The Fundamental Theorem of Trigonometry

Omar Hijab

Dedicated to Rani

1 Introduction

A mystifying feature of today’s calculus books is a lack of any discussion of exactly what is the measure of an angle. The subject is assumed to have been treated somehow in secondary school. However, this is not so, as axiomatic treatments [1] of angle measure bypass this issue by design.

Instead of treating angle measure in a pre-calculus course or early in a calculus course, within the framework of just-learned cartesian geometry, calculus books often go outside this framework by appealing to pictures not only for motivation, but also for justification. As a consequence, they forfeit the opportunity to present angle measure as a basic paradigm of calculus.

Typically, calculus books state the measure \( \theta \) of an angle is the length of the subtended arc along the unit circle, use this to define \( \sin \theta \), then go on to derive

\[
\lim_{\theta \to 0} \frac{\sin \theta}{\theta} = 1
\]  

by sandwiching a sector’s area between the areas of inscribed and circumscribed triangles. None of this is defined analytically when beginning cartesian geometry.

The book [5] comes close to defining the measure of an angle, as it discusses the history of the subject at length. The books [3], [4] ignore the issue completely, although the material here would fit perfectly there.

\[ \theta' \]
\[ \theta + \theta' \]

Figure 1: Additivity.

Whatever definition one takes for the measure of the angle, it should be additive: When angles are stacked, their measures should add (Figure 1).

\[ \theta + \theta' \]

*Temple University, Philadelphia, PA, USA, hijab@temple.edu
Additive angle measure $\theta$ was introduced by Archimedes, as part of his work\cite{2} leading to his estimate of the half-circumference of the unit circle,
\[
\frac{223}{71} < \pi < \frac{22}{7}.
\] (2)
By contrast, Hipparchus\cite{6} and Ptolemy\cite{7} used chord measure $\theta_1$, which is equivalent to $\theta$ but not additive, to build their trigonometric tables.
In this note, we explain how angle stacking leads to complex numbers, derive the additivity of Archimedes’ angle measure, and show how additivity leads to the fundamental theorem of trigonometry.
Apart from the completeness property of the reals, continuity, and the intermediate value theorem, our presentation remains within the confines of cartesian geometry, and is therefore accessible to the beginning calculus student.

2 Angle Stacking

In the cartesian plane, points are ordered pairs of real numbers $P = (x, y)$, $P' = (x', y')$, and points may be added, $P + P' = (x + x', y + y')$, and multiplied by real numbers $t$, $tP = (tx, ty)$.
An angle is an ordered pair of rays starting from a common point, the vertex of the angle. If the vertex is the origin $O = (0, 0)$, then an angle is determined by the intersections $P$, $P'$ of its rays with the unit circle. We say an angle is anchored if its vertex is $O$ and its first intersection is $I = (1, 0)$.
Complex multiplication and division are forced upon us as soon as we stack angles as in Figure 1, even before angle measure is defined.
This is clearest when the angles are anchored. Let $P$ and $P'$ be on the unit circle, and let $P''$ be obtained by stacking $P'$ atop $P$, as in Figure 2.

To make stacking precise, draw the circle with center $Q = x'P$ and radius $|y'|$, as in Figure 2. When $x'y' \neq 0$, this circle intersects the unit circle at
\[
P'' = (xx' - yy', x'y + xy') \quad \text{and} \quad P''' = (xx' + yy', x'y - xy').
\] (3)
Proof of (3). Let $\langle P, P' \rangle = xx' + yy'$, $|P|^2 = \langle P, P \rangle$, and $P^\perp = (-y, x)$. Then $P$ is on the unit circle iff $|P|^2 = 1$, and $P$, $P'$ on the unit circle satisfy $\langle P, P' \rangle = 0$.
iff $P' = \pm P'$. Since $P''$ is on the circle of center $Q$ and radius $|y'|$, we may write

$$P'' = Q + y'R,$$

for some $R$ on the unit circle. Then $1 = |P''|^2 = |x'P + y'R|^2$ iff $(P, R) = 0$ iff $R = \pm P'$, yielding $P'' = x'P \pm y'P'$, which is (3).

