MODULAR FIELDS*

SAUNDERS MAC LANE, Harvard University

1. Introduction. The general theory of modular fields, though elementary in its presuppositions, offers an instructive cross-section of modern algebraic methods. These fields exhibit the generality of subject-matter inherent in abstract algebra, and at the same time illustrate the intimate connection between algebraic and arithmetic problems.

Modular fields arise first in number theory in the consideration of congruences with a prime modulus p. For integers a and b the ordinary definition states that

$$a \equiv b \pmod{p} \qquad \text{means that} \qquad p \text{ divides } (a - b).$$

Any integer a on division by p yields a quotient q and a remainder r,

$$a = qp + r, \qquad 0 \leqq r < p;$$

hence $a \equiv r \pmod{p}$, where the remainder r is one of the integers

$$F_p\colon \quad 0, 1, 2, \cdots, p-2, p-1. \tag{1}$$

Any integer is congruent to one of those in this set of p numbers.

With these numbers alone one can still carry out algebraic operations, provided one adds and multiplies these numbers in the ordinary fashion, and then reduces the answer by congruence to one of the numbers (1). For example, if $p=5$, the product $2 \cdot 3 = 6$ should really be $2 \cdot 3 \equiv 6 - 5 = 1$. In this fashion one can make multiplication and addition tables for $p=5$, as shown. It is strange

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

·	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

that this idea has not appeared more† in texts on number theory, for the idea is an essentially simple one. One can introduce it by the intuitively natural algebra of the words "even" and "odd," as

$$\text{even} \cdot \text{even} = \text{even}, \qquad \text{even} \cdot \text{odd} = \text{even}, \qquad \text{odd} \cdot \text{odd} = \text{odd},$$
$$\text{even} + \text{even} = \text{even}, \qquad \text{even} + \text{odd} = \text{odd}, \qquad \text{odd} + \text{odd} = \text{even}.$$

This is just the algebra of integers modulo $p=2$.

* An address delivered before the Mathematical Association of America at Columbus, Ohio, December 30, 1939.

† *Cf.* remarks in Weiss [**26**].

A congruence modulo p has all the properties of an equation; congruences can be added and multiplied term by term, and the relation of congruence is reflexive, symmetric, and transitive. If the modulus p is fixed, one might just as well dub congruence "equality." Every integer is then "equal" to one of the p symbols, $0, \cdots, p-1$, and the sums and products of these symbols, so identified, give exactly the algebra of the integers modulo p, as described above.

If one objects to rebaptizing "congruence" by fiat, one may adopt the more sophisticated procedure* of replacing each remainder r modulo p by the *class* r_p of all integers $r, r+p, r+2p, \cdots$ congruent to it. Such "congruence classes" are then added and multiplied according to the rules

$$r_p + s_p = (r+s)_p, \qquad r_p \cdot s_p = (rs)_p. \tag{2}$$

Furthermore the congruence classes r_p and s_p will be equal (*i.e.*, will contain the same elements) if and only if the integers r and s are congruent, so the desired "equality" has now been properly introduced. In any event the *integers modulo p* form a finite set of objects (1) satisfying all rules of algebra.

The presence of such arithmetic objects, which are certainly not ordinary numbers but which still obey ordinary algebra, is the reason why modern algebra is abstract. To separately discuss the algebra of numbers, then the algebra of congruence classes, then the algebra of functions, and so on would be most inefficient. Instead, theorems are better proved for any (abstractly conceived) system of objects whatever to which the basic rules of algebra apply.

These laws of algebra for a set F of objects, such as the integers modulo p, are codified as follows: For a and b in F there is uniquely defined a *sum* $a+b$ and a *product* $a \cdot b$. This product is *commutative* $[ab=ba]$ and *associative* $[a(bc)=(ab)c]$, as is also the sum. The *distributive* law $a(b+c)=ab+ac$ holds for all a, b, and c. The set F contains a zero 0 and a unit 1, with the characteristic properties

$$a + 0 = a = 0 + a, \qquad 1 \cdot a = a = a \cdot 1,$$

respectively. Finally, *subtraction* and *division* are possible, which is to say that the equations $a+x=0$ and $b \cdot y=1$ have solutions x and y in F, except when $b=0$. Any set F of elements with all these properties is called a *field*. One may say that a field is any system of elements within which addition, subtraction, multiplication, and division (excluding division by zero) can be carried out in the usual fashion.

Well known fields are: (a) the set of all rational numbers; (b) the set of all real numbers; (c) the set of all complex numbers. The field (1) composed of the integers modulo p is often called the *Galois field* $GF[p]$. A *modular field* is any field containing such a $GF[p]$.

These fields $GF[p]$ are not the only finite fields. One may construct larger fields by simply adjoining to a $GF[p]$ the roots of certain algebraic equations. The process resembles the construction of the complex numbers from the field R

* *Cf.* Albert [1, p. 7]; van der Waerden [27, p. 13]; or Mac Lane [17, Chapter I].

of real numbers. Here one adjoins to R a symbol i representing a root of the equation $x^2+1=0$; the field C of all complex numbers $a+bi$ then contains everything which can be expressed rationally in terms of i and real numbers. The fact that C is generated over R by adjoining i is symbolized by $C=R(i)$. Note in particular that the polynomial x^2+1 used to generate this extension is *irreducible* over R, because it cannot be factored into polynomials of smaller degree with coefficients in R.

