AN EXISTENCE THEOREM FOR DIFFERENCE POLYNOMIALS[1]

R. M. COHN

Introduction. The abstract varieties (also called manifolds) of difference algebra [2], [3] have not heretofore had a realization as sets of functions comparable to the realization provided for differential manifolds by the analytic existence theorem for differential equations (see [4], particularly p. 23). It is the purpose of this Note to provide such a realization by means of an existence theorem yielding solutions of difference equations as complex-valued functions defined, except for isolated singularities, on the non-negative real axis. Since the abstract varieties consist by definition of elements lying in integral domains, these functions are required to generate difference rings which are integral domains. This distinguishes the existence theorem from the mere use of the difference equation as a recursion relation. In addition, the functions studied here are piecewise analytic, though discontinuous. There is no reason to regard the class of functions selected in this Note as definitive: other choices may better repay further study. It would, in particular, be interesting to know whether one can obtain continuous solutions. This question is discussed briefly in the next to last section. The concluding section provides a partial analogue of an important approximation theorem of differential algebra ([4, p. 123]). The notation and terminology are as in [3], and an earlier discussion of the problem with particular reference to meromorphic solutions will be found there, on pp. 114 and 242.

It might seem natural to seek solutions defined only on the integers, rather than on a half-line. But this is not possible. The only solutions of the difference polynomial yy_1-1 which are complex-valued functions on the integers and generate difference rings which are integral domains are also solutions of y^2-1. Yet the abstract variety of yy_1-1 is not contained in that of y^2-1. Alternatively, it is very likely possible to obtain suitable solutions as functions from the integers to a ring of functions of one or more variables. This amounts to finding solutions defined on the integers and involving arbitrary constants.

Definitions and statement. A complex-valued function $f(x)$ is to be called a *permitted function* if it is defined for $x \geqq 0$ except at a set $S(f)$

Received by the editors January 23, 1965.

[1] Research for this paper was supported by National Science Foundation Grant 1992.

which has no limit points, is analytic in each of the intervals into which the non-negative real axis is divided by omission of the points of $S(f)$, and is either identically 0, or is 0 at only finitely many points in any finite interval.

A *permitted difference ring* is a difference ring R whose elements are permitted functions, and whose transforming operator is the map $f(x)\rightarrow f(x+1)$, $f(x)\in R$. (More precisely, the elements of R are equivalence classes of permitted functions, with $f(x)$ equivalent to $g(x)$ if they coincide except on $S(f)\cup S(g)$. It follows from the definition of permitted functions that the map $f(x)\rightarrow f(x+1)$ is an isomorphism.) A permitted difference ring which is a field is called a *permitted difference field*. Evidently, a permitted difference ring has a quotient field which is a permitted difference field.

Henceforth, K_0 will denote the difference field of rational functions with complex coefficients and with the transform of $f(x)$ defined to be $f(x+1)$, $f(x)\in K_0$. Then K_0 may be regarded as a permitted difference field by restricting the domain of its members to $x\geqq 0$. Let $\mathcal{K}$ be the set of permitted difference fields containing K_0. Let $\mathcal{K}^*$ be the subset of $\mathcal{K}$ defined as follows: $K\in\mathcal{K}^*$ if $K\in\mathcal{K}$, and there exists an infinite set of functions analytic throughout $[0, 1]$ which is algebraically independent over K (here regarded as a field of functions over $[0, 1]$). One can easily obtain a nonenumerable set of functions analytic on $[0, 1]$ which is algebraically independent over K_0, for example by selecting functions with distinct isolated essential singularities not on $[0, 1]$. It follows that any member of $\mathcal{K}$ which is at most enumerably generated over K_0 is a member of $\mathcal{K}^*$.

A system $\mathfrak{M}$ of difference overfields of a difference field M is called *complete* over M [**3**, p. 114] if, given distinct perfect difference ideals Σ_1 and Σ_2 of a polynomial difference ring over M, there exists $N\in\mathfrak{M}$ such that the set of solutions of Σ_1 with coordinates in N is distinct from the set of solutions of Σ_2 with coordinates in N.

THEOREM I. *Let $K\in\mathcal{K}^*$. $\mathcal{K}$ is a complete system of difference overfields of K.*

It follows from the criterion for complete systems in [**3**, p. 242], that to prove Theorem I it is sufficient to prove the following existence theorem.

