ACTA ARITHMETICA
XXXIX (1981)

Value sets of functions over finite fields

by

S. D. COHEN (Glasgow)

1. Introduction and general results. Let F be a finite field of order q and characteristic p. Where necessary adjoin ∞ to F as a possible value of a variable or function in the obvious way (see [2], § 4). For any rational function $f = f_1/f_2$ in $F(x)$, where f_1 and f_2 are co-prime polynomials, define $V(f)$ to be the set of values taken by f in F and $\deg f$, the degree of f, to be $\max(\deg f_1, \deg f_2)$.

Our chief object in this paper is to discuss the extent to which a function f of bounded degree is determined by $V(f)$. More precisely, we consider when $V(g) \subseteq V(f)$ can hold for two functions f and g. In fact, we solve the problem completely for functions f of degree not exceeding 4. For details of the results see § 2.

The remainder of this section is devoted to a summary of the various general results which bring together and extend work discussed by the author in [1] and [2] and by M. Fried in [6], [8] and [10] and which form the abstract background from which the specific functions of § 2 will emerge.

Accordingly, let $h(x, y)$ be a polynomial in x with coefficients in $F(y)$. We shall say that h is *x-soluble* (in F) if, for every y in F, $h(x, y) = 0$ is soluble with x in F. For the application to value sets we shall set $h(x, y) = f_1(x) - g(y)f_2(x)$, where $f = f_1/f_2$ and g are rational functions in $F(x)$. (We shall frequently abuse notation and write $f(x) - g(y)$ for this polynomial or even for the numerator of the rational function $f(x) - g(y)$.)

Returning to the case of a general h, which need not even be irreducible, we outline a proof of the following result.

PROPOSITION 1.1. *Let $h(x, y)$ be a separable polynomial of degree m in x with coefficients in $F(y)$ of degree $\leqslant n$. Let $h(x, y) = 0$ have roots $x_1, \ldots, x_m$ in a splitting field K over $F(y)$. Let $\bar{F}$ denote the algebraic closure of F in K and $G^*(K, F(y))$, etc., the subset of the galois group $G(K, F(y))$*

of K over $F(y)$ comprising automorphisms whose restrictions to $\bar{F}$ fix precisely F. If

$$G^*(K, F(y)) = \bigcup_{i=1}^{m} G^*(K, F(x_i, y)), \tag{1.1}$$

then h is x-soluble. Conversely, if $q > c(m, n)$ and h is x-soluble, then (1.1) *holds.*

Proof. We can assume that h is square-free. For brevity, put $G = G(K, F(y))$, $G^* = G^*(K, F(y))$, $G^*(i) = G^*(K, F(x_i, y))$, $i = 1, \ldots, m$, $G_1^* = \bigcup_{i=1}^{m} G^*(i)$. Also, for any y_0 in F, let $A(y_0)$ denote the conjugacy class in G which has the defining property of the Frobenius automorphism of some prime in K dividing $y-y_0$. Certainly $A(y_0)$ exists; it is uniquely defined if $y-y_0$ is unramified in each $F(x_i, y)$, i.e. if $h(x, y_0)$ is square-free. Then, in fact, $A(y_0) \subseteq G^*$ and indeed, if $y-y_0$ is unramified, then $A(y_0) \subseteq G_1^*$ if and only if $h(x, y_0) = 0$ is soluble in F (see Lemma 3 of [1] and [2]). Moreover, even if $y-y_0$ is ramified, then $h(x, y_0) = 0$ is soluble in F provided $A(y_0) \subseteq G_1^*$ ([2], p. 55, or [10], p. 223). Hence, (1.1) implies that h is x-soluble. Conversely, if h is x-soluble, then, by the function field analogue of the Čebotarev density theorem ([2], Lemma 2, [10], Proposition 2), for large q, $G^* = G_1^*$ and (1.1) holds.

Actually, in the last sentence of the above proof, it suffices to assume that h is x-soluble with, at most, kq^δ, say, exceptions where $0 \leqslant \delta < 1$ (see [2], p. 59). Consequently, the hypothesis of the second part can be weakened as in the following theorem.

THEOREM 1.2. *In the situation of Proposition* 1.1, *suppose that h is x-soluble except for at most kq^δ values of y in F, where $k > 0$ and $0 \leqslant \delta < 1$. If $q > c(m, n, \delta, k)$, then* (1.1) *holds and actually h is x-soluble in F. In particular, if f and g are functions of degree $\leqslant m, n$, respectively and $q > c$, then $|V(g) \setminus V(f)| < kq^\delta$ implies that $V(g) \subseteq V(f)$.*

The final assertion of Theorem 1.2 completely resolves a conjecture and a conditional result of Fried ([5], Conjecture 2, [8], Corollary 2). It should have been proved in [2] but was obscured there by our not taking $h(x, y) = f(x)-g(y)$; in fact, the discussion was equivalent to putting $h(x, y) = (f(x)-y)(g(x)-y)$ so that values y_0 of y for which $g(x)-y_0$ were not square-free had to be left out of the considerations.

In discussing possible occurences of (1.1), it may be convenient to separate the cases in which h is irreducible (in $F[x, y]$) or reducible, respectively. Alternative conditions for an irreducible h to be x-soluble are provided in the next result (cf. [2], Lemma 4 and Theorem 3, [8], Proposition 1).

PROPOSITION 1.3. *In the situation of Proposition* 1.1, *let h be irreducible in $F(x, y)$. Suppose that $q > c(m, n)$. Then the following are equivalent:*

(i) *h is x-soluble in F;*

(ii) *$h(x, y) = 0$ has a unique solution x in F for all y in F for which the discriminant of h (as a polynomial in x) is non-zero;*

(iii) *$h(x, y)$ is absolutely irreducible in $F(x, y)$ but has no absolutely irreducible factors except $(x-z)$ in $F(x, y, z)$, where $h(z, y) = 0$.*

Indeed, for any q, (iii) *implies* (i) *and* (ii).

Proof. We use the notation employed in proving Proposition 1.1. Note that the condition that $h(x, y)$ be absolutely irreducible in $F[x, y]$ is equivalent to the condition that $F(x_i, y) \cap \bar{F} = F$ for all $i = 1, \ldots, m$ (cf. (4.10) of [2]). Hence, Lemma 4 of [2] shows that (iii) is equivalent to each of (a) $G^* = G_1^*$, and (b) the $G^*(i)$ are pairwise disjoint. By Proposition 1.1, (i) and (iii) are equivalent as required, while it follows from Theorem 1.2 that (ii)$\Rightarrow$(i). Finally, suppose (i) holds. Then (a) and (b) are true. Hence every member of $A(y_0)$ belongs to precisely one $G^*(i)$. By [1], Lemma 5, if $h(x, y_0)$ is square-free, then $h(x, y_0) = 0$ has a unique solution in F. This completes the proof.

Note. It will follow from the examples of Theorem 2.1 (II) below that the exceptional y in (ii) may definitely give rise to multiple solutions of $h(x, y) = 0$ in F. (Thus some modification appears to be necessary in statement (2.12) of [8].)

Condition (1.1) for h to be x-soluble is, at first sight, a very restrictive one. Indeed, it implies that $G = G(K, F(y))$ is *admissible* in the following sense: G can be represented as a permutation group on $(1, \ldots, m)$ and is a cyclic extension of a normal subgroup $\hat{G}$; moreover, if G^* is the subset of G every member of which generates $G/\hat{G}$, then $G^* = \bigcup_{i=1}^{m} G^*(i)$, where $G^*(i)$ denotes the stabilizer of i in G^*. Indeed, for h to be irreducible (and so absolutely irreducible, by Proposition 1.3 (iii)), an admissible G has additionally to be transitive, and, since $F \neq \bar{F}$, the cyclic extension $G/\hat{G}$ must be non-trivial. (In the irreducible case, Fried, [11], p. 153, has given a description of an admissible G corresponding to Proposition 1.3.)

Accordingly, in order to find all x-soluble h of given degree m in x, it is first necessary to find all admissible G contained in the symmetric group S_m. This is straightforward for $m \leqslant 4$; there are two non-trivial possibilities with G transitive and one with G intransitive; in effect, these are dealt with in §§ 5–7 below. More generally, the known examples of permutation polynomials, namely cyclic and Chebychev polynomials (see [9]), indicate that $\hat{G}$ may be a cyclic or metacyclic group. But there are other possibilities even when G is transitive. Indeed, $\hat{G}$ need not even be soluble, as shown by the following example pointed out to the author by J. Saxl, in which $m = 28$. Take $G = \mathrm{P\Gamma L}(2, 8)$, $\hat{G} = \mathrm{PGL}(2, 8)$ so that $|G| = 1512$, $\hat{G}$ is simple and $G/\hat{G}$ is cyclic of order 3. As for the intransitive (reducible) case, Fried ([10], pp. 211, 227) has announced examples (with $h(x, y) = f(x) - g(y)$, f, g *polynomials*) which imply the existence of admissible G with $G/\hat{G}$ trivial (so that $\bar{F} = F$).