In similar vein consider the polynomial $f(x)=x^2+x+1$ over the field F_2 with two elements (the integers modulo 2). Neither $f(1)$ nor $f(0)$ is zero, so this polynomial $f(x)$ has no roots in F_2, hence has no linear factors, hence is irreducible over F_2. Invent a symbol u to denote a root of $f(x)=0$, so that

$$u^2 + u + 1 = 0, \qquad u^2 = -u - 1 = u + 1.$$

(Recall that $-1=+1$, modulo 2.) All higher powers of u can thereby be successively reduced to linear expressions in u. Reciprocals can be similarly reduced, so that the field generated by u contains all told just four linear expressions: 0, 1, u, $u+1$. These combine under addition and multiplication

+	0	1	u	$u+1$
0	0	1	u	$u+1$
1	1	0	$u+1$	u
u	u	$u+1$	0	1
$u+1$	$u+1$	u	1	0

·	0	1	u	$u+1$
0	0	0	0	0
1	0	1	u	$u+1$
u	0	u	$u+1$	1
$u+1$	0	$u+1$	1	u

as shown in the tables. The process of obtaining this field by adjoining to the original F_2 a root u of x^2+x+1 is known as *algebraic extension* of F_2, and the resulting field $F_2(u)$ is called a Galois field of 4 elements.

For each prime p and each integral exponent n one may analogously extend the field of integers modulo p to a field consisting of exactly* p^n elements. As E. H. Moore first showed, *any* two fields with p^n elements each are algebraically indistinguishable (isomorphic). The arithmetic origin of all these finite fields is the study of algebraic integers. If $\mathfrak{p}$ is a prime ideal in a field K of algebraic numbers, then the congruences modulo this ideal behave as do ordinary congruences, and yield like them a finite field with p^n elements, where p^n is the so-called "norm" of the ideal $\mathfrak{p}$. The properties of the resulting finite fields play an essential rôle in the class field theory and in the study of rational division algebras (Albert [2, ch. 9]).

2. Characteristics. The integers modulo p have one peculiar property. The unit 1, added p times to itself, yields $p\equiv 0 \pmod{p}$ as answer; hence

$$1 + 1 + \cdots + 1 = 0, \qquad (p \text{ summands}). \tag{3}$$

* See detailed discussion of finite fields in van der Waerden [27, §31]; or Albert [1, p. 166].

On multiplying this equation by any integer a, one has

$$a + a + \cdots + a = 0, \qquad (p \text{ summands}), \tag{4}$$

in the Galois field F_p. Any field F, all of whose elements a have the property (4), is called a field of *characteristic* p, or a *modular field*. It can be shown* that any non-modular field has an infinite characteristic, in the sense that $a \neq 0$ entails $a+a+\cdots+a \neq 0$, for any number of summands. Any finite field of p^n elements essentially contains the integers modulo p, hence satisfies (3) and therefore (4). Thus any finite field is modular.

Watch the effect of (4) on the binomial expansion,

$$(a + b)^p = a^p + pa^{p-1}b + (p(p-1)/2)a^{p-2}b^2 + \cdots + pab^{p-1} + b^p.$$

According to the genesis of this expansion, the term $pa^{p-1}b$ second on the right really represents a sum of p products $a^{p-1}b+a^{p-1}b+\cdots+a^{p-1}b$. In a field of characteristic p this sum is zero. The other intermediate terms of the binomial expansion suffer the same fate, for each binomial coefficient $p(p-1)/2, \cdots, p$ is a multiple of the characteristic p. One has left only

$$(a + b)^p = a^p + b^p, \qquad (a, b \text{ in } F \text{ of characteristic } p). \tag{5}$$

As S. C. Kleene has remarked, a knowledge of the case $p=2$ of this equation would corrupt freshman students of algebra!

The pth power of a product is always a product of pth powers, so the rules

$$(a \pm b)^p = a^p \pm b^p, \qquad (ab)^p = a^p b^p, \qquad (a/b)^p = a^p/b^p \tag{6}$$

hold in any field of characteristic p. These rules state that the process of raising to a pth power leaves the operations of addition, division, *etc.*, unchanged. This process yields a correspondence

$$a \longleftrightarrow a^p, \qquad (\text{from } F \text{ to } F^p), \tag{7}$$

which carries the field F into the field F^p composed of all pth powers from F. The correspondence is one-to-one, for the equality of two pth powers $a^p=b^p$ would entail $0=b^p-a^p=(b-a)^p$, and hence $b=a$. To summarize, the correspondence $a \longleftrightarrow a^p$ is an isomorphism, where an *isomorphism* between two fields is defined to be any one-to-one correspondence which preserves sums and products.

Repeated application of the rules in (6) shows that the pth power of any rational expression can be computed by applying the exponent p to each term or factor in the expression. In particular,

$$(1 + 1 + \cdots + 1)^p = 1^p + 1^p + \cdots + 1^p = 1 + 1 + \cdots + 1 \tag{8}$$

holds in the field of integers modulo p. If we use m summands here, this is $m^p=m$. In terms of congruences this is $m^p \equiv m \pmod{p}$, which is the little Fermat Theorem!

---

* *Cf.* Albert [1, p. 30]; Mac Lane [17, §21]; van der Waerden [27, §25].

3. Algebraic and transcendental extensions. Our major concern is the structure of the general modular field, finite or infinite. In the analogous case of fields of numbers it is customary to distinguish the algebraic numbers, such as $\sqrt{3}$, which satisfy some polynomial equation with rational coefficients, from the transcendental numbers (e, π), which satisfy no such equation. In general, let a given field F be contained in any larger field K. An element u of K is *algebraic* over F if u is a root of a polynomial

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1 x + a_0 \tag{9}$$

with coefficients a_i in F. If this equation $f(x)=0$ for u be chosen with a degree n as small as possible, the polynomial $f(x)$ is *irreducible* over F. For, a reducible $f(x)$ would have factors $f(x)=f_1(x)f_2(x)$ with coefficients in F, and u would satisfy one of the equations $f_1(x)=0$, $f_2(x)=0$, of degree smaller than n. An element u in K not algebraic over F is called *transcendental*; for u transcendental, $f(u)=0$ implies that all the coefficients in $f(x)$ are zero.

