SYMMETRIC FUNCTIONS AND THE PHASE PROBLEM IN CRYSTALLOGRAPHY

J. BUHLER AND Z. REICHSTEIN

ABSTRACT. The calculation of crystal structure from X-ray diffraction data requires that the phases of the "structure factors" (Fourier coefficients) determined by scattering be deduced from the absolute values of those structure factors. Motivated by a question of Herbert Hauptman, we consider the problem of determining phases by direct algebraic means in the case of crystal structures with n equal atoms in the unit cell, with n small. We rephrase the problem as a question about multiplicative invariants for a particular finite group action. We show that the absolute values form a generating set for the field of invariants of this action, and consider the problem of making this theorem constructive and practical; the most promising approach for deriving explicit formulas uses SAGBI bases.

1. INTRODUCTION

If the unit cell of a crystal has n atoms, located at positions $\mathbf{r}_j$, $1 \leq j \leq n$, then the structure factor associated to a reciprocal lattice vector $\mathbf{v}$ is

$$E_{\mathbf{v}} := \sum_{j=1}^{n} a_j \exp(2\pi i \mathbf{v} \cdot \mathbf{r}_j)\,,$$

where the a_j are the scattering amplitudes determined by the electron charge distribution in the j-th atom. This is, in effect, a Fourier transform coefficient, and the structure of the crystal can be determined from the $E_{\mathbf{v}}$ by an inverse Fourier transform. However, in standard diffraction experiments, it is impossible to measure the $E_{\mathbf{v}}$ – only their absolute values are observable. The "phase problem" of crystallography is to determine the phases of the $E_{\mathbf{v}}$ given magnitudes $|E_{\mathbf{v}'}|$ for sufficiently many $\mathbf{v}'$; this problem is fundamental in the subject, and has received considerable attention ([Gia], [Ha2]). The problem of retrieving phase information from absolute values, together with other physical constraints, occurs in several other areas of physics, astronomy, and engineering.

1991 *Mathematics Subject Classification.* 05E05, 13A50, 13P99, 20C10.

Key words and phrases. Crystallography, structure factor, phase problem, symmetric function, group action, field of invariants, SAGBI basis, algorithmic computation, multiplicative invariant, rational invariant field.

Z. Reichstein was partially supported by NSF grant DMS-901675 and by an NSERC research grant.

In the crystallographic context, the phases can be determined in principle only up to an additive constant, which is equivalent, via a Fourier transform, to the indeterminacy of the origin of the crystal. It is natural to consider "structure invariants," which are multiplicative combinations of structure factors that are invariant under change of origin, i.e., additive translation of the phase. In addition, structure invariants play an important role in commonly used stochastic methods for phase retrieval. The function

$$E_{\mathbf{v_1}} E_{\mathbf{v_2}} \dots E_{\mathbf{v_m}} \tag{1}$$

is easily seen to be a structure invariant when the reciprocal lattice vectors $\mathbf{v_i}$ sum to zero. The most common case is $m = 3$, i.e., the "triplet-structure invariant"

$$E_{\mathbf{v_1}} E_{\mathbf{v_2}} E_{-\mathbf{v_1}-\mathbf{v_2}}.$$

Exact formulas for phases of triplet structure invariants are known in terms of magnitudes for $n = 1, 2, 3$ [Ha1]. Herbert Hauptman asked one of us for a formula for arbitrary n, and the purpose of this paper is to show that, at least in the case in which all atoms in the crystal have the same a_j, such a formula exists, and to explore techniques for finding such formulas explicitly.

In this paper we shall only consider crystals with equal atoms (or equal polyatomic clumps); we will set the identical scattering factors a_j equal to 1. In addition, it is convenient to assume that the space group is the most basic group P1 (isomorphic to the group $\mathbb{Z}^3$ of translations); see [Ha1, Appendix 1] for a description of how to generalize to arbitrary space groups.

Under these assumptions the triplet phase determination problem can be converted to a question about multiplicative invariant functions. Somewhat to our surprise, this question seems to be new.

To express the triplet phase problem to a problem in symmetric functions, start by noting that the phase $\phi_{\mathbf{v_1},\mathbf{v_2}}$ of the triplet structure factor $E_{\mathbf{v_1}} E_{\mathbf{v_2}} E_{-\mathbf{v_1}-\mathbf{v_2}}$ satisfies

$$E_{\mathbf{v_1}} E_{\mathbf{v_2}} E_{-\mathbf{v_1}-\mathbf{v_2}} = |E_{\mathbf{v_1}}|\ |E_{\mathbf{v_2}}|\ |E_{-\mathbf{v_2}-\mathbf{v_2}}| \exp(i\phi_{\mathbf{v_1},\mathbf{v_2}})\,.$$

Thus the cosine of the phase can be expressed in terms of absolute values and the sum of the triplet-structure invariant and its complex conjugate.

We want to express the phase $\phi_{\mathbf{v_1},\mathbf{v_2}}$ in terms only of absolute values

$$|E_{a\mathbf{v_1}+b\mathbf{v_2}}|^2 \;=\; E_{a\mathbf{v_1}+b\mathbf{v_2}} E^*_{a\mathbf{v_1}+b\mathbf{v_2}} \;=\; E_{a\mathbf{v_1}+b\mathbf{v_2}} E_{-a\mathbf{v_1}-b\mathbf{v_2}}$$

corresponding to reciprocal lattice vectors $a\mathbf{v_1} + b\mathbf{v_2}$, where a and b are integers.

Fix $\mathbf{v_1}$ and $\mathbf{v_2}$ and let

$$\begin{aligned} x_j &= \exp(2\pi i \mathbf{v_1}\cdot\mathbf{r}_j), \qquad & y_j &= \exp(2\pi i \mathbf{v_2}\cdot\mathbf{r}_j), \quad 1 \le j \le n \\ X &= (x_1, \cdots, x_n), & Y &= (y_1, \cdots, y_n)\,. \end{aligned}$$

Then

$$E_{\mathbf{v_1}} = \sum_{j=1}^{n} x_j \quad \text{and} \quad E_{\mathbf{v_2}} = \sum_{j=1}^{n} y_j \, .$$

Using the fact that $z + z^* = 2|z|\cos(\phi)$ if z is a complex number with absolute value $|z|$ and argument ϕ, we see that we need to express

$$\begin{aligned} E_2(X,Y) &:= E_{\mathbf{v_1}} E_{\mathbf{v_2}} E_{-\mathbf{v_1}-\mathbf{v_2}} + E_{-\mathbf{v_1}} E_{-\mathbf{v_2}} E_{\mathbf{v_1}+\mathbf{v_2}} \\ &= |E_{\mathbf{v_1}}| \; |E_{\mathbf{v_2}}| \; |E_{-\mathbf{v}-\mathbf{v_2}}| \; 2\cos(\phi_{\mathbf{v_1},\mathbf{v_2}}) \\ &= \sum_{j=1}^{n} x_j \sum_{k=1}^{n} y_k \sum_{l=1}^{n} \frac{1}{x_l y_l} \; + \; \sum_{j=1}^{n} \frac{1}{x_j} \sum_{k=1}^{n} \frac{1}{y_k} \sum_{l=1}^{n} x_l y_l \\ &= \sum_{j,k,l=1}^{n} \frac{x_j y_k}{x_l y_l} \; + \; \sum_{j,k,l=1}^{n} \frac{x_l y_l}{x_j y_k} \end{aligned}$$

in terms of the magnitudes

$$\begin{aligned} q_{a,b}(x_1, \ldots, x_n, y_1, \ldots, y_n) &:= E_{\mathbf{v}} E_{-\mathbf{v}} \\ &= \sum_{j=1}^{n} x_j^a y_j^b \cdot \sum_{j=1}^{n} x_j^{-a} y_j^{-b} = \sum_{i,j=1}^{n} \frac{x_i^a y_i^b}{x_j^a y_j^b} \, , \end{aligned}$$

for suitable integers a and b; here $\mathbf{v} = a\mathbf{v_1} + b\mathbf{v_2}$ is an arbitrary reciprocal lattice vector.

We will call $q_{a,b}$ an "observable" since it is (the square of) an absolute value and therefore it is possible to observe it physically. Thus our goal is to express E_2 as a rational function of the observables $q_{a,b}$. From now on we will treat this as a question about variables x_i and y_i, ignoring the fact that they are complex numbers of absolute value one. A simple Zariski density argument shows that this does not change the underlying problem, i.e., we are not going to "miss" any identities by assuming that x_i and y_i are arbitrary complex numbers, rather than just those of absolute value one.

Example 1.1. The reader can easily verify that for $n = 2$ the triplet phase invariant is a polynomial in three observables:

$$E_2(X,Y) = 2(q_{1,0} + q_{0,1} + q_{1,1}) - 8.$$

Example 1.2. The formula for $n = 3$ is considerably more elaborate, and was discovered by Hauptman [Ha1]; it takes the form $E_2(X,Y) = N/D$, where

$$D := q_{0,1} + q_{1,0} + q_{1,1} - 3 = -3 + \sum_{i,j=1}^{3} \left(\frac{x_i}{x_j} + \frac{y_i}{y_j} + \frac{x_i y_i}{x_j y_j} \right)$$

and

$$\begin{aligned} N &:= 135 - 31D + D^2 + 2(q_{1,0}q_{0,1} + q_{1,0}q_{1,1} + q_{0,1}q_{1,1}) \\ &\quad +(q_{0,1}q_{2,1} + q_{1,0}q_{1,2} + q_{1,-1}q_{1,1}) \\ &\quad -5(q_{1,2} + q_{2,1} + q_{1,-1}) - 2(q_{0,2} + q_{2,2} + q_{2,0})\,. \end{aligned}$$

(For another formula, see Example 6.5.)

Higher order generalizations of triplet structure invariants are defined as a product of structure factors E_v where the v sum to 0. If a structure invariant is the product of m factors (1) for arbitrary m, then this gives rise, in a similar manner, to the problem of expressing

$$E_m(X_1, \cdots, X_m) := \sum_{j,j_1,\ldots,j_m=1}^{n} \Big(\frac{x_{1j_1}x_{2j_2}\ldots x_{mj_m}}{x_{1j}\ldots x_{mj}} + \frac{x_{1j}\ldots x_{mj}}{x_{1j_1}x_{2j_2}\ldots x_{mj_m}} \Big) \tag{2}$$

as a rational function in observables

$$q_{r_1,\ldots,r_m} := \sum_{i,j=1}^{n} \frac{x_{1i}^{r_1}\ldots x_{mi}^{r_m}}{x_{1j}^{r_1}\ldots x_{mj}^{r_m}}\,, \tag{3}$$

where $r_1, \ldots, r_m$ are integers.

Note that the rational function $f = E_m$ has the following invariance properties.

a: f is of weight 0 in each n-variable vector

$$X_1 = (x_{11}, \ldots, x_{1n}), \quad \ldots, \quad X_m = (x_{m1}, \ldots, x_{mn})\,.$$

That is, $f(c_1X_1, \ldots, c_mX_m) = f(X_1, \ldots, X_m)$ for any non-zero scalars $c_1, \ldots, c_m$, or, more succinctly: $f(c_jx_{ij}) = f(x_{ij})$.

b: f is self-reciprocal in the sense that it remains unchanged if every variable x_{ij} is simultaneously replaced by x_{ij}^{-1}, i.e., $f(x_{ij}^{-1}) = f(x_{ij})$.

c: f is multi-symmetric in the sense that it remains unchanged if the variables in each array X_i are (simultaneously) permuted by the same permutation $\sigma \in \mathrm{S}_n$, i.e., $f(x_{i\sigma(j)}) = f(x_{ij})$.