Having found an admissible G, we would next like to find all h (if any) for which $G = G(K, F(y))$ (in the obvious correspondence). In the irreducible case, this includes what Fried [11] has called the "general Schur problem" and is very difficult. For general h we give a solution only in the case that the total degree of h (in x and y) does not exceed 3 (§ 8). However, if $h(x, y)$ is of the form $f(x) - g(y)$, we can invoke properties of the discriminant and, in this way, obtain a complete solution provided $\deg f \leqslant 4$. These are the results listed in § 2 and proved in §§ 3–7.

Finally, in § 9, we shall consider some non-trivial examples of sets of functions $\{f_i\}$ which cover F, i.e. for which $\bigcup V(f_i) = F$.

2. Results on value sets. We describe here our main results on the existence of pairs of function $f(x)$, $g(x)$ in $F(x)$ for which $V(g) \subseteq V(f)$.

Define a permutation function P over F to be one for which $V(P) = F$. Then, trivially, $V(g) \subseteq V(P)$ for any function g. Now obviously a non-singular, linear fractional transformation L in $F(x)$ is a permutation function. However, there are others, e.g. the monomials x^n provided $(n, q-1) = 1$ and the Chebychev polynomials T_n for certain values of n, see [9]. These can be included in a more general class of functions of the form $f = \hat{f}(Q)$ for which $V(f) = V(\hat{f})$. The main result, which follows, shows that, in addition to such functions, there are some interesting pairs of functions (f, g) with $\deg f \leqslant 4$ for which $V(g) \subseteq V(f)$. In its statement and throughout we use the following notation. L denotes a non-singular linear fractional transformation; P is a permutation function; λ is an arbitrary non-square in F; F^i denotes the field of order q^i.

THEOREM 2.1. *Let f, g be rational functions in $F(x)$. Then $V(g) \subseteq V(f)$ if either* (I) *or* (II) *below holds.*

(I). *$f = \hat{f}(Q)$, $g = \hat{f}(R)$ for some $\hat{f}$, Q, R in $F(x)$ with*

$$V(\hat{f}(Q)) = V(\hat{f}). \tag{2.1}$$

In particular, (2.1) *is satisfied whenever Q is a permutation function and $\hat{f}$ is any function.*

(II). *p $(= \operatorname{char} F) > 3$ and $f = L\circ f^*\circ P$, $g = L\circ g^*\circ R$, where L, P and R are in $F(x)$ and f^* and g^* are one of the following pairs:*

(i) $f^*(x) = x^3 - 3x + 2$, $g^*(x) = 4/(3\lambda x^2 + 1)$;

(ii) $f^*(x) = x^3 - 3\lambda x$,

$$g^*(x) = 2\lambda[\alpha(x^2+\lambda) + 2\beta\lambda x]/[\beta(x^2+\lambda) + 2\alpha x], \tag{2.2}$$

where (α, β) is a chosen pair in $F \times F$ for which $(-3\lambda)(\alpha^2 - \lambda\beta^2)$ is a non-zero square in F;

(iii) $q \equiv 1 \pmod 3$ *and*

$$f^*(x) = (x^4 + 4x^3)/(8x - 4), \qquad g^*(x) = \mu x^3$$

where μ is any non-cube in F;

(iv) $f^*(x) = x^4 + 4x^3$,

$$g^*(x) = \begin{cases} 108\mu x^3/(\mu x^3 - 1)^2, & \text{if } q \equiv 1 \pmod 3, \\ 108(x^2+3)^3/[\nu(x+\sqrt{-3})^3 - \nu^{-1}(x-\sqrt{-3})^3]^2, & \text{if } q \equiv -1 \pmod 3, \end{cases}$$

where μ is any non-cube in F and ν is any non-cube in F^2 whose conjugate over F is $\pm\nu^{-1}$;

(v) $a \neq 0, \frac{1}{4}, 1$ *and* $f^*(x) = [(x^2 + 3a - 3)^2/4(2x+3)] + 3a - 1$,

$$g^*(x) = \begin{cases} (\mu^2 x^6 + a^3)/\mu x^3, & \text{if } q \equiv 1 \pmod 3, \\ a^{3/2}[\nu(x+\sqrt{-3})^6 + \nu^{-1}(x-\sqrt{-3})^6]/(x^2+3)^3, & \text{if } q \equiv -1 \pmod 3, \end{cases}$$

where μ is any non-cube in F and ν is a non-cube in F^2 whose conjugate over F is ν^{-1} or $-\nu^{-1}$ according as a is, or is not, a square in F, respectively.

Conversely, suppose that $\deg f \leqslant 4$, $\deg g \leqslant n$ and that $q > c(n)$ with $p > 3$. Then $V(g) \subseteq V(f)$ implies that either (I) *or* (II) *holds.*

Remarks. (a) That (I) implies $V(g) \subseteq V(f)$ is obvious. The sufficiency of (II) will emerge during the demonstration of the converse which, of course, is the harder task. (Note that, if $\deg f \leqslant 4$, then, in (II), we must have $P = L_1$.) Actually, the case in which f and g are cubic polynomials was partially considered by McCann and Williams [14] who showed that, if $q = p > 7$, then $V(f) = V(g)$ implied that $g = f(L)$ or $f = P$.

(b) For some values of q we can explicitly simplify the form of the function g^* in (ii). For if -3 or -1 is a square in F we may choose $(\alpha, \beta) = (1, 0)$ or $(0, 1)$, respectively. Thus we may take for (2.2)

$$g^*(x) = \begin{cases} \lambda(x^2+\lambda)/x, & \text{if } q \equiv -1 \pmod 3, \\ 4\lambda^2 x/(x^2+\lambda), & \text{if } q \equiv 1 \pmod{12}. \end{cases}$$

(c) The g^* of (iv) and (v) are in $F(x)$ despite the fact that, if $q \equiv -1 \pmod 3$, then $\sqrt{-3}$ and ν lie in $F^2 \setminus F$. In (iv), for example, ν could also be described as one of the $\frac{2}{3}(q+1)$ non-cubes in F^2 which are $2(q+1)$-th roots of unity in F^2.

(d) Actually, in the excluded cases $a = 0, 1$, (v) remains valid but (after suitable linear transformations) reduces to (iii) and (iv), respectively.

(e) As regards (iii), since μ can be any non-cube in F, we also have $V(\mu g^*) \subseteq V(f^*)$. Indeed, we have

$$V(f^*) \supseteq V(\mu x^3) \cup V(\mu^2 x^3). \tag{2.3}$$

In particular $|V(f^*)| \doteqdot 3q/4$, $|V(g^*)| = (q+2)/3$.

(f) In case (II), the containments $V(g) \subseteq V(f)$ are all proper. This is apparent from (2.3) in case (iii). Otherwise, in cases (i) and (ii) we have, approximately, $|V(f)| = 2q/3$, $|V(g^*)| = \frac{1}{2}q$, while in cases (iv), (v), $|V(f^*)| = 5q/8$, $|V(g^*)| = \frac{1}{3}q$.

(g) If the degree of f is allowed to exceed 4, it remains to describe what other exceptional cases require to be added to (II). Certainly, if $f(x)-g(y)$ is reducible then, as mentioned in § 1, Fried [10] has asserted that there are algebraic number fields K and *polynomials* f and g, defined over K and not linearly related such that $V(f \pmod{\mathfrak{p}}) = V(g \pmod{\mathfrak{p}})$ for almost all prime ideals $\mathfrak{p}$ of K. On the other hand, if $f(x)-g(y)$ is irreducible, then, although there are additional admissible possibilities (in the sense of § 1) for the galois group of $h(x, y)$, these may never be realised by h of the form $f(x)-g(y)$.

Of course an explicit classification of all functions satisfying (I) is desirable. We provide such for $\deg f \leqslant 4$. First we describe the permutation functions. We show that the only non-trivial ones are of degree 3. (Of course, the non-existence of permutation *polynomials* of degrees 2 and 4 is well known.) In particular, we show that there is a class of permutation functions of degree 3 which includes no polynomials.

THEOREM 2.2. *Let f be a permutation function of degree $\leqslant 4$. Suppose that $q > c$ (absolute) and $p > 3$. Then $f = L$ or $f = L_1 \circ f^* \circ L_2$, where*

$$f^*(x) = \begin{cases} x^3, & \text{if} \quad q \equiv -1 \pmod 3, \\ (x^3+3\lambda x)/(3x^2+\lambda), & \text{if} \quad q \equiv 1 \pmod 3. \end{cases}$$

Next, we show that (I) may hold with $Q \neq P$ even when $H(f) \leqslant 4$. It is enough to suppose that $g = \hat{f}$ so that $V(f) = V(g)$.