Important is not the element u in K by itself, but the field $F(u)$ which it generates. The field consists of all rational combinations of u with coefficients in F, and is called a *simple extension* of F, "algebraic" or "transcendental" according as u is algebraic or transcendental over F. This dichotomy is the root of one of the basic results found by Steinitz in his pioneering investigations of fields (Steinitz [**23**]): *Any modular field can be obtained by successive transcendental and algebraic extensions of a field (isomorphic to the field) of integers modulo p.*

Such extensions can be used not only to build up a given field K from a subfield F, but also to manufacture new fields from old. Given a polynomial $f(x)$ irreducible over a field F, one can concoct a symbol u for a root of this polynomial and construct therewith an algebraic extension $F(u)$ generated by the root u. In point of fact, $F(u)$ consists of elements expressible as polynomials $b_0+b_1u+\cdots+b_{n-1}u^{n-1}$, with coefficients in F and of degree less than the degree n of the given $f(x)$.

Alternatively, a variable t over a modular field gives rise to rational functions

$$\frac{g(t)}{h(t)} = \frac{b_0 + b_1 t + \cdots + b_r t^r}{c_0 + c_1 t + \cdots + c_m t^m}, \qquad (c_i,\ b_j \text{ in } F, \text{ not all } c_i = 0). \tag{10}$$

Under the usual rules for adding and multiplying such expressions, the totality of these rational functions is a field $F(t)$ which is a simple transcendental extension of F. If F is a finite field, the resulting field $F(t)$ is the simplest instance of an infinite modular field.

4. Inseparable equations. Over the transcendental extension $F(t)$ there are in turn algebraic extensions, such as that generated by a root of the polynomial $f(x)=x^p-t$. This $f(x)$ is irreducible over $F(t)$, for if it could be factored, the denominators in t could be eliminated, and we could write $x^r-t=g(x, t)h(x, t)$, with factors which are polynomials in x and t. Since the product of these two polynomials is linear in t, one of them must be linear in t, while the other cannot

involve t at all! This is absurd unless one of the factors is a constant; hence $f(x)$ is indeed irreducible.

But trouble arises with the introduction of a root u for this equation $x^p-t=0$. Since this u is a pth root of t, we have a factorization

$$x^p - t = x^p - u^p = (x - u)^p, \tag{11}$$

according to the rule (6) for the pth power of a difference. This means that u is a p-fold root of x^p-t, so this irreducible polynomial has all its roots equal, and t has only one pth root.

This differs drastically from the usual situation with ordinary complex nth roots, for an irreducible polynomial $f(x)$ with *rational* coefficients can never have a multiple root. Let us trace the proof of this fact. If $f(x)$ has a complex number r as m-fold root, then $f(x)=(x-r)^m g(x)$, with $m>1$. The derivative is

$$f'(x) = (x - r)^{m-1}[mg(x) + (x - r)g'(x)]. \tag{12}$$

Since $m>1$, this insures that $f(x)$ and $f'(x)$ have a common factor $(x-r)^{m-1}$, not a constant. But the highest common factor of $f(x)$ and $f'(x)$ can be found by the euclidean algorithm, using only rational operations. This highest common factor then has rational coefficients, and its degree is at most that of $f'(x)$. It must divide $f(x)$, counter to the assumed irreducibility of that polynomial.

Can this contradiction be deduced for a polynomial $f(x)$, irreducible not over the rationals but over some modular field, and having a multiple root r in a larger field? The derivative $f'(x)$ of calculus is no longer available, but for any polynomial $f(x)$ as in (9) a "formal" derivative can still be defined as

$$f'(x) = na_n x^{n-1} + (n - 1)a_{n-1}x^{n-2} + (n - 2)a_{n-2}x^{n-3} + \cdots + a_1. \tag{13}$$

Here the coefficient ia_i of the term x^{i-1} is to denote the sum

$$ia_i = a_i + a_i + \cdots + a_i, \qquad (i \text{ summands}). \tag{14}$$

Apply this derivative to the troublesome polynomial x^p-t of (11). We find

$$(x^p - t)' = px^{p-1} = x^{p-1} + \cdots + x^{p-1} = 0, \qquad (p \text{ summands}).$$

No wonder that an argument on the H. C. F. of x^p-t and 0 runs aground! Looking back, one sees that the argument following (12) about multiple roots will work, except in such cases when $f'(x)$ vanishes.

When do all coefficients ia_i of $f'(x)$ vanish? In a modular field $ia_i=0$ means either that a_i itself is zero, or that the number i of summands, in (14), is a multiple of the characteristic p. A coefficient a_i can thus differ from zero only for terms $a_i x^i$ with exponent $i\equiv 0 \pmod{p}$. The vanishing of $f'(x)$ means therefore that $f(x)$ can involve x only as powers of x^p, so that $f(x)$ has the form

$$g(x) = b_m x^{mp} + b_{m-1}x^{(m-1)p} + \cdots + b_1 x^p + b_0. \tag{15}$$

An irreducible polynomial $g(x)$ of this form must always have p-fold roots. Such a polynomial is called *inseparable* (its roots cannot be "separated" into distinct roots). Many properties of ordinary equations fail for inseparable equations.

An element u algebraic over a modular field F is called *separable* over F if the irreducible equation for u is separable (*i.e.*, has no multiple roots). Of the inseparable algebraic elements the simplest examples are pth roots which satisfy inseparable equations $x^p=a$. Consider an arbitrary inseparable element u, root of an inseparable polynomial (15) of degree mp. This polynomial involves only pth powers of its variable, so u^p is a root of an equation

$$h(y) = b_m y^m + b_{m-1} y^{m-1} + \cdots + b_1 y + b_0, \tag{16}$$

of smaller degree m. The adjunction of the root u to our field F can then be effected in two stages

$$F \longrightarrow F(u^p) \longrightarrow F(u^p, \sqrt[p]{u^p}) = F(u).$$

The element u^p first adjoined may still belong to an inseparable equation $h(y)=0$; in that event the process can be reapplied to get u^{p^2} satisfying an equation of still smaller degree. The adjunction of an inseparable algebraic element to a modular field can be accomplished by adjoining successive pth roots of a suitable separable algebraic element (Steinitz [**23**]). This reduction of algebraic extensions to separable extensions followed by extensions by pth roots, indicates that the novel properties are concerned chiefly with the latter type of extension.*

