INTEGRATION: WHY YOU CAN AND WHY YOU CAN'T

by Rick Miranda
Colorado State University

At most colleges and universities, a large part of the second semester of calculus is devoted to the arcane subject commonly known as "techniques of integration". The basic problem is to find a closed-form expression for $\int f(x)dx$ where $f(x)$ is a specific function of the variable x. Typically, the following methods are discussed:

- 'forward' substitutions $x = g(u)$
- 'backward' substitutions $u = h(x)$
- integration by parts
- the use of exponentials and logarithms
- trigonometric substitutions
- inverse hyperbolic trig functions
- powers of sines and cosines
- integrals with quadratics
- partial fraction decompositions.

There are two logical reactions to this subject:

a) There is too much material here.

b) There is not enough material here.

For anyone who has taken or taught this course, (a) hardly needs explanation. Firstly, the mastery of all these techniques requires quite a bit of rote memorization of basic integrals, especially for the average student. Secondly, when faced with an integration problem, the 'menu' of possible techniques to try to apply is large enough to make the decision process fairly complicated. Finally, with extensive tables and (lately) computer programs which integrate all functions encountered in this course, the motivation to delve into this subject with one's sleeves rolled up is naturally diminished, and this is made worse by the amount there seem to be to know.

Have you ever heard (b) from a student of this subject? Well, now you have, and let me explain why. After a good solid course on the techniques of integration, including a thorough discussion of the topics listed above, I could well come away with the following broad classification of integrals:

i) The integrals which I can find.

ii) The integrals which I cannot find.

Statement (b) is one reaction to the existence of the second class.

Most of the integrals encountered in the course are of type (i) (or should be, by the end of the semester). A student, in fact, may never see an integral of type (ii), and may conclude that all integrals are of type (i), for the appropriate choice of "I"; since he (or she) knows in his gut that he can't possibly solve all integration problems, the conclusion is that he is not the appropriate choice for "I", and that the subject is much too complicated for mere mortals to think about.

If an integral of type (ii) is seen in this course, it is usually in one of the "set up but do not evaluate the integral which computes..." problems on an exam; when going over the questions on the next day, the teacher may make a remark to the effect that "we can't find this integral..." and the subject is embarrassingly dropped. Generally, no attempt is made to explain why some integrals can be found and some can't, and we're back to reaction (b) (on a slightly different level): There is something missing here.

In this article I'd like to discuss why there are integrals of types (i) and (ii), and try to explain the fascinating relationship between this apparently analytic subject and the much more geometric subject of algebraic plane curves.

Let me begin by stating a theorem.

Theorem. Let $R(t)$ be a rational function of the variable t, i.e., $R(t)$ is the ratio of two polynomials. Then

$$\int R(t)\,dt$$

can be found.

(Of course, actually finding a closed-form expression for it involves factoring polynomials and solving linear equations, and is a formidable task in itself -- but I won't address these problems here.)

In my view, it is not unfair to say that, even given the mass of material devoted to integration techniques, this is the only true theorem in this course; the other topics covered are really just methods to use as the occasion arises. This being the case, one would think that this would be the focal point of this course. However, it is hardly ever stated explicitly, and often the details of the process of partial fractions (which is the proof of this theorem) is given much more weight than the simple and obviously powerful statement itself. This is understandable, since carrying out the partial fraction decomposition is a complicated and cumbersome task, even in fairly simple situations, and requires some attention. However, I think it is a mistake not to rise above the fray and drive the point home that here is a large and common class of functions which are all "of type (i)" -- I can integrate them!

If you grant that this is the 'only' theorem of this type, then your mind should naturally turn to the following: can other integrals be brought to this form by clever substitutions, and can this theorem therefore achieve a wider scope of application? The well-known answer to this question is: Sometimes, if you get lucky.

Example. Integrate $\int\sqrt{1+x^2}\,dx$.

Solution. Substitute $x = \frac{2t}{(1-t^2)}$. Then $1 + x^2 = \left(\frac{(1+t^2)}{(1-t^2)}\right)^2$ and

$dx = \frac{(2+2t^2)}{(1-t^2)^2}$ so the above integral transforms to $\int\frac{2+4t^2+2t^4}{(1-t^2)^3}\,dt$,

and the theorem applies.

This seemed pretty lucky. What if I try $\int\sqrt{1+x^3}\,dx$? In this case I'm stuck for a clever substitution. What is going on here? In order to fix our attention on a certain general class of functions, consider the following.

Definition. A function $y = y(x)$ is algebraically dependent on x if there is a polynomial $f(x_1,x_2)$ in two variables, such that $f(x,y(x))$ is identically zero.

