



Candidates must complete this page and then give this cover and their final version of the extended essay to their supervisor.

Candidate session number

Candidate name

School number

School name

Examination session (May or November)

M

Year

2012

Diploma Programme subject in which this extended essay is registered: Mathematics

(For an extended essay in the area of languages, state the language and whether it is group 1 or group 2.)

Title of the extended essay:

The Evolution of Isaac
Newton's Numerical Method

Candidate's declaration

This declaration must be signed by the candidate; otherwise a grade may not be issued.

The extended essay I am submitting is my own work (apart from guidance allowed by the International Baccalaureate).

I have acknowledged each use of the words, graphics or ideas of another person, whether written, oral or visual.

I am aware that the word limit for all extended essays is 4000 words and that examiners are not required to read beyond this limit.

This is the final version of my extended essay.

Candidate's signature

Date:

01/12/2011

Supervisor's report and declaration

The supervisor must complete this report, sign the declaration and then give the final version of the extended essay, with this cover attached, to the Diploma Programme coordinator.

Name of supervisor (CAPITAL letters) _____

Please comment, as appropriate, on the candidate's performance, the context in which the candidate undertook the research for the extended essay, any difficulties encountered and how these were overcome (see page 13 of the extended essay guide). The concluding interview (viva voce) may provide useful information. These comments can help the examiner award a level for criterion K (holistic judgment). Do not comment on any adverse personal circumstances that may have affected the candidate. If the amount of time spent with the candidate was zero, you must explain this, in particular how it was then possible to authenticate the essay as the candidate's own work. You may attach an additional sheet if there is insufficient space here.

The candidate spent a good deal of time selecting a research question. The background reading served him well and once he had a suitable topic area, things fell into place easily.

The candidate proved to be an exemplary student meeting all deadlines and providing an insight into a historic and significant area of mathematics.

This declaration must be signed by the supervisor; otherwise a grade may not be issued.

I have read the final version of the extended essay that will be submitted to the examiner.

To the best of my knowledge, the extended essay is the authentic work of the candidate.

I spent hours with the candidate discussing the progress of the extended essay.

Supervisor's signature: _____

Date: 28/2/12

Assessment form (for examiner use only)

Candidate session number

Achievement level

Criteria	Examiner 1	maximum	Examiner 2	maximum	Examiner 3
A research question	2 ✓	2		2	
B introduction	2 ✓	2		2	
C investigation	4 ✓	4		4	
D knowledge and understanding	3 ✓	4		4	
E reasoned argument	1 ✓	4		4	
F analysis and evaluation	2 ✓	4		4	
G use of subject language	4 ✓	4		4	
H conclusion	2 ✓	2		2	
I formal presentation	4 ✓	4		4	
J abstract	2 ✓	2		2	
K holistic judgment	2 ✓	4		4	
Total out of 36	28 ✓				

Name of examiner 1: _____ Examiner number: _____
(CAPITAL letters)

Name of examiner 2: _____ Examiner number: _____
(CAPITAL letters)

Name of examiner 3: _____ Examiner number: _____
(CAPITAL letters)

IB Cardiff use only: B: _____

IB Cardiff use only: A: 99018 Date: 11-2

The Evolution
of
Isaac Newton's Numerical Method

Extended Essay

Mathematics

International Baccalaureate Program

Candidate:

Session: May 2012

Essay Supervisor:

December 2011

The
Evolution
of
Isaac Newton's Numerical Method

Extended Essay
Mathematics

Abstract

When Isaac Newton's "*De Analysi...*" was finished in 1669, it detailed a method for approximating solutions to polynomial equations. Today, the root-approximating algorithm referred to as the "Newton Method" differs significantly from what Newton described in "*De Analysi...*". In Newton's time it was a purely algebraic method for approximating roots of polynomials. Today it can approximate the roots for any type of function, and its rendition as the "Multivariate Newton Method" can be used to approximate the solution of an $n \times n$ system of multivariate, non-linear equations.

In Part I, this essay investigates the evolution of the single-variate formulation of the Method from 1669 to the modern day. It investigates **how, and thanks to whom, did Isaac Newton's Numerical Method arrive at its current formulation?** To arrive at an answer to the question, it focuses on Newton and his successors, Joseph Raphson and Thomas Simpson, during the 17th and 18th century. The investigative approach taken consisted of matching key stages in the Method's mathematical development to the work of each of these men. It analyses the mathematical importance and limitations of these key stages, and how each led to the next.

In a preliminary conclusion, it demonstrates that Newton simply conceptualized the Method; it was Joseph Raphson who then introduced direct iteration and later Thomas Simpson who linked it to calculus.

In Part II, the essay analyses Simpson's further development of the Method, proceeding to conclude that Simpson then set up the algebra from which the multi-variate formulation of the Method can be derived. This was something that Simpson himself was not able to do for at the time matrix algebra, essential to this 20th Century rendition of this method, had not yet been invented.

Words: 279

Contents

Abstract	ii
1 Introduction	1
I Single-Variate Formulation	2
2 The Newton Method Today	2
3 Isaac Newton's Work	5
4 Joseph Raphson's Work	10
5 Thomas Simpson's Work	13
6 Preliminary Conclusion	16
II Multi-Variate Formulation	17
7 Simpson's Breakthrough	17
8 The Multivariate Newton Method	21
9 Conclusion	24
10 Bibliography	25
A Appendix	26

1 Introduction

Newton's Numerical Method can be referred to as an iterative algorithm that employs differential calculus to arrive at successively closer approximations to the root of a function. Despite being named after Isaac Newton, a brief glance through Newton's "*De analysi per aequationes numero terminorum infinitas*"¹ (The analysis of Equations of Infinite Terms), where Newton first demonstrated his Method, will reveal that the Method he detailed was very different from what we today know as the Newton Method. This indicates that ~~the~~ it must have undergone substantial change since Newton's work in "*De Analysi...*".

This essay aims to clarify the road that the Method took from the 17th Century to the modern day. A road that is unknown even to many that use the Method every day. Understanding the Method's evolution can provide an understanding of its roots and the key stages which contributed to its development. Furthermore, this understanding can clarify who deserves what credit for the method's effectiveness today. In short; **how, and thanks to whom, did Isaac Newton's Numerical Method arrive at its current formulation?** To answer this question it is crucial to understand the Method's contemporary formulation and compare it to what Newton detailed in his "*De Analysi...*". Once this contrast is established, it is then possible to fill in the gaps and trace how the work of Newton's successors contributed to the Method's evolution.

