

# Origin and Evolution of the Secant Method in One Dimension

**Joanna M. Papakonstantinou and Richard A. Tapia**

**Abstract.** Many in the mathematical community believe that the secant method arose from Newton's method using a finite difference approximation to the derivative, most likely because that is the way that it is taught in contemporary texts. However, we were able to trace the origin of the secant method all the way back to the Rule of Double False Position described in the 18th-century B.C. Egyptian Rhind Papyrus, by showing that the Rule of Double False Position coincides with the secant method applied to a linear equation. As such, it predates Newton's method by more than 3,000 years. In this paper, we recount the evolution of the Rule of Double False Position as it spanned many civilizations over the centuries leading to what we view today as the contemporary secant method. Unfortunately, throughout history naming confusion has surrounded the Rule of Double False Position. This naming confusion was primarily a product of the last 500 years or so and became particularly troublesome in the past 50 years, creating confusion in the use of the terms Double False Position method, Regula Falsi method, and secant method. We elaborate on this confusion and clarify the names used.

**1. INTRODUCTION.** The goal of this paper is to present a historical development of the secant method in one dimension. We hope to enhance perspective and understanding by presenting the secant method in an environment that includes the closely-related algorithms of Newton's method, the Regula Falsi method, and the modified Regula Falsi method. Hence, we begin by describing these methods using current functional notation in §2. An implied convention in the literature that we subscribe to is that the secant method is an iterative procedure, while the Rule of Double False Position is not. The historical development of the Rule of Double False Position as a non-iterative method is presented in §3. Here we demonstrate first that the algebraic formula characterizing the Rule of Double False Position is exactly the algebraic formula that characterizes the secant method. Hence, the Rule of Double False Position can be viewed as the first iteration of the secant method. However, the original definition of the Rule of Double False Position was for a linear equation. Moreover, for a linear equation, the secant method converges in one iteration; in this application, the two methods coincide and the Rule of Double False Position should be considered the origin of the secant method. In §4 we present the various names that have been associated with the Rule of Double False Position throughout many civilizations and centuries. The first path to obtaining the secant method is complete once iteration has been incorporated into the Rule of Double False Position. This completion is discussed in §5. In §6 we consider the naming of the secant method and the first determination of its convergence rate. A second path to the secant method is the path that originates with Newton's method. This path is briefly discussed in §7. Finally, in §8 we make some concluding remarks and attempt to rationalize and perhaps clarify the naming confusion that evolved over the many years.

**2. FOUR BASIC ALGORITHMS.** In this section, we present the numerical methods using current functional notation that will be referred to throughout the paper. Consider $f : \mathbb{R} \to \mathbb{R}$ and let $x^* \in \mathbb{R}$ be a zero of the function $f$.

http://dx.doi.org/10.4169/amer.math.monthly.120.06.500
MSC: Primary 01A85, Secondary 65F10; 49M15

**2.1. Newton's Method.** The guiding principle in Newton's method is the use of a succession of zeros of tangent lines to better approximate a zero of the function $f(x)$. In Figure 1, $f(x)$ represents the nonlinear function whose zero we are trying to find.

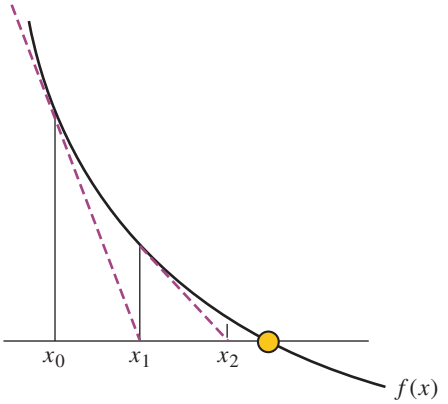


**Figure 1.** Newton's method begins by using the tangent line passing through the point $(x_0, f(x_0))$.

The point-slope form of the line passing through the point $(x_0, f(x_0))$ with slope $f'(x_0)$ is

$$l(x) = f(x_0) + f'(x_0)(x - x_0).$$

Now, letting $x_1$ be the $x$-intercept of this line, i.e., the point such that $l(x) = 0$, we obtain

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Hence, the iteration

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

is *Newton's method* and $x_k$ represents the $k$th approximation to the solution.

**2.2. The Secant Method.** In contrast to Newton's method, which uses a succession of zeros of tangent lines, the guiding principle in the secant method is the use of a succession of zeros of secant lines obtained by two-point interpolation to better approximate a zero of a function $f(x)$. In Figure 2, $f(x)$ represents the function whose zero we are trying to find.

The two-point form of the line passing through the points $(x_0, f(x_0))$ and $(x_1, f(x_1))$ is

$$l(x) = \frac{(x - x_0)}{(x_1 - x_0)} f(x_1) + \frac{(x - x_1)}{(x_0 - x_1)} f(x_0). \tag{2}$$

Now, letting $x_2$ be the $x$-intercept of this line, i.e., the point such that $l(x) = 0$, we obtain

$$x_2 = x_1 - \frac{x_1 - x_0}{f(x_1) - f(x_0)} f(x_0). \tag{3}$$

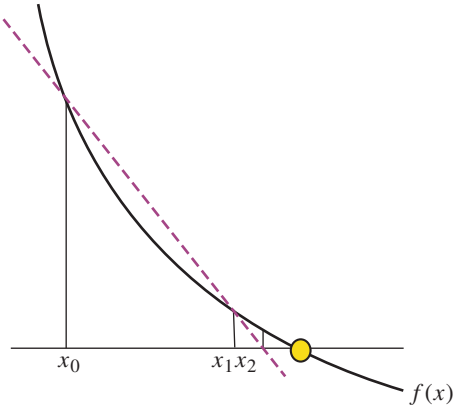


**Figure 2.** The secant method begins by using the secant line passing through the points $(x_0, f(x_0))$ and $(x_1, f(x_1))$.

The resulting iteration

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f(x_k) - f(x_{k-1})} f(x_k), \tag{4}$$

is the secant method and can also be written as

$$x_{k+1} = \frac{x_{k-1} f(x_k) - x_k f(x_{k-1})}{f(x_k) - f(x_{k-1})}. \tag{5}$$

As mentioned above, a popular way of obtaining the secant method in one dimension is to replace the derivative in the Newton iteration (1) with the difference quotient

$$\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}},$$

which can be viewed as an approximation to $f'(x_k)$. This is an interesting fact, but it should not be treated as a definition.

**2.3. The Regula Falsi Method.** The guiding principle of the Regula Falsi method is, like the secant method, the use of a succession of zeros of secant lines obtained from two-point interpolation to better approximate a zero of a function $f(x)$. In Figure 3, $f(x)$ represents the function whose zero we are trying to find.

