Associative rational functions in two variables*

Joel V. Brawley[1], Shuhong Gao[1] and Donald Mills[2]

[1] Department of Mathematical Sciences
Clemson University
Clemson, SC 29634-0975 USA
E-mail: brawley@CLEMSON.EDU, sgao@math.clemson.edu

[2] Department of Mathematics
Southeastern Louisiana University
Hammond, LA 70402 USA
E-mail: dmills@selu.edu

Abstract. A rational function $R(x, y)$ over a field is said to be associative if $R(R(x, y), z) = R(x, R(y, z))$. Associative rational functions over a field define group laws on subsets of the field (plus the point at infinity). In this paper, all the associative rational functions of two variables over an arbitrary field are determined and consequently all the groups obtainable from such functions are determined as well.

1 Introduction and main results

Let $\mathbb{F}$ be any field. Trivially the polynomial $x + y$ defines, via substitution, a group operation (or group law) on $\mathbb{F}$, and the polynomial xy defines a group operation on $\mathbb{F} \setminus \{0\}$. Further, it is well-known that the polynomial $x + y - 1$ also defines a group on $\mathbb{F}$ and that the polynomial $x + y + xy$ defines a group on $\mathbb{F} \setminus \{-1\}$. It is natural to ask for other such polynomials $f(x, y) \in \mathbb{F}[x, y]$ or even rational functions $R(x, y) \in \mathbb{F}(x, y)$ which define group operations on subsets of $\mathbb{F}$. More generally, one may consider formal power series in two variables and this leads to formal groups of dimension one which are closely related to elliptic curves [6]; see also [4] for formal groups of higher dimension.

We are interested in determining rational functions over $\mathbb{F}$ that can be used to define group laws on some subsets of $\mathbb{F}$. Let $R(x, y) \in \mathbb{F}(x, y)$ be any rational function. We say that $R(x, y)$ is *associative* if the equation

$$R(R(x, y), z) = R(x, R(y, z)) \tag{1}$$

is valid in $\mathbb{F}(x, y, z)$, the rational function field with distinct variables x, y, z. A rational function $R_1(x, y)$ is said to be *equivalent* to $R(x, y)$ if there is a linear fractional map $f(x) = (ax + b)/(cx + d) \in \overline{\mathbb{F}}(x)$, $ad - bc \neq 0$, such that

$$R_1(x, y) = f^{-1}(R(f(x), f(y))) \tag{2}$$

* The second author was supported in part by NSF under Grant #DMS9970637 and NSA under Grant #MDA904-00-1-0048. The current address for the third author: Donald Mills, Department of Mathematical Sciences, United States Military Academy, West Point, NY 10996; E-mail: ddmills@arl.mil.

where $\overline{\mathbb{F}}$ denotes the algebraic closure of $\mathbb{F}$ and $f^{-1}(x) = (-dx+b)/(cx-a)$ is the inverse of f. For example, $x+y+xy$ is equivalent to xy under $f(x) = x+1$. It is easy to check that $R_1(x, y)$ is associative iff $R(x, y)$ is. This defines an equivalence relation on associative rational functions over $\mathbb{F}$. The functions $x, y, x + y, x + y + xy$ are associative but not equivalent to each other. One of our main goals is to show that they represent all the associative rational functions over $\mathbb{F}$.

Theorem 1. *Let $\mathbb{F}$ be any field. Then any nonconstant associative rational function in $\mathbb{F}(x, y)$ is equivalent to exactly one of $x, y, x + y$ and $x + y + xy$.*

Remark. In defining the equivalence above, we could have allowed f to be any rational function in $\overline{\mathbb{F}}(x)$ whose compositional inverse is also a rational function in $\overline{\mathbb{F}}(x)$, that is, f is an automorphism of the field $\overline{\mathbb{F}}(x)$ fixing $\overline{\mathbb{F}}$. But the only automorphisms of $\overline{\mathbb{F}}(x)$ that fix $\overline{\mathbb{F}}$ are linear fractional maps.

With this theorem, we can determine all groups whose elements are from $\mathbb{F}$ and whose group laws can be defined by rational functions over $\mathbb{F}$. In the first place, if G is any finite group of order $\leq |\mathbb{F}|$ and S is any subset of $\mathbb{F}$ with $|S| = |\mathbb{F}|$, then it is well-known that any one-to-one correspondence between S and G can be used to define a group law on S for which S and G are isomorphic. Further, this group law for S can always be defined by means of a polynomial $f(x, y)$ obtained by Lagrange interpolation (for two variables); thus, every finite group of order $\leq |\mathbb{F}|$ is possible. However, the polynomial $f(x, y)$ may not be associative in the above sense and classifying such polynomials amounts to classifying finite groups, which is certainly a hard problem with a nature different from that of an infinite S. We do not intend to pursue the case of a finite S in this paper, but instead will concentrate on infinite S.