¹Newton, 1669 (p. 218-223)

Part I

Single-Variate Formulation

2 The Newton Method Today

Definition 1 If x_k is a close approximation to the root of a real function $f(x)$, then a closer approximation x_{k+1} can be obtained by:

$$\boxed{x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}} \quad (2.1)$$

This is the most common formulation of Newton's Numerical Method. The reasoning behind 2.1 can be derived from Taylor's Theorem. Given a function $f(x)$, the first order Taylor series approximation at a point x_k is a linear equation:

$$f(x) \simeq f(x_k) + f'(x_k)(x - x_k)$$

If the initial estimate x_k is accurate enough, the root of this linear approximation is close to the root x of the function (assuming convergence). Let us assume that the solution of this linear is in fact the root of $f(x)$. Now solving for the x :

$$\begin{aligned} f(x_k) + f'(x_k)(x - x_k) &= 0 \\ f'(x_k)(x - x_k) &= -f(x_k) \\ x - x_k &= -\frac{f(x_k)}{f'(x_k)} \\ x &= x_k - \frac{f(x_k)}{f'(x_k)} \end{aligned}$$

For successive iterations, the argument x is essentially x_{k+1} , making this result effectively equivalent to 2.1.

continuously differentiable

equation? which?

Not convincing
What does this show?

Nothing proved here

Example 1 Consider $y = x^2 + 3x - 1$

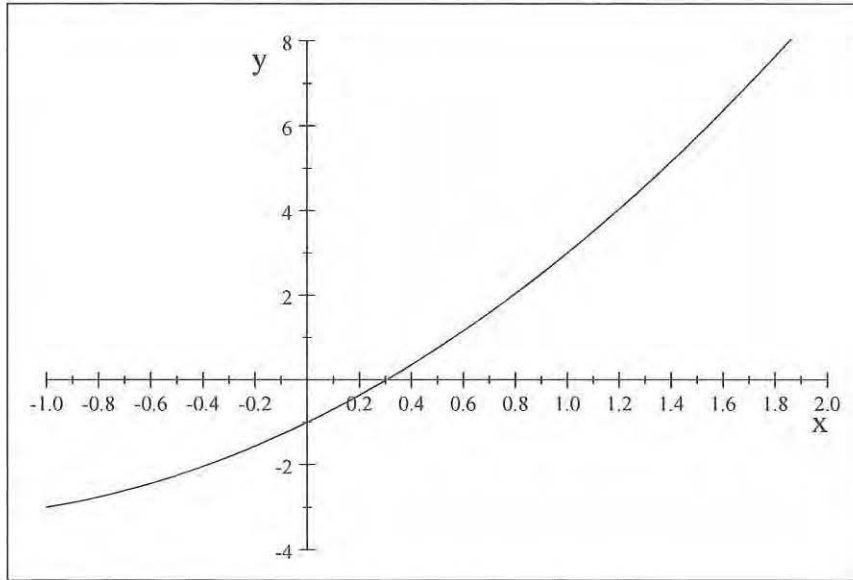


Illustration of $y = x^2 + 3x - 1$.

taking the first order Taylor series approximation at $x_k = 0.4$

$$\begin{aligned} f(x) &\simeq f(x_k) + f'(x_k)(x - x_k) \\ &\simeq f(0.4) + f'(0.4)(x - 0.4) \end{aligned}$$

equating to zero

$$\begin{aligned} f(0.4) + f'(0.4)(x - 0.4) &= 0 \\ f'(0.4)(x - 0.4) &= -f(0.4) \\ x - 0.4 &= -\frac{f(0.4)}{f'(0.4)} \\ x &= 0.4 - \frac{f(0.4)}{f'(0.4)} \\ &= 0.4 - \frac{(0.4)^2 + 3(0.4) - 1}{2(0.4) + 3} \\ &= 0.305\,263\,157\,9 \end{aligned}$$

*An example
is not a proof!*

giving a value closer to the real root of the equation. This can also be illustrated by drawing a tangent on the graph of $y = x^2 + 3x - 1$ at the estimated

root using the first order Taylor series approximation above:

$$\begin{aligned} f(x) &\simeq f(0.4) + f'(0.4)(x - 0.4) \\ &\simeq (0.4)^2 + 3(0.4) - 1 + (2(0.4) + 3)(x - 0.4) \\ y &\simeq 3.8x - 1.16 \end{aligned}$$

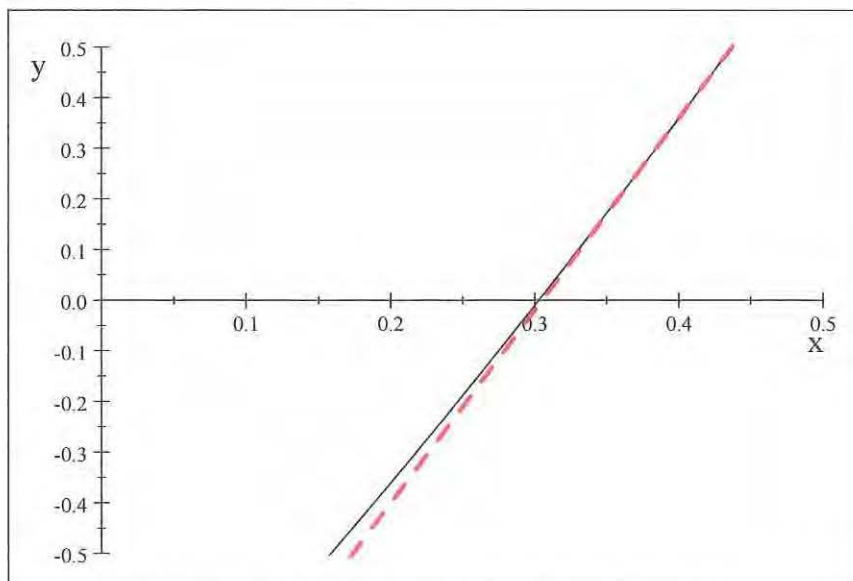


Illustration of $y = x^2 + 3x - 1$ (solid) and its Taylor series approximation at 0.4 (dashed) $y = 3.8x - 1.16$.

Visually, the new estimate for a root 0.305 is in fact much closer than the initial estimate 0.4. This can be repeated successively for closer approximations.

The "modern" Newton method has not been proved correct here (so far) //

However, the reasoning featured above and the formulation of 2.1 did not feature in any of Newton's works. Tracking its evolution can determine how the method went from being a purely algebraic method during Newton's time, to today being used recursively and incorporating calculus in order to approximate roots of an equation. Previous papers have been written on the subject of the evolution of Newton's Method, often either advocating skepticism of Newton's work (N. Kollerstrom) or arguing his superior contribution compared to his successors (T. Ypma). To approach this investigation it is important to take into account the conclusions of previous papers written about the Method. However, the goal is to ultimately draw a link between the work of Newton and the Method as we know it today, and analyzing that link such as to reach an objective conclusion.