5. Perfect fields. There are no inseparable algebraic elements over the field of integers modulo p, for this field already contains the pth roots of all of its elements—indeed, the Fermat Theorem, $a^p=a$, asserts that every element is its own pth root. A *perfect* field F of characteristic p is a field in which each element a has a pth root. Over such a field each pth root equation $x^p=a$ is reducible, as $x^p-a=(x-\sqrt[p]{a})^p$. More generally *any inseparable polynomial* $g(x)$ *involving only pth powers of* x *must be reducible over a perfect field.* For, each coefficient b_i of the polynomial $g(x)$ in (15) has in F a pth root $b_i^{1/p}$; according to the simple behavior of pth powers this gives a factorization

$$g(x) = (b_m^{1/p} x^m + b_{m-1}^{1/p} x^{m-1} + \cdots + b_1^{1/p} x + b_0^{1/p})^p.$$

Every finite field F *is perfect,* hence has no inseparable algebraic extensions. To prove this, recall the correspondence $a \longleftrightarrow a^p$ of (7), which is a one-to-one correspondence between *all* elements of F and those elements a^p which are pth powers. Since there are but a finite number of elements in F, there must be the same number of pth powers. This means that every element is a pth power.

A simple transcendental extension $F(t)$ of a modular field can never be perfect. To verify this we need only produce an element with no pth root in the field. The variable t itself is such an element, for if t had as pth root some rational function $g(t)/h(t)$ in the field, t would equal $[g(t)/h(t)]^p$, a pth power

* Technically, the least power $q=p^e$ such that u^q is separable over F is known as the *exponent* of u over F. The *degree* of u over F is the degree of its irreducible equation, while the degree of u^q is known as the *reduced degree* of u.

which can be calculated by the rule (6). In the notation of (10), the result is

$$(c_0^p + c_1^p t^p + \cdots + c_m^p t^{mp})t = b_0^p + b_1^p t^p + \cdots + b_r^p t^{rp},$$

an identity which clearly cannot hold good. For similar reasons a multiple transcendental extension $F(t_1, t_2, \cdots, t_n)$, consisting of all rational functions of n independent variables t_i, cannot be a perfect field.

6. Galois theory. To what extent can one generalize to modular fields the ordinary properties of fields of rational and algebraic numbers? A major topic is the Galois theory, which analyzes the solvability of a polynomial equation $f(x) = 0$ over a field F. The roots $r_1, \cdots, r_n$ of this equation generate over F a *root field*

$$K = F(r_1, r_2, \cdots, r_n), \qquad \text{where} \qquad f(x) = (x - r_1)(x - r_2) \cdots (x - r_n); \tag{17}$$

the Galois Theory studies K in terms of its group of automorphisms, each of which is an isomorphism of the field K with itself, induced by a permutation of the roots r_i. Should these roots all be equal, the only such permutation is the identity, and the theory breaks down. Only if one assumes that the roots are all distinct, *i.e.*, that $f(x)$ is separable, does the standard theory of root fields hold* over a modular F.

This straightforward generalization does not suffice for irreducible *in*separable polynomials. The first process to fail is the construction of a "Galois resolvent," which is an equation with a root u in K such that all the roots r_i can be rationally expressed in terms of this single quantity u. In terms of fields, this means that the multiple algebraic extension $K = F(r_1, \cdots, r_n)$ can be represented as a simple extension $F(u)$. Over an imperfect field F there may be multiple algebraic extensions which cannot be so represented. Consider for instance the rational function field,

$$F_0 = P(t_1, t_2), \qquad P \text{ perfect}, \tag{18}$$

in two independent variables t_1 and t_2. An adjunction of pth roots will yield an extended field

$$K_0 = F_0(u_1, u_2); \qquad u_1^p = t_1, \qquad u_2^p = t_2, \tag{19}$$

which consists of all elements expressible as polynomials

$$w = \sum_{i,j} a_{ij} u_1^i u_2^j = h(u_1, u_2), \qquad (i, j = 0, \cdots, p - 1), \tag{20}$$

with coefficients a_{ij} in F_0. This field K_0 is not a simple extension $K_0 = F_0(w)$ for any w. For, if there were a generator w, then by the rule for pth powers,

$$w^p = \sum_{i,j} a_{ij}^p u_1^{ip} u_2^{jp} = \sum_{i,j} a_{ij}^p t_1^i t_2^j$$

* *Cf.* Albert [1, ch. VIII]; van der Waerden [27, ch. 7]; Mac Lane [17, §68].

is in F_0, so w is a pth root of an element of F_0. That such a single pth root could generate the field K_0 containing two independent pth roots u_1 and u_2 is unreasonable. This hunch can be substantiated by an argument on the degree* of the extension K_0 of F_0.

If a multiple extension does not have one generator, what is then the *minimum* number of generators? Miriam Becker [6] has recently found the answer. Over the particular field $P(t_1, t_2)$ of (18) it appears that *any* multiple algebraic extension can be expressed by two generators, just as in the case of the special extension K_0 of (19). The underlying reason is the presence of just two independent pth roots, $\sqrt[p]{t_1}$ and $\sqrt[p]{t_2}$, not in the field $P(t_1, t_2)$; the pth root of any other rational function $g(t_1, t_2)$ in the field can be expressed by the rule (6) in terms of these two pth roots, together with pth roots of coefficients which already lie in the perfect base field P.

Over any modular field F one calls the r pth roots $a_1^{1/p}, a_2^{1/p}, \cdots, a_r^{1/p}$ *p-independent* if no one of them can be rationally expressed in terms of F and the others. Becker proves that *any multiple algebraic extension of an imperfect field F can be generated by m elements, where m is the maximum number of independent pth roots over F.* If $m=0$, F is perfect: if $m=1$, any multiple algebraic extension is simple, as shown by Steinitz.