Henceforth it is assumed that S is infinite (subset of $\mathbb{F}$) with its group law defined by a rational function $R(x, y)$ (i.e. for $a, b \in S$, $a \cdot b = R(a, b)$). Since a group law is associative, we have

$$R(R(a, b), c) = R(a, R(b, c)) \text{ for all } a, b, c \in S. \tag{3}$$

By using the fact that S is infinite, one can show that (3) implies (1). So $R(x, y)$ must be associative and, by Theorem 1, is equivalent to one of $x, y, x + y, x + y + xy$. But none of x or y can define a group so $R(x, y)$ is equivalent to $x + y$ or $x + y + xy$. The linear fractional map that defines the equivalence induces an isomorphism of S to some subgroup of the additive or multiplicative group of $\overline{\mathbb{F}}$. We will show that the linear fractional map can be defined over a quadratic extension of $\mathbb{F}$, so S is in fact isomorphic to a subgroup of the additive or multiplicative group of a quadratic extension of $\mathbb{F}$.

Theorem 2. *For any infinite subset S of a field $\mathbb{F}$, if S is a group defined by a rational function over $\mathbb{F}$ then S is isomorphic to a subgroup of the additive or multiplicative group of a quadratic extension of $\mathbb{F}$.*

We should remark that when using a rational function to define a group law, it is necessary to use ∞, the point at infinity (or use the projective space $\mathbb{P}^1(\mathbb{F})$). Here we adopt the usual convention that, for $a, b, c, d \in \mathbb{F}$,

$$\frac{a\infty + b}{c\infty + d} = \frac{a}{c} \quad \text{for } c \neq 0,$$
$$\frac{a\infty + b}{c\infty + d} = \infty \quad \text{for } ad \neq 0 \text{ and } c = 0,$$
$$\frac{a}{0} = \infty \quad \text{for } a \neq 0.$$

We give below two examples of groups which can be verified directly. We denote by $\mathbb{F}_q$ a finite field of q elements. Example 3 comes from the proof of Lemma 12, while Example 4 is a special case of Lemma 11.

Example 3. Let $\mathbb{F} = \mathbb{F}_2(t^2)$ where t is transcendental over $\mathbb{F}_2$, $S = \mathbb{F}$ and $R = (xy + t^2)/(x + y)$. Then R is equivalent to $x + y$ and R defines a group on S isomorphic to the additive group

$$\left\{ \frac{1}{h + t} : h \in \mathbb{F} \right\}$$

of $\mathbb{F}_2(t)$, a quadratic extension of $\mathbb{F} = \mathbb{F}_2(t^2)$. Note that the group isomorphism is defined by $f = 1/(x + t)$.

Example 4. Let $R = (xy - 1)/(x + y + 1) \in \mathbb{F}_q(x, y)$ where q is a power of a prime > 3. Let ω be a root of $X^2 + X + 1$. Then R is equivalent to xy under the linear fractional map $f(x) = (1 - \omega x)/(x + \omega + 1)$. Depending on whether ω lies in $\mathbb{F}_q$, we have the following two cases.

(a) Suppose $q - 1$ is divisible by 3. Let $S = (\mathbb{F}_q \cup \{\infty\}) \setminus \{\omega, \omega^2\}$. Then R defines a group on S isomorphic to the multiplicative group of $\mathbb{F}_q$ with the isomorphism defined by $f(x)$.
(b) Suppose $q - 1$ is not divisible by 3. Let $S = \mathbb{F}_q \cup \{\infty\}$. Then R defines a group on S isomorphic to the unique subgroup of order $q + 1$ of the multiplicative group of $\mathbb{F}_{q^2}$ with the isomorphism defined by $f(x)$.

We should mention the related results in formal groups. If one is allowed to use formal power series for f then $x + y$ and $x + y + xy$ can be obtained from each other. Indeed if the characteristic of $\mathbb{F}$ is zero then

$$x + y + xy = f^{-1}\left(f(x) + f(y)\right)$$

where

$$f(x) = \log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n-1}\frac{x^n}{n} + \cdots,$$
$$f^{-1}(x) = \exp(x) - 1 = x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots.$$

In general, a formal power series $R(x,y) \in \mathbb{F}[[x,y]]$ is called a *formal group* if $R(x,y) = x + y +$ (terms of orders ≥ 2) and $R(x,y)$ is associative (i.e. satisfying (1). When the characteristic of $\mathbb{F}$ is zero, it can be proved that every formal group $R(x,y) \in \mathbb{F}[[x,y]]$ is equivalent to $x+y$ via a formal power series $f \in \mathbb{F}[[x]]$ defined by some generalized logarithm associated with $R(x,y)$; see [6, Proposition 5.2, p. 122] or [4, Corollary 5.4.4, p. 31]. Our result shows that all the formal groups that are rational functions are equivalent to either $x+y$ or $x+y+xy$ with the equivalence defined by rational functions, instead of formal power series. Note that determining all the formal groups that are polynomials is much easier as indicated by the exercise in [4, p. 2] and Theorem 1.1 in [3, p. 12]; see also Lemma 5 in the next section.

Our interest in this paper was motivated by the work in [1,2,5] where it is desirable to construct irreducible polynomials of higher degrees from those of lower degrees. In [1], Brawley and Carlitz introduce a so-called diamond operation on a subset of a field and use this operation to form composition of polynomials. When the diamond operation defines a group law on the subset then composition of irreducible polynomials gives irreducible polynomials under certain conditions. In [2], it is shown that composition of polynomials can be computed efficiently if the diamond operation is defined by a polynomial or a rational function.

