AN ELEMENTARY INTRODUCTION TO COSET TABLE METHODS IN COMPUTATIONAL GROUP THEORY

J. Neubüser
RWTH Aachen, 5100 Aachen, West Germany

0. *PROLOGUE*

> "...; in fact the method can be reduced to a purely mechanical process, which becomes a useful tool with a wide range of application. ..., we venture to predict that our method will prove quite practicable for most groups (at any rate such as occur naturally in geometry or analysis) of order less than a thousand, and for many groups of much higher order."
> *J.A. Todd, H.S.M. Coxeter*, 1936, [57].

The paper 'A practical method for enumerating cosets of a finite abstract group' from which the quotation is taken, may very well be thought of as starting the subject of a series of 5 survey lectures which were given at "Groups - St. Andrews 1981" under the title "Computational methods in group theory". The quotation itself was the guiding principle for them; I neither dealt with the question of algorithmic solubility of problems - this will in fact often be obvious - nor with the use of computers for solving specific group-theoretic problems in an ad hoc fashion but restricted attention to methods which are designed (and have been implemented) for practical use in a variety of cases.

Of course in 1936 Todd and Coxeter proposed and used their method for hand calculations. As far as I know the first proposal to use a computer on a group-theoretic problem appeared in print in the 'Manchester University Computer Inaugural Conference' in 1951 [49] where M.H.A. Newman discussed how a computer could be used to investigate 2-groups along the lines of P. Hall's approach. Although this proposal apparently has never been followed in detail, computers have produced quite spectacular results about finite p-groups during the last 7 years.

The first actual implementation of a group theoretical program seems to have taken place about two years later (1953), when C.B. Haselgrove implemented a Todd-Coxeter method on the Cambridge EDSAC 1 computer. No documentation of that program seems to be left, but J. Leech gives some description of it and subsequent implementations including his own, in a later survey article [35].

In 1960 a first paper on an implementation of a group theoretical method, this time for finding the lattice of subgroups of a finite group, was published [45]. Since then the number of publications on the subject has grown steadily reporting on the invention of a wide range of "practicable methods" and applications of increasing relevance. While in surveys published about 12 years ago [15], [46] a rather complete description of all activities in the field could still be given, in those 5 lectures I could only introduce some main lines of development which may be indicated by the titles of the lectures: 'Coset Tables', 'Permutation Groups', 'Collection', 'Subgroup Structure', and 'Characters'. Also the bibliography distributed at the conference contained only a selection of titles. A much more complete bibliography which is kept current in Aachen by V. Felsch (of course on a computer), can be obtained on request [25]. The lectures were given with the intention of introducing "theoretical" group theorists to computational methods. So facts from group theory and representation theory were assumed to be known. On the other hand the description of computational methods started from scratch, assuming only very vague understanding of the way a computer works.

I have to apologize that this paper does not, as originally intended, cover (with some more details and care) all the topics touched in the lectures; when trying to write them up I soon found the manuscript growing beyond and the progress behind schedule. So in the end I have confined the paper mainly to the first topic. As an excuse I can offer that for all other four topics comprehensive treatments have been published recently or are being prepared for publication.

The state of the art with Sims' powerful techniques for the construction of large permutation groups from a set of generators is very clearly described in Leon's papers [39] and [40] of which the first is an easier introduction to the subject, the second a more detailed report. Of the literature quoted there [51] and [52] deal with methods for further calculation in the permutation groups thus constructed. A variation of these methods for matrix groups is given in Butler's thesis [6] and forthcoming joint papers of Butler and Cannon [7], [8], [9].

The study of collection methods and in particular the nilpotent quotient algorithm can best be started with the papers [29] of Havas and Nicholson and [48] of Newman, where also good references to the origin of these methods are given.

For the present state of methods for the closer investigation of the subgroup structure of a given group I have to refer to two papers that are planned for the proceedings of the August 1982 Durham Conference on computational group theory. One, by V. Felsch, will describe details of the newest implementation of a lattice-of-subgroups program, the other, a joint paper of several authors, largely interactive "top-down" methods for soluble and nilpotent groups which are based on collection techniques.

Finally for the same proceedings a report, again by several authors, is being prepared on a character table system and its use.

