^2 \) factor could, however, have been combined with the matrix \( B_h \) in order to give (3.9) the same representation as (3.4).

In the proofs, (3.4) will often be expressed in the form

\[
-1 \quad -1 
\]

\[ y_h = Z_h^{-1} f_h \]  

(3.10)

where \( Z_h^{-1} = h^2 A_h^{-1} B_h \). Let \( y = A^{-1} f \) represent the exact solution to the differential equation. Then \( \psi_h A^{-1} f \) represents values of the solution of the differential equation at the points \( x_k, k = 0, 1, \ldots, n \). Since \( y_h \) is the solution to (3.10), or equivalently (3.9),

\[
(\psi_h A^{-1} - Z_h^{-1} \psi_h) f 
\]

is the difference between the values of the exact and the approximate solutions of the differential equation. The suitability of the finite difference method (3.10) will be assessed on the basis of the magnitude of the norm of (3.11).

Let \( H \) be a Hilbert space whose elements are also elements of the space \( C \) of continuous functions (with maximum norm). Let \( H_h \) denote the subspace of \( H \) consisting of all functions which vanish at the points \( x_k = a + kh, k = 0, 1, \ldots, n \), \( H_h^1 \) the orthogonal complement of \( H_h \) in \( H \) and let \( P_{H_h^1} \) and \( P_{H_h} \) be the projectors of \( H \) on \( H_h^1 \) and \( H_h \) respectively. The norm and scalar product in \( H \) will be denoted by \( \| \cdot \| \) and \( ( \cdot, \cdot ) \) respectively.

We will now show that \( H_h^1 \) is a vector space of dimension at most \( n+1 \). Note that if \( f_1 \in H_h^1 \) and \( f_2 \in H_h^1 \) and if \( f_1 \) and \( f_2 \) assume the same
values at the points \( x_k, \ k = 0,1,\ldots,n \), then \( f_1 = f_2 \). This follows because \( f_1 - f_2 \in H_h \) and \( f_1 - f_2 \in H_h \) too, which implies that \( f_1 - f_2 = 0 \).

It is then clear that the set

\[
S_{n+1} = \{ f_k \mid f_k(x_i) = \delta_{k,i}, k,i = 0,1,\ldots,n \},
\]

where \( \delta_{k,i} \) is the Kronecker delta function, is a spanning set for \( H_h \).

Thus \( H_h \) can have dimension at most \( n+1 \). In what follows we will assume the \( H_h \) has dimension exactly \( n+1 \). (When we apply the general theory to a particular space \( H \), we will have to show that the dimension is actually \( n+1 \).) Also, since \( H_h \) is an \( n+1 \) dimensional vector space, \( H_h \) is isomorphic to the \( n+1 \) dimensional vector space \( E_h \). The map \( \psi_h \) will be an isomorphic mapping of \( H_h \) on \( E_h \). We will denote the inverse map of \( E_h \) onto \( H_h \) by \( \psi_h^{-1} \).

Next, we will define when a finite difference method will be said to be optimal.

**Definition.** A matrix \( O_{Z_h}^{-1} \) is the optimal matrix of a finite difference method for the problem (3.7) and (3.8) with respect to the space \( H \) if

\[
\sup_{f \in H} \frac{|(\psi_h A_h^{-1} - O_{Z_h}^{-1} \psi_h) f|}{\|f\|} = \inf_{Q \in H} \sup_{f \in H} \frac{|(\psi_h A_h^{-1} - Q^{-1} \psi_h) f|}{\|f\|}
\]

where the operation \( \inf \) refers to all matrices \( Q^{-1} \) of order \( n+1 \) and \( \| \cdot \| \) refers to the maximum norm in \( E_h \).

Note that this definition is the same type as we have used before, i.e., \( O_{Z_h}^{-1} \) is said to be optimal if it minimizes the worst possible error.
In order to show the existence of a particular matrix \( O_{Q}^{-1} \), which we will later prove to be optimal, let \( f_{i} \), \( i = 0,1,\ldots,n \), be a set of basis functions in \( H_{h}^{1} \). Then \( \psi_{h}f_{i}^{1} \) and \( \psi_{h}A^{-1}f_{i}^{1} \), \( i = 0,1,\ldots,n \), will both be sets of basis functions in \( E_{h}^{1} \). Here we are appealing to the linearity properties of \( \psi_{h} \) and \( A^{-1} \) and to the fact that if a function in \( H_{h}^{1} \) agrees with the zero function at the points \( x_{k} \), \( k = 0,1,\ldots,n \), then it is the zero function. From matrix theory, it follows (see [4, pp. 64, 80]) that there exists a linear map \( L \) such that

\[
L(\psi_{h}f_{i}) = \psi_{h}A^{-1}f_{i}, \quad i = 0,1,\ldots,n,
\]

and a matrix \( O_{Q}^{-1} \) of order \( (n+1) \) such that

\[
\psi_{h}A^{-1}f_{i} = O_{Q}^{-1}\psi_{h}f_{i}, \quad i = 0,1,\ldots,n.
\]

Therefore, we also have, for any \( g \in H_{h}^{1} \), that

\[
\psi_{h}A^{-1}g = O_{Q}^{-1}\psi_{h}g.
\]

We will show that the matrix \( O_{Q}^{-1} \) is optimal in

**Theorem 3.1.** The matrix \( O_{Q}^{-1} \) is an optimal matrix of the finite difference method for the problem (3.7) and (3.8) with respect to the space \( H \). Further, if \( O_{Z}^{-1} \) is an arbitrary optimal matrix (note that we have not claimed uniqueness of the optimal matrix \( O_{Q}^{-1} \)) then

\[
\sup_{f \in H} \frac{|(\psi_{h}A^{-1} - O_{Z}^{-1}\psi_{h})f|}{\|f\|} = \|\psi_{h}A^{-1}p_{h}\|^{*} \quad (3.12)
\]
where \[ \| \psi_h A^{-1} P_h \|_* = \sup_{f \in H} \frac{|(\psi_h A^{-1} P_h) f|}{\| f \|} \]

and \( \| \cdot \| \) denotes the maximum norm in \( E_h \).

**Proof.** Let \( g \in H \) and

\[
\alpha = \inf_{Q^{-1}} \sup_{f \in H} \frac{|(\psi_h A^{-1} - Q^{-1} \psi_h) f|}{\| f \|}
\]

Since \( Q^{-1} \psi_h g = 0 \), for \( g \in H \),

\[
\alpha = \inf_{Q^{-1}} \sup_{f \in H} \frac{|(\psi_h A^{-1} - Q^{-1} \psi_h) f|}{\| f \|} \geq \inf_{g \in H} \sup_{Q^{-1}} \frac{|(\psi_h A^{-1} - Q^{-1} \psi_h) g|}{\| g \|} \geq \sup_{g \in H} \frac{|\psi_h A^{-1} g|}{\| g \|} = \| \psi_h A^{-1} P_h \|_*
\]

Therefore \( \alpha \geq \| \psi_h A^{-1} P_h \|_* \). Let \( f \in H \), then \( Q^{-1} \psi_h f = 0 \) for any matrix \( Q^{-1} \). Consequently

\[
\| \psi_h A^{-1} - Q^{-1} \psi_h \|_* \]

\[
= \sup_{f \in H} \frac{|(\psi_h A^{-1} - Q^{-1} \psi_h) P_h f| + (\psi_h A^{-1} - Q^{-1} \psi_h) P_h f|}{\| f \|}
\]

\[
= \sup_{f \in H} \frac{|(\psi_h A^{-1} - Q^{-1} \psi_h) P_h f| + (\psi_h A^{-1} P_h f|}{\| f \|}
\]
Here we are using the decomposition of $f$ given by the Projection Theorem. By the definition of $Q^{-1}$, $\psi_h A^{-1} f = Q^{-1} \psi_h f$ for every $f \in H_h^1$. Thus follows
\[
\|\psi_h A^{-1} - Q^{-1} \psi_h\|_* = \sup_{f \in H} \frac{|(\psi_h A^{-1} - Q^{-1} \psi_h) P_h^1 + \psi_h A^{-1} P_h f|}{|f|}
\]
\[
= \sup_{f \in H} \frac{\|\psi_h A^{-1} P_h f\|}{\|f\|} = \|\psi_h A^{-1} P_h\|_* .
\]

Therefore, $\|\psi_h A^{-1} - Q^{-1} \psi_h\|_* = \|\psi_h A^{-1} P_h\|_* \leq \alpha$.

But $\alpha$ is the infimum over all matrices $Q^{-1}$.