3 Isaac Newton's Work

Conceptualization

The earliest printed account of Isaac Newton's (1642 - 1727) work on The Method is in John Wallis' "A Treatise..." of 1685.² Newton's own text developing the method ("De Analysi...") was only published later by William Jones.³ In spite of this, his method was known to various mathematicians at the time, mainly through circulations of copies of his manuscripts.

Newton took an algebraic approach to the problem of approximating roots to a function. He also credited the work of Viète as an inspiration for this method⁴:

Definition 2 For any $f(x)$ where X is a real root and z is a close approximation, $z + p$ is a closer approximation⁵. Solving for p and repeating for successively introduced variables will give consecutive "adjustments" to the initial estimate such that $z + p + q + r + \dots$ will result in an ever closer approximation to X .

Example 2 As first used by Newton to demonstrate the method (detailed in John Wallis's "A Treatise..."):

$$y = x^3 - 2x - 5 \tag{3.1}$$

By creating a sign table

x	$f(x)$	<i>Sign</i>
1	$1 - 2 - 5 = -6$	<i>Negative</i>
2	$8 - 4 - 5 = -1$	<i>Negative</i>
3	$27 - 6 - 5 = 16$	<i>Positive</i>
4	$64 - 8 - 5 = 51$	<i>Positive</i>

we can see a sign change (and hence a root) around $x = 2$, therefore $2+p = X$.

²Wallis, 1685

³Ypma, 1995 (p. 537)

⁴*Ibid.*, (p. 540)

⁵Note: In Part I of this essay, z denotes an initial estimate for the root, a constant.

Hence:

$$x^3 - 2x - 5 = 0$$

$$x = 2 + p$$

$$\Rightarrow (2 + p)^3 - 2(2 + p) - 5 = 0$$

$$p^3 + 6p^2 + 10p - 1 = 0 \quad (3.2)$$

The two terms p^3 and $6p^2$ can be ignored as p has such a small absolute value that when squared and cubed it becomes insignificant to the equation.

Hence:

$$10p - 1 \simeq 0$$

$$p \simeq 0.1$$

Newton then continued using the same logic, introducing a new variable at each iteration:

$$p^3 + 6p^2 + 10p - 1 = 0$$

$$p = 0.1 + q$$

$$\Rightarrow (0.1 + q)^3 + 6(0.1 + q)^2 + 10(0.1 + q) - 1 = 0 \quad (3.3)$$

$$q^3 + 6.3q^2 + 11.23q + .061 = 0$$

Again ignoring the cubic and quadratic terms:

$$11.23q + .061 \simeq 0$$

$$q \simeq -\frac{0.061}{11.23} = -0.0054$$

Iterating once more:

$$q^3 + 6.3q^2 + 11.23q + .061 = 0$$
$$q = -0.0054 + r \quad (3.4)$$

$$(3.5)$$

$$\implies (-0.0054 + r)^3 + 6.3(-0.0054 + r)^2 +$$
$$(3.6)$$

$$+ 11.23(-0.0054 + r) + .061 = 0$$

$$r^3 + 6.2838r^2 + 11.162r + 0.000541551 = 0 \quad (3.7)$$

$$11.162r + 0.000541551 \simeq 0$$

$$-0.00004852 \simeq r$$

The method can be continued for as many iterations as desired. However, it is evident that the method uses little more than polynomial manipulation and continuous summation to approximate the root. The final step is to sum all of the variables together to get the total offset from the initial estimation:

$$X \simeq z + p + q + r \quad (3.8)$$

$$= 2 + 0.1 - 0.0054 - 0.00004852$$

$$= 2.09455148 \quad (3.9)$$

$$\therefore X \simeq 2.09455148$$

This is correct to eight decimal places (for r , the last substituted variable, has eight decimal places). The rate of convergence of Newton's method is a rather complex theme, and outside the scope of this essay. However, let us go under the assumption that, despite the suggestions of Myron Pawley⁶, Maseres was right in saying that if "*one step of the Newton Method is right to n decimal places, then the next step will be right to $2n$* ".⁷ This suggests quadratic convergence (i.e. the number of correct decimal places doubles with each iteration).

Why?

Why?

One can now begin to understand that the method originally employed by Isaac Newton bears little resemblance with what is today known as the "Newton Method". Firstly, it is not **directly** iterative. There is no recursive

⁶Pawley, 1940 (p.113)

⁷*Ibid.* (p.114)

formula into which one reintroduces the approximation that was retrieved in the last iteration. Instead, a new expression must be formulated ([3.2], [3.3], [3.7]) in different variables every time we wish to iterate and then sum the value of these variables must be taken at the end [3.9]. Furthermore, there is absolutely no connection between the method and Newton's "Method of Fluxions", or today's differential calculus. These two main differences were highlighted by Nick Kollerstrom⁸ as the primary indicators that there were more important contributions to the method later on.

It is important to note that this method works with polynomials, but it is impossible to apply to non-polynomial equations without the implementation of other mathematical tools. Nonetheless, it was this formulation that was developed by Isaac Newton. It constituted the first step in this long journey of mathematical development. Despite its limitations, detailed below, it was still of some use at the time, and the start of the Method's development

Example 3

$$y = x^{1/3} + 3x - 5 \tag{3.10}$$

Again making a sign table

x	f(x)	Sign
1	$1 + 3 - 5 = -1$	Negative
2	$\sqrt[3]{2} + 6 - 5 = \sqrt[3]{2} + 1$	Positive

Root between 1 & 2, closer to 1:

$$\begin{aligned}
 1 + p &= x \\
 (1 + p)^{1/3} + 3(1 + p) - 5 &= 0
 \end{aligned}
 \tag{3.11}$$

At this point the first term of [3.11] can no longer be easily simplified without further techniques (such as a Binomial Expansion or Taylor Series approximation).