While preparing the lectures for Groups - St. Andrews 1981 as well as while writing this report, I have freely used the papers quoted as well as possibly other ones and private communications and notes of many of their authors. Acknowledging this I want to thank them for their help and cooperation over many years and ask their indulgence if I have missed out details, in particular historical ones.

1. *THE TODD-COXETER METHOD*

Let us start by discussing the method Todd and Coxeter proposed for working from a finite presentation:

$$G = \langle g_1,\ldots,g_n \mid r_1(g_1,\ldots,g_n) = 1,\ldots,r_m(g_1,\ldots,g_n) = 1 \rangle$$

or shorter $G = \langle E|R \rangle$ with $E = \{g_1,\ldots,g_n\}$ and $R = \{r_i(g_1,\ldots,g_n) \mid i=1,..,m\}$ where each "relator" $r_i(g_1,\ldots,g_n)$ is a word in $g_1,\ldots,g_n$. (For simplicity we shall use the same notation, e.g., $r_i(g_1,\ldots,g_n)$, for a *word* in elements $g_1,\ldots,g_n \in G$, i.e. formally a finite sequence $g_{i_1},g_{i_2},\ldots,g_{i_t}$ with $g_{i_j} \in \{g_1,\ldots,g_n, g_1^{-1},\ldots,g_n^{-1}\}$ and its value, i.e., the *element* $g_{i_1}g_{i_2}\ldots g_{i_t} \in G$; if we have to emphasize the distinction we shall do so by using the terms "word" or "element" in the text.) Such a presentation not only defines a group uniquely up to isomorphism, as the factor group of a free group F on free generators $f_1,\ldots,f_n$ by the normal closure under F of the set $\{r_1(f_1,\ldots,f_n),\ldots,r_m(f_1,\ldots,f_n)\}$, see e.g. [41], but it may arise naturally also in other branches of mathematics as the following example illustrates.

Example 1. From Fig. 1 we can read off that the fundamental group of the octahedron space has the presentation

$$G = \langle a,b,c,d \mid abc = bdc = bad = acd = 1 \rangle,$$

the relations coming from II, III, I and IV respectively. Eliminating $c = b^{-1}a^{-1}$, $d = a^{-1}b^{-1}$ and putting $A := a$, $B := b^{-1}$ we have

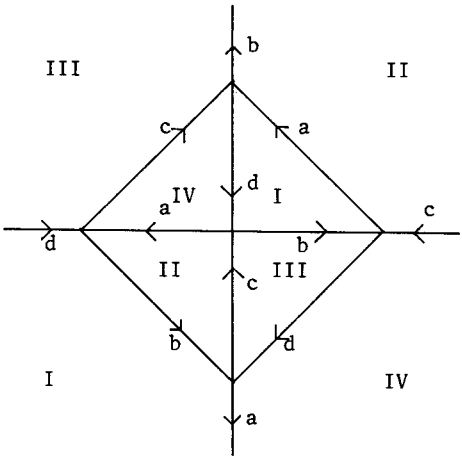
$$G = \langle A,B \mid A^3 = B^3 = (AB)^2 \rangle.$$

The relations $A^6 = B^6 = 1$ are consequences and elements of G have normal form A^ν, $0 \leqslant \nu \leqslant 5$; $A^\nu BA^\mu$, $0 \leqslant \nu \leqslant 2$, $0 \leqslant \mu \leqslant 5$. Hence $|G| \leqslant 24$.

So it is a question of relevance even outside group theory to ask, for any given finite presentation, if the group presented is finite, and if so what is its order.

