The inverse of secant and trigonometric substitutions*

Attila Máté
Brooklyn College of the City University of New York

January 19, 2017

Contents

1 The definition of the inverse secant

There are various definitions of the function arcsec used in basic calculus courses. For example, Stewart [3] gives the following definition:

$$\operatorname{arcsec}_1 x \overset{def}{=} y \quad \text{if } x = \sec y \text{ and } 0 \le y < \frac{\pi}{2} \text{ or } \pi \le y < \frac{3\pi}{2}; \tag{1.1}$$

we gave a subscript to arcsec to distinguish the different definitions. An older edition of CRC Mathematical Tables [2, p. 232] gives the definition

$$\operatorname{arcsec}_2 x \overset{def}{=} y \quad \text{if } x = \sec y \text{ and } 0 \le y < \frac{\pi}{2} \text{ or } -\pi \le y < -\frac{\pi}{2} \tag{1.2}$$

(CRC uses the notation Arcsec). A newer edition of [4, Section 6.3 Inverse circular functions, Subsection 6.3.2 Principal values of inverse circular functions] gives the definition

$$\operatorname{arcsec}_3 x \overset{def}{=} y \quad \text{if } x = \sec y \text{ and } 0 \le y < \frac{\pi}{2} \text{ or } \frac{\pi}{2} < y \le \pi. \tag{1.3}$$

*Written for the course Mathematics 1206 (Calculus 2) at Brooklyn College of CUNY.

This definition seems to be standard, since it is used by the major computer algebra software Maxima, Maple, and Mathematica, and it is also given by the NIST.[1.1] We will discuss the putative advantages of the functions arcsec_1 and arcsec_2, and whether there is any reason to keep using them. Using them without the subscripts, as usual, can create confusion with the standard use of the same notation in the in a different sense. These functions are used in calculating integrals of form

$$\int R(x, \sqrt{x^2-1})\,dx,$$

where R is a rational function. One of the usual ways to evaluate such an integral is to substitute $t = \operatorname{arcsec}_k x$ with $k = 1$, 2, or 3, but the choice of k makes a difference in these calculations. In fact, taking $t = \operatorname{arcsec}_k x$, we have

$$x = \sec t, \qquad \sqrt{x^2-1} = \tan t \quad (k=1,2), \qquad \operatorname{sgn} x\sqrt{x^2-1} = \tan t \quad (k=3),$$

where

$$\operatorname{sgn} x \overset{def}{=} \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0. \end{cases}$$

Thus, $k = 1$ or $k = 2$ seems to make no difference, while $k = 3$ introduces the inconvenience of apparently having to deal with the cases whether $x > 0$ or $x < 1$ separately.

In fact, there is a subtle difference between cases $k = 1$ and $k = 2$ in that $k = 2$ is easier to relate to the case $k = 3$. Accepting the definition

$$\operatorname{arcsec} x \overset{def}{=} \operatorname{arcsec}_3 x, \tag{1.4}$$

we have

$$\operatorname{arcsec}_2 x = \operatorname{sgn} x \; \operatorname{arcsec} x.$$

Hence, one can use the substitution represented by arcsec_2 without explicitly invoking the function arcsec_2. That is, using the standard meaning of $\operatorname{arcsec} x = \operatorname{arcsec}_3 x$, we use the substitution as follows:

$$\begin{gathered} x = \sec t \text{ where } 0 \le t < \frac{\pi}{2} \text{ or } -\pi \le t < -\frac{\pi}{2}, \\ dx = \sec t \, \tan t \, dt, \qquad \sqrt{x^2-1} = \tan t, \qquad t = \operatorname{sgn} x \; \operatorname{arcsec} x. \end{gathered} \tag{1.5}$$

Doing the substitution this way, there is no difference between using the substitution $t = \operatorname{arcsec}_2 x$ and $t = \operatorname{sgn} x \operatorname{arcsec} x$ until the time the integration is done and one is ready to write the result in terms of x.

There are other methods for calculating integrals of rational expressions of x and of the square root of a quadratic polynomial of x. One can use the Euler substitutions, or one can use hyperbolic functions [1].

