Trigonometry From Differential Equations

Per Vognsen

Contents

1 Introduction

Trigonometry is an area of mathematics taught to every high-school student. The student who goes on to study more advanced mathematics is faced with problems as the intuitive definitions of concepts such as angle, sine and cosine are often inadequate in producing rigorous proofs in calculus. For instance, the usual proof found in calculus textbooks of the limit

$$\lim_{x\to 0} \frac{\sin x}{x} = 1$$

can be shown to be circular upon detailed examination. This identity is at the foundation of essentially all calculus involving trigonometric functions and so this is a very unsatisfactory state of affairs. Rigor is the hallmark of modern mathematics and we should settle for nothing less for statements as fundamental as the above.

Fortunately there are ways of achieving the highest level of rigor. In this article most of the details pertaining to one such approach are worked out. All the arguments are simple if somewhat clever at times. Nothing more than elementary calculus is used, with only a single exception. Rather than starting with geometric definitions, we start with definitions based on differential calculus and proceed to show that all the familiar properties hold true. Finally, we make direct connection with the geometry by linking our definitions with the unit circle and then angles.

2 Definition

Consider the pair of coupled first-order differential equations

$$\begin{aligned} C'(t) &= -S(t) \\ S'(t) &= C(t) \end{aligned}$$

with initial conditions $C(0) = 1$ and $S(0) = 0$. For a brief moment we will assume that these equations have solutions C and S and derive some elementary consequences of the definitions. We will then prove their existence by constructing explicit solutions using power series.

3 The Theorem of Pythagoras

Consider the function $C(t)^2 + S(t)^2$. Its derivative is $2C(t)C'(t) + 2S(t)S'(t)$ which, by applying the definitions of C and S, reduces to 0. A function with vanishing derivative is constant. This implies $C(t)^2 + S(t)^2 = C(0)^2 + S(0)$ and thus

$$C(t)^2 + S(t)^2 = 1$$

An immediate corollary of this identity is that we have the bounds $C(t)^2 \leq 1$ and $S(t)^2 \leq 1$ and hence

$$\begin{aligned} |C(t)| &\leq 1, \\ |S(t)| &\leq 1. \end{aligned}$$

4 Construction by Power Series

If we differentiate C repeatedly, we find that $C'(t) = -S(t)$, $C''(t) = -C(t)$, $C'''(t) = S(t)$ and $C''''(t) = C(t)$. We are back with C where we started. This pattern repeats itself indefinitely and we therefore have $C^{(4n+0)}(t) = C(t)$, $C^{(4n+1)}(t) = -S(t)$, $C^{(4n+2)}(t) = -C(t)$ and $C^{(4n+3)}(t) = S(t)$ where n is any nonnegative integer.

Next we will perform a Maclaurin expansion of C. To accomplish this we need the values of the derivatives evaluated at 0. These are $C^{(4n+0)}(0) = 1$, $C^{(4n+1)}(0) = 0$, $C^{(4n+2)}(0) = -1$ and $C^{(4n+3)}(0) = 0$. The Maclaurin expansion of C is thus

$$1 - \frac{t^2}{2!} + \frac{t^4}{4!} + \cdots$$

with the Lagrange remainder

$$R_n = \frac{s^n\, C^{(n)}(s)}{n!}$$

for some s in the $(0, t)$ interval. To show that the Maclaurin series is convergent, we have to demonstrate that $R_n \to 0$ as $n \to \infty$. The derivative $C^{(n)}(s)$ must be $\pm C(s)$ or $\pm S(s)$. Thus $\left|C^{(n)}(s)\right| \leq 1$ by the bounds on C and S we derived earlier. We therefore have

$$|R_n| \leq \frac{|s|^n}{n!}.$$

Since half the factors of $n!$ are greater than or equal to $n/2$, we have the estimate $n! \geq (n/2)^{n/2}$ and so we get

$$\frac{|s|^n}{n!} \leq \frac{(s^2)^{n/2}}{(n/2)^{n/2}} = \left(\frac{s^2}{n/2}\right)^{n/2}.$$

