Inclusion-Exclusion calculations

Michael Vaughan-Lee

November 2018

Higman [1] proves the following theorem.

Theorem 1 *The number of ways of choosing a finite number of elements from GF(q^n) subject to a finite number of monomial equations and inequalities between them and their conjugates over GF(q), considered as a function of q, is PORC.*

Here we are choosing elements $x_1, x_2, \ldots, x_k$ (say) from the finite field $\mathrm{GF}(q^n)$ (where q is a prime power) subject to a finite set of equations and non-equations of the form

$$x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k} = 1$$

and

$$x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k} \neq 1$$

where $n_1, n_2, \ldots, n_k$ are integer polynomials in the Frobenius automorphism $x \to x^q$ of $\mathrm{GF}(q^n)$. Higman calls these equations and non-equations monomial. Higman's proof of Theorem 1 involves 5 pages of homological algebra, but a shorter more elementary proof can be found in [2] and in [3].

To prove Theorem 1 you actually only need to prove that the number of ways of choosing a finite number of elements from $\mathrm{GF}(q^n)$ subject to a finite number of monomial equations between them and their conjugates over $\mathrm{GF}(q)$, considered as a function of q, is PORC. To see this suppose that we have a set S of equations and a set T of non-equations. Let T^* be the set of equations obtained from T be replacing all the $\neq$'s by $=$'s. For each subset $U \subseteq T^*$ let n_U be the number of solutions to the equations $S \cup U$. Then the number of solutions to the equations S and the non-equations T is

$$\sum_{U \subseteq T^*} (-1)^{|U|} n_U. \tag{1}$$

In [2] and in [3] I show that to find the number of ways of choosing a finite number of elements from $\mathrm{GF}(q^n)$ subject to a finite number of monomial equations S we write the equations in S as the rows of a matrix. We also have to add in equations $x_i^{q^n-1} = 1$ to make sure that the solutions lie in $\mathrm{GF}(q^n)$. For example, we represent the equations

$$x_1^{q^2-1} = 1,\ x_1^{q+1}x_2^{-2} = 1,\ x_1^{q^n-1} = 1,\ x_2^{q^n-1} = 1$$

by the matrix

$$\begin{bmatrix} q^2-1 & 0 \\ q+1 & -2 \\ q^n-1 & 0 \\ 0 & q^n-1 \end{bmatrix}.$$

For any given value of q this matrix is an integer matrix and the number of solutions to the equations is the product of the elementary divisors in the Smith normal form of the matrix. In [3] I show that the the number of solutions to a set of monomial equations, when considered as a function of q, is PORC. In fact I show that the number of solutions can be expressed in the form $df(q)$ for some primitive polynomial $f(x) \in \mathbb{Z}[x]$, where

$$d = \alpha + \sum_{i=1}^{r} \alpha_i \gcd(q - n_i, m_i)$$

for some rational numbers $\alpha, \alpha_1, \alpha_2, \ldots, \alpha_r$, some integers $m_1, m_2, \ldots, m_r$ with $m_i > 1$ for all i, and for some integers n_i with $0 < n_i < m_i$ for all i. In addition I give an algorithm for computing d and f.

So we have an algorithm for computing the PORC function giving the number of ways of choosing a finite number of elements from $\mathrm{GF}(q^n)$ subject to a finite number of monomial equations and inequalities between them and their conjugates over $\mathrm{GF}(q)$. However, as described above, the algorithm involves computing n_U for every possible subset $U \subseteq T^*$ so the algorithm is only practical in this form if the set T of non-equations is relatively small. In this note we consider a particular calculation that arose in Bettina Eick's and my calculation of the PORC formulae giving the numbers of k-dimensional 7 generator class two Lie algebras over $\mathrm{GF}(q)$ for $8 \leq k \leq 28$. This calculation is described in Section 5 of our paper "Counting p-groups and Lie algebras using PORC polynomials". Here we expand on what was written in our paper.

We consider a diagonal matrix A in $\mathrm{GL}(7, q)$ with seven distinct eigenvalues $a_1, a_2, \ldots, a_7$. The exterior square of A is a diagonal matrix B in

GL$(21,q)$ with eigenvalues a_ia_j $(1 \le i < j \le 7)$. As described in our paper, we need to find the PORC formulae giving the number of matrices B of each possible type that arise as A ranges over all possible diagonal matrices in GL$(7,q)$ with seven distinct eigenvalues. For any given choice of $a_1, a_2, \ldots, a_7$ the type of B is determined by which equations $a_ia_j = a_ka_l$ hold (and which do not hold). Since $a_1, a_2, \ldots, a_7$ are all distinct, the equation $a_ia_j = a_ka_l$ is only possible if i, j, k, l are all distinct. So we let R be the set of all possible equations $a_ia_j = a_ka_l$ with $i < j, k$, $k < l$, i, j, k, l all distinct. (So $|R| = 105$.) Let T be the set of 21 equations $a_i = a_j$ with $1 \le i < j \le 7$. Then, to take just one example, the matrix B has 21 distinct eigenvalues if the eigenvalues of A satisfy *none* of the equations in $R \cup T$. So the PORC formula (1) for the number of B which have 21 distinct eigenvalues would be an alternating sum of 2^{126} terms. We describe below in some detail how we were able to compute this sum.

For each subset $S \subseteq R$ we let g_S be the PORC formula giving the number of choices of distinct elements $a_1, a_2, \ldots, a_7 \in \mathrm{GF}(q)$ satisfying the equations in S and satisfying none of the equations in $R \backslash S$. Permutations of the eigenvalues $a_1, a_2, \ldots, a_7$ give an action of Sym(7) on R, and if S and T are in the same orbit under this action then $g_S = g_T$. As described in our paper, R has more than 8×10^{27} orbits of subsets, but only 426 of these orbits contain subsets S with $g_S \neq 0$. We need to find representatives for these 426 orbits, together with the corresponding values of g_S.

