NOTES FOR MATH 555, Spring 1999: FINITE GROUP THEORY; version 3/30/99

CONTENTS

1. Jordan-Hölder Theorem; Solvable Groups

We begin with the definition of a group with operators.

DEFINITION 1.1. Let S be a set. An S-group is a group G together with a mapping $S \to \mathrm{End}(G)$. Equivalently, an S-group is a group G together with a mapping

$$S \times G \to G, \qquad (s, g) \mapsto g^s$$

such that $(gh)^s = g^s h^s$ for all $g, h \in G$ and all $s \in S$.

We may say that G is a group with operators S.

EXAMPLES.

1. A vector space V over a field F is an F-group (with the additional assumption that the mapping $F \to \mathrm{End}(V)$ is a ring homomorphism). Notice that normally $\mathrm{End}(G)$ has no ring structure, but if G is an abelian group then $\mathrm{End}(G)$ is naturally a ring with respect to pointwise addition and composition.
2. If G is a group then G is an $\emptyset$-group. All the results in this section apply just to groups, rather than S-groups, by taking this special case.

3. If G is a group then G is a G-group, under the operation
$$g^s = s^{-1}gs.$$
4. If G is an S-group then G is also a S_0-group for any $S_0 \subseteq S$; more generally if T is any set and $\phi : T \to S$ any mapping, then any S-group G inherits the structure of a T-group via ϕ.
5. If G is a group then G is naturally an $\mathrm{End}(G)$-group, as well as an $\mathrm{Aut}(G)$-group, and the natural mapping $G \to \mathrm{Aut}(G)$ then yields an inherited structure of G as a G-group, which is the same as that introduced above.

DEFINITION 1.2. If G is an S-group then an S-subgroup of G is a subgroup $H \leq G$ such that $h^s \in H$ for all $h \in H$ and all $s \in S$. An S-subgroup of G inherits the structure of S-group.

Note: If H is a normal S-subgroup of G, then the factor group inherits the structure of S-group via $(Hg)^s = Hg^s$, and the natural projection $\pi_H : G \to G/H$ is an S-homomorphism. This is trivial to verify. We say that G is S-simple if and only if $G \neq 1$ but the only normal S-subgroups of G are G and 1.

If G and H are S-groups then an S-homomorphism (isomorphism) from G to H is a group homomorphism (isomorphism) $\phi : G \to H$ such that
$$\phi(g^s) = (\phi(g))^s$$
for all $g \in G$ and $s \in S$. In the isomorphism case we may write
$$\phi : G \cong_S H \quad \text{or just} \quad G \cong_S H.$$

The Noether isomorphism theory goes through for S-groups. Fix a set S.

THEOREM 1.3.
1. *If $\phi : G \to H$ is an S-homomorphism, then $\ker\phi$ is a normal S-subgroup of G, and there is a unique injective S-homomorphism $\bar{\phi} : G/\ker\phi \to H$ such that $\phi = \bar{\phi} \circ \pi_{\ker\phi}$.*
2. *If H and K are S-subgroups of G, and $H \leq N_G(K)$ where*
$$N_G(K) = \{x \in G \mid K^x = K\},$$
then HK is an S-subgroup of G and the natural injection $H \to HK$ induces an S-isomorphism
$$HK/K \cong_S H/H \cap K.$$
3. *If $\phi : G \to H$ is a surjective S-homomorphism with kernel K, then the mappings $L \mapsto \phi(L)$ and $J \mapsto \phi^{-1}(J)$ are inverse bijective correspondences between the set of all subgroups of G containing K and the set of all subgroups of H. Under these correspondences, normal subgroups correspond to normal subgroups (and the corresponding factor groups are then S-isomorphic). Moreover, intersections and joins also correspond: the correpsondences are lattice isomorphisms.*

Part (2) reminds us to note the following useful facts about products of subgroups. First, for any two subgroups H and K of a group G, we have that
$$HK \leq G \iff HK = KH.$$

Indeed, if $HK = KH$ then $HKHK = HHKK = HK$ and $(HK)^{-1} = KH = HK$, and the converse is equally easy to prove. In particular, in an abelian group the product of subgroups is always a subgroup, but in general this is not the case; however, it is true as long as at least one of the subgroups is normal.

Second, if H and K are finite, then $|HK| = |H||K|/|H \cap K|$, since each element of HK can be represented as a product hk, $h \in H$, $k \in K$, in exactly $|H \cap K|$ ways ($hk = (hg)(g^{-1}k)$ for all $g \in H \cap K$; check that this is gives all such representations of hk as a product).

Now the Noether isomorphism theory, together with a simple induction argument, yields the Jordan-Hölder Theorem. We state it only for finite groups, but it is not hard to generalize to groups satisfying both the maximum and minimum condition on S-subgroups: i.e., that every strictly ascending or strictly descending chain of S-subgroups of G is finite. Consequently it includes, for instance, the invariance of dimension of finite-dimensional vector spaces.

DEFINITION 1.4. Let G be an S-group. A subnormal series (or filtration) for G is a finite chain of subgroups

$$1 = G_n \lhd G_{n-1} \lhd \cdots \lhd G_1 \lhd G_0$$

for some integer $n \geq 0$. The S-groups $G_0/G_1, G_1/G_2, \ldots, G_{n-1}/G_n$ are called the factors of the series, and the series is called a composition series if and only if the factors are all S-simple.

The unordered list $G_0/G_1, G_1/G_2, \ldots, G_{n-1}/G_n$ of factors is called the list of factors of the series. The list of factors is thus a "multiset".

THEOREM 1.5 (JORDAN-HÖLDER). *Let G be a finite S-group. Then G has a composition series. Moreover, any two composition series have the same list of factors.*

The list of factors of a composition series of G is called the "set" of composition factors of G (really a multiset).

PROOF. Induction on $|G|$. If G has prime order, or indeed is S-simple, then the assertions are obvious. Otherwise for existence, let K be a proper normal S-subgroup of G. By induction K and G/K have composition series. The series for K, together with the full preimage of the series for G/K under the projection $\pi_K : G \to G/K$, combine to form a composition series for G, because of Theorem 1.3(3). Note that the list of factors of this series is just obtained by attaching the list for K to the list for G/K.

For uniqueness, say two composition series for G begin $G = G_0 > G_1 > \cdots$ and $G = G_0 > G_1^* > \cdots$. If $G_1 = G_1^*$, then by induction the uniqueness follows. Otherwise $G_1 G_1^*$ is a normal S-subgroup of G containing either G_1 or G_1^* properly. Since G/G_1 and G/G_1^* are S-simple, $G_1 G_1^* = G$. Then by Theorem 1.3(2),

$$G/G_1 \cong G_1^*/G_1 \cap G_1^* \text{ and } G/G_1^* \cong G_1/G_1 \cap G_1^*. \tag{1A}$$

In particular, all of these groups are therefore S-simple.

Now take any composition series of $G_1 \cap G_1^*$ and append it to $G > G_1 > G_1 \cap G_1^*$ and $G > G_1^* > G_1 \cap G_1^*$ to give two composition series of G:

$$\text{(a) } G > G_1 > G_1 \cap G_1^* > \cdots \text{ and (b) } G > G_1^* > G_1 \cap G_1^* > \cdots.$$

These have the same list of factors by (1A). Moreover the list of factors of series (a) is obviousy G/G_1 together with the set of composition factors of G_1 (well-defined by induction). Since the given series $G > G_1 > \cdots$ also passes through G_1, it has the same list of factors as (a): G/G_1 together with the composition factors of G_1. In exactly the same way, G/G_1^* and the composition factors of G_1^* comprise the list of factors of both series (b) and the given series $G > G_1^* > G_2^* > \cdots$. Since (a) and (b) have the same list of factors, so do the two given series. □

Remarks.

1. It is clear that any subnormal series of G, if not already a composition series, can be refined to a composition series, by the addition of further terms.
2. Applied to the cyclic group Z_n this gives another proof of the fundamental theorem of arithmetic. One must know that the only simple abelian groups are those of prime order.
3. A (finite-dimensional) representation of a group G over a field F is simply a finite-dimensional vector space V which is also a G-group in such a way that $v^{gh} = (v^g)^h$ for all $v \in V$ and all $g, h \in G$. Thus it is an abelian group V which is also a $F \cup G$-group, subject to this axiom as well as the vector space axioms. In this context, one calls the representation irreducible if and only if V is $F \cup G$-simple. The Jordan-Hölder theorem then applies to show that every representation has a filtration in which the factors are irreducible representations, and the resulting list of irreducible representations is uniquely determined by the original representation.
4. If $N \lhd G$, then as observed above, the composition factors of G are those of N together with those of G. In particular, the composition factors of a normal subgroup or of a quotient group are among those of G.
5. For an arbitrary subgroup $H \leq G$, a weaker result is available: intersecting H with a composition series of $G = G_0 > G_1 > \cdots$ yields a subnormal series $H \cap G_0 \geq H \cap G_1 \geq \cdots$, with each $H \cap G_{i-1}/H \cap G_i$ isomorphic to the subgroup $(H \cap G_{i-1})G_i/G_i$ of G_{i-1}/G_i. Thus each composition factors of H is isomorphic to a section (subquotient) of a composition factor of G.

From this point on "group" will mean "finite group". Sometimes we shall say "finite group", but this is only for emphasis.

A finite group is solvable if and only if all its composition factors are abelian, or equivalently of prime order. The previous remarks imply that subgroups and quotients of solvable groups are solvable, and if $N \lhd G$, then G is solvable if and only if N and G/N are solvable. Furthermore,

Proposition 1.6. *If H and K are solvable subgroups of G, with $K \lhd G$, then HK is solvable.*

Proof. $K \lhd HK$, with both K and $HK/K \cong H/H \cap K$ solvable. □

The hypotheses that H and K are solvable and HK is a group are insufficient for this conclusion, as is shown by the example of the nonsolvable group $G = A_5$, which is the product $G = HK$, $H = A_4$, $K \cong Z_5$.

As a consequence:

Proposition 1.7. *Every finite group G has a normal solvable subgroup $Sol(G)$ which contains all its normal solvable subgroups. Moreover, $Sol(G/\,Sol(G)) = 1$.*

PROOF. The product of normal solvable subgroups is normal and solvable. The last statement holds since the full preimage of $\mathrm{Sol}(G/\mathrm{Sol}(G))$ in G is normal and solvable. □

An alternative notation for $\mathrm{Sol}(G)$ is $O_{(\infty)}(G)$.

Clearly an analogous result holds with solvability being replaced by the condition that the composition factors all lie in some specified set of simple groups. For instance, we may replace the set of all cyclic groups of prime order by the singleton $\{Z_p\}$ for some prime p. Notice that a finite group G has order p^n if and only if all its composition factors are isomorphic to Z_p. (One direction is trivial; and the other holds since whenever $|G| = p^n > 1$, the group G has a nontrivial center, so has a central subgroup of order p, from which the result follows inductively.) Thus every finite group G has a normal p-subgroup $O_p(G)$ which contains all its normal p-subgroups. Likewise, for any set of primes π, G has a normal π-subgroup $O_\pi(G)$ containing all its normal π-subgroups (use the set of all simple groups whose orders are divisible only by primes in π).

Dual to the preceding results, we also have:

PROPOSITION 1.8. *If H and K are normal subgroups of G and both G/H and G/K are solvable, then $G/H \cap K$ is solvable.*

PROOF. G/H and $H/H \cap K \cong HK/K \lhd G/K$ are solvable. □

Thus G has a unique smallest normal subgroup K such that G/K is solvable. This subgroup could be called $O^{(\infty)}(G)$.

Analogously, G has unique smallest subgroups $O^p(G)$ and $O^\pi(G)$ such that the respective quotients are a p-group and a π-group.

The subgroups just defined are all characteristic in the following sense.

DEFINITION 1.8. A subgroup K of G is said to be characteristic in G if and only if $K^\alpha = K$ for every automorphism α of G.

We write K char G.

Examples of characteristic subgroups include: $[G, G]$, the commutator subgroup of G; $Z(G)$, the center of G; $O_p(G)$, the largest normal subgroup of G whose order is a power of the prime p. (Such a largest normal subgroup exists, because if H and K are two normal p-subgroups of G, then HK is certainly a normal subgroup of G and $|HK|$ divides $|H||K|$ so is a power of p.)

LEMMA 1.9.
1. *If K char G, then $K \lhd G$, but not conversely.*
2. *If K char L char G, then K char G.*
3. *If K char $L \lhd G$, then $K \lhd G$.*

LEMMA 1.10. *$O^{(\infty)}(O^{(\infty)}(G)) = O^{(\infty)}(G)$. Likewise $O^p(O^p(G)) = O^p(G)$ and $O^\pi(O^\pi(G)) = O^\pi(G)$ for any set π of primes.*

PROOF. We have $O^{(\infty)}(O^{(\infty)}(G))$ char $O^{(\infty)}(G) \lhd G$, so $O^{(\infty)}(O^{(\infty)}(G)) \lhd G$. But the corresponding factor group is an extension of solvable groups, so is solvable, whence $O^{(\infty)}(G) \leq O^{(\infty)}(O^{(\infty)}(G))$. Hence equality holds. The other proofs are similar. □

2. Chief Factors; The F^*-Theorem

DEFINITION 2.1. A chief series of the finite group G is a G-composition series, where G is a G-group under conjugation: $g^h = h^{-1}gh$ for all $g, h \in G$. The factors are called the chief factors of G.

Thus the chief factors of G are G-groups. The fact that the set G of operators is a group here is reflected in a further identity:

$$g^{hk} = (g^h)^k$$

for all $g \in G$ and $h, k \in G$; it then follows that for any chief factor J of G, this identity is inherited for all $g \in J$ and $h, k \in G$.

A chief factor of G is called a top chief factor if it occurs as G/G_1 for some chief series of G. For such a top chief factor, the condition that G/G_1 have no normal G-invariant subgroups is just equivalent to the condition that G/G_1 be simple (as an abstract group). So a top chief factor is simple as an abstract group. However, other chief factors need not be simple, as we shall see.

Notice that if $N \lhd G$, then every chief factor of G/N "is" a chief factor of G, the only difference being that we consider a chief factor of G/N to be a G/N-group, while the corresponding chief factor of G is a G-group. However, they correspond under the mapping $G \to G/N$ of sets of operators. In addition, G has further chief factors "in N", that is, of the form K/L where $L \leq K \leq N$ and $L \lhd G$ and $K \lhd G$. These are not necessarily chief factors of N; if K/L happens to be N-simple, then it is a chief factor of N, but in general we only know that K/L is G-simple, and G-simplicity does not imply N-simplicity. (Example: $G = A_4$, $N = V \cong Z_2 \times Z_2$. Then N is a chief factor of G. However, N is not a chief factor of itself; the chief factors of N are Z_2 and Z_2, with trivial N-action.)

We wish to describe, when $N \lhd G$, the connection between chief factors of N and chief factors of G within N. This depends on the notion of when two N-groups (groups with operators N) are G-conjugate.

Suppose that $N \lhd G$ and that X is an N-group. We formally construct another N-group, labelled X^g, consisting of all symbols

$$x^g, x \in X.$$

The multiplication is $x^g y^g = (xy)^g$, and the operator action is

$$(x^g)^n = y^g, \text{ where } y = x^{n^{g^{-1}}} \text{ in the } N\text{-group } X. \tag{2.1}$$

If we identify X with X^g via $x \mapsto x^g$ (an isomorphism of abstract groups), and if the operator action of N on X is given by the mapping

$$\alpha : N \to \mathrm{End}(X),$$

then the operator action of N on X^g is given by the composite

$$N \to N \to \mathrm{End}(X), \; n \mapsto n^{g^{-1}} \mapsto \alpha(n^{g^{-1}}).$$

We say that the two N-groups X and Y are G-conjugate if and only if $Y \cong_N X^g$ for some $g \in G$. Notice that the only difference between them is the different way that the operators act; as abstract groups they are isomorphic.

LEMMA 2.2. *Let $N \lhd G$ and let K be a chief factor of G within N. Then there exist chief factors $K_1, \ldots, K_n$ of N, all G-conjugate, such that*

$$K \cong K_1 \times \cdots \times K_n \text{ as } N\text{-groups.}$$

PROOF. By definition $K = M/L$ where $L \lhd G$, $M \lhd G$, $M \leq N$. There is no loss in factoring out L and assuming that $L = 1$, since L acts trivially on everything in sight. Then M is a minimal normal subgroup of G contained in N.

Let M_1 be a minimal normal subgroup of N contained in M. Let $M_1, \ldots, M_r$ be all the G-conjugates of M_1. These are all minimal normal subgroups of N, so are chief factors of N. Notice that if $M_i = M_1^g$, then as an N-group M_i is isomorphic to the formal conjugate M_1^g defined above, since the equations in (2.1) hold in G. The M_1 all lie in N so all normalize each other. Therefore $M_1 \ldots M_r \lhd N$. Indeed G permutes $M_1, \ldots, M_r$ among themselves and so $M_1 \ldots M_r \lhd G$. By minimality of M,

$$M = M_1 \cdots M_r.$$

Now choose the smallest integer s such that (after renumbering the M_i if necessary) we have $M = M_1 \cdots M_s$. We argue that

$$M = M_1 \times \cdots \times M_s.$$

Indeed, all M_i are normal in M and so we need only check that

$$M_i \cap M_1 \ldots \hat{M}_i \ldots M_s = 1$$

for each i. But if this were false, this intersection would be normal in N, hence equal M_i by minimality, and so $M = M_1 \ldots \hat{M}_i \dot{M}_s$, contradicting our minimal choice. This proves the direct product decomposition. Considering both sides as N-groups we see that M is isomorphic to the direct product of the G-conjugate chief factors M_i of N. □

Now we can prove:

THEOREM 2.3. *Let J be a chief factor of a finite group G. Then J is the direct product of isomorphic simple groups: $J = J_1 \times \cdots \times J_n$.*

PROOF. Write $J = K/L$ with $L \lhd G$, $K \lhd G$. Then J is a chief factor of G/L and without loss we may assume that J is a minimal normal subgroup of G. By the lemma, J is isomorphic to the direct product of G-conjugates of some chief factor of J. So J is isomorphic as abstract group to the direct product of copies of some chief factor J_1 of J. If $J = G$, then G is a chief factor of itself, hence simple, and we are done. Otherwise $J < G$ and by induction J_1 is the direct product of isomorphic simple groups; again we are done. □

The structure of J is well understood when J_1 is of prime order. The structure for the nonabelian case is in a sense even simpler.

PROPOSITION 2.4. *In any direct product $K = K_1 \times \cdots \times K_s$ of nonabelian simple groups, the only normal subgroups are those subgroups which are products of some of the K_i.*

PROOF. We shall use the notation $[x, y]$ for the commutator of two elements of a group G:

$$[x, y] = x^{-1}y^{-1}xy = x^{-1}x^y = (y^{-1})^x y,$$

so that if either x or y lies in a normal subgroup X of G, then $[x, y] \in X$ as well.

Let $L \lhd K = K_1 \times \cdots \times K_s$. Let $\pi_i : K \to K_i$ be the projections associated to this direct sum decomposition. Then $\pi_i(L) \lhd \pi_i(K) = K_i$, so for each i we have either $\pi_i(L) = 1$ or K_i. If $\pi_i(L) = 1$ for any i, there is no loss in simply deleting that K_i, so we may assume that $\pi_i(L) = K_i$ for each i, and using the fact that $L \lhd K$ we must prove that $L = K$. Fix i, and choose $x = x_1 \cdots x_s \in L$ with $x_i \neq 1$. Since K_i is not abelian there is $y \in K_i$ not commuting with x_i. But then $[y, x] = [y, x_i]$ is a nonidentity element of $K_i \cap L$, so $K_i \cap L = K_i$ by simplicity of K_i. Since i is arbitrary, $K = \prod_i K_i \leq L$, as desired. □

A consequence of this is that any automorphism of K must permute the set $\{K_1, \ldots, K_s\}$, since these are all the minimal normal subgroups of K.

We now look at G from a slightly different point of view, regarding elements by the way they act on the chief factors of G.

DEFINITION 2.5.

$$F^*(G) = \{g \in G \mid g \text{ induces an inner automorphism on each chief factor of } G\}.$$

DEFINITION 2.6. A group G is chiefly inner if and only if $F^*(G) = G$, that is, each element of G induces an inner automorphism on each chief factor of G.

Note that every $g \in G$ automatically induces an inner automorphism on any top chief factor of G. So if G is the direct product of simple groups, $F^*(G) = G$. However, $F^*(G) \neq G$ in general; for instance, $F^*(\Sigma_3) = A_3$.

The following properties of $F^*(G)$ are then basic.

THEOREM 2.7.

1. $F^*(G)$ *char* G.
2. $F^*(G)$ *is chiefly inner, and is in fact the largest chiefly inner normal subgroup of* G.
3. (BENDER'S F^*-THEOREM) $C_G(F^*(G)) \leq F^*(G)$.

Notice that if $G \neq 1$, then $F^*(G) \neq 1$ by the last statement. Bender's F^*-Theorem may well be called the Fundamental Theorem of Finite Group Theory. Its importance derives from the fact, as we shall see presently, that chiefly inner groups have a relatively transparent structure; it states that every finite group is made up of two overlapping pieces: $F^*(G)$, a chiefly inner group, and $G/Z(F^*(G))$, a group of automorphisms of a chiefly inner group.

PROOF. The product of inner automorphisms is inner, so obviously $F^*(G)$ is a subgroup of G. Moreover, given $g \in G$, the action of g on each chief factor may be computed with reference to any particular chief series of G. Since chief series are carried by automorphisms into other chief series, it follows that every automorphism of G carries $F^*(G)$ into itself, hence onto itself. Thus $F^*(G)$ char G.

Refine $1 \leq F^*(G) \leq G$ to a chief series $1 = H_n < H_{n-1} < \cdots < H_0 = F^*(G) < \cdots < G$ of G, and then refine the series of $H_i's$ to a chief series of $F^*(G)$. Every $g \in F^*(G)$ induces an inner automorphism on each H_{i-1}/H_i, *a fortiori* on every

g-invariant section of H_{i-1}/H_i and hence on every chief factor of $F^*(G)$. This proves that $F^*(G)$ is chiefly inner.

Let N be a normal chiefly inner subgroup of G. Let $g \in N$ and let K/L be a chief factor of G; we argue that g induces an inner automorphism on K/L. Indeed, we may assume that either $N \leq L$ or $K \leq N$. In the first case, $[g, K] \leq N \leq L$ and so g acts trivially on K/L. In the second case, K/L is isomorphic as N-group to a direct product of chief factors on N, on each of which g induces an inner automorphism since N is chiefly inner. Therefore g induces an inner automorphism on K/L, as claimed. Then $g \in F^*(G)$ by definition. Hence $N \leq F^*(G)$ and (b) is proved.

To prove the fundamental theorem, set $F = F^*(G)$, $C = C_G(F)$ and $Z = C \cap F$, and assume that $C \not\leq F$, so that $Z < C$. These are normal subgroups of G, so we may choose Y such that $Z < Y \leq C$ and Y/Z is a chief factor of G. Thus Y/Z is the direct product of simple groups. Consequently every chief factor K/L of Y either lies in $Z \leq Z(Y)$ or is a top chief factor of Y. In either case Y induces inner automorphisms on K/L, so Y is chiefly inner. By (b), $Y \leq F$, so $Y \leq Z$, a contradiction. □

EXERCISES. If $N \triangleleft G$ and G is chiefly inner, then both N and G/N are chiefly inner. If $N \leq Z(G)$, then the image of $F^*(G)$ in G/N is $F^*(G/N)$.

3. Commutators; The Basic Decomposition of Chiefly Inner Groups

Recall that a group G is chiefly inner if and only if every $g \in G$ induces only inner automorphisms on every chief factor of G. Thus G is chiefly inner if and only if $G = F^*(G)$. Thus $F^*(F^*(G)) = F^*(G)$. It is also clear that quotients of chiefly inner groups are chiefly inner. Moreover, so are normal subgroups, since if $N \triangleleft G$ and G is chiefly inner, then every element of N induces inner automorphisms on every chief factor of G within N, *a fortiori* (by refinement) on every chief factor of N.

We shall investigate the structure of such groups in the next few sections. We begin with some basic facts about commutators.

DEFINITION 3.1. If H, K and L are subsets of a group G, then

$$[H, K] = \langle [h, k], \mid h \in H,\ k \in K \rangle \text{ and } [H, K, L] = [\,[H, K], L].$$

If $x, y, z \in G$, then

$$[x, y, z] = [\,[x, y], z].$$

Longer commutators will be assumed to be left-associated unless explicitly stated otherwise:

$$[x_1, x_2, x_3, \cdots, x_n] = [\,[\cdots[\,[x_1, x_2], x_3], \cdots], x_n].$$

Notice that

$$x^y = x[x, y].$$

This has the following useful and immediate consequence for arbitrary subgroups H and K of a group G:

$$H \text{ normalizes } K \text{ if and only if } [K, H] \leq K.$$

The property of solvability can be stated in terms of commutators. Given G, define

$$G^{(1)} = G \text{ and } G^{(i+1)} = [G^{(i)}, G^{(i)}] \text{ for all } i \geq 1.$$

Then each $G^{(i)}$ is a characteristic subgroup of G, and the series

$$G = G^{(1)} \geq G^{(2)} \geq \cdots$$

is called the derived series of G. Since G is finite, it eventually stabilizes, and the last term is written $G^{(\infty)}$ (parentheses important!).

PROPOSITION 3.2.
1. $G^{(\infty)} = O^{(\infty)}(G)$, *the smallest normal subgroup of* G *with a solvable quotient.*
2. G *is solvable if and only if* $G^{(\infty)} = 1$.
3. $G^{(\infty)}$ *is perfect, i.e., equals its own commutator subgroup.*

PROOF. The last statement is immediate from the definition, and the second from the first. Notice that the derived series, reduced modulo $G^{(\infty)}$, gives a normal series for $G/G^{(\infty)}$ with abelian factors, hence solvable factors. So $G/G^{(\infty)}$ is solvable, and $O^{(\infty)}(G) \leq G^{(\infty)}$. Then $G^{(\infty)}/O^{(\infty)}(G)$ is a quotient of $G/O^{(\infty)}(G)$ so is solvable. If it is nontrivial it then has a quotient group of prime order, so $G^{(\infty)}$ has such a factor group $G^{(\infty)}/N$ as well. But then $[G^{(\infty)}, G^{(\infty)}] \leq N$, contradicting the perfectness of $G^{(\infty)}$. □

The following properties are of great use.

LEMMA 3.3.
1. $[x, y]^{-1} = [y, x]$.
2. $[x, yz] = [x, z][x, y]^z$.
3. $[xy, z] = [x, z]^y[y, z]$.
4. (PHILIP HALL'S IDENTITY) $[x, y^{-1}, z]^y[y, z^{-1}, x]^z[z, x^{-1}, y]^x = 1$.

PROOF. The first part is immediate from the definition. Notice that $x^y = x[x, y]$. The next two parts follow from use of this fact and the respective identities $x^{yz} = (x^y)^z$ and $(xy)^z = x^z y^z$. In the second case we also need the trivial identity $ay = ya^y$. As for Philip Hall's identity, notice that

$$\begin{aligned}[x, y^{-1}, z]^y = y^{-1}(yx^{-1}y^{-1}x)z^{-1}x^{-1}yxy^{-1}zy &= (x^{-1}y^{-1}xz^{-1}x^{-1})(yxy^{-1}zy) \\ &= \phi(x, y, z)^{-1}\phi(y, z, x),\end{aligned}$$

where $\phi(x, y, z) = xzx^{-1}yx$. The left side of Hall's identity is then

$$\phi(x, y, z)^{-1}\phi(y, z, x)\phi(y, z, x)^{-1}\phi(z, x, y)\phi(z, x, y)^{-1}\phi(x, y, z) = 1.$$

□

The last identity is reminiscent of the Jacobi identity, and in certain circumstances one may use it to build a finite Lie algebra from a finite group.

Important consequences follow:

LEMMA 3.4. *If* $H \leq G$ *and* $K \leq G$, *then*
1. $[H, K] = [K, H]$.
2. $[H, K] \lhd \langle H, K \rangle$.

PROOF. The first is clear from the first part of the preceding lemma. The second holds since $[x,z]^y = [xy,z][y,z]^{-1}$; taking $x, y \in H$ and $z \in K$ shows that H normalizes $[H,K]$, and symmetry implies that K does as well. □

LEMMA 3.5. (THE THREE-SUBGROUPS LEMMA) *Let $N \triangleleft G$, and let H, K, and L be subgroups of G such that $[H,K,L] \leq N$ and $[K,L,H] \leq N$. Then $[L,H,K] \leq N$.*

PROOF. Let $\overline{G} = G/N$. Then $\overline{[H,K,L]} = [\overline{H}, \overline{K}.\overline{L}]$ and similarly for the cycled triple commutators. Passing modulo N we may therefore assume that $[H,K,L] = [K,L,H] = 1$ and must prove that every element of K commutes with every element of $[L,H]$. Since $[L,H]$ is generated by commutators $[z,x]$, $z \in L$, $x \in H$, it is enough to show that $[z,x,y] = 1$ for all $x \in H$, $y \in K$, $z \in K$. Noting that H is closed under inversion and using Philip Hall's identity with x^{-1} in place of x, we obtain the desired result. □

We can now obtain the basic factorization of $F^*(G)$.

DEFINITION 3.6. If G is any finite group, then

$$E(G) = O^{(\infty)}(F^*(G)) = (F^*(G))^{(\infty)} \text{ and } F(G) = O_{(\infty)}(F^*(G)) = \mathrm{Sol}(F^*(G)).$$

THEOREM 3.7. *Let G be any group. Then*

$$F^*(G) = E(G)F(G) \text{ and } [E(G), F(G)] = 1.$$

Moreover, $E(G)$ is perfect and $F(G)$ is solvable.

REMARKS.

1. The product $E(G)F(G)$ is not necessarily a direct product, however. For instance, if $G = SL_n(q)$, so that $G/Z(G)$ is simple and G is perfect, then $G = E(G)$ and $F(G) = Z(G) \leq E(G)$, and whenever $Z(G) \neq 1$, which occurs when $(n, q-1) \neq 1$, then $F(G) \neq 1$.
2. Much more stringent conclusions hold about the individual structures of both $E(G)$ and $F(G)$, as will be seen in the next sections.

PROOF. The last statement is immediate from the definitions. Moreover, since $F^*(F^*(G)) = F^*(G)$ there is no loss in replacing G by $F^*(G)$ and assuming G to be chiefly inner.

Consider the quotient $\overline{G} = G/F(G)$. As a quotient of G, its chief factors are among those of G, and so

$$\overline{G} \text{ is chiefly inner, with } \mathrm{Sol}(\overline{G}) = 1.$$

We claim that any group with these properties is the direct product of nonabelian simple groups. This claim is proved by induction. If $\overline{G} = 1$ then there is nothing to prove. Otherwise let $N/1$ be a chief factor of $\overline{G}$. Then N is the direct product of isomorphic simple groups, which must be nonabelian as $\mathrm{Sol}(\overline{G}) = 1$. The mapping $G \to \mathrm{Aut}(N)$ induced by conjugation has image $\mathrm{Inn}(N)$ since $\overline{G}$ is chiefly inner. Thus it is really a mapping $\phi : G \to \mathrm{Inn}(N)$. The restriction to N is therefore already onto, and so $\overline{G} = N\ker(\phi) = NC_{\overline{G}}(N)$. But $Z(N) = 1$ so $\overline{G} = N \times C$, where $C = C_{\overline{G}}(N)$. Now $C \cong \overline{G}/N$ is again chiefly inner, and $\mathrm{Sol}(C)$ char $C \triangleleft \overline{G}$

so $\mathrm{Sol}(C) \leq \mathrm{Sol}(\overline{G}) = 1$, so induction applies to C and completes the proof of the claim.

The claim implies that $\overline{G}$ is perfect, and so $\overline{G} = \overline{G}^{(\infty)} = \overline{G^{(\infty)}} = \overline{E(G)}$. Therefore $G = E(G)F(G)$.

Next, refine $F(G) > 1$ to a chief series of G, ending with $F(G) = F_0 > F_1 > \cdots > F_n = 1$. Each F_{i-1}/F_i is solvable, so is the direct product of groups of prime order, so is abelian and has trivial inner automorphism group. Since G is chiefly inner it follows that G acts trivially by conjugation on each F_{i-1}/F_i. Therefore $[G, F_{i-1}] \leq F_i$ for each i. We argue inductively that $[E(G), F_0] \leq F_i$ for each i. For $i = 1$ this has already been proved. If it holds for i then $[E(G), F_0, E(G)] \leq [F_i, E(G)] \leq F_{i+1}$. By the Three Subgroups Lemma, $[E(G), E(G), F_0] \leq F_{i+1}$, completing the induction as $E(G)$ is perfect. Taking $i = n$ we get $[E(G), F_0] = 1$.□

Any solvable chiefly inner normal subgroup of G lies in $F^*(G)$ and then in $F(G)$. Likewise any perfect chiefly inner normal subgroup of G lies in $F^*(G)$ and then in $E(G)$. So $F(G)$ and $E(G)$ are, the largest normal chiefly inner subgroups of G which are solvable and perfect, respectively. In particular

$$E(E(G)) = E(G) \text{ and } F(F(G)) = F(G).$$

4. $F(G)$; Nilpotent Groups

Definition 4.1. A group G is nilpotent if and only if $G = F(G)$. Equivalently, G is nilpotent if and only if G is solvable and chiefly inner. Equivalently, G is nilpotent if and only if G acts trivially on each of its chief factors.

The comments at the end of the last section show:

Proposition 4.2. *For any group G, $F(G)$ is the largest normal nilpotent subgroup of G.*

Examples.

1. Quotients of nilpotent groups are nilpotent. So are subgroups; this is less obvious, but true since every chief factor of a subgroup arises as a subquotient of a chief factor of G.
2. Direct products of nilpotent groups are nilpotent. Indeed the chief factors of $H \times K$ are those of H (considered as HK-groups with K acting trivially) together with those of K (considered as HK-groups with H acting trivially).
3. If $G/Z(G)$ is nilpotent, then so is G. For G acts trivially on the chief factors of $G/Z(G)$, and obviously also on the chief factors of G within $Z(G)$.
4. Groups of prime power order are nilpotent. For if nontrivial, they have nontrivial centers; the assertion then follows inductively from the previous remark.

Thus direct products of groups of prime power order are nilpotent. This turns out to be a characterization of nilpotence. There are many other characterizations as well.

We first prove a simple but far-reaching consequence of Sylow's Theorem.

Theorem 4.3. (The Frattini Argument) *Let $K \lhd G$ and let P be a Sylow p-subgroup of K for some prime p. Then $G = KN_G(P)$.*

PROOF. G acts by conjugation on K. Let $x \in G$. Then P^x is a Sylow p-subgroup of K and so by Sylow's Theorem, $P^x = P^y$ for some $y \in K$. Therefore $x = (xy^{-1})y$ with $y \in K$ and $xy^{-1} \in N_G(P)$. So $G = N_G(P)K = KN_G(P)$. □

Now for the connections between nilpotence and Sylow subgroups.

THEOREM 4.4. *Let G be a group. Then the following conditions are equivalent:*

1. *G is nilpotent.*
2. *Every Sylow subgroup of G is normal in G.*
3. *G is the direct product of its Sylow subgroups.*
4. *If $x, y \in G$ and the orders of x and y are relatively prime, then $xy = yx$.*
5. *Every maximal subgroup of G is normal in G.*
6. *If H is any subgroup of G, then $H < N_G(H)$.*

PROOF.

(1) $\implies$ (6) Take a chief series $G = G_0 > G_1 > \cdots$ of G, so that by definition of nilpotence, $[G, G_{i-1}] \leq G_i$ for all i. Let k be the smallest integer such that $G_k \leq H$. Since $H < G$, $k > 0$. Then $[G_{k-1}, H] \leq [G_{k-1}, G] \leq G_k \leq H$, so $G_{k-1} \leq N_G(H)$. Therefore $N_G(H) \neq H$.

