Ph 129
Numerical Analysis
Numerical Analysis

E4 : Introduction

E4 : Calculation and Representation of Functions

E4/1 : Difference Operators

E4/2 : Equidistant Interpolation Formulas

E4/3 : General Polynomial Interpolation

E4/4 : Chebyshev Series

E4/5 : Analyticity

E4/6 : Statistical Warming

E5 : Numerical Integration

E5/1 : Newton-Cotes Formulas

E5/2 : Romberg Integration

E5/3 : Gaussian Integration

E5/4 : Monte Carlo Methods (earlier)

The notes include numerical methods for differential equations. We will not cover these or linear algebra methods. These topics are well covered in Applied Mathematics Courses.
E6 Ordinary Differential Equations

The discussion in Mathews and Walker, page 353-355, is quite good. However, as we take relish in later, they perversely describe a very poor version of Predictor-Corrector method. As usual Noble is good and more complete while still being elementary (chapter 10). My bible, modern computing methods, is this time rather poor.

Remember that ordinary differential equations can be divided by their boundary conditions:

\[ A \quad \longrightarrow \quad B \]

\[ e.g. \quad y'' = f(x, y) \]

\[ y, y', \text{ given at } A \text{ is initial value problem} \]

\[ y \text{ given at } A \text{ and } B \text{ is a boundary value problem} \]

The first two sections cover initial value problems. Note that by putting
\[ z = y', \] the ODE becomes \[ z' = f(x, y), \] a first order ODE. This can always be done: we not only assume this form but also consider just one equation

\[ y' = f(x, y) \tag{6.1} \]

we always solve this by setting up a grid of equidistant points \( x_i \) with grid size \( h = \frac{AB}{n} \) where \( n \to \infty \) (a perfect integer). The next two sections describe the two alternative methods of solving this problem.

\( 6.1 \) Runge-Kutta Methods

let \( y_i = y(x_i) \), as usual. Then we trivially have:
\[ y_2 - y_0 = hf(x_0, y_0) + O(h^2) \]  \hspace{1cm} (E6.2)

\[ y_2 - y_0 \equiv \int_{x_0}^{x_2} f(x, y) = \frac{1}{2} h \left\{ f(x_0, y_0) + f(x_1, y_1) \right\} + O(h^3) \]  \hspace{1cm} (E6.3)

Now in (E6.3), we don't know \( y_2 \) on right-hand side but we can use the approximation (E6.2) and introduce an error of same order in \( h \) as intrinsic error in (E6.3).

We must now extend this rule to a higher accuracy by using Simpson's rule for \( y_2 - y_0 \equiv \int_{x_0}^{x_2} f(x, y) \). Thus, introducing \( x_{1/2} \), half way between \( x_0 \) and \( x_2 \), replace (E6.3) by:

\[ y_2 = y_0 + \frac{h}{6} \left\{ f(x_0, y_0) + 4f(x_{1/2}, y_{1/2}) + f(x_1, y_1) \right\} + O(h^5) \]  \hspace{1cm} (E6.4)

Unfortunately, we can no longer use (E6.2) to estimate \( y_{1/2} \) and \( y_2 \) in (E6.4) as this will give the old error \( O(h^2) \) and won't realize accuracy of Simpson's rule (E6.4). However, the choice - on right
hand side of (E6.4) - of
\[ y_{1/2} \Rightarrow y_{1/2}^* = y_0 + \frac{1}{2} h f(x_0, y_0) \] (E6.5)
\[ y_2 \Rightarrow y_2^* = y_0 + h \left[ -f(x_0, y_0) + 2f(x_{1/2}, y_{1/2}^*) \right] \]
gives an error of \( O(h^4) \) in (E6.6).

More generally, we construct scheme
\[ \Delta_1 = h f(x_0, y_0) \] (E6.6)
\[ \Delta_2 = h f(x_0 + \alpha h, y_0 + \alpha_2 A_2) \]
\[ \Delta_3 = h f(x_0 + \beta h, y_0 + \beta_2 A_2 + \beta_3 A_3) \]
\[ h_1, \alpha, \beta = \alpha_1, \beta_1, \ldots, A_i \]
\[ y_1 = y_0 + (\alpha_1 \Delta_1 + \cdots) + R \]
where all the coefficients \( \alpha, \beta, \alpha_i, \beta_i, \) and \( A_i \) are chosen to minimize error \( R \).

The Runge-Kutta fourth order rule is
\[ \Delta_1 = h f(x_0, y_0) \]
\[ \Delta_2 = h f(x_0 + \frac{1}{2} h, y_0 + \frac{1}{2} \Delta_2) \] (E6.7)
\[ \Delta_3 = h f(x_0 + \frac{1}{2} h, y_0 + \frac{1}{2} \Delta_2) \]
\[ \Delta_4 = h f(x_0 + h, y_0 + \Delta_3) \]
and \[ y_2 = \frac{1}{6} \left[ \Delta_1 + 2 \Delta_2 + 2 \Delta_3 + \Delta_4 \right] + R_5 + y_0 \]
and \( R_5 = O(h^5) \) has achieved the full accuracy possible with Simpson's rule.
6/2 Predictor-Corrector Methods

(1) These differ from Runge-Kutta in that they only use grid point values and not the intermediate (e.g. \((x_{i+1/2}, y_{i+1/2})\) in (6.4)) values \(x \in x_0 \leq x \leq x_1\) used in Runge-Kutta.

(6.2, 3) are already a simple predictor-corrector method i.e.

\[
y^*_{i+1} = y_i + h f_i \quad \text{predict, error } O(h) \\
y_{i+1} = y_i + \frac{1}{2} h (f_i + f^*_i) \quad \text{correct, error } O(h^2)
\]

Here \(f_i = f(x_i, y_i)\), \(f^*_i = f(x_{i+1/2}, y^*_i)\).

Generally we use some prediction formula for \(y^*_i\) with error \(O(h^0)\) and a corrector with an error \(O(h^2)\). As above, Substitution of predictor into corrector with keep \(O(h^{\hat{p}+2})\). However, as we will see, if predictor also has \(O(h^{\hat{p}+2})\) error, one can use difference between \(y_i\) and \(y_{i+1}\) to estimate error.
and in fact get a $O(h^{n+2})$ estimate of $y_n$

As mentioned above, in predictor-corrector methods we use previous values $y_0, y_{-1}$ to increase accuracy $p$; not extra values in $x_0 \rightarrow x_1$ which was Runge-Kutta device. We use our difference techniques, defining:

$$g(x) = f(x, y(x))$$

a function of $x$ only. Use backward differences

$$g(x) = g_0 + pg_0 + \frac{1}{2} p(p+1)g_0 + \ldots \quad (66.8)$$

where $p = (x-x_0)/h$

Now

$$y_1 = y_0 + \int_{x_0}^{x_1} g(x) dx$$

$$= y_0 + h \int_{0}^{p} f(p) dp$$

where $f(p) = g(x_0 + hp)$.

The $p$ integrals are easy:

This is prediction

$$y_1' = y_0 + h (1 + \frac{1}{2}D + \frac{1}{2}D^2 + \ldots) f_0 \quad (66.9)$$

where implicitly we take terms up to $D^{n-2}$ to get an $O(h^{n+2})$ error.
As is appropriate, (66.9) uses only \( f_0, f_1, f_2 \ldots \) which are known.

To derive the correction formula, we next use the backward difference formula for \( g(x) \) starting at \( x = x_3 \) where for \( x = x_3 \) we use \( f_3^* = f(x_2, y_2^* \) and \( y_2^* \) is prediction.

\[
g(x) = g_3 + 2Dg_2 + \frac{1}{2} g_1 (x-x_3) D^2 g_2 + \ldots
\]

where \( q = (x-x_3)/h \)

Then, as before, we do \( q \) integrals to find:

\[
y_3 = y_0 + h (1 - \frac{1}{2} D - \frac{1}{2} D^2 \ldots ) f_3^* \quad (66.10)
\]

and again we stop after terms of order \( D^{n-1} \) to get an \( O(h^{n+1}) \) error.

As an example, taking terms up to \( D^2 \) in prediction and correction we have
\[
predict \quad y^*_2 = y_0 + \frac{h}{12} \left[ 23y_0 - 16y_{-1} + 5y_{-2} \right] \quad (66.1.11)
\]
\[
correct \quad y_2 = y_0 + \frac{h}{12} \left[ 5y^*_2 + 8y_0 - y_{-1} \right]
\]

These were written differences out in terms of function values. The errors are:

- Predict: \( \frac{3}{8} h^4 f''_0 \)
- Correct: \( -\frac{h^4}{24} f''_0 \)

where derivatives are evaluated somewhere in range: but to get order in \( h \), we can replace them by \( f''_0 \).

