Term Rewrite Rules for Finite Fields

Stanley Burris* and John Lawrence†

Abstract

Let $\mathbf{F}_1, \ldots, \mathbf{F}_k$ be finite fields with distinct characteristics. We give a finite set of equations which axiomatize the equational theory of $\mathbf{F}_1, \ldots, \mathbf{F}_k$ and then use these axioms to find a finite AC-term rewrite system which is complete for this theory. In particular this gives finite complete AC-term rewrite systems for many instances of $x^m \approx x$ rings. However there are m for which no such term rewrite system exists — the smallest is $m = 22$.

A popular test problem for automated theorem proving has been the derivation of $xy \approx yx$ from ring axioms plus $x^m \approx x$ $(m > 1)$. At present this has been achieved for $m = 2, 3, 4, 6$ using versions of the Knuth-Bendix algorithm. Zhang [6] has given a polynomial gcd algorithm which can be used to prove commutativity for about three-fourths of the even exponents m (by reducing such cases to $x^2 \approx x$ or $x^4 \approx x$).

A related problem is to find a complete set of AC-term rewrite rules for $x^m \approx x$ rings. This has been solved for $m = 2$ by Hsiang [2] and for $m = 3$ by Stickel [5]. As a corollary to our theorem we find such rewrite rules for many $m \geq 2$ (about 94% of the m below 100,000). Our result also covers Nipkow's [4] AC-term rewrite rules for rings satisfying $x^p \approx x$ and $p \approx 0$, where p is a prime.

1 From Equations to Rewrite Rules.

Throughout this section let $\mathbf{F}_1, \ldots, \mathbf{F}_k$ be a fixed list of finite fields with distinct characteristics $p_1, \ldots, p_k$. The set of equations true of the fields $\mathbf{F}_i$

*Research supported by a grant from NSERC

†Research supported by a grant from NSERC

$(1 \le i \le k)$ is the *equational theory* of $\mathbf{F}_1, \ldots, \mathbf{F}_k$. Assume the respective sizes of these fields are $q_1 > \cdots > q_k$, let $c = p_1 \cdots p_k$, and let n be such that $n - 1$ is the least common multiple of $q_1 - 1, \ldots, q_k - 1$.

LEMMA 1.1 The equational theory of $\mathbf{F}_1, \ldots, \mathbf{F}_k$ is axiomatized by the axioms for rings plus

$$\begin{cases} x^n \approx x \\ c \approx 0 \\ \dfrac{c}{p_i} x^{q_i} \approx \dfrac{c}{p_i} x \qquad (1 \le i \le k). \end{cases} \tag{1}$$

Proof. Clearly these equations are satisfied by $\mathbf{F}_1, \ldots, \mathbf{F}_k$ as $\mathbf{F}_i$ satisfies $p_i \approx 0$ and $x^{q_i} \approx x$ $(1 \le i \le k)$. Now if $\mathbf{F}$ is a finite field satisfying these equations then the characteristic of $\mathbf{F}$ divides c since $c \approx 0$ holds, so the characteristic is some p_i $(1 \le i \le k)$. Next, $\mathbf{F}$ satisfies $\frac{c}{p_i} x^{q_i} \approx \frac{c}{p_i} x$ implies $\mathbf{F}$ satisfies $x^{q_i} \approx x$, so $\mathbf{F}$ embeds in $\mathbf{F}_i$. Thus we have proved that the fields satisfying (1) are the subfields of $\mathbf{F}_1, \ldots, \mathbf{F}_k$. By a result of Jacobson [3] we know that a ring equation follows from (1) iff it holds on the finite fields satisfying (1). Consequently we have found axioms for the equational theory of $\mathbf{F}_1, \ldots, \mathbf{F}_k$. ■

In the following when we refer to polynomial we mean an element of $\mathbf{Z}[X]$, where X is an infinite set of *commuting* variables. Since we are working in the language $+, \cdot, -, 0, 1$ of rings, we really mean that for $n > 0$ the constant monomial n is an abbreviation for $1 + \cdots + 1$, and the expression nA means $A + \cdots + A$ (in both cases the number of summands is n). Given a monomial $M = x_1^{s_1} \cdots x_l^{s_l}$ we define $\mathrm{vdeg}(M)$ to be the maximum of $s_1, \ldots, s_l$. If P is a polynomial define $\mathrm{vdeg}(P)$ to be the largest $\mathrm{vdeg}(M)$ for M a monomial in P.

LEMMA 1.2 If F is a finite field of characteristic p and size q, and if A and B are two polynomials such that vdeg(A) and vdeg(B) are less than q, then $A \approx B$ holds on $\mathbf{F}$ iff the corresponding coefficients are congruent modulo p.