A key difference between the Regula Falsi method and the secant method is that in the first step of the Regula Falsi method, the two initial estimates, $x_0$ and $x_1$, are chosen such that $f(x_0)$ and $f(x_1)$ are of opposite signs ($f(x_0) f(x_1) < 0$). This is unlike the secant method, where there is no restriction that the initial estimates bracket a zero. The iteration

$$x_{k+1} = \frac{\overline{x} f(x_k) - x_k f(\overline{x})}{f(x_k) - f(\overline{x})} \tag{6}$$

is the *Regula Falsi method*, where $\overline{x}$ is an endpoint of the original bracketing interval that remains fixed. At each step of the Regula Falsi method, the current approximation replaces the previous interval endpoint whose corresponding function value has the same sign as the current best estimate of $x^*$, while the other interval endpoint is retained. A new secant line of $f(x)$ is constructed and the process is continued in an

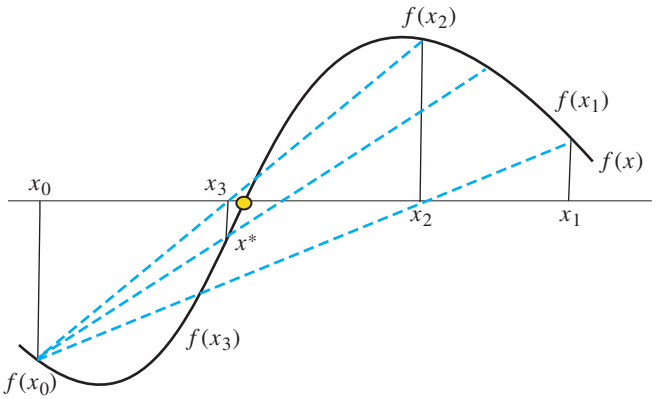


**Figure 3.** The Regula Falsi method: $x^*$ may not always remain bracketed at each step.

iterative fashion, always holding the same initial estimate, one endpoint of the original bracketing interval, fixed for all subsequent iterations while the other endpoint is always updated. Booth [**3**], in 1955, seems to be the first to refer to this method as Regula Falsi.

The key feature of the Regula Falsi method is that, instead of always using the two most recently computed iterates as in the secant method (5), one of the initial estimates is held fixed for all subsequent iterations while the other endpoint is always updated. However, it is important to mention that in the Regula Falsi method, as in the secant method, a zero does not necessarily remain bracketed by successive iterates at each step and, in some instances, the methods fail.

**2.4. The Modified Regula Falsi Method.** Modifications of the Regula Falsi method have been made that ensure that a zero remains bracketed at each step;[1] see Figure 4. For example, at each step, if instead of holding one of the interval endpoints fixed, the interval endpoints are changed to ensure that the new interval contains a zero of $f(x)$. This just means that the $x$-value corresponding to the function value that has opposite sign as the current function value is always retained, and not just in the first step as

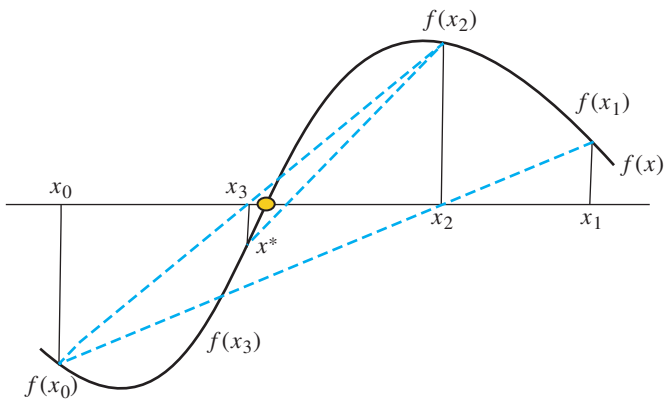


**Figure 4.** The Modified Regula Falsi method: $x^*$ remains bracketed at each step.

[1]Early examples of modifications of the Regula Falsi method may be found in Willers' 1948 book [**31**] and in Householder's 1953 book [**13**].

in the Regula Falsi method.[2] Stanton [**27**], in 1961, was the first author that we could find to explicitly describe our Modified Regula Falsi method. He referred to it as the Regula Falsi method.

The bracketing feature of the Modified Regula Falsi method should be viewed as a safeguarding procedure that enhances convergence of the approximating sequence to a zero of the function under consideration. However, it does so at a price, since it is well known that the Modified Regula Falsi method is much slower than the secant method. We digress briefly from the main purpose of the paper to qualify these remarks.

An iterative method for approximating a zero of a nonlinear equation is said to have convergence rate $r$ if it generates sequences $\{x_k\}$, that when they converge to a zero $x^*$, they satisfy the inequality

$$|x_{k+1} - x^*| \leq c\,|x_k - x^*|^r \tag{7}$$

for positive constants $c$ and $r$ and for no larger $r$. It is well known that the convergence rate for Newton's method is 2; for the secant method, the convergence rate is the golden mean $\frac{1}{2}(1+\sqrt{5}) \approx 1.62$; and for the Regula Falsi method and the Modified Regula Falsi method, the convergence is linear, i.e., a rate of 1 with a constant $c < 1$. Newton's method and the secant method, when they converge, are extremely fast; however, they are only guaranteed to converge if the initial guesses are sufficiently close to a zero. Newton's method requires the calculation of a function value and a derivative value per iteration. The secant method requires only one function value per iteration. Hence, if the calculation of a function and a derivative are comparable, then we could perform two iterations of the secant method with the same work needed to perform one iteration of Newton's method. Such a two-step method would have a convergence rate of the golden mean squared, which is substantially larger than 2. Thus, from a computational complexity point of view, the secant method is an optimal method. The safeguarding modification of the secant method that leads to the Modified Regula Falsi method promotes convergence, but destroys its fast convergence. Therefore, a good strategy would be to consider the hybrid method of starting out with Modified Regula Falsi and then switching to the secant method, and this is often done.

**3. THE FASCINATING LIFE OF THE RULE OF DOUBLE FALSE POSITION.** Our story begins with the so-called Rule of Single False Position and the Rule of Double False Position, which collectively we call the Rules of False Position. The history of the Rule of Double False Position spans many civilizations over many centuries. In ancient times, mathematics was used as a tool to answer questions that arose in daily life. The earliest evidence of two of these tools (the Rules of False Position) was found in Egyptian papyri and Babylonian clay tablets from the 18th century B.C.[3]

The Rules of False Position were always written rhetorically rather than using the language or notation of today's mathematics, as they were not known at the time. In addition, the problems that were solved using the Rules of False Position were often presented within the context of a real-life situation. As we describe the Rules of False Position, it is critically important to realize that the Egyptians and the Babylonians did not know algebra, indeed it did not exist at that time, nor did they have the notion of an equation; hence, they could not make obvious simplifications and they did not work with a general rule. There is no evidence of the use of a procedure instead, each prob-

[2]To learn more about modifications of the Regula Falsi method that ensure that each new interval contains a zero, see Bronson [**4**].

[3]For more on the history of mathematics in the Babylonian civilization, see Høyrup [**14**], Neugebauer [**19**], and Robson [**24**].

lem used specific numbers with the solution given as a set of instructions. So, problems that would be considered trivial today posed a high degree of difficulty in ancient times.

The most important mathematical text from ancient Egypt is the Ahmes Papyrus, written by the scribe Ahmes in about 1659 B.C. and derived from material dated approximately 2000–1800 B.C. [**8**]. Today it is called the *Rhind Mathematical Papyrus* after the Scottish Egyptologist and antiquarian Alexander Henry Rhind, who purchased it from a shop in Luxor while traveling in Egypt and brought it back to England in 1858. The Papyrus was donated by Rhind's estate to the British Museum in 1864, where it still resides today [**8, 21**]. The Rhind Mathematical Papyrus, written in hieratic notation[4] (see Figure 5), is a two-sided document containing a collection of 87 real-life word problems, with solutions on one side and tables to aid in computation on the other. The examples cover a wide range of mathematical ideas needed for a scribe to fulfill his duties. Thus, we deduce that this treatise was used in the training of scribes.