Note the correction formula has an error which is \(-\frac{1}{9}\) that in prediction formula. So estimate of error is: \( \frac{1}{10} (y^*_2 - y_2) \) in \( y_2 \).

Further

\[
y_2 \) final = \frac{1}{10} y^*_2 \) predict + \frac{9}{10} y_2 \) correct
\]

has an error of \( O(h^5) \).
(ii) Error Analysis

The formulae (66.8, 9) are called the Adams-Bashforth method but there are several other possible predictor-corrector methods. For instance, Mathews and Walker discuss one of Milne's methods which uses Simpson's rule to predict \( y_{2} \) by integrating not from \( y_{0} \) but from \( y_{-1} \):

\[
y_{2} = y_{-1} + \frac{h}{3} \left[ f_{2} + 4f_{0} + f_{-2} \right] \quad (66.12)
\]

There is a difficulty with this scheme for \( h \rightarrow 0 \). In (66.8, 9) we get

\[ Y_{n+1} = Y_{n} \]

But in (66.12), we find

\[ Y_{n+1} = Y_{n-2} \]

To solve last equations, set \( y = a^{n} \). Then \( a = 1 \) for AB (Adams-Bashforth) but \( a = \pm 1 \) for Milne. The latter has the desired solution plus an
oscillation $1 -1 1 -1 1 -1$
$n \to n+2 \to n+2 \to n+3 \ldots$

and so the method is unstable.

Generally, if you use a predictor-corrector formula

$$y_1 = a_0 y_0 + a_{-1} y_{-1} + a_{-2} y_{-2} + O(h)$$

then stability demands that if $2$ solve

$$z^m = a_0 z^{m-1} + a_{-1} z^{m-2} + \ldots + a_{-m}$$

this equation has

(a) One root $= 1$ (guaranteed if it is a correct interpolation formula)
(b) All other roots have modulus $< 1$. (This makes it certain that anomalous solutions to $y'' = 0$ equations will damp out as we write in $n$ like $x^n$ where $\lambda$ modulus of root).

Note that the $O(h)$ terms in equation will affect roots by $O(h)$ i.e., it can
in Milnes's method take nasty -1 out to $-1 + 2h$ and method is stable. However, you only have to change a sign in equation to get $-1 - 2h$ and due to

**E6/3 Boundary Value Problem**

\[ A \begin{pmatrix} x_1 \\
\vdots \\
\vdots \\
x_n \\
\end{pmatrix} = \begin{pmatrix} B \\
\vdots \\
\vdots \\
x_{n+1} \end{pmatrix} \]

We have two basic problems:

\[ y'' = f(x, y) \quad (E6.16) \]

\[ y = y_0 \quad \text{at} \quad x = x_0, \quad y = y_n \quad \text{at} \quad x = x_{n+1} \]

or the eigenvalue problem

\[ y'' + \lambda q(x)y = 0 \quad (E6.17) \]

\[ y = 0 \quad \text{at} \quad x = x_0 \quad \text{and} \quad x_{n+1} \]

In E6/3, we were solving recursion relations: (E6.16) gives matrix inversion and (E6.17) matrix eigenvalue problem.

Take to begin with, (E6.16):

We use finite difference methods to express \((D = d/dx)\)

\[ h^2 D^2 = (2 \sinh^{-1} \frac{1}{2} s)^2 \]
or 
\[ h^2 y'' = \delta^2 - \delta^4 + \delta^6 \quad (E6.18) \]

i.e. 
\[ y'' = h^2 \delta^2 y + O(h^4) \]

or applying it at each of grid points and writing differences in terms of function values we get:
\[ y''_r = y''(x_r) = h^2 \left[ y_{r+1} - 2y_r + y_{r-1} \right] \]

and so \((E6.16) \) becomes
\[ 2y_2 - y_2 + h^2 f_2 = y_0 \]
\[ -y_2 - y_3 + h^2 f_2 = 0 \]
\[ -y_3 + 2y_r - y_{r+1} + h^2 f_r = 0 \]
\[ -y_{r-1} + 2y_r - y_{r+1} + h^2 f_r = 0 \]
\[ -y_{n-1} + 2y_n + h^2 f_n = y_{n+1} \]

If \( f \) is a function of \( y \) only, we can solve the above in the form:
\[ A \ y = -h^2 f + B \quad (E6.19) \]

where 
\[ y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad f = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix} \quad B = \begin{bmatrix} y_0 \\ \vdots \\ y_{n+1} \end{bmatrix} \]
\[ A = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ \vdots & & & \vdots \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & \end{bmatrix} \]

is an \( n \times n \) matrix.

Solving (E6.19) gives us a solution which can be improved/checksed by either changing grid size or using approximate solution to calculate \( \delta \), \( \delta' \), \( \delta'' \) ... corrects in (E6.18).

In eigenvalue case, we also have (E6.14) with \( f(x,y) = x \cdot y \cdot q(x) \) and \( y_0 = y_{n+1} = 0 \).

Put \( Q = \begin{bmatrix} q_1 & 0 \\ 0 & q_2 \\ \vdots & \vdots \\ 0 & q_n \end{bmatrix} \) a diagonal matrix.

Then we get
\[ A \cdot y + h^2 \cdot \lambda \cdot Q \cdot y = 0 \]

an eigenvalue equation for \( Q^{-1}A \).

One could also use calculus of variation techniques (Mathews and Walker p.322 onwards) to solve eigenvalue problen
E7: Partial Differential Equations

These are not subject to very different numerical techniques from ordinary differential equations. We first remember the theory of characteristics (Mathews and Walker, pages 217-226; Modern Computing Methods, pages 101-104.)

This classified equations into:

Elliptic: no real characteristics
Parabolic: no degenerate characteristics
Hyperbolic: two distinct

The first two are particularly similar to ODE. Elliptic equations typically give boundary value problems and one gets matrix inversion/eigenvalues formulation after making finite difference approximations to, say, \( \nabla^2 \). A parabolic equation e.g. \( \frac{\partial^2 f}{\partial x^2} = \frac{\partial f}{\partial t} \) is
typically a mixture of matrix \((m \times n)\) and recursion \((m \times t)\) techniques. For instance one can take a grid in \(x\), \((x_0, \ldots, x_{n+1})\) and reduce PDE to a coupled set of ODEs for \(\frac{dy(x_n,t)}{dt}\) where you make standard finite difference approximation for \(\frac{\partial^2 y}{\partial x^2}\).

If \(h\) is \(x\)-grid size, these read:

\[
\begin{align*}
  h^2 \frac{dy_1}{dt} + 2y_1 - y_2 &= y_0 \quad \text{for all } t \\
  h^2 \frac{dy_2}{dt} - y_1 + 2y_2 - y_3 &= 0 \\
  h^2 \frac{dy_n}{dt} - y_{n-1} + 2y_n - y_{n+1} &= y_{n+1}
\end{align*}
\]

where \(y_n(t) = y(x_n,t)\).

Of course, boundary conditions give \(y_n(0)\) and we can use \(h \in \{1/2, 1\}\) to solve above.

More destructive are hyperbolic equations.
E7/1 Hyperbolic Equations

(i) To recapitulate characteristics:

Equation is \[ a \frac{\partial^2 y}{\partial x^2} + b \frac{\partial x}{\partial x} + c \frac{\partial^2 y}{\partial y^2} = 0 \] (E7.1)

define \[ p = \frac{\partial u}{\partial x}, \quad q = \frac{\partial u}{\partial y} \]

\[ r = \frac{\partial^2 u}{\partial x^2}, \quad s = \frac{\partial^2 u}{\partial x \partial y}, \quad t = \frac{\partial^2 u}{\partial y^2} \]

differentials \[ du = p \, dx + q \, dy \] (E7.2)

\[ dp = r \, dx + s \, dy \] (E7.3)

\[ dq = s \, dx + t \, dy \] (E7.4)

Characteristics come from putting determinant of (E7.1, 2, 3, 4) = 0

\[ a (dy)^2 - b dy dx + c (dx)^2 = 0 \] (E7.5)

and are curves Cauchy conditions \( (\text{given } u, p, q) \) do not specify. Solution.

Suppose \( AB \) have given boundary conditions and red lines are characteristics.
The method to be used, is most easily illustrated when we choose characteristics as axes, i.e. equation is:
\[ \frac{\partial u}{\partial x} = \frac{\partial u}{\partial y} = e \quad (\text{E7.1}) \]
characteristics \(-e \, dy \, dx = 0\) i.e. \(x = \text{const.}\) or \(y = \text{const.}\).