Proof. Since $\mathbf{F}$ satisfies $p \approx 0$, the direction $(\Leftarrow)$ is clear. The direction $(\Rightarrow)$ is a straight forward induction on the number of variables appearing in the equation. For one variable the result follows from the fact that $x^q - x$ is the minimal polynomial of $\mathbf{F}$. Now for the induction step suppose that the claim is true for any polynomials A and B with fewer than n variables. Let $A(x_1, \ldots, x_n)$ and $B(x_1, \ldots, x_n)$ be equal on $\mathbf{F}$, and suppose both have vdeg less than q. The corresponding coefficients are congruent modulo p iff the coefficients of $A - B$ (considered as a polynomial) are congruent to 0 modulo p. Writing $A - B$ in the form

$$\sum_{i=0}^{q-1} C_i(x_1, \ldots, x_{n-1}) x_n^i \tag{2}$$

we see that it suffices to show that each of the polynomials $C_i(x_1, \ldots, x_{n-1})$ vanishes on $\mathbf{F}$ — for then by induction all the coefficients of each C_i will be congruent to 0 modulo p, and hence the same is true of the coefficients of $A - B$. We proceed by assuming some C_i does not vanish on $\mathbf{F}$. Then one can choose elements $\vec{a}$ from $\mathbf{F}$ such that not all C_i vanish at $\vec{a}$. Substituting this into (2) gives the one-variable polynomial

$$\sum_{i=0}^{q-1} C_i(a_1, \ldots, a_{n-1}) x_n^i$$

with coefficients in $\mathbf{F}$, and not all coefficients are 0. As this is a non-zero polynomial of degree at most $q - 1$ it follows that it can have at most $q - 1$ roots in $\mathbf{F}$, and consequently cannot vanish on $\mathbf{F}$. With this contradiction we have proved the lemma. ■

Now we turn to the rewrite rules. Define

$$a_1 = 1, a_i = p_1 \cdots p_{i-1} \qquad (2 \le i \le k).$$

Since the g.c.d. of $\frac{c}{p_1}, \ldots, \frac{c}{p_k}$ is 1, we can find integers n_i such that $\sum_{i=1}^k n_i \frac{c}{p_i} = 1$. Now define a_{ij} and b_{ij} $(1 \le j \le i \le k)$ by

$$a_{ij} = [a_j n_i \frac{c}{p_i}]_c$$

(where $[\ ]_c$ means to reduce modulo c to a non-negative value less than c),

$$b_{ij} = q_j - q_i + 1.$$

THEOREM 1.3 The following AC-term rewrite rules are complete for the equational theory of $\mathbf{F}_1, \ldots, \mathbf{F}_k$:

R0:	$x + 0 \longrightarrow x$		
R1:	$x \cdot 1 \longrightarrow x$	R2:	$x \cdot 0 \longrightarrow 0$
R3:	$x + (-x) \longrightarrow 0$	R4:	$x \cdot (y+z) \longrightarrow x \cdot y + x \cdot z$
R5:	$-1 \longrightarrow c - 1$	R6:	$-x \longrightarrow (c-1)x$
R7:	$c \longrightarrow 0$	R8	$cx \longrightarrow 0$
S_j :	$a_j x^{q_j} \longrightarrow \sum_{i=j}^{k} a_{ij} x^{b_{ij}} \qquad (1 \leq j \leq k)$		
T_j :	$a_j x^{q_j} v \longrightarrow \sum_{i=j}^{k} a_{ij} x^{b_{ij}} v \qquad (1 \leq j \leq k)$,		

For those j such that $q_i - 1 | q_j - 1$ for $i \geq j$ we can replace S_j and T_j by the following particularly simple rewrite rules:

$$\mathrm{S}'_j : \quad a_j x^{q_j} \longrightarrow a_j x$$
$$\mathrm{T}'_j : \quad a_j x^{q_j} v \longrightarrow a_j x v.$$

This always applies to $j = k$.