If we identify cartesian points $P = (x, y)$, $P' = (x', y')$ with complex numbers $z = x + iy$, $z' = x' + iy'$, the points (3) are identified with the complex product $zz'$ and the complex quotient $z/z'$, at least when $z$, $z'$ lie on the unit circle.

Given this, it makes sense to replace points $P$, $P'$, $P''$ by complex numbers $z$, $z'$, $z''$, and to rewrite (3) as

$$z'' = zz' \quad \text{and} \quad z'' = z/z'.$$  \hspace{2cm} (4)

Below we define the Archimedes measure $\theta(z)$ of $z \neq -1$ on the unit circle. To take into account both cases in (4), we require $\theta(1/z) = -\theta(z)$. Then we interpret Figure 1 by writing

$$\theta(zz') = \theta(z) + \theta(z'), \quad x > 0, x' > 0.$$  \hspace{2cm} (5)

It is in this form we establish additivity of $\theta = \theta(z)$.

### 3 Archimedes Bisection

Let $z = x + iy$ be on the punctured unit circle $z \neq -1$, and define $m_1 = (z + 1)/2$ and

$$z_1 = \frac{m_1}{|m_1|} = \frac{z + 1}{\sqrt{2 + 2x}} = \frac{x + 1}{\sqrt{2 + 2x}} + i \frac{y}{\sqrt{2 + 2x}} = x_1 + iy_1.$$  \hspace{2cm} (6)

A short computation shows $z_1^2 = z$, thus $z_1$ is a square root of $z$; we write $\sqrt{z}$ for this square root. By (6), $\sqrt{z}$ maps the punctured unit circle $z \neq -1$ continuously and bijectively onto the right-half unit circle $x_1 > 0$, and the imaginary parts of $z$ and $\sqrt{z}$ have the same sign.

Since $(\sqrt{z}\sqrt{z'})^2 = zz'$,

$$\sqrt{zz'} = \sqrt{z}\sqrt{z'},$$  \hspace{2cm} (7)

up to sign, whenever both sides of (7) are defined. When $z$, $z'$, $zz'$ are in the upper-half unit circle, both sides are defined and have positive imaginary parts. Hence (7) is correct as written, when $z$, $z'$, $zz'$ are in the upper-half unit circle.
Let \( \theta_1 = \theta_1(z) = |z - 1| \) be as in Figure 3, with \( z \) in the unit circle first quadrant \( x > 0, y > 0 \). Then it is easy to check \( \theta_1 = 2y_1 \) and

\[
y < 2y_1 < \frac{2y_1}{x_1} < \frac{y}{x},
\]

Similarly, let \( z_2 = x_2 + iy_2 = \sqrt{z_1} \) and let \( \theta_2 = \theta_2(z) = 2\theta_1(z_1) \) be the chord-length sum in Figure 3. Then \( \theta_2 = 4y_2 \) and

\[
2y_1 < 4y_2 < \frac{4y_2}{x_2} < \frac{2y_1}{x_1},
\]  

when \( z \) is on the upper-half unit circle \( y > 0 \).

If we define \( \theta_n \) and \( z_n = x_n + iy_n \) recursively by \( \theta_{n+1}(z) = 2\theta_n(\sqrt{z}) \) and \( z_{n+1} = \sqrt{z_n}, n \geq 2 \), then \( \theta_1, \theta_2, \theta_3, \ldots \) are obtained by repeated bisection of the subtended arc, and it is easy to check

\[
\theta_n = 2^n y_n, \quad n \geq 1.
\]  

(9)

Iterating (8) and appealing to (9) yields the sequences

\[
\theta_1 < \theta_2 < \theta_3 < \cdots < \frac{\theta_3}{x_3} < \frac{\theta_2}{x_2} < \frac{\theta_1}{x_1},
\]  

(10)

when \( z \) is on the upper-half unit circle \( y > 0 \). Here the decreasing sequence consists of the chord-length sums of the circumscribed chords obtained by dilating the inscribed chords in Figure 3 away from the origin.

By (10), \( \theta_1, \theta_2, \theta_3, \ldots \) is bounded. By (9), \( y_n \to 0 \), hence \( x_n \to 1 \), as \( n \to \infty \). By the completeness property of the real numbers, the sequences in (10) have a common limit \( \theta = \theta(z) \). By construction,

\[
\theta(z) = 2\theta(\sqrt{z})
\]  

(11)

follows, when \( y > 0 \).