Consider \((\text{E7.3})\)
\[
dp = r \, dx + s \, dy = r \, dx + e^{\frac{1}{s}} \, dy
\]
So on line \(x = \text{const.}\)
\[
dp = e^{\frac{1}{s}} \, dy \quad (\text{E7.6})
\]
Similarly on \(y = \text{const.}\)
\[
dq = e^{\frac{1}{r}} \, dx \quad (\text{E7.7})
\]

Thus, in previous diagram we know \((u, \rho, q)\) at \(X\) and \(Y\). Integrate \((\text{E7.6})\) along \(XZ\) (suppose this is \(x = \text{const.}\)) \(\Rightarrow p\) at \(Z\). Similarly integrate \((\text{E7.7})\) along \(YZ\).
to give \( q \) at \( z \). (E7.2) then gives \( u \) at \( z \). This process can be continued and clearly, given boundary conditions along \( AB \), we can find \( u \) at all of \( APBP' \).

(iii) Let's generalize method to case where characteristics are not orthogonal axes. Eliminate \( s \) and \( t \) between (E7.1) \((ar + bs + ct = 2)\) (E7.2) and (E7.3)

\[
E_{7.3} dy + C_{7.4} dx = 0
\]

\[
adpdy + c dq dx = edxdy
\]

\[+ S \left\{ a (dy)^2 + c (dx)^2 - b dx dy \right\}
\]

Now along a characteristic, coefficient of \( S \) is identically zero, and we get

\[adpdy + c dq dx = edxdy \quad (E7.3)
\]

(whence \( dx/dy \) satisfies (E7.5)).
(67.8) is an ordinary differential equation and implies a functional relation between \( p \) and \( q \) along characteristics. So, as before, integrate (67.8) along \( xz \) to give one relation between \( p \) and \( q \) at \( z \), integrate (67.8) along \( yz \) to find another relation. Combine them to find \( p \) and \( q \) at \( z \). The procedure is then as in simple example.
As the error is proportional to \( nh^5 \) and \( n \) doubles and \( h \) halves each go, the error should go down by \( \approx 1/16 \) each iteration and convergence is fast. So, in practice, one halves \( h \) until change in integral between steps is less than some assigned value.

There is a nice discussion of this in Noble § 9.3 p. 231 onwards (Vol. II).

**5/3 Gaussian Integration**

We are evaluating the same formula

\[
I = \int_a^b q(z) f(x) \, dx
\]

in terms of the same ansatz

\[
I = \sum_{r=0}^N w_r f(x_r)
\]

Now we choose both \( w_r \) and \( x_r \) (a grand total of \( 2n+2 \) parameters).
where \( f(x) \) is assumed known at \( x_0, \ldots, x_n \) and \( g(x) \) is sufficiently simple so that integrals (ES.3) involving it, can be calculated. Of course, we split \( g(x) \) #1 to make \( f(x) \) smoother and more reasonably approximated as a polynomial. We assume the form:

\[
I = \sum_{r=0}^{n} \omega_r f(x_r) \tag{ES.2}
\]

where the nos. \( \omega_r \)'s are to be found by requiring \( I \) to be exact for all polynomials of degree \( \leq n \). This is easy - for put

\[
\int_{a}^{b} x^m g(x) \, dx \tag{ES.3}
\]

and our condition becomes:

\[
\int_{a}^{b} x^m \, dx = 0 \quad \text{if} \quad 0 \leq m \leq n \tag{ES.4}
\]

(linear nature of integral ensures that (ES.4) guarantees exactness of (ES.2) for all polynomials of degree \( \leq n \)). We can write (ES.4) as:

\[
A \omega = b \tag{ES.5}
\]
E4/6 Statistical Warning.

The techniques we are discussing in the previous sections are only appropriate for functions $f(x)$ known with negligible error. Suppose $f(x)$ is measured experimentally to be $y_i = y_j$ at $x = x_i$ $(i = 0, \ldots, n)$. If you wish to find $f(x)$ in $[x_0, x_n]$, one should not use E4/1.4 for $y$ and $y_j$ separately.

\[ \sum \frac{1}{x_i} \left( \frac{1}{x_i} - \frac{1}{x_j} \right)^2 \]

\[ \text{least squares} \]

\[ \text{interpolation} \]

is consistent with constant $f(x)$ and the third degree (as 4 points) interpolating polynomial is ridiculous. For such problems, one should least squares techniques discussed in statistics.

1.2. minimize $\sum_{i=0}^{n} \left[ P_m(x_i) - y_i/\sigma_{y_i} \right]^2$
For $g(x) = 1$, we get normal Gaussian formulae with:

$$x_i = i-th \ \text{zero of } P_{n+1}(z)$$

$$w_i = \frac{2}{[1-x_i^2]^{\frac{1}{2}}} \left[ P'_{n+1}(z_i) \right]^{-2}$$

Clearly Gaussian integration is best when $f(x)$ is very hard to evaluate and one must minimize value of $n$. However it is very hard to estimate error in calculation as the simple iterative scheme described in ES/2 for Simpson's rule cannot work. Not only is it hard to evaluate weights and $x_i$'s as $n$ increases $\to \infty$, but you have to recalculate all $n$ functions each time.
There is no overlap in $x_i$ between different $n$ values. (Note Gauss only saves factor of $2$ in functor evaluations and Romberg is simpler to code and saves same factor of $2$ if you have to iterate).

So use Romberg if you have a few integrals which are needed to prescribed accuracy. Use Gaussian integration, when you have a lot of integrals with similar $f(x)$ and the right value of $n$ can be determined beforehand.
to make (E5.2) exact for polynomials of degree $\leq 2n+1$ (2n+2 equations analogous to (E5.3)).

Let $P_n(x)$ be the polynomials orthogonal with weight function $g(x)$

$$\int_a^b g(x) P_n(x) P_m(x) \, dx = h_n \delta_{nm}$$

Let $x^{(i)}_c$ be the $i$th zero of $P_n(x)$

Define $V_n(x,y)$ by the formula

$$V_n(x,y) = \sum_{k=0}^{n} P_k(x) P_k(y) / h_k$$

If $T_n(x)$ is any polynomial of degree $n$, then one can easily show:

$$\int_a^b g(x) V_n(x,y) T_n(x) \, dx = T_n(y)$$

It is not difficult to see that

(E5.2) is satisfied by

$$x_i = x_c^{(n+1)}$$

$$\omega_i = 1 / \sqrt{V_n(x_i^{(n+1)}, x_c^{(n+1)})}.$$
Take as a complete set \( \mathcal{G} \) of polynomials of degree \( 2n+1 \).

a) \( P_{n+1}(x) \): 1

\[
\begin{align*}
& x \quad x^2 \\
& \quad \vdots \\
& x^n
\end{align*}
\]

\[\prod_{i=1}^{n} \lambda(x)\]

I is zero and is ensured by \( x_i = x_i^{(n+1)} \).

b) \( \lambda_{n+1}(x) \): \( (x-x_0^{(n+1)})(x-x_1^{(n+1)}) \ldots (x-x_n^{(n+1)}) \)

omitted \( (x-x_i^{(n+1)}) \)

By construction, the only term that contributes to \( I \) for this polynomial is that with \( r = i \) and this gives formula for \( w_i \).

\[I = \prod_{n} (x_i^{(n+1)})\]

\[= w_i \lambda_{n+1}(x_i^{(n+1)} x_i^{(n+1)}) \vartheta_n(x_i^{(n+1)})\]
to arbitrary \( n \) are called Newton-Cotes formulas. (See Abramowitz and Stegun and find them explicitly written out.) We can relax restrictions:

(a) \( x_i \) equidistant
(b) \( x_0 = a \) and/or \( x_n = b \)
(c) \( g(x) = 1 \)

The reader need only solve (E5.5) to find relevant formulas for any \( n \).

(i) The errors in (E5.6,7) are written in the form:

\[
\text{trapezoidal - (E5.6)} - \quad -\frac{h^3}{12} f''(\xi) \quad (E5.8)
\]

\[
\text{Simpson - (E5.7)} - \quad -\frac{h^5}{90} f''''(\xi) \quad (E5.9)
\]

where, as usual, \( \xi \) is some value of \( x \) in \([a, b] \).

