Session 2465

Mathematical Definitions: What is this thing?

Andrew Grossfield
College of Aeronautics

Abstract

One sure way to turn students from the study of mathematics is to give them a mathematical definition. These one liners which jump out of nowhere are intended to be the conceptual starting point for five and six hundred page texts. An intelligent student is provided no basis to contest these mandated definitions, even if he does not understand or is confused by them.

New mathematical ideas have always been viewed with suspicion and historically have only been accessible to insiders. Mathematicians have always been under the gun of other mathematicians to justify everything asserted. As a result, a format of mathematics presentation arose that would provide for this justification. Nothing had to be easy to understand, as long as it made justifying, proving easy. Accept as a principle that it makes no sense to prove anything that is not understood. Math teachers conventionally violate the principle. They define something and then state and prove a theorem in the next breath. That is the model to which all traditional math teachers aspire.

Any student who objected too strongly to the meaninglessness of the mathematics regimen was banished to study Social Sciences or Arts. Few texts departed from the program. Only teachers who were schooled in the texts and proud of their mastery of the texts were considered competent.

Almost everyone teaching math today is a survivor of that dictatorial system, a master of the mathematical party line. Now the vogue is reform. Reform of what? Will computers or applications provide new, understandable definitions? Will computers or applications say what we are studying and why?

This paper considers mathematical definitions: π, variables, trig functions, functions, limits and derivatives. It provides examples of good and bad definitions in search of principles by which all students, even those who teach, can glimpse whether they should accept or reject what is being put on their plate.

Introduction

Definitions, in general, are not easy. As an exercise select a familiar object and describe it in words. What is a chair? A dog? The technical definition of dog is a member of a set, a member of a species. A species is a set of animals which when mated with another member of the set produces a member of the set. Thus the set of horses comprises a species as does the set of donkeys. But mules, the offspring of the mating of a horse with a donkey, are incapable of reproduction; and therefore do not comprise a species. However, children play with dogs every day without a thought of the definition of a dog and certainly know what a dog is. To the children the technical definition is irrelevant. We must distinguish between a technical definition and what the thing, itself, is.

π

π is the ratio of the circumference of a circle to it's diameter. Does this definition have any flaws? Yes. First, the statement is based on a principle: the fact that this ratio is a constant number, independent of the size of the circle. The definition cannot make sense to anyone who does not comprehend the principle. To make sense, the principle must be clearly stated and understood prior to the definition.

Second, the number may not be of a type with which the student is familiar. The student at this stage may be familiar with integers, rational fractions and even algebraic numbers, but this may be his/her first confrontation with a transcendental number. Why does the number representing the ratio merit a name? Why don't we just write the number down, e.g. 3 1/7, 3.14 or 3.1416? A student could think if these values for π are not equal, then π must not be a constant.

To summarize this path to understanding, first should come the recognition of the marvelous principle that a process under some variety of conditions always produces the same value, that is, the result is constant. Remaining to be explored is the nature and value of the constant. Only after these concerns are addressed can the definition be said to provide a serious attempt to be understandable.

Variables

The variable, as a concept, is not treated well by mathematicians. The mathematician defines a variable as a member of a set, but everything is a member of some set, while not everything is a mathematical variable. The sets which fill a mathematician's mind are sets like the cantor and mendelbrot sets (interesting constructions but premature in an introductory calculus class) while all that is needed at most is simple intervals, include endpoints or not, as suits the case.

Variables are letters or symbols that represent quantitative; that is, measurable attributes of an object or system under study. Variables, as they are used in calculus, are a notational invention that enables the writing of physical and mathematical laws and relationships. It would be extremely difficult to convey Newton's Law, $F = ma$, the pythagorean right triangle theorem, $a^2 + b^2 = c^2$, or any other relationship without variable notation. The essence of the concept of variable is not in the upper and lower bounds of the set to which the variable belongs, but in the attributes that it represents and in the relationships in which the variable is involved.

Trig Functions

The sine of an angle in a right triangle is the ratio of the length of the opposite side to the length of the hypotenuse. What's wrong? This definition like that of π is based on a principle of geometry regarding similar triangles. Leave out the principle and the student will be confused. The principle is that the ratio of sides of similar triangles is independent of the size of the triangle. This means that the ratios of sides of a right triangle are variables that depend only on the angles and not on the size of the triangle. The student might memorize the words but without the principle the idea will not be understood.

Functions

"Functions are sets of ordered pairs." "Functions are mappings from a domain to a range." "Functions are by definition single valued." These statements may provide good starting points for mathematical proofs, but they fail to convey the idea. I cannot imagine any calculus student expressing a desire to study and acquire information about ordered pairs or mappings from a domain to a range. This certainly is not the image in the minds of engineers who use functions.

The functions that are meaningful to engineering students are primarily studied in calculus. **These functions are relationships between two variables which either track continuously together or where one variable continuously controls the other.** A simple form for describing these relationships is one equation in the two variables. These functions might also be described by graphs or tables. Mathematicians want to define everything as abstractly and as generally as possible. In so doing they fail to convey the idea and thus lose the attention of the students.