7. Derivatives. The solution of an ordinary equation $f(x)=0$ by radicals (if possible) proceeds in successive stages which correspond to successive fields lying between the coefficient field F and the root field K. For a separable equation the whole array of possible intermediate fields is finite—but not so for some inseparable extensions. Between the fields F_0 and K_0 of (19) lie infinitely many distinct fields $F_0((t_1+t_2^m)^{1/p})$, with $m=1, p+1, 2p+1, \cdots$. For a separable equation the fields intermediate between K and F can be put into one-to-one correspondence with the sub-groups of the Galois group of automorphisms of K over F. This certainly fails for an inseparable extension like (19), for in that case the Galois group of K_0 over F_0 consists of the identity alone and so has no proper sub-groups to correspond to intermediate fields. Specifically, the Galois group consists of all isomorphisms of K_0 with itself which leave fixed each element in the base field F_0; but an isomorphism leaving fixed the elements t_1 and t_2 of F_0 must likewise leave fixed their *unique* pth roots u_1 and u_2 and hence must leave all elements of K_0 fixed.

For this description of intermediate fields by the Galois group Jacobson has found a substitute, in the special case of extensions K obtained by adjoining any number of pth roots to a modular field F, as

$$K = F(a_1^{1/p}, a_2^{1/p}, \cdots, a_n^{1/p}), \qquad \text{each } a_i \text{ in } F. \tag{21}$$

By a piece of poetic justice, his solution depends on exploiting the very formal

* This degree is the maximum number of elements of K_0 "linearly independent" over F_0. This maximum is p^2, for any w is linearly dependent on the p^2 elements $u_1^i u_2^j$ of (20). For a simple extension $F_0(w)$ the degree would be only p. Hence $F_0(w)$ cannot equal K_0.

derivatives whose misbehavior (*cf.* §4) is at the root of inseparability. For example, in the field K_0 of (19) one has two "derivative" operators D_1 and D_2, defined for the arbitrary element $w=h(u_1, u_2)$ of (20) by

$$(22) \qquad h(u_1, u_2)D_1 = \partial h(u_1, u_2)/\partial u_1, \qquad h(u_1, u_2)D_2 = \partial h(u_1, u_2)/\partial u_2.$$

This time the properties of pth powers are fortunate, for $u_1^pD_1 = pu_1^{p-1} = 0$, as it ought to be, for $u_1^p = t_1$ is in the base field and so should have derivative 0 according to the definition (22). These derivatives can be used to characterize sub-fields of K_0; for example, the sub-field $F_0(u_1)$ consists of everything annihilated by the operator D_2 (*i.e.*, of all w with $wD_2=0$).

In general, Jacobson considers [12] all *formal differentiation operators* D which map K into itself by a correspondence $w \to wD$ which carries elements of F into zero and which obeys the usual formal rules for differentiation:

$$(v + w)D = vD + wD, \qquad (vw)\,D = v(wD) + (vD)w.$$

From any two such operators D_1 and D_2 one may construct new differentiations $D_1 \pm D_2$, D_1^p, and D_1c, for c in F. Furthermore, the commutator $[D_1, D_2] = D_1D_2 - D_2D_1$ is again a formal differentiation. This commutator satisfies the identity

$$[[D_1, D_2], D_3] + [[D_2, D_3], D_1] + [[D_3, D_1], D_2] = 0,$$

which is one of the essential postulates for a Lie algebra. The set $\mathfrak{L}$ of all differentiations is in fact a Lie algebra over the base field F. This algebra acts as a substitute for the Galois group of a field K of type (21), in the sense that *there is a one-to-one correspondence between the fields intermediate between K and F and the restricted Lie sub-algebras of the algebra $\mathfrak{L}$ of all formal differentiations of K over F.* For this purpose a *restricted* sub-algebra of $\mathfrak{L}$ is a sub-set $\mathfrak{L}'$ of $\mathfrak{L}$ which is itself a Lie algebra and which is restricted to contain D^p for each D of $\mathfrak{L}'$.

8. Algebraic geometry. A skew curve can be represented as the intersection of two surfaces, which may often be taken as cylinders

$$(23) \qquad f(x, y) = 0, \qquad g(x, z) = 0$$

with axes parallel to the z and y coördinate axes, respectively. If f and g are polynomials, the intersection of these cylinders is an algebraic curve. Alternatively, x may be viewed as a quantity transcendental over the field C of complex numbers; the polynomial equations then make the quantities y and z algebraic over the field $C(x)$ of rational functions of x. All told they give a field $C(x, y, z)$ generated by "algebraic functions" y and z of x. This field is the algebraic invariant of the curve (23). The ordinary analytic theory of these algebraic function fields can be developed, without using the geometry of the Riemann surface, if the base field C of complex numbers is replaced by a perfect modular field P or even by an imperfect one.*

* *Cf.* general discussion of these abstract algebraic functions in Mac Lane-Nilson [19] or Schilling [21]. Especially interesting is the introduction of a Riemann Zeta function when P is finite (Hasse [9]), the peculiar behavior of the Weierstrass points whenever P is modular (Schmidt [22]), and the generalizations of Abelian functions (Schilling [20]).

In an n-dimensional euclidean space an r-dimensional algebraic manifold can be described as the set of points common to $n-r$ suitable algebraic hypersurfaces. These hypersurfaces may be taken, as in (23), in the form of "cylinders"

$$(24)\quad f_1(y_1, \cdots, y_r, y_{r+1}) = f_2(y_1, \cdots, y_r, y_{r+2}) = \cdots = f_{n-r}(y_1, \cdots, y_r, y_n) = 0,$$

where each f_i is an irreducible polynomial actually containing y_{r+i}. As coefficients in (24) we use not complex numbers but elements from a perfect modular field P. If this field P is finite, this means that we are considering a manifold in some finite affine (or projective) geometry, consisting of a finite number of "points" specified by coördinates in P. Algebraically, the symbols $y_1, \cdots, y_n$ related by (24) generate a field $K=P(y_1, \cdots, y_r, y_{r+1}, \cdots, y_n)$, consisting of all rational functions of these quantities, subject only to the rules of algebra and the special conditions (24). This field is obtained from the base field P by r successive simple extensions by the transcendentals $y_1, \cdots, y_r$, followed by $n-r$ successive algebraic extensions by the roots $y_{r+1}, \cdots, y_n$ of the polynomial equations (24). In a sense, the geometry of the manifold depends on the structure of this field.