THEOREM 2.3. *Suppose that $\deg f \leqslant 4$ and $\deg g \leqslant n$. If $q > c(n)$ and $p > 3$, then $V(f) = V(g)$ if and only if $g = f(P)$ or $f = L \circ f^* \circ L_1$, $g = L \circ g^* \circ P$, where f^* and g^* are one of the following pairs:*

(i) $f^*(x) = x^4$, $g^*(x) = x^2$ *and* $q \equiv -1 \pmod 4$;
(ii) $f^*(x) = (x^4+\lambda)/x^2$, $g^*(x) = (x^2+\lambda)/x$;
(iii) $f^*(x) = (x^2+\lambda)^2/2(x^3-\lambda x)$, $g^*(x) = (x^2+\lambda)/x$.

3. Auxiliary results. When h has degree $\leqslant 4$ (in x) some of the results of § 1 can be rephrased in a manner involving its discriminant. In fact, if $h(x, y) = f(x)-g(y)$, we shall find that, by considering the shape of the discriminant, the functions f and g can be normalized, thereby greatly simplifying the argument.

Accordingly, let $h(x, y)$ be a square-free polynomial of degree m ($\geqslant 2$) with coefficients in $F(y)$ and zeros $x_1, \ldots, x_m$ in a splitting field. Let $D_h(y)$ denote the discriminant $a^{2m-2} \prod_{i \neq j} (x_i - x_j)$ of h, where $a = a(y)$ is its leading coefficient. Further, for any f in $F(x)$, we shall also, without fear of ambiguity, use $D_f(y)$ to denote the polynomial $D_{f_1(y)-f_2(y)}(y)$ (in $F[y]$), where $f(x) = f_1(x)/f_2(x)$ and f_1 and f_2 are co-prime polynomials with f_1 monic. We summarise some relevant properties of D_f which are due essentially to the fact that the extension $F(x, y)$ of $F(y)$, where $f(x) = y$, has genus 0. They are actually valid for any field F whose characteristic $> m$. In our case, assume $p > m$.

In the first place, $\deg D_f \leqslant 2m-2$. Put $r_\infty = 2m-2-\deg D_f$. Suppose that D_f has prime decomposition $a \prod_{i=1}^{s} \mathscr{P}_i^{r_i}$ in $F[y]$ where $a\ (\neq 0) \in F$ and the $\mathscr{P}_i$ are monic irreducibles. Formally adjoin a *linear* polynomial denoted (temporarily) by $\mathscr{P}_\infty$ which vanishes at ∞ and put $\mathscr{D}_f = \left(\prod_{i=1}^{s} \mathscr{P}_i^{r_i}\right) \mathscr{P}_\infty^{r_\infty}$. Refer to the set of ordered pairs of the form $(\deg \mathscr{P}, r)$ (with multiplicities) as included in the *ramification data* of f over F. Its significance is as follows. Let γ be any root of $\mathscr{P}_i(y) = 0$ in $\bar{F}$, the algebraic closure of F. Let the zeros of $f_1 - \gamma f_2$ in $\bar{F}$ have multiplicities $e_1, e_2, \ldots$, with the convention that, if $e_\infty = m - \deg f_2$ is non-zero, then e_∞ is included. Then, of course, $\sum e_j = m$, but in fact, we also have $\sum (e_j - 1) = r_i$. The collections $E(\mathscr{P}_i) = \{e_1, e_2, \ldots\}$ complete the ramification data of f. Note that $|E(\mathscr{P}_i)| = m - r_i$. Since $F(\gamma) = F(L(\gamma))$ for any L in $F(x)$ and γ in $\bar{F}$ (adjoining ∞ to F, if necessary), it is clear from the above interpretation of the ramification data, that it is preserved under compositions of the form $L_1 \circ f \circ L_2$ with L_1, L_2 in $F(x)$. Further, if the pair $(1, r)$ is included in the ramification data, then by replacing f by $L(f)$ for appropriate L, we can assume that $\mathscr{P}_\infty^r$ appears in $\mathscr{D}_f$, so that f_2 has prime decomposition of the form $f_2 = \beta P_1^{e_1} P_2^{e_2} \ldots$ $(e_j > 0)$ where $\sum (e_j - 1) \leqslant r$. In this situation, if $\deg P_1 = 1$, we can replace f by $f(L)$ and assume that $\deg f_2 = m - e_1$.

To complete the preliminaries we state a vital lemma, which follows immediately from a more general result of the author [3].

LEMMA 3.1. *Suppose that $r = 2$ or 3 and $p > 3$. Let A and B be rational functions in $F(x)$ with A not an r-th power in $F(x)$. Suppose that $A(B)$ is an r-th power in $F(x)$. If $r = 2$, then $A = QA_1^2$, where $A_1 \in F(x)$ and Q is a polynomial of degree $\leqslant 2$ in $F[x]$. If $r = 3$, then $A = LA_1^3$, where L, $A_1 \in F(x)$.*

An explicit description of those A, B for which $A(B)$ is an rth power (for any r) is given in [3].

4. The quadratic case. If $\deg f = 1$ then, of course, the results of § 2 are trivial. The case $\deg f = 2$ is disposed of in the following theorem.

THEOREM 4.1. *Let $h(x, y)$ in $F(x, y)$ have degree 2 in x and degree n in y. Suppose $q > c(n)$ and $p > 2$. Then h is x-soluble in F if and only if it is reducible in $F(x, y)$. In particular, suppose f and g are functions in $F(x)$ with $\deg f = 2$, $\deg g = n$. Then the following are equivalent:*

(i) $V(g) \subseteq V(f)$;

(ii) $g = f(R)$ *for some R in $F(x)$.*

If also $\deg g = 2$, then the following are each equivalent to (i) *or* (ii).

(iii) $V(g) = V(f)$;

(iv) $g = f(L)$;

(v) $D_g(y) = D_f(y)v^2(y)$, *where* $v(y) \in F(y)$.

Proof. Condition (iii) of Proposition 1.3 can never hold if $m = 2$ and the first part is clear. If $h(x, y) = f(x) - g(y)$, then reducibility of h is equivalent to (ii) so that (i) and (ii) are equivalent. Finally, suppose that $\deg g = 2$. The following implications are obvious: (ii)$\Leftrightarrow$(iv)$\Rightarrow$(iii) $\Rightarrow$(i)$\Rightarrow$(ii). Hence (i)–(iv) are equivalent. Moreover, (iv)$\Rightarrow$(v) while (v)$\Rightarrow$(iii) is an easy property of the discriminant.

5. Functions of degree 3. In the cubic case we use Proposition 1.3 in the following form.

PROPOSITION 5.1. *Suppose, in Proposition 1.3, that h is a cubic in x and $p > 3$. Then the following can be added to the list of equivalent conditions* (i)–(iii):

(5.1) (iv) $D_h(y) = \lambda v^2(y)$, *where* $v(y) \in F(y)$.

Proof. For a given y in F, $h(x, y)$ has a unique zero (of multiplicity 1) in F if and only if $D_h(y)$ is a non-square in F. Thus (iv)$\Rightarrow$(ii) while (ii)$\Rightarrow$(iv) (for large q) follows from a result of Perel'muter [18].

We now take $h(x, y) = f(x) - g(y)$ and proceed to prove the results of § 1 with $\deg f = 3$. Trivially, in this case, (I) of Theorem 2.1 occurs if and only if $Q = P$. We can assume h irreducible.

Suppose therefore that $V(g) \subseteq V(f)$. This property clearly survives the operation of replacing f and g by $L \circ f \circ L_1$ and $L \circ g \circ L_2$, respectively. By Proposition 5.1, $\lambda D_f(g(y))$ is a square in $F(y)$. Hence, by Lemma 3.1, $\lambda D_f(y)$ is a square apart from a factor of degree at most 2. If $\deg(D_f(y)) > 2$, then $D_f(y)$ must have a square factor, while if $\deg(D(y)) = 2$, then certainly $\mathscr{L}_\infty^2$ divides $\mathscr{D}_f$, where now $\mathscr{L}_\infty$ denotes the infinite linear factor $\mathscr{P}_\infty$ of § 3. Hence, in either case, $\mathscr{D}_f$ has a square factor. We use $\mathscr{L}$, $\mathscr{L}_1$, $\mathscr{L}_2$ to denote distinct linear polynomials (possibly $\mathscr{L}_\infty$) and $\mathscr{Q}$ to denote an irreducible quadratic polynomial in $F[y]$ and consider the three possibilities for $\mathscr{D}_f$.

(a) $\mathscr{D}_f = \mathscr{L}_1^2\mathscr{L}_2^2$. As in § 3, we may assume that, in fact, $\mathscr{L}_2 = \mathscr{L}_\infty$ and that f is a polynomial. Indeed, by linear transformations we may take $\mathscr{D}_f(y) = y^2$, whence $L \circ f \circ L_1 = x^3$. So assume $f(x) = x^3$. Then $D_f(y) = -27y^2$ and hence -3 is a non-square in F (i.e. $q \equiv -1 \pmod 3$) and f is a permutation polynomial.