Let's say that equations in two variables, which represent functions, also represent curves. Tell students that calculus is a continuation of high school geometry where the emphasis was on polygons and circles. Now we are engaged in the study of curves. If geometry is the study of shapes, then this study of curves is a new development in geometry. The functions of calculus enable us to study a wide class of curves, particularly to locate intersections, directions of tangent lines, points of inflection, areas, arc lengths and radii of curvature. All these properties of curves are evident in the graphs of the functions and appear much more tenable than ordered pairs and mappings.

Another problem in the definition of functions is the premature introduction of the concepts of domains and ranges. Most elementary functions of calculus have infinite extent and even when an application limits the extent we generally want to view the function in its entirety. The only natural limitations on the domains of calculus functions are zeroes in the denominator and negative values under an even root. Why don't we say this and go on to graph the function where the important features of the curves can be viewed. Instead I believe the concept is conveyed to uncertain students that mysterious points randomly emerge where the functions can't be evaluated and their calculators will indicate "ERROR." What is most important about functions is how the variables track, the shape of their curves, not the limitations of range and domain.

The last problem with the definition of functions is the emphasis and requirement of being single valued. The curves of algebraic equations may not be single valued. Does this mean we have to cut them up to consider them as functions? If we understand our equations, the condition of being multi-valued is not a catastrophe. Algebraic curves can have tangent lines and derivatives and in parametric form are not treated very differently from single valued curves. Let's not encumber our definition at the start with unnecessary and irrelevant baggage.

To sum up, the problems here are:

1) not conveying the idea, and
2) prematurely introducing and emphasizing secondary considerations such as domains and ranges and the property of being single-valued.

Limits

Limits are for every ε there exists a δ …. Where did this come from? Do you think your students believe it? There must be something wrong here. Here, too, missing principles, well known to the teacher, confuse the student. The whole story is not being conveyed.

Principle: The function u(x) = (x-a)/(x-a) has the value 1 at every x except when x = a where the function u(x) is undefined. The graph of u(x) coincides with the horizontal line y = 1 with the exception that the point corresponding to x = a has been removed from the line. This property of u(x) is described as a gap discontinuity with a limit of 1 at x = a.

Principle: If g(x) is any continuous function, then h(x) = g(x) u(x) will remain unchanged from g(x) except at the point x = a where the product will be undefined. Again we could say that this multiplication has caused a point to be removed from the graph of g(x). We see that this multiplication has created a gap discontinuity in the function h(x) at x = a.

Say the function g(x) is a polynomial or rational function with no discontinuity at x = a. If h(x) = g(x) (x-a)/(x-a) is written in expanded form than it may not be obvious that h(x) will be undefined when evaluated at x = a. Of course the student may be perplexed when he discovers that h(x) is undefined at x = a.

Principle: Factoring the numerator and denominator of h(x) = g(x) (x-a)/(x-a) enables (x-a)/(x-a) to be cancelled and discloses the continuous function g(x) which can be evaluated at x= a. Knowing g(a) enables us to insert the missing point back into the graph of the function h(x). Multiplying and canceling factors such as (x-a)/(x-a) and (x-b)/(x-b) enables us to insert and remove point gaps in rational functions at will. The gap is simply a point with a vertical coordinate and a horizontal coordinate.

Definition: The limit of the function h(x) with a gap discontinuity is the vertical coordinate of the gap.

This definition recognizes the existence of gap discontinuities and clearly describes the concept under discussion and avoids the constructions of infinitesimals, neighborhoods and epsilons, which may have been introduced prematurely. A last point which might be made is that if h(x) can be expressed as

$$g(x)(x-a)^m/(x-a)^n \quad \text{where } g(x) \text{ is continuous then}$$

the limit of h(x) as $x \rightarrow a$ will be

0	if $m > n$,
g(a)	if $m = n$, and
infinity	if $m < n$.

What is the pedagogical principle here? Don't bring in the heavy artillery too soon. Let the student become accustomed to the idea in familiar circumstances (such as the rational functions here) before introducing machinery capable of dealing with every possibility ever dreamed up by those inventive mathematicians. If a student is not familiar with rational functions and their properties then he/she is not prepared for the study of limits.

Derivatives

The derivative is the limit of the difference quotient. That's like saying that a pot of gold is what is located at the end of the rainbow. That is like saying that chicken soup is the result of placing the ingredients in a pot and cooking. No. Distinguish the concept from a technique of construction. Do this, do this, do this, do it forever and what you get is the derivative. The concept is not a recipe. A butterfly is not something you catch in the net.

Straight lines have slopes. The slope of a line is a measure of the direction in which the line points. Every non-vertical line has a unique slope. Lines that point in the same direction have the same slope. The slope of a line is a number. This number can be easily converted with a calculator into the angle of inclination of the line and vice versa. Curves do not have slopes. However, curves have directionally related features and the derivative is a wonderful invention to approach these features.

