## Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Introduction</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Reading List</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td><strong>Chapter 1. The approximation problem and existence of best approximations</strong></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Study Session 1: Approximation in a metric space</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Study Session 2: Approximation in a normed linear space</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td><strong>Chapter 2. The uniqueness of best approximation</strong></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Study Session 1: Convexity</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Study Session 2: Best approximation operators</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td><strong>Chapter 3. Approximation operators and some approximating functions</strong></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Study Session 1: ‘Good’ versus ‘best’ approximation</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Study Session 2: Types of approximating functions</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td><strong>Chapter 4. Polynomial interpolation</strong></td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Study Session 1: Polynomial interpolation</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Study Session 2: Chebyshev interpolation</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td><strong>Chapter 5. Divided differences</strong></td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Study Session 1: Basic properties of divided differences</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Study Session 2: Numerical considerations and Hermite interpolation</td>
<td>49</td>
</tr>
</tbody>
</table>
Introduction

The subject of Approximation Theory lies at the frontier between Applied Mathematics and Pure Mathematics. Practical problems, such as the computer calculation of special functions like $e^x$, lead naturally to theoretical problems, such as ‘how well can we approximate by a given method?’ or ‘how fast does a given algorithm converge?’.

Powell’s book Approximation Theory and Methods (hereafter referred to as ‘Powell’) provides an excellent introduction to these theoretical problems, covering the basic theory of a wide range of approximation methods. Professor Powell is an expert on both pure and applied approximation theory, and the book contains a very detailed list of references to and discussion of the research literature.

This course is based on a treatment of fifteen chapters of Powell. Do not be misled by the statement that this is an undergraduate textbook. Much of the material can be taught at that level, but when looked at in detail many parts of it are quite demanding. These course notes will guide you through the book telling you which sections to read, explaining difficult parts, correcting errors (mercifully few!) and setting SAQs and Problems to test your understanding of the material. You should attempt all the SAQs and as many Problems as you have time for: full solutions are given at the end of the notes for each chapter.

You will find the exercises in Powell quite varied. Many are routine, but others are rather hard and some are very hard (particularly those which contain the word ‘investigate’). I have resisted the temptation to attach ‘stars’ to harder exercises and instead tried to provide ‘hints’, where appropriate. In general I feel that, at this level, it is a good idea for you to try and make your own judgement about the difficulty of a given problem.

Many of the exercises require the use of a good scientific calculator (one with special functions, including hyperbolics, and a memory). Some require the solution of non-linear equations of the form $f(x) = 0$ by using, for example:

the bisection method (finding an interval $[a, b]$ such that $f(a), f(b)$ have opposite signs, testing $f(c)$, where $c = \frac{1}{2}(a + b)$, and then repeating the process with either $[a, c]$ or $[c, b]$);

Newton’s method (making a good initial guess $x_0$ at a solution and then calculating the sequence $x_n$ given by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \ldots.$$ Such methods can be implemented on a basic scientific calculator (especially if only a rough answer is needed), but it will obviously save time if you have access to a computer. You will not be expected to determine accurate solutions by such methods in the examination. On the matter of accuracy, I have tended to present calculations as they appeared on my own calculator, and have sometimes given final answers correct to only three significant digits.

In order to pace you through the course, there are four Tutor-Marked Assignments (TMAs). These are compulsory in that you cannot pass the course without obtaining a reasonable average grade on them. Your three best TMAs carry 50% of the total marks for the course, the remaining 50% coming from the three-hour examination at the end of the course. Please note that TMAs cannot be accepted after their cut-off dates, other than in exceptional circumstances.

Although you should have plenty to do reading Powell and these course notes, I have added a reading list after this introduction. This splits into books covering the background material which is assumed in Powell (Linear Algebra, Metric Spaces, etc.) and other books on Approximation Theory.
I should be grateful to receive any comments you may have on the course notes and on the set book. The course notes have already benefited greatly from close reading by Mick Bromilow here at the OU and Martin Stynes of University College, Cork. Their help has been invaluable. Finally, I should like to thank all those who have helped prepare these Course Notes, including members of the Desktop Publishing Unit, Alison Cadle who edited them, and the many M832 students who have supplied corrections to earlier versions.

Phil Rippon
Milton Keynes, August 2006

Reading List

Background

(A concise introduction to real analysis, including metric spaces, integration and functions of several variables, as well as basic linear algebra — available in a paperback International Student Edition.)

(An introduction to metric spaces, emphasising the importance of iteration. Plenty of explanation.)

(An introduction to the use of vector spaces of functions in solving linear differential equations — lots of worked exercises.)

M203 *Introduction to Pure Mathematics*
MST204 *Mathematical Models and Methods*
M386 *Metric and Topological Spaces* (now part of M435)

Approximation Theory

(Cheap and covers very similar material to Powell, with less on splines and more on rational approximation.)

(Cheap, but a classic text which overlaps Powell considerably, though with a much greater emphasis on complex approximation.)

(Recent and sophisticated, this book examines the more difficult non-linear theory which Powell largely avoids.)
Chapter 1  The approximation problem and existence of best approximations

The book begins with a discussion of the types of problems which are to be solved and several fundamental results. Powell assumes that the reader is quite familiar with metric spaces and so the commentary below includes a short refresher course on these, in case you are rusty on this subject.

This chapter splits into TWO study sessions:

**Study session 1**: Sections 1.1–1.2.

**Study session 2**: Sections 1.3–1.5.

---

**Study Session 1: Approximation in a metric space**

**Read** Sections 1.1 and 1.2

**Commentary**

1. Section 1.1 describes the three ingredients of an approximation problem:
   (a) a function \( f \) to be approximated, lying in some underlying (background) set \( B \);
   (b) a set of functions \( A \subseteq B \) from which we wish to choose an approximation \( g \) to \( f \);
   (c) a means of measuring how close together \( g \) and \( f \) are.

   ![Diagram showing function \( f \), set \( A \), and approximation]

   In a **continuous** approximation problem \( f \) is typically a real function, such as \( f(x) = e^x \), and the set \( A \) is a finite-dimensional vector space of real functions, such as the set \( \mathcal{P}_n \) of polynomials of degree at most \( n \). One measure of how closely a function \( g \) approximates to \( f \) on an interval \([a, b]\) is
   \[
   \max_{a \leq x \leq b} |g(x) - f(x)|.
   \]

   In a **discrete** approximation problem \( f \) is typically a vector of function values \((f(x_1), \ldots, f(x_n))\), where \( g \) belongs to some set of approximating functions. Note that \( f \) and \( g \) are used here to represent both a function and the corresponding vector of function values. One measure of how closely \( g \) approximates to \( f \) is
   \[
   \max_{1 \leq i \leq n} |g(x_i) - f(x_i)| \quad \text{or} \quad \max_{1 \leq i \leq n} |g(x_i) - f_i|.
   \]
A **metric space** \((B, d)\) is a set \(B\) and a **metric** (or distance function) \(d(a, b)\), \(a, b \in B\), such that for all \(a, b, c \in B\):

1. \(d(a, b) \geq 0\), with equality if and only if \(a = b\);
2. \(d(a, b) = d(b, a)\);
3. \(d(a, c) \leq d(a, b) + d(b, c)\).

The most familiar metric spaces are \(\mathbb{R}\) with the metric \(d(a, b) = |a - b|\) and \(\mathbb{R}^2\) with the metric
\[
d(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}, \quad a = (a_1, a_2), \ b = (b_1, b_2).
\]

This example explains why (M3) is called the **triangle inequality**.

More generally \(\mathbb{R}^n\) is a metric space with
\[
d(a, b) = \left( \sum_{i=1}^{n} (a_i - b_i)^2 \right)^{\frac{1}{2}}, \quad a = (a_1, \ldots, a_n), \ b = (b_1, \ldots, b_n). \tag{1}
\]

For general \(n\) it is not quite so obvious that (M3) holds. The proof is given later when we introduce a large family of metric spaces. Before that we recall a number of definitions and results for future reference. No proofs are given as these results are quite standard.

**Convergence** A sequence \(a_n, n = 1, 2, \ldots\), in \(B\) is **convergent** with limit \(a^*\) if \(d(a_n, a^*) \to 0\) as \(n \to \infty\).

**Closed set** A subset \(F\) of \(B\) is **closed** if every convergent sequence \(a_n, n = 1, 2, \ldots\), in \(F\) has its limit in \(F\).

For example, the closed ball
\[
\{ b \in B : d(a, b) \leq r \}, \quad r > 0,
\]
is a closed set.

**Open set** A subset \(E\) of \(B\) is **open** if \(B \setminus E\) is closed.

For example, the open ball
\[
\{ b \in B : d(a, b) < r \}, \quad r > 0,
\]
is an open set.

**Compact set** A subset \(K\) of \(B\) is **compact** if every sequence \(a_n, n = 1, 2, \ldots\), in \(K\) has a convergent subsequence \(a_{n_k}, k = 1, 2, \ldots\), whose limit \(a\) is in \(K\).

For example, every finite set is compact. In \(\mathbb{R}^n\) with the metric \(d\) given by equation (1), every closed set which is also bounded (i.e. lies inside some fixed closed ball) is compact. Note that **every** compact set is closed.
Continuous function  A function \( \phi : (B, d) \to (B', d') \) is continuous at \( a \in B \) if for each \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that
\[
d(a, b) < \delta \quad \Rightarrow \quad d'(\phi(a), \phi(b)) < \varepsilon
\]
(equivalently: for each sequence \( a_n \to a \) in \( B \), we have \( f(a_n) \to f(a) \)). We say that \( \phi : (B, d) \to (B', d') \) is continuous if \( \phi \) is continuous at each \( a \in B \).

Uniformly continuous function  A function \( \phi : (B, d) \to (B', d') \) is uniformly continuous on \( B \) if for each \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that, for all \( a, b \in B \),
\[
d(a, b) < \delta \quad \Rightarrow \quad d'(\phi(a), \phi(b)) < \varepsilon.
\]

Extreme Value Theorem  If \( \phi : (B, d) \to (\mathbb{R}, d') \) is continuous (where \( d'(a, b) = |a - b| \)), then \( \phi \) attains a maximum value and a minimum value on any compact subset \( K \) of \( B \).

Uniform Continuity Theorem  If \( \phi : (B, d) \to (B', d') \) is continuous then \( \phi \) is uniformly continuous on any compact subset \( K \) of \( B \).

3. Theorem 1.1. The proof can be shortened. You can omit the second sentence and the word ‘Otherwise’ from the third sentence, and then use the notation \( a^* \) in place of \( a^+ \). Note that Powell uses ‘limitpoint’ to mean the limit of a convergent subsequence.

The following picture may be helpful.

4. The set \( \mathcal{A} \) discussed after Theorem 1.1 is not compact; for example, the sequence \( (1 - \frac{1}{n}, 0) \) lies in \( \mathcal{A} \) but has no subsequence which converges to a limit in \( \mathcal{A} \). If \( f = (2, 0) \), say, then for each \( a \in \mathcal{A} \) we can find an \( \pi \in \mathcal{A} \) which is closer to \( f \) than \( a \).

Self-assessment questions

S1 Consider the problem of fitting the data in Figure 1.2 by a straight line. Show that the set \( \mathcal{A} \) of vectors \((p(x_1), \ldots, p(x_5))\), arising from functions \( p(x) = c_0 + c_1(x) \), forms a 2-dimensional subspace of \( \mathbb{R}^5 \).

S2 Prove the following generalisation of Theorem 1.1. If \( \mathcal{A}_1 \) and \( \mathcal{A}_2 \) are compact subsets of \( B \), then there exist \( a_1^* \) in \( \mathcal{A}_1 \) and \( a_2^* \) in \( \mathcal{A}_2 \) such that
\[
d(a_1^*, a_2^*) = \inf\{d(a_1, a_2) : a_1 \in \mathcal{A}_1, a_2 \in \mathcal{A}_2\}.\]

\[6\]
Study Session 2: Approximation in a normed linear space

Read Sections 1.3, 1.4 and 1.5

Commentary

1. Almost every metric space in Powell arises as a **normed linear space** (n.l.s.). This is a linear space \( B \) (also called a vector space) with an associated **norm** \( \|a\|, a \in B \), such that, for all \( a, b \in B \) and \( \lambda \in \mathbb{R} \):

   - **(N1)** \( \|a\| \geq 0 \), with equality if and only if \( a = 0 \);
   - **(N2)** \( \|\lambda a\| = |\lambda| \|a\| \);
   - **(N3)** \( \|a + b\| \leq \|a\| + \|b\| \).

Roughly speaking, the norm measures how large the element \( a \) is, that is, how far \( a \) lies from the zero element of the space.

By defining

\[ d(a, b) = \|a - b\|, \]

we find that \( (B, d) \) is a metric space. Properties (M1) and (M2) are immediate, as is (M3), since

\[ \|a - c\| = \|(a - b) + (b - c)\| \quad \text{(by linearity)} \]

\[ \leq \|a - b\| + \|b - c\|. \quad \text{(by (N3))} \]

For this reason, (N3) is also called the **triangle inequality**.

Powell gives some important examples of norms in Section 1.4. Two of these have useful geometric interpretations.

\[ \|f\|_{\infty} = \max_{a \leq x \leq b} |f(x)| \]

\[ \|f\|_1 = \int_a^b |f(x)| \, dx \]

The corresponding metrics have similar geometric interpretations.

\[ \|f - g\|_{\infty} = \max_{a \leq x \leq b} |f(x) - g(x)| \]

\[ \|f - g\|_1 = \int_a^b |f(x) - g(x)| \, dx \]
The 2-norm

\[ \|f\|_2 = \left( \int_a^b f(x)^2 \, dx \right)^{\frac{1}{2}} \]

has no convenient geometric interpretation.

For each of these norms, properties (N1) and (N2) are evident, but (N3) requires some work. We give here the argument for the continuous 2-norm above. The proof is in two stages.

(I) Cauchy–Schwarz Inequality

\[
\left| \int_a^b f(x)g(x) \, dx \right| \leq \|f\|_2 \|g\|_2, \quad \text{where } f, g \in C[a, b].
\]

**Proof** If \( \|f\|_2 = 0 \) or \( \|g\|_2 = 0 \), then the result is clear. Otherwise, we use the inequality

\[ \sqrt{AB} \leq \frac{A + B}{2}, \quad \text{for } A, B \geq 0, \quad (2) \]

with \( A = f(x)^2/\|f\|_2^2 \) and \( B = g(x)^2/\|g\|_2^2 \). Integration gives

\[
\frac{1}{\|f\|_2 \|g\|_2} \int_a^b |f(x)g(x)| \, dx \leq \frac{1}{2} \left\{ \frac{1}{\|f\|_2^2} \int_a^b f(x)^2 \, dx + \frac{1}{\|g\|_2^2} \int_a^b g(x)^2 \, dx \right\} = 1.
\]

The desired inequality now follows from

\[
\left| \int_a^b f(x)g(x) \, dx \right| \leq \int_a^b |f(x)g(x)| \, dx.
\]

**Remark** Another proof of the Cauchy–Schwarz inequality appears in the notes for Chapter 2.

(II) Minkowski’s Inequality

\[ \|f + g\|_2 \leq \|f\|_2 + \|g\|_2, \quad \text{where } f, g \in C[a, b]. \]

**Proof** If \( \|f + g\|_2 = 0 \), then the result is clear. Otherwise, note that

\[
\|f + g\|_2^2 = \int_a^b (f(x) + g(x))^2 \, dx
\]

\[
\leq \int_a^b |f(x) + g(x)| |f(x)| \, dx + \int_a^b |f(x) + g(x)| |g(x)| \, dx
\]

\[
\leq \|f + g\|_2 \|f\|_2 + \|f + g\|_2 \|g\|_2,
\]

by the Cauchy–Schwarz inequality. The desired inequality now follows on dividing by \( \|f + g\|_2 \).

As you should have realised, Minkowski’s inequality is just (N3) for the 2-norm \( \|f\|_2 \). The proof of (N3) for the discrete 2-norm is similar, proceeding via the discrete version of the Cauchy–Schwarz inequality:

\[
\left| \sum_{i=1}^m a_i b_i \right| \leq \left( \sum_{i=1}^m a_i^2 \right)^{\frac{1}{2}} \left( \sum_{i=1}^m b_i^2 \right)^{\frac{1}{2}}.
\]

The metric on \( \mathbb{R}^n \) which arises from the discrete 2-norm is precisely that defined in equation (1) of the commentary for Study Session 1.