What of the presence of inseparable equations in the definition (24) of such a manifold? Suppose, for instance, that the equation $f_1=0$ is inseparable in y_{r+1}, so that this variable appears only as a pth power. Certainly this could not simultaneously be the case for all the variables $y_1, \cdots, y_r, y_{r+1}$ in f_1, for in that event we could extract the pth root of every term in the equation $f_1=0$, thus making $f_1=(g_1)^p$, counter to the assumed irreducibility of f_1 over the perfect field P. Suppose then that y_1 is one of the variables which does not appear in $f_1(y_1, \cdots, y_r, y_{r+1})$ only as a pth power. The equation $f_1(y_1, \cdots, y_{r+1})$, which originally defined y_{r+1} inseparably over the field $P(y_1, \cdots, y_r)$, can be turned about and viewed as a definition of y_1 as a quantity *separable* and algebraic over the field $P(y_2, \cdots, y_r, y_{r+1})$, generated by the r independent transcendentals $y_2, \cdots, y_{r+1}$. A further juggling of the independent variables can then be applied to any subsequent equations of (24) which may be inseparable. Hence the result: *If a field $K=P(y_1, \cdots, y_n)$ is obtained from a perfect field P by adjoining a finite number of elements $y_1, \cdots, y_n$, one can find for K a generation $K=P(t_1, \cdots, t_r; u_1, \cdots, u_{n-r})$ involving r simple transcendental extensions by variables t_i, followed by $n-r$ separable algebraic extensions.* Whenever independent transcendents t_i in K have this property, that every element in K is *separable* and algebraic over $P(t_1, \cdots, t_r)$, we say that the $t_1, \cdots, t_r$ form a *separating transcendence basis* for K over P.

This construction of separating transcendence bases was discovered independently for different purposes: by the author, in connection with Albert's theory of pure forms (Albert [**4**]); by van der Waerden [**28**], for a new proof of the theorem that two distinct irreducible algebraic manifolds M_r and M_{n-r} in projective n-space intersect in a finite number of points, and, moreover, that the "number" of points, properly counted, is the product of the degrees of M_r and M_{n-r}.

9. Preservation of independence. The troubles of inseparable equations can be avoided whenever we find a separating transcendence basis for the field under consideration. Unfortunately this cannot always be done. Suppose, for instance, that the base field is the field $F_0=P(t_1, t_2)$ of all rational functions of two transcendents t_1 and t_2 over a perfect field P, and construct a larger field L by adjoining first a new transcendent z and then an algebraic element u, with

$$u^p = t_1 + t_2 z^p, \qquad L = F_0(z, u). \tag{25}$$

Since the pth root u is inseparable over $F_0(z)$, this z is surely not a *separating* transcendence basis for L over F_0. The order of adjunction might have been inverted, adding u first as a transcendent to L and then z, but the equation (25) indicates that z would then be a pth root. The same trouble would always arise: one can prove that L has over F_0 *no* separating transcendence basis.* The same troublesome example arises in Krull's general ideal theory [13].

To find the reason for this absence of separability one must look at the possible independent pth roots in the base field F_0. In §6 we saw that the pth roots $\sqrt[p]{t_1}$ and $\sqrt[p]{t_2}$ were p-independent there, because neither can be expressed in terms of F_0 and the other. These pth roots are no longer p-independent in the top field L, for the defining equation (25) of that field gives an expression $\sqrt[p]{t_1}=u-z\sqrt[p]{t_2}$. This suggests that we restrict attention to those extensions L over F which *preserve p-independence*, in the sense that any set of p-independent pth roots over F remains p-independent over L. The relevance of this concept is indicated by the following alternative description: *a field L preserves p-independence over F if and only if the adjunction to F of any finite set of elements $y_1, \cdots, y_n$ from L yields a field $F(y_1, \cdots, y_n)$ which has over F a separating transcendence basis.*

This concept also makes it possible to find explicit conditions that given extensions have separating transcendence bases (Mac Lane [16]). One simply stated result is this: *If a field K has a finite separating transcendence basis over a sub-field M, then any field L between K and M also has a finite separating transcendence basis over M.* In other words, one can find a set S of independent transcendents in L, such that every element of L satisfies over $M(S)$ an algebraic irreducible equation without multiple roots.

10. General field towers. What can be said of the structure of arbitrarily complicated modular fields? The fields $P(y_1, \cdots, y_n)$ associated with algebraic manifolds had separating transcendence bases over a perfect field P. Does every modular field have a separating transcendence basis T over a suitable perfect sub-field?

The answer is no. A simple counterexample may be built from the extension $P(t)$ of a finite field P by a transcendental t. We saw in §5 that $P(t)$ is imperfect because t has in it no pth root. If we try to embed $P(t)$ in a larger field P' which

* Even though, according to the Theorem of §8, L has over the *original* perfect field P a separating transcendence basis consisting of u, z, and t_1.

will be perfect, we must have in P' a pth root $t^{1/p}$ and hence the whole rational function field $P(t^{1/p})$ generated by this root. In this field $t^{1/p}$ has no pth root, so we add $t^{p^{-2}}$, and so on, till we have the "tower"

$$P(t) \subset P(t^{p^{-1}}) \subset P(t^{p^{-2}}) \subset P(t^{p^{-3}}) \subset \cdots . \tag{26}$$

The field enveloping everything in this tower may be called $P(t^{p^{-\infty}})$; it consists of all elements lying in any one of the fields (26). Furthermore this sum field $P(t^{p^{-\infty}})$ is perfect, for an element in any one of the fields of (26) does have a pth root in the next field of the tower.