**3.1. The Rule of Single False Position.** While the problems of the Rhind Mathematical Papyrus were written rhetorically, scholars are in agreement that they represent what can be expressed today as algebraic equations. Problems 24–34 of the Rhind Mathematical Papyrus are examples of problems in one unknown of the first degree, which can be represented using contemporary algebraic notation as finding a number $x$ such that

$$a_1x + \cdots + a_nx = c. \tag{8}$$

Of course, using algebra, we would simplify such problems to

$$ax = c, \tag{9}$$

where $a = a_1 + a_2 + \cdots + a_n$. Also, from a current mathematical point of view, this problem is simple to solve. We need only to sum the $a_i$s in (8) and divide $c$ by $a$ to get the solution $x = \frac{c}{a}$. However, remember that the people of the time could not perform algebraic simplifications; hence, the value of the coefficient $a$ was not known.

The first step of the method they used to solve for $x$ in the rhetorical analog of the linear equation (8), was to choose a so-called false position, i.e., an initial guess of the solution. The initial guess was not so arbitrary. Instead, the false position was chosen with the aim of operating with whole numbers, since calculation with fractions could present difficulties [**8**]. Keep in mind that they were working only with the rhetorical analog of equation (8) and not with the simplified algebraic equation (9). The Rule of Single False Position is the following.

Choose a false position (initial guess) $x = x_0$, and calculate $c_0$ where

$$ax_0 = c_0.$$

Now, calculate

$$x = \frac{(c)(x_0)}{c_0}, \tag{10}$$

[4]The hieroglyphic form was pictorial, where each character represented an object. The hieratic form was symbolic. It replaced frequently-used symbols with new symbols that made it more economical. As time passed and writing came into general use in Egypt, even the hieratic form proved to be too cumbersome. This led to the invention of a type of shorthand, the demotic (popular) notation [**5**].

which we recognize as the solution

$$x = (c)\left(\frac{1}{a}\right).$$

Hence, the solution was obtained (10) without determining $a$ or dividing by $a$.[5] (See Table 1 for an example.) This method was later called *Simple False Position* [**8**], *Process of Supposition*, or most commonly, the *Rule of Single False Position* [**16**]. In essence, this method was a way of using an initial guess to obtain the solution to a specific problem and was not a general rule for solving other problems of the same kind.

Today, knowing the form of the equation and knowing simple algebra, if we were asked to solve Equation (8) not knowing the value of $a$, knowing the value of $c$, and being able to calculate $cx$ for any given $x$, most clever students would come up with the Rule of Single False Position. It is a reasonable solution technique and it is quite remarkable that it was discovered in that day and age without the knowledge of equations or algebra.

We now present an example problem solved by using the Rule of Single False Position. Figure 5 illustrates Problem 26 from the Rhind Mathematical Papyrus, written in hieratic notation.[6]

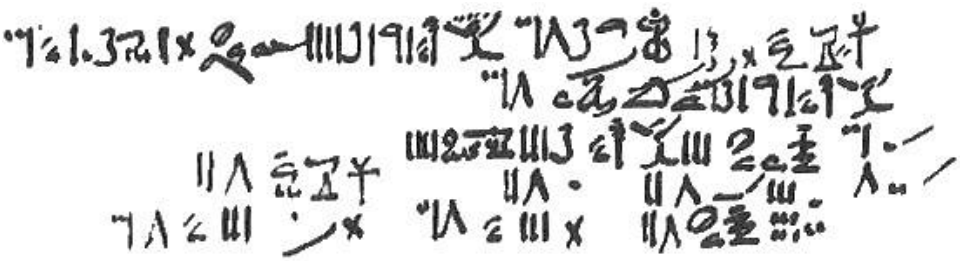

**Figure 5.** (Taken from Chabert [**8**].)

To demonstrate how difficult the notation and the technique of calculation was at that time, we transcribe Problem 26 from hieroglyphic notation into algebraic notation and describe the enumerated steps to solve the problem using the Rule of Single False Position (see Table 1).

**3.2. The Rule of Double False Position.** Since the Egyptians had the Rule of Single False Position to solve for $x$ in $ax = c$, they quite naturally tried to apply it to other real-life word problems, which we would represent today using algebraic notation as finding a number $x$ such that

$$ax + b = c, \tag{11}$$

where $b \neq 0$. Having no knowledge of algebra at the time, people did not know how to move terms from one side of an equation to the other [**16**]. Furthermore, they considered (9) and (11) to represent two different mathematical phenomena.

[5]For a discussion of the differing schools of thought on how this problem was solved, see math historian Eleanor Robson's book [**24**].

[6]A discussion of this problem can be found in various texts. To read more on Egyptologist Thomas E. Peet's comments on this problem, see [**22**].

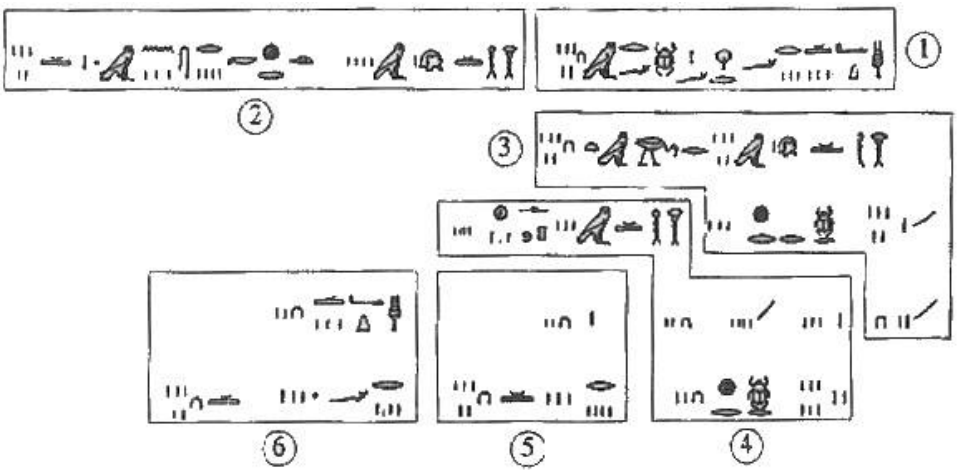

**Table 1.** Description of Problem 26 of the Rhind Mathematical Papyrus. Adapted from Chabert [8].

| Step | Transcription of hieroglyphics | Description using algebraic notation |
|---|---|---|
| 1 | A quantity, $\frac{1}{4}$ of it added to it, becomes 15 | $x + \frac{1}{4}x = 15$ |
| 2 | Operate on 4; make thou $\frac{1}{4}$ of them, namely 1<br>The total is 5. | Guess $x = 4$<br>$4 + 1 = 5$. |
| 3 | Operate on 5 for the finding of 15<br>\1 5<br>\2 10<br>There becomes 3. | Divide: $\frac{15}{5} = 3$ |
| 4 | Multiply: 3 times 4.<br>1 3<br>2 6,<br>\4 12<br>There becomes 12. | Multiply wrong answer ($x = 4$) by 3:<br>$3 \times 4 = 12$. |
| 5 | 1 12,<br>$\frac{1}{4}$ 3<br>Total 15 | $12 + \frac{1}{4}(12) = 15$ |
| 6 | The quantity is 12. $\frac{1}{4}$ of it is 3; the total is 15. | Thus, $x = 12$. |