Therefore,
\[
\|\psi_h A^{-1} - Q^{-1} \psi_h\|_* = \alpha . \quad (3.13)
\]

Thus, $Q^{-1}$ is an optimal matrix. Further, since (3.13) implies that
\[
\|\psi_h A^{-1} P_h\|_* = \alpha ,
\]
the definition of an optimal matrix gives us (3.12).

The convergence of the optimal finite-difference method is insured by
Theorem 3.2. Let the space $H$ have the property

$$\lim_{h \to 0} \| \psi_h A^{-1} p_h \| = 0$$

(3.14)

and $o_{Z_h^{-1}}$ denote the sequence of optimal matrices of the finite-difference method for the problem (3.7) and (3.8) with respect to the space $H$. Then

$$\lim_{h \to 0} \| \psi_h A^{-1} - o_{Z_h^{-1}} \psi_h \| = 0.$$  

(3.15)

The proof of this theorem is an immediate consequence of Theorem 3.1, since

$$\| \psi_h A^{-1} - o_{Z_h^{-1}} \psi_h \| = \| \psi_h A^{-1} p_h \|.$$  

The construction of optimal matrices is sometimes difficult. Therefore, we will consider the asymptotically optimal matrices given in the following

Definition. A sequence of matrices $Z_h^{-1}$, $h = (b-a)/n$, $n = 1, 2, \ldots$, which characterizes a finite-difference method for the problem (3.7) and (3.8) with respect to the space $H$ subject to (3.15), will be said to be an asymptotically optimal sequence of matrices with respect to $H$, if

$$\| \psi_h A^{-1} p_h - Z_h^{-1} \psi_h \| = o(\| \psi_h A^{-1} p_h \|).$$  

(3.16)
By appealing to the projection decomposition for $A^{-1}$ and the reverse triangle inequality, it is clear that (3.16) implies

$$\|\psi_h A^{-1} - Z_h^{-1} \psi_h\|_* \leq \|\psi_h A^{-1} P_h\|_* + o(\|\psi_h A^{-1} P_h\|_*).$$  (3.17)

Thus, if the space $H$ satisfies the hypothesis of Theorem 3.2, we can again prove convergence of the finite-difference method as $h \to 0$. The next theorem is a statement of this result.

**Theorem 3.3.** Let $Z_h^{-1}$, $h = (b-a)/n$, $n = 1, 2, \ldots$, denote a sequence of matrices which is asymptotically optimal for the problem (3.7) and (3.8) with respect to $H$. Then

$$\|\psi_h A^{-1} - Z_h^{-1} \psi_h\|_* \to 0.$$

Observe that this theorem is the analogue for asymptotically optimal matrices of Theorem 3.2. Its proof is an immediate consequence of (3.17).

In the next theorem, 3.4, we will give conditions under which a sequence of matrices can be shown to be optimal. As an example of an interpretation of Theorem 3.4, think of the operators $A$, $B$, and $L$ as defined in the section on preliminary notation.

**Theorem 3.4.** Given a boundary-value problem for the self-adjoint, second-order equation (3.7) with boundary conditions $y(a) = y(b) = 0$ which will be written symbolically in the form

$$Ay = f,$$  (3.18)
consider a related boundary-value problem

\[ By = f \]  

(3.19)

where \( A = B + L \) and \( B \) is a self-adjoint second-order equation of the type (3.7) with the same boundary conditions. Note that \( L \) is well defined once \( A \) and \( B \) are given. Suppose \( A^{-1} \), \( LA^{-1} \), \( B^{-1} \), \( B^{-1}L \), and \( (I + B^{-1}L)^{-1} = A^{-1}B \) are continuous mappings of \( H \) into itself, and let

\[ o_{B^{-1}} \]

be a sequence of matrices for which \( \psi_{B^{-1}}p_{h} = o_{B^{-1}}\psi_{h} \) and \( \phi_{h} \) be a sequence of matrices such that \( I + \phi_{B^{-1}L}o_{B^{-1}} \) is non-singular. Finally, let

\[ \| \psi_{B^{-1}}(\psi_{h}^{-1}\psi_{h} - P_{h}^{-1}L)A^{-1}\|_{\ast} = o(\| \psi_{h}A^{-1}p_{h} \|_{\ast}) \]  

(3.20)

\[ \| \psi_{B^{-1}}p_{h}LA^{-1}\|_{\ast} = o(\| \psi_{h}A^{-1}p_{h} \|_{\ast}) \]  

(3.21)

\[ | \psi_{B^{-1}}(\psi_{h}^{-1}\psi_{h} - P_{h}^{-1}L)(I + B^{-1}L)^{-1}\psi_{h}^{-1} | = o(1) \]  

(3.22)

\[ | \psi_{B^{-1}}p_{h}L(I + B^{-1}L)^{-1}\psi_{h}^{-1} | = o(1) \]  

(3.23)

\[ | \psi_{h}(I + B^{-1}L)^{-1}\psi_{h}^{-1} | = o(1). \]  

(3.24)

Then, for sufficiently small values of \( h \), the matrices

\[ (I + \phi_{B^{-1}L}o_{B^{-1}}) \]

form a sequence which is asymptotically optimal for the problem (3.18) with respect to \( H \).

We shall prove this theorem by use of...
**Lemma 3.4.** Let $V$ be a linear mapping of the Banach space $X$ into the Banach space $Y$ and assume that there exists, for every $y \in Y$, an $x \in X$ such that

$$
||x|| \leq N ||y||
$$

and

$$
||V(x) - y|| \leq q ||y||, \quad 0 \leq q \leq 1.
$$

Then, for arbitrary $y \in Y$, the equation $V(x) = y$ has the solution $x \in X$ such that

$$
||x|| \leq \frac{N}{1 - q} ||y||.
$$

**Proof.** Choose $y \in Y$. Then there exists an $x_0$ such that

$$
||x_0|| \leq N ||y|| \quad \text{and} \quad ||y - V(x_0)|| \leq q ||y||.
$$

Since $y - V(x_0) \in Y$, there exists an $x_1$ such that

$$
||x_1|| \leq N ||y - V(x_0)|| \quad \text{and} \quad ||y - V(x_0) - V(x_1)|| \leq q ||y - V(x_0)||
$$

so

$$
||y - V(x_0) - V(x_1)|| \leq q^2 ||y||.
$$

Using this process repeatedly, we find that there exists a sequence $x_0, x_1, \ldots, x_n$, where $n$ is arbitrary, such that

$$
||x_n|| \leq N ||y - \sum_{i=0}^{n-1} V(x_i)|| \quad \text{and} \quad ||y - \sum_{i=0}^{n} V(x_i)|| \leq q^{n+1} ||y||.
$$

(3.25)
Appealing to the fact that $V$ is linear,

$$
\|y - \sum_{i=0}^{n} V(x_i)\| = \|y - V(\sum_{i=0}^{n} x_i)\|
$$

so

$$
\|y - V(\sum_{i=0}^{n} x_i)\| \leq q^{n+1} \|y\|.
$$

Since $X$ and $Y$ are Banach spaces,

$$
\lim_{n \to \infty} \|y - V(\sum_{i=0}^{n} x_i)\| \leq \lim_{n \to \infty} q^{n+1} \|y\| = 0,
$$

which implies that

$$
y = V(x) \text{ where } x = \sum_{i=0}^{\infty} x_i.
$$

Now from (3.25) it follows that

$$
\|x_j\| \leq N\|y - \sum_{i=0}^{j-1} V(x_i)\| \leq Nq^j \|y\|.
$$

Therefore,

$$
\|x_0 + x_1 + \ldots + x_n\| \leq \|x_0\| + \ldots + \|x_n\| \leq \sum_{i=0}^{n} Nq^i \|y\|,
$$

and

$$
\|x\| = \lim_{n \to \infty} \|\sum_{i=0}^{n} x_i\| \leq \lim_{n \to \infty} \sum_{i=0}^{n} Nq^i \|y\|.
$$

Taking the limit of the geometric series, we conclude that the equation

$V(x) = y$ has the solution $x \in X$ such that

$$
\|x\| \leq \frac{N}{1 - q} \|y\|.
$$
We will now give the

Proof of Theorem 3.4.

Part I. Let \( C_h = \psi_h A^{-1}_h P_h - (I + o_B^{-1}_h L_h)^{-1} o_B^{-1}_h \psi_h \).