Note that, as expected, trapezoidal formula is exact for linear functions (error \( \propto f''(\xi) \) vanishes for \( f(x) = ax + b \)) but Simpson's rule - which as derived is exact for quadratic \( f(x) \) - is unexpectedly exact for cubics. In other words, its error is small and it is a widely used formula
w is column vector of w's, \( \mathbf{b} \) is column vector of \( x_i \)'s and \( A \) is \((n+1) \times (n+2)\) matrix of coefficients \( A_{mr} = (x_r)^m \). It is easy to show that \( A \) is nonsingular (if \( x_r \) distinct) and hence that (ES.5) can be solved and (ES.2) established for any \( g(x) \) in (ES.1).

Usually (ES.5) is solved once and for all for given \( g(x) \) and then (ES.2) can be used quickly for any \( f(x) \) that crops up. Examples:

(a) Trapezoidal Rule

\[ n=1, \ g(x) = 1 \quad x_0 = a, \ x_2 = b \]
\[ \int_a^b f(x) \, dx \approx \frac{1}{2} h \left[ f(a) + f(b) \right] \quad h = \frac{b-a}{2} \quad (ES.6) \]

(b) Simpson's Rule:

\[ n=2, \ g(x) = 1 \quad x_0 = a, \ x_1 = \frac{1}{2}(a+b), \ x_2 = b \]
\[ \int_a^b f(x) \, dx = \frac{1}{3} h \left[ f(x_0) + 4f(x_1) + f(x_2) \right] \quad (ES.7) \]

where \( h = \frac{1}{2}(b-a) \)

Both these are equi-spaced (in \( x_i \)) rules with \( g(x) = 1 \) and the generalization
\[ \int_{x_0}^{x_{2n}} f(x) \, dx = \frac{h}{3} \left[ f_0 + 4f_2 + 2f_4 + 4f_6 + \ldots + f_{2n} \right] - \frac{n^5}{90} f^{(iv)}(3) \]  

(E5.9)

where \( f_i = f(x_i) \) and \( x_i \) are equidistant.

Get this by applying Simpson to intervals \( x_0, x_2, x_4, \ldots, x_{2n-2} \) to \( x_{2n} \) in succession.

One could also generate a formula in terms of the Newton-Cotes values \( f(x_i) \) by using the Newton-Cotes formula appropriate to \( 2n+1 \) points. This gains an extra derivative in error just like Simpson, and error is \( c h^{2n+3} f^{(2n+2)}(3) \), where \( c \) is some known number. In practice, this is not often used because

(a) The implied interpolating of \( f(x) \) over a wide range by a high order polynomial can produce disastrous results when \( f(x) \) is not polynomial. (e.g.
has singularities) on the other hand, local interpolation by \( f(x) \) by a low order polynomial (used in (E5.9)) is much safer.

(b) It is much easier to estimate error in (E5.9) as we describe in Sect. E5/2.

(iii) Points of Technique

Note, when evaluating \( I = \int_a^b f(x)g(x)dx \), one can choose between

(a) Change of variable e.g. \( dy = g(x)dx \) converts integral to \( g(y)=1 \) form.

(b) Newton-Cotes (or Gauss - see E5/3) formula with weight function \( g(x) \).

This remark also hold for change of integration range (e.g. to take \( b \) from \( -\infty \) to finite value).

In each case, one must simply use criterion.

Is \( f(x) \) or \( f(y) \) better approximated by a polynomial? One can use same criterion to
both choose \( g(x) \) and/or subtract off some particularly nasty part (e.g. Singularity) of \( f(x) \).

\[
I = \int_a^b \left( f_1(x) + f_2(x) \right) \, dx = I_1 + I_2
\]

\[
I_1 = \int_a^b f_1(x) g(x) \, dx \quad \text{done numerically}
\]

\[
I_2 = \int_a^b f_2(x) g(x) \, dx \quad \text{done analytically}
\]

\[E5/2 \quad \text{Romberg Integration}\]

A very convenient way of evaluating integrals is to repeatedly apply Simpson's rule - halving interval each time.

\[a \quad x \quad x \quad x \quad b \quad \text{Step 1}\]

\[a \quad x \quad x \quad x \quad x \quad 2\]

\[a \quad x \quad x \quad x \quad x \quad x \quad x \quad 3\]

The point is that, as diagram indicates, one need only evaluate half as many (\( x \) in diagram) functionals as you expect because you use old values (\( 0 \) in figure).
Ghost Question: Problem Set 3 Question 4

The answer from grader assumes that probability of "no ghosts" is independent of probability of "no observer". This is false.

The problem does address probability of a T.O. (trained observer) entering house if ghost is there. We will assume that T.O.'s never enter the house if a ghost is there and indeed there is also only one ghost. You could assume T.O.'s will enter if ghost is present but this will make no difference to final answer on ghost probabilities. It will affect answer on T.O. probabilities but question didn't ask for these.
Problem Set 5

3

walls(2)  
\text{target (1)}

F = \text{Full} \quad \varepsilon = 2 = \text{Empty}
1 = \text{target} \quad F = 1 + 2
\text{t} = \text{time}
\text{c} = \text{cross section}
\sigma = \text{standard deviation (NOT cross section!)}
N = \text{Number of events}

N_F = (\theta_1 + \theta_2) \, t_F
N_E = c_2 \, t_E

G_1 = \text{desired} = \frac{N_F}{t_F} - \frac{N_E}{t_E}

N_E \text{ and } N_F \text{ are independent}

\sigma^2(c_1) = \frac{\sigma^2(N_F) + \sigma^2(N_E)}{t_F^2} = \frac{N_F}{t_F^2} + \frac{N_E}{t_E^2}

\text{as } \text{N}_E, \text{N}_F \text{ Poisson
}

\sigma^2(c_1) = \frac{(\theta_1 + \theta_2) + c_2}{t_F} + \frac{c_2}{t_E}

\text{to be minimized subject to } b_E + b_F = T

\text{which is (differentiate wrt } \delta t_F = -\delta t_E\text{)}

\frac{t_F^2}{t_E^2} = \frac{c_1 + c_2}{c_2}

\frac{\text{Time Full}}{\text{Time Empty}} = \sqrt{\frac{\text{Rate Full}}{\text{Rate Empty}}}

\text{Time Full}}
Under these assumptions, there are only three possibilities:

1. Nobody in house $X(t)$
2. T.O. in house $Y(t)$
3. Ghost in house $Z(t)$

The differential equations are:

$$X(t + dt) = X(t)\left[1 - 2dt - 3dt\right] + Y(t)2dt$$

or

$$\frac{dX}{dt} = 2Y - 5X$$

Similarly,

$$\frac{dY}{dt} = 2X - 2Y$$
$$\frac{dZ}{dt} = 3X$$

with $X(0) = 1, Y(0) = Z(0) = 0$.

One can solve by say Laplace Transform and find:

$$X(t) = \left[\frac{4e^{-6t} + e^{-t}}{5}\right]$$
$$Z(t) = \left[\frac{5 - 2e^{-6t} - 3e^{-t}}{5}\right]$$
$$Y(t) = \left[\frac{2e^{-t} - 2e^{-6t}}{5}\right]$$
Problem Set 5

4. Let us take some examples and use

\[
\frac{f_1}{f_2} = \frac{\Pr [\text{value} = \text{mean} + \frac{\text{std deviation}}{\sigma}]}{\Pr [\text{value} = \text{mean}]}
\]

Question: Suggests take ln of this but it doesn't matter.

For Gaussian \( f_1 = f_2 = \exp(-1/2) = 0.61 \)

For Poisson: \( \sigma = \sqrt{m} = 2 \)

\[
\begin{array}{ccc}
\sigma = 2 & 0.75 & 0.53 & m = 4 \\
3 & 0.69 & 0.55 & 9 \\
4 & 0.67 & 0.56 & 16 \\
5 & 0.65 & 0.57 & 25 \\
\end{array}
\]

Note that for a Gaussian, 68% of events are in \( \pm 1\sigma \) and 95% in \( \pm 2\sigma \).

We Poisson is broader or less sure than a Gaussian and that ratios \( f_1/f_2 \) are reasonably in accord with

Gaussian for \( \sigma \approx 4 \) but even \( \sigma = 3 \) is not terrible.

\( m \sim 10 \rightarrow 30 \) is a reasonable lower limit.
Problem Set 5

5. The number of events in region 
\[ t_1 > t > t_2, \]
is
\[ N_{\text{bin}} = C \int_{t_1}^{t_2} dt \, e^{10t} \]
\[ = \frac{C}{10} \left[ e^{10t_1} - e^{10t_2} \right] \]

Put \( N = 10^5 \), \( t_1 = 0 \), \( t_2 = -2 \)

Guess \( C = 10^6 \)

Now the smallest bin of interest
is such that \( (t_1 = t_0 + st, t_2 = t_0 - st, \) 
\( st \) positive) \( e^{10t_1} \) and \( e^{10t_2} \) differ by \( \leq 10\% \).