The argument to prove property (N3) for

\[ \|f\|_p = \left( \int_a^b |f(x)|^p \, dx \right)^{1/p}, \]

if \( 1 < p < \infty \), is similar to the case \( p = 2 \). Instead of the Cauchy–Schwarz inequality, we need the more general Hölder inequality.
\[ \left| \int_a^b f(x)g(x) \, dx \right| \leq \|f\|_p \|g\|_q, \]

where \( p^{-1} + q^{-1} = 1 \), whose proof is based on the inequality

\[ A^{1/p} B^{1/q} \leq \frac{A}{p} + \frac{B}{q}, \quad A, B \geq 0. \]

Since Powell uses only the 1-norm, 2-norm and \( \infty \)-norm, we omit the details.

2. Theorem 1.2 is of fundamental importance, and the proof looks straightforward (note the use of the ‘backwards’ form of the triangle inequality \( \|a - f\| \geq \|a\| - \|f\| \) in (1.15), which is equivalent to \( \|a\| \leq \|a - f\| + \|f\| \)). However, the second sentence is not quite so transparent as it may appear. A closed ball in \( \mathbb{R}^n \) is certainly compact, but this does not immediately imply that a closed ball in a finite-dimensional subspace of an n.l.s. is compact. The proof of this fact is a little tricky but it illustrates how ‘analysis’ and ‘linear algebra’ interact in this subject.

We prove that if \( A \) is a finite-dimensional subspace of an n.l.s. and \( M \geq 0 \), then

\[ A_M = \{ a \in A : \|a\| \leq M \} \]

is compact, by showing that any sequence \( a_m, m = 1, 2, \ldots \), in \( A_M \) has a convergent subsequence, whose limit must be in \( A_M \) (since \( A_M \) is closed).

Let \( b_1, \ldots, b_n \) be a basis for \( A \) and write each \( a_m \) in the form

\[ a_m = \lambda_{1m} b_1 + \cdots + \lambda_{nm} b_n, \quad \lambda_{1m}, \ldots, \lambda_{nm} \in \mathbb{R}. \]

The result follows if we can show that the sequence \( \lambda_m = (\lambda_{1m}, \ldots, \lambda_{nm}) \) of coefficient vectors has a convergent subsequence in \( \mathbb{R}^n \), since the function \( \phi: \mathbb{R}^n \to A \) given by

\[ \phi(\lambda) = \lambda_1 b_1 + \cdots + \lambda_n b_n, \quad (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n, \]

is continuous.

The normalised coefficient vectors \( \mu_m = \lambda_m / \|\lambda_m\|_2 \) lie on the unit sphere \( \{ \mu \in \mathbb{R}^n : \|\mu\|_2 = 1 \} \) and so have a convergent subsequence \( \mu_m \) with limit \( \mu^* \), say, where \( \|\mu^*\|_2 = 1 \). By the continuity of \( \phi \), once again,

\[ \mu^*_1 b_1 + \cdots + \mu^*_n b_n = \lim_{k \to \infty} \left( \mu_{1m} b_1 + \cdots + \mu_{nm} b_n \right) = \lim_{k \to \infty} a_{mk} / \|\lambda_{mk}\|_2. \]

Hence \( \|\lambda_{mk}\|_2 \to \infty \) as \( k \to \infty \) (for otherwise, since \( \|a_{mk}\| \leq M \), the above limit is 0, which contradicts the linear independence of \( b_1, \ldots, b_n \)). We may assume, therefore (by taking a further subsequence), that \( \|\lambda_{mk}\|_2 \) is convergent and deduce that \( \lambda_{mk} = \|\lambda_{mk}\|_2 \mu_{mk} \) is convergent, as required.

3. In the proof of Theorem 1.3, the Cauchy–Schwarz inequality is applied with \( f = |e| \) and \( g = 1 \).
Self-assessment questions

S3 Prove that the function \( \phi \) in the above proof is continuous.

S4 Prove that \( \| f \|_1, f \in C[a, b] \), satisfies (N3).

S5 Prove that \( \| f \|_\infty, f \in C[a, b] \), satisfies (N3).

S6 Give an alternative proof of Theorem 1.2 by considering 
\[ A_0 = \{ a \in A : \| a - f \| \leq \| f \| \} \]

S7 Verify equations (1.24), (1.25) and (1.26).

S8 Let \( A \) and \( f \) be as in Figure 1.4. Determine \( \inf_{a \in A} \| f - a \| \) for the 1-norm, the 2-norm and the \( \infty \)-norm.

Problems for Chapter 1

P1 Use first principles to find:
(a) the best approximations to \( f(x) = e^x \) on \([0, 1]\) by a constant, in \( L_1, L_2 \) and \( L_\infty \);
(b) the best approximation to \( f(x) = x^2 \) on \([0, 1]\) by a linear function \( p(x) = ax \), in \( L_\infty \).

P2 Powell Exercise 1.5

P3 Powell Exercise 1.6

P4 Powell Exercise 1.7

P5 Powell Exercise 1.1 (Hint: choose a suitable compact subset of \( A_1 \) and apply SAQ S2.)

Solutions to SAQs in Chapter 1

S1 Since
\[ p(x_i) = c_0 + c_1 x_i, \quad i = 1, 2, \ldots, 5, \]
we have
\[ (p(x_1), p(x_2), p(x_3), p(x_4), p(x_5)) = c_0(1, 1, 1, 1, 1) + c_1(x_1, x_2, x_3, x_4, x_5). \]
Now \((1, 1, 1, 1, 1)\) and \((x_1, x_2, x_3, x_4, x_5)\) are fixed vectors in \( \mathbb{R}^5 \), which are not linearly dependent, and \( c_0, c_1 \) can take any real values. Hence the set of vectors \((p(x_1), p(x_2), p(x_3), p(x_4), p(x_5))\) forms a 2-dimensional subspace of \( \mathbb{R}^5 \).

S2 Let \( d^* = \inf \{ d(a_1, a_2) : a_1 \in A_1, a_2 \in A_2 \} \) and choose sequences \( a_{1n} \in A_1, a_{2n} \in A_2, n = 1, 2, \ldots, \) such that
\[ \lim_{n \to \infty} d(a_{1n}, a_{2n}) = d^*. \]
By the compactness of \( A_1 \) and \( A_2 \), we can choose common subsequences \( a_{1nk}, a_{2nk}, k = 1, 2, \ldots, \) such that
\[ \lim_{k \to \infty} a_{1nk} = a_1^* \quad \text{and} \quad \lim_{k \to \infty} a_{2nk} = a_2^*. \]
with $a_i^* \in A_1$ and $a_i^* \in A_2$ (first choose a convergent subsequence $a_{1n_k}$ of $a_{1n}$ and then, if necessary, a subsequence of $a_{2nk}$).

Now, by the triangle inequality,

$$d(a_i^*, a_j^*) \leq d(a_i^*, a_{1n_k}) + d(a_{1n_k}, a_{2nk}) + d(a_{2nk}, a_j^*).$$

Letting $k \to \infty$, we deduce that $d(a_i^*, a_j^*) \leq d^*$, so that $d(a_i^*, a_j^*) = d^*$, as required.

S3 If $\lambda = (\lambda_1, \ldots, \lambda_n)$, $\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$, then

$$\phi(\lambda) - \phi(\mu) = (\lambda_1 - \mu_1)b_1 + \cdots + (\lambda_n - \mu_n)b_n.$$ Hence

$$\|\phi(\lambda) - \phi(\mu)\| \leq |\lambda_1 - \mu_1|\|b_1\| + \cdots + |\lambda_n - \mu_n|\|b_n\|$$

$$\leq \|\lambda - \mu\|_2 \max_{1 \leq i \leq n} \|b_i\|$$

say. Hence

$$\|\lambda - \mu\|_2 < \varepsilon / K \Rightarrow \|\phi(\lambda) - \phi(\mu)\| < \varepsilon,$$

which proves that $\phi$ is (uniformly) continuous on $\mathbb{R}^n$.

S4 Let $f, g \in C[a, b]$. Since

$$|f(x) + g(x)| \leq |f(x)| + |g(x)|,$$
for all $x \in [a, b]$, we deduce that

$$\int_a^b |f(x) + g(x)| \, dx \leq \int_a^b |f(x)| \, dx + \int_a^b |g(x)| \, dx,$$

that is,

$$\|f + g\|_1 \leq \|f\|_1 + \|g\|_1,$$

as required.

S5 Let $f, g \in C[a, b]$ and suppose that

$$\max_{a \leq x \leq b} |f(x) + g(x)| = |f(c) + g(c)|,$$
where $c \in [a, b]$ (the maximum is attained because the function $x \mapsto |f(x) + g(x)|$ is continuous on $[a, b]$). Then

$$|f(c) + g(c)| \leq |f(c)| + |g(c)|$$

$$\leq \max_{a \leq x \leq b} |f(x)| + \max_{a \leq x \leq b} |g(x)|,$$

that is,

$$\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty,$$

as required.

S6 The set $A_0$ is compact, being the intersection of a closed ball in $B$ with $A$ and hence a closed subset of a compact set. Thus we can, by Theorem 1.1, choose $a^* \in A_0$ such that

$$\|a - f\| \geq \|a^* - f\|, \quad a \in A_0.$$

To see that

$$\|a - f\| \geq \|a^* - f\|, \quad a \in A,$$

note that if $a \in A \setminus A_0$, then

$$\|a - f\| > \|0 - f\| \geq \|a^* - f\|,$$

since $0 \in A_0$. 11
S7 Since \( \lambda > 0 \), \( 1 - x^\lambda \geq 0 \) for \( 0 \leq x \leq 1 \), so that

\[
\|e\|_1 = \int_0^1 (1 - x^\lambda) \, dx = \left[ x - \frac{x^{\lambda+1}}{\lambda+1} \right]_0^1 = \frac{\lambda}{\lambda + 1},
\]

\[
\|e\|_2^2 = \int_0^1 (1 - x^\lambda)^2 \, dx = \int_0^1 (1 - 2x^\lambda + x^{2\lambda}) \, dx
\]

\[
= \left[ x - \frac{2x^{\lambda+1}}{\lambda+1} + \frac{x^{2\lambda+1}}{2\lambda+1} \right]_0^1 = \frac{2\lambda^2}{(\lambda + 1)(2\lambda + 1)},
\]

\[
\|e\|_\infty = \max_{0 \leq x \leq 1} |1 - x^\lambda| = 1.
\]

S8 \( \inf_{a \in A} \|f - a\|_1 = 1; \inf_{a \in A} \|f - a\|_2 = 1/\sqrt{2}; \inf_{a \in A} \|f - a\|_\infty = 1/2. \)

**Solutions to Problems in Chapter 1**

P1 (a) Let \( p(x) = c \) be a constant approximation to \( f(x) = e^x \). It is clear that the minimum of \( \|f - p\|_1 \) occurs when \( 1 \leq c \leq e \). Since \( e^x - c = 0 \) when \( x = \log c \), with \( e^x - c < 0 \) for \( x < \log c \), we have

\[
\|f - p\|_1 = \int_0^1 |e^x - c| \, dx
\]

\[
= \int_0^{\log c} (e - e^x) \, dx + \int_{\log c}^1 (e^x - c) \, dx
\]

\[
= 2e\log c - 3c + e + 1.
\]

The minimum of this expression occurs when

\[
2(\log c + 1) - 3 = 0 \Rightarrow c = \sqrt{e}.
\]

Hence \( p(x) = \sqrt{e} \) is the best \( L_1 \) approximation, with \( \|f - p\|_1 = (\sqrt{e} - 1)^2 \).

To find the best approximation in \( L_2 \) we minimise

\[
\|f - p\|_2^2 = \int_0^1 (e^x - c)^2 \, dx
\]

\[
= e^2 - 2c(e - 1) + \frac{e^2 - 1}{2}
\]

\[
= (c - (e - 1))^2 + \frac{e^2 - 1}{2} - (e - 1)^2.
\]

Evidently the minimum occurs for \( c = e - 1 \), and so \( p(x) = e - 1 \) is the best \( L_2 \) approximation, with \( \|f - p\|_2 = \sqrt{\frac{1}{2}(3 - e)(e - 1)} \).

The best \( L_\infty \) approximation again occurs when \( 1 \leq c \leq e \), and the maximum error occurs at the ends of the interval, so that

\[
\max_{x \in [0,1]} |e^x - c| = c - 1 = e - c \Rightarrow c = \frac{1}{2}(e + 1).
\]

Hence \( p(x) = \frac{1}{2}(e + 1) \) is the best \( L_\infty \) approximation, with \( \|f - p\|_\infty = \frac{1}{2}(e - 1) \).

(b) The best \( L_\infty \) approximation to \( f(x) = x^2 \) on \([0, 1]\) by \( p(x) = ax \) occurs when \( 0 < a < 1 \). For a given value of \( a \), \( 0 < a < 1 \), there are two candidates for the point \( x \) which maximises \( |x^2 - ax| \), namely the point \( x = 1 \) and the point where \( x^2 - ax \) is at a minimum, which satisfies

\[
2x - a = 0 \Rightarrow x = a/2.
\]
Now if $x = 1$, then
\[ |x^2 - ax| = |1 - a| = 1 - a, \]
and if $x = \frac{1}{2}a$, then
\[ |x^2 - ax| = \frac{1}{2}a\frac{1}{2}a - a = \frac{1}{4}a^2. \]
Since $1 - a$ is decreasing and $\frac{1}{4}a^2$ is increasing, we deduce that the best $L_\infty$ approximation occurs when
\[ 1 - a = \frac{a^2}{4} \Rightarrow a = 2(\sqrt{2} - 1) \approx 0.8284. \]
Hence the best $L_\infty$ approximation is $p(x) = 2(\sqrt{2} - 1)x$, with $\|f - p\|_\infty = 3 - 2\sqrt{2}.$

**P2** Here $\mathcal{A}$ is the set of continuous piecewise-linear functions on $[a,b]$. Given $f \in C[a,b]$ and $\varepsilon > 0$ we must construct such a piecewise-linear function $g$, with
\[ \|f - g\|_\infty < \varepsilon. \]
(Note that the letter $a$ has two meanings in the question.)

Since $f$ must be uniformly continuous on $[a,b]$, there exists $\delta > 0$ such that
\[ |x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon. \]
(3)

Now choose $n > (b - a)/\delta$ and define $x_k = a + (b - a)(k/n)$, for $k = 0, 1, \ldots, n$. Notice that $x_{k+1} - x_k = (b - a)/n < \delta$.

Next define $g(x_k) = f(x_k)$, for $k = 0, 1, \ldots, n$, and extend the function $g$ linearly to each interval $[x_k, x_{k+1}]$ by the formula
\[ g(x) = \frac{(x_{k+1} - x)f(x_k) + (x - x_k)f(x_{k+1})}{x_{k+1} - x_k}. \]

We claim that $\|f - g\|_\infty < \varepsilon$, that is,
\[ |f(x) - g(x)| < \varepsilon, \quad x_k \leq x \leq x_{k+1}, \quad k = 0, 1, \ldots, n - 1. \]

This holds because
\[ \begin{align*}
|f(x) - f(x_k)| < \varepsilon \\
|f(x) - f(x_{k+1})| < \varepsilon 
\end{align*} \]
by condition (3), and the values of $g(x)$, $x_k \leq x \leq x_{k+1}$, lie between $f(x_k)$ and $f(x_{k+1})$. Thus if $f \in C[a,b]$ but $f \notin \mathcal{A}$, then there is no best approximation to $f$ from $\mathcal{A}$ because for any $\varepsilon > 0$ there exists $g \in \mathcal{A}$ such that $\|f - g\|_\infty < \varepsilon$. This example shows that Theorem 1.2 is false if we drop the hypothesis that $\mathcal{A}$ be finite-dimensional.
Let us take $[a, b] = [0, 1]$ for simplicity. The example can always be adapted to $[a, b]$ by a translation.

Consider first the example $e(x) = 1 - x^\lambda$, $0 \leq x \leq 1$. Equations (1.24) and (1.25) give
\[
\frac{\|e\|_2}{\|e\|_1} = \sqrt{\frac{2\lambda + 2}{2\lambda + 1}},
\]
which shows that
\[
1 \leq \frac{\|e\|_2}{\|e\|_1} \leq \sqrt{2}, \quad 0 \leq \lambda < \infty.
\]
However, if we allow negative values of $\lambda$ then $\frac{\|e\|_2}{\|e\|_1}$ becomes unbounded as $\lambda$ tends to $-\frac{1}{2}$ from above. Of course, $f(x) = x^\lambda$ is not continuous on $[0, 1]$ for negative values of $\lambda$, but this observation suggests a possible ‘shape’ for our example.