(6) $\implies$ (5) We have $M < N_G(M)$ so $N_G(M) = G$.

(5) $\implies$ (2) Let P be a Sylow p-subgroup in G and $H = N_G(P)$. If $H = G$ we are done. Otherwise, H lies in some maximal subgroup M of G. By (5), $M \lhd G$, so by the Frattini argument, $G = MN_G(P) = MH$, contradiction.

(2) $\implies$ (3) Write $|G| = p_1^{a_1} \cdots p_n^{a_n}$ where $p_1, \ldots, p_n$ are the distinct prime divisors of $|G|$ and let $P_1, \ldots, P_n$ be corresponding Sylow subgroups of G, so that for each i, $|P_i| = p_i^{a_i}$ and P_i is unique. By assumption $P_i \lhd G$ for each i. Set $G_k = P_1 \cdots P_k$ for each $k = 1, \ldots, n$. Clearly $G_k \lhd G$ for each k. We claim that $G_k = P_1 \times \cdots \times P_k$, and in particular $|G_k| = p_1^{a_1} \cdots p_k^{a_k}$, for any k. This is clear for $k = 1$, and inductively $G_k = G_{k-1}P_k$, the product of two normal subgroups of relatively prime order, so $G_k = G_{k-1} \times P_k$, proving the claim. In particular for $k = n$ we obtain $G_n = P_1 \times \cdots \times P_n = G$, as required.

(3) $\implies$ (4) We have $G = P_1 \times \cdots \times P_n$, and writing $x = x_1 \cdots x_n$ and $y = y_1 \cdots y_n$ accordingly, we see that the the order $|x|$ of x is $|x_1| \cdots |x_n|$ and similarly for $|y|$, so the condition that x and y have relatively prime orders amounts to saying that for each i, either $x_i = 1$ or $y_i = 1$. Therefore $[x_i, y_i] = 1$ for all i and so $[x, y] = 1$.

(4) $\implies$ (2) Choose one Sylow p-subgroup for each prime divisor of G, obtaining a set $\{P_1, \ldots, P_n\}$ of Sylow subgroups of G. Then $\langle P_1, \ldots, P_n \rangle$ has order divisible by every prime power divisor of $|G|$ so equals G. By hypothesis it normalizes every P_i, so $P_i \lhd G$.

(3) $\implies$ (1) We have already seen that p-groups, and direct products of nilpotent groups, are nilpotent.

□

Several other characterizations of nilpotence exist; we mention a couple.

DEFINITION 4.5. For any group G, the series

$$G = G^1 \geq G^2 \geq \cdots G^i \geq \cdots$$
$$\text{and} \quad 1 = Z_0(G) \leq Z_1(G) \leq Z_2(G) \leq \cdots \leq Z_i(G) \leq \cdots$$

are defined by the conditions

$$G^{i+1} = [G^i, G]$$
$$\text{and} \quad Z_i(G)/Z_{i-1}(G) = Z(G/Z_{i-1}(G))$$

for all $i \geq 1$. These series of characteristic subgroups are the lower and upper central series of G. Each terminates as soon as two successive terms are equal, and this final term is called G^∞ or $Z_\infty(G)$, respectively.

EXERCISES.

1. G is nilpotent $\iff G^\infty = 1 \iff Z_\infty(G) = G$.
2. If G is nilpotent, then the length of the upper and lower central series (i.e., the number of nontrivial factors) is the same; it is called the nilpotence class of G.
3. G has nilpotence class 1 if and only if it is abelian. G has nilpotence class 2 if and only if $[xy, z] = [x, z][y, z]$ for all $x, y, z \in G$, in which case it is also true that $[x, yz] = [x, y][x, z]$ for all $x, y, z \in G$.

An important example of a nilpotent group is the group $G = U_n(R)$ of upper triangular unipotent matrices over a commutative ring R. The elements of the subgroup G^k then vanish in the diagonals which are $1, 2, \ldots, k-1$ places to the right of the main diagonal, and so $G^n = 1$, whence G has nilpotence class at most $n-1$. (Exercise: it's exactly $n-1$.)

There are close connections between nilpotence and the Frattini subgroup.

DEFINITION 4.6. For any group G, the Frattini subgroup of G is

$$\Phi(G) = \bigcap_{M \text{ maximal in } G} M.$$

Obviously $\Phi(G)$ is a characteristic subgroup of G.

LEMMA 4.7. *If $X \leq G$ and $G = X\Phi(G)$, then $G = X$.*

PROOF. Otherwise, X lies in some maximal subgroup M of G, and hence so does $X\Phi(G)$, contradiction. □

THEOREM 4.8. (FRATTINI) *For any finite group G, $\Phi(G)$ is nilpotent. More generally, if K is a normal subgroup of G containing $\Phi(G)$, and $K/\Phi(G)$ is nilpotent, then K is nilpotent. In particular if $G/\Phi(G)$ is nilpotent, then so is G.*

PROOF. The second statement is what requires proof; the first statement is the special case $K = \Phi(G)$, and the last statement is the special case $K = G$.

In the second statement, let P be a Sylow p- subgroup of K. Then $P\Phi(G)/\Phi(G)$ is a Sylow p-subgroup of $K/\Phi(G)$, hence normal in $K/\Phi(G)$, hence the unique Sylow p-subgroup of $K/\Phi(G)$, so characteristic in $K/\Phi(G)$ and hence normal in $G/\Phi(G)$. Therefore $P\Phi(G) \lhd G$. Also since $P\Phi(G) \leq K$, P is a Sylow p-subgroup of $P\Phi(G)$. Therefore by the Frattini argument, $G = P\Phi(G)N_G(P) = N_G(P)\Phi(G)$ so $G = N_G(P)$, i.e., $P \lhd G$. *A fortiori*, $P \lhd K$. As P was arbitrary, K is nilpotent.□

COROLLARY 4.9. *$\Phi(G) \leq F(G)$ and $F(G/\Phi(G)) = F(G)/\Phi(G)$ for any group G.*

PROOF. This follows from the characterization of $F(G)$ as the largest normal nilpotent subgroup of G. Since $\Phi(G)$ is nilpotent, the first statement follows. Also writing $F(G/\Phi(G)) = F_0/\Phi(G)$, we deduce from the previous result that F_0 is nilpotent; it is obviously normal in G, so lies in $F(G)$. The converse inclusion is obvious since quotients of nilpotent groups are nilpotent. □

In the previous situation we can say a bit more;

LEMMA 4.10. $\Phi(F(G)) \leq \Phi(G)$. *More generally, if* $X \lhd Y$, *then* $\Phi(X) \leq \Phi(Y)$.

PROOF. Note that $\Phi(X)$ char $X \lhd Y$, so $\Phi(X) \lhd Y$. It suffices to prove that any maximal subgroup M of Y contains $\Phi(X)$. But otherwise, the subgroup $M\Phi(X)$ would contain M properly and so equal Y. Then $X = X \cap Y = X \cap M\Phi(X) = (X \cap M)\Phi(X)$ and so $X = X \cap M \leq M$, contradicting $\Phi(X) \not\leq M$. □

EXERCISE. $\Phi(G \times H) = \Phi(G) \times \Phi(H)$.

The structure of $F(G)/\Phi(G)$, which is a quotient of $F(G)/\Phi(F(G))$, is given as follows.

PROPOSITION 4.11. *If* G *is nilpotent, then* $\Phi(G)$ *is the smallest normal subgroup* K *of* G *such that* G/K *is the direct product of groups of prime order.*

PROOF. Since G is nilpotent, every maximal subgroup of G is normal, and we set

$$\overline{G} = \prod_{M \text{ maximal in } G} G/M.$$

Then the obvious mapping $G \to \overline{G}$ has kernel $\cap_{M \text{ maximal}} M = \Phi(G)$, so $G/\Phi(G)$ embeds in $\overline{G}$. Now each G/M has no proper subgroups by maximality, so has prime order. Thus $G/\Phi(G)$ embeds in the direct product of groups of prime order, so has square-free exponent, and by the fundamental theorem on finite abelian groups is itself such a direct product.

Conversely suppose that $N \lhd G$ and G/N is the direct product of groups H_i of prime order. For each such direct factor, the product K_i of the remaining direct factors is a maximal subgroup. The preimage M_i of K_i in G is maximal, and then $N = \cap_i M_i \geq \Phi(G)$. □

5. $E(G)$; Quasisimple and Semisimple Groups

DEFINITION 5.1. A group G is semisimple if and only if $G = E(G)$. Equivalently, G is semisimple if and only if G is perfect and chiefly inner.

The terminology is not standard; perhaps "semiquasisimple" would be a more accurate word, but it is cumbersome. Some authors use "semisimple" for groups which are the direct product of simple groups, a somewhat different notion.

EXAMPLES.

1. For any group G, $E(G)$ is semisimple.
2. If G is perfect and $G/Z(G)$ is the direct product of nonabelian simple groups, then G is semisimple. In particular if G is the product of quasisimple groups $G_1, \ldots, G_n$ and $[G_i, G_j] = 1$ for all $i \neq j$, then $G/Z(G_1) \cdots Z(G_n)$ is the direct product of simple groups and $Z(G_i) \leq Z(G)$ for all i, so G is semisimple.

The next theorem shows that this last example is the only example.

THEOREM 5.2. *If G is semisimple, then there exist subgroups $G_1, \ldots, G_n$ of G such that*

1. *Each G_i is quasisimple.*
2. *$G = G_1 \cdots G_n$.*
3. *$[G_i, G_j] = 1$ for all $i \neq j$.*

Moreover, the list of subgroups $G_1, \ldots, G_n$ is uniquely determined, up to order.

DEFINITION 5.3. The groups G_i in the preceding theorem are called the components of G. For an arbitrary (not necessarily semisimple) group G, the components of G are defined to be the components of $E(G)$.

PROOF. Assume that G is semisimple and set $Z = Z(G)$ and $\overline{G} = G/Z$. Then $F(G) \leq Z$ by the decomposition theorem for $F^*(G)$, so equality holds. As in the proof of that result, $\overline{G}$ is the direct product $\overline{G} = \overline{G}_1 \times \cdots \times \overline{G}_n$ of nonabelian simple groups. For each i let H_i be the full preimage in G of $\overline{G}_i$, and set $G_i = H_i^{(\infty)}$. Then since $\overline{G}_i$ is perfect, the image of G_i in $\overline{G}$ is still $\overline{G}_i$.

We argue that the required conditions hold. First of all, $[\overline{G}_i, \overline{G}_j] = 1$ for all $i \neq j$, and so $[G_i, G_j] \leq Z(G)$. Consequently $[G_i, G_j, G_i] = 1$, and the Three Subgroups Lemma and the fact G_i is perfect implies that $[G_i, G_j] = 1$. In particular, if we set $H = G_1 \cdots G_n$ then H is a subgroup of G, mapping onto $\overline{G}_1 \cdots \overline{G}_n = \overline{G}$, so $G = HZ(G)$. Thus $H \lhd G$ and so $G/H \cong Z(G)/Z(G) \cap H$ is abelian, whence $H = G$ as G is perfect by assumption. Finally G_i is perfect, $G_i \cap Z(G) \leq Z(G_i)$, and $G_i/G_i \cap Z(G) \cong \overline{G}_i$ is nonabelian simple, so G_i is quasisimple.

To complete the proof we must prove the uniqueness of the G_i. Suppose then that $K_1, \ldots, K_m$ satisfy the same conditions. Then the $\overline{K}_i$ are normal subgroups of $\overline{G}$, so are products of some of the $\overline{G}_i$; but K_i is quasisimple, so $\overline{K}_i$ has at most one nonsolvable composition factor. Hence every $\overline{K}_i$ is some $\overline{G}_k$. As $[\overline{K}_i, \overline{K}_j] = 1$ for $i \neq j$, distinct K_i's give distinct $\overline{K}_i$'s. Also $\overline{G} = \overline{K}_1 \cdots \overline{K}_m$ so every $\overline{G}_k$ is some $\overline{K}_i$. Hence $m = n$ and we may renumber so that $\overline{K}_i = \overline{G}_i$ for all i. Thus $K_i Z = G_i Z$. But then K_i and G_i are both perfect subgroups of this group with abelian factor groups, so both contain and then equal the commutator subgroup of this group. Hence $K_i = G_i$. □

As observed at the end of Section 3, we have:

PROPOSITION 5.4. *$E(G)$ is the largest normal semisimple subgroup of G.*

To summarize:

THEOREM 5.5.

1. *A group is semisimple if and only if it is the elementwise-commuting product of quasisimple groups.*
2. *A group is nilpotent if and only if it is the elementwise-commuting product of groups of prime power order.*
3. *A group is chiefly inner if and only if it is the elementwise-commuting product of a semisimple group and a nilpotent group.*
4. *$F^*(G)$ is the largest chiefly inner normal subgroup of G.*
5. *$C_G(F^*(G)) \leq F^*(G)$.*

PROOF. (1) and (2) follow from 5.2 and 4.4. If G is chiefly inner, then $G = F^*(G) = E(G)F(G)$, a commuting product, by 3.7. Semisimple and nilpotent groups are chiefly inner by definition and the commuting product of two chiefly inner groups is easily seen to be chiefly inner, proving (3). Every chiefly inner normal subgroup of G lies in $F^*(G)$ by the decomposition 3.7 and the uniqueness results 4.2 and 5.4. Now 2.7 completes the proof. □

The components of a semisimple group can be characterized as the normal quasisimple subgroups, or slightly more valuably as the subnormal quasisimple subgroups.

DEFINITION 5.6. The subgroup H of G is subnormal in G (notation: $H \lhd\lhd G$) if and only if there is a chain $H = H_0, H_1 \ldots, H_n = G$ of subgroups of G such that $H_{i-1} \lhd H_i$ for each $i = 1, \ldots, n$.

The following properties are then easy to verify:

(a) If $H \lhd G$ then $H \lhd\lhd G$.
(b) If $K \lhd\lhd H \lhd\lhd G$ then $K \lhd\lhd G$.
(c) If $K \lhd\lhd G$ and $H \leq G$, then $K \cap H \lhd\lhd H$.
(d) If $K \lhd\lhd G$ and $K \leq H \leq G$, then $K \lhd\lhd H$.
(e) G is nilpotent if and only if $H \lhd\lhd G$ for every $H \leq G$.

However, there is no "subnormalizer" of an arbitrary subgroup; if $H \leq G$, $K, L \leq G$ and $H \lhd\lhd K$ and $H \lhd\lhd L$, it does not follow that $H \lhd\lhd \langle K, L\rangle$. Example: Let $G = S_5$, $H = (12)$, $K = \langle (12), (13)(24)\rangle \in \mathrm{Syl}_2(G)$, $L = H \times \langle (345)\rangle$. Then $H \lhd\lhd K$ and $H \lhd\lhd L$ but $\langle K, L\rangle = G$ and H is not subnormal in G (see the next lemma).

The following property is proved trivially by induction:

LEMMA 5.7. *If $H \lhd\lhd G$ then $F(H) \leq F(G)$, $E(H) \leq E(G)$ and $F^*(H) \leq F^*(G)$. In particular if $H \lhd\lhd G$ and H is nilpotent (resp. semisimple) then $H \leq F(G)$ (resp. $H \leq E(G)$).*

PROOF. To prove the first statement, induction trivially reduces us to the case $H \lhd G$, in which case the statements have been proved. The second statement follows immediately as $H = F(H)$ (resp. $H = E(H)$). □

PROPOSITION 5.8. *Let $G = E(G) = G_1 \cdots G_n$ be semisimple, as in Theorem 5.5. Then the G_i are subnormal quasisimple subgroups of G and are the only such subgroups.*

PROOF. Clearly $G = G_1 \cdots G_n$ normalizes each G_i. It remains to show that if $X \lhd\lhd G$ and X is quasisimple, then $X = G_i$ for some i. But the image $\overline{X}$ of X in $\overline{G} = G/Z(G) = \overline{G}_1 \times \cdots \times \overline{G}_n$ is subnormal. But the only normal subgroups in this direct product of nonabelian simple groups are products of some factors, and by repetition of this argument, these are the only subnormal subgroups. Since $\overline{X}$ has a unique nonabelian composition factor, $\overline{X} = \overline{G}_i$ for some i. Then $XZ(G) = G_iZ(G)$, and then as at the end of the proof of 1.5, $X = O^{(\infty)}(XZ(G)) = O^{(\infty)}(G_iZ(G)) = G_i$. □

PROPOSITION 5.9. *The components of an arbitrary group G are the subnormal quasisimple subgroups of G.*

PROOF. The components of G are those of $E(G)$, so are subnormal in $E(G)$, hence subnormal in G. Conversely if $K \lhd\lhd G$ and K is quasisimple, then $K = E(K) \leq E(G)$ by 5.7, and so K is a component of $E(G)$ by 5.8, hence one of G. $\square$

COROLLARY 5.10. *If $H \lhd\lhd G$ then every component of H is a component of G.*

At this point we introduce some examples showing the power of $F^*(G)$ as a descriptive tool for a finite group G.

First, let us begin an analysis which when carried a bit further would give a classification up to isomorphism of all groups of order 24. Let G be a group of order 24 and let S and T be Sylow 3- and 2-subgroups of G. Then $|G : T| = 3$ and so the action of G on the right coset space $T\backslash G$ shows that G has a normal subgroup N (the kernel of this action) such that G/N embeds in Σ_3 and N is a 2-group. Consequently G is solvable.

Set $F = F(G)$. By the F^*-theorem and the solvability of G,

$$C_G(F) \leq F.$$

We also know that F is nilpotent, so is the direct product of a 2-group and a 3-group. This focusses our attention on the relevant questions: what are all the nontrivial 2- and 3-subgroups P of orders dividing 24, and what are their automorphism groups? The answer is:

P	Z_3	Z_2	Z_4	$Z_2 \times Z_2$	Z_8	$Z_4 \times Z_2$	$Z_2 \times Z_2 \times Z_2$	D_8	Q_8
$\mathrm{Aut}(P)$	Z_2	1	Z_2	S_3	$Z_2 \times Z_2$	8	$GL_3(2)$	D_8	S_4

Here the only assertion being made about $\mathrm{Aut}(Z_4 \times Z_2)$ is that it has order 8. This can be seen as follows: define a basis of $Z_4 \times Z_2$ to be an order pair (x, y) such that $Z_4 \times Z_2 = \langle x\rangle \times \langle y\rangle$ and such that $\langle x\rangle \cong Z_4$ and $\langle y\rangle \cong Z_2$. It is easy to check that $\mathrm{Aut}(Z_4 \times Z_2)$ permutes regularly the set of such bases. But $Z_4 \times Z_2$ has exactly 4 elements of order 4, and for each such element x there exist exactly two elements y of order 2 such that $y \notin \langle x\rangle$. So there exist 8 bases and hence 8 automorphisms. (The same procedure can be used to make calculations about automorphisms of any finite abelian group.) The structure of $Z_2 \times Z_2 \times Z_2$ as abelian group is identical to the structure of a 3-dimensional vector space over $\boldsymbol{F}_2$, explaining that entry.

6. A Refined F^*-Theorem

According to the F^*-Theorem, every element of $G - F^*(G)$ induces a non-inner automorphism on $F^*(G)$. Control of the (often large) factor group $G/F^*(G)$, or even $G/Z(F^*(G))$, is therefore dependent on knowledge of the automorphism group of $F^*(G)$. This control actually depends only on the knowledge of the automorphism group of a quotient group of $F^*(G)$ which is the direct product of simple groups.

The decomposition

$$F^*(G) = E(G)F(G) = (G_1 \cdots G_m)(P_1 \times \cdots P_n)$$

as the elementwise-commuting product of the components G_i and the Sylow subgroups $P_1, \ldots, P_n$ of $F(G)$, together with the fact that $E(G)$ and P_i are characteristic in $F^*(G)$, making them normal in G, together with the fact that the set of components of $E(G)$ is uniquely determined, immediately yields:

PROPOSITION 6.1. *Let the G_i and P_i be as above and let G_0 be the subgroup of G consisting of all elements normalizing each of $G_1, \cdots, G_m$. Then $F^*(G) \leq G_0 \lhd G$, and the permutation action of G on $\{G_1, \ldots, G_m\}$ yields an embedding*

$$G/G_0 \leq \Sigma_m.$$

Moreover, if we set $\overline{G}_i = G_i/Z(G_i)$ and $\overline{P}_i = P_i/\Phi(P_i)$ for each i, then the actions of G_0 on the G_i and the P_i yield embeddings

$$G_0/Z(F^*(G)) \leq Aut(\overline{G}_1) \times \cdots \times Aut(\overline{G}_m) \times Aut(P_1) \times \cdots \times Aut(P_n) \text{ and}$$
$$G_0/F(G) \leq Aut(\overline{G}_1) \times \cdots \times Aut(\overline{G}_m) \times Aut(\overline{P}_1) \times \cdots \times Aut(\overline{P}_n).$$

The proof relies on the following elementary but basic fact about automorphisms of quasisimple groups.

PROPOSITION 6.2. *Let K be a quasisimple group and set $\overline{K} = K/Z(K)$, a simple group. Then the obvious action of $Aut(K)$ on $\overline{K}$ yields an embedding $Aut(K) \to Aut(\overline{K})$. In other words, any automorphism of K centralizing $K/Z(K)$ is trivial.*

PROOF. Let α be such an automorphism. For any $x \in K$, $x^\alpha = xx^\beta$, where $x^\beta \in Z(K)$. Thus $\beta : K \to Z(K)$, and since α and the identity are homomorphisms, we have

$$(xy)(xy)^\beta = (xy)^\alpha = x^\alpha y^\alpha = xx^\beta yy^\beta.$$

But x^β commutes with y and it follows that β is a homomorphism. Since $Z(K)$ is abelian, $\ker\beta \geq [K, K] = K$. Therefore $x^\beta = 1$ for all $x \in K$ and $\alpha = 1$. □

Examination of the proof shows that simplicity of $\overline{K}$ is not needed.

There is an analogue of this proposition for automorphisms of p-groups. We defer the proof to later sections.

PROPOSITION 6.3. *Let P be a p-group and α an automorphism of P of order relatively prime to p. If α acts trivially on $P/\Phi(P)$, then $\alpha = 1$.*

PROOF OF PROPOSITION 6.1. The first embedding is obvious. Now by Proposition 6.2, $\mathrm{Aut}(G_i) \cong \mathrm{Aut}(G_i/Z(G_i))$. Hence the second embedding follows immediately by the F^*-Theorem. The third embedding is the strongest statement.

Let C be the set of all elements of G acting trivially by conjugation on each $G_i/Z(G_i)$ and on each $P_i/\Phi(P_i)$. The embedding we want will follow once we prove that $C = F(G)$. By an exercise in Section 4, since $F(G) = P_1 \times \cdots \times P_n$, we have

$$\Phi(F(G)) = \Phi(P_1) \times \cdots \times \Phi(P_n),$$

the direct product of elementary abelian groups. In particular $F(G)$ acts trivially on each $P_i/\Phi(P_i)$. Since also $[F(G), E(G)] = 1$, the definition of C yields that

$$F(G) \leq C.$$

To establish equality it will suffice to show that C is nilpotent. Thus by Theorem 4.8 it will suffice to show that $C/\Phi(C)$ is nilpotent. But $\Phi(F(G)) \leq \Phi(C)$ since $F(G) \lhd C$, and so it will suffice to show that $C/\Phi(F(G))$ is nilpotent. Set $\overline{C} = C/\Phi(F(G))$.

By the definition of C, we know that $[C, P_i] \leq \Phi(P_i)$ for each i, and hence $[C, F(G)] \leq \Phi(F(G))$. Thus

$$[\overline{C}, \overline{F(G)}] = 1, \text{ that is, } \overline{F(G)} \leq Z(\overline{C}).$$

Hence every chief factor of $\overline{C}$ within $\overline{F(G)}$ is certainly acted on trivially by $\overline{C}$. A group is nilpotent if and only if it acts trivially on all its chief factors, so it suffices to show that $\overline{C}$ acts trivially on all its chief factors between $\overline{F(G)}$ and $\overline{C}$. Thus we get our last reduction: it is enough to show that

$$\overline{C}/\overline{F(G)} \text{ is nilpotent, that is, } C/F(G) \text{ is nilpotent.}$$

To demonstrate this and thereby complete the proof, we use characterization (4) of nilpotence in Theorem 4.4: we must show that if x and y are elements of C whose images modulo $F(G)$ have relatively prime orders, then $[x, y] \in F(G)$. We may replace x and y by other elements with the same images modulo $F(G)$, and by doing this we may assume that the orders of x and y are relatively prime. (See the exercise below.) Now since $x, y \in C$, both x and y act trivially on each $G_i/Z(G_i)$, hence on each G_i by Proposition 6.2, hence on $E(G)$. Therefore certainly

$$[x, y] \text{ centralizes } E(G).$$

Similarly for any i, the elements x and y act trivially on $P_i/\Phi(P_i)$. Since x and y have relatively prime orders, one of them has order relatively prime to $|P_i|$, and so one of them acts trivially on P_i by Proposition 6.3. Consequently $[x, y]$ centralizes P_i. Since i is arbitrary,

$$[x, y] \text{ centralizes } F(G) \text{ and therefore centralizes } E(G)F(G) = F^*(G).$$

By the F^*-Theorem, $[x, y] \in C_G(F^*(G)) = Z(F(G)) \leq F(G)$. □

EXERCISE. Suppose that $N \lhd X$ and set $\overline{X} = X/N$. If $x \in X$ then there exists $y \in X$ such that $\overline{y} = \overline{x}$ and such that the orders of y and $\overline{x}$ are divisible by the same primes (although not necessarily to the same powers). In fact, y may be chosen to be a suitable power of x.

7. Automorphisms of p-Groups; Coprime Action

Study of the action of G on $F(G)$ opens up a theory of the action of p'-groups on p-groups. (A p'-group is one whose order is not divisible by p.) A formal semidirect product construction makes it possible to get some of these theorems from the F^*-Theorem. The construction is as follows.

PROPOSITION 7.1. *Let H and K be groups and let H act on K via automorphisms, i.e., assume that there is given a homomorphism*

$$\phi : H \to Aut(K).$$

Then there is a group G containing H and K as disjoint subgroups, and such that $K \lhd G$, $G = HK$, and $k^h = k^{\phi}(h)$ for all $k \in K$ and $h \in H$.

PROOF. Let G set-theoretically be $H \times K$, the set of all ordered pairs (h, k) with $k \in K$ and $h \in H$. Define

$$(h, k)(h', k') = (hh', k^{\phi(h')}k').$$

It is then a routine matter to check that this is an associative binary operation, $(1, 1)$ is an identity element, and the inverse of (h, k) is $(h^{-1}, k^{\phi(h^{-1})})$. Moreover the injections $k \mapsto (1, k)$ and $h \mapsto (h, 1)$ of K and H into G are homomorphisms, and identifying H and K with their images we obtain the result. □

One could use the notation $G = K \rtimes_\phi H$.

REMARKS.

1. An arbitrary group M is said to be a semidirect product of subgroups A and B if and only if $M = AB$, $A \cap B = 1$ and $A \triangleleft M$. It is not necessarily the case that a group X with a normal subgroup A is a semidirect product of A by anything. The "splitting" question is nontrivial.
2. If a triple (H, K, ϕ) is "isomorphic" to another triple (H^*, K^*, ϕ^*), then the corresponding semidirect products are isomorphic. However, the semidirect products may be isomorphic without the triples being isomorphic. For example, if G is any group and we consider two actions of G on G: one by the mapping $\phi : G \to G$ taking $g \mapsto 1$ for all g, the other by the mapping $\psi : G \to G$ taking g to conjugation by G, then both semidirect products are isomorphic to $G \times G$, although the triples aren't isomorphic unless G is abelian.
3. If H is any subgroup of $\mathrm{Aut}(K)$, then the inclusion mapping ϕ of H into $\mathrm{Aut}(K)$ leads to a semidirect product KH, in which $C_{KH}(K) = Z(K)$.

THEOREM 7.2. *Let H be a p'-group acting on the p-group P. If H acts trivially on $P/\Phi(P)$, then it acts trivially on P.*

PROOF. It is enough to show this for a single element of H, thus for the case that H is a cyclic subgroup of $\mathrm{Aut}(P)$; we form the semidirect product $G = HP$ and must show that $C_H(P/\Phi(P)) = 1$. Since $H \leq \mathrm{Aut}(P)$, and H is a cyclic p'-group, $F^*(G) = F(G) = P$. Therefore $\Phi(P) \leq \Phi(G)$ and $F^*(G/\Phi(P)) = F^*(G)/\Phi(P) = P/\Phi(P)$. By the F^*-Theorem, $C_G(P/\Phi(P)) = P/\Phi(P)$, and the result follows. □

THEOREM 7.3. *Let H be a p'-group acting on the p-group P. Then $P = C_P(H)[P, H]$.*

PROOF. The proof is by induction on $|P|$. Set $Q = [P, H]$ and $C = C_P(H)$. We have $Q \triangleleft HP$, so CQ is a group. By induction every proper H-invariant subgroup of P lies in CQ. In particular $\Phi(P) \leq CQ$.

If $\Phi(P) \neq 1$, then by induction $\overline{P} = C_{\overline{P}}(H)[\overline{P}, H]$. Both of these subgroups are H-invariant, so if they are proper then their inverse images lie in CQ by induction, and $P = CQ$. So we may assume that $[\overline{P}, H] = 1$ or $\overline{P}$. In the first case, H centralizes P by the previous result, so $C = P$. In the second case, $[P, H]\Phi(P) = P$, so $Q = P$. In either case we are done.

So assume that $\Phi(P) = 1$, that is, P is elementary abelian. This time let $\overline{P} = P/Q$. Therefore $[\overline{P}, H] = 1$. Let $x \in P$; it suffices to find $y \in C$ such that $y \equiv x \mod Q$. The element $y_1 \prod_{g \in H} x^g$ almost satisfies the requirements. Namely, since P is abelian it is clear that $y_1 \in C$. Also $x^g = x[x, g] \in xQ$, soi $\overline{y}_1 = \overline{x}^{|H|}$.

Since H is a p'-group, $\overline{x}$ is therefore a power $\overline{y}_1^m$ of y_1. Then $y = y_1^m$ satisfies the requirements. □

PROPOSITION 7.4. *Suppose that H is a p'-group acting on a p-group P. Let Q be an H-invariant normal subgroup of P. Then any element of $C_{P/Q}(H)$ is the image of an element of $C_P(H)$.*

PROOF. Without loss we may replace P/Q by the preimage of $C_{P/Q}(H)$ and assume that H centralizes P/Q. We must then argue that $C_P(H)Q = P$. But $[P, H] \leq Q$ so this holds by the previous result. □

PROPOSITION 7.5. *Let H be a p'-group acting on a p-group P. Then*

$$[P, H, H] = [P, H].$$

PROOF. $[P, H, H]$ is normalized by H and $[P, H]$. It also is clearly normalized by $C_P(H)$, and hence by $[P, H]C_P(H) = P$. Passing to $P/[P, H, H]$ we may assume that $[P, H, H] = 1$, and must prove that $[P, H] = 1$ already. Since $[P, H, H] = 1$, we have $[P, H] \leq C_P(H)$. But then $P = [P, H]C_P(H) = C_P(H)$ so $[P, H] = 1$. □

PROPOSITION 7.6. *Let the p'-group H act on the p-group P and set $C = C_P(H)$. If $C_P(C) \leq C$, then $C = P$.*

PROOF. Let $N = N_P(C)$. Then $[N, C] \leq C$ so $[N, C, H] = 1$. Also $[C, H] = 1$ so $[C, H, N] = 1$. By the Three Subgroups Lemma $[H, N, C] = 1$, so $[H, N] \leq C_P(C) = leC$. Therefore $[N, H, H] = 1$ so $[N, H] = 1$, so $N \leq C$. Therefore $N = C$. But P is nilpotent, so $C = P$. □

THEOREM 7.7 (THOMPSON'S $A \times B$ LEMMA. *Suppose that A is a p'-group, B is a p-group, and $A \times B$ acts on the p-group P in such a way that $[A, C_P(B)] = 1$. Then $[A, P] = 1$.*

PROOF. Forming the semidirect product $P(A \times B)$ and setting $Q = PB$, we have the p'-group A acting on the p-group B. Let $C = C_Q(A)$. Then $B \leq C$ and $C_P(B) \leq C$ by assumption. Let $x \in C_Q(C)$ and write $x = yz$, $y \in P$, $z \in B$. For any $b \in B$, x and b commute so $x = x^b = y^b z^b = yz$, and by the uniqueness of expression in semidirect products, $y^b = y$. Therefore $y \in C_P(B) \leq C$, and as $z \in B \leq C$ we have $x \in C$. By the previous result, A centralizes Q and *a fortiori* centralizes P. □

An analogue of 7.6 can be proved for arbitrary groups.

THEOREM 7.8. *Suppose that G is an arbitrary group and α is an automorphism of G such that*

1. *α centralizes $F^*(G)$, and*
2. *The order of α is relatively prime to $|F(G)|$.*

Then $\alpha = 1$.

PROOF. Assume that $\alpha \neq 1$. We may replace α by a power of itself and assume that α has prime order p. Form the semidirect product $H = G\langle\alpha\rangle$.

Every component of $E(H)$ lies in G, hence in $E(G)$. Therefore $F^*(H) \cap G = E(H)(F(H) \cap G) \leq F^*(G)$, and so equality holds. If $F^*(H) = F^*(G)$, then $\alpha \in$

$C_H(F^*(H)) \leq F^*(H) = F^*(G)$, contradiction. Therefore $F^*(H) > F^*(G)$, so $F(H) > F(G)$. Since $p = |H : G|$ does not divide $|F(G)|$, it follows that $F(H) = F(G) \times Z$, with $Z \cong Z_p$ and $Z \not\leq G$. But then $Z = O_p(F(H)) \lhd H$ and so $H = Z \times G$. It follows that α is an inner automorphism, say corresponding to $x \in G$. But then $x \in C_G(F^*(G)) \leq F(G)$, so the order of x is relatively prime to p. Therefore so is the order of α and so $\alpha = 1$.

8. Local Subgroups; Groups of Characteristic p-Type

In this section we use the preceding results to illustrate some of the techniques of local analysis of simple groups. The main idea of local analysis is to take a simple group (satisfying some side conditions) and use the simplicity to deduce information about its local subgroups. If the side conditions are strong enough, then the local subgroup information obtained might be enough to allow one to identify the isomorphism type of the simple group. Thus there are two aspects of local analysis: the passage from global information (simplicity) to local information about the group, and the passage from local information to global information (the isomorphism type). Much of the classification of finite simple groups is a successful translation of the condition of simplicity to information about the local structure of a minimal counterexample, and then to information about its isomorphism type.

DEFINITION 8.1. A p-local subgroup of a group G is a subgroup of the form

$$N = N_G(P)$$

for some p-subgroup P of G with $P \neq 1$.

THEOREM 8.2. *If $F^*(G)$ is a p-group for some prime p, then $F^*(N)$ is also a p-group for every p-local subgroup N of G.*

PROOF. Say $N = N_G(P)$ and let $F = F^*(G)$ and $F_0 = F^*(N)$. Let x be any p'-element of F_0. We show that

$$[x, F] = 1,$$

whence $x \in F$ by the F^*-Theorem. Since x is a p'-element and F is a p-group, $x = 1$. Since x is an arbitrary p'-element of F_0, it will follow by Sylow's Theorem that F_0 is a p-group.