⁸Kollerstrom, 1992 (p. 347)

Example 4 Let us observe the method at work with a trigonometric function.
For Example, $y = \sin x$; Solution around $x = 3$:

$$y = \sin x \quad (3.12)$$

$$3 + p = x$$

$$\therefore y = \sin(3 + p)$$

$$0 = \sin(3 + p)$$

$$= \sin 3 \cos p + \sin p \cos 3$$

$$- \sin p \cos 3 = \sin 3 \sqrt{1 - \sin^2 p}$$

$$\sin^2 p \cos^2 3 = \sin^2 3 (1 - \sin^2 p)$$

$$\sin^2 p \cos^2 3 = \sin^2 3 - \sin^2 3 \sin^2 p$$

$$\sin^2 p \cos^2 3 + \sin^2 3 \sin^2 p = \sin^2 3$$

$$\sin^2 p (\cos^2 3 + \sin^2 3) = \sin^2 3$$

$$\sin^2 p = \sin^2 3$$

$$\sin p = \sin 3$$

$$p = 3 + k2\pi \text{ or } \pi - 3 + k2\pi \quad (3.13)$$

If we choose $p = 3 + k2\pi$, the method diverges instead of converging. Alternatively, while it is possible to find the root with $p = \pi - 3 + k2\pi$, we are having to refer to the periodicity of trigonometric functions (i.e. recognizing that p can equal $\pi - 3 + k2\pi$):

$$0 = \sin(3 + p) = \sin(3 + 2\pi - 3)$$

$$0 = \sin 2\pi$$

Ultimately we are given a solution in terms of π , not a numerical solution. Since the method's purpose is finding the numerical value of π , it **does not work**. This is an example of circular reasoning: an answer can only be found if it is already known. Additionally, it must be recognized that the steps are heavily algebraic and it takes a rather long time to set up each step (the whole of 3.12 – 3.13 being only one iteration). From this perspective, it is simply not practical, and would be more complex to program into a modern day computer than the Newton Method known today.

4 Joseph Raphson's Work

Direct Iteration

Direct Iteration was introduced into the method by Joseph Raphson (1648 - 1715). Although he published his work before Newton (1697), his work could have been inspired by Newton's unpublished work. Regardless, Raphson published his "*Analysis...*"⁹ with an almost insignificant reference to Newton, and without any definite credit of having based his method on his contemporary's work¹⁰. Raphson's method for finding the roots of equations bears its similarities to Newton's, but makes a very important breakthrough; it is directly iterative.

Proposition 1 *Suppose that for the function*

$$ax^3 + bx^2 + cx + d = 0 \quad (4.1)$$

There is a real root at x_0 . Suppose a close estimate to this root is z . Then $z + p$ is a closer approximation. Substitute $z + p$ for x , and use binomial expansion to arrive at an expanded form.

$$a(z + p)^3 + b(z + p)^2 + c(z + p) + d = 0 \quad (4.2)$$

$$az^3 + 3az^2p + 3azp^2 + ap^3 + bz^2 + 2bzp + bp^2 + cz + cp + d = 0 \quad (4.3)$$

At this point we can implement similar logic as that used in the previous section. Because p is a minute difference, powers of p will be relatively insignificant. Hence the expression can be reduced to:

$$\begin{aligned} az^3 + 3az^2p + bz^2 + 2bzp + cz + cp + d &\simeq 0 \\ -(3az^2p + 2bzp + cp) &= az^3 + bz^2 + cz + d \\ \frac{az^3 + bz^2 + cz + d}{3az^2 + 2bz + c} &= p \end{aligned}$$

Conclusion 1 *If z is a close approximation to the real root of $ax^3 + bx^2 + cx + d = 0$, then a closer approximation is:*

Why? This has not been shown!

$$z - \frac{az^3 + bz^2 + cz + d}{3az^2 + 2bz + c} \quad (4.4)$$

⁹Raphson, 1690 (p. 5, 7)

¹⁰Kollerstrom, 1992 (p. 348)

In fact, it is not true without restrictions

Example 5 Newton's $y = x^3 - 2x - 5$ with an initial estimation of $x_0 = 2$:

$$\begin{aligned}x_1 &= x_0 - \frac{x^3 - 2x - 5}{3x^2 - 2} \\x_0 = 2 \implies x_1 &= 2 - \frac{x^3 - 2x - 5}{3x^2 - 2} \\&= 2 - \frac{2^3 - 2 \times 2 - 5}{3 \times 2^2 - 2} \\&= 2 - \frac{-1}{10} \\&= 2.1\end{aligned}$$

An example is /
NOT a proof .

$$\begin{aligned}x_2 &= x_1 - \frac{x^3 - 2x - 5}{3x^2 - 2} \\&= 2.1 - \frac{2.1^3 - 2 \times 2.1 - 5}{3 \times 2.1^2 - 2} \\&= 2.1 - .0054 = 2.0946 \\x_3 &= x_2 - \frac{x^3 - 2x - 5}{3x^2 - 2} \\&= 2.0946 - \frac{2.0946^3 - 2 \times 2.0946 - 5}{3 \times 2.0946^2 - 2} \\&= 2.0946 - .00004851 \\&= 2.09455149\end{aligned}$$

This has resulted in a general recursive formula for approximating any cubic. Indeed it is possible to create a recursive formula for every degree polynomial. Raphson created these formulae for polynomials up to the tenth power. Raphson did not, however, attempt to create these sorts of general formulae for other (transcendental) types of equations. This is possibly because his method relies on binomial expansion, which is not present in transcendental functions.

The binomial expansion featured in Raphson's work is also in Newton's (note similarity between [3.2] and [4.2] or [4.3]). However, Raphson instead opted to find a general solution in terms of z and p instead of attempting to solve for a constant at every iteration and then reintroducing a variable. Although Newton tackled (and successively solved) Kepler's Equation in his Principia using a formulation¹¹ of Raphson's work, he did not link it to previous renditions of his own Method. In essence, it was a one-off use of a

¹¹Ypma, 1995 (p. 542)

particular approach to a problem, and not the formulation of a method for iteratively approximating roots (something only Raphson was able to do).

One instantly strikes a connection between the numerator and the denominator in [4.4]. The latter is the derivative of the former. This was, however, a connection that was not made until many years later by Joseph Simpson. An attempted explanation as to why Raphson did not spot the calculus in the method is that many of the calculus developments of the 1690s (When Raphson worked on this method) were made in mainland Europe whereas Raphson was in England. Perhaps the reference of 1690s Leibnizian calculus developments was De L'Hôpital's "*Analyse d'Infiniments Petits*", published in 1696¹². This work certainly contained the required calculus to draw resemblance between the algebra employed by Raphson in his Method and Differential Calculus. However, it seems as though Raphson was completely unaware of it; his only calculus reference being Newton's work featured in J. Wallis' *Opera Mathematica* of 1693, it was insufficient to link the two fields of algebra and differential calculus in this particular method.¹³

It must be noted that as mentioned by Kollerstrom¹⁴, Raphson was not the only English mathematician who failed to appreciate Leibniz's calculus; Edmond Halley too failed to link his algebraic method to fluxions. This is probably due to the fact that "new ideas take a while to become accepted"¹⁵. Even years later when Simpson drew the connection between the algebra of the method and differential calculus, some might have argued that it was so revolutionary that it might be wrong to connect the two fields so directly.

