The Miracle Substitution: How and Why It Works

William M. Faucette

September 2002

When I learned the techniques of integration in my college calculus class, my classmates and I learned all the usual techniques—u-substitution, integration by parts, trigonometric substitutions, partial fractions, and so forth. All these techniques are, of course, still part of every second semester calculus course, but there seems to be one technique of integration which has lost favor over the years. Among recent texts, [2, pp. 542–543], [3, p. 570], and [9, pp. 498–499] relegate this method of integration to problems at the end of a section. Of the calculus texts I have examined, only [1, pp. 548–549] still treats this method within the text itself.

The method to which I refer is the standard technique for integrating rational functions of $\sin\theta$ and $\cos\theta$ using what I have always known as "the miracle substitution". The substitution $t = \tan(\theta/2)$ will always reduce any rational function of $\sin\theta$ and $\cos\theta$ to a rational function of t, which can then be solved by the technique of partial fractions. According to [9, pp. 498–499], this substitution was first noticed by Karl Weierstrass (1815–1897). Of course, in freshman level calculus classes, the reason it is called "the miracle substitution" is that the substitution is completely unmotivated ... but it works.

In this paper, I will recall "the miracle substitution," and how it works, but I will then proceed to explain *why* the miracle substitution works using the fundamentals of rational curves from algebraic geometry.

I would like to express my thanks to Ted Shifrin who so clearly explained these notions in his March, 2001, talk at the Southeastern regional meeting of the Mathematical Association of America held at Huntingdon College in Montgomery, Alabama. He made me realize—for the first time—why "the miracle substitution" works.

1 What Is The Miracle Substitution?

Let $r(\theta)$ be any rational function of $\sin(\theta)$ and $\cos(\theta)$. The miracle substitution sets

$$t = \tan\left(\frac{\theta}{2}\right) \ .$$

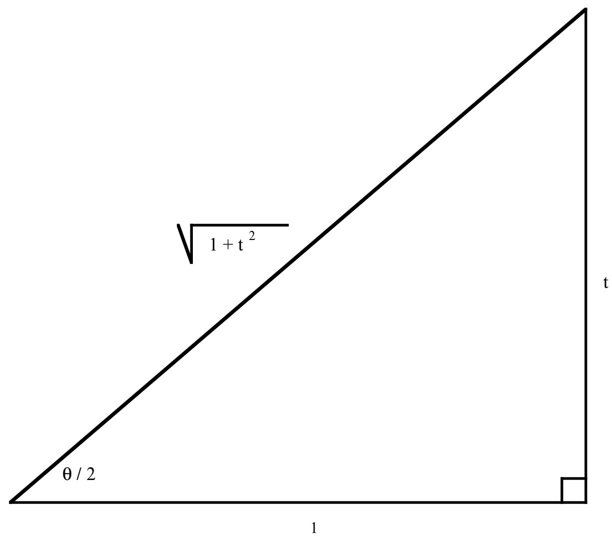

Figure 1.

From Figure 1, we see that

$$\begin{aligned} \cos\left(\frac{\theta}{2}\right) &= \frac{1}{\sqrt{1+t^2}} \\ \sin\left(\frac{\theta}{2}\right) &= \frac{t}{\sqrt{1+t^2}} \ . \end{aligned}$$

Now, using the standard double-angle formulas from plane trignometry, we have

$$\begin{aligned} \cos\theta &= \cos^2(\theta/2) - \sin^2(\theta/2) \\ &= \left[\frac{1}{\sqrt{1+t^2}}\right]^2 - \left[\frac{t}{\sqrt{1+t^2}}\right]^2 \\ &= \frac{1}{1+t^2} - \frac{t^2}{1+t^2} \\ &= \frac{1-t^2}{1+t^2} \ . \\ \sin\theta &= 2\sin(\theta/2)\cos(\theta/2) \\ &= 2\frac{1}{\sqrt{1+t^2}}\frac{t}{\sqrt{1+t^2}} \\ &= \frac{2t}{1+t^2} \ . \end{aligned}$$

Finally,

$$\begin{aligned}
t &= \tan\left(\frac{\theta}{2}\right) \\
\theta &= 2\arctan(t) \\
d\theta &= 2\frac{dt}{1+t^2} \\
&= \frac{2\,dt}{1+t^2},
\end{aligned}$$

abusing the notation, as is customary in freshman calculus courses.