Consider $\langle x \rangle \times P$ acting on F. To achieve our goal it is enough to show that $[x, C_F(P)] = 1$, for then the $A \times B$ Lemma finishes the proof. Notice that $F = O_p(G)$, so $N_F(P)$ is a normal p-subgroup of $N = N_G(P)$. Thus $N_F(P) \leq O_p(N) = O_p(F_0)$. But also the p'-element $x \in F_0 = F(N)$ satisfies $x \in E(F_0)O_{p'}(F_0)$, since $F_0/E(F_0)O_{p'}(F_0)$ is the image of $O_p(F_0)$ and hence is a p-group. Therefore

$$[x, N_F(P)] \leq [O_p(F_0), E(F_0)O_{p'}(F_0)] = 1.$$

A fortiori $[x, C_F(P)] = 1$, completing the proof. □

DEFINITION 8.3. A nonabelian simple group G is of characteristic p-type if and only if $F^*(N) = O_p(N)$ for every p-local subgroup N of G. We may say that p is a characteristic prime for G.

For instance, A_5 is of characteristic 2-type, 3-type and 5-type. So is A_6. But A_7 is not of characteristic 2- or 3-type.

The reason for the terminology will shortly be clear. First notice that 8.2 means that the definition can be simplified.

LEMMA 8.4. *Let G be a nonabelian simple group. Then G is of characteristic p-type if and only if $F^*(N) = O_p(N)$ for every* **maximal** *p-local subgroup of G.*

PROOF. One direction is obvious. In the other direction, let N be an arbitrary p-local subgroup and expand N to a maximal p-local subgroup M of G. We have $N = N_G(P)$ for some p-group $P \leq N$, and so $N = N_M(P)$ with $P \leq M$. Therefore $F^*(N)$ is a p-group by the preceding result.

As an example we consider the groups $PSL_n(q)$, $q = p^m$.

THEOREM 8.5. *The groups $PSL_n(q)$, $n \geq 2$, $q = p^m$, are of characteristic p-type.*

To prove this we first make a couple of simplifying remarks. Set $G = SL_n(q)$ and $Z = Z(G)$, so that $G/Z \cong PSL_n(q)$ and Z has order dividing $q-1$; in particular Z is a p'-group.

LEMMA 8.6. *Suppose that $Y \lhd X$ and Y is a p'-group. Set $\overline{X} = X/Y$. Then for any p-subgroup P of X we have*

$$N_{\overline{X}}(\overline{P}) = \overline{N_X(P)}.$$

Consequently the mapping $N \mapsto \overline{N}$ is a bijection between the set of p-local subgroups of X and the set of p-local subgroups of $\overline{X}$.

PROOF. If x normalizes P then clearly $\overline{x}$ normalizes $\overline{P}$. Thus $N_{\overline{X}}(\overline{P})$ contains $\overline{N_X(P)}$ (and this inclusion is true without any extra hypotheses about Y or P). Conversely, let N be the full preimage in X of $N_{\overline{X}}(\overline{P})$. We must show that $N \leq N_X(P)Y$. But since $\overline{P} \lhd \overline{N}$ and PY is the full inverse image of $\overline{P}$, it must be that $PY \lhd N$. Since Y is a p'-group, $P \in \mathrm{Syl}_p(PY)$. By the Frattini argument, $N = PYN_N(P)$ and so $N \leq YN_X(P)$. This proves the displayed equation, and the second statement follows immediately. □

Now setting $\overline{G} = G/Z \cong PSL_n(q)$, we know that any maximal p-local subgroup of $\overline{G}$ has the form $\overline{N}$ for some maximal p-local subgroup N of G. Moreover, by the exercise at the end of Section 2, since $Z = Z(G)$ we have

$$F^*(\overline{N}) = \overline{F^*(N)}.$$

Thus to prove the theorem it suffices to prove:

PROPOSITION 8.7. *If N is a maximal p-local subgroup of $G = SL_n(q)$, $n \geq 2$, $q = p^m$, then $F^*(N) = Q \times Z$, where $Q = O_p(N)$, the largest normal p-subgroup of Q.*

We identify G with $SL(V)$, where V an n-dimensional vector space over $\boldsymbol{F}_q$. For each proper subspace W of V, we have the stabilizer G_W, a subgroup of G. We shall show that

(a) For any W, G_W is a maximal subgroup of G;
(b) For any W, $F^*(G_W) = O_p(G_W) \times Z$, with $O_p(G_W) \neq 1$; and
(c) Every p-local subgroup of G lies in some G_W.

Since we know that G has no nontrivial normal p-subgroup, (b) will imply that G_W is a p-local subgroup, and then (a) and (c) wil imply that the G_W are all the maximal p-local subgroups of G, and then (b) will yield the theorem.

To prove (a) we set $d = \dim W$ and consider the natural action of G on the set Ω of d-dimensional subspaces of V. It is quite clear that this action is transitive, and G_W is a point stabilizer. So to prove (a) we must show that the action of G on Ω is primitive.

Suppose that Ψ is a block for the action of G on Ω, with $|\Psi| > 1$. We must show that $\Psi = \Omega$. Replacing Ψ by Ψ^g for some $g \in G$ we may assume that $W \in \Psi$. Choose a second $W_1 \in \Psi$ and set $i = \dim(W \cap W_1)$. Consider the graph whose vertex set is Ω, with two vertices X, Y joined by an (undirected) edge if and only if $\dim(X \cap Y) = i$. (We speak of X and Y as "neighbors".) Thus W and W_1 are neighbors. It is obvious that the action of G on Ω preserves the graph we have just defined. Moreover, it is not difficult to show that

$$G_W \text{ is transitive on the set of neighbors of } W,$$

that is, G_W is transitive on the set of d-dimensional subspaces of V meeting W in an i-dimensional subspace. But G_W, fixing W, must carry Ψ to itself, and since $W_1 \in \Psi$ it follows that all neighbors of W in Ω lie in Ψ.

Indeed, for any $W_0 \in \Psi$, there is $g \in G$ carrying W to W_0; then g must carry the block Ψ to itself and so all neighbors of W_0 in Ω lie in Ψ. We have proved that

$$\text{the connected component of } W \text{ in } \Omega \text{ lies entirely in } \Psi.$$

We leave as an exercise for the reader to prove the following fact about subspaces of V:

$$\Omega \text{ is connected.}$$

The consequence is that $\Psi = \Omega$, and so assertion (a) above is proved.

To prove (b), we choose an appropriate basis of V so that G_W consists of all block matrices of the form

$$g = \begin{matrix} A & B \\ 0 & C \end{matrix}.$$

Here A and C have to be nonsingular, of course, with $\det A \det C = 1$; but otherwise there are no restrictions on A, B and C. There is a homomorphism

$$\phi : G_W \to GL_d(q) \times GL_{n-d}(q),\ g \mapsto (A, C)$$

its kernel is the subgroup Q defined by the condition $B = 0$, and has order $q^{d(n-d)}$. In particular Q is a p-group. In fact when two elements of Q are multiplied, their upper right blocks just add, so Q is isomorphic to the direct product of $d(n-d)$ copies of $\boldsymbol{F}_q^+$, and so is an elementary abelian p-group.

As g varies over G_W, A and C vary over the corresponding general linear groups (although not independently). It follows that

$$O_p(\phi(G_W)) = 1.$$

Indeed if R is a normal p-subgroup of $\phi(G_W)$ then its projections are normal p-subgroups of $GL_d(q)$ and $GL_{n-d}(q)$, respectively; but we know that there are no such normal subgroups. Consequently

$$Q = O_p(G_W).$$

Furthermore, for the matrix g above to lie in $C(Q)$, we find by a matrix calculation that it is necessary and sufficient for A and C to be scalar matrices with the same eigenvalue. Consequently

$$C_{G_W}(Q) = Q \times Z.$$

The preceding two equations imply that

$$F^*(G_W) = QC_{F^*(G_W)}(Q) = Q \times Z,$$

proving (b).

Finally the proof of (c) rests on the following fact about representations of p-groups:

LEMMA 8.8. *Let V be a finite-dimensional vector space over a field $\boldsymbol{F}$ of characteristic $p > 0$. Let P be a finite p-group (for the same prime p). Then P fixes a nonzero vector in V.*

PROOF. The proof is by induction on $|P| + \dim V$. Using induction we reduce to the case that P acts faithfully on V and irreducibly as well, that is, P leaves invariant no proper nonzero subspace of V. If $P = 1$ there is nothing to prove; otherwise choose $z \in Z(P)$ of order p. Let Z be the corresponding linear transformation of V. Then $Z^p = 1$, so $(Z-1)^p = 1$ as $\boldsymbol{F}$ has characteristic p. Then 1 is the only eigenvalue of Z, and the existence of a corresponding eigenvector shows that $C_V(z)$ is nontrivial. But $C_V(z)$ is left invariant by P, since $z \in Z(P)$. Therefore $C_V(z) = V$ by irreducibility, contradicting the faithfulness of P on V.

Now in (c) above, if $N = N_G(P)$ is a maximal p-local subgroup of G, then the subspace W of vectors in V fixed by each element of P is therefore a proper nontrivial subspace. But it is left invariant by N and so $N \leq G_W$. This completes the proof of Theorem 8.5.

However, notice that if r is a prime different from p, for instance r dividing $q-1$, then a diagonal matrix x of order r and with d eigenvalues equal to 1 will have the property that $C_{\hat{G}_0}(x)$ has a subnormal subgroup isomorphic to $SL_d(q)$. If $d \geq 3$, or even if $d \geq 2$ and $q > 3$, this implies that $E(C_G(x^*)) \neq 1$, where x^* is the image of x. An easily seen example of this is for $n = 6$, say, with $d = 4$. Then x is conjugate to no other element of xZ and so $C_G(x^*)$ is precisely the image of $C_{\hat{G}_0}(x)$.

Notice also that in the alternating groups $G = A_n$, as long as $n \geq 2p$ and $n - p \geq 5$, the centralizer C of a p-cycle satisfies $E(C) \cong A_{n-p}$, so G is not of characteristic p-type.

However, there is some pressure on simple groups to be of characteristic p-type for some prime p.

Another important remark is that for the simple groups $PSL_n(q)$, the p-local structure determines the group in the following specific sense. We have seen that the maximal p-local subgroups of $PSL_n(q)$ are in bijective correspondence with the set of proper subspaces of the underlying vector space, that is, with the elements of the corresponding projective geometry $\mathcal{P}$. The incidence relation in this geometry is also definable group-theoretically: two subspaces W and X of V are incident (i.e. one contains the other) if and only if the intersection $G_W \cap G_X$ lies in no other G_Y except for $Y = W$ and $Y = X$. The point we wish to emphasize is, in this case, called the fundamental theorem of projective geometry: the only automorphisms

of this geometry arise from $PGL_n(q)$, automorphisms of the field $\boldsymbol{F}_q$, and polarities interchanging subspaces of dimensions d and $n-d$ for all d and arising from identifications of V with its dual space. This leads to the characterization

$$PSL_n(q) = E(\mathrm{Aut}(\mathcal{P})).$$

So if one can get the p-local structure of an arbitrary simple group G to look like that of $PSL_n(q)$, then G embeds in $PSL_n(q)$, and this should lead without much trouble to $G \cong PSL_n(q)$.

We shall not prove the fundamental theorem of projective geometry, except that below we shall compute the automorphism group of $PSL_n(q)$, which is much the same thing.

In view of the above example, one might ask whether it is ever feasible to show that a group has the same p-local structure as $PSL_n(q)$. Surely one needs plenty of extra hypotheses for this. On the other hand, not only $PSL_n(q)$ but also all the finite classical matrix groups, and indeed all the finite simple groups of Lie type, have characteristic p-type for some prime p (the characteristic of the underlying field). Thus with the hindsight provided by the classification theorem, every finite simple group except most alternating groups and some of the sporadic groups have characteristic p-type for some prime p. Nothing like a proof of this is known that doesn't use the classification; a conceptual proof would be as exciting a development as one might imagine in this subject. Not much is known about characteristic primes of arbitrary finite simple groups, but we shall give two theorems (one without proof).

The first theorem, due to Bender, may begin to explain the penchant that simple groups have for having a characteristic prime, at least if their maximal subgroups have significant overlap.

Definition 8.9 (Bender). Let G be a nonabelian simple group and let M and N be maximal subgroups of G. We say that

$$M \rightsquigarrow N$$

if and only if $F^*(M) \cap N$ is "comparatively large" in the sense that it contains $N_{F^*(M)}(D)$ for some subgroup $D \leq F(M)$.

We may write

$$F(M) = P_1 \times P_2 \times \cdots \times P_r$$

where the P_i are the Sylow subgroups of $F(M)$, groups whose orders are powers of the distinct primes $p_1, \ldots, p_r$. Recall also that $F^*(M) = F(M)E(M)$ with $[F(M), E(M)] = 1$. The subgroup D of $F(M)$ in this definition is nilpotent, and so it has its own direct product decomposition:

$$D = D_1 \times D_2 \times \cdots \times D_r,$$

where $D_i \leq P_i$ for each i. Some or all of the D_i may be trivial. It then follows that if we set $W = N_{F^*(M)}(D)$ and $W_i = N_{P_i}(D_i)$ for each i, then

$$W = (W_1 \times W_2 \times \cdots \times W_r)E(M).$$

The definition asserts that $W \leq F^*(M) \cap N$.

LEMMA 8.6. *Suppose that $M \rightsquigarrow N$. Then*
1. *$E(M) \leq N$.*
2. *$Z(F(M)) \leq N$.*
3. *For any prime p, any p'-element of M centralizing $O_p(M) \cap N$ also centralizes $O_p(M)$.*

PROOF. We have $E(M) \leq W \leq N$, and likewise for each i, $Z(P_i) \leq N_{P_i}(D_i) = W_i \leq N$, proving the first two parts. Moreover, $D_i \leq W_i$ so $C_{P_i}(W_i) \leq C_{P_i}(D_i) \leq N_{P_i}(D_i) = W_i$. Moreover $P_i \cap N \geq W_i$ so any p'-element of M centralizing $P_i \cap N$ centralizes W_i, and then centralizes P_i by 7.6.

Now we are ready for the theorem.

THEOREM 8.10 (BENDER). *Suppose that G is a nonabelian simple group and that M and N are maximal subgroups such that*

$$M \rightsquigarrow N \text{ and } N \rightsquigarrow M.$$

Then either $M = N$, or else $F^*(M)$ and $F^*(N)$ are both p-groups for the same prime p.

Notice that in $G = L_n(q)$, any two upper block diagonal subgroups M and N are maximal subgroups with $F^*(M)$ and $F^*(N)$ contained in N and M respectively, and both $F^*(M)$ and $F^*(N)$ are p-groups.

PROOF. First a few observations. The hypotheses are symmetric in M and N; this symmetry will be invoked several times. The maximality of M implies that if $K \lhd M$, then $N_G(K) = M$ or G; but the latter can only occur if $K = 1$, by simplicity. Thus

$$M = N_G(K) \text{ for any nonidentity normal subgroup of } M.$$

In particular,

$$M = N_G(Z(O_p(M))) \text{ for any prime divisor } p \text{ of } |F(M)|,$$

and $M = N_G(E(M))$, etc. Similar statements hold for N.

We assume that

(8.1) $M \neq N$, whence M and N have no common nonidentity normal subgroup.

We now argue that the various "pieces" of $F^*(M)$ commute with all the "pieces" of $F^*(N)$, except for pairs of corresponding pieces.

LEMMA 8.11. *The following conditions hold:*
(a) $[E(M), F(N)] = [F(M), E(N)] = 1$*;*
(b) *For any distinct primes p and q, $[O_p(M), O_q(N)] = 1$; and*
(c) *A component of $E(M)$ and a component of $E(N)$ either coincide or commute elementwise.*

PROOF. To prove (a) it suffices to prove the first statement, by symmetry; thus we must show for each p that $[E(M), O_p(N)] = 1$. We do know that $E(M) \leq N$, so $E(M)$ acts on $O_p(N)$. In particular the commutator $[E(M), O_p(N) \cap M]$ lies in $O_p(N)$ so is a p-group. But it also lies in $E(M)$ since M normalizes $E(M)$. And this commutator is normalized by $E(M)$ (in general $[X, Y] \lhd \langle X, Y \rangle$). Therefore $[E(M), O_p(N) \cap M] \leq O_p(E(M))$. But $E(M)/Z(E(M))$ is the direct product of nonabelian simple groups, so $[E(M), O_p(N) \cap M] \leq Z(E(M))$. Thus

$$[E(M), O_p(N) \cap M, E(M)] = 1, \text{ so } [E(M), O_p(N) \cap M] = 1$$

by a usual application of the Three Subgroups Lemma. Consequently by the preceding lemma, every p'-element of $E(M)$ centralizes $O_p(N)$. Thus

$$[E, O_p(N)] = 1$$

where E is the subgroup of $E(M)$ generated by all its p'-elements. Clearly $E \lhd E(M)$, and E contains all Sylow subgroups of $E(M)$ for primes other than p, so $E(M)/E$ is a p-group and hence solvable. Since $E(M) = [E(M), E(M)]$ it follows that $E(M) = E$, and we have proved (a).

To prove (b), there is no loss in assuming that $O_p(M)$ and $O_q(N)$ are both nontrivial. Consider

$$[O_p(M) \cap N, O_q(N) \cap M] \leq O_p(M) \cap O_q(N) = 1. \tag{8.2}$$

Thus by the preceding lemma, $O_p(M) \cap N$ centralizes $O_q(N)$. But $Z(O_p(M)) \leq O_p(M) \cap N$, so

$$O_q(N) \leq C_G(O_p(M) \cap N) \leq C_G(Z(O_p(M))) \leq N_G(Z(O_p(M))) = M.$$

Thus $O_q(N) \cap M = O_q(N)$, and by symmetry $O_p(M) \cap N = O_p(M)$, so (8.2) is the statement we wished to prove.

As for (c), notice that $E(M) \leq M \cap N \leq M$; the components of $E(M)$ are therefore subnormal quasisimple subgroups of $M \cap N$ so are components of $E(M \cap N)$. So are the components of $E(N)$, by symmetry. Therefore any two such objects either coincide or commute elementwise, completing the proof. □

We can now prove that

$$E(M) = E(N) = 1, \text{ so that } F^*(M) = F(M) \text{ and } F^*(N) = F(N). \tag{8.3}$$

Let K be a component of $E(M)$. Then $K \leq N$, and by 8.11a,c, K commutes elementwise with or induces inner automorphisms on every component of $E(N)$ as well as on $F(N)$. Therefore K induces inner automorphisms on $F^*(N)$. But the F^*-Theorem gives an injection

$$N/Z(F(N)) \to \mathrm{Aut}(F^*(N)),$$

and the full preimage of $\mathrm{Inn}(F^*(N))$ is then clearly just $F^*(N)$. Consequently $K \leq F^*(N)$. But $K = [K, K]$, while $F^*(N)/E(N)$ is solvable, so $K \leq E(N)$. We have proved that

$$E(M) \leq E(N).$$

By symmetry equality holds, and then $E(M) = 1$ by (8.1). This proves (8.3).

Next we show that

$$|F(M)| \text{ and } |F(N)| \text{ have the same prime divisors.} \tag{8.4}$$

If q were a prime divisor of $|F(M)|$ but not of $|F(N)|$, it would follow from 8.11b that $O_q(M) \le C_G(F(N)) = C_N(F(N)) = Z(F(N))$ by the F^*-Theorem, so $O_q(M) = 1$, contradiction. Thus (8.4) holds.

We may now write

$$F(M) = P_1 \times \cdots \times P_r,\ F(N) = Q_1 \times \cdots \times Q_r,$$

where all the P_i and Q_i are nontrivial, and P_i and Q_i are p_i-groups with $p_1, \ldots, p_r$ distinct primes.

If $r = 1$, then the conclusion of the theorem holds, so we assume that $r > 1$ and derive a contradiction. Then $[P_i, Q_j] = 1$ for all $i \ne j$ by 8.11b, so $P_i \le C_G(Q_j) = N$. Thus $F(M) \le N$.

The group $P_1F(N)$ is nilpotent. (Indeed, it has a Sylow p_1-subgroup P_1Q_1, and the other Sylow subgroups are $Q_2, \ldots, Q_r$. Because $F(N)$ is nilpotent and because of 8.11b, these Sylow subgroups all commute elementwise, whence the nilpotence.) Next consider the groups P_1 and $C_N(Q_1)$. The first one centralizes $Q_2, \ldots, Q_r$, and the second one centralizes Q_1, so their commutator $[P_1, C_N(Q_1)]$ centralizes $Q_1Q_2 \cdots Q_r = F(N) = F^*(N)$ and hence lies in $Z(F(N))$ by the F^*-Theorem. Therefore modulo $F(N)$, these groups commute elementwise. Therefore $C_N(Q_1)$ normalizes $P_1F(N)$. So it normalizes $O_p(P_1F(N)) = P_1Q_1$. Now

$$[P_1, C_N(Q_1)] \le F(N) \cap P_1Q_1 = Q_1, \text{ whence } [P_1, C_N(Q_1), C_N(Q_1)] = 1.$$

Because of this and 7.5, $[P_1, x] = 1$ for any p'-element $x \in C_N(Q_1)$. Thus any such p'=element lies in $C_G(P_1) = C_M(P_1)$. We have shown that

$$\text{Every } p'\text{-element of } C_N(Q_1) \text{ lies in } C_M(P_1).$$

By symmetry, the set $\mathcal{A}$ of p'-elements of $C_N(Q_1)$ coincides with the set of p'-elements of $C_M(P_1)$. Since $Q_1 \lhd N$, $C_N(Q_1) \lhd N$ and similarly $C_M(P_1) \lhd M$. But then $\langle \mathcal{A} \rangle$ is a common normal subgroup of M and N, so is trivial by (8.1). But $P_2 \subseteq C_M(P_1)$ so $P_2 = 1$, a contradiction. This completes the proof.

A number of basic questions about the characteristic of a finite simple group remain open to this day. Which finite simple groups are not of characteristic p-type for some prime p? Only most alternating groups, and some sporadic groups. No proof of this independent of the classification is known. On the other hand, the characteristic is usually unique in the sense of the following theorem.

DEFINITION 8.12. If p is a prime, E is an elementary abelian p-group and G an arbitrary group, then $m_p(E)$ and $m_p(G)$ are defined by

$$p^{m_p(E)} = |E|$$
$$m_p(G) = \max\{m_p(E) \,|\, E \le G,\ E \text{ elem ab } p\text{-gp}\}.$$

THEOREM (KEN KLINGER & GEOFF MASON 1974). *Let G be a finite simple group which is of characteristic p-type and characteristic 2-type for some odd prime p. Then either $m_2(G) \le 2$ or $m_p(G) \le 2$.*

9. Involutions

In the previous section we mentioned the main problems of local analysis: global to local, using the simplicity of G to obtain information about local structure, and then local to global, using local information to determine the isomorphism type of G. We take a brief side-trip to mention the special importance of 2-local information in the local to global problem. Elements of order 2 are called involutions, and the most important 2-local subgroups are the centralizers of involutions. We prove a couple of results here, due to Brauer and Fowler and Thompson, which demonstrate the powerful effect of the structure of centralizers of involutions on the structure of a simple group. What's special about involutions is the following elementary fact, which has no analogue for other primes.

LEMMA 9.1. *Let x and y be involutions in a group. Then*

$$(xy)^x = (xy)^y = (xy)^{-1}.$$

The group $\langle x, y\rangle$ is a dihedral group of order $2n$, where $n = |xy|$, unless $x = y$ in which case $\langle x, y\rangle = \langle x\rangle \cong Z_2$.

PROOF. Since $x^{-1} = x$ and $y^{-1} = y$, $(xy)^x = xxyx = yx = (xy)^{-1}$. Therefore the dihedral group $a^n = 1$, $b^2 = 1$, $a^b = a^{-1}$ of order $2n$ maps onto $\langle x, y\rangle$, with $a \mapsto xy$ and $b \mapsto x$. This mapping is an isomorphism unless $x \in \langle xy\rangle$; in this latter case, then also $y \in \langle xy\rangle$ so $x = y$ as $\langle xy\rangle$ is cyclic.

Consequently

Any two involutions of a finite group G lie in a local subgroup of G.

THEOREM 9.2 (BRAUER-FOWLER). *Let G be a finite simple group of even order, and x an involution of G. Then*

$$|G| \leq (2|C_G(x)|^2)!$$

The precise formula is too crude to be valuable except in the philosophical sense that $|G| \leq f(|C_G(x)|)$ for some function f, independent of G.

If the simplicity assumption is dropped, then no such result holds, as is demonstrated by the family of dihedral groups. In a dihedral group, which can have arbitrarily large order, the centralizer of a reflection (i.e., non-central involution) has order just 4.

PROOF. Let $g = |G|$ and $c = |C_G(x)|$, and let m be the minimum of the numbers $|G : C_G(y)|$ as y ranges over $G^\# = G - \{1\}$. Since G is simple, $m > 1$, and the permutation action of G on the conjugacy class of the corresponding element y gives an injection $G \to S_m$. It therefore suffices to show that

$$m \leq 2c^2.$$

Let C be the conjugacy class of x in G, so that $|C| = g/c$ and the number of pairs (u, v) with $u, v \in C$ is just g^2/c^2. On the other hand, such a pair (u, v) determines and is determined by the pair (uv, u), consisting of an element of G and an involution inverting it. For each $y \in G$ let $I(y)$ be the set of involutions inverting y. Thus

$$\frac{g^2}{c^2} = \sum_{y \in G} |I(y)|.$$

But for $v, w \in I(y)$, the element vw^{-1} centralizes y, so $I(y)$ lies in a single coset in G of $C_G(y)$. Thus

$$\frac{g^2}{c^2} \leq \sum_{y \in G} |C_G(y)| = g + \sum_{y \in G^\#} |C_G(y)|.$$

For each y in the sum, $m \leq |G : C_G(y)|$ by definition of m, so $|C_G(y)| \leq g/m$. Therefore

$$\frac{g^2}{c^2} \leq g + (g-1)\frac{g}{m} \leq g\frac{g}{m} + (g-1)\frac{g}{m} \leq \frac{2g^2}{m}.$$

This yields the desired inequality $m \leq 2c^2$. □

Brauer developed a highly effective method for getting much stronger results when the specific structure of C is given. His method is based on not just counting the set $C \times C$, but summing nontrivial complex characters of G over that set. In this way precise order formulae for G are obtained.

In the case that G has two or more conjugacy classes of involutions, Thompson revealed an elementary formula for the order of G, which has frequently been found to be effective. We set up the situation now. Suppose that G is a group of even order.

For any involutions u, v, g of G let

$$\#(u^{\cdot}v^{\cdot} \to g)$$

be the number of ordered pairs (u_1, v_1) such that

$$g \in \langle u_1 v_1 \rangle,\ u_1 \text{ is } G\text{-conjugate to } u, \text{ and } v_1 \text{ is } G\text{-conjugate to } v.$$

Notice that by the lemma, the pairs being counted here consist of elements of $N_G(\langle g \rangle)$. In particular, the numbers $\#(u^{\cdot}v^{\cdot} = g)$ are determined by a) the structure of $N_G(g)$ and b) knowledge of which elements of $N_G(g)$ are G-conjugate to u and v.

Furthermore, the count is invariant under conjugation and so depends only on the conjugacy classes of u, v and g.

THEOREM 9.3 (THOMPSON ORDER FORMULA). *Let G be a group of even order in which the conjugacy classes of involutions are represented by elements $g_1, \ldots, g_r$. Assume that $r \geq 2$. Then*

$$|G| = |C_G(g_1)||C_G(g_2)| \sum_{i=1}^{r} \frac{\#(g_1^{\cdot}g_2^{\cdot} \to g_i)}{|C_G(g_i)|}.$$

PROOF. Let C_i be the conjugacy class of g_i. Thompson counts the set

$$C_1 \times C_2$$

in two ways. Obviously its cardinality is $|G|^2/|C_G(g_1)||C_G(g_2)|$. On the other hand, for any $(u, v) \in C_1 \times C_2$ we argue that the element uv has even order. Otherwise the group $\langle u, v \rangle$ would be dihedral of order $2n$, n odd, and so u and v would be conjugate in $\langle u, v \rangle$ (by Sylow!), contradicting the assumption that they are in different G-conjugacy classes. Thus each pair (u, v) determines an involution z,

namely the involution lying in $\langle uv\rangle$. By definition, moreover, (u, v) is one of the pairs counted in $\#(g_1^{\cdot}g_2^{\cdot} \to z)$. It follows that

$$|C_1 \times C_2| = \sum_z \#(g_1^{\cdot}g_2^{\cdot} \to z),$$

where in the sum, z ranges over the set of involutions of G. Sorting the sum by conjugacy classes we get

$$\frac{|G|^2}{|C_G(g_1)||C_G(g_2)|} = \sum_{i=1}^{r} \frac{|G|}{|C_G(g_i)|}\#(g_1^{\cdot}g_2^{\cdot} \to g_i),$$

the desired equation. □

We emphasize that the numbers $\#(g_1^{\cdot}g_2^{\cdot} \to g_i)$ can be calculated in $C_G(g_i)$; one needs to know the structure of this centralizer and the manner in which it intersects the G-conjugacy classes of g_1 and g_2.

We give a simple example demonstrating the power of this formula, and observe that for more complicated examples it is generally a matter of intestinal fortitude to complete the desired calculation. When Griess, for example, was making his initial investigations of the potential group which turned out to be the Monster, he may have used this formula to compute its order[1].

For our example let us assume that G is a finite group containing a subgroup

$$H \cong S_4$$

of odd index. Assume that for any involution $x \in H$,

$$C_G(x) \cong D_8 \text{ if } x \in O_2(H), \text{ and } C_G(x) \cong Z_2 \times S_3 \text{ otherwise.}$$

We shall show that

$$G \cong S_5,$$

the main step being to show that $|G| = 120$. The involutions in H fall into two H-conjugacy classes, represented by elements $x \in O_2(H)$ and $y \in H - O_2(H)$ which we now fix. Since H contains a Sylow 2-subgroup of G, every involution of G is conjugate to x or y; on the other hand x and y, having nonisomorphic centralizers, are not isomorphic. So G has two conjugacy classes of involutions.

To calculate $\#(x^{\cdot}y^{\cdot} \to x)$, we may calculate in $S = C_G(x)$, a Sylow 2-subgroup of H. Now $S \cong D_8$ and the G-conjugates of x in S are precisely the involutions in the subgroup $O_2(H) \cong Z_2 \times Z_2$ of S. Take an involution $v \in S - O_2(H)$, a typical G-conjugate of y in S (there are two such). Then a pair (u, v) gets counted in $\#(x^{\cdot}y^{\cdot} \to x)$ if and only if $u \in O_2(H)^{\#}$ and x is a power of uv. Thus the pair (x, v) does not get counted, but the other two pairs do. Since there were two choices for v, we have

$$\#(x^{\cdot}y^{\cdot} \to x) = 4.$$

To calculate $\#(x^{\cdot}y^{\cdot} \to y)$, we may calculate in $C_G(y) = \langle y\rangle \times A$, where $A \cong S_3$. Now $x \in C_G(y)$. Thus $\langle x, y\rangle$ is a Sylow subgroup of $C_G(y)$. We may replace A by $O_3(A)\langle x\rangle$, and assume that $x \in A$. We know from S that xy and y are conjugate, so the G-conjugates of x in A are then the three involutions $x = x_1, x_2, x_3$ of A. Now x_iy is an involution which is not y, so the pairs (x_i, y) do not contribute to

[1] $2^{46}3^{20}5^{9}7^{6}11^{2}13^{3}17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71$

$\#(x^{\cdot}y^{\cdot} \to y)$. But $(x_i, x_j y)$ does contribute, since $x_i x_j$ has odd order k and so $(x_i x_j y)^k = y$. Therefore

$$\#(x^{\cdot}y^{\cdot} \to y) = 9.$$

Now the Thompson Order Formula reads

$$\begin{aligned} |G| &= |C_G(x)||C_G(y)| \left(\frac{\#(x^{\cdot}y^{\cdot} \to x)}{|C_G(x)|} + \frac{\#(x^{\cdot}y^{\cdot} \to y)}{|C_G(y)|} \right) \\ &= |C_G(x)|\#(x^{\cdot}y^{\cdot} \to y) + |C_G(y)|\#(x^{\cdot}y^{\cdot} \to x) = 8 \cdot 9 + 12 \cdot 4 = 120. \end{aligned}$$

Thus $|G : H| = 5$. But H contains no nontrivial normal subgroup of G ($O_2(H)$ can't be normal in G, since x has $|G : C_G(x)| = 15$ conjugates), and so $G \cong S_5$.

In conclusion we mention that Bender has managed to refine methods for getting order formulae in the case of one conjugacy class of involutions. The idea is to take a maximal subgroup M containing the centralizer of an involution and get a second count of the number $|G : C_G(x)|$ of involutions of G by counting them one coset of M at a time. When such a coset contains two involutions x, y, the product xy is an element of M inverted by x, and so information about normalizers of subgroups of M can be brought to bear on the order of the group G.

10. Extraspecial p-Groups and Sporadic Subgroup Structure

We shall give another example of the Thompson Order Formula, this time to the local situation arising in the sporadic group J_2, the "Hall-Janko" group of order 604800. In the process we introduce some of the features common to the local structure of a number of the sporadic simple groups. Important among these are "extra-special" p-groups.

DEFINITION 10.1. Let p be a prime. A p-group P is special if and only if $[P, P] = \Phi(P) = Z(P)$ and $P \neq 1$. A special p-group P is extraspecial if and only if in addition, $Z(P) \cong Z_p$.

LEMMA 10.2. *If P is a special p-group, then $Z(P)$ is elementary abelian. Moreover the symbol $[x, y]$ is multiplicative in each argument.*

PROOF. By assumption every commutator lies in $Z(P)$, so

$$[xy, z] = [x, z]^y [y, z] = [x, z][y, z],$$

and similarly $[x, y]$ is multiplicative in the second argument. Consequently for any $x, y \in P$ we have $[x, y]^p = [x^p, y]$. But $x^p \in \Phi(P) = Z(P)$, so $[x^p, y] = 1$. Thus every commutator is trivial or has order p. Since $[P, P] = Z(P)$ is abelian, it is therefore elementary abelian. □

LEMMA 10.3. *A nonabelian p-group P is extra-special if and only if $Z(P) \cong Z_p$ and $P/Z(P)$ is elementary abelian.*

PROOF. One direction is clear. Conversely suppose that $Z(P) \cong Z_p$ and $P/Z(P)$ is elementary abelian. Since $\Phi(P)$ is the smallest normal subgroup with elementary abelian quotient, and $[P, P]$ the smallest with abelian quotient, we have

$$1 \neq [P, P] \leq \Phi(P) \leq Z(P).$$

Since $Z(P)$ has order p, all these groups are equal. □

LEMMA 10.4. *Any nonabelian group of order p^3 is extraspecial.*

PROOF. In a nonabelian group P, the quotient $P/Z(P)$ must be noncyclic, since otherwise an element y whose image generated $P/Z(P)$ would satisfy $P = \langle y, Z(P)\rangle$ and hence P would be abelian. Now if $|P| = p^3$ to boot, then $Z(P)$ being nontrivial (P is nilpotent), we must have $Z(P) \cong Z_p$ and $P/Z(P)$ a noncyclic group of order p^2. But groups of order p^2 are abelian so $P/Z(P) \cong Z_p \times Z_p$ and the previous lemma finishes the proof. □

We next give a way to combine extraspecial groups to form a larger one.

DEFINITION 10.5. Let P and Q be nonabelian groups such that $Z(P) \cong Z(Q) \cong Z_p$. Choose an isomorphism

$$\phi : Z(P) \to Z(Q).$$

The central product $P \circ Q$, or more properly $P \circ_\phi Q$, is defined as the group

$$P \times Q/R,$$

where $R = \{(z^{-1}, \phi(z)) \mid z \in Z(P)\}$.

The definition seems to depend on ϕ, but we temporarily ignore this since it will turn out that for the situations we care about, this dependence will disappear.

If we write $P \circ Q = \overline{P \times Q}$, then we see that

(a) $\overline{P} \cong P$ and $\overline{Q} \cong Q$, since $P \cap R = Q \cap R = 1$;
(b) $P \circ Q = \overline{PQ}$ and $[\overline{P}, \overline{Q}] = 1$;
(c) $\overline{P} \cap \overline{Q} = Z(\overline{P}) = Z(\overline{Q}) = Z(P \circ Q)$; and
(d) $P \circ Q/Z(\overline{P}) \cong (P \times Q)/(Z(P) \times Z(Q)) \cong P/Z(P) \times Q/Z(Q)$.

