Without a doubt, the birth of calculus is a glorious yet traumatic time for mathematics. It is appropriate we start the course by visiting there since it is both the power of the subject as well as the open wounds it left behind that drove mathematical activity for several consequent decades if not centuries.

Its two creators-discoverers Isaac Newton (1642-1727) and Gottfried Leibniz (1646-1716) would probably not recognize what we today refer to as the fundamental theorem of calculus as such.

If one does a Google search of the FTC, one usually encounters some kind of statement connecting the notions of integration and differentiation as inverse operations. Or equivalently, the connection between the tangents of the graph of one function and the area under the curve of another. Yet, by the time Newton and Leibniz come around, Fermat, Descartes and Huygens among many others had already discussed derivatives and their connection with tangents and extremal values of functions, and the areas under curves of varied complexity had been computed by basically doing Riemann sums integration. Finally, the connection between the two processes of integration and differentiation had been foreseen by Gregory and Barrow, and Newton had been exposed to these ideas in Barrow's lectures. As we will see below, they (Leibniz and Newton) had their own ideas on what calculus was all about, and they are different from what a student receives in their freshman course.

**Leibniz & the FTC**

Like Fermat, Gottfried Leibniz, was not a mathematician by trade. He was a diplomat who traveled widely, and as such came to meet and discuss mathematics with all the best-known mathematicians and scientists of his time, including Huygens and Newton.

It is, in a sense, unfortunate that Leibniz met and corresponded with Newton, since he will, many years after the meeting, be accused of plagiarizing his ideas on calculus from Newton. A long, scandalous dispute followed, and although his name was eventually cleared, the dispute left a bitter taste in the soul of mathematicians on both sides of the English Channel. This led to a partial isolation of English mathematicians from those in the Continent, where calculus, and its consequent disciplines such as differential equations, will explode into a massive and powerful discipline. Leibniz would die unbeknownst to the world outside of mathematics and in relative poverty.

It is fortunate, however, that Leibniz met Huygens, since it is a question posed to him by Huygens that possibly stimulated Leibniz's discovery of the connection between integration and differentiation.

Huygens asked what the sum of the reciprocals of the triangular numbers added to:
Several centuries prior, Oresme had proven the divergence of the harmonic series. We pause to give a different proof of this divergence that was given shortly after Leibniz:

**A Different Argument for the Divergence of the Harmonic Series**

First we need a simple fact. Let \( n \) be a positive integer. Then \( \frac{1}{n-1} + \frac{1}{n+1} > \frac{2}{n} \) since the LFS equals \( \frac{2n}{n^2-1} > \frac{2n}{n^2} \). Thus the sum of three consecutive reciprocals is greater than 3 times the middle reciprocal: \( \frac{1}{n-1} + \frac{1}{n} + \frac{1}{n+1} > \frac{3}{n} \). For example, if \( n = 7 \), then

\[
\frac{1}{6} + \frac{1}{7} + \frac{1}{8} > \frac{3}{7},
\]

since the left-hand side is \( \frac{73}{168} > \frac{72}{168} = \frac{3}{7} \).

Assume then, by way of contradiction, that the harmonic series converges. Let \( S \) denote the total sum. We know then that

\[
S = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \frac{1}{10} + \frac{1}{11} + \frac{1}{12} + \frac{1}{13} + \frac{1}{14} + \frac{1}{15} + \frac{1}{16} + \cdots
\]

\[
= \frac{1}{1} + \left( \frac{1}{2} + \frac{1}{3} + \frac{1}{4} \right) + \left( \frac{1}{5} + \frac{1}{6} + \frac{1}{7} \right) + \left( \frac{1}{8} + \frac{1}{9} + \frac{1}{10} \right) + \left( \frac{1}{11} + \frac{1}{12} + \frac{1}{13} \right) + \left( \frac{1}{14} + \frac{1}{15} + \frac{1}{16} \right) + \cdots
\]

\[
> 1 + \frac{3}{3} + \frac{3}{9} + \frac{3}{3} + \frac{3}{9} + \frac{3}{15} + \cdots = 1 + \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots = 1 + S
\]

which is certainly nonsensical. Very elegant indeed!

As it turned out, the question Huygens asked had been answered already, but unknown to Leibniz, he plunged ahead into the problem.

He understood that from a given sequence: \( \alpha : a_1, a_2, a_3, a_4, \ldots \), one could obtain two other ones, the sum and the difference. The sum (or, as we would say today, the series), \( \Sigma(\alpha) \), is defined as follows \( \Sigma(\alpha) : a_1, a_1 + a_2, a_1 + a_2 + a_3, a_1 + a_2 + a_3 + a_4, \ldots \). For example, if we start with the simple sequence \( 1,1,1,\ldots \) and take consecutive sums of it, we obtain the following configuration:

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Sigma(\alpha) )</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>( \Sigma^2(\alpha) )</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>10</td>
<td>15</td>
<td>21</td>
<td>28</td>
<td>36</td>
<td>45</td>
<td>55</td>
<td>66</td>
<td>78</td>
</tr>
<tr>
<td>( \Sigma^3(\alpha) )</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>20</td>
<td>35</td>
<td>56</td>
<td>84</td>
<td>120</td>
<td>165</td>
<td>220</td>
<td>286</td>
<td>364</td>
</tr>
<tr>
<td>( \Sigma^4(\alpha) )</td>
<td>1</td>
<td>5</td>
<td>15</td>
<td>35</td>
<td>70</td>
<td>126</td>
<td>210</td>
<td>330</td>
<td>495</td>
<td>715</td>
<td>1001</td>
<td>1365</td>
</tr>
</tbody>
</table>
If we think of this as a matrix with the rows and columns labeled by 0,1,2,…, then we see that we are dealing with Pascal’s triangle since the \( i,j \)–entry is nothing but
\[
\binom{i+j}{i} \quad \text{It is in this form that Pascal originally wrote his triangle, and we can see the diagonals of our table are indeed the rows of the traditional Pascal triangle.}
\]

The second one: the difference, \( \Delta(\alpha) \), is defined as follows, \( \Delta(\alpha) : b_1, b_2, b_3, \ldots \) where:
\[
b_1 = a_2 - a_1, \quad b_2 = a_3 - a_2, \quad b_3 = a_4 - a_3, \quad \text{etcetera.}
\]
Thus, for example, if \( \alpha \) is the sequence of triangular numbers, \( \alpha: 1,3,6,10,15,\ldots \) then \( \Delta(\alpha) \) is \( 2,3,4,5,\ldots \).