Assume \( x > 0, y > 0 \). Since \( \sqrt{2 + 2x} < 2 \) and by (6) \( x_1 > y_1 \),

\[
2(1 - x_1 + y_1) = 2 - 2(x_1 - y_1) < 2 - \sqrt{2 + 2x(x_1 - y_1)} = 1 - x + y.
\]

Iterating this, \( 2^n(1 - x_n + y_n) < 1 - x + y \), hence \( \theta_n < 2(1 - x_1 + y_1), n \geq 1 \), as suggested by the dashed lines in Figure 3. Passing to the limit,

\[
y < \theta(z) < 1 - x + y, \quad z = x + iy, x > 0, y > 0.
\]  

(12)

Extend \( \theta(z) \) to the lower-half unit circle \( y < 0 \) by

\[
\theta(z) = -\theta(1/z),
\]  

(13)

and set \( \theta(1) = 0 \). Then (11) and (13) are valid on the punctured unit circle \( z \neq -1 \), and \( \theta(z) = \theta \left( \sqrt{1 - y^2} + iy \right) \) is an odd function of \( y \) on the right-half unit circle.
Proof of additivity. We establish (5) for $z = x + iy$, $z' = x' + iy'$ in the right-half unit circle, $x > 0$ and $x' > 0$. Let $z'' = zz' = x'' + iy''$, and let $\theta = \theta(z)$, $\theta' = \theta(z')$, and $\theta'' = \theta(z'')$. Since (5) is immediate when $yy'y'' = 0$, we may assume $yy'y'' \neq 0$. There are two cases.

First assume $yy' > 0$. By (13), (5) is valid for $z$, $z'$ iff (5) is valid for $1/z$, $1/z'$. Hence, in this case, we may assume $y > 0$ and $y' > 0$. Let $z'_n = x'_n + iy'_n$, $z''_n = x''_n + iy''_n$, $n \geq 1$, be the corresponding sequences starting from $z'$, $z''$ respectively, and let $\theta''_n$, $\theta'_{n'}$, $n \geq 1$, be the corresponding chord-length sums. Then, by (7), $z''_n = z_n z'_n$, thus $y''_n = x'_n y_n + x_n y'_n$, hence, by (9),

$$
\theta''_n = x'_n \theta_n + x_n \theta'_n, \quad n \geq 1.
$$

Now send $n \to \infty$. Since $x_n \to 1$, $x'_n \to 1$, we obtain $\theta'' = \theta + \theta'$, which is (5).

Second assume $yy' < 0$. Then we have $x'' = xx' - yy' > 0$. Since $yy' < 0$, $y''(-y)$ and $y''(-y')$ have opposite signs. By switching the roles of $z$ and $z'$ if necessary, we may assume $y''(-y) > 0$. Applying the first case to $z''$ and $1/z = x - iy$, 

$$
\theta(z') = \theta(z''/z) = \theta(z'') + \theta(1/z) = \theta(z'z) - \theta(z),
$$

which is (5). $\Box$

Let $z$ be in the unit circle first quadrant. From (6), $y_1$ is an increasing function of $y$. Similarly, with $y_2$ playing the role of $y_1$, $y_2$ is an increasing function of $y_1$, hence an increasing function of $y$. Continuing in this manner, $y_n$, $n \geq 1$, are increasing functions of $y$. By (9), $\theta_n$, $n \geq 1$, are increasing functions of $y$. Passing to the limit, and since $\theta(z)$ is an odd function of $y$, it follows $\theta(z)$ is an increasing function of $y$, when $z$ is in the right-half unit circle.