Notice that if $x \in P$ and $y \in Q$, and $\overline{xy} \in Z(P \circ Q)$, then $\overline{xy}$ commutes elementwise with $\overline{P}$; but $\overline{y}$ commutes with $\overline{P}$ by (b), so $\overline{x} \in Z(\overline{P})$. Similarly $\overline{y} \in Z(\overline{Q})$, proving that $Z(\overline{P}) = Z(\overline{Q}) = Z(P \circ Q)$. The remaining statements are fairly easy to check.

Consequently we have:

LEMMA 10.6. *If P and Q are extraspecial of orders p^m and p^n, respectively, then $P \circ Q$ is extraspecial of order p^{m+n-1}.*

Indeed, (c) and (d) show that the criterion of Lemma 10.3 is satisfied.

The next result will imply that every extraspecial group arises as the central product of groups of order p^3. Notice that if P is a group with subgroups Q and R such that

$$P = QR,\ [Q, R] = 1 \text{ and } Q \cap R = Z(Q) = Z(P) \cong Z_p, \tag{10.1}$$

then $Z(R) \leq Z(P)$ so $Z(R) = Z(Q)$ and then $P \cong Q \circ R$ (where ϕ is taken to be the identity mapping $Z(Q) \to Z(R)$.)

LEMMA 10.7. *Let P be an extraspecial group. Then there exist nonabelian subgroups Q of P of order p^3. For any such Q, $P \cong Q \circ R$ where $R = C_P(Q)$. Moreover, if $|P| > p^3$ then R is extraspecial.*

PROOF. Choose any $x \in P - Z(P)$. There exists $y \in P$ such that $[x, y] \neq 1$, and we choose any such y and set $Q = \langle x, y \rangle$. Since $P/Z(P)$ is elementary abelian but Q is not abelian, $Q \cap Z(P) \neq 1$ and so $Z(P) \leq Q$; then $Q/Z(P)$ is generated by the images of x and y, so has order p^2, whence $|Q| = p^3$. This proves the first statement. Conversely given Q it is clear that Q arises from such elements x and y.

Notice that

$$|P : C_P(x)| = p.$$

Indeed, this index is the number of conjugates x^g, $g \in P$, or equivalently the number of elements $x^{-1}x^g = [x, g]$, so it's at most $|[P, P]| = p$. But $C_P(x) < P$ so the index is exactly p. Similarly $|P : C_P(y)| = p$. Moreover $C_P(x) \neq C_P(y)$, since $C_P(x)$ contains x but not y. Therefore

$$P = C_P(x)C_P(y), \text{ whence } C_P(Q) = C_P(x) \cap C_P(y) \text{ has index } p^2 \text{ in } P.$$

But also $Q \cap C_P(Q) = Z(Q)$ has order p, so has index p^2 in Q. It follows that $P = QC_P(Q) = QR$. Obviously $[Q, R] = 1$. The conditions of (10.1) are therefore satisfied, so $P \cong Q \circ R$. Then $Z(R) \leq Z(P) \cong Z_p$, so $Z(R) = Z(P)$, and $R/Z(R)$ embeds in $P/Z(P)$ so is elementary abelian, whence R is extraspecial. □

It follows that extraspecial groups are precisely those groups which are central products of nonabelian groups of order p^3. It's about time that we determine what those basic building blocks are.

For $p = 2$, the dihedral group D_8 and quaternion group Q_8 are examples, and are nonisomorphic, having different numbers of elements of various orders. Any nonabelian group P of order 8 is isomorphic to one of these two. Indeed P must contain an element x of order 4, since otherwise $x^2 = 1$ for all $x \in P$ would imply that P would be abelian; then $P = \langle x, y \rangle$ for any $y \in P - \langle x \rangle$; also $\langle x \rangle \triangleleft P$ since $\langle x \rangle$ is a maximal subgroup and P is nilpotent; since P is nonabelian, $x^y \neq x$ and so $x^y = x^{-1}$; finally $y^2 \in \langle x \rangle$ and y^2 commutes with y, so $y^2 = 1$ or x^2. These two possibilities lead directly to $P \cong D_8$ or Q_8, respectively.

For $p > 2$, it turns out again that there are exactly two isomorphism types of nonabelian groups of order p^3. One of them is most easily described as the subgroup

$$A_{p^3} = \left\{ \begin{bmatrix} 1 & * & * \\ 0 & 1 & * \\ 0 & 0 & 1 \end{bmatrix} \right\} \leq GL_3(p).$$

Notice that for any $g \in A_{p^3}$, $(g - 1)^3 = 0$ (in the algebra of matrices), and so $(g - 1)^p = 0$ as $p > 2$. Therefore $g^p = 1$, so

$$A_{p^3} \text{ has exponent } p.$$

Moreover, taking $x = 1 + e_{12}$, $y = 1 + e_{23}$ and setting $z = [x, y]$, where the e_{ij} are the unit matrices, we see that x and y generate A_{p^3} and

$$x^p = y^p = z^p = 1, \ [x, z] = [y, z] = 1, \ [x, y] = z.$$

These relations clearly define a group of order at most p^3, so form defining relations for A_{p^3}.

In fact if $p > 2$ and A is any nonabelian group of order p^3 and exponent p, then $A \cong A_{p^3}$. To see this, take a maximal subgroup B of A; thus $B = \langle y \rangle \times \langle z \rangle$, and

contains a subgroup of $Z(P)$ of order p, which we may take to be $\langle z\rangle$. The group $P/\langle z\rangle$, having order p^2, is abelian, so if we let $x \in A - B$; then $[x, y] \in \langle z\rangle$. Since P is generated by x, y and z, and is nonabelian, $[x, y] \neq 1$; replacing z by a power of itself we may assume $[x, y] = z$, and so $A \cong A_{p^3}$.

The other isomorphism type H_{p^3} of nonabelian group of order p^3, $p > 2$, is the semidirect product of $\langle x\rangle \cong Z_{p^2}$ by $\langle y\rangle \cong Z_p$, with $x^y = x^{1+p}$. Notice that $\mathrm{Aut}(Z_{p^2})$ is isomorphic to the group of units of the ring Z/p^2Z, so is abelian of order $p(p-1)$ and in particular has a unique subgroup of order p. So any two semidirect products of Z_{p^2} by Z_p, if nonabelian, are isomorphic. Moreover, if H is any nonabelian group of order p^3 containing $\langle x\rangle \cong Z_{p^2}$, then the image of H in $\mathrm{Aut}(\langle x\rangle)$ is the unique subgroup of $\mathrm{Aut}(\langle x\rangle)$ of order p and so some $y \in H - \langle x\rangle$ satisfies $x^y = x^{1+p}$. To show that $H \cong H_{p^3}$, which we shall do, it is enough to show that y can be chosen so that $y^p = 1$. In doing this we are free to replace y by $x^i y$ for any i. Notice also that in any case, $y^p \in \langle x\rangle$ and y^p commutes with y, so $y^p \in C_{\langle x\rangle}(y) = \langle x^p\rangle$. Thus $y^p = x^{jp}$ for some j. Now $xy = yx^y = yxx^p$, and so

$$(x^i y)^p = y^p x^i x^{p(i+2i+\cdots+pi)} = x^{jp+i} x^{p\binom{p+1}{2}} = x^{jp+i},$$

the sum representing the number of y's to be pulled to the left across an x. Here we also use the fact that $x^{p^2} = 1$ and that p divides $\binom{p+1}{2}$, owing to the oddness of p. The upshot is that we only have to take $i = -jp$ to get a complement to $\langle x\rangle$ in H, so $H \cong H_{p^3}$.

For brevity, let's write D and Q for D_8 and Q_8, and A and H for A_{p^3} and H_{p^3}.

PROPOSITION 10.8.

(a) *A p-group P is extraspecial if and only if it is the central product of nonabelian groups of order p^3. The number of such groups is called the width of P.*

(b) *Given p and given a width n, there are exactly two extraspecial p-groups of width n, up to isomorphism. If $p = 2$ these are $D * D * \cdots * D * D$ and $D * D * \cdots * D * Q$. If $p > 2$ these are $A * A * \cdots * A * A$ and $A * A * \cdots * A * H$.*

(c) *For $p = 2$, we have $D * D \cong Q * Q \ncong D * Q$. For $p > 2$ we have $H * H \cong A * H \ncong A * A$.*

PROOF. The first assertion follows from the previous lemma and analysis. In (c), notice that $Q * Q$ has a subgroup isomorphic to D; if the first factor is $\langle x, y\rangle$ and the second is $\langle u, v\rangle$, then $(yu)^2 = y^2u^2 = 1$ and $[x, yu] = [x, y] \neq 1$, so $\langle x, yu\rangle \cong D$. Therefore $Q * Q \cong D * D$ or $D * Q$, by 10.7. But $Q * Q \ncong D * Q$, as can be seen by counting elements of $P - Z(P)$ of order 2; $Q * Q$ has $3 \cdot 3 \cdot 2$ of them, namely the elements not in either central factor, while $D * Q$ has 4 in the D factor and $2 \cdot 3$ outside the D factor, for a total of 10. This proves (b) and (c) for $p = 2$. The proof for $p > 2$ is similar; notice that $H * H$ contains a subgroup isomorphic to A, and that $A * A$ is of exponent p so is not isomorphic to $A * H$ or $H * H$. □

EXERCISES. Verify that for a nonabelian group P of order p^3, $\mathrm{Aut}(P)$ permutes transitively the nonidentity elements of $Z(P)$.

Compute the numbers of elements of order p in the two extraspecial p-groups of width n.

Extraspecial p-groups arise naturally in two places: in linear groups over the field $\boldsymbol{F}_p$, and in many of the sporadic groups.

For instance, in $G = PSL_n(p)$, if we let z be a transvection, say the image of $I + e_{1n}$, and $C = C_G(z)$, then $Q = F^*(C)$ is the image of the group

$$\begin{bmatrix} 1 & * & * & \cdots & * & * \\ 0 & 1 & 0 & \cdots & 0 & * \\ 0 & 0 & 1 & \cdots & 0 & * \\ 0 & 0 & 0 & \cdots & 1 & * \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

which is extraspecial of order p^{2n-1}, i.e., width $n-1$.

EXERCISE. If $p > 2$, then $Q \cong A * A * \cdots * A$. If $p = 2$, then $Q \cong D * D * \cdots * D$.

DEFINITION 10.9. A simple group G is of $\boldsymbol{F}_p$-type if and only if there is an element $z \in G$ of order p such that $F^*(C_G(z))$ is extraspecial.

Notice that in this situation, if we set $C = C_G(z)$ and let $P \in \mathrm{Syl}_p(C)$, then $Q = F^*(C) \le P$, so $Z(P) \le C_C(Q) \le Q$ by the F^*-theorem, and then $Z(P) \le Z(Q) = \langle z \rangle$. Indeed if we let $P \le R \in \mathrm{Syl}_p(G)$, then the same reasoning shows that $Z(R) = \langle z \rangle$. Thus the z in the definition in fact is a generator of a Sylow p-center.

As an example, $PSL_n(p)$ is of $\boldsymbol{F}_p$-type. A number of sporadic groups are as well. While $PSL_n(p)$ is normally not of $\boldsymbol{F}_q$-type for any other prime $q \neq p$, some sporadic groups are remarkably "versatile". The most startling example is the Monster, which is of $\boldsymbol{F}_p$-type for $p = 2$, 3, 5, 7 and 13 (i.e. if and only if p is a prime such that $p - 1$ divides 24). No satisfactory explanation for this is known.

In a group of $\boldsymbol{F}_p$-type, $C = C_G(z)$ and $Q = F^*(C)$ satisfy the condition that $Z(Q) = C_C(Q)$, and hence $C/Z(Q)$ embeds in $\mathrm{Aut}(Q)$. Therefore

$$C/Q \text{ embeds in } \mathrm{Out}(Q) = \mathrm{Aut}(Q)/\mathrm{Inn}(Q).$$

Furthermore, $Z(Q) = \langle z \rangle \le Z(C)$, so $F^*(C/\langle z \rangle) = F^*(C)/\langle z \rangle = Q/\langle z \rangle$. (This is because the preimage of $F^*(C/\langle z \rangle)$ in C is chiefly inner; we are only tacking on a central chief factor at the bottom.) Thus C/Q also embeds in $\mathrm{Aut}(Q/\langle z \rangle \cong GL_{2n}(p)$, where Q has width n. But the fact that C acts on Q restricts how it can act on $Q/\langle z \rangle$.

Indeed let P be any extraspecial p-group of width n. Set $\overline{P} = P/Z(P)$. We think of $\overline{P}$ as a vector space over $\boldsymbol{F}_p$, and of $Z(P)$ as a copy of (the additive group of) $\boldsymbol{F}_p$. Then the mapping $(x, y) \mapsto [x, y]$ from $P \times P \to Z(P)$ induces a mapping

$$\beta : \overline{P} \times \overline{P} \to Z(P), \quad (\overline{x}, \overline{y}) \mapsto [x, y]$$

which is well-defined because different preimages of $\overline{x}$ differ only by elements of $Z(P)$, which do not affect commutators; and which is bilinear because of the identity $[uv, w] = [u, w]^v[v, w] = [u, w][v, w]$ in P, the latter because $[P, P] \le Z(P)$.

Moreover, the mapping $x \mapsto x^p$ induces a well-defined mapping

$$q : \overline{P} \to Z(P), \quad \overline{x} \mapsto x^p.$$

This is well-defined because for any $y \in Z(P)$, $(xy)^p = x^p y^p = x^p$.

LEMMA 10.10.

(a) *β is a nondegenerate alternating form, that is, $\beta(\overline{x}, \overline{x}) = 0$ for all $\overline{x} \in \overline{P}$, and no element other than the identity is orthogonal to all of $\overline{P}$;*

(b) *If $p = 2$, then q is a quadratic form on $\overline{P}$ with associated bilinear form β, that is,*

$$q(\overline{x}^c) = c^2 q(\overline{x}) \text{ for all } c \in \boldsymbol{F}_2, \text{ and } q(\overline{xy}) = q(\overline{x})q(\overline{y})\beta(\overline{x},\overline{y})$$

for all $\overline{x}$ and $\overline{y}$;

PROOF. If $\overline{x}$ were orthogonal to all of $\overline{P}$, then $x \in Z(P)$, and so $\overline{x} = 1$. The only other nontrivial statement is the last equation, which follows from

$$(xy)^2 = xyxy = xxy[y,x]y = xxyy[y,x] = x^2y^2[y,x] = x^2y^2[x,y],$$

the last because when $p = 2$, $[P,P]$ has exponent 2, so $[y,x] = [x,y]^{-1} = [x,y]$. □

In the above situation, we let $Sp(\overline{P}) = Sp(\overline{P},\beta)$ be the full group of all linear transformations of $\overline{P}$ preserving the form β. Likewise we let $O(\overline{P}) = O(\overline{P},q)$ be the full group of all linear transformations of $\overline{P}$ preserving q.

PROPOSITION 10.11.

(a) *Suppose that p is odd and P is an extraspecial p-group of exponent p. Let $Aut_0(P) = C_{\mathrm{Aut}(P)}(Z(P))$. Then by mapping any $\alpha \in Aut_0(P)$ to the induced mapping $\overline{\alpha} \in Aut(\overline{P})$, we obtain an exact sequence*

$$1 \to Inn(P) \to Aut_0(P) \to Sp(\overline{P}) \to 1.$$

(b) *Suppose that P is an extraspecial 2-group. Then there is a similar exact sequence*

$$1 \to Inn(P) \to Aut(P) \to O(\overline{P}) \to 1.$$

PROOF. In the first case since α is an automorphism, $\overline{\alpha}$ must preserve β. Likewise $\overline{\alpha}$ preserves q in the second case, so the mapping γ

$$\mathrm{Aut}_0(P) \to Sp(\overline{P}) \quad \text{or} \quad \mathrm{Aut}(P) \to O(\overline{P}),$$

respectively, is well-defined. The kernel of γ is the group of automorphisms of P which act trivially on both $P/Z(P)$ and $Z(P)$. If τ is such an automorphism, then for any $x \in P$ we have

$$\tau(x) = x\zeta(x),\ \zeta(x) \in Z(P).$$

Since τ is a homomorphism and $\zeta(x) \in Z(P)$, it follows that ζ is a homomorphism as well. Since $\tau|Z(P)$ is trivial, ζ lifts to a homomorphism $\overline{P} \to Z(P)$ which we also call ζ. But since β is nondegenerate, there exists $\overline{y} \in \overline{P}$ such that $\zeta(\overline{x}) = \beta(\overline{x},\overline{y})$ for all $x \in P$. A preimage y of $\overline{y}$ then satisfies $\tau(x) = x[x,y]$ for all $x \in P$. Therefore $\tau(x) = x^y$ and τ is inner. This proves that $\ker\gamma = \mathrm{Inn}(P)$.

It remains to show that γ is onto. Choose any basis $\overline{x}_1,\dots,\overline{x}_m$ of $\overline{P}$ and any preimages $x_1,\dots,x_m$ of these basis vectors. Then every element $g \in P$ has a unique expression as

$$g = x_1^{a_1}x_2^{a_2}\cdots x_n^{a_m}z^a,\ 0 \le a_i < p,\ 0 \le a < p.$$

Given $\tau \in Sp(\overline{P})$, choose any $y_i \in P$ such that $\overline{\delta(x_i)} = \overline{y}_i$, and define

$$\sigma(g) = y_1^{a_1}y_2^{a_2}\cdots y_n^{a_m}z^a.$$

We leave it as an exercise to verify that $\sigma \in \mathrm{Aut}(P)$, unless possibly $p = 2$; in that case if $\tau \in O(\overline{P},q)$, then $\sigma \in \mathrm{Aut}(P)$. □

The case we are interested in is a small one: $Q = F^*(C_G(z)) \cong Q_8 * D_8$, and $C_G(z)/Q \cong A_5$. By contrast, for the Monster, $Q = D_8 * D * \cdots * D$ has width 12, and $C_G(z)/Q \cong Co_1$, another sporadic group, the projective reduction of the automorphism group of a 24-dimensional lattice (the "Leech lattice") in Euclidean space.

COROLLARY 10.12. *Out*$(D_8 * Q_8) \cong \Sigma_5$.

PROOF. Let $P = D_8 * Q_8$ and $\overline{P} = P/Z(P)$ as above. Then P has 10 involutions outside $Z(P)$, so the quadratic form q takes the value 0 on exactly 5 nonidentity elements of $\overline{P}$. It is easy to check that P is generated by its involutions outside $Z(P)$ (the product of an element of order 4 in D_8 with one in Q_8 is an involution) and so $\mathrm{Out}(P) \cong O(\overline{P}, q)$ embeds in Σ_5. Indeed, $O(\overline{P}, q)$ is precisely the set of elements of $\mathrm{Aut}(\overline{P}) \cong GL_4(2)$ preserving the set of 5 "singular" vectors. These 5 elements satisfy a single relation in the 4-dimensional space $\overline{P}$, and it is again easy to check that this relation is that their product is 1 (i.e., additively, their sum is 0 in the vector space $\overline{P}$). This relation being completely symmetric, any permutation of the 5 singular vectors extends to a linear transformation on $\overline{P}$, proving the result.□

We are now ready to give the involution pattern of the Hall-Janko group J_2, from which we shall determine its order.

DEFINITION 10.13. Let G be a simple group. We say that G has the involution pattern of J_2 if and only if the following conditions hold:

(a) G has $\boldsymbol{F}_2$-type;
(b) G has two conjugacy classes of involutions, represented by z and t;
(c) $C_G(z)$ is a semidirect product of $Q = O_2(C_G(z))$ by $A \cong A_5$;
(d) All involutions in Q are G-conjugate to z, and all other involutions in $C_G(z)$ are G-conjugate to t;
(e) $C_G(t) = V \times B$ where $V \cong Z_2 \times Z_2$, $t \in V$, and $B \cong A_5$;
(f) All involutions of B are G-conjugate to z, and all other involutions of $C_G(t)$ are G-conjugate to t.

THEOREM 10.14. *If G is a simple group whose involution pattern is that of J_2, then $|G| = 604800$.*

PROOF. We have $|C_G(z)| = 2^5 \cdot 60$ and $|C_G(t)| = 2^2 \cdot 60$. To apply the Thompson Order Formula we must compute

$$a = \#(z^{\cdot}t^{\cdot} \to z) \text{ and } b = \#(z^{\cdot}t^{\cdot} \to t).$$

There is no loss in replacing z and t by conjugates. We may assume then that a Sylow 2-subgroup T of $C_G(t)$ lies in a Sylow 2-subgroup S of $C_G(z)$ (which is one of G).

We first calculate b, which is easier. In order for an element $u \in C_G(t) = V \times B$ to have $t = u^n$ as a power, we must have $u = u_V u_B$ and so $t = u^n = u_V^n u_B^n$. Thus $u_V^n = t$ and $u_B^n = 1$. This means, since V has exponent 2, that $u_V = t$, n is odd, and u_B then has odd order. Thus the elements u which are roots of t are precisely products of t with elements of B of odd order. Now b counts the number of pairs (y, v) such that y is an involution of B (by condition 10.13f), v is an involution of $C_G(t) - B$, and $yv = u$ is a root of t. Writing $v = wx$ with $w \in V$, $x \in B$, we see that $x^2 = 1$, w is an involution, $yv = w(xy)$, and so $w = t$ and xy has odd

order. The last condition implies that $x \neq 1$, so x is an involution. Thus b is the number of pairs $(x, y) \in B \times B$ such that $x^2 = 1$, y is an involution, and xy has odd order. There are 15 involutions $y \in B$, and for each of them there are only 2 other involutions x such that xy does not have odd order. So

$$b = 15 \cdot 13.$$

Now consider a. For a pair (x, y) with x conjugate to z and y to t to have the product xy a root of z, we first must have $x, y \in C_G(z)$ and so $x \in Q$, $y \notin Q$, by condition (d). If $[x, y] = 1$, then $(xy)^2 = 1$ and so the only possibility is that $xy = z$, which is impossible since $x, z \in Q$ but $y \notin Q$. Thus the pairs we are counting must satisfy $[x, y] \neq 1$ as a necessary condition.

We argue that it is also sufficient: that is, if $x \in Q$ and $y \in C_G(z) - Q$ are involutions such that $[x, y] \neq 1$, then z is a power of xy. For in the quotient $C_G(z)/Q$, the image of xy is the same as that of y, so is an involution. Thus xy has even order, so $(xy)^2 = u$ is a nonidentity element of Q. Since Q is extraspecial every nonidentity element of Q is a root of z, except for the 10 noncentral involutions. So we must rule out the possibility that $(xy)^2 = u$ is a noncentral involution of Q. But in that case, u would be the central involution of the dihedral group $\langle x, y\rangle$, and so $x \in C_Q(u)$. But the only involutions in $C_Q(u)$ are those in $\langle u, z\rangle$ (to see this, observe that $\mathrm{Aut}(Q)$ is transitive on the 10 noncentral involutions, so it's enough to prove it for one particular noncentral involution; since $Q = D_8 * Q_8$, clearly there's an involution in the D_8 whose centralizer is $Z_2 \times Q_8$, which has just three involutions). Therefore $x \in \langle u, z\rangle$. As $u \in Z(\langle x, y\rangle)$, $[y, \langle u, z\rangle] = 1$ and so $[x, y] = 1$, contradiction. We have proved that

$$a = \#\text{ of pairs of involutions } (x, y) \in Q \times C_G(z) - Q \text{ such that } [x, y] \neq 1.$$

Let c be the number of involutions of $C_G(z) - Q$, that is, the number of G-conjugates of t commuting with z. We can count c by considering the set

$$S = \{(u, v) \mid u \in z^G,\ v \in t^G,\ [u, v] = 1\}.$$

Clearly G acts on S. For a given second coordinate, say t, the pairs in S are the pairs (u, t) with $u \in B$, by condition (f). Since $B \cong A_5$ and all involutions of A_5 are conjugate, $C_G(t)$ acts transitively on the pairs in S for which $v = t$. Consequently G acts transitively on S, and so $C_G(z)$ acts transitively on the set of pairs in S for which $u = z$. That is, $C_G(z)$ is transitive on the set of all $v \in C_G(z)$ which are conjugate to t. Thus c counts the size of the $C_G(z)$-conjugacy class of t. Furthermore, the stabilizer of (z, t) can be calculated in $C_G(t)$ as $C_G(\langle z, t\rangle) = C_{C_G(t)}(z) = V \times C_B(z) \cong Z_2 \times Z_2 \times Z_2 \times Z_2$. This equals $C_{C_G(z)}(t)$, and so

$$c = |C_G(z) : C_G(\langle z, t\rangle)| = 2^3 \cdot 15.$$

The fact that all involutions of $C_G(z) - Q$ are $C_G(z)$-conjugate also implies that $a = cd$, where d is the number of involutions of Q not commuting with t. But the involutions in $C_Q(t)$ are involutions in $C_G(t)$ conjugate to z, and so there's a four-group worth of them, i.e., three of them. Since Q has 11 involutions, $d = 11 - 3 = 8$ and so

$$a = cd = 2^6 \cdot 15.$$

Now the Thompson Order Formula implies that

$$|G| = a|C_G(t)| + b|C_G(z)| = 2^6 \cdot 15 \cdot 2^4 \cdot 15 + 13 \cdot 15 \cdot 2^7 \cdot 15 = 2^7 3^3 5^2 7 = 604800.$$

□

11. Automorphisms of Semisimple Groups

In this section we make some remarks about automorphisms of semisimple groups. The basic Proposition 6.2 has already been proved and merits a second mention. We make another general observation and then compute the automorphism group of a couple of simple groups.

PROPOSITION 11.1. *Let $H \leq G$. Then*

$$E(G) = C[E(G), H],$$

where C is the product of all components of G centralized by H. Moreover the commutator $[E(G), H]$ is the product of all components of $E(G)$ not centralized by H.

PROOF. $E(G)$ is the product $G_1 \cdots G_n$ of all components. We may assume notation chosen so that H does not centralize any of $G_1, \ldots, G_m$, but does centralize $G_{m+1}, \ldots, G_n$. Then H normalizes $G_0 = G_1 \cdots G_m$, and centralizes $E(G)/G_0$, so $[E(G), H] \leq G_0$. Let $D = [E(G), H]$.

Let $Z = Z(E(G))$ and $\overline{E(G)} = E(G)/Z$. Then $\overline{E(G)} = \overline{G}_1 \times \cdots \times \overline{G}_n$, the direct product of nonabelian simple groups, and $\overline{D} = [\overline{E(G)}, H] \lhd \overline{E(G)}$, so $\overline{D}$ is the product of some $\overline{G}_i$. If $\overline{G}_1 \not\leq \overline{D}$, then H centralizes $G_1/Z(G_1)$ and hence G_1, contradiction. Therefore $\overline{G}_1 \leq \overline{D}$. So $G_1 \leq DZ$. But then $G_1 = [G_1, G_1] \leq [DZ, DZ] = [D, D] \leq D$. The same argument may be applied to $G_2, \ldots, G_m$. □

Next we look at the specific example of the alternating groups A_n, $n \geq 5$. We have

$$F^*(S_n) = A_n \text{ and so } S_n \text{ embeds in } \mathrm{Aut}(A_n).$$

THEOREM 11.2. *Let $n \geq 5$. Then $Aut(A_n) \cong S_n$ unless $n = 6$. Moreover, $Aut(A_6)$ contains S_6 as a subgroup of index 2.*

Consequently $\mathrm{Out}(A_n)$ is small: it is generally Z_2, and has order 4 in the exceptional case.

Let $\alpha \in \mathrm{Aut}(A_n)$. The strategy is first to argue, by counting sizes of conjugacy classes, that α must carry the set of 3-cycles to itself (unless $n = 6$). Then we characterize the underlying set $\Omega = \{1, \ldots, n\}$ in terms of group-theoretic properties of the set of 3-cycles, so α arises from an "automorphism" of Ω, i.e. an element of S_n.

LEMMA 11.3. *Let $x \in A_n$ and suppose that x has cycle shape $1^{r_1} 2^{r_2} \cdots n^{r_n}$, that is, x is the product of r_1 1-cycles, r_2 2-cycles, etc., all disjoint. (Thus $\sum_i i r_i = n$ and the r_i are nonnegative integers.) Then*

$$|C_{A_n}(x)| = \frac{1}{d} \prod_{i=1}^{n} i^{r_i} (r_i)!,$$

where $d = 1$ or 2. More precisely $d = 2$ unless x is the product of disjoint cycles of distinct odd lengths.

PROOF. Let $C = C_{S_n}(x)$. Write $x = x_1 x_2 \cdots x_n$ where x_i is the product of all the i-cycles in the cycle decomposition of x. Thus if we let Ω_i be the support of x_i (and Ω_1 the fixed point set of x), we see that $\Omega = \{1, \ldots, n\}$ is partitioned by $\Omega_1, \Omega_2, \cdots$. Clearly C stabilizes each Ω_i, and so $C = C_1 \times C_2 \times \cdots$, where $C_i = C_{S_{\Omega_i}}(x_i)$. Now C_i permutes the r_i orbits of x_i on Ω_i (each of length i), and any permutation of these orbits is achievable by a suitable element of C_i. The subgroup of C_i leaving each of these orbits invariant is just generated by the cycles in the decomposition of x_i and so has order i^{r_i}. Thus $|C|$ is the displayed product above, and the displayed equation holds with $d = |C : C_{A_n}(x)|$.

Obviously $d = 1$ or 2, the latter if and only if x centralizes some odd permutation. The product of two i-cycles is the square of some $2i$-cycle so is centralized by an odd permutation; an i-cycle for i even is itself odd. On the other hand if x is the product of cycles of odd and distinct lengths, then C is just the product of the $\langle x_i \rangle$ and so $d = 1$. This completes the proof. □

It is then a straightforward exercise to prove:

LEMMA 11.4. *Let $n \geq 5$, let $x \in A_n$ be a 3-cycle and let $y \in A_n$ have order 3. If $|C_{A_n}(x)| = |C_{A_n}(y)|$, then either y is also a 3-cycle, or else $n = 6$ and y is the product of two disjoint 3-cycles.*

The hypothesis that y has order 3 simplifies matters greatly. Now automorphisms preserve orders of elements and orders of centralizers, so we immediately deduce:

LEMMA 11.5. *If $n \geq 5$ and $\alpha \in Aut(A_n)$, and $x \in A_n$ is a 3-cycle, then either x^α is a 3-cycle or else $n = 6$.*

We digress to show that reality is responsible for the exception in this lemma.

LEMMA 11.6. *A_6 and S_6 both have automorphisms carrying 3-cycles to products of two disjoint 3-cycles.*

PROOF. By inspection (or Sylow's Theorem), S_5 has six Sylow 5-subgroups. The conjugation action on them is transitive by Sylow's theorem and induces a homomorphism

$$S_5 \to S_6$$

whose image H is a transitive subgroup of S_6. Indeed, a Sylow normalizer in S_5 has order 20, so has no element of order 3. So the elements of H of order 3 have no fixed points and hence have cycle shape 3^2.

The above homomorphism is also clearly injective, since S_5 has no nontrivial normal subgroup of index more than 2. Therefore

$$|S_6 : H| = 6.$$

Let Ψ be the set of right cosets of H in S_6. The action of S_6 on Ψ by translation therefore induces a homomorphism

$$\phi : S_6 \to S_\Psi$$

whose kernel is the intersection of all conjugates of H. Again this forces ϕ to be injective. Identifying S_Ψ with S_6 we can interpret ϕ as an automorphism of S_6. But since H is the stabilizer of a point in the action underlying ϕ, $\phi(H)$ is the stabilizer of a point of Ψ. Taking $x \in H$ of order 3, we see that x has cycle shape 3^2, whereas

$\phi(x)$ has a fixed point and so has cycle shape 3. We have thus constructed the required automorphism of S_6, and also of A_6 by restriction. □

In view of this existence result, Theorem 11.2 will follow from the following easier result:

THEOREM 11.7. *Let $\alpha \in Aut(A_n)$, $n \geq 5$, and suppose that α carries 3-cycles to 3-cycles. Then α is conjugation by some $g \in S_n$.*

PROOF. Let x and y be 3-cycles in A_n. Let their supports be Ω_x and Ω_y, and set $n_{x,y} = |\Omega_x \cap \Omega_y|$, the size of the overlap. One may check that $n_{x,y}$ is determined by the product xy, as follows: $n_{x,y} = 0$ if and only if xy has order 3 and is not a 3-cycle; $n_{x,y} = 1$ if and only if xy has order 2 or is a 3-cycle different from both x and y; $n_{x,y} = 2$ if and only if xy has order 5, and $n_{x,y} = 3$ if and only if $xy = x$, y or 1.

These conditions are preserved by α, and so we conclude that

$$n_{x,y} = n_{x^\alpha, y^\alpha} \tag{11.1}$$

for all 3-cycles x, y.

For each $i = 1, \ldots, n$ let G_i be the stabilizer in A_n of the point i, so that $G_i \cong A_{n-1}$. We use (11.1) to argue that

$$G_i^\alpha = G_{i'} \text{ for some (unique) } i'. \tag{11.2}$$

For simplicity of notation, consider the case $i = n$. Then setting $x_1 = (123)$, $x_2 = (234)$, $x_3 = (345)$, etc., we have

$$G_n = \langle x_1, x_2, \ldots, x_{n-3} \rangle.$$

Moreover, $n_{x_{i-1}, x_i} = 2$ for each i. Consequently

$$G_n^\alpha = \langle y_1, y_2, \ldots, y_{n-3} \rangle$$

where we have put $y_i = x_i^\alpha$, and we know that $n_{y_{i-1}, y_i} = 2$ for each i. It follows that as we add generators y_i, each y_i (except the initial y_1) adds at most one new point to the support of G_n^α. Since y_1 is a 3-cycle, the support of G_n^α has cardinality at most $3 + 1 + \cdots + 1 = n - 1$. Thus $G_n^\alpha \leq G_n'$ for some n', and equality holds as the orders coincide. This proves (11.2).

Now define $g \in S_n$ by $G_i^\alpha = G_{i^g}$. (Since the G_i are distinct, this clearly is a permutation.) Let β be the composite of α with conjugation by g^{-1}. We complete the proof by showing that $\beta = 1$.

We know that $G_i^\beta = G_i$ for all i. By induction on n, $\beta | G_i$ is then trivial for each i, so β fixes all 3-cycles; since these generate A_n, $\beta = 1$. (Of course one must start the induction, by direct consideration of the case $n = 5$. Indeed, again inductively, it suffices to show that an automorphism β of A_4 carrying the stabilizer of each point to itself is trivial. Thus $(123)^\beta = (123)$ or (132). Likewise $(234)^\beta = (234)$ or (243). Now $(123)(234)$ has order 2 while $(123)(243)$ has order 3. Hence if $(123)^\beta = (123)$ then β fixes (234) also, and hence acts trivially on $\langle (123)(234) \rangle = A_4$. Thus we may assume that $(123)^\beta = (132)$, and similarly that β inverts every element of A_4 of order 3. The equation $(123)(243) = (124)$ is then not invariant under β, contradiction.) □

EXERCISE. Show that $\mathrm{Out}(A_6) \cong Z_2 \times Z_2$.