Our main results are summarized in the following theorem. Both parts answer questions posed by H. Hauptman.

Theorem 1.3. *(1) (Theorem 5.1 or Theorem 6.1) Every rational function $f(x_{ij})$ in mn variables, satisfying the invariance properties* **a**, **b**, **c**, *can be expressed as a rational function of the observables $q_{c_1,\ldots,c_m}$. In particular, E_m can be expressed as a rational function of the observables $q_{c_1,\ldots,c_m}$ for any $m \geq 1$.*

(2) (Proposition 7.1(b)) Suppose $n \geq 4$. Then E_m is not a polynomial in the observables $q_{c_1,\ldots,c_m}$ for any $m \geq 2$.

Our proof of Theorem 1.3 requires the consideration of all f satisfying the invariance properties **a**, **b**, **c**, even if one is only interested in the structure invariants E_m. This is true of most of our other proofs and algorithms.

The overall outline of the paper is as follows. In Section 2 we prove (a slightly more precise version of) Theorem 1.3(2) for $m = 1$ using basic Galois theory and some combinatorial arguments; the proof is not obviously constructive. In Section 3 we consider two possible approaches to making the argument constructive. In the subsequent section we give a fast algorithm for $n \leq 4$, based on SAGBI bases. In Section 5 we turn our attention to the multi-array case, i.e., $m > 1$. In Section 6 we reduce the problem of computing the invariant E_m to that of expressing certain invariant functions in terms of observables in the single array case ($m = 1$). We then use the SAGBI basis algorithm of Section 4 to obtain new expressions for E_2 in the case where $n = 3$ and 4; see Examples 6.5 and 6.6. In Section 7 we prove Theorem 1.3(2). Finally, in Section 8 we study the structure of the field of rational functions f satisfying the invariance properties (**a**) - (**c**) as an abstract field, without reference to the observables.

In the course of our work on this paper, we have encountered a phenomenon that often arises in the interstices between mathematics and its applications. Depending on the context, solving a mathematical problem can mean many different things, e.g.,

(a) proving a theorem,

(b) giving a constructive proof, or

(c) giving an algorithmic proof, suitable for practical computations.

As one moves down this list, the problem can become more difficult, requiring different techniques and ways of thinking. However, there is usually a subtle but important interplay between these different modes of solution.

Acknowledgments

The authors would like to thank H. Hauptman for bringing this problem to their attention, M. Lorenz for contributing Proposition 8.1, and D. Eisenbud, N. Elkies, J. Friedman, I. Laba, H. Lenstra, G. Martin, D. Peterson, C. Procesi, I. Swanson, and R. Thomas for helpful conversations.

2. One set of variables

Fix a field k of characteristic 0. Let $X = (x_1, \ldots, x_n)$ be an n-tuple of independent variables over k. We will operate on n-tuples as if they were diagonal matrices, so $X^{-1} = (x_1^{-1}, \ldots, x_n^{-1})$, $\mathrm{tr}(X) = x_1 + \cdots + x_n$, etc.

Let $k(X)_0 \subset k(x_1, \cdots, x_n)$ be the field of rational functions in the x_i of total degree 0; in other words, an element $f \in k(X)_0$ is a quotient of homogeneous polynomials of the same degree. Equivalently, $k(X)_0$ is the field generated by the x_i/x_j:

$$k(X)_0 = k(x_i/x_j : 1 \leq i, j \leq n, i \neq j) \ .$$

We note that the field $k(X)_0$ can also be viewed as the function field of the projective space $\mathbb{P}^{n-1}$.

The symmetric group S_n acts on $k(X)_0 = k(x_1, \cdots, x_n)_0$ by permuting the variables $x_1, \cdots, x_n$ in the natural way. In addition, we let τ denote the automorphism that takes X to X^{-1}, i.e.,

$$\tau(x_i) = \frac{1}{x_i}, \quad 1 \le i \le n.$$

This automorphism is obviously of order two, and we let $\mathrm{T} = \{1, \tau\}$ denote the corresponding group. The actions of S_n and T commute so that the group

$$G := \mathrm{S}_n \times \mathrm{T}$$

acts on $k(X)_0$. This action is faithful for $n \ge 3$. If $n = 2$, $G = \mathrm{S}_2 \times \mathrm{T}$ has order 4, and the kernel of its action on $k(x_1, x_2)_0$ is the subgroup of order 2 generated by (σ, τ), where σ is the nontrivial element of S_2.

The main theorem of this section is that the observables

$$q_r := \mathrm{tr}(X^r)\mathrm{tr}(X^{-r}) = \sum_{i,j=1}^{n} \left(\frac{x_i}{x_j}\right)^r \tag{4}$$

generate the invariant field $k(X)_0^G$. (Note that here $m = 1$, so that the observables (3) have only one subscript.)

Theorem 2.1.

$$k(X)_0^G = k(q_r \,|\, 1 \le r \le n(n-1)/2)\ .$$

Before proving the theorem, we prove two lemmas, one combinatorial and the other algebraic.

Let $n \ge 2$ be an integer, $N = n(n-1)$, and Λ be the N-element set

$$\Lambda = \{(i,j) \mid i, j = 1, \ldots, n \text{ and } i \ne j\}. \tag{5}$$

When convenient we will sometimes omit the comma and write (ij) instead of (i,j). In addition, we tend to visualize the elements of Λ as the off-diagonal positions in an n by n matrix.

There is a natural action of $G = \mathrm{S}_n \times \mathrm{T}$ on Λ, where τ acts by transposition

$$\tau(i,j) = (j,i)$$

and the symmetric group S_n acts simultaneously on the rows and columns:

$$\sigma(i,j) = (\sigma(i), \sigma(j)), \quad \sigma \in \mathrm{S}_n.$$

This gives a map from G to the symmetric group $\mathrm{S}_N = \mathrm{S}_{n^2-n} = \mathrm{Sym}(\Lambda)$ of all permutations of Λ, and this map is an injection for $n \ge 3$, in which case we will usually just choose to regard G as a subgroup of S_N. In the case $n = 2$ the map is surjective with kernel of order two as described earlier.

We will say that elements $x = (i,j)$ and $x' = (i',j')$ of Λ are *opposite* if x is the transpose of x', i.e., $i = j'$ and $j = i'$. If exactly one of these equalities holds, i.e., x' is not opposite to x, but it lies in the same row or column as the transpose of x, then we say that x and x' are *adjacent*. Note that opposite pairs are not also adjacent.

We will say that $g \in \mathrm{S}_N$ preserves adjacency (respectively opposition) if for any adjacent (respectively, opposite) elements $x, x' \in \Lambda$, the images $g(x)$ and $g(x')$ are also adjacent (respectively, opposite). Our combinatorial lemma says that for $n > 2$, $G = \mathrm{S}_n \times \mathrm{T}$ is precisely the subgroup of S_N that preserves both of these relations.

We note that for $n = 2$ the situation is simple: G maps onto $\mathrm{S}_N = \mathrm{S}_2$ and the nontrivial element of S_N preserves both adjacency and opposition. So from now on we consider $n \geq 3$.

Lemma 2.2. *Let $n \geq 3$. Then $h \in \mathrm{S}_N$ preserves both adjacency and opposition if and only if $h \in G \subset \mathrm{S}_N$.*

Proof. It is immediate from the definition that every h in G preserves both adjacency and opposition, so we only need to prove that any element preserving these relations lies in G.

Suppose $h \in \mathrm{S}_N$ preserves both adjacency and opposition. To show that $h \in G$, we will multiply h by elements of G until we arrive at the identity permutation of Λ.

It is easy to see that S_n acts transitively on Λ. Thus, after composing h with an element of $\mathrm{S}_n \subset G$, we may assume $h(12) = (12)$.

We claim that we may also assume that $h(13) = (13)$. Indeed, since h preserves opposition, $h(12) = (12)$ implies $h(21) = (21)$. Suppose $h(13) = (ij)$. Since (21) and (13) are adjacent, so are (21) and (ij), i.e., either $i = 1$ or $j = 2$. If $i = 1$ then $j \geq 3$; thus after replacing h by $[3, j]h$, we obtain $h(12) = (12)$ and $h(13) = (13)$, as desired. (Here $[3, j]$ denotes the transposition in S_n that interchanges 3 and j.) On the other hand, if $j = 2$ then $i \neq 1, 2$ and, after replacing h by $[3, i][1, 2]\tau h$, we once again obtain $h(12) = (12)$ and $h(13) = (13)$. This proves the claim.

Since h preserves opposition, $h(13) = (13)$ implies that $h(31) = (31)$. Now, since (23) is the unique pair adjacent to both (31) and (12), and $h(23)$ is the unique pair adjacent to both $h(31) = (31)$ and $h(12) = (12)$, we conclude that $H(23) = (23)$. Since h preserves opposition, we also have $h(32) = (32)$. Thus h fixes (ij) for $1 \leq i, j \leq 3$. This completes the proof of Lemma 2.2 for $n = 3$; from now on we will assume that $n \geq 4$.

Suppose $h(1i) = (ab)$ for some $i \geq 4$. Since (21) and $(1i)$ are adjacent, so are (21) and (ab). That is, either $a = 1$ or $b = 2$. Repeating this argument with (31) in place of (21), we see that either $a = 1$ or $b = 3$. Since b cannot be equal to both 2 and 3, we conclude that $a = 1$. In other words, $h(1i) = (1\sigma(i))$, where σ is a permutation of $4, 5, \ldots, n$. After replacing h by $\sigma^{-1}h$, we reduce to the case where h fixes $(1i)$ (and thus $(i1)$), for every $i = 2, \ldots, n$.

We claim that h is the identity permutation, i.e., that $h(ab) = (ab)$ for every $(ab) \in \Lambda$. Since we know this in the cases where $a = 1$ or $b = 1$, we may assume $a, b \geq 2$. In this case (ab) is the unique element of Λ that is adjacent to both $(1a)$ and $(b1)$. Hence, $h(ab) = (ab)$, as claimed. $\square$

Next, we prove an algebraic lemma from which the theorem will follow easily. Let

$$f(t) = \prod_{\substack{i,\,j\,=\,1 \\ i \neq j}}^{n} \left(t - \frac{x_i}{x_j}\right) = \sum_{i=0}^{N} (-1)^i c_i t^{N-i}$$

be the polynomial of degree $N = n(n-1)$ whose roots are the x_i/x_j, $i \neq j$. The coefficients c_i are the elementary symmetric functions in those roots, and since the reciprocals of roots are themselves roots, the polynomial f satisfies $t^N f(1/t) = f(t)$, which is equivalent to

$$c_i = c_{N-i}, \quad 0 \leq i \leq N.$$

The elements $q_r = \mathrm{tr}(X^r)\mathrm{tr}(X^{-r})$ are symmetric functions of the x_i/x_j are hence polynomials in the c_i; we let

$$K := k(q_r \,|\, r = 1, 2, \cdots)$$

be the field generated by the observables q_r.

As we will see, the proof of Theorem 2.1 basically comes down to determining the Galois group of f over the field K.

