Accuracy vs Efficiency of Numerical Methods

How to program a Calculator

Jake Darby

Abstract

This document will discuss and analyse various numerical methods for computing functions commonly found on calculators. The aim of this paper is to, for each set of functions, compare and contrast several algorithms in regards to their effiency and accuracy.

4 Trigonometric Functions

This section will focus on trigonometric functions, which are commonly used cyclic functions. These functions have been studied for hundreds of years, and can be challenging to calculate. We will discuss several methods of calculating them below before comparing methods.
TODO: Extend and Eloquate introduction

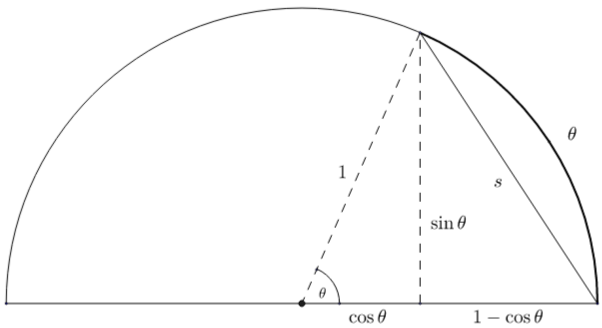
4.1 Calculating π

Several of the methods in this section require that we already know the value of π, for example when we are applying several trig identities. Here we will briefly discuss several methods for calculating the value of π, so that we may use this value in later subsections.
TODO: Fill out with content methods

4.2 Geometric Method

The first method I will be discussing is a method based on geometric properties that are derived on a circle, and we will start by considering values of $\cos$ in the range $[0, \frac{\pi}{2}]$. To do this we will consider the following figure of the unit circle:

Figure 4.2.1: Diagram showing angles to be dealt with

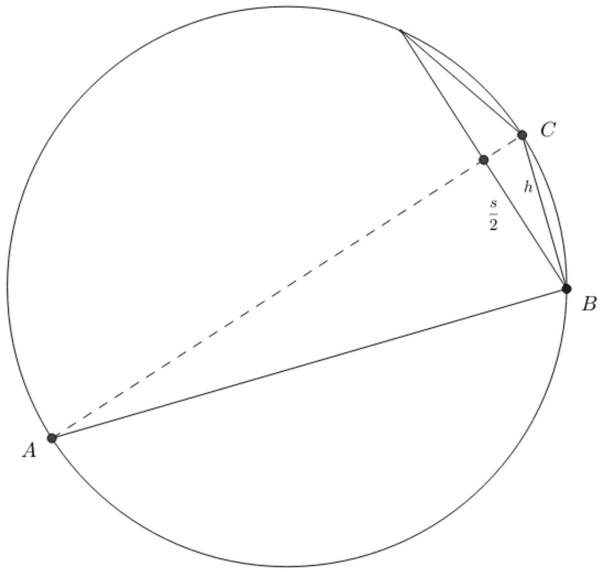


Here theta will be given in radians, and we can note that the labelled arc has length θ due the formula for the circumference of a circle. By using the following derivation we can find a formula for θ in terms of s:

$$\begin{aligned} s^2 &= \sin^2\theta + (1-\cos\theta)^2 \\ &= (\sin^2\theta + \cos^2\theta) + 1 - 2\cos\theta \\ &= 2 - 2\cos\theta \qquad \text{By using } \sin^2\theta + \cos^2\theta = 1 \\ \cos\theta &= 1 - \frac{s^2}{2} \end{aligned}$$

We will now consider a second diagram which will allow us to calculate an approximate value of s.
We will first note that by an elementary geometry result we can know that the angle ABC is a right-angle; also we can consider that h is an approximation of $\frac{\theta}{2}$, which will become relevant later. Now because AC is a diameter of our circle then it's length is 2 and thus, by utilising Pythagarus' Theorem, we get that the length of AB is $\sqrt{AC^2 - BC^2} = \sqrt{4 - h^2}$.