The first step of the method they used to solve for $x$ in the rhetorical analog of the linear equation (11) was to choose two different initial guesses (false positions) of the solution. There were no restrictions on the initial guesses. Suppose that the first guess of the solution is $x = x_0$; then we get the corresponding residual error $e_0$, where

$$ax_0 + b - c = e_0. \tag{12}$$

Suppose that the second guess of the solution is $x = x_1$; then we get the corresponding residual error $e_1$, where

$$ax_1 + b - c = e_1. \tag{13}$$

They gave instructions[7] on how to obtain the solution using the relation

$$x = \frac{x_0 e_1 - x_1 e_0}{e_1 - e_0}. \tag{14}$$

This method for solving for $x$ is now most commonly referred to as the *Rule of Double False Position* [**16**].[8]

**3.3. Double False Position as the Secant Method.** If for the two arbitrary false positions, $x_0$ and $x_1$, we write $e_0 = f(x_0)$ and $e_1 = f(x_1)$, where $f(x) = ax + b - c$, then we obtain the solution[9]

$$x = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)}, \tag{15}$$

which coincides with the first step of the secant method (5) applied to the equation $f(x) = 0$. It is most interesting that the secant method applied to a linear equation converges in one step; hence, it is correct to say that the Rule of Double False Position is the secant method applied to a linear equation.

When compared to the construction of the Rule of Single False Position, it is much more remarkable that the ancients were able to construct the Rule of Double False Position. It should be considered a great mathematical milestone. On the one hand, we are quite surprised that what they constructed coincides with the secant method for a linear equation. On the other hand, from a purely philosophical point of view, we reason that if we are given a linear equation and two points of interpolation and were able to write down an expression for the solution, then the procedure must be equivalent to constructing a line through these two points and giving the $x$-intercept of this line as the solution, i.e., the guiding principle of the secant method. Since the function is linear, the line interpolating the two points is the original function, and we obtain the exact solution without iterating. It is satisfying that there is beautiful and wonderful consistency in mathematics.

The Egyptian papyri contained rhetorical examples that represent linear equations solved using the Rule of Double False Position. While many of the problems dealt with the sale and distribution of properties, inheritance, or for the purpose of portion control and the prediction of production, some of the problems presented seem inconsequential in comparison. For example, consider the following problem.[10]

> *When asking someone his age he answers: if my age were doubled and added to this* $\frac{1}{2}$, $\frac{1}{3}$, *and* $\frac{1}{4}$ *part of my age and* 6 *years, then all together should equal* 80. *How old is he?*

This example can be written using algebraic notation as the linear equation

$$2x + \frac{1}{2}x + \frac{1}{3}x + \frac{1}{4}x + 6 = 80.$$

[7]Instructions resemble: "Multiply the second error by the first guess and multiply the first error by the second guess. Subtract whichever product is smaller from the larger and divide this result by the difference of the smaller error subtracted from the larger error."

[8]In the next section, we explain that this rule was first given an English name, the Rule of Two (False) Positions, by Chuquet in 1484.

[9]In 1978, Smeur [**26**] described how Frisius' in 1540 solved $1\frac{1}{2}x^2 = 200$ using the "Rule of Double False." Smeur explained that $x$ is calculated from $x = \frac{x_1^2 f_2 - x_2^2 f_1}{f_2 - f_1}$, which lends itself to the notation we use in (15).

[10]Taken from p. 67 of Smeur [**26**] but originally appeared on p. 186 v. of J. van der Scheure's 1611 edition of his 1600 *Arithmetica, oft Rekenconst, Haarlem*.

To solve this problem using the Rule of Double False Position, first let $x_0 = 36$, then $e_0 = 2x_0 + \frac{1}{2}x_0 + \frac{1}{3}x_0 + \frac{1}{4}x_0 + 6 - 80 = 37$.[11] Next, let $x_1 = 16$, then $e_1 = 2x_1 + \frac{1}{2}x_1 + \frac{1}{3}x_1 + \frac{1}{4}x_1 + 6 - 80 = -24\frac{2}{3}$. Thus, the solution to this problem is

$$x = \frac{x_0 e_1 - x_1 e_0}{e_1 - e_0} = 24.$$

The Egyptian papyri also contained rhetorical examples that represent systems of two linear equations in two unknowns, which are solved using the Rule of Double False Position. The fact that similar problems were solved using the same method in different civilizations in the same time frame provides evidence that these problems reflect the problems of that time. Although the literature suggests that each civilization independently invented the same method to solve these problems, we are of the opinion that traders carried stories, hence transferring information (such as the explanation of this process used to answer the problems that arose) along trade lines between Egypt and Babylonia. Consider the following problem.[12]

*Let a* 1 *mŭ of good field cost* 3 *hundred; and* 7 *mŭ of poor field cost* 5 *hundred. Now* 1 *qĭng field is bought together, the price is* 1 *myriad. Of the good and poor fields, how much is there each?*

This example can be written using algebraic notation as the system of linear equations

$$g + p = 100 \ (m\check{u})$$
$$300g + \frac{500}{7}p = 10000 \text{ (coins)}$$

where $g$ and $p$ represent the areas (in $m\check{u}$) of the good and poor fields, respectively.[13] To solve this problem using the Rule of Double False Position, first let $g_0 = 20$, then $p_0 = 80$, and $e_0 = 300g_0 + \frac{500}{7}p_0 - 10000 = 1714\frac{2}{3}$. Next, let $g_1 = 10$, then $p_1 = 90$, and $e_1 = 300g_1 + \frac{500}{7}p_1 - 10000 = -571\frac{3}{7}$. Thus, the solution to this problem is

$$g = \frac{g_0 e_1 - g_1 e_0}{e_1 - e_0} = 12\frac{1}{2} \quad \text{and} \quad p = 87\frac{1}{2}.$$

They cleverly eliminated one variable, in turn, reducing the system to a linear equation of the form (11). As a result, they were able to use the Rule of Double False Position to obtain the exact solution to the system of linear equations.

There is evidence that the Egyptians extended the Rule of Double False Position to quadratics but did not use the rule in an iterative manner. They performed only one step and were aware that when the problem was more complicated (e.g., quadratic), the solution they obtained using the Rule of Double False Position was only approximate. Consider the following problem.[14]

*Divide* 40 *into two numbers so that the sum of both squares is* 850.

[11] At that time, they used the initial guess, $x_0$, to calculate $2x_0 + \frac{1}{2}x_0 + \frac{1}{3}x_0 + \frac{1}{4}x_0 + 6 = 117$. Then, they evaluated $117 - 80$ to determine the corresponding error $e_0 = 37$.

[12] Taken from p. 37 of Lun [**17**].

[13] $Q\check{i}ng$ and $m\check{u}$ are units of area measure such that 1 $q\check{i}ng$ = 100 $m\check{u}$.