Part I of this proof will consist of getting a bound on \( |C_h|_F \). Since \( A = B + L, B^{-1} A = I + B^{-1} L \), so that

\[
A^{-1} = (I + B^{-1} L)^{-1} B^{-1}.
\]

Using the fact that \( o_B^{-1}_h \psi_h = \psi_h B^{-1}_h \psi_h \), we also find

\[
o_B^{-1}_h = \psi_h B^{-1}_h \psi_h^{-1} = \psi_h B^{-1}_h \psi_h^{-1},
\]

and

\[
C_h = \psi_h (I + B^{-1} L)^{-1} B^{-1}_h P_h - (I + o_B^{-1}_h L_h) \psi_h B^{-1}_h P_h. \tag{3.27}
\]

Therefore,

\[
C_h = \{\psi_h (I + B^{-1} L)^{-1} - (I + o_B^{-1}_h L_h) \psi_h \} B^{-1}_h P_h
\]

\[
= (I + o_B^{-1}_h L_h)^{-1} \{ (I + o_B^{-1}_h L_h) \psi_h - \psi_h (I + B^{-1} L) \} (I + B^{-1} L)^{-1} B^{-1}_h P_h
\]

\[
= (I + o_B^{-1}_h L_h)^{-1} \{ o_B^{-1}_h L_h \psi_h - \psi_h B^{-1} L \} (I + B^{-1} L)^{-1} B^{-1}_h P_h.
\]

Hence, from (3.26),

\[
C_h = (I + o_B^{-1}_h L_h)^{-1} \{ \psi_h B^{-1}_h \psi_h^{-1} \psi_h - \psi_h B^{-1} L \} A^{-1}_h P_h
\]

\[
= (I + o_B^{-1}_h L_h)^{-1} \psi_h B^{-1}_h \{ \psi_h^{-1} \psi_h - L \} A^{-1}_h P_h.
\]
Thus,
\[
C_h = (I + O_{B_{h_{h_{h}}}^{-1}})^{-1} \psi_{B_{h_{h}}}^{-1} \{ \psi_{B_{h_{h}}}^{-1} \psi_{h_{h_{h}}} \psi_{P_{h}} - \psi_{P_{h}} - P_{h} \} A_{P_{h}}^{-1}
\]
\[
= (I + O_{B_{h_{h}}}^{-1})^{-1} \{ \psi_{B_{h_{h}}}^{-1} (\psi_{h_{h_{h}}}^{-1} \psi_{P_{h}} - P_{h}) A_{P_{h}}^{-1} - (\psi_{B_{h_{h}}}^{-1})(P_{h} A_{P_{h}}^{-1}) \} P_{h}.
\]

Thus,
\[
\|C_h\|_* \leq \|I + O_{B_{h_{h}}}^{-1}\|_* \|\psi_{B_{h_{h}}}^{-1} (\psi_{h_{h_{h}}}^{-1} \psi_{P_{h}} - P_{h}) A_{P_{h}}^{-1}\|_*
\]
\[+ \|\psi_{B_{h_{h}}}^{-1} P_{h} A_{P_{h}}^{-1}\|_* \).
\]

Part II. Next, we will prove that \(|(I + O_{B_{h_{h}}}^{-1})^{-1}|\) is a uniformly bounded sequence of numbers. This fact will be sufficient to finish the proof of Theorem 3.3, for if \(|(I + O_{B_{h_{h}}}^{-1})^{-1}| \leq M\), for some \(M\) and all \(h\), then from the hypothesis of Theorem 3.4, i.e., from assumptions (3.20) and (3.21),
\[
\|C_h\|_* \leq M \{ \|\psi_{B_{h_{h}}}^{-1} P_{h}\|_* + \|\psi_{B_{h_{h}}}^{-1} P_{h}\|_* \},
\]
which implies that
\[
\lim_{h \to 0} \|\psi_{B_{h_{h}}}^{-1} P_{h} \cdot (I + O_{B_{h_{h}}}^{-1})^{-1} O_{B_{h_{h}}} \psi_{B_{h_{h}}}^{-1} \|_* = 0.
\]

Consequently, we will then have shown that
\[
(I + O_{B_{h_{h}}}^{-1})^{-1} O_{B_{h_{h}}}^{-1}
\]
is an asymptotically optimal sequence of matrices with respect to \(H\).
We will show that
\[
|(I + oB_{-1}h_{-1})^{-1}| = \sup_{x \in E_h} \frac{|(I + oB_{-1}h_{-1})^{-1}x|}{|x|}
\]
is a uniformly bounded sequence of numbers by appealing to Lemma 3.4.

On the basis of Lemma 3.4 it will be sufficient to show that
\[
|\psi_h(I + B^{-1}L)^{-1} \psi^{-1}_h| \leq N \tag{3.28}
\]
and
\[
|(I + oB_{-1}h_{-1}) \psi_h(I + B^{-1}L)^{-1} \psi^{-1}_h - I| \leq q < 1. \tag{3.29}
\]

To see why these inequalities imply the desired conclusion, let
\[
y \in E_h,
\]
\[
x = (\psi_h(I + B^{-1}L)^{-1} \psi^{-1}_h)y,
\]
and
\[
V = I + oB_{-1}h_{-1}.
\]

Then, (3.28) and (3.29) imply that for \( y \in E_h \) there exists an
\[x = (\psi_h(I + B^{-1}L)^{-1} \psi^{-1}_h)y\]
such that
\[
|x| \leq N |y| \quad \text{and} \quad |V(x) - y| \leq q|y|.
\]

From Lemma 3.4 it will then follow that the equation \( V(x) = y \) has the solution \( x \in E_h \) such that
\[
|x| \leq \frac{N}{1 - q} |y|.
\]

Thus the operator \( V^{-1} \) satisfies the relation
\[
\frac{|x|}{|y|} = \frac{|V^{-1}(y)|}{|y|} = \frac{|(I + oB_{-1}h_{-1})^{-1}(y)|}{|y|} \leq \frac{N}{1 - q},
\]
for all \( y \in E_h, y \neq 0 \). Hence
\[
|\left(I + o_B^{-1}_{h}L_h\right)^{-1}| = \sup_{y \in E_h} \frac{|(I + o_B^{-1}_{h}L_h)y|}{|y|} < \frac{N}{1 - q}.
\]

Therefore, if we can show that (3.28) and (3.29) are correct, we will have proven that \(|(I + o_B^{-1}_{h}L_h)^{-1}|\) is uniformly bounded by \(\frac{N}{1 - q}\).

Inequality (3.28) is clearly a part of our hypothesis since (3.24) states that
\[
|\psi_h(I + B^{-1}L)^{-1}\psi^{-1}_h| = O(1).
\]

We will now show that (3.29) has also been hypothesized.

\[
(I + o_B^{-1}_{h}L_h)\psi_h(I + B^{-1}L)^{-1}\psi^{-1}_h - I
\]
\[
= (I + o_B^{-1}_{h}L_h)\psi_h(I + B^{-1}L)\psi^{-1}_h - \psi_h(I + B^{-1}L)(I + B^{-1}L)^{-1}\psi^{-1}_h
\]
\[
= \{(I + o_B^{-1}_{h}L_h)\psi_h - \psi_h(I + B^{-1}L)\} (I + B^{-1}L)^{-1}\psi^{-1}_h.
\]

Using the facts, in order, that
\[
I\psi_h = \psi_h,
\]
\[
o_B^{-1}_{h} = \psi_B^{-1}\psi_h
\]
and
\[
L = P_hL + P^1hL,
\]
we find
\[
\{(I + o_B^{-1}_{h}L_h)\psi_h - \psi_h(I + B^{-1}L)\} (I + B^{-1}L)^{-1}\psi^{-1}_h
\]
\[
= o_B^{-1}_{h}L_h\psi_h - \psi_hB^{-1}L\psi^{-1}_h.
\]
\[ = \{\psi_h B^{-1} \psi_h^{-1} \} (I + B^{-1}L)^{-1} \psi_h^{-1} \]
\[ = \{\psi_h B^{-1} (\psi_h^{-1} \psi_h^{-1} - \frac{1}{h} P_h L) - \psi_h B^{-1} P_h L\} (I + B^{-1}L)^{-1} \psi_h^{-1} . \]

Therefore, from the triangle inequality and the assumptions (3.22) and (3.23),
\[
| (I + \theta B^{-1} L)^{-1} \psi_h (I + B^{-1}L)^{-1} \psi_h^{-1} - I | \\
\leq | \psi_h B^{-1} (\psi_h^{-1} \psi_h^{-1} - \frac{1}{h} P_h L) (I + B^{-1}L)^{-1} \psi_h^{-1} | \\
+ | (\psi_h B^{-1} P_h L) (I + B^{-1}L)^{-1} \psi_h^{-1} | \leq o(1) + o(1). \]

Consequently, for sufficiently small \( h \), inequality (3.28) holds.