In the context of sequences, the relation between the difference and the sum, and their concatenation, is easily understood. If \( \alpha : a_1, a_2, a_3, \ldots \) Then
\[
\Delta(\Sigma(\alpha)) = a_2, a_3, a_4, \ldots,
\]
which is almost \( \alpha \) (all we would need to recover \( \alpha \) is attach \( a_1 \) at the beginning. Also
\[
\Sigma(\Delta(\alpha)) = a_2 - a_1, a_3 - a_2, a_4 - a_3, \ldots,
\]
and again \( \alpha \) is easily recoverable from this sequence, once \( a_1 \) is given.

Leibniz set out to a parallel construction to the table above, but instead of consecutive sums he took consecutive differences. Since he was starting with a decreasing sequence, he needed to modify the difference to mean:
\[
b_1 = a_1 - a_2, \quad b_2 = a_2 - a_3, \quad b_3 = a_3 - a_4, \quad \text{etcetera.}
\]

His first (row) sequence was the sequence of harmonic numbers:
\[
\alpha: \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots
\]

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( \Delta(\alpha) )</th>
<th>( \Delta^2(\alpha) )</th>
<th>( \Delta^3(\alpha) )</th>
<th>( \Delta^4(\alpha) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/2</td>
<td>1/3</td>
<td>1/4</td>
<td>1/5</td>
</tr>
<tr>
<td>1/2</td>
<td>1/6</td>
<td>1/12</td>
<td>1/20</td>
<td>1/30</td>
</tr>
<tr>
<td>1/3</td>
<td>1/12</td>
<td>1/30</td>
<td>1/60</td>
<td>1/105</td>
</tr>
<tr>
<td>1/4</td>
<td>1/20</td>
<td>1/60</td>
<td>1/140</td>
<td>1/280</td>
</tr>
<tr>
<td>1/5</td>
<td>1/30</td>
<td>1/105</td>
<td>1/280</td>
<td>1/630</td>
</tr>
</tbody>
</table>

Then he readily observed that the second row consisted of the halves of the reciprocals of the triangular numbers. Hence if we let \( \alpha : a_1, a_2, a_3, a_4, \ldots \) and \( \Delta(\alpha) := b_1, b_2, b_3, \ldots \), then we know \( b_1 = a_1 - a_2, \quad b_2 = a_2 - a_3, \quad b_3 = a_3 - a_2, \quad \text{etcetera.} \) So, in a similar fashion to the discussion above, \( \Sigma(\Delta(\alpha)) := a_1 - a_2, a_1 - a_3, \ldots \) So the sum of all the reciprocals of the triangular numbers, which is
\[
2 \lim_{n \to \infty} \Sigma(\beta) = 2 \lim_{n \to \infty} \Sigma(\Delta(\alpha)) = 2a_1 = 2
\]
since \( a_n \to 0 \) as \( n \to \infty \).

This was very exciting to him, since he realized he could adequately add one sequence by
simply taking differences of another. Although certainly that reminds one of the basic ideas behind the fundamental theorem of calculus, it did not become real calculus until he pushed it further.

The notion of **differential** showed up exactly as an application of the difference of a sequence applied to variables—still in a sequential mode of thinking. **Function were nonexistent**—only variables were considered. This of course had a positive and a negative effect—on the positive side one could consider arbitrary equations such as the **folium**, with equation \( x^3 + y^3 = 6xy \), which was originally a challenge to the previous generation (that of Fermat & Descartes) for finding its tangents.

Thus, the variable \( x \) would be considered as taking a sequence of values, and its difference would be considered a new variable called \( dx \). As the sequence for \( x \) would get more crowded, the variable \( dx \) would become **infinitesimal** in nature. Leibniz developed much of our modern notation for the calculus such as \( dx \), \( \frac{dy}{dx} \) and \( \int \) and it is this notation (as opposed to Newton’s fluxion notation) that will be adopted in the rest of Europe.

Similarly, **integrals** would occur as the sum of the sequence of \( x \), and would be referred to as \( \int x \). Of course these would become infinitely large, but since most of the time the integral would be applied to infinitesimal quantities, they would become just real quantities.

His basic belief was that there were levels of magnitude:

| \( x, y \) were variables which were at the real level; | level 0 |
| \( dx, dy \) were differentials which were infinitesimals; | level \(-1\) |
| \( \int x, \int y \) were integrals which were very large; | level \(1\) |
| \( d^2x = d(dx), d^2y \) were second differentials (lower level). | level \(-2\) |
| \( \int dx, \int dy \) were real quantities again | up from \(-1\) to 0 |

Thus the differential lowered the level while the integral raised the level—and hence even today we use \( \int dx \) to denote a real quantity, and every integral has to have a differential inside to make sure the levels are correct.

His key belief was that quantities at different levels were not comparable. For example \( x + dy \approx x \) since \( x \) is at a higher level than \( dy \), and \( dx + d^2y \approx dx \) for the same reason.

Also multiplication would have an effect on the level: two real quantities multiplied would
remain real—think of $0 + 0 = 0$, but a real quantity multiplied by a differential would be a differential: $0 + (-1) = -1$ while the product of two differentials would be a second differential: $-1 + (-1) = -2$.

To compute the differential of a variable was simple: $dx$ was the differential of $x$, $dy$ was the differential of $y$, etcetera. Now if $z = x + y$ then it was essential that $dz = dx + dy$.

Naturally this is very true for sequences.

The following is still known today as **Leibniz’s Rule**: $d(xy) = xdy + ydx$.

His technique for computing differentials consisted of substituting every variable by itself added to its differential, and then subtracting the original expression and isolating the level $-1$ quantity:

$$d(xy) = (x + dx)(y + dy) - xy = xy + xdy + ydx + dx(dy) - xy = xdy + ydx + dx(dy)$$

and the end result would be (**Leibniz’ Rule**) $d(xy) = xdy + ydx$ since $dx(dy)$ is at level $-2$ and hence of no consequence to a differential. The notion of limit would have to wait another century before it was developed as an essential ingredient to differential calculus, and the word derivative was nowhere to be found.

Algebra was the instrument of choice. For example, suppose we wanted the quotient rule: $dz$ where $z = \frac{x}{y}$. Then a way to proceed was to consider $zy = x$ and take differentials obtaining $zdy + ydz = dx$. Isolating $dz$ we get

$$dz = \frac{dx - zdy}{y} = \frac{dx - \frac{x}{y}dy}{y} = \frac{ydx - xdy}{y^2}.$$  

The use of infinitesimals is not considered acceptable today (although is being reevaluated presently), yet it is quite effective and Euler used it effectively and unworriedly.

Considered even with less regard is a method used by Euler and others involving infinite quantities.