This perfect field $P(t^{p^{-\infty}})$ can have over P no separating transcendence basis. Any such basis would consist of a single transcendent t', which must lie in some one of the fields $P(t^{p^{-e}})$ of the tower (26). The generating element $t^{p^{-(e+1)}}$ of the next field is then a quantity inseparable over $P(t')$, so t' cannot have been the desired separating basis.

The tower (26) as written shows $P(t^{p^{-\infty}})$ generated by a transcendental extension followed by successive (inseparable) extractions of pth roots. Nevertheless each field of this tower, considered by itself, is a simple transcendental extension of P by $t^{p^{-e}}$. The whole field is thereby approximated by a tower of fields, each of which has a separating transcendence basis over the base field P, and each of which consists of pth powers of elements in the next field. F. K. Schmidt has shown that any perfect field P' has a similar "separating tower" over any one of its perfect sub-fields. He also stated without proof an analogous tower theorem for an imperfect field, but it was later shown by examples* that this general theorem could not hold. Recently F. K. Schmidt and the author have jointly [18] found a modified tower theorem: *If a modular field K is generated from a perfect sub-field P by a denumerable number of elements, then there is a sub-field L with a separating transcendence basis over P and a tower of fields $L \subset M_0 \subset M_1 \subset \cdots$ which collectively exhaust K, such that each M_i has over L a separating transcendence basis and is generated over L by pth powers from M_{i+1}.* The non-denumerable cases can then be broken down into a transfinite sequence of denumerable steps, each of which "preserves p-independence" in the sense discussed in §9.

The separability of these field towers is essential to get polynomials with distinct roots, in order to apply an implicit function theorem.† This is used in the proof of the structure theorem for p-adic fields (*cf.* Hasse-Schmidt [10]). These p-adic fields are fields topologically complete with respect to a suitable norm (or "absolute value"), obtained by extending the norm for the p-adic numbers of Hensel.‡ These p-adic fields are not themselves modular fields, but they determine a congruence relation $a \equiv b \pmod{p}$ from which modular fields can be obtained by the standard arithmetic device.

* *Cf.* Mac Lane [15]. Curiously enough, these examples involve a use of the modular law of lattice theory!

† The so-called Hensel-Rychlik theorem; *cf.* Albert [1] or Mac Lane-Nilson [19, §11].

‡ See the description in C. C. MacDuffee [14].

11. Troublesome examples. The extent of our ignorance of general modular fields can be forcibly illustrated by various startling examples. The field $P(t^{p^{-\infty}})$ used to illustrate §10 was still manageable, for though it had no separating transcendence basis, it at least was itself perfect. But can there be an imperfect field K which has no separating transcendence basis over some perfect sub-field P? There is indeed such a K, for which P may even be chosen as the maximum perfect sub-field. Over a finite field P choose a countable set of indeterminates $t_1, t_2, \cdots$, and then introduce additional algebraic elements in accord with the inseparable relations

$$y_1^p = t_1 + t_2 t_3^p, \qquad y_2^p = t_2 + t_3 t_4^p, \qquad y_3^p = t_3 + t_4 t_5^p, \cdots. \tag{27}$$

Our example is the field $K = P(t_1, t_2, \cdots ; y_1, y_2, \cdots)$. Since the y's are pth roots, the t's clearly cannot form a separating transcendence basis. One might try to invert the equations (27) to define everything in terms of the basis $t_1, t_2, y_1, y_2, y_3, \cdots$, but that still leaves the pth roots such as $t_3^p = (y_1^p - t_1)/t_2$. It can be shown that no method of picking a transcendence basis for K over P will yield a basis which is separating, and this example is but a taste of the trouble possible (*cf.* [15], [16]).

12. p-Algebras. The relevance of the study of inseparable extensions to other algebraic questions is clearly illustrated by the p-algebras, which are defined* as linear algebras over a field F of characteristic p which have as degree some power of the characteristic. The theory of these algebras, which culminates in the theorem that every such algebra is "similar" to a cyclic algebra, depends essentially on the construction of inseparable fields contained in the algebra (in technical parlance, every p-algebra has a purely inseparable splitting field). To illustrate this, choose as the base field the field $P(t)$ of all rational functions of t with coefficients in a perfect field P. Introduce a pth root u, with $u^p = t$, and a a quantity v with $v^p = v + t$. The set of all sums

$$w = \sum_{i,j} a_{ij} u^i v^j, \qquad (i = 0, \cdots, p-1; j = 0, \cdots, p-1; a_{ij} \text{ in } F),$$

then forms a linear algebra of degree p over F, if one uses the multiplication table

$$u^p = t, \qquad v^p = v + t, \qquad vu = u(v+1).$$

The essential point for the theory is that this algebra contains both the inseparable extension $F(u)$ and the cyclic separable extension $F(v)$ of the base field F.

There are many further ways in which modular fields can arise in other algebraic investigations. We mention here only the use of fields of characteristic 2 in discussing Boolean algebras (Stone [24]), the theory of matrices over a modular field (Albert [5]), the definition of modular fields by special polynomials (Carlitz [7]), and the quasi-algebraic closure of finite fields (Chevalley [8]).

* *Cf.* Albert [2, ch. 7]; and also Jacobson [11], Teichmüller [25].

13. Summary. Modular fields include finite fields, Galois extensions of fields, algebraic function fields, and fields for algebraic manifolds, as well as for more bizarre types. The study of such fields is suggested by their origin in arithmetic questions about congruences, p-adic numbers, and ideal theory. On the other hand, an independent survey of their structure is indicated by the program of abstract algebra: first the development of the abstract concept ("field") in order to cover the variegated known examples, then the derivation of general theorems touching this concept, and lastly a classification of the types of systems which fall under the concept. We have seen that the straightforward generalization of the known properties of number fields is but one phase of our structure theory. There is also the investigation of characteristic new phenomena, of inseparability, of p-independence and the like, which distinguish the modular fields from the non-modular. The presence of curious examples of fields, which must at present still be given individual treatment, indicates that the present situation abounds in new questions, and that abstract algebra can very well give rise to concrete conundrums.

