

Mathematical Foundations of the Slide Rule

Joseph Pasquale

Department of Computer Science and Engineering
University of California, San Diego

June 26, 2011

Abstract

We present a mathematical principle that provides a precise formulation for the types of functions a slide rule can be designed to calculate. In addition to providing the mathematical foundations for how and why a slide rule works, this principle establishes its mathematical possibilities and limits. We also present the theory of how slide rule scales are constructed.

Introduction

The slide rule works because of a very basic property of Euclidean geometry, which is that line lengths are additive: Given two consecutive line segments \overline{AB} and \overline{BC} that are aligned, the sum of their lengths equals the length of their sum, i.e., $\text{length } \overline{AB} + \text{length } \overline{BC} = \text{length } \overline{AC}$, where $\overline{AC} = \overline{AB} + \overline{BC}$. This simple property, along with the ability to establish different scales on these line segments that can represent a variety of single-variable functions, lead to a surprisingly large class of multi-variable functions that can be calculated on a slide rule. An interesting question is: What types of functions are possible? We answer this question by developing a principle for their general form.

While Oughtred's invention of the slide rule dates to the early 1600's [1], a general mathematical theory of the slide rule came much later. The basis for such a theory can be found in d'Ocagne's development of nomography in the late 1800's [2], which focused on the design of graphical charts and scales for various types of calculation. A general mathematical formulation of slide rule scales and calculation was given by Runge in his lecture notes on graphical methods in the early 1900's [3]. The key part of this formulation was captured in what he described as a "Principle of the Slide Rule," on which our work is based. Stokes also described a "Mathematical Principle of the Slide Rule" that was a generalization of Runge's formulation [4].

Later descriptions of a "Principle of the Slide Rule," especially those that appeared in many slide rule books and manuals, lost the generality of Runge's and Stoke's formulations. They focused mainly on the property of logarithms (that $\log x + \log y = \log xy$) and how it could be exploited to perform multiplication and other calculations on a slide rule. This is understandable as most slide rules have logarithmic scales.

Shortly after the work of d'Ocagne and Runge, Lipka developed a more complete formulation of "graphical and mechanical computation" [5]. Later refinement and streamlining of methods in graphical calculation, with applications to the slide rule, can be found in the works of Davis [6], Mackey [7], and Hoelscher [8].

An earlier version of this paper appeared in the Proceedings of the International Meeting for Collectors of Historical Calculating Instruments (IM 2011), Cambridge, MA, Sept. 23-25, 2011, pp. 1-8, published by the Oughtred Society (www.ough tred.org).

Slide Rule Scales

A slide rule has three basic parts: a body, a slide, and an indicator. On the body and slide are scales that are parallel to each other. These scales represent various functions, and by moving the slide relative to the body, a calculation involving multiple functions can be performed, with the indicator allowing corresponding values on the various scales to be matched so that intermediate and final results can be set and read. While there are a wide variety of different types of slide rules – linear, circular, cylindrical, multi-slide, general vs. special purpose, etc. – the concepts we describe remain essentially the same for all slide rules.

To see how multiple scales are used to perform a calculation, we need to first understand what a scale is and how it is constructed. A scale is simply a finite line with graduated marks, each corresponding to a value x and located at a distance given by $f(x)$, called the *function of the scale*, relative to an origin x_0 . An example of a scale is shown in Figure 1.

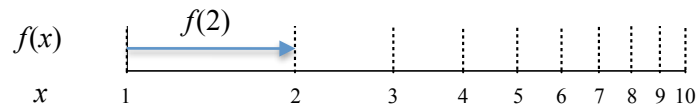


Figure 1: Example of a scale. Taking the origin to be $x_0 = 1$ and locating it at the left end of the scale's line, then $f(2)$ is the distance from the left end to $x = 2$.

Note that $x_0 = f^{-1}(0)$ by definition, i.e., the *value* of the origin x_0 is that for which $f(x_0) = 0$. The actual *location* of the origin on the line, i.e., the mark corresponding to the value x_0 , may be chosen for convenience. It is typically located at the left end of the line, but in general, it may be located anywhere on it, even on a projection of the line beyond its edges. We will assume that the origin is at the left end unless indicated otherwise.