(b) $\mathscr{D}_f = \mathscr{Q}^2$. In this case we may assume that $\mathscr{Q}(y) = y^2 - \lambda$. It follows that, if $f = f_1/f_2$, then

$$f_1(x) + \sqrt{\lambda} f_2(x) = (a_1 + \sqrt{\lambda} a_2)\big(v_1(x) + \sqrt{\lambda} v_2(x)\big)^3,$$

where $a_1, a_2 \in F$ and v_1 and v_2 are linear or constant polynomials in $F[x]$. Thus, replacing f by $L \circ f \circ L_1$, where $L_1^{-1}(x) = (a_1 x + \lambda a_2)/(a_2 x + a_1)$, $L_1^{-1} = v_1/v_2$, we obtain

$$f_1(x) + \sqrt{\lambda} f_2(x) = (x + \sqrt{\lambda})^3,$$

whence $f(x) = (x^3 + 3\lambda x)/(3x^2 + \lambda)$. Accordingly, $D_f(y) = -108\lambda(y^2 - \lambda)^2$, -3 is a square in F and $q \equiv 1 \pmod 3$. Moreover, by Proposition 4.1, this f is a permutation function.

(c) $\mathscr{D}_f = \mathscr{L}^2\mathscr{L}_1\mathscr{L}_2$ or $\mathscr{L}^2\mathscr{Q}$. As before we may assume that $\mathscr{L} = \mathscr{L}_\infty$ and indeed that f is a polynomial. In fact, by a linear transformation of x, we may take $f(x) = x^3 - 3\eta x$, where $\eta = 1$ or λ. Put $g(y) = 2\eta u$. Then $D_f(2\eta u) = -108\eta^2(u^2 - \eta)$. Hence $(-3\lambda)(u^2 - \eta)$ is a square in $F(y)$ and

$$V(g) \subseteq S := \{2\eta a\colon (-3\lambda)(a^2 - \eta) \text{ is a square in } F\} \subseteq V(f). \tag{5.2}$$

Now, for the moment suppose $\eta = 1$ and put $g_0(y) = 4(3\lambda y^2 + 1)^{-1} - 2$. Then $D_{g_0}(2y) = (-12\lambda)(y-1)/(y+1)$ and evidently $V(g_0) = S$. Hence

$$V(g) \subseteq V(f) \Leftrightarrow V(g) \subseteq V(g_0) \Leftrightarrow g = g_0(R),$$

for some R in $F(x)$, by Theorem 4.1 (ii). The necessity and sufficiency of (i) of Theorem 2.1 (II) follows.

Next suppose that $\eta = \lambda$ and that a and β in F are such that $(-3\lambda)(a^2 - \lambda\beta^2)$ is a non-zero square in F. Put

$$g_0(y) = 2\lambda[a(y^2 + \lambda) + 2\beta\lambda y]/[\beta(y^2 + \lambda) + 2ay]. \tag{5.3}$$

Then $D_{g_0}(y)(2\lambda y)/(-3\lambda)(y^2 - \lambda)$ is a square in $F(y)$. By comparing this with (5.2) and using the argument of the $\eta = 1$ case, we see that $V(g) \subseteq V(f) \Leftrightarrow g = g_0(R)$. To complete the proof, it remains to show that, if g^* is also given by (5.3) with another pair (a, β), then $g^* = g_0(L)$. By Theorem 4.1 (iii)–(v), this is so.

6. Functions of degree 4, the irreducible case. We suppose now that $h(x, y) = \sum_{i=0}^{4} h_i x^i$ ($h_4 \neq 0$) is a quartic polynomial in x with coefficients in $F(y)$. Its classical cubic resolvent, namely

$$x^3 - h_2' x^2 + (h_3' h_1' - 4h_0')x - h_3'^2 h_0' + 4h_2' h_0' - h_1'^2,$$

where $h_i' = h_i/h_4$, will be denoted by $\mathscr{R}_h(x, y)$. In the first instance we suppose that h is irreducible; thus, for example, Proposition 1.3 is applicable. Recall that F^i denotes the field of order q^i.

Proposition 6.1. *In the situation of Proposition* 1.1, *suppose that h has degree* 4 *in x and is irreducible over $F(x, y)$ and that $p > 3$. Then h is x-soluble in F if and (when $q > c(n)$) only if $\mathscr{R}_h(x, y)$ is irreducible in $F(x, y)$ but reducible in $F^3(x, y)$. In particular, if h is x-soluble in F and $q > c(n)$, then*

(i) *h is x-soluble in F^2 and*

(ii) *$D_h(y) = h_4^6 D_{\mathscr{R}_h}(y)$ is a square in $F(y)$.*

Proof. We use the notation of Proposition 1.1. Suppose h is x-soluble in F. Since h is irreducible, then $F \neq \bar{F}$ and $|G(K, F(y))|$ is divisible by 4. Indeed, by Proposition 1.3 (iii), $|G(K, F(y))|$ is also divisible by 3. In fact, since $G(K, F(y))$ is a cyclic extension of $G(K, \bar{F}(y))$, we must have $G(K, F(y)) = A_4$ and $G(K, \bar{F}(y)) = V := \{1, (12)(34), (13)(24), (14)(23)\}$. Thus $\bar{F} = F^3$. Accordingly (ii) holds and (i) follows from Proposition 1.3 (iii). Moreover, by [13], Theorem 43, $\mathscr{R}_h$ is irreducible in $F(x, y)$ but reducible in $F^3(x, y)$. Conversely, if this last fact holds then $G(K, F(y)) = A_4$ and $G(K, F^3(y)) = V$. But $A - V$ comprises only 3-cycles and so (2.1) is satisfied and consequently h is x-soluble. This completes the proof.

From Proposition 6.1, the x-solubility of h depends on the reducibility of the cubic $\mathscr{R}_h$. Accordingly, we need a result which follows easily from "Cardan's formulas" for the solution of a cubic equation (see [12], p. 258).

Lemma 6.2. *Let $\mathscr{R}(x) = x^3 + ax + \beta$ where a, β (not both zero) belong to a field Ω of characteristic > 3. Suppose that the discriminant D of $\mathscr{R}$ is a square in Ω. Then $\mathscr{R}(x) = 0$ has one (and so all) solutions in Ω if and only if for ome choice of the square root $\sqrt{D}$,*

$$\theta = \{-\tfrac{1}{2}(\beta + \sqrt{(D/-27)})\}^{1/3} \tag{6.1}$$

belongs to $\Omega(\sqrt{-3}) \setminus \{0\}$.

We now specialise to the case $h(x, y) = f(x) - g(y)$, where f has degree 4, but still assume f irreducible. As far as the results of § 2 are concerned, we now show that $V(g) \subseteq V(f)$ if and only if f and g are given by one of (iii)–(v) of Theorem 2.1 (II).

Assume then that $V(g) \subseteq V(f)$. By Proposition 6.1, $D_f(g(y))$ is a square in $F(y)$ and so, by Lemma 3.1, $D_f(y)$ (which has degree $\leqslant 6$) is also a square apart from a factor of degree at most 2. Indeed, a quick survey of the various possibilities reveals that actually $\mathscr{D}_f$ is a square apart from a factor of degree at most 2. We consider the various possibilities according to the factorisation of $\mathscr{D}_f$, using $\mathscr{L}, \mathscr{L}_1, \ldots, \mathscr{Q}, \mathscr{Q}_1, \ldots, \mathscr{C}$, for distinct linear, quadratic and cubic irreducibles, respectively. As in § 5, we pass freely from f and g to equivalent pairs $L \circ f \circ L_1$ and $L \circ g \circ L_2$. Some preliminary observations are helpful. First, if $\mathscr{P}$ divides $\mathscr{D}_f$ with, in the notation of § 3, $r = r_{\mathscr{P}} = 2$, then $E(\mathscr{P})$ can be either $\{1, 3\}$ or $\{2, 2\}$. Again if $\mathscr{L}_1$ and $\mathscr{L}_2$ both divide $\mathscr{D}_f$ and $E(\mathscr{L}_1)$ and $E(\mathscr{L}_2)$ are either $\{2, 2\}$ or $\{4\}$ (although possibly unequal), then replacing f by $L \circ f \circ L_1$ as appropriate, we get $f = af_0^2$, where $H(f_0) = 2$, so that for any g, certainly $|G(K, F(y))| \leqslant 8$ which is impossible, granted that we are discussing only the irreducible case meantime.

