Numerical Analysis (10th ed)

R.L. Burden, J. D. Faires, A. M. Burden

Chapter 1

Mathematical Preliminaries and Error Analysis

Chapter 1.1: Review of Calculus*

The definitions and theorems you will encounter in this section are from Calculus. They are necessary building blocks for much of what is to come in the course. The main "take-aways" in this section are:

- How do we know that a function, has at least one solution on a given interval? This concept will be used to identify intervals where solutions to functions exist.
- How do we find the maximum of the absolute value of a function over a given interval? This concept will be used extensively in determining error bounds.
- How can we approximate a functional value around a region of interest? This concept will be used extensively throughout the course.

To answer the first bullet point let's consider the following example: How do we go about showing that a function, $f(x) = x - 1 - (\ln(x))^x$ has at least one solution on the interval [3,4]? We can use the Intermediate Value Theorem to show that if f is continuous on [3,4] and the functional value changes from positive to negative at the end points, then there must be at least one c value in (3,4) for which $f(c) = 0$.

How do we know that the given function f is continuous on [3,4]? That is easy, each term is defined on that interval so the sum of the terms is defined as well. Now, let's check the functional values at the endpoints of the interval by defining the function and then computing the functional values at $x = 3$ and $x = 4$:

$f := x \to x - 1 - (\ln(x))^x$

$$x \to x - 1 - \ln(x)^x \quad \textbf{(1.1)}$$

$f(3)$

$$2 - \ln(3)^3 \quad \textbf{(1.2)}$$

at 5 digits →

$$0.6741 \quad \textbf{(1.3)}$$

$f(4)$

$$3 - 16\ln(2)^4 \quad \textbf{(1.4)}$$

at 5 digits →

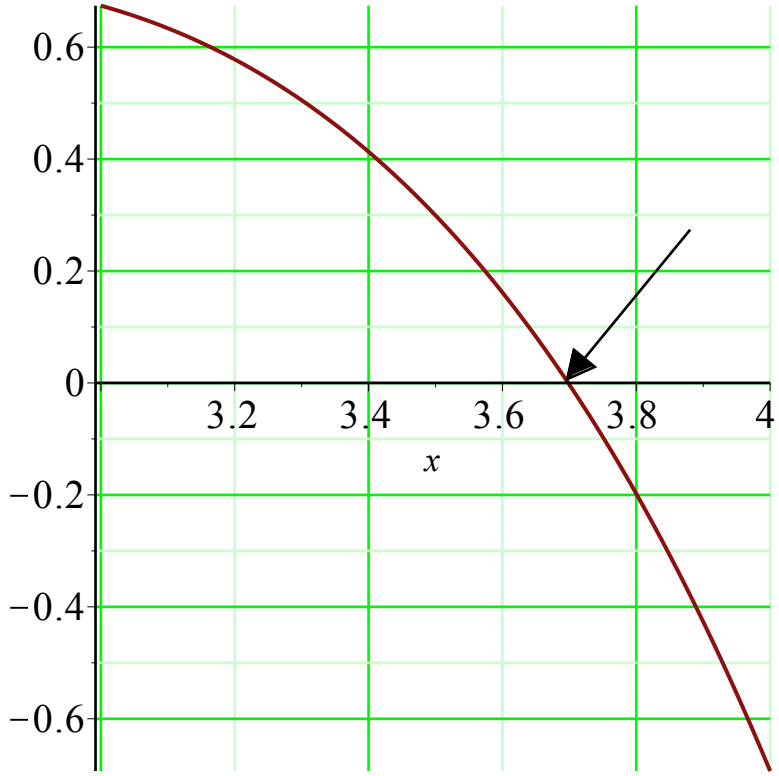
$$-0.6934 \quad \textbf{(1.5)}$$

Notice how the functional value changes sign at the endpoints. Certainly, $f(4) < 0 < f(3)$. Since the

function is continuous on [3,4] (no jumps, holes, or breaks), the Intermediate Value Theorem guarantees that there is at least one value, c, in the interval (3,4) for which the functional value at $x = c$ is zero, that is, $f(c) = 0$.

