

## IMPLICITLY ELEMENTARY INTEGRALS

ROBERT H. RISCH

**ABSTRACT.** The explicitly elementary functions of complex variables  $z_1, \dots, z_n$  are those functions built up from  $\mathbf{C}(z_1, \dots, z_n)$  by exponentiation, taking logarithms, and algebraic operations. The implicitly elementary functions are obtained by solving, via the implicit function theorem, for some of the variables in terms of the others, in systems of equations formed by setting a set of explicitly elementary functions equal to 0. Here we prove a 1923 conjecture of J. F. Ritt to the effect that if the indefinite integral of an explicitly elementary function is implicitly elementary, then it is explicitly elementary. The method features a geometrization of the concepts involved.

**1. Introduction.** The subject matter of this paper is related to the old problem of determining which elementary functions have elementary indefinite integrals. Usually the elementary functions of variables  $z_1, \dots, z_n$  are taken to be those functions built up from the rational functions of  $z_1, \dots, z_n$  by successively taking logarithms, exponentiating, and performing algebraic operations (i.e., solving polynomial equations in a single variable with previously defined functions as coefficients). For example:  $(\cos \tan^{-1}(\exp z_1 + z_2)^{1/2})^{1/2}$ .

Of course we obtain the trigonometric and inverse trigonometric functions this way, since the complex numbers are available. Also, it can be shown by induction that the union over  $n$ , of the class of functions obtained, is closed under taking real part, absolute value, etc., i.e., the operation of forming the function of  $2n$  variables that when restricted to real arguments is the real part or absolute value of a given function of  $n$  variables. Also, composition keeps us in this class.

However  $w(z)$ , defined by  $w - a \sin w = z$ ,  $a \in \mathbf{C}^*$ , is not elementary in the above sense; see [4, p. 56]. Thus it seems reasonable to refer to the class above as the explicitly elementary functions, while there is also the larger class of implicitly elementary functions; viz., the closure of the explicitly elementary functions under the operation of forming the  $r$  functions of  $k$  variables that are defined by setting previously defined functions of  $r + k$  variables equal to zero (proper Jacobian condition being satisfied). This class is closed under taking inverses, in particular.

In 1835, Joseph Liouville proved (essentially) that if  $\mathcal{F}$  is a field of meromorphic functions of one variable  $z$ , that is closed under  $d/dz$ , then a function  $f \in \mathcal{F}$  has an antiderivative that is explicitly elementary over  $\mathcal{F}$  (see §2) iff there are functions  $v_0, v_1, \dots, v_m$  in  $\mathcal{F}$  and  $c_1, \dots, c_m$  in  $\mathbf{C}$  so that  $\int f = v_0 + \sum c_i \log v_i$ . Since it can be shown that  $\exp z^2$  and  $(\sin z)/z$  do not have integrals of this special form, it follows that they do not have explicitly elementary integrals.

Received by the editors June 6, 1975.

AMS (MOS) subject classifications (1970). Primary 12H05, 34A05.

© American Mathematical Society 1976

However the question remains as to whether these functions might be integrable if implicitly elementary functions were allowed. In 1923, J. F. Ritt gave a partial negative answer in [3]. He proved that if the integral of an explicitly elementary function can be defined by setting a single explicitly elementary function of two variables equal to zero, then that integral must be explicitly elementary and so satisfy the conclusion of Liouville's theorem. As an immediate consequence of this and the nonelementary character of elliptic integrals of the first kind, he obtains the result that the nonconstant elliptic functions are not explicitly elementary.

As for the general question, he writes (p. 212):

"While the formal elements of our proof given in §IV can be extended to settle this question affirmatively [negatively in our formulation—R. H. R.], we see no way of avoiding certain function theoretic assumptions which, though light in almost any other case, would be out of place in a problem of this kind. We shall therefore not write now on this question."

However in 1948, when referring to the first clause of the above claim, he writes [4, p. 94]:

"I wish to withdraw this statement. I do not have the details now, if indeed I ever did. In particular, the question as to whether the integral of an elementary function may be represented parametrically by elementary expressions may be of interest."