We have now shown that \( \| C_h \| \leq \frac{N}{1 - q} \{ o(1) + o(1) \} \). This implies that for sufficiently small values of \( h \) the set of matrices \( (I + \theta B^{-1} L)^{-1} \theta B^{-1} \) forms a sequence which is asymptotically optimal for the problem (3.18) with respect to \( H \). Thus, we have completed the proof of Theorem 3.4.

We shall now discuss applications of the preceding theory. Let \( H \) be the Hilbert space \( W^{(1)}_2 \) of all continuous functions, the first derivatives of which are square integrable in \([a,b] \) with the scalar product
\[
(u,v) = \int_a^b u'(x)v'(x) \, dx + u(a)v(a). \]
By the derivative of a function \( h(x) \) we mean a function \( g(x) \) for which

\[
    h(x) = \int_a^x g(x) \, dx + c.
\]

Note that there are no equivalence classes in \( W_2^{(1)} \) since

\[
    0 = (u-v,u-v) = \int_a^b (u'(x) - v'(x))^2 \, dx + (u(a) - v(a))^2
\]

implies that \( u(x) = v(x) \) for all \( x \in [a,b] \). Also observe that the functions in \( W_2^{(1)} \) are absolutely continuous and have first derivatives which belong to \( L^2 \). Thus, it is clear that \( W_2^{(1)} \) is a subset of \( C \); however, the topologies in \( W_2^{(1)} \) and \( C \) are not equivalent.

Next, consider the structure of the space \( H_h \). Let \( g \in H_h \) and let \( z(x) \) be an arbitrary function with compact support in \( (x_i,x_{i+1}) \) with square integrable first derivatives. Then \( z \in H_h \), and, by definition,

\[
    (g,z,) = 0
\]

which implies that

\[
    \int_a^b g'(x)z'(x) \, dx + g(a)z(a) = \int_{x_{i+1}}^{x_{i+1}} g'(x)z'(x) \, dx = 0.
\]

Thus

\[
    \int_{x_i}^{x_{i+1}} g'(x)z'(x) \, dx = \int_{x_i}^{x_{i+1}} g''(x)z(x) \, dx = 0.
\]

Since \( z(x) \) was arbitrary, this implies that \( g''(x) = 0 \). Therefore \( g(x) \) is piecewise linear. Since \( g(x) \) is continuous, it is fully
defined by its values at the points \( x_i, \ i = 0, 1, \ldots, n \). Conversely, if \( g(x) \) is continuous and linear in every interval \((x_i, x_{i+1})\), then for every \( v \in \mathcal{H}_h\),

\[
(g, v) = \int_a^b g'v' \, dx + g(a)v(a) = \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} c_i v'(x) \, dx
\]

where \( c_i = g'(x) \) over the interval \((x_i, x_{i+1})\). Thus

\[
(g, v) = \sum_{j=0}^{n-1} c_j [v(x_{i+1}) - v(x_i)]
\]

and since \( v \in \mathcal{H}_h \), \( v(x_n) = 0 \) for all \( n \), so that

\[
(g, v) = 0.
\]

Therefore, \( \mathcal{H}_h \) consists of those functions which are linear over the intervals \((x_i, x_{i+1})\). Consequently, it is evident that the set of piecewise linear functions

\[
B = \{ f_j(x) \mid f_j \in \mathcal{H}_h, f_j(x_i) = \delta_{i,j}, \ i, j = 0, 1, \ldots, n \}
\]

forms a basis for \( \mathcal{H}_h \) and that \( \mathcal{H}_h \) has dimension \( n+1 \). We have now shown that the space \( \mathcal{W}_2^{(1)} \) provides the proper setting in which to apply our previous theory (in particular, \( \mathcal{W}_2^{(1)} \) satisfies the conditions of the two paragraphs which follow (3.11)).

Let

\[
Ay = (p(x)y')' - q(x)y = f(x)
\]

(3.30)
and assume that $p(x) \geq p_0 > 0$, $q(x) \geq 0$, $p \in C^1[a,b]$, and $q \in C^2[a,b]$.

Let $f \in H_h$. Then $f(x_i) = 0$, $i = 0, 1, \ldots, n$. Therefore, for $x_i \leq x \leq x_{i+1}$,

$$f(x) = \int_{x_i}^x f'(t) \, dt = \int_{x_i}^x f'(t) \, dt.$$  

Since $f'(t)$ and 1 are elements of $L^2$, the Schwartz inequality implies that

$$|f(x)| \leq \int_{x_i}^x |f'(t)| \, dt \leq \left\{ \int_{x_i}^x (f'(t))^2 \, dt \right\}^{1/2} \left\{ \int_{x_i}^x 1^2 \, dt \right\}^{1/2}$$

$$\leq \left\{ \int_{x_i}^x (f'(t))^2 \, dt + (f(a))^2 \right\}^{1/2} \sqrt{h}$$

$$\leq \|f\|_{W^1_2} \sqrt{h}. \quad (3.31)$$

Thus, it follows that the Hilbert space $W^1_2$ satisfies (3.14).

Consider first the case $q(x) = 0$, i.e., the boundary-value problem

$$By = (p(x)y')' = f(x), \quad y(a) = y(b) = 0. \quad (3.32)$$

In order to define an optimal matrix $o^{-1}_{B_h}$ such that

$$\psi B^{-1}_h f = o^{-1}_{B_h} \psi f, \quad f \in H_h,$$

we will appeal to one of Marcuk's identities. This result will be paraphrased from Babuska [1, p. 138].
Let $y(x)$ be a solution to the self-adjoint second-order equation

$$(p(x)y'(x))' - q(x)y(x) = f(x)$$

over the interval $[a,b]$ for the boundary conditions

$$\alpha_1 y'(a) - \beta_1 y(a) = \gamma_1$$
$$\alpha_2 y'(b) + \beta_2 y(b) = \gamma_2$$

where $p(x)$, $q(x)$, $f(x)$ are assumed to be bounded functions, which satisfy on $[a,b]$ the conditions $p(x) \geq p_0 > 0$, $q(x) \geq 0$, and $\alpha_1, \beta_1$ non-negative numbers such that $\alpha_1 + \beta_1 > 0$, $i = 1,2$; let $a = x_{k-1} < x_{k-(1/2)} < x_k < x_{k+(1/2)} \leq x_{k+1} \leq b$ be given points over the interval $[a,b]$; then one has the identity

$$\begin{align*}
&\frac{y(x_{k+1}) - y(x_k)}{\int_{x_k}^{x_{k+1}} (1/p(x)) \, dx} - \frac{y(x_{k}) - y(x_{k-1})}{\int_{x_{k-1}}^{x_k} (1/p(x)) \, dx} - \int_{x_{k-1}}^{x_{k+(1/2)}} (qy + f) \, dx \\
&- \frac{1}{\int_{x_k}^{x_{k+1}} (1/p(x)) \, dx} \int_{x_k}^{x_k+(1/2)} \left\{ (1/p(x)) \, dx \right\} \int_{x_k}^{x} (qy + f) \, dt \, dx \\
&+ \frac{1}{\int_{x_{k-1}}^{x_k} (1/p(x)) \, dx} \int_{x_{k-1}}^{x_k} \left\{ (1/p(x)) \, dx \right\} \int_{x_{k}-(1/2)}^{x} (qy + f) \, dt \, dx.
\end{align*}$$
This identity applies to the boundary-value problem which we are considering.

It has been shown that if $f \in H^1_h$, then $f$ is piecewise linear. We now claim that the solution of (3.32), for $f \in H^1_h$, satisfies the system of equations

$$p_{k-(1/2)}y(x_{k-1}) - (p_{k-(1/2)} + p_{k+(1/2)})y(x) + p_{k+(1/2)}y(x_{k+1})$$

$$= h^2 [\gamma_{k-1} f(x_{k-1}) + \gamma_k f(x_k) + \gamma_{k+1} f(x_{k+1})],$$

$k = 1, 2, \ldots, n-1, y(x_0) = y(x_n) = 0,$

(3.33)

where

$$p_{k-(1/2)} = \frac{1}{\int_{x_{k-1}}^{x_k} \frac{1}{p(x)} dx}, \quad p_{k+(1/2)} = \frac{1}{\int_{x_k}^{x_{k+1}} \frac{1}{p(x)} dx},$$

$$\gamma_k = (3/4) + \left(\frac{p_{k+(1/2)}}{2h^2}\right) \int_{x_k}^{x_{k+1}} \frac{1}{p(x)} (x-x_{k-(1/2)}) \left(1 + \frac{h}{x-x_{k-(1/2)}}\right) dx,$$