It is essential to analyze the importance of his method from a recursive perspective; it is now much easier to perform successive iterations and arrive at a root to the polynomial.

¹²Ypma, 1994. (p. 543)

¹³Kollerstrom, 1992 (p. 349)

¹⁴*Ibid.* (p. 350)

¹⁵*Ibid* (p. 349)

5 Thomas Simpson's Work

The Introduction of Calculus

Thomas Simpson's (1710 - 1761) contributions to the method made it what is today referred to as the "Newton" method for one single-variate equation. Thus far, there is not a hint of calculus in the root-finding methods displayed; only algebra is employed. Thomas Simpson, making no reference to his predecessors whatsoever, published a method described so simply, he correctly claimed it could be "of considerable use [compared to contemporary methods]"¹⁶. It was in his "Essays ... in ... Mathematicks"¹⁷ of 1740 that he detailed a method (for approximating the roots of a function) equivalent¹⁸ to (see Appendix for original text):

What is $R(x)$? The function?

$$A(x_k) = \frac{d}{dx} R(x_k) \implies x_{k+1} = x_k - \frac{R(x_k)}{A(x_k)}$$
$$\therefore x_{k+1} = x_k - \frac{R(x_k)}{R'(x_k)}$$

such that if x_k is a close enough approximation to the root of a function R , $x_k - \frac{R(x_k)}{R'(x_k)}$ is a closer approximation x_{k+1} .

Why? Not true in general!

This is a differential calculus-based, directly iterative approach to approximating transcendental equations; or what is referred to as the Newton Method. Note how it is essentially the same as [2.1] in Definition 1. As can be seen in the original text, it is the first published ^{ration} mentioning of the method of fluxions in connection with the approximation method. The first piece of evidence linking the work of the late 17th century and early 18th century mathematicians, and the essential characteristic of the Newton Method today. Not only did Simpson create a general formula linking the concepts of calculus and algebra within the method, but he also made it possible to apply the method to any type of continuously differentiable function, and not just the algebraic functions explored by Newton and Raphson. It is now possible to use this new formulation of the method to approximate the roots of transcendental and even mixed functions.

¹⁶Simpson, 1740 (p. vii)

¹⁷*Ibid* (p. 81)

¹⁸Note: it wasn't until Fourier further developed more modern mathematical notation a century after Simpson's publishing that this formulation of the Method was published. (Fourier, 1830)

Example 6

$$\begin{aligned}
 f(x) &= \sin(x^2 - 3x) + x^2 \\
 \implies f'(x) &= \cos(x^2 - 3x)(2x - 3) + 2x \\
 x_{k+1} &= x_k - \frac{R(x_k)}{R'(x_k)} \\
 \implies x_{k+1} &= x_k - \frac{\sin(x_k^2 - 3x_k) + x_k^2}{\cos(x_k^2 - 3x_k)(2x_k - 3) + 2x_k}
 \end{aligned}$$

		Iteration Number									
		1	2	3	4	5	6	7	8	9	10
Initial x Value		1.00000	0.96246	0.96181	0.96181	0.96181	0.96181	0.96181	0.96181	0.96181	0.96181
Variables	f(x)	0.09070	0.00152	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	f'(x)	2.41615	2.33391	2.33240	2.33240	2.33240	2.33240	2.33240	2.33240	2.33240	2.33240
New x Value		0.96245 98257 84455	0.9618 10038 035607	0.9618 09829 001757	0.9618 09829 001736	0.9618 09829 001736	0.9618 09829 001736	0.9618 09829 001736	0.9618 09829 001736	0.9618 09829 001736	0.9618 09829 001736

Table 1: Iterations of the function $y = \sin(x^2 - 3x) + x^2$. Made using Microsoft Excel. Precision of 15 decimal places according to IEEE floating point number standards. Note: By the 4th iteration there is already a 15 d.p. accuracy.

Example 7

$$f(x) = \ln x + x^2$$

$$\Rightarrow f'(x) = \frac{1}{x} + 2x = \frac{2x^2 + 1}{x}$$

$$x_{k+1} = x_k - \frac{R(x_k)}{R'(x_k)}$$

$$\Rightarrow x_{k+1} = x_k - x_k \frac{\ln x_k + x_k^2}{2x_k^2 + 1}$$

		Iteration Number									
		1	2	3	4	5	6	7	8	9	10
Initial x Value		1.00000	0.66667	0.65291	0.65292	0.65292	0.65292	0.65292	0.65292	0.65292	0.65292
Variables	f(x)	1.00000	0.03898	0.00003	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	f'(x)	3.00000	2.83333	2.83742	2.83742	2.83742	2.83742	2.83742	2.83742	2.83742	2.83742
New x Value		0.66666 66666 666667	0.6529 09253 842097	0.6529 18640 413836	0.6529 18640 419205	0.6529 18640 419205	0.6529 18640 419205	0.6529 18640 419205	0.6529 18640 419205	0.6529 18640 419205	0.6529 18640 419205

Table 2: Iterations of the function $y = \ln x + x^2$. Made using Microsoft Excel. Precision of 15 decimal places according to IEEE floating point number standards. Note: By the 4th iteration there is already a 15 d.p. accuracy.

6 Preliminary Conclusion

Thus far it can be concluded that the Newton method, while bearing the name of only one of its contributors, was the result of the efforts of multiple men. It was thanks to Isaac Newton that the method was first conceived, but in its crude shape it was of little use compared to the recursive definition demonstrated by Raphson. Even so, it was not until Simpson introduced calculus that the method was truly of great use for approximating the roots of all types of functions. Interestingly, Newton was the only man who divulged where the inspiration for his method came from. The other two men failed to acknowledge any links between their methods mathematics of their peers, suggesting they either came up with their methods themselves or where not inclined to credit their predecessors

The reason for Newton's name being associated with the Method is probably due to the fact that when leading mathematicians like Joseph Louis Lagrange (1736 - 1813) and Jean Baptiste Joseph Fourier (1768 - 1830) wrote their papers over half a century later, formulating the modern mathematical notation of the Method, they referred to it by Newton's name, never referencing the other contributors.¹⁹ With regards to the work of Lagrange and Fourier, it did not contribute to the method as much as it contributed to mathematics itself, and the method inherently benefitted from these developments. However, at its core, and in terms of its efficiency, it did not change, it was simply reformulated. While these men's contributions to the Method must not be overlooked, they were secondary to the work of the aforementioned others, and hence beyond the scope of this essay. Nonetheless, the influential weight of these two men and their publications amongst the scientific world sheds some light as to why today we usually credit Newton alone for this method's development.