THEOREM II. *If Σ is a reflexive prime difference ideal of the simple polynomial difference ring $K\{y\}$, $K\in\mathcal{K}^*$, and $1\notin\Sigma$, then Σ has a generic zero in one of the members of $\mathcal{K}$.*

Proof of the existence theorem. Suppose $\Sigma\neq\{0\}$, and let $A^{(0)}$,

$A^{(1)}, \cdots$ denote a characteristic sequence (defined in [**3**, Chapter VI, §20], and in [**2**, p. 145], where they are called basic sequences) of Σ. Let the order of Σ and, hence, of $A^{(0)}$ be r. It will be shown that there exist functions $f^{(i)}$, $i=0, 1, \cdots$, such that:

(a) Each $f^{(i)}$ is defined and complex-valued on $[0, 1]$, except at the points of a finite set $T(i)$.

(b) Each $f^{(i)}$ is analytic in each of the subintervals of $[0, 1]$ in which it is defined.

(c) Let $P \in K\{y\}$ be of order s. If $P \in \Sigma$, then the substitution $y_i = f^{(i)}$, $i=0, \cdots, s$, annuls P at each point of $[0, 1]$ at which $f^{(0)}, \cdots, f^{(s)}$ and the coefficients of P are all defined. If $P \notin \Sigma$, this substitution fails to annul P except at a finite subset of $[0, 1]$.

The $f^{(i)}$ will be constructed inductively. If $r>0$, choose $f^{(0)}, \cdots, f^{(r-1)}$ to be functions analytic on $[0, 1]$ and algebraically independent over K (regarded as a field of functions over $[0, 1]$). This is possible since $K \in \mathfrak{K}^*$. Then (a), (b), and (c) hold for $s<r$.

Suppose $f^{(0)}, \cdots, f^{(h)}$, $h \geqq r-1$, have been selected so as to satisfy (a), (b), and (c) with $s \leqq h$. Let B denote the polynomial obtained from $A^{(h+1-r)}$ when the y_i, $0 \leqq i \leqq h$, are replaced by the corresponding $f^{(i)}$. Since the initial and the y_{h+1}-discriminant of $A^{(h+1-r)}$ are not in Σ ([**3**, Introduction, Theorem XXXII]), B is a polynomial of positive degree in y_{h+1} whose initial and discriminant vanish at only finitely many points of $[0, 1]$. Let $T(h+1)$ denote the points of $[0, 1]$ at which either some coefficient of B is undefined or the initial or discriminant vanishes. There exists an analytic solution of $B=0$ on each of the open or half-open subintervals into which $[0, 1]$ is divided by the points of $T(h+1)$. Let $f^{(h+1)}$ be the function which is composed of these solutions. Then $f^{(h+1)}$ satisfies (a) and (b).

Let $P \in K\{y\}$ be of order $h+1$. If $P \in \Sigma$, $JP = L$, where L is a linear combination of the polynomials $A^{(0)}, \cdots, A^{(h+1-r)}$ and J is a product of powers of the initials of these polynomials. It follows that the substitution $y_i = f^{(i)}$, $i=0, \cdots, h+1$, annuls P at all points of $[0, 1]$ where the relevant functions are defined except, perhaps at those which are zeros of J. Since J is of order at most h and is not in Σ, the induction hypothesis shows that this set is finite and P is annulled at these points also. Hence, the first statement of (c) is satisfied for $s=h+1$. If $P \notin \Sigma$, there exists a polynomial $H \notin \Sigma$ of order at most h which is a linear combination of P and $A^{(0)}, \cdots, A^{(h+1-r)}$. By the induction hypothesis the substitution $y_i = f^{(i)}$, $i=0, \cdots, h$, annuls H at only finitely many points. Hence, the substitution $y_i = f^{(i)}$, $i=0, \cdots, h+1$, annuls P at only finitely many points. This completes the proof of (c).

Let $\lfloor x \rfloor$ denote as usual the greatest integer not exceeding x, where x is non-negative; and let $x^* = x - \lfloor x \rfloor$. Let $f(x) = f^{(\lfloor x \rfloor)}(x^*)$. Then $f(x)$ is defined for $x \geqq 0$, except at a set $S(f)$ which has no limit points, and is analytic in each of the intervals into which the non-negative real axis is divided by omission of the points of $S(f)$. Let R be the ring formed by adjoining $f(x), f(x+1), \cdots$, (all restricted to $x \geqq 0$) to K. If $P \in K\{y\}$, and t is non-negative, then the result of substituting $f(x+i)$ for y_i, $i=0, 1, \cdots$, in P and evaluating at t is the same as the result of substituting $f^{(i)}$ for y_i, $i=0, 1, \cdots$, in $P_{\lfloor t \rfloor}$ and evaluating at t^*. It follows from (c) that if $P \in \Sigma$ one obtains 0 for every $t \geqq 0$ for which the relevant functions are defined; but if $P \notin \Sigma$, one obtains 0 at not more than a finite number of points in any finite interval.