This is \( 2st \sim 0.01 \) GeV\(^{-2} \)
So we will choose bins such that \( 2st \geq 0.01 \)

We will determine if to make
then larger than. This by requiring that \( N_m > 20 \) evauls (as in question 4).

\[
\frac{C}{10} \left[ e^{10t_1} - e^{10t_2} \right] > 20
\]

or by expanding exponentials

\[
\frac{C e^{10t_0}}{10} \left[ e^{10s} - e^{-10s} \right] > 20
\]

\[
28t > 20 \times 10^6 \times e^{-10t_0}
\]

\[
t_0 = -0.5
\]

\[
= -0.5
\]

\[
= -1
\]

\[
= -1.5
\]

\[
\text{Smaller than } 0.01 \text{ - ignore}
\]

\[
\text{Large than } 0.01
\]

\[
66 \text{ - expansion works down}
\]

Change line to around \( -2 \) at \( t = -1 \)

and maybe use \( \ln \)

\[-1 > t > -2 \text{ GeV}^2\]
E1: Introduction (Things to keep you awake at night).

Suppose we have to solve numerically the recurrence relation:

\[ f_{r+1} = 8f_r - 12f_{r-1} \quad (E1.1) \]

plus the boundary conditions

\[ f_1 = 1 \quad f_2 = 2 \quad (E1.2) \]

Then this is a splendid well posed mathematical problem and the general solution is

\[ f_r = 2^r \quad (E1.3) \]

and it was not necessary to come to Caltech to find this out.

However, suppose we had taken

\[ f_1 = 1 \quad f_2 = 2.01 \quad (E1.4) \]

Then solving \((E1.1)\) gives

\[
\begin{array}{cccccccc}
  r & 1 & 2 & 3 & 4 & 5 & 6 \\
  (E1.4) & 1 & 2.01 & 4.08 & 8.52 & 19.2 & 51.36 \\
  (E1.1) & 1 & 2 & 4 & 8 & 16 & 32 \\
\end{array}
\]

and using the appropriate solution bears no resemblance to the real
one. The reason for this is that until you impose boundary conditions, the general solution of (E1.1) is:

\[ f_1 = A \cdot 2^t + B \cdot 6^t \]

the divergence between the exact and approximate solutions is just a reflection of the fact that any deviation from (E1.1) (e.g. in exact boundary conditions (E1.4) or from rounding error when splitting (E1.1)) will immediately diverge through (E1.1) and cause divergence we saw in example. This illustrates two points:

(a) If (E1.1) was an equation of physics and (E1.2) was measured experimental data for which we get

\[ f_1 = 1 \pm 0.01 \quad f_2 = 2 \pm 0.01 \]

Then (E1.4) shows that the problem of finding f20 is insoluble and the numerical calculation impossible.
(1) If \((E1.1, 2)\) are both laws of physics, then problem is now soluble mathematically but just poorly set up numerically and impossible to solve on any computer which has finite number of digits and will give rounding errors to propagate as \(6^5\). In this case one must use mathematical ingenuity to recast problem in a form which is numerically stable and so rounding errors will not cause divergence. For instance \(6f_{r+1} = 13f_r - 2f_{r-1}\) (\(E1.5\)) with same boundary conditions \((E1.2)\) has general solution \(A2^r + B3^r\) and any \(B \neq 0\) will disappear and not increase at nauseam.

Alternately \(f_{r+1} = 2f_r, \quad f_0 = 1\) \((E1.6)\) is an even simpler reformulation which is mathematically identical but numerically distinct from \((E1.4)\).
(iii) (a) Similar difficulties occur with matrices. For instance, suppose we wish to solve:

\[ x + 2y = 4 + R_1 \]
\[ x + 2.001y = 4.003 + R_2 \]  \hspace{1cm}(61.6)

Then for \( R_1 = R_2 = 0 \), we have exact solution \( y = 3 \), \( x = 2 \). However, the small change in rhs (which might be data again) to \( R_1 = 0 \), \( R_2 = -0.006 \) changes solution drastically to \( x = 10 \), \( y = -3 \). Thus (61.6) is very unstable to rounding and measurement errors.

For (61.6) the source of the difficulty is quite obvious - the rhs's are almost identical - alternately we can phrase it as the determinants of matrix \( \begin{bmatrix} 1 & 2 \\ 1 & 2.001 \end{bmatrix} \) being 0.001 and much smaller than values of elements (cofactors).
(b) A less blatant example is the notorious Hilbert matrix which is so pathological it is standard test case to prove out matrix inversion failure.

In 4 dimensions:

\[
A = \begin{bmatrix}
\frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\
\frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\
\frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \\
\frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8}
\end{bmatrix}
\]

\[
A^{-1} = \begin{bmatrix}
200 & -1200 & 2100 & -1120 \\
-1200 & 8400 & -15120 & 8400 \\
2100 & -15120 & 29400 & -16800 \\
-1120 & 8400 & -16800 & 9800
\end{bmatrix}
\]

Solve \(Ax = b\) with \(b_1 = b_2 = b_3 = b_4 = 1\).

Then \(x_1 = -20\)
\(x_2 = 180\)
\(x_3 = -420\)
\(x_4 = 280\)

However elements of \(A^{-1}\) are very large.

maximum is 30,000. So error in \(x_1\):

\(8x \approx 30,000 \|b\r\) where \(\|b\r\) error in \(b\). This can clearly be very large compared to value of \(x_1\) even for modest \((10^{-3})\) errors in \(b\).
(c) One can formulate these difficulties, rather neatly with eigenvalues. Let $\lambda \sigma$.

Set $\quad c = |\lambda_{\text{max}}| / |\lambda_{\text{min}}|$

If $\quad c \approx 1$ problem well behaved.

$\quad c >> 1$ problem (matrix) ill-conditioned.

Thus $\quad Ax = b$

$\quad A = (A^{-1} D q P) \quad P^{-1} = P^T$

($A$ symmetric, $Dq$ diagonal) with elements $\quad b^T = P x \quad x = P^T y$

$\quad Dq y = b^T$

and $\quad s y_i = s b_i / \lambda$

If $\quad \lambda$ very small, (compared to $\lambda_{\text{max}}$)

then a small error in $b^T$ ($s b_i$) will give a large error in $y$ and

problem is poorly posed.

However, this illustrates another

point. Suppose we do have a poorly

posed problem $c >> 1$, then we needn'
throw up our hands. Although all the $x$'s are undetermined (as $x = P^T y$ is l.c. of all $y$'s and one $y$ at least badly determined) we can state the numerical solution nearly in terms of $y$'s, i.e. not all of $n$ physical parameters $x_i$ are determined by our measurements of $b_i$, rather $m$ ($< n$) l.c.'s of $x$ $y_i$ (with large eigenvalues) are well determined and $n - m$ l.c.'s are badly determined. The latter require a new experiment.

(d) In any numerical calculation one has to be very careful about undetermined parameters because computer will give answers for all one has to either check theoretically problem well-posed or check numerically that $|\text{final answer}|$ not $< |\text{individual terms of sum}|$. 
E4 Calculation and Representation of Functions

E4/1 Difference Operators

(i) Difference operators are not as important now as in good old days because on digital computers, functions are evaluated by summing series and not by table look-up (the latter takes too much space core). However, they are still important for linear and partial differential equations.

Mathews and Walker introduce them on page 345. Construct a table of $\sin x$ as on following page. $\sin x$ is a typical (analytic) function whose differences get smaller until they become random things of same size as rounding error - the latter then diverge (we give an example of this latter under propagation of errors - it is not shown in $\sin x$ table).
<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>y = \sin x</th>
</tr>
</thead>
<tbody>
<tr>
<td>0°</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10°</td>
<td>0.1736</td>
<td>-0.0052</td>
</tr>
<tr>
<td>20°</td>
<td>0.342</td>
<td>-0.0104</td>
</tr>
<tr>
<td>30°</td>
<td>0.5</td>
<td>-0.0152</td>
</tr>
<tr>
<td>40°</td>
<td>0.6428</td>
<td>-0.0196</td>
</tr>
<tr>
<td>50°</td>
<td>0.766</td>
<td>-0.0232</td>
</tr>
<tr>
<td>60°</td>
<td>0.866</td>
<td>-0.0263</td>
</tr>
<tr>
<td>70°</td>
<td>0.9397</td>
<td>-0.0286</td>
</tr>
<tr>
<td>80°</td>
<td>0.9848</td>
<td>-0.0299</td>
</tr>
<tr>
<td>90°</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

Rounding errors.