For example in 1734, Euler would argue as follows. Given that

$$\sum_{n=1}^{N} \frac{1}{n} = \ln N + C_N ,$$

and that for all infinitely large $N$, $C_N = C \approx 0.577...$, where $C$ is the so called Euler constant, (the modern day usage is $\gamma$ to denote this number) then one can argue that
\[
\sum_{n=1}^{2N} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots + \frac{1}{2N-1} - \frac{1}{2N} \\
= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots + \frac{1}{2N-1} + \frac{1}{2N} - 2 \left( \frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \ldots + \frac{1}{2N} \right) \\
= \sum_{n=1}^{2N} \frac{1}{n} - \sum_{n=1}^{N} \frac{1}{n} = \ln(2N) + \ln(2) - \ln N \approx \ln \left( \frac{2N}{N} \right) = \ln 2.
\]

By a similar argument, Euler would obtain,

\[1 + \frac{1}{2} - \frac{2}{3} + \frac{1}{4} + \frac{1}{5} - \frac{2}{6} + \ldots = \ln 3\]

and the two tables below show he was perhaps justified in his calculations:

<table>
<thead>
<tr>
<th>(n)</th>
<th>(\sqrt[n]{n})</th>
<th>Coefficient in (\ln 2)</th>
<th>Coefficient in (\ln 3)</th>
<th>Sum</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000000</td>
<td>1</td>
<td>1</td>
<td>1.00000</td>
<td>1.00000</td>
</tr>
<tr>
<td>2</td>
<td>0.500000</td>
<td>-1</td>
<td>1</td>
<td>0.50000</td>
<td>1.50000</td>
</tr>
<tr>
<td>3</td>
<td>0.333333</td>
<td>1</td>
<td>-2</td>
<td>0.83333</td>
<td>0.83333</td>
</tr>
<tr>
<td>4</td>
<td>0.250000</td>
<td>-1</td>
<td>1</td>
<td>0.58333</td>
<td>1.08333</td>
</tr>
<tr>
<td>5</td>
<td>0.200000</td>
<td>1</td>
<td>1</td>
<td>0.78333</td>
<td>1.28333</td>
</tr>
<tr>
<td>6</td>
<td>0.166667</td>
<td>-1</td>
<td>-2</td>
<td>0.61667</td>
<td>0.95000</td>
</tr>
<tr>
<td>7</td>
<td>0.142857</td>
<td>1</td>
<td>1</td>
<td>0.75952</td>
<td>1.09286</td>
</tr>
<tr>
<td>8</td>
<td>0.125000</td>
<td>-1</td>
<td>1</td>
<td>0.63452</td>
<td>1.21786</td>
</tr>
<tr>
<td>9</td>
<td>0.111111</td>
<td>1</td>
<td>-2</td>
<td>0.74563</td>
<td>0.99563</td>
</tr>
<tr>
<td>10</td>
<td>0.100000</td>
<td>-1</td>
<td>1</td>
<td>0.64563</td>
<td>1.09563</td>
</tr>
<tr>
<td>11</td>
<td>0.090909</td>
<td>1</td>
<td>1</td>
<td>0.73654</td>
<td>1.18654</td>
</tr>
<tr>
<td>12</td>
<td>0.083333</td>
<td>-1</td>
<td>-2</td>
<td>0.65321</td>
<td>1.01988</td>
</tr>
<tr>
<td>13</td>
<td>0.076923</td>
<td>1</td>
<td>1</td>
<td>0.73013</td>
<td>1.09680</td>
</tr>
<tr>
<td>14</td>
<td>0.071429</td>
<td>-1</td>
<td>1</td>
<td>0.65871</td>
<td>1.16823</td>
</tr>
<tr>
<td>15</td>
<td>0.066667</td>
<td>1</td>
<td>-2</td>
<td>0.72537</td>
<td>1.03490</td>
</tr>
<tr>
<td>16</td>
<td>0.062500</td>
<td>-1</td>
<td>1</td>
<td>0.66287</td>
<td>1.09740</td>
</tr>
<tr>
<td>17</td>
<td>0.058824</td>
<td>1</td>
<td>1</td>
<td>0.72170</td>
<td>1.15622</td>
</tr>
<tr>
<td>18</td>
<td>0.055556</td>
<td>-1</td>
<td>-2</td>
<td>0.66614</td>
<td>1.04511</td>
</tr>
<tr>
<td>19</td>
<td>0.052632</td>
<td>1</td>
<td>1</td>
<td>0.71877</td>
<td>1.09774</td>
</tr>
<tr>
<td>20</td>
<td>0.050000</td>
<td>-1</td>
<td>1</td>
<td>0.66877</td>
<td>1.14774</td>
</tr>
<tr>
<td>21</td>
<td>0.047619</td>
<td>-1</td>
<td>-2</td>
<td>0.71639</td>
<td>1.05250</td>
</tr>
<tr>
<td>22</td>
<td>0.045455</td>
<td>-1</td>
<td>1</td>
<td>0.67094</td>
<td>1.09796</td>
</tr>
<tr>
<td>23</td>
<td>0.043478</td>
<td>1</td>
<td>1</td>
<td>0.71441</td>
<td>1.14143</td>
</tr>
<tr>
<td>24</td>
<td>0.041667</td>
<td>-1</td>
<td>-2</td>
<td>0.67275</td>
<td>1.05810</td>
</tr>
<tr>
<td>25</td>
<td>0.040000</td>
<td>1</td>
<td>1</td>
<td>0.71275</td>
<td>1.09810</td>
</tr>
<tr>
<td>26</td>
<td>0.038462</td>
<td>-1</td>
<td>1</td>
<td>0.67429</td>
<td>1.13656</td>
</tr>
<tr>
<td>27</td>
<td>0.037037</td>
<td>1</td>
<td>-2</td>
<td>0.71322</td>
<td>1.06249</td>
</tr>
<tr>
<td>28</td>
<td>0.035714</td>
<td>-1</td>
<td>1</td>
<td>0.67561</td>
<td>1.09820</td>
</tr>
<tr>
<td>29</td>
<td>0.034483</td>
<td>1</td>
<td>1</td>
<td>0.71009</td>
<td>1.13269</td>
</tr>
<tr>
<td>30</td>
<td>0.033333</td>
<td>-1</td>
<td>-2</td>
<td>0.67676</td>
<td>1.06602</td>
</tr>
</tbody>
</table>
Abel (in the early part of the 19th century) would be one of the first to give an acceptable argument for this fact.