Examples. $y = \sqrt{x}$ $(f(x_1,x_2) = x_1 - x_2^2)$

$$y = x^{4/5} \quad (f(x_1,x_2) = x_1^4 - x_2^5)$$

$$y = \sqrt{1 + x^3} \quad (f(x_1,x_2) = x_1^3 + 1 - x_2^2)$$

The integrals $\int\sqrt{1 + x^2}\, dx$, $\int\sqrt{1 + x^3}\, dx$, etc., are examples of integrals which involve functions of x which are algebraically dependent on x, and this is the class of functions which I want to focus on. Our general problem can be formulated as follows.

<u>*The General Problem of Integration of Algebraic Functions.*</u>

Let $R(x_1, x_2)$ be a rational function of two variables.

Let $y = y(x)$ be algebraically dependent on x.

Can $\int R(x, y(x))dx$ be found?

The answer is again: sometimes. But it doesn't have anything to do with luck. Let's try to think about this systematically. If $y = y(x)$ is algebraically dependent on x, then there is this polynomial $f(x_1,x_2)$ such that $f(x, y) \equiv 0$. Now the equation $f(x_1,x_2) = 0$ defines a so-called "algebraic curve" in the (x_1,x_2)-plane, and $(x, y(x))$ always lies <u>on</u> this curve. The properties of this curve should therefore be important in studying $y(x)$. Central for us is the following property.

<u>*Definition.*</u> Let $f(x_1,x_2)$ be a polynomial in two variables. The curve $C = \{(x_1,x_2) | f(x_1,x_2) = 0\}$ is <u>rationally parametrized</u> if there are rational functions $x_1 = x_1(t)$, $x_2 = x_2(t)$, such that $f(x_1(t), x_2(t))$ is identically zero as a function of t.

In this case the point $(x_1(t), x_2(t))$ will lie on the curve C for all values of t. Let's look at any easy example.

<u>*Example.*</u> Let $f(x_1,x_2) = x_1^2 + x_2^2 - 1$, so that the curve C is the unit circle. Then C is rationally parametrized by $x_1(t) = \frac{(1 - t^2)}{(1 + t^2)}$, $x_2(t) = \frac{2t}{(1 + t^2)}$. (Check this!) This is not magic. Note that the point $P = (-1,0)$ is on C. Let L_t be the line through P with slope t; an equation for L_t is $x_2 = t(x_1 + 1)$. For any t, this line L_t will intersect the circle C in two points, one of which is, of course, P. Call

the other point P_t. A little algebra will convince you that $P_t = \left(\frac{(1 - t^2)}{(1 + t^2)}, \frac{2t}{(1 + t^2)}\right)$, giving the explicit parametrization above.

The importance of a rational parametrization for the curve C is demonstrated by the following.

Theorem. Let $R(x_1,x_2)$ be a rational function of two variables and let $y = y(x)$ be algebraically dependent on x, with $f(x, y(x))$ identically zero. Assume that the curve $C = \{(x_1,x_2) | f(x_1,x_2) = 0\}$ can be rationally parametrized. Then $\int R(x, y(x))dx$ can be found.

Proof. Let $x_1 = x_1(t)$, $x_2 = x_2(t)$ be the parametrization of C. Note that $x = x_1(t)$, $y = x_2(t)$ in this case; make this substitution into the integral. One gets $\int R(x_1(t), x_2(t))\left(\frac{dx_1}{dt}\right)dt$, which has a rational integrand. We can now apply the theorem.

The above proposition seems to be constructive, too; the only hitch is in parametrizing the curve C. In particular, the immediate question is: Which curves C can be rationally parametrized, and how? If $f(x_1,x_2) = ax_1 + bx_2 + c$, so that the degree of f is one and C is a line, then clearly C may be rationally parametrized; $x_1 = bt + z_1$, $x_2 - at + z_2$, where (z_1,z_2) is any point on C. In this case $y(x) = x_2(x_1) = -(\frac{a}{c})x - (\frac{b}{c})$ is a linear function of x and any rational expression in x and y can be immediately reduced to a rational function of x alone, so the above process is not too enlightening.

Fortunately, there is one other large class of curves which can be parametrized.

Proposition. Any conic C (i.e., defined by $f(x_1,x_2) = 0$ where $f(x_1,x_2)$ is of degree 2) can be rationally parametrized.