We next turn to an apparently much more complicated group:

$$G = PSL_n(q),\ n \geq 2.$$

We include the exceptional cases $PSL_2(2)$ and $PSL_2(3)$ since they are not exceptional in this calculation. First we observe the existence of automorphisms of certain types, then we prove that every automorphism is a product of automorphisms of those types. The four types are:

1. Inner automorphisms.
2. Diagonal automorphisms. A diagonal automorphism is one induced by conjugation by a diagonal matrix (in $PGL_n(q)$), or by any conjugate. Since diagonal matrices have arbitrary determinants, the group generated by all inner and diagonal automorphisms is $PGL_n(q)$.
3. Field automorphisms. Any automorphism of the underlying field $\boldsymbol{F}_q$ induces in an obvious manner an automorphism of the matrix group $GL_n(q)$ and hence an automorphism of G. Such an automorphism—and any $PGL_n(q)$-conjugate, to make the notion coordinate-free—is called a field automorphism.
4. Graph automorphisms. If $n \geq 3$, the automorphism $X \mapsto {}^TX^{-1}$ is called a graph automorphism, as are its $PGL_n(q)$-conjugates. Of course this automorphism also exists when $n = 2$, but then it is inner, induced by the companion matrix of the polynomial $t^2 + 1$.

Let $I = PSL_n(q)$, let $ID = PGL_n(q)$, and let IDF be the group generated by ID and all field automorphisms, and let IDG be the group generated by ID and all graph automorphisms. Let $IDFG$ be the group generated by all automorphisms of all four types.

THEOREM 11.8. *$Aut(PSL_n(q)) = IDFG$. The subgroups I, ID, IDF are all normal in $Aut(PSL_n(q))$, and*

$$ID/I \cong Z_{(n,q-1)},\ IDF/ID \cong Aut(\boldsymbol{F}_q),\ IDG/ID \cong Z_2,\ (n \geq 3)$$
$$\text{and } IDFG/ID = IDF/ID \times IDG/ID.$$

Since $\mathrm{Aut}(\boldsymbol{F}_q)$ is cyclic, as are ID/I and IDG/ID, we conclude:

COROLLARY 11.9. *$Out(PSL_n(q))$ is solvable.*

The remainder of this section is devoted to a proof of Theorem 11.8. It is due to R. Steinberg. Roughly speaking the idea of the proof is to show that an automorphism, when multiplied if necessary by a graph automorphism, acts on the underlying projective geometry $P^{n-1}(\boldsymbol{F}_q)$, and then to show that any such automorphism is the product of a collineation (induced by an element of $GL_n(q)$) and a field automorphism. This latter statement, sometimes called the fundamental theorem of projective geometry, can be given a very geometric proof. However we shall give a group-theoretic version of it. The proof either way can be generalized to compute the automorphism group of any classical group and indeed any finite simple group of Lie type. The IDFG statement still holds, for suitable definitions of diagonal, field and graph automorphisms.

Set

$$K = GL_n(q), G = SL_n(q),\ \text{and } \overline{G} = PSL_n(q). \tag{11.3}$$

Let α be an arbitrary automorphism of G[2]. Our proof will develop some of the machinery of B, N-pairs, and we shall pause from time to time to make deductions about α, particularly conclusions that modifying α by an inner, diagonal, graph or field automorphism, we may assume that α has some additional property. A chain of such reductions will eventually allow us to assume that $\alpha = 1$, whence the conclusion that the original α is a product of involutions of the four types.

Let B be the upper triangular subgroup of G, H the diagonal subgroup of G, and U the upper triangular unipotent subgroup of G (i.e., with 1's on the diagonal), so that $U \lhd B$ and indeed

$$B = UH, \text{ a semidirect product.} \tag{11.4}$$

Let V be the lower triangular unipotent subgroup of G. Thus VH is the lower triangular subgroup of G.

An order calculation shows that $U \in \mathrm{Syl}_p(K)$ and hence $U \in \mathrm{Syl}_p(G)$, so that $\overline{U} \in \mathrm{Syl}_p(\overline{G})$. Furthermore, a simple calculation shows that U leaves invariant exactly one subspace of the underlying vector space of each dimension, and so $N_G(U)$ leaves those subspaces invariant, leading to the conclusion

$$N_G(U) = B.$$

By Sylow's theorem, we may multiply α by an inner automorphism in order to be able to assume that

$$\alpha \text{ leaves } U \text{ (or } \overline{U}) \text{ invariant.}$$

We wish to argue that such a modification is possible so that α leaves V invariant, and indeed leaves certain subgroups P_i and P_i^- invariant. Here $P_1, \ldots, P_{n-1}$ contain U and P_i is defined by the vanishing of all entries below the diagonal, with the exception of the $(i+1, i)$ entry. Likewise P_i^- contains V and is defined by the vanishing of all entries above the diagonal except $(i, i+1)$. In order to see that this modification is possible, we introduce the elements of the "Bruhat decomposition" of G.

For each ordered pair (i, j) of distinct integers from 1 to n, define the corresponding *root subgroup* X_{ij} and root elements $x_{ij}(c)$ as follows:

$$x_{ij}(c) = I + ce_{ij}; \qquad X_{ij} = \{x_{ij}(c) \,|\, c \in \boldsymbol{F}_q\} \tag{11.12}$$

which is an H-invariant subgroup of U. The root subgroups are additively isomorphic to the field:

$$X_{ij} \cong \boldsymbol{F}_q^+, \qquad x_{ij}(c) \mapsto c$$

and so are elementary abelian groups of order q. We emphasize that

$$x_{ij}(c) = 1 \iff c = 0;$$

this is true not only in G, but also in $\overline{G}$, and it allows our proof to go over to $\overline{G}$, in spite of the fact that certain elements of H become trivial in $\overline{G}$.

Call an ordered pair (i, j) *positive* (resp., negative) if and only if $i < j$ (resp., $i > j$). Then U is generated by the positive X_{ij}, and V by the negative ones.

Let N be the group of all monomial matrices in G, that is, the group of all matrices with a single nonzero entry in each row and column. The intersection

[2]The proof for G adapts without serious difficulty, and practically without change, to a proof for $\overline{G}$.

of this with B is just H. Moreover, each element of N determines a certain row permutation; this is a homomorphism onto S_n, with kernel H. Thus

$$B \cap N = H \lhd N, \tag{11.5}$$

and mapping each element of N to the corresponding row permutation,

$$N/H \cong S_n. \tag{11.6}$$

(Mapping to the column permutation gives an anti-isomorphism.) A coset of the form Bn or nB, with $n \in N$, is determined simply by the coset $nH = Hn$, since $H \leq B$. So if we define

$$W = N/H, \tag{11.7}$$

then for $w \in W$ the notation Bw and wB for a certain coset of B in G is well-defined. For any distinct i, j, the notation X_{ij}^w is similarly well-defined since H normalizes X_{ij}. Indeed

$$X_{ij}^w = X_{i^w j^w}.$$

Let w_0 be the element $(1\,n)(2\,n-1)\cdots$ of W. Then $w_0^2 = 1$, and

$$w_0 U w_0 = V, \text{ so that } w_0 U = V w_0. \tag{11.8}$$

Now let $g \in G$ be given. Multiplying g on the left by an element of U amounts to applying a sequence of row operations, adding multiples of various rows to higher rows. By a suitable such multiplication we may reduce one column to zero in all rows but the last, another column to zero in all rows but the last two, etc. We do this moving from left to right. That is we choose the leftmost nonzero entry in the bottom row and use it to clear out the column above it. In the next row up we again choose the leftmost nonzero entry and use it to clear out the rest of the column above it. As a result we obtain a matrix bg, $b \in B$, which can be reduced to a monomial matrix by column operations, replacing columns by themselves plus multiples of columns to their left. The result is that $bgb' = n \in N$ for some $b, b' \in B$. Therefore $g \in BnB$. We have proved that

$$G = \bigcup_{w \in W} BwB. \tag{11.9}$$

Exercise. Distinct elements $w \in W$ give distinct (and therefore disjoint) double cosets in (11.9).

In particular

$$G = \langle B, N \rangle.$$

Now a conjugate of B by an arbitrary element $g = bwb' \in G$ may be written as $B^g = B^{bwb'} = (B^w)^{b'}$. Therefore the intersection of B with this conjugate is

$$B \cap B^g = (B \cap B^w)^{b'},$$

a conjugate of $B \cap B^w$. If w happens to carry some positive pair (i, j) to a positive pair (k, l), then $B \cap B^w$ contains X_{kl} and so has order divisible by p. If w carries no positive pair to a positive pair, then $w = w_0$ and $B \cap B^w = B \cap B^{w_0} = UH \cap VH = H$, of order relatively prime to p. It follows that the only Sylow p-subgroups U^* of G such that $U \cap U^* = 1$ are those of the form $U \cap U^g$ with $g \in Bw_0B$, and these are all B-conjugate to V. Thus B permutes transitively the set of all Sylow

p-subgroups of G disjoint from U. Since B normalizes U, we may further modify α by an inner automorphism in order to be able to assume that

$$\alpha \text{ normalizes both } U \text{ and } V.$$

But $N_G(U) = B$ and similarly $N_G(V) = VH$, so

$$\alpha \text{ normalizes } UH \cap VH = H.$$

We next consider what happens when two double cosets BwB and $Bw'B$ are multiplied. The result, $BwBw'B$, is of course invariant under left and right multiplication by B, and so is a union of B-double cosets. We do not answer the question in general, but consider it only when w or w' is a "fundamental involution" in $W \cong S_n$. We choose $w_{ij} \in W$ to be the element corresponding to the transposition $(ij) \in S_n$. The fundamental involutions are then

$$w_{12}, w_{23}, \ldots, w_{n-1\,n}. \tag{11.10}$$

EXERCISE. These involutions generate W.

We next set

$$n_{i-1\,i} = \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix}$$

and argue that

$$\text{For any } i = 1, \ldots, n-1,\ P_i = B \cup Bn_{i-1\,i}B.$$

Clearly $n_{i-1\,i} \in P_i$ and $B \leq P_i$. Furthermore, if we let L_i be the block-diagonal $SL_2(q)$ subgroup of G supported in the 2×2 square formed by rows and columns $i-1$ and i, then we have $P_i = BL_i$ and so it suffices to show that if we put $B_i = B \cap L_i$, then

$$L_i = B_i \cup B_i n_{i-1\,i} B_i. \tag{11.11}$$

But in the action of L_i on the corresponding projective space $P^1(q)$, the stabilizer of a point is B_i. The desired equation (11.11) is then equivalent to the assertion that B_i acts transitively on the set of remaining points, that is, that the action on $P^1(q)$ is 2-transitive. But this fact has already been observed.

Now

$$Q_i = \prod \{X_{jk} \mid j < k,\ (j,k) \neq (i-1, i)\} \tag{11.14}$$

is the subgroup of U defined by vanishing in the $(i-1, i)$ coordinate. It is easily checked by matrix calculation that

$$Q_i \lhd P_i = Q_i L_i \tag{11.15}$$

and

$$U = X_{i-1\,i} Q_i, \tag{11.16}$$

a semidirect product.

Now we can prove the Bruhat Lemma:

BRUHAT LEMMA. *Let $w \in W$, and let w_i be a fundamental involution. Then*

$$Bw_iBwB = Bw_iwB \text{ or } Bw_iwB \cup BwB,$$

according as $w((i-1,i))$ is positive or negative.

PROOF. First suppose that $w((i-1,i))$ is positive. Then writing $B = HU = HX_{i-1\,i}Q_i$ and using the fact that $n_{i-1\,i}$ normalizes Q_i and H (by (11.15) and (11.5)), we have

$$w_{i-1\,i}Bw = w_{i-1\,i}HQ_iX_{i-1\,i}w = HQ_iw_{i-1\,i}wX^w_{i-1\,i} \leq Bw_{i-1\,i}wB,$$

the last step since $(i-1^w, i^w)$ is assumed positive, and thus $X^w_{i-1\,i} \leq U \leq B$.

Next suppose that $(i-1^w, i^w)$ is negative. Set $w_1 = w_{i-1\,i}w$. Then $(i-1^{w_1}, i^{w_1})$ is positive, so

$$Bw_{i-1\,i}Bw_1B = BwB$$

by the previous case. Now

(11.17)
$$w_{i-1\,i}Bw = w_{i-1\,i}Bw_{i-1\,i}w_1 \subseteq P_iw_1 = (B \cup Bw_{i-1\,i}B)w_1 \subseteq Bw_1B \cup BwB.$$

Moreover, the conjugate $w_{i-1\,i}Bw_{i-1\,i}$ meets both of the double cosets B and $Bw_{i-1\,i}B$ nontrivially (it doesn't lie in B, and contains 1), so the left side of (11.17) meets both the double cosets on the right side nontrivially. Multiplying by B on left and right yields the desired equality. □

Using the Bruhat lemma we can prove:

PARABOLIC LEMMA. *Any subgroup of G containing B is generated by B and some subset of $\{w_{12}, w_{23}, \ldots, w_{n-1\,n}\}$.*

(By w_{ij} we really mean some preimage of this element in N; which preimage is irrelevant, since $H \leq B$.)

PROOF. Let $B \leq P \leq G$. By (11.9), $P = \cup_{w\in S}BwB$ for some subset S of W containing 1. Since P is a subgroup of G, S is a subgroup of W. We must show that S is generated by some $w_{i-1\,i}$'s. If $S = 1$, then there is nothing to prove. Let $1 \neq w \in S$. Since $w \neq 1$, w cannot preserve the natural order on $\{1, 2, \ldots, n\}$, and so we may fix i such that $(i-1^w, i^w)$ is negative. Set $w_1 = w_{i-1\,i}w$. We also define $N(w)$ to be the number of positive pairs taken by w to a negative pair. Notice that the only positive pair taken by $w_{i-1,i}$ to a negative pair is $(i-1, i)$ itself, and so

$$N(w_1) = N(w) - 1. \tag{11.18}$$

We shall argue that

$$w_1 \in S. \tag{11.19}$$

It will follow that $w_{i-1\,i} \in S$. Inductively (on $N(w)$) it will also follow that we can obtain a factorization of w as a product of fundamental involutions all of which lie in S. Varying w over S we obtain a collection of fundamental involutions which together clearly will generate W, as required.

To prove (11.19), we us the Bruhat Lemma to conclude that $w_{i-1\,i}Bw$ meets BwB nontrivially. Therefore $w^{-1}Bw = w_1^{-1}w_{i-1\,i}Bw$ meets $w_1^{-1}BwB$ nontrivially. But the left side lies in P, and $B \leq P$ and $w \in S$, so $w_1 \in S$. This completes the proof. □

As an immediate corollary we find that

(11.20)

$P_1, \ldots, P_{n-1}$ are the only subgroups of G containing B as a maximal subgroup.

This is clear since $P_i = \langle B, w_i \rangle$

Therefore α permutes $\{P_1, \ldots, P_{n-1}\}$. But α normalizes V, and $P_i \cap V = X_{i\,i-1}$. For simplicity let us write

$$X_{i\,i-1} = X_{-i} \text{ and } X_{i-1,i} = X_i.$$

Thus α permutes $\{X_{-1}, \ldots, X_{-(n-1)}\}$. Similarly, it permutes $X_1, X_2, \ldots X_{n-1}$ among themselves, and so we obtain permutations ρ and σ of $\{1, \ldots, n-1\}$ such that

$$X_i^\alpha = X_{i^\rho} \text{ and } X_{-i}^\alpha = X_{-i^\sigma},\ i = 1, \ldots, n-1. \tag{11.21}$$

In fact

$$\rho = \sigma.$$

This follows from the following two facts, which can be verified by an easy matrix calculation:

$$[X_i, X_{-j}] = 1 \text{ if } i \neq j,$$

while

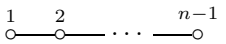
$$\langle X_i, X_{-i} \rangle = L_i, \text{ a nonabelian group.}$$

Thus $[X_i, X_{-j}]$ is nonabelian if and only if $i = j$; applying α we find that $i = j$ if and only if $i^\rho = j^\sigma$, so $\rho = \sigma$.

In the same vein we can calculate that

$$[X_i, X_j] \neq 1 \iff |i - j| = 1.$$

We can form the "Dynkin diagram" of G; it is a graph whose nodes are the X_i, joined in this case by a single edge if and only if the corresponding X_i do not commute elementwise. Thus the Dynkin diagram for G is just



Because this graph is defined by group theoretic properties of the X_i, which properties are preserved by α, the permutation ρ of the vertices must be a symmetry of the Dynkin diagram. There are thus two possibilities for ρ: it is the identity or else ($n \geq 2$ and) it flips the diagram over.

But if $n \geq 3$, we can produce an automorphism which flips the diagram over: namely, the graph automorphism

$$\gamma : g \mapsto {}^T g^{-1}$$

carries each X_{ij} to X_{ji}, and hence the product $\alpha_0 = \gamma\omega_0$ of γ with ω_0, the inner automorphism corresponding to an arbitrarily chosen preimage of w_0 in N, satisfies $X_{ij}^{\alpha_0} = X_{n-j\,n-i}$. Now α_0 satisfies all the conditions that our α has been reduced to having at this point: it leaves U and V invariant, and clearly flips the Dynkin diagram. Moreover, α_0 is the product of a graph automorphism and an inner automorphism.

Thus, replacing α if necessary by $\alpha\alpha_0$, we may now assume that

$$\rho = 1, \text{ that is, } \alpha \text{ leaves each } X_i \text{ and } X_{-i} \text{ invariant.}$$

We are beginning to close in on α. Write

$$x_i(c) = 1 + ce_{i-1\,i} \text{ and } x_{-i}(c) = 1 + ce_{i\,i-1}.$$

Whether we are in G or $\overline{G}$, we have

$$x_i(c) = x_{i'}(c') \iff i = i' \text{ and } c = c'.$$

There thus exist bijections ϕ_i and ψ_i from $\boldsymbol{F}_q$ to itself such that

$$x_i(c)^\alpha = x_i(c^{\phi_i}) \text{ and } x_{-i}(c)^\alpha = x_{-i}(c^{\psi_i}).$$

Since $x_{\pm i}(0) = 1$, we certainly have $\phi_i(0) = 0 = \psi_i(0)$ for each i. Letting h be the diagonal matrix

$$h = \operatorname{diag}(1, \phi_1(1)^{-1}, \phi_1(1)^{-1}\phi_2(1)^{-1}, \dots, \phi_1(1)^{-1}\phi_2(1)^{-1}\cdots\phi_{n-1}(1)^{-1}),$$

we modify α by the diagonal automorphism which is conjugation by h and achieve

$$\phi_i(1) = 1,\ i = 1, \dots, n-1.$$

Moreover, since

$$x_i(c)x_i(c') = x_i(c + c')$$

and similarly for $x_{-i}(c + c')$, while α is an automorphism,

$$\phi_i \text{ and } \psi_i \text{ preserve addition, for each } i.$$

Our next goal is to show that the ϕ_i and ψ_i are all equal, and that they are in fact an automorphism of $\boldsymbol{F}_q$. This will imply that α is a field automorphism.

To accomplish this we consider the equation (for a fixed i)

$$x_i(t)x_{-i}(u)x_i(t) = x_{-i}(u)x_i(t)x_{-i}(u),\ t, u \in \boldsymbol{F}_q^\times.$$

A matrix calculation shows that this equation holds if and only if $tu + 1 = 0$. Applying α we find that

$$tu + 1 = 0 \iff t^{\phi_i}u^{\psi_i} + 1 = 0, \tag{11.22}$$

and if we set

$$n_i(t) = x_i(t)x_{-i}(-t^{-1})x_i(t),$$

then

$$n_i(t)^\alpha = n_i(t^{\phi_i}).$$

We also set

$$h_i(t) = n_i(t)n_i(1)^{-1}.$$

Calculating again with matrices we see that

$$h_i(t) = n_i(t)n_i(1)^{-1} = \operatorname{diag}(1, \dots, 1, t, t^{-1}, 1, \dots, 1),$$

with t in the i-th place. Since $\phi_i(1) = 1$ it follows that

$$h_i(t)^\alpha = h_i(t^{\phi_i}).$$

Another calculation with matrices shows that

$$x_i(t)^{h_j(u)} = x_i(tu^{n_{ij}})$$

where

$$n_{ij} = 2 \text{ if } i = j,\ n_{ij} = \pm 1 \text{ if } i = j \pm 1,\ n_{ij} = 0 \text{ otherwise.}$$

Thus

$$(tu^{n_{ij}})^{\phi_i} = t^{\phi_i}(u^{\phi_j})^{n_{ij}}$$

for arbitrary $t, u \in \boldsymbol{F}_q^\times$.

Taking $i = j$ we find that $(tu^2)^{\phi_i} = t^{\phi_i}(u^{\phi_i})^2$. Thus ϕ_i is multiplicative when restricted to the set of squares, or indeed by additivity, to the set of sums of squares. But in a finite field every element is a sum of squares and so ϕ_i is multiplicative, hence

$$\phi_i \in \mathrm{Aut}(\boldsymbol{F}_q).$$

Taking $i = j \pm 1$ then shows that $\phi_{i-1} = \phi_i$ for all i, so the ϕ_i are all equal. Since they are all automorphisms, (11.22) then shows that $\phi_i = \psi_i$ for all i. Let ϕ be this common automorphism of $\mathrm{Aut}(\boldsymbol{F}_q)$. Modifying α by the inverse of the corresponding field automorphism of G (or $\overline{G}$), we may now assume that

$$\alpha \text{ fixes each element of each } X_i \text{ and each } X_{-i}.$$

Thus α fixes each element of the group $G_0 = \langle X_i, X_{-i} \,|\, 1 \le i < n\rangle$. But G_0 contains all $n_i(t)$, hence contains a monomial matrix corresponding to any given permutation, hence by conjugation contains all X_{ij}; containing all $n_i(t)$ it also contains all $h_i(t)$ and hence all diagonal matrices (of determinant 1); so it contains B and N, so is the entire group.

$$\alpha = 1.$$

This completes the proof that every automorphism is the product of inner, diagonal, field and graph automorphisms. The rest of the theorem—the structure assertions about sections of $\mathrm{Aut}(G)$—is left as an exercise.

Long before the classification, Schreier formulated the following conjecture, which has now been verified is a corollary of the classification (and observation of the automorphism groups of the finite simple groups, all of which have been calculated).

Schreier Conjecture. *If G is a finite simple group, then Out(G) is solvable.*

No proof exists outside of the classification. Notice that $\mathrm{Out}(G)$ is "very solvable"—the derived length of $\mathrm{Out}(G)$ is quite small—in the cases we have computed, and this is typical. In the most complicated case the derived length is just 3, and this only occurs rarely; most of the time $\mathrm{Out}(G)$ is either abelian or metabelian (an extension of an abelian group by an abelian group). A conceptual proof might be aided by a sharper "conjecture" than this. However, remarkably little progress has been made on this question in the abstract.

12. Group Extensions and Cohomology; Schur-Zassenhaus Theorem

We change direction now and discuss the extension problem for finite groups. Roughly speaking it is the question: given groups A and B, how many groups G are there with a normal subgroup $N \cong A$ such that $G/N \cong B$, and how can one distinguish them?

As observed above, given an action $\phi : B \to \mathrm{Aut}(A)$ of B on A, there is the corresponding "trivial" or "split" extension, that is, the semidirect product of A by B. Any extension G such that N has a complement in G (necessarily isomorphic to B) and yielding the given action of B on A is isomorphic to the semidirect product; this is a routine exercise.

Thus another much weaker question is: for which actions is every extension split? One may also ask: in a semidirect product, how many conjugacy classes of complements are there? These questions lead to the study of cohomology, which has several applications to the study of the structure of groups.

A general answer to the extension problem is far away. For the weaker questions, there are numerous cases (i.e., triples (A, B, ϕ)) for which the answer is known to be positive or negative. Perhaps the most striking is the Schur-Zassenhaus Theorem:

THEOREM 12.1 (SCHUR-ZASSENHAUS). *Suppose that $N \lhd G$ and that $|N|$ and $|G : N|$ are relatively prime. Then N has a complement in G, and any two such complements are N-conjugate.*

As stated this theorem contains a trap which makes it very deep. Namely, what we can prove at this stage is a "weak" Schur-Zassenhaus theorem: N has a complement in G, and assuming that either N or G/N is solvable, then any two complements are N-conjugate. The only known way to complete the proof of Theorem 12.1 is then to quote Feit-Thompson Theorem, which is a very deep result whose proof is outside the scope of this course:

THEOREM 12.2 (WALTER FEIT & JOHN THOMPSON, 1963). *Every finite group of odd order is solvable.*

Since either $|N|$ or $|G : N|$ is odd, Theorem 12.1 then follows.

In order to prove the "weak" Schur-Zassenhaus Theorem, we shall proceed in two steps:

1. Reduction to the case that N is abelian.
2. The proof in the case that N is abelian.

The reduction is rather routine, using the Frattini argument and induction. The second step, when N is abelian, will be accomplished by the introduction of the cohomology groups $H^i(G/N, N)$, $i = 1, 2$, the interpretation of these cohomology groups as obstructions to the two assertions of the theorem, and then a proof that if $|N|$ and $|G/N|$ are relatively prime, then these two cohomology groups vanish.

We now begin Step 1. Let G be a minimal counterexample to the weak Schur-Zassenhaus Theorem. Notice that whenever $K \lhd G$ with $K \leq N$, then G/K satisfies the hypotheses of the theorem with respect to the normal subgroup N/K. Also whenever $M \leq G$ with $G = MN$, then M satisfies the hypotheses with respect to the normal subgroup $M \cap N$.

LEMMA 12.3. *N is a minimal normal subgroup of G.*

PROOF. Suppose not, and let $K \lhd G$ with $1 < K < N$. Set $\overline{G} = G/K$. Then $\overline{G}$ and $\overline{N}$ satisfy the hypotheses, but $\overline{G}$ has smaller order than G so satisfies the conclusion by minimality. Let $\overline{M}$ be a complement to $\overline{N}$ in $\overline{G}$. Then the full preimage M of $\overline{M}$ in G has order $|K| \cdot |G/N|$ and has the normal subgroup K, so by minimality again, K has a complement in M, which is then a complement to N in G. Furthermore, if S and T are complements to N in G, then $\overline{S}$ and $\overline{T}$ are complements to $\overline{N}$ in $\overline{G}$, so by minimality are conjugate in $\overline{G}$. In proving that S and T are conjugate we may therefore replace S by a conjugate and assume that $\overline{S} = \overline{T}$. Then the full preimage W of $\overline{S}$ again satisfies the hypotheses, relative to K, and S and T are complements to K in W, so are conjugate again by minimality. Thus G satisfies the conclusions of the theorem, contradiction. □

LEMMA 12.4. *If G is a minimal counterexample to the existence of complements, then N is an elementary abelian p-group for some prime p.*

PROOF. In view of the preceding lemma, it is enough to show that N is solvable, and we show in fact that N is nilpotent. To do this it is enough to show that $N \leq \Phi(G)$.

But if $N \not\leq \Phi(G)$, then there is a maximal subgroup M of G such that $N \not\leq M$. Therefore $G = NM$. Then M and $M \cap N$ satisfy the hypotheses of the theorem, so by minimality there is a complement S to $M \cap N$ in M, and then $|S| = |M/M \cap N| = |G/N|$ so S is a complement to N in G, contradiction.

LEMMA 12.5. *If G is a minimal counterexample to the conjugacy of complements, then N is an elementary abelian p-group for some prime p.*

PROOF. By hypothesis, either N or G/N is solvable. If N is solvable we are done. So assume that G/N is solvable. Set $\overline{G} = G/N$.

Let $\overline{M} = M/N$ be a minimal normal subgroup of $\overline{G}$. Thus $\overline{M}$ is a q-group for some prime q. Let S_1 and S_2 be complements to N in G. Set $T_i = S_i \cap M$. Notice that $G = NS_i = MS_i$ so $|T_i| = |S_i|M|/|G| = |M|/|N|$, so the T_i are complements to N in M. Hence they are Sylow q-subgroups of M, so are $M = NT_i$-conjugate and hence N-conjugate. Replacing S_1 by an N-conjugate we may therefore assume that $T_1 = T_2$.

Since $M \lhd G$, we have $T_i \lhd S_i$ for each i. Set $W = N_G(T_1)$. By a Frattini argument, $G = MW$. By the last paragraph, $\langle S_1, S_2 \rangle \leq W$. Thus S_1 and S_2 are complements to $N \cap W$ in W. We reach a contradiction whether $W < G$ or $W = G$ by showing that S_1 and S_2 are N-conjugate. If $W < G$ then S_1 and S_2 are $N \cap W$-conjugate by minimality. If $W = G$, then $T_1 \lhd G$ and we may pass to G/T_1 and again get by minimality that S_1/T_1 and S_2/T_1 are NT_1/T_1-conjugate, so S_1 and S_2 are N-conjugate, contradiction. □

This completes Step 1 of the proof.

For Step 2 we now introduce cohomology groups[3]. We start with a group H acting on an abelian group A, via an action

$$\phi : H \to \mathrm{Aut}(A), \quad g \mapsto (a \mapsto a^g).$$

We shall consider extensions of A by H (or H by A, in more popular terminology). By definition such an extension is a short exact sequence

$$1 \to A \xrightarrow{\iota} E \xrightarrow{\psi} H \to 1$$

of finite groups such that the conjugation action of E on A, when factored through $E/A \cong H$, is the given mapping ϕ, that is, $a^g = a^{\phi(\psi(g))}$ where the left side is computed in E.

We identify A with its image in E, so that ι is just the inclusion mapping.

By definition this extension splits if and only if A has a complement C in E. Such a complement C is isomorphic to H via ψ, and the inverse of $\psi|C$ yields a mapping $\sigma : H \to E$:

$$E \underset{\sigma}{\overset{\psi}{\rightleftarrows}} H \qquad \text{such that } \psi \circ \sigma = 1_H.$$

[3]Finiteness of the groups is irrelevant to this discussion, except for the places where it is obviously relevant!.

Indeed, splitting is equivalent to the existence of such a mapping σ, since the image of σ would be the required complement.

Now we consider the conjugacy question. Suppose that E is a split extension, and that $E = AC$ for some complement. We seek to parametrize the set of all complements in such a way that we can identify when complements are conjugate. We write $C = \{\sigma(h) \mid h \in H\}$.

If X is any complement, then X consists of elements $\tau(h)$, $h \in H$, where τ is another splitting homomorphism, and we may write

$$\tau(h) = \sigma(h)f(h),\ h \in H$$

for some function

$$f : H \to A.$$

The fact that τ and σ are both homomorphisms implies that

$$\begin{aligned} \sigma(h)\sigma(k)f(hk) = \sigma(hk)f(hk) = \tau(hk) = \tau(h)\tau(k) = \sigma(h)f(h)\sigma(k)f(k) \\ \text{so } f(hk) = \sigma(k)^{-1}f(h)\sigma(k)f(k),\ \text{that is,} \\ f(hk) = f(h)^k f(k) \text{ for all } h, k \in H \end{aligned}$$

This condition is called the 1-cocycle condition, and any function satisfying it is called a 1-cocycle of H in A. Conversely, any function f satisfying this condition leads to a splitting homomorphism τ and hence to a complement. Thus the set of complements is parametrized by the set of all such functions f.

Notice that if f and g are both 1-cocycles, then their pointwise product is as well (A is abelian by assumption!).

Thus the set of all 1-cocycles is a group, notated

$$Z^1(H, A),$$

and it is in bijective correspondence with the set of complements to A in the (arbitrary) split extension E.

When is the complement X above actually A-conjugate to C? The condition is that there exist $a \in A$ such that

$$\sigma(h) = \tau(h)^a \text{ for all } h \in H.$$

But $\tau(h)^a = \sigma(h)^a f(h)^a = \sigma(h)^a f(h)$, so the condition is that

$$\sigma(h) = \sigma(h)^a f(h),\ \text{or equivalently } f(h) = a^{-1}a^h.$$

Here a has to be independent of h. Notice that any function f in this form (for some $a \in A$; actually f is called δa) is really a 1-cocycle; this is just the familiar commutator identity $[a, \sigma(hk)] = [a, \sigma(h)]^k[a, \sigma(k)]$. In the correspondence between $Z^1(H, A)$ and the set of complements to A in E, the set of conjugates of C therefore corresponds to the set

$$B^1(H, A)$$

of 1-cocycles of the form $f = \delta a$ for some $a \in A$. Such cocycles are called 1-coboundaries. It is easy to check that $B^1(H, A)$ is a subgroup of $Z^1(H, A)$.

DEFINITION 12.6. $H^1(H, A) = Z^1(H, A)/B^1(H, A)$.

We have proved:

PROPOSITION 12.7. *In a split extension of an abelian group A by a group H, all complements are conjugate if and only if $H^1(H, A) = 1$.*

More generally:

PROPOSITION 12.8. *For a group H acting on an abelian group A, the group $H^1(H, A)$ is in bijective correspondence with the set of conjugacy classes of complements to A in an arbitrary split extension of A by H.*

The proof is as straightforward as the preceding discussion and is left to the reader.

The following result therefore has the conjugacy part of the Schur-Zassenhaus Theorem as an immediate consequence, for it implies that $H^1(H, A) = 1$ if $|H|$ and $|A|$ are relatively prime.

PROPOSITION 12.9. *The exponent of $H^1(H, A)$ divides both the exponent of A and the order of H.*

PROOF. Since the operation on $Z^1(H, A)$ is pointwise, it is immediate that the exponent of $Z^1(H, A)$ divides that of A, and the exponent of $H^1(H, A)$ therefore does as well.

Let $f \in Z^1(H, A)$. To complete the proof we must show that $f^{|H|} \in B^1(H, A)$. For any $h, k \in H$ we have

$$f(k) = f(hk)f(h)^{-k}.$$

Fixing k and taking the product over all $h \in H$ we obtain

$$f(k)^{|H|} = b^{-1}b^k, \text{ where } b = \prod_{h \in H} f(h)^{-1}$$

Therefore $f^{|H|} \in B^1(H, A)$. □

For the existence part of the Schur-Zassenhaus Theorem, we investigate an arbitrary extension

$$1 \to A \to E \to H \to 1$$

to see whether it splits. Again the idea is to parametrize the set of all extensions somehow, and distinguish the split extension among them.

Given such an extension, a section of H in E is defined to be any injective mapping $\sigma : H \to E$. This amounts to a set $\{\sigma(h) \mid h \in H\}$ of coset representatives for A in E. Such a set of coset representatives satisfies the condition that $\sigma(h)\sigma(k)$ is in the coset of $\sigma(hk)$, and so there exists a (unique) function $f : H \times H \to A$ such that

$$\sigma(h)\sigma(k)f(h, k) = \sigma(hk) \text{ for all } h, k \in H.$$

Now f is not arbitrary since $\sigma(g)\sigma(h)\sigma(k)$ can be associated two ways, and correspondingly we work out

$$\begin{aligned}\sigma(ghk) &= \sigma(gh)\sigma(k)f(gh, k) = \sigma(g)\sigma(h)f(g, h)\sigma(k)f(gh, k)\\ &= \sigma(g)\sigma(hk)f(g, hk) = \sigma(h)\sigma(h)\sigma(k)f(h, k)f(g, hk),\end{aligned}$$

whence

$$f(g, h)^k f(gh, k) = f(g, hk)f(h, k) \text{ for all } g, h, k \in H.$$

This is called the 2-cocycle condition, and the set of all functions f satisfying it is

$$Z^2(H, A),$$

again a group under pointwise operations.

Thus every extension determines several 2-cocycles, one for each section σ. How are all these cocycles (for a given extension) related? Given a section σ for an extension, all the sections τ for that extension are the functions satisfying

$$\sigma(h) = \tau(h)b(h)$$

for some function $b : H \to A$. If f and g are the cocycles arising from σ and τ, then we find that

$$\sigma(h)\sigma(k)f(h,k) = \sigma(hk) = \tau(hk)b(hk) = \tau(h)\tau(k)g(h,k)b(hk), \text{ so}$$
$$\tau(h)\tau(k)g(h,k)b(hk)f(h,k)^{-1} = \sigma(h)\sigma(k) = \tau(h)b(h)\tau(k)b(k) = \tau(h)\tau(k)b(h)^k b(k),$$
$$\text{whence } g(h,k)f(h,k)^{-1} = b(h)^k b(hk)^{-1} b(k).$$

Thus if for any function $b : H \to A$ we define

$$(\delta b)(h,k) = b(h)^k b(hk)^{-1} b(k),$$

we have

$$gf^{-1} = \delta b.$$

Notice that for any b, the function δb satisfies the cocycle condition, that is, $\delta b \in Z^2(H,A)$. (Check this!) Also $\delta(b_1 b_2) = \delta b_1 \cdots \delta b_2$. Thus the set of all functions of the form δb for some $b : H \to A$ is a subgroup of $Z^2(H,A)$ is a subgroup, which is called

$$B^2(H,A).$$

DEFINITION 12.10. $H^2(H,A) = Z^2(H,A)/B^2(H,A)$.