We will give here a proof of the general case. Our argument, however, is not a formal extension of Ritt's. Use will be made, without references, of some basic algebraic geometry, for which [1] is more than adequate.

I wish to thank Professor Hans Lewy, who, independent of Ritt, suggested this problem to me over eight years ago.

**2. Definitions and the theorem.** The implicitly elementary functions of one variable are a subclass of the functions defined by "algebraic differential equations". That is, the elements of some field  $\mathfrak{D}$  of meromorphic functions on a Riemann surface  $R$ , covering a region of the  $z$  plane, with  $\mathfrak{D}$  being both closed under  $d/dz$  and finitely generated over  $\mathbf{C}$ .

Any set of functions,  $y_1, \dots, y_m$ , generating  $\mathfrak{D}$ , satisfy an autonomous system  $dy_i/dz = f_i(y_1, \dots, y_m)$ ,  $i = 1, \dots, m$ , with  $f_i \in \mathbf{C}(Y_1, \dots, Y_m)$ . They also determine an affine model  $V$ , in  $\mathbf{C}^m$ , of the abstract field, isomorphic to  $\mathfrak{D}$ .  $V$  is the Zariski closure of the curve  $y(z) = (y_1(z), \dots, y_m(z))$  and  $\mathfrak{D} = y^*(\mathbf{C}(V))$  where  $y^*f = f \circ y$ .

The system of differential equations yields a rational vector field  $X = \sum_{i=1}^m f_i(\partial/\partial Y_i)$  in  $\mathbf{C}^m$ , that is tangent to  $V$ .  $y(z)$  is an integral curve of  $X$ , lying on  $V$ . We write  $X$  for  $X|_V$  from now on. Under the field isomorphism  $y^*$ ,  $X$  and  $d/dz$  commute; i.e.,  $y^*(Xf) = d(y^*(f))/dz$ . Consequently  $X$  has no nonconstant first integrals; i.e., kernel  $X = \mathbf{C}$ .

The isomorphism, just sketched, shows that we need not work with a differential field of meromorphic functions. Instead we will work with pairs  $(\mathfrak{D}, X)$  where  $\mathfrak{D}$  is a field of characteristic 0 that is finitely generated, of transcendence degree  $k$  over an algebraically closed field  $K$ .  $X$  is a derivation of  $\mathfrak{D}/K$ , so it is of the form  $\sum_{i=1}^k g_i(\partial/\partial x_i)$ ,  $g_i \in \mathfrak{D}$  and the  $x_i$  comprising a transcendence basis for  $\mathfrak{D}/K$ . We assume kernel  $X = K$ . Let  $\Omega^1(\mathfrak{D})$  be the vector space of one forms over  $\mathfrak{D}$ .

$$\begin{aligned} \Omega^1(\mathfrak{D}) &= \left\{ \sum_{i=1}^k g_i dx_i : g_i \in \mathfrak{D}, dx_i \left( \frac{\partial}{\partial x_j} \right) = \delta_j^i \right\} \\ &= \text{dual space to the derivations of } \mathfrak{D}/K. \end{aligned}$$

We let  $\omega_1, \dots, \omega_{k-1}$  be linearly independent in  $\Omega^1(\mathfrak{D})$  and satisfy  $\omega_i(x) = 0, i = 1, \dots, k - 1$ .

The assumption that  $\mathfrak{D}$  is finitely generated over  $K$  is used when we convert from a vector field statement to one involving differential forms. The work here could probably be reformulated to avoid it.

Under the above assumptions we have:

DEFINITION.  $(\mathfrak{F}, Y)$  is *explicitly elementary* over  $(\mathfrak{D}, X)$  iff

$$\mathfrak{F} = \mathfrak{D}(\theta_1, \dots, \theta_n), X = Y|_{\mathfrak{D}},$$

where each  $\theta_i$  satisfies either

(a)  $\theta_i$  is algebraic over  $\mathfrak{D}(\theta_1, \dots, \theta_{i-1})$ , or

(b)  $Y\theta_i/\theta_i = Yb$  or  $Yb/b = Y\theta_i$

for some  $b \in \mathfrak{D}(\theta_1, \dots, \theta_{i-1})$ .