Let us consider some examples. For the identity function $f(x) = x$, we would have a line marked 0 at the left end and uniformly spaced marks labeled with the numbers 1, 2, 3, ... such that 1 is at distance 1, 2 at distance 2, 3 at distance 3, and so on. For the square function $f(x) = x^2$, again we would have a line marked 0 at the left end, but with marks labeled 1, 2, 3, 4, ... located at distances 1, 4, 9, 16, ... respectively.

The origin need not necessarily correspond to the value $x_0 = 0$; the only requirement is that $f(x_0) = 0$. Thus, taking as an example the inverse function $f(x) = 1/x$, $x_0 = \infty$ (a result of $f(x_0) = 0$), and so the line is marked ∞ at the left end, 8 at distance $1/8$, 4 at distance $1/4$, 2 at distance $1/2$, 1 at distance 1, and so on. For the logarithmic function $f(x) = \log x$ (using common, or base 10, logs), $x_0 = 1$, thus locating 1 at the left end (since $\log 1 = 0$), 2 at distance 0.301 (since $\log 2 \approx 0.301$), 3 at distance 0.477, and more generally, number x at distance $\log x$. The scale shown in Figure 1 is that of $f(x) = \log x$.

Note that our notion of distance has direction. Adopting the convention that distance increases from left to right on a horizontal line, if a value x is located to the right of the origin, then it is a positive distance away; to the left of the origin, a negative distance away. Consequently, in constructing a scale, it is important that distance does indeed increase from left to right when determining locations of values. In the example for $f(x) = 1/x$, the values for x *decrease* from left to right ($\infty, 8, 4, 2, 1, \dots$) on the scale so that their distances, given by $1/x$, *increase* from left to right (0, $1/8, 1/4, 1/2, 1, \dots$).

As an example where the origin is beyond the end of the line of the scale, consider $f(x) = 1 - \sin x$ over the domain $x = 0$ to 45 . To ensure that distance increases from left to right, the left end is assigned the value 45 so that x values will decrease from 45 to 0 , resulting in the following sequence of $(x, f(x))$ pairs: $(45, 0.293)$, $(30, 0.5)$, $(15, 0.741)$, $(0, 1)$. While distance does indeed increase (from 0.293 to 1), the leftmost value, 45 , is not located at distance 0 . Consequently, the origin (whose value is $x_0 = 90$ since $f(90) = 0$) is not at the left end of the scale, but beyond it to its left. Note that this is not a problem as the theory we develop allows the origin to be anywhere.

From Two Dimensions to One

This way of representing a function effectively compresses a two-dimensional graph of a function into a one-dimensional scale, i.e., a line of graduated marks. Consider the logarithmic function, $f(x) = \log x$. Its two-dimensional and one-dimensional graphical representations are shown in Figure 2.