Lemma 2.3. *With the above notation,*

(a) $k[c_1, \ldots, c_r] = k[q_1, \ldots q_r]$ *for any* r, $1 \leq r \leq N$.

(b) $K = k(c_1, \ldots, c_{N/2}) = k(q_1, \ldots, q_{N/2})$.

(c) $k(X)_0$ *is the splitting field of* $f(t)$ *over* K.

Proof. Since

$$q_r - n = \mathrm{tr}(X^r)\mathrm{tr}(X^{-r}) - n = \sum_{i \neq j} (x_i/x_j)^r$$

is the sum of the m-th powers of the roots of $f(t)$, part (a) follows from Newton's formulas that express the symmetric polynomials $c_1, \cdots, c_r$ in terms of the power sums $q_1, \cdots, q_r$.

Part (b) follows from part (a) and the symmetry of the c_i. Part (c) follows from (b) and the fact that $k(X)_0 = k(x_i/x_j)$. □

We are now ready to finish the proof of Theorem 2.1. Clearly, $K \subset k(X)_0^G$. Consider the tower

$$\begin{array}{c} k(X)_0 \\ | \\ k(X)_0^G \\ | \\ K \end{array}$$

of field extensions. By the lemma, $k(X)_0$ is a Galois extension of K. Identify the set of roots of $f(t)$ with the set $\Lambda = \{(i,j) \,|\, i \neq j\}$, letting $x_i/x_j \leftrightarrow (i,j)$. The action of $G = \mathrm{S}_n \times \mathrm{T}$ on the set of roots is the same as its action on Λ described earlier.

The notions of adjacency and opposition in Λ have a natural interpretation in this context. If $x = (ab)$ and $x' = (cd)$ are elements of Λ let $r = x_a/x_b$ and $r' = x_c/x_d$ be the corresponding roots of $f(t)$. Then x and x' are adjacent if and only if rr' is again a root of $f(t)$ and opposite if and only if $rr' = 1$. Thus any $\mathrm{Gal}(k(X)_0/K)$ acts on the set of roots in a way that preserves both adjacency and opposition. Lemma 2.2 now tells us that $\mathrm{Gal}(k(X)_0/K) = G$. Thus $k(X)_0^G = K$. This completes the proof of Theorem 2.1.

Remark 2.4. Since $\mathrm{S}_n \subset G$ acts transitively on the roots of $f(t)$, we conclude that $f(t)$ is irreducible over $k(X)_0^G$.

3. Constructive proofs

The proof in the last section ultimately relies on a fundamental and beautiful result in Galois theory: anything fixed by all elements of a Galois group lies in the ground field. We will now discuss a constructive proof, which could be viewed as the result of tracing through the argument of the previous section, rendering the underlying Galois theory explicit at each step. We note that both proofs rely on Lemma 2.2.

We begin by letting $k[X^{\pm 1}]_0$ be the k-algebra whose elements are k-linear combinations of Laurent monomials $x_1^{a_n} \dots x_n^{a_n}$ of total degree 0, i.e., where the a_i are integers whose sum is 0. Note that $k[X^{\pm 1}]_0$ is generated, as a k-algebra, by elements of the form x_i/x_j; in particular, the field $k(X)_0$ is the field of fractions of $k[X^{\pm 1}]_0$. Note also that $k[X^{\pm 1}]_0$ is a G-invariant subring of $k(X)_0$.

Let z_{ij} be a set of $N = n(n-1)$ algebraically independent variables over k, where i and j are distinct integers between 1 and n. For notational convenience, we also set $z_{ii} = 1$ for every $i = 1, \dots, n$. We now define a surjective k-algebra homomorphism

$$\phi\colon k[z_{ij}] \longrightarrow k[X^{\pm 1}]_0$$

by $\phi(z_{ij}) = x_i/x_j$. Let s_r be the rth elementary symmetric polynomial in the N variables z_{ij} and $p_r = \sum_{i \neq j} z_{ij}^r$ be the sum of the rth powers of these variables. Here $s_0 = 1$ and $\phi(s_r)$ is the element of $k[X^{\pm 1}]_0$ we called c_r in the statement of Lemma 2.3. We let S_N denote the group of all permutations of $\{z_{ij} : i \neq j\}$, and identify $G = \mathrm{S}_n \times \mathrm{T}$ with the subgroup that acts on the z_{ij} by $\sigma(z_{ij}) = z_{\sigma(i)\sigma(j)}$, for $\sigma \in \mathrm{S}_n$, and $\tau(z_{ij}) = z_{ji}$.

Define three polynomials $D_1, D_2, D \in k[z_{ij}]$ of N variables by

$$D_1(z_{ij}) = \prod_{\substack{(ab) \text{ and } (cd) \in \Lambda \\ \text{are not opposite}}} (z_{ab} z_{cd} - 1)\,,$$

$$D_2(z_{ij}) = \prod_{\substack{(ab) \text{ and } (cd) \in \Lambda \\ \text{are not adjacent}}} (z_{ab} z_{cd} - z_{ef})\,.$$

and $D = D_1 D_2$. As we shall see below, $\phi(D)$ is a "universal denominator", such that if $f \in k[X^{\pm 1}]_0^G$ then $\phi(D) f$ is a polynomial in the observables.

Lemma 3.1. *(a) Suppose $g \in G$. Then $\phi(gD) = \phi(D) \neq 0$*

(b) Suppose $g \notin G$. Then $\phi(gD) = 0$.

Proof. (a) Clearly $gD_1 = D_1$ and $gD_2 = D_2$, because g preserves both adjacency and opposition. Thus $gD = D$, so that $\phi(gD) = \phi(D)$. To show that $\phi(D) \neq 0$, note that if the image under ϕ of a factor of D_1 is zero then

$$\frac{x_a}{x_b}\frac{x_c}{x_d} = 1\,.$$

This implies that (ab) and (cd) are opposite elements of Λ, which is excluded by the definition of D_1. This shows that $\phi(D_1) \neq 0$. Similarly, $\phi(z_{ab}z_{cd} - z_{ef}) = 0$ if and only if (ab) and (cd) are adjacent in Λ. Thus $\phi(D_2) \neq 0$, and consequently, $\phi(D) \neq 0$.

(b) If $g \in \mathrm{S}_n$ is not in G then by Lemma 2.2 at least one of the following holds: (i) g does not preserve opposition or (ii) g does not preserve adjacency.

If (i) holds then g^{-1} does not preserve opposition either. In other words, there exists a pair of non-opposite elements (ab) and (cd) such that $g(ab)$ and $g(cd)$ are opposite, say, $b_1 = c_1$. Then $z_{ab}z_{cd} - 1$ is a factor of D_1 and $\phi(g(z_{ab}z_{cd} - 1)) = 0$. Hence, $\phi(g(D_1)) = 0$ and thus $\phi(g(D)) = 0$.

Similarly, if (ii) holds then there exists a pair of non-adjacent elements (ab) and (cd) such that $g(ab) = (a_1b_1)$ and $g(cd) = (c_1d_1)$ are adjacent, say, $b_1 = c_1$. Now

$$\phi(g(z_{ab}z_{cd} - z_{g^{-1}(a_1d_1)})) = \phi(z_{a_1b_1}z_{c_1d_1} - z_{a_1d_1})) = \frac{x_{a_1}}{x_{b_1}}\frac{x_{c_1}}{x_{d_1}} - \frac{x_{a_1}}{x_{d_1}} = 0$$

so that $\phi(g(D_2)) = 0$ and thus $\phi(g(D)) = \phi(g(D_1))\phi(g(D_2)) = 0$, as claimed. □

With the universal denominator $\phi(D)$ in hand, we can now state our algorithm.

Algorithm 3.2.

Input: A function $f(x_1, \dots, x_n)$ in $k(X)_0^G$.

Output: A rational function in the q_r representing f.

Step 1: Write $f = f_1/f_2$ of a quotient of elements f_i that are in $k[X^{\pm 1}]_0^G$; apply each of the subsequent steps to f_1 and f_2 (to simplify the notation we just assume from now on that $f \in k[X^{\pm 1}]_0^G$).

Step 2: Find an element $F \in k[z_{ij}]$ such that $\phi(F) = f$.

Step 3: Set $A = \sum_{g \in \mathrm{S}_N} g(DF)$, write the numerator and denominator of

$$\frac{\phi(A)}{2n!\,\phi(D)}$$

as polynomials in the q_r and output the result.

We comment on each step in turn.

Step 1: To express an invariant degree 0 rational function f as a quotient of invariant polynomials, recall that f is a quotient of two polynomials of equal degree, say of degree d. Dividing top and bottom by x_1^d, we can write

f in the form $f_1 = a/b$, where a and b are in $K[X^{\pm 1}]_0$ (but a and b may not be G-invariant). Since f is G-invariant,

$$f = \frac{1}{2n!}\sum_{g\in G} g(f) = \frac{1}{2n!}\sum_{g\in G}\frac{g(a)}{g(b)} = \frac{f_1}{f_2}$$

where the numerator and denominator f_i are G-invariant polynomials in $k[X^{\pm 1}]_0^G$. It suffices to express f_1 and f_2 as rational functions in the q_r, and we can therefore assume that f is in $k[X^{\pm 1}]_0^G$ from now on.

Step 2: To lift f to an element $F \in k[z_{ij}]$, write $f(x_1, \ldots, x_n)$ as a k-linear combination of degree-0 Laurent monomials $x_1^{a_1} \ldots x_n^{a_n}$, with $a_1 + \cdots + a_n = 0$. Any such monomial is a product of a finite number of terms of the form $\frac{x_i}{x_j}$. Write all of the monomials in f in this form and replace each $\frac{x_i}{x_j}$ by z_{ij} to obtain the desired $F \in k[z_{ij}]$.

Step 3: By Lemma 3.1 the only non-zero terms in the sum

$$\phi(A) = \sum_{g\in \mathrm{S}_N} \phi(g(D))\phi(g(F))$$

correspond to $g \in G$. Thus

$$\phi(A) = \sum_{g\in G} \phi(g(D))\phi(g(F)) = \sum_{g\in G} \phi(D) g(\phi(F)) = 2n!\,\phi(D)\,f\ ,$$

as claimed.

The only remaining thing that needs to be done is to explain how to write $\phi(A)$ and $\phi(D)$ as polynomials in the q_r. Clearly A is a symmetric polynomial in the z_{ij}, where $1 \le i, j \le n$. It is therefore a polynomial in the elementary symmetric functions s_i, $1 \le i \le N$, and a polynomial in the power sums $p_r = \sum z_{ij}^r$ by Newton's formulas. Since $\phi(p_r) = q_r - n$ it follows that $\phi(A)$ can be written as a polynomial in the q_r, as desired.

From the formula

$$2n!\phi(D) = \sum_{g\in G} \phi(g(D)) = \sum_{g\in \mathrm{S}_N} \phi(g(D)) = \phi\left(\sum_{g\in \mathrm{S}_N} g(D)\right)$$

it follows that, once D is symmetrized over S_N, the same procedure can be used to express $2n!\phi(D)$ in terms of the observables.

Remark 3.3. The above procedure can be modified to produce polynomials in the q_r with $r \le N/2$. Indeed, recall that $\phi(s_{N-i}) = c_{N-i} = c_i = \phi(s_i)$. Thus the image under ϕ of a polynomial $P(s_1, \cdots, s_N)$ in the elementary symmetric functions is unchanged if s_i is replaced by s_{N-i} for $i > N/2$.