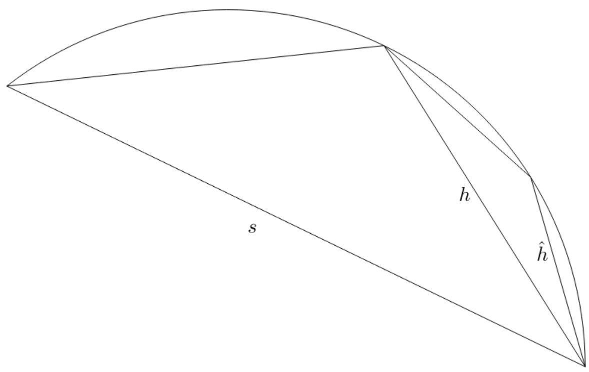
Figure 4.2.2: Diagram detailing how to calculate s



From here we consider the area of triangle ABC, which can be calculated as $\frac{1}{2} \cdot h \cdot \sqrt{4 - h^2}$ and as $\frac{1}{2} \cdot 2 \cdot \frac{s}{2}$; by equating these two, squaring both sides and re-arranging we get that $s^2 = h^2(4 - h^2)$. Now we have the basis for a method that will allow us to calculate $\cos\theta$.

To complete our method we will consider introducing a new line that is to h what h is to s as shown in the diagram below:

Figure 4.2.3: Detailing the recursive steps



It is easy to see that if we repeat the steps above we get that $h^2 = \hat{h}^2(4 - \hat{h}^2)$, and it also follows that $\hat{h} \approx \frac{\theta}{4}$. Using this we can take an initial guess of $h_0 := \frac{\theta}{2^k}$, for some $k \in \mathbb{N}$, and then calculate $h_{n+1}^2 = h_n^2(4 - h_n^2)$ where $n \in [0, k] \cap \mathbb{Z}$; finally we calculate $\cos\theta = 1 - \frac{h_k^2}{2}$, giving the following algorithm:

Algorithm 4.2.1: Geometric calculation of $\cos$

```
geometric_cos(θ ∈ [0, π/2], k ∈ ℕ)
    h_0 := θ/2^k
    n := 0
    while n < K:
```

```
        h_{n+1}^2 := h_n^2 · (4 − h_n^2)
        n ↦ n + 1
    return 1 − h_k^2 / 2
```

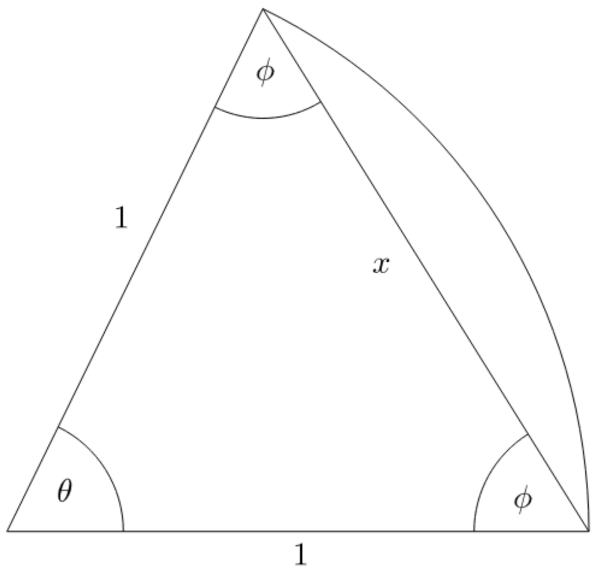
Now we can use the above pseudocode to calculate any trigonometric function value by using various trigonometric identities. First we suppose $\theta \in \mathbb{R}$, then we can repeatedly apply the identity $\cos\theta = \cos(\theta \pm 2\pi)$ to either add or subtrack 2π until we have a value $\theta' \in [0, 2pi)$. Once we have this value we can utilise the following assignment to calculate $\cos\theta$:

$$\cos\theta = \begin{cases} \cos\theta' & : \theta' \in [0, \frac{\pi}{2}] \\ -\cos(\pi - \theta') & : \theta' \in [\frac{\pi}{2}, \pi] \\ -\cos(\theta' - \pi) & : \theta' \in [\pi, \frac{3\pi}{2}] \\ \cos(2\pi - \theta') & : \theta' \in [\frac{3\pi}{2}, 2\pi) \end{cases}$$

Using Algorithm 4.2.1 we can also easily calculate both $\sin\theta$ and $\tan\theta$, by further use of trigonometric identities. In particular we note that $\sin\theta = \cos(\theta - \frac{\pi}{2}$ and $\tan\theta = \frac{\sin\theta}{\cos\theta}$. Hence we can now calculate the trigonometric function value of any angle.