Leibniz accomplished a lot using infinitesimals—one of his key devices was the **infinitesimal triangle** where the hypotenuse was labeled $ds$ where $s$ stood for length. Until today we use the crucial fact that $(ds)^2 = (dx)^2 + (dy)^2$ and so $ds = \sqrt{(dx)^2 + (dy)^2}$ and naturally $\int ds = s$ in order to compute the lengths of curves.

From the same infinitesimal triangle, Leibniz realized that the slope of tangent lines was nothing but $\frac{dy}{dx}$ and naturally that the area under the curve was $\int ydx$.

He clearly stated that $\int dx = x$ and $d\int x = x$.

Thus, it is possible that Leibniz would not have been that impressed with the FTC as we know it. Instead we discuss a different idea that Leibniz found much more intriguing than $\int dx = x$. But first we do an example from the first course in calculus.

**Example 1.** Today we use a very useful technique called the **substitution method** with little geometrical thought of its implications.

On the picture on the left, the green line represents the graph of the function $y = e^x$, while the red line is the graph of $y = 2xe^{x^2}$.

If we were to ask for the area under the red graph (and above the $x$–axis), we would be asking for $\int_0^1 2xe^{x^2}dx$, and we would find this definite integral by using the substitution $u = x^2$, which would immediately leads us to the integral $\int_0^1 e^{u}du$ and the area under the green line, and although we can accept that the areas are the same, it is not readily discernible why it is so.

It was Leibniz who, in the 1670's, discovered this general principle or technique to evaluate areas. He referred to it by **transmutation of areas**. A technique basically equivalent to the **Fundamental Theorem of Calculus**, which we now give in some detail, and in, more or less, modern notation and ideology. It is important to observe that Leibniz is still rooted in Greek geometrical methods such as similarity of triangles, even if it is at an infinitesimal level.
Suppose we have an interval \([a, b]\) and we have a function \(y = f(x)\) defined on this interval. We are interested in the area under the curve of this function.

Start by considering two neighboring (very close) points on the graph of function, has coordinates \(x\) and \(y = f(x)\), and \(Q\) coordinates are \(x + dx\) and \(y + dy\), is a small change in \(x\), and \(dy\) is the corresponding small change in \(y\). We denote the origin.

Continuing in the language of indivisibles, we let the length of the curve from point \(P\) to point \(Q\) be denoted by \(ds\) (recall that \(s\) usually denotes length of a curve in Calculus).

Consider the tangent line to the curve at the point \(P\) and suppose it intersects the \(y\)-axis at \(T = (0, z)\).

Since the right triangle \(\Delta TUP\) is similar to the right triangle \(\Delta PRQ\), we have that

\[
\frac{y - z}{x} = \frac{dy}{dx},
\]

and solving for \(z\) we get,

\[
z = y - x \frac{dy}{dx}.
\]

We can use this expression to define a new function \(z\) of \(x\), whose graph is given by:

At the origin \(O\), draw the perpendicular to the tangent line \(TP\) and let it intersect this line at point \(S\), which has hypoteneuse \(z\), and let \(k\) be the distance from \(S\) to the origin.

Since
\( \angle STO + \angle PTU = 90^\circ \),

and \( \angle PTU = \angle QPR \),

we have \( \angle STO = \angle PQR \).

Hence right triangle \( \triangle OST \) is similar to right triangle \( \triangle PRQ \), and we have

\[
\frac{dx}{k} = \frac{ds}{z}.
\]

Consider now the infinitesimal triangle \( \triangle OPQ \):

Its base is \( ds \) and its height is \( k \), hence its area is \( \frac{kds}{2} \),

which by the similarity above equals \( \frac{zdx}{2} \).

But as the picture on the left illustrates, \( \frac{zdx}{2} \) equals half of the area in the indicated infinitesimal rectangle under the function \( z \).

Hence we have that the area under the graph of \( z \), which is given by

\[
\int_a^b zdx
\]

is twice the area of the shape made from all the infinitesimal triangles. Hence, in the picture on the right, the area on the left is half of the area on the right.
But by cutting and pasting in the picture below, we get that the area under $f(x)$ equals the area

$$\text{OCD} + \text{OBC} - \text{OAD}.$$ 

But by the result above, $\text{OCD}$ equals one half the area under $z$. Easily, $\text{OBC}$ equals $\frac{1}{2} b f(b)$ and $\text{OAD}$ is $\frac{1}{2} a f(a)$, so symbolically we have:

$$\int_a^b ydx = \frac{1}{2} \left( \int_a^b zdx + bf(b) - af(a) \right)$$

and we have exchanged one computation of areas for another that may turn out to be simpler than the original as we will exemplify below.

But before we do that let's make a couple of observations. First, $b f(b) - af(a)$ can be simplified using standard evaluation notation to $[xy]^b_a$. Second, we can use the fact that $z = y - x \frac{dy}{dx}$ to substitute in the equation above to obtain

$$\int_a^b ydx = \frac{1}{2} \left( \int_a^b zdx + bf(b) - af(a) \right)$$
\[
\int_a^b \left( y - x \frac{dy}{dx} \right) dx + \left[ xy \right]_a^b
\]

\[
= \frac{1}{2} \left( \int_a^b \frac{dy}{dx} dx - \int_a^b y dy + \frac{1}{2} \left[ xy \right]_a^b \right)
\]

which after clearing and multiplying by 2 yields the integration by parts formula due to Leibniz:

\[
\int_a^b y dy = \left[ xy \right]_a^b - \int_a^b x dy,
\]

which is tantamount—in the picture—to the three shaded areas filling in the rectangle—confirming the geometric reasoning he had employed in the more sophisticated equation 1 above.

Leibniz himself was pleased with the following application of his ideas.

Consider a circle of radius 1 centered at the point (1,0) so that it is tangent to the \( y \)-axis.

Then its upper semicircle has the equation \( y = \sqrt{2x - x^2} \).

Since \( y^2 = 2x - x^2 \), differentiating, we get

\[
2ydy = 2dx - 2xdx
\]

and so

\[
\frac{dy}{dx} = \frac{1 - x}{y},
\]

hence

\[
z = y - x \frac{1 - x}{y} = \frac{x}{\sqrt{2 - x}}.
\]

Solving for \( x \), we have

\[
x = \frac{2z^2}{1 + z^2}.
\]

But, then we have

\[
\frac{\pi}{4} = \int_0^1 y dx
\]

\[
= \frac{1}{2} \left( \int_0^1 z dx + \left[ x \sqrt{2x - x^2} \right]_0^1 \right) \text{ by 1 above}
\]

\[
= \frac{1}{2} \left( \int_0^1 z dx + \left[ x \sqrt{2x - x^2} \right]_0^1 \right) \text{ by 1 above}
\]
\[
\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots
\]

However, although one can find the expression

\[
\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots
\]

quite beautiful, it is not very practical for doing computation—it converges too slowly. For example, if we use the fact that any alternating series toggles between exceeding and being less than the true sum, we have the not very good estimate

\[
3.108269 < \pi < 3.173842 \text{ even after 30 terms.}
\]

Although if we averaged the two terms, we would obtain the superior 3.1410555.