Consider instead the continuous function
\[
f_\varepsilon(x) = \begin{cases} 1 - x/\varepsilon, & 0 \leq x \leq \varepsilon, \\ 0, & \varepsilon < x \leq 1, \end{cases}
\]
where $0 < \varepsilon < 1$. We have
\[
\|f_\varepsilon\|_1 = \int_0^\varepsilon (1 - x/\varepsilon) dx = \left[ x - \frac{x^2}{2\varepsilon} \right]_0^\varepsilon = \varepsilon/2,
\]
\[
\|f_\varepsilon\|_2^2 = \int_0^\varepsilon (1 - x/\varepsilon)^2 dx = \left[ x - \frac{x^2}{\varepsilon} + \frac{x^3}{3\varepsilon^2} \right]_0^\varepsilon = \varepsilon/3.
\]
Hence $\frac{\|f_\varepsilon\|_2}{\|f_\varepsilon\|_1} = 2/\sqrt{3\varepsilon} \to \infty$ as $\varepsilon \to 0$.

(i) The unit ball for the 1-norm in $\mathbb{R}^3$ is a regular octahedron centred at the origin. The part of its boundary in the first octant has equation $x + y + z = 1$. Thus, as $r$ increases,
\[
\{ a : \|a\|_1 \leq r \}
\]
first meets $3x + 2y + z = 6$ at the point $(2, 0, 0)$, which is the closest point to the origin with respect to the 1-norm.

(ii) The unit ball for the 2-norm in $\mathbb{R}^3$ is the ordinary ball centred at the origin. As $r$ increases,
\[
\{ a : \|a\|_2 \leq r \}
\]
first meets $3x + 2y + z = 6$ at a point $(x, y, z)$ whose normal (to the plane) passes through the origin. Since the line $\{(3k, 2k, k) : k \in \mathbb{R}\}$ is normal to the plane we solve
\[
3(3k) + 2(2k) + k = 6 \quad \Rightarrow \quad k = 3/7.
\]
Thus $(9/7, 6/7, 3/7)$ is the closest point to the origin with respect to the 2-norm.

(iii) The unit ball for the $\infty$-norm in $\mathbb{R}^3$ is the cube with vertices $(\pm 1, \pm 1, \pm 1)$. As $r$ increases,
\[
\{ a : \|a\|_\infty \leq r \}
\]
first meets $3x + 2y + z = 6$ at a point of the form $(k, k, k)$, $k > 0$. Thus $(1, 1, 1)$ is the closest point to the origin with respect to the $\infty$-norm.
The idea, as in Theorem 1.2, is to choose a compact subset $A_2$ of $A_1$ which must contain the point of $A_1$ which is closest to $A_0$. For example, we can choose

$$A_2 = A_1 \cap \{a : \|a\| \leq 2M\},$$

where $M$ is so large that

$$A_0 \subseteq \{a : \|a\| \leq M\}.$$

Then choose $a_0^* \in A_0$ and $a_1^* \in A_2$ (see SAQ S2) such that

$$\|a_0^* - a_1^*\| \leq \|a_0 - a_1\|, \quad a_0 \in A_0, \quad a_1 \in A_2.$$

To prove that

$$\|a_0^* - a_1^*\| \leq \|a_0 - a_1\|, \quad a_0 \in A_0, \quad a_1 \in A_1,$$

note that if $a_0 \in A_0$ and $a_1 \in A_1 \setminus A_2$, then

$$\begin{align*}
\|a_0 - a_1\| &\geq \|a_1\| - \|a_0\| \\
&> 2M - M \\
&\geq \|a_0^*\| \\
&= \|a_0^* - 0\| \\
&\geq \|a_0^* - a_1^*\|,
\end{align*}$$

as required.
Chapter 2  The uniqueness of best approximation

Ideally the method used to choose an approximation from a set $A$ to a given function $f$ should give a unique answer. This chapter is devoted to the study of those conditions under which a best approximation from $A$ to $f$ is unique. Important new concepts introduced include ‘convexity’ and ‘scalar product’.

This chapter splits into TWO study sessions:

**Study session 1**: Sections 2.1 and 2.2.

**Study session 2**: Sections 2.3 and 2.4.

---

**Study Session 1: Convexity**

**Read**  Sections 2.1 and 2.2

**Commentary**

1. The following diagrams illustrate the notion of a convex set and a strictly convex set.

A point $a$ is **interior** to a set $A$ in a metric space if the open ball $\{b : d(a, b) < r\} \subset A$, for some $r > 0$.

2. In the proof of Theorem 2.1, there is no need for modulus signs around $\theta$ and $1 - \theta$, since both quantities are positive.

3. The following diagram illustrates the proof of Theorem 2.3.

\[ s = \frac{1}{2}(s_0 + s_1) + \lambda \left( f - \frac{1}{2}(s_0 + s_1) \right) \]

Note that the number $\lambda \geq 0$, which appears in (2.6), does not need to be maximal. All that is required is $\lambda > 0$ and $s \in A$. 
4. The following diagram illustrates the proof of Theorem 2.4.

Note that any finite-dimensional subspace $\mathcal{A}$ is convex, but not compact (consider the sequence $na$, $n = 1, 2, \ldots$, where $a \neq 0$).

Self-assessment questions

S1 Which of the unit balls in Figure 1.5 (page 10) are strictly convex?

S2 Prove that
(a) the intersection of two convex sets is convex;
(b) the intersection of two strictly convex sets is strictly convex.

S3 Show that the norms (a) $\| \cdot \|_1$ and (b) $\| \cdot \|_\infty$ are not strictly convex on $C[0, 1]$.

Study Session 2: Best approximation operators

Read Sections 2.3 and 2.4

Commentary

1. The following diagram illustrates the definition of the best approximation operator $X$.

A projection operator is one for which $X(X(f)) = X(f)$, that is, the best approximation from $\mathcal{A}$ to a point $a \in \mathcal{A}$ is $a$ itself.

2. The final comment in Section 2.3 relates to the earlier comment on the importance of the continuity of the best approximation operator to computer calculations.
3. A **scalar product** (or inner product) on a linear space \( \mathcal{B} \) is a real-valued function \((a, b), a, b \in \mathcal{B}\), such that for all \( a, b, c \in \mathcal{B} \) and \( \lambda, \mu \in \mathbb{R} \):

(S1) \((a, a) \geq 0\), with equality if and only if \(a = 0\);

(S2) \((a, b) = (b, a)\);

(S3) \((a, \lambda b + \mu c) = \lambda(a, b) + \mu(a, c)\).

Two important scalar products are given in Section 2.4.

In any linear space with a scalar product we can define a norm by the equation

\[ \|a\| = \sqrt{(a, a)}. \]

As usual, only the proof of the triangle inequality requires any work; it follows from a version of the Cauchy–Schwarz inequality (see SAQ S6):

\[ |(a, b)| \leq \|a\| \|b\|, \quad a, b \in \mathcal{B}, \]

together with the identity

\[ \|a + b\|^2 = \|a\|^2 + 2(a, b) + \|b\|^2, \]

which you can easily verify.

We shall meet other examples of scalar products in Chapter 11.

4. The proof of Theorem 2.7 is perhaps more clearly written as follows.

If \( f \neq g \) and \( \|f\|_2 = \|g\|_2 = 1 \), then we have

\[ \|f - g\|^2_2 = 1 - 2(f, g) + 1 > 0 \quad \Rightarrow \quad (f, g) < 1, \]

and hence

\[ \|\theta f + (1 - \theta)g\|^2_2 = \theta^2 + 2\theta(1 - \theta)(f, g) + (1 - \theta)^2 \]

\[ < \theta^2 + 2\theta(1 - \theta) + (1 - \theta)^2 \]

\[ = 1. \]

5. The norms

\[ \|f\|_p = \left( \int_a^b |f(x)|^p \, dx \right)^{1/p}, \quad 1 < p < \infty, \]

are all strictly convex. The proof (for \( p \neq 2 \)) depends on a careful study of the possibility of equality in Hölder’s inequality.

**Self-assessment questions**

**S4** Prove Theorem 2.6. (Hint: consider \( A_0 = \{ a \in A : \|a\| \leq 4\|f\| \} \).

**S5** Let \( w \) be a positive function in \( \mathcal{C}[a, b] \). Prove that

\[ (f, g) = \int_a^b w(x)f(x)g(x) \, dx, \]

is a scalar product on \( \mathcal{C}[a, b] \).

**S6** Prove the Cauchy–Schwarz inequality for scalar products by considering the discriminant of the quadratic expression

\[ (a + \lambda b, a + \lambda b), \quad \lambda \in \mathbb{R}. \]

**S7** Draw figures to illustrate the non-uniqueness of best approximation in the four examples on pages 18–19.
Problems for Chapter 2

P1 Powell Exercise 2.4
P2 Powell Exercise 2.5
P3 Powell Exercise 2.6
P4 Powell Exercise 2.8
P5 Powell Exercise 2.1

Solutions to SAQs in Chapter 2

S1 Only the unit ball in the 2-norm is strictly convex.

S2 (a) Let $S$ and $T$ be convex sets. If $s_0, s_1 \in S \cap T$, then the points

$$s = \theta s_0 + (1 - \theta)s_1, \quad 0 < \theta < 1,$$

also lie in both $S$ and $T$ (since $S$ and $T$ are convex) and hence in $S \cap T$. Thus $S \cap T$ is convex.

(b) The proof is as above, but in addition we observe that if $s$ is an interior point of both $S$ and $T$ then $s$ is an interior point of $S \cap T$.

S3 (a) Consider $f(x) = 2x$ and $g(x) = 2(1 - x)$ on $[0, 1]$. Clearly $\|f\|_1$ and $\|g\|_1 = 1$, but

$$\left( \frac{1}{2}(f + g) \right)(x) = 1, \quad 0 \leq x \leq 1,$$

so that $\| \frac{1}{2}(f + g) \|_1 = 1$. Hence $\| \cdot \|_1$ is not strictly convex.

(b) Consider $f(x) = 1$ and $g(x) = x$ on $[0, 1]$. Clearly $\|f\|_\infty = 1$ and $\|g\|_\infty = 1$, but

$$\left( \frac{1}{2}(f + g) \right)(x) = \frac{1}{2}(1 + x), \quad 0 \leq x \leq 1,$$

so that $\| \frac{1}{2}(f + g) \|_\infty = 1$. Hence $\| \cdot \|_\infty$ is not strictly convex.

S4 The idea, as in Theorem 1.2, is to consider a compact subset of $A$ which is large enough to contain best approximations to all points in a neighbourhood of $f$, and then apply Theorem 2.5.

For example, if we take

$$A_0 = \{ a \in A : \|a\| \leq 4 \|f\| \},$$

then $A_0$ contains the best approximations (from $A$) to all points $g$ such that

$$\|g\| \leq 2\|f\|$$

(because $A_0 \supseteq \{ a \in A : \|a\| \leq 2\|g\| \}$, for all such $g$).

Thus if $\|g\| \leq 2\|f\|$, then the best approximation operator $X_0$ with respect to $A_0$ coincides with the best approximation operator $X$ with respect to $A$.

Applying Theorem 2.5, we find that $X_0$ is continuous at $f$, and so therefore is $X$. 

19
(S1) \( (f, f) = \int_a^b w(x)f(x)^2 \, dx \geq 0 \), since \( w(x)f(x)^2 \geq 0 \), \( a \leq x \leq b \).

Equality can occur only if \( f(x)^2 = 0 \), \( a \leq x \leq b \), since \( w(x) > 0 \), \( a \leq x \leq b \).

(S2) Obvious, by definition.

(S3) \( (f, \lambda g + \mu h) = \int_a^b w(x)f(x)(\lambda g(x) + \mu h(x)) \, dx \)

\[ = \lambda \int_a^b w(x)f(x)g(x) \, dx + \mu \int_a^b w(x)f(x)h(x) \, dx \]

\[ = \lambda (f, g) + \mu (g, h). \]

(S6) Since, for all \( \lambda \in \mathcal{R} \),

\[ 0 \leq (a + \lambda b, a + \lambda b) = \|a\|^2 + 2\lambda (a, b) + \lambda^2 \|b\|^2, \]

we must have \( B^2 - 4AC \leq 0 \), where

\[ A = \|b\|^2, \quad B = 2(a, b), \quad C = \|a\|^2. \]

Thus

\[ 4(a, b)^2 \leq 4\|a\|^2 \|b\|^2 \Rightarrow |(a, b)| \leq \|a\| \|b\|, \]

as required.

(S7) The figures are as follows.

\[ \|f - a\|_1 = 2, \quad |\lambda| \leq 1 \]

\[ \|f - a\|_1 = 5, \quad |\lambda| \leq 1 \]

\[ \|f - a\|_{\infty} = 1, \quad 0 \leq \lambda \leq 1 \]

\[ \|f - a\|_{\infty} = 1, \quad 0 \leq \lambda \leq 1 \]
Solutions to Problems in Chapter 2

P1 Since the unit ball in the $\infty$-norm is a square with sides parallel to the axes, the best approximation in $A$ to a point $f \in \mathbb{R}^2 \setminus A$ is found as follows.

(a) If $f$ lies in one of the shaded sets, then $X(f)$ lies on the circle $\{a : \|a\|_2 = 1\}$ and on a (projection) line through $f$ at $45^\circ$ to the axes.

(b) If $f$ does not lie in one of the shaded sets, then $X(f)$ is the nearest of the four points $(\pm 1, 0), (0, \pm 1)$.

To prove directly (that is, without the help of Theorem 2.6) that $X(f)$ is continuous, suppose first that $f_1, f_2$ lie in the shaded set in the first quadrant. Then

$$\|f_1 - f_2\|_\infty \geq d/\sqrt{2},$$

where $d$ is the (ordinary) distance between the $45^\circ$ projection lines through $f_1$ and $f_2$.

Furthermore,

$$\|X(f_1) - X(f_2)\|_\infty \leq \sqrt{2}d,$$

since the line segment joining $X(f_1)$ to $X(f_2)$ makes an angle of more than $45^\circ$ with the projection lines from $f_1, f_2$. Hence

$$\|X(f_1) - X(f_2)\|_\infty \leq 2\|f_1 - f_2\|_\infty.$$ 

It follows that $X$ is continuous on the shaded sets. Since $X$ is constant on the four unshaded sets in $\mathbb{R}^2 \setminus A$ (and these constant values agree with the values of $X$ on the boundaries between the shaded and unshaded sets in $\mathbb{R}^2 \setminus A$) and $X$ is the identity on $A$ itself, we deduce that $X$ is continuous on the whole of $\mathbb{R}^2$. 

To prove that $X(f) = f/\|f\|$, if $\|f\| > 1$, we have to show that

$$\|f - g\| \geq \|f - f/\|f\|\|, \quad g \in A.$$  

But, by the ‘backwards’ form of the triangle inequality,

$$\|f - g\| \geq \|f\| - \|g\|  
\geq \|f\| - 1 \quad \text{(since } g \in A)  
= \|f\|(1 - 1/\|f\|)  
= \|f - f/\|f\|\|,$$

as required. (Where did we use the fact that $\|f\| > 1$?)

To prove that

$$\|X(f_1) - X(f_2)\| \leq 2\|f_1 - f_2\|, \quad f_1, f_2 \in B,$$

it is sufficient to consider three cases.

**Case 1** $\|f_1\| \leq 1, \|f_2\| \leq 1$.

In this case $X(f_1) = f_1$ and $X(f_2) = f_2$, so that

$$\|X(f_1) - X(f_2)\| = \|f_1 - f_2\|.$$  

**Case 2** $\|f_1\| \leq 1, \|f_2\| > 1$.

In this case $X(f_1) = f_1$ and $X(f_2) = f_2/\|f_2\|$, so that

$$\|X(f_1) - X(f_2)\| = \left\|f_1 - f_2 \frac{1}{\|f_2\|} \right\|  
= \left\|f_1 - f_2 + f_2 \left(1 - \frac{1}{\|f_2\|}\right)\right\|  
\leq \|f_1 - f_2\| + \|f_2\| \left(1 - \frac{1}{\|f_2\|}\right) \quad \text{(since } \|f_2\| > 1)  
= \|f_1 - f_2\| + \|f_2\| - 1  
\leq \|f_1 - f_2\| + \|f_2\| - \|f_1\| \quad \text{(since } \|f_1\| \leq 1)  
\leq 2\|f_1 - f_2\|.$$  

**Case 3** $\|f_1\| > 1, \|f_2\| \geq \|f_1\|$.  

In this case $X(f_1) = f_1/\|f_1\|$ and $X(f_2) = f_2/\|f_2\|$, so that

$$\|X(f_1) - X(f_2)\| = \left\|\frac{f_1}{\|f_1\|} - \frac{f_2}{\|f_2\|}\right\|  
= \frac{1}{\|f_1\|} \left\|f_1 - f_2 + f_2 \left(1 - \frac{\|f_1\|}{\|f_2\|}\right)\right\|  
\leq \|f_1 - f_2\| + \|f_2\| - \|f_1\| \quad \text{(since } \|f_1\| > 1)  
\leq 2\|f_1 - f_2\|.$$  

First we remark that the sum of two norms on a linear space is also a norm on that space.