1st & 2nd differences of values in previous column.

(ii) The numbers in this table can be labelled according to 3 different schemes.

Let \( y_n = (x_0 + nh) \) where \( h \) is some fixed interval and \( n \) an integer.

(a) Forward Differences

\[ \Delta y_n = y_{n+1} - y_n \quad \text{def.} \]

e.g.

\[ \Delta y_0 = y_1 - y_0 = 0.1232 \quad \text{above} \]

\[ \Delta^2 y_n = \Delta y_{n+1} - \Delta y_n \quad \text{def.} \]
(a) e.g. \( \Delta^2 y_0 = \Delta y_1 - \Delta y_0 = -0.0232 \) alive

The pattern is

\[ y_{-2} \quad \Delta y_{-2} \quad \Delta^2 y_{-2} \quad \Delta^3 y_{-2} \quad \Delta^4 y_{-2} \]

\[ y_{-1} \quad \Delta y_{-1} \quad \Delta^2 y_{-1} \quad \Delta^3 y_{-1} \quad \Delta^4 y_{-1} \quad \text{same subscript on diagonals} \]

\[ y_0 \quad \Delta y_0 \quad \Delta^2 y_0 \quad \Delta^3 y_0 \quad \Delta^4 y_0 \]

\[ y_1 \quad \Delta y_1 \quad \Delta^2 y_1 \quad \Delta^3 y_1 \]

\[ y_2 \quad \Delta y_2 \quad \Delta^2 y_2 \quad \Delta^3 y_2 \]

(b) Backward Differences

\( \Delta y_{n+1} = y_{n+1} - y_n \) def.

\( \Delta y_{n+1} = \Delta y_n \) of course, and pattern is

\[ y_{-2} \quad \Delta y_{-1} \quad \Delta^2 y_{-1} \quad \Delta^3 y_{-1} \quad \Delta^4 y_{-1} \quad \text{same subscript on diagonal} \]

\[ y_{-1} \quad \Delta y_0 \quad \Delta^2 y_0 \quad \Delta^3 y_0 \quad \Delta^4 y_0 \]

\[ y_0 \quad \Delta y_1 \quad \Delta^2 y_1 \quad \Delta^3 y_1 \]

\[ y_1 \quad \Delta y_2 \quad \Delta^2 y_2 \quad \Delta^3 y_2 \]

\[ y_2 \quad \Delta y_3 \quad \Delta^2 y_3 \quad \Delta^3 y_3 \]
(c) Central Differences

\[ \delta^2 y_{n+1/2} = y_{n+1} - y_n \]

\[
\begin{array}{cccccc}
  y_{-2} & \delta y_{-3/2} & \delta^2 y_{-1} & \delta^3 y_0 & \delta^4 y_{1/2} & \delta^5 y_1 \\
  y_{-1} & \delta y_{-1/2} & \delta^2 y_0 & \delta^3 y_{1/2} & \delta^4 y_1 \\
  y_0 & \delta y_{1/2} & \delta^2 y_0 & \delta^3 y_{1/2} & \delta^4 y_1 \\
  y_1 & \delta y_{3/2} & \delta^2 y_1 & \delta^3 y_{1/2} & \delta^4 y_1 \\
  y_2 & \delta^2 y_2 & \delta^3 y_1 & \delta^4 y_{1/2} & \delta^5 y_1 \\
\end{array}
\]

Some numbers on top lines but only appear every other order.

So \( \delta y_0 \) is not on table but only \( \delta^{2n} y_0 \) \((n \text{ integer})\).

(iii) Two Small points

(a) Polynomials

Note that like the \((m+1)\)'th derivative, the \((m+1)\)'th differences of a polynomial of degree \(n\) are zero. One can prove this by expressing \( y_n = f(x_0 + nh) \) as power series in \( h \). Then it is easy to see that one goes down one in highest power of \( x \) each difference.

Alternatively,
the formal methods we will soon discuss, prove this immediately.

(a) Error Propagation

Suppose we add to \( f(x) \) an error in one entry. Then superimposed on differences of \( f(x) \) we will have pattern:

\[
\begin{array}{cccc}
0 & 1 & 1 & 1 \\
0 & 1 & 1 & 4 \\
0 & -1 & 3 & 6 \\
0 & 1 & 1 & 1 \\
\end{array}
\]

The \( \binom{n}{k} \) a fan-out: this can be used to detect errors. Note rounding errors will give thin as they are a small sort of error!
(iv) **Symbolic Methods**

(a) Define \( E y_n = y_{n+1} \)

\[ p y_n = \frac{1}{2} \left( y_{n+2} + y_{n-2} \right) \]

where \( E \) = displacement, \( y(x_0 + h(x+h)) \) see below.

\( p \) = averaging operator

put \( E^s y(x) = y(x+sh) \) any \( s \), not necessarily integer.

Note this is consistent as \( E^{s_1} E^{s_2} = E^{s_1+s_2} \)

Then \( p = \frac{1}{2} \left( E^{1/2} + E^{-1/2} \right) \).

We can now express difference operators

\[ \Delta = E - 1 \]
\[ \nabla = 1 - E^{-1} \]
\[ \delta = E^{1/2} - E^{-1/2} \]

(b) More interestingly, we can relate the differential operator \( D \) to \( E \) and then to \( \Delta, \nabla \) and \( \delta \). This will be important for differential equations.

\[ Dy = \frac{dy}{dx} \quad \text{def.} \]

\[ E y(x) = y(x+h) \quad \text{use Taylor} \]

\[ = y(x) + h y' + \frac{h^2}{2!} y'' + \ldots \]

\[ = (1 + h D + \frac{h^2}{2} D^2 + \ldots) y(x) \]
or \( E \ y(x) = \exp(hD) \ y(x) \)

So \( E = \exp(hD) \)

Taking \( \delta = E^{1/2} - E^{-1/2} \)

we get \( \delta = \exp(1/2 hD) - \exp(-1/2 hD) \)

or \( \delta = 2 \sinh 1/2 hD \)

or \( hD = 2 \sinh^{-1} \delta \)

Taylor expanding \( hD = \delta - \delta^3/24 \) (using \( \sinh x = x - x^3/6 \))

Similarly \( hD = \log(1 + \delta) \)

\[ = \delta + \delta^2/2 \] and this converges more slowly than central difference formula.

Note that \( \delta^{(n)} \propto h^n D^n \)

i.e. \( \delta^{(n)} f(x) \propto h^n f^n(x) \)

and as \( \delta \) is higher orders in \( h \), this proves that \( (n+1)^{th} \) differences of an \( n^{th} \) degree polynomial vanish.

(c) One can use these formal method to derive some useful summation formulas.
For instance:

\[ S' = \sum_{j=0}^{p-1} f(x+jh) = (1 + E + \ldots + E^{p-1}) f(x) = \frac{E^p - 1}{E - 1} f(x) \]

now \[ \frac{x}{e^x - 1} = \sum_{n=0}^{\infty} B_n \frac{x^n}{n!} \]

defines Bernoulli numbers \( B_n \) (see page 48 of Mathews and Walker). Note \( B_3 = -\frac{1}{2}, B_2 = \frac{1}{6} \) etc.

So \[ S' = \frac{e^{ph} - 1}{hD} \sum_{n=0}^{\infty} B_n \left(\frac{hD}{n!}\right) f(x) \]

\[ = \frac{e^{ph} - 1}{hD} \sum_{n=0}^{\infty} B_n \frac{(hD)^n}{n!} f(x) \]

\[ = \frac{1}{h} \int_{x}^{x+ph} f(y) dy + \sum_{n=0}^{\infty} B_n \frac{h^{n-1}}{n!} \left\{ f^{(n-1)}(x+ph) - f^{(n-1)}(x) \right\} \]

which is the Euler-Maclaurin summation formula (Mathews and Walker—page 365). It relates \( S' \) and integral \( \frac{1}{h} \int_{x}^{x+ph} f(y) dy \) with error in form of derivatives. One can get error in form of differences by using
\[ \frac{\hbar D}{e^{\hbar D} - 1} = \frac{2 \sinh^{-1} \frac{\sqrt{2}}{2} \delta}{\frac{1}{2} \delta^2 + 8 \sqrt{1 + \delta^2 / 4}} \]

getting denominator from \( e^{\hbar D} - 1 = e - 1 \)

and \( \delta^2 = (E^2 - e^{-2})^2 = E - 2 + \frac{1}{e} \)

or \( E^2 - (2 + \delta^2) E + 1 = 0 \)

\[ E = 1 + \delta^{1/2} \pm \sqrt{(1 + \delta^{1/2})^2 - 1} \]

use + sign to get \( E \geq 2 \delta \)

or \( E = 1 + \frac{1}{2} \delta^2 + 8 \sqrt{1 + \delta^2 / 4} \)

In the problem set, we show how Euler's transformation (Mathews and Walker - page 54) can be derived by using similar formal manipulation.
In the previous pages, we interpreted
\[
\frac{1}{D} f(x) \quad \text{as} \quad \int_x^1 f(y) \, dy
\]
This is correct as
\[
D \left[ g(x) = \left( \frac{1}{D} f(x) \right) \right] = \left( D \cdot \frac{1}{D} \right) f(x) = f(x)
\]
and \[ \int_x^1 f(y) \, dy \] is the correct solution of this differential equation.