Bibliography

Albert, A. A., [1] Modern Higher Algebra, Chicago, 1937. [2] Structure of Algebras, American Mathematical Society Colloquium Publications, vol. XXIV, New York, 1939. [3] p-Algebras over a field generated by one indeterminate, Bulletin of the American Mathematical Society, vol. 43, 1937, pp. 733–736. [4] Quadratic null forms over a function field, Annals of Mathematics, vol. 39, 1938, pp. 494–505. [5] Symmetric and alternate matrices in an arbitrary field, I, Transactions of the American Mathematical Society, vol. 43, 1938, pp. 386–436.

Becker, M. F., and Mac Lane, Saunders, [6] The minimum number of generators for inseparable algebraic extensions, Bulletin of the American Mathematical Society, vol. 46, 1940, pp. 182–186.

Carlitz, L., [7] A class of polynomials, Duke Mathematical Journal, vol. 43, 1938, pp. 167–182.

Chevalley, C., [8] Demonstration d'une hypothèse de M. Artin, Abhandlungen aus dem mathematischen Seminar der Universität Hamburg, vol. 11, 1936, pp. 73–75.

Hasse, H., [9] Theorie der Kongruenzzetafunktionen, Sitzungsbericht der Preussischen Akademie der Wissenschaften, Berlin, 1934, pp. 250–255.

Hasse, H., and Schmidt, F. K., [10] Die Struktur diskret bewerteter Körper, Journal für die reine und angewandte Mathematik, vol. 170, 1934, pp. 4–63.

Jacobson, N., [11] p-Algebras of exponent p, Bulletin of the American Mathematical Society, vol. 43, 1937, pp. 667–670. [12] Abstract derivation and Lie algebras, Transactions of the American Mathematical Society, vol. 42, 1937, pp. 206–224.

Krull, W., [13] Beiträge zur Arithmetik kommutativer Integritätsbereiche VII, Inseparable Grundkörpererweiterung, Bemerkungen zur Körpertheorie, Mathematische Zeitschrift, vol. 45, 1939, pp. 319–334.

MacDuffee, C. C., [14] The p-adic numbers of Hensel, this MONTHLY, vol. 45, 1938, pp. 500–508.

Mac Lane, Saunders, [15] Steinitz field towers for modular fields, Transactions of the American Mathematical Society, vol. 46, 1939, pp. 23–45. [16] Modular fields *I*, Separating transcendence bases, Duke Mathematical Journal, vol. 5, 1939, pp. 372–393. [17] Notes on Higher Algebra, (planographed) Ann Arbor, 1939. With Schmidt, F. K., [18], Ueber inseparable Körper, forthcoming in Mathematische Zeitschrift. With Nilson, E. N., [19] Algebraic functions, (planographed) Ann Arbor, 1940.

Schilling, O. F. G., [20] Foundations of an abstract theory of Abelian functions, American

Journal of Mathematics, vol. 61, 1939, pp. 59–80. [21] Modern Aspects of the Theory of Algebraic Functions, Mimeographed, Chicago, 1938.

Schmidt, F. K., [22] Zur arithmetischen Theorie der algebraischen Funktionen II, Allgemeine Theorie der Weierstrasspunkte, Mathematische Zeitschrift, vol. 45, 1939, pp. 75–97.

Steinitz, E., [23] Algebraische Theorie der Körper, Journal für die reine und angewandte Mathematik, vol. 137, 1910, pp. 167–308; also edited by R. Baer and H. Hasse, Berlin, 1930.

Stone, M. H., [24] The theory of representations for Boolean algebras, Transactions of the American Mathematical Society, vol. 40, 1936, pp. 37–111.

Teichmüller, O., [25] p-Algebren, Deutsche Mathematik, vol. 1, 1936, pp. 362–388.

Weiss, Marie J., [26] Algebra for the undergraduate, this MONTHLY, vol. 46, 1939, pp. 635–642.

van der Waerden, B. L., [27] Moderne Algebra, vol. I, First edition, Berlin, 1930 (also second edition, 1938). [28] Zur algebraischen Geometrie XIV, Schnittpunktszahlen von algebraischen Mannigfaltigkeiten, Mathematische Annalen, vol. 115, 1938, pp. 619–644.

PROPER CONTINUED FRACTIONS

WALTER LEIGHTON, The Rice Institute

This paper generalizes the so-called "regular" continued fraction expansion of a real number. The treatment includes as a special case the "continued cotangent" expansion of Lehmer [2].

1. The expansion of a real number into a proper continued fraction. Let y_0 be any real number and $a_1, a_2, a_3, \cdots$ an arbitrary sequence of positive integers. If y_0 is an integer we shall say that its expansion into a *proper* continued fraction *terminates* and is given by

$$y_0 \sim b_0,$$

where $b_0 = y_0$. If y_0 is not an integer, let b_0 be the greatest integer $\leqq y_0$ (in symbols $b_0 = [y_0]$) and define real numbers $y_1, y_2, y_3, \cdots$ and positive integers $b_1, b_2, b_3, \cdots$ by the relations

$$y_n = \frac{a_n}{y_{n-1} - b_{n-1}}, \qquad b_n = [y_n], \tag{1.1}$$

successively for $n = 1, 2, 3, \cdots$. It is clear from (1.1) that each $b_n \geqq a_n$, $(n = 1, 2, 3, \cdots)$. If, eventually, some y_n, say y_k, is itself an integer, we shall say the expansion terminates, and that the proper expansion of y_0 into a continued fraction is given by

$$y_0 \sim b_0 + \frac{a_1}{b_1} \frac{a_2}{+ b_2} + \cdots + \frac{a_{k-1}}{b_{k-1}} \frac{a_k}{+ b_k}, \qquad (b_k = y_k). \tag{1.2}$$

We note that $b_k > a_k$. If no y_n is an integer, the expansion will not terminate and the proper continued fraction expansion of y_0 will be given by

$$y_0 \sim b_0 + \frac{a_1}{b_1} \frac{a_2}{+ b_2} + \cdots. \tag{1.3}$$