

How to Avoid the Inverse Secant (and Even the Secant Itself)

S. A. Fulling



Stephen A. Fulling (fulling@math.tamu.edu) is Professor of Mathematics and of Physics at Texas A&M University (College Station, TX 77843). He has an A.B. from Harvard and a Ph.D. from Princeton, both in physics. He has written a book on quantum field theory in curved space-time and a textbook on linear algebra with emphasis on applications to vector calculus and differential equations. He has done research in various areas of mathematical physics and differential equations, and he has taken part in a variety of curriculum-reform efforts at the undergraduate level.

An introductory polemic

An irritating technicality that must be handled somehow in a standard calculus course is the definition of the inverse secant function at negative arguments. There is no particularly natural choice of principal branch from among the infinitely many candidates, and the derivatives of different branches can differ in sign (see the figure below). Some textbook authors adopt the convention

$$\pi \leq \sec^{-1} x < \frac{3\pi}{2} \quad \text{for } x \leq -1, \quad \frac{d}{dx} \sec^{-1} x = \frac{1}{x\sqrt{x^2-1}}, \quad (1)$$

while others prefer

$$\frac{\pi}{2} < \sec^{-1} x \leq \pi \quad \text{for } x \leq -1, \quad \frac{d}{dx} \sec^{-1} x = \frac{1}{|x|\sqrt{x^2-1}}. \quad (2)$$

Then authors and lecturers have a responsibility to warn students about the existence of the other convention.

One might have expected that after more than a decade of calculus reform, the secant function and its inverse would have been de-emphasized to the vanishing point, along with its even less useful siblings, cosecant and cotangent, and their inverses. The persistence of \sec^{-1} presumably stems from the perceived need to provide a formula (see (17)) for the indefinite integral

$$\int \frac{dx}{x\sqrt{x^2-1}} \quad (3)$$

by inverting whichever of the differentiation formulas (1) and (2) one adopts. More generally, it is argued that the secant and tangent functions unavoidably arise when algebraic functions are integrated by trigonometric substitutions, so students should have at least a nodding acquaintance with the variety of integrals involving them.

There is, however, an alternative approach to integrals involving $\sqrt{x^2-1}$ or $\sqrt{x^2+1}$, which has unaccountably fallen out of favor in recent decades: *hyperbolic substitution*.

The hyperbolic sine and cosine are defined by

$$\cosh u \equiv \frac{1}{2}(e^u + e^{-u}), \quad \sinh u \equiv \frac{1}{2}(e^u - e^{-u}). \quad (4)$$

Their usefulness as replacements for trigonometric functions in integration by substitution stems from the identities

$$\cosh^2 u - \sinh^2 u = 1, \quad \frac{d}{du} \cosh u = \sinh u, \quad \frac{d}{du} \sinh u = \cosh u. \quad (5)$$

The hyperbolic functions (4) are important in their own right, constituting the natural basis of solutions of the differential equation

$$\frac{d^2 y}{du^2} = y$$

satisfying unit Dirichlet and Neumann initial data at $u = 0$; in this role they are indispensable in upper-division courses in applied mathematics. In a linear algebra course, (4) provides a beautiful example of a change of basis, of undeniable practical importance, in a real vector space with no intrinsic inner product or geometrical interpretation. Most important in the present context, hyperbolic substitutions are much simpler and nicer than trigonometric substitutions of the tangent and secant varieties. For one thing, the inverse hyperbolic functions can be expressed in terms of more elementary functions (logarithms and square roots—see (11) and (12)). Furthermore, the branch structure of these inverses is very simple: \sinh is bijective, and \cosh has a two-branched inverse (just like the square root), with no multiples of $\frac{\pi}{2}$ to be memorized. Nevertheless, many teachers of calculus have a strange dislike for the hyperbolic functions and prefer not to cover them at all. (An independent case in favor of the hyperbolics has been made by Gearhart and Shultz [1].)

In this article I hope to convince the reader that there is nothing that the secant and inverse secant do in the traditional “techniques of integration” chapter that cannot be done better by the hyperbolic sine and cosine and their inverses. *It is time for sec, csc, cot, sec⁻¹, csc⁻¹, and cot⁻¹ to be retired from our calculus syllabus, replaced by sinh and cosh. Our students will learn about two elegant and useful transcendental functions while being freed from six complicated and boring ones.*

The rest of the article has two parts. First, we derive several alternative antiderivative formulas, (13), (14), (16), with clear advantages over the traditional formulas (17). Then we’ll see that hyperbolic substitution provides an easy way of evaluating or evading (depending on context) the difficult integrals of powers of the secant that crop up in trigonometric substitutions.