Note also that if one only wants to express $f(x_1, \ldots, x_n)$ as a rational function in q_r for $1 \le r \le N$, rather than $1 \le r \le N/2$, then the algorithm of [St_1, Proof of Proposition 1.1.2] can be used to write A and the symmetrization of D directly as polynomials in the power sums p_i without going through elementary symmetric polynomials.

4. Initial terms and SAGBI bases

It is natural to try to apply Gröbner bases to our problem. Although Gröbner bases are usually applied to rings, they can be adapted to solve problems in function fields. This was first noticed in detail by Sweedler [Sw], and has since been extended in several theoretical and practical ways; see, e.g., [M]. In our context we would have a G-invariant polynomial $f(X)$ in x_i/x_j that, by Theorem 2.1, lies in the field $k(q_k) \subset k(X)_0$. By introducing extra variables and calculating the Gröbner basis of a suitably chosen ideal, an explicit expression can be found for f as a rational function in the q_i.

Unfortunately, all of our implementations of this idea suggest that it has the same trouble as implementations of the constructive algorithm given in the previous section: they are far too slow. The purpose of this section is to introduce a faster algorithm, for $n \leq 4$, using a variant of Gröbner bases called SAGBI bases.

We shall always assume that $f \in k[X]_0^G$; the general case reduces to this one (cf. Algorithm 3.2, Step 1).

The subduction algorithm. Given an element

$$p(x_1, \dots, x_n) = \sum c_a x^a \in k[X^{\pm 1}] = k[x_1^{\pm 1}, \dots, x_n^{\pm 1}]\,,$$

where $a = (a_1, \dots, a_n) \in \mathbb{Z}^n$, $x^a = x_1^{a_1} \dots x_n^{a_n}$ and $c_a \in k$. We will write $\mathbf{in}(p)$ for the initial exponent p, i.e., the lexicographically largest exponent a such that $c_a \neq 0$. If R is a subalgebra of $k[X^{\pm 1}]$ then $\{\mathbf{in}(p) \,|\, p \in R\}$ is clearly a subsemigroup of $\mathbb{Z}^n$; this semigroup is usually denoted by $\mathbf{in}(R)$. We are interested in the case where $R = k[X^{\pm 1}]_0^G$; in this case $\mathbf{in}(R)$ consists of elements $a = (a_1, \dots, a_n) \in \mathbb{Z}^n$ satisfying the following conditions:

$$\begin{array}{ll} & \text{(i) } a_1 + \dots + a_n = 0, \\ (6) & \text{(ii) } a_1 \geq \dots \geq a_n, \text{ and} \\ & \text{(iii) } (a_1, \dots, a_n) \succeq (-a_n, \dots, -a_1). \end{array}$$

Here and in the sequel, $\succ$ denotes the lexicographic order on $\mathbb{Z}^n$.

Proposition 4.1. *Suppose B is subset of $R = k[X^{\pm 1}]_0^G$ chosen so that the elements $\mathbf{in}(b)$ generate $\mathbf{in}(R)$ as a semigroup, as b ranges over B. Then $R = k[B]$.*

Our proof below is based on the subduction algorithm of Robbiano-Sweedler [RS] and Kapur-Madlener [KM] for expressing a given element $\alpha \in R$ as a polynomial in elements of B.

Proof. We want to write $\alpha \in R$ as a polynomial in elements of B. If $\alpha = 0$, we are done. Otherwise write $\mathbf{in}(\alpha) = e_1\mathbf{in}(b_1) + \dots + e_r\mathbf{in}(b_r)$, where $b_1, \dots, b_r \in B$ and $e_1, \dots, e_m$ are non-negative integers. Then α and $b_1^{e_1} \dots b_r^{e_r}$ have the same leading exponent; thus for some $c \in k$,

$$\alpha_1 = \alpha - c b_1^{e_1} \dots b_r^{e_r}$$

has a lexicographically smaller leading monomial than α. If $\alpha_1 = 0$, we are done. If not, we can replace α by α_1 and apply the same procedure. That is, after subtracting a monomial in elements of B from α_1, we obtain $\alpha_2 \in R$ with a smaller initial exponent, etc.

In order to complete the proof of the proposition, it is enough to show that the resulting sequence $\alpha = \alpha_0, \alpha_1, \alpha_2, \dots$ in R will terminate, i.e., $\alpha_r = 0$ for some $r \geq 0$. This is a very special case of [Re, Proposition 6.5]; for the sake of completeness we give a direct proof below.

By our construction $\mathbf{in}(\alpha_0) \succ \mathbf{in}(\alpha_1) \succ \mathbf{in}(\alpha_2) \succ \dots$. Thus it suffices to prove that for any given $a = (a_1, \dots, a_n) \in \mathbf{in}(R)$ there are only finitely many $a' = (a'_1, \dots, a'_n) \in \mathbf{in}(R)$ such that $a \succeq a'$. Indeed, if $a \succeq a'$ then $0 \leq a'_1 \leq a_1$. Now condition (iii) says that $a'_n \geq -a'_1 \geq -a_1$, and condition (ii) says that $-a_1 \leq a'_n \leq a'_i \leq a'_1 \leq a_1$. Thus a'_i may assume only finitely many values for every $i = 1, \dots, n$. This completes the proof of Proposition 4.1. □

For computational purposes, we are interested in those cases, where the set B in Proposition 4.1 can be chosen to be finite, i.e., $\mathbf{in}(R)$ is a finitely generated semigroup. In such cases we shall refer to B as a SAGBI basis of G; cf. [Re, Introduction]. (Here SAGBI stands for "subalgebra analog to Gröbner bases for ideals"; this term is due to Robbiano and Sweedler [RS].)

Unfortunately, by [Re, Theorem 1.6] $k[X^{\pm 1}]_0^G$ has a SAGBI basis only for $n = 2$, 3 and 4; see also [Re, Example 7.3]. Moreover, the situation cannot be remedied by replacing the lexicographic order with a different term order. On the other hand, for $n \leq 4$ the subduction algorithm is much faster than Algorithm 3.2.

Explicit SAGBI bases. Let c_i be the ith elementary symmetric polynomial in x_i/x_j, as in Lemma 2.3. (Here i and j are distinct integers ranging from 1 to n.) Recall that $c_1 = q_1 - n = \mathrm{tr}(X)\mathrm{tr}(X^{-1}) - n$.

Lemma 4.2. *The following elements form a SAGBI basis of* $k[X^{\pm 1}]_0^G$.

(a) c_1, *if* $n = 2$.

(b) c_1 *and* c_2, *if* $n = 3$.

(c) c_1, c_2, c_3 *and* p, *if* $n = 4$. *Here* $p = s_2(X)s_2(X^{-1})$, *where*

$$s_2(X) = x_1x_2 + x_1x_3 + \dots + x_3x_4$$

is the second symmetric polynomial in $X = (x_1, \dots, x_4)$.

Proof. Let $S = \mathbf{in}(k[X^{\pm 1}]_0^G)$ be the subsemigroup of $\mathbb{Z}^n$ given by (6).

(a) If $n = 2$ then S is clearly generated by $(1, -1) = \mathbf{in}(c_1)$ as a semigroup.

(b) For $n = 3$, S is generated, as a semigroup, by the elements $\lambda_1 = (1, 0, -1) = \mathbf{in}(c_1)$ and $\lambda_2 = (2, -1, -1) = \mathbf{in}(c_2)$. Indeed, every element of $\mu \in S$ is of the form $\mu = (a, -c, -b)$, where $b \geq c \geq 0$ and $a = b + c$. Thus $\mu = c\lambda_1 + (b - c)\lambda_2$ lies in the semigroup generated by λ_1 and λ_2.

(c) We want to show that any $\mu = (a, b, c, d) \in S$ can be written as a non-negative integer linear combination of

$$\begin{aligned}\lambda_1 &= (1, 0, 0, -1) = \mathbf{in}(c_1)\,,\\ \lambda_2 &= (2, 0, -1, -1) = \mathbf{in}(c_2)\,,\\ \lambda_3 &= (3, -1, -1, -1) = \mathbf{in}(c_3) \quad \text{and}\\ \lambda_4 &= (1, 1, -1, -1) = \mathbf{in}(p)\,.\end{aligned}$$

If $b \le 0$ then the desired linear combination is given by

$$\mu = (c - d)\lambda_1 + (b - c)\lambda_2 + (-b)\lambda_3\,.$$

If $b > 0$ then, after replacing μ by $\mu - b\lambda_4$, we can assume $b = 0$ and apply the above formula. □

Corollary 4.3. *(a) If $n = 2$ then $k[X^{\pm 1}]_0^G = k[c_1] = k[q_1]$.*

(b) If $n = 3$ then $k[X^{\pm 1}]_0^G = k[c_1, c_2] = k[q_1, q_2]$.

(c) If $n = 4$, $k[X^{\pm 1}]_0^G = k[c_1, c_2, c_3, p] = k[q_1, q_2, q_3, p]$.

Proof. Immediate from Lemma 4.2, Proposition 4.1 and Newton's formulas; cf. Lemma 2.3(a). □

Example 4.4. Let $n = 3$. Corollary 4.3 tells us that c_1 and c_2 form a SAGBI basis for $k[X^{\pm 1}]_0^G$. For instance, applying the subduction algorithm, we obtain:

$$E_2(X, X) = 2(c_1^2 + c_1 - c_2)\,.$$

Similarly,

$$E_2(X, X^2) = 2c_1^3 + 5c_1^2 - 5c_1c_2 + 9c_1 - 12c_2 + 18\,.$$

These identities will be used in Section 6.

Example 4.5. $n = 4$. Using the subduction algorithm of Proposition 4.1 to express c_2^2 and c_4 in terms of the SAGBI basis c_1, c_2, c_3, p, we obtain the following relations in the ring $k[X^{\pm}]_0^G$:

$$\begin{aligned}c_2^2 &= 2c_1^2p - 16c_1^2 - 8c_1c_2 - c_1p^2 + 15c_1p - 48c_1 +\\ &\quad 3c_2p - 12c_2 + c_3p - 2p^2 + 18p - 36\\ c_4 &= 3c_1^2 + c_1c_2 - 3c_1p + 17c_1 + c_2 - 3c_3 + p^2 - 10p + 21\,.\end{aligned}$$

Eliminating p^2 and solving for p gives

$$p = \frac{6 + 7c_1 + 7c_1^2 + 3c_1^3 - 10c_2 - 5c_1c_2 + c_1^2c_2 - c_2^2 - 6c_3 - 3c_1c_3 - 2c_4 - c_1c_4}{2 + c_1 + c_1^2 - 3c_2 - c_3}.$$

This shows that $p \in k(c_1, c_2, c_3, c_4)$ or, equivalently, $p \in k(q_1, \dots, q_4)$. Thus by Corollary 4.3 $k(X)_0^G = k(q_1, q_2, q_3, q_4)$. Recall that Theorem 2.1 asserts only that $k(X)_0^G = k(q_1, \dots, q_6)$; we have thus shown that the last two of these generators are not needed.