Proof. Let $\mathcal{F} = \{\mathbf{F}_1, \ldots, \mathbf{F}_k\}$. From

$$1 = \sum_{i=1}^{k} n_i \frac{c}{p_i}$$

follows

$$a_j = \sum_{i=1}^{k} a_j n_i \frac{c}{p_i},$$

and thus

$$a_j \equiv \sum_{i=j}^{k} a_j n_i \frac{c}{p_i} \qquad (\mathrm{mod}\ c).$$

This leads to

$$\mathcal{F} \models a_j x^{q_j} \approx \sum_{i=j}^{k} a_j n_i \frac{c}{p_i} x^{q_j}.$$

Now from

$$\mathbf{F}_i \models x^{q_i} \approx x, \qquad \mathbf{F}_{i'} \models \frac{c}{p_i} \approx 0 \qquad (i' \neq i)$$

we have

$$\mathcal{F} \models \frac{c}{p_i} x^{q_i} \approx \frac{c}{p_i} x,$$

and thus

$$\mathcal{F} \models \frac{c}{p_i} x^{q_j} \approx \frac{c}{p_i} x^{b_{ij}} \qquad (1 \leq j \leq i \leq k).$$

Consequently

$$a_j x^{q_j} \approx \sum_{i=j}^{k} a_j n_i \frac{c}{p_i} x^{b_{ij}}$$

holds in $\mathcal{F}$; and hence so does the equation

$$a_j x^{q_j} \approx \sum_{i=j}^{k} a_{ij} x^{b_{ij}}.$$

Thus the proposed rewrite rules correspond to valid equations in $\mathbf{F}_1, \ldots, \mathbf{F}_k$. Now observe that the rules in S_j and T_j are degree decreasing. Consequently the proposed set of rewrite rules is terminating, and any term $t(x_1, \ldots, x_n)$ can be reduced to a normal form

$$A_1(x_1, \ldots, x_n) + \cdots + A_k(x_1, \ldots, x_n)$$

where each A_i is a polynomial; and furthermore if $A_i \neq 0$

i. the monomials M in A_i satisfy

$$q_{i+1} \leq \mathrm{vdeg}(M) < q_i \qquad (1 \leq i \leq k-1)$$

$$0 \leq \mathrm{vdeg}(M) < q_k \qquad \text{if } i = k;$$

ii. the coefficients d in A_i satisfy

$$1 \leq d < a_{i+1} = p_1 \cdots p_i \qquad (1 \leq i \leq k-1)$$

$$1 \leq d < c = p_1 \cdots p_k \qquad (i = k).$$

It only remains to show that two AC-distinct normal forms are not equivalent on $\mathbf{F}_1, \ldots, \mathbf{F}_k$. So suppose $\sum A_i$ and $\sum B_i$ are two normal forms with $\sum A_i \approx \sum B_i$ true of $\mathbf{F}_1, \ldots, \mathbf{F}_k$. Clearly the vdeg's of $\sum A_i$ and $\sum B_i$ are less than q_1. As the coefficients in A_1 and B_1 are non-negative and less than p_1 it follows from Lemma 1.2, using $\mathbf{F}_1$, that A_1 and B_1 are AC-equivalent. Thus $\sum_{i\geq 2} A_i \approx \sum_{i\geq 2} B_i$ holds on $\mathbf{F}_1, \dots, \mathbf{F}_k$. As the coefficients in A_2 and B_2 are non-negative and less than $p_1 \cdot p_2$ it follows from Lemma 1.2, using $\mathbf{F}_1$ and $\mathbf{F}_2$, that A_2 and B_2 are AC-equivalent. Continuing we obtain $\sum A_i$ and $\sum B_i$ are AC-equivalent.

To see that one can use the simple forms S'_j and T'_j when $q_{i-1}|q_{j-1}$ for $i \geq j$, we only need to observe that $a_j x^{q_j} \approx a_j x$ holds on the fields $\mathbf{F}_1, \dots, \mathbf{F}_{j-1}$ because their characteristics divide a_j, and on the fields $\mathbf{F}_j, \dots, \mathbf{F}_k$ because they satisfy $x^{q_j} \approx x$ (since $q_{i-1}|q_{j-1}$ implies $x^{q_i} - x | x^{q_j} - x$ in $\mathbf{Z}[x]$). ■

Thus we have not only found a finite and complete AC-term rewriting system for the equational theory of $\mathbf{F}_1, \dots, \mathbf{F}_k$, but we have also given an explicit description of the normal forms.