To prove that

$$\|f\| = \|f\|_1 + \|f\|_\infty, \quad f \in C[-\pi, \pi],$$

is not strictly convex, let $A$ be the 1-dimensional subspace of functions of the form

$$g(x) = \lambda \sin^2 x, \quad -\pi \leq x \leq \pi,$$

where $\lambda \in \mathbb{R}$, and let $f(x) = x, -\pi \leq x \leq \pi$.

For $|\lambda| \leq 1$, the graph $y = \lambda \sin^2 x$ meets $y = x$ only at the origin since

$$|\lambda \sin^2 x| \leq |\sin x| < |x|, \quad \text{for } x \neq 0.$$
Hence, by the evenness of $\sin^2 x$, $\|f - g\|_1 = \pi^2$, for $|\lambda| \leq 1$.

Also, for $|\lambda| < 1$, the function

$$e(x) = x - \lambda \sin^2 x$$

has no local maximum or minimum on $\mathcal{R}$, since

$$e'(x) = 1 - \lambda \sin 2x > 0, \quad x \in \mathcal{R}.$$  

Hence

$$\|f - g\|_\infty = \max_{-\pi \leq x \leq \pi} |e(x)| = \max \{e(\pi), -e(-\pi)\} = \pi.$$

Thus

$$\|f - g\| = \pi^2 + \pi,$$

for $g(x) = \lambda \sin^2 x$, $|\lambda| \leq 1$.

Since it is also clear that

$$\|f - g\| \geq \pi^2 + \pi, \quad \lambda \in \mathcal{R},$$

we deduce that $f$ does not have a unique best approximation in $\mathcal{A}$. Hence, by Theorem 2.4, this norm is not strictly convex.

P4 Recall that the unit ball of $\mathbb{R}^3$ in the 1-norm is the regular octahedron whose face in the first octant lies on $x + y + z = 1$, and the unit ball of $\mathbb{R}^3$ in the $\infty$-norm is the cube with vertices $(\pm 1, \pm 1, \pm 1)$.

Thus the plane $x + y = 1$ meets the boundary of the unit ball in the 1-norm in a line segment and also meets the boundary of the ball of radius $\frac{1}{2}$ in the $\infty$-norm in a line segment.
This question shows that any closed, bounded, convex subset $A$ of a linear space $B$, with the property that $f \in A \Rightarrow -f \in A$, is the unit ball of some norm, namely that given by

$$
\|f\| = \begin{cases} 
0, & \text{if } f = 0, \\
\min \{ r \in (0, \infty) : f/r \in A \}, & \text{if } f \neq 0.
\end{cases}
$$

First note that the minimum in this definition is attained. Indeed, if we first define

$$
\|f\| = \inf \{ r \in (0, \infty) : f/r \in A \}, \quad f \neq 0,
$$

then $\|f\| > 0$, otherwise $A$ is unbounded. Also, there is a sequence $r_n \to \|f\|$, such that $f/r_n \in A$, and since $f/r_n \to f/\|f\|$ we deduce that $f/\|f\| \in A$, as required.

(N1) Certainly $\|f\| \geq 0$, by definition and we have just seen that $\|f\| > 0$ for $f \neq 0$.

(N2) $\|\lambda f\| = |\lambda| \|f\|$, for $\lambda \in \mathbb{R}$.

It is clear that this holds if $f = 0$ or $\lambda = 0$.

If $f \neq 0$ and $\lambda \neq 0$, then

$$
\|\lambda f\| = \min \{ r \in (0, \infty) : \lambda f/r \in A \}
$$

$$
= \min \{ r \in (0, \infty) : |\lambda| f/r \in A \} \quad \text{(since } f \in A \iff -f \in A) 
$$

$$
= \min \{ |\lambda| r \in (0, \infty) : |\lambda| f/(r|\lambda|) \in A \}
$$

$$
= |\lambda| \min \{ r \in (0, \infty) : f/r \in A \}
$$

$$
= |\lambda| \|f\|,
$$

as required.

(N3) $\|f + g\| \leq \|f\| + \|g\|$

At first sight this looks difficult. However, by the definition of $\|f + g\|$, it is sufficient to prove that

$$
\frac{f + g}{\|f\| + \|g\|} \in A. \quad (1)
$$

We know that $f/\|f\| \in A$ and $g/\|g\| \in A$ so that, by the convexity of $A$,

$$
\theta \frac{f}{\|f\|} + (1 - \theta) \frac{g}{\|g\|} \in A, \quad 0 < \theta < 1.
$$

If we now choose

$$
\theta = \frac{\|f\|}{\|f\| + \|g\|} \quad \text{so that} \quad 1 - \theta = \frac{\|g\|}{\|f\| + \|g\|},
$$

then we obtain (1).

The above argument breaks down if $\|f\| = \|g\| = 0$, but in this case $f = g = 0$, and so $\|f + g\| = 0$ also.

Hence $\|f\|, f \in B$, is indeed a norm on $B$. 

24
Chapter 3  Approximation operators and some approximating functions

Calculating a best approximation from a subspace $A$ of $B$ to $f$, with respect to some norm on $B$, may not be as easy as calculating an approximation using some other operator, such as an interpolation operator. To judge how good an approximation is obtained by such an operator $X$, we use the ‘norm’ $\|X\|$ of the operator, which is exploited in Sections 3.1 and 3.2. The other two sections of the chapter contain a discussion of the problems involved in approximating by polynomials and an introduction to piecewise polynomial approximation.

This chapter splits into TWO study sessions:

Study session 1: Sections 3.1 and 3.2.
Study session 2: Sections 3.3 and 3.4.

Study Session 1: ‘Good’ versus ‘best’ approximation

Read  Sections 3.1 and 3.2

Commentary

1. Powell defines $\|X\|$ to be the smallest real number such that

$$\|X(f)\| \leq \|X\| \|f\|, \quad f \in B.$$ 

Otherwise stated,

$$\|X\| = \sup \{ \|X(f)\|/\|f\| : f \in B, f \neq 0 \}.$$ 

Thus to determine the value of $\|X\|$, we must find a number $M$ such that

$$\|X(f)\| \leq M\|f\|, \quad f \in B,$$

and such that, whenever $M' < M$, there exists $f \in B$ with

$$\|X(f)\| > M'\|f\|.$$ 

If $\|X(f)\|/\|f\|$ is unbounded on $B$, then we say that the operator $X$ is unbounded.

Notice that if $X$ is a linear operator, then

$$\frac{\|X(\lambda f)\|}{\|\lambda f\|} = \frac{\|X(f)\|}{\|f\|}, \quad f \neq 0, \lambda \neq 0,$$

so that

$$\|X\| = \sup \{ \|X(f)\| : \|f\| = 1 \}.$$ 

In general, the supremum may not be attained because the unit sphere $\{f \in B : \|f\| = 1\}$ need not be compact (see, for example, Problem P1).
2. The following diagram may help you with Theorem 3.1.

![Diagram](image)

[Note the word ‘a’ in the first line of the proof of Theorem 3.1.]

3. Page 25, line 13. The reason why \( p^*(x) = x - \frac{1}{2} \) is the best \( L_\infty \) approximation by a linear polynomial to \( f(x) = x^2 \) on \([0, 1]\) will become clear in Chapter 7.

4. The point of the final paragraph of Section 3.2 is that algorithms for calculating best \( L_\infty \) approximations from \( P_n \) are more involved than those for applying linear (projection) operators \( X : \mathcal{B} \to \mathcal{A} \), such as interpolation.

If we determine \( Xf \) and compute \( f(x) - Xf(x) \) at various points, finding an \( x \) for which

\[
|f(x) - Xf(x)| > (1 + \|X\|)\varepsilon,
\]

then, by Theorem 3.1, the best approximation \( p^* \) will satisfy \( \|f - p^*\| > \varepsilon \). Thus a larger value of \( n \) may be required.

Self-assessment questions

S1 Show that the interpolation operator \( X \) defined at the bottom of page 23 is unbounded with respect to (a) the 1-norm, (b) the 2-norm.

S2 Explain why the application of this operator \( X \) to \( f(x) = x^2 \) (see equation (3.15)) shows that we can have equality in (3.11).

Study Session 2: Types of approximating functions

---

Read Sections 3.3 and 3.4

Commentary

1. Page 26, line 9. The promised technique appears in equation (3.23).

2. The space \( C^{(k)}[a,b] \). An example of a function \( f \) which belongs to \( C^{(k)}[a,b] \), but not to \( C^{(k+1)}[a,b] \) is

\[
f(x) = \begin{cases} 
x^{k+1}, & x \geq 0, 
-x^{k+1}, & x < 0.
\end{cases}
\]

For this function,

\[
f^{(k)}(x) = (k + 1)!|x|, \quad x \in \mathcal{R},
\]

which is continuous but not differentiable at 0.
3. The identity (3.23) holds because \( q \in P_n \) so that, as \( p \) varies over the whole space \( P_n \), so \( q + p \) varies over the whole of \( P_n \).

4. Table 3.1. For \( k = 1 \), the terms \( d_n^*(f) \) scale by a factor of approximately \( \frac{1}{2} \) as \( n \) doubles, whereas, for \( k = 3 \), they scale by approximately \( \frac{1}{8} \). This gives

\[
\begin{align*}
d_n^*(f) &\approx \frac{C_1}{2^n}, \quad k = 1, \\
d_n^*(f) &\approx \frac{C_3}{8^n} = \frac{C_3}{(2^n)^3}, \quad k = 3,
\end{align*}
\]

which suggests that \( d_n^*(f) \approx C_k/n^k \) in both cases.

Notice that in (3.20), for a fixed value of \( k \),

\[
\frac{(n-k)!}{n!} = \frac{1}{n(n-1) \ldots (n-k+1)} \sim \frac{1}{n^k} \text{ as } n \to \infty.
\]

(We say that \( a_n \sim b_n \) as \( n \to \infty \) if \( \lim_{n \to \infty} \left( \frac{a_n}{b_n} \right) = 1 \).)

5. Page 29, line 8. An analytic function is one which has a power series expansion about each point of its domain of definition. Such functions are completely determined by their values on any given open interval, no matter how short.

6. Page 29, line 6-. The spline function \( s \) is a piecewise polynomial on \([a, b]\), such that

\[
s(x) = p_j(x), \quad \xi_{j-1} \leq x \leq \xi_j, \quad j = 1, \ldots, n,
\]

where each \( p_j \in P_k \), and

\[
p^{(i)}_j(\xi_j) = p^{(i)}_{j+1}(\xi_j), \quad i = 0, 1, \ldots, k - 1, \quad j = 1, \ldots, n - 1.
\]

---

Formula (3.31) can be explained as follows. First note that

\[
s(x) = p_1(x), \quad \xi_0 \leq x \leq \xi_1.
\]

Next put

\[
q_1(x) = p_2(x) - p_1(x), \quad x \in R,
\]

so that

\[
q_1(\xi_1) = q_1'(\xi_1) = \cdots = q_1^{(k-1)}(\xi_1) = 0.
\]

Since \( q_1 \) is a polynomial of degree \( k \),

\[
q_1(x) = \frac{d_1}{k!}(x - \xi_1)^k, \quad \text{where } d_1 = q_1^{(k)}(\xi_1).
\]
Hence
\[ s(x) = p_2(x) = p_1(x) + q_1(x), \quad \xi_1 \leq x \leq \xi_2, \]
and so
\[ s(x) = p_1(x) + \frac{d_1}{k!} (x - \xi_1)^k, \quad \xi_0 \leq x \leq \xi_2. \]
Here
\[ (x - \xi_1)^+ = \begin{cases} 0, & x < \xi_1, \\ x - \xi_1, & x \geq \xi_1. \end{cases} \]
Now put
\[ q_2(x) = p_3(x) - p_2(x), \quad x \in \mathcal{R}, \]
and continue in this manner to obtain
\[ s(x) = p_1(x) + \frac{1}{k!} \sum_{j=1}^{n-1} d_j (x - \xi_j)^k, \quad a \leq x \leq b, \]
where \( d_j = q_j^{(k)}(\xi_j) \) and \( q_j = p_{j+1} - p_j \). Thus \( d_j \) is the jump in \( s^{(k)} \) at \( \xi_j \).

7. Page 30, line 9--. The 'big oh' notation used here needs some explanation. We say that
\[ f(x) = O(g(x)), \quad x \in S, \]
for some subset \( S \) of \( \mathbb{R} \), if
\[ |f(x)| \leq M |g(x)|, \quad x \in S, \]
where the constant \( M \) does not depend on \( x \). For example,
\[ x^2 + x = O(x^2), \quad x \geq 1, \]
whereas \( x^2 + x = O(x), \quad 0 \leq x \leq 1 \).

### Self-assessment questions

**S3** The best \( L_\infty \) approximation from \( \mathcal{P}_2 \) to \( f(x) = |x| \) on \([-1, 1]\) is \( p^*(x) = x^2 + \frac{1}{3} \) (this will become clear in Chapter 7). Verify that
\[ ||f - p^*||_\infty = \frac{1}{3}, \]
thus confirming one of the entries in Table 3.1.

**S4** Express the quadratic spline
\[ s(x) = \begin{cases} -x, & -1 \leq x \leq 0, \\ x^2 - x, & 0 \leq x \leq 1, \\ 3x^2 - 5x + 2, & 1 \leq x \leq 2, \end{cases} \]
in the form of equation (3.31).

### Problems for Chapter 3

**P1** Powell Exercise 3.2 (The last part is rather hard and can safely be ignored!)

**P2** Powell Exercise 3.3

**P3** Powell Exercise 3.4

**P4** Powell Exercise 3.6 (Use Theorem 3.1 and be content to get the lower estimate 0.048.)

**P5** Powell Exercise 3.8
Solutions to SAQs in Chapter 3

S1 Consider

\[ f_\varepsilon(x) = \begin{cases} 
 1 - x/\varepsilon, & 0 \leq x \leq \varepsilon, \\
 0, & \varepsilon < x < 1 - \varepsilon, \\
 1 + (x - 1)/\varepsilon, & 1 - \varepsilon \leq x \leq 1. 
\end{cases} \]

(Remember Problem P3, Chapter 1.)

Since \( f_\varepsilon(0) = f_\varepsilon(1) = 1 \), the interpolating function \( p = X(f_\varepsilon) \) is simply \( p(x) = 1, \ 0 \leq x \leq 1 \), and so \( \|p\|_1 = \|p\|_2 = 1 \). However,

\[
\|f_\varepsilon\|_1 = \int_0^\varepsilon (1 - x/\varepsilon) \, dx + \int_{1-\varepsilon}^1 (1 + (x - 1)/\varepsilon) \, dx = \varepsilon,
\]

so that

(a) \[ \frac{\|X(f_\varepsilon)\|_1}{\|f_\varepsilon\|_1} = \frac{1}{\varepsilon} \] is unbounded.

Also

\[
\|f_\varepsilon\|_2^2 = \int_0^\varepsilon (1 - x/\varepsilon)^2 \, dx + \int_{1-\varepsilon}^1 (1 + (x - 1)/\varepsilon)^2 \, dx = 2\varepsilon/3,
\]

so that

(b) \[ \frac{\|X(f_\varepsilon)\|_2}{\|f_\varepsilon\|_2} = \sqrt{\frac{3}{2\varepsilon}} \] is unbounded.

S2 If \( f(x) = x^2, \ 0 \leq x \leq 1 \), and \( p^*(x) = x - 1/8, \ 0 \leq x \leq 1 \), then there are 3 candidates for the point \( x \in [0, 1] \) such that \( \|f - p^*\|_\infty = |f(x) - p^*(x)| \).

These are 0, 1, and the point \( x \) where

\[ e(x) = f(x) - p^*(x) = x^2 - (x - 1/8) \]

is a minimum, which satisfies

\[ e'(x) = 2x - 1 = 0 \quad \Rightarrow \quad x = \frac{1}{2}. \]

Thus

\[ \|f - p^*\|_\infty = \max \{|e(0)|, |e(1)|, |e(1/2)|\} = \frac{1}{3}. \]

(The fact that \( f(x) - p^*(x) \) has the same absolute value, but alternating signs, at each of these 3 points is characteristic of best \( L_\infty \) approximation from certain subspaces — see Chapter 7.)

