CHAPTER 1

Mathematical Preliminaries and Error Analysis

2. Review of Calculus

NOTATION.

$C(X)$ — all functions continuous on the set $X$.

$C[a, b]$ — all functions continuous on the interval $[a, b]$.

$C^n(X)$ — all functions that have $n$ derivatives continuous on $X$.

$C^\infty(X)$ — all functions that have derivatives of all orders on $X$.

THEOREM (Mean Value Theorem). If $f \in C[a, b]$ and $f$ is differentiable on $(a, b)$, then a number $c$ in $(a, b)$ exists with

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

or

$$f(b) = f(a) + f'(c)(b - a).$$
**Theorem** (Extreme Value Theorem). If \( f \in C[a, b] \), then \( c_1, c_2 \in [a, b] \) exist with \( f(c_1) \leq f(x) \leq f(c_2) \) for all \( x \in [a, b] \). In addition, if \( f \) is differentiable on \((a, b)\), then \( c_1 \) and \( c_2 \) occur either at the endpoints of \([a, b]\) or where \( f' \) is zero.

![Graph of the Extreme Value Theorem](image)

**Theorem** (Intermediate Value Theorem). If \( f \in C[a, b] \) and \( K \) is any number between \( f(a) \) and \( f(b) \), then there exists \( c \in (a, b) \) such that \( f(c) = K \).

![Graph of the Intermediate Value Theorem](image)

**Corollary** (Intermediate Value Theorem). If \( f \in C[a, b] \) and \( f(a)f(b) < 0 \), then there exists \( c \in (a, b) \) such that \( f(c) = 0 \).

**Note.** This corollary is often used in approximating roots of equations.
**Theorem (Taylor’s Theorem).** Let \( f \in C^n[a, b] \) with \( f^{(n+1)} \) existing on \([a, b]\) and \( x_0 \in [a, b] \). For every \( x \in [a, b] \), there exists a number \( \xi(x) \) between \( x_0 \) and \( x \) with

\[
f(x) = P_n(x) + R_n(x)
\]

where

\[
P_n(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n
\]

\[
= \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k
\]

and

\[
R_n(x) = \frac{f^{(n+1)}(\xi(x))}{(n + 1)!}(x - x_0)^{n+1}.
\]

**Note.**

1. \( f \) is the sum of a Taylor polynomial of order \( n \) (also called a Maclaurin polynomial if \( x_0 = 0 \)) and a remainder term (or truncation error).
2. The case \( n = 0 \) is just the Mean Value Theorem (MVT).

\[
f(x) = P_0(x) + R_0(x) = \underbrace{f(x_0)}_{P_0(x)} + \underbrace{f'(\xi(x))(x - x_0)}_{R_0(x)}
\]

3. If \( f \) is a polynomial of degree \( r \), then for \( n \geq r \), \( f(x) = P_n(x) \).

**Maple.** See [taylor polynomials.mw](taylor.polynomials.mw) and/or [taylor polynomials.pdf](taylor.polynomials.pdf)
EXAMPLE. Let \( f(x) = 2x \cos(2x) - (x - 2)^2, x_0 = 0. \)

(a) Find \( P_3(x) \) and \( P_4(x) \) and use them to approximate \( f(0.4). \)

\[
\begin{align*}
f'(x) &= 2 \cos(2x) - 4x \sin(2x) - 2(x - 2) \\
f''(x) &= -4 \sin(2x) - 4 \sin(2x) - 8x \cos(2x) - 2 \\
&= -8 \sin(2x) - 8x \cos(2x) - 2 \\
f'''(x) &= -16 \cos(2x) - 8 \cos(2x) + 16x \sin(2x) \\
&= -24 \cos(2x) + 16x \sin(2x) \\
f^{(4)}(x) &= 48 \sin(2x) + 16 \sin(2x) + 32x \cos(2x) \\
&= 64 \sin(2x) + 32x \cos(2x) \\
f^{(5)}(x) &= 128 \cos(2x) + 32 \cos(2x) - 64x \sin(2x) \\
&= 160 \cos(2x) - 64x \sin(2x)
\end{align*}
\]

\[
\begin{align*}
f(0) &= -4; \quad f'(0) = 6; \quad f''(0) = -2; \quad f'''(0) = -24; \quad f^{(4)}(0) = 0 \\
P_3(x) &= -4 + 6x + \left( \frac{-2}{2} \right) x^2 + \left( \frac{-24}{6} \right) x^3 = -4x^3 - x^2 + 6x - 4 \\
f(0.4) &\approx P_3(0.4) = -2.016
\end{align*}
\]