Let's verify this with a graph:

$plot\left(x - 1 - (\ln(x))^{x}, x = 3..4, axis = [gridlines = [colour = green, majorlines = 2]]\right)$



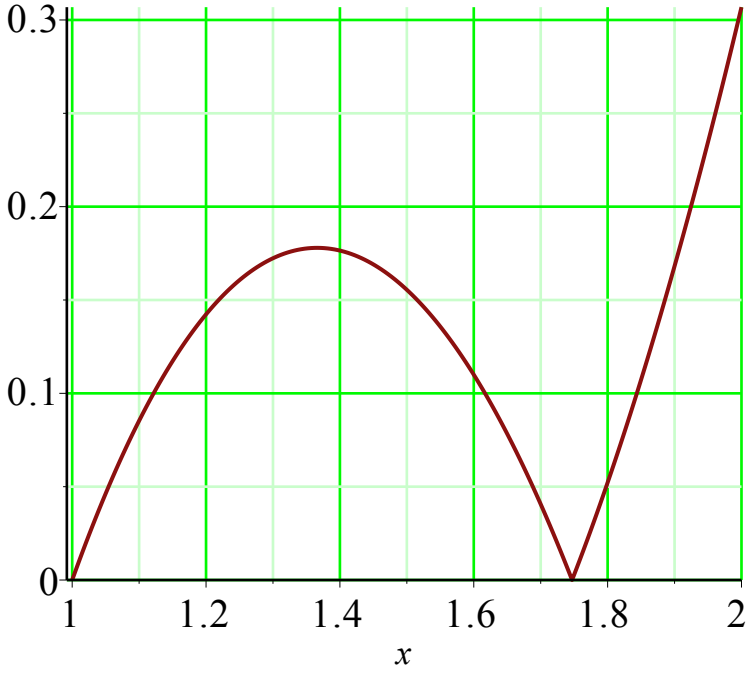
To answer the second bullet point we will use the Extreme Value Theorem in the following example to address how do we go about maximizing $\left|(x-1)^2 - \ln(x)\right|$over the interval [1,2]? We can use the Extreme Value Theorem to show that if f is continuous on [1,2] then not only is there a value c for which $g(x) < g(c)$ for all values of $x \in (1,2)$, but also if $g(x)$ is continuously differentiable on (1,2), then the maximum will either occur at one of the endpoints $g(1)$ or $g(2)$ or at a point, c, within the interval (1,2) where the derivative $g'(c) = 0$.

How do we know that the given function g is continuous on [1,2]? As before, each term is defined on that interval so the sum of the terms is defined as well.

$g := x \rightarrow (x-1)^2 - \ln(x)$:

If we look at the graph of $g(x) = \left|(x-1)^2 - \ln(x)\right|$ we can see that we have an absolute maximum at the right endpoint $x = 2$.

$plot\left(\left|(x-1)^2 - \ln(x)\right|, x = 1..2, axis = [gridlines = [colour = green, majorlines = 2]]\right)$



However, we need to show this algebraically. In order to do that, we will need to find the first derivative of the given function:

$gp := \mathrm{D}(g)(x)$

$$2x - 2 - \frac{1}{x} \qquad \textbf{(1.6)}$$

So we see that $g'(x) = 2x - 2 - \frac{1}{x}$ which is clearly differentiable on the given interval. We now need to find where this derivative is zero. To do that, we set $g'(x) = 2x - 2 - \frac{1}{x} = \frac{(2x^2 - 2x - 1)}{x} = 0$ and find x. Working some algebra magic,

$solve(gp = 0, x)$:

$\frac{1}{2} + \frac{1}{2}\sqrt{3} \xrightarrow{\text{at 5 digits}} 1.3660$

$\frac{1}{2} - \frac{1}{2}\sqrt{3} \xrightarrow{\text{at 10 digits}} -0.3660254040$

Although we find that $f'(x) = 0$ when $x = \frac{(1 \pm \sqrt{3})}{2}$ or $f'(x)$ is undefined when $x = 0$. Since 0 and $\frac{1}{2} - \frac{1}{2}\sqrt{3}$ are not in the domain of our function, we can ignore those values of x.