Define $\pi = 2\theta(i)$. Since $2\theta_1(i) = 2\sqrt{2}$ and $x_1(i) = 1/\sqrt{2}$, by (10), $2\sqrt{2} < \pi < 4$. To achieve (2) using (10), Archimedes effectively calculated

$$
2\theta_1(i) = 64 \sqrt{2 - \sqrt{\sqrt{2 + 2 + 2 + 2 + 2}}.}
$$

By (12), and since $\theta(z)$ is an odd function of $y$, $\theta(z)$ is continuous at $z = 1$ and $\theta(z) \neq 0$ when $z \neq 1$. If $z \neq z'$ are in the right-half unit circle and close to each other, then $z/z'$ is $\neq 1$ and close to 1. By (5),

$$
\theta(z/z') = \theta(z) - \theta(z'), \quad x > 0, x' > 0.
$$

It follows $\theta(z)$ is continuous and injective on the right-half unit circle. Since $\theta(z)$ is an increasing function of $y$ and $\theta(\pm i) = \pm \pi/2$, $\theta(z)$ maps the right-half unit circle into $(-\pi/2, \pi/2)$. By the intermediate value theorem, $\theta(z)$ is a continuous bijection from the right-half unit circle onto $(-\pi/2, \pi/2)$. By (11), $\theta(z)$ is a continuous bijection from the punctured unit circle $z \neq -1$ onto $(-\pi, \pi)$. 5
4 Trigonometry

The basic result of trigonometry is the existence of a continuous map \( z(\theta) = \cos \theta + i \sin \theta \) of the real line into the unit circle satisfying the addition formula

\[
z(\theta)z(\theta') = z(\theta + \theta')
\]  
(14)

for all \( \theta, \theta' \). By (14), any such map satisfies \( z(0) = 1 \) and \( z(-\theta) = 1/z(\theta) \).

Given a continuous map \( z(\theta) \) satisfying (14), we say a real \( \alpha \) is a period if \( z(\alpha) = 1 \). Then 0 is a period, every integer multiple of a period is a period, and a limit of periods is a period.

Since the constant map \( z(\theta) \equiv 1 \) satisfies (14) trivially, we only seek non-constant maps satisfying (14).

**Theorem** (Fundamental Theorem of Trigonometry). There is a non-constant continuous map \( z(\theta) \) of the real line into the unit circle, unique up to rescaling, satisfying (14).

**Proof of Existence.** Since \( \theta(z) \) is a continuous bijection, there is a continuous inverse \( z(\theta) \) on \((-\pi, \pi)\). Then \( z(\theta) \) can be extended uniquely to all reals with (14) valid for all \( \theta, \theta' \), and we write \( z(\theta) = \cos \theta + i \sin \theta \).

Since \( \theta(i) = \pi/2, z(\pi/2) = i \). By (5), \( z(\theta) \) satisfies (14) on \((-\pi/2, \pi/2)\), hence, for \( \theta \) in \((-\pi, \pi)\),

\[
z(\theta) = z(\theta/2)^2.
\]  
(15)

For \( \alpha \geq \pi \), suppose \( z(\theta) \) is defined on \((-\alpha, \alpha)\) and satisfies (14) on \((-\alpha/2, \alpha/2)\). If \( Z(\theta) \) extends \( z(\theta) \) to \((-2\alpha, 2\alpha)\) and satisfies (14) on \((-\alpha, \alpha)\), then \( Z(\theta) = Z(\theta/2)^2 = z(\theta/2)^2 \) on \((-2\alpha, 2\alpha)\), hence \( Z(\theta) \) is uniquely determined. Conversely, define \( Z(\theta) = z(\theta/2)^2 \) on \((-2\alpha, 2\alpha)\). Then \( Z(\theta) \) satisfies (14) on \((-\alpha, \alpha)\), since

\[
Z(\theta)Z(\theta') = z(\theta/2)^2z(\theta'/2)^2 = z((\theta + \theta')/2)^2 = Z(\theta + \theta')
\]

for \( \theta, \theta' \) in \((-\alpha, \alpha)\), and \( Z(\theta) \) extends \( z(\theta) \), since \( Z(\theta) = z(\theta/2)^2 \) on \((-\alpha, \alpha)\).

Iterating this, \( z(\theta) \) can be extended uniquely to \((-2\pi, 2\pi), (-4\pi, 4\pi), (-8\pi, 8\pi), \ldots, \) with the extensions satisfying (14) on \((-\pi, \pi), (-2\pi, 2\pi), (-4\pi, 4\pi), \ldots, \) resulting in a continuous map \( z(\theta) \) satisfying (14) for all reals. Since \( z(\pi) = z(\pi/2)^2 = -1 \) and \( z(2\pi) = z(\pi)^2 = 1 \), \( z(\theta) \) is surjective and \( 2\pi \) is a period.