The integral formerly known as \sec^{-1}

Since the integration problem (3) is the inverse secant’s alleged reason for existence, let us see what a hyperbolic substitution does to it. Recall first that (3) makes no sense (in real analysis) unless $|x| \geq 1$. We assume temporarily that $x > 0$, hence $x \geq 1$, and set

$$x = \cosh u \quad (u \geq 0); \quad dx = \sinh u \, du, \quad x^2 - 1 = \sinh^2 u. \quad (6)$$

Therefore,

$$\int \frac{dx}{x\sqrt{x^2-1}} = \int \frac{du}{\cosh u}. \quad (7)$$

(As an aside, note that not much is gained here by introducing the name $\operatorname{sech} u$ for the integrand of (7).) To continue we need

$$\int \frac{du}{\cosh u} = 2 \tan^{-1}(e^u) + C, \quad (8)$$

which can be either verified by differentiation or “discovered” through these intermediate steps:

$$\int \frac{2 du}{e^u + e^{-u}} = \int \frac{2e^u du}{e^{2u} + 1} = \int \frac{2 dv}{v^2 + 1} = 2 \tan^{-1} v + C.$$

Formula (8) is arguably less recondite than

$$\int \sec \theta d\theta = \ln |\sec \theta + \tan \theta| + C, \quad (9)$$

which inevitably plays a leading role in traditional treatments of trigonometric substitution. (And alas, there is seldom time in class to present the interesting history [3] of (9).)

Remark. The function

$$\operatorname{gd} u \equiv 2 \tan^{-1}(e^u) - \frac{\pi}{2}, \quad (10)$$

which satisfies the convenient initial condition $\operatorname{gd} 0 = 0$, is called the *Gudermannian* and can be used to express hyperbolic functions as trigonometric functions (of a different variable) and vice versa [2, Secs. 1.48 and 1.49]; for example, if $\theta = \operatorname{gd} u$, then $\sec \theta = \cosh u$, a formula quite pertinent to the equivalence of (14) and (16) below. For more on the history and applications of gd , see Robertson [4].

As previously remarked, one of the great charms of hyperbolic functions (as opposed to trig functions) is that their inverses can be expressed in terms of already familiar functions:

$$\sinh^{-1} x = \ln(x + \sqrt{x^2 + 1}) \quad \text{for all } x; \quad (11)$$

$$\cosh^{-1} x = \ln(x + \sqrt{x^2 - 1}) \quad \text{for } x \geq 1, \quad (12)$$

where \cosh^{-1} denotes the positive branch. Note that the right-hand side of (11) is indeed an odd function, though it may not look like one. To prove (11) and (2), simply apply the definitions (4) to their right-hand sides and simplify down to x .

Combining (7), (8), (6), and (12), one arrives at

$$\int \frac{dx}{x\sqrt{x^2-1}} = 2 \tan^{-1}(x + \sqrt{x^2-1}) + C, \quad (13)$$

at least for $x \geq 1$. It is now an elementary, though lengthy, exercise in differentiation to verify (13) for $x \leq -1$ as well. However, an alternative approach leads to a neater result. For negative x we can let $x = -\cosh u$ and repeat the previous calculation to obtain

$$\int \frac{dx}{x\sqrt{x^2-1}} = 2 \tan^{-1}(|x| + \sqrt{x^2-1}) + C, \quad (14)$$

which of course agrees with (13) in the positive case. The formula (14) is *even in x* , whereas (13) is not. This is not a contradiction: The constants of integration on the two disconnected domains, $x \geq 1$ and $x \leq -1$, are independent. At negative x , (13) and (14) with the same C simply differ by a constant (see the figure).

To relate these formulas to the traditional ones based on (1) and (2), we start with a known but nontrivial trigonometric identity,

$$\sec \theta + \tan \theta = \tan \left(\frac{\theta}{2} + \frac{\pi}{4} \right), \quad (15)$$

whose proof we omit. Now let $|x| = \sec \theta$ with θ in the first quadrant. Then $\tan \theta = \sqrt{x^2-1}$, and (14) becomes, up to the arbitrary constant,

$$\begin{aligned} \int \frac{dx}{x\sqrt{x^2-1}} &= 2 \tan^{-1}(\sec \theta + \tan \theta) \\ &= 2 \left(\frac{\theta}{2} + \frac{\pi}{4} \right) \\ &= \sec^{-1} |x| + \frac{\pi}{2}. \end{aligned}$$

In other words,

$$\int \frac{dx}{x\sqrt{x^2-1}} = \sec^{-1} |x| + C \quad (16)$$

is an alternative formula for the indefinite integral (for either sign of x). Similarly, after appropriate bookkeeping with quadrants, one can relate (13) or (14) at negative x to one's favorite definition of \sec^{-1} there.