2 Examples.

1. Rings satisfying $x^p \approx x$ and $p \approx 0$, where p is a prime.

By Lemma 1.2 these equations, along with defining equations for rings, axiomatize the equational theory of $\mathbf{GF}(p)$. Thus $k = 1$, $p_1 = p$, $q_1 = p$, $c = p$, $a_1 = 1$. Hence the AC-term rewrite rules are R0–R4 plus

$-1 \longrightarrow p-1$	$-x \longrightarrow (p-1)x$
$p \longrightarrow 0$	$px \longrightarrow 0$
$x^p \longrightarrow x$	$x^p v \longrightarrow xv.$

The equational theory of $x^m \approx x$ rings, for $m > 1$, is given by the equational theory of the finite fields $\mathbf{F}$ satisfying $x^m \approx x$, i.e., by those finite fields such that $|\mathrm{F}| - 1$ divides $m - 1$, where $|\mathbf{F}|$ is the cardinality of $\mathbf{F}$. (As noted by Zhang [6], if m is even only one characteristic appears, namely 2; so for even m we need the fields $\mathbf{GF}(2^k)$ such that $2^k - 1$ divides $m - 1$). Our rewrite rules will thus apply to those m such that for each characteristic p of a field satisfying $x^m \approx x$ there is a largest field (under embedding) of characteristic p which satisfies $x^m \approx x$. It turns out that for approximately 94% of the $m < 100,000$ we have this situation. The only exceptions for $m < 1,000$ are :

22 43 85 94 105 106 148 169 187 209 211 218 232 274 280 295 313
316 337 358 373 382 400 417 421 435 463 466 484 521 526 547
559 589 610 625 631 652 673 715 736 745 763 778 799 833 838
841 862 869 890 904 925 931 937 946 967 969 988

These are the numbers m below 1,000 for which there exists a prime p and two positive integers a and b such that both $p^a - 1$ and $p^b - 1$ divide $m-1$, but $p^{lcm(a,b)} - 1$ does not divide $m-1$. Now we give a table describing the maximal fields satisfying $x^m \approx x$ for $2 \leq m \leq 10$, and a set of rewrite rules for each case (note that with one exception we can use the rules S'_j, T'_j).

m	q_i	p_i	c	a_i
2	2	2	2	1
3	3,2	3,2	6	1,3
4	4	2	2	1
5	5,3,2	5,3,2	30	1,5,15
6	2	2	2	1
7	7,4,3	7,2,3	42	1,7,14
8	8	2	2	1
9	9,5,2	3,5,2	30	1,3,15
10	4	2	2	1

The rules R0–R4 are to be added to each of the following:

$\underline{m = 2, 6}$ **(Boolean rings)**

$-1 \longrightarrow 1$	$-x \longrightarrow x$
$2 \longrightarrow 0$	$2x \longrightarrow 0$
$x^2 \longrightarrow x$	$x^2 v \longrightarrow xv.$

$\underline{m = 3}$

$-1 \longrightarrow 5$	$-x \longrightarrow 5x$
$6 \longrightarrow 0$	$6x \longrightarrow 0$
$x^3 \longrightarrow x$	$x^3 v \longrightarrow xv$
$3x^2 \longrightarrow 3x$	$3x^2 v \longrightarrow 3xv.$

$\underline{m = 4, 10}$

$-1 \longrightarrow 1$	$-x \longrightarrow x$
$2 \longrightarrow 0$	$2x \longrightarrow 0$
$x^4 \longrightarrow x$	$x^4 v \longrightarrow xv.$

$\underline{m = 5}$

$-1 \longrightarrow 29$	$-x \longrightarrow 29x$
$30 \longrightarrow 0$	$30x \longrightarrow 0$
$x^5 \longrightarrow x$	$x^5 v \longrightarrow xv$
$5x^3 \longrightarrow 5x$	$5x^3 v \longrightarrow 5xv$
$15x^2 \longrightarrow 15x$	$15x^2 v \longrightarrow 15xv.$

$\underline{m = 7}$

We have $k = 3$, $p_1 = 7$, $p_2 = 2$, $p_3 = 3$, $q_1 = 7$, $q_2 = 4$, $q_3 = 3$ (so we can use S_1', T_1', S_2, T_2, S_3', T_3'), $c = 42$, $n_1 = -1$, $n_2 = 1$, $n_3 = -1$, $a_1 = 1$, $a_2 = 7$, $a_3 = 14$, $a_{22} = 21$, $a_{32} = 28$, $b_{22} = 1$, $b_{32} = 2$.