We now wish to analyse the error of our approximation for $\cos$, as the other methods have errors that are derivative of the error for approximating $\cos$. Now Figure 4.2.4 shows an arc of a circle which creates chord x, with this we will be able to calculate the exact length of the chord and thus work on the error of our approximations.

Figure 4.2.4: Diagram to find actual arc approximation



To start we will note that $\phi = \frac{\pi - \theta}{2} = \frac{\pi}{2} - \frac{\theta}{2}$, and then by using the Sine Law we get

$$\frac{x}{\sin\theta} = \frac{1}{\sin\phi} \implies x = \frac{\sin\theta}{\sin\phi}$$

Now we can recall the double angle formula for $\sin$, which gives $\sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2}$, and also $\sin\phi = \cos\frac{\theta}{2}$. This allows us to see that

$$x = \frac{2\sin\frac{\theta}{2}\cos\frac{\theta}{2}}{\cos\frac{\theta}{2}} = 2\sin\frac{\theta}{2}$$

Therefore we see that h_n is approximating the chord length associated with angle $\theta 2^{n-k}$, and thus $\epsilon_n = |h_n - 2\sin(\theta 2^{n-k-1})|$. Now as $h_0 = \theta 2^{-k} \approx 2\sin(\theta 2^{-k-1})$ then if follows that $\exists \phi$ such that $h_0 = 2\sin(\phi 2{-}k-1)$, from this we can see that $\phi = 2^{k+1}\sin^{-1}(\theta 2^{-k-1})$. We will uses these facts to prove a couple of propositions.

Proposition 4.2.1. $h_n = 2\sin(\phi 2^{n-k-1}) \forall n \in [0,k] \cap \mathbb{Z}$ *where* $\phi := 2^{k+1}\sin^{-1}(\theta 2^{-k-1})$.

Proof. Proceed by induction on $n \in [0,k] \cap \mathbb{Z}$.

$\mathbf{H}(n)$**:** $h_n = 2\sin(\phi 2^{n-k-1})$

$\mathbf{H}(0)$**:**

$$\begin{aligned} 2\sin(\phi 2^{-k-1}) &= 2\sin(\sin^{-1}(\phi 2^{-k-1})) \\ &= \theta 2^{-k} \\ &= h_0 && \text{by definition of } h_0 \end{aligned}$$

$\mathbf{H}(n) \implies \mathbf{H}(n+1)$**:**

$$\begin{aligned} h_{n+1} &= h_n\sqrt{4-h_n^2} \\ &= 2\sin(\phi 2^{n-k-1})\sqrt{4-4\sin^2(\phi 2^{n-k-1})} && \text{by H}(n) \\ &= 4\sin(\phi 2^{n-k-1})\cos(\phi 2^{n-k-1}) \\ &= 2\sin(\phi 2^{n-k}) && \text{by the use of double angle formulas} \end{aligned}$$

□

Proposition 4.2.2. $h_n > 2\sin(\theta 2^{n-k-1}) \forall n \in [0,k] \cap \mathbb{Z}$

Proof. We start by considering the expansion of the exact value of h_n.

$$\begin{aligned} h_n &= 2\sin(\phi 2^{n-k-1}) \\ &= 2\sin(2^{n-k-1}(2^{k+1}\sin^{-1}(\theta 2^{-k-1}))) \\ &= 2\sin(2^n \sin^{-1}(\theta 2^{-k-1})) \\ &= 2\sin(\theta 2^{n-k-1} + \tfrac{1}{6}\theta^3 2^{n-3k-3} + \mathcal{O}(2^{-5k})) \quad \text{Detailed in section } \textbf{??} \end{aligned}$$

Now as we know that $n \leq k$, then it follows that $\theta 2^{n-k-1} \leq \frac{1}{2}\theta$.

Also as $\theta \leq \frac{\pi}{2}$ we know that $\theta 2^{n-k-1} \leq \frac{\pi}{4}$.

We can also show that $\frac{1}{6}\theta^3 2^{n-3k-3} + \mathcal{O}(2^{-5k}) \leq \frac{\pi}{4}$, though the proof is ommited here for brevity; therefore we see that $\phi 2^{n-k-1} \leq \frac{\pi}{2}$, and obviously that $\phi 2^{n-k-1} > \theta 2^{n-k-1}$.