For the rest of the paper, we prove Theorem 1 in Section 2 and Theorem 2 in Section 3 .

2 Associative rational functions

We start with the easy case when $R(x,y)$ is a polynomial; its proof involves the basic ideas that are used in the general case when $R(x,y)$ is an arbitrary rational function.

Lemma 5. *Let $R(x,y) \in \mathbb{F}[x,y]$ be a nonconstant polynomial. If $R(x,y)$ is associative then $R(x,y)$ is equivalent to one of $x, y, x+y, x+y+xy$.*

Proof. Suppose $R(x,y) = \sum c_{ij}x^iy^j$. The associativity of R implies

$$\sum c_{ij}(R(x,y))^i z^j = \sum c_{ij}x^i(R(y,z))^j.$$

Let m be the degree in x of $R(x,y)$. Then comparing the degrees in x in the above equation gives $m^2 = m$, so m must be 0 or 1. Similarly, the degree in y of $R(x,y)$ is also at most 1. Hence we may assume that $R(x,y) = axy + bx + cy + d$ where $a,b,c,d \in \mathbb{F}$. Then

$$R(R(x,y),z) = a^2xyz + abxy + abxz + acyz + b^2x + bcy + (ad+c)z + (b+1)d$$
$$R(x,R(y,z)) = a^2xyz + abxy + acxz + acyz + (ad+b)x + bcy + c^2z + (c+1)d.$$

Comparing the coefficients gives

$$ab = ac, \quad b^2 = ad + b, \quad ad + c = c^2, \quad (b+1)d = (c+1)d.$$

First assume $a \neq 0$. Then $b = c$ and $d = (c^2 - c)/a$, so

$$R(x, y) = axy + c(x + y) + (c^2 - c)/a.$$

Let $f(x) = (x - c + 1)/a$. Then $x + y + xy = f^{-1}(R(f(x), f(y)))$. This means that in this case $R(x, y)$ is equivalent to $x + y + xy$ under $f^{-1}(x) = ax + c - 1$.

Assume now $a = 0$. Then $b^2 = b$ and $c^2 = c$. If $b \neq c$ then $d = 0$. Hence the only choices for R are $x, y, x + y + d$. The latter is equivalent to $x + y$ under the map $f = x + d$. □

In general let $R(x, y)$ be any associative rational function over $\mathbb{F}$, say $R(x, y) = F(x, y)/H(x, y)$ where $F(x, y), H(x, y) \in \mathbb{F}[x, y]$ and $\gcd(F(x, y), H(x, y)) = 1$. The associativity of R gives a polynomial equation. Similar to the above argument, we shall compare the degrees of x and y in this equation and prove that both $F(x, y)$ and $H(x, y)$ have degree at most one in x and y separately. Then we reduce R to special forms and eventually to polynomials by linear fractional maps. Since the proof is a little lengthy, we break it into several lemmas.

The following lemma seems interesting by itself, as it says that relatively prime polynomials remain relatively prime when their variables are substituted by other polynomials.

Lemma 6. *Let $F(x, y), H(x, y) \in \mathbb{F}[x, y]$ with $\gcd(F(x, y), H(x, y)) = 1$ and degrees f_x, h_x in x, respectively. Let $x_1, \ldots, x_n, z$ be distinct variables and let $a, b \in \mathbb{F}[x_1, \ldots, x_n]$ with $a/b \notin \mathbb{F}$ and $\gcd(a, b) = 1$. Then*

$$b^{f_x} F(a/b, z), \quad b^{h_x} H(a/b, z)$$

are polynomials in $\mathbb{F}[x_1, \ldots, x_n, z]$ and are relatively prime.

Proof. We can view $F(x, y), H(x, y)$ as univariate polynomials in y with coefficients in $\mathbb{F}(x)$, the rational function field in x. Since $\gcd(F(x, y), H(x, y)) = 1$ in $\mathbb{F}[x, y]$, we still have $\gcd(F(x, y), H(x, y)) = 1$ in $\mathbb{F}(x)[y]$. By the Euclidean algorithm, there are polynomials $U_0, V_0 \in \mathbb{F}(x)[y]$ such that

$$U_0 F + V_0 H = 1.$$

By eliminating the denominators in U_0, V_0, we have

$$UF + VH = W \tag{4}$$

for some $U, V \in \mathbb{F}[x, y]$ and $W \in \mathbb{F}[x]$ with $W \neq 0$. Let $\ell = f_x$ and $m = h_x$ be the degrees in x of $F(x, y)$ and $H(x, y)$, respectively. In (4), replacing x by a/b and y by z and eliminating the denominators, which are powers of b, we have

$$U_1 A + V_1 B = b^i W(a/b) \tag{5}$$

where $A = b^\ell F(a/b, z)$, $B = b^m H(a/b, z)$, $U_1, V_1 \in \mathbb{F}[x_1, \dots, x_n, z]$ and $b^i W(a/b) \in \mathbb{F}[x_1, \dots, x_n]$. It follows that $C = \gcd(A, B)$ must divide $b^i W(a/b)$, as polynomials in $\mathbb{F}[x_1, \dots, x_n, z]$. Since a/b is not a constant in $\mathbb{F}$, we have $b^i W(a/b) \neq 0$. So $C \in \mathbb{F}[x_1, \dots, x_n]$ and thus C divides all the coefficients of A, B as polynomials in $\mathbb{F}[x_1, \dots, x_n][z]$.