(a) [1] *$\mathscr{L}^3$ divides $\mathscr{D}_f$.* Using the ideas of § 3 (as already employed in § 5), we may assume that $\mathscr{L} = \mathscr{L}_\infty$ and, indeed, that f is a polynomial with $\mathscr{D}_f = \mathscr{L}_1^2 \mathscr{L}_2 \mathscr{L}_\infty^3$. Applying a linear transformation to x and multiplying by a suitable constant, we may assume even that $f(x) = x^4 + 4x^3$. Put $u = g(y)$. The cubic resolvent of $f(x) - u$ is

$$\mathscr{R}(x) = x^3 + 4ux + 16u.$$

Moreover, since $D_f(u) = D_{\mathscr{R}}(u) = -256u^2(u+27)$ is a square in $F(y)$, then $u = -R^2 - 27$ where $R(y) \in F(y)$. By Lemma 6.2, $\mathscr{R}(x)$ is reducible in $F^3(x, y)$ but not in $F(x, y)$ if and only if

$$(2/\sqrt{-3})\{(R + \sqrt{-27})(R^2 + 27)\}^{1/3} \quad \text{belongs to} \quad F^3(\sqrt{-3}, y) \setminus F(\sqrt{-3}, y). \tag{6.2}$$

Suppose, for the moment that $q \equiv 1 \pmod 3$ so that $\sqrt{-3} \in F$. Then evidently (6.2) holds if and only if

$$(R - \sqrt{-27})/(R + \sqrt{-27}) = \mu S^3, \tag{6.3}$$

where μ is a non-cube in F and $S \in F(y)$. However (6.3) is equivalent to

$$R = \sqrt{-27}(\mu S^3 + 1)/(\mu S^3 - 1). \tag{6.4}$$

Hence $g(y) = 108\mu S^3/(\mu S^3 - 1)^2 = g^*(S(y))$, where g^* is defined in (iv) of Theorem 2.1 (II) (with $q \equiv 1 \pmod 3$).

Alternatively suppose that $q \equiv -1 \pmod 3$ so that $F(\sqrt{-3}) = F^2$. We require (6.3) to hold with μ a non-cube in F^2 and S in $F^2(y)$ but R (given by (6.4)) in $F(y)$. This occurs if and only if the product of μS^3

[1] Actually, case (a) can be treated along with case (c).

and $\overline{\mu S^3}$, its conjugate over $F(y)$, is 1. In fact, we must have $\mu S^3 = \delta(T+\sqrt{-3})^3/(T-\sqrt{-3})^3$, where δ is a non-cube in F^2 such that $\delta\bar{\delta} = \delta^{q+1} = 1$ and $T \in F(y)$. This leads to

$$\begin{aligned} g(y) &= 108\delta(T^2+3)^3/\big(\delta(T+\sqrt{-3})^3-(T-\sqrt{-3})^3\big)^2 \\ &= 108(T^2+3)^3/\big(\nu(T+\sqrt{-3})^3-\nu^{-1}(T-\sqrt{-3})^3\big)^2, \end{aligned}$$

where $\nu^2 = \delta$ and ν is as described in (iv) of Theorem 2.1 (II); in particular $\nu \in F^2$ since $\delta^{q+1} = 1$. Hence $g(y) = g^*\big(T(y)\big)$, as required. Since the steps are reversible, this completes the proof in this case.

(b) *$\mathscr{L}_1$ and $\mathscr{L}_2$ divide $\mathscr{D}_f$ with $E(\mathscr{L}_1) = E(\mathscr{L}_2) = \{1, 3\}$.* In the usual way take $\mathscr{L}_2 = \mathscr{L}_\infty$ and $\deg f_2 = 1$. A linear transformation in x and multiplication by a constant enable us to concentrate our attention on the function $f(x) = (x^4+4x^3)/(4x+a)$, where $a\,(\in F) \neq 0, 16$ (otherwise f is not in its lowest terms). Put $g(y) = u$ and let $\mathscr{R}(x)$ be the resolvent cubic of $f(x)-u$. We have

$$\mathscr{R}(x) = x^3+4(a-4)ux+16u(a-u) \tag{6.5}$$

and

$$D_f(u) = D_{\mathscr{R}}(u) = 256u^2[-27(a-u)^2-u(a-4)^3] = 256u^2Q(u),$$

say. Moreover, taking $\mathscr{R}$ as the polynomial (6.5) in Lemma 6.2, we have

$$\theta = -2\{u[(a-u)+\sqrt{(Q(u)/-27)}]\}^{1/3}. \tag{6.6}$$

Now $Q(u)$ has a repeated factor (in $F(u)$) if and only if $a = -2$, 4 or 16. However, $a = 16$ has been excluded. Moreover, if $a = 4$ and $\sqrt{Q(u)}$ is taken to be $\sqrt{-27}(4-u)$, then, in fact, $\theta = 2[2u(u-4)]^{1/3}$. But Lemma 3.1 with $r = 3$ implies that $2u(u-4)$ can never be a cube in $F^3(y)$ for any g and so, by Lemma 6.2 and Proposition 6.1, we cannot have $V(g) \subseteq V(f)$. Next, putting $a = -2$, we obtain $Q(u) = -27(u-2)^2$. Since $D_f(u)$ is a square in $F(u)$, we must then have $\sqrt{-3} \in F$, i.e. $q \equiv 1 \pmod 3$. Further, taking $\sqrt{Q(u)} = \sqrt{-27}(u-2)$ in (6.6), we require $(-32u)^{1/3}$ to be in $F^3(y)$ but not in $F(y)$. Clearly, this is the case if and only if $u = 2\mu R^3$ for some R in $F(y)$ and non-cube μ in F, i.e. if and only if $u = 2g^*(R)$, where g^* is given by (iii) of Theorem 2.1 (II). Conversely, for f^* and g^* as given there, the above argument shows that $V(g^*) \subseteq V(f^*)$ for any q (with $p > 3$) and that actually, $\mathscr{D}_{f^*} = \mathscr{L}_1^2\mathscr{L}_2^2\mathscr{L}_3^2$, where $E(\mathscr{L}_1) = E(\mathscr{L}_2) = \{1, 3\}$ and $E(\mathscr{L}_3) = \{2, 2\}$ (since $x^4+4x^3-8x+4 = (x^2+2x-2)^2$).

To conclude this case, it suffices to show that $Q(u)$ cannot be square-free. For suppose $Q(u) = -27(u-a)(u-b)$, where $a, b \in F^2$ with $a \neq b$. Then $(u-a)/(u-b) = v^2$, where $v \in F(y)$. Thus $u = (bv^2-a)/(v^2-1)$ and

we may take $\sqrt{(Q(u)/-27)} = (b-a)v/(v^2-1)$. From (6.6) it follows that

$$\theta = -2\big\{\big[(bv^2-a)\big((a-b)v+a-a\big)\big]/(v-1)(v+1)^2\big\}^{1/3} \tag{6.7}$$

belongs to $F^6(y)$. If $2a \neq a+b$, the rational function in braces in (6.7) is in its lowest terms and so has no cube root in $F^6(y)$ for any $v(y)$ by Lemma 3.1. Indeed, even if $2a = a+b$, then $a \neq b$ and we would require $(bv^2-a)/(v+1)^2$ to have a cube root in $F^6(y)$ which again contradicts Lemma 3.1 since $a \neq b$. Hence $Q(u)$ is not square-free.

(c) *$\mathscr{L}_1$, $\mathscr{L}_2$ divide $\mathscr{D}_f$ with $E(\mathscr{L}_1) = \{2, 2\}$, $E(\mathscr{L}_2) = \{1, 3\}$.* Taking $\mathscr{L}_2 = \mathscr{L}_\infty$ and proceeding with the usual normalization process we may assume that

$$f(x) = (x^2+3\alpha-3)^2/4(2x+3),$$

where $\alpha\,(\in F) \neq \frac{1}{4}$ (otherwise f is not in its lowest terms). Put $u = g(y)$. When x is replaced by $x+2u-2$ in the cubic resolvent of $f(x)-u$, we obtain

$$\mathscr{S}(x) = x^3+48\big(u-(\alpha-1)^2\big)x-64\big(u^2+3(\alpha-1)u-2(\alpha-1)^3\big).$$

Thus

$$D_f(u) = D_{\mathscr{S}}(u) = -27.2^{12}u^2[(u+3\alpha-1)^2-4\alpha^3] = -27.2^{12}u^2Q(u),$$

say. Now, if $\alpha = 0$, we have case (b) again. So assume that $\alpha \neq 0$, thus $Q(u)$ is square-free. Put $u = R^{-1}(R^2+\alpha^3)-(3\alpha-1)$. Then $Q(u) = (R^2-\alpha^3)^2/R^2$. If θ is given by (6.1) with $\mathscr{R} = \mathscr{S}$ and $\sqrt{Q(u)} = (R^2-\alpha^3)/R$, we find that

$$\theta = 4(R-\alpha^2)R^{-2/3}. \tag{6.8}$$

Suppose that $q \equiv 1 \pmod 3$ so that $\sqrt{-3} \in F$. We require $\theta \in F^3(y) \setminus F(y)$ and u, $\sqrt{D_f(u)} \in F(y)$, whence $R = \mu S^3$, where μ is a non-cube in F and $S \in F(y)$. Thus f and g are determined by (v) of Theorem 2.1 (II).