After 100 terms we get the improved 3.131593 < \pi < 3.151493. And after 500, we have 3.139593 < \pi < 3.143589 Finally at 1,000 terms we get two-digit accuracy:

\[
3.140593 < \pi < 3.142592,
\]

with the average being 3.1415925, which is quite satisfying.

**Newton & the FTC**

One could say that Newton and Leibniz did understand thoroughly the fundamental theorem of calculus (as we call it today), and also both appreciated the power and range of the subject. Certainly, Newton used Calculus-type thinking to push the frontiers of mechanics and physics.

Born on Christmas Day, 1642, to a relatively poor widow, Isaac Newton showed promise as a student, and thus a brother of his mother agreed to support him in college. He attended Cambridge University. In 1665, during an outbreak of the plague, he was sent home, and it was during that period that he developed some of his best ideas. Soon after that, his teacher, Isaac Barrow resigned his position so that Newton could be appointed to follow
him. For the next 30 years Newton was a professor at Cambridge—alas, a terrible lecturer, hardly any one would attend his lectures, but a widely known scholar. In 1693, he suffered a nervous breakdown, partially caused by the stress suffered during the dispute with Leibniz. After he recovered, he was appointed in charge of the Royal Mint where he spent the remainder of his life. When he died, he was the most famous scientist in the world, and was buried with all the glory and ceremony at Westminster Abbey.

Sir Isaac Newton is one of the most distinguished names in the history of mathematics and science. He can be considered one of the founders of modern science, and his book *Philosophiae Naturalis Principia Mathematica* (1687) (often referred simply as the *Principia*) is a major book in Western civilization. He is still considered one of the most influential thinkers of all time, and Newton would be in the top three list of anyone’s choice of mathematicians (or physicists for that matter).

Being a physicist, he was interested in the notion of instantaneous velocity (which is closer to the limiting idea for a derivative), and it is in those explorations that he came with the idea of fluxion (his word for the derivative with respect to time). In fact, one of his famous law of motion, the second one, states that

\[ \text{Force} = \text{mass} \times \text{velocity} \]

where the dot stood for fluxion. In the more familiar Leibniz's notation the claim becomes:

\[ F = \frac{d(m \times v)}{dt}, \]

so if mass is a constant, we get the more common version of the second law: \( F = m \times a \) where \( a \) stands for acceleration. However, if mass is not constant, we get a different law—one that is valid at the very high speeds of atomic particles of modern physics.

A very easy application, when this law is linked with Galileo's conclusion that gravity is constant, is the calculation of the path followed by a projectile such as a cannon ball: *Suppose a projectile is shot from the origin at angle \( \theta \) with speed \( v_0 \). What is the path of the projectile?* If we separate the force into its two components, one in the x-direction and one in the y-direction (this idea is older than Newton), we get that \( F_x = 0 \) while \( F_y = -g \), a constant. But then \( a_x = 0 \) while \( a_y = \frac{-g}{m} \), a constant. But, by integrating accelerations, we get velocities, \( v_x = v_0 \cos(\theta) \) and \( v_y = \frac{-g}{m} t + v_0 \sin(\theta) \), if we assume time is measured so \( t = 0 \) is when we shot the projectile. Hence, \( s_x = v_0 \cos(\theta) t \) and \( s_y = \frac{-g}{2m} t^2 + v_0 \sin(\theta) t \), and if we graph this path, we get a parabola.

This kind of thinking allowed Newton to use mathematics to conclude physical facts based on a few premises. In a similar fashion
It was during that hiatus from school when he was in his early twenties that Newton may have independently arrived to his law of gravitation:

**two objects attract each other with a force proportional to their masses and inversely proportional to the square of their distance.**

Others had proposed it (Hooke for one). But he was definitely the first one to have used it mathematically to prove physical claims. He readily used mathematics (calculus ideas) to infer Kepler’s first two laws of planetary motion from the law of gravitation, and naturally this served as a major piece of evidence of support for this law.

But both of these results were published much later than they were realized, and in fact closer to Newton’s appreciation for calculus, and the power to calculate is his binomial theorem which we look at next. It was one of his first successes, and a definite step toward calculus. It is the extension of the traditional (or finite) binomial theorem to the case for other exponents besides positive integers.

**Newton Extends the Binomial Theorem**

What probably started as a technique to improve the computation of square roots, and other roots, became a broader weapon, and made him a superb manipulator of series—, which was critical to his whole view of calculus. We recall that

\[(1 + x)^n = \sum_{i=0}^{n} \binom{n}{i} x^i,\]

where \(n\) is an arbitrary positive integer.

Recall that \(\binom{n}{i} = \frac{n!}{i!(n-i)!}\) and if we cancel the \((n-i)!\) in the denominator, we arrive at

\(\binom{n}{i} = \frac{n \times n - 1 \times \cdots \times n - i + 1}{i \times i - 1 \times \cdots \times 1}\). Although the generalization was not so obvious in Newton’s original and more cumbersome notation, it is very clear in our present notation. Let \(\alpha\) be an arbitrary number now:

\[(1 + x)^\alpha = \sum_{i=0}^{\infty} \binom{\alpha}{i} x^i,\]

where of course \(\binom{\alpha}{i}\) is defined by

\[\frac{\alpha \times (\alpha - 1) \times (\alpha - 2) \times \cdots \times (\alpha - (i-1))}{i \times (i-1) \times (i-2) \times \cdots \times 1} .\]

Make the following observations:
- there are \(i\) factors in the numerator and \(i\) factors in the denominator.
- if \(n\) is a positive integer, then for \(i > n\), \(\binom{n}{i} = 0\), and thus Newton’s version
is a true extension of the finite case.

• Note the recursion, \( \binom{\alpha}{i+1} = \frac{\alpha - i}{i+1} \times \binom{\alpha}{i} \).

• \( \binom{\alpha}{i} \) is a polynomial in \( \alpha \) of degree \( i \)—with roots \( 0, 1, \ldots, i - 1 \).