This solution is undefined up to a constant, but this is irrelevant as operator $\mathcal{L}^0 - 1$ annihilates a constant.
even powers) we assume an even number of grid points which we label

\[ x_{-n}, x_{-n+1}, \ldots, x_0, x_1, \ldots, x_n, x_{n+1} \]

and after some algebra (use \( \varepsilon = e^A \),
\( A = \frac{1}{2} \sinh^{-1} \theta \) and express \( y_0 \),

\[ y_p = \frac{\sinh pA}{\sinh A} y_1 + \frac{\sinh qA}{\sinh A} y_0, \quad q = 1 - p. \]

Then write \( \sinh [z p, A/2] \) in terms of power series in \( \sinh A/2 \), we find a formula of

form

\[ y_p = (1-p) y_0 + p y_1 + \text{usual linear interpolation} + E_2 \delta^2 y_0 + E_4 \delta^4 y_0 + \ldots \]

\[ + F_2 \delta^2 y_1 + F_4 \delta^4 y_2 + \ldots \]

where \( E \) and \( F \) are called Everett coefficients. Explicitly

\[ E_{2n} = \left(1 - \frac{p+n}{2n+2}\right) \quad F_{2n} = \left(\frac{p+n}{2n+4}\right) \]

just like Newton's formula, \( (E4.2) \)

truncated after \( 8^n \) terms is equivalent to fitting a \((2n+4)^{th}\) degree polynomial
through \( x_n \ldots x_{n+1} \).
with polynomial interpolation

(i) we now generalize the above formulae to non-equidistant points \( x_0 \ldots x_n \). So we have \( n+1 \) arbitrary \( x \) values \( x_0 \ldots x_n \) and \( n+1 \) values \( y_m = f(x_m) \) and wish to estimate \( f(x) \) for some \( x \) value. We construct the unique \( n \)-th order polynomial \( P_n(x) \) which satisfies

\[
P_n(x_m) = y_m \quad m = 0 \ldots n \tag{E4.3}
\]

The polynomial \( P_n \) has \( n+1 \) coefficients \( a_i \), which are to be determined by the \( n+1 \) conditions \( (E4.3) \). So the problem looks solvable in principle. In fact we can write down the answer explicitly as:

\[
P_n(x) = \sum_{k=0}^{A} L_k(x) \cdot y_k
\]

\[
L_k(x) = q_k(x) / q_k(x_k)
\]

\[
q_k(x) = (x-x_0) \ldots (x-x_{k-1})(x-x_{k+1}) \ldots (x-x_n).
\]

\((E4.4)\) clearly satisfies \((E4.3)\) and is known as Lagrange's interpolation
$x_0 \ldots x_n$ and $x$. (we cover extrapolation and so $R$ may be bigger than the range $x_0 \ldots x_n$). Fix $x$, and let $X$ vary over $R$ and put

$$F(X) = f(X) - P_n(X) - P_{n+2}(X) \frac{S(x)}{(n+1)!}$$

Then $F(X) = 0$ at $n+2$ values of $X$ i.e. $X = x$ and $x_0, x_1 \ldots x_n$

Apply Rolle's theorem between each pair of zeros of $F(X)$; we deduce that $F'(X) = 0$ at $n+1$ values of $X$ in $R$. We can continue process, showing recursively that $F^{(m+2)}(X)$ vanishes at least once in $R$. Let this be at $X = \xi$.

**Polynomial of degree** $n$,

$$0 = F^{(m+3)}(\xi) \equiv f^{(m+1)}(\xi) - 0 - (n+1)! \frac{S(x)}{(n+1)!}$$

or

$$S(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

and total error is $P_{n+1}(x) f^{(n+1)}(\xi) / (n+1)!$

$\xi$ is implicitly a function of $x$ but this formula indicates that error is bounded.
by $C P_{n+1}(x)$ and so controlled by $P_{n+1}(x)$.

For equidistant $x_m = m$ and $n = 5$

This $P_{n+1}(x)$ (like all) rapidly grows as $x$ ventures outside $0 \to 5$; thus extrapolation is difficult. However, the error is clearly smallest in $[2, 3]$ of intervals in $0 \to 5$. This a manifestation of our comment that Newton's formula was best for $p$ in middle of range. However we will now discuss a choice of $x_i$ such that the $l_i$ are all of equal modulus and interpolation is uniformly good (bad) throughout range. *Note: extrapolation - see section 2 of Keller p. 270.
4.14 Chebyshev Series

(i). Interpolation assumes that function given at ordained points outside our control. We now consider what happens if we are also allowed to choose the positions of the points.

Definition: Let \( f(x) \) be a function to be approximated in \( x \in [a, b] \) by a polynomial \( p_n(x) \) of degree \( n \). Then the best polynomial approximation to \( f(x) \) is such that \( \max_{x \in [a, b]} |f(x) - p_n(x)| \) is as small as possible. This called the minimax principle.

One can prove that if \( f(x) \) is continuous, then \( p_n(x) \) exists satisfying the following property:

**Theorem:** The error \( E_n(x) = f(x) - p_n(x) \) has property that it attains its maximum modulus \( |E_n(x)| \) value = \( L \) at, at least, \( n+2 \) distinct points in \( [a, b] \) and is alternately positive and negative there.

**Note:** 1 or 2 of these points could be end points \( x=a \) or \( x=b \).
Proof: we cannot do this rigorously. We must assume that there does exist a \( p_n(x) \) satisfying conditions of theorem. This is not unreasonable as we have a grand total of \( 2(n+2) \) conditions (derivative = 0, value = \( \pm L \) at \( n+2 \) points) on \( 2(n+2) \) parameters (\( n+2 \) \( x \)-values, one \( L \) value, \( n+1 \) coefficients of polynomial).

Then suppose theorem is not true and the best polynomial \( q_n(x) \neq p_n(x) \). Then let \( \varepsilon_n(x) = f(x) - q_n(x) \) be error and trivially \( |\varepsilon_n(x)| < L \) for all \( L \).

Consider \( r(x) = \varepsilon_n(x) - \varepsilon'_n(x) = q_n(x) - p_n(x) \) at \( x \) values where \( \varepsilon_n(x) = \pm L \), \( r(x) \) has same sign as \( \varepsilon_n(x) \). i.e. \( r(x) \) has \( n+2 \) sign changes or zeroes. But \( r(x) \) is a polynomial of degree \( n \) and this is impossible.

\( \text{Q.E.D.} \)

Isaacson and Keller cover the theory of best polynomials rigorously; here we
We expand functions

\[ f(x) = \sum_{r=0}^{\infty} a_r \text{Tr}(x) \]

where \( a_r = \frac{2}{\pi} \int_{-1}^{1} f(x) \text{Tr}(x) \frac{1}{\sqrt{1-x^2}} \)

and, on \( \Sigma' \) indicates we should halve first term to account for anomaly in orthogonality condition.

Note \( \frac{\partial \text{Tr}}{\partial x} = -\sin (r \cos^{-1} x) \frac{1}{\sqrt{1-x^2}} \)

\[ = \frac{1}{\sqrt{1-x^2}} \sin [r \cos^{-1} x] \]

\[ = 0 \text{ if } r \cos^{-1} x = \pi \]

(\( s=1 \ldots r-1 \) or \( s=0, r \) illegal due to \( \sqrt{1-x^2} \)).