To obtain explicit expressions for c_5 and c_6 as rational functions in $c_1, \dots, c_4$, we use the subduction algorithm once again:

$$\begin{aligned} c_5 &= c_1^3 - 6c_1^2 - 5c_1c_2 + 7c_1p - 38c_1 + c_2p - 7c_2 + 6c_3 - 2p^2 + 20p - 42 \\ c_6 &= -2c_1^3 + c_1^2p + 6c_1c_2 - 5c_1p + 27c_1 - 2c_2p + 9c_2 - 7c_3 + \\ &\quad 2p^2 - 18p + 34\,, \end{aligned}$$

then substitute the above formula for p.

Note also that for $n = 3$, Theorem 2.1 says that $k(X)_0^G = k(q_1, q_2, q_3)$, but q_3 is not needed by Corollary 4.3(b). We do not know whether or not any of the generators listed in Theorem 2.1 can be left out for $n \geq 5$.

5. More sets of variables

The purpose of this section is to generalize Theorem 2.1 to the multi-array case. That is, instead of considering a single array of independent variables $X = (x_1, \dots, x_n)$, we shall consider m arrays:

$$\begin{aligned} X_1 &= (x_{11}, \dots, x_{1n})\,, \\ X_2 &= (x_{21}, \dots, x_{2n})\,, \\ &\dots\dots\dots\dots\dots\dots \\ X_m &= (x_{m1}, \dots, x_{mn})\,. \end{aligned}$$

We shall view such n-tuples as diagonal $n \times n$-matrices and operate with them as we did in Section 2. For example,

$$X_iX_j = (x_{i1}x_{j1}, \dots, x_{in}x_{jn})\,, \quad \operatorname{tr}(X_i) = x_{i1} + \dots + x_{in}\,,$$

the observables (3) can be written as

$$q_{r_1,\dots,r_m}(X_1, \dots, X_m) = \operatorname{tr}(X_1^{r_1} \dots X_m^{r_m})\operatorname{tr}(X_1^{-r_1} \dots X_m^{-r_m}) \tag{7}$$

and the functions E_m that arise in the phase transition problem (2) as

$$\begin{aligned} E_m(X_1, \dots, X_m) = \operatorname{tr}(X_1) \dots \operatorname{tr}(X_m)\operatorname{tr}(X_1^{-1} \dots X_m^{-1}) + \\ \operatorname{tr}(X_1^{-1}) \dots \operatorname{tr}(X_m^{-1})\operatorname{tr}(X_1 \dots X_m)\,. \end{aligned} \tag{8}$$

Let $k(X_1, \dots, X_m)_0$ be the subfield of $k(x_{11}, x_{12}, \dots, x_{mn})$ whose elements are rational functions in x_{ij} homogeneous of degree 0 in each n-tuple of variables X_i. In other words, $k(X_1, \dots, X_m)_0$ is the function field of the variety $(\mathbb{P}^{n-1})^m$. In Section 2 we studied the action of the group $G = \mathrm{S}_n \times \mathrm{T}$ on $\mathbb{P}^{n-1}$; this action extends to an action of $G \times \dots \times G = G^m$ on $(\mathbb{P}^{n-1})^m$. We shall be interested in the invariants for the action of the diagonal subgroup of G^m which we shall also denote by G. In concrete terms, the symmetric group S_n acts on the function field $k(X_1, \dots, X_m)_0$ of $(\mathbb{P}^{n-1})^m$ by simultaneously permuting the variables $x_{i1}, \dots, x_{in}$ for each i. The 2-element group $\mathrm{T} = \{1, \tau\}$ acts on $k(X_1, \dots, X_m)_0$ by

$$\tau\colon x_{ij} \longrightarrow \frac{1}{x_{ij}} \text{ for every } i = 1, \dots, m \text{ and } j = 1, \dots, n. \tag{9}$$

These two actions commute and thus induce an action of $G = \mathrm{S}_n \times \mathrm{T}$ on $k(X_1, \dots, X_m)_0$.

Theorem 5.1. *The field $k(X_1, \dots, X_m)_0^G$ is generated (over k) by the following elements:*

$$q_{1,\dots,1} := \mathrm{tr}(X_1 \dots X_m)\mathrm{tr}\,(X_1 \dots X_m)^{-1}$$

and

$$\begin{aligned} q_{r,0,\dots,0} &:= \mathrm{tr}(X_1^r)\mathrm{tr}(X_1^{-r})\,, \\ q_{0,r,\dots,0} &:= \mathrm{tr}(X_2^r)\mathrm{tr}(X_2^{-r})\,, \\ &\vdots \\ q_{0,\dots,0,r} &:= \mathrm{tr}(X_m^r)\mathrm{tr}(X_m^{-r})\,, \end{aligned}$$

where r ranges from 1 to $(n(n-1))/2$.

We will give a proof that generalizes the proof in Section 2, and then discuss the prospects for a more constructive proof. Note that our earlier results for $m = 1$ will be used in the proof.

We begin by disposing of the case $n = 2$. This case is anomalous in that G does not act effectively on $k(X_1, \dots, X_m)_0$; the kernel of this action is the 2-element subgroup of $G = \mathrm{S}_2 \times \mathrm{T}$ generated by (σ, τ), where σ is the non-trivial element of S_2 and τ is the non-trivial element of T. Thus for $n = 2$

$$k(X_1, \dots, X_m)_0^G = k(X_1, \dots, X_m)_0^\tau\,;$$

here τ acts via the involution (9). Setting

$$t_i = \frac{x_{i1}}{x_{i2}}$$

we see that $k(X_1, \dots, X_m)_0 = k(t_1, \dots, t_m)$ and τ acts on this field by taking t_i to $\frac{1}{t_i}$ for each $i = 1, \dots, m$. It is now a simple exercise in Galois theory to show that in this case

$$k(t_1, \dots, t_m)^\tau = k(t_1 + \frac{1}{t_1}, \dots, t_m + \frac{1}{t_m}, t_1 \dots t_m + \frac{1}{t_1 \dots t_m})\,.$$

Since $t_i + \frac{1}{t_i} = \mathrm{tr}(X_i)\mathrm{tr}(X_i^{-1}) - 2$ and

$$t_1 \dots t_m + \frac{1}{t_1 \dots t_m} = \mathrm{tr}(X_1 \dots X_m)\mathrm{tr}\,(X_1 \dots X_m)^{-1} - 2\,,$$

this proves Theorem 5.1 for $n = 2$.

From now on we will assume that $n \geq 3$. Let K be the field generated over k by the observables listed in the theorem. For $h = 1, \dots, n$ let

$$f_h(t) = \prod_{i \neq j} (t - \frac{x_{hi}}{x_{hj}})\,. \tag{10}$$

By Lemma 2.3 the coefficients of each $f_h(t)$ lie in K. The field $k(X_1, \dots, X_m)_0$ is clearly the splitting field of the product

$$f(t) := f_1(t) \dots f_m(t)$$

over K. The Galois group of f contains G and is naturally embedded in the product G^m of the Galois groups of the f_i.

Lemma 5.2. *Assume $n \geq 3$. An element $g = (g_1, \ldots, g_m)$ of G^m fixes*

$$q_{1,\ldots,1} = \mathrm{tr}(X_1 \ldots X_m)\mathrm{tr}\,(X_1 \ldots X_m)^{-1} \in k(X_1, \ldots, X_m)_0^G$$

if and only if $g_1 = \cdots = g_m$.

Proof. One direction is obvious: $(g_1, \ldots, g_1)$ fixes $q_{1,\ldots,1}$ for every $g_1 \in G$. For the purpose of proving the converse, we may replace g by $(g_1^{-1}, \ldots, g_1^{-1})g$ and thus assume $g_1 = id$. In other words, we want to prove that if $g = (id, g_2, \ldots, g_n)$ fixes $q_{1,\ldots,1}$ then $g_2 = \ldots g_m = id$ in G.

For $i = 2, \ldots, m$, let $g_i = (\sigma_i, \epsilon_i) \in G = \mathrm{S}_n \times \mathrm{T}$. Here $\mathrm{T} = \{1, \tau\}$ is written multiplicatively, i.e., τ is written as -1 and each $\epsilon_i = \pm 1$. In particular,

$$(1, g_2, \ldots, g_m) \cdot (\frac{x_{hi}}{x_{hj}}) = (\frac{x_{h\sigma_h(i)}}{x_{h\sigma_h(j)}})^{\epsilon_h}\,.$$

Writing out $q_{1,\ldots,1}$ explicitly in terms of the x_{ij} (cf. (3)), we obtain

$$q_{1,\ldots,1} - n = \sum_{i \neq j} \frac{x_{1i}}{x_{1j}} \frac{x_{2i}}{x_{2j}} \ldots \frac{x_{mi}}{x_{mj}} \in K\,,$$

and

$$q_{1,\ldots,1} - n = g(q_{1,\ldots,1} - n) = \sum_{i \neq j} \frac{x_{1i}}{x_{1j}} (\frac{x_{2\sigma_2(i)}}{x_{2\sigma_2(j)}})^{\epsilon_2} \ldots (\frac{x_{m\sigma_m(i)}}{x_{m\sigma_m(j)}})^{\epsilon_m}\,.$$

Comparing the terms of the last two equations, we see that for each $h = 2, \ldots, m$, either (i) $\epsilon_h = 1$ and $\sigma_h(i) = i$ for every $i = 1, \ldots, n$, i.e., $\sigma_h = id$, or (ii) $\epsilon_h = -1$ and $\sigma_h(i) = j$ for every pair of distinct integers i, j between 1 and n. In case (ii), $\sigma(i)$ assumes every value between 1 and n other than i, which is impossible for $n \geq 3$. We conclude that $g_2 = \cdots = g_n$, as claimed. □

This shows that G is the Galois group of $k(X_1, \cdots, X_m)_0$ over K. Examining the tower

$$\begin{array}{c} k(X_1, \ldots, X_m)_0 \\ | \\ k(X_1, \ldots, X_m)_0^G \\ | \\ K \end{array}$$

we conclude that $k(X_1, \ldots, X_m)_0^G = K$, thus completing the proof of the theorem.

We now remark that Algorithm 3.2 (for $m = 1$) can be extended to the multi-array case by means of a suitable universal denominator. The generalization is not entirely straightforward because of the "new" generator

$$q_{1,\ldots,1} = \mathrm{tr}(X_1 \ldots X_m)\mathrm{tr}\,(X_1 \ldots X_m)^{-1}$$

that connects the different X_i.

Let $k[X_1^{\pm 1}, \dots, X_m^{\pm 1}]_0$ be the k-linear combinations of Laurent monomials $x_{11}^{a_{11}} \dots x_{mn}^{a_{mn}}$ such that $a_{i1} + \dots + a_{in} = 0$ for every $i = 1, \dots, m$. Note that $k[X_1^{\pm 1}, \dots, X_m^{\pm 1}]_0$ is generated, as a k-algebra, by elements of the form $\frac{x_{ij}}{x_{il}}$; in particular, the field $k(X_1, \dots, X_m)_0$ is the field of fractions of $k[X_1^{\pm 1}, \dots, X_m^{\pm 1}]_0$. Note also that $k[X_1^{\pm 1}, \dots, X_m^{\pm 1}]_0$ is a G-invariant subring of $k(X_1, \dots, X_m)_0$.

