Classical Groups and Geometry

J.I. Hall

6 May 2015

Contents

Chapter 1

Introduction

1.1 Unique factorization

We are familiar with

(1.1). THEOREM. *For n a positive integer at least 2, let*

$$n = \prod_{i=1}^{k} p_i$$

and

$$n = \prod_{j=1}^{m} q_j$$

where each p_i and q_j is prime. Then $k = m$ and there is a permutation π with $p_i = q_{\pi(j)}$, for all i. □□

Here a prime is a positive integer not 1 and only divisible by 1 and itself.

A group S is simple if it is not 1 and has only 1 and itself as homomorphic images. The appropriate unique factorization theorem for finite groups is then

(1.2). THEOREM. (JORDAN-HÖLDER) *For G a nontrivial finite group, let*

$$G = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_k = 1$$

and

$$G = H_0 \trianglerighteq H_1 \trianglerighteq \cdots \trianglerighteq H_m = 1$$

where each $P_i = G_{i-1}/G_i$ and $Q_j = H_{j-1}/H_j$ is simple. Then $k = m$ and there is a permutation π with $P_i \simeq Q_{\pi(j)}$, for all i. □□

A big difference between the two factorization results is that the first admits the natural converse. Two positive integers with the same multiset of prime divisors are equal, but two groups (even abelian groups) with the same multiset of composition factors might not be isomorphic. So while number theory can focus on properties of the prime numbers, in finite group theory we must not only examine the simple groups but also study in what ways they can be glued together.

A major result identifies the possible factors in a Jordan-Hölder composition series:

(1.3). Theorem. (Classification of Finite Simple Groups) (1983, 2004) *A finite simple group is isomorphic to one of:*

(1) *a cyclic group of prime order p: Z_p;*

(2) *an alternating group:* $\mathrm{Alt}(n)$*;*

(3) *a classical group:* $\mathrm{PSL}_n(q)$, $\mathrm{PSp}_n(q)$, $\mathrm{PSU}_n(q)$, $\mathrm{P}\Omega_n^\epsilon(q)$*;*

(4) *an exceptional Lie type group* ${}^2\mathrm{B}_2(q)$, ${}^3\mathrm{D}_4(q)$, $\mathrm{E}_6(q)$, ${}^2\mathrm{E}_6(q)$, $\mathrm{E}_7(q)$, $\mathrm{E}_8(q)$, $\mathrm{F}_4(q)$, ${}^2\mathrm{F}_4(q)'$, $\mathrm{G}_2(q)$, ${}^2\mathrm{G}_2(q)$*;*

(5) *a sporadic simple group, of which there are twenty-six.* □□

Here n is an integer at least 2 and q is a prime power. Most choices of these two parameters do in fact give simple groups. The classical groups provide four two-parameter infinite families, so one could say loosely that most finite simple groups are classical. These are the families of groups that will be of greatest interest to us, but there will rarely be any advantage in restricting our attention to those that are finite.

With CFSG in hand, to find all finite groups, we are faced with the problem of gluing things together: extension theory. One of the reasons to study the examples (and especially the classical groups) is that additional knowledge can aid us is solving such questions. For instance Theorem (1.1) can be refined to:

(1.4). Theorem. *If A is a finite abelian group, then A is isomorphic to a direct sum $\bigoplus_{i=1}^k A_i$, for cyclic subgroups A_i of order $p_i^{a_i}$, with p_i prime and a_i a positive integer. Furthermore, if A is also isomorphic to $\bigoplus_{j=1}^m B_j$ for cyclic subgroups B_j of order $q_j^{b_j}$, with q_j prime and b_j a positive integer, then $k = m$ and there is a permutation π with $p_i = q_{\pi(j)}$ and $a_i = b_{\pi(j)}$, for all i.* □□

This is clearly a "unique factorization" result, but it is also a Jordan-Hölder theorem. Indeed theorems of Jordan-Hölder type are valid for many lattices, provided certain properties hold and appropriate definitions are made. So, for Theorem (1.4) we require that all factors are indecomposable and that all extensions split.

1.2 Some notation

Let G be a group. If H is a subgroup of G, then we write $H \leq G$ and $G \geq H$. We write $H \trianglelefteq G$ and $G \trianglerighteq H$ if H is a normal subgroup of G. For the subset X of G, $\langle X \rangle$ is the subgroup of G generated by X.

For $x, h \in G$ we write x^h for $h^{-1}xh$, the *conjugate* of x by h. More generally, for subsets X and H of G, we set $X^H = \{\, x^h \mid x \in X \,,\, h \in H \,\}$. Be warned: in some group theory texts [Rob82], X^H is defined to be the subgroup $\langle X^H \rangle$. That is not our convention.

1.3 Categories

We shall not focus on categories and their role in modern algebra (which is large), but they provide us for a convenient language for setting up our work.

A *category* C is a class $\mathrm{Obj}(\mathsf{C})$ of *objects*. For each pair $A, B \in \mathrm{Obj}(\mathsf{C})$, there is a set $\mathrm{Hom}_{\mathsf{C}}(A, B)$, pairwise disjoint, of C-*morphisms* For each triple of objects A, B, C, there is a *composition map*

$$\circ \colon \mathrm{Hom}_{\mathsf{C}}(A,B) \times \mathrm{Hom}_{\mathsf{C}}(B,C) \longrightarrow \mathrm{Hom}_{\mathsf{C}}(A,C)\,,$$

the image $\circ(f, g)$ usually being written $f \circ g$ or just fg.[1] The following are required:

(i) Always for $a \in \mathrm{Hom}_{\mathsf{C}}(A,B)$, $b \in \mathrm{Hom}_{\mathsf{C}}(B,C)$, and $c \in \mathrm{Hom}_{\mathsf{C}}(C,D)$ we have

$$(a \circ b) \circ c = a \circ (b \circ c) \in \mathrm{Hom}_{\mathsf{C}}(A,D)\,.$$

(ii) For every object X there is a unique morphism $1_X \in \mathrm{Hom}_{\mathsf{C}}(X,X)$, such that always for $a \in \mathrm{Hom}_{\mathsf{C}}(A,B)$ and $b \in \mathrm{Hom}_{\mathsf{C}}(B,C)$ we have

$$a \circ 1_B = a \text{ and } 1_B \circ b = b\,.$$

The notation 1_A can be confusing in those situations where the objects themselves have identity elements; for instance, we use 1_G to denote the identity element of the group G. Usually the usage will be clear from the context, but for clarity we will at times use Id_A to denote an identity map on some object A.

A motivating model for a category has Obj consisting of all sets with the morphism set $\mathrm{Hom}(A,B)$ be all set maps (functions) from A to B. Composition is then the usual composition of maps, and (i) observes that composition is associative. Then (ii) records the properties of the identity map $1_X = \mathrm{Id}_X$ from the set X to itself. We will denote this category Set.

Two objects A and B of the category C are *isomorphic* if there are morphisms $a \in \mathrm{Hom}_{\mathsf{C}}(A,B)$ and $b \in \mathrm{Hom}_{\mathsf{C}}(B,A)$ with $ab = 1_A$ and $ba = 1_B$. For instance, two sets are isomorphic in Set precisely when they are in bijection.

It is crucial to note that the definition of a category does not require $\mathrm{Obj}(\mathsf{C})$ to be a set. This is important since, for instance, we know that there is no such thing as the set of all sets. A category in which the class of objects is actually

[1] An alternative and common convention is to write $g \circ f$ and gf for the composition of f followed by g; see Section 1.5.

a set is a *small category*. Our definition is somewhat restrictive. By requiring that each $\mathrm{Hom}_{\mathsf{C}}(A, B)$ actually be a set, we have confined ourselves to those categories that are *locally small*.

A *subcategory* D of C is a category for which every object D of D is also an object of C and such that, for $D, E \in \mathrm{Obj}(\mathsf{D})$, we have $\mathrm{Hom}_{\mathsf{D}}(D, E) \subseteq \mathrm{Hom}_{\mathsf{C}}(D, E)$. The subcategory D of C is *full* if we always have the equality $\mathrm{Hom}_{\mathsf{D}}(D, E) = \mathrm{Hom}_{\mathsf{C}}(D, E)$. The subcategory D of C is *dense* if, for every object C of C there is an object D of D that is isomorphic to C in C. For instance, the full subcategory FSet of finite sets within Set has, in turn, the full and dense subcategory $\mathbb{Z}\mathsf{FSet}$ whose objects are the finite subsets of the integers.

A category C is *concrete* if it is a subcategory of Set. That is, if the objects of C are sets, perhaps decorated with additional structure, and the morphisms of $\mathrm{Hom}_{\mathsf{C}}(A, B)$ are set maps, perhaps with additional, required properties. Our prime example is Grp, the category of all groups in which the morphisms are the group homomorphisms.

We have already introduced the two important categories Set and Grp. For us the third category of primary interest is ${}_D\mathsf{Vec}$, the category of all left vector spaces over the division ring D. (Of course, we should really precede this by discussion of the categories Fld of fields and $\mathsf{DivRing}$ of *division rings*—not necessarily commutative fields.) Here the morphisms are the D-linear transformations from one D-space to another.

These three main categories have important full subcategories: the category FSet of finite sets (mentioned above); the category FGrp of finite groups and the category AbGrp of abelian groups; and the category ${}_D\mathsf{FVec}$ of finite dimensional left D-spaces.

Other categories of vector spaces will also be important to us. The categories Vec_D, and FVec_D are the right-space counterparts to the left-space categories ${}_D\mathsf{Vec}$ and ${}_D\mathsf{FVec}$. A more subtle example is the category Vec of *all* left vector spaces. (It has righthanded counterpart RVec.) Here the objects are pairs (D, V) (or ${}_DV$) of a division ring D and a left D-space V. The morphisms must then also be pairs: $[\sigma, s] \in \mathrm{Hom}_{\mathsf{Vec}}({}_DV, {}_EW)$ where σ is a homomorphism from D to E (a morphism from $\mathsf{DivRing}$) and s is an abelian group homomorphism (from AbGrp) that are compatible, in that for $a, b \in D$ with

$$a \xrightarrow{\sigma} a' \quad \text{and} \quad b \xrightarrow{\sigma} b'$$

and $u, v \in V$ with

$$u \xrightarrow{s} u' \quad \text{and} \quad v \xrightarrow{s} v'$$

we always have

$$au + bv \xrightarrow{[\sigma,s]} a'u' + b'v' \,.$$

Such maps are *semilinear*. It is important to realize that ${}_D\mathsf{Vec}$ is a subcategory of Vec that is typically not full, since, even for two spaces V and W both over $D = E$, the latter allows σ to be a nontrivial automorphism of the division ring D.

There are other categories we may encounter: rings Ring, modules ${}_R\mathsf{Mod}$ and Mod_R, and associative algebras Assoc.[2] These concrete categories are additionally *additive categories*. This means that each object set has a natural structure as abelian group and that each morphism is an abelian group homomorphism.

For categories C and D a *functor* $\mathbf{F}$ from C to D associates to each object C of C an object $\mathbf{F}C$ of D and to each morphism $f \in \mathrm{Hom}_{\mathsf{C}}(A, B)$ a morphism $\mathbf{F}(f)$ of $\mathrm{Hom}_{\mathsf{D}}(\mathbf{F}A, \mathbf{F}B)$, such that always

$$\mathbf{F}(f)\mathbf{F}(g) = \mathbf{F}(fg) \quad \text{and} \quad \mathbf{F}(1_A) = 1_{\mathbf{F}A}\,.$$

An obvious functor from C to itself is the *identity functor* $\mathbf{1}_{\mathsf{C}}$ with $\mathbf{1}_{\mathsf{C}}A = A$ and $\mathbf{1}_{\mathsf{C}}(f) = f$. The two categories C and D are *isomorphic* provided there are functors $\mathbf{F}\colon \mathsf{C} \longrightarrow \mathsf{D}$ and $\mathbf{G}\colon \mathsf{D} \longrightarrow \mathsf{C}$ with $\mathbf{FG} = \mathbf{1}_{\mathsf{D}}$ and $\mathbf{GF} = \mathbf{1}_{\mathsf{C}}$.

Although isomorphism gives us an equivalence relation on the collection of all categories, it is not a terribly helpful one. We have observed above that the category FSet of finite sets has the full, dense subcategory $\mathbb{Z}\mathsf{FSet}$ of finite subsets of the integers. These two categories are certainly not isomorphic, since the second is a small category while the first is not. On the other hand it seems relatively clear that the two categories do not differ in any other substantive way. More useful than isomorphism is *category equivalence.* Two categories are equivalent provided they have isomorphic full, dense subcategories. In particular FSet and $\mathbb{Z}\mathsf{FSet}$ are equivalent categories.

In the category C, the morphisms of $\mathrm{Hom}_{\mathsf{C}}(A, A)$ are the C-*endomorphisms*, and we will write $\mathrm{End}_{\mathsf{C}}(A)$ for $\mathrm{Hom}_{\mathsf{C}}(A, A)$. Those endomorphisms of A that are invertible, that is, are isomorphisms of A with itself, are the C-*automorphisms*, written $\mathrm{Aut}_{\mathsf{C}}(A)$ even $\mathrm{Aut}(A)$. As we see in the next section, these automorphism groups will play a central role for us. For instance $\mathrm{Aut}_{\mathsf{Set}}(A) = \mathrm{Sym}(A)$.

1.4 Representation and action

In group theory as in most parts of mathematics, in order to study an object carefully we wish to have a description of it that is easy to work with. For groups, this is done by representing them through their action upon something. Such actions may also be the reasons we are studying the groups in the first place. For instance, Galois was the first to consider finite groups seriously, and he encountered these groups as permutations of polynomial roots.

Let A be an object of the category C. Then a C-*representation* of the group G on A is a homomorphism $\rho\colon G \longrightarrow \mathrm{Aut}_{\mathsf{C}}(A)$. In this case we say that G is a *group of operators* on A and that G *acts on* A *via* ρ. This *action* of G on A is *faithful* if the kernel of ρ is the identity.

Three types of categories C and the associated representations within $\mathrm{Aut}_{\mathsf{C}}(\cdot)$ will be of particular interest to us:

[2] Hooray for the Oxford comma.