It is to be hoped that any authors and lecturers who are still unconvinced of the virtues of hyperbolic substitution will at least adopt (16) in place of either of the traditional formulas,

$$\int \frac{dx}{x\sqrt{x^2-1}} = \sec^{-1} x + C, \quad \int \frac{dx}{|x|\sqrt{x^2-1}} = \sec^{-1} x + C. \quad (17)$$

For (16) is analogous to the familiar formula

$$\int \frac{dx}{x} = \ln |x| + C, \quad (18)$$

it is even in x (unlike either of (17)), and it avoids any reference to the inverse secant function at negative arguments!

Our conclusions so far are summarized and made more precise in the following theorem and figure.

Theorem. Define

$$f_1(x) = 2 \tan^{-1}(|x| + \sqrt{x^2 - 1}),$$

$$f_2(x) = 2 \tan^{-1}(x + \sqrt{x^2 - 1}),$$

$$f_3(x) = \sec^{-1} |x|.$$

Further, let $f_4(x)$ be the branch of the inverse secant determined by (1), and $f_5(x)$ the branch determined by (2) (with $0 \leq \sec^{-1} x < \pi/2$ for $x \geq 1$). Then

- (a) f_1 and f_3 (and only they) are even functions, and $f_1 = f_3 + \pi/2$.
- (b) On the interval $[1, \infty)$, all five functions are antiderivatives of $1/(x\sqrt{x^2 - 1})$, and $f_1 = f_2$, $f_3 = f_4 = f_5$.
- (c) On the interval $(-\infty, -1]$, the first four functions and $-f_5$ are antiderivatives of $1/(x\sqrt{x^2 - 1})$, and $f_2 = f_1 - \pi$, $f_4 = f_3 + \pi$, $f_5 = \pi - f_3$.

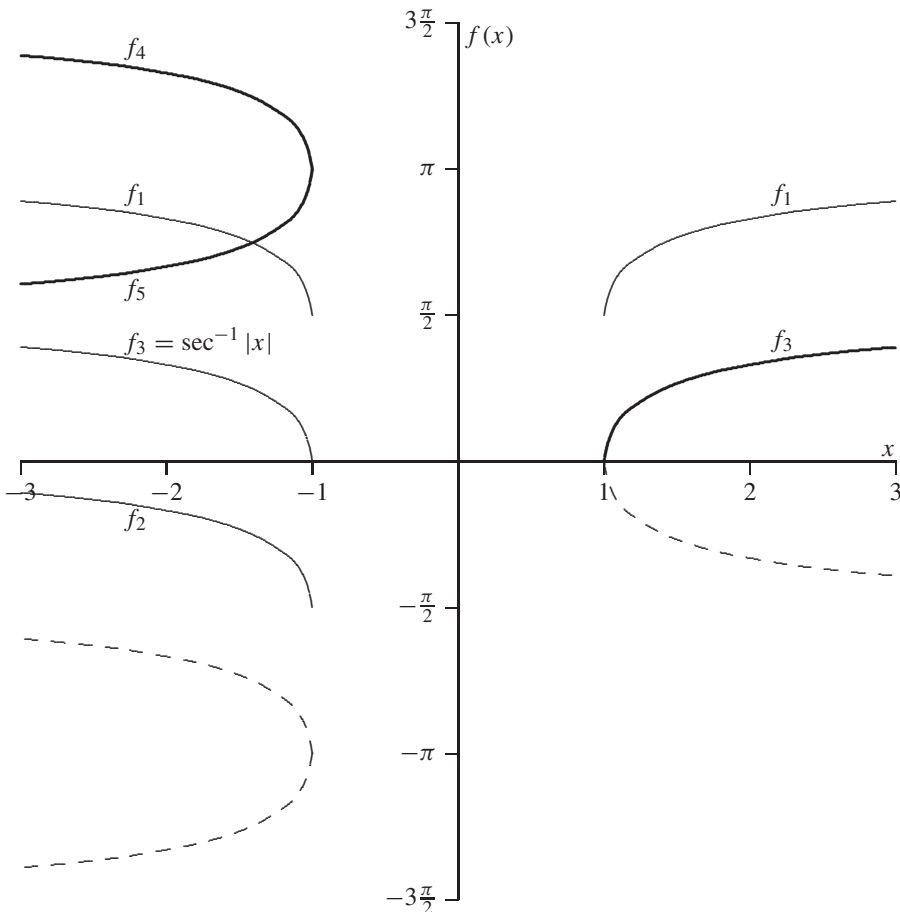


Figure 1. Graphs of the functions f_1 – f_5 discussed in the theorem. (Redundant names for the curves on the right are omitted.) Thick curves are the traditional rival branches of the inverse secant. Dashed curves are graphs of other logically possible definitions of \sec^{-1} . Thin solid curves are not branches of $\sec^{-1} x$, but nevertheless are useful as antiderivatives of $1/x\sqrt{x^2 - 1}$.