Let's examine the coefficients. Suppose

$$\begin{aligned} F(x,y) &= F_\ell(x)y^\ell + \cdots + F_1(x)y + F_0(x) \\ H(x,y) &= H_m(x)y^m + \cdots + H_1(x)y + H_0(x). \end{aligned}$$

Then

$$\begin{aligned} A &= b^\ell F_\ell(a/b)z^\ell + \cdots + b^\ell F_1(a/b)z + b^\ell F_0(a/b) \\ B &= b^m H_m(a/b)z^m + \cdots + b^m H_1(a/b)z + b^m H_0(a/b). \end{aligned}$$

Since $\gcd(F(x,y), H(x,y)) = 1$ in $\mathbb{F}[x,y]$, the coefficients

$$F_\ell(x), \dots, F_1(x), F_0(x), H_m(x), \dots, H_1(x), H_0(x)$$

must be relatively prime. Replacing x by a/b in these polynomials and eliminating denominators gives the coefficients of A and B. By a similar argument above via the Euclidean algorithm, we see that any common divisor of the coefficients of A and B together must divide a power of b. Since $\ell = \deg F_i(x)$ for some $1 \le i \le \ell$ and $\gcd(a, b) = 1$, we have $\gcd(b^\ell F_i(a/b), b) = 1$. Hence the coefficients of A and B must be relatively prime and C must be a constant in $\mathbb{F}$. □

Lemma 7. *Let $R(x,y) = F(x,y)/H(x,y)$ where $F(x,y), H(x,y) \in \mathbb{F}[x,y]$ with $\gcd(F(x,y), H(x,y)) = 1$. Denote by f_x, f_y the degrees of $F(x,y)$ in x, y separately, and similarly h_x, h_y for that of $H(x,y)$. If $R(x,y)$ is associative then $h_x \le f_x \le 1$ and $h_y \le f_y \le 1$.*

Proof. The associativity of $R(x,y)$ means that

$$\frac{F(R(x,y),z)}{H(R(x,y),z)} = \frac{F(x, R(y,z))}{H(x, R(y,z))}$$

i.e.

$$F(R(x,y),z)H(x,R(y,z)) = H(R(x,y),z)F(x,R(y,z)).$$

To eliminate the denominators, let

$$A = H(x,y)^{f_x} F(R(x,y),z), \quad B = H(y,z)^{h_y} H(x, R(y,z)) \tag{6}$$
$$C = H(x,y)^{h_x} H(R(x,y),z), \quad D = H(y,z)^{f_y} F(x, R(y,z)). \tag{7}$$

Then $A, B, C, D \in \mathbb{F}[x,y,z]$ and

$$H(x,y)^{h_x} H(y,z)^{f_y} A\, B = C\, D\, H(x,y)^{f_x} H(y,z)^{h_y}. \tag{8}$$

We shall prove that $h_x \le f_x \le 1$; the proof of $h_y \le f_y \le 1$ is similar.

Since $\gcd(F(x,y), H(x,y)) = 1$, we have

$$\gcd(A, H(x,y)) = 1 \quad \text{and} \quad \gcd(C, H(x,y)) = 1.$$

By Lemma 6, we also have $\gcd(A, C) = 1$. It follows from (8) that

$$A \mid D\,H(y,z)^{h_y} \quad \text{and} \quad C \mid B\,H(y,z)^{f_y}. \tag{9}$$

Consider the degrees in x of the polynomials in (9). Certainly,

$$\deg_x D\,H(y,z)^{h_y} = \deg_x D = f_x \quad \text{and} \quad \deg_x B\,H(y,z)^{f_y} = \deg_x B = h_x. \tag{10}$$

To find the degrees of A and C in x, suppose

$$F(x,y) = \sum_{i=i_0}^{f_x} a_i(y)x^i, \quad H(x,y) = \sum_{j=j_0}^{h_x} b_j(y)x^j \tag{11}$$

where $a_i(y), b_j(y) \in \mathbb{F}[y]$, $a_i(y) \ne 0$ for $i = i_0$ or f_x, and $b_j(y) \ne 0$ for $j = j_0$ or h_x. Then

$$A = \sum_{i=i_0}^{f_x} a_i(z)F(x,y)^i H(x,y)^{f_x-i}, \quad C = \sum_{j=j_0}^{h_x} b_j(z)F(x,y)^j H(x,y)^{h_x-j} \tag{12}$$

Now the degree in x of $F(x,y)^i H(x,y)^{f_x-i}$ is

$$if_x + h_x(f_x - i) = h_x f_x + i(f_x - h_x), \quad i_0 \le i \le f_x, \tag{13}$$

and that of $F(x,y)^j H(x,y)^{h_x-j}$ is

$$jf_x + h_x(h_x - j) = h_x^2 + j(f_x - h_x), \quad j_0 \le j \le h_x. \tag{14}$$

First assume $f_x > h_x$. Then $\deg_x A = f_x^2$. By (9) and (10), we have $f_x^2 \le f_x$, so $f_x \le 1$. In this case, we have $h_x = 0$ and $f_x = 1$.