Our discussion shows that the sections corresponding to a given extension form a coset of $B^2(H,A)$ in $Z^2(H,A)$, that is, an element of $H^2(H,A)$.

Now a split extension has a section which is a homomorphism, and the corresponding cocycle is the identity mapping. Conversely if an extension corresponds to the trivial element of $H^2(H,A)$, then the identity mapping is a corresponding cocycle, so that some section is a homomorphism and so the extension splits. Summarizing:

PROPOSITION 12.11. *Every extension of A by H determines a unique element of $H^2(H,A)$. This element is the identity element if and only if the extension splits.*

The following further statements can be proved in a straightforward manner:

PROPOSITION 12.12. *Associating an element of $H^2(H,A)$ to an extension, as described above, gives a bijection between $H^2(H,A)$ and the set of equivalence classes of extensions.*

Here two extensions are called equivalent if there is an isomorphism $\psi : E \to E_1$ such that the following diagram commutes:

$$\begin{array}{ccccccccc} 1 & \to & A & \to & E & \to & H & \to & 1 \\ & & \downarrow 1_A & & \downarrow \psi & & \downarrow 1_H & & \\ 1 & \to & A & \to & E_1 & \to & H & \to & 1 \end{array}$$

Thus if two extensions are equivalent, then $E \cong E_1$. However, the converse is false: equivalence is a finer relation than just isomorphism between E and E_1. For example, let $H = \langle x \rangle = Z_3$ act on $A = \langle a \rangle \times \langle b \rangle \cong Z_3 \times Z_3$ by $b^h = b$, $a^h = ab$.

One may check that up to isomorphism there are only two extensions, in which, respectively, a preimage x of h satisfies $x^3 = 1$ (the split extension) or $x^3 = b$. (For example, an extension with $x^3 = b^{-1}$ is isomorphic to the second of these...they are both split extensions of Z_9 by Z_3, relative to the same action.) On the other hand, $|H^2(H, A)| > 2$, since $H^2(H, A)$ is a nontrivial group (nonsplit extensions exist!) of exponent 3 (see the following result).

EXERCISE. Prove the assertions in the previous paragraph.

PROPOSITION 12.13. *For any H and A, the exponent of $H^2(H, A)$ divides the exponent of A and the order of H.*

PROOF. The first statement is proved as for $H^1(H, A)$. Let $f \in Z^2(H, A)$. Then

$$f(h, k) = f(gh, k)^{-1} f(g, hk) f(g, h)^{-k}$$

for all $g, h, k \in H$. Setting $b(h) = \prod_{g \in H} f(g, h)^{-1}$ and multiplying this equation over all $g \in H$, we get $f^{|H|}(h, k) = b(k)b(hk)^{-1}b(h)^k$, so $f^{|H|} = \delta b \in B^2(H, A)$. $\square$

In particular if H and A are relatively prime then $H^2(H, A) = 1$, which completes the proof of the weak form of the Schur-Zassenhaus Theorem.

Another situation in which $H^i(H, A)$ is not too difficult to compute is the case in which H is (finite) cyclic. Supposing that $H = \langle x \rangle$ has order n, we define mappings $\sigma = x - 1 : A \to A$ and $\tau = x^{n-1} + \cdots + x + 1 : A \to A$ by

$$a^\sigma = a^x a^{-1} \text{ and } a^\tau = \prod_{i=0}^{n-1} a^{x^i}.$$

Notice that since $x^n = 1$, it follows that

$$\sigma\tau = \tau\sigma = 0,$$

that is, $a^{\sigma\tau} = a^{\tau\sigma} = 1$ for all $a \in A$. Thus $\operatorname{im}\sigma \leq \ker\tau$ and $\operatorname{im}\tau \leq \ker\sigma$. Notice that $\ker\sigma = \{a \in A \mid a^x = a\}$.

PROPOSITION 12.14. *If H is cyclic and σ and τ are as just defined, then*

$$H^1(H, A) \cong \ker\tau / \operatorname{im}\sigma \text{ and } H^2(H, A) \cong \ker\sigma / \operatorname{im}\tau.$$

PROOF. A 1-cocycle $f : H \to A$ is determined by $f(x)$, since $f(x^{i+1}) = f(x^i)^x f(x)$. Indeed this relation implies that $f(x^m) = f(x)^{1+x+\cdots+x^{m-1}}$ for each m. Since $x^n = 1$, a 1-cocycle must satisfy $f(x) \in \ker\tau$. Conversely, it is easy to check that for any $a \in \ker\tau$, $f(x^i) = a^{1+x+\cdots+x^{i-1}}$ and $f(1) = 1$ defines a 1-cocycle. Thus $Z^1(H, A) \cong \ker\tau$ under $f \mapsto f(x)$. Now f is a coboundary if and only if there is $b \in A$ such that

$$a^{1+x+\cdots+x^{i-1}} = b^{-1}b^{x^i}$$

for each i. But this condition is equivalent to the condition $a = b^{-1}b^x$. Therefore $B^1(H, A)$ corresponds to $\operatorname{im}\sigma$ and so $H^1(H, A) \cong \ker\tau / \operatorname{im}\sigma$.

For $H^2(H, A)$, we can argue similarly. Or, given any extension of A by H we choose a preimage $y = \sigma(x)$ of x and consider the element y^n of A. This is obviously fixed by $\sigma(x)$, hence belongs to $\ker\sigma$. If we take a different preimage ya of x, we get instead $(ya)^n = y^n a^{1+x+\cdots+x^{n-1}} = y^n\tau(a)$. Thus each extension is associated to a coset of $\operatorname{im}\tau$, and we obtain a mapping $H^2(H, A) \to \ker\sigma / \operatorname{im}\tau$ taking the

coset of a 2-cocycle f to the image of the element $f(x,x^2)f(x^2,x^3)\cdots f(x^{n-1},1)$ (just compute that this is $\sigma(x)^n$). If two extensions correspond to the same coset of $\operatorname{im}\tau$, then it is easy to check that the corresponding extensions are isomorphic. Thus the mapping is injective. It is surjective because given any $b \in \ker\sigma$, we may construct an extension E as the set of all formal products $x^i a$, $0 \le i < n$, $a \in A$, with the product

$$x^i a x^j a' = \begin{cases} x^{i+j} a^{x^j} a' & \text{if } i+j < n \\ x^{i+j-n} a^{x^j} a' b & \text{if } i+j \ge n \end{cases}$$

In other words, we have sort of a semidirect product, but not really; there is the unique wrinkle that we have set $x^n = b$ instead of $x^n = 1$.

One must check that this binary operation yields a group; this is easily done. Furthermore, it cleary gives an extension which maps to the coset of b, since $\sigma(x^i) = x^i$, $0 \le i < n$, is a section. □

EXERCISES. If H and A are finite and H is cyclic then

$$|H_1(H,A)| = |H^2(H,A)|.$$

If A is a vector space over a finite field of characteristic p and $H = \langle x \rangle$ where x has order p^n, then $|H^1(H,A)| = |H^2(H,A)| = p^j$ where j is the number of Jordan blocks of x of order less than p^n.

13. Quasisimple Groups, Universal Central Extensions and Schur Multipliers

A quasisimple group G is an extension of $Z = Z(G)$ by the nonabelian simple group $\overline{G} = G/Z$. It is definitely not split (unless $Z = 1$, in which case the question of splitting is trivial). Indeed it is as "unsplit" as possible: $Z \le \Phi(G)$. For if M were a maximal subgroup of G not containing Z, then $G = MZ$ so $M \lhd G$, $G/M \cong Z/Z \cap M$ is abelian, and then $[G,G] \le M < G$, a contradiction.

One natural way to understand quasisimple groups is to ask, for each simple group $\overline{G}$, what the possible quasisimple groups G are which have $\overline{G}$ as their simple factor. This leads to the following notions.

DEFINITION 13.1. Let K be a finite group. A covering of K is a homomorphism

$$L \to K$$

which is surjective and whose kernel N satisfies $N \le [L,L] \cap Z(L)$.

Alternatively the requirement is that L is perfect and K is isomorphic to a central factor group of L.

In the above definition L must be finite, whether this is assumed or not. The following lemma implies this and underlies all of the succeeding discussion:

LEMMA 13.2. *Let G be a (possibly infinite) group and suppose that $G/Z(G)$ is finite. Then $[G,G]$ is finite.*

PROOF. If we let $g_1, \ldots, g_n$ be a set of coset representatives for $Z = Z(G)$ in G, then $H = \langle g_1, \ldots, g_n \rangle$ satisfies $HZ = G$, so $H \lhd G$ and as usual $[G, G] \leq H$. Indeed $G/[H, H]$ is abelian so $[G, G] = [H, H]$. There is no loss then in assuming that $G = \langle g_1, \ldots, g_n \rangle$. Now for each i and j we may write $g_i g_j = g_k z_{ij}$ for a unique k and unique $zij \in Z$. If we let $Z_0 = \langle z_{ij} \,|\, i, j \rangle$, then in $\overline{G} = G/Z_0$, the elements $\overline{g}_i$ form a complement and so $\overline{Z}$ is a direct factor of $\overline{G}$. This implies that $H \cap Z = Z_0$. Replacing G by H we may assume that Z is finitely generated. The torsion subgroup T of Z is finite, so passing to G/T we may assume that $T = 1$. Therefore Z is a free abelian group. If $[G, G]$ is not finite, then $Z \cap [G, G]$ contains a nontrivial element, which then lies in an infinite cyclic direct summand of Z. We are thus reduced to the case that Z is infinite cyclic and G/Z is finite.

Let Y be an infinite cyclic group containing Z with index n. Form the central product $G^* = GY$, identifying the subgroups Z of G and Z of Y, and letting $[G, Y] = 1$. (Explicitly, $GY = G \times Y/W$ where $W = \{(x, x^{-1}) \,|\, x \in Z\}$.) Thus G^* is an extension of Y by $\overline{G}$. Using the g_i for a section, we find that the z_{ij} form a corresponding cocycle: $f(g_i, g_j) = z_{ij}$. But there are unique $y_{ij} \in Y$ such that $y_{ij}^n = z_{ij}$, and by the uniqueness the y_{ij} give a cocycle whose n-th power is f. Since the exponent of $H^2(\overline{G}, Y)$ divides $|\overline{G}|$ by 12.13, G^* splits over Y: $G^* = YW$, $W \cap Y = 1$. It follows that $W \lhd G^*$ and in particular $[G^*, G^*] \leq W$. Therefore $[G, G] \leq W$ and as W is finite the proof is complete. □

DEFINITION 13.3. A covering

$$L \to K$$

of the perfect finite group K is a universal covering if and only if any covering $M \to K$ factors through a covering $\phi : L \to M$:

$$L \xrightarrow{\phi} M \to K.$$

This means therefore that any group covering K is isomorphic to a quotient group of L.

It is not at all obvious that universal coverings exist. Indeed, they don't exist for non-perfect groups.

EXERCISE. If K is not perfect, then K has no universal covering.

However, they do exist for perfect groups, as we shall see. Here is an example, although the proof will have to wait. Let $K = A_5$, the simple group of order 60. Let $L = SL_2(5)$. Notice that $PSL_2(5)$ is a simple group of order 60. Hence it is isomorphic to A_5 (see the exercise below). So the projection $SL_2(5) \to PSL_2(5)$, whose center $\langle \pm 1 \rangle$ has order 2, is a covering $SL_2(5) \to A_5$. It will turn out to be a universal covering. Consequently any quasisimple group whose simple quotient is A_5 is isomorphic either to A_5 or $SL_2(5)$; there are no intermediate quotients.

EXERCISE. Show that A_5 is the unique simple group of order 60. (Sketch: Let G be a simple group of order 60. By considering Sylow 5-subgroups, obtain an embedding $G \to A_6$. Then by considering the action of A_6 on the cosets of G, and using the simplicity of A_6, obtain an embedding $A_6 \to \Sigma_6$ in which G maps into the stabilizer Λ of a point, so that $\Lambda \cong \Sigma_5$.)

In the definition, uniqueness of ϕ is not required. However, the perfectness of L implies that ϕ is unique.

LEMMA 13.4. *Let $L \xrightarrow{\phi} M$ and $L \xrightarrow{\psi} M$ be coverings which become equal when composed with the projection $M \to M/Z(M)$. Then $\phi = \psi$.*

PROOF. The hypothesis means that for each $x \in L$ there is $\alpha(x) \in Z(M)$ such that $\phi(x) = \psi(x)\alpha(x)$. Since the image of α lies in $Z(M)$, and both ϕ and ψ are homomorphisms, α is a homomorphism. Hence $\ker \alpha \geq [L, L] = L$, so α is trivial and $\phi = \psi$.

This implies the "unique uniqueness" of universal covering groups.

PROPOSITION 13.5. *If K is perfect and $L_1 \to K$ and $L_2 \to K$ are universal covering, then they are intertwined by a unique isomorphism $L_1 \to L_2$.*

PROOF. Given the preceding lemma, this is the usual category-theoretic argument.

EXAMPLE. As an example of the failure of existence of universal covering groups in the non-perfect case, consider the group $G = Z_2 \times Z_2$. A covering of G is a surjective homomorphism $\phi : X \to G$ with $Z = \ker \phi \leq Z(X) \cap [X, X]$. Any Sylow subgroup P of Z of odd order is a direct factor of X by the Schur-Zassenhaus Theorem and the fact that $Z \leq Z(X)$. Thus $P \cap [X, X] = 1$, whence $P = 1$. So Z is a 2-group. Let $a, b \in X$ be such that $\phi(a)$ and $\phi(b)$ generate G. Then $\langle a, b, Z\rangle = X$. The basic commutator identities show that every commutator of two elements of X is in the group generated by $[a, b]$, and that $[a, b]^2 = [a^2, b] = 1$. Hence $[X, X]$ has order at most 2, so Z does as well, and then $|X| \leq 8$. Thus either $X \cong G$ or else $X \cong D_8$ or $X \cong Q_8$. Since there are two nonisomorphic covering groups of maximal order, there can't be a universal one.

It will turn out that the kernel of a universal covering of the perfect group G will be isomorphic to $H^2(G, \boldsymbol{Q}/\boldsymbol{Z})$. The important properties of $\boldsymbol{Q}/\boldsymbol{Z}$ are that its torsion subgroup is locally cyclic and divisible and contains elements of all orders. Thus the multiplicative group of an algebraically closed field of characteristic 0 would do just as well.

The *natural* isomorphism is with the dual group of $H^2(G, \boldsymbol{Q}/\boldsymbol{Z})$; thus dual groups arise naturally in the basic discussions of universal coverings, and so we pause to remind the reader of the essentials of duality for finite abelian groups.

For any finite abelian group A, we define its dual $\hat{A}$ to be

$$\hat{A} = \mathrm{Hom}_{\boldsymbol{Ab}}(A, \boldsymbol{Q}/\boldsymbol{Z}).$$

Now $\boldsymbol{Q}/\boldsymbol{Z}$ has only finitely many elements of a given order, indeed its solutions of $x^n = 1$ form a copy of Z_n. In particular $\hat{A}$ is another finite abelian group, and we have

$$\hat{(}A \times B) \cong \hat{A} \times \hat{B}$$

and

$$\text{If } A \cong Z_n, \text{ then } \hat{A} \cong Z_n.$$

Consequently $A \cong \hat{A}$ for every finite abelian group A. There is no natural isomorphism $\hat{A} \cong A$, however. For instance for $A \cong Z_n$, an isomorphism arises from the choice of a generator a of A via the correspondence $\phi \mapsto \phi(a)$.

LEMMA 13.6. *Let $L \stackrel{\phi}{\to} K$ be a covering, with $Z = \ker \phi$. Then there is an injective homomorphism*

$$\Psi = \Psi_\phi : \hat{Z} \mapsto H^2(K, \boldsymbol{Q}/\boldsymbol{Z})$$

which is an isomorphism if K is perfect and ϕ is universal.

PROOF. Notice that the action of K on both $\boldsymbol{Q}/\boldsymbol{Z}$, like Z, is understood to be trivial.

Let $f \in Z^2(K, Z)$ be a cocycle (unique up to a coboundary) corresponding to $1 \to Z \to L \to K \to 1$. For any homomorphism $\alpha : Z \to \boldsymbol{Q}/\boldsymbol{Z}$ of abelian groups, the mapping $\alpha \circ f : K \times K \to Z \to \boldsymbol{Q}/\boldsymbol{Z}$ is easily seen to lie in $Z^2(K, \boldsymbol{Q}/\boldsymbol{Z})$. Since the operations in $\hat{Z}$ and $Z^2(K, \boldsymbol{Q}/\boldsymbol{Z})$ are pointwise, we get a homomorphism from $\hat{Z}$ to $Z^2(K, \boldsymbol{Q}/\boldsymbol{Z})$ and hence a homomorphism Ψ.

To check the statements about injectivity and surjectivity, we interpret Ψ in terms of extension theory. Let $\alpha \in \hat{Z}$. From our given extension $1 \to Z \stackrel{\iota}{\to} L \to K \to 1$ we a construct an extension corresponding to the cocycle $\alpha \circ f$. It is the "pushout" M of

$$\begin{array}{ccc} Z & \stackrel{\iota}{\to} & L \\ \downarrow \alpha & & \\ \boldsymbol{Q}/\boldsymbol{Z} & & \end{array}$$

that is, $M = (\boldsymbol{Q}/\boldsymbol{Z} \times L)/M_0$, where M_0 is the subgroup consisting of all pairs $(\alpha(x), x^{-1})$, $x \in Z$. The mapping $j : \boldsymbol{Q}/\boldsymbol{Z} \to M$ arising from injection into the first factor is then injective and we obtain a "morphism of extensions":

$$\begin{array}{ccccccccc} 1 & \to & Z & \to & L & \to & K & \to & 1 \\ & & \downarrow \alpha & & \downarrow \beta & & \downarrow 1_K & & \\ 1 & \to & \boldsymbol{Q}/\boldsymbol{Z} & \to & M & \stackrel{\gamma}{\to} & K & \to & 1 \end{array} \tag{13A}$$

where γ arises from the projection of L to K and the trivial mapping $\boldsymbol{Q}/\boldsymbol{Z} \to K$ (the resulting mapping $\boldsymbol{Q}/\boldsymbol{Z} \times L \to K$ clearly annihilates M_0, so lifts to M).

Now f arose from some section σ of the upper extension. Then $\beta \circ \sigma$ is a section of the lower extension, and the corresponding cocycle is clearly $\alpha \circ f$. Thus $\Psi(\alpha)$ is the element of $H^2(K, \boldsymbol{Q}/\boldsymbol{Z})$ corresponding to the lower extension.

Now if $\Psi(\alpha) = 1$, then the lower extension splits, and so $[M, M] \cong K$ with $[M, M] \cap \boldsymbol{Q}/\boldsymbol{Z} = 1$. Consequenty $\beta([L, L] \cap Z) = 1$. But $Z \leq [L, L]$ by definition of covering, so $\alpha = 1$ and Ψ is injective.

Finally suppose that L is perfect and ϕ is universal. Given any element $\Gamma \in H^2(K, \boldsymbol{Q}/\boldsymbol{Z}$, form the corresponding extension $1 \to \boldsymbol{Q}/\boldsymbol{Z} \to N \to K \to 1$. Let $N_0 = [N, N]$, so that N_0 is perfect. Thus we have an extension $1 \to A \to N_0 \stackrel{\alpha}{\to} K \to 1$ for some $A \leq \boldsymbol{Q}/\boldsymbol{Z}$, and α is a covering. Hence by universality there is a commutative diagram

$$\begin{array}{ccccccccc} 1 & \to & Z & \to & L & \stackrel{\phi}{\to} & K & \to & 1 \\ & & \downarrow \gamma & & \downarrow & & \downarrow 1_K & & \\ 1 & \to & A & \to & N_0 & \stackrel{\alpha}{\to} & K & \to & 1 \end{array}$$

This leads to a homomorphism $\beta : Z \stackrel{\gamma}{\to} A \leq \boldsymbol{Q}/\boldsymbol{Z}$ and one checks that $\Psi(\beta) = \Gamma$. Thus Ψ is surjective. $\square$

We now prove the existence of universal coverings for perfect groups. Schur proved a more general result, with a weaker conclusion, for all finite groups, but we confine ourselves to the perfect case.

THEOREM 13.7 (ISSAI SCHUR, ∼1910). *Every finite perfect group K has a universal covering $L \to K$. A covering $L \to K$ is universal if and only if its kernel is isomorphic to $H^2(K, \boldsymbol{Q}/\boldsymbol{Z})$, which is a finite group.*

PROOF. Take a presentation

$$1 \to R \to F \to K$$

of G with F a finitely generated free group. Consider the diagram of injections:

$$\begin{array}{ccc} [F,F] & \to & F = R[F,F] \\ \uparrow & & \uparrow \\ D & \to & R \\ \uparrow & & \\ [R,F] & & \end{array} \tag{13B}$$

where $D = [F,F] \cap R$. Set $\overline{F} = F/[R,F]$, so that we have the diagram

$$\begin{array}{ccc} [\overline{F},\overline{F}] & \to & \overline{F} = \overline{R}[\overline{F},\overline{F}] \\ \uparrow & & \uparrow \\ \overline{D} & \to & \overline{R} \\ \uparrow & & \\ \overline{1} & & \end{array} \tag{13C}$$

Since $K \cong F/R$ is perfect we have $F = [F,F]R$. Thus $K \cong [F,F]/D$. Moreover since $D \leq R$, $[[F,F], D] \leq [F,R]$ so the diagram gives a homomorphism

$$\phi : [\overline{F},\overline{F}] \to K$$

with central kernel $\overline{D}$. By 13.2, $\overline{D}$ is therefore finite. Furthermore, $R/D \cong F/[F,F]$, a finitely generated free abelian group. Therefore $\overline{R}$ is a finitely generated abelian group, so its torsion subgroup $\overline{D}$ has a (free) complement $\overline{C}$ in $\overline{R}$. We have $\overline{C} \cap [\overline{F},\overline{F}] \leq \overline{C} \cap \overline{D} = 1$, and so $\overline{F} = \overline{C} \times [\overline{F},\overline{F}]$. Taking the commutator subgroup of both sides shows that $[\overline{F},\overline{F}]$ is perfect; thus ϕ is a covering.

A diagram chase, using the universal properties of free groups, now shows that ϕ is universal. Namely, given a covering $\psi : M \to K$ we may lift the obvious mapping $F \to K$ to a homomorphism $F \to M$. This homomorphism takes R into $\ker \psi$ and so takes $[F,R]$ to 1, so leads to a mapping $[\overline{F},\overline{F}] \to M$ which is the required lifting.

The assertion about universality and $H^2(K, \boldsymbol{Q}/\boldsymbol{Z})$ follows from the preceding lemma, and 13.2 implies the finiteness. □

In the non-perfect case most of the preceding result goes through. If K is any finite group, we define a representation group of K to be an extension

$$1 \to Z \to E \to K \to 1$$

such that $Z \leq [E,E] \cap Z(E)$ and such that $|Z| = |H^2(K, \boldsymbol{Q}/\boldsymbol{Z}$.

THEOREM 13.8 (SCHUR). *Let K be any finite group and repeat the construction of* (13.C). *Then*

1. $\overline{D} \cong H^2(K, \boldsymbol{Q}/\boldsymbol{Z}$.
2. $\overline{D}$ *is the torsion subgroup of the finitely generated abelian group* $\overline{R}$, *and for any complement* $\overline{C}$ *to* $\overline{D}$ *in* $\overline{R}$, *the group* $\overline{F}/\overline{C}$ *is a representation group of* K.
3. *For any covering* $\phi : L \to K$, *there is a representation group* $E \to K$ *which factors through* ϕ.

PROOF. The proof is practically the same. Again $\overline{D}$ is the torsion subgroup of the finitely generated group $\overline{R} \leq Z(\overline{F})$. So complements $\overline{C}$ exist. It is clear that $\overline{F}/\overline{C}$ then covers K, and the kernel is $\overline{D}$. This proves (2). For any covering $\phi : L \to K$, we obtain as a mapping $F \to L$ taking R into $\ker\phi$ and so annihilating $[F, R]$. So we get a homomorphism $\beta : \overline{F} \to L$. Now $\overline{R}$ maps into $\ker\phi$, but since $\ker\phi \leq [L, L]$, $\overline{D}$ maps onto $\ker\phi$. Therefore $\ker\beta$ contains a complement to $\overline{D}$ in $\overline{R}$, proving (3).

To prove (1) we repeat the proof of 13.6 but apply it to the central extension $1 \to \overline{R} \to \overline{F} \to K \to 1$. We obtain a homomorphism $\hat{\overline{R}} \to H^2(K, \boldsymbol{Q}/\boldsymbol{Z})$, whose kernel consists of homomorphisms annihilating $\overline{R} \cap [\overline{F}, \overline{F}] = \overline{D}$. In fact it is not hard to show that the kernel is precisely the full annihilator of $\overline{D}$: in (13A), if α annihilates $\overline{D}$, then in the lower extension, the kernel is disjoint from the commutator subgroup and is divisible, which forces the extension to split (since divisible abelian groups are injective). Again by injectivity, the restriction mapping $\hat{\overline{R}} \to \hat{\overline{D}}$ is surjective, and so we get an injection

$$j : \hat{\overline{D}} \to H^2(K, \boldsymbol{Q}/\boldsymbol{Z}).$$

We only have to show that it is surjective. But the same proof works: an element in H^2 corresponds to an extension, which arises from some mapping on F, and hence is in the image of j.

DEFINITION 13.9. The Schur multiplier $M(G)$ of a finite group G is the (dual of the) group $H^2(G, \boldsymbol{Q}/\boldsymbol{Z})$.

The Schur multipliers of all the finite simple groups have been calculated. Interestingly, they tend to be very small; they are all the direct product of at most two cyclic groups, and in most cases they are cyclic. For each infinite family of simple groups, such as the alternating groups A_n or the linear groups $PSL_n(q)$, there is a simple formula for the Schur multiplier which is valid for all but a finite number (no more than a half dozen) of groups, and all the exceptions are small. For these two families, for instance, we have

$$M(A_n) \cong \begin{cases} Z_2 \text{ if } n = 5 \text{ or } n \geq 8 \\ Z_6 \text{ if } n = 6 \text{ or } 7 \end{cases}$$

$$M(L_n(q)) \cong Z_{\gcd(n,q-1)} \text{ if } n \geq 2, \text{ with the exceptions}$$

$$M(L_2(4)) \cong M(L_3(2)) \cong Z_2$$

$$M(L_2(9)) \cong Z_6$$

$$M(L_3(4)) \cong Z_4 \times Z_{12}$$

$$M(L_4(2)) \cong Z_2$$

Thus the "canonical" universal covering of $L_n(q)$ is the standard mapping $SL_n(q) \to L_n(q)$.

Moreover, "sporadic" Schur multipliers are often connected with the existence of some sporadic simple group, or with some sporadic isomorphism. For example, $L_2(4) \cong L_2(5)$ (the simple group of order 60), and $M(L_2(5))$ is "canonically" Z_2. Likewise $L_3(2) \cong L_2(7)$, $L_2(9) \cong A_6$ and $L_4(2) \cong A_8$, which "explains" their Schur multipliers.

EXERCISES. Establish isomorphisms:[4]
1. $L_3(2) \cong L_2(7)$
2. $L_2(9) \cong A_6$
3. $L_4(2) \cong A_8$

14. More on Cohomology

Until further notice we fix a group G and a subgroup H of G.

If A is an abelian group on which G acts, then H acts on it as well. Restriction clearly gives a mapping

$$Z^i(G, A) \to Z^i(H, A),\ i = 1, 2$$

carrying $B^i(G, A)$ to $B^i(H, A)$, and thus the restriction mapping:

$$\rho_i : H^i(G, A) \to H^i(H, A).$$

A mapping can be defined in the opposite direction:

PROPOSITION 14.1. *There exists a "transfer" homomorphism*

$$\tau_i : H^i(H, A) \to H^i(G, A)$$

and the composite $\tau_i \circ \rho_i : H^i(G, A) \to H^i(G, A)$ *is just the power mapping* $\alpha \mapsto \alpha^{|G:H|}$.

To find this homomorphism we need a way of producing elements of H from elements of G, i.e. mappings $G \to H$.

For our fixed subgroup H we fix an ordering on the set of left cosets of H in G:

$$G = \cup_{i=1}^n G_i,\ n = |G : H|.$$

We then choose and fix a set of left coset representatives in G:

$$G_i = g_i H, \text{ so that } G = \cup_{i=1}^n g_i H.$$

[4]Hints. These all take a considerable amount of work. In (1), one approach is to show that both groups have subgroups of index 7, so both embed in A_7. (To get a subgroup of index 7 in $L_3(2)$, consider its action on the underlying 3-dimensional vector space. To get one in $L_2(7)$, find an embedding of $SL_2(3)$ in $SL_2(7)$ and use the normalizer of a subgroup of $L_2(7)$ isomorphic to $Z_2 \times Z_2$.) Show also that in both groups G a Sylow 7-normalizer S has order 21 and a Sylow 3-normalizer T is isomorphic to Σ_3, and S and T may be chosen so that if we put $R = S \cap T$, then $G = \langle S, T \rangle$ and R is a Sylow 3-subgroup of G. Then show that up to conjugacy in Σ_7, there is only one way to pick the pair (S, R) of subgroups of A_7, and given S and R there are (again up to conjugation in Σ_7) precisely two possibilities for T. However, for one of these two possibilities, $\langle S, T \rangle$ contains a 5-cycle so cannot have order 168. Therefore the embedded images in A_7 of the two simple groups of order 168 in must be conjugate in Σ_7, so they are isomorphic.

Then any element of G has a unique expression in the form $g_i h$. Now G acts on this coset space; we write

$$gg_iH = g_{g(i)}H$$

for any i and any $g \in G$. The mapping $i \mapsto g(i)$ is then a permutation of $\{1, \dots, n\}$, and $(gg')(i) = g(g'(i))$.

Now for any $g \in G$, there exist unique elements $h_1(g), \dots, h_n(g) \in H$ such that

$$gg_i = g_{g(i)}h_i(g). \tag{14A}$$

This equation will also be used in the form

$$g_{g(i)}^{-1} g = h_i(g) g_i^{-1}. \tag{14B}$$

Furthermore, the h_i are not quite multiplicative. If $g, x \in G$ we can compute xgg_i as both

$$(xg)g_i = g_{xg(i)}h_i(xg) \text{ and } x(gg_i) = xg_{g(i)}h_i(g) = g_{xg(i)}h_{g(i)}(x)h_i(g),$$

whence

$$h_i(xg) = h_{g(i)}(x)h_i(g). \tag{14C}$$

These mappings $h_1, \dots, h_n : G \to H$ are the basis of the transfer mapping. They are not canonical, but depend on the choice of coset representatives.

More precisely their order, like that of our coset decomposition, is arbitrary, and somewhat more seriously, we could have chosen different representatives $g_i' = g_i k_i,\ k_i \in H$. This would have led to the different functions

$$h_i'(g) = k_{g(i)}^{-1} h_i(g) k_i. \tag{14D}$$

We take a brief detour to construct the "transfer homomorphism for subgroups of finite index" (German: *Verlagerung*).

DEFINITION 14.2. Let G be a group and H a subgroup of finite index. We define

$$V = V_{G \to H} : G \to H/[H,H],\quad V(g) = [H,H]\prod_{i=1}^{n} h_i(g),$$

PROPOSITION 14.3. *V is a homomorphism depending only on G and H and not on the choice of representatives g_i.*

PROOF. Indeed since the image group is abelian and we are taking the product over all n, $(14C)$ shows that V is a homomorphism, and $(14D)$ shows that it is independent of the choice of representatives.

The transfer mapping V yields a homomorphism $G/[G,G] \to H/[H,H]$ which as we shall see below is useful for analyzing $G/[G,G]$ in terms of local subgroups of G.

Returning to cohomology, let $f \in Z^1(H, A)$. We define $tf \in Z^1(G, A)$ as follows:

$$(tf)(g) = \prod_{i=1}^{n} f(h_i(g))^{g_i^{-1}}.$$

LEMMA 14.3.
1. *t is a homomorphism $Z^1(H,A) \to Z^1(G,A)$.*
2. *t maps $B^1(H,A) \to B^1(G,A)$.*
3. *Up to multiplication by an element of $B^2(G,A)$, tf is independent of the choice of g_i.*

PROOF. Obviously t is multiplicative. We shall show that for any $g, x \in G$, and any i, we have

$$f(h_{g(i)}(x))^{g_{g(i)}^{-1} g} f(h_i(xg))^{-g_i^{-1}} f(h_i(g))^{g_i^{-1}} = 1.$$

This suffices for (1), for if we take the product of this equation over all i, and use the fact that $i \mapsto x(i)$ is a permutation of the indices, we obtain the 1-cocycle condition for f.

Using (14B) and (14C) we see that this equation is equivalent to

$$\Big[f(h_{g(i)}(x))^{h_i(g)} f(h_{g(i)}(x) h_i(g))^{-1} f(h_i(g))\Big]^{g_i^{-1}} = 1,$$

and so follows from the fact that $f \in Z^1(H,A)$. Thus (1) holds.