The corresponding interpolating polynomial is \( p(x) = x, \ 0 \leq x \leq 1 \), and, in this case,

\[ \|f - p\|_\infty = \max \{|x^2 - x| : 0 \leq x \leq 1\} = \frac{1}{8}, \]

since the maximum is taken at \( x = \frac{1}{2} \). Hence \( \|f - p\|_\infty = 2\|f - p^*\|_\infty \) and since \( \|X\|_\infty = 1 \) we do have equality in (3.11).
S3 If \( f(x) = |x|, -1 \leq x \leq 1, \) and \( p^*(x) = x^2 + \frac{1}{x}, -1 \leq x \leq 1, \) then there are 5 candidates for the point \( x \in [-1, 1] \) such that \( \|f - p^*\|_\infty = \|f(x) - p^*(x)\|_\infty \). These are 0, ±1, and the points ±\( x \) where
\[
e(x) = f(x) - p^*(x) = |x| - (x^2 + \frac{1}{x})
\]
is a maximum. As in SAQ S2, we find that \( x = \pm \frac{1}{2}, \) so that
\[
\|f - p^*\|_\infty = \frac{1}{5} = 0.125,
\]
which confirms the first entry in the \( k = 1 \) column of Table 3.1.

S4 Following the proof of (3.31) given in the commentary, put \( p_1(x) = -x, p_2(x) = x^2 - x, p_3(x) = 3x^2 - 5x + 2. \) Then
\[
q_1(x) = p_2(x) - p_1(x) = x^2 - x - (-x) = x^2,
q_2(x) = p_3(x) - p_2(x) = 3x^2 - 5x + 2 - (x^2 - x)
\]
\[= 2x^2 - 4x + 2\]
\[= 2(x - 1)^2.
\]
Hence \( s(x) = -x + (x)^2 + 2(x - 1)^2, -1 \leq x \leq 2. \)

**Solutions to Problems in Chapter 3**

P1 First we seek an expression \( M, \) depending on \( K \) but not on \( x, \) such that
\[
|X f(x)| \leq M \|f\|_\infty, \quad a \leq x \leq b.
\]
For example
\[
|X f(x)| \leq \int_a^b |K(x, y)| |f(y)| \, dy
\]
\[\leq \left( \int_a^b |K(x, y)| \, dy \right) \|f\|_\infty,
\]
so that
\[
\|X f\|_\infty \leq \left( \max_{a \leq x \leq b} \int_a^b |K(x, y)| \, dy \right) \|f\|_\infty.
\]
Hence
\[
\|X\|_\infty \leq \max_{a \leq x \leq b} \int_a^b |K(x, y)| \, dy = \int_a^b |K(x_0, y)| \, dy,
\]
say. To prove that \( \|X\|_\infty \) can be no less than this, we should like to find a function \( f \in C[a, b] \) such that \( \|f\|_\infty = 1 \) and
\[
\|X f\|_\infty = \int_a^b |K(x_0, y)| \, dy.
\]
The ideal function would be
\[
f(y) = \text{sgn}(K(x_0, y)) = \begin{cases} 1, & \text{if } K(x_0, y) > 0, \\ -1, & \text{if } K(x_0, y) < 0, \end{cases}
\]
so that
\[
(X f)(x_0) = \int_a^b K(x_0, y) f(y) \, dy = \int_a^b |K(x_0, y)| \, dy,
\]
which implies that \( \|X\|_\infty \geq \int_a^b |K(x_0, y)| \, dy. \) Unfortunately, however, this function \( f \) is not continuous (unless \( K(x_0, y) \) never vanishes).

Instead we take a continuous approximation \( f_\varepsilon, \varepsilon > 0, \) which differs from \( \text{sgn}(K(x_0, y)) \) only on a set of length less than \( \varepsilon. \)
It follows that
\[
\left| X f_\varepsilon(x_0) - \int_a^b |K(x_0, y)| \, dy \right| = \left| \int_a^b K(x_0, y) (f_\varepsilon(y) - \text{sgn}(K(X_0, y))) \, dy \right|
\]
\[
\leq K \int_a^b |f_\varepsilon(y) - \text{sgn}(K(x_0, y))| \, dy
\]
\[
\leq K \varepsilon,
\]
where \( K = \max_{a \leq y \leq b} |K(x_0, y)| \). Hence
\[
\| X f_\varepsilon \|_\infty \geq \int_a^b |K(x_0, y)| \, dy - K \varepsilon.
\]
Since \( \| f_\varepsilon \|_\infty = 1 \) and \( K \varepsilon \) can be taken arbitrarily small, we deduce that
\[
\| X \|_\infty = \int_a^b |K(x_0, y)| \, dy.
\]

**Remark** If \( \| X \|_\infty = 1 \) and \( X f = f \), then \( f \) need not be a constant. For example, if
\[
K(x, y) = \begin{cases} 4(1 - 2y), & 0 \leq x, y \leq \frac{1}{2}, \\ 8(1 - x)(1 - 2y), & \frac{1}{2} \leq x \leq 1, 0 \leq y \leq \frac{1}{2}, \\ 0, & \frac{1}{2} \leq y \leq 1, \end{cases}
\]
and
\[
f(x) = \int_0^1 K(x, y) \, dy = \begin{cases} 1, & 0 \leq x \leq \frac{1}{2}, \\ 2(1 - x), & \frac{1}{2} \leq x \leq 1, \end{cases}
\]
then \( \| X \|_\infty = 1 \) and
\[
f(x) = \int_0^1 K(x, y) f(y) \, dy, \quad 0 \leq x \leq 1.
\]

**P2** To prove that \( X \) is a projection (it is obviously linear) we have to show that \( X f(x) = f(x) \) when \( f(x) = a + bx \). But
\[
X f(x) = \frac{1}{2} \int_{-1}^1 (1 + 3xy) f(y) \, dy
\]
\[
= \frac{1}{2} \int_{-1}^1 f(y) \, dy + \frac{3x}{2} \int_{-1}^1 yf(y) \, dy
\]
\[
= \frac{1}{2} \int_{-1}^1 (a + by) \, dy + \frac{3x}{2} \int_{-1}^1 (ay + by^2) \, dy
\]
\[
= \frac{1}{2} (2a) + \frac{3x}{2} \left( \frac{2b}{3} \right)
\]
\[
= a + bx,
\]
as required.
Now, by Problem P1,
\[ \|X\|_\infty = \max_{-1 \leq x \leq 1} \frac{1}{2} \int_{-1}^{1} |1 + 3xy| dy \]
\[ = \frac{1}{2} \int_{-1}^{1} |1 + 3y| dy \quad \text{(or } \frac{1}{2} \int_{-1}^{1} |1 - 3y| dy \text{)} \]
To see this, consider the change in the graph \( 1 + 3xy \), \(-1 \leq y \leq 1\), as \( x \) increases from 0 to 1.

Hence
\[ \|X\|_\infty = \frac{1}{2} \left[ \frac{1}{2} (4 \times \frac{3}{4}) + \frac{1}{2} (2 \times \frac{3}{4}) \right] = \frac{5}{8}. \]

**P3** First note that \( X \) is a projection (it is clearly linear), since if \( f(x) = a + bx \), then
\[ Xf(x) = 2 \int_{0}^{1} (a + bt) \, dt + (x - \frac{1}{4}) (a + b - a) \]
\[ = a + b/4 + b(x - \frac{1}{4}) \]
\[ = a + bx, \]
as required.

Now, for \( f \in C[0,1] \),
\[ |Xf(x)| \leq \left| 2 \int_{0}^{1} f(t) \, dt \right| + |x - \frac{1}{4}| |f(1) - f(0)| \]
\[ \leq 2 \int_{0}^{1} |f(t)| \, dt + \frac{3}{4} |f(1)| + |f(0)| \]
\[ \leq \|f\|_\infty + \frac{3}{4} \cdot 2 \|f\|_\infty \]
\[ = \frac{5}{8} \|f\|_\infty. \]

Hence
\[ \|Xf\|_\infty \leq \frac{5}{8} \|f\|_\infty \quad \Rightarrow \quad \|X\|_\infty \leq \frac{5}{2}. \]

Thus, by Theorem 3.1,
\[ \|f - Xf\|_\infty \leq (1 + \frac{5}{2}) \|f - p^*\|_\infty, \]
where \( p^* \) is the best \( L_\infty \) approximation to \( f \) from \( P_1 \), and so
\[ \|f - Xf\|_\infty \leq \frac{7}{2} \|f - p\|_\infty, \quad \text{for } p \in P_1. \]

**Remark** In fact \( \|X\|_\infty = 5/2 \) in this problem, as you can see by considering, for \( 0 < \varepsilon < 1 \),
\[ f_\varepsilon(x) = \begin{cases} -1 + 2x/\varepsilon, & 0 \leq x \leq \varepsilon, \\ 1, & \varepsilon < x \leq 1. \end{cases} \]
There is rather more to this question than meets the eye! First, if \( p(x) = a + bx + cx^2 \), then
\[
p(0) = a, \quad p(1) = a + b + c, \quad p(3) = a + 3b + 9c, \quad p(4) = a + 4b + 16c,
\]
and it is true that
\[
a + 3b + 9c = -\frac{1}{2}a + (a + b + c) + \frac{1}{2}(a + 4b + 16c).
\]

Now
\[
\min_{p \in \mathcal{P}_2} \max_{0 \leq x \leq 4} |f(x) - p(x)| = \|f - p^*\|_\infty,
\]
where \( p^* \) is the best \( L_\infty \) approximation to \( f \) from \( \mathcal{P}_2 \). Thus, by Theorem 3.1,
\[
\|f - p^*\|_\infty \geq \frac{\|f - Xf\|_\infty}{1 + \|X\|_\infty},
\]
where \( X \) is any linear projection from \( C[0,4] \) to \( \mathcal{P}_2 \).

Given the first part of the question, it seems a good idea to let \( X \) be the interpolation operator \( X(f) = p \), defined by \( p(0) = f(0) \), \( p(1) = f(1) \), \( p(4) = f(4) \). This is certainly a linear projection operator. Since
\[
f(3) = -\frac{1}{2}f(0) + f(1) + \frac{1}{2}f(4) \pm 0.15,
\]
we have
\[
\|f - Xf\|_\infty \geq |f(3) - p(3)|
\]
\[
= |\pm 0.15 + (-\frac{1}{2}f(0) + f(1) + \frac{1}{2}f(4)) - (-\frac{1}{2}p(0) + p(1) + \frac{1}{2}p(4))|
\]
\[
= 0.15.
\]

Thus
\[
\|f - p^*\|_\infty \geq \frac{0.15}{1 + \|X\|_\infty}.
\]

To obtain the desired lower estimate for \( \|f - p^*\|_\infty \), we need to show, therefore, that \( \|X\|_\infty \leq 2 \). Unfortunately, it turns out that \( \|X\|_\infty = \frac{17}{8} \). Nevertheless, we indicate the argument.

Recall that, since \( X \) is linear,
\[
\|X\|_\infty = \max \{ \|Xf\|_\infty : \|f\|_\infty = 1 \}.
\]

Thus we seek to maximise the \( L_\infty \) norm of \( X(f) = p \), subject to the constraints \( |p(0)| \leq 1 \), \( |p(1)| \leq 1 \), \( |p(4)| \leq 1 \). It seems likely that \( \|p\|_\infty \) will be largest if we choose \( p(0) \), \( p(1) \) and \( p(4) \) to be \( \pm 1 \). Taking this for granted, there remains only an analysis of the (non-trivial) cases.

As you can easily check, case (c) gives the largest value of \( \|p\|_\infty \). In this case
\[
p(x) = \frac{17}{8} - \frac{1}{2}(x - \frac{5}{2})^2 \quad \Rightarrow \quad \|p\|_\infty = p\left(\frac{5}{2}\right) = \frac{17}{8}.
\]

Hence \( \|X\|_\infty = \frac{17}{8} \), so that
\[
\|f - p^*\|_\infty \geq \frac{0.15}{1 + \frac{17}{8}} = 0.048.
\]
Two questions remain.
(I) How do we justify taking \(p(0), p(1), p(4)\) to be \(\pm 1\)?
(II) Can we in fact obtain the better estimate 0.05?

There are various ways to answer Question (I). For example, we could argue from basic principles, examining the effects of taking \(p(0) = 1, |p(1)| < 1, |p(4)| < 1\), and so on. This would be tedious, and would not generalise to other problems.

More generally, we can use a linear programming argument. This sounds very grand, but it is really quite a simple idea. We want to maximise

\[ |p(x)| = |a + bx + cx^2|, \quad 0 \leq x \leq 4, \]

subject to the constraints

\[ |p(x_i)| = |a + bx_i + cx_i^2| \leq 1, \]

where \(x_1 = 0, x_2 = 1, x_3 = 4\). Now any equation of the form

\[ Xa + Yb + Zc = k, \] where \(X, Y, Z, k\) are constant, defines a plane in \(\mathbb{R}^3\).

Hence the above 3 constraints define a parallelopiped \(P\), centred at the origin, of possible values of \((a, b, c)\) in \(\mathbb{R}^3\).

The required maximum \(M\) of \(|p(x)|\) occurs for some \((a, b, c) \in P\) and \(x_0 \in [0, 4]\), so that

\[ M = \max_{(a, b, c) \in P} |a + bx_0 + cx_0^2|. \]

Since \(x_0\) is now fixed, we can find \(M\) by moving the plane \(a + bx_0 + cx_0^2 = k\) as far as possible from the origin, while still meeting \(P\); at this point \(M = |k|\). Now, however, the plane must pass through at least one vertex of \(P\), so that \(|p(x_i)| = 1\), for \(i = 1, 2, 3\), as required.

We shall see another approach in Chapter 4 which contains a formula for \(\|X\|_{\infty}\), where \(X\) is an interpolation operator from \(C[a, b]\) to \(P_n\).

To answer Question (II) we look again at the proof of Theorem 3.1. Using equation (3.12), we have

\[ 0.15 = |f(3) - (X(f))(3)| \\
= |(f - p^*)(3) - (X(f - p^))(3)| \\
\leq \|f - p^*\|_{\infty} + \|(X(f - p^))(3)\| \]

Now consider the problem of maximising \(|(Xg)(3)|\), for \(g \in C[0, 4]\), while keeping \(\|g\|_{\infty}\) constant. Once again this is a linear programming problem so that the maximum occurs for \(g(0) = \pm\|g\|_{\infty}, g(1) = \pm\|g\|_{\infty}, g(4) = \pm\|g\|_{\infty}\).

Examining cases (a), (b), (c) given earlier, we find that

\[ |(Xg)(3)| \leq 2\|g\|_{\infty}, \quad g \in C[0, 4] \]

((c) is again the extreme case). Hence, with \(g = f - p^*\),

\[ 0.15 \leq \|f - p^*\|_{\infty} + 2\|f - p^*\|_{\infty} = 3\|f - p^*\|_{\infty}, \]

and so \(\|f - p^*\|_{\infty} \geq 0.05\), as required.
This one is a little easier! First, since every quadratic spline is differentiable at points of \((-1, 1)\), we cannot have \(\|f - s\|_\infty = 0\), that is \(s(x) = f(x), -1 \leq x \leq 1\), because \(f\) is not differentiable at 0.

However, we can make \(\|f - s\|_\infty < \varepsilon\) by defining

\[
s(x) = \begin{cases} 
-x, & -1 \leq x \leq -\varepsilon, \\
p(x) = a + bx^2, & -\varepsilon < x < \varepsilon, \\
x, & \varepsilon \leq x \leq 1.
\end{cases}
\]

To guarantee that \(s\) is a quadratic spline, we require

\[
p(\pm\varepsilon) = \varepsilon, \quad \text{that is, } a + b\varepsilon^2 = \varepsilon
\]

\[
p'(\pm\varepsilon) = \pm 1, \quad \text{that is, } 2b\varepsilon = 1
\]

\[
\Rightarrow b = \frac{1}{2\varepsilon}, \quad a = \frac{\varepsilon}{2}
\]

Since the worst error occurs at \(x = 0\), we have

\[
\|f - s\|_\infty = |f(0) - s(0)| = \frac{\varepsilon}{2} < \varepsilon,
\]

as required.
Chapter 4  Polynomial interpolation

This chapter begins a detailed investigation of the interpolation of continuous functions by polynomials. It turns out that the choice of interpolation points makes a considerable difference to the accuracy of the interpolating approximation; for example, we see in this chapter that equally-spaced interpolation points make a rather poor choice. This investigation of interpolation continues in Chapter 5.