But you have not proved that — under some conditions — the method works

¹⁹Cajori, 1911 (p. 29-32)

Part II

Multi-Variate Formulation

7 Simpson's Breakthrough

Extending the Investigation

Despite having formulated a partial conclusion for the initial question, and having described the evolution of the method to its formulation [2.1], this conclusion is potentially incomplete. To understand why, further analysis of Thomas Simpson's work is required. Although we have explained the evolution of the method's singlevariate formulation, Simpson's work hints towards another formulation of the method which was not particularly significant in its time, but evolved to much greater importance in the 20th Century. This would become a multivariate version of [2.1], today known as the Multivariate Newton Method.

2 Equations in 2 Variables

Simpson's "*Essays...*"²⁰ was a significant publication for the development of the Newton Method as we know it. His "Case I", as detailed above, handles the root-approximation of single non-linear equations in one variable. Simpson did not, however, stop at this point; he proceeded to describe a similar method for the approximation of the intersection of 2 implicit functions in 2 variables. Albeit more complex, it too is a significant achievement - not on its own, but *for the questions it raises and the path it leads to*.

Simpson made no reference as to where he might have discovered inspiration for this particular method, and leaves the reader to presume he intuitively followed it through from his "Case I". The definition below is interpreted using modern mathematical notation from Simpson's own work. (See Appendix for original text).

²⁰Simpson, 1740 (p. 82)

Definition 3 Take the partial derivatives with respect to each variable of the two functions to be approximated. Giving them a variable name, "A" represents the partial derivative of f_1 with respect to x . Similarly, "B" represents the partial derivative of f_1 with respect to y . Lowercase "a" and "b" are the same but for f_2 .

$$\frac{\partial}{\partial x} f_1 = A \quad (7.1)$$

$$\frac{\partial}{\partial y} f_1 = B \quad (7.2)$$

$$\frac{\partial}{\partial x} f_2 = a \quad (7.3)$$

$$\frac{\partial}{\partial y} f_2 = b \quad (7.4)$$

The final step is to combine the above variables into two ad-hoc "multiples" Δx and Δy , to arrive at the value by which to adjust x_k and y_k , the initial estimated coordinates of the intersection.²¹

$$\frac{Br - bR}{Ab - aB} \stackrel{(\ominus)}{=} \Delta x$$

$$\frac{aR - Ar}{Ab - aB} \stackrel{(\ominus)}{=} \Delta y$$

Explanation?
Is that a definition of Δx and Δy ?

where R and r are the two equations being intersected:

$$f_1(x, y) = R$$

$$f_2(x, y) = r$$

Then, estimating initial values of x_k and y_k of the intersection, the closer values x_{k+1} and y_{k+1} can be attained by: Why? This is not shown!

$$x_{k+1} = x_k + \Delta x_k = x_k + \frac{Br - bR}{Ab - aB} \quad (7.5)$$

$$y_{k+1} = y_k + \Delta y_k = y_k + \frac{aR - Ar}{Ab - aB} \quad (7.6)$$

The great benefit of this method is allowing us to find the intersection between two functions in their implicit form. For example, in the intersection of [7.7] and [7.8]. Not shown!!

²¹Note: In the Part II of this essay, subscript k denotes the argument to which the subscript belongs evaluated at the k^{th} iteration.

Example 8 Intersection of $f_1(x, y) = x^2 + y^2 - 10$ and $f_2(x, y) = 2x^3 - y^2$:

$$f_1(x, y) = x^2 + y^2 - 10 \quad (7.7)$$

$$f_2(x, y) = 2x^3 - y^2 \quad (7.8)$$

taking the partial derivatives of each function with respect to each variable:

$$\frac{\partial}{\partial x} f_1 = 2x = A \quad (7.9)$$

$$\frac{\partial}{\partial y} f_1 = 2y = B \quad (7.10)$$

$$\frac{\partial}{\partial x} f_2 = 6x^2 = a \quad (7.11)$$

$$\frac{\partial}{\partial y} f_2 = -2y = b \quad (7.12)$$

taking variables R_k and r_k , the deviation from zero at the k^{th} iteration resultant from x_k and y_k such that:

$$f_1(x_k, y_k) = R_k$$

$$f_2(x_k, y_k) = r_k$$

and setting up the multiples as described by Simpson:

$$\begin{aligned} \frac{Br - bR}{Ab - aB} &= \frac{2yr + 2yR}{-4xy - 12x^2y} = \frac{4x^3 + 2x^2 - 20}{-4x - 12x^2} \\ \frac{aR - Ar}{Ab - aB} &= \frac{6x^2R - 2xr}{-4xy - 12x^2y} = \frac{2x^3 + 6xy^2 - 60x + 2y^2}{-4y - 12xy} \end{aligned}$$

we can formulate recursively that:

$$\Rightarrow x_{k+1} = x_k + \frac{Br - bR}{Ab - aB} = x_k + \frac{4x^3 + 2x^2 - 20}{-4x - 12x^2} \quad (7.13)$$

$$\Rightarrow y_{k+1} = y_k + \frac{aR - Ar}{Ab - aB} = y_k + \frac{2x^3 + 6xy^2 - 60x + 2y^2}{-4y - 12xy} \quad (7.14)$$

We have a recursive formula for finding the intersection of these particular functions. Below are results from a computer model created based on this recursive formula to demonstrate the changing of the variables over each iteration.

An example is not a proof

		Iteration Number									
		1	2	3	4	5	6	7	8	9	10
Initial Estimate x Value		2.000	1.643	1.562	1.559	1.559	1.559	1.559	1.559	1.559	1.559
Initial Estimate y Value		2.000	2.857	2.753	2.752	2.752	2.752	2.752	2.752	2.752	2.752
Variables	R	-2.000	0.862	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	r	12.000	0.705	0.052	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	A	4.000	3.286	3.125	3.117	3.117	3.117	3.117	3.117	3.117	3.117
	B	4.000	5.714	5.505	5.503	5.503	5.503	5.503	5.503	5.503	5.503
	a	24.000	16.194	14.647	14.574	14.574	14.574	14.574	14.574	14.574	14.574
	b	-4.000	-5.714	-5.505	-5.503	-5.503	-5.503	-5.503	-5.503	-5.503	-5.503
Change in x Value		-0.357	-0.080	-0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Change in y Value		0.857	-0.105	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 3: Iteration of the formulas $x_{n+1} = x_n + \frac{4x^3+2x^2-20}{-4x-12x^2}$ and $y_{n+1} = y_n + \frac{2x^3+6xy^2-60x+2y^2}{-4y-12xy}$ for the intersection of the functions $f_1 = x^2 + y^2 - 10$ and $f_2 = 2x^3 - y^2$. Made using Microsoft Excel. Precision of 15 decimal places according to IEEE floating point number standards.