> (1) In Set each object A is a set (with no further structure). Its automorphism group $\mathrm{Aut}_{\mathsf{C}}(A) = \mathrm{Aut}_{\mathsf{Set}}(A)$ is then $\mathrm{Sym}(A)$, the *symmetric group* of all permutations on the set A. The associated representations are *permutation representations.*
>
> (2) For the category $\mathsf{C} = {}_D\mathsf{Vec}$ of vector spaces V over the division ring D, the automorphism group is the *general linear group* $\mathrm{GL}_D(V)$, also written $\mathrm{GL}({}_DV)$ or $\mathrm{GL}(V)$. The associated representations are the *D-linear representations.*[3] If instead we take the vector space ${}_DV$ as an object in the category Vec (with no uniformly specified coefficient ring) then we have $\mathrm{Aut}_{\mathsf{Vec}}({}_DV) = \Gamma\mathrm{L}_D(V)$ or $\Gamma\mathrm{L}({}_DV)$ or (with some abuse) $\Gamma\mathrm{L}(V)$. The representations are now *semilinear*, since they may involve nontrivial automorphisms of D.
>
> (3) If $\mathsf{C} = \mathsf{Grp}$, the category of groups, then we are concerned with *group automorphisms.*

In concrete categories, such as these, if G acts on the object A and B is a subset of A with $B^g \subseteq B$ for all $g \in G$, then we say that B is *G-invariant.* If B is actually a subobject of A, then G acts on B via the restriction of g to B, which we write as $g|_B$.

If the object B of C is isomorphic to A via the morphism $f\colon A \longrightarrow B$, then the commutative diagram:

$$\begin{tikzcd}
A \arrow[r, "h"] \arrow[d, "f"] & A \arrow[d, "f"] \\
B \arrow[r, "h^*"] & B
\end{tikzcd}$$

provides an isomorphism $f^*\colon \mathrm{Aut}_{\mathsf{C}}(A) \longrightarrow \mathrm{Aut}_{\mathsf{C}}(B)$ given by

$$h \xrightarrow{f^*} h^* = f^{-1}hf\,.$$

Any representation ρ_A of G on A then has a naturally associated representation $\rho_B = \rho_A f^*$ on B; the following diagram commutes:

$$\begin{tikzcd}
G \arrow[r, "\rho_A"] \arrow[dr, "\rho_B"] & \mathrm{Aut}_{\mathsf{C}}(A) \arrow[d, "f^*"] \\
 & \mathrm{Aut}_{\mathsf{C}}(B)
\end{tikzcd}$$

In this case ρ_A and ρ_B are said to be *equivalent* representations. If G is isomorphic to the group H via the map $\gamma\colon G \longrightarrow H$, and the following diagram

[3] In many places, a linear representation is required to be acting on a finite dimensional space from ${}_D\mathsf{FVec}$. This restriction will not be made here.

commutes:

$$\begin{tikzcd}
G \arrow[r, "\rho_A"] \arrow[d, "\gamma"] & \mathrm{Aut}_{\mathsf{C}}(A) \arrow[d, "f^*"] \\
G \arrow[r, "\rho_B"] & \mathrm{Aut}_{\mathsf{C}}(B)
\end{tikzcd}$$

then the two representations ρ_A and ρ_B are said to be *semi-equivalent* with $\rho_B = \rho_A^{[\gamma,f]}$.

The equivalence f^* is just the semi-equivalence $[\mathrm{Id}_G, f]$:

$$\begin{tikzcd}
G \arrow[r, "\rho_A"] \arrow[d, "\mathrm{Id}_G"] & \mathrm{Aut}_{\mathsf{C}}(A) \arrow[d, "f^*"] \\
G \arrow[r, "\rho_B"] & \mathrm{Aut}_{\mathsf{C}}(B)
\end{tikzcd}$$

where we have used Id_G to denote the identity endomorphism of G in Grp.

If C is an additive category, then the set of C-endomorphisms $\mathrm{End}_{\mathsf{C}}(A) = \mathrm{Hom}_{\mathsf{C}}(A, A)$ has a natural structure as a ring under composition and (pointwise) addition. In this case, we may also define a C-representation of a ring R as a ring homomorphism $\varphi\colon R \longrightarrow \mathrm{End}_{\mathsf{C}}(A)$ for A and object of C, where we must require the identity of the ring R to map to 1_A. The group $\mathrm{Aut}_{\mathsf{C}}(A)$ is the group of units of $\mathrm{End}_{\mathsf{C}}(A)$, so for additive categories a representation of the group G leads to a representation of the group ring $\mathbb{Z}G$. This can provide a powerful tool for study of G. We have concepts of equivalence and semiequivalence similar to those above.

1.5 Right versus left

Action on the left is exhibited by function composition,

$$f(g(x)) = (fg)(x)\,,$$

while action on the right is modeled by exponentiation,

$$(x^f)^g = x^{fg}\,.$$

The distinction is largely a matter of convention. Analysts almost always favor left action, since they are regularly dealing with functions and their properties. Algebraists often prefer right action, and that is usually the case here. This is evidenced by our composition map for morphisms; our definitions $h^* = f^{-1}hf$ in the previous section and of conjugation by $x^g = g^{-1}xg$; and our chosen convention for permutation multiplication:

$$(1,2,3,4,5)(2,4,6,8) = (1,4,5)(2,3,6,8)\,.$$

We do not demand right action always. For instance, above we have written functors as functions and compose them as functions. And even when our action is on the right, we sometimes use function notation, for instance in characterizing a group homomorphism $\varphi\colon G \longrightarrow H$ by $\varphi(g)\varphi(h) = \varphi(gh)$ (but here we must be careful to avoid casual composition).

Our preferred notation for right action is exponentiation

$$a \xrightarrow{f} a^f ,$$

but there may be times when is it helpful to use af or even $a.f$ rather than a^f. (In particular, nested exponentiation can look very confusing.)

In any event, there are situations where the correct thing is to use both right and left action. For instance our preferences for right action of morphisms and left scalar action on vector spaces go hand in hand:

(1.5). PROPOSITION. *Let V and W be left vector spaces over the division ring D, and let g be a D-endomorphism (linear transformation) from V to W. That is, for $u, v \in V$ and $x, w \in W$ with*

$$u \xrightarrow{g} x \text{ and } v \xrightarrow{g} w ,$$

and, for $\alpha, \beta \in D$, we have

$$\alpha u + \beta v \xrightarrow{g} \alpha x + \beta w .$$

Let $e_1, \dots, e_i, \dots, e_m$ be a basis of V, so that $v = \sum_{i=1}^m v_i e_i$ of V is represented in the left D-space of row vectors D^m by

$$\vec{v} = (v_1, \dots, v_i, \dots, v_m) .$$

Similarly let $f_1, \dots, f_j, \dots, f_n$ be a basis of W with $w = \sum_{j=1}^n w_j f_j$ of W represented in D^n by

$$\vec{w} = (w_1, \dots, w_j, \dots, w_n) .$$

Define the scalars g_{ij} from D by

$$e_i \xrightarrow{g} \sum_{j=1}^n g_{ij} f_j ,$$

and then let $G = (g_{ij})_{ij}$ be the matrix of $D^{m,n}$ with (i,j)-entry g_{ij}. Then

$$\vec{v}G = \vec{w} . \qquad \Box$$

The point here is the action of the morphism g taking the *left* D-space V to the *left* D-space W is naturally represented via *right* multiplication by the matrix G.

Let V and W be left D-spaces with $\mathrm{Hom}(V, W)$ the set of abelian group homomorphisms from V to W and $\mathrm{Hom}_D(V, W)$ $(= \mathrm{Hom}_{{}_D\mathsf{Vec}}(V, W))$ the corresponding set of linear transformations—homomorphisms as left D-spaces. In this situation one often says that the f of $\mathrm{Hom}_D(V,W)$ are those homomorphisms whose group action "commutes" with the scalar action of D. Instead it might be better to say that these actions "associate," since the condition that f of $\mathrm{Hom}(V,W)$ belongs to $\mathrm{Hom}_D(V,W)$ is

$$(\alpha x)f = \alpha(xf)\,,$$

for all $\alpha \in D$ and $x \in V$.

Of course, the matrix spaces

$$D^m (= D^{1,m} = \mathrm{Mat}_{1,m}(D)) \quad \text{and} \quad D^{m,n}\,(= \mathrm{Mat}_{m,n}(D))$$

have natural structure as both left and right D-spaces. But if D is a noncommutative division ring, care must be taken since, in general, $\alpha\vec{x}$ and $\vec{x}\alpha$ are different as are γG and $G\gamma$. The identity (from Proposition (1.5))

$$(\alpha\vec{v})(G\gamma) = \alpha(\vec{v}G)\gamma$$

tells us that $D^{m,n} = \mathrm{Hom}_{_D\mathsf{Vec}}(D^m, D^n)$, taking left D-space D^m to left D-space D^n, has its natural structure as a *right* D-space. (This is particularly important in discussion of dual spaces $V^* = \mathrm{Hom}_{_D\mathsf{Vec}}(V, D)$.)

(1.6). Problem.

(a) *Verify Proposition (1.5).*

(b) *In the situation of Proposition (1.5), let h be a second D-endomorphism from V to W and H the corresponding matrix representing h with respect to the bases $e_1, \dots, e_i, \dots, e_m$ and $f_1, \dots, f_j, \dots, f_n$. Prove that the D-endomorphism $g+h$ is represented by the matrix $G+H$ and so that, as abelian groups,* $\mathrm{Hom}_D(V,W)$ *and* $\mathrm{Mat}_{m,n}(D)$ *are isomorphic.*

(c) *Consider the special case $V = W$ and $e_i = f_i$, for all i. Prove that* $\mathrm{End}_D(V)$ *and* $\mathrm{Mat}_n(D)$ *are isomorphic as rings.*

Remark. *When rephrased appropriately, these remarks remain valid for arbitrary rings R with identity where V and W are free R-modules.*

Chapter 2

Basic group theory

2.1 Cosets and double cosets

If X and Y are subsets of G, then

$$XY = \{ xy \mid x \in X, y \in Y \}.$$

In particular, if $X \leq G$ and $Y = \{y\}$ then Xy is a *coset* of X in G (as is yX).[1] The number of distinct cosets Xy in G is the *index* of X in G, written $[G{:}X]$. (It is equal to the number of distinct cosets yX in G; see Problem (2.36).)

The basic result on cosets is:

(2.1). LEMMA. *Let $H \leq G$ and $x, y \in G$.*

(a) *$Hx \cap Hy$ is either $Hx = Hy$ or is empty.*

(b) *$xH \cap yH$ is either $xH = yH$ or is empty.*

(c) $|Hx| = |Hy| = |xH| = |yH|$.

(d) *For $H \leq G$, $Hx = Hy$ if and only if $yx^{-1} \in H$; and $xH = yH$ if and only if $xy^{-1} \in H$.* □

We immediately have the central result of finite group theory.

(2.2). THEOREM. (LAGRANGE'S THEOREM) *If $G \geq H$ then $|G| = [G{:}H]|H|$.* □

Also of interest are *double cosets* HxK for subgroups $H, K \leq G$. The double coset HxK is a union of various cosets Hxk of H and of various cosets hxK of K. As with cosets, two double cosets are either equal or disjoint, but unlike

[1] These days Hy is usually called a right coset and yH a left coset [Asc00], but in the past [Hll59] it was often the other way around.

cosets, double cosets can have varying orders. For instance $H1H = H$ but other double cosets HxH will likely contain more than one coset of H.

The set $\{ Hx \mid x \in G \}$ is often denoted $H\backslash G$ and correspondingly we write $\{ xH \mid x \in G \} = G/H$.[2] Similarly $\{ HxK \mid x \in G \} = H\backslash G/K$.

The product of cosets $HxHy$ is the disjoint union of the cosets $Hxhy$ and always contains the coset $Hxy = Hx1y$. The identity $HxHy = Hxy$ characterizes H as a normal subgroup of H, as discussed below.

Double coset multiplication will also be of interest. Especially $HxH.HyH$ is always the union of the double cosets $HxhyH$ and contains the double coset $HxyH$.

2.2 Quotients and isomorphism

Two groups A and B are isomorphic provided they are the same group only with names changed. That is, there is a bijection $\varphi\colon A \longrightarrow B$ with

$$a_1 \cdot a_2 = a_3 \text{ if and only if } \varphi(a_1) \cdot \varphi(a_2) = \varphi(a_3)\,.$$

If we are not concerned about the specific map φ, we write $A \simeq B$ to indicate that A and B are isomorphic.

Homomorphisms are more complicated. They are maps $\varphi\colon A \longrightarrow B$ still satisfying $\varphi(a_1)\varphi(a_2) = \varphi(a_1a_2)$, but they need no longer be injective or surjective. Surjectivity can be forced by replacing B with the image $\varphi(A)$. Still, we are not just renaming elements; we may be ignoring the distinctions between certain elements and identifying them with each other.

An arbitrary subgroup K of the group G is a *normal subgroup* of G if it is the kernel of some homomorphism. This is a qualitative definition. There are equivalent quantitative statements. In particulat the subgroup K is normal if, for every $x \in G$, $Kx = xK$. Thus K is normal in G precisely when $K^x = K$, for all $x \in G$.

It is helpful to realize that the subgroup K is normal in G precisely when $SK = KS$, for every subset S of G. In particular, when K is normal, the cosets $G/K = \{ Kx \mid x \in G \}$ do multiply naturally as sets:

$$KxKy = Kxy\,.$$

The group G/K (with this multiplication) is the *factor group* or *quotient group* of G by K.

For any homomorphism $\varphi\colon G \longrightarrow H$, the subgroup $\{ g \in G \mid \varphi(g) = 1_H \}$ is the *kernel* of φ, written $\ker\varphi$, and is normal in $\varphi(G)$.

(2.3). THEOREM. (FIRST ISOMORPHISM THEOREM) *Let $\varphi\colon G \longrightarrow H$ be a homomorphism of groups. Then the image $\varphi(H)$ is isomorphic to the factor group G/K, where $K = \ker\varphi = \{ g \in G \mid \varphi(g) = 1_H \}$. The isomorphism is given by $\varphi(g) \mapsto Kg$.* □

[2] Care must be taken with this, since for many it implies that H is a normal subgroup.

The First Isomorphism Theorem is critical in all group theory. It says that every homomorphism has a canonical model.

(2.4). Theorem. (Second Isomorphism Theorem) *Let $H \leq G$ and $K \trianglelefteq G$. Then $HK = KH \leq G$, and $HK/K \simeq H/H \cap K$ via the map $Kh \mapsto (H \cap K)h$.* □

(2.5). Theorem. (Third Isomorphism Theorem) *Let $G \trianglerighteq N$ and $G \geq K \geq N$. Then K is normal in G if and only if K/N is normal in G/N. In that case, $G/K \simeq (G/N)/(K/N)$.* □

The Second and Third Isomorphism Theorems discuss the lattices of subgroups and normal subgroups of a group. In particular the First and Third Isomorphism Theorems tell us that, for any surjective homomorphism $\varphi \colon G \longrightarrow H$, there is a natural bijection between the lattice of normal subgroups of H and the lattice of normal subgroups of G that contain $\ker \varphi$.

The First and Third Isomorphism Theorems are often invoked without mention, while use the Second often occasions remark.