We may generalize the situation slightly: let an additional finite set $S = \{s_1(g_1,\ldots,g_n),\ldots,s_p(g_1,\ldots,g_n)\}$ of words in the generators be given and let U be the subgroup of G generated by the elements $s_1(g_1,\ldots,g_n),\ldots,s_p(g_1,\ldots,g_n) \in G$. We may ask if U is of finite index in G and if so what is this index G:U. The Todd-Coxeter method attempts to find the index, if it is finite, by enumerating cosets of U in G in a systematic trial and error procedure. This procedure is based on two simple facts: if $s(g_1,\ldots,g_n) \in S$ then clearly $Us(g_1,\ldots,g_n) = U$, and if $r(g_1,\ldots,g_n) \in R$ then for any coset Uh, $h \in G$

Fig. 1

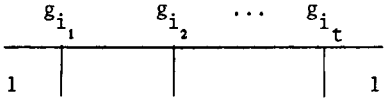


one has $Uh\, r(g_1,\ldots,g_n) = Uh$. Hence if $r(g_1,\ldots,g_n) = g_{i_1} \cdots g_{i_t}$, $g_{i_j} \in \bar{E} := \{g_1,\ldots,g_n, g_1^{-1},\ldots,g_n^{-1}\}$, and if a sequence of cosets

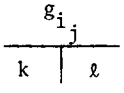
$$U_0 := Uh,\ U_1 := U_0 g_{i_1},\ldots,U_j := U_{j-1} g_{i_j},\ldots$$

is defined, then $U_t = U_0$ and the analogous statement holds for $U_0 = U$ and $s(g_1,\ldots,g_n) \in S$.

This is used in the following way: for each $s(g_1,\ldots,g_n) = g_{i_1} \cdots g_{i_t}$, $g_{i_j} \in \bar{E}$, a "subgroup table"

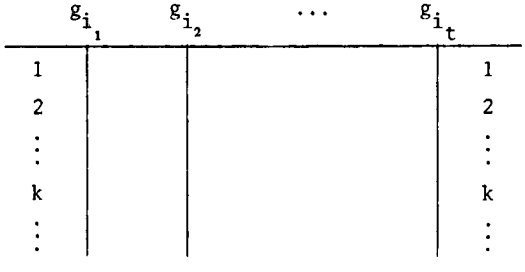


providing a single line for entering $t+1$ numbers representing cosets is set up, the entry



meaning that the coset of U which has got number k in the Todd-Coxeter enumeration procedure multiplied by g_{i_j} from the right yields the coset with number ℓ. Let the subgroup U itself be given the number 1, then the above remark makes clear that the first and the last entry in that line of coset numbers are 1.

Further for each relator $r(g_1,\ldots,g_n) = g_{i_1} \cdots g_{i_t}$, $g_{i_j} \in \bar{E}$, a "relation table" is set up organized analogously but providing one line for each coset given a number in the enumeration process. For the reason given above, the k-th line then starts and closes with k:



Finally to facilitate the bookkeeping of how cosets are numbered a "coset table" is set up, listing for each coset C that has been given a number k in the process, and each generator g_i and its inverse which number is given to Cg_i and to Cg_i^{-1}, respectively:

	g_1	...	g_n	g_1^{-1}	...	g_n^{-1}
1						
2						
⋮						
k						
⋮						

A Todd-Coxeter procedure then consists of defining in some sequence numbers for cosets of U, starting with 1 := U, then proceeding e.g. with $2 := 1g_1 (=Ug_1)$, $3 := 1g_2 (=Ug_2)$, $4 := 2g_2^{-1} (=Ug_1 g_2^{-1})$,..., with the only rule that a coset number ℓ must be defined by an equation $\ell = kg$, $k < \ell$, $g \in \bar{E}$, where the place of kg in the coset table is still vacant. As soon as the coset number ℓ has been defined the ℓ-th rows of the coset table and the relation tables are initialized and then this definition and its trivial consequence $\ell g^{-1} = k$ are filled into all possible vacant places of the various tables. The aim of the process is to obtain information about equality or inequality of cosets that have been given different numbers, i.e. whose coset representatives are given by different words in the generators, from the fact that lines in the subgroup tables or relation tables close up. Whenever that happens, an equality of the kind $kg = \ell$ for some $g \in \bar{E}$ is obtained and such an equality is called a "deduction". When such a deduction is reached three possibilities can occur: either

(i) the places of both kg and ℓg^{-1} in the coset table are still empty. In this case we fill the number ℓ into the place of kg and k into the place of ℓg^{-1} in the coset table and also insert this information into all other relevant places in the other tables; or

(ii) the place of kg in the coset table is already filled by the number ℓ (and hence then the place of ℓg^{-1} by the number k). In this case our deduction brings no new information; or

(iii) at least one of the places in the coset table is filled with a number different from that given by the deduction. In this case we conclude that we have given different numbers, a and b, say, to the same coset.