Since \( f^{(4)}(0) = 0, \)

\[
P_4(x) = P_3(x) = -4x^3 - x^2 + 6x - 4.
\]
(b) Use the error formula in Taylor’s Theorem to find an upper bound for the error $|f(0.4) - P_3(0.4)|$ and $|f(0.4) - P_4(0.4)|$.

By graphing, for $0 \leq x \leq 0.4$, $|f^{(4)}(x)| \leq 55.1$ and $|f^{(5)}(x)| \leq 160$. Then

$$|R_3(.4)| = |f(0.4) - P_3(0.4)| = \frac{|f^{(4)}(\xi(x))|}{4!}|.4 - 0|^4 \leq \frac{55.1}{24}(.4)^4 \leq .05878$$

by rounding up. Similarly,

$$|R_4(.4)| = |f(0.4) - P_4(0.4)| = \frac{|f^{(5)}(\xi(x))|}{5!}|.4 - 0|^5 \leq \frac{160}{120}(.4)^5 \leq .01366,$$

a better bound than for $|R_3(.4)|$.

The actual absolute error is

$$|f(0.4) - P_3(0.4)| = |(.8 \cos(.8) - 2.56) - (-2.016)| \approx .01337.$$

Notice that in this problem we are dealing with error at a point. □
\textbf{Example.} Approximate \(\sin(42^\circ)\) with error less than or equal to \(10^{-6}\).

\[
x = 42^\circ = 45^\circ - 3^\circ = \frac{\pi}{4} - \frac{\pi}{60} = \frac{7\pi}{30}
\]

Use Taylor polynomial with \(x_0 = \frac{\pi}{4}\), so \(x - x_0 = -\frac{\pi}{60}\).

To find \(n\), with \(f(x) = \sin(x)\) and \(\frac{7\pi}{30} \leq \xi(x) \leq \frac{\pi}{4}\),

\[
|R_n\left(\frac{7\pi}{30}\right)| = \frac{|f^{(n+1)}(\xi(x))|}{(n+1)!} \left|\frac{7\pi}{30} - \frac{\pi}{4}\right|^{n+1} \leq \frac{1}{(n+1)!} \left(\frac{\pi}{60}\right)^{n+1} \approx \frac{.05236^{n+1}}{(n+1)!}.
\]

Then \(\left|R_2\left(\frac{7\pi}{30}\right)\right| \approx \frac{.05236^3}{3!} = .0000239\) and

\[
\left|R_3\left(\frac{7\pi}{30}\right)\right| \approx \frac{.05236^4}{4!} = .000000313 < 10^{-6}, \text{ so take } n = 3.
\]

\[
f(x) = \sin(x), \quad f'(x) = \cos(x), \quad f''(x) = -\sin(x), \quad f'''(x) = -\cos(x),
\]

so

\[
f\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2}, \quad f'\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2}, \quad f''\left(\frac{\pi}{4}\right) = -\frac{\sqrt{2}}{2}, \quad f'''\left(\frac{\pi}{4}\right) = -\frac{\sqrt{2}}{2}.
\]

Thus

\[
\sin(42^\circ) = \sin\left(\frac{7\pi}{30}\right) \approx P_3\left(\frac{7\pi}{30}\right)
\]

\[
= f\left(\frac{\pi}{4}\right) + f'\left(\frac{\pi}{4}\right)(x - x_0) + \frac{f''\left(\frac{\pi}{4}\right)}{2!}(x - x_0)^2 + \frac{f'''\left(\frac{\pi}{4}\right)}{3!}(x - x_0)^3
\]

\[
= \frac{\sqrt{2}}{2} \left( -\frac{\pi}{60}\right) - \frac{\sqrt{2}}{4} \left( -\frac{\pi}{60}\right)^2 - \frac{\sqrt{2}}{12} \left( -\frac{\pi}{60}\right)^3 = .6691304
\]