Next assume $f_x = h_x$. Then all the values in (13) and (14) are equal to h_x^2. Denote $n = h_x^2$. We need to find the coefficient of x^n in A and C. Note that the leading coefficient of x in $F(x,y)$ is $a_{f_x}(y)$, denoted by $a(y)$, and the leading coefficient of x in $H(x,y)$ is $b_{h_x}(y)$, denoted by $b(y)$. By (12), the coefficient of x^n in A is

$$\sum_{i=i_0}^{f_x} a_i(z)a(y)^i b(y)^{f_x-i} = b(y)^{f_x} F\left(\frac{a(y)}{b(y)}, z\right) \tag{15}$$

and the coefficient of x^n in C is

$$\sum_{j=j_0}^{h_x} b_j(z)a(y)^j b(y)^{h_x-j} = b(y)^{h_x} H\left(\frac{a(y)}{b(y)}, z\right). \tag{16}$$

Since $\gcd(F(x,y), H(x,y)) = 1$, the polynomials in (15) and (16) can not both equal to zero. Hence A or C has degree n in x. In either case, it follows from (9) and (10) that $n = h_x^2 \le h_x$. So $f_x = h_x \le 1$.

Finally assume $f_x < h_x$. We have from (13) that

$$\deg_x A = i_0 f_x + h_x(f_x - i_0). \tag{17}$$

This and (9) imply that

$$i_0 f_x + h_x(f_x - i_0) \le f_x.$$

Since $h_x > f_x$, it follows that $f_x - i_0 = 0$ and thus $f_x^2 \le f_x$. Hence $f_x \le 1$ and $\deg_x A = \deg_x D = f_x$. Considering the degrees of x in (8), we have

$$h_x^2 + \deg_x B = \deg_x C + h_x f_x.$$

By (9), $\deg_x B \ge \deg_x C$, thus $h_x f_x \ge h_x^2$. But $h_x > f_x \ge 0$, so $f_x \ge h_x$, impossible as $f_x < h_x$ by our assumption in this case. ☐

Lemma 8. *Suppose*

$$R(x,y) = \frac{u_1xy + u_2x + u_3y + u_4}{v_1xy + v_2x + v_3y + v_4} \in \mathbb{F}(x,y) \tag{18}$$

is associative. Then $R(x,y)$ is equivalent to a polynomial or one of the following

$$\frac{u_1xy + u_2x + u_3y + u_4}{v_2x + v_3y}, \qquad \frac{u_1xy + u_2x + u_3y + u_4}{x - y + 1}. \tag{19}$$

(With possibly different values of u's and v's.)

Proof. We first assume that $v_1 \neq 0$. Let $f(x) = a + 1/x$. Then $f^{-1}(x) = 1/(x-a)$. Note that

$$\begin{aligned} R(f(x), f(y)) &= \frac{u_1(a+1/x)(a+1/y) + u_2(a+1/x) + u_3(a+1/y) + u_4}{v_1(a+1/x)(a+1/y) + v_2(a+1/x) + v_3(a+1/y) + v_4} \\ &= \frac{u_1(a^2xy + a(x+y) + 1) + u_2y(ax+1) + u_3x(ay+1) + u_4xy}{v_1(a^2xy + a(x+y) + 1) + v_2y(ax+1) + v_3x(ay+1) + v_4xy} \\ &= \frac{(u_1a^2 + u_2a + u_3a + u_4)xy + (u_1a + u_3)x + (u_1a + u_2)y + u_1}{(v_1a^2 + v_2a + v_3a + v_4)xy + (v_1a + v_3)x + (v_1a + v_2)y + v_1}, \end{aligned}$$

and so

$$f^{-1}\left(R(f(x), f(y))\right) = \frac{(v_1a^2 + v_2a + v_3a + v_4)xy + (v_1a + v_3)x + (v_1a + v_2)y + v_1}{(-v_1a^3 + (u_1 - v_2 - v_3)a^2 + (u_2 + u_3 - v_4)a + u_4)xy + E}$$

where

$$E = (-v_1a^2 + (u_1 - v_3)a + u_3)x + (-v_1a^2 + (u_1 - v_2)a + u_2)y + u_1 - av_1. \tag{20}$$

Taking a to be a root of the polynomial

$$-v_1X^3 + (u_1 - v_2 - v_3)X^2 + (u_2 + u_3 - v_4)X + u_4 \tag{21}$$

will yield a new R with $v_1 = 0$.