This phenomenon is called a "coincidence" a = b. When a coincidence is found, we have to replace the bigger one of the two numbers a and b by the smaller one in all our tables. We postpone the discussion how to do this.

The above description does not determine in which order a definition or a deduction is inserted into the various places in the tables into which it fits. In fact this order is irrelevant for conclusions about the Todd-Coxeter method that we shall draw, although it may influence the efficiency of the method. In the following examples we shall adopt the rule to take first the subgroup tables and then the relation tables in turn and fill into them line by line, first from the front and then from the end whatever information is available in the coset table. As soon as new information is gained (cases (i) and (iii)) this is entered into the coset table and the process is started all over again.

In the coset table, a number will be underlined, if it has been defined in this place; in the subgroup table and the relation tables deductions are underlined with a full line if they yield new information that can directly be put into the coset table (case (i) above), by a dotted line, if not (case (ii)) and by a wavy line, if they lead to a coincidence (case (iii)). When we want to keep a record of the sequence in which deductions occurred, we shall number them and put the number beneath the underlining.

Let us follow these rules in our example (in which case (iii) above does not occur). Putting A := a, B := b^{-1} one has c = $b^{-1}a^{-1}$ = BA^{-1}, d = $a^{-1}b^{-1}$ = $A^{-1}B$ and the shorter presentation:

$$G = \langle A,B \mid ABA^{-2}B = 1,\ B^{-1}A^{-1}B^2A^{-1} = 1 \rangle \quad \text{or}$$

$$G = \langle A,B \mid BABA^{-1}A^{-1} = 1,\ ABAB^{-1}B^{-1} = 1 \rangle.$$

With S = $\{A^2\}$ one has one subgroup table, two relation tables and the coset table, as shown below with the first definition 2 := 1A and the deduction 2A = 1, which follows from the closing of the subgroup table, being inserted.

Subgroup table

A	A
1	2 | 1
	1

Relation tables

B	A	B	A^{-1}	A^{-1}
1			1	2 | 1
2			2	1 | 2

A	B	A	B^{-1}	B^{-1}
1	2			1
2	1			2

The coset table

	A	B	A^{-1}	B^{-1}
1	2		2	
2	1		1	

Proceeding further by the sequence of definitions 3 := 1B, 4 := $1B^{-1}$, 5 := 2B, 6 := $2B^{-1}$, 7 := $3A^{-1}$, 8 := 4A we get the following picture:

Subgroup table

A	A
1	2
	1

Relation tables

B	A	B	A^{-1}	A^{-1}
1	3	4	1	2	1
2	5	6	2	1	2
3	6	7	8	4	3
4	1	2	5	8	4
5	4	8	7	6	5
6	2	1	3	7	6
7	8	5	4	3	7
8	7	3	6	5	8

A	B	A	B^{-1}	B^{-1}
1	2	5	6	3	1
2	1	3	4	5	2
3	4	1	2	6	3
4	8	7	3	1	4
5	6	2	1	4	5
6	7	8	5	2	6
7	3	6	7	8	7
8	5	4	8	7	8

The coset table

	A	B	A^{-1}	B^{-1}
1	2	3	2	4
2	1	5	1	6
3	4	6	7	1
4	8	1	3	5
5	6	4	8	2
6	7	2	5	3
7	3	8	6	8
8	5	7	4	7

All lines of all our tables have closed simultaneously. We shall show later that when this happens all coset numbers represent different cosets and all cosets have been numbered.

Before we do this let us look at another example which shows that coincidences (case (iii) above) cannot generally be avoided.