Let $g(x) \in R$, and assume that there is a bounded infinite subset I of $x \geqq 0$ such that $g(x)=0$, $x \in I$. There is a difference polynomial $P \in K\{y\}$ such that $g(x)$ is found by substituting $f(x)$ for y in P. By the second part of the conclusion of the preceding paragraph, $P \in \Sigma$. Hence, by the first part of the conclusion, $g(x)=0$. Therefore R consists of permitted functions and is a permitted difference ring. The quotient field L of R is in $\mathcal{K}$. The conclusion of the preceding paragraph also shows that L contains a generic zero, namely $f(x)$, of Σ.

If $\Sigma = \{0\}$, the construction used above for the $f^{(i)}$, $i<r$, is extended to all i, and $f(x)$ is then defined as before.

If the functions of K can be extended to $x<0$ with the obviously needed properties, as when $K=K_0$, then it is not at all difficult to modify the procedure for constructing $f(x)$ so as to obtain a function defined on the entire real line except for isolated points. Thus one finds a generic zero of Σ in an inversive difference field [**3**, p. 57] consisting of functions defined except for singularities on the real line.

Continuity. A function $f(x)$ defined on the non-negative real axis except at a set S which has no limit points will be called *essentially continuous* (e.c.) if either it is identically 0 or it is continuous at each point in the complement of S, and $1/f(x)$ has limit 0 at each point of S. The preceding construction may fail to yield e.c. functions. Though one could meet the requirements for essential continuity in the interiors of the intervals between integers, there seems to be no control over the behavior at the integers themselves. It would certainly be interesting to remove this deficiency.

Some idea of the problem involved in obtaining essential continuity can be obtained by considering the special case in which Σ is of order 0 and the coefficient field is K_0. We retain the earlier notation and also use $\mathfrak{M}$ to denote the variety of Σ, and a to denote a solution of Σ in the abstract sense. (Since the order is 0, a is a generic zero of Σ.)

The polynomial $A^{(0)}$ is necessarily of order 0. Its variety $\mathfrak{N}$ may have more than one irreducible component. ([**3**, Chapter VIII, Example 2], with rational coefficients replaced by complex coefficients. Let $A^{(0)}$ have the solution $\sqrt{x}+\sqrt{(x+t)}$.) Let us assume for the moment that $\mathfrak{N}$ has only the component $\mathfrak{M}$. In this case the $A^{(i)}$ are transforms of $A^{(0)}$, and it is easy to see that each $f^{(i)}$ may be obtained from the Riemann surface of $A^{(0)}$. In particular, choosing a continuous curve in this surface which covers the non-negative real axis, the corresponding values of y, where finite, furnish values of $f(x)$. The function $f(x)$ so obtained is essentially continuous and is continuous for all sufficiently large x, and this is, indeed, true of every element of $K_0\langle f(x)\rangle$.

It will be shown that, in particular, if a is normal over K, then $\mathfrak{N}$ has only one component. Let $\mathfrak{M}^*$ be a component of $\mathfrak{N}$ and a^* a generic zero of $\mathfrak{M}^*$ in some abstract difference overfield of K. Let L be the inversive closure of $K_0\langle a\rangle$ and M the inversive closure of $K_0\langle a^*\rangle$. Of course, the underlying fields of L and M are K_0-isomorphic and normal over K_0. Let $K_0 \subset K_1 \subset \cdots \subset K_r$ $(=L)$ be the Babbitt's decomposition [**1**, Theorem 2.3] or [**3**, Chapter VII, Theorem VII] for L. Suppose that for some i, $0 \leqq i < r$, it has been shown that there exists a K_0-isomorphism ϕ_i of K_i into M. By definition, if $i>0$, and from either the proof of Theorem 2.9 of [**1**] or from Theorem XIX of Chapter IX of [**3**] if $i=0$, K_{i+1}/K_i is equivalent to a benign extension with, say, minimal standard generator m_i. We may choose $m_i \in K_{i+1}$. Now ϕ_i extends to an isomorphism of the underlying field of L onto the underlying field of M. Let n_i be the image of m_i under this isomorphism. It follows from the Corollary to Theorem VI of Chapter VII of [**3**] that ϕ_i extends to a difference isomorphism of $K_i\langle m_i\rangle$ onto $(\phi_i K_i)\langle n_i\rangle$, and, hence, to a difference isomorphism of K_{i+1} into M. Since ϕ_0 exists (it is the identity) it follows by induction that L and M are K_0-isomorphic (in the difference field sense). Then $\mathfrak{M}=\mathfrak{M}^*$, and $\mathfrak{M}$ is the only component of $\mathfrak{N}$.