If there is $a \in A$ such that $f(g) = a^{-1} a^g$ for all $g \in H$, then for $g \in G$ we have

$$f(g) = \prod_i f(h_i(g))^{g_i^{-1}} = \prod_i a^{-g_i^{-1}} a^{h_i(g) g_i^{-1}} = \prod_i a^{-g_i^{-1}} a^{g_{g(i)}^{-1} g} = b^{-1} b^g,$$

where $b = \prod_i a^{g_i}$. This proves (2). The proof of (3) is left to the reader. □

By (1) and (2), t induces a homomorphism

$$\tau : H^1(H,A) \to H^1(G,A).$$

LEMMA 14.4. *$\tau \circ \rho$ is the $|G:H|$-power mapping on $H^1(G,A)$.*

PROOF. Let $f \in Z^1(G,A)$. We must show that $t(\rho(f))$ and $f^{|G:H|}$ are equal modulo $B^1(G,A)$. Now

$$t(\rho(f))(g) = \prod_i f(h_i(g))^{g_i^{-1}}.$$

But since $f \in Z^1(G,A)$, the following calculation is valid:

$$f(h_i(g))^{g_i^{-1}} f(h_i(g) g_i^{-1})^{-1} f(g_i^{-1}) = 1,$$

so in view of (14B) and the cocycle condition,

$$f(h_i(g))^{g_i^{-1}} = f(g_{g(i)}^{-1} g) f(g_i^{-1})^{-1} = f(g_{g(i)}^{-1})^g f(g) f(g_i^{-1})^{-1}.$$

Interchanging the last two factors and taking the product over all i, and setting $a = \prod_i f(g_i)^{-1}$, we get

$$t\rho f(g) = a^g a^{-1} f(g)^{|G:H|}.$$

□

This proves Proposition 14.1 in the case $i = 1$. The case $i = 2$ may be done by a similar, although somewhat lengthier, calculation. Here one defines

$$(tf)(x,y) = \prod_{i=1}^{n} f(h_{y(i)}(x), h_i(y))^{g_i^{-1}}, \ f \in Z^2(H,A),$$

and lifts this to a homomorphism $\tau : H^2(H,A) \to H^2(G,A)$. We leave most of the details to the reader. In the check that $\tau \circ \rho$ is the $|G:H|$-power mapping, use the fact that $f \in Z^2(G,A)$ and pull g_i^{-1} through to the front and out using the cocycle condition and $(14B,C)$. What remains will amount to a coboundary. Watch:

$$\begin{aligned}
&f(h_{y(i)}(x), h_i(y))^{g_i^{-1}} f(x,y)^{-1} \\
&= f(h_{y(i)}(x), h_i(y)g_i^{-1}) f(h_{y(i)}(x)h_i(y), g_i^{-1})^{-1} f(h_i(y), g_i^{-1}) f(x,y)^{-1} \\
&= f(h_{y(i)}(x), g_{y(i)}^{-1} y) f(h_i(xy), g_i^{-1})^{-1} f(h_i(y), g_i^{-1}) f(x,y)^{-1} \\
&= f(h_{y(i)}(x), g_{y(i)}^{-1})^y f(h_{y(i)}(x) g_{y(i)}^{-1}, y) \cdot \\
&\qquad f(g_{y(i)}^{-1}, y)^{-1} f(h_i(xy), g_i^{-1})^{-1} f(h_i(y), g_i^{-1}) f(x,y)^{-1} \\
&= f(h_{y(i)}(x), g_{y(i)}^{-1})^y [f(g_{xy(i)}^{-1} x, y) f(x,y)^{-1}] \cdot \\
&\qquad f(g_{y(i)}^{-1}, y)^{-1} f(h_i(xy), g_i^{-1})^{-1} f(h_i(y), g_i^{-1}) \\
&= f(h_{y(i)}(x), g_{y(i)}^{-1})^y [f(g_{xy(i)}^{-1}, xy) f(g_{xy(i)}^{-1}, x)^{-y}] \cdot \\
&\qquad f(g_{y(i)}^{-1}, y) f(h_i(xy), g_i^{-1})^{-1} f(h_i(y), g_i^{-1}) \\
&= f(h_{y(i)}(x), g_{y(i)}^{-1})^y f(h_i(xy), g_i^{-1})^{-1} f(h_i(y), g_i^{-1}) \cdot \\
&\qquad f(g_{y(i)}^{-1}, y)^{-1} f(g_{xy(i)}^{-1}, xy) f(g_{xy(i)}^{-1}, x)^{-y}.
\end{aligned}$$

When we take the product over all i and use the fact that $i \mapsto y(i)$ and $i \mapsto xy(i)$ are permutations of $\{1, \ldots, n\}$, we find that right side becomes the coboundary $(\delta b)(x,y)$, where $b(g) = \prod_{i=1}^{n} f(h_i(g), g_i^{-1}) f(g_i^{-1}, g)^{-1}$. This leads to $\tau \circ \rho(\overline{f}) = \overline{f}^{|G:H|}$, where $f \mapsto \overline{f} \in H^2(G,A)$.

COROLLARY 14.5. *If H contains a Sylow p-subgroup of G, then the restriction mappings $\rho : H^i(G,A) \to H^i(H,A)$ are injective on a Sylow p-subgroup of $H^i(G,A)$. If furthermore A has exponent a power of p, then ρ is injective.*

PROOF. $\tau\rho$ is injective on p-elements, so ρ is as well.

This leads to a generalization of the Schur-Zassenhaus Theorem:

THEOREM 14.6 (GASCHÜTZ). *Suppose that $N \lhd G$ and N is an abelian p-group. Let P be a Sylow p-subgroup of G. If P splits over N, then G splits over N.*

PROOF. The restriction mapping $H^2(G/N, N) \to H^2(P/N, N)$ is injective by the corollary. From the way cocycles are associated with extensions, it is clear that the cohomology classes in these cohomology groups corresponding to the groups G and P correspond to each other under the restriction mapping. Since one of these cohomology classes is trivial, so is the other.

THEOREM 14.7 (MASCHKE). *Suppose that G is a p'-group acting on an abelian p-group A. Then any G-invariant subgroup of A has a G-invariant complement.*

PROOF. Let N be a G-invariant subgroup of A. Apply the previous result to N and the semidirect product AG.

EXERCISE. If G is a p'-group and G acts on the abelian p-group A, then $A = [A, G] \times C_A(G)$.

15. Transfer, abelian quotients and fusion

Given a finite group G and subgroup H, we constructed in the previous section the "transfer homomorphism"[5]

$$V = V_{G\to H} : G \to H/[H,H], V(g) = \prod_i h_i(g)[H,H],$$

where the g_i are a set of left coset representatives for H in G, and the h_i are the functions $G \to H$ such that

$$gg_i = g_{g(i)}h_i(g) \text{ where } gg_iH = g_{g(i)}H. \tag{15A}$$

As observed in the previous section, this mapping is independent of the choice of coset representatives. Another important property is the invariance under automorphisms: if $\alpha \in \mathrm{Aut}(G)$, then

$$(V_{G\to H}(g))^\alpha = V_{G\to H^\alpha}(g^\alpha),$$

as follows directly from the definitions.

The corresponding mapping

$$\hat{V} : \hat{H} \to \hat{G}$$

on the dual groups $\hat{G} = \mathrm{Hom}(G, Q/Z)$ and $\hat{H} = \mathrm{Hom}(G, Q/Z)$ is given by

$$\hat{V}(f)(g) = f(\prod_i h_i(g))$$

and so it's the transfer mapping defined in the previous section

$$\hat{V} = \tau : H^1(H, Q/Z) \to H^1(G, Q/Z).$$

Important insights about the transfer come from both points of view.

We derive a useful formula for helping to compute $V_{G\to H}(g)$ for $g \in G$. With no more effort we can get a more general and very useful result, computing $V_{G\to H}|K$ for any subgroup K of G. This is the "Mackey formula." Applying the Mackey formula to $K = \langle g \rangle$ will yield a formula for $V(g)$.

Continuing the above notation, we consider subgroups H and K of G and the action of G on the left coset space of H in G given by (15.A). The K-orbits are in one-to-one correspondence with the K, H-double cosets:

$$G = \cup_{j\in J} K g_j H, \ Kg_jH = \cup_{i\in I_j} g_i H$$

[5]More generally, given any $K \lhd H$ with H/K abelian, so that $[H,H] \le K$, we may follow $V_{G\to H}$ by the projection onto H/K to obtain the homomorphism $V_{G\to H/K}$ defined by $V_{G\to H/K}(g) = \prod_i h_i(g)K$.

where J is a subset of $\{1,\dots,n\}$, and the I_j form a partition of $\{1,\dots,n\}$, with $j \in I_j$ for each j. The set J is not uniquely determined, but the partition is.

The Mackey lemma is based on the possibility of choosing the g_i nicely with respect to K.

LEMMA 15.1. *For each double coset KgH, choose a set C_g of left coset representatives of $K^g \cap H$ in K^g. Then*

$$\cup_{j\in J} g_j C_{g_j}$$

is a disjoint union, and is a set of left coset representatives of H in G.

PROOF. Suppose that $y \in C_{g_j}$, $z \in C_{g_{j'}}$ and $g_j y H = g_{j'} z H$. By choice $g_j y \in g_j K^{g_j} = K g_j$, so the left side lies in $K g_j H$. A similar argument applies to the right side, and so since double cosets are disjoint or equal, $j = j'$. Then $z \in y \in H \cap K^{g_j}$, so by choice, $z = y$. Thus the union is disjoint and is part of a set of left coset representatives. On the other hand any coset gH lies in $K g_j H$ for some j, so has the form $k g_j H = g_j k^{g_j} H$ for some $k \in K$. Then $k g_j H = g_j c H$ for some $c \in C_{g_j}$, completing the proof. □

PROPOSITION 15.2 (MACKEY LEMMA). *Let H and K be subgroups of the finite group G, and continue the above notation. Then for any $k \in K$,*

$$V_{G\to H}(k) = \prod_{j\in J} (V_{K\to K\cap H^{g_j^{-1}}}(k))^{g_j} [H,H]$$

The formula can be rewritten in slightly inaccurate language as the equality of an "up-down" mapping with a product of "down-up" mappings. Let ι_g be conjugation by g. The formula is then equivalent to

$$V_{G\to H}|K = \prod_{j\in J} \pi_{[H,H]} \circ \iota_{g_j} \circ V_{K\cap H^{g_j^{-1}}},$$

or inaccurately as

$$V_{G\to H}|K = \prod_{j\in J} (V_{K\cap H^{g_j^{-1}}})^{g_j}.$$

Dually we get the equality of mappings $H^1(H, Q/Z) \to H^1(K, Q/Z)$:

$$\rho_{G\to K} \circ \tau_{H\to G} = \prod_{j\in J} \tau_{K\cap H^{g_j}\to K} \circ \rho_{H^{g_j}\to H^{g_j}\cap K} \circ \iota_{g_j}.$$

The Mackey lemma in representation theory makes an entirely analogous assertion. Moreover, the last statement happens also to be true in each dimension, i.e. on each $H^n(H, Q/Z)$.

PROOF. Choose coset representatives as in the preceding lemma. Then

$$V_{K^{g_j}\to K^{g_j}\cap H}(y) = \prod_i h_{ij}(y),$$

where the product is over all $i \in I_j$, and I_j indexes the set C_{g_j} of coset representatives. Thus for any $x_i \in C_{g_j}$ and $y \in K^{g_j}$, $y x_i = x_{k(i)} h_{ij}(y)$. Consequently for any

$k \in K$, $kg_jx_i = g_jk^{g_j}x_i = g_jx_{i'}h_{ij}(k^{g_j})$. Thus

$$V_{G\to H}(k) = \prod_i h_i(k)[H,H] = \prod_{j\in J}\prod_{i\in I_j} h_{ij}(k^{g_j})[H,H]$$
$$= \prod_{j\in J} V_{K^{g_j}\to K^{g_j}\cap H}(k^{g_j})[H,H].$$

This immediately yields the desired formula, since conjugation by g_j is an automorphism of G.

COROLLARY 15.3. *Let $H \le G$ and $g \in G$. Then*

$$V_{G\to H}(g) = \prod_j g_j g^{m_j} g_j^{-1}[H,H],$$

where the product is indexed by the $\langle g\rangle, H$ double cosets, and each such double coset $\langle g\rangle g_jH$ contributes the term $g_jg^{m_j}g_j^{-1}$, where m_j is the least positive integer such that this conjugate lies in H, and m_j is also the number of cosets of H in this double coset. Thus $\sum_j m_j = |G:H|$.

PROOF. Notice that if $G = \langle a\rangle$ is cyclic and $H \le G$ with $|G:H| = n$, then $h_i(x) = x$ for all i and all $x \in G$ and so $V_{G\to H}(x) = x^n$. Now applying the previous result with $K = \langle g\rangle$ yields this corollary.

Thus $V_{G\to H}(g)$ is the product of elements of $H/[H,H]$ which are the images of G-conjugates of powers of g. The study of such conjugates is known as the fusion of g in H with respect to G, and its bearing on abelian quotients of G is now clear in a general way. We make it precise as follows.

THEOREM 15.4 (FOCAL SUBGROUP THEOREM OF D. G. HIGMAN). *Let P be a Sylow p-subgroup of G. Define the focal subgroup by*

$$F(G,P) = \langle a^{-1}a^g \mid a \in P, g \in G, a^g \in P\rangle.$$

Then

$$F(G,P) = [G,G]\cap P = \ker V_{G\to P}.$$

This implies in particular that the transfer digs out every abelian p-quotient of G, since the largest abelian p-quotient of G is the same as that of $G/[G,G]$, which is isomorphic to a Sylow p-subgroup of $G/[G,G]$, namely $P[G,G]/[G,G] \cong P/P\cap[G,G]$.

PROOF. Clearly

$$F(G,P) \le [G,G]\cap P \le \ker V_{G\to P}|P$$

since $a^{-1}a^g = [a,g]$ and V, mapping G to an abelian group, has $[G,G]$ in its kernel. Finally let $g \in \ker V_{G\to P}$ with $g \in P$. Then $1 = V_{G\to P}(g) = \prod_j g_jg^{m_j}g_j^{-1}$ as in the preceding corollary. Modulo $F(G,P)$, we have

$$g^{|G:P|} = g^{\sum_j g_j} = \prod_j g^{m_j} = \prod_j g_jg^{m_j}g_j^{-1} = 1.$$

Since P is a Sylow p-subgroup and $g \in P$, g is a p-element but p does not divide $|G:P|$. Therefore $g = 1$ modulo $F(G,P)$, so $g \in F(G,P)$, completing the proof.

A key part of the classification of simple groups is the "local control of transfer", by which is meant (for a given prime p) the existence of one or more p-local subgroups N_i of G such that the largest abelian p-quotient of G is computable from those of the groups N_i. In general local control occurs, because of the focal subgroup theorem and the Alperin Fusion Theorem below. But there are several striking results about situations in which a single p-local subgroup "controls transfer". We give two such results, one about 100 years old and due to Burnside, and the other relatively new (late 70's) and due to Yoshida. Although Yoshida's result is much more striking, the proofs are not that different.

DEFINITION 15.5. Let P be a Sylow p-subgroup of G, and N a subgroup of G containing P. Then N is said to control p-transfer in G if and only if the largest abelian p-factor groups of G and N are isomorphic.

Here are a few results on local control of transfer.

THEOREM 15.6 (BURNSIDE). *Suppose that $P \in Syl_p(G)$ and P is abelian. Let $N = N_G(P)$. Then N controls p-transfer in G. Moreover, the largest abelian p-quotient of N (and therefore also the largest abelian p-quotient of G) is isomorphic to*

$$P \cap Z(N) \cong P/[N,P].$$

PROOF. First we argue, using the commutativity of P, that

(15.1) if $x \in P$, $g \in G$ and $x^g \in P$, then x and x^g are N-conjugate.

Indeed x^g lies in the intersection $P \cap P^g$ of the abelian groups P and P^g, so $C_G(x^g)$ contains both P and P^g. Then P and P^g are Sylow subgroups of $C_G(x^g)$ so there is $h \in C_G(x^g)$ such that $P^{gh} = P$. Then $x^{gh} = (x^g)^h = x^g$ and $gh \in N_G(P)$, proving (15.1).

From the displayed statement and the Focal Subgroup Theorem, N controls p-transfer. Indeed the largest abelian p-quotient of G is isomorphic to $P/F(G,P)$, while that of N is isomorphic to $P/F(N,P)$. But $F(G,P)$ is generated by all elements $x^{-1}x^g$ such that $x \in P$, $g \in G$ and $x^g \in P$; and $F(N,P)$ has a similar characterization. Hence by (15.1), $F(G,P) = F(N,P)$, so N controls p-transfer. Since $P \lhd N$, furthermore, it is clear that $F(N,P)$ is generated by all $x^{-1}x^g$ with $x \in P$ and $g \in N$, so $F(N,P) = [N,P]$.

Finally if we consider the action of the p'-group N/P on the abelian p-group P, we know that $P = [N,P] \times C_P(N)$. Hence $P/[N,P] \cong C_N(P) = P \cap Z(N)$. $\square$

This is a fairly strong result. For instance, it implies:

COROLLARY 15.7. *In a nonabelian simple group G, if p is the smallest prime divisor of $|G|$, then Sylow p-subgroups of G cannot be cyclic.*

PROOF. The proof uses the fact that

$$|\mathrm{Aut}(Z_n)| = \phi(n),$$

where ϕ is the Euler ϕ-function. (Proof: $\phi(n)$ is the number of generators of Z_n, and if x is a given generator then for every generator y there exists a unique automorphism taking x to y.)

Now if a Sylow p-subgroup P were cyclic, say of order p^n, then $N_G(P)$ would act on P, with P itself acting trivially, so $N_G(P)/P$ would be a p'-group acting

on P. The image of $N_G(P)/P$ in $\mathrm{Aut}(P)$ would have order dividing $|N_G(P) : P|$ and divisible only by primes greater than p; but $|\mathrm{Aut}(P)| = \phi(p^n) = p^{n-1}(p-1)$ is divisible only by primes less than p. So this image must be trivial. This means that $N_G(P)$ acts trivially on P, so $P = P \cap Z(N_G(P))$, so the largest abelian p-factor group of G is isomorphic to P. By assumption P is nontrivial, so G is not simple, contradiction.

Burnside's theorem also makes it easy to rule out certain numbers as orders of simple groups, for instance 924 (use $p = 11$).

Yoshida found a strong generalization of Burnside's Theorem. To state it we introduce wreath products. If B and W are groups, and W acts on a set Ω, one may define the wreath product $B \wr_\Omega W$ as the semidirect product

$$B \wr_\Omega W = B^* W,$$

where B^* is the set of all Ω-tuples of elements of B and W acts on B^* by permuting coordinates the way it permutes Ω. Formally we may take B^* to be the group of all functions $f : \Omega \to B$, with pointwise multiplication, and for $f \in B^*$ and $w \in W$ we put $f^w(\alpha) = f(\alpha^{w^{-1}})$, for all $\alpha \in \Omega$.

In the notation

$$B \wr W$$

it is to be understood implicitly that $\Omega = W$ and W acts regularly on Ω (i.e. by right multiplication).

Such wreath products, and closely related structures, pop up commonly as Sylow subgroups of alternating or classical groups (but not for the prime which is the characteristic!). For example, Sylow p-subgroups of S_{p^2} have order p^{p+1}, and one of them, call it P, is generated by the p p-cycles $x_1, x_2, \cdots, x_p$, where

$$x_i = ((i-1)p+1\,(i-1)p+2 \cdots ip)$$

and the additional element

$$w = (1\,p+1\,2p+1 \cdots (p-1)p+1) \cdots (p\,2p \cdots p^2)$$

of order p. Then there is an isomorphism

$$P \cong Z_p \wr Z_p$$

carrying $\langle x_1, \ldots, x_p \rangle$ to the base group B^* and $\langle w \rangle$ to W.

THEOREM 15.8 (YOSHIDA). *Suppose that $P \in Syl_p(G)$ and that P has no quotient isomorphic to $Z_p \wr Z_p$. Then $N_G(P)$ controls p-transfer in G.*

PROOF. Set $N = N_G(P)$ and consider the transfer mapping

$$V = V_{G \to N} : G \to \overline{N} = N/[N, N].$$

The group $\overline{P}$ is a Sylow p-subgroup of $\overline{N}$, and we let $\overline{Q}$ be a Sylow p-subgroup of the image of V. Thus G has an abelian p-quotient isomorphic to $\overline{Q}$. What we need to prove is that

$$\overline{Q} = \overline{P}. \tag{15.2}$$

For then G will have an abelian p-quotient isomorphic to $\overline{P}$, and so $|G/[G,G]|_p \geq |N/[N,N]|_p$. This means that $|P \cap [G,G]| \leq |P \cap [N,N]|$. But $[N,N] \leq [G,G]$ so $P \cap [N,N] = P \cap [G,G]$, as required. Proceeding by contradiction we assume that

$$\overline{Q} < \overline{P},$$

and let Q be the full preimage of $\overline{Q}$ in $\overline{P}$, so that $Q < P$. Choose a maximal subgroup R of P containing Q, and choose $x \in P - R$ of minimal order. We compute $V(x)$ using the Mackey lemma with $K = P$, and obtain

$$V(x) = \prod_i V_{P \to P \cap N^{g_i^{-1}}}(x)^{g_i}[N,N]. \tag{15.3}$$

As the double coset representative corresponding to the trivial double coset N we are free to choose $g_1 = 1$, and for that term we have

$$V_{P \to P}(x)[N,N] = x[N,N].$$

Since $x \notin R$, certainly $\overline{x} \notin \overline{Q}$ so this contribution to $V(x)$ lies outside $\overline{Q}$. However, by definition $V(x) \in \overline{Q}$, and so another term from (15.3) must lie outside $\overline{Q}$. If we call the corresponding g_i just g, for simplicity, we get

$$V_{P \to P \cap N^{g^{-1}}}(x)^g [N,N] \notin \overline{Q}, \text{ and } g \notin N. \tag{15.4}$$

Set $M = P \cap N^{g^{-1}}$. Then $M < P$, for otherwise $P \leq N^{g^{-1}}$, whence $P^g \leq N$, so $P^g = P$ (as $P \lhd N$) and so $g \in N_G(P) = N$, contrary to our choice of g. Because of (15.4), $V_{P \to M}(x)$ does not lie in the image of $Q \cap M \pmod{[M,M]}$. Therefore it does not lie in the image of $R \cap M$. (Recall that R was a maximal subgroup of P containing Q.) It follows that

$$V_{P \to M/M \cap R}(x) \neq 1. \tag{15.5}$$

Since $|P : R| = p$, the nontrivial quotient $M/M \cap R$ must also have order p. The proof is completed by the following lemma, which contradicts the assumption about $Z_p \wr Z_p$:

LEMMA 15.9 (YOSHIDA). *Let P be a p-group, M a proper subgroup of P, and R a maximal subgroup of P. Let x be an element of $P - R$ of minimal order and suppose that* (15.5) *holds. Then P has a quotient isomorphic to $Z_p \wr Z_p$.*

PROOF. Since (15.5) holds, $M \not\leq R$ and so $P = RM$. Let A be a maximal subgroup of P containing M. Thus $P = RA$, so $A/A \cap R$ has order p. We consider

$$\overline{y} = V_{P \to A/\Phi(A)}(x).$$

If $\overline{y} \in A \cap R/\Phi(A)$, then $V_{P \to A}(x)$ would be the image mod $[A,A]$ of an element $z \in R$. Since $R \lhd P$, all conjugates of powers of z lie in R, so by 15.3, $V_{A \to M}(V_{P \to A}(x))$ would be the image of an element of R. But this composite is $V_{P \to M}(x)$ (see the exercise below) and so $V_{P \to M/M \cap R}$ would be trivial. This contradicts (15.5). Consequently $\overline{y} \notin A \cap R/\Phi(A)$.

We argue that $x \in A$. If not, then $P = \langle x \rangle A$ and there is a unique $\langle x \rangle$, A-double coset in P. By Corollary 15.3, $V_{P \to A}(x)$ is the image in $A/[A,A]$ of a P-conjugate of x^p lying in A. But x^p has lower order than x, and there are no elements of lower order in $P - R$ by assumption, so $V_{P \to A}(x) \in (A \cap R)[A,A]/[A,A]$. Therefore $\overline{y} \in A \cap R/\Phi(A)$, contradicting the previous paragraph. Thus, $x \in A$.

Now if we choose any $g \in P - A$, the elements g^i, $0 \leq i < p$, are a set of A, $\langle x \rangle$-double cosets and so by 15.3,

$$1 \neq V_{G \to A/\Phi(A)}(x) = \prod_{i=0}^{p-1} x^{g^i} \Phi(A).$$

Let γ be the linear transformation on $A/\Phi(A)$ induced by conjugation by g. Let $\hat{x}$ be the image in $A/\Phi(A)$ of x. The element $\hat{x}^{1+\gamma+\gamma^2+\cdots+\gamma^{p-1}}$ is thus nontrivial, which means that $\hat{x}$ generates a γ-direct summand of $A/\Phi(A)$ on which γ induces a Jordan block of (maximal) size p. Let B be a γ-invariant complement to this summand. Then P/B is an extension of a p-dimensional vector space by an element of order p acting as a single Jordan block. Our calculations of cohomology for cyclic groups show that $H^2(\langle \gamma \rangle, A/B)$ is trivial, so P/B splits over A/B and hence $P/B \cong Z_p \wr Z_p$. □

Exercise. Make sense of the following assertion, and then prove it:

$$\text{If } H \leq K \leq G, \text{ then } V_{G \to H} = V_{K \to H} \circ V_{G \to K}.$$

A. Some Simple Groups

It is well-known that the alternating groups A_n are simple, $n \geq 5$. To furnish some more easy examples of simple groups, we prove the simplicity of the special linear groups here. For any field F and integer $n \geq 1$, $GL_n(F)$ is the group of nonsingular linear transformations on an n-dimensional vector space over F; choice of an ordered basis leads to a coordinatization of $GL_n(F)$, unique up to inner automorphism. The kernel of the determinant mapping $GL_n(F) \to F^\times$ is $SL_n(F)$, and we write Z for the group of scalar multiples of the identity. Thus $SL_n(F) \lhd GL_n(F)$ with $GL_n(F)/SL_n(F) \cong F^\times$, and $Z \leq Z(GL_n(F))$ with $Z \cong F^\times$. It is pretty clear (check it!) that $Z = Z(GL_n(F))$. Set $Z_0 = Z \cap SL_n(F)$, and

$$PSL_n(F) = L_n(F) = SL_n(F)/Z_0 \text{ and } PGL_n(F) = GL_n(F)/Z,$$

so that $L_n(F) \lhd PGL_n(F)$. The case we are interested in here is $|F| = q$, in which case we write $L_n(q)$, $GL_n(q)$, etc. In that case

$$|GL_n(q) : SL_n(q)| = |Z| = q - 1, \qquad |Z_0| = \gcd(n, q-1) = |PGL_n(q) : PSL_n(q)|.$$

(Check it.)

THEOREM A.1. *For any $n \geq 2$ and any prime power q, $L_n(q)$ is nonabelian and simple, with the exceptions*

$$L_2(2) \cong \Sigma_3 \text{ and } L_2(3) \cong A_4.$$

There are many proofs of this; we give one of the most direct, which also generalizes to the other "classical groups", with some difficulty. I believe that it is due to Iwasawa. It depends on the notion of a primitive group action.

DEFINITION A.2. Let G act on the set Ω. Assume that the action is transitive. A block, or set of imprimitivity, is a nonempty subset $\Psi \subseteq \Omega$ such that for each $g \in G$,

$$\Psi^g \cap \Psi = \emptyset \text{ or } \Psi.$$

The action is imprimitive if and only if there exists a block other than Ω and a singleton $\{\alpha\}$, $\alpha \in \Omega$. The action is primitive if and only if it is not imprimitive.

LEMMA A.3. *Let G act transitively on Ω and let $\alpha \in \Omega$. Then the action is primitive if and only if the isotropy subgroup G_α is a maximal subgroup of G. Furthermore, if G acts doubly transitively on Ω (i.e., transitively on the set of ordered pairs of distinct elements of Ω), then it acts primitively.*

PROOF. If there exists a nontrivial set Ψ of imprimitivity, let $H = \{g \in G \mid \Psi^g = \Psi\}$. Clearly H is a subgroup of G and whenever $\alpha^g \in \Psi$, then $g \in H$. In particular since G is transitive on Ω, H is transitive on Ψ. Moreover $G_\alpha \leq H$. Then $1 < |\Psi| = |H : G_\alpha| < |\Omega| = |G : G_\alpha|$, and so G_α is not maximal. Conversely, if there is a subgroup H such that $G_\alpha < H < G$, then $\Psi = \alpha^H$ is easily checked to be a set of imprimitivity, nontrivial because it has size $|H : G_\alpha|$.

Finally suppose that G acts doubly transitively. Let Ψ be a block, containing $\alpha \in \Omega$, say. Suppose that $|\Psi| > 1$ and choose $\beta \in \Psi - \{\alpha\}$. Then for any $\gamma \in \Omega$ with $\gamma \neq \alpha$, there is $g \in G$ such that $\alpha^g = \alpha$ and $\beta^g = \gamma$. The first implies that g stabilizes Ψ, and so the second implies that $\gamma \in \Psi$. Therefore $\Psi = \Omega$ and G acts primitively. □

LEMMA A.4. *Suppose that G acts primitively on Ω and $N \lhd G$. Then the action of N on Ω is either trivial or transitive.*

PROOF. For any $\alpha \in \Omega$, the N-orbit α^N satisfies

$$(\alpha^N)^g = \alpha^{Ng} = \alpha^{gN} = (\alpha^g)^N.$$

Thus the images of α^N under elements of g are the N-orbits, which partition Ω. Therefore α^N is a set of imprimitivity, so equals $\{\alpha\}$ or Ω. Since α is arbitrary the lemma follows. □

THEOREM A.5 (IWASAWA). *Let G act on Ω, with kernel Z. Let $\alpha \in \Omega$. Let K be a normal subgroup of G_α and assume that the following conditions hold:*

1. *G acts primitively on Ω.*
2. *K is solvable.*
3. *The G-conjugates of K generate G.*
4. *$G = [G, G]$.*

Then G/Z is nonabelian simple.

PROOF. Without loss we may pass to G/Z; the hypotheses all carry over and so inductively we may assume that $Z = 1$, so that G acts faithfully on Ω.

Let $N \lhd G$ with $N \neq 1$. By the preceding lemma, N acts transitively on Ω. Therefore $G = NG_\alpha$. Since $K \lhd G_\alpha$ it follows that $NK \lhd G$. Thus all the G-conjugates of K lie in NK, so $NK = G$ by assumption. But then $G/N \cong K/K \cap N$ is solvable. Also G/N is perfect by assumption, so it is trivial, i.e., $N = G$. Thus G is simple. Since $G = [G, G]$, G is nonabelian. □

We apply this to $G = SL_n(q)$, with $n \geq 2$. We take a matrix representation of G as $n \times n$ matrices over the field of q elements, and consider the action of G on the set Ω of 1-dimensional subspaces of the underlying n-dimensional space of column vectors.

LEMMA A.6. *G acts doubly transitively on Ω. The kernel of the action of G on Ω is Z_0.*

PROOF. Distinct elements of Ω correspond to subspaces generated by two linearly independent elements. Clearly then $GL_n(q)$ acts doubly transitively. Moreover for a given $\alpha, \beta \in \Omega$ and a given $c \in F^\times$, there is $g \in GL_n(q)$ fixing α and β and such that $\det g = c$ (indeed scaling α by c and acting trivially on a complement including β). Then given distinct γ and δ, there is $h \in GL_n(q)$ taking (α, β) to (γ, δ); there also is $g \in GL_n(q)$ such that $\det g = \det h$ and g fixes (α, β); and then $g^{-1}h \in G$ carries (α, β) to (γ, δ), proving double transitivity.

The final statement is left to the reader. □

The stabilizer G_α of the 1-dimensional subspace $\alpha = W$ of the underlying vector space V has a normal subgroup K defined by

$$K = \{g \in G_\alpha \mid g \text{ induces the identity mapping on } V/W\}.$$

If we take $\alpha = \langle [1, 0, \cdots, 0] \rangle$ then K is the group of all matrices which are equal to the identity matrix in all positions except the first column (off the diagonal). This easily implies that

LEMMA A.7. *K is abelian.* □

LEMMA A.8. *The G-conjugates of K generate G.*

PROOF. The G-conjugates of elements of K include all matrices $x_{i,j}(c) = I + ce_{i,j}$, $i \neq j$, differing from the identity matrix in exactly one position. We show that every element of G is a word in such matrices. Pre- and post-multiplying by $x_{i,j}(c)$ effects row and column operations "of the third kind" on a matrix, and since $x_{i,j}(c)^{-1} = x_{i,j}(-c)$, what we need to show is that by a sequence of such operations, an arbitrary matrix g of determinant 1 may be reduced to the identity matrix. Notice that these row and column operations do not affect the determinant.

But it is quite clear that g can be reduced to a "monomial" matrix h, in which every row and column has just one nonzero entry. It then suffices to show that any monomial matrix can be reduced to another monomial matrix which has 1 somewhere on the diagonal; for then the desired result follows inductively on the size of the matrix (and by the observation that at the last stage, the very last entry must also be 1 because $\det g = 1$.

This is easiest if h is not diagonal, by using three row operations as illustrated by the following simple case:

$$\begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix} \to \begin{bmatrix} 1 & a \\ b & 0 \end{bmatrix} \to \begin{bmatrix} 1 & a \\ 0 & -ab \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & -ab \end{bmatrix}.$$

If h is diagonal, a similar sequence of three row operations transforms it to a non-diagonal monomial matrix, reducing us to the previous case. □

We have now verified the first three conditions in Theorem A.5. To complete the proof we use the third condition and the following lemma to verify the fourth condition:

LEMMA A.9. *Unless $n = 2$ and $q \leq 3$, every element of K is a commutator of elements of G.*

PROOF. All nonidentity elements of K are conjugate in $GL_n(q)$, and since $G \triangleleft GL_n(q)$ it is enough to show that one such element is a commutator. Letting $h_1(c) = \mathrm{diag}(c, c^{-1}, 1, \cdots)$ we find that

$$[x_{2,1}(1), h_1(c)] = x_{2,1}(c^2 - 1).$$

If $q > 3$ then we may choose c so that $c^2 \neq 1$, and are done. Note also that if $n \geq 3$, we have

$$[x_{1,2}(1), x_{2,3}(1)] = x_{1,3}(1),$$

which is conjugate to an element of K, as desired. This completes the proof.

But whenever $K \leq [G, G]$, then since G is generated by conjugates of K, it follows that $G = [G, G]$. This completes the proof of Theorem A.2 except for the assertions about $L_2(2)$ and $L_2(3)$. But these groups have order 6 and 12, respectively, and embed in Σ_Ω, with $|\Omega| = 3$ or 4, respectively, determining their isomorphism type.

EXERCISES.. Suppose that G acts transitively on Ω, and consider the action of G on $\Omega \times \Omega$; each orbit of G on $\Omega \times \Omega$ other than the "diagonal" orbit $\{(\alpha, \alpha) \mid \alpha \in \Omega\}$ gives rise to a "directed graph" structure on Ω. Show that G acts primitively if

and only if for each "non-diagonal" orbit $\mathcal{O}$ of G on $\Omega \times \Omega$, the corresponding graph is "connected" in the sense that for any distinct $\alpha, \beta \in \Omega$, there is a chain $\alpha = \alpha_0, \alpha_1, \ldots, \alpha_n = \beta$ such that either $(\alpha_{i-1}, \alpha_i) \in \mathcal{O}$ or $(\alpha_i, \alpha_{i-1}) \in \mathcal{O}$ for each i.

Use this to prove that $G = SL_n(q)$ acts primitively not only on the set of 1-dimensional subspaces of the underlying vector space, but also on the set of all subspaces of any particular proper dimension.

Show that each of these primitive actions leads to a simplicity proof via Theorem A.5.

B. The Classical Groups

Besides the linear groups $PSL_n(q)$ there are three other families of "classical" simple matrix groups. Each of these arises from a finite-dimensional vector space V over a field $\boldsymbol{F}$, with certain extra "geometry" Γ on V. In this situation we can first form the group $I = \mathrm{Isom}(V, \Gamma)$ of all $\boldsymbol{F}$-linear transformations of V preserving Γ; normally this yields a simple group G just as the general linear group $GL(V)$ yields $PSL(V)$, namely $G = [I, I]/Z \cap [I, I]$ where Z is the group of scalar transformations of V. Indeed the linear groups $PSL(V)$ fall under this rubric, with the geometry Γ being empty, so that $I = GL(V)$. In the other examples leading to finite simple groups, the geometries are also highly homogeneous, which is to say that any two isometric subspaces of V are equivalent under an element of I; this fundamental result is known as "Witt's Lemma". However, in the examples other than $PSL(V)$, the geometries are also "nondegenerate".

The extra structure Γ always[6] includes a nondegenerate biadditive form $(\cdot, \cdot)$ on V, such that $(v, w) = 0$ if and only if $(w, v) = 0$. When this equation holds we say that $v \perp w$. Thus perpendicularity is always a symmetric relation. It will also always be the case that $v \perp w$ if and only if $cv \perp w$ for all $c \in \boldsymbol{F}$. For any subset U of V, we define

$$U^\perp = \{v \in V \mid v \perp w \ \forall w \in W\}.$$

Our conditions imply that $U^\perp$ is always a subspace of V. By definition, nondegeneracy of the form means

$$\mathrm{rad}\, V = V^\perp = 0.$$

For any subspace W of V, the geometry Γ will restrict in an obvious manner to give a geometry on W. However, there is no reason why such subspaces W need be nondegenerate; that is, we shall consider subspaces W such that

$$\mathrm{rad}\, W = W \cap W^\perp \text{ may not be } 0.$$

More specifically there are three kinds of geometry:

(a) *Symplectic.* Here the nondegenerate form is assumed to be $\boldsymbol{F}$-bilinear and alternating, that is,

$$(v, v) = 0 \ \forall v \in V.$$

Notice that for a symplectic form, $(v, w) = -(w, v)$, as can be seen from expanding $0 = (v + w, v + w)$. Conversely, in any characteristic but 2, the alternating condition $(v, v) = 0$ is equivalent to $(v, w) = -(w, v)$. But in

[6] In the case of odd-dimensional orthogonal groups in characteristic 2, the form is actually degenerate. But for finite groups anyway, we shall ignore these groups since it turns out that they appear elsewhere in our list, as symplectic groups.

characteristic 2, the condition $(v, v) = 0$ is stronger, and that is the one which is imposed.