$$- \left(\frac{p_{k-(1/2)}}{2h^2}\right) \int_{x_{k-1}}^{x_k} \frac{1}{p(x)} (x-x_{k-(1/2)}) \left(1 + \frac{h}{x-x_{k-(1/2)}}\right) dx,$$

$$\gamma_{k-1} = (1/8) - \left(\frac{p_{k-(1/2)}}{2h^2}\right) \int_{x_{k-1}}^{x_k} \frac{1}{p(x)} (x-x_{k-(1/2)}) \left(1 - \frac{h}{x-x_{k-(1/2)}}\right) dx,$$

$$\gamma_{k+1} = (1/8) + \left(\frac{p_{k+(1/2)}}{2h^2}\right) \int_{x_k}^{x_{k+1}} \frac{1}{p(x)} (x-x_{k+(1/2)}) \left(1 + \frac{h}{x-x_{k+(1/2)}}\right) dx.$$
In order to verify the claim that the solution of (3.32) satisfies the system of equations (3.33), for \( f \in \mathcal{H}_h^1 \), we will solve Marcuk's identity for \( f \in \mathcal{H}_h^1 \). Since \( q(x) = 0 \), Marcuk's identity can be reduced to the system

\[
\begin{align*}
p_{k-(1/2)}y(x_{k-1}) - (p_{k-(1/2)} + p_{k+(1/2)})y(x_k) + p_{k+(1/2)}y(x_{k+1}) \\
= h \int_{x_{k-(1/2)}}^{x_{k+(1/2)}} f(x) \, dx + p_{k+(1/2)} \int_{x_k}^{x_{k+1}} \{ \frac{1}{p(x)} \int_{x_k}^{x} f(t) \, dt \} \, dx \\
- p_{k-(1/2)} \int_{x_{k-1}}^{x_k} \{ \frac{1}{p(x)} \int_{x_{k-1}}^{x} f(t) \, dt \} \, dx.
\end{align*}
\]

Observe that for \( f \in \mathcal{H}_h^1 \),

\[
f(x) = f(x_{k-1}) \frac{(x-x_k)}{h} + (f(x_k) \frac{(x-x_{k-1})}{h})
\]

\[(3.34)\]

for \( x_{k-1} \leq x \leq x_k \), and

\[
f(x) = f(x_k) \frac{(x_{k+1}-x)}{h} + f(x_{k+1}) \frac{(x-x_k)}{h}
\]

\[(3.35)\]
for $x_k \leq x \leq x_{k-1}$. Therefore

$$\int_{x_{k+1/2}}^{x} f(t) \, dt = \frac{1}{2}(x-x_{k+1/2})[f(x) + f(x_{k+1/2})], \quad (3.36)$$

$x_{k+1/2} \leq x \leq x_{k+1}$, which is just the area of a trapezoid. Consequently, from (3.35),

$$\int_{x_{k+1/2}}^{x} f(t) \, dt = \frac{1}{2}(x-x_{k+1/2}) \left\{ f(x_k) \left[ \frac{1}{2} - \frac{(x-x_{k+1/2})}{h} \right] 
+ f(x_{k+1}) \left[ \frac{1}{2} + \frac{(x-x_{k+1/2})}{h} \right] \right\} \right\} + (1/2) \left[ f(x_{k+1}) + f(x_k) \right]

= \frac{1}{2}(x-x_{k+1/2}) \left\{ f(x_k) \left[ 1 - \frac{(x-x_{k+1/2})}{h} \right] 
+ f(x_{k+1}) \left[ 1 + \frac{(x-x_{k+1/2})}{h} \right] \right\}.

Similarly (using the formula for the area of a trapezoid and (3.34)), it can be shown that

$$\int_{x_{k-1/2}}^{x} f(t) \, dt = \frac{1}{2}(x-x_{k-1/2}) \left\{ f(x_{k-1}) \left[ 1 - \frac{(x-x_{k-1/2})}{h} \right] 
+ f(x_k) \left[ 1 + \frac{(x-x_{k-1/2})}{h} \right] \right\} \left\{ x_{k-1/2} \leq x \leq x_k \right\}. \quad (3.37)$$
Using the same techniques used above, we find that

\[
\int_{x_{k-(1/2)}}^{x_{k+(1/2)}} f(x) \, dx = \frac{1}{2}((1/2)h)(f(x_{k-(1/2)}) + f(x_k))
\]

\[+ \frac{1}{2}((1/2)h)(f(x_{k+(1/2)}) + f(x_k))
\]

\[= \frac{h}{4} \left\{ f(x_{k-1}) + f(x_k) \right\} + \frac{h}{2} \left\{ f(x_k) + f(x_{k+1}) \right\} + f(x_k).
\]

\[= \frac{h}{8}f(x_{k-1}) + \frac{3h}{4}f(x_k) + \frac{h}{8}f(x_{k+1}). \quad (3.38)
\]

Combining (3.36), (3.37), and (3.38), we see that, for \( f \in \mathcal{H}_h \), the solution of (3.32) satisfies the system (3.33).

Therefore, if \( Ay = f \), then the vectors \( \bar{y} \) and \( \bar{f} \) satisfy the matrix equation

\[ -A_h \bar{y} = -h^2 \mathcal{G}_h \bar{f}, \]

where

\[
-A_h = \begin{bmatrix}
1, & 0, & 0, & 0, & 0, & 0, & 0, & 0 \\
-p_{1/2}, & p_{1/2}, & p_{3/2}, & -p_{3/2}, & 0, & 0, & 0, & 0 \\
0, & -p_{3/2}, & p_{3/2}, & p_{5/2}, & -p_{5/2}, & 0, & 0, & 0 \\
0, & 0, & 0, & 0, & 0, & 0, & 0, & 1
\end{bmatrix}, \quad (3.39)
\]
and

\[ G_h = \begin{bmatrix}
0, 0, 0, 0, 0, \\
\gamma_0^{(1)}, \gamma_1^{(1)}, \gamma_2^{(1)}, 0, \\
0, \gamma_2^{(1)}, \gamma_2^{(1)}, \gamma_3^{(1)}, \\
0, 0, 0, 0, 0, \\
\gamma_n^{(n-1)}
\end{bmatrix}. \quad (3.40) \]

Thus, if

\[-\frac{1}{h^2} \frac{\partial^2}{\partial x^2} g \big|_{x=0} = -h^{-1} \gamma_n^{(n-1)} g, \quad (3.41)\]

then, for \( g \in \mathcal{H}_h \),

\[ \psi_h A^{-1} g = o_B^{-1} \psi_h g, \]

which implies that \( o_B^{-1} \) is an optimal matrix of the finite difference method for the problem (3.32). Note that \( o_B^{-1} \) is well defined since \( A_h^{-1} \) is an irreducibly diagonally dominant matrix and thus has an inverse (see [8, p. 23]). Therefore, the method of finite differences, defined by the system (3.33), is the optimal method for the solution of (3.32) for this choice of the Hilbert space \( \mathcal{H} = W_2^{(1)} \), and since (3.31) implies that the space \( \mathcal{H} \) satisfies the hypothesis of Theorem 3.2, the convergence of this optimal method is ensured by Theorem 3.2, i.e.,

\[ \lim_{h \to 0} \| \psi_h A^{-1} - o_B^{-1} \psi_h \|_* = 0. \]
Next, consider the construction of an asymptotically optimal method for the general case of (3.7) and (3.8) when \( q(x) \geq 0 \). In accordance with Theorem 3.4, decompose the operator \( A \) in (3.30) in the form \( A = B + L \), where

\[
B y = (p(x)y')',
\]

\[
L y = -q(x)y.
\]

As a preliminary result, we will prove

**Lemma 3.5.** If \( Ay = f \), where \( y \) satisfies the boundary conditions \( y(a) = y(b) = 0 \), then

\[
|y| \leq C |f|
\]

where \( | \cdot | \) denotes the maximum norm and \( C \) is some constant.

**Proof:** Let \( K = \max_{x \in [a,b]} |f(x)| \) and define

\[
w = y + Kv
\]

where \( v \) is a solution of

\[
Av = 1
\]

which satisfies the boundary conditions \( v(a) = v(b) = 0 \). Then,

\[
Aw = Ay + K \geq 0.
\]

Since \( w(a) = w(b) = 0 \), the Maximum Principle (see [6, p. 7]) implies that

\[
w < 0 \text{ for } x \in (a,b)
\]
Thus, for all \( x \in [a, b] \),

\[
y + Kv \leq 0 \leq K \max_{x \in (a, b)} |v|.
\]

Define \( C_1 = \max_{x \in (a, b)} |v| \). Since \( K = \max_{x \in [a, b]} |f| \),

\[
y - K(C_1 - v) \leq 2C_1 |f|.
\]

On the other hand, by defining

\[ w_1 = y - Kv, \]

we can use a similar argument to conclude that

\[ -2C_1 |f| \leq y. \]

Thus

\[ |y| \leq |f| \]

where \( C = 2C_1 \). This ends the proof of Lemma 3.5.