We now relinquish the hypothesis that a is normal over K_0. It follows from [**3**, p. 216] that $K_0\langle a\rangle$ is contained in a difference field $K_0\langle c\rangle$, where c is algebraic and normal over K_0. Let Π be the difference ideal of $K_0\{y\}$ with generic zero c. The results of the two preceding paragraphs show the existence of a generic zero $\bar{c}$ of Π in a member of K whose elements are continuous for sufficiently large x and essentially continuous. Let ϕ be the K_0-isomorphism from $K_0\langle c\rangle$ onto $K_0\langle\bar{c}\rangle$ such that $\phi c=\bar{c}$. Then ϕa is a generic zero of Σ. This proves the following theorem.

THEOREM III. *If Σ is a reflexive prime difference ideal of $K_0\{y\}$, and the order of Σ is 0, then Σ has a generic zero in a member of K each of whose elements is continuous for sufficiently large x and essentially continuous.*

Approximation. Let $\mathcal{S}$ denote a set of permitted functions, and let $g(x)$ be a permitted function. Then $g(x)$ is said to *adhere* to $\mathcal{S}$ at $p\geqq 0$, if $g(p+i)$ is defined, $i=0, 1, \cdots$, and for each $\epsilon>0$ and positive integer t there exists $h(x)\in\mathcal{S}$ such that $h(p+i)$ is defined and $|g(p+i)-h(p+i)|<\epsilon$, $i=0, \cdots, t$. If p is a point of adherence, so is $p+t$, $t=0, 1, \cdots$.

Suppose Σ is a difference ideal of $K\{y\}$, $K\in\mathcal{K}$, $\mathcal{S}$ a set of solutions of Σ in overfields of K which are members of $\mathcal{K}$, and $g(x)$ a permitted function such that the set C of points of adherence of $g(x)$ to $\mathcal{S}$ is dense in $x\geqq 0$. Then $g(x)$ is a solution of Σ. To prove this we note that without loss of generality we may assume $g(x)=0$. Let $A\in\Sigma$. Then $A=B+b(x)$, where B admits the solution 0 and $b(x)$ is a permitted function. If $c\in C$ is a point at which the coefficients of A are defined it follows from the definition of adherence that $b(c)=0$. Then $b(x)$ is 0 at a set of points dense in some interval. By the definition of permitted function $b(x)=0$, so that 0 is a solution of A.

If in particular $K=K_0$, $g(x)\in K_0$, then it is sufficient to know that $g(x)$ adheres to $\mathcal{S}$ at just one point p. For then $b(x)$ is a rational function such that $b(p+t)=0$, $t=0, 1, \cdots$, except possibly at the set of points at which some coefficient of A is undefined. This set is finite, so that $b(x)=0$.

The following approximation theorem permits an implication in the opposite direction.

THEOREM IV. *Let Σ be a reflexive prime difference ideal of $K\{y\}$, $K\in\mathcal{K}^*$, and let $g(x)$ be a solution of Σ in a member of $\mathcal{K}$ containing K. Let $\mathcal{S}$ denote the set of generic zeros of Σ in members of $\mathcal{K}$ containing $K\langle g(x)\rangle$. Then $g(x)$ adheres to $\mathcal{S}$ at a set of points dense in $x\geqq 0$.*

PROOF. If $\Sigma=[1]$ there is nothing to prove. If $\Sigma=[0]$ and $g(x)$ is a generic zero of Σ, then $g(x)\in\mathcal{S}$, so that the conclusion is trivial. In the remaining cases, $g(x)$ satisfies a nonzero difference polynomial of $K\{y\}$ so that the degree of transcendence of $K\langle g(x)\rangle$ over K is finite [**3**, Chapter II, Theorem VIII]. Hence, $K\langle g(x)\rangle\in\mathcal{K}^*$. We assume henceforth that Σ is neither $[0]$ nor $[1]$. The case $\Sigma=[0]$, $g(x)$ not a generic zero, is treated by a modification as at the end of the proof of Theorem II.