**Proof of Uniqueness step 1.** Let \( z(\theta) \) be any non-constant continuous map \( z(\theta) \) of the real line into the unit circle satisfying (14). Then \( z(\theta) \) has a positive period.

Write \( z(\theta) = x(\theta) + iy(\theta) \). Since 0 is a period, \( x(0) = 1 \). Since \( z(\theta) \) is non-constant, there is an \( \alpha \neq 0 \) with \( x(\alpha) < 1 \). Since \( \cos 0 = 1 \), by continuity, there is an \( n \geq 1 \) with \( x(\alpha) < \cos(2\pi/n) < 1 \). By the intermediate value theorem, there is a \( \beta \neq 0 \) with \( x(\beta) = \cos(2\pi/n) \), hence \( y(\beta) \) equals one of \( \pm \sin(2\pi/n) \). By (14),

\[
z(n\beta) = z(\beta)^n = (\cos(2\pi/n) \pm i \sin(2\pi/n))^n = \cos(2\pi) \pm i \sin(2\pi) = 1,
\]

hence \( z(\pm n\beta) = 1 \), hence there is a positive period.
Proof of Uniqueness step 2. \( z(\theta) \) has a least positive period \( \alpha \).

Since a limit of periods is a period, the infimum \( \alpha \) of all positive periods is a period. For every real \( \theta \) and every period \( \beta > 0 \), for some integer \( n \), we have \( n\beta \leq \theta < (n+1)\beta \). Hence for every period \( \beta > 0 \), every real \( \theta \) lies within distance \( \beta \) of some period. If \( \alpha = 0 \), there would be arbitrarily small positive periods \( \beta \).

This would imply every real \( \theta \) is the limit of some sequence of periods, and thus \( z(\theta) \equiv 1 \). Since by assumption this is disallowed, \( \alpha > 0 \).

Proof of Uniqueness step 3. By rescaling \( z(\theta) \) to \( z(\alpha \theta/2\pi) \), we may assume \( \alpha = 2\pi \). Then \( z(\pi) = -1 \), so \( z(\pi/2) = i \) or \( z(-\pi/2) = i \). By rescaling \( z(\theta) \) to \( z(-\theta) \) if necessary, we may assume \( z(\pi/2) = i \). Then \( z(\theta) \) is uniquely determined.

Since \( 2\pi \) is the least positive period, \( z(\theta) \neq \pm 1 \) for \( 0 < |\theta| < \pi \). In particular, \( \sqrt{z(\theta)} \) is defined for \( -\pi < \theta < \pi \). By (14), we have (15), hence

\[
z(\theta/2) = \sqrt{z(\theta)}, \quad -\pi < \theta < \pi,
\]

up to sign. We claim (16) is correct as written. This is immediate when \( \theta = 0 \), so assume \( \theta \neq 0 \). If the imaginary parts of \( z(\theta) \) and \( z(\theta/2) \) have opposite signs, then, by the intermediate value theorem, for some \( \theta' \) between \( \theta \) and \( \theta/2 \), we have \( z(\theta') = 1 \) or \( z(\theta') = -1 \). Since this can’t happen, the imaginary parts of \( z(\theta) \) and \( z(\theta/2) \) must have the same sign.

It follows the imaginary parts of \( \sqrt{z(\theta)} \) and \( z(\theta/2) \) have the same sign, establishing the claim. Using (14) and (16) repeatedly, for all dyadic rationals \( \theta = k/2^n \), \( z(\theta\pi/2) = (i^k)^{2^{-n}} = i^\theta \). Since the dyadic rationals are dense and \( z(\theta) \) is continuous, this determines \( z(\theta) \).

Inserting \( z = \cos \theta + i \sin \theta \) in (12) leads to (1).

Acknowledgements

The author thanks Sheldon Axler, Eric Grinberg, Munther Hindi, and Daniel Klain for helpful remarks. The author also thanks Rawa for their hospitality.

References