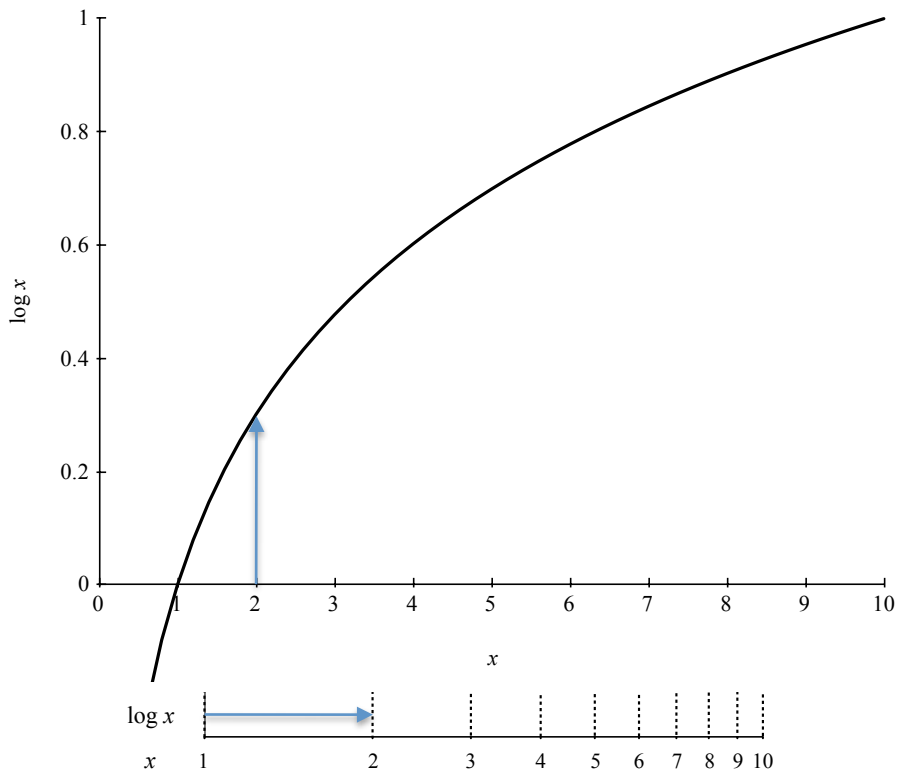


Figure 2: The upper two-dimensional graph plots the function $y = \log x$. The lower one-dimensional graph is a scale for $\log x$. Notice that the vertical rise at $x = 2$ in the upper graph equals the horizontal distance to $x = 2$ in the lower graph.

The key property tying the two representations is that the vertical distance relative to the x -axis over which the curve rises at a certain value x in the two-dimensional graph is the same as the distance from the origin x_0 (e.g., the left end of the line) to the same value x in the one-dimensional graph. It is this ingenious way of graphically representing functions that allows so much information to be packed in a small amount of space on the slide rule, and allows these functions to be part of calculations through the simple translational movement of the slide.

There is a major constraint that results from this compression. The function *must be monotonic*, at least over the domain of values x represented on the scale, so that each distance measured by $f(x)$ corresponds to a unique value x . Examples include all of the functions mentioned so far: $f(x) = x$, $f(x) = x^2$, $f(x) = 1/x$, and $f(x) = \log x$. To illustrate the problem that arises if the function is not monotonic, consider $f(x) = \sin x$ over the domain $x = 0$ to 180 . Since $\sin x = \sin (180 - x)$, *two* values correspond to a single distance, e.g., 45 and 135 at the distance 0.707 since $\sin 45 = \sin 135 \approx 0.707$. This can be resolved by simply using two scales, each over a portion of the domain where the function is monotonic. Some slide rules solve this by using a single marked line for the scale, but with two values per mark distinguished by using color-coded labels, thus interpreting the single line as two scales.

Scale Construction

To construct a scale, it is useful to draw the two dimensional graph of $f(x)$, using the left-end and right-end domain values as the extremes of the x -axis. For these and intermediate x values, vertical lines are drawn to the curve corresponding to $f(x)$. Where the vertical lines meet the curve, horizontal lines are drawn to the y -axis. Marks are placed where the horizontal lines meet the y -axis, and are labeled with the corresponding x values. The portion of the y -axis with these marks then becomes a slide rule scale.

For $f(x) = x^2$, this is shown in Figure 3, with the added vertical and horizontal lines drawn as dotted. The x -axis domain labels are shown on the right side of the y -axis where the corresponding horizontal lines meet it.