The isomorphism theorems together allow us to do a great deal of group theory. If in a category C we have some counterpart to them, then C becomes a much more manageable place to work. In particular, concepts like simplicity and results of Jordan-Hölder type are more approachable. Concrete categories like ${}_R\mathsf{Mod}$ and Mod_R have them, and the definition of abelian categories is designed to guarantee them in an appropriate form.

2.3 Subgroups and action

Each g of the group G acts on G via conjugation:

$$(xy)^g = g^{-1}xyg = g^{-1}xgg^{-1}yg = x^g y^g \,.$$

Our conjugacy definition $x^g = g^{-1}xg$ makes conjugacy into a right action[3]:

$$(x^g)^h = h^{-1}(g^{-1}xg)h = (gh)^{-1}x(gh) = x^{gh} \,.$$

The induced automorphism ι_g is called an *inner automorphism* of G. The image of the representation $\iota \colon G \longrightarrow \mathrm{Aut}(G)$ $(= \mathrm{Aut}_{\mathsf{Grp}}(G))$ is the *inner automorphism group* $\mathrm{Inn}(G)$ and is normal in $\mathrm{Aut}(G)$. The kernel of the representation ι consists of everything fixed by conjugation. That is the *center* $\mathrm{Z}(G)$ of G:

$$\mathrm{Z}(G) = \{\, z \in G \mid zg = gz \text{ for all } g \in G \,\}\,;$$

so $\mathrm{Inn}(G)$ is isomorphic to $G/Z(G)$. The quotient $\mathrm{Aut}(G)/\mathrm{Inn}(G)$ is the *outer automorphism group* of G, denoted $\mathrm{Out}(G)$.

[3] Those who prefer left action will define the conjugate of x by g to be gxg^{-1}, which can be denoted ${}^g x$ giving ${}^h({}^g x) = {}^{hg}x$.

For any subset H of G, the *normalizer* of H in G denoted $\mathrm{N}_G(H)$, is

$$\mathrm{N}_G(H) = \{\, g \in G \mid H^g = H \,\}\,.$$

This is is the largest subgroup of G within which H is a normal subset. Therefore a subgroup H is normal in G precisely when $G = \mathrm{N}_G(H)$.

The group $\mathrm{N}_G(H)$ acts on H by conjugation, and the kernel of this action $\mathrm{C}_G(H)$ is the *centralizer* of H in G:

$$\mathrm{C}_G(H) = \{\, g \in G \mid gh = hg \text{ for all } h \in H \,\}\,.$$

The normalizer of a subgroup H always contains H itself, but the centralizer of H can be very small even if H is large. In any event, $\mathrm{C}_G(H) \trianglelefteq \mathrm{N}_G(H)$. If the set H contains the single element h, then $\mathrm{N}_G(H) = \mathrm{C}_G(h)$ (which we write in place of $\mathrm{C}_G(\{h\})$).

(2.6). LEMMA. *Let H be a subset of group G. Then the number of distinct conjugates of H in G is $[G{:}\mathrm{N}_G(H)]$.*

PROOF. Let g_i, for $i \in I$ be a complete set of representatives for distinct cosets of $\mathrm{N}_G(H)$ in G. Then for every $g \in G$ there are unique $n \in N$ and $i \in I$ with $g = ng_i$. Then $H^g = H^{ng_i} = H_{g_i}$, so the H^{g_i} give all conjugates. On the other hand $H^{g_i} = H^{g_j}$ gives $g_i(g_j)^{-1} \in \mathrm{N}_G(H)$, hence $i = j$. □

(2.7). PROPOSITION. (THE CLASS EQUATION) *Let g_i, for $1 \le i \le n$, be representatives for the conjugacy classes of noncentral elements in the finite group G. Then*

$$|G| = |\mathrm{Z}(G)| + \sum_{i=1}^{n} [G{:}\mathrm{C}_G(g_i)]\,.$$

PROOF. The group G is the disjoint union of its distinct conjugacy classes. A class contains one element if and only if it is in the center of G; that is, the union of the classes of size 1 has cardinality $|\mathrm{Z}(G)|$. The noncentral class containing each g_i has cardinality $[G{:}\mathrm{C}_G(g_i)]$ by Lemma (2.6). □

If A acts on G and H is a subset of G with $H^a = H$ for all $a \in A$ then H is *A-invariant.* (See Section 1.4.) For instance N is normal in G precisely when it is $\mathrm{Inn}(G)$-invariant. The nonidentity group G is a *simple group* provided 1 and G are the only normal subgroups of G, that is, the only $\mathrm{Inn}(G)$-invariant subgroups of G.

A subgroup of G is *characteristic* in G when it is $\mathrm{Aut}(G)$-invariant. Clearly this is a stronger requirement than normality. A group G that is $\mathrm{Aut}(G)$-simple is *characteristically simple.* That is, its only characteristic subgroups are 1 and G itself.

(2.8). PROPOSITION.

(a) *If P char Q char R, then P char R.*

(b) *If P char R and Q/P char R/P, then Q char R.*

(c) *If P char $Q \trianglelefteq R$, then $P \trianglelefteq R$.* □

If M and N be subgroups of G with $N \trianglelefteq M$, then the quotient group M/N is called a *section* of G. Sections form the basis of an important group theoretical technique call *internal representation theory.*

(2.9). Lemma. *Let M and N be A-invariant subgroups of G with $N \trianglelefteq M$. (This happens, for instance, if $A \leq \mathrm{N}_G(M)$ and N char M.) Then A acts on the section M/N by $(Nm)^a = Nm^a$ for $m \in M$ and $a \in A$.* □

2.4 Commutator theory

We have $gh = hg.g^{-1}h^{-1}gh$; so we define the *commutator*

$$[g, h] = g^{-1}h^{-1}gh$$

to gauge the extent to which g and h commute[4]. Obviously, they commute if and only if $[g, h] = 1$. We iterate by defining $[g, h, k] = [[g, h], k]$. For $H, K \leq G$, we set $[H, K] = \langle\, [h, k] \mid h \in H, k \in K \,\rangle$ (= $[K, H]$ by Lemma (2.11)(b)). Furthermore $[H, K, L] = [[H, K], L]$.

(2.10). Lemma. *If φ is a homomorphism from G to M then $[x, y]^\varphi = [x^\varphi, y^\varphi]$. Thus if $I \leq H \leq G$ and $J \leq K \leq G$, then $[I, J] \leq [H, K]$ and $\varphi([H, K]) = [\varphi(H), \varphi(K)]$. In particular, if H and K are characteristic subgroups of G, then $[H, K]$ is also characteristic in G.* □

(2.11). Lemma. *Let $x, y, z \in G$.*

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

(b) $[x, y] = (y^{-1})^x y$; $[x, y] = x^{-1}x^y$.

(c) $[xy, z] = [x, z]^y[y, z]$; $[x, yz] = [x, z][x, y]^z$.

(d) $[x, y^{-1}, z]^y[y, z^{-1}, x]^z[z, x^{-1}, y]^x = 1$. □

(2.12). Corollary.

Let $H, K \leq G$.

(a) $\mathrm{N}_G(H) \geq K$ *if and only if* $H \geq [H, K]$.

(b) $[K, H] = [H, K] \trianglelefteq \langle H, K\rangle$.

(c) $\langle H^G\rangle = H[H, G]$.

Proof. (a) We have $\mathrm{N}_G(H) \geq K$ if and only if $h^k \in H$, for all $h \in H$ and $k \in K$. This happens precisely when, for all $h \in H$ and $k \in K$, we have (using Lemma (2.11)(b)) $h^{-1}h^k = [h, k] \in H$, which is the case if and only if $H \geq [H, K]$.

[4] In places where left action is preferred to right action, the opposite notation $[g, h] = ghg^{-1}h^{-1}$ is often used.

(b) From Lemma (2.11)(a) we find $[K,H]=[H,K]$. For all $g,h\in H$ and $k\in K$, we have

$$[h,k]^g=[hg,k][g,k]^{-1}\in[H,K]$$

by Lemma (2.11)(e). Hence $H\le \mathrm{N}_G([H,K])$, and similarly for K.

(c) By (b) the subgroup $[H,G]$ is normal in G. In particular, $H[H,G]$ is a subgroup. As $h^g=h[h,g]$, we have

$$\langle H^G\rangle\le H[H,G]\le\langle\, h,h^g\mid h\in H, g\in G\,\rangle=\langle H^G\rangle\,. \qquad\square$$

(2.13). LEMMA. (THREE SUBGROUPS LEMMA) *Let $X,Y,Z\le G$. If $[X,Y,Z]=[Y,Z,X]=1$, then $[Z,X,Y]=1$.*

PROOF. By Lemma (2.11)(e), for all $x\in X$, $y\in Y$, and $z\in Z$, we have

$$[x,y^{-1},z]^y[y,z^{-1},x]^z[z,x^{-1},y]^x=1\,.$$

Since by hypothesis $[x,y^{-1},z]^y=[y,z^{-1},x]^z=1$, we have

$$[z,x^{-1},y]=[z,x^{-1},y]^x=1\,,$$

for all appropriate x,y,z. That is, every $y\in Y$ commutes with all $[z,x^{-1}]\in[Z,X]$. As these constitute a generating set for $[Z,X]$, in fact $[Z,X,Y]=1$, as desired. □

The proof of the next corollary makes use of a very standard group theoretic practice: a *bar convention*. In these situations, the image $\varphi(H)$ of the homomorphism φ from the group H is denoted $\bar{H}$ (or $\tilde{H}$ or $\hat{H}$, and so forth); then for each subset A of H, the subset $\varphi(A)$ of $\bar{H}$ is denoted $\bar{A}$ (respectively, $\tilde{H}$, $\hat{H}$).

(2.14). COROLLARY. *Let $X,Y,Z\le G$ and $N\trianglelefteq G$. If $[X,Y,Z]\le N$ and $[Y,Z,X]\le N$, then $[Z,X,Y]\le N$.*

PROOF. We make use of the bar convention $\bar{G}=G/N$. By hypothesis, $[\bar{X},\bar{Y},\bar{Z}]=\bar{1}$ and $[\bar{Y},\bar{Z},\bar{X}]=\bar{1}$. Therefore by the Three Subgroups Lemma (2.13), we have $[\bar{Z},\bar{X},\bar{Y}]=\bar{1}$. That is, $[Z,X,Y]\le N$. □

(2.15). COROLLARY. *If $[X,X]=X$ and $[Y,X,X]=1$, then $[Y,X]=1$.*

PROOF. We have, by assumption, $[X,Y,X]=[Y,X,X]=1$; so from the Three Subgroups Lemma (2.13), we get $1=[X,X,Y]=[[X,X],Y]=[X,Y]=[Y,X]$. □

The subgroup $[G,G]=G'$ is the *derived subgroup* of G, a normal, indeed characteristic, subgroup of G (see Lemma (2.10)).

(2.16). THEOREM. *If N is normal in G with G/N abelian, then $N\ge G'$. Conversely if $N\ge G'$, then N is normal in G and G/N is abelian.*

Proof. Let[5] $\bar{G} = G/N$. Thus $\bar{G}$ is abelian if and only if $[\bar{g}, \bar{h}] = \bar{1}$ (for all $\bar{g}, \bar{h} \in \bar{G}$) if and only if $[g, h] \in N$ (for all $g, h \in G$) if and only if $G' \leq N$. In the abelian group[6] $\tilde{G} = G/G'$, the subgroup $\tilde{N}$ is certainly normal; therefore N is normal in G. □

(2.17). Proposition. *If $G = \langle H^G \rangle$, then G/G' is isomorphic to $H/G' \cap H$. In particular, G/G' is a quotient of H/H'.*

Proof. We have $G = \langle H^G \rangle = H[H, G]$ by Corollary (2.12). Thus $[H, G]$ is a normal subgroup of G with $H' = [H, H] \leq [H, G] \leq [G, G] = G'$. By the Dedekind modular law we have $G' = (H \cap G')[H, G]$, so the result follows from the isomorphism theorems. □

A group G with $G = G'$ is a *perfect group.* By the theorem, G is perfect if and only if its only abelian quotient is 1. A group G is *quasisimple* if it is perfect and $G/\mathrm{Z}(G)$ is simple.

(2.18). Proposition. *A simple group is either perfect or cyclic of prime order.*

Proof. If G is not perfect, then $G > G'$. For simple G, this forces $G' = 1$; so G is abelian. Let $1 \neq g \in G$. Then $\langle g \rangle$ is normal in abelian $G = \langle g \rangle$. If $|g| = |G|$ was not prime then some $\langle g^e \rangle$ would be a nontrivial proper subgroup of G, also normal; so G is cyclic of prime order. □

2.5 Extensions

An arbitrary group G with $N \trianglelefteq G$ and $G/N \simeq H$ is an *extension* of N by H. Returning to unique factorization, we investigate the extent to which knowledge of N and H determines G in the extension

$$1 \longrightarrow N \longrightarrow G \longrightarrow H \longrightarrow 1$$

There are two basic issues:

(a) What is the action of H on N; that is, the homomorphism $\varphi\colon H \longrightarrow \mathrm{Aut}(N)$?

(b) Given an action of H on N, what are the possible extension types G?

Here is an important observation: given any action of a group H on a group N, there is always *at least one solution* G to the extension problem described above.

Let $\varphi\colon H \longrightarrow \mathrm{Aut}(N)$ be a homomorphism, and define on the set $H \times N$ the multiplication

$$(h_1, n_1)(h_2, n_2) = (h_1 h_2, n_1^{\varphi(h_2)} n_2)$$

[5] a first bar convention

[6] a second bar convention

We write $H \ltimes_\varphi N$ for this set endowed with this multiplication. It is called the *semidirect product* of N by H or the *split extension* of N by H.

(2.19). Theorem.

(a) *The semidirect product $M = H \ltimes_\varphi N$ is a group. The inverse of (h,n) is $(h^{-1}, (n^{-1})^{\varphi(h^{-1})})$.*

(b) *$H_0 = \{\, (h,1) \mid h \in H \,\}$ is a subgroup of M isomorphic to H.*

(c) *$N_0 = \{\, (1,n) \mid n \in H \,\}$ is a normal subgroup of M isomorphic to N.*

(d) *$H_0 \cap N_0 = 1$, $M = H_0 N_0$, and $M/N_0 \simeq H_0$.* □□

This is the "external" semidirect product. The corresponding "internal" semidirect product motivates the external definition.