A useful corollary to Lemma 3.5 is that

\[ |y| \leq C |f| \leq C \{ \sqrt{b - a} + 1 \} \|f\| \quad (3.42) \]

This result follows immediately from Lemma 3.5 and the fact that for any \( f \in H \)

\[ |f(x)| \leq \|f\|_{W_2^1} \{ \sqrt{b - a} + 1 \} \quad (3.43) \].
(Inequality (3.43) can be proven by methods similar to those which we used to prove inequality (3.31) for \( f \in H_h \).

Next, we will show that the operators \( A^{-1}, L A^{-1}, B^{-1}, B^{-1} L \), and \( (I + B^{-1} L)^{-1} = A^{-1} B = I - A^{-1} L \) are continuous mappings of the space \( H \) into itself. Since the operator \( L \) is multiplication by \(-q(x)\) it is continuous. If we can show that \( A^{-1} \) and \( B^{-1} \) are also continuous, it will follow that all of the above operators are continuous.

To show that \( A^{-1} \) and \( B^{-1} \) are continuous, we will appeal to the sequential definition of continuity: if the convergence to zero of any sequence \( f \in H \) guarantees the convergence to zero of the sequence \( A^{-1} f \in H \), then \( A^{-1} \) is continuous. As part of the proof, we will need to establish that if \( A y = f \) then

\[ |y'| \leq C_o |f| \]

where \( C_o \) is some constant. The proof of this inequality is omitted since it is done in the same way as the proof of Lemma 3.6 which appears later in this paper.

Let \( f_n(x) \in H \) be a sequence of functions which converges to zero, and let \( y_n \) denote the sequence

\[ y_n = A^{-1} f_n. \]

Using inequalities (3.42) and (3.44) it then follows that

\[ \|y_n\| = \int_a^b (y'_n)^2 \, dx + y_n^2(a) \leq \|f_n\|^2 \{ (b-a) + 1 \}. \]
Therefore, \( |y_n| \) converges to zero if \( |f_n| \) does. Thus \( A^{-1} \) is a continuous transformation from \( H \) into \( H \). The same arguments apply to \( B^{-1} \). Consequently, \( A^{-1}, LA^{-1}, B^{-1}, B^{-1}L \) and \((I + B^{-1}L)^{-1} = A^{-1}B = I - A^{-1}L \) are continuous mappings of the space \( H \) into itself.

An optimal matrix \( O_{B_h}^{-1} \) has already been constructed for the boundary-value problem (3.32), namely \( O_{B_h}^{-1} = h^{2}A_h^{-1}G_h \) where \( A_h \) is given by (3.39) and \( G_h \) is given by (3.40). This matrix, \( O_{B_h}^{-1} \), fulfills the required conditions of the operator \( O_{B_h}^{-1} \) in Theorem 3.4.

Let

\[
\begin{pmatrix}
q(x_0), & 0, & 0, & 0, & 0, & 0, & 0
\end{pmatrix}
\begin{pmatrix}
0, & q(x_1), & 0, & 0, & 0, & 0, & 0
\end{pmatrix}
\begin{pmatrix}
0, & 0, & q(x_2), & 0, & 0, & 0, & 0
\end{pmatrix}
\begin{pmatrix}
0, & 0, & 0, & q(x_n), & 0, & 0, & 0
\end{pmatrix}
\]

We have now defined the operators which correspond to the operators in Theorem 3.4. It is our claim that these operators satisfy the hypothesis to Theorem 3.4.

First, note that

\[
\psi^{-1}_h \psi = \psi^{-1}_h
\]

since both sides define the piecewise linear function in \( h \) whose values at the points \( x_i \), \( i = 0,1,\ldots,n \), are \( (qf)(x_i) \), for any given function \( f \). This then implies that the assumptions (3.20) and (3.22) in Theorem 3.4 are satisfied.
In order to show that assumption (3.21) is satisfied, i.e., that
\[ \| \psi_h B^{-1} P_h A^{-1} \|_* = o (\| \psi_h A^{-1} P_h \|_*) , \]
we will repeatedly be using the Green's function, \( K(x,t) \), for the operator \( A \) with the boundary conditions \( y(a) = y(b) = 0 \). From Kreider [3, p. 498], we find that some basic properties of \( K(x,t) \) are as follows:

1. \( K(x,t) \) is defined and continuous for \( a < x < b, a < t < b \), and, as a function of \( x \), is twice continuously differentiable except when \( x = t \);

2. For each fixed \( t_0 \) in \([a,b]\), \( K(x,t_0) \) belongs to the subspace \( S \) (i.e. satisfies the boundary conditions imposed on the problem), and in addition, is a solution of the equation \( Ly = 0 \) (\( Ay = 0 \)) except at the point \( x = t_0 \):

3. \[
\frac{d}{dx} K(x,t_0) \bigg|^{t_0^+}_{t_0^-} = 1/p(t_0).
\]

One further result we will need to show that (3.21) is satisfied is

**Lemma 3.6.** If \( f \in H \) and \( y = A^{-1} f \), then

\[
| (q(x)y)' | \leq K_0 | f | . \tag{3.45}
\]

**Proof:**

\[
(q(x)y)' = q(x)y' + 2q'(x)y' + q''(x)y, \tag{3.46}
\]

and from Kreider [3, p. 501] we find that
\[ y = \int_{a}^{b} K(x,t)f(t) \, dt, \quad (3.47) \]

\[ y' = \int_{a}^{b} \frac{\partial}{\partial x} K(x,t)f(t) \, dt, \]

and

\[ y'' = \int_{a}^{b} \frac{\partial^2}{\partial x^2} K(x,t)f(t) \, dt + f(x)/p(x) \]

where

\[ \int_{a}^{b} \frac{\partial^2}{\partial x^2} K(x,t)f(t) \, dt \]

means

\[ \int_{a}^{x} \frac{\partial^2}{\partial x^2} K(x,t)f(t) \, dt + \int_{x}^{b} \frac{\partial^2}{\partial x^2} K(x,t)f(t) \, dt \]

since \( \frac{\partial}{\partial x} K(x,t) \) is discontinuous at \( t = x \) and a second derivative cannot be taken. Since \( K(x,t) \) and \( \frac{\partial}{\partial x} K(x,t) \) are bounded (refer to properties (1) and (3) of \( K(x,t) \)), it follows that

\[ C_1 = \max_{x,t \in [a,b]} |K(x,t)|, \]

and

\[ C_2 = \sup_{x,t \in [a,b]} |\frac{\partial}{\partial x} K(x,t)| \]

are finite numbers. Further, let

\[ M_0 = \max_{x \in [a,b]} |1/p(x)|, \]

\[ M_1 = \max_{x \in [a,b]} |q''(x)|, \]
\[ M_2 = \max_{x \in [a,b]} |q'(x)| , \]

\[ M_3 = \max_{x \in [a,b]} |q(x)| , \]

and \[ M_4 = \max_{x \in [a,b]} |p'(x)| . \]

Observe that all of these numbers are finite because of the assumed properties of \( p(x) \) and \( q(x) \). Let

\[ C_3 = M_0 \left( M_3 C_1 + M_4 C_2 \right) . \]

From property (2) of \( K(x,t) \),

\[ p(x) \frac{\partial^2}{\partial x^2} K(x,t) + p'(x) \frac{\partial}{\partial x} K(x,t) - q(x) K(x,t) = 0, \]

except when \( x = t \). Therefore

\[ \left| \frac{\partial^2}{\partial x^2} (x,t) \right| \leq \frac{1}{|p(x)|} \left( \left| q(x) K(x,t) \right| + \left| p'(x) \frac{\partial}{\partial x} K(x,t) \right| \right) \]

\[ \leq M_0 \left( M_3 C_1 + M_4 C_2 \right) \]

\[ \leq C_3 . \]

Thus

\[ \left| \frac{\partial^2}{\partial x^2} K(x,t) \right| \leq C_3, \]

except when \( x = t \). This implies that, for each \( x \),

\[ \lim_{t \to x^+} \left| \frac{\partial^2}{\partial x^2} K(x,t) \right| \leq C_3 \]

and \[ \lim_{t \to x^-} \left| \frac{\partial^2}{\partial x^2} K(x,t) \right| \leq C_3 . \]
Thus, from (3.46)

\[
|q(x,y)^n| \leq M_3 \int_a^b \frac{\partial^2}{\partial x^2} K(x,t)f(t) \, dt + M_0 |f|
\]

\[
+ 2M_2 \int_a^b K(x,t)f(t) \, dt + M_1 \int_a^b |K(x,t)f(t)| \, dt
\]

\[
\leq M_3 \{2C_3(b-a)|f| + M_0 |f|
\]

\[
+ 2M_2 C_2(b-a)|f| + M_1 C_1(b-a)|f|
\]

\[
\leq [(b-a)\{2M_3 C_3 + 2M_2 C_2 + M_1 C_1\} + M_0]|f|
\]

\[
\leq K_0 |f|
\]

\[
\leq K_0 \|f\|
\]

where \(K_0 = (b-a)\{2M_3 C_3 + 2M_2 C_2 + M_1 C_1\} + M_0\). This completes Lemma 3.6.