Alternatively, suppose that $q \equiv -1 \pmod 3$. For u and $\sqrt{D_f(u)}$ to be in $F(y)$ we require R in $F^2(y)$ and $R\bar{R} = \alpha^3$ (where $\bar{R}$ is the conjugate of R over $F(y)$). Together with the fact that θ (given by (6.8)) is in $F^6(y) \setminus F^2(y)$, this implies that $R = \nu\alpha^{3/2}(S+\sqrt{-3})^3/(S-\sqrt{-3})^3$, where $S \in F(y)$ and ν is a non-cube in F^2 with $\nu\bar{\nu} = 1$ if $\sqrt{\alpha} \in F$ and $\nu\bar{\nu} = -1$ if $\sqrt{\alpha} \notin F$. This gives the second part of (v) of Theorem 2.1 (II). Once again the steps are reversible.

Note finally in this case that

$$\mathscr{D}_{f^*} = \begin{cases} \mathscr{L}_1^2\mathscr{L}_2^2\mathscr{L}_3\mathscr{L}_4,^{(2)} & \text{if} \quad \sqrt{\alpha} \in F, \\ \mathscr{L}_1^2\mathscr{L}_2^2\mathscr{Q}, & \text{if} \quad \sqrt{\alpha} \notin F. \end{cases}$$

(2) If $\alpha = 1$, then $R_3 = \mathscr{L}_1$ (case (a)).

(d) $\mathscr{D}_f = \mathscr{L}^2\mathscr{Q}^2$. By Proposition 6.1 (ii), even in F^2 we have $V(g) \subseteq V(f)$. Moreover, in F^2, $\mathscr{D}_f = \mathscr{L}^2\mathscr{L}_1^2\mathscr{L}_2^2$, say. So, by (b) and (c), there exist L, L_1, L_2 in $F^2(x)$ such that, if $f^*(x) = (x^4+4x^3)/\mu(8x-4)$ (where μ is a non-cube in F^2) and $g^*(x) = x^3$, then

$$f = L\circ f^*\circ L_1, \quad g = L\circ g^*\circ L_1. \tag{6.9}$$

Now f, g, μf^* and g^* are actually in $F(x)$. Consequently, (6.9) yields

$$f^* = L^*\circ(\mu\bar{\mu}^{-1})f^*\circ L_1^*, \quad g^* = L^*\circ g^*\circ L_2^*, \tag{6.10}$$

where $L^* = L^{-1}\circ\bar{L}$, $L_i^* = \bar{L}_i\circ L_i^{-1}$, $i = 1, 2$ and, typically, $\bar{L}$ is the conjugate of L over $F(x)$. It follows from (6.10) that, in F^2, $V(g^*) = V(L^*(g^*))$. However, $|F^2| \equiv 1 \pmod 3$ and so $g^*(x)$ $(= x^3)$ is not a permutation polynomial in F^2. For large q, the only other possibility permitted by Theorem 2.1 (with $m = 3$) is that $L^*(x^3) = (L_3(x))^3$ for some L_3 in F^2. Clearly this implies that $L^*(x) = \beta x$ or $1/\beta x$ for some non-zero β in F^2. But then, from (6.10) again, either $f^* = \gamma f^*(L_1^*)$ or $f^* = 1/\gamma f^*(L_1^*)$, where $\gamma = \beta\mu\bar{\mu}^{-1}$. It is a simple exercise to show that the latter alternative is impossible for any L_1^* and the former implies that L_1^* is the identity and $\gamma = 1$. Thus L_1 and L_4 (where $L_4(x) = L(x/\mu)$) are actually in $F(x)$. However, by (6.9), $f = L_4\circ(\mu f^*)\circ L_1$ which implies that over F, f and μf^* have the same ramification data which by case (b) contradicts the assumption that $\mathscr{D}_f = \mathscr{L}\mathscr{Q}^2$. Hence this form is impossible.

(e) $\mathscr{D}_f = \mathscr{Q}^2\mathscr{L}_1\mathscr{L}_2$ *or* $\mathscr{Q}_1^2\mathscr{Q}_2$. Using Proposition 6.2 (ii) to work in F^2, we have $\mathscr{D}_f = \mathscr{L}_3^2\mathscr{L}_4^2\mathscr{L}_1\mathscr{L}_2$, where necessarily $E(\mathscr{L}_3) = E(\mathscr{L}_4) = \{1, 3\}$. But this is impossible by case (c).

(f) $\mathscr{D}_f = \mathscr{C}^2$. We must have $E(\mathscr{C}) = \{1, 3\}$. Replace F by F^6 so that now $\sqrt{-3} \in F$ and $\mathscr{D}_f = \mathscr{L}_1^2\mathscr{L}_2^2\mathscr{L}_3^2$, $E(\mathscr{L}_i) = \{1, 3\}$, $i = 1, 2, 3$. Then although $V(g) \subseteq V(f)$ will now be false, we still must have $\sqrt{D_f(u)}$ (where $u = g(y)$) in $F(y)$. Further, as in case (b), $f = L\circ f^*\circ L_1$, where $f^*(x) = (x^4+4x^3)/(4x+a)$ and θ (given by (6.6)) is in $F(y)$. The argument of case (b) forces $a = -2$. But then $E(\mathscr{L}_3)$ (say) must be $\{2, 2\}$ and we have a contradiction.

This exhausts the possibilities for $\mathscr{D}_f$. Hence the discussion of Theorem 2.1 for $\deg f = 4$ is complete in the "irreducible case".

7. Functions of degree 4, the reducible case. We may suppose that $h(x, y)$ has degree 4 in x and is reducible yet does not have a linear factor. Thus h must be the product of two irreducible quadratics. We use Proposition 1.1 in the following form.

PROPOSITION 7.1. *Suppose that, in the situation of Proposition* 1.1, $h = h_1h_2$, *where both* h_1 *and* h_2 *are irreducible quadratics in* x *over* $F(y)$. *Then* (1.1) *holds if and only if* $D_{h_1}(y)/D_{h_2}(y)$ *is a non-square in* F *itself.*

Proof. Here (1.1) is equivalent to the fact that h_1 and h_2 have different splitting fields over $F(y)$ but the same splitting field over $F^2(y)$. The result follows.

Now take $h(x, y) = f(x) - g(y)$. In our situation the next assertion is not hard to see and, in any case, follows from a result of Fried (Proposition 2 of [7]). It is that there exist rational functions $\hat{f}$, $\hat{g}$, f_1, g_1 in $F(x)$ such that $f = \hat{f}(f_1)$, $g = \hat{g}(g_1)$, $\hat{f}(x) - \hat{g}(y)$ is also the product of two irreducible factors in $F(x, y)$ and the splitting field of $\hat{f}(x) - t$ over $F(t)$ (where t is an indeterminate) is the same as that of $\hat{g}(x) - t$ over $F(t)$. Clearly, $\deg\hat{f} = 2$ or 4. We consider each case in turn and determine precisely when (1.1) or its equivalent in Proposition 7.1 is satisfied.

(a) $\deg\hat{f} = 2$. Clearly, $\hat{g} = \hat{f}(R)$ for some R in $F(x)$. Replacing g_1 by $R(g_1)$ we may assume that $\hat{g} = \hat{f}$. Further, replacing f_1 by $L\circ f_1\circ L_1$ and $\hat{f}$ by $\hat{f}(L^{-1})$ for appropriate L, L_1 in $F(x)$, we may take $f_1(x) = x^2$ or $(x^2-\lambda)/x$, where, as always, λ is a non-square in F. Indeed, we may then replace $\hat{f}$ by $L_2(\hat{f})$, say, and assume that $\hat{f}(x) = x^2 + ax$ $(a \in F)$ or $\hat{f}(x) = (x^2+a)/2(x+\beta)$ $(a, \beta \in F$, not both 0 and $\beta^2 \neq a)$. However, if, for instance, $f(x) = x^2 + ax$, $f_1(x) = x^2$, then

$$f(x) - g(y) = (x^2 - g_1(y))(x^2 + g_1(y) + a)$$

and clearly, by Proposition 7.1, (1.1) can hold only if $a = 0$ and $\sqrt{-1} \notin F$, i.e. $q \equiv -1 \pmod 4$. In this way, it is a straightforward exercise to reduce the possibilities to one of the following (i)–(iv).

(i) $\hat{f}(x) = x^2 = f_1(x)$, $q \equiv -1 \pmod 4$. Here

$$f(x) - g(y) = (x^2 - g_1(y))(x^2 + g_1(y))$$

and (1.1) holds for any g_1 by Proposition 7.1. Moreover, $V(g) = V(f)$ if and only if for all x in F, either $g_1(y) = x^2$ or $g_1(y) = -x^2$ is soluble for y in F, i.e. if and only if $g_1 = P$.

(ii) $\hat{f}(x) = (x^2+\lambda)/x$, $f_1(x) = x^2$. Here

$$f(x) - g(y) = (x^2 - g_1(y))(x^2g_1(y) - \lambda),$$

giving rise (as in (i)) to the pair (ii) of Theorem 2.3.