Hence we may assume that $v_1 = 0$ for our original R. We may further assume that $v_4 \neq 0$, otherwise it is of the first type in (19) already. Let $f(x) = ax + b$. Then

$$\begin{aligned}
R(f(x), f(y)) &= \frac{u_1(ax+b)(ay+b) + u_2(ax+b) + u_3(ay+b) + u_4}{v_2(ax+b) + v_3(ay+b) + v_4} \\
&= \frac{u_1a^2xy + \cdots}{v_2ax + v_3ay + (v_2+v_3)b + v_4}, \\
f^{-1}\left(R(f(x), f(y))\right) &= \frac{u_1axy + \cdots}{v_2ax + v_3ay + (v_2+v_3)b + v_4}.
\end{aligned}$$

If $v_2 + v_3 \neq 0$ then we can take $b = -v_4/(v_2 + v_3)$ and R is equivalent to the first type in (19). Suppose $v_2 + v_3 = 0$, i.e. $v_2 = -v_3$. If $v_2 = 0$ then R is already a polynomial. But if $v_2 \neq 0$ then we can take $a = v_4/v_2$ and R is equivalent to the second type in (19). □

Lemma 9. *Let R be associative of the first type in (19). If R is not a polynomial then R must be of the form*

$$\frac{xy + u}{x + y}$$

which is indeed associative.

Proof. Assume $v_2 \neq 0$; the proof is similar if $v_3 \neq 0$. Then R can be rewritten as

$$R = \frac{u_1xy + u_2x + u_3y + u_4}{x + vy} = \frac{F(x,y)}{H(x,y)}$$

where $F(x, y) = u_1xy + u_2x + u_3y + u_4$ and $H(x, y) = x + vy$. Since R is not a polynomial, $\gcd(F(x, y), H(x, y)) = 1$. Note that

$$\begin{aligned}
R(R(x,y), z) &= \frac{u_1zF(x,y) + u_2F(x,y) + u_3zH(x,y) + u_4H(x,y)}{F(x,y) + vH(x,y)} = \frac{A}{C} \\
R(x, R(y,z)) &= \frac{u_1xF(y,z) + u_2xH(y,z) + u_3F(y,z) + u_4H(y,z)}{xH(y,z) + vF(y,z)} = \frac{B}{D}
\end{aligned}$$

where A, B, C, D are the corresponding numerators and denominators. By Lemma 6, $\gcd(A, C) = 1$ and $\gcd(B, D) = 1$. So the equation $A/C = B/D$ implies that $C = dD$ for some constant $d \in \mathbb{F}$. If $v = 0$ then $D = xH(y, z) = xy$ and $C = F(x, y) = x$, impossible for $C = dD$. Hence we may assume that $v \neq 0$. As

$$\begin{aligned}
C &= u_1xy + vxz + v^2yz + u_2x + u_3y + u_4 \\
D &= xy + vxz + vu_1yz + vu_2y + vu_3z + vu_4.
\end{aligned}$$

Comparing the coefficients of x, y gives $u_2 = 0$ and $u_3 = 0$. Then by the coefficients of xy, xz and yz we have

$$u_1 = d, \quad v = dv, \quad v^2 = dvu_1.$$

Since $v \neq 0$, it follows that $d = u_1 = 1$ and hence $v = 1$. Therefore R is the required form.

Lemma 10. *Let R be associative of the second type in (19). If R is not a polynomial then R must be of the form*

$$\frac{xy + u}{x + y + 1}$$

which is indeed associative.

Proof. Let $F(x, y) = u_1xy + u_2x + u_3y + u_4$ and $H(x, y) = x - y + 1$. Since R is not a polynomial, $\gcd(F(x, y), H(x, y) = 1$. Similar to above,

$$\begin{aligned} R(R(x, y), z)) &= \frac{u_1zF(x, y) + u_2F(x, y) + u_3zH(x, y) + u_4H(x, y)}{F(x, y) - zH(x, y) + H(x, y)} = \frac{A}{C} \\ R(x, R(y, z)) &= \frac{u_1xF(y, z) + u_2xH(y, z) + u_3F(y, z) + u_4H(y, z)}{xH(y, z) - F(y, z) + H(y, z)} = \frac{B}{D}. \end{aligned}$$

By Lemma 6, $\gcd(A, C) = \gcd(B, D) = 1$. So $C = dD$ for some $d \in \mathbb{F}$. As

$$\begin{aligned} C &= u_1xy - xz + yz + (u_2 + 1)x + (u_3 - 1)y - z + u_4 + 1 \\ D &= xy - xz - u_1yz + x + (1 - u_2)y - (u_3 + 1)z - u_4 + 1, \end{aligned}$$

comparing the coefficients of xy, xz, yz gives

$$u_1 = d, \quad -1 = -d, \quad 1 = -du_1.$$

So $d = u_1 = 1$ and $2 = 0$, which means that the characteristic of $\mathbb{F}$ is 2. Now the coefficients of x, y give

$$u_2 + 1 = d, \quad u_3 - 1 = d(-u_2 + 1).$$

So $u_2 = u_3 = 0$ and R is the required type. □

Lemma 11. *Every rational function of the form*

$$\frac{xy + u}{x + y + v}$$

is equivalent to xy.

Proof. Let a, b be the two roots of the polynomial $X^2 - vX - u$ and $f = (x + a)/(x + b)$. Then one can check directly that the function in the lemma is equal to $f^{-1}(f(x) \cdot f(y))$. □

Proof of Theorem 1. Let $R \in \mathbb{F}(x, y)$ be any associative rational function. By Lemma 7, R must be of the form (18), hence equivalent to a polynomial or one of the two types in (19). In the latter case, by Lemmas 9 and 10, R must be of the form in Lemma 11, which is equivalent to xy thus to $x+y+xy$. Therefore R is equivalent to $x, y, x + y$ or $x + y + xy$ by Lemma 5.