Let $A^{(0)}, A^{(1)}, \cdots$ be a characteristic sequence of Σ. Let r denote the order of Σ. For each positive integer s let $D^{(s)}$ denote the product of the initials and y_{i+r}-discriminants of the $A^{(i)}$, $i \leqq s-r$, or let $D^{(s)}=1$, if $s<r$.

Let J be any open set in $[0, 1]$. We must find in J a point of adherence of $g(x)$ to $\mathcal{S}$. Let I_0 be a nondegenerate closed interval contained in J. We shall construct inductively a sequence $I_0, I_1, \cdots$ of nondegenerate closed intervals each contained in the preceding. Suppose that, for some $s>0$, I_{s-1} has been obtained. Let q be an interior point of I_{s-1} such that $g(x)$ and the coefficients of $A^{(0)}, \cdots, A^{(s-r)}$ are defined and analytic at q. Let C be a circle in the complex plane with center q such that these functions are defined and analytic throughout C, and the intersection of C and the real axis is contained in I_{s-1}. The prime ideal $\Sigma_s \cap K[y, \cdots, y_s]$ of the ring $K[y, \cdots, y_s]$ does not contain $D^{(s)}$ and admits the solution $y_i=g(x+i)$, $i=0, \cdots, s$. Let K' denote the field generated by the coefficients of $A^{(0)}, \cdots, A^{(s-r)}$ as functions in C, and Σ_s' the prime ideal $\Sigma_s \cap K'[y, \cdots, y_s]$ of $K'[y, \cdots, y_s]$. Then Σ_s' admits the solution $y_i=g(x+i)$, $i=0, \cdots, s$, does not contain $D^{(s)}$, and does contain $A^{(0)}, \cdots, A^{(s-r)}$. By a theorem due to Ritt [**4**, p. 103] there exists an open subset C' of C such that Σ_s' has a solution $y_i=h_i(x)$, $i=0, \cdots, s$, analytic throughout C' and not annulling $D^{(s)}$ at any point of C', such that $|h_i(x)-g(x+i)| \leqq 1/s$, $x \in C'$, $i=0, \cdots, s$. Furthermore, an examination of the proof of this theorem shows that C' may be chosen to intersect the real axis. Let I_s be a nondegenerate closed interval contained in this intersection.

Let $p \in \bigcap_s I_s$. Given a positive integer t there exist complex numbers $\bar{y}_0, \cdots, \bar{y}_t$ such that:

(1) $|\bar{y}_i - g(p+i)| < 1/t$, $i=0, \cdots, t$.

(2) The substitution $x=p$, $y_i=\bar{y}_i$, $i=0, \cdots, t$, annuls $A^{(0)}, \cdots, A^{(t-r)}$.

(3) The substitution in (2) does not annul $D^{(t)}$.

It will be shown that the functions $f^{(i)}$ of the proof of Theorem II may be chosen so that $f^{(i)}(p)=\bar{y}_i$, $0 \leqq i \leqq t$. This is clearly possible for $i<r$, since to the functions selected as in the proof of Theorem II to be algebraically independent over $K\langle g(x)\rangle$ one may add suitable constants to obtain the desired values at p. Suppose $f^{(0)}, \cdots, f^{(h)}$ have been so selected, where $r-1 \leqq h < t$. Then on replacing the y_i by the $f^{(i)}$, $0 \leqq i \leqq h$, in $A^{(h+1-r)}$ and $D^{(h+1)}$ we obtain polynomials B and E respectively in y_{h+1}, with coefficients analytic at x, such that B is annulled and E is not on putting $x=p$, $y_{h+1}=\bar{y}_{h+1}$. Hence, $f^{(h+1)}$ may be chosen to assume the value $\bar{y}_{h+1}$ at p. Completing the construction of Theorem II there results a generic zero $f(x)$ of Σ in a member of $\mathcal{K}$ such that $|f(p+i)-g(p+i)|<1/t$, $i=0, \cdots, t$. This is sufficient to show that $g(x)$ adheres to $\mathcal{S}$ at p.

References

1. A. E. Babbitt, Jr., *Finitely generated pathological extensions of difference fields*, Trans. Amer. Math. Soc. **102** (1962), 63–81.

2. R. M. Cohn, *Manifolds of difference polynomials*, Trans. Amer. Math. Soc. **64** (1948), 133–172.

3. ———, *Difference algebra*, Wiley, New York, 1965.

4. J. F. Ritt, *Differential Algebra*, Amer. Math. Soc. Colloq. Publ. Vol. 33, Amer. Math. Soc., Providence, R. I., 1950.

Rutgers, The State University and
University of Puerto Rico