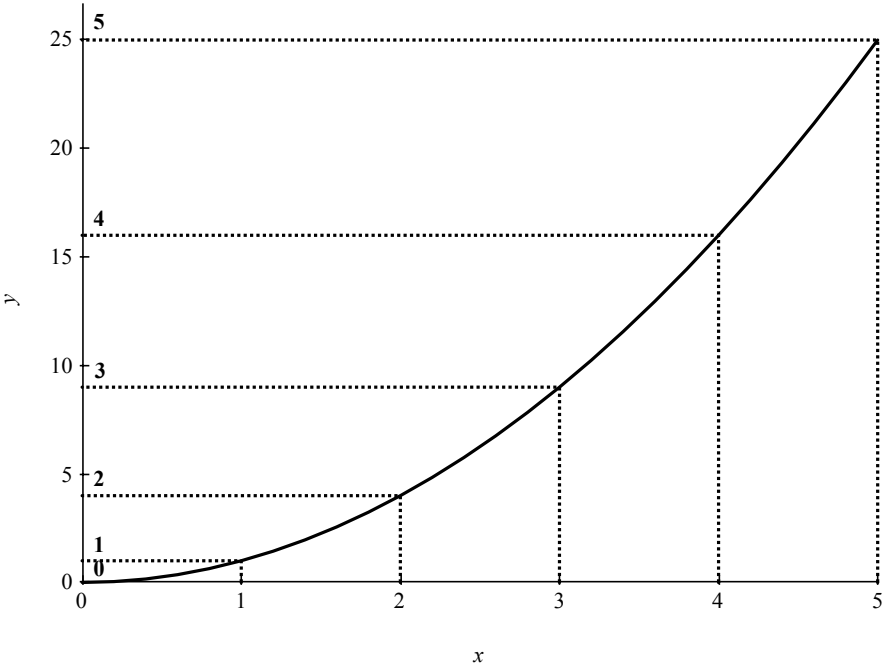


Figure 3: A two-dimensional graph of $f(x) = x^2$, with x values graphically located to the y -axis.

Next, the graph is flipped about its diagonal so the x -axis is vertical and y -axis is horizontal, resulting in Figure 4.

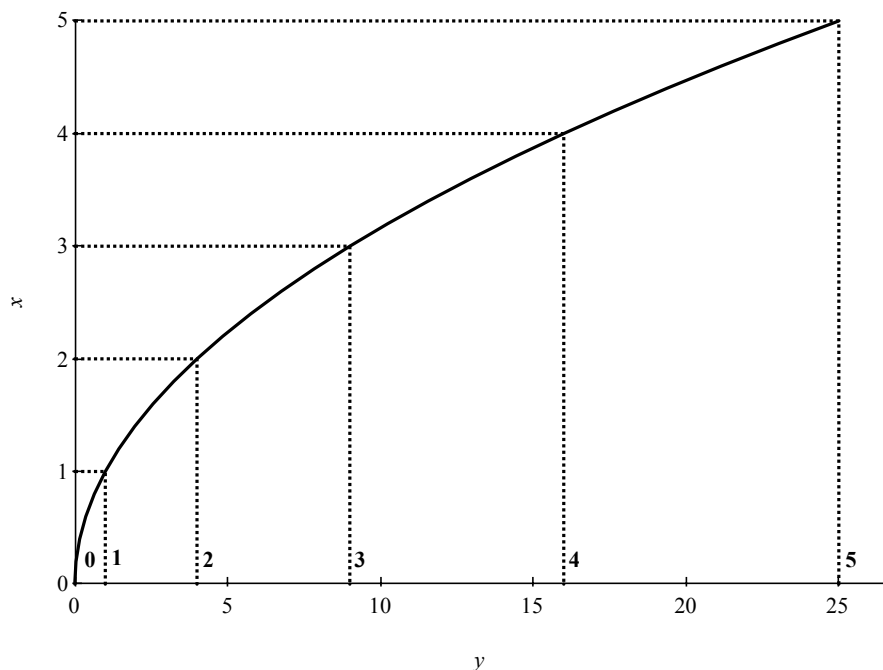


Figure 4: The two-dimensional graph of $f(x) = x^2$ flipped about its diagonal.

Finally, keeping only the portion of the graph that contains the horizontal axis, we get Figure 5.



Figure 5: The resulting scale for $f(x) = x^2$.

This, then, is the compressed one-dimensional graph of, or scale for, $f(x) = x^2$, which gives the distance from the origin at the left end to a mark corresponding to x . As a check, we can visually observe that the distance between consecutive marks is the difference between the squares of their labeled values. For example, the distance from 0 to 1 is 1 (the result of $1^2 - 0^2$), the distance from 1 to 2 is 3 (the result of $2^2 - 1^2$), the distance from 2 to 3 is 5 (the result of $3^2 - 2^2$), and more generally, the distance from x to $x + 1$ is $2x + 1$ (the result of $(x + 1)^2 - x^2$).

Calculating using Scales

If we take two x^2 scales and place them side by side and parallel to each other, and allow one to slide relative to the other along the same axis, we can calculate the length of a right triangle's hypotenuse given sides x and y , which is $\sqrt{x^2 + y^2}$. For example, to calculate $\sqrt{3^2 + 4^2}$, we would configure the scales as shown in Figure 6.