(2.20). Corollary. *Let G have subgroups H and N with $H \le \mathrm{N}_G(N)$ and $H \cap N = 1$. Then $\langle H, N\rangle = HN \simeq H \ltimes_\varphi N$, where $\varphi\colon H \longrightarrow \mathrm{Aut}(N)$ is conjugation, given by $n^{\varphi(h)} = n^h = h^{-1}nh$.*

Proof. $h_1 n_1 \cdot h_2 n_2 = h_1(h_2 h_2^{-1}) n_1 h_2 n_2 = h_1 h_2\, n_1^{h_2} n_2$. □

Thus the semidirect product is a tool for realizing automorphisms of groups via inner automorphisms in larger groups. In it, the action of H on N is always given by conjugation. Generally if A acts on K, then we often write $[a,x]$ for $x^a x^{-1}$ and $[x,a]$ for $x^{-1}x^a$, since these are the natural commutators in the semidirect product $A \ltimes K$. Also when we write $[A,K]$ and $[K,A]$, the calculations are done within $A \ltimes K$.

As just done, we often drop the subscript and write

$$\langle H, N\rangle = HN \simeq H \ltimes N$$

when the intended action is understood. Since N is normal in HN, we have $HN = NH$. This suggests that we can also write internal semidirect products as $NH = N \rtimes H$. The multiplication is

$$n_1 h_1 \cdot n_2 h_2 = n_1 n_2^{h_1^{-1}}\, h_1 h_2\,.$$

This, in turn, suggests the definition for another external semidirect product: $N \rtimes_\varphi H$ is defined on the set $N \times H$ with multiplication given by

$$(n_1, h_1) \cdot (n_2, h_2) = (n_1 n_2^{\varphi(h_1)^{-1}}, h_1 h_2)\,,$$

where, as before $\varphi\colon H \longrightarrow \mathrm{Aut}(N)$ is a homomorphism.

(2.21). Lemma. *The subgroup isomorphisms $(h,1) \mapsto (1,h)$, for $h \in H$, and $(1,n) \mapsto (n,1)$, for $n \in N$, extend to an isomorphism of $H \ltimes_\varphi N$ and $N \rtimes_\varphi H$.* □

2.6 Solvable groups

Subnormality is the transitive extension of normality. The finite series[7]

$$G = G_0 \geq G_1 \geq \cdots \geq G_i \geq \cdots \geq G_n = 1$$

is a finite *subnormal series* if each G_{i+1} is normal in G_i. Similarly it is a *normal series* or *characteristic series* if all its members are normal and then characteristic in G. A finite subnormal series is more frequently called *finite series*. (We refrain from defining general series.) A subgroup H is *subnormal* in G if it is a member of some subnormal series starting at G. The *defect* of subnormal H is the length of the shortest subnormal series from G and finishing at H. For instance G has defect 0 in G while any normal subgroup N has defect 1.

The group G is *solvable* if it has a finite subnormal series

$$G = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_i \trianglerighteq \cdots \trianglerighteq G_n = 1$$

in which all the factors G_i/G_{i+1} are abelian. It is important that we not assume the series to be a composition series; an abelian group is always solvable, but only finite abelian groups have finite composition series.

Set $G = G^{(0)}$ and, for $i \geq 1$,

$$G^{(i)} = [G^{(i-1)}, G^{(i-1)}] .$$

Then $\{ G^{(i)} \mid i \geq 0 \}$ is the *derived series* of G. This is a characteristic series by Lemma (2.10). We set $G^{(\infty)} = \bigcup_{i \geq 0} G^{(i)}$, a characteristic and perfect subgroup of G.

We encountered $G^{(1)} = [G, G] = G'$, the derived subgroup of G earlier. By Theorem (2.16), each factor $G^{(i)}/G^{(i+1)}$ is abelian. Therefore, if $G^{(n)} = 1$, for some n, then G is solvable. Indeed we see next that a group is solvable if and only if there is an n with $G^{(n)} = 1$.

(2.22). Theorem. *Let $G = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_k$ with all factors G_{i-1}/G_i abelian. Then $G_i \geq G^{(i)}$ for all i.*

Proof. The case $k = 1$ is contained in Theorem (2.16). The general case follows by induction. □

(2.23). Corollary. *G is solvable if and only if $G^{(k)} = 1$ for some k.* □

Calculation of the derived series can thus be thought of as a "greedy algorithm" for verifying solvability. One consequence is that, if n is chosen minimal subject to $G^{(n)} = 1$, then any subnormal series testifying to the solvability of G must have length at least n; and there is such a series of length exactly n (namely, the derived series). This n is called the *derived length* of the solvable group G.

[7] We are actually describing *chains* rather than *series*, but the terminology is standard.

(2.24). Theorem.

(a) *Subgroups and quotient groups of solvable groups are solvable.*

(b) *For $N \trianglelefteq G$, if N and G/N are solvable then G is solvable.*

(c) *If A and B are normal and solvable in G, then AB is normal and solvable in G.*

Proof. If $H \leq G$, then $H^{(k)} \leq G^{(k)}$; so subgroups of solvable groups are solvable. Also, for any homomorphism, $\varphi(G') = \varphi(G)'$ (by Lemma (2.10)); so quotients of solvable groups are solvable.

We can stick together series with abelian factors for G/N and N to produce one for G. In particular this is possible when $G = AB$ and $N = A$. □

(2.25). Corollary. *If G is a finite group, then it has a unique maximal normal solvable subgroup.*

Proof. By the last part of the theorem, in a finite group all normal solvable subgroups generate a solvable normal subgroup, clearly the largest. □

The normal subgroup of the corollary is the *solvable radical* of G and is, in fact, characteristic in G.

2.7 Nilpotent groups

Let $\mathrm{L}_0(G) = G$ and, for $i \geq 1$,

$$\mathrm{L}_i(G) = [\mathrm{L}_{i-1}(G), G]\,.$$

By Lemma (2.10) this gives a characteristic series $\{\, \mathrm{L}_i(G) \mid i \geq 0 \,\}$, which is called the *lower central series* for G. By design each $\mathrm{L}_{i-1}(G)/\mathrm{L}_i(G)$ is central in $G/\mathrm{L}_i(G)$, but again it is not certain that there is some n with $\mathrm{L}_n(G) = 1$.

The lower central series is initially the same as the derived series. Indeed $\mathrm{L}_0(G) = G = G^{(0)}$ and $\mathrm{L}_1(G) = [G, G] = G^{(1)}$; but then they can diverge, since $\mathrm{L}_2(G) = [[G, G], G]$ will generically have $G^{(2)} = [[G, G], [G, G]]$ as a proper subgroup.

Again, let $\mathrm{Z}_0(G) = 1$, $\mathrm{Z}_1(G) = \mathrm{Z}(G)$, and, for $i \geq 1$,

$$\mathrm{Z}_{i+1}(G)/\mathrm{Z}_i(G) = Z(G/\mathrm{Z}_i(G))\,;$$

that is, $\mathrm{Z}_{i+1}(G)$ is the preimage in G of the center of $G/\mathrm{Z}_i(G)$. The series $\{\, \mathrm{Z}_i(G) \mid i \geq 0 \,\}$ is the *upper central series* for G. By Proposition (2.8)(b) each $\mathrm{Z}_{i+1}(G)$ is characteristic in G. In this case the subgroup 1 belongs to the series, but G might not.

(2.26). Theorem. *Let*

$$G = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_i \trianglerighteq \cdots \trianglerighteq G_n = 1$$

be a normal series in the group G with, for each $i \geq 0$,

$$G_i/G_{i+1} \leq Z(G/G_{i+1})\,.$$

(a) $\mathrm{L}_i(G) \leq G_i$, *hence* $\mathrm{L}_n(G) = 1$.

(b) $\mathrm{Z}_i(G) \geq G_{n-i}$, *hence* $\mathrm{Z}_n(G) = G$.

Proof. (a) Induct on i, with the case $i = 0$ being clear. If $\mathrm{L}_i(G) \leq G_i$ then

$$L_{i+1}(G) = [\mathrm{L}_i(G), G] \leq [G_i, G] \leq G_{i+1}\,.$$

(b) Induct on i, with the case $i = 0$ clear. Set $\bar{G} = G/\mathrm{Z}(G)$. Then by the definition of the upper central series,

$$\mathrm{Z}_i(\bar{G}) = \overline{\mathrm{Z}_{i+1}(G)} = \mathrm{Z}_{i+1}(G)/\mathrm{Z}(G)\,.$$

By the Third Isomorphism Theorem (2.5) we have $\bar{G}_i \trianglelefteq \bar{G}$ with

$$\bar{G}_i/\bar{G}_{i+1} \leq Z(\bar{G}/\bar{G}_{i+1})\,.$$

In particular, $\bar{G}_{n-1} = \bar{1}$ as $G_{n-1} \leq \mathrm{Z}(G)$. By induction $\bar{G}_{(n-1)-i} \leq \mathrm{Z}_i(\bar{G})$. Taking preimages in G, we find $G_{n-(i+1)} \leq \mathrm{Z}_{i+1}(G)$, as desired. □

A normal series as in the theorem is a *central series* for G. Thus G has a central series if and only if its lower central series reaches 1 in a finite number n of steps if and only if its upper central series reaches G in a finite number m of steps. In this case, the group G is *nilpotent*. The theorem goes on to tell us that the smallest such n and m are equal. This number $n = m$ is the *nilpotence class* of the group G.

We next have a nilpotent counterpart to Theorem (2.24), although part (b) should really be viewed as the failure within nilpotent groups of a basic property of solvable groups: an extension of a solvable group by a solvable group by a solvable group is solvable, while a (noncentral) extension of a nilpotent group by a nilpotent group is rarely nilpotent, as Sym(3) and Alt(4) testify. This also means that extracting (c) from (b) is more complicated.

(2.27). Theorem.

(a) *Subgroups and quotient groups of nilpotent groups are nilpotent.*

(b) *For $N \trianglelefteq G$, if $N \leq \mathrm{Z}(G)$ and G/N is nilpotent then G is nilpotent.*

(c) *If A and B are normal and nilpotent in G, then AB is normal and nilpotent in G.*

(d) *Let G be nilpotent and $1 \neq N \trianglelefteq G$. Then $\mathrm{Z}(G) \cap N \neq 1$.*

Proof. (a) By Lemma (2.10) always $\mathrm{L}_i(H) \leq \mathrm{L}_i(G)$ and $\varphi(\mathrm{L}_i(G)) = \mathrm{L}_i(\varphi(G))$.

(b) By (a) we may assume $N = \mathrm{Z}(G)$. But then the preimage of the upper central series for G/N is the upper central series for G.

(c) By induction on nilpotence class in $G/\mathrm{Z}(G)$, we have $AB\mathrm{Z}(G)/\mathrm{Z}(G)$ nilpotent. Then by (b) the preimage $AG\mathrm{Z}(G)$ is nilpotent as is its subgroup AB (by (a)).

(d) Choose the smallest i with $N \cap \mathrm{Z}_i(G) \neq 1$. As N is normal

$$[G, N \cap \mathrm{Z}_i(G)] \leq N \cap \mathrm{Z}_{i-1}(G) = 1\,,$$

and $1 \neq N \cap \mathrm{Z}_i(G) \leq \mathrm{Z}(G)$. □

We are used to doing induction in finite groups on order. Nipotent groups allow induction on class even when infinite. Similarly solvable groups allow induction on derived length even when infinite.

(2.28). Proposition. *Let G be a nilpotent group.*

(a) *$U < \mathrm{N}_G(U)$ for all $U < G$.*

(b) *U is subnormal in G for all $U \leq G$.*

(c) *$[N, G] < N$ for all nonidentity normal N.*

(d) *$\mathrm{Z}(G/N)$ is nontrivial for all proper normal N.*

Proof. Part (d) follows from Theorem (2.27)(a).
(a) Choose the largest i with $\mathrm{Z}_i(G) \leq U$. Then

$$[U, \mathrm{Z}_{i+1}(G)] \leq [\mathrm{Z}_{i+1}(G), G] \leq \mathrm{Z}_i(G) \leq U\,.$$

Thus $\mathrm{Z}_{i+1}(G) \leq \mathrm{N}_G(U)$ by Corollary (2.12), but $\mathrm{Z}_{i+1}(G)$ is not in U.
(a) By our proof of (a), the subnormal series

$$U < \mathrm{N}_G(U) < \mathrm{N}_G(\mathrm{N}_G(U)) < \cdots$$

reaches G in a number of steps at most the nilpotence class of G.
(c) Choose the largest i with $N \leq \mathrm{L}_i(G)$. Then

$$[N, G] \leq [\mathrm{L}_i(G), G] = \mathrm{L}_{i+1}(G)\,.$$

As N is not contained in $\mathrm{L}_{i+1}(G)$, we must have $[N, G] < N$. □

For finite groups G, each of these properties actually characterizes G as being nilpotent. (See Problem (2.41).)

(2.29). Corollary. *If G is a finite group, then it has a unique maximal normal nilpotent subgroup.* □

This subgroup, called the *Fitting subgroup* of G, is then clearly characteristic in G.

2.8 Finite p-groups and Sylow's First Theorem

Let p be a prime. The finite group G is a *p-group* if it has order a power of the prime p.[8]

[8] Sylow's First Theorem and Lagrange's Theorem imply that this is equivalent to requiring all elements to have order a power of p. This second version is a better definition in that it has content for infinite groups as well.

(2.30). LEMMA. *Let G be a finite p-group. Then G is nilpotent. Let $P \leq G$ with $|P| = p^a$ and $|G| = p^b$. Then for each integer c with $a \leq c \leq b$ there is a subgroup Q with $P \leq Q \leq G$ and $|Q| = p^c$.*

PROOF. Recall the Class Equation of Proposition (2.7):

$$|G| = |\mathrm{Z}(G)| + \sum_{i=1}^{n} [G{:}\mathrm{C}_G(g_i)]\,,$$

where the g_i are representatives for the distinct conjugacy classes of noncentral elements in G. Each $[G{:}\mathrm{C}_G(g_i)]$ is a multiple of p, as is $|G|$. Therefore $|\mathrm{Z}(G)|$ is also a multiple of p, which is to say that $\mathrm{Z}(G)$ is nontrivial. Therefore P is nilpotent by Theorem (2.27)(b) and induction. The rest then follows by Proposition (2.28)(a) since, for proper Q, the p-group $\mathrm{N}_G(Q)/Q$ has a nontrivial center. □

(2.31). THEOREM. (SYLOW'S FIRST THEOREM) *If the finite group G has order $|G| = p^a m$ with p prime, $a \in \mathbb{N}$, and $\gcd(p, m) = 1$, then G contains subgroups of order p^a and index m.*

PROOF. Consider again the Class Equation or Proposition (2.7):

$$|G| = |\mathrm{Z}(G)| + \sum_{i=1}^{n} [G{:}\mathrm{C}_G(g_i)]\,.$$

If any of the indices $[G{:}\mathrm{C}_G(g_i)]$ is not a multiple of p, then $\mathrm{C}_G(g_i)$ is a proper subgroup of G whose order is a multiple of p^a, and we are done by induction.