3 Determining the groups

This section is devoted to proving Theorem 2.

Before we proceed to the proof, we make some general remarks. Suppose that $f : S \longrightarrow G$ is any bijection of sets and $(G, \cdot)$ is a group. Then one can define a group law on S as follows:

$$a \cdot b = f^{-1}(f(a) \cdot f(b)), \quad \forall a, b \in S. \tag{22}$$

Here f is automatically a group isomorphism from S to G. On the other hand, if S already has a group law that satisfies (22) for some bijection $f : S \longrightarrow G$ then f is a group isomorphism from S to G.

Now let S be an infinite subset of $\mathbb{F} \cup \{\infty\}$ where $\mathbb{F}$ is a field. Suppose that a group law on S is defined by a rational function $R \in \mathbb{F}(x, y)$, i.e. $a \cdot b = R(a, b)$ for $a, b \in S$. By Theorem 1, R is equivalent to $x + y$ or xy via some linear fractional map $f \in \overline{\mathbb{F}}(x)$, i.e. $R(x, y) = f^{-1}(f(x) + f(y))$ or $f^{-1}(f(x)f(y))$. Hence

$$a \cdot b = R(a, b) = f^{-1}(f(a) + f(b)), \quad \forall a, b \in S$$

or

$$a \cdot b = R(a, b) = f^{-1}(f(a)f(b)), \quad \forall a, b \in S.$$

Let $G = f(S) \subseteq \overline{\mathbb{F}}$. Then G is an additive subgroup of $\overline{\mathbb{F}}$ in the first case or a multiplicative group of $\overline{\mathbb{F}}$ in the second case. Obviously G is in the same extension of $\mathbb{F}$ as that defining the map f. From the polynomial (21), it seems that one may need to go to a cubic extension of $\mathbb{F}$ to find the coefficients of f. We prove that this is not the case and indeed a quadratic extension of $\mathbb{F}$ suffices. Consequently S is isomorphic to a subgroup (additive or multiplicative) of a quadratic extension of $\mathbb{F}$, hence Theorem 2 is proved.

It remains to prove that the equivalence of R can be realized by a linear fractional map over a quadratic extension of $\mathbb{F}$. This is done by the next two lemmas.

Lemma 12. *Suppose $R \in \mathbb{F}(x, y)$ is equivalent to $x + y$. Then the equivalence can be defined by a linear fractional map over $\mathbb{F}$, except when the characteristic of $\mathbb{F}$ is 2 and $\mathbb{F}$ is not perfect in which case a required linear fractional map can be found in a quadratic extension of $\mathbb{F}$.*

Proof. Suppose $R(x, y) = f^{-1}(f(x) + f(y))$ where $f = (ax + b)/(cx + d)$ with $a, b, c, d \in \overline{\mathbb{F}}$ and $ad - bc \neq 0$. First assume that $c = 0$. Then f can be rewritten as $f = ax + b$ for some $a, b \in \overline{\mathbb{F}}$ with $a \neq 0$. In this case,

$$R = f^{-1}(f(x) + f(y)) = x + y + b/a.$$

As $R \in \mathbb{F}(x, y)$, we have $b/a \in \mathbb{F}$. Let $g = x + b/a$. Then we have $R = g^{-1}(g(x) + g(y))$ so the equivalence is defined by a linear map over $\mathbb{F}$.

Now assume $c \neq 0$. Then f can be rewritten as as $f = (ax+b)/(x+d)$ for some $a, b, d \in \overline{\mathbb{F}}$ with $ad - b \neq 0$. Note that

$$R = f^{-1}(f(x)+f(y)) = \frac{(b-2ad)xy - ad^2(x+y) - bd^2}{axy + b(x+y) + 2bd - ad^2}.$$

There are two cases depending on $a = 0$ or not. Suppose $a = 0$. Then $b \neq 0$ (as $ad - b \neq 0$), and

$$R = \frac{bxy - bd^2}{b(x+y) + 2bd} = \frac{xy - d^2}{x+y+2d} \in \mathbb{F}(x, y).$$

So $2d, d^2 \in \mathbb{F}$. If $\mathrm{char}(\mathbb{F}) \neq 2$ then $2d \in \mathbb{F}$ implies that $d \in F$. If $\mathrm{char}(\mathbb{F}) = 2$ and $\mathbb{F}$ is perfect then $d^2 \in \mathbb{F}$ implies that $d \in \mathbb{F}$. In both cases f is defined over $\mathbb{F}$. If $\mathrm{char}(\mathbb{F}) = 2$ but $\mathbb{F}$ is not perfect then d may not be in $\mathbb{F}$ but always in a quadratic extension of $\mathbb{F}$; See Example 3 for an instance.