(iii) $\hat{f}(x) = (x^2+\lambda)/2x$, $f_1(x) = (x^2-\lambda)/2x$. Here

$$f(x) - g(y) = (x^2g_1(y) - 2\lambda x - \lambda g_1(y))(x^2 - 2xg_1(y) - \lambda),$$

the first quadratic having discriminant $4\lambda(g_1^2+\lambda)$ and the second $4(g_1^2+\lambda)$. This leads to pair (iii) of Theorem 2.3. Further, $V(g) = V(f)$ if and only if for all x in F either $g_1(y) = 2\lambda x/(x^2-\lambda)$ or $g_1(y) = (x^2-\lambda)/2x$ is soluble. But, easily,

$$V(2\lambda x/(x^2-\lambda)) \cup V((x^2-\lambda)/2x) = F$$

so $V(g) = V(f)$ if and only if $g_1 = P$.

(iv) $\hat{f}(x) = x^2$, $f_1(x) = (x^2-\lambda)/x$. Here

$$f(x)-g(y) = \big(x^2-xg_1(y)-\lambda\big)\big(x^2+xg_1(y)-\lambda\big),$$

the two factors having identical discriminants so that Proposition 7.1 cannot be satisfied.

(b) $\deg \hat{f} = 4$. Here $f_1 = L$ and we may assume, in fact, that $f = \hat{f}$. Let K be the common splitting field of $f(x)-t$ and $\hat{g}(x)-t$ over $F(t)$ with corresponding isomorphic galois groups $G(f)$, $G(\hat{g})$, respectively. Let y be a zero of $g(x)-t$ and put $v = g_1(y)$. Thus $\hat{g}(v) = t$ and so $v \in K$. By Proposition 7.1, $[K(y):F(y)] = 4$ (although $[K(y):F^2(y)] = 2$). By the theorem of natural irrationalities, r is divisible by 4. But $|G(\hat{g})| = rH(\hat{g})$ where $r\,\big|\,\big(H(\hat{g})-1\big)!$. On the other hand, $|G(f)|\,\big|\,24$. The only consistent conclusion is that $r = 4$, $\deg\hat{g} = 6$ and $G(f) = S_4$, the symmetric group. Thus $G(\hat{g})$ must be a transitive subgroup of S_6 isomorphic to S_4. The situation just described seems unlikely; nevertheless there are circumstances where it would occur save for the assumption that $f(x)-\hat{g}(y)$ be reducible, namely when $\hat{g}(x)$ is $\mathscr{R}(x^2)$, where $\mathscr{R}$ is the cubic resolvent of f. However, the additional hypothesis that $f(x)-\hat{g}(y)$ be reducible enables us to reach a contradiction as follows. Consider the subgroup V of $G(f)$ whose members fix a prescribed root of $f(x) = t$. Then $V \cong S_3$. Regarding V as a subgroup of $G(\hat{g})$ and using the fact that $f(x)-\hat{g}(v)$ is reducible, we see that for suitable numbering of the roots of $\hat{g}(x) = t$ we have

$$V = \{(123)(456),\ (132)(465),\ (12)(34),\ (13)(46),\ (23)(56),\ (1)\}.$$

However, there is no way V could be one of precisely four conjugate subgroups of any transitive subgroup of S_6. So this case is, after all, impossible.

It may be helpful to point out that, in the above, the known example [4] of a pair (f, g) with $f(x)-g(y)$ reducible and $\deg\hat{f} = 4$ (namely, $f(x) = (x^2-1)^2$, $g(x) = -4x^2(x^2-1)$) is eliminated by the demand that $[K(y):F(y)] = 4$.

8. x-soluble polynomials of total degree 3. For general polynomials $h(x, y)$ of degree 3 or 4 in x, the normalisation procedure achieved in §§ 5–6 for the case $h(x, y) = f(x)-g(y)$ is not available. However, we can characterise those polynomials h of *total* degree 3 in $F[x, y]$ which are x-soluble, thus extending work of Mordell [15]. We use Proposition 5.1 in the following form.

LEMMA 8.1. *Suppose that in Proposition* 5.1, *h has the form*

$$h(x, y) = x^3+h_1(y)x+h_0(y), \quad h_0 \neq 0. \tag{8.1}$$

Then (5.1) *holds if and only if* $q \equiv -1 \pmod 3$ *and* $h_1 = 0$ *or*

$$h_1 = -3(A^2+3\lambda B^2), \quad h_0 = 2A(A^2+3\lambda B^2),$$

where $A, B\ (\neq 0) \in F(y)$.

Proof. If $h_1 = 0$, then $D_h = -27h_0^2$ and (5.1) holds if and only if $\sqrt{-3} \notin F$.

If $h_1 \neq 0$, put $h_2 = -2h_1A/3$. Then

$$D_h(y) = 12h_1^2(-\tfrac{1}{3}h_1-A^2)$$

and (5.1) holds if and only if $-\frac{1}{3}h_1-A^2 = 3\lambda B^2$. The result follows.

Before stating our theorem, we note that if h is x-soluble, then so is

$$h_1(x, y) = ah(bx+cy+d,\ ey+f), \quad abe \neq 0, \tag{8.2}$$

and we say that h and h_1 are *x-equivalent*.

THEOREM 8.2. *Let $h(x, y)$ in $F[x, y]$ be a polynomial of total degree* 3 *and suppose that $q > q_0$ (absolute) and $p > 3$. Then h is x-soluble if and only if it has a factor linear in x or is x-equivalent to a polynomial of one of the following types:*

$$\left.\begin{array}{rl} \text{I.} & x^3-g(y), \\ \text{II.} & (x+y+1)^3-27xy, \end{array}\right\} \textit{with } q \equiv -1 \pmod 3,$$

$$\left.\begin{array}{rl} \text{III.} & x^3+3\eta x+y(3x^2+\eta), \\ \text{IV.} & x^3+3\eta(y+1)^2x+y\big(3x^2+\eta(y+1)^2\big), \end{array}\right\} \textit{with } \eta = \begin{cases} 1, & \textit{if } q \equiv -1 \pmod 3, \\ \lambda, & \textit{if } q \equiv 1 \pmod 3, \end{cases}$$

$$\text{V.}\quad x^3-(3x-2)(3\lambda y^2+1).$$

Remark. Actually, apart from II, all the above x-soluble h derive from functions of the form $f(x)-g(y)$. For essentially III is $[(x+\sqrt{\eta})/(x-\sqrt{\eta})]^3-(y-\sqrt{\eta})/(y+\sqrt{\eta})$, the transformation $x \to x/(y+1)$, $y \to y/(y+1)$ sends III onto IV and V is $(L\circ f^*\circ L_1)(x)-L\big(g^*(y)\big)$, where f^* and g^* are given by (i) of Theorem 2.1 (II), $L(x) = 4/x$ and $L_1(x) = -2/x$.

Proof. Suppose that $h(x, y)$ is x-soluble. From Theorem 4.1 we may assume that h is irreducible of degree 3 in x and so is x-equivalent to a polynomial of the form (8.1) also of total degree 3. By Lemma 8.1, either h is x-equivalent to I for some g or is x-equivalent to

$$x^3-3C^{-2}(A^2+3\lambda B^2)x+2AC^{-3}(A^2+3\lambda B^2), \tag{8.3}$$

where A, B, C and the coefficients of (8.3) are all non-zero polynomials in $F[y]$. We may suppose also that A, C and $E = A^2+3\lambda B^2$ are co-prime. Since $C^2|E$, then A and C are co-prime and so $C^3|E$. However, A and B are also co-prime, for, if not, we would have A linear and both AB^{-1} and C in F. But then

$$A^{-3}C^3h(AC^{-1}x, y) = x^3-3ax+2a, \quad a = 1+3\lambda A^{-2}B^2 \in F,$$

which has discriminant $\lambda(18B/A)^2$ and so is reducible, whence h is reducible.

Now write (8.3) as $x^3-3CGx+2AG$, where C and $G:=C^{-3}E\in F(y)$ and $\deg C+\deg G\leqslant 2$, $\deg A+\deg G\leqslant 3$. We consider the various possibilities for C and G. Let $\bar{F}$ be the algebraic closure of F and put $\delta=\sqrt{-3\lambda}$ in $\bar{F}$.

(a) $\deg G=0$. Thus $\deg C=1$ or 2 and $\deg E=3$ or 6. Since A and B are co-prime, then so are $A+\delta B$ and $A-\delta B$. Therefore, in $\bar{F}(y)$, $A+\delta B=C_1^3$ where C_1 divides C. But then, since $G\in F$, $h(x,y)$ has a factor in $\bar{F}[x,y]$ of $x+G^{1/3}C_1+CG^{2/3}/C_1$ (by Cardan's Formula) which contradicts the fact that h is absolutely irreducible (Proposition 1.2 (iii)).

(b) $\deg C=\deg G=1$. This case is impossible for it would imply that E has degree 4 yet is divisible by C^3.