(b) So maxima and minima of \( \text{Tr}(x) \) occur for \( x = \cos \frac{s \pi}{r} \) and value

at these points is \( \cos \left( \frac{s \pi}{r} \right) \) i.e. \( (-1)^s \)

meanwhile, derivative does vanish there, but \( \text{Tr}(x) = 1 \) at \( x = 1 \) and \( (-1)^r \)

at \( x = -1 \). Thus \( \text{Tr}(x) \) has \( r+1 \) maxima and minima in \([-1,1]\) and is alternately \( \pm 1 \) at these points. Now we can easily prove:
leave the general theory and just show why Chelyshev polynomials are sometimes good approximations to the best polynomial.

2) Chelyshev Polynomials

(a) Mathematically, the Chelyshev polynomial is defined by

$$T_r(x) = \cos(r \cos^{-1} x)$$

as a polynomial of degree $r$ in $x$.

The first few polynomials are

$$T_0(x) = 1$$
$$T_1(x) = x$$
$$T_2(x) = 2x^2 - 1$$
$$T_3(x) = 4x^3 - 3x$$

They are examples of orthogonal polynomials we studied earlier this year.

The orthogonality relation is:

$$\int_{-1}^{1} T_r(x) T_s(x) \frac{dx}{\sqrt{1-x^2}} = \begin{cases} \pi & r = s > 0 \\ \frac{1}{2} \pi & r = s < 0 \\ 0 & r \neq s \end{cases}$$

as is obvious transforming to $x = \cos \theta$. 
(ii) Newton’s Formula

Suppose we are given \( n+1 \) equidistant points \( x_m \) (\( x_m = x_0 + mh, m = 0, 1, \ldots, n \)) and we know \( y_m = f(x_m) \), what is our best estimate of \( y_p = y(x_0 + ph) \)?

\[
y_p = \sum_{k=0}^{n} E^k y_0 = (1 + \Delta)^n y_0 \\
= y_0 + n \Delta y_0 + \frac{n(n-1)}{2!} \Delta^2 y_0 + \ldots \tag{E4.1}
\]

which is still exact. However, the values \( y_m \) (\( 0 \leq m \leq n \)) are sufficient to calculate the first \( n \) differences \( y_0, \Delta y_0, \ldots, \Delta^n y_0 \). If we truncate (E4.1) after \( \Delta^n y_0 \) term, then we get an approximate formula for \( y_p \) known as Newton’s interpolating formula.

As the truncation \( \Delta^m y = 0 \), \( m > n + 1 \) would
be exact for a polynomial of degree $n$, it is clear that Newton's formula is entirely equivalent to taking for $f(x_0)$ the value of the polynomial approximation to $f(x)$ that coincides with the function at the $n+1$ points $x_0, \ldots, x_n$.

(ii) Everett's Formula

Now it is clear intuitively that such interpolation is most appropriate for values of $p \approx \frac{1}{2}n$ near the midpoint of the range of the $x$-values that $f(x)$ is known. (For $p=0$, it the tail is wagging the dog.) However the formula (64.1) is not very convenient for such large $p \approx \frac{1}{2}n$ near the optimal value. However we can rearrange (64.1) by using central $\delta$ and not forward $\Delta$ differences.

i.e. use $\delta = E^{\frac{1}{2}A} - E^{-\frac{1}{2}A} = 2\sinh A/2$

where $E = e^A, A = hD$

To get calculable powers of $\delta$ (i.e.
formula. If $x_m$ were equidistant, $x_m = x_0 + mh$, then Lagrange is just a rearrangement of Newton's or Hermite's formula in terms of function values - not differences.

(ii) Error Analysis

The analysis of the $x$ dependence of the truncation error in (64.4) is rather neat. First we remind you of:

**Rolle's Theorem**: If $g(x)$ is a suitably nice function of $x$ in $[a, b]$ which satisfies $g(a) = g(b)$, then there is a $\xi$ in $[a, b]$ such that $g'(\xi) = 0$.

\[ \begin{array}{c}
\xi \\
\hline
\xi
\end{array} \]

Define $P_{n+1}(x) = (x-x_0)(x-x_1) \ldots (x-x_n)$ and as the error in (64.4) is clearly zero at $x_0, x_1 \ldots x_n$, we write

\[ f(x) = P_n(x) + P_{n+1}(x) S(x) \]

and try to find $S(x)$.

Define $R$ as the range containing
Theorem: If $f(x)$ is a polynomial of degree $n+1$, then $\sum_{r=0}^{n} a_r T_r(x)$ is the best polynomial approximation of degree $n$ in $[-1,1]$.

Proof: $f(x) = \sum_{r=0}^{n} a_r T_r(x)$ exactly

$= p_n(x) + a_{n+1} T_{n+1}(x)$

where $p_n(x) = \sum_{r=0}^{n} a_r T_r(x)$.

So $e_n(x) = f(x) - p_n(x)$ satisfies condition of our earlier theorem on p.60 and so $p_n(x)$ is best polynomial q.e.d.

The above shows that, for arbitrary $f(x)$, if error dominated by next term in series then Chebyshev expansion will be near best polynomial approximation. Of course, Chebyshev expansion is much easier to find than the best expansion which is why its approximate usefulness is of interest. We can see the same effect in
the error analysis of sect. 4/3. We found error \( s(x) = E_n(x) = \hat{p}_n(x) f^{n+1}(x) \) where \( \hat{p}_n(x) \) is polynomial vanishing at \( n+1 \) points where \( f(x) = p_n(x) \), the interpolating polynomial. Choosing \( \hat{p}_n(x) \) as the zeroes of \( T_{n+1}(x) \) gives \( p_{n+1}(x) = T_{n+1}(x) \) and the error is not of equidistant formula.

Again this is only precise if \( f^{n+1}(x) \) is constant i.e. \( f(x) \) is a polynomial and we're back with the theorem on page 64.

(c) We quote a rather pretty theorem to be found on page 32 of Snyder. This Chebyshev polynomials give best peak convergence of all ultraspherical polynomials. Put expansion polynomial

\[
e_n(x) = c_n (1-x^2)^{-\alpha} \frac{d^n}{dx^n} [1-x]^{n+\alpha}
\]
where $\alpha = \infty$ is Taylor series $e_n(x) \sim x^n$.
$\alpha = -\frac{1}{2}$ is Chebyshev.
$\alpha = 0$ is Legendre.

Then expand arbitrary $f(x)$
\[
f(x) = \sum_{r=0}^{\infty} a_r e_r(x) + e_n
\]

Synder shows that
\[
|e_n(x)| \leq \frac{f^{(n+1)}(\xi)}{\beta_n (n+1)!}
\]

$\beta_n = 1$ : Taylor \hspace{1cm} \text{worsT}

$= 2^n \sqrt{\frac{\pi}{\Gamma(n+1)}}$ : Legendre
$= 2^n$ : Chebyshev \hspace{1cm} \text{best}
Previously we indicated that choice of interpolation points and expansion polynomials was not an art but had a respectable theory behind it. We now overthrow the last of "artistic" assumptions i.e. why use a polynomial?

The answer is that, in general, there is no reason and there are better expansion functions. However, unlike Chebyshev series which are famous in numerical analysis, this point seems to not to be covered in the standard books on numerical analysis. In fact, it is a classical mathematical theory with rigorous theorems and probably should be used more. It is just being discovered in high-energy physics but hasn't had fantastic success yet.

I will indicate method by a trivial example.
Suppose we wish to approximate \( f(x) = 1/(1+x) \) over the range \( x = -0.9 \) to 
-0.5. Polynomials are possible but quite a few will be needed as, for instance, 
Taylor series about \( x = -0.7 \) has a radius of convergence 0.3 i.e. in 
\[
  f(x) = \sum_{n} a_n (x+0.7)^n, \quad a_n \sim (1/3)^n.
\]

On the other hand, if we tried to expand \( 1/(2+x) \) in some range we would find faster convergence \( a_n \sim (1/3)^n \).

These examples indicate that for analytic functions, the position of 
Singularity in complex plane control convergence of polynomial expansion and 
hence accuracy of series truncated after a given term. Now conformal maps can move 
Singularity and hence improve convergence.
For instance to approximate $f(x) = \frac{1}{1+2x}$ we move singularly at $x = -1$ to $\infty$ by transforming $u = \frac{1}{1+2x}$ so $f(x)$ becomes $f(u) = u$ and polynomials in $u$ certainly converge fast (note these are not polynomials in $\overline{z}$).

So theory is: given domain of analyticity i.e. singularity, structure and domain $D$ of approximation: map former to achieve fastest possible rate of convergence of polynomial approximations in $D$. There is a unique answer for given $D$ and $D'$.

Warning: in practice, the behaviour at $\infty$ of the original function is of importance. This point gets mapped to finite $u$ and can produce a nasty (essential) singularity.