We compute a few of these polynomials (using mainly the recursion given above).

\[
\begin{align*}
\binom{\alpha}{0} &= 1, \text{ by definition or agreement;} \\
\binom{\alpha}{1} &= \alpha; \\
\binom{\alpha}{2} &= \frac{\alpha \times (\alpha - 1)}{2!} = \frac{\alpha^2 - \alpha}{2}; \\
\binom{\alpha}{3} &= \frac{\alpha \times (\alpha - 1) \times (\alpha - 2)}{3!} = \frac{\alpha^3 - \frac{\alpha^2}{2} + \frac{\alpha}{3}}{3}; \\
\binom{\alpha}{4} &= \frac{\alpha - 3}{4} \times \binom{\alpha}{3} = \frac{\alpha^4}{24} - \frac{\alpha^3}{4} + \frac{11\alpha^2}{24} - \frac{\alpha}{4}; \\
\binom{\alpha}{5} &= \frac{\alpha - 4}{5} \times \binom{\alpha}{4} = \frac{\alpha^5}{120} - \frac{\alpha^4}{12} + \frac{7\alpha^3}{24} - \frac{5\alpha^2}{12} + \frac{\alpha}{5}; \\
\binom{\alpha}{6} &= \frac{\alpha^6}{720} - \frac{\alpha^5}{48} + \frac{17\alpha^4}{144} - \frac{5\alpha^3}{16} + \frac{137\alpha^2}{360} - \alpha; \\
\binom{\alpha}{7} &= \frac{\alpha^7}{5040} - \frac{\alpha^6}{240} + \frac{5\alpha^5}{144} - \frac{7\alpha^4}{48} + \frac{29\alpha^3}{90} - \frac{7\alpha^2}{20} + \frac{\alpha}{7}. 
\end{align*}
\]

So, for example, if we are interested in taking square roots, we let \( \alpha = \frac{1}{2} \) and we get the following coefficients,

\[
\begin{align*}
\frac{1}{0} &= 1, & \frac{1}{1} &= \frac{1}{2}, & \frac{1}{2} &= -\frac{1}{8}, & \frac{1}{3} &= \frac{1}{16} \\
\frac{1}{4} &= -\frac{5}{128}, & \frac{1}{5} &= \frac{7}{256}, & \frac{1}{6} &= -\frac{21}{1024}, & \frac{1}{7} &= \frac{33}{2048} 
\end{align*}
\]

So if we put them together with the binomial theorem, we get that

\[
\sqrt{1 + x} \approx 1 + \frac{x}{2} - \frac{x^2}{8} + \frac{5x^3}{128} + \frac{7x^4}{256} - \frac{21x^5}{1024} + \frac{33x^6}{2048} + \text{higher order terms},
\]

so if \( x \) is small, we should have a reasonable approximation.

For example, if we are interested in \( \sqrt{7} \), then we can handle it this way,

\[
\sqrt{7} = \sqrt{9 - 2} = 3\sqrt{1 - \frac{2}{9}}, \text{ so we let } x = -\frac{2}{9}, \text{ and we get } \frac{11248487}{12754584} \text{ for } \sqrt{1 - \frac{2}{9}}, \text{ which}
\]
approximates to 0.8819171993, which when multiplied by 3, gives 2.645751598, a good estimate for $\sqrt{7}$.

Sometimes, a closed expression for the coefficients is desired (and can be found), although it may be difficult to find the pattern at first. Let us revisit the coefficients we just have computed: $1, \frac{1}{2}, -\frac{1}{8}, \frac{1}{16}, -\frac{5}{128}, \frac{7}{256}, -\frac{21}{1024}, \frac{33}{2048}$ and at first the pattern does elude us. But let us go back to the definition:

$$\binom{1}{i} = \frac{\frac{1}{2} \times \frac{1}{2} - 1 \times \frac{1}{2} - 2 \times \cdots \times \frac{1}{2} - (i - 1)}{i \times (i - 1) \times (i - 2) \times \cdots \times 1} = \frac{\frac{1}{2} \times \frac{-1}{2} \times \frac{-3}{2} \times \frac{-5}{2} \times \cdots \times \frac{-(2i - 3)}{2}}{i!} = \frac{(-1)^{i-1} 1 \times 3 \times 5 \times \cdots \times (2i - 3)}{2^{i} i!}.$$  

Can we simplify this further? Perhaps simplify is the wrong word, and what we are trying to do is reduce the expression to more familiar functions. If we could reduce everything to the factorial, we would be at peace. We need to understand then the product of consecutive odds:

$$1 \times 3 \times 5 \times \cdots \times (2i - 3) = \frac{(2i - 3)!}{2 \times 4 \times 6 \times \cdots \times (2i - 4)} = \frac{(2i - 3)!}{(2 \times 1) \times (2 \times 2) \times (2 \times 3) \times \cdots \times (2 \times (i - 2))} = \frac{(2i - 3)!}{2^{i-2} (i - 2)!},$$

and so we conclude that

$$\binom{1}{i} = \frac{(-1)^{i-1} (2i - 3)!}{2^{2i-2} (i - 2)! i!}$$

for any $i$. Now whether this expression is computationally acceptable depends very much on our control of the factorial. But we do have what is called a closed expression.

**Newton Solves Equations**

Another early mathematical contribution using calculus-type reasoning is his useful way to finding roots of functions. The pursuit of finding solutions to equations is an ancient part of our subject, and methods have been developed through the ages. To wit, the Mesopotamians had a process for approximating square roots that we now revisit.

Let $n$ be a positive number. The idea behind the computation of its square root is simple. Take any guess, let us call it $a$. Then a companion guess is $\frac{n}{a}$ since $a \times \frac{n}{a} = n$, and that is the property we are looking for. So any time we have a guess, we really have two guesses. When will we have succeeded? When our two guesses are close to one another, for then we are indeed close to a square root. What should we do with the two guesses then? What is more reasonable than to average them? Indeed, that is what we do. This is an iteration, a never-ending process. This idea did not particularly seem troublesome to the Babylonians,
but to their successors, the Greeks, with their more rigorous requirements, infinite processes indeed seemed untidy. Returning to the algorithm, more precisely, given \( n \) take \( a_0 = a \) to be any (positive) initial guess for its square root. As mentioned above, we are going to keep refining this guess by defining a sequence that actually does converge to \( \sqrt{n} \).

Define \( a_{k+1} = \frac{1}{2} \left( a_k + \frac{n}{a_k} \right) \).