As can be seen, from about the fourth iteration the error variables R and r become insignificant (Smaller than 10^{-9} marked as 0) very quickly. The intersection arrived at is point (1.559, 2.752).

This method, in spite of its significance for the mathematics of 1740, leaves the modern day mathematician with two questions. Firstly, why did Simpson not attempt to make a general method for all systems of equations, instead leaving it in a formulation that requires an *ad-hoc* algebraic manipulation for each set of equations? More importantly, why did he not extend his finding to systems of more than two equations?

8 The Multivariate Newton Method

n Equations in n Variables

Thus far we have seen how Simpson not only developed the Newton Method of today for single non-linear equations, but also for systems of two equations in two variables. However, his impact upon the field of numerical analysis went deeper, and can still be seen today. It led to a method for solving n functions in n variables. To illustrate this, let us employ a tool that Simpson did not have in his lifetime: Matrices. The key lies in the unexplained multiples of the previous section:

indeed!

$$\frac{Br - bR}{Ab - aB} \text{ and } \frac{aR - Ar}{Ab - aB}$$

We can organize these terms as matrices:

$$\Delta x = \frac{Br - bR}{Ab - aB} = \begin{bmatrix} \frac{b}{Ab - aB} & -\frac{B}{Ab - aB} \end{bmatrix} \begin{bmatrix} -R \\ -r \end{bmatrix} \quad (8.1)$$

$$\Delta y = \frac{aR - Ar}{Ab - aB} = \begin{bmatrix} -\frac{a}{Ab - aB} & \frac{A}{Ab - aB} \end{bmatrix} \begin{bmatrix} -R \\ -r \end{bmatrix} \quad (8.2)$$

and if we join the two matrices in 8.1 and 8.2

$$\begin{bmatrix} \frac{b}{Ab - aB} & -\frac{B}{Ab - aB} \\ -\frac{a}{Ab - aB} & \frac{A}{Ab - aB} \end{bmatrix} = \mathbf{M}$$
$$\implies \mathbf{M}^{-1} = \begin{bmatrix} A & B \\ a & b \end{bmatrix}$$

This astonishing yet simple result shows that the inverse of \mathbf{M} is the matrix of partial derivatives with respect to the functions' variables such that evaluated at the k^{th} iteration:

$$\boxed{\mathbf{M}_k \times \mathbf{R}_k = \Delta \mathbf{X}_k}$$

What is X_k ?
What is R_k ?

Poorly presented

Where \mathbf{R}_k is the matrix of the residuals, or values of the functions at k^{th} iteration:

$$\mathbf{R}_k = \begin{bmatrix} -R \\ -r \end{bmatrix} = \begin{bmatrix} f_1(x_k, y_k) \\ f_2(x_k, y_k) \end{bmatrix}$$

and where $\Delta\mathbf{X}_k$ is the matrix of values to be added to x and y at the k^{th} iteration:

$$\begin{aligned} \mathbf{X}_k &= \begin{bmatrix} \text{Value of } x \text{ at iteration } k \\ \text{Value of } y \text{ at iteration } k \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} \\ \Delta\mathbf{X}_k &= \begin{bmatrix} \text{Change in value of } x \text{ at iteration } k \\ \text{Change in value of } y \text{ at iteration } k \end{bmatrix} = \begin{bmatrix} \frac{Br-bR}{Ab-aB} \\ \frac{aR-Ar}{Ab-aB} \end{bmatrix}_k \end{aligned}$$

such that²²:

$$\mathbf{X}_{k+1} = \mathbf{X}_k + \Delta\mathbf{X}_k \quad (8.3)$$

Using the result [8.3] we can recursively approximate closer values of variables x and y with each iteration. This successfully explains the multiples in Simpson's "Case II", showing that both the algebraic and matrix approach are *linked*. Hence, patterns developed in the matrix object should hold true for the algebraic counterpart.

Proposition 2 Consider matrix \mathbf{M}^{-1} . It is a matrix of partial derivatives organized as follows:

$$\mathbf{M}^{-1} = \begin{bmatrix} A & B \\ a & b \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x} f_1 & \frac{\partial}{\partial y} f_1 \\ \frac{\partial}{\partial x} f_2 & \frac{\partial}{\partial y} f_2 \end{bmatrix}$$

such that

$$\mathbf{M}_k^{-1} \times \Delta\mathbf{X}_k = \mathbf{R}_k$$

Now if we were presented with a case where instead of 2 functions in 2 variables, we had n functions in n variables

$$\begin{aligned} f_1(x, y) &\rightarrow f_1(x_1, x_2 \dots x_n) \\ f_2(x, y) &\rightarrow f_2(x_1, x_2 \dots x_n) \\ &\dots \\ &f_n(x_1, x_2 \dots x_n) \end{aligned}$$

²²Keffer, 1998

then a similar scenario could be constructed. The matrix of partial derivatives of the functions \mathbf{M}^{-1} can be expanded following the same principle such that

$$\begin{bmatrix} \frac{\partial}{\partial x} f_1 & \frac{\partial}{\partial y} f_1 \\ \frac{\partial}{\partial x} f_2 & \frac{\partial}{\partial y} f_2 \end{bmatrix} \rightarrow \begin{bmatrix} \frac{\partial}{\partial x_1} f_1 & \dots & \frac{\partial}{\partial x_n} f_1 \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} f_n & \dots & \frac{\partial}{\partial x_n} f_n \end{bmatrix}$$

Here, the number of rows and columns is no longer restricted to 2×2 as in the \mathbf{M}^{-1} matrix as before, but is now $n \times n \forall n \in \mathbb{Z}$. This matrix of partial derivatives is known as the Jacobian Matrix \mathbf{J} at k^{th} iteration for a system of functions whose value at place j, i can be defined²³:

$$(J_{j,i}) = \frac{\partial(f_j)}{\partial(x_i)}$$

similarly, the matrices $\Delta \mathbf{X}_k$ and \mathbf{R}_k can be redefined for n functions in n variables:

$$\Delta \mathbf{X}_k = \begin{bmatrix} \text{Change in } x_1 \\ \dots \\ \text{Change in } x_n \end{bmatrix} = \begin{bmatrix} \Delta x_1 \\ \dots \\ \Delta x_n \end{bmatrix}_k$$

$$\mathbf{R}_k = \begin{bmatrix} -f_1 \text{ value at } k^{\text{th}} \text{ iteration} \\ \dots \\ -f_n \text{ value at } k^{\text{th}} \text{ iteration} \end{bmatrix} = \begin{bmatrix} -f_1(x_1, x_2, \dots, x_n) \\ \dots \\ -f_n(x_1, x_2, \dots, x_n) \end{bmatrix}_k$$