Our first presentation is obviously equivalent to

$$< A,B \mid BAB = A^2,\ ABA = B^2 >.$$

We change this, seemingly only slightly, to

$$G_1 = \langle A,B \mid BAB^{-1} = A^2,\ ABA^{-1} = B^2 \rangle$$

and enumerate again cosets of $U_1 = \langle A^2 \rangle$. From the definition 2 := 1A we obtain the deduction 2A = 1 and from the further definition 3 := 1B in turn the deductions 3A = 3, $1B^{-1} = 2$ and $2B^{-1} = 2$. At the time, when this last one is found, the state of the tables is the following:

	A	A
1	2	1
		1

	B	A	B^{-1}	A^{-1}	A^{-1}
1	3	3	1	2	1
2	1	2	2	1	2
3			3	3	3

	A	B	A^{-1}	B^{-1}	B^{-1}
1	2			3	1
2	1	3	3	1	2
3					3

	A	B	A^{-1}	B^{-1}
1	2	3	2	2
2	1	1	1	
3	3		3	1

Trying to insert 2B = 2 into the coset table we discover the coincidence 1 = 2. Multiplying this by B, we get immediately 3 = 1B = 2B = 1, hence the numbers 1, 2 and 3 all denote the same coset, we have encountered what is called a "total collapse".

It is clear that in this example we could not avoid defining some redundant coset numbers, in fact it is not difficult to construct examples for which the number of redundant cosets exceeds any given bound see, e.g. [33, p.93]. Therefore we have to describe a systematic procedure for the elimination of redundant coset numbers when coincidences have been found. When two coset numbers a and b represent the same coset, of course also the cosets ag and bg are equal for each element $g \in G$. Taking for g the generators g_i and their inverses, we see that from a comparison of the entries in the places of ag and bg in the coset table we may find deductions and in particular further coincidences, which we have to keep in mind while still dealing with the first one.

We do this (in principle) by establishing, as soon as we encounter a coincidence, an equivalence relation between coset numbers, calling two coset numbers equivalent, when they have been shown to represent the same coset. So this equivalence relation will change with the progress of the procedure which we now describe.

For this description we no longer assume that we have just found a first coincidence but rather that the coset numbers are already sorted into equivalence classes and that we want to eliminate one of a

pair of equivalent coset numbers. We choose some pair of coset numbers a and b, with b > a from some equivalence class and do the following:

C1. We replace each entry b in all the tables by a.

C2. For each $g \in \bar{E}$ we compare the entries in the g-column of the coset table in lines a and b (i.e. we compare the entries in the places of ag and bg).

a) If the place for bg is still empty, we do nothing.

b) If the places for ag and bg are both filled and if the entries are equal or are unequal but in the same equivalence class, we do nothing.

c) If the place for bg contains some entry c, while the place for ag is empty, we copy c into that place (i.e. we put ag = c as a new deduction).

d) If the place of ag contains the entry c, the place of bg the entry c′ with c and c′ not in the same equivalence class then we join the two classes to which c and c′ belong.

C3. When we have executed C2 for all $g \in \bar{E}$ we delete the b-th lines in all tables.

Let us discuss the effect of this procedure on the entries of the coset table. Before a first coincidence is found, the entries in the coset table satisfy the following property $\mathfrak{P}$:

($\mathfrak{P}$) Let a and b be coset numbers. The coset table lists that ag = b for some $g \in \bar{E}$ iff it also lists $bg^{-1} = a$.

At that time of course our equivalence classes consist of single coset numbers and whenever this is so, property $\mathfrak{P}$ may evidently be reformulated as

($\mathfrak{P}'$) Let A and B be equivalence classes of coset numbers and $g \in \bar{E}$. There exist $a \in A$ and $b \in B$ for which the coset table lists ag = b iff there exist $a' \in A$ and $b' \in B$ for which the coset table lists $b'g^{-1} = a'$.

We claim that if we enter our coincidence procedure with an equivalence relation satisfying $\mathfrak{P}'$, the new equivalence relation produced by the procedure will also satisfy $\mathfrak{P}'$. This is clear for steps C1 and C2(c) because a and b are supposed to be in the same equivalence class; it is trivial for steps C2(a) and C2(b) and it is also true for step C2(d), because joining two equivalence classes of a relation satisfying $\mathfrak{P}'$ in any case produces a relation satisfying $\mathfrak{P}'$. Finally step C3 preserves $\mathfrak{P}'$ because after step C2 for each non-empty place in line b the corresponding entry of line a is filled with an equivalent entry.

Our procedure diminishes the number of coset numbers by one and is applicable whenever there exists an equivalence class consisting of more than one coset number, hence after applying the procedure