This chapter splits into TWO study sessions:

Study session 1: Sections 4.1 and 4.2.
Study session 2: Sections 4.3 and 4.4.

Study Session 1: Polynomial interpolation

Read  Sections 4.1 and 4.2

Commentary

1. Equation (4.2) represents \((n + 1)\) linear equations (one for each interpolation point) with \(n + 1\) unknowns (the coefficients of the required polynomial). Theorem 4.1 shows that the corresponding \((n + 1) \times (n + 1)\) matrix

\[
\begin{pmatrix}
1 & x_0 & x_0^2 & \ldots & x_0^n \\
1 & x_1 & x_1^2 & \ldots & x_1^n \\
\vdots \\
1 & x_n & x_n^2 & \ldots & x_n^n
\end{pmatrix}
\]

must be non-singular. In fact, this Vandermonde matrix, as it is called, has determinant

\[
\prod_{0 \leq i < j \leq n} (x_j - x_i),
\]

which is clearly non-zero if and only if the \(x_i\) are distinct points.

2. Below is the graph of a typical Lagrange function

\[
\ell_k(x) = \prod_{j=0}^{n} \left( \frac{x - x_j}{x_k - x_j} \right)
\]

with \(n = 10\) and \(k = 6\).

![Graph of a typical Lagrange function](image)

Note that if all the \(x_i\) are kept fixed except for \(x_k\) and \(x_{k+1}\) which both tend to a number \(c\), then \(\ell_k(x) \to \infty\) for any \(x \neq x_0, x_1, \ldots, x_{k-1}, c, x_{k+2}, \ldots, x_n\).

3. The useful symbol \(\delta_{ki}\) in equation (4.11) is called the **Kronecker delta**.
4. The remarks before the statement of Theorem 4.2 provide a way of remembering that it is the \((n + 1)\)th derivative of \(f\) which appears in the error formula (4.13).

5. Equation (4.15) can be written as

\[
g^{(n+1)}(\xi) = f^{(n+1)}(\xi) - c(x)(n + 1)! \prod_{i=0}^{n} \frac{1}{(x - x_i)} = 0.
\]

Self-assessment questions

S1 Powell Exercise 4.1

S2 Verify equation (4.11). (This identity is required in Chapter 10 and in Chapter 19.)

S3 Powell Exercise 4.8

Study Session 2: Chebyshev interpolation

Read Sections 4.3 and 4.4

Commentary

1. Table 4.1. Here is the graph of the Runge example together with its Lagrange interpolating polynomial \(p\) of degree 10.

Notice that \(p(4.5) \simeq 1.6\), as indicated by the 5th entry in the middle column of Table 4.1. Here is the graph of the corresponding function

\[
\text{prod}(x) = \prod_{j=0}^{10} (x - x_j).
\]

As you can see, there is a close relationship between \(\text{prod}(x)\) and the size of the error function in the above interpolation.
2. Chebyshev polynomials (pronounced Cheby’shov’ in Russian).

We have

\[ \cos \theta = \cos \theta \quad \Rightarrow \quad T_1(x) = x, \]
\[ \cos 2\theta = 2 \cos^2 \theta - 1 \quad \Rightarrow \quad T_2(x) = 2x^2 - 1 \]
and \[ \cos 3\theta = 4 \cos^3 \theta - 3 \cos \theta \quad \Rightarrow \quad T_3(x) = 4x^3 - 3x. \]

The graphs of these Chebyshev polynomials appear below.

Formula (4.25) shows that \( T_n(x) \) is a polynomial of degree \( n \) in which the coefficient of \( x^n \) is \( 2^{n-1} \). It has zeros at the points \( \cos \left( \frac{(2i-1)\pi}{2n} \right) \), \( i = 1, \ldots, n \), since

\[
T_n \left( \cos \left( \frac{(2i-1)\pi}{2n} \right) \right) = \cos \left( n \cos^{-1} \left( \cos \left( \frac{(2i-1)\pi}{2n} \right) \right) \right) = \cos \left( \frac{(2i-1)\pi}{2} \right) = 0
\]

(note that \( 0 < (2i-1)\pi/(2n) < \pi \), for \( i = 1, \ldots, n \)).

Thus the function \( T_{n+1}(x) \) has \( n + 1 \) (simple) zeros which are, in increasing order,

\[ x_i = \cos \left( \frac{2(n-i)+1}{2(n+1)} \frac{\pi \cdot i}{2(n+1)} \right), \quad i = 0, 1, \ldots, n, \]

and we deduce that

\[ T_{n+1}(x) = 2^n \prod_{i=0}^{n} (x - x_i) \quad \Rightarrow \quad \max_{-1 \leq x \leq 1} |\text{prod}(x)| = \frac{1}{2^n}. \]

The graphs below shows the points \( x_i \) in the case \( n = 10 \) and the graph \( y = T_{11}(x) \).
If \( x_i \) are the Chebyshev points for the interval \([a, b]\), defined by (4.28) and (4.30), and \( t_i \) are the Chebyshev points for \([-1, 1]\), defined by (4.27), then, for \( a \leq x \leq b \) with \( x = \lambda + \mu t \),

\[
\prod(x) = \prod_{i=0}^{n} (x - x_i) \\
= \prod_{i=0}^{n} ((\lambda + \mu t) - (\lambda + \mu t_i)) \\
= \mu^{n+1} \prod_{i=0}^{n} (t - t_i) \\
= \mu^{n+1} \prod(t),
\]

where the latter product is defined with respect to the \( t_i \). Thus

\[
\max_{a \leq x \leq b} |\prod(x)| = \mu^{n+1} \max_{-1 \leq t \leq 1} |\prod(t)| \\
= \left( \frac{b - a}{2} \right)^{n+1} \cdot \frac{1}{2^n} \\
= 2 \left( \frac{b - a}{4} \right)^{n+1}.
\]

For example, with \( n = 10 \) and \([a, b] = [-5, 5]\), this maximum is 47683.7, which is considerably smaller than the corresponding maximum for equally-spaced points. Finally, we plot the interpolating polynomial of degree 10 to the Runge example using these Chebyshev interpolation points.

This graph confirms the fifth entry in the third column of Table 4.4, which gives the maximum error in the above interpolation of approximately 0.1.

3. Theorem 4.3 provides an alternative method of calculating \( \|X\| \) in the solution of Problem P4, Chapter 3. Also, the formula for \( \|X\| \) should remind you of Problem P1, Chapter 3. In the proof, some comments are required about the last two equalities of (4.32).

First, it is legitimate to change the order of the sup and the max because, quite generally, we have

\[
\sup_{x \in X} \sup_{y \in Y} \phi(x, y) = \sup_{y \in Y} \sup_{x \in X} \phi(x, y),
\]

for any real-valued function \( \phi(x, y) \), \( x \in X, y \in Y \). Indeed, for any fixed \( x \in X, y \in Y \),

\[
\phi(x, y) \leq \sup_{\xi \in X} \phi(\xi, y)
\]

so that

\[
\sup_{y \in Y} \phi(x, y) \leq \sup_{y \in Y} \sup_{\xi \in X} \phi(\xi, y).
\]
Thus
\[
\sup_{x \in X} \sup_{y \in Y} \phi(x, y) \leq \sup_{y \in Y} \sup_{x \in X} \phi(x, y),
\]
and the reverse inequality is proved similarly.

Next, note that in the final step of (4.32) it is clear that
\[
\max_{a \leq x \leq b} \|f\| \leq \sum_{k=0}^{n} |\ell_k(x)|,
\]
and equality is seen to hold by taking \(f(x_k) = \text{sgn}(\ell_k(x^*))\), where
\[
\sum_{k=0}^{n} |\ell_k(x^*)| = \max_{a \leq x \leq b} \sum_{k=0}^{n} |\ell_k(x)|.
\]

Note that all the norms in Theorem 4.3 are the \(\infty\)-norm \(\|\cdot\|_\infty\).

4. Table 4.5. It is natural to ask at what rate the norms in the right-hand column are increasing. It can be shown that these grow like \((\frac{2\pi}{\log e}) \log_n n\).

5. The ‘optimal nodes problem’, to find the interpolating points which minimise \|
X\| (see Powell Exercise 4.10) was solved only comparatively recently; see the remarks in Appendix B. Note that the two independent papers [28] and [89], referred to in Appendix B, appeared ‘back-to-back’ in the Journal of Approximation Theory in 1978.

Self-assessment questions

S4 By calculating the interpolating polynomial from \(P_2\) to the Runge example, with suitable interpolation points \(x_0, x_1, x_2\), confirm the first entry in each column of Table 4.1.

S5 (a) Use formula (4.25) to calculate \(T_4(x)\) and \(T_5(x)\).
(b) Prove that \(T_{2n}(x) = 2T_n(x)^2 - 1\).

S6 Check the first entry in both columns of Table 4.5.

Problems for Chapter 4

P1 Powell Exercise 4.2

P2 Powell Exercise 4.3  (Hint: find the maxima of \(|(x - a)(x - b)|\) and \(|(x - a)(x - \frac{1}{2}(a + b))(x - b)|\) on \([a, b]\).)

P3 Powell Exercise 4.4  (Hint: try to find a substitute for the function \(g\) of (4.14).)

P4 Powell Exercise 4.5  (Hint: decide first where the maximum and minimum gaps occur.)

P5 Powell Exercise 4.6
Solutions to SAQs in Chapter 4

S1 Using equation (4.7),
\[ p(x) = f(0) \frac{(x-1)(x-2)(x-3)}{(0-1)(0-2)(0-3)} + f(1) \frac{(x-0)(x-2)(x-3)}{(1-0)(1-2)(1-3)} + f(2) \frac{(x-0)(x-1)(x-3)}{(2-0)(2-1)(2-3)} + f(3) \frac{(x-0)(x-1)(x-2)}{(3-0)(3-1)(3-2)} = -\frac{1}{6}f(0)(x-1)(x-2)(x-3) + \frac{1}{2}f(1)x(x-2)(x-3) - \frac{1}{2}f(2)x(x-1)(x-3) + \frac{1}{6}f(3)x(x-1)(x-2). \]

Hence
\[ p(6) = -10f(0) + 36f(1) - 45f(2) + 20f(3). \]

If \( f(x) = (x-3)^3 \), then \( f(0) = -27, f(1) = -8, f(2) = -1, \) and \( f(3) = 0 \), so
\[ p(6) = -10(-27) + 36(-8) - 45(-1) + 20(0) = 27. \]

This is correct since \( f(6) = 27 \) and \( f \) is a cubic, so the interpolation is exact.

The uncertainty of \( p(6) \), if that of each function value is \( \pm \varepsilon \), is
\[ \pm \sum_{k=0}^{3} \varepsilon |\ell_k(6)| = \pm (10\varepsilon + 36\varepsilon + 45\varepsilon + 20\varepsilon) = \pm 111\varepsilon. \]

S2 Substituting (4.3) in (4.9) gives
\[ \sum_{k=0}^{n} x_k^n \prod_{j=0}^{n} \left( \frac{x-x_j}{x_k-x_j} \right) = x^i, \quad i = 0, 1, \ldots, n, \]

since the interpolation is exact for polynomials of degree \( i \leq n \). The coefficient of \( x^n \) on the right is 1 if \( i = n \) and 0 otherwise, whereas on the left it is
\[ \sum_{k=0}^{n} x_k^n \prod_{j=0}^{n} \frac{1}{(x_k-x_j)}. \]

Hence (4.11) follows.

S3 The Lagrange interpolation formula can be written as follows:
\[ \sum_{k=0}^{n} f(x_k)\ell_k(x) = \sum_{k=0}^{n} f(x_k) \prod_{j=0}^{n} \left( \frac{x-x_j}{x_k-x_j} \right) = \prod_{j=0}^{n} (x-x_j) \sum_{k=0}^{n} \left( \frac{f(x_k)}{\prod_{j=0}^{n} (x_k-x_j)} \right) \frac{1}{x-x_k}. \]

Each of the quantities
\[ \mu_k = \frac{f(x_k)}{\prod_{j=0, j \neq k}^{n} (x_k-x_j)} \]

depends only on the data points and the function values, and so can be calculated beforehand. Hence
\[ p(x) = \prod_{j=0}^{n} (x-x_j) \sum_{k=0}^{n} \frac{\mu_k}{x-x_k}. \]

can be evaluated using \( n+1 \) multiplications, \( n \) additions, \( n+1 \) divisions and \( n+1 \) subtractions. Altogether this is \( 4n+3 \leq 5n \) arithmetic operations \((n \geq 3)\).
S4 First calculate
\[ f(-5) = \frac{1}{26}, \quad f(0) = 1, \quad f(5) = \frac{1}{26}. \]
The unique quadratic \( p(x) \) taking these values is of the form \( p(x) = 1 - ax^2 \), \( a > 0 \). To find \( a \) we use
\[ \frac{1}{26} = 1 - a \cdot 5^2 \Rightarrow a = \frac{1}{26}. \]
Now, with \( n = 2 \) we have \( x_{\frac{5}{2}} = \frac{5}{2} \) and
\[ f\left( \frac{5}{2} \right) = \frac{1}{1 + \frac{25}{4}} = \frac{1}{26} = 0.03807934, \]
\[ p\left( \frac{5}{2} \right) = 1 - \frac{1}{26} \cdot \frac{25}{4} = \frac{79}{100} = 0.799615384. \]
Thus
\[ f\left( \frac{5}{2} \right) - p\left( \frac{5}{2} \right) = -0.62168435, \]
and the verification is complete.

S5 (a) \[ T_2(x) = 2xT_3(x) - T_2(x) \]
\[ = 2x(4x^3 - 3x) - (2x^2 - 1) \]
\[ = 8x^4 - 8x^2 + 1, \]
\[ T_5(x) = 2x(8x^4 - 8x^2 + 1) - (4x^3 - 3x) \]
\[ = 16x^5 - 20x^3 + 5x. \]
(b) \[ T_2n(x) = \cos(2n \cos^{-1} x) \]
\[ = 2 \cos^2(n \cos^{-1} x) - 1 \]
\[ = 2T_n(x)^2 - 1. \]

S6 Let \( X \) be the interpolation operator for the equally-spaced points \( x_0 = -5, x_1 = 0, x_2 = 5 \). Then
\[ \|X\|_\infty = \max \{ \|Xf\|_\infty : \|f\|_\infty = 1 \} \]
\[ = \max \{ \|p\|_\infty : p \in P_2, |p(x_i)| \leq 1, i = 0, 1, 2 \}. \]
As in Problem P4, Chapter 3, the maximum is attained when \( p(x_0), p(x_1), p(x_2) = \pm 1 \), and so we need consider only these cases. It is easy to see that the maximum must occur for \( p(x_0) = -1, p(x_1) = 1, p(x_2) = 1 \), with the polynomial
\[ p(x) = \frac{5}{2} - \frac{1}{26}(x - \frac{5}{2})^2 \Rightarrow \|p\|_\infty = p\left( \frac{5}{2} \right) = 1.25. \]
This confirms the first entry in the left-hand column of Table 4.5.

In the right-hand column, Powell uses the Chebyshev interpolation points given by (4.27), scaled so that the initial and final points map to \(-5\) and 5, respectively (see the discussion before Table 4.5). In the first entry the points given by (4.27) are \(-\sqrt{3}/2, 0, \sqrt{3}/2\), so \( \|X\|_\infty \) is calculated using \( x_0 = -5, x_1 = 0, x_2 = 5 \) again. Hence \( \|X\|_\infty = 1.25 \) again, as required.

Note that if \( \|X\|_\infty \) is calculated on \([-1, 1]\) using the points in (4.27), then the values obtained are different from those in the right-hand column. This is because the points in (4.27) do not include \pm 1. A formula for \( \|X\|_\infty \), calculated in the latter way, is given in Powell Exercise 12.6.
Alternatively, Theorem 4.3 can be used. Here is the calculation for equally-spaced points \( x_0 = -5, \ x_1 = 0, \ x_2 = 5 \). First
\[
\ell_0(x) = \frac{(x - 0)(x - 5)}{(-5 - 0)(-5 - 5)} = \frac{1}{30}x(x - 5),
\]
\[
\ell_1(x) = \frac{(x - (-5))(x - 5)}{(0 - (-5))(0 - 5)} = -\frac{1}{25}(x^2 - 25),
\]
\[
\ell_2(x) = \frac{(x - (-5))(x - 0)}{(5 - (-5))(5 - 0)} = \frac{1}{50}x(x + 5).
\]

Now, for \( 0 \leq x \leq 5 \),
\[
\sum_{k=0}^{2} |\ell_k(x)| = \frac{1}{30}x(5 - x) + \frac{1}{25}(25 - x^2) + \frac{1}{50}x(x + 5)
\]
\[
= \frac{1}{60}(25 + 5x - x^2),
\]
and the maximum of this expression occurs when \( x = \frac{5}{2} \). Hence
\[
\max_{0 \leq x \leq 5} \sum_{k=0}^{2} |\ell_k(x)| = \frac{1}{60}(25 + 5(0.5) - (0.5)^2) = 5/4.
\]

By symmetry, the maximum of this sum will be the same for \(-5 \leq x \leq 0\) and so \( \|X\|_\infty = 5/4 \), as required.