[14] Taken from p. 71 of Smeur [**26**] but originally appeared on p. 186 v. of J. van der Scheure's 1611 edition of his 1600 *Arithmetica, oft Rekenconst, Haarlem*. This example can also be found on p. 7 of Ma [**17**].

This example can be written using algebraic notation as the system of two equations (one of which is second order) with two unknowns,

$$x + y = 40,$$
$$x^2 + y^2 = 850.$$

To attempt the solution of this problem using the Rule of Double False Position, first let $x_0 = 30$, then $y_0 = 10$, and $e_0 = x_0^2 + y_0^2 - 850 = 150$. Next, let $x_1 = 20$, then $y_1 = 20$, and $e_1 = x_1^2 + y_1^2 - 850 = -50$. Thus, the answer to this problem obtained using the Rule of Double False Position is

$$x = \frac{x_0 e_1 - x_1 e_0}{e_1 - e_0} = \frac{45}{2} = 22.5 \quad \text{and} \quad y = \frac{35}{2}.$$

However, we see that $x^2 + y^2 = (\frac{45}{2})^2 + (\frac{35}{2})^2 = 813\frac{1}{2} \neq 850$; thus the solution is only approximate. Therefore, the application of the Rule of Double False Position to a quadratic can be viewed as taking one step of the secant method on the given quadratic. To justify this algebraically, first simplify the above system to the quadratic equation (with one unknown) as

$$f(x) = x^2 - 40x + 375.$$

Let $x_0 = 30$, so $f(x_0) = 75$. Now, let $x_1 = 20$, so $f(x_1) = -25$. After performing one step of the secant method, the approximate solution obtained is

$$x = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)} = 22.5,$$

which is the same approximate solution obtained from using the Rule of Double False Position.

The use of the Rule of Double False Position appeared in the texts of many civilizations in the centuries following. For example, the earliest surviving Chinese mathematics text, *Jiŭ Zhāng Suàn Shù* (Computational Prescriptions in Nine Chapters) [**8**], also known as *The Nine Chapters on the Mathematical Art* [**16**], dates back to the Hàn Dynasty around 200 B.C., and represents the collective efforts of many scholars over several centuries. It contains 246 problems in nine chapters, with each chapter containing practical problems connected with everyday life, their solutions, and brief descriptions of the methods used to solve them.[15] In Chapter 7 (the title, in English, translates to "Excess and Deficit"), twenty problems were solved using *yíng bù zú shu*, which literally means "too much and not enough" and can be recognized as the Rule of Double False Position [**8**]. This is the first evidence of the Rule of Double False Position being considered a general rule to be used on particular problems and given a name.

In the 9th century, the Arab mathematician Abu Jafar Mohammad ibn-Mūsa al-Khwārizmī wrote two influential books, which were translated into Latin in the 12th century and circulated throughout Europe [**2**].[16] Also in the 9th century, Abū Kāmil

[15]The purpose of *The Nine Chapters on the Mathematical Art* was similar to that of the Rhind Mathematical Papyrus—to serve as a practical handbook with problems that the ruling officials of the state were likely to encounter [**5**].

[16]Only John of Seville's 12th-century Latin translation of Al-Khwārizmī's second book, *Algorithmi de numero indorum*, "Calculation within Indian Numerals," still exists [**5**]. Latin was the *lingua franca* of the scientific world [**28**].

wrote *Kitāb fil-jabr w'al muqābalah*, 'Book of Algebra' (a commentary on, and elaboration of, Al-Khwārizmī's work). This was entirely devoted to *Hitāb al Khāta'ayn*, which literally means 'rule of the two errors', and can be recognized as the Rule of Double False Position [**8**].

The first evidence of the use of the Rule of Double False Position in 12th-century India is found in an anonymous Latin book, *Liber Augmenti et Diminutionis*, which literally means "Book of Increase and Decrease." This Latin book was translated from Arabic and presented the rule Hisab al Khāta'ayn, (which in Latin translates to "Regula Augmenti et Diminitionis," and in English translates to "Rule of Increase and Decrease") to solve a linear equation. In 12th-century India, problems were posed simply for the pleasure of solving them instead of for utilitarian function (unlike the texts that previously contained mathematics).

In 1202, Leonardo Pisano, also known as Fibonacci, wrote *Liber Abaci*, "Book of the Abacus," which contained 15 chapters dealing with arithmetic and algebra, including a mixture of Indian arithmetic methods and Arab algebraic methods.[17] In Chapter 13 of *Liber Abaci*, Fibonacci described the Arabic rule, Elchataytm (which can be recognized as the Rule of Double False Position) and referred to it as the Augmented and Diminished method. He applied this rule to rhetorical problems that represent linear equations, and 29 of these problems were reproduced with little or no change from the Kamil's Arabic "Book of Algebra" [**5, 25**].

In 1484, Chuquet completed a three-part mathematical manuscript[18] entitled "Triparty en la science des nombres." In the final section of the first part of his "Triparty," Chuquet describes what he called "the rule of two false positions", which can be recognized as the Rule of Double False Position. This is the first time an English title is given to what is now most commonly called the Rule of Double False Position.

In the 16th century, Latin names and terms were introduced to describe existing mathematical methods. In 1527, Bienewitz [**1**], also known as Petrus Apianus, introduced the term "Regula Falsi," his Latin translation of the Rule of Double False Position, and defined it as a method that "learns to produce truth from two lies" [**18**].[19] (The term "Regula Falsi" literally translates to "rule of falseness.") Bienewitz explained that the term 'false' is used because the solution is produced from two 'false' initial estimates and not because the method is wrong or false [**1**].

The fact that Bienewitz introduced the term Regula Falsi to refer to the method already known as the Rule of Double False Position (and named "Rule of Two False Positions" by Chuquet in 1484) arguably explains how the Rule of Double False Position acquired the name the Regula Falsi method. From this point on, the Rule of Double False Position was referred to as not only the Rule of Double False Position but also as Regula Falsi.[20] This marks the start of the naming confusion involving Regula Falsi, which we elaborate on later.

Recall that the Rule of Double False Position was originally defined for rhetorical examples that represent linear equations, but it was also used, in a non-iterative manner, to obtain approximate solutions to rhetorical examples that represent quadratic

[17]Fibonacci wrote *Liber Abaci* after returning from extensive travel about the Mediterranean, visiting Egypt, Syria, Greece, Sicily, and Provence, to receive a solid mathematical foundation. *Liber Abaci* was revised in 1228, circulated in manuscript form until it was printed in Italy in 1857, and was not translated into English until 2002 [**5**].

[18]According to Nordgaard [**20**], Chuquet was little known outside of France yet had a considerable influence in France. Chuquet's "Triparty" was circulated only in manuscript and was not printed until 1880. A study [**10**], published in 1985, includes an extensive translation of Chuquet's mathematical manuscript.

[19]This is information from Maas [**18**]. The original text by Bienewitz is in German.