Hence, as $\sin$ is an increasing function in the range $[0, \frac{\pi}{2}]$, we conclude that

$$h_n = 2\sin(\phi 2^{n-k-1}) > 2\sin(\theta 2^{n-k-1})$$

. □

With these two propositions we can now consider the error of our approximation of $\cos$. First we will prove the following proposition regarding the error of the approximation of s:

Proposition 4.2.3. *If* $\epsilon_n := |h_n - 2\sin(\theta 2^{n-k-1})| \forall n \in [0,k] \cap \mathbb{Z}$*, then* $\epsilon_k < 2^k \epsilon_0)$.

Proof. $\epsilon_n = h_n - 2\sin(\theta 2^{n-k-1})$ as $h_n > 2\sin(\theta 2^{n-k-1})$ by Proposition 4.2.2.

Now we see that:

$$\begin{aligned}\epsilon_{n+1} &= h_{n+1} - 2\sin(\theta 2^{n-k}) \\ &= h_n\sqrt{4-h_n^2} - 4\sin(\theta 2^{n-k-1})\cos(\theta 2^{n-k-1})\end{aligned}$$

If we consider the equation $\alpha\beta - \gamma\delta = (\alpha - \gamma) + \alpha(\beta - 1) - \gamma(\delta - 1)$ and apply it to our current formula we get:

$$\begin{aligned}\epsilon_{n+1} &= (h_n - 2\sin(\theta 2^{n-k-1})) + h_n(\sqrt{4-h_n^2} - 1) - 2\sin(\theta 2^{n-k-1})(2\cos(\theta 2^{n-k-1}) - 1) \\ &= \epsilon_n + h_n(\sqrt{4-h_n^2} - 1) - 2\sin(\theta 2^{n-k-1})(2\cos(\theta 2^{n-k-1}) - 1) \\ &= 2\epsilon_n + h_n(\sqrt{4-h_n^2} - 2) - 2\sin(\theta 2^{n-k-1})(2\cos(\theta 2^{n-k-1}) - 2) \\ &= 2\epsilon_n + h_n(\sqrt{4-h_n^2} - 2) + 2\sin(\theta 2^{n-k-1})(2 - 2\cos(\theta 2^{n-k-1})) \\ &< 2\epsilon_n + h_n(\sqrt{4-h_n^2} - 2\cos(\theta 2^{n-k-1})) \\ &< 2\epsilon_n + h_n(\sqrt{4-4\sin^2(\theta 2^{n-k-1})} - 2\cos(\theta 2^{n-k-1})) \\ &= 2\epsilon_n + h_n(2\cos(\theta 2^{n-k-1}) - 2\cos(\theta 2^{n-k-1})) \\ &= 2\epsilon_n\end{aligned}$$

The inequalities in the above derivation arrise from the fact that $h_n > 2\sin(\theta 2^{n-k-1})$ by Proposition 4.2.2.

Hence as we now know that $\epsilon_{n+1} < 2\epsilon_n$, we then see that $\epsilon_n < 2^n\epsilon_0$. Therefore we prove our statement that

$$\epsilon_k < 2^k\epsilon_0$$

□

Obviously $\epsilon_k = |h_k - s|$, and we can now use this to find the error of our final answer. First we will start by letting $\mathcal{C} := 1 - \frac{1}{2}h_k^2$ and note that analytically $cos\theta = 1 - \frac{1}{2}s^2$. Therefore we will now consider $\epsilon_{\mathcal{C}} = |\mathcal{C} - \cos(\theta)|$:

$$
\begin{aligned}
\epsilon_{\mathcal{C}} &= |1 - \frac{h_k^2}{2} - 1 + \frac{s^2}{2}| \\
&= \frac{1}{2}|h_k^2 - s^2| \\
&= \frac{1}{2}|h_k h_k - 2\sin(\frac{\theta}{2})2\sin(\frac{\theta}{2})| \\
&= \frac{1}{2}(h_k h_k - 2\sin(\frac{\theta}{2})2\sin(\frac{\theta}{2}) && \text{as } 2\sin(\frac{\theta}{2}) < h_k \\
&= \frac{1}{2}(2\epsilon_k + h_k(h_k - 2) - 2\sin(\frac{\theta}{2})(2\sin(\frac{\theta}{2}) - 2) \\
&< \frac{1}{2}(2\epsilon_k + h_k(h_k - 2\sin(\frac{\theta}{2}))) \\
&= \frac{1}{2}(2 + h_k)\epsilon_k \\
&= \frac{1}{2}(2 + 2\sin(\frac{\phi}{2}))\epsilon_k \\
&= (1 + \sin(\frac{\phi}{2}))\epsilon_k \\
&\leq 2\epsilon_k
\end{aligned}
$$