Solutions to Problems in Chapter 4

P1 By Theorem 4.2, applied on \([0, 1]\) to the interpolation points \( \{0, 0.7\} \) and \( \{0.7, 1\} \), there exist \( \xi_0, \ \xi_1 \in [0, 1] \) such that
\[
e_0(x) = f(x) - p_0(x) = \frac{1}{2}x(x - 0.7)f^{(2)}(\xi_0), \ x \in [0, 1],
\]
\[
e_1(x) = f(x) - p_1(x) = \frac{1}{2}(x - 0.7)(x - 1)f^{(2)}(\xi_1), \ x \in [0, 1],
\]
where \( p_0, \ p_1 \in P_1 \) with
\[
p_0(0) = 0, \ p_0(0.7) = p_1(0.7) = 0.7, \ p_1(1) = 0.1.
\]
Now note that
\[
|x(x - 0.7)| \leq |(x - 0.7)(x - 1)| \iff |x| \leq |x - 1| \iff x \leq \frac{1}{2}.
\]
Therefore (assuming that we have no information about \( f^{(2)} \)) it is best to use \( p_0(x) \) for \( 0 < x < \frac{1}{2} \) and \( p_1(x) \) for \( \frac{1}{2} < x < 1 \).

Applying these error estimates at \( x = 0.5 \) itself, where \( p_0(0.5) = 0.5 \) and \( p_1(0.5) = 1.1 \), we obtain
\[
f(0.5) - 0.5 = -0.05f^{(2)}(\xi_0) \leq 0.05\|f^{(2)}\|_\infty,
\]
\[
f(0.5) - 1.1 = 0.05f^{(2)}(\xi_1) \geq -0.05\|f^{(2)}\|_\infty.
\]
Hence
\[
1.1 - 0.05\|f^{(2)}\|_\infty \leq f(0.5) \leq 0.5 + 0.05\|f^{(2)}\|_\infty,
\]
as required. Thus
\[
\|f^{(2)}\|_\infty \geq \frac{1.1 - 0.5}{0.05 + 0.05} = 6.
\]
According to Theorem 4.2, the error in interpolating \( f(x) = \cos x \) over
\([k\pi/n_1, (k+1)\pi/n_1]\) by \( p_1 \in P_1 \), such that \( p_1(k\pi/n_1) = f(k\pi/n_1) \) and
\( p_1((k+1)\pi/n_1) = f((k+1)\pi/n_1) \), is at most
\[
\max_{\frac{k\pi}{n_1} \leq x \leq \frac{(k+1)\pi}{n_1}} \left| \frac{1}{2} \left( x - \frac{k\pi}{n_1} \right) \left( x - \frac{(k+1)\pi}{n_1} \right) \right| \| f^{(2)} \|_{\infty}.
\]

Since \( \| f^{(2)} \|_{\infty} \leq 1 \) and the maximum of \( |(x-a)(x-b)| \) on \([a, b]\) is
\( ((b-a)/2)^2 \), we deduce that
\[
\| f - p_1 \|_{\infty} \leq \frac{1}{4} \left( \frac{\pi}{2n_1} \right)^2 = \frac{\pi^2}{8n_1}.
\]

To guarantee that this error is less than \( 10^{-6} \) it is, therefore, sufficient for \( n_1 \) to satisfy
\[
n_1 > \frac{10^3 \pi}{\sqrt{8}} = 1110.7 \ldots.
\]

Again by Theorem 4.2, the error in interpolating \( f(x) = \cos x \) over
\([k\pi/n_2, (k+1)\pi/n_2]\) by \( p_2 \in P_2 \), such that \( p_2(k\pi/n_2) = f(k\pi/n_2) \),
\( p_2((k+1)\pi/n_2) = f((k+1)\pi/n_2) \), \( p_2((k+2)\pi/n_2) = f((k+2)\pi/n_2) \), is at most
\[
\max_{\frac{k\pi}{n_2} \leq x \leq \frac{(k+2)\pi}{n_2}} \left| \frac{1}{6} \left( x - \frac{k\pi}{n_2} \right) \left( x - \frac{(k+1)\pi}{n_2} \right) \left( x - \frac{(k+2)\pi}{n_2} \right) \right| \| f^{(3)} \|_{\infty}.
\]

Since \( \| f^{(3)} \|_{\infty} \leq 1 \) and the maximum of \( |(x-a)(x-b)(x-c)| \) on
\([a, b, c]\) is \( (2\sqrt{3}/9)((b-a)/2)^3 \), we deduce that
\[
\| f - p_2 \|_{\infty} \leq \frac{1}{6} \frac{2\sqrt{3}}{9} \left( \frac{\pi}{n_2} \right)^3 = \frac{\pi^3}{3^{5/2}n_2^3}.
\]

To guarantee that this error is less than \( 10^{-6} \) it is, therefore, sufficient for \( n_2 \) to satisfy
\[
n_2 > \frac{10^2 \pi}{3^{5/2}} = 125.7 \ldots.
\]

Clearly this exercise is related to Theorem 4.2. To solve it we must find a substitute for the function \( g(t) \) of (4.14). It is natural to consider the function
\[
g(t) = f(t) - p(t) - e(x) \frac{t^n(t-1)^n}{x^n(x-1)^n}, \quad 0 \leq t \leq 1,
\]
and to try and show that \( g^{(2n)}(\xi) = 0 \), for some \( \xi \in [0, 1] \), since this would give
\[
0 = f^{(2n)}(\xi) - e(x) \frac{(2n)!}{x^n(x-1)^n},
\]
as required.

To prove that \( g^{(2n)}(\xi) = 0 \) for some \( \xi \) it is sufficient to prove that \( g^{(n)} = 0 \) for
\( n+1 \) distinct points in \([0, 1]\), since we can then apply Rolle’s theorem \( n \) times, as in the proof of Theorem 4.2.

By the definition of \( g \) we have \( g(x) = 0 \), and
\[
g^{(k)}(0) = 0, \quad g^{(k)}(1) = 0, \quad k = 0, 1, \ldots, n-1.
\]

If we apply Rolle’s theorem to \( g \) on \([0, x]\) and on \([x, 1]\), then we deduce that there are two distinct points \( x_0 \in (0, x) \) and \( x_1 \in (x, 1) \) such that
\[
g'(x_0) = g'(x_1) = 0.
\]
Now apply Rolle’s theorem to \( g' \) on \([0, x_0], [x_0, x_1], [x_1, 1]\), to deduce that there are three distinct points in \((0, 1)\) at which \( g^{(2)} = 0 \). Continuing in this way, we apply Rolle’s theorem \( n \) times to deduce that there are indeed \( n+1 \) distinct points in \((0, 1)\) at which \( g^{(n)} = 0 \), as required.
Since

\[ x_i = \cos \left( \frac{[2(n-i)+1]\pi}{2(n+1)} \right), \quad i = 0, 1, \ldots, n, \]

and the function \( f(x) = \cos x \) is concave on \([0, \pi/2]\) and convex on \([\pi/2, \pi]\), the maximum gap occurs in the middle of the range and the minimum gap occurs at the ends.

If \( n \) is even, then the maximum gap \((i = \frac{1}{2}n, \ i + 1 = \frac{1}{2}(n + 1))\) is

\[
\cos \left( \frac{(n-1)\pi}{2(n+1)} \right) - \cos \left( \frac{(n+1)\pi}{2(n+1)} \right) = 2 \sin \left( \frac{n\pi}{2(n+1)} \right) \sin \left( \frac{\pi}{2(n+1)} \right) < 2 \sin \left( \frac{\pi}{2(n+1)} \right) < \frac{\pi}{n+1},
\]

since \( \sin x < x \), for \( x > 0 \).

If \( n \) is odd, then the maximum gap \((i = \frac{1}{2}(n-1), \ i + 1 = \frac{1}{2}(n+1))\) is

\[
\cos \left( \frac{n\pi}{2(n+1)} \right) - \cos \left( \frac{(n+2)\pi}{2(n+1)} \right) = 2 \sin \left( \frac{(n+1)\pi}{2(n+1)} \right) \sin \left( \frac{\pi}{2(n+1)} \right) = 2 \sin \left( \frac{\pi}{2(n+1)} \right) < \frac{\pi}{n+1}.
\]

Since the gap for \( n + 1 \) equally-spaced points is \( 2/n \), the desired factor is indeed less than \( \pi/2 \) in both cases.

The minimum gap \((i = n-1, \ i + 1 = n)\) is

\[
\cos \left( \frac{\pi}{2(n+1)} \right) - \cos \left( \frac{3\pi}{2(n+1)} \right) = 2 \sin \left( \frac{\pi}{2(n+1)} \right) \sin \left( \frac{\pi}{n+1} \right).
\]

Thus the ratio of the maximum to the minimum gap is

\[
\begin{cases} 
\sin \left( \frac{\frac{n\pi}{2(n+1)}}{\frac{\pi}{n+1}} \right), & n \text{ even}, \\
1/\sin \left( \frac{\pi}{n+1} \right), & n \text{ odd}.
\end{cases}
\]

It is evident that

\[
\frac{1}{\sin \left( \frac{\pi}{n+1} \right)} > \frac{1}{\sin \left( \frac{\pi}{n+1} \right)} = \frac{n + 1}{\pi},
\]

so the required lower estimate clearly holds for \( n \) odd. For \( n \) even, there is a little more work to do, since \( \sin(n\pi/2(n+1)) < 1 \). However,

\[
\sin \left( \frac{n\pi}{2(n+1)} \right) = \sin \left( \frac{\pi}{2} - \frac{\pi}{2(n+1)} \right) = \cos \left( \frac{\pi}{2(n+1)} \right)
\]

and

\[
\sin \left( \frac{\pi}{n+1} \right) = 2 \sin \left( \frac{\pi}{2(n+1)} \right) \cos \left( \frac{\pi}{2(n+1)} \right),
\]

so that

\[
\frac{\sin \left( \frac{n\pi}{2(n+1)} \right)}{\sin \left( \frac{\pi}{n+1} \right)} = \frac{1}{\sin \left( \frac{\pi}{2(n+1)} \right)} > \frac{1}{2 \sin \left( \frac{\pi}{2(n+1)} \right)} = \frac{n + 1}{\pi},
\]

which completes the solution.
We give a solution along the lines of that used to find $\|X\|_\infty$ in Problem P4, Chapter 3. Note that Theorem 4.3 cannot be used because we are not interpolating by a general element of $P_3$. To find $\|X\|_\infty$, we must find the maximum on $[0,3]$ of $|p(x)| = |c_0 + c_1 x + c_3 x^3|$, where $|p(0)| \leq 1$, $|p(2)| \leq 1$ and $|p(3)| \leq 1$; once again, by the linear programming argument, we need consider only the cases $p(0), p(2), p(3) = \pm 1$. The critical cases are sketched below.

In fact, case (c) gives the greatest value of $|p(x)|$ in $[0,3]$. In this case we have $p(0) = 1$, $p(2) = 1$ and $p(3) = -1$, so that

$$p(x) = 1 + \frac{8}{15} x - \frac{2}{15} x^3 \quad \Rightarrow \quad \|p\|_\infty = p(2/\sqrt{3}) = 1 + \frac{32}{45\sqrt{3}},$$

as required.
Chapter 5  Divided differences

In Chapter 4 we found that interpolation can provide a good method of
determining a polynomial approximation to a given function. This chapter is
devoted to a good method of calculating such an interpolating polynomial using a
formula due to Newton which involves divided differences.

This chapter splits into TWO study sessions:
Study session 1: Sections 5.1, 5.2 and 5.3.
Study session 2: Sections 5.4 and 5.5.

Study Session 1: Basic properties of divided
differences

Read  Sections 5.1, 5.2 and 5.3

Commentary

1. The definition of the divided difference given in Section 5.1 makes it clear
that \( f[x_0, x_1, \ldots, x_n] \) is independent of the order in which the points
\( x_0, x_1, \ldots, x_n \) appear. For example
\[
f[x_0, x_1, x_2, x_3] = f[x_1, x_3, x_0, x_2].
\]

2. The remarks at the bottom of page 47 will make more sense after you have
read how to calculate divided differences in Section 5.3.

3. The key features of the Newton formula (5.12) are that, for
\( k = 0, 1, \ldots, n - 1 \),
   (a) the first \( k + 1 \) terms comprise the polynomial \( p_k \in \mathcal{P}_k \) which interpolates
   \( f \) at \( x_0, x_1, \ldots, x_k \);
   (b) the \((k+2)\)th term is an estimate for the error in the approximation of \( f \)
   by \( p_k \).

If a large number of function values are available, therefore, Newton’s
formula should give better and better approximations to \( f \) by choosing more
and more interpolation points. By checking the size of each additional term
calculated, one can decide when further interpolation points are of no help.

4. Some special cases of (5.14) are
\[
f[x_j, x_{j+1}] = \frac{f[x_{j+1}] - f[x_j]}{x_{j+1} - x_j},
\]
\[
f[x_j, x_{j+1}, x_{j+2}] = \frac{f[x_{j+1}, x_{j+2}] - f[x_j, x_{j+1}]}{x_{j+2} - x_j}.
\]
The following diagram may help to interpret (5.14) in general.

The \((k + 1)\)th divided difference is found using the two adjacent terms in the previous column and the corresponding \(x\) values at the ends of the diagonals.

5. The method of calculating divided differences given in Theorem 5.3 explains the remarks at the bottom of page 47. For example, if the data \(f(x_0), f(x_1), \ldots, f(x_n)\) is given and \(h_i = x_i - x_{i-1}, i = 1, 2, \ldots, n\), then adding \(\varepsilon\) to \(f(x_2)\) has the following effect on the table (note how the errors grow rapidly if the \(h\) values are small).

\[
\begin{array}{c|c|c}
 x_0 & 0 & 0 \\
 x_1 & 0 & +\varepsilon / (h_1 h_2 + h_2^2) \\
 x_2 & +\varepsilon / h_2 & -\varepsilon / (h_2 h_3) \\
 x_3 & 0 & +\varepsilon / (h_3^2 + h_3 h_4) \\
 x_4 & 0 & \\
\end{array}
\]

The pattern which emerges in the case of equally-spaced data is investigated in Powell Exercise 5.3 and exploited in Powell Exercise 5.8.

6. When evaluating (5.12) it is sometimes convenient to use the nested form

\[
p_n(x) = f(x_0) + (x - x_0)(f[x_0, x_1] + (x - x_1)(f[x_0, x_1, x_2] + \cdots \\
\cdots (f[x_0, x_1, \ldots, x_{n-1}] + (x - x_{n-1})f[x_0, x_1, \ldots, x_n]) \cdots)).
\]

Self-assessment questions

S1 Powell Exercise 5.1
S2 Powell Exercise 5.2
Study Session 2: Numerical considerations and Hermite interpolation

Read Sections 5.4 and 5.5

Commentary

1. The method of interpolation by calculating the coefficients $c_i$, $i = 0, 1, \ldots, n$, in $p(x) = \sum_{i=0}^{n} c_i x^i$ is, of course, convenient for interpolating by very low degree polynomials, where the coefficients can be found exactly. For higher degree polynomials, however, it is difficult to calculate the coefficients with sufficient accuracy because the corresponding matrix equation may be ill-conditioned.

2. The assertion at the bottom of page 54 that $p$ is a multiple of $\prod_{i=0}^{m} (x - x_i)^{\ell_i+1}$ is true because $f^{(j)}(x_i) = 0, \quad j = 0, 1, \ldots, \ell_i, \quad i = 0, 1, \ldots, m$. This implies that the Taylor expansion of $p$ about each $x_i$ begins

$$p(x) = \frac{f^{(\ell_i+1)}(x_i)}{\ell_i!} (x - x_i)^{\ell_i+1} + \cdots,$$

so that $(x - x_i)^{\ell_i+1}$ is a factor of $p(x)$, for each $i$.

3. The word ‘suitable’ at the top of page 56 can be interpreted to mean ‘valid’.

4. The proof of Theorem 5.5 shows that Hermite interpolation is the limiting case of Newton’s formula (5.12), which is obtained when various adjacent interpolation points merge together.