The derivative will enable us to determine whether a curve rises, falls or remains stationary. It provides a numerical measure of the direction toward which a curve heads and a method of locating extreme points. It will enable us to determine which of two curves rises faster. The derivative concept will enable us to determine if a curve is turning up or turning down. It will enable us to locate points of inflection.

There are two concepts of derivative which texts often do not distinguish. Students may be confused by the use of a single word to describe two different concepts. The first concept is the point derivative, a number.

Definition: The point derivative of a smooth function at a point is the slope of the tangent line at that point.

This definition describes the point derivative in the way it is conceived of and used. It is not something that we must carefully creep up on and then wonder whether we have caught it or not. At this stage how the value of the derivative is computed or arrived at is of secondary importance.

In addition, engineers are not constantly worried about the existence of derivatives. Mathematicians have established that those functions are differentiable which are produced by adding, subtracting, multiplying and dividing differentiable functions, except where a denominator is zero. Mathematicians have established that differentiable functions of differentiable functions are differentiable. And the engineers trust the mathematician's proofs. The functions most encountered by engineers are the elementary algebraic and transcendental functions that do not present the exotic obstacles that worry mathematicians. Engineers embarking on a study of calculus need the concepts of slope, rate of change and differential variation and control. Should the unstated fears of mathematicians be allowed to govern the presentation of these important concepts?

The second definition of a derivative is the functional derivative, a function.

Definition: The functional derivative of a smooth function is the function that describes the values of the point derivative as a function of position. One way to obtain the value of the point derivative at a point is to evaluate the functional derivative at that point.

Picture this scene occurring in calculus classes once a semester across the nation. The teacher says the "The derivative is the limit … etc." The student thinks but does not say, "So what!" The teacher senses the student's discomfort and says "If you throw a ball into the air and consider height as a function of time, then the derivative will provide the instantaneous speed of the ball. The teacher's statement is true but does it address the student's bewilderment? The teacher believes the mention of the physical application will dispel all confusion. Does it? The student may have no interest in physics and he may be more unfamiliar with and more fearful of physics than he is of geometry. In addition, the speed of a falling projectile is too fast to be easily observable or measurable by a student in a math class. I am not saying that physics is not a marvelous application of calculus. I am questioning its value in explaining the concepts of calculus. If an application is needed, one should be provided which is more observable and controllable by the student. Elementary mechanics and geometry offer many such examples. The derivative is the appropriate tool to explain instantaneous velocity not the other way around.

What pedagogical principles are indicated here?

1) The definition must be more that a sentence. It should be a story conveying the major aspects of the object under study.
2) If a student is unfamiliar or uncertain about limits, a definition based on or in terms of limits will provide the student little value.
3) Do not define an object as a result of a construction.
4) If a word has more than one meaning, indicate these meanings and when the differing meanings are intended.
5) Unnecessary applications may only compound a student's confusion. Experts in their own disciplines may provide better treatment of these applications, including projectile motion, and the topics will be provided the attention and respect that they deserve.

Where does the limit of a difference quotient idea fit in? The idea of computing the derivative by a limiting process is a fantastic idea. Yet these limiting processes serve only as computational techniques. We drop them as soon as we discover simpler methods. True, the derivations of all the derivative rules are based on these constructions involving limits. And it is not that the mathematical constructions are not ingenious. They certainly are. But are these constructions the best point to initiate the exposition of the ideas?

Summary

We marvel at the creations of the minds of mathematicians, proposed amid uncertainty and doubt, distrusted and abandoned by most who would come in marginal contact with them. These conceptions have more than proved their value to our modern technological society. I wonder why the mathematicians, those who are most familiar with the objects of mathematics, have not shown more care in conveying their interests to the general populace.

The definition is not an impediment, an obstacle to be glossed over in order to arrive at the important course material. **The definition is the important course material** and should be accorded better treatment. The definition is the starting point. The place where the most care should be taken to attract the imagination of the student. The definition is the place where it should be made clear to the student that something of real value is being studied and in what directions the study might lead.

I have offered only a few examples but the problem of definition pervades mathematical writing. How are polynomials defined at the eighth grade level? What are discriminants? What are partial differential equations? What are tensors? Are vectors "directed line segments"? Then what is the direction of the components of a mixture in a linear mixture problem? What are eigenvectors? i is not the square root of –1. Real negative numbers do not have square roots. Stop telling how to construct it. Say what it is.

References:

1. Grossfield, Andrew "What are Differential Equations? A Review of Curve Families" Proceedings of the 1997 ASEE Annual Conference, Session 2665

2. Grossfield, Andrew "What is College Algebra?" Proceedings of the 1998 Fall Regional Conference of the Middle Atlantic Section of the ASEE, Session VI (158-166)

ANDREW GROSSFIELD
Throughout his career, Dr. Grossfield has combined an interest in engineering design and mathematics. He earned a BSEE at the City College of New York. During the early sixties, he obtained an M.S. degree in mathematics at night while working full time during the day, designing circuitry for aerospace/avionics companies. He is licensed in New York as a Professional Engineer. Ai207@freenet.buffalo.edu is his email address.