Integrating powers of the secant

Trigonometric substitution has an unpleasant habit of resulting in integrals like the one in (9), or, more generally, $\int \sec^p \theta d\theta$ with p a positive integer, that look at least as hard as the original algebraic integration problem. Making the natural substitution $x = \tan \theta$ ($-\frac{\pi}{2} < \theta < \frac{\pi}{2}$) yields

$$\int \sec^p \theta d\theta = \int (x^2 + 1)^{(p-2)/2} dx. \quad (19)$$

The right-hand side of (19) is elementary if p is even, but what if p is odd? The advice given to the student by the traditional textbook is, “Make the trigonometric substitution $x = \tan \theta$,” which takes us straight back to the left-hand side of (19). Velleman [5] (see also [3]) shows that the substitution $y = \sin \theta$ turns the left-hand side into the integral of a rational function, which can be integrated by partial fractions. Here we investigate what hyperbolic substitution in the right-hand side has to offer.

The appropriate substitution this time is

$$x = \sinh u; \quad dx = \cosh u du, \quad \sqrt{x^2 + 1} = \cosh u. \quad (20)$$

It turns (19) into $\int \cosh^{p-1} u du$.

Let's concentrate first on the case $p = 1$ (that is, (9)):

$$\begin{aligned} \int (x^2 + 1)^{-1/2} dx &= \int du = u + C \\ &= \sinh^{-1} x + C \\ &= \ln(x + \sqrt{x^2 + 1}) + C \end{aligned} \quad (21)$$

by (11). If our original interest was in integrating the algebraic function, we are done; if we really cared about the secant integral for its own sake, we now use $x = \tan \theta$ to go from (21) to (9) in the quadrants where $\sec \theta > 0$. (How to handle the other quadrants is left to the reader's taste.)

For $p = 3$ we have

$$\int \sqrt{x^2 + 1} dx = \int \cosh^2 u du. \quad (22)$$

There are two ways to proceed, depending on taste. First, the hyperbolic functions obey identities in close parallel to those for trigonometric functions; one can memorize, look up, or rederive the identity exactly corresponding to the one one would use to evaluate $\int \cos^2 \theta d\theta$. On the other hand, it is again a great charm of hyperbolic functions that they can always be eliminated through (4) in favor of the exponential function, which obeys a much simpler and shorter, but equally powerful, list of identities. (Indeed, many students learn, even if their calculus books never tell them, that the best way to recover trig identities is to use the function $e^{i\theta}$ in this same way.) Thus we have

$$\begin{aligned} \int \cosh^2 u du &= \frac{1}{4} \int (e^{2u} + e^{-2u} + 2) du \\ &= \frac{1}{4} \left(\frac{1}{2} e^{2u} - \frac{1}{2} e^{-2u} + 2u \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{4}(\sinh 2u + 2u) \\
&= \frac{1}{2}(\sinh u \cosh u + u) \\
&= \frac{1}{2}(x\sqrt{x^2 + 1} + \sinh^{-1} x),
\end{aligned}$$

hence

$$\int \sqrt{x^2 + 1} dx = \frac{1}{2}x\sqrt{x^2 + 1} + \frac{1}{2}\ln(x + \sqrt{x^2 + 1}) + C, \quad (23)$$

and ultimately

$$\int \sec^3 \theta d\theta = \frac{1}{2} \tan \theta \sec \theta + \frac{1}{2} \ln |\sec \theta + \tan \theta| + C. \quad (24)$$

Larger odd values of p can in principle be treated in the same way, although, as always in this type of problem, the complexity increases.

Acknowledgments. I thank Philip Yasskin for comments on the manuscript, and a referee for contributing the more heuristic proof of (8).

References

1. W. B. Gearhart and H. S. Shultz, Tugging a barge with hyperbolic functions, *College Math. J.* **34** (2003) 42–49.
2. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, 1965.
3. V. F. Rickey and P. M. Tuchinsky, An application of geography to mathematics: History of the integral of the secant, *Math. Mag.* **53** (1980) 162–166.
4. J. S. Robertson, Gudermann and the simple pendulum, *College Math. J.* **28** (1997) 271–276.
5. D. J. Velleman, Partial fractions, binomial coefficients, and the integral of an odd power of $\sec \theta$, *Amer. Math. Monthly* **109** (2002) 746–749.

Where there are numbers, there is beauty.—Proclus

$$\begin{aligned}
\sqrt{2} &= \frac{1 + 1/3 - 1/5 - 1/7 + 1/9 + 1/11 - \dots}{1 - 1/3 + 1/5 - 1/7 + 1/9 - 1/11 + \dots} \\
&= \frac{1 + 1/3 - 1/5 - 1/7 + 1/9 + 1/11 - \dots}{\pi/4} \\
&= e^{1/2 - 1/4 + 1/6 - 1/8 + 1/10 - \dots}.
\end{aligned}$$

Submitted by Sidney Kung (sidneykung@yahoo.com) of Jacksonville, FL.