[20]The term "Regula Falsi" came into use long before the Regula Falsi method was developed (in the 1950s) and it was used to describe the Rule of Double False Position.

equations. In 1540, Frisius [**11**] claimed that he was the first to apply the Rule of Double False Position (which he called Regula Falsi) to quadratic equations of the form $ax^2 = b$.[21] Frisius' application of (his slightly modified version of) the Rule of Double False Position was an exercise performed strictly out of theoretical interest, since by this time, algebra was known and practiced. Consider the following problem.[22]

> *From a rectangle of* 200 *square yards the length is one and a half times the width. What are the length and the width.*

This example can be written using algebraic notation as the system of two equations,

$$l \times w = 200,$$
$$l = 1\frac{1}{2}w,$$

where $l$ and $w$ represent the length and width of the rectangle, respectively. Frisius was aware that he could have used direct substitution to simplify this system to $1\frac{1}{2}w^2 = 200$, a quadratic equation in one variable, and solve for $w^2$. However, he wanted to demonstrate that it was possible to solve the problem using the Rule of Double False Position. To do this, he first let $w_0 = 4$; then the length is 6, the area is 24, and $e_0 = 1\frac{1}{2}w_0^2 - 200 = -176$. Next, he let $w_1 = 20$, which resulted in a length of 30, an area of 600, and $e_1 = 1\frac{1}{2}w_1^2 - 200 = 400$.

At this step, Frisius modified the Rule of Double False Position. Instead of solving for $w$ as

$$\frac{w_0e_1 - w_1e_0}{e_1 - e_0},$$

he evaluated $w^2$ as

$$\frac{w_0^2e_1 - w_1^2e_0}{e_1 - e_0}$$

and took the square root of this result. (At the time, knowledge of square and cube roots was known and root tables existed for quick reference.) Frisius calculated the width to be $11\frac{27}{55}$ and the length to be $15\frac{77}{100}$; however, $11\frac{27}{55} \times 15\frac{77}{100} \approx 181 \neq 200$. Frisius knew, as he stated, that the Rule of Double False Position is correct only for linear equations. He realized that the solution he attained from using his modified Rule of Double False Position on his example problem described above was only approximate, since he stated that it is impossible to get the exact answer using this method.

**4. NAMES ASSOCIATED WITH THE RULE OF DOUBLE FALSE POSITION.** Even though the Rule of Double False Position dates back to the 18th century

[21] In 1525, German mathematician Christoff Rudolff wrote the first German algebra book *Die coss*, which means "the variable," where the Rules of Coss (where $ax^2 = b$ represents the second Rule of Coss, $ax^3 = b$ represents the third Rule of Coss, etc.) are presented and the modern symbol for the square root is introduced. Gemma Frisius made his claim in response to Rudolff's comment that it was impossible to solve the second, third, and fourth Rules of Coss using Rules of False (which consist of the Rules of Single and Double False Position). Frisius solved the third and fourth Rules of Coss using a modification of the Rule of Single False Position.

[22] Taken from p. 72 of Smeur [**26**] but originally appeared in Frisius [**11**].

B.C., it was not thought of as a general rule or a method at that time, and therefore was not given a specific name. It was not until 200 B.C., in China, that it was considered a general rule and given a name—*yíng bù zú shu*. Since then, it has been given different names (see Table 2) but has most commonly been referred to as the Rule of Double False Position since the 11th century A.D.

**Table 2.** Evolution of the naming of the Rule of Double False Position.

| Country | Century | Rule name |
|---|---|---|
| Egypt | 18th B.C. | — |
| Babylonia | 18th B.C. | — |
| China | 2nd B.C. | yíng bù zú (too much and not enough) |
| Arab | 9th A.D. | hisab al-Khataayn (rule of two errors) |
| Europe | 11th A.D. | elchataym (two errors) |
| Africa | 13th A.D. | method of scales |
| Europe | 15th,16th A.D. | rule of two false positions/regula falsi/ rule of double false position/regula positionum |
| U.S. | 20th A.D. | rule of double false position/method of false position/ regula falsi/secant method |

**5. THE END OF THE FIRST PATH TO THE SECANT METHOD.** In 1545, Cardano [**7**], in his *Artis Magnae*, demonstrated that the Rule of Double False Position (calling it "De Regula Liberae Positionis," which literally translates to "(Concerning) the rule of free position") could be used as an iterative procedure.[23] He described the rule as an iterative process, where multiple steps must be performed in order to improve the approximation [**6**]. He solved quadratic and cubic equations using the rule and included explanations of how he solved the problems using the rule [**2**]. We have now traveled the path from the Rule of Double False Position to what we know today as the secant method for a nonlinear equation. Cardano called it "De Regula Liberae Positionis." Of course, this awkward name never achieved acceptance in the literature.

Before we leave this section, we make a few comments. Newton was not fond of publishing his work, including the origin of the calculus. However, he kept a fairly detailed notebook of his scientific and mathematical ideas. These unpublished papers, dated early 1665, remained in the possession of the family estate until 1872, when the fifth Earl of Portsmouth donated many of the papers to Cambridge University, where they still reside. Whiteside's collection of Newton's unpublished notes, entitled *Newton's Waste Book* [**30**], includes in Volume I (covering the period 1664–1669) an illustration of a geometric technique based on similar triangles that Newton used to approximate a zero of a nonlinear equation. Newton's technique (which he did not refer to by any name) is equivalent to the Rule of Double False Position applied to a nonlinear equation. We point out that in describing both methods, Newton did not mention iteration, even though by this time Cardano's work on iterating the Rule of Double False Position was over one hundred years old. We believe that if Newton had been familiar with Cardano's work, then he would not have proposed his version of the Rule of Double False Position. Moreover, we believe that Newton did not discover

[23] A copy of the original Latin text was made by scanning microfiche.

the secant method by taking a finite difference approximation to the derivative as discussed earlier. This will be revisited in §6. In formulating what today we call Newton's method, Newton did not state the method in terms of the formal derivative, although it was undoubtedly known to him at the time. Newton worked with polynomial equations and presented the correction term as an algebraic quantity, which would be the same as we would obtain using the correction term defined in (1) in terms of the derivative. He did consider several equations that involved trigonometric expressions. Here, he cleverly treated the trigonometric terms as infinite polynomials by replacing them with their power series expansions. To what extent Newton knew that his correction term could be stated in terms of the derivative is open to speculation. We believe that he knew, since he often gained motivation from drawing pictures. An equally profound question is to ask to what extent Newton understood the notion of iteration. We know that he did not iterate either of the two proposed algorithms in his writings. One school of thought is that, since the Rule of Double False Position was initially applied to a linear equation and did not have to be iterated, the notion of iteration was sufficiently foreign that it took Cardano to hammer it home as a break in mathematical algorithmic tradition. Another school of thought is that, once the first iteration is defined, it is a rather straightforward realization that the procedure could be continued in the obvious fashion. Perhaps the truth is somewhere in between, with individuals belonging to each school. However, it is our considered opinion that Newton belonged to the latter school and believed that iteration was an obvious extension of his methods. Hence, if we credit Newton with the so-called Newton's method, even though he did not specifically state iteration in its description, then it is fair to say that Newton discovered the secant method independently, around the same time that he proposed what we today call Newton's method.

**6. NAMING AND CONVERGENCE RATE FOR THE SECANT METHOD.** Thomas Fincke [**9**] introduced the word "secant" in his 1583 treatise on geometry [**2**]. The word "secant" is from the Latin root "secare", which means to cut. This term makes sense, since Fincke depicted cutting a circle. It was in the same treatise that Fincke introduced the secant line.