From the TI-89, \(\sin\left(\frac{7\pi}{30}\right) = .6691306\),

so \(\left|\sin(42^\circ) - P_3\left(\frac{7\pi}{30}\right)\right| = .0000002 = 2 \times 10^{-7}. \square\)

\textbf{Maple.} See taylor's error.mw and/or taylor's error.pdf
3. Round-off Error and Computer Arithmetic

**Maple.** See computernumbers.mw and/or computernumbers.pdf

There are problems in using a finite subset of the reals (rational numbers, actually) to represent all reals.

Suppose each machine number can be represented as a $k$-digit decimal machine number in floating-point form as

$$\pm 0.d_1 d_2 \cdots d_k \times 10^n, \quad 1 \leq d_1 \leq 9, \quad 0 \leq d_i \leq 9, \quad i = 2, \ldots, k.$$  

Suppose $y = 0.d_1 d_2 \cdots d_k d_{k+1} d_{k+2} \times 10^n$ is in the numerical range of the machine.

Two methods for getting the floating-point form of $k$ digits $fl(y)$:

1. $fl(y) = 0.d_1 d_2 \cdots d_k \times 10^n$ is chopping.

2. **rounding:** add $5 \times 10^{n-(k+1)}$ and then chop to get $0.\delta_1 \delta_2 \cdots \delta_k \times 10^m$.

We refer to both cases as **round-off error**.

**Example** (of rounding). Suppose we wish to round $0.372648 \times 10^3$ and $0.372653 \times 10^3$ with $k = 4$ and $n = 3$, so

$$5 \times 10^{n-(k+1)} = 5 \times 10^{-2} = 5 \times 10^{-5} \times 10^3 = .00005 \times 10^3.$$

<table>
<thead>
<tr>
<th>$0.372648 \times 10^3$</th>
<th>$0.372653 \times 10^3$</th>
<th>$0.00005 \times 10^3$</th>
<th>$0.00005 \times 10^3$</th>
<th>$0.372698 \times 10^3$</th>
<th>$0.372703 \times 10^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.372 \times 10^3$</td>
<td>$0.3727 \times 10^3$</td>
<td>$0.372 \times 10^3$</td>
<td>$0.3727 \times 10^3$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Now chop:

$0.372 \times 10^3$ $\Box$
Measuring approximation errors:

Suppose $p^*$ approximates $p$:

**absolute error** $= |p - p^*|$

**relative error** $= \frac{|p - p^*|}{|p|}$

**Problem** (p.20, #1a). $p = \pi, \quad p^* = \frac{22}{7}$.

absolute error $= |\pi - \frac{22}{7}| \approx .001264489267$.

relative error $= \frac{|\pi - \frac{22}{7}|}{|\pi|} \approx 4.02499434755 \times 10^{-4}$. □

Since 4 is the largest nonnegative value of $t$ for which
relative error $< 5 \times 10^{-t}$

we say $p^*$ approximates $p$ to **4 significant digits**.

An approximation to floating-point arithmetic:

$x \oplus y = \text{fl}(\text{fl}(x) + \text{fl}(y)) \quad x \otimes y = \text{fl}(\text{fl}(x) \times \text{fl}(y))$

$x \ominus y = \text{fl}(\text{fl}(x) - \text{fl}(y)) \quad x \oslash y = \text{fl}(\text{fl}(x) \div \text{fl}(y))$

In Maple, “Digits:=t;” causes all arithmetic to be rounded to $t$ digits (10 is the default), and, for example,

$x \oplus y$ corresponds to evalf(evalf(x)+evalf(y)).
There are problems to be noted with subtraction:

**Example.** Use 4 digit chopping

\[
\text{fl}(\pi) = 3.141 \quad \text{relative error} \approx 1.89 \times 10^{-4}
\]

\[
\text{fl}(\frac{22}{7}) = 3.142 \quad \text{relative error} \approx 2.73 \times 10^{-4}
\]

Both representations have 4 significant digits, but

\[
\frac{22}{7} \Theta \pi = \text{fl}(3.142 - 3.141) = 0.001 \text{ with}
\]

\[
\text{relative error} = \frac{|\left(\frac{22}{7} - \pi\right) - 0.001|}{|\left(\frac{22}{7} - \pi\right)|} \approx 2.09 \times 10^{-1},
\]

yielding a result with only 1 significant digit. \(\square\)

Thus subtracting nearly equal numbers cancels significant digits, and these cannot be recovered. We get small absolute error, but large relative error.
4. Errors in Scientific Computation

Sequencing of operations can make a difference:

**Problem** (p.27, #1b). Use four digit rounding to find the roots of \( \frac{1}{3}x^2 + \frac{123}{4}x - \frac{1}{6} = 0 \). Note that the exact roots round to \(-92.2554197358\) and \(-.005419735789\).

The quadratic formula for \( ax^2 + bx + c = 0 \) gives

\[
x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}
\]
as solutions.

\[
a = .3333, \quad b = 30.75, \quad c = -.1667, \quad b^2 = 945.6, \quad ac = -.0556, \quad 4ac = -.2224
\]
\[
b^2 - 4ac = 945.8, \quad \sqrt{b^2 - 4ac} = 30.75, \quad -b + \sqrt{b^2 - 4ac} = 0
\]
\[
-b - \sqrt{b^2 - 4ac} = -61.50, \quad \text{and} \quad 2a = .6666, \quad \text{so}
\]
\[
x_1 = \frac{0}{.6666} = 0 \quad \text{and} \quad x_2 = \frac{-61.50}{.6666} = -92.26, \quad \text{and}
\]
relative error=1 and relative error=4.96 \times 10^{-5}.

Thus we have no and 5 significant digits. But rationalizing numerators,

\[
x_1 = \frac{-2c}{b + \sqrt{b^2 - 4ac}} \quad \text{and} \quad x_2 = \frac{-2c}{b - \sqrt{b^2 - 4ac}},
\]
so \( 2c = -.3334 \), \( b + \sqrt{b^2 - 4ac} = 61.5 \), and \( b - \sqrt{b^2 - 4ac} = 0 \), giving

\[
x_1 = \frac{-.3334}{61.5} = .005421 \quad \text{and} \quad x_2 = \frac{-.3334}{0}, \quad \text{undefined.}
\]

Relative error for \( x_1 \) is \( 2.3 \times 10^{-4} \), giving 4 significant digits. □

**Maple.** See [nested.mw](nested.mw) and/or [nested.pdf](nested.pdf)
algorithm — a procedure that unambiguously describes a finite, ordered sequence of steps to solve a problem or approximate a solution.

stable algorithm — small changes in initial data produce small changes in the final results

unstable algorithm — not stable

conditionally stable algorithm — stable only for certain choices of initial data

Growth of round-off error:

Suppose an error \( E_0 > 0 \) is introduced at some stage of the calculations and the error after \( n \) subsequent operations is \( E_n \).

linear growth of (relative) error — \( E_n \approx CnE_0 \), where \( C \) is a constant independent of \( n \). This reflects a stable algorithm.

exponential growth of (relative) error — \( E_n \approx C^mE_0 \) for some \( C > 1 \). This reflects an unstable algorithm.