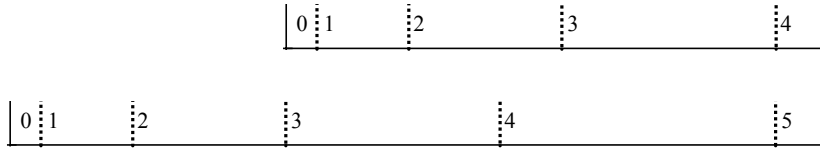


Figure 6: Calculating $\sqrt{3^2 + 4^2} = 5$ using two x^2 scales.

Since we are adding the distance from 0 to 3 on the lower scale, to the distance from 0 to 4 on the upper scale, the total distance coincides with the distance from 0 to 5 on the lower scale, which gives the solution, $\sqrt{3^2 + 4^2} = 5$.

This example demonstrates how a slide rule works based on the geometric property that line lengths are additive: $\text{length } \overline{AB} + \text{length } \overline{BC} = \text{length } (\overline{AB} + \overline{BC})$. It also shows that scales do not have to be logarithmic to work.

Mathematical Principle of the Slide Rule

We now consider the general form of a calculation with a pair of scales defined by the functions $f(x)$ and $g(y)$. By placing one scale next to another, we arrive at the following key property:

$$f(x'') - f(x') = g(y'') - g(y') \quad (1)$$

where x' and y' are vertically aligned, as are x'' and y'' , as shown in Figure 7. This property appeared in [3] (in a slightly different form) and in [4].

By allowing one scale to *slide*, i.e., be repositioned, relative to the other, the relative locations of the origins of the upper and lower scales can change, so there is no fixed relationship between $f(x')$ and $g(y')$, nor between $f(x'')$ and $g(y'')$. And yet, the property stated in (1) is key because it remains invariant, regardless of where the sliding scale is set relative to the other.

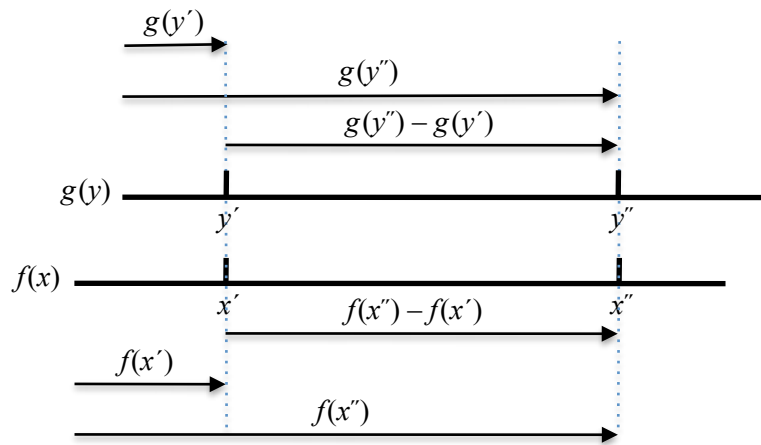


Figure 7: Two sliding scales defined by the functions $f(x)$ and $g(y)$ and marked with values x' , x'' , y' , and y'' , with the property that $f(x'') - f(x') = g(y'') - g(y')$.

We can take advantage of the ability of sliding one scale relative to the other, as this allows us to effectively select the values of any three of x' , x'' , y' , and y'' , which will determine the fourth. For example, say that $f(x) = x^2$ and $g(y) = y^2$. To calculate $\sqrt{3^2 + 4^2}$, we would let $x' = 3$, $y' = 0$, and $y'' = 4$, with the result that $x'' = 5$, just as was shown in Figure 6.

Equation (1) leads to the following main result.

Mathematical Principle of the Slide Rule: A slide rule with two scales that are defined by the functions $f(x)$ and $g(y)$ can calculate any function of the form:

$$h(x, y, z) = f^{-1}(f(x) + g(y) - g(z)) \quad (2)$$

The form of (2) is quite versatile in determining the types of functions that can be calculated. We begin with the simplest: $x + y$, i.e., addition of two numbers. Defining $f(x) = x$ and $g(y) = y$, we get the desired result: $h(x, y, 0) = f^{-1}(f(x) + g(y) - g(0)) = f^{-1}(x + y - 0) = x + y$. To calculate the hypotenuse of a right triangle with sides x and y , we define $f(x) = x^2$ and $g(y) = y^2$: $h(x, y, 0) = f^{-1}(f(x) + g(y) - g(0)) = f^{-1}(x^2 + y^2 - 0) = \sqrt{x^2 + y^2}$. To calculate the resistance of two resistors R_1 and R_2 in parallel, we define $f(x) = 1/x$ and $g(y) = 1/y$: $h(R_1, R_2, \infty) = f^{-1}(f(R_1) + g(R_2) - g(\infty)) = f^{-1}(1/R_1 + 1/R_2 - 0) = 1/(1/R_1 + 1/R_2)$.