Let z_{hij} be a set of mN algebraically independent variables over k, where $N = n(n-1)$, i and j are distinct integers between 1 and n, and h ranges from 1 to m. Let t be another independent variable and let

$$\phi\colon k[z_{hij}, t] \longrightarrow k[X_1^{\pm 1}, \dots, X_m^{\pm 1}]_0$$

be the surjective k-algebra homomorphism given by

$$\phi(z_{hij}) = \frac{x_{hi}}{x_{hj}} \text{ and } \phi(t) = q_{1,\dots,1} - n.$$

We note that S_N^m acts on $k[z_{hij}, t]$ and where the h-th component permutes the z_{hij} and fixes t.

Our universal denominator is a polynomial $E \in k[z_{hij}, t]$ that has the property that for $\sigma_1, \dots, \sigma_m \in \mathrm{S}_N$

$$\phi((\sigma_1, \dots, \sigma_m)E) = \begin{cases} \phi(E) & \text{if } \sigma_1 = \dots = \sigma_m \in G, \\ 0 & \text{in all other cases.} \end{cases} \tag{11}$$

Here, as before, we view G as a subgroup of S_N for $n \geq 3$.

The polynomial E is defined by

$$E(z_{hij}, t) := D(z_{1ij}) \cdot \dots \cdot D(z_{mij})\, E_1(z_{hij}, t)\,,$$

where D is the polynomial defined in Section 3 and

$$E_1(z_{hij}, t) := \prod (t - \sum_{i \neq j} z_{1ij} z_{2g_2(ij)} \dots z_{mg_m(ij)})\,;$$

the product is taken over all $g_2, \dots, g_m \in G$ such that at least one $g_i \neq 1$, and the sum is taken over all pairs of distinct integers i and j between 1 and n.

As in the $m = 1$ case one can verify that E satisfies (11). The algorithm for expressing invariants in terms of observables is now similar to the $m = 1$ case, and we leave the details to the reader.

6. Reduction to one set of variables

We now return to the problem of writing a G-invariant multihomogeneous rational function $f(x_{11}, x_{12}, \dots, x_{mn})$ of total degree 0 in each group of variables $X_1 = (x_{11}, \dots, x_{1n})$, …, $X_m = (x_{m1}, \dots, x_{mn})$, as a rational function in the "observables" $q_{r_1,\dots,r_m}$.

In principle, the algorithm sketched in the preceding section is a solution to this problem. As one might expect, it is too slow to be of practical significance. On the other hand, the approach we took in Section 4 cannot

be extended to $m \geq 2$, even for small n, because suitable (finite) SAGBI bases do not exist; see [Re, Theorem 1.6 and Example 7.3]. This motivated our search for an algorithm that would reduce computations in the multi-array case ($m \geq 2$) to computations in the single-array case ($m = 1$). In this section we discuss such an algorithm and use it to generate explicit formulas for $E_2(X, Y)$ for $n = 3$ and 4.

Another generating set of observables. We begin by proving another variant of Theorem 1.3(a).

Theorem 6.1. *The field $k(X_1, \dots, X_m)_0^G$ is generated (over k) by the elements*

$$q_r := q_r(X_1) = q_{r,0,\dots,0} := \mathrm{tr}(X_1^r)\mathrm{tr}(X_1^{-r})$$

and

$$\begin{aligned} q_s^{(2)} &:= q_{s,1,0,\dots,0} := \mathrm{tr}(X_1^s X_2)\mathrm{tr}(X_1^{-s}X_2^{-1})\,, \\ &\vdots \\ q_s^{(m)} &:= q_{s,0\dots,0,1} := \mathrm{tr}(X_1^s X_m)\mathrm{tr}(X_1^{-s}X_m^{-1})\,. \end{aligned}$$

where $r = 1, \dots, \frac{n(n-1)}{2}$ and $s = 0, \dots, n(n-1) - 1$.

Informally speaking, the element $q_{1,\dots,1}$ of Theorem 5.1 ties $X_1, \dots, X_m$ together, where as the elements $q_r^{(i)}$ of Theorem 6.1 relate X_i to X_1 for each $i = 2, \dots, m$.

Proof. Let $K = k(X_1, \dots, X_m)_0 = k(x_{ij}/x_{il})$, and let K^N be a vector space of dimension $N = n(n-1)$ over K. We shall write elements of K^N as (z_{ij}), where $(i, j) \in \Lambda$, i.e., i and j are distinct integers between 1 and n, as in (5). The natural action of $G = \mathrm{S}_n \times T$ on Λ induces a permutation action on K^N.

Given an n-tuple $A = (a_1, \dots, a_n)$, we will denote the N-tuple of ratios $\frac{a_i}{a_j}$ by $\rho(A)$. Then $\rho\colon K^n \longrightarrow K^N$ is a G-equivariant map. Finally, for $h = 1, \dots, m$ let Z_h be the N-tuple of variables (z_{hij}), where $(i, j) \in \Lambda$.

We want to show that any $f(X_1, \dots, X_m) \in k(X_1, \dots, X_m)^G$ can be written as a rational function in the observables listed in the statement of the theorem. We begin with several reductions. First of all, we may assume without loss of generality that $f(X_1, \dots, X_m) \in k[X_1^{\pm 1}, \dots, X_m^{\pm 1}]_0^G$, by writing an invariant rational function as a quotient of invariant polynomials. Secondly, $f(X_1, \dots, X_m)$ can be lifted to a G-invariant polynomial $F(Z_1, \dots, Z_m)$ in z_{hij} such that

$$f(X_1, \dots, X_m) = F(\rho(X_1), \dots, \rho(X_m))\,.$$

Thirdly, we may assume without loss of generality that $F(Z_1, \dots, Z_m)$ is a homogeneous polynomial in the arrays of variables $Z_1, \dots, Z_m$ of multi-degree $(d_1, \dots, d_m)$. Indeed, any G-invariant F can be written as a sum of G-invariant multihomogeneous components, say, $F = F_1 + \cdots + F_r$, and we may replace f by $f_i = F_i(\rho(X_1), \dots, \rho(X_m))$.

Multilinearizing F, we obtain a G-invariant multilinear polynomial M in $d = d_1 + \cdots + d_m$ N-variable arrays such that

$$f(X_1, \dots, X_m) = M(\underbrace{\rho(X_1), \dots, \rho(X_1)}_{d_1 \text{ times}}, \dots, \underbrace{\rho(X_m), \dots, \rho(X_m)}_{d_m \text{ times}}) . \tag{12}$$

Next we observe that by the Vandermonde argument

$$\rho(I), \rho(X_1), \dots, \rho(X_1^{N-1}) \tag{13}$$

form a K-basis of K^N; here I stands for the identity n-tuple $(1, \dots, 1)$. In particular, for every $2 \le i \le m$, we can write

$$\rho(X_i) = \lambda_{i0}\rho(I) + \lambda_{i1}\rho(X_1) + \cdots + \lambda_{i,N-1}\rho(X_1^{N-1}) . \tag{14}$$

for some $\lambda_{i0}, \dots, \lambda_{i\,N-1} \in K^N$. Substituting this into (12) and expanding, we see that $f(X_1, \dots, X_m)$ can be written as a sum of terms of the form

$$(\text{monomial in } \lambda_{ij})\ M(\rho(X_1^{i_1}), \dots, \rho(X_1^{i_d})).$$

(In fact, $i_1 = \cdots = i_{d_1} = 1$ in each term, but we shall not use this in the sequel.) Since M is G-invariant, each

$$M(\rho(X_1^{i_1}), \dots, \rho(X_1^{i_d}))$$

is an element of $k[X_1]_0^G$, and thus, by Theorem 2.1, it can be written as a rational function in the observables $q_r = q_r(X_1)$.

Thus it remains to show that each λ_{ij} lies in the field L generated by the elements listed in the statement of the theorem. Note that by Theorem 2.1 the observables q_r lie in L for every r (and not just for $r = 1, \dots N/2$). Taking the dot product of both sides of (14) with $\rho(I), \rho(X_1), \dots, \rho(X_1^{N-1})$, and remembering that

$$\rho(X_1^i) \cdot \rho(X_1^j) = q_{i+j} - n \text{ and } \rho(X_i) \cdot \rho(X_1^j) = q_j^{(i)} - n,$$

we obtain the following linear system:

$$\sum_{i=1}^{N} (q_{i+j} - n)\lambda_i = q_j^{(i)} - n, \text{ where } i = 0, 1, \dots, N-1. \tag{15}$$

Since $\rho(I), \rho(X_1), \dots, \rho(X_1^{N-1})$ form a basis of K^N, the matrix

$$Q = \begin{pmatrix} q_0 - n & q_1 - n & \dots & q_{N-1} - n \\ q_1 - n & q_2 - n & \dots & q_{N-1} - n \\ \dots & & & \\ q_{N-1} - n & q_N - n & \dots & q_{2N-2} - n \end{pmatrix} . \tag{16}$$

of this system is non-singular. Solving the linear system (15), we conclude that each λ_j lies in L, as claimed. □

Multilinear invariants. Our proof of Theorem 6.1 reduces the computation of a given element f of $k(X_1,\dots,X_m)_0^G$ to computing finitely many elements in $k[X_1]_0^G$. Note that this reduction is constructive. The resulting algorithm is cumbersome in general; however, it simplifies considerably in the case where the polynomial $F(Z_1,\dots,Z_m)$ is itself multilinear, i.e., $d_1=\dots=d_m=1$ and $M=F$. This is exactly what happens in that case of greatest interest to us, namely, $f=E_m$ (cf. (2)); here

$$F(Z_1,\dots,Z_m)=\sum_{j_1,\dots,j_m,j=1}^{n} z_{1j_1j}\dots z_{mj_mj}+z_{1jj_1}\dots z_{mjj_m}\,.$$

Proposition 6.2.

$$E_m(X_1,\dots,X_m)=\sum_{i_2,\dots,i_m=0}^{N-1}\lambda_{2i_1}\dots\lambda_{mi_m}E_m(X_1,X_1^{i_2},\dots,X_1^{i_m})\,,$$

where

$$\begin{pmatrix}\lambda_{i0}\\ \lambda_{i1}\\ \dots\\ \lambda_{i\,N-1}\end{pmatrix}=Q^{-1}\begin{pmatrix}q_0^{(i)}-n\\ q_1^{(i)}-n\\ \dots\\ q_{N-1}^{(i)}-n\end{pmatrix}.$$

and Q is the $N\times N$-matrix (16).

Proof. The first formula is obtained by substituting (14) into

$$E_m(X_1,\dots,X_m)=F(\rho(X_1),\dots,\rho(X_m))$$

and expanding the right hand side. (Note that the specific form of F is not used here; we only use the fact that F is multilinear.) The second formula comes from solving the linear system (15). □

Remark 6.3. Proposition 6.2 remains valid if E_m is replaced by any f in $k[X_1^{\pm1},\dots,X_m^{\pm1}]_0^G$ such that $f(X_1,\dots,X_m)=F(\rho(X_1),\dots,\rho(X_m))$ for some G-invariant multilinear polynomial $F(Z_1,\dots,Z_m)$ in m N-variable arrays $Z_1,\dots,Z_m$.