Fixing s, we have $n/2 > s^2$ for all $n \geq N$ where $N = 2s^2 + 1$. Letting $r = \frac{s^2}{N/2}$ we therefore have

$$\left(\frac{s^2}{n/2}\right)^{n/2} \leq r^{n/2}$$

where $r < 1$. We have $r^{n/2} \to 0$ as $n \to 0$ and since $|R_n| \leq r^{n/2}$ it follows that $|R_n| \to 0$ as well. A similar analysis shows that the Maclaurin series for S,

$$-t + \frac{t^3}{3!} - \frac{t^5}{5!} + \cdots\ ,$$

converges to S everywhere.

We have thus constructed power series solutions to the pair of differential equations. In fact, it can be shown that these solutions are the only ones. This is a special case of a general uniqueness result for first-order differential equations.

5 Even and Odd

A simple but very important consequence of the construction of solutions by power series is that we can easily see that C is even, $C(-t) = C(t)$, and that S is odd, $S(-t) = -S(t)$.

6 Addition Formulas

Define the function f by

$$f(s) = C(sx)C(x+y-sx) - S(sx)S(x+y-sx)$$

where x and y are real numbers. We have

$$f(0) = C(0)C(x+y) - S(0)S(x+y) = C(x+y)$$

and

$$f(1) = C(x)C(y) - S(x)S(y).$$

We now differentiate f with respect to s. The first term differentiates to

$$-xS(sx)C(x+y-sx) + xC(sx)S(x+y-sx).$$

Proceeding to the second term, we find that its derivative is

$$-xC(sx)S(x+y-sx) + xS(sx)C(x+y-sx).$$

These two terms cancel each other out. As a consequence f has a vanishing derivative and is therefore constant. In particular, we have $f(0) = f(1)$ and therefore

$$C(x+y) = C(x)C(y) - S(x)S(y).$$

It can be shown by analogous methods that

$$S(x+y) = C(x)S(y) + C(y)S(x)$$

by using the function

$$g(s) = C(sx)S(x+y-sx) + C(x+y-sx)S(sx)$$

in place of f.

7 Existence of Zeros and Definition of $\frac{\pi}{2}$

Suppose for the sake of contradiction that C were to have no positive zeros. Since $|C(t)| \leq 1$, it follows from the hypothesis that $0 < C(t) \leq 1$ for $t > 0$. Suppose that L is the greatest lower bound on C for positive values of t. We then have $0 < L \leq C(t)$ for all $t > 0$. Since $S'(t) = C(t)$, we have

$$S(t) = \int_0^t C(t')\, dt'.$$

Using $C(t) \geq L$ for $t > 0$, we get the estimate

$$S(t) \geq \int_0^t L\, dt' = tL.$$

Note that $tL > 1$ for $t > 1/L$. By the above estimate, we thus have $S(t) > 1$ for $t > 1/L$. But this contradicts the bound $|S(t)| \leq 1$. The initial hypothesis must therefore have been incorrect and we may conclude that C indeed has a positive zero. The evenness of C implies that C also has a negative zero.

Now we can define $\frac{\pi}{2}$ to be the smallest positive zero of C. We have just demonstrated that C has a positive zero so this definition is in fact meaningful.

8 Function Values at Multiples of $\frac{\pi}{2}$

The purpose of this section is to calculate the values of C and S at integerm multiples of $\frac{\pi}{2}$. Firstly, since $C(\frac{\pi}{2})^2 + S(\frac{\pi}{2})^2 = 1$ we have $S(\frac{\pi}{2})^2 = 1$ because $\frac{\pi}{2}$ is a zero of C. Thus $S(\frac{\pi}{2})$ is either -1 or 1.

Since $C(0) = 1$ and $\frac{\pi}{2}$ is the smallest positive zero of C, it follows that C is strictly positive on the $\left(0, \frac{\pi}{2}\right)$ interval. Hence S is strictly increasing on $\left(0, \frac{\pi}{2}\right)$ since $S'(t) = C(t)$. We have $S(0) = 0$ and thus S must be positive on $\left(0, \frac{\pi}{2}\right)$. Therefore $S(\frac{\pi}{2}) = 1$.