Thus we can assume that $\sum_{i=1}^{n}[G{:}\mathrm{C}_G(g_i)]$ is a multiple of p, and hence (as in the previous lemma) $\mathrm{Z}(G)$ has order a multiple of p. By the structure theory of finite abelian groups (see Theorem (1.4)) $\mathrm{Z}(G)$ contains a subgroup Z of order p. Now we are done by induction in G/Z. □

Such subgroups are *Sylow subgroups* or *Sylow p-subgroups* or *p-Sylow subgroups* of G.

(2.32). LEMMA. *Let P be a Sylow subgroup of the group G. Then $\mathrm{N}_G(\mathrm{N}_G(P)) = \mathrm{N}_G(P)$.*

PROOF. The Sylow subgroup P is not only normal in $\mathrm{N}_G(P)$ but also characteristic. Therefore $P \trianglelefteq \mathrm{N}_G(\mathrm{N}_G(P))$, forcing $\mathrm{N}_G(\mathrm{N}_G(P)) = \mathrm{N}_G(P)$. □

2.9 Direct products and sums

For any two groups H and N there is always the trivial map $\varphi\colon H \longrightarrow \mathrm{Aut}(N)$ that sends each element of H to the trivial automorphism of N. In that case the semidirect product is actually the *direct product*

$$H \ltimes_\varphi N = H \times N\,.$$

In this case $[H,N]=1$, and N also acts trivially on the normal subgroup H:

$$H \times N \simeq N \ltimes H = N \times H\,.$$

Once we have defined the direct product of two groups, we immediately have the direct product of any finite set of groups via

$$H_1 \times H_2 \times H_3 = (H_1 \times H_2) \times H_3 \simeq H_1 \times (H_2 \times H_3)\,.$$

More generally, let G_i, $i \in I$, be a set of groups. The *direct product* of the G_i, written $\bigotimes_{i\in I} G_i$ is the group consisting of those sequences $(g_i)_{i\in I}$ with multiplication is defined pointwise:

$$(g_i)_{i\in I}(h_i)_{i\in I} = (g_ih_i)_{i\in I}\,.$$

For $|I|=2$, indeed for finite I, this is just the direct product defined above.

The direct product is sometimes referred to as the *external direct product* since it is a group constructed from the collection of groups G_i, initially external to the product. The next result produces a direct product from inside a group:

(2.33). Theorem. (Chinese Remainder Theorem) *Let N_i, for $i \in I$, be normal subgroups of the group G with $G = \langle\, N_i \mid i \in I\,\rangle$, and set $G_i = G/N_i$. Then the map*

$$g \longrightarrow (N_ig)_{i\in I}$$

is a homomorphism of G into the direct product $\bigotimes_{i\in I} G_i$ with kernel $\bigcap_{i\in I} N_i$.

Proof. The map $g \longrightarrow (N_ig)_{i\in I}$ is certainly a homomorphism, since each of its projections $\pi_i \colon g \mapsto N_ig$ is the natural factor map $\pi_i \colon G \longrightarrow G/N_i = G_i$. The kernel is then

$$\{\, g \in G \mid N_ig = N_i,\ \text{for all } i \in I\,\} = \bigcap_{i\in I} N_i\,. \qquad \square$$

Indeed, more is true. For each j, the described coordinate projection $\pi_j \colon g \mapsto (N_ig)_{i\in I} \mapsto N_jg$ is onto G_j. In general, a subgroup of $H \le \bigotimes_{i\in I} G_i$ for which each projection $\pi_i(H)$ is onto G_i, is a *subdirect product* of the G_i. An important example of a subdirect product is the the embedding of the group G on the diagonal of $G \times G$ via $g \mapsto (g,g)$.

The Chinese Remainder Theorem reveals the direct product as the categorical product in Grp relative to the various projections π_i. The categorical coproduct in Grp is the free product. This shall not be of much direct interest to us, but it does have a quotient that is important.

For each i we have the natural injection ι_i of G_i into $\bigotimes_{i\in I} G_i$ that takes $g \in G_j$ to the I-tuple $(g_i)_{i\in I}$ with $g_j = g$ and $g_i = 1_{G_i}$ for $i \neq j$. The group $\bigoplus_{i\in I} G_i$ is the subgroup $\langle\, \iota_i(G_i) \mid i \in I\,\rangle$ of $\bigotimes_{i\in I} G_i$. This is then the normal subgroup of $\bigotimes_{i\in I} G_i$ consisting of those $(g_i)_{i\in I}$ with $g_i = 1_{G_i}$ for all but a finite number of i. In particular if I is finite, then $\bigoplus_{i\in I} G_i = \bigotimes_{i\in I} G_i$, but if I is infinite then the containment is proper.

The group $\bigoplus_{i\in I} G_i$ is a kind of coproduct in Grp in that is the natural quotient of the coproduct subject to the additional relations

$$[\iota_i(G_i), \iota_j(G_j)] = 1 \text{ for all } i \neq j\,.$$

In particular, in the category of abelian groups AbGrp this is the coproduct. For this reason, we call $\bigoplus_{i\in I} G_i$ the *direct sum* of the G_i.

The reason we make these distinctions here is that, just as with the Chinese Remainder Theorem and direct products, there is a natural context in which we encounter direct sums.

(2.34). THEOREM. *If, in the group G, the subgroups G_i, for $i \in I$, normalize each other and, for all $i \in I$ satisfy*

$$G_i \cap \langle\, G_j \mid j \neq i \,\rangle = 1\,,$$

then the subgroup $\langle\, G_i \mid i \in I \,\rangle$ is isomorphic to the direct sum $\bigoplus_{i\in I} G_i$.

PROOF. The isomorphism from $\bigoplus_{i\in I} G_i$ to G is given by

$$(g_i)_{i\in I} \mapsto \prod_{i\in I} g_i\,.$$

This is well-defined since all but a finite number of the g_i are 1_G and for $i \neq j$ the elements g_i and g_j of G commute. □

This theorem explains why the direct sum is often called the *internal direct product.* There are various different terms for these two products, and that can be confusing. The (external) direct product is also the *Cartesian product* and the *unrestricted direct product* while the direct sum (internal direct product) is also the *restricted direct product* and sometimes, even, the *direct product.* We shall only refer to the direct sum as the direct product when I is finite so they are equal, as in the next theorem.

(2.35). THEOREM. *A finite group is nilpotent if and only if it is the direct product of its Sylow subgroups.*

PROOF. A finite group that is the direct product of p-groups is nilpotent by Theorem (2.27) and Lemma (2.30).

Now let G be finite and nilpotent. For each Sylow subgroup P, we have $\mathrm{N}_G(\mathrm{N}_G(P)) = \mathrm{N}_G(P)$ by Lemma (2.32). As G is nilpotent, this forces $\mathrm{N}_G(P) = P$ by Proposition (2.28)(a). That is, every Sylow subgroup of G is normal. Let P_i, for $1 \leq i \leq n$, be the (unique) Sylow p_i-subgroup (with $p_i \neq p_j$, for $i \neq j$). For each i the subgroup $\langle\, P_j \mid j \neq i \,\rangle = \prod_{j\neq i} P_j$ has p_i' order and so is disjoint from the p_i-group P_i. Therefore G is the direct sum, hence product, of the P_i by Theorem (2.34). □

2.10 Problems

(2.36). Problem. *Prove that the inverse map $g \mapsto g^{-1}$ is an anti-automorphism of every group G that, for each subgroup H of G, induces a bijection between the set of right cosets of H and the set of left cosets of H.*

(2.37). Problem. *Let $(G, \cdot)$ be a group. Let $a \in G$. Define, on the set G, a new multiplication $\circ$ by*

$$x \circ y = x \cdot a \cdot y\,.$$

Prove that there is a bijection $\varphi\colon G \longrightarrow G$ with $\varphi(g) \circ \varphi(h) = \varphi(g \cdot h)\,.$ for all $g, h \in G$. What is the identity element of $(G, \circ)$?

(2.38). Problem. (Reidermeister Quadrangle Condition) *Let A be a Latin square with entries from the set X. That is, A is an $X \times X$ array in which every element of X occurs exactly once in each row and exactly once in each column.*

Prove that the rows and columns of A can be relabeled to make it into the multiplication table of some group if and only if, for all $a, b, c \in X$, whenever the pattern

$$\begin{bmatrix} & \vdots & & \vdots & \\ \cdots & a & \cdots & b & \cdots \\ & \vdots & & \vdots & \\ \cdots & c & \cdots & d & \cdots \\ & \vdots & & \vdots & \end{bmatrix}$$

occurs in rows i, j and columns m, n and

$$\begin{bmatrix} & \vdots & & \vdots & \\ \cdots & a & \cdots & b & \cdots \\ & \vdots & & \vdots & \\ \cdots & c & \cdots & d' & \cdots \\ & \vdots & & \vdots & \end{bmatrix}$$

is in rows i', j' and columns m', n', then $d = d'$.
(Let $\mathcal{C}$ be the set all triples $(r, c, e) \in X^3$ for which e is the entry to be found in the cell of A at the intersection of row r and column c. Then the hypothesis is: for all $i, j, m, n, i', j', m', n' \in X$,

$$(i, m, a), (i, n, b), (j, m, c), (j, n, d), (i', m', a), (i', n', b), (j', m', c), (j', n', d') \in \mathcal{C}$$

implies $d = d'$.)

(2.39). Problem. *Let $B \leq G$. Prove that $BxB.ByB \supseteq BxyB$, for all $x, y \in G$, and that equality always holds if and only if B is normal in G.*

(2.40). Problem.

(a) *Prove that if G is solvable then $U' < U$ for all $1 \neq U \leq G$.*

(b) *Let G be finite. Prove that G is solvable if and only if $U' < U$ for all $1 \neq U \leq G$.*

(2.41). Problem. *Let G be a finite group with any one of the following properties:*

(a) $U < \mathrm{N}_G(U)$ *for all* $U < G$.

(b) U *is subnormal in* G *for all* $U \leq G$.

(c) $[N, G] < N$ *for all nonidentity normal* N.

(d) $\mathrm{Z}(G/N)$ *is nontrivial for all proper normal* N.

Prove that G *is nilpotent. (That is, for finite groups each of the properties of Proposition (2.28) characterizes nilpotent groups. This is false for each when we include infinite groups.)*

(2.42). Problem. *Prove that in arbitrary* G *we have* $L_i(G) \geq G^{(i)}$ *for all* $i \geq 0$.

(2.43). Problem. *Let* P *be nilpotent of class at most* 2. *(That is,* $P' \leq \mathrm{Z}(P)$*). Set* $\bar{P} = P/\mathrm{Z}(P)$. *Prove*

(a) $[a, b][a, c] = [a, bc]$.

(b) $[a, c][b, c] = [ab, c]$

(c) $a^k b^k = (ab)^k [a, b]^{\binom{k}{2}}$.

(d) *The map* $f(\bar{x}, \bar{y}) = [x, y]$ *is a well-defined biadditive map from the abelian group* $\bar{P}$ *into the abelian group* $\mathrm{Z}(P)$.

(e) *If* $\bar{P}$ *and* $\mathrm{Z}(P)$ *are elementary abelian* p*-groups, then the map* $q(\bar{x}) = x^p$ *is a well-defined map from the* $\mathbb{Z}_p$*-space* $\bar{P}$ *to the* $\mathbb{Z}_p$*-space* $\mathrm{Z}(P)$ *that is linear for* p *odd and for* $p = 2$ *satisfies* $f(\bar{x}, \bar{y}) = q(\bar{x} + \bar{y}) + q(\bar{x}) + q(\bar{y})$.

(2.44). Problem. *Prove that in a finite group, any subnormal nilpotent subgroup is contained in the Fitting subgroup.*

(2.45). Problem. *Let* G *contain the normal abelian subgroups* N_i *for* $1 \leq i \leq n$. *Set* $N = \langle N_i \mid 1 \leq i \leq n \rangle$, *which is nilpotent by Theorem (2.27)(c). Prove that* N *has nilpotence class at most* n.

(2.46). Problem.

(a) *Let* $G = \bigoplus_{i \in I} G_i$ *be the direct sum of the nonabelian simple groups* G_i. *Show that* H *is subnormal in* G *if and only if* $H = \bigoplus_{j \in J} G_j$, *for some subset* J *of* I.

Remark. *This result is not true if two of the simple groups* G_i *are abelian and isomorphic.*

(b) *Prove that direct sum of isomorphic simple groups is characteristically simple.*

Chapter 3

Permutation and Linear Groups

Groups are often encountered as permutation groups or linear groups. These are often the right places to look for factorization (reduction) methods.

3.1 Permutation groups

3.1.1 Basics

Recall that for the set Ω, the *symmetric group* $\mathrm{Sym}(\Omega)$ is the group $\mathrm{Aut}_{\mathsf{Set}}(\Omega)$ of all bijections of Ω with itself. For finite $|\Omega| = n$, the group $\mathrm{Sym}(\Omega)$ is isomorphic to $\mathrm{Sym}(\{1, 2, \ldots, n\})$, which we usually write as $\mathrm{Sym}(n)$.

(3.1). Lemma. *For n a positive integer,* $|\mathrm{Sym}(n)| = n!$. □

If $\varphi\colon G \longrightarrow \mathrm{Sym}(\Omega)$ is a *permutation representation* (that is, a Set-representation), then we say that Ω is a *G-space.* We write $\ker_G(\Omega)$ for the kernel of the representation φ, and the representation φ is *faithful* if its kernel is trivial. If $G \le \mathrm{Sym}(\Omega)$ then we say that (G, Ω) is a *permutation group.* We also abuse this terminology by extending it to include faithful permutation representations $\varphi\colon G \longrightarrow \mathrm{Sym}(\Omega)$.

If f is a bijection (that is, Set-isomorphism) of the two sets Ω and Δ, then we have the induced isomorphism f^* of $\mathrm{Sym}(\Omega) = \mathrm{Aut}_{\mathsf{Set}}(\Omega)$ and $\mathrm{Sym}(\Delta) = \mathrm{Aut}_{\mathsf{Set}}(\Delta)$, as in Section 1.4

$$\begin{tikzcd}
\Omega \arrow[r, "a"] \arrow[d, "f"] & \Omega \arrow[d, "f"] \\
\Delta \arrow[r, "a^*"] & \Delta
\end{tikzcd}$$

with

$$a \mapsto a^{f^*} = a^* = f^{-1}af\,.$$

For the permutation representation $\rho_\Omega\colon G \longrightarrow \mathrm{Sym}(\Omega)$ this provides us with the equivalent representation $\rho_\Delta\colon G \longrightarrow \mathrm{Sym}(\Delta)$ given by $\rho_\Delta = \rho_\Omega f^*$:

$$\begin{tikzcd}
G \arrow[r, "\rho_\Omega"] \arrow[dr, "\rho_\Delta"] & \mathrm{Sym}(\Omega) \arrow[d, "f^*"] \\
 & \mathrm{Sym}(\Delta)
\end{tikzcd}$$

In this case, Ω and Δ are said to be *isomorphic G-spaces*. If further (G, Ω) and (H, Δ) are permutation groups, and there is a group isomorphism $\varphi\colon G \longrightarrow H$ for which the following diagram commutes:

$$\begin{tikzcd}
G \arrow[r, "\rho_\Omega"] \arrow[d, "\varphi"] & \mathrm{Sym}(\Omega) \arrow[d, "f^*"] \\
G \arrow[r, "\rho_\Delta"] & \mathrm{Sym}(\Delta)
\end{tikzcd}$$

then the semi-equivalence (φ, f) is a *permutation isomorphism* of (G, Ω) and (H, Δ).