In 1958, T. A. Jeeves [**15**] seems to be the first to use the term "secant method" to refer to the algorithm under discussion. Jeeves explained that it is "the secant modification of Newton's method". Jeeves also presented the first proof that we can locate of the golden mean convergence rate of the secant method (in one dimension). He proved that at each iteration of the secant method, the increase in the number of significant digits is $\frac{1}{2}(1+\sqrt{5}) \approx 1.62$ (the golden mean) times the previous increase.

**7. A SECOND PATH TO THE SECANT METHOD.** Clearly, the second path that led to the contemporary secant method is through obtaining the secant method as a modification of Newton's method. Our historical journey would not be complete without attempting to determine who first proposed such a modification of Newton's method and who first realized that the two paths—the Cardano Rule of Double False Position path and the modified Newton's method path—led to the same destination, the contemporary secant method.

The answer to these two questions can be found in the literature associated with the naming of the secant method. Jeeves included a footnote referencing the previous work of Wegstein, who in a 1958 paper [**29**], referred to what Jeeves called the secant method as a "modified form of Newton's method" and explained that this method was contained implicitly in Willers' 1948 book [**31**] as "the method of false position." Willers described only the first iteration, so we do not know exactly what he meant by

his method of false position. However, in doing so he captured the guiding principle of the secant method, i.e., two-point linear interpolation. In summary, Wegstein in 1958 stated that the secant method was a modified form of Newton's method and that it was contained implicitly in Willer's book as the method of false position. It follows then that we should credit Wegstein for connecting our two paths to the secant method, understanding that this connection is somewhat sketchy.

**Table 3.** Some examples of the inconsistencies in naming the Regula Falsi method, the Modified Regula Falsi method, and the secant method.

| Date | Authors | Regula Falsi (R.F.) | Modified R.F. | Secant Method |
|---|---|---|---|---|
| 1944 | Whittaker and Robinson | rule of false position | | |
| 1948 | Willers | | method of false position | |
| 1953 | Householder | | regula falsi | |
| 1955 | Booth | regula falsi/ rule of false position | | |
| 1958 | Wegstein | | | modified form of Newton's method |
| 1958 | Jeeves | | | secant method |
| 1960 | Ostrowski | regula falsi | | iteration with successive adjacent points |
| 1961 | Stanton | | regula falsi | |
| 1962 | Hochstrasser | rule of false position | | |
| 1964 | Traub | | regula falsi | secant iteration function (I.F.) |
| 1964 | Henrici | | | regula falsi |
| 1964 | Fröberg | | | regula falsi |
| 1966 | Isaacson and Keller | classical regula falsi method | | method of false position |
| 1970 | Ortega and Rheinboldt | regula falsi | | secant method |
| 1972 | Blum | | | method (or rule) of false position/ regula falsi |
| 1974 | Dahlquist and Björck | | regula falsi | |
| 1975 | Smeur | rule of false | | |
| 1977 | Gellert, Hellwich Küstner and Kästner | fixed point method | secant method | method of false position |
| 1978 | Atkinson | | regula falsi | |
| 1981 | Gill, Murray and Wright | regula falsi/ method of false position | | secant method/ method of linear interpolation |
| 2008 | Dahlquist and Björck | false-position method/ regula falsi | | |

**8. NAMING CONFUSION.** Table 3 represents some of the different names that have been used to describe the Regula Falsi method, the Modified Regula Falsi method, and the secant method. The blank spaces in the table indicate that either the method was not presented, or the method was not referred to by a specific name by that particular author(s).

References to the actual Rule of Double False Position became a part of the naming confusion. In 1978, Smeur [**26**] described the Rule of Double False Position, explained that it is called Regula Falsi or Rule of False, and stated that the rule is only correct for linear equations.[24] In 1991, Hämmerlin and Hoffman [**12**] stated that the Regula Falsi method was one step of the secant method and that the secant method was a result of iterating the Regula Falsi method. It seems to be implicit that they understood that the Rule of Double False Position (which they called Regula Falsi, the name originally introduced by Bienewitz) was used for linear equations, in turn implying that the secant method was used for nonlinear equations and was iterated.

We have shown that the terms Regula Falsi and Rule of False Position have been used interchangeably to describe the Regula Falsi method and the Modified Regula Falsi method, as well as the secant method. In addition, we presented some of the many inconsistencies in the naming of the Regula Falsi method, the Modified Regula Falsi method, and even the secant method. Our respect for the authors mentioned is such that, while collectively they created mass confusion in the naming of these various methods, we believe that each must have been working within an implicit understanding. We think that they viewed a "false position" as an initial approximation to the solution, and "Regula Falsi" or "Rule of False Position" as any method that uses linear interpolation based on two false positions to obtain a new approximation to the solution. There is naming consistency within this understanding; however, the confusion unfortunately remains. We must admit that in recent years, primary sources have been overlooked. As a result, contemporary usage is as follows. The Modified Regula Falsi method described in §2.4 is now what current mathematics texts and popular websites call the Regula Falsi method, and it is presented in sections on bracketing methods because it is a natural extension of the method of bisection (which is a bracketing method). Furthermore, the Regula Falsi method described in §2.3 is ignored in current mathematical texts. However, today everyone calls the secant method the secant method.

**ACKNOWLEDGMENTS.** The authors would like to thank three anonymous referees, and our colleague Michael Trosset, for suggestions that greatly improved the paper.

REFERENCES

1. P. Apianus, *Eyn Newe Unnd wolgegründte underweysun aller Kauffmanss Rechnung*, Ingolstadt, 1527.
2. W. P. Berlinghoff, F. Q. Gouvêa, *Math through the Ages—A Gentle History for Teachers and Others*, Expanded Edition, Oxten House and Mathematical Association of America, Washington, DC and Farmington, ME, 2004.
3. A. D. Booth, *Numerical Methods*, Academic Press, New York, 1955.
4. G. J. Bronson, *C++ for Engineers and Scientists*, PWS Publishing, University of Michigan, 1999.
5. D. M. Burton, *The History of Mathematics: An Introduction*, sixth edition. McGraw Hill, New York, 2007.
6. M. V. Cantor, *über Geschichte der Mathematik*, Leipzig, 1900.
7. G. Cardano, *Artis magnae, sive de regulis algebraicis (Ars magna)*, Nuremberg, 1545.
8. J.-L. Chabert, *A History of Algorithms: From the Pebble to the Microchip*, Springer-Verlag, Italy, 1998.

[24] Smeur stated that this rule could be found in the 1537 Dutch book by G.V. Hoecke. Like Bienewitz, Hoecke and Smeur both use the term Regula Falsi to refer to the Rule of Double False Position.