(b) *Unitary*. Here $\boldsymbol{F}$ is assumed to have an automorphism σ of order 2 (for finite $\boldsymbol{F}$ this means that the cardinality of $\boldsymbol{F}$ is a square, and then σ is uniquely determined). We write $c^\sigma = \overline{c}$ for $c \in \boldsymbol{F}$. In the unitary case the nondegenerate form is assumed to be sesquilinear and hermitian, which means that

$$(v, w) = \overline{(w, v)}\ \forall v, w \in V, \text{ and}$$
$$\text{For fixed } v \in V,\ (v, w) \text{ is } \boldsymbol{F}\text{-linear in } w.$$

Consequently for fixed $w \in V$, (v, w) is conjugate linear in v, that is, additive in v, and $(cv, w) = \overline{c}(v, w)$.

(c) *Orthogonal*. Here a finer structure than a bilinear form is imposed: that of a quadratic form q on V. By definition, this is a function $q : V \to \boldsymbol{F}$ such that

(1) $q(cv) = c^2 q(v)$ for all $v \in V$, $c \in \boldsymbol{F}$; and
(2) The form $(v, w) = q(v + w) - q(v) - q(w)$ is bilinear.

It is immediate from (2) that the associated bilinear form (v, w) is symmetric, that is, $(v, w) = (w, v)$ for all $v, w \in V$. We call q nondegenerate if and only if the associated bilinear form (v, w) is nondegenerate. Notice that in characteristic 2, we have $(v, v) = q(2v) - 2q(v) = 0$, so the associated bilinear form is alternating.

There is a close connection between quadratic forms and symmetric bilinear forms. Namely if $(\cdot, \cdot)$ is a symmetric bilinear form, then we may set $q(v) = (v, v)$. Provided the characteristic of $\boldsymbol{F}$ is not 2, we see that q is a quadratic form whose associated bilinear form is $2(\cdot, \cdot)$. Consequently except in characteristic 2, the theories of nondegenerate quadratic forms and nondegenerate symmetric bilinear forms are equivalent, and there is a common prejudice in favor of the latter. But in characteristic 2, the theories differ, and quadratic forms have no substitute. As noted earlier, we shall only consider the case that $\dim V$ is even when considering orthogonal groups in characteristic 2.

In all three cases, the bilinear form $(\cdot, \cdot)$ sets up an identification of V with its dual space V^*, via the mapping

$$\iota : V \to V^*, \quad v \mapsto (v, \cdot).$$

The mapping ι is $\boldsymbol{F}$-linear except in the unitary case, in which case it is $\boldsymbol{F}$-"semilinear", that is $\iota(cv) = \overline{c}\iota(v)$. But in any case nondegeneracy implies that ι is injective, whence by a dimension argument it is surjective.

Because of this, and from familiar facts about the pairing between V and V^*, we see that for any subspace W of V,

$$\dim W^\perp + \dim W = \dim V.$$

However, as already observed, the two subspaces on the left can intersect nontrivially, so this equation does not come from a direct sum decomposition. However,

$$V = W \oplus W^\perp \text{ if } W \text{ is nondegenerate},$$

for then $W \cap W^\perp = 0$. We shall write

$$V = W \perp X$$

to mean that $V = W \oplus X$ and $X \subseteq W^\perp$. Of course this condition is then equivalent to $X = W^\perp$ and to $W = X^\perp$, and such a decomposition is only possible if W is nondegenerate, in which case X is too.

We shall speak of V, equipped with a nondegenerate alternating form, nondegenerate unitary form, or nondegenerate quadratic form, as a "classical $\boldsymbol{F}$-space". For good measure we also include plain vector spaces V, with no additional structure. Thus we have four "types" τ of classical $\boldsymbol{F}$-spaces: $\tau = S$, U, O or L. Occasionally the relationship between two spaces of different types, or two spaces over different fields, comes up, but for the most part all the spaces in a given discussion are of the same type and over the same field. This should be assumed by the reader unless explicitly stated otherwise.

The previous discussion of orthogonal decompositions can be reversed: if V and W are classical $\boldsymbol{F}$-spaces of the same type, we can clearly construct an orthogonal sum $V \perp W$ which is a classical $\boldsymbol{F}$-space of the same type. For type L this is just $V \oplus W$; for the other types it is again $V \oplus W$, with the forms extended in the only way possible such that $(v, w) = 0$ for all $v \in V$ and $w \in W$. In the orthogonal case this means for example that $q(v + w) = q(v) + q(w)$ for all such v and w.

An isometry between two classical $\boldsymbol{F}$-spaces V, W (degenerate or not) of the same type is simply an $\boldsymbol{F}$-isomorphism between V and W preserving the extra structure Γ appropriate to the type, as specified above. When Γ is determined by the form $(\cdot, \cdot)$, that is, in all cases but orthogonal spaces in characteristic 2, the existence of an isometry α is equivalent to the existence of bases $v_1, \ldots, v_n$ of V and $w_1, \ldots, w_n$ of W such that $(v_i, v_j) = (w_i, w_j)$ for all i and j; when these exist one need only take $\alpha(v_i) = w_i$, $i = 1, \ldots, n$.

There is a slightly weaker notion: a similarity between V and W is an $\boldsymbol{F}$-isomorphism between V and W transporting the extra structure Γ on V to a scalar multiple of the structure Γ on W.

The self-isometries of V then obviously form a group $\mathrm{Isom}(V)$, and isometric spaces have isomorphic isometry groups. Indeed:

PROPOSITION B.1. *Similar spaces have isomorphic isometry groups.*

□

So one question which suggests itself is to sort out the different isometry classes—and similarity classes—of classical $\boldsymbol{F}$-spaces of a given type. For finite fields the answer to this question is far simpler than for fields in general.

PROPOSITION B.2. *Nondegenerate symplectic spaces exist only in even dimensions. In each even dimension, and over an arbitrary field $\boldsymbol{F}$, there is a unique isometry class of nondegenerate symplectic $\boldsymbol{F}$-spaces.*

PROPOSITION B.3. *Over a finite field $\boldsymbol{F}$ whose cardinality is a perfect square, there is a unique isometry class of nondegenerate unitary $\boldsymbol{F}$-spaces in each positive dimension.*

PROPOSITION B.4. *Let $\boldsymbol{F}$ be a finite field of odd characteristic. In every odd dimension there are exactly two isometry classes of nondegenerate orthogonal $\boldsymbol{F}$-spaces, and all such spaces are similar.*

PROPOSITION B.5. *Let $\boldsymbol{F}$ be a finite field. In every even dimension there are exactly two isometry classes of nondegenerate orthogonal $\boldsymbol{F}$-spaces.*

In the analysis of these propositions, the notions of isotropic vector and hyperbolic plane are central. In a classical $\boldsymbol{F}$-space of type S, U or O, an isotropic vector is one for which $(v, v) = 0$. Thus in a symplectic space, all vectors are isotropic. A hyperbolic plane U is a two-dimensional subspace which is nondegenerate and which is generated by two isotropic vectors v and w. Since $v \perp v$ but $v \notin U^{\perp}$, $(v, w) \neq 0$. Rescaling w we may take v and w such that $(v, w) = 1$. Then $(w, v) = -1$ if V is of type S, and $(w, v) = 1$ in the other two cases. In particular, any two hyperbolic planes of type S or of type U are isometric. In the case of type O, hyperbolic planes have equivalent bilinear forms, but the quadratic forms giving rise to them may differ (in characteristic 2, only).

Proposition B.2 follows directly from the following lemma and the fact that in the S-type case, all hyperbolic planes are isometric:

LEMMA B.6. *Every nondegenerate symplectic space is the orthogonal sum of hyperbolic planes.*

PROOF. Let V be such a space. Choose any $0 \neq v \in V$. Then $v \notin V^{\perp}$ so there is $w \in V$ such that $(v, w) \neq 0$. Since $(v, v) = 0$, the subspace U spanned by v and w is 2-dimensional and so is a hyperbolic plane. Consequently $V = U \perp U^{\perp}$ and the result follows by induction. □

The proof of Proposition B.3 is more complicated, necessarily so because it fails over some infinite fields.

In the unitary case the fixed field of σ is a subfield $\boldsymbol{F}_0$ of $\boldsymbol{F}$ such that $\boldsymbol{F}$ is a Galois extension of $\boldsymbol{F}_0$ with Galois group $\langle\sigma\rangle \cong Z_2$. We let Tr and N be the trace and norm mappings from $\boldsymbol{F}$ down to $\boldsymbol{F}_0$, so that

$$Tr(c) = c + \overline{c},\ N(c) = c\overline{c} \text{ for all } c \in \boldsymbol{F}.$$

Thus Tr is a nonzero $\boldsymbol{F}_0$-linear mapping. Furthermore, since $\boldsymbol{F}$ is finite, say of cardinality q^2, we have $\overline{c} = c^q$, so $N(c) = c^{1+q}$ and in particular N is surjective.

LEMMA B.7. *Let V be a nondegenerate unitary $\boldsymbol{F}$-space over any field, or a nondegenerate orthogonal space over a field of odd characteristic. Then V is the orthogonal sum of nondegenerate* 1*-dimensional subspaces.*

PROOF. We first argue that V contains a nonisotropic vector. Suppose false and choose any $v, w \in V$ and any $c \in \boldsymbol{F}$. Then in the unitary case we have

$$0 = (v + cw, v + cw) = (v, v) + (cw, cw) + (v, cw) + \overline{(v, cw)} = Tr(c(v, w)).$$

Since c is arbitrary but Tr is a nonzero mapping this implies that $(v, w) = 0$. Since v and w are arbitrary this contradicts the nondegeneracy assumption.

Similarly in the orthogonal case we get $0 = (v + cw, v + cw) = 2c(v, w)$, and as the characteristic is odd this again leads to the impossible conclusion $(v, w) = 0$.

Thus in any case V contains a nonisotropic vector v. Then $V_1 = \boldsymbol{F}v$ is a nondegenerate subspace, so $V = V_1 \perp V_1^{\perp}$, and the lemma follows by induction. □

LEMMA B.8. *Continuing the previous lemma, in the unitary case, if $\boldsymbol{F}$ is finite then V has an orthonormal basis.*

PROOF. By the previous lemma it suffices to prove this in the case $\dim V = 1$. But if $V = \boldsymbol{F}v$ then $(v,v) = \overline{(v,v)}$ is a nonzero element of $\boldsymbol{F}_0$. Since N is onto there is $c \in \boldsymbol{F}$ such that $N(c)^{-1} = (v,v)$, and then $(cv, cv) = N(c)(v,v) = 1$. □

This lemma immediately implies Proposition B.3.

For the orthogonal case the finiteness of $\boldsymbol{F}$ again comes into play, as follows:

LEMMA B.9. *Let a and b be nonzero elements of $\boldsymbol{F}$. Then every element of $\boldsymbol{F}$ may be written $ax^2 + by^2$ for some $x, y \in \boldsymbol{F}$.*

PROOF. If $\boldsymbol{F}$ has characteristic 2 then every element of $\boldsymbol{F}$ is a square and ax^2 already is enough. In odd characteristic, the nonzero squares form a subgroup of $\boldsymbol{F}^\times$ of index 2. So if a is a square and b is not, or vice-versa, then the forms ax^2 and by^2 together are enough. So we may assume that a/b is a square. Multiplying by a^{-1} we see that it is enough to show that every element of $\boldsymbol{F}^\times$ is the sum of two squares. The set of such sums is closed under multiplication by squares, so it is enough to show that some nonsquare is a sum of two squares. But otherwise the set of squares would be an additive subgroup of $\boldsymbol{F}^+$, an absurdity since its cardinality is $(q+1)/2$, which does not divide q. □

LEMMA B.10. *Let V be a nondegenerate orthogonal $\boldsymbol{F}$-space, $\boldsymbol{F}$ finite, and assume that the characteristic of $\boldsymbol{F}$ is odd. Then V has an orthogonal basis in which all but one basis vector satisfies $(v,v) = 1$.*

PROOF. It suffices to show that if $\dim V > 1$, then V contains a vector v such that $(v,v) = 1$; then $V = W \perp W^\perp$ where $W = \boldsymbol{F}v$ and induction completes the proof. We know that V has an orthogonal basis $v_1, \ldots, v_n$. Let $a_i = (v_i, v_i)$. Then $(av_1 + bv_2, av_1 + bv_2) = a_1a^2 + a_2b^2$, and so the preceding lemma, furnishes the desired unit vector. □

If V is an n-dimensional $\boldsymbol{F}$-space with a nondegenerate symmetric bilinear form $(\cdot,\cdot)$, then by taking a basis $B = \{v_1, \ldots, v_n\}$ of V we obtain an $n \times n$ matrix

$$A_B = \big[(v_i, v_j)\big].$$

A different basis $C = \{w_1, \ldots, w_n\}$, related to B by the matrix $X = [x_{ij}]$, so that $w_i = \sum_j x_{ij}v_j$, yields

$$A_C = XA_BX^T.$$

Thus $\det A_C = (\det X)^2 \det A_B$, and so

$$\det A_B(\boldsymbol{F}^\times)^2 \in \boldsymbol{F}^\times/(\boldsymbol{F}^\times)^2$$

is an invariant of the form. We shall call $\det A_B$ the discriminant of the form; it is thus well-defined only as an element of the group $\boldsymbol{F}^\times/(\boldsymbol{F}^\times)^2$ of "cosquares of $\boldsymbol{F}$". Thus in the finite case the discriminant of the form is well-defined only as either a square or a nonsquare. We may use the notation

$$\operatorname{disc} V \in \boldsymbol{F}^\times/(\boldsymbol{F}^\times)^2.$$

LEMMA B.11. *Let V and W be nondegenerate orthogonal subspaces of equal dimension over the field $\boldsymbol{F}$ of odd characteristic. Then V and W are isometric if and only disc V =disc W (in $\boldsymbol{F}^\times/(\boldsymbol{F}^\times)^2$). Moreover, if $\dim V$ is odd, then V and W are similar.*

PROOF. If V and W are isometric, then the discriminants of their forms are clearly the same. Conversely, choose orthogonal bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$ of V and W such that $(v_i, v_i) = (w_i, w_i) = 1$ for all $1 \leq i < n$. Set $a = (v_n, v_n)$ and $b = (w_n, w_n)$. Then the discriminants of the forms, using these bases, are a and b, respectively. Replacing v_n by the scalar multiple cv_n changes a to c^2a. Hence if a and b give equal cosquares, then bases can be chosen so that the two forms correspond exactly, and so the spaces are isometric. This proves the first statement.

For any $c \in \boldsymbol{F}^\times$, let V_c be the same space V but with the new form $(\cdot, \cdot)_c$ defined by $(v, w)_c = c(v, w)$. Then the identity mapping $V \to V_c$ is a similarity, and disc $V_c = [c^n]$ disc V, where $[c]$ is the image of c in the group of cosquares. If n is odd, then disc $V_c = [c]$ disc V, and so spaces similar to V have all possible discriminants. Thus by the first statement, spaces similar to V have all possible isometry types, proving the second statement.

This lemma immediately implies Proposition B.4, and also Proposition B.5 in the case of odd characteristic.

Before completing the proof of Proposition B.5 by considering the case of characteristic 2, we mention an important homogeneity feature of unitary and orthogonal spaces over finite fields.

LEMMA B.12. *Let V be a nondegenerate unitary or orthogonal space. Then any nonzero isotropic vector lies in a hyperbolic plane. If V has dimension at least 2 or 3, respectively, over the finite field $\boldsymbol{F}$, then V contains nonzero isotropic vectors, and hence contains hyperbolic planes.*

PROOF. Let $v \in V$ be any nonzero isotropic vector. Then choosing any vector $w \in V - v^\perp$, we embed v in the nondegenerate 2-dimensional subspace $W = \boldsymbol{F}v + \boldsymbol{F}w$. We search for a second isotropic vector $w' = w + cv$, setting $(v, w) = a$ and $(w, w) = b$, and find that $(w', w') = b + ca + (ca)'$, where $(ca)' = \overline{ca}$ or ca in the unitary or orthogonal case, respectively. In the unitary case $b = \overline{b}$ and so we may find c such that $Tr(ca) = -b$; in the orthogonal case we can certainly solve $2ca = -b$ in odd characteristic, and in characteristic 2 we already know that $(\cdot, \cdot)$ is alternating. Thus our search is successful and $W = \boldsymbol{F}v + \boldsymbol{F}w'$ is a hyperbolic plane.

To prove that isotropic vectors exist, observe again that in the orthogonal case in characteristic 2, the associated bilinear form is alternating so the assertion is trivial. Otherwise, by Lemmas B.8 and B.10, V contains an orthogonal basis $v_1, \ldots, v_n$ such that $(v_1, v_1) = (v_2, v_2) = 1$. An arbitrary linear combination $v = av_1 + bv_2$ then satisfies $(v, v) = N(a) + N(b)$ or $a^2 + b^2$ (in the unitary and orthogonal cases, respectively). Since N is surjective, we may find an isotropic such linear combination in the unitary case. In the orthogonal case we can use Lemma B.9 to find a and b so that v satisfies $(v, v) = -(v_3, v_3)$, and then $v + v_3$ is isotropic.

Now return to orthogonal spaces in characteristic 2. For such a space V, associated bilinear form is alternating and so V is the direct sum of hyperbolic planes. We therefore need to classify 2-dimensional spaces, and then consider isometry types of orthogonal sums of such spaces. In this analysis singular vectors play an important role; by definition, a vector $v \in V$ is singular if and only if $q(v) = 0$. A subspace W of V is called totally singular if and only if all its vectors are singular. Since $q(cv) = c^2 q(v)$, a singular vector generates a totally singular 1-space, usually called a singular 1-space for brevity.

For a 2-dimensional space V with basis v_1 and v_2, one obvious quadratic form is

$$q_0(\xi_1 v_1 + \xi_2 v_2) = \xi_1 \xi_2.$$

The associated bilinear form satisfies $(v_1, v_2) = 1$; there are exactly two singular 1-subspaces, namely those generated by v_1 and v_2.

A second quadratic form arises from identifying V with some field $\boldsymbol{F}^*$ which is an extension of $\boldsymbol{F}$ of degree 2 and setting

$$q_{\boldsymbol{F}^*}(v) = N_{\boldsymbol{F}^*/\boldsymbol{F}}(v).$$

The associated bilinear form is $(v, w) = Tr_{\boldsymbol{F}^*/\boldsymbol{F}}(vw')$, where $w \mapsto w'$ is the nontrivial element of the Galois group of $\boldsymbol{F}^*$ over $\boldsymbol{F}$. There are no singular 1-spaces. In particular this form is not isometric to q_0.

LEMMA B.13. *Any nondegenerate 2-dimensional orthogonal $\boldsymbol{F}$-space, $\boldsymbol{F}$ a perfect field of characteristic 2, is isometric to q_0 or to $q_{\boldsymbol{F}^*}$ for some field extension $\boldsymbol{F}^*$ of $\boldsymbol{F}$ of degree 2.*

PROOF. Suppose first that V contains a singular vector $v \neq 0$. Choose a vector $w \notin \boldsymbol{F}v$; replacing it by a scalar multiple we may assume that $(v, w) = 1$. Writing $q(w) = b$ and setting $w' = v + cw$, $c \in \boldsymbol{F}$, we find that

$$q(w') = q(v + cw) = q(v) + q(cw) + (v, cw) = c^2 b + c.$$

Thus if $b \neq 0$ we may take $c = -b^{-1}$ and in any event we have found a second singular vector w'. Again replacing w' by a scalar multiple, which we call w, we have in the end found a basis v, w of V such that $q(v) = q(w) = 0$ and $(v, w) = 1$. Then $q(av + bw) = (av, bw) = ab$ so q is isometric to q_0.

We now can assume that V contains no singular vectors, and must identify V with some field extension of $\boldsymbol{F}$ of degree 2. Choose any nonzero $v \in V$ and set $a = q(v) \in \boldsymbol{F}$. Since $\boldsymbol{F}$ is perfect, $a = c^2$ for some $c \in \boldsymbol{F}$, and replacing v by $c^{-1}v$ we may assume that $q(v) = 1$. Now choose $w \in V - \boldsymbol{F}v$ such that, as usual, $(v, w) = 1$, and set $b = q(w) \neq 0$. Then for any $x = \xi v + w \in V$,

$$q(x) = \xi^2 q(v) + q(w) + \xi(v, w) = \xi^2 + \xi + b.$$

The polynomial $p(\xi) = \xi^2 + \xi + b$ therefore has no zeros, so is irreducible (and separable); we identify V with its splitting field $\boldsymbol{F}^*$ by identifying v with 1 and w with a root ω of p, and extending $\boldsymbol{F}$-linearly. Let ω' be the other root of p Then for any $\xi \in \boldsymbol{F}$, $N_{\boldsymbol{F}^*/\boldsymbol{F}}(\xi + \omega) = (\xi + \omega)(\xi + \omega') = p(\xi) = q(\xi v + w)$. Thus q coincides with $q_{\boldsymbol{F}^*}$ on vectors of the form $\xi v + w$, and therefore on all vectors. □

We shall call the forms $q_{\boldsymbol{F}^*}$ asingular. Over a finite field $\boldsymbol{F}$ there is a unique extension $\boldsymbol{F}^*$ and so there are exactly two isometry types of two-dimensional orthogonal spaces in characteristic 2: one is asingular, the other isn't.

In analogy with Lemma B.12, we have:

LEMMA B.14. *Let V be a nondegenerate orthogonal space of dimension at least 4 over a finte field $\boldsymbol{F}$ of characteristic 2. Then V contains a nondegenerate 2-dimensional subspace which is not asingular.*

PROOF. By Lemma B.6, V is the orthogonal sum of hyperbolic planes. It suffices to show that the orthogonal sum $W = X \perp Y$ of two asingular hyperbolic planes contains a nondegenerate 2-dimensional subspace which is not asingular. We write elements of W as $w = \langle x, y\rangle$, $x \in \boldsymbol{F}^*$, $y \in \boldsymbol{F}^*$. (Since $\boldsymbol{F}$ is finite, $\boldsymbol{F}^*$ is unique and so X and Y can be identified). Choose a basis x_1, x_2 of $\boldsymbol{F}^*$ over $\boldsymbol{F}$; then the element $\langle x_1, x_1\rangle$ is singular, and together with $\langle x_2, 0\rangle$ generates a 2-dimensional nondegenerate subspace, which is not asingular, as required. □

As an immediate consequence we have:

LEMMA B.15. *Over a finite field $\boldsymbol{F}$ of characteristic 2, let W and X be 2-dimensional orthogonal spaces equipped with the quadratic forms q_0 and $q_{\boldsymbol{F}^*}$ defined above. Then every $2n$-dimensional nondegenerate orthogonal $\boldsymbol{F}$-space is isometric to one of the two spaces*

$$V_{2n}^+ = W \perp W \perp \cdots \perp W \perp W \text{ or } V_{2n}^- = W \perp W \perp \cdots \perp W \perp X.$$

To distinguish these spaces we introduce an invariant $\Delta(V)$ of V in characteristic 2 analogous to disc V in odd characteristic. This was introduced by Arf in 1941, and is called the "quasidiscriminant" by Dieudonné[7]. We shall call it the Dickson discriminant, since in characteristic different from 2 the discriminant of the form and the determinant of a transformation are dual operations, while in characteristic 2 the Dickson discriminant of the form is dual to an invariant of an orthogonal transformation defined by Dickson and now called the "Dickson invariant".

In characteristic different from 2, the squaring map $\boldsymbol{F}^\times \to \boldsymbol{F}^\times$ is at the heart of things; the discriminant of an orthogonal space is an element of the cokernel, and the determinant of an orthogonal transformation is an element of the kernel. In characteristic 2 the analogue of this squaring map is the mapping $\wp : \boldsymbol{F}^+ \to \boldsymbol{F}^+$, $\wp(x) = x^2 + x$, whose kernel is just $\boldsymbol{F}_2$; the Dickson discriminant is an element

$$\Delta(V) \in \boldsymbol{F}^+/\wp(\boldsymbol{F}^+) \cong Z_2 \tag{BA}$$

while the Dickson invariant takes its values in $\boldsymbol{F}_2$. Moreover it turns out that

$$\Delta(V_1 \perp V_2) = \Delta(V_1) + \Delta(V_2),\ \Delta(W) = 0,\ \Delta(X) \neq 0, \tag{BB}$$

where W and X are as in Lemma B.15, and of course $\Delta(V)$ depends only on the isometry type of V.

Consequently the Dickson discriminant distinguishes V_{2n}^+ and V_{2n}^-, and so we obtain the following proposition, completing the proof of Proposition B.5.

PROPOSITION B.16. *In Lemma B.15, every nondegenerate space is isometric to exactly one of V_{2n}^+, V_{2n}^-. If V_1 and V_2 are nondegenerate orthogonal spaces of the same dimension over a finite field of characteristic 2, then V_1 and V_2 are isometric if and only if $\Delta(V_1) = \Delta(V_2)$.*

The Dickson discriminant is defined as follows. Choose any decomposition $V = V_1 \perp \cdots \perp V_n$ as an orthogonal sum of hyperbolic planes, and choose a basis v_i, w_i of V_i such that $(v_i, w_i) = 1$. Then set

$$\Delta(V) = \sum Q(v_i)Q(w_i). \qquad \bmod \wp(\boldsymbol{F}).$$

[7] J. Dieudonné, *La Géometrie des Groupes Classiques*

This apparently depends on the choice of hyperbolic basis, but a different choice of basis leads to the same value $\mod \wp(\boldsymbol{F})$.

EXERCISE. Check this.

The properties (BB) are then all easy to verify.

Another useful invariant of a classical space V of type S, U or O is its **Witt index** $w(V)$, defined as follows:

(a) If V is of type S or U, or of type O in odd characteristic, then $w(V)$ is the maximum dimension $\dim U$ among all totally isotropic subspaces U of V.
(b) If V is of type O in characteristic 2, then $w(V)$ is the maximum dimension $\dim U$ among all totally singular subspaces U of V.

Here U is totally isotropic if the restriction of $(\cdot,\cdot)$ to U is identically 0, and U is totally singular if the restriction of q to U is identically 0. Since $q(v+w) = q(v)+q(w)+(v,w)$, any totally singular subspace is also totally isotropic. Moreover, any totally isotropic subspace U satisfies $U \subseteq U^\perp$, and so $\dim U \le \dim V/2$. Notice also that if $V = V_1 \perp V_2$ and U_i is a totally isotropic (or totally singular) subspace of V_i, $i = 1, 2$, then $U_1 \oplus U_2$ is again totally isotropic (or totally singular). Hence w is superadditive: $w(V_1 \perp V_2) \ge w(V_1) + w(V_2)$.

To compute $w(V)$ for all nondegenerate V, we use the following fundamental lemma:

LEMMA B.17. *Let V be a classical space of type S, O or U. Let U be a totally isotropic subspace of V. Let $x_1, \dots, x_m$ be a basis of U, with x_1 singular in the case of type O. Then there exists $y_1 \in V$ such that if we put $V_1 = \boldsymbol{F}x_1 + \boldsymbol{F}y_1$, then V_1 is a hyperbolic plane orthogonal to $x_2, \dots, x_m$, and in the case of type O and even characteristic, V_1 is not asingular.*

PROOF. Set $X_1 = x_1^\perp$, so that $x_1 \in X_1$ and X_1 is a hyperplane of V. Indeed $\boldsymbol{F}x_1 = X_1^\perp$ and so $\boldsymbol{F}x_1 = \operatorname{rad} X_1$. Choose any complement Y_1 to $\boldsymbol{F}x_1$ in X_1 containing $x_2, \dots, x_m$, so that $X_1 = \boldsymbol{F}x_1 \perp Y_1$ and Y_1 is nondegenerate. Then set $W_1 = Y_1^\perp$, so that $\dim W_1 = 2$ and $x_1 \in W_1$. Since Y_1 is nondegenerate, so is W_1. By our classification, since W_1 contains isotropic vectors, W_1 is a hyperbolic plane; and in the orthogonal, characteristic 2 case, W_1 contains singular vectors so is not asingular. Therefore there exists $y_1 \in W_1 - \boldsymbol{F}x_1$ which is also isotropic (or singular). □

This determines the Witt index $w(V)$ for all our nondegenerate spaces V.

PROPOSITION B.18. *The Witt index $w(V)$ of a nondegenerate classical space of type S, U or O of dimension m over a finite field $\boldsymbol{F}$ is as follows:*

(a) *If V is of type S or U, or if V is of type O and m is odd, then $w(V) = [m/2]$.*
(b) *If V is of type O and m is even, then $w(V) = m/2$ if V is the orthogonal sum of 2-dimensional spaces which are hyperbolic planes (or not asingular in the characteristic 2 case). Otherwise $w(V) = (m/2) - 1$.*

PROOF. As observed above, $w(V) \le m/2$ and so $w(V) \le [m/2]$. The assertion in (a) follows from the fact that V contains the orthogonal sum of $[m/2]$ hyperbolic planes, by Lemmas B.6 and B.12. For the assertion of (b), the preceding lemma, used inductively, allows us to construct, from any r-dimensional totally isotropic (or totally singular, in characteristic 2) subspace, a set of mutually orthogonal hyperbolic planes (which are not asingular in characteristic 2). Thus $w(V)$ is the

largest number of mutually orthogonal such planes which can be found in V. This number is $m/2$ or $m/2-1$ by the first assertion of Proposition B.16. □

One neat deduction from the development can be stated:

COROLLARY B.19. *The similarity class of a classical space over a finite field is determined by its type, dimension and Witt index.*

Having sorted out the different types of nondegenerate spaces, we proceed to the key result of Witt. One part of it is a cancellation law.

PROPOSITION B.20. *Let V be a nondegenerate classical $\boldsymbol{F}$-space, $\boldsymbol{F}$ a finite field. Suppose that W and X are nondegenerate subspaces of V and that W and X are isometric. Then $W^\perp$ and $X^\perp$ are isometric.*

PROOF. We have $V = W \perp W^\perp = X \perp X^\perp$, all of these spaces being nondegenerate. In particular since W and X have the same dimension, so do $W^\perp$ and $X^\perp$. In the symplectic and unitary cases, the assertion follows by Propositions B.2 and B.3. In the orthogonal case in odd characteristic, we have disc V =disc W disc $W^\perp$ =disc X disc $X^\perp$, so disc $W^\perp$ =disc $X^\perp$ and Lemma B.11 completes the proof. Likewise in characteristic 2, consideration of the Dickson discriminant Δ provides the proof, using Proposition B.16. □

THEOREM B.21 ("WITT'S LEMMA"). *Let V_1 and V_2 be isometric nondegenerate classical $\boldsymbol{F}$-spaces, and let U_1 and U_2 be isometric subspaces of V_1 and V_2, respectively. Then any isometry from U_1 to U_2 extends to an isometry from V_1 to V_2.*

The statement of the theorem is equivalent to the combination of the following two corollaries.

COROLLARY B.22. *Isom(V) permutes transitively the set of all subspaces of V of a given isometry type.*

COROLLARY B.23. *Let $U \subseteq V$. Then the stabilizer in Isom(V) of U induces all of Isom(U) on U (although it generally does not act faithfully).*

PROOF OF WITT'S LEMMA. Let $\phi : U_1 \to U_2$ be an isometry between subspaces of the isometric nondegenerate spaces V_1 and V_2. We argue that ϕ can be extended to an isometry $V_1 \to V_2$, by induction on $\dim(V_1/U_1) + \dim V_1$. Thus it suffices only to extend ϕ to an isometric embedding of some subspace of V_1 properly containing U_1 into V_2.

We may write $U_1 = \operatorname{rad} U_1 \perp X_1$, for some nondegenerate space X_1. Set $X_2 = \phi(X_2)$; then $U_2 = \operatorname{rad} U_2 \perp X_2$. Also set $W_i = X_i^\perp$, $i = 1, 2$. Since $V_i = X_i \perp W_i$ and ϕ already induces an isometry from X_1 to X_2, we have that W_1 and W_2 are isometric by the cancellation law. It is enough to show that $\phi|\operatorname{rad} U_1$ extends to an isometry from W_1 to W_2, and this follows by induction unless $W_i = V_i$, which we may henceforth assume.

Thus $X_1 = 0$ and so U_1 is totally isotropic. Let $x_1, \ldots, x_m$ be a basis of U_1, chosen in the type O case so that if possible, x_1 is singular. Set $x_i' = \phi(x_i)$. Except in the case of orthogonal groups in even characteristic, we apply Lemma B.17 to find isotropic vectors $y_1 \in V_1$ and $y_1' \in V_2$ such that $\boldsymbol{F}x_1 + \boldsymbol{F}y_1$ and $\boldsymbol{F}x_1' + \boldsymbol{F}y_1'$ are hyperbolic planes, with $(y_1, x_i) = (y_1', x_i') = 0$ for all $i > 1$. Then if we put

$\phi(y_1) = y_1'$, and extend linearly. we have extended ϕ to an isometric embedding of the larger subspace $U_1 + \boldsymbol{F}y_1$, and are done by induction. We may therefore assume that the characteristic is 2 and the type is O. If x_1 is singular, then the previous argument may be imitated, again using Lemma B.17.

We are reduced to the case of type O, characteristic 2, and U_1 contains no singular vectors. Since U_1 is totally isotropic, $q(v + w) = q(v) + q(w)$, i.e., q is additive on U_1, and is semilinear as $q(cv) = c^2q(v)$. Since $q|U_1$ maps U_1 to the 1-dimensional space $\boldsymbol{F}^+$, its kernel has codimension 1; but the kernel consists of singular vectors so is 0. Thus we are finally reduced to the case $U_1 = \boldsymbol{F}x_1$, with x_1 nonsingular. If $U_1^\perp$ and $U_2^\perp$ contain singular vectors x_2 and x_2', respectively, we may extend ϕ by sending x_2 to x_2', and are done by induction. Thus we But $U_1^\perp = U_1 \perp T_1$, with T_1 nondegenerate. If T_1 contains a nonsingular vector x, then some linear combination $x_1 + cx$ is singular, since $q(x_1 + cx) = q(x_1) + c^2q(x)$. Thus as long as $T_1 \neq 0$, i.e., $\dim V_1 > 2$, the vector x_2 is available, and similarly x_2' is as well.

Finally we must consider the case $\dim V = 2$. In this case Witt's Lemma is equivalent to the statement that for any $c \in \boldsymbol{F}$, $\mathrm{Isom}(V)$ is transitive on the set of vector v such that $q(v) = c$. For the asingular case, we identify V with $\boldsymbol{F}^*$ and have $q(x) = N(x)$ where $N = N_{\boldsymbol{F}^*/\boldsymbol{F}}$. Then for any $y \in \boldsymbol{F}^*$ such that $N(y) = 1$, left multiplication λ_y by y lies in $O(V)$. But if $N(x) = N(x')$, then $N(x'x^{-1}) = 1$ and $\lambda_{x'x^{-1}}$ carries x to x', as required. Finally for the remaining case, taking independent singular vectors x, y such that $(x, y) = 1$, the transformation $h(c) : x \mapsto cx$, $y \mapsto c^{-1}y$ lies in $O(V)$ for each $c \in \boldsymbol{F}^\times$. As c varies we find that $O(V)$ has all orbits of length at least $|\boldsymbol{F}| - 1$ on $V - \boldsymbol{F}x - \boldsymbol{F}y$. But in this set there are just $|\boldsymbol{F}| - 1$ vectors with a given value of q, so $O(V)$ is transitive on them. As for singular vectors, the $h(c)$ permute transitively the scalar multiples of x; and the transformation interchanging x and y lies in $O(V)$, so $O(V)$ is transitive as well on singular vectors. □