As one more step in our discussion of assumption (3.21), we will show that

\[
|\psi_h B^{-1} p_h A^{-1} f| \leq C'h^2 |f| ,
\]

(3.48)

where \(C'\) is some constant. In our proof we will appeal to the following theorem from Moursund and Duris [5, p. 124]:

Let \(f(x) \in C^2[a,b]\), where \(a < x_0 < b\). Then for \(x \in [a,b]\) there exists a number \(\xi\), depending upon \(x\) and satisfying

\[
\min(x, x_1, x) < \xi < \max(x, x_1, x)
\]
such that
\[ f(x) = \frac{x-x_i}{x_o-x_i} f(x_o) + \frac{x-x_o}{x_i-x_o} f(x_i) + \frac{(x-x_o)(x-x_i)}{2} f''(\xi). \]

Applying this theorem to \((qy)(x)\), for \(x_i \leq x \leq x_{i+1}\), we see that
\[ (qy)(x) = \frac{x-x_i}{h} (qy)(x_i) + \frac{x-x_i}{h} (qy)(x_{i+1}) + \frac{(x-x_i)(x-x_{i+1})}{2 h} (qy)''(\xi) \]

so that the difference between \((qy)(x)\) and the straight line connecting \((qy)(x_i)\) and \((qy)(x_{i+1})\) is given by the second derivative term. Using the above interpolation theorem and the fact that \(P_h(qy)(x_i) = 0\), \(i = 0,1,\ldots,n\), we see that
\[ |P_h(qy)(x)| \leq \frac{h^2}{2} |(qy)'| \]

where \(|(qy)''| = \sup_{x \in [a,b]} |(qy)''(x)|\).

Therefore, by Lemma 3.6,
\[ |P_h(qy)(x)| \leq \frac{h^2}{2} |(qy)'| \leq \frac{k_0 h^2}{2} |f|. \tag{3.49} \]

Let \(g \in H\), and define \(u = B^{-1}g\). Then, from Lemma 3.5 and inequality (3.49), it follows that
\[ |B^{-1}P_h L A^{-1} f| \leq |B^{-1}P_h L y| \leq |B^{-1}P_h (-qy)| \leq \frac{k_0 K_1}{2} h^2 |f|. \]
Thus,
\[
|\psi_h^{-1}P_h L A^{-1}f| \leq \frac{K_0 K_1}{2} h^2 |f|,
\]
and we have completed the proof of inequality (3.48).

As the final step in our discussion of (3.21), we will show that there exists a function \( \phi_h(t) \in H_h \) such that
\[
(3.50)
\]
where \( C'' \) is some constant. Then, by combining results (3.48) and (3.50), it will follow that (3.21) is satisfied.

Define a sawtooth function \( \phi_h(x) = \)
\[
\begin{cases}
  x - x_i & \text{for } x_i \leq x \leq x_i + (1/2)h, \\
  x_{i+1} - x & \text{for } x_i + (1/2)h \leq x \leq x_{i+1}, i = 0,1,\ldots,n-1.
\end{cases}
\]
Then \( \phi_h(x) \in H_h \), and, for all \( h \), \( |\phi_h(x)|_{W^2_2} = b-a \).

In order to show that (3.50) is valid, we will be appealing to the following properties of the Green's function. For the operator \( A \) and the boundary conditions \( y(a) = y(b) = 0 \), the Green's function is defined as
\[
K(x,t) = \begin{cases}
  \frac{y_1(x)y_2(t)}{p(t)[y_1(t)y_2'(t) - y_2(t)y_1'(t)]}, & x \leq t, \\
  \frac{y_1(t)y_2(x)}{p(t)[y_1(t)y_2'(t) - y_2(t)y_1'(t)]}, & x \geq t,
\end{cases}
\]
where \( y_1(x) \) and \( y_2(x) \) are independent solutions of the homogeneous equation \( Ay = 0 \), chosen so that \( y_1 \) satisfies the boundary condition imposed at \( x = a \), and \( y_2 \) satisfies the boundary condition imposed at \( x = b \) (see [3, p. 499]). Note that \( K(x,t) \) is independent of the right hand side of the differential equation. By using the facts that

1. \( y_1(x) \neq 0 \) for \( x \neq a \) and \( y_2(x) \neq 0 \) for \( x \neq b \) (Here we are appealing to the Sturm comparison theorem, which is given in [3, p. 232], and the facts that (i) the equation \( Ay = f \) can be put in the form required by the comparison theorem (see [3, p. 238]), and (ii) \( q(x) \geq 0 \).

Example 3 in [3, p. 233] illustrates the way in which the Sturm comparison theorem can be used to give (1), (2) the Wronskian, \( W(y_1, y_2, t) = y_1(t)y_2'(t) - y_1'(t)y_2(t) \), never vanishes, and (3) \( p(x) \geq p_0 > 0 \), we are able to conclude that \( K(x,t) \) never changes sign. Without loss of generality, assume that \( K(x,t) \geq 0 \).

Therefore, considering the continuity and the above stated properties of \( K(x,t) \), we are able to construct a function

\[
K_2(x,t) = \begin{cases} 
K_3 & \text{for } a + \delta \leq x, t \leq b - \delta, \\
0 & \text{otherwise},
\end{cases}
\]

where \( \delta \) and \( K_3 \) are positive constants chosen so that \( \delta \) is of the form

\[
0 < \delta = \frac{(b-a)}{j} \leq \frac{(b-a)}{4}, \quad (3.51)
\]

for some integer \( j \geq 4 \), and, for \( a + \delta \leq x, t \leq b - \delta \),

\[
K_2(x,t) = K_3 < K(x,t).
\]
In order to show that (3.50) is valid, define

\[ y_h(x) = A^{-1} \phi_h(x). \]

Then, for \( a + \delta \leq x \leq b - \delta \),

\[ y_n(x) = \int_a^b K(x,t) \phi_h(t) \, dt. \]

Thus, for \( a + \delta \leq x \leq b - \delta \),

\[
\begin{aligned}
&\int_a^b K(x,t) \phi_h(t) \, dt \\
&\geq \int_a^b K_2(x,t) \phi_h(t) \, dt \\
&\geq K_3 \int_{a+\delta}^{b-\delta} \phi_h(t) \, dt. 
\end{aligned}
\] (3.52)

The next step is to compute

\[
\int_{a+\delta}^{b-\delta} \phi_h(t) \, dt.
\]

This is done by considering the possible places in which the interval \((a, a+\delta)\) could intersect some subinterval of the form \((x_i, x_{i+1})\). The three important cases are (1) \( \frac{h}{2} \leq x_{i+1} - (a+\delta) \leq h \), (2) \( 0 \leq x_{i+1} - (a+\delta) \leq \frac{h}{2} \), and (3) \( x_{i+1} = a+\delta \). These three cases and the respective forms of the integral of the sawtooth function \( \phi_h(t) \) are reflected in the following formula:
\[
\int_{a+\delta}^{b-\delta} \phi_h(t) \, dt =
\begin{cases}
\frac{h^2}{4} + \frac{h^2}{2} - (h - (Nh-\delta))^2, & \frac{h}{2} < Nh-\delta < h, \\
\frac{Nh-\delta}{4}^2 + (Nh-\delta)^2, & 0 < Nh-\delta \leq \frac{h}{2} \\
\frac{Nh+\delta}{4}, & Nh+\delta.
\end{cases}
\]

Here \( n^* \) represents the number of complete intervals \((x_i, x_{i+1})\) contained in \((a+\delta, b-\delta)\), and \(2N = n - n^*\). Therefore, since \( n = (b-a)/h \) represents the total number of intervals \((x_i, x_{i+1})\) in \((a,b)\), \(N\) indicates the number of complete intervals plus the number of partial intervals in \((a,a+\delta)\) or \((b-\delta,b)\).