As $\epsilon_{\mathcal{C}} \leq 2\epsilon_k$, then by Proposition 4.2.3 we see that $\epsilon_{\mathcal{C}} < 2^{k+1}\epsilon_0$. Now to consider ϵ_0 we first observe that $\epsilon_0 = \theta 2^{-k} - 2\sin\theta 2^{-k-1}$, and therefore we can conclude that:

$$\epsilon_{\mathcal{C}} < 2\theta - 2^{k+2}\sin(\theta 2^{-k-1})$$

If we then wish to calculate $\cos\theta$ accurate to N decimal places then we are looking to find $k \in \mathbb{N}$ such that

$$2\theta - 2^{k+2}\sin(\theta 2^{-k-1}) < 10^{-N} \implies 2^{k+2}\sin(\theta 2^{-k-1}) > 2\theta - 10^{-N}$$

For an example of the above in action we will be taking $\theta = 0.5$. The table below shows the minimum $k \in \mathbb{N}$ to guarantee N digits of accuracy in the result:

N	k
5	6
10	14
50	80
100	163
1000	1658

As can be seen the value of k required to acheive N digits of accuracy increases roughly linearly when $\theta = 0.5$. Testing for other values of θ reveals them to have similar required values for k, at least within the same order of each other.

Another consideration for Algorithm 4.2.1 is that we could "run it in reverse" to attain an algorithm for the inverse cosine function. To start take line 7 which is $\mathcal{C} = 1 - \frac{1}{2}h_k^2$, which can be re-arranged to give $h_k^2 = 2 - 2\mathcal{C}$, where we know $\mathcal{C}$ as our initial value.

Line 5 is a little more difficult, but by re-arranging we see that $h_n^4 - 4h_n^2 + h_{n+1}^2 = 0$, which can be solved via the quadratic formula to give $h_n^2 = 2 \pm \sqrt{4 - h_{n+1}^2}$. Now we can make the observation that if $x \in \mathbb{R}_0^+$, then $\cos^{-1}(-x) = \pi - \cos^{-1}(x)$ and so we can restric our algorithm to only consider $x \in [0, 1]$. With this we know that $\theta \in [0, \frac{\pi}{2}]$, and thus $h_k \leq \sqrt{2}$. Therefore as $h_{n+1} > h_n \forall n \in [0, k-1] \cap \mathbb{Z}$ we see that $h_n^2 \leq 2 \forall n \in [0, k] \cap \mathbb{Z}$. This allows us to ascertain that to reverse Line 5 we perform $h_n^2 = 2 - \sqrt{4 - h_{n+1}^2}$.

Finally line 2 is reversed by returning the value $2^k h_0$; therefore we get the following algorithm for $\cos^{-1}(x)$ where $x \in [0, 1]$:

Algorithm 4.2.2: Geometric calculation of $\cos^{-1}$

```
1  geometric_aCos(x ∈ [0,1], k ∈ ℕ)
2      h_k := 2 − 2x
3      n := k − 1
4      while n ≥ 0:
5          h_n^2 := 2 − √(4 − h_{n+1}^2)
6          n ↦ n − 1
7      return 2^k h_0
```

Similar to the regular trigonometric functions we can use trigonometric identities to calculate the inverse trigonometric functions from $\cos^{-1}$. To start we recall that $\cos^{-1}(-x) = -\cos(x)$ where $x \in [0, 1]$, then we can use the identities that $\sin^{-1}(x) = \frac{\pi}{2} - \cos^{-1}(x)$ and $\tan^{-1}(x) = \sin^{-1}(\frac{x}{\sqrt{x^2+1}})$.

If we suppose that all operations in the method are accurately computed then Algorithm 4.2.2 is a computation with high accuracy. This is because there is no initial guess, such as in Algorithm 4.2.1, and so the only introduction of error is assuming that $2^k h_0 \approx \theta$. However as we discuss in detail in Section **??**, calculating square roots is not a simple task and thus will introduce error to the method in general; therefore the accuracy of the method is roughly as accurate as our method of calculating square roots.