Conclusion 2

$$\mathbf{J}_k \times \Delta \mathbf{X}_k = \mathbf{R}_k$$

$$\mathbf{X}_{k+1} = \mathbf{X}_k + \Delta \mathbf{X}_k$$

$$\therefore \mathbf{X}_{k+1} = \mathbf{X}_k + \mathbf{J}_k^{-1} \mathbf{R}_k \quad (8.4)$$

We are hence presented with the Multivariate version of the Newton Method. The result summarized in [8.4] is a single line of notation linking the concepts of Newton's single-variate Method and matrix algebra for approximating roots of n equations in n variables. It is crucial to mark how this was derived from Thomas Simpson's work – something he was not able to do at the time simply because he did not have the mathematical tool to do so: matrices.

²³Keffer, 1998 (Lecture Notes)

But nothing has been proved.

too many definitions included in the statement of the proposition which is unclear

This EE is a good historical account of the derivation of the Newton method for approximating roots of a scalar or vectorial function but unfortunately it is purely descriptive: it does not contain a single proof.

Nor does it contain any mention of the conditions necessary for the sequence of approximations to converge. This is a serious omission since it gives the impression that the sequence always converges which is incorrect.

A little mathematical activity (as opposed to purely descriptive activity) and reasoning would have considerably improved this essay.

9 Conclusion

Newton's work was critical to the development of the method – it was his thought experiments that sparked it. Hence, to him should be attributed the success of **conceptualizing** the method. However, we should not disregard the work of his successors; Raphson and Simpson, who made the contemporary application of the method explained in Part I of this essay possible. It was Raphson who developed its **direct iteration**, and it was Simpson who **linked it with calculus** making it possible for the Method to then develop in the twentieth century, as shown in Part II, to approximate solutions for **systems of n equations in n variables**. These four critical points represent the four steps that the Method went through: Conceptualization, Development of Direct Iteration, Link with Calculus and Link with Matrices & systems of equations. An evolution that took over 300 years.

However, when it comes to the Multivariate Newton Method, it was Simpson's work that most significantly contributed to it. The *ad-hoc* multiples were a foreshadowing of the work to come in the 20th Century, and inherently the multivariate version of the algorithm should be named after the Thomas Simpson, the man who first hinted at it, just as the Newton Method is named after Isaac Newton.

Perhaps for this reason, and for pragmatic purposes, the Method is rightly named after Isaac Newton. But he was not the Method's sole father. The method is instead the offspring of centuries of mathematical development and holistic cooperation.

Ultimately it was not the work of one man, but the successive development of the method throughout the ages that makes it so useful today. This paper does not intend to designate one man as the master behind the method (as others have before), but instead highlight how it was overlapping and continuous work of all these men that contributed to the evolution of Isaac Newton's Numerical Method.

From an algebraic method for approximating roots of polynomials, to a recursive algorithm for approximating solutions of multivariate non-linear systems using matrices, this method is a story of true mathematical continuity. The knowledge continuum that moves science forward.

Words: 3800

10 Bibliography

References

- Cajori, F. (1911). *Historical Note on the Newton-Raphson Method of Approximation*. The American Mathematical Monthly.
- Encyclopædia Britannica. (2011). *Brook Taylor*. Retrieved September 11, 2011, from Encyclopædia Britannica Online: www.britannica.com/EBchecked/topic/584793/Brook-Taylor
- Fourier, J. (1830). *Analyse des Équations Déterminées*. Paris: Firmin Didot Frères, Libraires.
- Keffer, D. (1998). *Rootfinding in systems of Equations*. Knoxville, Tennessee.
- Kollerstrom, N. (1992, 25). *Thomas Simpson and "Newton's method of approximation": an enduring myth*. British Journal of Historical Science.
- Newton, I. (1711). *De analysi per aequationes numero terminorum infinitas*. In D. T. Whiteside's, *The Mathematical Papers of Isaac Newton*. London.
- Pawley, M. G. (1940). *New Criteria for Accuracy in Approximating Real Roots by the Newton- Raphson Method*. National Mathematics Magazine.
- Raphson, J. (1690). *Analysis Aequationum Universalis*. London.
- Simpson, T. (1740). *Essays On Several Curious And Useful Subjects, In Speculative And Mixed Mathematics*. London.
- Wallis, J. (1685). *A Treatise of Algebra both Historical and Practical*. London.
- Ypma, T. J. (1995, December Vol. 37, No. 4). *Historical Development of the Newton-Raphson Method*. Society for Industrial and Applied Mathematics Review.

A Appendix

Extracts from Thomas Simpson's "Essays On Several Curious And Useful Subjects, In Speculative And Mixed Mathematics."

Page 81:

Case I, When only one Equation is given, and one Quantity (x) to be determined.

Take the **fluxion** of the given Equation (be it what it will) supposing x , the unknown, to be the variable Quantity; and having divided the whole by x' , let the Quotient be represented by A . Estimate the value of x pretty near the Truth, substituting the same in the Equation, as also in the Value of A , and let the Error R , or resulting Number in the former, be divided by this numerical Value of A , and the Quotient be subtracted from the said former Value of x ; and from thence will arise a new Value of that Quantity much nearer to the Truth than the former, where-with proceeding as before, another new Value may be had, and so an-other, etc. 'till we arrive to any Degree of Accuracy desired.

Page 82:

Case II, When there are two Equations given, and as many Quantities (x and y) to be determined.

Take the Fluxions of both the Equations, considering x and y as variable, and in the former collect all the Terms, affected with x' , under their proper Signs, and having divided by x' , put the Quotient A ; and let the remaining Terms, divided by y' , be represented by B : In like manner, having divided the Terms in the latter, affected with x' , by x' , let the Quotient be put $= a$, and the rest, divided by y' , $= b$. Assume the Values of x and y pretty near the Truth, and substitute in both the Equations, marking the Error in each, and let these Errors, whether positive or negative, be signified by R and r respectively: Substitute likewise in the values of $A B a b$, and let $\frac{(Br-bR)}{(Ab-aB)}$ and $\frac{(aR-Ar)}{(Ab-aB)}$ be converted into Numbers, and respectively added to the former Values of x and y ; and thereby new Values of those Quantities will be obtained; from whence, by repeating the Operation, the true Values may be approximated. ad libitum.