From these computations, it is clear that any integrand containing a rational function of $\sin\theta$ and $\cos\theta$ can be converted to an integrand containing a rational function of t.

2 Example

Since this technique of integration seems to be neglected recently, let's see how it works in an example. This example is taken from [9, p. 499, #59].

EXAMPLE: Compute

$$\int \frac{1}{3\sin\theta - 4\cos\theta}\,d\theta.$$

SOLUTION: Using the miracle substitution,

$$\begin{aligned}
\cos\theta &= \frac{1-t^2}{1+t^2} \\
\sin\theta &= \frac{2t}{1+t^2} \\
d\theta &= \frac{2\,dt}{1+t^2},
\end{aligned}$$

we have

$$
\begin{aligned}
\int \frac{1}{3\sin\theta - 4\cos\theta}\, d\theta &= \int \frac{1}{3\left(\frac{2t}{1+t^2}\right) - 4\left(\frac{1-t^2}{1+t^2}\right)} \frac{2\,dt}{1+t^2} \\
&= \int \frac{2}{3\,(2t) - 4\,(1-t^2)}\, dt \\
&= \int \frac{2}{4t^2 + 6t - 4}\, dt \\
&= \int \frac{1}{2t^2 + 3t - 2}\, dt \\
&= \int \frac{1}{2t^2 + 3t - 2}\, dt \\
&= \int \frac{1}{(2t-1)(t+2)}\, dt \\
&= \int \frac{2/5}{2t-1} + \frac{-1/5}{t+2}\, dt \\
&= \frac{1}{5}\ln|2t-1| - \frac{1}{5}\ln|t+2| + C \\
&= \frac{1}{5}\ln\left|\frac{2t-1}{t+2}\right| + C \\
&= \frac{1}{5}\ln\left|\frac{2\tan\left(\frac{\theta}{2}\right) - 1}{\tan\left(\frac{\theta}{2}\right) + 2}\right| + C.
\end{aligned}
$$

3 Why Does The Miracle Substitution Work?

Referring to Figure 2, we consider the unit circle, $x^2 + y^2 = 1$, and an angle θ in standard position. Let the point of intersection of the terminal side of angle θ and the unit circle have coordinates (x, y). Draw a segment from the point $(-1, 0)$ to the point of (x, y). The acute angle formed at $(-1, 0)$ then has measure $\theta/2$. Let t be the distance from the origin to the point of intersection of the y-axis with the terminal side of the ang1e $\theta/2$. Notice that t is the slope of the line from $(-1, 0)$ to (x, y), so

$$t = \frac{y}{x+1}\ .$$

Now, we wish to find the coordinates x and y in terms of t. To do this, we simply solve the equations of the line $t = y/(x+1)$ and the circle $x^2 + y^2 = 1$