(c) $\deg C=0$, $\deg G=1$. For this $\deg A=\deg B=\deg E=1$ so that $\delta\in F$, i.e. $q\equiv -1 \pmod 3$. Replacing x by $C^{-1}x$ any y by $ay+b$ for suitable a ($\neq 0$), b in F, we may take $A(y)=y$, $B(y)=(y+1)/\sqrt{-3}$, so that h is x-equivalent to $x^3+(3x-2y)(2y+1)$. Now the transformation $x\to-\frac{1}{3}(x+y+1)$, $y\to-\frac{1}{2}(y+1)$ shows that h is x-equivalent to II.

(d) $\deg C=0$, $\deg G=2$. Then $\deg E=2$ and $\deg A\leqslant 1$. If $\deg A=\deg B=1$, then, as in (c), we may set $A(y)=y$, $B(y)=y+1$ and h is x-equivalent to

$$x^3-3\big(y^2+3\lambda(y+1)^2\big)x+2y\big(y^2+3\lambda(y+1)^2\big).$$

A further transformation $x\to x+y$ indicates that h is x-equivalent to IV but with $\eta=-3\lambda(=\delta^2)$. To get $\eta=1$ in the case $q\equiv -1 \pmod 3$, apply the extra transformation $x\to\delta x$, $y\to\delta(y+1)-1$.

A similar discussion reveals that if $\deg A=1$, $\deg B=0$, then h is x-equivalent to III, while if $\deg A=0$, $\deg B=1$, then h is x-equivalent to V.

The sufficiency of I–V is obvious from the above and Lemma 8.1. Thus the proof is complete.

From Theorem 8.2, it is easy to guess which cubics h are both x-soluble and y-soluble in F. Formal verification of Mordell's result [15] (stated below) is indeed possible from this starting point and probably represents a shorter and less intricate method than that of Mordell. Nevertheless, the proof is not actually immediate and, for brevity, is omitted.

THEOREM 8.3 (Mordell). *Let $h(x,y)$ in $F[x,y]$ have total degree 3. Suppose that $q>q_0$ and $p>3$. Then h is both x-soluble and y-soluble in F if and only if h has a factor linear in x and a factor (possibly the same) linear in y or one of $h(x,y)$ and $h(y,x)$ is of the form (8.2) with $c=0$, where h_1 is one of I–III in Theorem 8.2 with $g(y)=y$ or y^3+1 in I.*

9. Covering sets. We shall call a set of functions $\{f_i(x)\}$ in $F(x)$ a *covering set* if $\bigcup_i V(f_i)=F$. Here are some simple examples.

(i) $\{f_i\}$, $f_i=P$ for some i.

(ii) $\{x^r,\gamma x^r,\ldots,\gamma^{r-1}x^r\}$, where $r\mid(q-1)$ and γ in F is a non dth power for any divisor d of r with $d>1$.

(iii) $\{(x^2-\lambda)/2x,\ 2\lambda x/(x^2-\lambda)\}$ (see § 7).

Using exponential sums, Mordell [16], [17], has constructed a non-trivial covering set comprising a function of degree 4 and a function of degree 3. However, the natural approach to covering sets may be to use the following result which follows immediately from Proposition 1.1. (For related work on the more general problem $V(f)\subseteq\bigcup_i V(f_i)$, see [5], [9], § 4.)

PROPOSITION 9.1. *In the situation of Proposition* 1.1, *let*

$$h(x,y)=\prod_{i=1}^{m}\big(f_i(x)-y\big).$$

Suppose that

$$G^*\big(K,F(y)\big)=\bigcup_{i=1}^{m}\Big(\bigcup_{x_i}G^*\big(K,F(x_i)\big)\Big),$$

where the inner union is over all roots x_i of $f_i(x)=y$. Then $\{f_i\}$ is a covering set for F.

Using Proposition 9.1 and previous results in this paper, we can demonstrate some examples of covering sets valid for any F with $p>3$.

(iv) Mordell's covering set

$$\{f_1,f_2\}=\{x^4+ax^2+bx,\ (x^3+2ax^2-a^2x-b^2)/4x\},\quad b\neq 0,$$

follows easily from Proposition 9.1 since $f_2(x)-y$ is the cubic resolvent of $f_1(x)-y$. Another covering set arising in this way is

$$\big\{(x^4+ax+b)/x^2,\ x-\big(a^2/(x^2-4b)\big)\big\},\quad ab\neq 0.$$

(v) From (i) of Theorem 2.1 (II) (cf. Theorem 8.2 (v)), we get the pair

$$\{(x^3-3x+2)/(3x-2),\ 3x^2\}.$$

(vi) From (iii) of Theorem 2.1 (II) we get the pair

$$\{f_1,f_2\}=\{(x^4+4x^3)/(8x-4),\ x^3\}.$$

However, although this is a covering set for all q, the manner of the covering depends on q. For, of course, if $q\equiv -1 \pmod 3$, then $f_2=P$ and we have a trivial covering set of type (i). On the other hand, if $q\equiv 1 \pmod 3$, then $|V(f_1)\cap V(f_2)|\doteqdot q/12$.

(vii) Finally, we exhibit a non-trivial covering set of three functions. Put

$$\{f_1,f_2,f_3\}=\{(x^2-1)^3,\ -4x^2(x^2-1),\ \tfrac{1}{2}(x-1)^2/(x^2+1)\}.$$

(As noted earlier (§ 7), $f_1(x)-f_2(y)$ is reducible.) Then the roots of $f_i(x) = y$, $i = 1, 2, 3$, can be written as $\{\alpha_1, -\alpha_1, \alpha_2, -\alpha_2\}$, $\{\beta_1, -\beta_1, \beta_2, -\beta_2\}$, $\{\gamma, \gamma^{-1}\}$, respectively, where $\beta_i = \frac{1}{2}(\alpha_1 \pm \alpha_2)$, $i = 1, 2$, and

$$\gamma = -(2y-1)^{-1}\big(1-2\alpha_1\alpha_2(\alpha_1^2-1)\big).$$

Moreover, $G^*(K, F(y))$ is the whole galois group G, say, and has order 8. If the roots of each $f_i(x) = y$ are numbered in the order given, then the action of G as a permutation of these roots is as follows:

f_1	f_2	f_3
(1)	(1)	(1)
(12)	(14) (23)	(12)
(34)	(13) (24)	(12)
(12) (34)	(12) (34)	(1)
(13) (24)	(34)	(12)
(14) (23)	(12)	(12)
(1423)	(1324)	(1)
(1324)	(1423)	(1)

Thus $\{f_1, f_2, f_3\}$ is a covering set by Proposition 9.1.

References

[1] S. D. Cohen, *The distribution of polynomials over finite fields*, Acta Arith. 17 (1970), pp. 255–271.
[2] — *The distribution of polynomials over finite fields, II*, ibid. 20 (1972), pp. 53–62.
[3] — *Composite rational functions which are powers*, Proc. R. Soc. Edinburgh A, 83 A (1979), pp. 11–16.
[4] H. Davenport, D. J. Lewis and A. Schinzel, *Equations of the form $f(x) = g(y)$*, Quart. J. Math. Oxford (2) 12 (1961), pp. 304–312.
[5] M. Fried, *Arithmetical properties of value sets of polynomials*, Acta Arith. 15 (1969), pp. 91–125.
[6] — *On a conjecture of Schur*, Michigan Math. J. 17 (1970), pp. 41–55.
[7] — *The field of definition of function fields and a problem in the reducibility of polynomials in two variables*, Illinois J. Math. 17 (1973), pp. 128–146.
[8] — *On a theorem of MacCluer*, Acta Arith. 25 (1974), pp. 121–126.
[9] — *Arithmetical properties of function fields, II; The generalized Schur problem*, ibid. 25 (1974), pp. 225–258.
[10] — *On Hilbert's irreducibility theorem*, J. Number Theory 6 (1974), pp. 211–231.
[11] — *Galois groups and complex multiplication*, Trans. Amer. Math. Soc. 235 (1978), pp. 141–163.
[12] N. Jacobson, *Basic algebra I*, Freeman, San Francisco 1974.
[13] I. Kaplansky, *Fields and rings*, Chicago lectures in Mathematics, Chicago 1969.

[14] K. McCann and K. S. Williams, *Cubic polynomials with the same residues* (mod p), Proc. Cambridge Phil. Soc. 64 (1968), pp. 655–658.
[15] L. J. Mordell, *Cubic polynomials with the same residues* (mod p), Proc. London Math. Soc. (3) 21 (1970), pp. 129–144.
[16] — *Rational functions representing all residues* mod p, J. London Math. Soc. 5 (1972), pp. 166–168.
[17] — *Rational functions representing all residues* mod p, *II*, Proc. Amer. Math. Soc. 35 (1972), pp. 411–412.
[18] G. I. Perel'muter, *On certain sums of characters*, Uspekhi Matem. Nauk 18 (1963), pp. 145–149.

UNIVERSITY OF GLASGOW
Glasgow G12 8QW
Scotland

Received on 30.8.1978
and in revised form on 9.11.1978 (1097)