Geometrically, we have irreducible varieties  $W$ , corresponding to  $\mathfrak{F}$ , and  $V$  corresponding to  $\mathfrak{D}$ . There is a rational vector field  $Y$  on  $W$  and a dense morphism  $\pi: W \rightarrow V$  such that  $\pi_*(Y) = X$ .  $\pi$  may be factored:

$$W = V_n \rightarrow V_{n-1} \rightarrow \dots \rightarrow V_0 = V,$$

each  $V_i$  being a branched covering of a Zariski open subset of  $W_{i-1}$  or else  $W_i = W_{i-1} \times K$  or  $W_{i-1} \times K^*$  and the morphism being projection on the first factor.

To motivate the definition of implicitly elementary, we first remark that any system of elementary equations over  $\mathfrak{D} (= \mathbb{C}(z))$  of the introduction,  $z$  the “independent variable”) can be replaced by a system in additional variables in which only two types of equations occur:

(1) polynomials in several variables with coefficients in  $\mathfrak{D}$ ;

(2) a sequence in distinct variables,  $e^{y_i} - z_i, i = 1, \dots, n$ .

We arrange such a system in an array of three nested blocks: Any set of meromorphic functions on a common domain, satisfying the old system, is part of a larger set satisfying the new one. We begin with a system whose Jacobian, with respect to the dependent variables, does not identically vanish. It is easy to see that the new system also has this property.

Each displayed variable occurs in at least one equation in each of the blocks underneath that variable. The rank of the associated matrix of differentials is maximal so the rank of the matrix of differentials of each block will also be maximal and determine the “dependent variables” occurring in it.

In the next definition, think of  $\mathfrak{D}$  as the field generated by all those functions either occurring in the coefficients of these equations or determined by the first block;  $\mathfrak{F}$  is formed from  $\mathfrak{D}$  by adding the functions determined by the second block.

DEFINITION.  $(\mathfrak{F}, X)$  is *basic implicitly elementary* over  $(\mathfrak{D}, X)$  iff  $\mathfrak{F} = \mathfrak{D}(y_1, z_1, \dots, y_n, z_n)$ ,  $\mathfrak{F}$  being of transcendence degree  $r$  over  $\mathfrak{D}, X, y_i$



written  $\alpha$  for  $\pi^* \alpha$ ,  $f$  for  $\pi^* f$ . (We use this sort of imprecision, without mention, from now on.) We may work with the weaker (more general) hypothesis that there are  $w_0, w_i \in K(W)$ ,  $d_i \in K$ ,  $\sigma \in \Omega^1(V)$  such that

$$(1) \quad \left( dw_0 + \sum d_i \frac{dw_i}{w_i} - \sigma \right) (X) = 0$$

holds on  $W$ . Thus we are proving a result about the elementary integrability of line integrals.

The conclusion of the theorem can now be stated: There is an  $\omega \in \Omega^1(V)$ ,  $\omega = dv_0 + \sum_{i=1}^m c_i dv_i/v_i$ ,  $v_0, v_i$  in  $K(V)$ ,  $c_i \in K$  such that  $\omega - \sigma$  vanishes on  $X$ .  $\omega$  looks like the pull back of an invariant one form on  $(K \times K^*)^n$  under a rational map of  $V$  into this group. The proof we give may be viewed as showing that first we can find a multiple-valued mapping of  $V$  into the group (by the process in Case 2) and then construct a single-valued map (by the process in Case 1).

CASE 1.  $K(W)$  is algebraic over  $K(V)$ .

Without loss of generality we can assume  $K(W)$  Galois over  $K(V)$ .