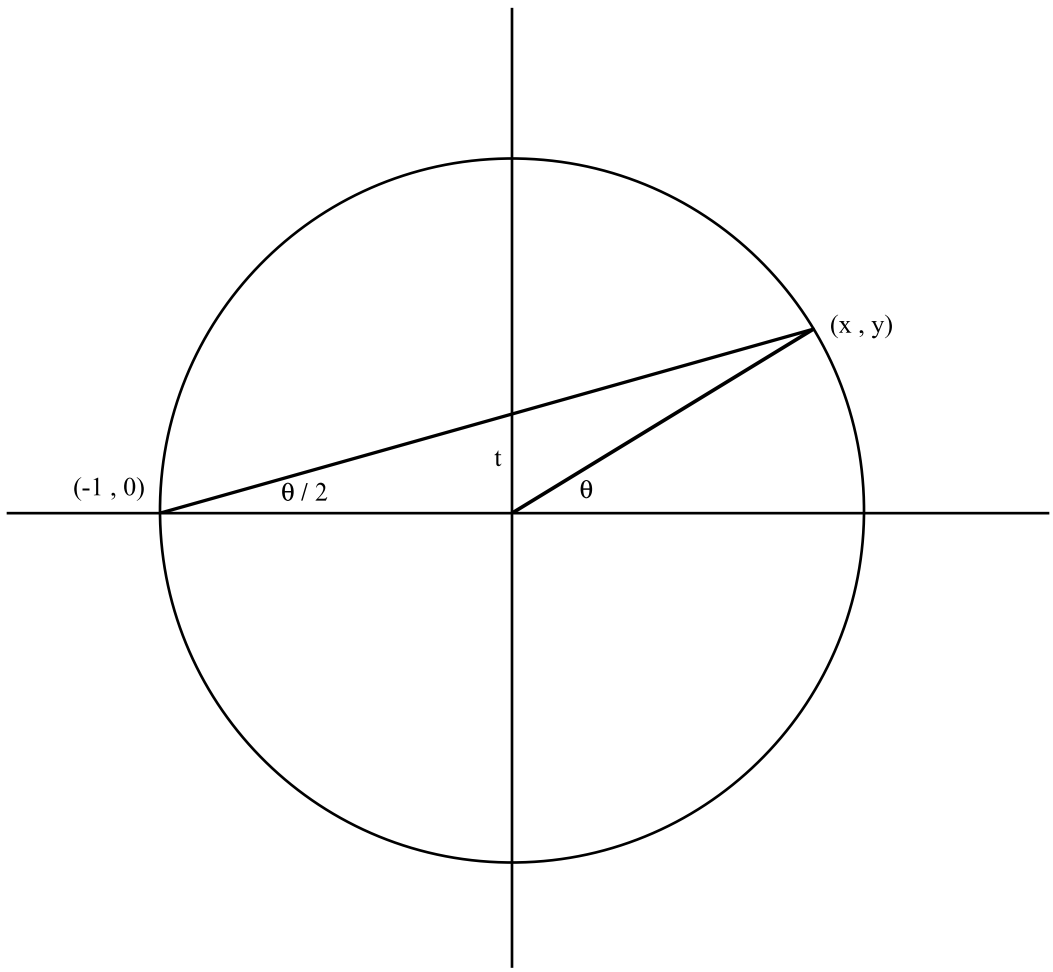

Figure 2.

simulataneously:

$$\begin{aligned} x^2 + y^2 &= 1 \\ x^2 + (t(x+1))^2 &= 1 \\ x^2 + t^2x^2 + 2t^2x + t^2 &= 1 \\ x^2 + t^2x^2 + 2t^2x + t^2 - 1 &= 0 \\ x^2(1+t^2) + 2t^2x + (t^2-1) &= 0\ . \end{aligned}$$

Using the quadratic formula, we have

$$\begin{aligned} x &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\ &= \frac{-2t^2 \pm \sqrt{(2t^2)^2 - 4(1+t^2)(t^2-1)}}{2(1+t^2)} \\ &= \frac{-2t^2 \pm \sqrt{4t^4 - 4(t^4-1)}}{2(1+t^2)} \\ &= \frac{-2t^2 \pm 2}{2(1+t^2)} \\ &= \frac{-t^2 \pm 1}{1+t^2} \\ &= -1 \text{ or } \frac{1-t^2}{1+t^2}\ . \end{aligned}$$

Ignoring the root -1, which corresponds to the point $(-1, 0)$, we see that

$$x = \frac{1-t^2}{1+t^2}$$

and

$$y = t(x+1) = t\left(\frac{1-t^2}{1+t^2} + 1\right) = \frac{2t}{1+t^2}\ .$$

Notice that we have defined functions

$$\begin{aligned} f : S^1 \setminus \{(-1,0)\} &\to \mathbb{R} \\ (x,y) &\mapsto \frac{y}{x+1} \end{aligned}$$

and

$$\begin{aligned} g : \mathbb{R} &\to S^1 \setminus \{(-1,0)\} \\ t &\mapsto \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right) \end{aligned}$$

which are inverse functions. Using the terminology of algebraic geometry, this may be viewed as a birational map of a smooth conic with the affine line.