[1.1] National Institute of Standards and Technology. For their definition of arcsec, see NIST Digital Library of Mathematical Functions, formulas 4.23.2, 4.23.5 and 4.23.8. From the viewpoint of complex functions, this definition has a clear advantage in that it can be extended to a definition complex plane with the aid of branch cuts, see loc.cit, §4.23(ii). This cannot be done for the definition given in (1.1) and (1.2).

2 An example

As an illustration, calculate the integral

$$I = \int \frac{dx}{x^3\sqrt{x^2-1}}. \tag{2.1}$$

2.1 Solution via arcsec substitution

We will give a solution via an arcsec substitution as described in (1.5). We have

$$I = \int \frac{\sec t \tan t\, dt}{\sec^3 t \tan t} = \int \frac{dt}{\sec^2 t} = \int \cos^2 t\, dt. \tag{2.2}$$

This integral can be evaluated either by the half-angle formula

$$\cos^2 t = \frac{\cos 2t - 1}{2}$$

or by the integration by parts formula

$$\int f(t)g'(t)\, dt = f(t)g(t) - \int f'(t)g(t)\, dt \tag{2.3}$$

with $f(t) = \cos t$ and $g'(t) = \cos t$, in which case $f'(t) = -\sin t$ and $g(t) = \sin t$ (in the latter, we took the constant of integration to be 0, which is the simplest choice). Choosing integration by parts, we have

$$\begin{aligned} I = \int \cos^2 t &= \cos t \sin t + \int \sin^2 t\, dt = \cos t \sin t + \int (1 - \cos^2 t)\, dt \\ &= \cos t \sin t + t - \int \cos^2 t\, dt = \cos t \sin t + t - (I - 2C), \end{aligned}$$

where C is an arbitrary constant, necessary since the constant of integration occurring in the integral I in the left-hand side need not be the same as the constant of integration on the right-hand side. Thus,

$$I = \frac{1}{2}(\cos t \sin t + t) + C = \frac{\tan t}{2\sec^2 t} + \frac{t}{2} + C = \frac{\sqrt{x^2-1}}{2x^2} + \frac{1}{2}\operatorname{sgn} x \operatorname{arcsec} x + C. \tag{2.4}$$

For the constant of integration one need to keep in mind that the expression on the right-hand side is undefined if $-1 < x < 1$, and so there is no connection between the expression for $x < 0$ and for $x > 0$; hence one should use two different constants of integration for $x < 0$ and for $x > 0$. Nevertheless, for the sake of simplicity, we will retain the form of the result, and will not indicate explicitly that two different constants of integration are involved.[2.1]

[2.1] This is common practice. For example, one routinely writes that

$$\int \frac{dx}{x} = \ln|x| + C,$$

and in fact the value of C for $x > 0$ has no relation to its value for $x < 0$. This is because the integrand is undefined at $x = 0$, and so the integral consists of two separate parts, one for $x > 0$ and one for $x < 0$. If one does not want to follow this practice, one needs to write

$$\int \frac{dx}{x} = \begin{cases} \ln x + C_1 & \text{if } x > 0, \\ \ln(-x) + C_2 & \text{if } x < 0 \end{cases}$$

for some constants C_1 and C_2, and this is almost never done.

Using the identity

$$\operatorname{arcsec} x = \arccos\frac{1}{x} = \frac{\pi}{2} - \arcsin x$$

(the first equation here is true only for the definition given by equation (1.3) for $\operatorname{arcsec} x$, and is not true for the definitions given by (1.1) and (1.2)), we obtain

$$\begin{aligned} I &= \frac{1}{2}(\cos t \sin t + t) + C = \frac{\sqrt{x^2-1}}{2x^2} + \frac{1}{2}\operatorname{sgn} x \arccos\frac{1}{x} + C. \\ &= \frac{\sqrt{x^2-1}}{2x^2} + \frac{\pi}{4}\operatorname{sgn} x - \frac{1}{2}\operatorname{sgn} x \arcsin\frac{1}{x} + C. \end{aligned}$$

The term $(\pi/4)\operatorname{sgn} x$ can be incorporated into the arbitrary constant of integration C. The fact that this term has different values according as $x < 0$ or $x > 0$ makes no difference if we agree that the "constant" C need not be the same for $x < 0$ and $x > 0$.[2.2] Thus, we can write the final result as

$$\begin{aligned} I &= \frac{\sqrt{x^2-1}}{2x^2} - \frac{1}{2}\operatorname{sgn} x \arcsin\frac{1}{x} + C = \frac{\sqrt{x^2-1}}{2x^2} - \frac{1}{2}\arcsin\frac{1}{x \operatorname{sgn} x} + C \\ &= \frac{\sqrt{x^2-1}}{2x^2} - \frac{1}{2}\arcsin\frac{1}{|x|} + C. \end{aligned}$$

This is in fact the result given by Maxima. In giving the result of integration, Maxima usually avoids the use of the function arcsec, because there is no general agreement on its definition, even though Maxima uses the definition given by (1.3).