We can calculate the values of C and S at π and 2π using the addition formulas. At π, we have $C(\pi) = C(\frac{\pi}{2}+\frac{\pi}{2}) = C(\frac{\pi}{2})C(\frac{\pi}{2}) - S(\frac{\pi}{2})S(\frac{\pi}{2}) = 1$ and since $C(\pi)^2 + S(\pi)^2 = 1$ we have $S(\pi)^2 = 1 - C(\pi)^2 = 0$. Hence $S(\pi) = 0$.

Calculating the values of the two functions at 2π, we find that $C(2\pi) = C(\pi+\pi) = C(\pi)C(\pi) - S(\pi)S(\pi) = 1$ and $S(2\pi) = S(\pi + \pi) = 2C(\pi)C(\pi) = 0$.

9 Periodicity

Using the addition formulas and the values of C and S we calculated in the previous section, we get

$$C(t + 2\pi) = C(t)C(2\pi) - S(t)S(2\pi) = C(t)$$

and

$$S(t + 2\pi) = C(t)S(2\pi) + C(2\pi)S(t) = S(t).$$

We thus see that C and S are periodic with period 2π.

10 Co-relations

We next derive an important pair of relations between C and S. By the addition formula,

$$\begin{aligned} C\left(t - \frac{\pi}{2}\right) &= C(t)C\left(-\frac{\pi}{2}\right) - S(t)S\left(-\frac{\pi}{2}\right) \\ &= S(t)S\left(\frac{\pi}{2}\right) \\ &= S(t). \end{aligned}$$

Adding $\frac{\pi}{2}$ to t, we arrive at the dual identity

$$S\left(t + \frac{\pi}{2}\right) = C(t).$$

11 Relationship to the Unit Circle

Define the mapping

$$\begin{aligned} T : \mathbb{R} &\to \mathbb{R}^2 \\ t &\mapsto (C(t), S(t)) \end{aligned}$$

from the real number line to the plane. It is clearly continuous since C and S are differentiable. The identity $C(t)^2 + S(t)^2 = 1$ implies that the image of T is a subset of the unit circle.

Consider the behavior of T on the interval $\left[0, \frac{\pi}{2}\right]$. Earlier we noted that S is increasing on $(0, \frac{\pi}{2})$ and therefore injective on that interval. At 0, the mapping takes on the value $(0, 1)$ and at $\frac{\pi}{2}$ its value is $(0, 1)$. Thus the image of T on $\left[0, \frac{\pi}{2}\right]$ is a subset of the first quarter of the unit circle.

Since $\left[0, \frac{\pi}{2}\right]$ is connected and T is continuous, it follows that its image is likewise connected. The only connected subset of the first quarter unit circle that contains $(1, 0)$ and $(0, 1)$ is the entire first quarter of the unit circle. Hence S must be a homeomorphism of $\left[0, \frac{\pi}{2}\right]$ onto the first quarter unit circle.

Similar reasoning shows that T maps $\left[\frac{\pi}{2}, \pi\right]$ homeomorphically onto the second quarter unit circle, $[\pi, 3\pi/2]$ homeomorphically onto the third quarter unit circle and $[3\pi/4, 2\pi]$ homeomorphically onto the fourth quarter unit circle. Thus T is seen to be a homeomorphism of $\left[0, \frac{\pi}{2}\right)$ onto the unit circle with the point $(1, 0)$ removed.

12 Defining Angles

Given a point P on the unit circle, we define its angle $\theta(P)$ to be the inverse of T when restricted to $\left[0, \frac{\pi}{2}\right)$. This inverse exists due to the surjectivity of T and its injectivity when restricted to this interval, as was established in the previous section. Since T restricted to $\left[0, \frac{\pi}{2}\right)$ is a homeomorphism it has a continuous inverse and thus θ is continuous. We can define the angle between two points on the unit circle by the difference of their angles modulo π.