Permutation representations are important for at least two reasons. They are relatively easy to work with and calculate in, and every group has a faithful representation as a permutation group. We have discussed G acting on itself by conjugation, but a more elementary action exists.

(3.2). Theorem. (Cayley's Theorem) *Every group G is faithfully represented in* $\mathrm{Sym}(G)$ *via right translation:*

$$g \mapsto \rho(g) \quad \textit{where} \quad x^{\rho(g)} = xg \quad \textit{for } x \in G\,.$$

Proof. We have

$$(x^{\rho(g)})^{\rho(h)} = (xg)h = x(gh) = x^{\rho(gh)}\,.$$

The action is faithful since $1^{\rho(g)} = g$ implies that $\rho(g)$ is nontrivial when g is not the identity. □

The associated representation $\rho\colon G \longrightarrow \mathrm{Sym}(G)$ is called the *right regular representation.*

3.1.2 Transitivity

For ω in the G-space Ω, we set

$$\omega^G = \{\, \omega^g \mid g \in G \,\}\,,$$

the *orbit* (or *G-orbit*) of ω for this action. If the only orbit is Ω, then we say that G is *transitive* on Ω. Otherwise G is *intransitive*.

In most situations, questions about permutation groups can be reduced to questions about transitive permutation groups.

(3.3). PROPOSITION. *If G is a permutation group on Ω with orbits Ω_i, for $i \in I$, then G is isomorphic to a subdirect product of the groups $G_i = G/\ker_G(\Omega_i)$, each acting faithfully and transitively on Ω_i.*

PROOF. This follows from the Chinese Remainder Theorem (2.33). □

The subgroup

$$G_\omega = \{\, g \in G \mid \omega^g = \omega \,\}\,,$$

is the *stabilizer* of ω in G. We sometimes also use the notation $\mathrm{Stab}_G(\omega)$.

If Δ is a subset of Ω, then

$$G_\Delta = \mathrm{Stab}_G(\Delta) = \{\, g \in G \mid \Delta^g = \Delta \,\}\,,$$

the *global stabilizer* of Δ, while

$$G_{[\Delta]} = \bigcap_{\delta \in \Delta} G_\delta = \{\, g \in G \mid \delta^g = \delta,\ \text{all } \delta \in \Delta \,\}$$

is the *pointwise stablizer* of Δ. It probably would be better to denote (as some do) the global stabilizer of Δ by $\mathrm{N}_G(\Delta)$ and the pointwise stabilizer by $\mathrm{C}_G(\Delta)$.

A permutation representation $\varphi\colon G \longrightarrow \mathrm{Sym}(\Omega)$ is *semiregular* if $G_\omega = 1$, for all $\omega \in \Omega$. (In particular, it is faithful.) It is *regular* if it is semiregular and transitive (as in the right regular representation).

Just as factor groups give us canonical models for homomorphic images, so coset spaces give us canonical models for transitive permutation spaces.

For $H \leq G$, consider the coset space $H\backslash G = \{\, Hx \mid x \in G \,\}$,[1] which is naturally a G-space under the action of right translation:

$$\rho_H\colon G \longrightarrow \mathrm{Sym}(H\backslash G) \text{ given by } (Hx)^{\rho_H(g)} = Hxg\,,$$

so that

$$((Hx)^{\rho_H(g)})^{\rho_H(k)} = (Hxg)k = Hx(gk) = Hx^{\rho_H(gk)}\,.$$

Of course, the right regular representation is ρ_{1_G}.

(3.4). THEOREM. *Let G be transitive on Ω. For a fixed $\omega \in \Omega$, set $H = G_\omega$ and $H\backslash G = \{\, Hg \mid g \in G \,\}$. Then the set Ω and the coset space $H\backslash G$ are isomorphic as G-spaces. In particular $|\omega^G| = [G{:}G_\omega]$.*

PROOF. For hx in the coset Hx we have $\omega^{hx} = (\omega^h)^x = \omega^x$. Conversely, if $\omega^x = \omega^y$, then $yx^{-1} \in H$ and $Hx = H(yx^{-1})x = Hy$. Therefore the map $f\colon \Omega \longrightarrow H\backslash G$ given by

$$\alpha \xrightarrow{f} Hx \iff \alpha = \omega^x$$

[1] This can be viewed as the set of H-orbits for in the *left* regular representation of G; see Problem (3.25).

is a well-defined bijection. If g is in G then $\alpha^g = (\omega^x)^g = \omega^{xg}$, so we have the commutative diagram

$$\begin{tikzcd}
\alpha \arrow[r, "g^{\rho_\Omega}"] \arrow[d, "f"] & \alpha^g \arrow[d, "f"] \\
Hx \arrow[r, "g^{\rho_H}"] & Hxg
\end{tikzcd}$$

That is, $\rho_H = \rho_\Omega f^*$; and the two G-spaces are isomorphic, as claimed.

As G is transitive on Ω,

$$|\omega^G| = |\Omega| = |H\backslash G| = [G{:}G_\omega]\,. \qquad \square$$

(3.5). Lemma. *If $\alpha^h = \beta$, then $(\alpha^g)^{h^g} = \beta^g$.*

Proof. $(\alpha^g)^{h^g} = (\alpha^g)^{g^{-1}hg} = (\alpha^h)^g = \beta^g$. That is, the diagram

$$\begin{tikzcd}
\alpha \arrow[r, "h"] \arrow[d, "g"] & \beta \arrow[d, "g"] \\
\alpha^g \arrow[r, "h^g"] & \beta^g
\end{tikzcd}$$

commutes. □

So if h has cycle representation

$$h = \dots(\dots,\alpha,\beta,\dots)\dots$$

then h^g has cycle representation

$$h^g = \dots(\dots,\alpha^g,\beta^g,\dots)\dots$$

The permutation $h \in \mathrm{Sym}(\Omega)$ has *cycle type* $1^{a_1}2^{a_2}\dots i^{a_i}\dots$, where a_i is the number of orbits of length i that h has in Ω. Those terms with $a_i = 0$ are always deleted. The term 1^{a_1} is also usually deleted as well (although for infinite Ω this can cause confusion).

The lemma tells us that conjugacy in $\mathrm{Sym}(\Omega)$ preserves cycle type. Indeed, two elements of $\mathrm{Sym}(\Omega)$ are conjugate if and only if they have the same cycle type.

(3.6). Corollary. *If Ω is a G-space and $\omega \in \Omega$, then $(G_\omega)^g = G_{\omega^g}$ for all $g \in G$.* □

If H is a subgroup of G, then the *core* $\ker_G(H)$ of H in G is the largest normal subgroup of G contained in H.

(3.7). Corollary. *If Ω is a transitive G-space and $\omega \in \Omega$, then the core of G_ω is* $\ker_G(\Omega)$. □

(3.8). Lemma. (A Frattini argument) *Let G be transitive on Ω and $\omega \in \Omega$. For $N \subseteq G$, we have $\omega^N = \Omega$ if and only if $G = G_\omega N$.*

Proof. Assume $\omega^N = \Omega$. For $g \in G$, we have $\omega^g = \alpha = \omega^n$, for some $n \in N$. Therefore $gn^{-1} \in G_\omega$, and $g = (gn^{-1})n \in G_\omega N$. Conversely, if $G = G_\omega N$, then for any $\alpha \in \Omega$, there is a $g \in G$ with $\omega^g = \alpha$. Then $g = g_1 n_1$, for $g \in G_\omega$ and $n_1 \in N$, so that $\omega^{n_1} = \omega^{g_1 n_1} = \omega^g = \alpha$. Hence $\omega^N = \Omega$. □

If Ω is a G-space, then so is each Ω^k, with action in each coordinate. We are particularly interested in the case $k = 2$.

Any subset of Ω^2 is a relation and any union of orbits of G on Ω^2 is an *invariant relation.* An orbit of G on Ω^2 is an *orbital* of G. We may think of an invariant relation as a directed graph whose automorphism group contains G. We may even think of it as an edge-colored graph, with the edges from orbital $\mathcal{O}_i$ colored with i.

To each orbital $\mathcal{O}$, there is associated a *paired orbital* $\mathcal{O}^\cup$ given by

$$(a, b) \in \mathcal{O} \qquad \Longleftrightarrow \qquad (b, a) \in \mathcal{O}^\cup .$$

Every G-invariant relation Γ has an underlying G-invariant undirected graph $\Gamma \cup \Gamma^\cup$, and any G-invariant relation $\Gamma = \Gamma^\cup$ is naturally an undirected graph.

The *rank* of transitive G on Ω is the number of G-orbits on Ω^2. The diagonal

$$\Delta = \{\, (\alpha, \alpha) \mid \alpha \in \Omega \,\}$$

is always an orbital, so G has rank 1 on Ω if and only if $|\Omega| = 1$. The transitive group (G, Ω) has rank 2 precisely when $|\Omega| > 1$ and G is transitive on

$$\Omega^2 \setminus \Delta = \begin{bmatrix} \Omega \\ 2 \end{bmatrix} .$$

That is, for all pairs $\alpha \neq \beta$ and $\alpha' \neq \beta'$ from Ω, there is a $g \in G$ with $\alpha.g = \alpha'$ and $\beta.g = \beta'$. In this case G is said to be 2*-transitive* or *doubly transitive* on Ω.

(3.9). Proposition.

(a) *If G is transitive on Ω then the rank of G on Ω is the number of orbits of of G_ω on Ω.*

(b) *If G is transitive on Ω then the rank of G on Ω is the number of $H - H$ double cosets in $H \backslash G / H$.*

Proof. By Theorem (3.4) these are equivalent.

For (a), let Σ be an orbit of of G_ω in Ω. Then there is a unique orbital $\mathcal{O}$ with

$$\mathcal{O} \cap \{\, (\alpha, \omega) \mid \alpha \in \Omega \,\} = \{\, (\alpha, \omega) \mid \alpha \in \Sigma \,\} .$$

This gives a bijection between orbitals and orbits of G_ω. □

(3.10). Corollary. *G is 2-transitive on Ω if and only if G is transitive on Ω and G_ω is transitive on $\Omega \setminus \omega$.* □

More generally, $G \leq \mathrm{Sym}(\Omega)$ is said to be *k-transitive* on Ω if it is transitive on $\left[{\Omega \atop k}\right]$. If G is k-transitive on Ω, then it is also k'-transitive, for all $k' \leq k$.

Additionally, G is *k-homogeneous* on Ω if it is transitive on $\binom{\Omega}{k}$. Of course k-transitivity implies k-homogeneity, but the converse is far from true. Indeed, for $|\Omega| = k$ finite, the only k-transitive subgroup of $\mathrm{Sym}(\Omega)$ is $\mathrm{Sym}(\Omega)$ itself, while every subgroup of $\mathrm{Sym}(\Omega)$ is k-homogeneous. In particular, an arbitrary G that is k-homogeneous may not be $(k-1)$-homogeneous.

3.1.3 Primitivity

In the last section we discussed G-invariant relations—graphs. Now we specialize to the case of G-invariant equivalence relations, where the corresponding graphs are disjoint unions of complete graphs.

There are three obvious invariant equivalence relations $\sim$ on Ω:

(i) $\alpha \sim \omega$ for all $\alpha, \omega \in \Omega$, with the single equivalence class Ω;

(ii) $\alpha \sim \omega$ if and only if $\alpha = \omega$, with $|\Omega|$ equivalence classes, all of size 1;

(iii) $\alpha \sim \omega$ if and only if there is a $g \in G$ with $\alpha^g = \omega$, with equivalence classes the orbits of Ω.

The first two are the *trivial equivalence relations.*

We say that G is *primitive* in its action on Ω if the only invariant equivalence relations are the trivial ones. If G is primitive on Ω, then either G is transitive on Ω or G is trivial on Ω of size 2. Therefore one usually requires of primitive groups that they be transitive.

If G is transitive on Ω but not primitive, then it is *imprimitive.* A *block of imprimitivity* for G acting on Ω is an equivalence class for some nontrivial invariant equivalence relations.

(3.11). THEOREM. *If G is 2-transitive on Ω, then G is primitive on Ω.*

PROOF. Suppose Σ is a block with $\alpha, \beta \in \Sigma$ with $\alpha \neq \beta$. For every $\gamma \neq \alpha$ there is a $g \in G_\alpha$ with $\beta^g = \gamma$. Then

$$\gamma \in \{\alpha, \gamma\} = \{\alpha, \beta\}^g \subseteq \Sigma^g = \Sigma\,,$$

as $\alpha \in \Sigma \cap \Sigma^g$. Therefore $\Sigma = \Omega$ and G is primitive. □

(3.12). PROPOSITION. *Let G be transitive on Ω with $\omega \in \Omega$.*

(a) *G is primitive on Ω if and only if G_ω is maximal in G.*

(b) *The map $N \longrightarrow \omega^N$ gives an isomorphism of the lattice of subgroups N of G with $G_\omega \leq N \leq G$ with the lattice of blocks of imprimitivity from ω upto Ω.*

PROOF. (b) The bijection is given by

$$N \longrightarrow \omega^N \text{ and its inverse } \Delta \longrightarrow \mathrm{Stab}_G(\Delta)\,.$$

Let $G_\omega \leq N \leq G$, and set $\Delta = \omega^N$. If $\beta \in \Delta \cap \Delta^g$ then there are $n_1, n_2 \in N$ with $\omega^{n_1} = \beta = (\omega^{n_2})^g$. Therefore $\omega^{n_2 g n_1^{-1}} = \omega$ and $n_2 g n_1^{-1} \in G_\omega \leq N$. Thus $g \in N$ and $\Delta^g = (\omega^N)^g = \alpha^N = \Delta$. That is, Δ is a block of imprimitivity. Furthermore, as N is transitive on Δ we have $\mathrm{Stab}_N(G) = G_\omega N = N$ by the Frattini argument of Lemma (3.8).