Logarithmic Scales and Slide Rule Calculations

We have purposely avoided using logarithms in the above examples to emphasize the generality of (2) and that it does not depend on their properties. However, most slide rules have logarithmic scales. The reason is that by using logarithmic functions for $f(x)$ and $g(y)$, a wide range of very useful calculations can be performed (especially when used in conjunction with additional scales that are functions or inverse functions of $f(x)$ and $g(y)$, which we discuss below).

For example, consider multiplication of two numbers x and y . The goal is to express $h(x, y, z)$ so that it equals xy . This can be done by using logarithms, taking $f(x) = \log x$ and $g(y) = \log y$, so that $h(x, y, 1) = f^{-1}(f(x) + g(y) - g(1)) = f^{-1}(\log x + \log y - \log 1) = 10^{\log x + \log y} = xy$. On a slide rule, the C and D scales, corresponding to $g(y)$ and $f(x)$, respectively, are typically used for multiplication.

We can also exploit the fact that functions $f(x)$ and $g(y)$ can be different. Consider raising x to the y power, i.e., x^y . By defining $f(x) = \log \log x$ and $g(y) = \log y$, we obtain the desired result: $h(x, y, 1) = f^{-1}(f(x) + g(y) - g(1)) = f^{-1}(\log \log x + \log y - \log 1) = 10^{y \log x} = x^y$. Leaving the third parameter z as a variable produces $h(x, y, z) = x^{y/z}$. On a “log log” slide rule, the LL scales, corresponding to $f(x) = \log \log x$ over a number of continuous and non-overlapping domains, and the C scale corresponding to $g(y) = \log y$, are used for x^y .

Functions of Scales

Many of the other scales on a general-purpose slide rule can act as either $f(x)$ or $g(y)$ in other types of calculations. They can also act as functions of the C or D scales (or even other scales), or as inverse functions, thus allowing the C or D scales to be functions of them. When used in this way, the C and D scales act as *reference scales* with respect to these other scales. The mathematical formulas that capture these other scales when they act as functions or inverse functions of their reference scales are as follows.

Consider the case where a scale $z(w)$ is a function $u(x)$ of the domain values of reference scale $f(x)$. The goal is to find a formula for $z(w) = f(x)$ such that $w = u(x)$. The result is $z(w) = f(u^{-1}(w))$. For example, say we want to determine the function for the A scale, which gives squares of the D scale, its reference scale. Since the D scale's function is $f(x) = \log x$, and $u(x) = x^2$ (i.e., we want squares of the D scale), then the A scale's function is $z(w) = f(u^{-1}(w)) = \log \sqrt{w} = \frac{1}{2} \log w$. By inspecting a slide rule, one can see that the location of 4 on the A scale is indeed at the same distance as the location of 2 on the D scale, since $\frac{1}{2} \log 4 = \log 2$.

The other case is where a reference scale $f(x)$ is a function $v(w)$ of the domain values of scale $z(w)$. The goal is to find a formula for $z(w) = f(x)$ such that $v(w) = x$. The result is $z(w) = f(v(w))$. For example, say we want to determine the function for the S scale, which gives inverse sines of the C scale, its reference scale; alternatively, the C scale gives sines of the S scale. The C scale's function is $f(x) = \log 10x$ (since we want to interpret its x values as going from 0.1 to 1, rather than 1 to 10, thus providing sine values for angles from 5.74 to 90 degrees) and $v(w) = \sin w$ (i.e., we want sines of the S scale), then the S scale's function is $z(w) = f(v(w)) = \log (10 \sin w)$. Checking some values, the location of 90 on the S scale is at the same distance as that of the location of 1 on the C scale, at the right end, since $\log (10 \sin 90) = \log (10 \times 1.0)$; and the location of 45 on the S scale is at the same distance as that of the location of 0.707 on the C scale, since $\log (10 \sin 45) = \log (10 \times 0.707)$.