For example, let \( n = 257,941 \), \( a = a_0 = 1,000 \). An initial guess that is very bad will just prolong the algorithm, but not deter it. The following table illustrates how quickly we will have quite a few decimal places correctly computed. Remember, what we want is the two guesses to be close to each other. What we see in these numbers of course depends on who we are. A Mesopotamian would perhaps accept this answer as finished, while some Greeks would perhaps not accept it as ever finished.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( a_0 )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( a_3 )</th>
<th>( a_4 )</th>
<th>( a_5 )</th>
<th>( a_6 )</th>
<th>( a_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>257,941</td>
<td>1000</td>
<td>628.9705</td>
<td>519.5354074</td>
<td>508.0096871</td>
<td>507.8789394</td>
<td>507.8789226</td>
<td>507.8789226</td>
<td>507.8789226</td>
</tr>
</tbody>
</table>

By Newton’s time, polynomials of arbitrary degree and infinite series have become acceptable. Two techniques for solving such equations were already widely used. These stem from at least the time of Islam if not earlier. They are the bisection method and the method of false position—from which another variant came about later called the secant method.

To find a solution to \( p(x) = 0 \) both methods start with an \( a \) and a \( b \) given where \( p(a) \) and \( p(b) \) disagree in sign, one is positive, one is negative. Hence one expects a root between \( a \) and \( b \). Both procedures shrink the interval where the root is located. The bisection method simply goes to the midpoint between \( a \) and \( b \), \( \frac{a+b}{2} \), and decides by checking the sign at the midpoint in which of the two halves the root is located in. We will illustrate this technique first (the example is due to Newton).

We are solving \( x^3 - 2x - 5 = 0 \)

We start with \( a = 2 \) and \( b = 3 \) as our original guess, and then go from there. The table is fairly self-explanatory.
The method of false position is more complicated in that it uses the secant to the graph given by the two points \( a \) and \( b \). In more detail, it takes as its new guess the intersection of the \( x \)-axis with the line going through the points \( (a, p(a)) \) and \( (b, p(b)) \). Thus, for the method of false position method, the new point is given by the expression:

\[
c = \frac{bp(a) - ap(b)}{p(b) - p(a)}.
\]

Obviously, the computation of the new guess is harder, but as we can see, at least in this case, the method of false position works considerably faster than the bisection method.

Tackling the same cubic, we get

<table>
<thead>
<tr>
<th>Iteration</th>
<th>New</th>
<th>New Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( 2 )</td>
<td>2.058823529 -0.390799919</td>
</tr>
<tr>
<td></td>
<td>( 3 )</td>
<td>2.08126366 -0.147204060</td>
</tr>
<tr>
<td></td>
<td>( 3 )</td>
<td>2.092739574 -0.02020866</td>
</tr>
<tr>
<td></td>
<td>( 3 )</td>
<td>2.093883708 -0.00745673</td>
</tr>
<tr>
<td></td>
<td>( 3 )</td>
<td>2.094305451 -0.001011574</td>
</tr>
<tr>
<td></td>
<td>( 3 )</td>
<td>2.094460846 -0.000372653</td>
</tr>
</tbody>
</table>

Newton took this one step further because he could compute tangents efficiently—and one could refer to Newton's Method as the Tangent Method. The reason for this name is because it finds its new guess by following the tangent line to the function at a guess.

Start with one guess \( x_0 \) (as opposed to two guesses necessary in the other methods), and then follow one's nose by using the tangent line at the point of the original guess. How do we simply find the new guess, \( x_1 \)?

We follow the tangent line at the point of the graph corresponding to our initial guess:

Geometrically then, we have an idea of where to locate our new approximation. But how do we find the point efficiently?

Easily—use the slope:

\[
slope = f'(x_0)
\]

and solving for \( x_1 \), we get
Once we get $x_1$, we can use the same expression to get $x_2$, and then $x_3$, etcetera, continuing the iteration. When do we stop? **If there is not much change from one $x$ to the next $x$, most probably we are close to a root, and we can stop.**

When we apply it to the cubic we looked at before we get the following very short table.

| Newton's |       |
|----------|--|---|
| $x$      | $f(x)$  |
| 2        | 1     |
| 2.1      | 0.061 |
| 2.094568121 | 0.000185723 |
| 2.094551482 | 0.0000000017 |

But we also should remark that depending on the nature of the equation, the method can be unstable, and not lead to a root at all—the curious reader may attempt to find a root of $x^3$ by Newton's method. Nevertheless, the method is so useful it is programmed in most hand-held calculators.

Needless to say, the method just described has been polished through time, and we now spend some time describing what Newton originally did. He used one of the original ideas behind calculus. That idea is **ignoring terms of higher order than 1**, in other words, ignoring everything except linear terms—actually, that is what approximating a curve by the tangent line is all about. Newton was an expert at that technique.

We illustrate, analytically, his original thinking behind his method with the same example, the cubic $x^3 - 2x - 5 = 0$. We start with one approximation, 2, the same as above, and so we let $x = 2 + p$, and obtain $p^3 + 6p^2 + 10p - 1 = 0$—remember $x$ is supposed to be a root. Ignoring all but the linear term, we get $10p - 1 = 0$, $p = 0.1$ and we have a better approximation $x = 2.1$. Again, we let $x = 2.1 + p$, and substituting, we get $p^3 + 6.3p^2 + 11.23p + 0.61 = 0$, and thus, by again considering only the linear term, we have that $p = -\frac{61}{11230} \approx -0.00543187$, which quickly gives us an estimate for $x = 2.1 - 0.00543187 = 2.09456813$—just as before.

Newton would use this technique in indefatigable and unrelenting ways.

*Newton finds the Series for the Sine & the Cosine*
Most of us take Taylor's series expansions very much for granted. Yet this most wonderful and powerful theorem could not have even been thought of without enough examples to hint at its existence.

Before he could find the series for the sine function, Newton computed the series for the arcsine. Consider the circle \( x^2 + y^2 = 1 \) on the first quadrant. And take an arbitrary point \( P \) on the \( x \)-axis, at distance \( x \) from the origin \( O \). Then we know the height of the circle at that point is \( \frac{1}{\sqrt{1-x^2}} \), which by his binomial theorem Newton could expand into:

\[
\sqrt{1-x^2} = 1 - \frac{x^2}{2} - \frac{x^4}{8} - \frac{x^6}{16} - \frac{5x^8}{128} - \frac{7x^{10}}{256} - \frac{21x^{12}}{1024} - \frac{33x^{14}}{2048} + \cdots.
\]

He could then integrate this series to obtain the total shaded area \( \triangle OPQR \), consisting of the triangle \( \triangle OPQ \) and the section of the circle \( OQR \):

\[
\Delta + \angle = x - \frac{x^3}{6} - \frac{x^5}{40} - \frac{x^7}{112} - \frac{5x^9}{1152} - \frac{7x^{11}}{2816} - \frac{21x^{13}}{13312} - \frac{33x^{15}}{30720} + \cdots.
\]