Explicit determination of the triplet phase invariant. For $m=2$ the formula of Proposition 6.2 can be rewritten in the matrix form:

$$E_2(X,Y)=(e_0,\dots,e_{N-1})\,Q^{-1}\begin{pmatrix}q_{0,1}-n\\ q_{1,1}-n\\ \dots\\ q_{N-1,1}-n\end{pmatrix}, \tag{17}$$

where $e_i=e_i(X)=E_2(X,X^i)$. Here, as usual, we write X for X_1 and Y for X_2; cf. (2). We have thus reduced the computation of $E_2(X,Y)$ to the computation of $e_i=E(X,X^i)$ for $i=0,\dots,N-1$. Note that $e_0=2nq_1$,

so there are only $N-1$ elements $e_1, \dots, e_{N-1}$ we need to compute. We can further reduce this number by using the basis

$$\rho(X^{-N/2}), \rho(X^{-N/2+1}), \dots, \rho(X^{N/2-1})$$

of K^n instead of (13). This has the effect of shifting the subscripts in (17) as follows:

Proposition 6.4.

$$E_2(X,Y) = (e_{-N/2}, e_{1-N/2}, \dots, e_{N/2-1}) \cdot R^{-1} \cdot \begin{pmatrix} q_{-N/2,1} - n \\ q_{-N/2+1,1} - n \\ \dots \\ q_{N/2-1,1} - n \end{pmatrix},$$

where $q_r = q_{r,0}$, $e_i = E_2(X, X^i)$, *and* R *is the* $N \times N$*-matrix*

$$R = \begin{pmatrix} q_{-N} - n & q_{-N+1} - n & \dots & q_{-1} - n \\ q_{-N+1} - n & q_{-N+2} - n & \dots & q_0 - n \\ \dots & & & \\ q_{-1} - n & q_0 - n & \dots & q_{N-2} - n \end{pmatrix}.$$

□

Keeping in mind the identities $e_i = e_{-1-i}$ and $e_0 = 2nq_1$, we see that the formula of Proposition 6.4 reduces the computation of $E_2(X,Y)$ to the computation of $e_i = E_2(X, X^i)$ for $i = 1, \dots, (N/2) - 1$. We also note that since $q_{-r} = q_r$, the matrix R involves q_r only for $r = 1, \dots, N$, as opposed to $r = 1, \dots, 2N-2$ for the matrix (16).

The calculation of the e_i involves only one set of variables. For $n = 3$ we have the following explicit results.

Example 6.5. Let $n = 3$. Then $N = 6$, $q_0 = 9$, and Proposition 6.4 gives

$$E_2(X,Y) = (e_2, e_1, 6q_1, 6q_1, e_1, e_2)\, R^{-1} \begin{pmatrix} q_{-3,1} - 3 \\ q_{-2,1} - 3 \\ \dots \\ q_{2,1} - 3 \end{pmatrix},$$

where R is the 6×6-matrix

$$R = \begin{pmatrix} q_6 - 3 & q_5 - 3 & q_4 - 3 & q_3 - 3 & q_2 - 3 & q_1 - 3 \\ q_5 - 3 & q_4 - 3 & q_3 - 3 & q_2 - 3 & q_1 - 3 & 6 \\ q_4 - 3 & q_3 - 3 & q_2 - 3 & q_1 - 3 & 6 & q_1 - 3 \\ q_3 - 3 & q_2 - 3 & q_1 - 3 & 6 & q_1 - 3 & q_2 - 3 \\ q_2 - 3 & q_1 - 3 & 6 & q_1 - 3 & q_2 - 3 & q_3 - 3 \\ q_1 - 3 & 6 & q_1 - 3 & q_2 - 3 & q_3 - 3 & q_4 - 3 \end{pmatrix}.$$

Recall that e_1 and e_2 were computed in Example 4.4 by using SAGBI basis techniques:

$$\begin{aligned} e_1 &= 2c_1^2 + 2c_1 - 2c_2 \\ e_2 &= 2c_1^3 + 5c_1^2 - 5c_1c_2 + 9c_1 - 12c_2 + 18\,. \end{aligned}$$

These can be easily expressed in terms of q_1 and q_2 by using Newton's identities, which here amount to

$$c_1 = q_1 - 3 \qquad \text{and} \qquad c_2 = ((q_1-3)^2 - (q_2-3))/2.$$

The explicit formula for E_2 that results is quite different from, and considerably more elaborate than, the formula given in Example 1.2.

Example 6.6. Let $n = 4$. Then $N = 12$, $q_0 = 16$, and the formula of Proposition 6.4 reduces to

$$E_2(X,Y) = (e_5, e_4, e_3, e_2, e_1, 8q_1, 8q_1, e_1, e_2, e_3, e_4, e_5) \cdot R^{-1} \cdot \begin{pmatrix} q_{-6,1} - 4 \\ q_{-2,1} - 4 \\ \cdots \\ q_{5,1} - 4 \end{pmatrix},$$

where R is the 12×12-matrix

$$R = \begin{pmatrix} q_{12}-4 & q_{11}-4 & \ldots & q_3-4 & q_2-4 & q_1-4 \\ q_{11}-4 & q_{10}-4 & \ldots & q_2-4 & q_1-4 & 12 \\ q_{10}-4 & q_9-4 & \ldots & q_1-4 & 12 & q_1-4 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ q_1-4 & 12 & \ldots & q_9-4 & q_{10}-4 & q_{11}-4 \end{pmatrix}.$$

The elements $e_1, \ldots, e_5$ can again be explicitly determined using the SAGBI basis subduction algorithm; the result is:

$$\begin{aligned}
e_1 &= 2c_1^2 + 8c_1 - 2c_2 - 4p + 20 \\
e_2 &= 2c_1^3 + 4c_1^2 - 5c_1c_2 - c_1p + 7c_1 - 5c_2 + 3c_3 + 2p + 2 \\
e_3 &= 2c_1^4 + 4c_1^3 - 7c_1^2c_2 + 3c_1^2p - 41c_1^2 - 31c_1c_2 \\
&\quad +7c_1c_3 - 2c_1p^2 + 40c_1p - 154c_1 + 8c_2p - 28c_2 \\
&\quad +2c_3p + 24c_3 - 8p^2 + 72p - 136 \\
e_4 &= 2c_1^5 + 4c_1^4 - 9c_1^3c_2 + 13c_1^3p - 131c_1^3 - 80c_1^2c_2 + 9c_1^2c_3 \\
&\quad -7c_1^2p^2 + 144c_1^2p - 651c_1^2 + 24c_1c_2p - 186c_1c_2 + 7c_1c_3p \\
&\quad +43c_1c_3 - 29c_1p^2 + 379c_1p - 1136c_1 - 5c_2c_3 + 21c_2p \\
&\quad -120c_2 + 3c_3p + 78c_3 - 34p^2 + 334p - 688 \\
e_5 &= 2c_1^6 + 4c_1^5 - 11c_1^4c_2 + 27c_1^4p - 243c_1^4 - 142c_1^3c_2 \\
&\quad +11c_1^3c_3 - 14c_1^3p^2 + 289c_1^3p - 1317c_1^3 + 46c_1^2c_2p \\
&\quad -429c_1^2c_2 + 14c_1^2c_3p + 53c_1^2c_3 - 66c_1^2p^2 + 892c_1^2p \\
&\quad -2830c_1^2 - 17c_1c_2c_3 + 81c_1c_2p - 471c_1c_2 + 19c_1c_3p \\
&\quad +135c_1c_3 + 2c_1p^3 - 137c_1p^2 + 1286c_1p - 3039c_1 - 2c_2^3 \\
&\quad -35c_2c_3 + 27c_2p - 167c_2 + 3c_3^2 - 2c_3p^2 - c_3p \\
&\quad +171c_3 + 8p^3 - 150p^2 + 856p - 1402
\end{aligned}$$

Then we eliminate p, by using the formula in Example 4.5. to express $e_1, \dots, e_5$ as rational functions in c_1, c_2 and c_3. Finally, we use Newton's identities to rewrite each e_i as a rational function of q_1, q_2 and q_3.

Remark 6.7. The SAGBI basis algorithms were implemented explicitly in magma [BCP], and are very efficient, though of course they are limited to $n \leq 4$.

7. Regular invariants

Theorems 5.1 and 6.1 tell us that if we allow $r_1, \dots, r_m$ to range over the integers then the observables $q_{r_1,\dots,r_m}(X_1, \dots, X_m)$ generate $k(X_1, \dots, X_m)_0^G$ as a field extension of k. H. Hauptman asked us whether the functions E_m in (8) that arise in the phase determination problem (2) can be written as *polynomials*, rather than rational functions in the $q_{r_1,\dots,r_m}$. More generally, we can ask whether the the observables $q_{r_1,\dots,r_m}$ generate the k-algebra $k[X_1^{\pm 1}, \dots, X_m^{\pm 1}]_0^G$. In this section, we will show that the answer is generally "no."

Proposition 7.1. *Assume $n \geq 4$.*

(a) The function

$$f(X) = E_2(X, X) = \mathrm{tr}^2(X)\mathrm{tr}(X^{-2}) + \mathrm{tr}^2(X^{-1})\mathrm{tr}(X^2)$$

cannot be written as a polynomial in the observables q_r, as r ranges over the integers.

(b) Suppose $m \geq 2$. Then function $E_m(X_1, \dots, X_m)$ given by (8) cannot be written as a polynomial in the observables $q_{r_1,\dots,r_m}$, as $r_1, \dots, r_m$ range over the integers.

(c) For any $m \geq 1$ the k-algebra $k[X_1^{\pm 1}, \dots, X_m^{\pm 1}]_0^G$ is not generated by the observables $q_{i_1,\dots,i_m}$, as $i_1, \dots, i_m$ range over the integers.

By Corollary 4.3(a) and (b), f can be written as a polynomial in the observables if $n = 2$ or 3; i.e., part (a) is not true for $n = 2$ or 3. Also, part (b) is not true for $m = 1$ (indeed, E_1 is itself an observable).

Curiously, as we saw in the Section 5, the observables $q_{i_1,\dots,i_m}$ come close to generating the k-algebra $k[X_1^{\pm 1}, \dots, X_m^{\pm 1}]_0^G$ in the following sense: every $\alpha \in k[X_1^{\pm 1}, \dots, X_m^{\pm 1}]_0^G$ can be written in the form

$$\alpha = \frac{\beta}{\phi(E)}$$

for some $\beta \in k[q_{i_1,\dots,i_m}]$. Here $\phi(E)$ is a fixed nonzero element of $k[q_{i_1,\dots,i_m}]$, independent of α.

Proof. Part (b) is easily deduced from part (a) by specializing X_1 and X_2 to a single array of indeterminates $X = (x_1, \dots, x_n)$, and $X_3, \dots, X_m$ to the "identity array" $I = (1, \dots, 1)$. Part (c) is a consequence of parts (a) (for $m = 1$) and (b) (for $m \geq 2$), since $f \in k[X]_0^G$ and $E_m \in k[X_1^{\pm 1}, \dots, X_m^{\pm 1}]_0^G$ for every $m \geq 2$.

Thus we only need to prove (a). Let

$$\rho\colon \mathbb{A}^n - \{x_1 \dots x_n = 0\} \longrightarrow \mathbb{A}^N$$

be the map given by

$$\rho(x_1, \dots, x_n) = (\frac{x_1}{x_2}, \frac{x_1}{x_3}, \dots, \frac{x_{n-1}}{x_n})\,. \tag{18}$$

To prove part (a), we claim that it suffices to exhibit n-tuples x and y with nonzero coordinates such that

$$\text{(i) } f(x) \neq f(y) \qquad \text{and} \qquad \text{(ii) } \rho(x) \sim \rho(y),$$

where two N-tuples u and v are equivalent, written $u \sim v$, if one is a permutation of the other.