Thus  $W$  is a  $q$ -sheeted covering of a Zariski open subset of  $V$  and the automorphism group  $\{\phi_1 = e, \dots, \phi_q\}$  of  $K(W)/K(V)$  acts transitively on each fibre of the covering map  $W \rightarrow V$ . When stated entirely in terms of differential forms, (1) reads in this situation:

$$\begin{aligned} dw_0 + \sum d_i dw_i/w_i - \sigma &= \sum_{i=1}^{k-1} g_i \omega_i, & g_i \in K(W), \\ &\vdots \\ d(\phi_j w_0) + \sum d_i d(\phi_j w_i)/\phi_j w_i - \sigma &= \sum_{i=1}^{k-1} (\phi_j g_i) \omega_i, \\ &\vdots \\ d(\phi_q w_0) + \sum d_i d(\phi_q w_i)/\phi_q w_i - \sigma &= \sum_{i=1}^{k-1} (\phi_q g_i) \omega_i. \end{aligned}$$

By adding these conjugate equations, we obtain:

$$\left[ d(\text{Trace } w_0) + \sum d_i \frac{d(\text{Norm } w_i)}{\text{Norm } w_i} - q\sigma \right] = \sum (\text{Trace } g_i) \omega_i,$$

i.e.,

$$\left[ \frac{1}{q} d(\text{Trace } w_0) + \sum \frac{d_i}{q} \frac{d(\text{Norm } w_i)}{\text{Norm } w_i} - \sigma \right] (X) = 0$$

as desired.

CASE 2.  $K(W)$  is basic implicitly elementary over  $K(V)$ .

Here (1) is equivalent to

$$(2) \quad dw_0 + \sum d_i \frac{dw_i}{w_i} - \sigma = \sum_{i=1}^r f_i \left( \frac{dz_i}{z_i} - dy_i \right) + \sum_{j=1}^{k-1} g_j \omega_j,$$

with  $f_i, g_j$  els. of  $K(W)$  and the forms occurring in the sums on the right-hand side of the equation being linearly independent over  $K(W)$ .

Applying  $d$  to both sides of (2), we obtain

$$(3) \quad -d\sigma = \sum_{i=1}^r df_i \wedge \left( \frac{dz_i}{z_i} - dy_i \right) + \sum_{j=1}^{k-1} \sigma_j \wedge \omega_j$$

where  $\sigma_j \in \Omega^1(W)$ ,  $d\sigma \in \Omega^2(V)$ . We multiply both sides of (3) by

$$\lambda_i = \omega_1 \wedge \cdots \wedge \omega_{k-1} \wedge \prod_{j \neq i} \left( \frac{dz_j}{z_j} - dy_j \right),$$

and for each  $i$  we have an equation

$$df_i \wedge (dz_i/z_i - dy_i) \wedge \lambda_i = 0.$$

Thus,  $df_i = \sum_1^r h_j (dz_j/z_j - dy_j) + \sum_1^{k-1} i_j \omega_j$ ,  $h_j, i_j$  els. of  $K(W)$ . So  $Xf_i = df_i(X) = 0$  and therefore  $f_i \in K$ .

We rewrite (2) as

$$(4) \quad \left[ d(w_0 + \sum f_i y_i) + \sum d_i \frac{dw_i}{w_i} - \sum f_i \frac{dz_i}{z_i} \right] - \sigma = \sum_{j=1}^{k-1} g_j \omega_j.$$

Let  $Z$  be a  $k$  ( $=$  dimension  $V$ ) dimensional subvariety of  $W$  such that  $K(Z) \supseteq K(V)$  and all terms occurring in (4) are in the exterior algebra over the local ring of  $Z$  in  $W$ .  $Z$  is the graph of the "multiple-valued map" of  $V$  into  $(K \times K^*)^n$  mentioned previously. We restrict (4) to  $Z$  and are in Case 1.

Since an implicitly elementary extension consists of a sequence of the kinds occurring in Cases 1 and 2, we have established the result in general.

**3. Appendix: Liouville's original proof of his theorem.** For the benefit of those who may wish to struggle with [4], we sketch a presentation, in the spirit of the preceding, of Liouville's proof of his theorem for explicitly elementary functions. The technique involved occurs repeatedly in Ritt's book. It is suggested there, on p. 27, that the argument might derive from Liouville's acquaintance with Abel's work on the insolvability of quintic equations. More pertinent, we think, was Liouville's familiarity with the calculus of variations. We use some of the formulas and concepts of the first two chapters of [2]. What we use clearly holds in the algebraic situation at hand.