Clearly $g_S = 0$ if the relations in S imply a relation $a_i = a_j$, or if they imply a relation in $R \backslash S$. Furthermore it is easy to see whether the relations in S imply a relation r — compute the PORC formula for the number of choices of $a_1, a_2, \ldots, a_7$ which satisfy the relations in S, and also compute the PORC formula for the number of choices of $a_1, a_2, \ldots, a_7$ which satisfy the relations in $S \cup \{r\}$. If the two formulae are the same then the relations in S imply r. We computed a set of representatives for the Sym(7)-orbits of subsets $S \subseteq R$ with the property that the relations in S do not imply any relations $a_i = a_j$ or any relations in $R \backslash S$. There were 483 of these orbits. For each of these 483 orbit representatives S we computed the PORC formula f_S giving the number of choices of distinct elements $a_1, a_2, \ldots, a_7 \in \mathrm{GF}(q)$ satisfying the relations in S (and possibly also some relations in $R \backslash S$). We will describe below how we computed the functions f_S. It turned out that $f_S = 0$ for 56 of these representatives, leaving us with 427 orbit representatives for which we still needed to compute g_S. It may seem paradoxical that we can have $f_S = 0$ when S does not imply any relation $a_i = a_j$, but consider the three relations $a_1^2 = a_2^2$, $a_1^2 = a_3^2$, $a_2^2 = a_3^2$. You cannot find solutions to these three equations with a_1, a_2, a_3 all distinct, but there are solutions with $a_1 \neq a_2$, other solutions with $a_1 \neq a_3$, and other solutions with $a_2 \neq a_3$.

So we were left with 427 representatives S with $f_S \neq 0$, and we needed to compute g_S for each of these representatives. We sort these 427 representatives into a list

$$S_1, S_2, \ldots, S_{427}$$

chosen so that if $i < j$ then $|S_i| \geq |S_j|$, and we store the values of f_{S_i} for $i = 1, 2, \ldots, 427$. Then we apply the following piece of pseudo code.

```
for i in [1..426] do
 let O be the Sym(7)-orbit of S_i
 for j in [i+1..427] do
   if |S_i| = |S_j| then continue; end if;
   for T in O do
     if S_j ⊂ T then f_{S_j} := f_{S_j} − f_{S_i}; end if;
   end for;
 end for;
end for;
```

This procedure replaces f_{S_i} by g_{S_i} for $i = 1, 2, \ldots, 427$.

It remains to describe how to compute f_S when $S \subseteq R$. Let V be the set of 21 equations $a_i = a_j$ ($1 \leq i < j \leq 7$), and for each subset $U \subseteq V$ let n_U be the number of choices of $a_1, a_2, \ldots, a_7$ which satisfy the relations in $S \cup U$ (as well, possibly, as other relation in $R \cup V$). Then, as in equation (1),

$$f_S = \sum_{U \subseteq V} (-1)^{|U|} n_U.$$

There are 2^{21} terms in this sum, so it actually quite feasible to compute f_S in this way. But there is a much more efficient way. First we compute a list of subsets U of V with the property that the relations in U do not imply any relations in $V \backslash U$. There are 877 of these subsets, corresponding to the 877 equivalence relations on $\{a_1, a_2, \ldots, a_7\}$. Let $\mathcal{R}$ be the set of these 877 subsets. For each subset $U \in \mathcal{R}$ we let n_U be the number of choices of $a_1, a_2, \ldots, a_7$ satisfying the relations in $S \cup U$ (as well, possibly, as other relation in $R \cup V$). Then, as in the computation of g_S, we take each subset $T \in \mathcal{R}$ in turn, starting with the largest subsets, and then the next largest, and so on, and iteratively replacing n_U by $n_U - n_T$ whenever $U \in \mathcal{R}$ is a proper subset of T. At the end of this process $n_{\{\}}$ will have been replaced by f_S. Note that this procedure gives

$$f_S = \sum_{U \in \mathcal{R}} m_U n_U$$

for some integer coefficients m_U which are independent of S. We can determine the coefficients m_U as follows. For each $i = 1, 2, \ldots, 877$ let w_i be the row vector of length 877 with i^{th} entry 1, and with all other entries 0. Order the elements of $\mathcal{R}$ in a sequence $U_1, U_2, \ldots, U_{877}$ chosen so that if $i < j$ then $|U_i| \geq |U_j|$. (So $U_1 = V$ and $U_{877} = \{\}$.) Then apply the following piece of pseudo code.

```
for i in [1..876] do
 for j in [i + 1..877] do
   if U_j ⊂ U_i then w_j := w_j − w_i; end if;
 end for;
end for;
```

At the end of this process w_{877} will have been replaced by

$$(m_{U_1}, m_{U_2}, \ldots, m_{U_{877}}).$$

So the functions f_S can each be computed as a linear combination of 877 functions n_U, rather than as a linear combination of 2^{21} functions n_U.

References

[1] G. Higman, *Enumerating p-groups. II: Problems whose solution is PORC*, Proc. London Math. Soc. **(3) 10** (1960), 566–582.

[2] Michael Vaughan-Lee, *On Graham Higman's famous PORC paper*, Internat. J. Group Theory **1** (2012), 65–79.

[3] Michael Vaughan-Lee, *Choosing elements from finite fields*, arXiv.1707.09652 (2017).