Indeed, assume that (i) and (ii) are true. Since the observables q_r are symmetric functions in $\{x_i/x_j\}$, where $i, j = 1, \dots, n$, $i \neq j$, the fact that $\rho(x) \sim \rho(y)$ implies that $q_r(x) = q_r(y)$ for every integer r. Thus $q(x) = q(y)$ for every $q \in k[q_1, q_2, \dots]$; given that $f(x) \neq f(y)$, we immediately deduce that $f \notin k[q_1, q_2, \dots]$ as desired. Note that it is sufficient if the coordinates of x and y lie in the algebraic closure of the field k.

First consider the case $n > 4$. Let z be a primitive n-th root of unity. (Recall that we are assuming that $\mathrm{char}\,(k) = 0$, so that z exists in the algebraic closure of k.) We claim that the points

$$\begin{aligned} x &= (1, 1, z^3, z^3, z^4, \cdots, z^{n-1}) \\ y &= (z, z, z^2, z^2, z^4, \cdots, z^{n-1}) \end{aligned}$$

satisfy conditions (i) and (ii). Here x is obtained from the point $(1, z, z^2, \cdots, z^{n-1})$ by replacing z by 1, and z^2 by z^3, and y is obtained from $(1, z, z^2, \cdots, z^{n-1})$ by a similar alteration of two coordinates. Note that

$$\mathrm{tr}(x^i) + \mathrm{tr}(y^i) = 2(1^i + z^i + z^{2i} + \cdots + z^{(n-1)i}) = 0 \tag{19}$$

for any $i \not\equiv 0 \pmod{n}$. In particular, $\mathrm{tr}(x^i) = -\mathrm{tr}(y^i)$ for $i = \pm 1, \pm 2$ and consequently, $f(x) = -f(y)$. Thus, in order to check (i), we only need to verify that $f(x) \neq 0$. This is easily done; substituting

$$\mathrm{tr}(x^i) = 1 - z^i - z^{2i} + z^{3i} = (1 - z^i)(1 - z^{2i})$$

into $f(x) = \mathrm{tr}^2(x)\mathrm{tr}(x^{-2}) + \mathrm{tr}^2(x^{-1})\mathrm{tr}(x^2)$, we see that

$$f(x) = 2z^{-6}(1 - z)^2(1 - z^2)^3(1 - z^4) \neq 0$$

for any $n > 4$.

Now we have to show that $\rho(x) \sim \rho(y)$. To do this let $G_x(t)$ denote the formal generating function

$$G_x(t) = 2 + 2t^3 + t^4 + \cdots + t^{n-1} \in \mathbb{Z}[t]/(t^n - 1)$$

in which the coefficient of t^i is the number of times that z^i appears as a coordinate in x. Then $G_x(t)G_x(t^{-1}) \in \mathbb{Z}[t]/(t^n - 1)$ is the formal generating function of $\rho(x)$. Thus (ii) is equivalent to

$$G_x(t)G_x(t^{-1}) = G_y(t)G_y(t^{-1}) \text{ in } \mathbb{Z}[t]/(t^n - 1). \tag{20}$$

Using the factorization $t^n-1=(t-1)(t^{n-1}+\cdots+1)$, we get an isomorphism

$$\mathbb{Z}[t]/(t^n-1)\simeq\mathbb{Z}\oplus R, \text{ where } R:=\mathbb{Z}[t]/(t^{n-1}+\cdots+1).$$

Here t maps to 1 in the first summand. We see that in order to prove (20) it suffices to check that the two sides have the same images in $\mathbb{Z}$ and in R. The image in $\mathbb{Z}$ is obtained by evaluating at 1, and $G_x(1)=G_y(1)=n$, so the desired identity is immediate in $\mathbb{Z}$. If g_x and g_y denote the images of G_x and G_y in R, then a calculation similar to (19) show that $g_y(t)=-g_x(t)$.

It follows immediately that $g_x(t)g_x(t^{-1})=g_y(t)g_y(t^{-1})$ as desired. This finishes the proof for $n>4$.

For $n=4$ we have to consider a more carefully crafted example. Let z be a primitive 13-th root of unity, and set

$$x=(1,z,z^4,z^6),\qquad y=(1,z,z^4,z^{11})\,. \tag{21}$$

An easy calculation, which we will leave to the reader, shows that $f(x)=-f(y)\neq 0$, and a generating function argument shows that $\rho(x)\sim\rho(y)$. (This choice of x and y is based on the fact that $\{0,1,4,6\}$ and $\{0,1,4,11\}$ are inequivalent planar difference sets for the cyclic group $\mathbb{Z}/13\mathbb{Z}$; cf. [Ry, Chapter 9].)

This completes the proof of Proposition 7.1. □

8. Rationality

In this section we investigate the structure of the field $k(X_1,\dots,X_m)_0^G$, without specific reference to the observables. The main result, Proposition 8.1 below, was communicated to us by M. Lorenz.

Recall that a field extension K/k is called *rational* (or equivalently, K is said to be rational over k) if $K=k(t_1,\dots,t_r)$ for some elements $t_1,\dots,t_r\in K$, algebraically independent over k. The extension K/k is called *stably rational* if there exists a field L, containing K, which is rational over both K and k. In other words, for some finite collection of indeterminates $s_1,\dots,s_r$ the field $L=K(s_1,\dots,s_r)$ is rational over k.

Proposition 8.1. *(a) For $m=1$, $k(X)_0^G$ is rational over k.*

(b) For any $m\geq 1$, $k(X_1,\dots,X_m)_0^G$ is stably rational over k.

Proof. (a) First assume $n=2$. Then $k(X)^G\subset k(X)_0=k(x_1/x_2)$, and the desired conclusion follows from Lüroth's theorem. For $n\geq 3$, part (a) is a special case of a theorem of N. Lemire; see [L, Theorem 7.7]. To see how Lemire's theorem applies, note that the elements x_i/x_j (viewed additively, with the natural S_n-action) form the root system A_{n-1}. The group $G=\mathrm{S}_n\times T$ is the automorphism group of this root system, and $k(X)_0=k(M)$, where M is the lattice (i.e., the multiplicative subgroup of $k(X)_0^*$) generated by $\{x_i/x_j\,|\,i,j=1,\dots,n\}$.

(b) We argue by induction on m. The base case, $m=1$, is part (a). For the induction step, assume $m\geq 2$. Let $K_m=k(X_1,\dots,X_m)_0$ and $L=K_{m-1}(x_1,\dots,x_n)$, where $X_m=(x_1,\dots,x_n)$. By the induction assumption,

K_{m-1}^G is stably rational over k; thus it is enough to show that K_m^G is stably rational over K_{m-1}^G. We will do this by proving that L^G is rational over both K_m^G and K_{m-1}^G.

Note that since $L = K_m(x_1) = K_{m-1}(x_1, \dots, x_n)$ and the G-action on K_{m-1} is faithful, the desired conclusion would follow from the no-name lemma, if G-acted linearly (or, more precisely, semi-linearly) on the variables $x_1, \dots, x_n$; see e.g., [EM, Proposition 1.1] or [Sh, Appendix 3]. However, the no-name lemma cannot be used directly in this setting because our G-action is not semi-linear (τ acts by inversion!).

Fortunately, our action can be easily linearized, following an argument of Hajja and Kang [HK, Lemma 2.3(i)]. Let $y_i = \frac{1-x_i}{1+x_i}$. Then S_n permutes $y_1, \dots, y_n$ and τ sends each y_i to $-y_i$. Since $x_i = \frac{1-y_i}{1+y_i}$, we have

$$L = K_m(y_1) = K_{m-1}(y_1, \dots, y_n)\,.$$

Now the no-name lemma tells us that L^G is rational over both K_m^G and K_{m-1}^G, as claimed. □

Remark 8.2. Let H be the subgroup of index 2 in $G = \mathrm{S}_n \times T$ consisting of pairs of the form $(\sigma, \tau^{\operatorname{sign}\sigma})$. Hajja and Kang [HK, Theorem 3.2] have shown that $k(X_1, \dots, X_m)_0^H$ is rational over k.

References

[BCP] W. Bosma, J. Cannon, C. Playoust, *The Magma algebra system. I. The user language*, Computational algebra and number theory (London, 1993). J. Symbolic Comput. **24** (1997), no. 3-4, 235–265.

[Da] J. Dalbec, *Multisymmetric Functions*, Beiträge zur Algebra und Geometrie, **40** (1999), 27-51.

[EM] S. Endô, T. Miyata, *Quasi-permutation modules over finite groups*, J. Math. Soc. Japan, **25**, no. 3, (1973), 397–421.

[F] D. R. Farkas, *Reflection groups and multiplicative invariants*, Rocky Mountain J. Math, **16**, no. 2 (1986), 215-222.

[Gia] C. Giacovazzo, *Direct Phasing in Crystallography*, Oxford University Press, 1998.

[Ha1] H. Hauptman, D. Y. Guo, H. Xu, R. H. Blessing, *Algebraic Direct Methods for Few-Atom Structure Models*, Acta Crystallographica A**58** (2002), 361–369.

[Ha2] H. Hauptman, The phase problem of x-ray crystallography, Rep. Prog. Phys. (1991), 1427-1454.

[HK] M. Hajja, M.-C. Kang, *Twisted actions of symmetric groups*, J. Algebra **188** (1997), no. 2, 626–647.

[KM] D. Kapur, K. Madlener, *A completion procedure for computing a canonical basis for a k-subalgebra*, Computers and mathematics (Cambridge, MA, 1989), 1–11, Springer, New York, 1989.

[L] N. Lemire, *Reduction in the rationality problem for multiplicative invariant fields*, J. Algebra **238** (2001), no. 1, 51–81.

[M] J. Müller-Quade, R. Steinwandt, *Gröbner bases applied to finitely generated field extensions*, J. Symbolic Comput. **30** (2000), 469–490.

[Re] Z. Reichstein, *SAGBI bases in rings of multiplicative invariants*, Comment. Math. Helv. **78** (2003), no. 1, 185–202.

[RS] L. Robbiano, M. Sweedler, *Subalgebra bases*, Commutative algebra (Salvador, 1988), Lecture Notes in Math., 1430, Springer, Berlin, 1990, 61–87.

[Ry] H. J. Ryser, *Combinatorial mathematics*, The Carus Mathematical Monographs, No. 14, published by the Mathematical Association of America, 1963.

[Sh] I. R. Shafarevich, *Basic Algebraic Geometry*, vol. 1, second edition, Springer-Verlag, 1994.

[St_1] B. Sturmfels, Algorithms in Invariant Theory, Texts and Monographs in Symbolic Computation, Springer-Verlag, 1993.

[St_2] B. Sturmfels, Gröbner Bases and Convex Polytopes, University Lecture Series, vol. 8, Amer. Math. Soc., 1995.

[Sw] M. Sweedler, *Using Groebner Basis to Determine the Algebraic and Transcendental Nature of Field Extensions: Return of the Killer Tag Variables*, in Applied algebra, algebraic algorithms and error-correcting codes (San Juan, PR, 1993), Lecture Notes in Comput. Sci., 673, Springer, Berlin, 1993, 66–75.

[vW] B. L. van der Waerden, Algebra, 8th edition, Springer-Verlag, 1971.

Reed College, Portland, OR 97202
E-mail address: jpb@@reed.edu

Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2, Canada
E-mail address: reichst@@math.ubc.ca