Now, let's compute the functional value at $x = \frac{1}{2} + \frac{1}{2}\sqrt{3}$ and at the endpoints of the interval [1,2].

$|g(1)|$

$$0 \tag{1.7}$$

at 5 digits $\longrightarrow$

$$0. \tag{1.8}$$

$$\left|g\left(\frac{1}{2}+\frac{1}{2}\sqrt{3}\right)\right|$$

$$-\left(-\frac{1}{2}+\frac{1}{2}\sqrt{3}\right)^2+\ln\left(\frac{1}{2}+\frac{1}{2}\sqrt{3}\right) \tag{1.9}$$

at 5 digits $\longrightarrow$

$$0.17790 \tag{1.10}$$

$|g(2)|$

$$1-\ln(2) \tag{1.11}$$

at 5 digits $\longrightarrow$

$$0.30685 \tag{1.12}$$

Of the three functional values, $|f(2)| = 0.30685$ is the largest and therefore, the maximum value of the function is 0.30685 which occurs at the right endpoint of the interval when x=2. Remember, that the maximum may not occur at an endpoint so all of the work must be done. Clearly, the graph above supports our algebra.

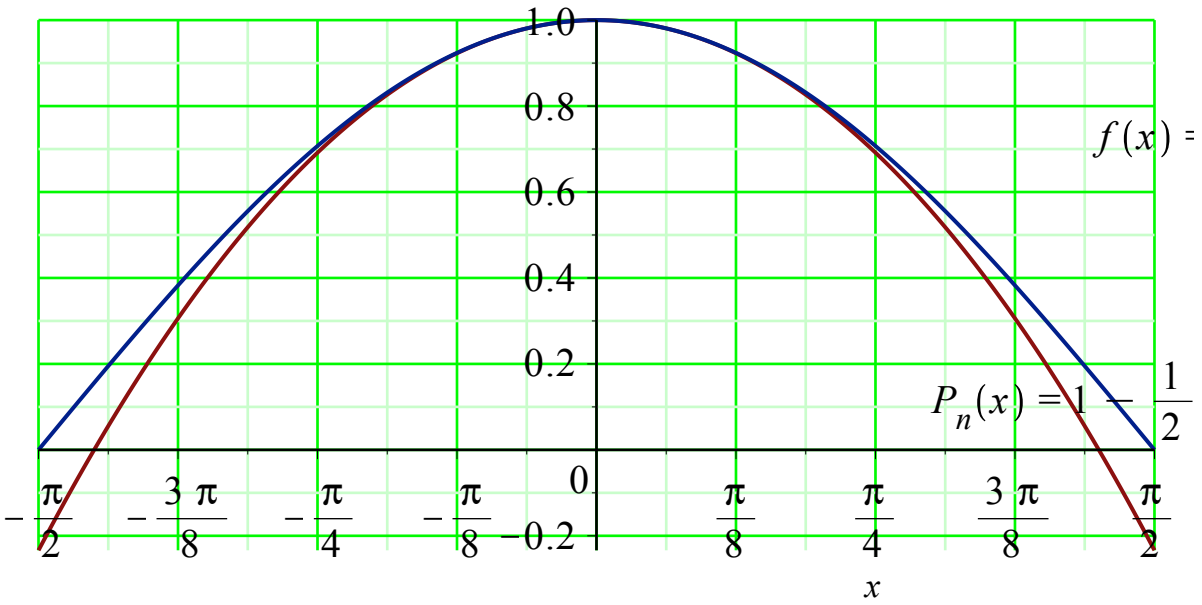
Finally, let's consider how we can approximate a functional value around some are of interest. As long as the function is n-times continuously differentiable on [a, b], the $(n+1)^{st}$ derivative exists on [a,b], and $x_0 \in [a, b]$, then we know that a number $\varsigma(x)$ between x_0 and x exists for which the n-th degree Taylor polynomial about x_0 is given by $P_n(x) = \sum_{i=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^n$ with error term $R_n(x) = \frac{f^{(n+1)}(\varsigma(x))}{(n+1)!}(x-x_0)^{n+1}$.