Determining Distances in Measurement Units

Throughout the paper, we have referred to distances simply as pure numbers. To express them in some unit of measure (e.g., inches) and in a practical form so that an actual scale line can be demarcated using a ruled measure and assigned values, the following formula can be used. It determines the distance *from the left end of the scale* (which need not be the origin, thus the formula is general), to the point that corresponds to x :

$$d(x) = m (f(x) - f(x_L)) / (f(x_R) - f(x_L)) \quad (3)$$

where x_L and x_R are the desired x values at the left and right ends of scale, respectively, and m is the desired length of the scale in a chosen unit of measure. Checking the formula's correctness at the ends: $d(x_L) = 0$, and $d(x_R) = m$.

To illustrate with the function $f(x) = x^2$, say that $x_L = 0$ and $x_R = 5$, and the total length of the scale is 10 in. Then, $d(x) = 0.4 x^2$ in. For $f(x) = 1 - \sin x$ over the domain $x = 0$ to 90, $x_L = 90$ and $x_R = 0$ since $f(90) = 0 < f(0) = 1$, and $d(x) = 10 (1 - \sin x)$ in. Checking values, $d(90) = 0$ in., $d(30) = 5$ in., and $d(0) = 10$ in., which correspond to the left end, middle, and right end of the scale, respectively, as expected. Again taking $f(x) = 1 - \sin x$ but limiting the domain to $x = 0$ to 45, the formula works correctly: $x_L = 45$ and $x_R = 0$ since $f(45) = 0.293 < f(0) = 1$, and $d(x) = 14.14 (0.707 - \sin x)$ in. Checking values, $d(45) = 0$ in., $d(30) = 2.93$ in., $d(15) = 6.34$ in., and $d(0) = 10$ in.

Conclusion

In addition to its overwhelming practical success as the engineer's calculating tool, the slide rule is also an interesting theoretical object in its own right, whose mathematics is worthy of study. We have presented some of the mathematical foundations of the slide rule, culminating in a Mathematical Principle of the Slide Rule that allows one to determine what calculations become possible as a result of the simple geometrical addition of line lengths and the construction of scales based on one-dimensional graphs.

Upon further study, one can only marvel at the wide variety of calculations made possible by the slide rule, a simple mechanical device of few moving parts, manipulated by the hand, but powered by the mind.

Acknowledgment

The author wishes to express his deep gratitude to Dr. J. Robert Beyster for his generous support of the author's educational and research activities, an example being this work.

References

1. F. Cajori, *A History of the Logarithmic Slide Rule and Allied Instruments*, Originally published in 1910, reprinted by Astragal Press, 1994.
2. M. d'Ocagne, *Le Calcul Simplifie, Graphical and Mechanical Methods for Simplifying Calculation*, 3rd Edition, Originally published in 1928, reprinted and translated into English by MIT Press, 1986.
3. C. Runge, *Graphical Methods; A Course of Lectures Delivered in Columbia University, New York, October, 1909 To January, 1910*, Originally published by Columbia University Press, 1912, reprinted by Nabu Press, 2010.
4. G. D. C Stokes, *The Slide Rule*, In *Handbook of the Napier Tercentenary Celebration Or Modern Instruments and Methods of Calculation*, edited by E. M. Horsburgh, Originally published by G. Bell and Sons, and the Royal Society of Edinburgh, 1914, Reprinted by Tomash Publishers, 1982.
5. J. Lipka, *Graphical and Mechanical Computation*, John Wiley & Sons, 1918.
6. D. S. Davis, *Empirical Equations and Nomography*, McGraw-Hill Book Company, 1943.
7. C. Mackey, *Graphical Solutions*, 2nd Edition, John Wiley & Sons, 1947.
8. R. P. Hoelscher, J. N. Arnold, and S. H. Pierce, *Graphic Aids in Engineering Computation*, McGraw-Hill Book Company, 1952.

Joe Pasquale is the J. Robert Beyster Professor of Computer Science and Engineering at the University of California, San Diego, where he has been on the faculty since 1987. His research is in operating systems, distributed systems and networks, focusing on performance and reliability of Internet-scale systems with highly decentralized control. He received his Ph.D. from the University of California, Berkeley, and his S.B. and S.M. from MIT, where as an undergraduate in 1976 he was one of the few students still using a slide rule.