Maple. See convergence.mw and/or convergence.pdf.
1. MATHEMATICAL PRELIMINARIES AND ERROR ANALYSIS

Rates of convergence for sequences:

Our algorithms often generate sequences of values. Suppose \( \{ \alpha_n \} \to \alpha \) and \( \beta_n \to 0 \). Recall
\[
\{ \alpha_n \} = \alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n, \ldots
\]
and
\[
\{ \alpha_n \} \to \alpha \text{ means } \lim_{n \to \infty} \alpha_n = \alpha.
\]
If there exists \( K > 0 \) such that
\[
|\alpha_n - \alpha| \leq K|\beta_n| \text{ for large } n,
\]
we say \( \{ \alpha_n \} \) converges to \( \alpha \) with a rate of convergence \( O(\beta_n) \) ("big oh of \( \beta_n \)"") or \( \{ \alpha_n \} \to \alpha \) with rate of convergence \( O(\beta_n) \). Usually, \( \beta_n = \frac{1}{n^p} \) for some \( p > 0 \). We want the largest such value \( p \) so that \( \alpha_n = \alpha + O\left(\frac{1}{n^p}\right) \).

**Example.** Consider \( \left\{ \frac{2n + 3}{n + 7} \right\} \to 2 \) and \( \left\{ \frac{2n^3 + 3}{n^3 + 7} \right\} \to 2 \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \frac{2n + 3}{n + 7} )</th>
<th>( \frac{2n^3 + 3}{n^3 + 7} )</th>
<th>( \frac{1}{n} )</th>
<th>( \frac{1}{n^3} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.625000</td>
<td>.625000</td>
<td>1.000000</td>
<td>1.000000</td>
</tr>
<tr>
<td>2</td>
<td>.777778</td>
<td>1.266667</td>
<td>.500000</td>
<td>.125000</td>
</tr>
<tr>
<td>3</td>
<td>.900000</td>
<td>1.676471</td>
<td>.333333</td>
<td>.037037</td>
</tr>
<tr>
<td>4</td>
<td>1.000000</td>
<td>1.845070</td>
<td>.250000</td>
<td>.015625</td>
</tr>
<tr>
<td>5</td>
<td>1.083333</td>
<td>1.916667</td>
<td>.200000</td>
<td>.008000</td>
</tr>
<tr>
<td>6</td>
<td>1.153846</td>
<td>1.950673</td>
<td>.166667</td>
<td>.004630</td>
</tr>
</tbody>
</table>

\[
\left| \frac{2n + 3}{n + 7} - 2 \right| = \left| \frac{2n + 3 - 2n - 14}{n + 7} \right| = \left| \frac{-11}{n + 7} \right| = \frac{11}{n + 7} \leq 11 \cdot \frac{1}{n} \quad (\beta_n = \frac{1}{n})
\]

so \( \frac{2n + 3}{n + 7} = 2 + O\left(\frac{1}{n}\right) \) and the rate of convergence is \( O\left(\frac{1}{n}\right) \) since \( \lim_{n \to \infty} \frac{1}{n} = 0 \).
\[
\left| \frac{2n^3 + 3}{n^3 + 7} - 2 \right| = \left| \frac{2n^3 + 3 - 2n^3 - 14}{n^3 + 7} \right| = \left| \frac{-11}{n^3 + 7} \right| = \frac{11}{n^3 + 7} \leq \frac{11}{n^3} \quad (\beta_n = \frac{1}{n^3})
\]

so \( \frac{2n^3 + 3}{n^3 + 7} = 2 + O\left(\frac{1}{n^3}\right) \) and rate of convergence is \( O\left(\frac{1}{n^3}\right) \) since \( \lim_{n \to \infty} \frac{1}{n^3} = 0 \).

\( \square \)

**Problem** (p.29 #10b). Find the rate of convergence of \( \{ \sin \frac{1}{n^2} \} \).

We know \( \lim_{n \to \infty} \frac{1}{n^2} = 0 \). From calculus, \( \lim_{h \to 0} \frac{\sin h}{h} = 1 \Rightarrow \lim_{k^2 \to 0} \frac{\sin k^2}{k^2} = 1 \Rightarrow \)

\( \left( \text{letting } k = \frac{1}{n} \right) \lim_{n \to \infty} \frac{\sin \frac{1}{n^2}}{\frac{1}{n^2}} = 1 < 2 \). Then, for large \( n \), \( \sin \frac{1}{n^2} < 2 \cdot \frac{1}{n^2} \).