Proof. Let me present two proofs of this statement, one algebraic and one geometric in spirit. The first step of the algebraic proof is to change coordinates from (x_1,x_2) to (x,y) so that $f(x_1,x_2)$ becomes $g(x,y) = \frac{x^2}{a^2} \pm \frac{y^2}{b^2} - 1$, the "standard form" for a conic. This is a linear change of coordinates, so that if we can parametrize $g(x,y) = 0$ by rational functions, we will be able to transport this parametrization to $f(x_1, x_2)$. The second step is to explicitly parametrize the standard conic $g(x,y) = 0$. Here is one way.

$$x = a\frac{b^2 \mp a^2t^2}{b^2 \pm a^2t^2}, \quad y = \frac{2ab^2t}{b^2 \pm a^2t^2} .$$

A more geometric proof is afforded by following the hint of the circle example. Pick any point P on the conic C. Parametrize the lines through P by their slopes: if $P = (x_0, y_0)$, let L_t be the line $y - y_0 = t(x - x_0)$ through P with slope t. Now intersect L_t with the conic C; one will get two points, one of which is P, the other is $P_t = (x(t), y(t))$; it is not hard to see that $x(t)$ and $y(t)$ are rational parametrizations of the conic C.

Q.E.D.

Note that in the above argument, one might want to use a vertical line sometimes where the slope "is infinity". This leads naturally into some elementary concepts of projective geometry, which I do not wish to discuss at this time.

As promised by our theorem, a proposition about parametrizing curves should give us a nice application to integrals. Here's the result for conics restated for this purpose:

Corollary. For any numbers a, b and c, the integral

$$\int R(x, \sqrt{ax^2 + bx + c})dx$$

can be found (where $R(x_1, x_2)$ is a rational expression in two variables).

Proof. If $y = \sqrt{ax^2 + bx + c}$, then y is algebraically dependent on x; $f(x,y) = y^2 - ax^2 - bx - c$ is identically zero. Since $f(x,y)$ has degree 2, the curve $f(x_1, x_2) = 0$ defines a conic, and therefore may be rationally parametrized. Now the theorem applies.

Q.E.D.

In our course on techniques of integration, a lot of time is spent developing methods for handling integrals involving $\sqrt{ax^2 + bx + c}$, but the general result above is very rarely brought out into the open -- I think it should be.

As long as we're here...

Parametrizing conics has been fun for millenia. Let us recall our parametrization of the circle

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

Note that if t is a rational number, then x and y will both be rational numbers also. So what? Well, write $t = \frac{u}{v}$, with u and v integers. Then, clearing denominators, we see that

$$x = \frac{v^2 - u^2}{u^2 + v^2}, \qquad y = \frac{2uv}{u^2 + v^2},$$

and $x^2 + y^2 = 1$ means that $(v^2 - u^2)^2 + (2uv)^2 = (u^2 + v^2)^2$. In other words, $(v^2 - u^2, 2uv, u^2 + v^2)$ is a Pythagorean triple. Moreover, it is an elementary theorem from number theory that all Pythagorean triples come this way. This very geometric approach to number theory was pioneered by the Greek Diophantus, and has been refined into some amazing results relating the geometry of solutions to equations and the number theory which naturally arises.

But back to integration. Recall the following magic trick for integrating an expression involving $\sin\theta$ and $\cos\theta$: make the substitution $\theta = 2\arctan(t)$. Why does this work? A little trigonometry and differentiation formulas (including the dreaded half-angle formulas) will produce

$$\cos\theta = \frac{1 - t^2}{1 + t^2}, \ \sin\theta = \frac{2t}{1 + t^2}, \qquad d\theta = \frac{2dt}{1 + t^2}$$

and so this substitution replaces the trigonometric integrand with a rational integrand, and now we use the theorem. From our vantage point, this amazing and ad hoc substitution, which at first glance works "because it works", is seen as exactly substituting the rational parametrization of the circle which we've become quite familiar with for the trigonometric parametrization $x = \cos\theta$, $y = \sin\theta$. Hence we have the following (without any magic!):

Corollary. If $R(x_1,x_2)$ is a rational expression in two variables, then $\int R(\cos\theta, \sin\theta)d\theta$ can be found.

Recall the hyperbolic functions sinh(x) and cosh(x), so called because they give a parametrization of the hyperbola $x_1^2 - x_2^2 = 1$; $\cosh^2(x) - \sinh^2(x) = 1$ for any x, so $(\cosh(x), \sinh(x))$ always lies on the unit hyperbola. We now know that the unit hyperbola can also be rationally parametrized by

$$x_1 = \frac{1 + t^2}{1 - t^2}, \quad x^2 = \frac{2t}{1 - t^2} \quad .$$

Our main theorem now yields the following immediately.

Corollary. If $R(x_1,x_2)$ is any rational expression in two variables, then $\int R(\cosh(x), \sinh(x))dx$ can be found. (Using the chain rule it is easy to see that $dx = \frac{2dt}{1 - t^2}$ using the above substitutions for cosh(x) and sinh(x).)