In the formula for

\[
\int_{a+\delta}^{b-\delta} \phi_h(t) \, dt,
\]

\(\frac{h^2}{2} - (h - (Nh-\delta))^2\) and \((Nh-\delta)^2\) are positive. Thus, from (3.52), we obtain the bound

\[
y_h(x) \geq K_3(n^*\frac{h^2}{4}).
\]  

(3.53)

In order to obtain a more useful bound on \(y_h(x)\), observe that since \(\delta\) divides the interval \((a,b)\) into \(j\) equal parts (see definition (3.51)), it follows that \(\delta\) divides \(n\) into \(j\) 'nearly equal' parts—that is if we partition \((a,b)\) into the subintervals
(a, a+δ), (a+δ, a+2δ), ..., (a+(j-1)δ, a+jδ = b),
then each such subinterval contains 'almost' the same number of subintervals (xᵢ, xᵢ₊₁). The possibility that an element of the δ-partition contains only part of a subinterval (xᵢ, xᵢ₊₁) is the reason we state that δ divides n into j 'nearly equal' parts. Although these parts may not be exactly equal, we can write, however, that

\[ n' > \frac{n}{j} - 2 \]

where n' represents the minimum number of subintervals (xᵢ, xᵢ₊₁) contained in any element of the δ-partition. Also, since \( j > 4 \), it follows that

\[ n^* > n' > \frac{n}{j} - 2 \geq \frac{n}{2j} \]

and, for \( n > 4j \), that

\[ n^* > n' > \frac{n}{j} - 2 \geq \frac{n}{2j} \]  \hspace{1cm} (3.54)

Thus, in order to obtain the final bound on \( y_n(x) \), we combine (3.53) and (3.54), to conclude that

\[ y_n(x) \geq K_3 \left( n^* \frac{h^2}{4} \right) \]

\[ \geq K_3 \frac{nh^2}{8j} \]

\[ \geq K_3 \frac{(b-a)h}{8j} \]

\[ \geq C''h \]
where \( C'' = \frac{(b-a)}{b-a} \). Therefore,

\[
y_h(x) \geq C''h, \text{ for } a+\delta \leq x \leq b-\delta,
\]

which implies that (3.50) is valid for the function \( \phi_h(x) \).

Further, by combining (3.48) and (3.50), it is seen that

\[
\frac{|\psi B^{-1} p L A^{-1} f|}{C'h|f|} \leq h \leq \frac{|\psi A^{-1} \phi|}{C''}
\]

This then implies that

\[
|\psi B^{-1} p L A^{-1}|_\psi = o\left(\frac{|\psi A^{-1} p|}{h}\right),
\]

so assumption (3.21) in Theorem 3.4 is satisfied.

Next, we will show that assumption (3.23) is also satisfied.

Decompose the operator \((I + B^{-1} L)^{-1}\) in the form

\[
(I + B^{-1} L)^{-1} = I + (-A^{-1} L) = I + R. \tag{3.55}
\]

Clearly the operator \( R = -A^{-1} L \) renders a continuous mapping of \( H \) into \( H \). One may therefore write

\[
\psi B^{-1} p L(I + B^{-1} L)^{-1} \psi^{-1} = \psi B^{-1} p L \psi^{-1}
\]

\[
+ \psi B^{-1} p L \psi^{-1}.
\]

Let \( u(x) = \psi^{-1}_h \bar{u} \) where \( \bar{u} \) is any \( n+1 \) dimensional vector. Then, \( u(x) \) is the linear function connecting the components of \( \bar{u} \). Now, for \( x \in (x_i, x_{i+1}) \), inequality (3.49) gives
\[ |p_h \nabla^2_h u| \leq |P_h q u(x)| \leq \frac{h^2}{2} \sup_{x \in (x_i, x_{i+1})} \{ |q''(x)u(x)| + 2u'(x)q'(x) \} . \]  \tag{3.57}

Here, we have dropped the term \( q(x)u''(x) \) because \( u''(x) = 0 \) for \( x \in (x_i, x_{i+1}) \). Since
\[ u(x) = \frac{x-x_i}{h} u_{i+1} + \frac{x_{i+1}-x}{h} u_i , \]
for \( x_i \leq x \leq x_{i+1}, i = 0,1, \ldots, n-1 \), where \( u_i \) designates the \( i \)th component of \( \vec{u} \), it is clear that
\[ \max_{x \in [a,b]} |u(x)| \leq \frac{2(b-a)}{h} |\vec{u}| , \]
and, for \( i = 0,1, \ldots, n-1 \), that
\[ \sup_{x \in (x_i, x_{i+1})} |u'(x)| \leq \frac{2}{h} |\vec{u}| . \]
Thus, by considering inequality (3.57) and the maximum values of \( q''(x) \) and \( z'(x) \), we are able to find a constant \( C_2 \) such that, for \( x \in (x_i, x_{i+1}), i = 0,1, \ldots, n-1 \),
\[ |P_h \nabla^2_h u| \leq C_2 h |\vec{u}| . \] \tag{3.58}
But \( P_h \nabla^2_h u \) is a continuous function, so inequality (3.58) must hold for all \( x \in [a,b] \).

Using (3.57) and Lemma 3.5, it is found that
\[ |\psi_h^{-1} B_p \nabla^2_h u| \leq C_2 K h |\vec{u}| \]
where \( K \) is the constant given by Lemma 3.5. An analogous result applies to the term

\[
\psi_h B^{-1} P_h L R \psi_h^{-1}.
\]

Therefore, from (3.56), we obtain directly (3.23), i.e.,

\[
|\psi_h B^{-1} P_h L (I + B^{-1} L)^{-1} \psi_h^{-1}| = o(1).
\]

Assumption (3.24) follows from the definition of the operator \( R \): since \( R = -A^{-1} L \), Theorem 3.5 implies that

\[
|R y| \leq |-A^{-1} L y| \leq |A^{-1} q y| \leq K |q y|,
\]

where \( K \) is some constant. The triangle inequality gives us

\[
|\psi_h (I + B^{-1} L)^{-1} \psi_h^{-1}| \leq |\psi_h (I + R) \psi_h^{-1}|
\]

\[
\leq |\psi_h I \psi_h^{-1}| + |\psi_h R \psi_h^{-1}|.
\]

Hence, by applying (3.59) and taking the maximum value of \( |q(x)| \), it is clear that

\[
|\psi_h R \psi_h^{-1} u| \leq K_1 |u|,
\]

where \( K_1 \) is some constant. A similar result holds for \( \psi_h I \psi_h^{-1} \).

Therefore we obtain (3.24), i.e.,

\[
|\psi_h (I + B^{-1} L)^{-1} \psi_h^{-1}| = O(1).
\]
Next, we will prove that the matrix $I + \sigma_{B, h}^{-1} \cdot I$ is non-singular.

Observe that

$$I + \sigma_{B, h}^{-1} \cdot I = -A_h^{-1} \{ -A_h - G_h \cdot h^2 \}$$

where $A_h$ is given by (3.39) and $G_h$ is given by (3.40). Thus, if $I + \sigma_{B, h}^{-1} \cdot I$ were singular, then the following matrix would also be singular:

$$-A_h - G_h \cdot h^2 =$$

$$= \begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
-P(1/2) + h^2 q(x_1) y_1^{(1)} & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
-P(3/2) + h^2 q(x_2) y_2^{(1)} & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 1
\end{bmatrix}$$

However, for sufficiently small $h$, $-A_h - G_h \cdot h^2$ is irreducibly diagonally dominant and has an inverse. Therefore, $-A_h - G_h \cdot h^2$ is non-singular. This implies that, for sufficiently small $h$, $I + \sigma_{B, h}^{-1} \cdot I$ is also non-singular.
We have now shown that the given operators satisfy the hypothesis to Theorem 3.4. Hence it follows from Theorem 3.4 that the sequence of matrices

$$(I + \beta_{\neq h}^{-1} \frac{\partial}{\partial h})^{-1} \beta_{\neq h}^{-1}$$

forms an asymptotically optimal sequence, i.e., that the system

$$[p_{k-(1/2)} - h^2 \gamma_{k-1} q(x_{k-1})] y_{k-1}$$

$$- [p_{k-(1/2)} + p_{k+(1/2)} + h^2 \gamma_k q(x_k)] y_k$$

$$+ [p_{k+(1/2)} - h^2 \gamma_{k+1} q(x_{k+1})] y_{k+1}$$

$$+ h^2 [\gamma_{k-1} f(x_{k-1}) + \gamma_k f(x_k) + \gamma_{k+1} f(x_{k+1})]$$

yields an asymptotically optimal method of solution.
BIBLIOGRAPHY