2.2 Trying to avoiding trigonometric substitutions

In the example given in (2.1), there is a simple way to avoid trigonometric substitutions, or other common substitutions used to integrate expressions with a quadratic polynomial under square root. The reason is the occurrence of x with an odd exponent outside the square root. We can write

$$I = \int \frac{x\,dx}{x^4\sqrt{x^2-1}}.$$

We can use the substitution

$$t = \sqrt{x^2-1}, \tag{2.5}$$

when $t^2 = x^2 - 1$, so $2t\,dt = 2x\,dx$, i.e., $t\,dt = x\,dx$; further, $x^2 = t^2 + 1$. When we want to retrace this substitution, it is important to remember that $|x| = \sqrt{t^2-1}$, and so

$$x = |x| \operatorname{sgn} x = \operatorname{sgn} x \sqrt{t^2-1}. \tag{2.6}$$

Using this substitution, we have

$$I = \int \frac{t\,dt}{(t^2+1)^2 t} = \int \frac{dt}{(t^2+1)^2}. \tag{2.7}$$

[2.2] See footnote 2.1 on p. 3.

There are at least two ways of evaluating this integral. One way uses the trigonometric substitution $t = \tan u$, in which case $\sqrt{t^2+1} = \sec u$ and $dt = \sec^2 u\, du$. Hence

$$I = \int \frac{\sec^2 u\, du}{\sec^4 u} = \int \frac{du}{\sec^2 u} = \int \cos^2 u\, du$$

In fact, this integral has been calculated in another context (cf. (2.2) and (2.4). This is not surprising, since by a circuitous route we got back to the same place we were with equation (2.2). Indeed, retracing the substitutions, we have $\sec^2 u = t^2 + 1 = x^2$, so $x = \pm \sec t$; so this in fact corresponds to the $x = \sec t$ substitution used before. According to (2.4), we have

$$I = \int \cos^2 u\, du = \frac{1}{2}(\cos u \sin u + u) + C.$$

We need to express this in terms of the variable x. We have $\tan u = t$ and $t = \sqrt{x^2-1}$, so

$$\sec u = \sqrt{\tan^2 u + 1} = \sqrt{t^2+1} = \sqrt{\left(\sqrt{x^2-1}\right)^2 + 1} = \sqrt{x^2} = |x|.$$

Thus

$$\begin{aligned} I &= \frac{1}{2}(\cos^2 u \tan u + u) + C = \frac{1}{2}\left(\frac{\tan u}{\sec^2 u} + \operatorname{arcsec}|x|\right) + C \\ &= \frac{1}{2}\left(\frac{\sqrt{x^2-1}}{x^2} + \operatorname{arcsec}|x|\right) + C. \end{aligned} \tag{2.8}$$

This result differs in form from the result obtained in (2.4), but the difference is only in an additive constant, which means that the C here is different from the C in (2.4). More precisely, if we write C' instead of C for the constant in (2.4), and retain the name C for the constant at present, then $C' = C$ if $x > 0$ and $C' = C - \pi$ if $x < 0$; it is important to recall that the integral for $x > 0$ and for $x < 0$ can use different values of C. This is because for $x < 0$ we have

$$\operatorname{sgn} x \operatorname{arcsec} x = -\big(\pi - \operatorname{arcsec}(-x)\big) = -(\pi - \operatorname{arcsec}|x|) = \operatorname{arcsec}|x| + \pi$$

according to equations (1.3) and (1.4).