Self-assessment questions

S3 Verify equation (5.19).

S4 Calculate the value $p(1.8)$ given by (5.20) and confirm the value $p(1.8)$ given by (5.21). (Evaluate these polynomials by nested multiplication:

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = a_0 + x(a_1 + \cdots + x(a_{n-1} + a_n x)\cdots)).$$

S5 Verify that the polynomial (5.29) satisfies the last two interpolation conditions in (5.28).

Problems for Chapter 5

P1 Powell Exercise 5.3

P2 Powell Exercise 5.4

P3 Powell Exercise 5.5

P4 Powell Exercise 5.7

P5 Powell Exercise 5.8
Solutions to SAQs in Chapter 5

S1 The table is as follows.

<table>
<thead>
<tr>
<th>$x_i$</th>
<th>$f(x_i)$</th>
<th>Order 1</th>
<th>Order 2</th>
<th>Order 3</th>
<th>Order 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2</td>
<td>3.28</td>
<td>14.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>17.36</td>
<td>-3.72</td>
<td>-0.8</td>
<td>1.0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>14.96</td>
<td></td>
<td>4.32</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>19.28</td>
<td></td>
<td>6.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>36.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Thus

$$p_4(x) = 3.28 + 14.08(x + 2) - 3.72(x + 2)(x + 1) + (x + 2)(x + 1)(x - 2)$$

and so

$$p_4(4) = 3.28 + 14.08 \times 6 - 3.72 \times 6 \times 5 + 6 \times 5 \times 2 = 36.16,$$

as expected. Note that $p_4(x) = p_4(x)$ in this example.

S2 The required formula for $p'_n(x_0)$ follows from (5.12) by noting that, for $k = 1, 2, \ldots, n - 1,$

$$\frac{d}{dx} (x - x_0) \ldots (x - x_k) = (x - x_1) \ldots (x - x_k) + (x - x_0) \frac{d}{dx} (x - x_1) \ldots (x - x_k)$$

and so

$$\frac{d}{dx} (x - x_0) \ldots (x - x_k)_{|x=x_0} = (x_0 - x_1) \ldots (x_0 - x_k).$$

Thus

$$p'(2) = f[2, 3] + (2 - 3)f[2, 3, 4] + (2 - 3)(2 - 4)f[2, 3, 4, -1]$$
\[+ (2 - 3)(2 - 4)(2 + 1)f[2, 3, 4, -1, -2].\]

By Comment 1 on page 47 and the above table,

$$f[2, 3] = f[3, 2] = 4.32,$$
$$f[2, 3, 4] = f[4, 3, 2] = 6.28,$$
$$f[2, 3, 4, -1] = f[4, 3, 2, -1] = 1,$$

Thus

$$p'(2) = 4.32 + (2 - 3)6.28 + (2 - 3)(2 - 4) = 0.04.$$
In *exact* arithmetic

\[ p(1.8) = 0.0823 - 0.2 \times 0.23633 + 0.2 \times 0.17 \times 0.329 \]
\[ - 0.2 \times 0.17 \times 0.1 \times 0.32887 + 0.2 \times 0.17 \times 0.1 \times 0.04 \times 0.5008 \]
\[ = 0.0823 - 0.047266 + 0.011186 - 0.001118158 + 0.0000681088 \]
\[ = 0.0451699508. \]

(Note that using nested multiplication here may lose you a couple of digits at the end.)

Using (5.20),

\[ p(1.8) = 6.70098 + 1.8(-13.36021 + 1.8(10.3856 + 1.8(-3.69241 + 1.8 \times 0.50272))) \]
\[ = 0.04516435, \]

which agrees with the data to 4 places of decimals. Using (5.21),

\[ p(1.8) = 6.701 + 1.8(-13.36 + 1.8(10.386 + 1.8(-3.6924 + 1.8 \times 0.50272))) \]
\[ = 0.046916672, \]

which agrees with the data to only 2 places of decimals.

Since \(1.8 - 1.6 = 0.2\), \(1.8 - 1.7 = 0.1\) and \(1.8 - 1.8 = 0\),

\[ p(1.8) = 0.08297 + 0.2(-0.246892 + 0.2(0.33592 + 0.1 \times (-0.29735))) \]
\[ = 0.045166, \]

as required. Now

\[
\frac{d}{dx}(x - 1.6)^2 = 2(x - 1.6),
\]

\[
\frac{d}{dx}(x - 1.6)^2(x - 1.7) = 2(x - 1.6)(x - 1.7) + (x - 1.6)^2,
\]

\[
\frac{d}{dx}(x - 1.6)^2(x - 1.7)(x - 1.8) = 2(x - 1.6)(x - 1.7)(x - 1.8) + (x - 1.6)^2(2x - 3.5).
\]

Hence

\[ p'(1.8) = -0.246892 + 0.33592 \times 2 \times 0.2 \]
\[ -0.29735(2 \times 0.2 \times 0.1 + 0.2 \times 0.2) \]
\[ + 0.20375(0.2 \times 0.2 \times 0.1) \]
\[ = -0.135497, \]

as required.

**Remark** In more complicated examples, one might use a more systematic approach to calculate \(p'(x)\), where

\[ p(x) = a_0 + (x - x_0)(a_1 + (x - x_1)(a_2 + \cdots (x - x_{n-1})(a_n + a_{n+1}(x - x_n)) \cdots)). \]

Put

\[ p(x) = q_0(x) = a_0 + (x - x_0)q_1(x) = a_0 + (x - x_0)(a_1 + (x - x_1)q_2(x)) = \ldots, \]

so that

\[ q_n(x) = a_n + a_{n+1}(x - x_n) \quad \text{and} \quad q_{n+1}(x) = a_{n+1}. \]

Then

\[ q_k(x) = a_k + (x - x_k)q_{k+1}(x) \]

and so

\[ q'_k(x) = q_{k+1}(x) + (x - x_k)q'_{k+1}(x). \]

Hence, by induction,

\[ p'(x) = q_1(x) + (x - x_1)(q_2(x) + \cdots (q_n(x) + (x - x_n)q_{n+1}(x)) \cdots)). \]
Solutions to Problems in Chapter 5

P1  First note that with the given values of \( x_i \), we have
\[
\prod_{j=0}^{n} (x_k - x_j) = k!h^k(n-k)!) (-h)^{n-k} = (-1)^{n-k}h^n(k!)(n-k)!,
\]
so that
\[
f[x_0, x_1, \ldots, x_n] = h^{-n} \sum_{k=0}^{n} (-1)^{n-k} \frac{f(x_k)}{k!(n-k)!}.
\]
To verify that this formula is consistent with Theorem 5.3 we note that
\[
f[x_j, \ldots, x_{j+k+1}] = h^{-k-1} \sum_{i=0}^{k+1} (-1)^{k+1-i} \frac{f(x_{i+j})}{i!(k+1-i)!},
\]
\[
f[x_j, \ldots, x_{j+k}] = h^{-k} \sum_{i=0}^{k} (-1)^{k-i} \frac{f(x_{i+j})}{i!(k-i)!},
\]
\[
f[x_{j+1}, \ldots, x_{j+k+1}] = h^{-k} \sum_{i=0}^{k} (-1)^{k-i} \frac{f(x_{i+j+1})}{i!(k-i)!},
\]
and
\[x_{j+k+1} - x_j = (k+1)h.\]
Now, the coefficient of \( f(x_{i+j}) \) in
\[
\frac{f[x_{j+1}, \ldots, x_{j+k+1}] - f[x_j, \ldots, x_{j+k}]}{x_{j+k+1} - x_j}
\]
is
\[
\frac{1}{(k+1)h^{k+1}} \left[ \sum_{i=0}^{k+1} \frac{(-1)^{k+1-i} f(x_{i+j})}{i!(k+1-i)!} - \sum_{i=0}^{k} \frac{(-1)^{k-i} f(x_{i+j})}{i!(k-i)!} \right]
\]
\[
= \frac{(-1)^{k+1-i}}{(k+1)h^{k+1}} \left[ \sum_{i=0}^{k+1} \frac{1}{i!(k+1-i)!} + \sum_{i=0}^{k} \frac{1}{i!(k-i)!} \right]
\]
\[
= \frac{(-1)^{k+1-i}}{h^{k+1}i!(k+1-i)!},
\]
which is the coefficient of \( f(x_{i+j}) \) in \( f[x_j, \ldots, x_{j+k+1}] \). Hence the recurrence relation (5.14) does indeed hold.

P2  The table can be reconstructed from the first entry in each column using (5.14) in the form
\[
f[x_{j+1}, \ldots, x_{j+k+1}] = f[x_j, \ldots, x_{j+k}] + (x_{j+k+1} - x_j) f[x_j, \ldots, x_{j+k+1}].
\]
For example,
\[
f[1.8, 1.76, 1.7, 1.63] = f[1.76, 1.7, 1.63, 1.6] + (1.8 - 1.6) f[1.8, 1.76, 1.7, 1.63, 1.6]
\]
\[= -0.32887 + 0.2 \times 0.50080 \]
\[= -0.22871, \]
\[
f[1.76, 1.7, 1.63] = f[1.7, 1.63, 1.6] + (1.76 - 1.6) f[1.76, 1.7, 1.63, 1.6]
\]
\[= 0.329 + 0.16 \times (-0.32887) \]
\[= 0.2763808, \]
\[
f[1.8, 1.76, 1.7] = f[1.76, 1.7, 1.63] + (1.8 - 1.63) f[1.8, 1.76, 1.7, 1.63]
\]
\[= 0.2763808 + 0.17 \times (-0.22871) \]
\[= 0.2375001. \]
Continuing in this manner, we find
\[f[1.7, 1.63] = -0.23633 + (1.7 - 1.6) \times 0.329 = -0.20343,\]
\[f[1.76, 1.7] = -0.20343 + (1.76 - 1.63) \times 0.2763808 = -0.16750049,\]
\[f[1.8, 1.76] = -0.16750049 + (1.8 - 1.7) \times 0.2375001 = -0.14375048,\]
\[f[1.63] = 0.08230 + (1.63 - 1.6) \times (-0.23633) = 0.0752101,\]
\[f[1.7] = 0.0752101 + (1.7 - 1.63) \times (-0.20343) = 0.06097,\]
\[f[1.76] = 0.06097 + (1.76 - 1.7) \times (-0.16750049) = 0.05091997,\]
\[f[1.8] = 0.05091997 + (1.8 - 1.76) \times (-0.14375048) = 0.045169951.

### P3

The required table is as follows.

\[
\begin{array}{llllll}
 f(0) & f'(0) & f''(0) & f[1,0] & f'(0) - f''(0) & f'(1) - 3f[1,0] + 2f'(0) + \frac{1}{2}f''(0) \\
 f(0) & f'(0) & f''(0) & f[1,0] & f'(0) - f''(0) & f'(1) - 2f[1,0] + f'(0) \\
 f(1) & f'(1) & f'(1) & f'(1) & & \\
 f(1) & & & & & \\
\end{array}
\]

Hence
\[
p(x) = f(0) + xf'(0) + \frac{x^2}{2} f''(0) + x^3 (f[1,0] - f'(0) - \frac{1}{2} f''(0)) \\
+ x^3 (f'(1) - 3f[1,0] + 2f'(0) + \frac{1}{2} f''(0)).
\]

If \( f(x) = (x + 1)^4 \), then \( f(0) = 1, f'(0) = 4, f''(0) = 12, f(1) = 16, \\
f''(1) = 32 \), so that
\[
p(x) = 1 + 4x + 6x^2 + x^3 (15 - 4 - 6) + x^3 (32 - 45 + 8 + 6) \\
= 1 + 4x + 6x^2 + 5x^3 + x^3 (x - 1) \\
= 1 + 4x + 6x^2 + 4x^3 + x^4 \\
= (1 + x)^4,
\]
as required.

### P4

It is easy to prove that if \( f^{(k)} \) is strictly increasing, then the \( k \)th-order differences are increasing. Indeed, by Theorem 5.1 and Theorem 5.3,
\[
f[x_j, \ldots, x_{j+k}] = (x_{j+k} - x_j) f[x_j, \ldots, x_{j+k+1}] \\
= (x_{j+k+1} - x_j) f^{(k+1)}(\xi) \frac{1}{(k+1)!},
\]
for some \( \xi \in [x_j, x_{j+k+1}] \). Since \( f^{(k)} \) is strictly increasing, we have
\( f^{(k+1)}(\xi) \geq 0 \). Thus
\[
f[x_j, \ldots, x_{j+k+1}] \geq f[x_j, \ldots, x_{j+k}].
\]

Some extra work is required to prove that this inequality must be strict. If it is not, then \( f[x_j, \ldots, x_{j+k+1}] = 0 \), and so the polynomial \( p \) in \( \mathcal{P}_{k+1} \) which interpolates \( f \) at \( x_j, \ldots, x_{j+k+1} \) actually lies in \( \mathcal{P}_k \). Therefore, \( e = f - p \) has (at least) \( k + 2 \) zeros and so, by Rolle’s Theorem, \( e^{(k)} \) has (at least) \( 2 \) zeros. Hence \( f^{(k)} = p^{(k)} \) at (at least) 2 points. But \( p^{(k)} \) is a constant and \( f^{(k)} \) is strictly increasing — a contradiction. Hence the \( k \)th-order differences are strictly increasing.
The difference table is as follows.

<table>
<thead>
<tr>
<th>$x_i$</th>
<th>$f(x_i)$</th>
<th>Order 1</th>
<th>Order 2</th>
<th>Order 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.11978</td>
<td>0.00957</td>
<td>0.00018</td>
</tr>
<tr>
<td>0.1</td>
<td>0.11978</td>
<td>0.129348</td>
<td>0.009588</td>
<td>-0.002982</td>
</tr>
<tr>
<td>0.2</td>
<td>0.249126</td>
<td>0.138936</td>
<td>0.006606</td>
<td>0.009015</td>
</tr>
<tr>
<td>0.3</td>
<td>0.38862</td>
<td>0.145542</td>
<td>0.015621</td>
<td>-0.008982</td>
</tr>
<tr>
<td>0.4</td>
<td>0.533604</td>
<td>0.161163</td>
<td>0.006639</td>
<td>0.003013</td>
</tr>
<tr>
<td>0.5</td>
<td>0.694767</td>
<td>0.167802</td>
<td>0.009652</td>
<td>0.000005</td>
</tr>
<tr>
<td>0.6</td>
<td>0.862569</td>
<td>0.177454</td>
<td>0.009698</td>
<td>0.000041</td>
</tr>
<tr>
<td>0.7</td>
<td>1.040023</td>
<td>0.187111</td>
<td>0.009689</td>
<td>-0.000015</td>
</tr>
<tr>
<td>0.8</td>
<td>1.227134</td>
<td>0.196809</td>
<td>0.009683</td>
<td>0.206492</td>
</tr>
<tr>
<td>0.9</td>
<td>1.423943</td>
<td>0.206492</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.630435</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The second-order differences are irregular. Most noticeably, the numbers 0.006606, 0.015621, 0.006639 are substantially different from the other entries, and this can be traced to an error in the value of $f(0.4)$. Indeed, if the above value of $f(0.4)$ is increased by $\varepsilon$, then the increases in the second-order differences are $\varepsilon$, $-2\varepsilon$, $\varepsilon$ respectively, so that their average remains constant at $(0.006606 + 0.015621 + 0.006639)/3 = 0.009622$.

For the middle one of these three differences to equal 0.009622, we require $2\varepsilon = 0.015621 - 0.009622$, that is, $\varepsilon = 0.0029995$. It appears likely, therefore, that $f(0.4)$ should actually be 0.536604. (Alternatively: note that the increases in the corresponding third-order differences are $\varepsilon$, $-3\varepsilon$, $3\varepsilon$, $-\varepsilon$, so that $\varepsilon \simeq 0.003$.)

Once this error is corrected, notice that the second-order differences are increasing, apart from the last one, and that the increase from 0.009657 to 0.009698 is abnormally large. This can be traced to an error in the value of $f(0.8)$. Once again an increase of $\varepsilon$ in the above value of $f(0.8)$ leads to increases in the last three second-order differences of $\varepsilon$, $-2\varepsilon$, $\varepsilon$ respectively. The average of these differences is 0.0096793, which suggests that $2\varepsilon = 0.009698 - 0.0096793$, that is, $\varepsilon = 0.0000093$. It appears likely, therefore, that $f(0.8)$ should actually be 1.227143.

Once these corrections are made, the third-order differences are (giving only the significant digits)

18, 18, 15, 18, 13, 14, 14, 12,

which is quite regular, since the final digit of the data is probably rounded.