Let's look at a simple example. Suppose we want to find the second degree Taylor polynomial for $f(x) = \sin(x)$ about x_0= 0 and then use it to approximate $f(0.35)$.
Since we are looking for the second degree Taylor polynomial, we need to find the first and second derivatives for $P_n(x)$ and then the 3rd derivative for the error $R_n(x)$.

$$P_2(x) = \frac{y0\cdot(x-0)^0}{0!}+\frac{y1\cdot(x-0)^1}{1!}+\frac{y2\cdot(x-0)^2}{2!}$$

$$P_n(x) = 1-\frac{1}{2}x^2 \tag{1.13}$$

$$plot\left(\left\{\cos(x), 1-\frac{1}{2}x^2\right\}, x=-\frac{\pi}{2}..\frac{\pi}{2}, axis=[gridlines=[colour=green, majorlines=2]]\right)$$



Now suppose we want to approximate $\cos(0.01)$ using the Taylor polynomial of degree 2. Then we have $P_2(0.01) = 1 - \frac{1}{2}(0.01)^2$

$$P_2(0.01) = 0.9999500000 \qquad \textbf{(1.14)}$$

Graphically, we can see that this is very close to the exact value of $\cos(0.01) \xrightarrow{\text{at 10 digits}}$ 0.9999500004. To find the error bound, we use the Extreme Value theorem to maximize $|R_2(x)|$

$= \left|\frac{f3}{3!}(x-0)^3\right| = \left|-\frac{1}{6}\sin(\varsigma)\cdot x^3\right|$ on the interval (0, 0.01). We know that the derivative is, $\frac{1}{6}\sin(\varsigma)\cdot x^3$ so that we have a bound of:

$\frac{1}{6}\cos(\varsigma)\cdot(0.01)^3 < 0.1666666667 \text{ x } 10^{-6}\ \cos(\varsigma) < 0.1666666667 \text{ x } 10^{-6}(1) = 0.\overline{16} \text{ x } 10^{-6}$

The actual error is $|\cos(0.01) - P_2(0.01)| = |0.9999500004 - 0.9999500000| = 4 \text{ x } 10^{-10}$

Chapter 1.2: Round-off Errors and Computer Arithmetic*

The main "take-aways" in this section are:
• What is the difference between absolute error and relative error?
• How do we properly compute expressions using rounding or chopping arithmetic?

Before we begin our discussion, note that it is important to write numbers in normalized decimal floating-point form: $\pm 0.d_1d_2\ldots d_k$, $1 \le d_1 \le 9$, $0 \le d_i \le 9$, $i = 2,\ldots,k$ as we proceed with any computation in this section.

Suppose p^* is an approximation to $p \neq 0$. We define

Actual Error: $p - p^*$ represents the signed difference between the accepted exact value and the approximate value, that is how much your approximation is above or below the exact value.

Absolute Error: $|p - p^*|$ represents the amount of physical error in a measurement, that is, how much different is your approximation to the exact value.

Relative Error: $\frac{|p - p^*|}{|p|}$ represents how good the error in a measurement is relative to what is being measured, that is, what is the percent difference between your approximation and the exact value.

There are numerous examples of this in the text and on the internet, so additional examples will not be provided here.

The next bullet point often gives students problems because they forget to round or chop at EACH step of a computation. Students must be aware that rounding (chopping) is NOT done in the final stage of computation, but at each and every step of the computation. For example, suppose we want to use four-digit chopping on the following computation:
$18\,\pi - 3.120128$. To get what is considered the exact value, take out your calculator and compute $p = 18\,\pi - 3.120128 \approx 53.42853976$.