Since \( \lim_{n \to \infty} \frac{1}{n^2} = 0 \), \( \sin \frac{1}{n^2} = 0 + O\left(\frac{1}{n^2}\right) \), so rate of convergence is \( O\left(\frac{1}{n^2}\right) \). \( \square \)

We can also use the big oh notation to show how some divergent sequences grow as \( n \) becomes large. If positive constants \( p \) and \( K \) exist with

\[ |\alpha_n| \geq Kn^p \] for all large values of \( n \),

we say that \( \{ \alpha_n \} \to \infty \) with rate of convergence \( O(n^p) \).

Finally, we can use big oh for functions.

Suppose \( \lim_{h \to 0} G(h) = 0 \) and \( \lim_{h \to 0} F(h) = L \). If there is \( K > 0 \) such that

\[ |F(h) - L| \leq K|G(h)| \] for \( h \) small enough, then \( F(h) = L + O(G(h)) \).

Usually, we like to use \( G(h) = h^p \), and want the largest \( p \) such that

\[ F(h) = L + O(h^p) \].
Recall.

\[ e^h = 1 + h + \frac{h^2}{2} + \cdots + \frac{h^n}{n!} + \frac{e^\xi}{(n+1)!} h^{n+1} \]

\[ \sin h = h - \frac{h^3}{3!} + \frac{h^5}{5!} - \cdots + (-1)^n \frac{h^{2n+1}}{(2n+1)!} + \frac{(-1)^{n+1} \cos \xi}{(2n+3)!} h^{2n+3} \]

\[ \cos h = 1 - \frac{h^2}{2!} + \frac{h^4}{4!} - \cdots + (-1)^n \frac{h^{2n}}{(2n)!} + \frac{(-1)^{n+1} \cos \xi}{(2n+2)!} h^{2n+2} \]

\[ \frac{1}{1+h} = 1 - h + h^2 - \cdots + (-1)^n h^n + \frac{(-1)^{n+1}}{(1+\xi)^{n+2}} h^{n+1} \]

**Problem** (p.29 #11b). Find the rate of convergence for \( \lim_{h \to 0} \frac{1 - e^h}{h} = -1 \).

\[ e^h = 1 + h + \frac{e^\xi}{2} h^2 \Rightarrow 1 - e^h = -h - \frac{e^\xi}{2} h^2 \Rightarrow \frac{1 - e^h}{h} = -1 - \frac{e^\xi}{2} h \Rightarrow \]

\[ \left| \frac{1 - e^h}{h} + 1 \right| = \left| -\frac{e^\xi}{2} h \right| < \frac{2}{2} |h| \text{ for } h \text{ small enough. Thus } \frac{1 - e^h}{h} = -1 + O(h). \]

**Example.** Find \( \lim_{h \to 0} \frac{\cos h - 1 + \frac{1}{2} h^2}{h^4} \) and its rate of convergence.

\[ \cos h = 1 - \frac{1}{2} h^2 + \frac{1}{24} h^4 - \frac{\cos \xi}{720} h^6 \text{ for } 0 < \xi < h \Rightarrow \]

\[ \cos h - 1 + \frac{1}{2} h^2 = \frac{1}{24} h^4 - \frac{\cos \xi}{720} h^6 \Rightarrow \]

\[ \frac{\cos h - 1 + \frac{1}{2} h^2}{h^4} = \frac{1}{24} \underbrace{- \frac{\cos \xi}{720} h^2} \Rightarrow \]

\[ \text{Looks like } L + O(h^2) \]

\[ \left| \frac{\cos h - 1 + \frac{1}{2} h^2}{h^4} - \frac{1}{24} \right| = \left| \frac{\cos \xi}{720} h^2 \right| \leq \frac{1}{720} |h^2|. \]
Thus \( \lim_{h \to 0} \frac{\cos h - 1 + \frac{1}{2}h^2}{h^4} = \frac{1}{24} \) with rate of convergence \( O(h^2) \) or

\[
\frac{\cos h - 1 + \frac{1}{2}h^2}{h^4} = \frac{1}{24} + O(h^2). \qed
\]