Let's expand this view a bit. Viewing the unit circle as the real part of the affine conic, C, with equation $z^2 + w^2 = 1$ and the real line as the real part of the complex line, we have

$$\begin{aligned} g : \mathbb{C} \setminus \{\pm i\} &\rightarrow C \\ t &\mapsto \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2} \right) . \end{aligned}$$

Expanding a bit further by viewing these two curves as living in projective space, we have

$$\begin{aligned} \tilde{g} : \mathbb{P}^1 &\rightarrow C \\ [t, s] &\mapsto \left[s^2 - t^2, 2st, s^2 + t^2 \right] , \end{aligned}$$

which is the standard biregular morphism of algebraic varieties between the projective line and the smooth conic, there being only one up to projective equivalence. This biregular morphism induces an isomorphism

$$\tilde{g}^* : \Omega(C) \rightarrow \Omega(\mathbb{P}^1)$$

of the algebras of rational differential forms. For more information on affine and projective plane curves, rational functions, and differential forms, see [5], [6], [8], [4], and [7].

4 Let's Get Back to the Problem

Let's say you are given a rational function of $\sin\theta$ and $\cos\theta$. Referring to Figure 2, we see that $x = \cos\theta$ and $y = \sin\theta$, so we may view an integrand containing a rational function of $\sin\theta$ and $\cos\theta$ as a rational differential form of x and y defined on the unit circle. Now, as in the last section, the map $\tilde{g}$ induces an isomorphism

$$\tilde{g}^* : \Omega(C) \rightarrow \Omega(\mathbb{P}^1)$$

of the algebras of rational differential forms. Hence, we can take our rational function of x and y defined on the unit circle and pull it back to a rational differential form on $\mathbb{P}^1$.

What do these rational differential forms look like? It is well known that if t is an affine coordinate on $\mathbb{P}^1$, then any rational differential form on $\mathbb{P}^1$ is a rational function of t times dt.

5 Ahha!

So, the miracle substitution is not nearly as miraculous as it appears to freshman calculus classes. The substitution $t = \tan(\theta/2)$ transforms a rational integral of $\sin\theta$ and $\cos\theta$ into a rational integral of t because it is an implementation of standard results in algebraic geometry: A nonsingular conic in the plane is isomorphic as a projective algebraic variety to the the projective line and all rational differential forms on the projective line are rational functions of a local coordinate.

References

[1] H. Anton. *Calculus (Brief Edition), Sixth Edition.* John Wiley & Sons, 1999.

[2] C. H. Edwards and D. E. Penney. *Calculus, Sixth Edition.* Prentice-Hall, New York, 2002.

[3] R. L. Finney, M. D. Weir, and F. R. Giordano. *Thomas' Calculus, Tenth Edition.* Addison Wesley, 2001.

[4] W. Fulton. *Algebraic Curves.* Benjamin/Cummings Publishing Company, Reading, Massachusetts, 1974.

[5] P. A. Griffith and J. Harris. *Principles Algebraic Geometry.* John Wiley & Sons, 1978.

[6] P. A. Griffiths. *Introduction to Algebraic Curves.* American Mathematical Society, 1989.

[7] R. Miranda. *Algebraic Curves and Riemann Surfaces.* American Mathematical Society, 1995.

[8] I. R. Shafarevich. *Basic Algebraic Geometry: Varieties in Projective Space.* Springer-Verlag, New York, 1994.

[9] J. Stewart. *Calculus: Early Transcendentals, Fourth Edition.* Brooks/Cole, 1999.