This just about exhausts the applications of the existence of rational parametrizations for conics to the theory of integration. Can we proceed to higher degree curves? Well, there are curves which are not conics, but which can still be rationally parametrized:

Example. $y = x^{p/q}$ satisfies $f(x,y) = y^q - x^p \equiv 0$. This is parametrized by $x = t^q$, $y = t^p$. Hence,

Corollary. $\int R(x,x^{p/q})dx$ can be found, where $R(x_1,x_2)$ is any rational expression in two variables.

Example. The lemniscate $f(x,y) = (x^2 + y^2)^2 - (x^2 - y^2) \equiv 0$ (draw this!) has a rational parametrization

$$x = \frac{2t(t + 1)}{4t^2 + 1}, \quad y = \frac{2t(4t^2 - 1)}{(4t^2 + 1)(2t + 1)} \quad .$$

To find this, one intersects the lemniscate C with a circle C_t centered at $(t, -t)$ of radius $\sqrt{2}t$, so that $\underline{0} = (0, 0)$ is on C_t. In fact, $C \cap C_t$ consist of $\underline{0}$ and one other point P_t, which has the above coordinates.

The above example looks like I'm just showing off -- maybe that's right. Finding parametrizations for plane curves is not easy, and in fact most curves $\{f(x,y) = 0\}$ cannot be rationally parametrized! One example is $y^2 - x^3 - 1 = 0$, which defines the algebraic function $y = \sqrt{1 + x^3}$, which I got stuck on earlier. (If you're good with polynomials, you might try to prove that $y^2 - x^3 - 1 = 0$ can't be rationally parametrized.) One corollary of our discussion, then, is that $\int\sqrt{1 + x^3}dx$ can't be expected to be found with our present techniques. In general, the integrals involving the square root of a cubic polynomial in x are classically called elliptic (they arise in computing various quantities associated to an ellipse, e.g., arclength, etc.) and can't be solved in closed form using elementary functions. Now we know why: behind the whole problem lies an unparametrizable curve!

The problem of parametrizing curves actually led to the invention of topology. Assume $\{f(x,y) = 0\}$ is parametrized. This gives a nice continuous function from {t-space} to {solutions to $f(x,y) = 0$}, sending a typical t to $(x(t),y(t))$. There's nothing in all of the above discussion which says that t can't be a complex number instead of just a real number; after all, we went "backward" to rational t's for a number-theoretic application -- why not go "forward" to complex t's? Recall that {complex t-space} is a 2-sphere, if you add the point at ∞ (which, again, we saw earlier was not unreasonable). So the above parametrization can be viewed as a nice continuous function from the 2-sphere to complex solutions (x,y) to $f(x,y) = 0$. Therefore, intuitively, these complex solutions better look pretty much like a sphere. However, in lots of examples, this solutions set doesn't look anything like a sphere. For example, the complex solutions to $y^2 = 1 + x^3$ made up, topologically, a torus. So there seems to be a real topological obstruction here to parametrizing this curve, and the attempt to understand this phenomenon led to the development of modern topology.

It turns out that the general curve of degree at least 3 (i.e., $f(x,y)$ has degree ≥ 3) cannot be rationally parametrized; however, there are special curves which can be, as the examples above illustrate. The general problem of the existence of rational parametrizations of plane curves ultimately led to the flowering of the field of algebraic geometry, and is quite complicated.

Have we then simply substituted one field of ignorance for another? No, not really. I think we have isolated the essential problem, which is one of parametrization, not integration, and along the way elucidated many of the standard results of integration theory, all in terms of one basic idea. This kind of overview can only benefit any student of this subject, can put into its proper perspective the more mundane aspects of the techniques of integration, and hopefully motivate both student and teacher with a broader picture of the field.

One last highly beneficial side effect to this approach is that, on the horizon of this subject, which seems to some, at first glance, to be a "dead end" mathematically, we see the following topics rising tantalizingly out of the mist:

- the theory of conics
- number theory, and diophantine equations
- topology
- complex variables
- higher analysis
- algebraic geometry.

This is a large part of modern mathematics! Do all hard problems (like why I can't integrate everything) lead to such unexpected, diverse areas? I don't know, but even one example is an occasion for celebration by a lover of mathematics.

REGIONAL MEETINGS

Many regional meetings of the Mathematical Association of America regularly have sessions for undergraduate papers. If two or more colleges and at least one local chapter help sponsor or participate in such undergraduate sessions, financial help is available up to $50. Write to:

Dr. Richard Good
Department of Mathematics
University of Maryland
College Park, Maryland 20742