4.3 Taylor Series

If we consider our definition of a McClaurin Series from Section **??**, we can use this to approximate our Trigonometric Functions. Consider first $\cos\theta$, for which we know that $\frac{d}{d\theta}\cos\theta = -\sin\theta$; it then follows that $\frac{d^2}{d\theta^2}\cos\theta = -\cos\theta$, $\frac{d^3}{d\theta^3}\cos\theta = \sin\theta$ and $\frac{d^4}{d\theta^4}\cos\theta = \cos\theta$.

If we let $f(x) = \cos x$ and use the known values $\cos(0) = 1$ and $\sin(0) = 0$, then we see that:

$$f^{(n)}(0) = \begin{cases} 1 & : \ 4 \mid n \\ 0 & : \ 4 \mid n-1 \\ -1 & : \ 4 \mid n-2 \\ 0 & : \ 4 \mid n-3 \end{cases}$$

By simplifying this by ommitting the 0 coefficient terms we get the following series:

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots \tag{4.3.1}$$

By using similar working we can get that the series associated with $\sin x$):

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots \tag{4.3.2}$$

Before we go any further we need to consider when Equations 4.3.1 and 4.3.2 converge to their respective functions. To do this we will use the ratio test for series as defined in **??**, using Equation 4.3.1 we see that

$$\begin{aligned} L_{\mathcal{C}} &= \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| \\ &= \lim_{n\to\infty} \left| \frac{\frac{(-1)^{n+1}}{(2n+2)!} x^{2n+2}}{\frac{(-1)^n}{(2n)!} x^{2n}} \right| \\ &= \frac{(2n)!}{(2n+2)!} |x|^2 \\ &= \frac{1}{(2n+2)(2n+1)} |x|^2 \end{aligned}$$

Now it is easy to see that, $L_{\mathcal{C}} = 0$ for all values of x as the fractional component decreases as n increases and $|x|^2$ is a constant. Therefore we can conclude that Equation 4.3.1 converges to $\cos(x)$ for all values of x. We can use a very similar deduction to show that Equation 4.3.2 converges to $\sin(x)$ for all values of x.

The above means that $\cos$ and $\sin$ can be approximated using Taylor Polynomials, in particular for a given $N \in \mathbb{N}$:

$$\cos x \approx \sum_{n=0}^{N} \tfrac{(-1)^n}{(2n)!} x^{2n} \quad \text{and} \quad \sin x \approx \sum_{n=0}^{N} \tfrac{(-1)^n}{(2n+1)!} x^{2n+1}$$

This allows us to create the following two methods for computing $\cos x$ and $\sin x$:

Algorithm 4.3.1: Taylor computation of $\cos$ and $\sin$

```
taylor_cos(x ∈ ℝ, N ∈ ℕ)
    C := 0
    n := 0
    while n < N:
        C ↦ C + (-1)^n · 1/(2n)! x^(2n)
        n ↦ n + 1
    return C

taylor_sin(x ∈ ℝ, N ∈ ℕ)
    S := 0
    n := 0
    while n < N:
        S ↦ S + (-1)^n · 1/(2n+1)! x^(2n+1)
        n ↦ n + 1
    return S
```

4.4 CORDIC

CORDIC is an algorithm that stands for COrdinate Rotation DIgital Computer and can be used to calculate many functions, including Trigonometric Values. The CORDIC algorithm works by utilising Matrix Rotations of unit vectors. This algorithm is less accurate than some other methods but has the advantage of being able to be implemented for fixed point real numbers in efficient ways using only addition and bitshifting.

CORDIC works by taking an initial guess of $\mathbf{x}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ which can be rotated through an anti-clockwise angle of γ by the matrix

$$\begin{pmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{pmatrix} = \frac{1}{\sqrt{1+\tan\gamma^2}} \begin{pmatrix} 1 & -\tan\gamma \\ \tan\gamma & 1 \end{pmatrix}$$

By taking taking smaller and smaller values of γ we can create an iterative process to find $\mathbf{x}_n$ which converges, for a given $\beta \in (-\frac{\pi}{2}, \frac{\pi}{2})$, to

$$\begin{pmatrix} \cos\beta \\ \sin\beta \end{pmatrix}$$

7 Preliminary References

http://math.exeter.edu/rparris/peanut/cordic.pdf
Inside your Calculator by Gerald R Rising
Wolfram Alpha