Let $\omega \in \Delta$, a block of imprimitivity for G on Ω. For each $\beta \in \Delta$ there is a $g \in G$ with $\omega^g = \beta$ as G is transitive. Then $\beta \in \Delta \cap \Delta^g$, so $\Delta = \Delta^g$ and $g \in N = \mathrm{Stab}_G(\Delta)$ with $\Delta = \omega^N$.

(a) The subgroup G_α is maximal in G if and only if there are no subgroups in between G_α and G if and only if there are no blocks in between α and Ω if and only if G is primitive on Ω. □

(3.13). PROPOSITION. *Let G be transitive on Ω and N a normal subgroup of G. The N-orbits in Ω are blocks of imprimitivity for G, and G/N is transitive on the set of these blocks.*

PROOF. We have $G_\alpha \leq G_\alpha N \leq G$ and $\alpha^{G_\alpha N} = \alpha^N$, so this follows from the previous proposition.

□

In a sense, every block of imprimitivity arises in this manner. The block set for an imprimitive group is an *equipartition* of Ω, a partition into parts of equal size. The global stabilizer in $\mathrm{Sym}(\Omega)$ of this equipartition then contains G and has these blocks as its orbits under a normal subgroup.

(3.14). COROLLARY. *If G is primitive on Ω and N is normal in G then N is either trivial or transitive on Ω.* □

(3.15). LEMMA. (BREY-IWASAWA-WILSON LEMMA) *Let G be primitive on Ω and $A \trianglelefteq G_\omega$ with $\langle A^G \rangle = G$. For $N \trianglelefteq G$, either N is trivial on Ω or $G/N \simeq A/A \cap N$.*

PROOF. Assume that N is not trival on Ω. Thus normal N is not in the core of the maximal subgroup G_ω, and $G = G_\omega N = \mathrm{N}_G(A)N$. In particular $A^G = A^N$ and $G = \langle A^G \rangle = \langle A^N \rangle = AN$. But then $G/N = AN/N \simeq A/A \cap N$, as claimed. □

(3.16). LEMMA. (IWASAWA'S LEMMA) *Let the perfect group G be primitive on Ω and abelian $A \trianglelefteq G_\omega$ with $\langle A^G \rangle = G$. For $N \trianglelefteq G$, either N is trivial on Ω or $G = N$.*

PROOF. As A is abelian, its quotient $A/A \cap N$ is also abelian. But G is perfect; so by the previous lemma if N is not in the kernel, then $G/N \simeq A/A \cap N$ is trivial. □

3.1.4 Sylow's Theorems

Assume that the finite group G has order $|G| = p^a m$ with p prime and $\gcd(p, m) = 1$. Then, as discussed in Section 2.8, a subgroup of order p^a is a *Sylow p-subgroup.* Two of Sylow's three theorems are really results about permutation groups. We emphasize that by proving them first here without reference to Sylow's First Theorem.

(3.17). THEOREM.

(a) (SYLOW'S SECOND THEOREM) *Any two Sylow p-subgroups of the finite group G are conjugate.*

(b) (SYLOW'S THIRD THEOREM) *Assume that the finite group G contains a nontrivial Sylow p-subgroup. Then for any p-subgroup Q of G the number of Sylow p-subgroups containing Q is congruent to* 1 *modulo p. In particular, the number of Sylow p-subgroups is congruent to* 1 *modulo p, and every p-subgroup is in at least one Sylow p-subgroup.*

PROOF. Of course, if there are no Sylow p-subgroups, then the Second Sylow Theorem is true trivially. Now assume that P is a Sylow p-subgroup. Let $\Omega = \{\, P^g \mid g \in G \,\}$, the set of conjugates of P in G and clearly a transitive G-space under conjugation.

For any p-subgroup Q of G

$$G_P \cap Q = \mathrm{N}_Q(P) \leq P\,,$$

as the Sylow p-subgroup P is contained in hence equal to the p-group $P\mathrm{N}_Q(P)$. In particular, P is the only Sylow p-subgroup normalized by P. Therefore in its action on Ω, the P-orbits all have length a multiple of p except for the single orbit $\{P\}$. In particular $|\Omega|$ is congruent to 1 modulo p. This in turn implies that for any p-subgroup Q, its number of orbits of length 1 in $|\Omega|$ is congruent to 1 modulo p.

If we apply the last argument to any Sylow p-subgroup Q, we learn that there is a conjugate P^g that contains Q. That is, $Q = P^g$ for some $g \in G$. This is the Second Sylow Theorem. Thus Ω consists of all Sylow p-subgroups, and the previous arguments give the Third Sylow Theorem. □

(3.18). COROLLARY. (THE FRATTINI ARGUMENT) *Let p be a prime. If N is normal in the finite group G, then $G = N_G(P)N$, where P is a Sylow p-subgroup of N.*

PROOF. Set $\Omega = \{$Sylow p-subgroups of $N\}$, on which G acts by conjugation. Then N is transitive on Ω by Sylow's Second Theorem and the point stabilizer is $G_P = N_G(P)$. Therefore this follows from our general Frattini Argument, Lemma (3.8). □.

We proved Sylow's First Theorem as Theorem (2.31). We now give a more permutation theoretic proof[2] due to Wielandt.

(3.19). THEOREM. (SYLOW'S FIRST THEOREM) *If the finite group G has order $|G| = p^a m$ with p prime, $a \in \mathbb{N}$, and* $\gcd(p, m) = 1$, *then G contains subgroups of order p^a and index m.*

[2] Of course, the Class Equation used in the earlier proof is really a permutation argument. In the action of G on itself by conjugation, the conjugacy classes are the orbits and the length of each orbit is given by Theorem (3.4).

Proof. Consider $\Omega = \binom{G}{p^a}$, the set of all subsets of G having cardinality p^a. This is a G-space via right translation.

Let Δ be a G-orbit in Ω and $D \in \Delta$, a p^a-subset of G. For every $g \in G$, there is at least one $h \in G$ with $g \in Dh \in \Delta$. Therefore the length $|\Delta|$ of the orbit must be at least $|G|/p^a = m$ (and with equality, every g is in exactly one Dh).

Assume for the moment that Δ is one of these short orbits. Then $|\mathrm{Stab}_G(D)| = |G|/|\Delta| = p^a$; the stabilizer is a Sylow p-subgroup P. We can further assume that D was chosen to contain 1_G, so $P \subseteq DP = D$ as P stabilizes D. This forces $P = D$, and $\Delta = \{\, Ph \mid h \in G \,\}$ is the coset space $P\backslash G$. We conclude that each Sylow p-subgroup P_i contributes to Ω exactly one orbit of length m, namely $P_i\backslash G$.

Every orbit not of length m has greater length, so its stabilizer cannot have order a multiple of p^a. In particular, the orbit length is a multiple of p. We conclude that

$$|\Omega| = k\,m + pn\,,$$

where k is the number of distinct Sylow p-subgroups in G and n is some integer. If we can prove that $|\Omega|$ is not a multiple of p, then k cannot be zero. In that case, there is at least one Sylow p-subgroup and our proof of the First Sylow Theorem will be done.

The trick here is that the needed calculation regarding $|\Omega|$ does not really depend on the group G but only on its order. If we replace G by any other group G_0 of the same order, then we have

$$|\Omega| = |\Omega_0| = k_0 m + pn_0$$

where k_0 is the number of distinct Sylow p-subgroups in G_0. With a careful choice of G_0, this calculation may be easier to make than the earlier one. And that is indeed the case: let G_0 be the cyclic group $Z_{|G|}$. Then G_0 has a unique Sylow p-subgroup. Thus $k_0 = 1$, and

$$|\Omega| = |\Omega_0| = m + pn_0\,.$$

This is not a multiple of p as m is not, and we are done! □

3.1.5 Problems

(3.20). Problem. *Prove that a group of order $2^3 3^4$ is solvable.*

(3.21). Problem. *For the permutation group (G, Ω), prove that the following are equivalent:*

(1) *(G, Ω) is transitive and $G_\gamma = 1$, for some $\gamma \in \Omega$.*

(2) *(G, Ω) is permutation isomorphic to the right regular representation.*

(3) *For all $\alpha, \beta \in \Omega$ there is a unique $g \in G$ with $\alpha^g = \beta$.*

(3.22). Problem. *Prove that a transitive, faithful permutation representation of an abelian group is regular.*

(3.23). PROBLEM. *Prove:*
(a) *A subgroup of a semiregular group is itself semiregular.*
(b) *Each orbit of a semiregular permutation representation of G is isomorphic to the right regular representation ρ.*
(c) *If H is finite and semiregular on Ω_1 and Ω_2 with $|\Omega_1| = |\Omega_2|$, then Ω_1 and Ω_2 are isomorphic H-spaces.*

(3.24). PROBLEM. *Suppose $H_1, H_2 \leq G$ with finite $H_1 \simeq H_2$. Then for the right regular representation of G, $\rho\colon G \longrightarrow \mathrm{Sym}(G)$, there is an $s \in \mathrm{Sym}(G)$ with $\rho(H_1)^s = \rho(H_2)$.*

(3.25). PROBLEM. *Let N be a group and let the image N^ρ of N under the right regular representation $\rho\colon N \longrightarrow \mathrm{Sym}(N)$.*
(a) *Let $\lambda\colon N \longrightarrow \mathrm{Sym}(N)$ be the* left regular representation *of N given by*

$$k \mapsto \lambda(k) \text{ with } g^{\lambda(k)} = k^{-1}g\,,$$

for all $g \in N$.
   (i) *Prove that λ is an isomorphism of N with its image N^λ in $\mathrm{Sym}(N)$.*
   (ii) *Prove that $C_{\mathrm{Sym}(N)}(N^\rho) = N^\lambda$.*
(b) *Let $M = N_{\mathrm{Sym}(N)}(N^\rho)$ and $A = \mathrm{Stab}_M(1_N)$, the stabilizer of 1_N in M. Prove that $M = AN^\rho$ and that this is the internal semidirect product of N^ρ by A.*
(c) *Prove that $A \simeq \mathrm{Aut}(N)$.*

(3.26). PROBLEM. *Let $G \leq \mathrm{Sym}(\Omega)$. Prove that G is k-transitive on Ω if and only if it is transitive on Ω and, for $a \in \Omega$, G_a is $(k-1)$-transitive on $\Omega \setminus \{a\}$.*

(3.27). PROBLEM. (D.G. HIGMAN) *A graph is* connected *if you can get from any vertex to any other by walking along a finite length path of edges, disregarding edge direction. It is* strongly connected *if the path can always be chosen to be directed.*
(a) *Prove that transitive (G, Ω) is primitive if and only if all nondiagonal orbital graphs are strongly connected*
(b) *Let G be transitive on finite Ω, and let Γ be a G-invariant graph on Ω. Prove that, Γ is connected if and only if it is strongly connected. ("If you can walk from a to b, then you can drive.")*

REMARK. *An infinite directed path shows that this can be false for infinite Ω.*

(3.28). PROBLEM. *Let the group G contain the set $R = \{\, r_i \mid i \in I \,\}$. The* Cayley graph $\mathrm{C}(G\,; R)$ *has as vertex set the elements of G and edges*

$$x \longrightarrow y \quad \Longleftrightarrow \quad yx^{-1} \in R\,.$$

We can make it a colored graph via

$$x \xrightarrow{i} y \quad \Longleftrightarrow \quad yx^{-1} = r_i\,.$$

G acts by translation as a regular group of automorphisms on $\mathrm{C}(G\,; R)$ since

$$xg \xrightarrow{i} yg \quad \Longleftrightarrow \quad yg(xg)^{-1} = r_i \quad \Longleftrightarrow \quad yx^{-1} = r_i \quad \Longleftrightarrow \quad x \xrightarrow{i} y\,.$$

Thus the Cayley graph is an orbital graph for the right regular permutation representation. If we wish it to be undirected, we must require R to be closed under inverses. The full automorphism group could be much larger than G. (Imagine $R = G \setminus 1$.)
(a) *Prove that $\mathrm{C}(G\,; R)$ is connected if and only if $G = \langle R \rangle$.*
(b) *Assume $G = \langle R \rangle$. Prove that the elements of G are the only automorphisms of $\mathrm{C}(G\,; R)$ that respect the edge coloring.*

3.2 Linear groups

3.2.1 Basics

A *K-linear representation* of the group G is a homomorphism $\varphi\colon G \longrightarrow \mathrm{GL}_K(V)$, where V is a vector space over the division ring K. For $g \in G$ and $v \in V$, we usually write $v^{\varphi(g)}$ in the more compact form v^g. The K-space V is a *G-module* over K.

If φ is faithful, then G is said to be a *linear group*. The *degree* of the representation is $\dim_K(V)$. A related concept is a *projective representation*, that being a homomorphism $\varphi\colon G \longrightarrow \mathrm{PGL}_K(V)$. As can be seen in Theorem (1.3), most of the finite simple groups are realized best as projective linear groups. That is a prime motivation for studying linear and projective representations.

Care must be taken with the terminology, since we are not requiring V to have finite dimension. In the literature, the term "linear group" is usually reserved for those groups with faithful representations of finite degree, the *finite dimensional linear groups*.

If K is a field, then a ${}_K\mathsf{Vec}$-representation $\varphi\colon G \longrightarrow \mathrm{GL}_K(V)$ extends by linearity to a representation of the group algebra $KG \longrightarrow \mathrm{End}_K(V)$. This powerful observation allows the methods of associative algebras (such as the Wedderburn-Artin theory) to be applied in group theory. This becomes more unwieldy for division rings K; so we avoid this sharp tool, although it does leave its mark in our "module" terminology.

If f is an invertible K-linear transformation $f\colon V_1 \longrightarrow V_2$ (that is, a ${}_K\mathsf{Vec}$-isomorphism) of the two K-spaces V_1 and V_2, then we have the induced isomorphism f^* of $\mathrm{GL}_K(V_1) = \mathrm{Aut}_{{}_K\mathsf{Vec}}(V_1)$ and $\mathrm{GL}_K(V_2) = \mathrm{Aut}_{{}_K\mathsf{Vec}}(V_2)$, as in Section 1.4:

$$\begin{tikzcd}
V_1 \arrow[r, "a"] \arrow[d, "f"] & V_1 \arrow[d, "f"] \\
V_2 \arrow[r, "a^*"] & V_2
\end{tikzcd}$$

with

$$a \mapsto a^{f^*} = a^* = f^{-1}af\,.$$

Two K-linear representations $\varphi_1\colon G \longrightarrow \mathrm{GL}_K(V_1)$ and $\varphi_2\colon G \longrightarrow \mathrm{GL}_K(V_2)$ are *equivalent* in ${}_K\mathsf{Vec}$ if, for some isomorphism $f\colon V_1 \longrightarrow V_2$ and its induced isomorphism f^*, the diagram

$$\begin{tikzcd}
G \arrow[r, "\varphi_1"] \arrow[dr, "\varphi_2"] & \mathrm{GL}_K(V_1) \arrow[d, "f^*"] \\
& \mathrm{GL}_K(V_2)
\end{tikzcd}$$

commutes. In this case we say that V_1 and V_2 are *isomorphic G-modules*.