Let  $V$  be a model of  $\mathfrak{D}$ ,  $W$  a model of  $\mathfrak{D}(\theta_1, \dots, \theta_n)$ .  $\pi: W \rightarrow V$ .  $Y$  and  $X$  are the vector fields involved.  $\pi_*(Y) = X$ . We have

$$(\omega - \sigma)(Y) = (dw_0 + \sum d_i dw_i/w_i - \sigma)(Y) = 0 \quad \text{holding on } W.$$

The task is to obtain, by induction, the situation

$$(dv_0 + \sum c_i dv_i/v_i - \sigma)(X) = 0 \quad \text{holding on } V.$$

The case where  $\theta_i$  is algebraic over  $\mathfrak{D}(\theta_1, \dots, \theta_{i-1})$  is handled as in Case 1 in §2. Let us consider the case where  $\mathfrak{D}(\theta)$  is an "exponential extension" of  $\mathfrak{D}$ ; thus  $Y\theta/\theta = Xb$ ,  $b \in \mathfrak{D}$ . We will take  $V \times K^*$  as our model of  $\mathfrak{D}(\theta)$ .  $\theta$  becomes our coordinate on  $K^*$ . We have

$$(5) \quad (d\theta/\theta - db)(Y) = 0.$$

Thus  $Y$  on  $V \times K^*$  may be considered the horizontal lift, given by the flat connection determined by the form  $-db$  on  $V$ , of the vector field  $X$ .

Let  $A = \theta(\partial/\partial\theta)$ . It is the fundamental vector field on  $V \times K^*$ . It commutes with  $Y$ , so  $L_A Y = [A, Y] = 0$ .

With  $C$  standing for the contraction operator we get, since  $\omega(Y)$  is independent of  $\theta$ ,

$$\begin{aligned} 0 &= L_A(\omega(Y)) = L_A[C(\omega \otimes Y)] = C[(L_A \omega) \otimes Y + \omega \otimes (L_A Y)] \\ &= (L_A \omega)(Y). \end{aligned}$$

But by the well-known formula

$$L_A = d \circ i_A + i_A \circ d,$$

we obtain  $d(\omega(A))(Y) = (L_A \omega)(Y) = 0$ . Thus  $\omega(A)$  is constant along the trajectories of  $Y$  and therefore  $= c \in K$ .

Our use of  $A$  corresponds to replacing  $\theta$  by  $\mu\theta$  and differentiating with respect to  $\mu$ , as is done in [4].

We now restrict (5) to  $V \times \{1\}$  in  $V \times K^*$ . Then  $Y$  agrees with  $i_* X + db(X)(\theta(\partial/\partial\theta))$  along  $V \times \{1\}$ . Thus there,

$$(6) \quad \omega(Y) = \omega(i_* X) + cdb(X) = (i^* \omega + cdb)(X),$$

which yields an equation of the desired form.

The logarithmic case is similar with the equation corresponding to (6) being

$$\omega(Y) = \omega(i_* X) + c \frac{db}{b}(X) = \left( i^* \omega + c \frac{db}{b} \right)(X).$$

#### REFERENCES

1. S. Lang, *Introduction to algebraic geometry*, Interscience, New York, 1958. MR 20 #7021.
2. S. Kobayashi and K. Nomizu, *Foundations of differential geometry*, Vol. I, Interscience, New York, 1963. MR 27 #2945.
3. J. F. Ritt, *On the integrals of elementary functions*, Trans. Amer. Math. Soc. 25 (1923), 211-222.
4. ———, *Integration in finite terms. Liouville's theory of elementary methods*, Columbia Univ. Press, New York, 1948. MR 9, 573.

MATHEMATICAL SCIENCES DEPARTMENT, IBM THOMAS J. WATSON RESEARCH CENTER, YORK-TOWN HEIGHTS, NEW YORK 10598