2.3 A recursive formula

We will establish the equation

$$\int \frac{dx}{(1+x^2)^{\alpha+1}} = \frac{x}{2\alpha(1+x^2)^\alpha} + \frac{2\alpha-1}{2\alpha}\int \frac{dx}{(1+x^2)^\alpha}, \tag{2.9}$$

valid for any real $\alpha \neq 0$. This formula can be used recursively to evaluate integrals of the form

$$\int \frac{dx}{(1+x^2)^n}$$

for integers $n > 1$; in particular, it can be used to evaluate the integral in equation (2.7). This formula can be established as follows. We have

$$\int \frac{dx}{(1+x^2)^\alpha} = \int \frac{x^2+1}{(1+x^2)^{\alpha+1}}\, dx = \int \frac{x^2}{(1+x^2)^{\alpha+1}}\, dx + \int \frac{1}{(1+x^2)^{\alpha+1}}\, dx. \tag{2.10}$$

Here, the first integral on the right-hand side can be evaluated by integration by parts. We have

$$\int \frac{x^2}{(1+x^2)^{\alpha+1}}\,dx = \int x\,\frac{x}{(1+x^2)^{\alpha+1}}\,dx = -\frac{x}{2\alpha(1+x^2)^\alpha} + \frac{1}{2\alpha}\int \frac{dx}{(1+x^2)^\alpha}, \tag{2.11}$$

where the second equation was obtained by the integration by parts formula (2.3) with $f(x) = x$ and

$$g'(x) = \frac{x}{(1+x^2)^{\alpha+1}}$$

after calculating $g(x)$ as follows: by using the substitution $t = 1+x^2$, when $dt = 2x\,dx$, we find that

$$g(x) = \int \frac{x}{(1+x^2)^{\alpha+1}} = \frac{1}{2}\int \frac{dt}{t^{\alpha+1}} = -\frac{1}{2\alpha t^\alpha} + C = -\frac{1}{2\alpha(1+x^2)^\alpha} + C;$$

we take $C = 0$ here.[2.3] Replacing the first term on the right-hand side of (2.10) with the right-hand side of (2.11), we obtain

$$\int \frac{dx}{(1+x^2)^\alpha} = -\frac{x}{2\alpha(1+x^2)^\alpha} + \frac{1}{2\alpha}\int \frac{dx}{(1+x^2)^\alpha} + \int \frac{dx}{(1+x^2)^{\alpha+1}}.$$

Rearranging this equation, we obtain formula (2.11).

2.4 Avoiding trigonometric substitution

Using formula (2.9) with $\alpha = 1$, we can calculate the integral on the right-hand side of (2.7) without using a trigonometric substitution. We have

$$I = \int \frac{dt}{(t^2+1)^2} = \frac{t}{2(1+t^2)} + \frac{1}{2}\int \frac{dt}{1+t^2} = \frac{t}{2(1+t^2)} + \frac{1}{2}\arctan t + C.$$

Noting that $t = \sqrt{x^2-1}$ according to (2.5), for the integral given in formula (2.1) we have

$$I = \int \frac{dx}{x^3\sqrt{x^2-1}} = \frac{\sqrt{x^2-1}}{2x^2} + \frac{1}{2}\arctan\sqrt{x^2-1} + C.$$

This agrees with the result given in (2.8), since

$$\arctan\sqrt{x^2-1} = \operatorname{arcsec}|x|.$$

Indeed, if the write t for the left-hand side then $0 \le t < \pi/2$, and so

$$\sec t = \sqrt{\tan^2 t - 1} = \sqrt{\left(\sqrt{x^2-1}\right)^2 + 1} = \sqrt{x^2} = |x|.$$

[2.3] When integrating by parts, we would always pick a specific value of the constant of integration C, deemed most useful in the calculation. Most often, this value is $C = 0$.

References

[1] Attila Máté. Hyperbolic functions.
`http://www.sci.brooklyn.cuny.edu/ mate/misc/hyperbolic_functions.pdf`, 2016.

[2] Samuel M. Selby, editor. *Standard Mathematical Tables*. The Chemical Rubber Co., Cleveland, Ohio 44128, 21st edition, 1973.

[3] James Stewart. *Single Variable Calculus/Early Transcendentals*. Cengage Learning, USA, 8th edition, 2012.

[4] Daniel Zwillinger, editor. *CRC Standart Mathematical Tables and Formulae*. Chapman & Hall/CRC, Boca Raton, New York, 31st edition, 2003.