9. T. Fincke, *homae Finkii Flenspurgensis Geometriae rotundi libri XIIII*, 1583.
10. G. Flegg, C. Hay, B. Moss, *Nicolas Chuquet, Renaissance Mathematician*, D. Reidel Publishing Company, Dodrecht/Boston/Lancaster, 1985.
11. G. Frisius, *Arithmeticae Practicae Methodus Facilis*, Antwerp, 1540.
12. G. Hämmerlin, K.-H. Hoffmann, *Numerical Mathematics*, Springer-Verlag, New York, 1991.
13. A. S. Householder, *Principles of Numerical Analysis*, McGraw Hill, New York, 1953.
14. J. Høyrup, *Algebra and Native Geometry: An Investigation of Some Basic Aspects of Old Babylonian Mathematical Thought*, Altorientalische Forschungen, **17** (1990).
15. T. A. Jeeves, *Secant Modification of Newton's Method*, Westinghouse Research Laboratories, New York, 1958.
16. S. Kangshen, J. N. Crossley, A. W.-C. Lun, *The Nine Chapters on the Mathematical Art: Companions and Commentary*, Oxford University Press and Science Press, Beijing, New York, 1999.
17. L. Ma, *The Rule of False: Early Applications and Conjectured Transmissions*, Technical Paper 1993-15, Chalmers University of Technology, University of Götebrog, 1993.
18. C. Maas, *Was ist das Falsche an der Regula Falsi?*, Mitteilungen der Mathematischen Gesellschaft in Hamburg 11, **3** (1985).
19. O. E. Neugebauer, *Mathematical Cuneiform Texts*, American Oriental Society, New Haven, CT, **29** (1974).
20. M. A. Nordgaard, *A historical survey of algebraic methods of approximating the roots of numerical higher equations up to the year 1819*, second edition, Teachers College, Columbia University, New York, 1922.
21. National Council of Teachers of Mathematics, *Historical Topics for the Mathematics Classroom*, The National Council of Teachers of Mathematics, Reston, VA, 1998.
22. T. E. Peet, *The Rhind Mathematical Papyrus*, Hodder and Stoughton, New York, 1923.
23. G. Robins, C. Shute, *The Rhind Mathematical Papyrus: An Ancient Egyptian text*, British Museum Publications, London, 1987.
24. E. Robson, *Mathematics in Ancient Iraq: A Social History*, Princeton University Press, Princeton, NJ, 2008.
25. L. Sigler, *Fibonacci's Liber Abaci: A Translation into Modern English of Leonardo Pisano's Book of Calculation*, Springer, New York, 2002.
26. A. E. Smeur, *The Rule of False Applied to the Quadratic Equation, in Three Sixteenth Century Arithmetics*, Archives Internationales d'Histoire des sciences, **28** (1978).
27. R. G. Stanton, *Numerical Methods for Science and Engineering*, Prentice Hall, Englewood Cliffs, NJ, 1961.
28. D. J. Thomas, *Joseph Raphson, Fellow of the Royal Society*, Notes and Records of the Royal Society of London, July **44** no. 2 (1990).
29. J. H. Wegstein, *Accelerating Convergence of Iterative Processes*, Communications of the ACM, Notes and Records of the Royal Society of London, **1** (1958).
30. D. T. Whiteside, *The Mathematical Papers of Isaac Newton, Volumes I-VII*, Communications of the ACM, Cambridge University Press, Cambridge, 1967-1976.
31. F. A. Willers, *Practical Analysis: Graphical and Numerical Methods*. A Translation by Robert T. Beyer, Dover Publications, New York, 1948.

**JOANNA M. PAPAKONSTANTINOU** received her B.A., M.A., and M.A.T., as well as her Ph.D., from Rice University, under the direction of the second author. After completing her postdoctoral work at Rice University in the Computational and Applied Mathematics Department, she joined PROS Revenue Management as a Senior Associate where she served as a science expert in the Center of Excellence. Currently, she works as the Senior Science Consultant at Advanous. She remains involved in research in the field of optimization, writes curriculum and teaches, and actively participates in outreach activities.
*Advanous, 2425 Dorrington Unit A, Houston, TX 77030*
*joanna.papa@gmail.com*

**RICHARD A. TAPIA,** 2011 recipient of the National Medal of Science, holds the rank of University Professor in the Rice Department of Computational and Applied Mathematics. He is also the Director of the Rice University Center for Excellence and Equity in Education. Due to Tapia's efforts, Rice has received national recognition for its educational outreach programs and has become a national leader in producing women and underrepresented minority Ph.D. recipients in the mathematical sciences.
*Rice University, 6100 Main Street, Houston, TX 77005*
*rat@rice.edu*

## Another Proof of Young's Inequality for Products

Several different proofs of Young's inequality can be found in numerous articles and textbooks. An elementary proof for the generalized form is presented here in two steps. We first state the generalized Young's inequality on $R^n$. Let $p_1, p_2, \ldots, p_n$ be real numbers such that $p_k > 1$ for all $k = 1, 2, \ldots, n$, and $\sum_{k=1}^{n} \frac{1}{p_k} = 1$. For nonnegative real numbers $a_1, a_2, \ldots, a_n$, we have

$$\sum_{k=1}^{n} \frac{a_k^{p_k}}{p_k} \geq a_1 a_2 \ldots a_n.$$

**Step 1.** For any given $n$ positive integers $m_1, m_2, \ldots, m_n$ with $\mathrm{GCD}(m_1, m_2, \ldots, m_n) = 1$, we define $M = \sum_{k=1}^{n} m_k$ and $p_k = \frac{M}{m_k}$. Hence, we have

$$p_k > 1 \text{ for every } k = 1, 2, \ldots, n; \text{ and } \sum_{k=1}^{n} \frac{1}{p_k} = 1. \tag{1}$$

For any $n$ nonnegative numbers $a_1, a_2, \ldots, a_n$, we have

$$\sum_{k=1}^{n} \frac{a_k^{p_k}}{p_k} = \frac{1}{M} \sum_{k=1}^{n} m_k a_k^{p_k} = \frac{1}{M} \left[ \underbrace{(a_1^{p_1} + a_1^{p_1} + \cdots + a_1^{p_1})}_{m_1\text{-times}} + \cdots + \underbrace{(a_n^{p_n} + a_n^{p_n} + \cdots + a_n^{p_n})}_{m_n\text{-times}} \right].$$

By using the inequality

$$\text{Arithmetic Mean} = \text{A} \geq \text{G} = \text{Geometric Mean}$$

we will get Young's inequality for rational numbers:

$$\sum_{k=1}^{n} \frac{a_k^{p_k}}{p_k} \geq \left[ \prod_{k=1}^{n} \left(a_k^{p_k}\right)^{m_k} \right]^{1/M} = \prod_{k=1}^{n} a_k. \tag{2}$$

**Step 2.** To prove this inequality for any real numbers $p_k$ satisfying (1), we take sequences of rational $p_{j,k}$, $j = 1, 2, \ldots$, converging to $p_k$ for those irrational $p_k$'s. Inequality (2) holds for $p_{j,k}$'s for each $j = 1, 2, \ldots$ and $k = 1, 2, \ldots, n$. By taking the limit of both sides as $j$ goes to infinity, the continuity of the function $f(x) = c^x/x$, for any nonnegative constant $c$ and $x > 1$, gives us the generalized form of Young's inequality for any $p_k > 1, k = 1, 2, \ldots, n$. And $n = 2$ gives us the more standard form of Young's inequality

$$\frac{a^p}{p} + \frac{b^q}{q} \geq ab,$$

where $p > 1, q > 1, 1/p + 1/q = 1$, and $a, b$ are nonnegative numbers.

—Submitted by M. Reza Akhlaghi,
Big Sandy Community & Technical College

http://dx.doi.org/10.4169/amer.math.monthly.120.06.518
MSC: Primary 43A15