$-1 \longrightarrow 41$	$-x \longrightarrow 41x$
$42 \longrightarrow 0$	$42x \longrightarrow 0$
$x^7 \longrightarrow x$	$x^7 v \longrightarrow xv$
$7x^4 \longrightarrow 28x^2 + 21x$	$7x^4 v \longrightarrow (28x^2 + 21x)v$
$14x^3 \longrightarrow 14x$	$14x^3 v \longrightarrow 14xv.$

$\underline{m = 8}$

$-1 \longrightarrow 1$	$-x \longrightarrow x$
$2 \longrightarrow 0$	$2x \longrightarrow 0$
$x^8 \longrightarrow x$	$x^8 v \longrightarrow xv.$

$\underline{m = 9}$

$-1 \longrightarrow 29$	$-x \longrightarrow 29x$
$30 \longrightarrow 0$	$30x \longrightarrow 0$
$x^9 \longrightarrow x$	$x^9 v \longrightarrow xv$
$3x^5 \longrightarrow 3x$	$3x^5 v \longrightarrow 3xv$
$15x^2 \longrightarrow 15x$	$15x^2 v \longrightarrow 15xv.$

We note that the only m in the examples above for which the $q_i - 1$ are not linearly ordered by divisibility is $m = 7$; hence the need for S_2, T_2. For the other m above we use the simple rewrite rules S'_j and T'_j.

3 Concluding Remarks.

We thought we could use the methods of this paper on any finite collection of finite fields. However the obvious attempt to generalize Lemma 1.2 fails. One can see the problem quite clearly if one considers the two fields $\mathbf{GF}(4)$ and $\mathbf{GF}(8)$ (this is precisely the set of maximal fields satisfying $x^{22} \approx x$). The minimum polynomial for these two fields is the least common multiple of $x^4 + x$ and $x^8 + x$ in $\mathbf{GF}(2)[x]$, i.e., $x^{10} + x^9 + x^8 + x^3 + x^2 + x$. The single AC-rewrite rule

$$x^{10} \longrightarrow x^9 + x^8 + x^3 + x^2 + x$$

(along with R0–R4) suffices to reduce every polynomial in *one* variable to a polynomial of degree less than 10, and hence to the desired normal form as the minimum polynomial has degree 10. Now applying this rule to two variable polynomials would give 2^{10^2} distinct normal forms. However the free algebra in the variety generated by these two fields has size 2^{76}, and thus we need more AC-term rewrite rules. As there are no degree reducing one variable AC-term rewrite rules which apply to polynomials of degree less than 10 we are forced to look for two variable AC-term rewrite rules. From the fact that $(x^8 + x)(y^4 + y)$ vanishes on both fields we obtain the rule

$$x^8y^4 \longrightarrow x^8y + xy^4 + xy$$

which reduces the number of normal forms to 2^{80}. Having approached this close to the desired result, we now claim that *there is no set of AC-term rewrite rules for $x^{22} \approx x$.* Our justification hinges on the consideration of what the normal form $q(x, y)$ of the polynomial $p(x, y)$ defined to be $x^8y^2 + x^8y + x^2y$ could be. Since $p(x, y) \approx p(y, x)$ is a consequence of $x^{22} \approx x$ and since $p(x, y)$ is not AC-equivalent to $p(y, x)$ it follows that $p(x, y)$ cannot be in normal form. Suppose $q(x, y)$ is the normal form of $p(x, y)$ using some AC-term rewrite rules. Then the (monomial) degree of $q(x, y)$ is no more than 10, the degree of $p(x, y)$. Since $p(x, y) \approx q(x, y)$ must be a consequence of $x^{22} \approx x$, some calculations show that either $p(x, y)$ or $p(y, x)$ must be a part of $q(x, y)$; but this would not be possible with a terminating set of AC-term rewrite rules.

Generalizing this we suspect that one cannot find AC-term rewrite rules for $x^n \approx x$ if there is a characteristic p such that there is no maximum field of characteristic p satisfying $x^n \approx x$.

References

1. Arens, R.F., and Kaplansky, I., *Topological representations of algebras.* Trans. Amer. Math. Soc. **63** (1948), 457–481.
2. Hsiang, J. *Topics in automated theorem proving and program generation.* PhD Thesis, Report R-82-1113. Dept of Computer Science, University of Illinois, Urbana, IL., 1982.
3. Jacobson, N., *Structure theory for algebraic algebras of bounded degree.* Ann. of Math. **46** (1945), 695–707.
4. Nipkow, T., *Unification in primal algebras, their powers and their varieties.* J. Assoc. Comp. Mach. **37** (1990), 742–776.
5. Stickel, M.E. *A case study of theorem proving by the Knuth-Bendix method: discovering that $x^3 \approx x$ implies ring commutativity.* 7^{th} International Conf. on Automated Deduction, Napa, California. Lecture Notes in Computer Science **170** (1984), 248–258.
6. Zhang, H. *Automated proof of ring commutativity problems by algebraic methods.* J. Symbolic Computation **9** (1990), 423–427.

Department of Pure Mathematics
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

(e-mail: snburris@thoralf.waterloo.edu)
(e-mail: snburris@thoralf.uwaterloo.ca)