Suppose $a \neq 0$. Then R can be rewritten as

$$R = \frac{(\frac{b}{a} - 2d)xy - d^2(x+y) - \frac{b}{a}d^2}{xy + \frac{b}{a}(x+y) + 2d\frac{b}{a} - d^2} \in \mathbb{F}(x, y).$$

Hence

$$\frac{b}{a},\ \frac{b}{a} - 2d,\ d^2 \in \mathbb{F}.$$

This implies that $d \in \mathbb{F}$ except when $\mathrm{char}(\mathbb{F}) = 2$ and $\mathbb{F}$ is not perfect. The proof is finished by noting that $R = g^{-1}(g(x) + g(y))$ where $g = (x + b/a)/(x+d)$. $\square$

Lemma 13. *Suppose that $R \in \mathbb{F}(x, y)$ is equivalent to xy. Then the equivalence is defined by a linear fractional map over a quadratic extension of $\mathbb{F}$.*

Proof. Suppose $R(x, y) = f^{-1}(f(x)f(y))$ where $f = (ax+b)/(cx+d)$ with $a, b, c, d \in \overline{\mathbb{F}}$ and $ad - bc \neq 0$. If $c = 0$ then f is of the form $f = ax + b$ with $a \neq 0$ and

$$R = f^{-1}(f(x)f(y)) = axy + b(x+y) + (b^2 - b)/a.$$

Since $R \in \mathbb{F}(x, y)$, we have $a, b, \in \mathbb{F}$, so $f \in \mathbb{F}(x)$.

Hence we may assume that $c \neq 0$. Then f can be rewritten as $f = (ax+b)/(x+d)$ for some $a, b, d \in \overline{\mathbb{F}}$ with $ad - b \neq 0$. One can check directly that

$$R = f^{-1}(f(x)f(y)) = \frac{(b - a^2d)xy - bd(a-1)(x+y) - bd(b-d)}{(a^2-a)xy + a(b-d)(x+y) + b^2 - ad^2}. \tag{23}$$

If $a = 0$ then we see easily that $b, d \in \mathbb{F}$ and so f is defined over $\mathbb{F}$. If $a = 1$ then

$$R = \frac{xy - bd}{x+y+b+d} \in \mathbb{F}(x, y).$$

Hence $bd, b+d \in \mathbb{F}$ and thus b, d are in a quadratic extension of $\mathbb{F}$. We may now assume that $a \neq 0$ or 1. Then (23) may be rewritten as

$$R = \frac{\frac{b-a^2d}{a^2-a}xy - \frac{bd}{a}(x+y) - \frac{bd}{a}\frac{b-d}{a-1}}{xy + \frac{b-d}{a-1}(x+y) + \frac{b^2-ad^2}{a^2-a}}.$$

Since $R \in \mathbb{F}(x,y)$, we have

$$\frac{b-d}{a-1}, \ \frac{bd}{a}, \ \frac{b-a^2d}{a^2-a}, \ \frac{b^2-ad^2}{a^2-a} \in \mathbb{F}.$$

If $a=-1$ then $bd, b-d \in \mathbb{F}$ so b, d are in a quadratic extension of $\mathbb{F}$. Hence we may further assume that $a \neq -1$. Let

$$b - d = c_1(a-1), \quad a^2d - b = c_2(a^2 - a)$$

where $c_1, c_2 \in \mathbb{F}$. Then

$$b = \frac{a}{a+1}(c_1a + c_2), \quad d = \frac{1}{a+1}(c_1 + c_2a).$$

Note

$$\frac{bd}{a} = \frac{1}{(a+1)^2}(c_1a+c_2)(c_1+c_2a) \in \mathbb{F}.$$

Let $c_3 = bd/a \in \mathbb{F}$. Then we see that a is a root of the quadratic polynomial

$$c_3(X+1)^2 - (c_1X + c_2)(c_1 + c_2X) \in \mathbb{F}[X].$$

So a, b, d all lie in a quadratic extension of $\mathbb{F}$. This proves that f is always defined over a quadratic extension of $\mathbb{F}$. □

References

1. J. V. Brawley and L. Carlitz, "Irreducibles and the composed product for polynomials over a finite field", *Discrete Math.* **65** (1987), 115-139.
2. J. V. Brawley, S. Gao and D. Mills, "Computing composed products of polynomials," in *Finite Fields: Theory, Applications, and Algorithms* (Waterloo, ON, 1997, R. C. Mullin and G. L. Mullen, Ed.), 1–15, *Contemporary Mathematics*, **225**, Amer. Math. Soc., Providence, RI, 1999.
3. L. Childs, D. Moss and J. Sauerberg, "Dimension one polynomial formal groups", p. 11–19, in *Hopf Algebras, Polynomial Formal Groups, and Raynaud Orders* (L. N. Childs, C. Greither, D. J. Moss, J. Sauerberg and K. Zimmermann), Memoirs AMS, no. 651, 1998.
4. M. Hazewinkel, *Formal Groups and Applications*, Academic Press, New York, 1978.
5. A. J. Menezes, I. F. Blake, X. Gao, R. C. Mullin, S.A. Vanstone and T. Yaghoobian, *Applications of Finite Fields*, Kluwer Academic Publishers, Boston-Dordrecht-Lancaster, 1993.
6. J. H. Silverman, *The Arithmetic of Elliptic Curves*, Springer-Verlag, New York, 1986.