But in four-digit chopping we would compute as follows:
$a = 18 = 0.18 \times 10^2,\ b = \pi = 3.14159265 = 0.3141 \times 10^1;\ c = 3.120128 = 0.3120 \times 10^1$
Step 1: $(0.18 \times 10^2)(0.3141 \times 10^1) = 0.5653 \times 10^2 = 0.5653 \times 10^2 = 56.53$.
Step 2: $56.53 - 3.120 = 53.41 = p^*$

The actual error is $p - p^* = 53.42853976 - 53.41 = 0.01853976$

The absolute error is $|p - p^*| = |53.42853976 - 53.41| = 0.01853976$

The relative error is $\frac{|p - p^*|}{|p|} = \frac{|53.42853976 - 53.49|}{53.42853976} = 0.000347$

Chapter 1.3: Algorithms and Convergence*

The main "take-aways" in this section are:
• How do we determine the stability of an algorithm?
• How do determine the rate of convergence of an iterative process?

So how do we determine the stability of an algorithm? We focus on the sequence that is generated by the algorithm and look at the relative error between the formula and the finite-digit approximation arithmetic at each step. We note that if small changes in the initial data produce correspondingly small changes in the final results, then an algorithm is considered to be stable.

Formally, what this means is that for error $E_0 > 0$ introduced at some stage in the calculation after n subsequent operations we have **linear** error growth if $E_n \approx CnE_0$ for some constant C and **exponential** error growth if $E_n \approx C^nE_0$ for some constant $C > 1$.

Let's take a closer look at the illustration in the text.

For any constants c_1 and c_2, we can show that
$p_n = c_1\left(\frac{1}{3}\right)^n + c_2 3^n$ is a solution to the recursive equation $p_n = \frac{10}{3}p_{n-1} - p_{n-2}$ for $n = 2, 3,...$

To show this, we substituted the solution at p_{n-1} and p_{n-2} into the recursive equation to verify that we get a true statement. This is shown in the text. Now, suppose that we are given $p_0 = 1$ and $p_1 = \frac{1}{3}$.

We will use $p_n = c_1\left(\frac{1}{3}\right)^n + c_2 3^n$ to determine unique values for the constants c_1 and c_2 as follows:

$$p_0 \rightarrow 1 = c_1\left(\frac{1}{3}\right)^0 + c_2 3^0 = c_1 + c_2$$
$$p_1 \rightarrow \frac{1}{3} = c_1\left(\frac{1}{3}\right)^1 + c_2 3^1 = \frac{1}{3}c_1 + 3\,c_2$$

Solving the system formed for the constants c_1 and c_2 yields $c_1 = 1$, $c_2 = 0$. Therefore, we have that $p_n = \left(\frac{1}{3}\right)^n$ for all n.

Using this same process and five-digit rounding arithmetic and $p_0 = 1.0000$ and $p_1 = 0.33333$ to compute c_1 and c_2 we will find that
$c_1 = 1.0000$, $c_2 = -0.12500 \times 10^{-5}$.
The sequence $\{\hat{p}_n\}$, $n = 0 ... \infty$ generated is then given by $\hat{p}_n = 1.0000\left(\frac{1}{3}\right)^n - 0.12500 \times 10^{-5}\left(3^n\right)$ has round-off error
$p_n - \hat{p}_n = \left(\frac{1}{3}\right)^n - \left(1.0000\left(\frac{1}{3}\right)^n - 0.12500 \times 10^{-5}\left(3^n\right)\right) = 0.12500 \times 10^{-5}\left(3^n\right)$. Thus, this procedure is unstable because the error grows **exponentially** with n.

To address the second bullet, we suppose that $\{\beta_n\}$, $n = 1,.. \infty$ is a sequence known to converge to zero, and $\{\alpha_n\}$, $n = 1,.. \infty$ converges to a number α. If a positive constant K exists with $|\alpha_n - \alpha| \le K|\beta_n|$ for large n, then we say that $\{\alpha_n\}$, $n = 1,.. \infty$ converges to α with **rate, or order, of convergence** $O(\beta_n)$.
The example in the text is straight forward, and will not be discussed here.

We can also use the O notation to describe the rate at which functions converge as follows:
If $\lim_{h \to 0} G(h) = 0$ and $\lim_{h \to 0} F(h) = \text{L}$ and a positive constant K exists with $|F(h) - L| \leq K|G(h)$ for sufficiently small h, then we say that $F(h) = L + OG(h))$.

Again, the example in the text is straight forward, and will not be discussed here.