But the area of the triangle \( \triangle OPQ \) is known,

\[
\Delta = \frac{x\sqrt{1-x^2}}{2} = x - \frac{x^3}{2} - \frac{x^5}{4} - \frac{x^7}{16} - \frac{5x^9}{32} - \frac{7x^{11}}{128} - \frac{21x^{13}}{512} - \frac{33x^{15}}{2048} + \cdots.
\]

And thus the section of the circle \( \angle QOR \) is given by the difference of the two series:

\[
\angle = \frac{x^3}{2} - \frac{3x^5}{80} + \frac{5x^7}{224} + \frac{35x^9}{2304} + \frac{63x^{11}}{5632} + \frac{231x^{13}}{26624} + \frac{143x^{15}}{20480} + \cdots.
\]

But \( \angle = \frac{\theta}{2} \), and \( x = \cos\left(\frac{\pi}{2} - \theta\right) = \sin(\theta) \), and so \( \theta = \arcsin(x) \), and we have that

\[
\theta = \arcsin(x) = x + \frac{x^3}{6} + \frac{3x^5}{40} + \frac{5x^7}{112} + \frac{35x^9}{1152} + \frac{63x^{11}}{2816} + \frac{231x^{13}}{13312} + \frac{143x^{15}}{10240} + \cdots.
\]

One of the best early examples is provided by the series expansion of the sine (and also the cosine)—which Newton developed. Start with the series of the Arcsine that we have above:

\[
\theta = \arcsin(x) = x + \frac{x^3}{6} + \frac{3x^5}{40} + \frac{5x^7}{112} + \frac{35x^9}{1152} + \frac{63x^{11}}{2816} + \frac{231x^{13}}{13312} + \frac{143x^{15}}{10240} + \cdots.
\]

The idea is to solve for \( x \) in terms of \( \theta \), \( x = \sin(\theta) \). At all times, we will eliminate all nonlinear terms, continuing with the basic idea behind his method.

\[
x + \frac{x^3}{6} + \frac{3x^5}{40} + \frac{5x^7}{112} + \frac{35x^9}{1152} + \frac{63x^{11}}{2816} + \frac{231x^{13}}{13312} + \frac{143x^{15}}{10240} + \cdots = \theta = 0.
\]
Dropping all nonlinear terms, we have as a first approximation, 
\( x \approx \theta \), which as the picture indicates fits the sine function for small \( x \)'s.

We are going to illustrate the method by using the first four terms of the series for the Arcsine:
\[
 x + \frac{x^3}{6} + \frac{3x^5}{40} + \frac{5x^7}{112} - \theta = 0.
\]

We let now \( x = \theta + p \), and we get:
\[
 (\theta + p) + \frac{(\theta + p)^3}{6} + \frac{3(\theta + p)^5}{40} + \frac{5(\theta + p)^7}{112} - \theta = 0,
\]

Thus,
\[
 \frac{2800\theta^3 + 1260\theta^5 + 750\theta^7}{1680} + \left( \frac{16 + 8\theta^2 + 6\theta^4 + 5\theta^6}{16} \right) p + \text{higher order terms} = 0.
\]

Ignoring the higher order terms and solving for \( p \), we get that:
\[
 p = -\frac{2800\theta^3 + 1260\theta^5 + 750\theta^7}{1680 + 16\theta^2 + 6\theta^4 + 5\theta^6} = -\frac{\theta^3}{6} + \text{higher order terms},
\]

and so
\[
 x \approx \theta - \frac{\theta^3}{6}.
\]

Now we let \( x = \theta - \frac{\theta^3}{6} + p \), and substitute in \( \circledast \), and we obtain:
\[
 \left( \theta - \frac{\theta^3}{6} + p \right) + \frac{\left( \theta - \frac{\theta^3}{6} + p \right)^3}{6} + \frac{3\left( \theta - \frac{\theta^3}{6} + p \right)^5}{40} + \frac{5\left( \theta - \frac{\theta^3}{6} + p \right)^7}{112} - \theta = 0.
\]

And so when we expand,
\[
 \frac{1,306,360\theta^7 + \text{higher order terms in } \theta^1}{156,764,160} + \left( \frac{756,496 + \text{higher order terms in } \theta}{746,496} \right) p
\]

\[
 + \text{higher order terms in } p = 0.
\]

Simplifying, we have

\[\text{In fact, the higher order terms in } \theta \text{ are } 622,080\theta^7 + 5,019,840\theta^9 - 3,538,080\theta^{11} + 1,088,640\theta^{13} - 187,488\theta^{15} + 18,900\theta^{17} - 1,050\theta^{19} + 250\theta^{21} \text{. Similarly, for the other fraction.}\]
\[
\frac{\theta^3 + \text{higher order terms in } \theta}{120} + (1 + \text{higher order terms in } \theta)p + \cdots = 0.
\]

When solving for \( p \) and ignoring anything but the lowest term order we get \( p = \frac{\theta^3}{120} \), and so we have

\[
x \approx \theta - \frac{\theta^5}{6} + \frac{\theta^7}{120}.
\]

Continuing in this way, he eventually predicted the correct pattern:

\[
\sin(\theta) = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \frac{\theta^9}{9!} - \frac{\theta^{11}}{11!} + \cdots.
\]

From this series he used the fact that

\[
\cos(\theta) = \sqrt{1 - \sin^2(\theta)},
\]

and substituted the series for the sine into the series for \( \sqrt{1 - x^2} \)—an absolutely formidable calculation.

Newton did this in order to establish

\[
\cos(\theta) = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \frac{\theta^8}{8!} - \frac{\theta^{10}}{10!} + \cdots.
\]

Did he know that the derivative of the sine function is the cosine? Perhaps not, since he would most probably have used the derivative to find the series for the cosine instead. To be strictly accurate historically, we have given a modern representation of Newton’s actual calculations. At that time it was more customary to view the sine as corresponding to an arc more than an angle—so in fact radians were not needed.
Next we give the pictures for the next eight approximations of the sine series. And we see a steady improvement in our approximations.

As we end this chapter, we need to mention that the foundations of calculus were problematic in both sides of the English Channel. But, ironically, it was the arrogance of Halley (of Halley’s Comet and a friend of Newton’s) that would prompt Bishop Berkeley to utter these words to discredit calculus:

He who can digest a second or third fluxion, a second or third difference, need not, methinks, be squeamish about any point in Divinity. And what are these fluxions? The velocities of evanescent increments. And what are these evanescent increments? They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them ghosts of departed quantities?

These criticisms would resonate with the mathematical community for a long time—yet they proceeded to use calculus in powerful and surprising ways.