In the larger category Vec the two linear representations $\varphi_1\colon G \longrightarrow \mathrm{GL}_{K_1}(V_1)$ and $\varphi_2\colon G \longrightarrow \mathrm{GL}_{K_2}(V_2)$ are equivalent if the diagram

$$\begin{tikzcd}
G \arrow[r, "\varphi_1"] \arrow[dr, "\varphi_2"] & \mathrm{GL}_{K_1}(V_1) \arrow[d, "f^*"] \\
& \mathrm{GL}_{K_2}(V_2)
\end{tikzcd}$$

commutes, where now f^* is induced by a Vec-isomorphism: an invertible semilinear map $f\colon {}_{K_1}V_1 \longrightarrow {}_{K_2}V_2$. A *semilinear map* (or τ-semilinear map) $[\tau, t]$ from ${}_KV$ to ${}_FW$ is a homomorphism of additive groups $t\colon (V, +) \longrightarrow (W, +)$ that additionally, for the field embedding $\tau\colon K \longrightarrow F$, satisfies

$$(av)^t = a^\tau v^t\,,$$

for all $a \in K$, $v \in V$.

Permutation representations give rise to linear representations. Let G act as permutations on the set Ω. For an arbitrary division ring K, let the *K-permutation module* $K\Omega$ be the K-vector space with basis $\mathcal{B} = \{\, e_\omega \mid \omega \in \Omega\,\}$. For each $g \in G$ we define the linear transformation $\varphi(g) \in \mathrm{GL}_K(K\Omega)$ by

$$e_\omega^g = e_{\omega^g}$$

for all $\omega \in \Omega$. The map $\varphi\colon g \longrightarrow \mathrm{GL}_K(K\Omega)$ is then a representation of G, the *K-permutation representation.*

The matrix representing $\varphi(g)$ in the basis $\mathcal{B}$ has a unique nonzero entry in each row and each column, that entry being a 1. Such a matrix M is called a *permutation matrix* and is orthogonal in the sense that its transpose is its inverse: $MM^\top = I$.

As an immediate consequence of Cayley's Theorem (3.2), we have

(3.29). THEOREM. *Every group has a faithful representation as a K-linear group.* □

This is false if we restrict ourselves to representations of finite degree.

An elementary abelian p-group can be thought of as a vector space over the field $\mathbb{F}_p$. Therefore linear representations also play an important role in the internal representation theory discussed earlier in Lemma (2.9).

(3.30). PROPOSITION. *Let V be an elementary abelian p-group. Then* $\mathrm{Aut}(V) = \mathrm{GL}_{\mathbb{F}_p}(V)$. □

3.2.2 Irreducibility

With permutation groups we progressed from intransitive to transitive to primitive groups. We attempt similar reductions for linear groups. For permutation groups, we initially factored intransitive groups into transitive groups. For linear groups, the corresponding move is from reducible groups to irreducible groups, but it turns out that there is more than one flavor of each.

The representation $\varphi\colon G \longrightarrow \mathrm{GL}_K(V)$ is *reducible* if there is a G-invariant (actually, $\varphi(G)$-invariant) K-subspace W with $0 < W < V$. If φ is not reducible, then it is *irreducible*. We then also say that G is irreducible on the V and that the G-module V is irreducible.

(3.31). LEMMA. *Let W be a submodule of the G-module V over K with $0 < W < V$. Further, let $V^+ = W \oplus V/W$, the K-space direct sum of W and V/W.*

(a) *The quotient K-space V/W has a natural structure as G-module given by $(v+W)^g = v^g + W$.*

(b) *V^+ has a natural structure as G-module $W \oplus V/W$ given by $(u,(v+W))^g = (u^g, v^g + W)$.* □

If in the lemma there is in V a G-submodule X with $V = W \oplus X$, then V/W and X are isomorphic G-modules. We say that the extension is *split* and that V is *G-decomposable* (or just *decomposable*) with G-decomposition given by $V = W \oplus X$. If V has no proper decompositions, then it is *indecomposable*. We can think of decomposability as a strong form of reduciblity. Decomposability implies reduciblity, and irreduciblity implies indecomposablity. If Ω is an intransitive permutation space for G, then the associated permutation module is decomposable, being the direct sum of smaller permutation modules.

In Proposition (3.3) we used the Chinese Remainder Theorem to reduce the study of intransitive permutation groups to that of transitive groups. We can use the present lemma to attempt a similar reduction from reducible to irreducible representations of linear groups. There are two problems with the approach. In the permutation case, we always knew that an orbit provides a transitive "subrepresentation," but here there is no guarantee that a given module has an irreducible submodule. More fundamentally, and again unlike the permutation case, even when the map from G into $\mathrm{GL}_K(V)$ is faithful, there is no guarantee that the related representation in $\mathrm{GL}_K(V')$ is also faithful. The kernel will be trivial if the extension is split (in which case V and V' are isomorphic G-modules), but not in general. We will return to the case of nontrivial kernels in Section 3.2.4.

It is possible that a G-module over K is irreducible only because we are considering it over the wrong division ring. If K is a subdivision ring of L, then the L-space W has a natural structure as K-space of dimension $\dim_K(W) = [L\colon K]\dim_L(W)$. For example, the correspondence between $a + bi \in \mathbb{C}$ and $(a,b) \in \mathbb{R}^2$ allows us to think of a complex representation of degree n as a real representation of degree $2n$. As a partial converse, one can think of any G-module X of dimension m over K as a G-module ${}_LX$ of dimension m over L by extending coefficients to the tensor product module ${}_LX = L \otimes_K X$. In this case the new module ${}_LX$ may be G-reducible even though the original module ${}_KX$ was G-irreducible. In the example, the real module ${}_{\mathbb{R}}W$ of dimension $m = 2n$

when extended to ${}_{\mathbb{C}}W$ is a direct sum of two complex modules of dimension n, one isomorphic to the original W and the second isomorphic to W twisted by complex conjugation. An irreducible G-module over K that remains irreducible over all extensions L of K is *absolutely irreducible*. In most places, the discussion of absolute irreducibility is restricted to the situation in which K and L are both fields.

3.2.3 Primitivity

The G-module V is *completely reducible* if it is a sum of irreducible G-modules. In this case, for each irreducible G-submodule W of V we set $H_G(W)$ to be the sum in V of all irreducible submodules of V that are isomorphic to W as G-modules. The module $H_G(W)$ is the *homogeneous component* corresponding to W.

(3.32). PROPOSITION. *Let the completely reducible G-module V over K be the sum $\sum_{e\in E} W_e$ of irreducible G-modules W_e.*

(a) *There is a subset J of E with $V = \bigoplus_{j\in J} W_j$.*

(b) *Let I be a subset of J such that $\{\, W_i \mid i \in I \,\}$ is a maximal set of pairwise nonisomorphic irreducible G-modules. Then every irreducible G-submodule of V is isomorphic to one of the W_i and $V = \bigoplus_{i\in I} H_G(W_i)$.*

PROOF. (a) Consider the partially ordered set $\mathcal{I}$ of subsets S of E with $\sum_{s\in S} W_s = \bigoplus_{s\in S} W_s$, ordered by containment. For every chain in $\mathcal{I}$, the union of the chain is also in $\mathcal{I}$; so by Zorn's Lemma there is a maximal subset J. If $V' = \bigoplus_{j\in J} W_j$ is proper in V, then there is a W_e not contained in V_0, since $V = \sum_{e\in E} W_e$. By irreduciblity of W_e we have $W_e \cap V' = 0$, hence $W_e + V' = W_e \oplus V'$. But then $J \cup \{e\} \in \mathcal{I}$, against maximality of J. Therefore $V = V' = \bigoplus_{j\in J} W_j$.

(b) Let U be an irreducible G-submodule of V. For the nonzero element $u \in U$, there is a finite subset J_0 of J with $u \in V_0 = \bigoplus_{j\in J_0} W_j$, hence by irreducibility $U \leq V_0$. Choose such a J_0 of minimal cardinality. Clearly J_0 is nonempty; let $k \in J_0$. Then U is not in $X_0 = \bigoplus_{j\in J_0\setminus k} W_j$, and by irreduciblity U meets this partial sum trivially. Therefore we have $U + X_0 = U \oplus X_0$, and by the Second Isomorphism Theorem $U \simeq (U \oplus X_0)/X_0 \leq V_0/X_0 \simeq W_k$. By irreducibility of W_k, the modules U and W_k are isomorphic.

For each $i \in I$, let J_i be the subset of J consisting of those W_j isomorphic to W_i. Then $V_j = \bigoplus_{j\in J_i} W_j$ is certainly in $H_G(W_i)$ and $V = \bigoplus_{i\in I} V_i$. But the argument of the previous paragraph shows than any irreducible G-submodule U isomorpic to W_i is contained within a sum of irreducible modules W_j with $j \in J_i$. Therefore $U \leq V_i$ and so $V_i = H_G(W_i)$. □

Let V be a G-module over K and N a normal subgroup of G. For each $g \in G$ and X an N-submodule of V, the image X^g is a new N-submodule of V. Indeed, for each $n \in N$ and $x^g \in X^g$, we have

$$(x^g)^n = (x^{gng^{-1}})^g \in X^g\,,$$

as N is normal in G and X is an N-module. We also have $X = (X^g)^{g^{-1}}$, so X is irreducible if and only if X^g is irreducible.

(3.33). THEOREM. (CLIFFORD'S THEOREM) *Let V be an irreducible G-module over K and N a normal subgroup of G. Assume that V contains an irreducible N-submodule W. (This will always be the case when* $\dim_K(V) < \infty$*). Then we have:*

(a) *V is completely reducible as an N-module, and every irreducible N-submodule of V is isomorphic to W^g, for some $g \in G$.*

(b) *V is the direct sum of the distinct homogeneous components $H = H_N(W^g)$, and G/N permutes these transitively under $H \mapsto H^g$.*

(c) *If $H = H_N(W^g)$ is an N-homogeneous component of V and A is the stabilizer of N in G, then H is an irreducible A-module.*

PROOF. (a,b) As W is N-irreducible, each W^g is N-irreducible The sum $\sum_{g \in G} W^g$ is then a G-submodule of V. As V is an irreducible G-module, $V = \sum_{g \in G} W^g$, and the rest follows by the proposition.

(c) For each A-submodule L of H, the direct sum of the distinct L^g is a G-submodule of V. As G is irreducible on G, this forces L to be either 0 or L. That is, A is irreducible on H. □

If there are two or more N-homogeneous components in Clifford's Theorem, then we have an important instance of a frequent occurence. An irreducible G-module V is *imprimitive* if $V = \oplus_{i \in I} H_i$ is the direct sum of subspaces H_i (of fixed dimension) permuted transitively by G. If the G-module V is not imprimitive, then it is *primitive*.

3.2.4 Unipotent linear groups

As promised, we return to Lemma (3.31) to study the kernel of the stabilizer of a subspace.

(3.34). LEMMA. *The stabilizer of the subspace W in $\mathrm{GL}_K(V)$ is the semidirect product of $U = \{\, M \in \mathrm{GL}_K(V) \mid M|_W = 1_W,\, M|_{V/W} = 1_{V/W} \,\}$ by $L = \mathrm{GL}_K(W) \times \mathrm{GL}_K(X)$, where X is some complement to W in V.*

If $V = W \oplus X$ is a decomposition into G-invariant subspaces W and X of V, then the subgroup stabilizing each subspace is $L = \mathrm{GL}_K(W) \times \mathrm{GL}_K(X)$.

PROOF. In matrices:

$$\left\{\begin{pmatrix} A & 0 \\ * & B \end{pmatrix}\right\} = \left\{\begin{pmatrix} I & 0 \\ * & I \end{pmatrix}\right\} \rtimes \left\{\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}\right\}$$

where

$$U = \left\{\begin{pmatrix} I & 0 \\ * & I \end{pmatrix}\right\} \text{ and } L = \left\{\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}\right\}. \quad \square$$

More generally, if $V = V_0 \geq V_1 \geq \cdots \geq V_k = 0$ is an G-invariant series in the K-space V, then the *stability group* of the series is the subgroup

$$\left\{ g \in G \;\middle|\; g|_{V_{i-1}/V_i} = 1_{V_{i-1}/V_i},\ 1 \leq i \leq k \right\}.$$

We shall see below that such a group is nilpotent of class at most k. (See Problem (6.36).)

(3.35). Lemma. *Let $V = V_0 \geq V_1 \geq \cdots \geq V_k = 0$ be a G-invariant series for the faithful G-module V over K. Then $V^+ = \bigoplus_{i=1}^k V_{i-1}/V_i$ is a G-module over K.* □

If we restrict to the finite dimensional case, $\dim_K(V) = n < \infty$, then, for an appropriate choice of basis, the kernel N of the representation of G on V^+ consists of all matrices of the form

$$\begin{pmatrix} I & 0 & 0 & 0 & 0 \\ * & I & 0 & 0 & 0 \\ * & * & \ddots & 0 & 0 \\ * & * & * & I & 0 \\ * & * & * & * & I \end{pmatrix}$$

where identity matrices have degree $\dim_K(V_{i-1}/V_i)$ for $k \geq i \geq 1$, going from top to bottom.

The group of all $n \times n$ matrices over K with 1's on the diagonal and 0's above the diagonal is the *lower unitriangular group*, which we will denote $\mathrm{U}(n)K$. All such kernels as N above are all *block lower unitriangular* (with k blocks) and are contained in $\mathrm{U}(n)K$.

The unipotent elements of $\mathrm{GL}_n(K)$ are those that are conjugate to an element of $\mathrm{U}(n)K$. More generally, a *unipotent element* g of $\mathrm{GL}_K(V)$ (arbitrary dimension) is one with $(g-1)^m = 0$ (in $\mathrm{End}_K(V)$), for some integer m. A *unipotent subgroup* of $\mathrm{GL}_K(V)$ is one all of whose elements are unipotent.

(3.36). Lemma.

(a) *For* $\mathrm{char}K = 0$*, every nonidentity unipotent element of* $\mathrm{GL}_K(V)$ *has infinite order.*

(b) *For* $\mathrm{char}K = p > 0$*, an element of* $\mathrm{GL}_K(V)$ *is unipotent if and only if it is a p-element. In this case, if $(g-1)^m = 0$, then $g^{p^{m-1}} = 1$.*

(c) *A unipotent element $g \in \mathrm{GL}_n(K)$ is conjugate to an element of* $\mathrm{U}(n)K$ *and, if* $\mathrm{char}K = p > 0$*, has $g^{p^{n-1}} = 1$.*

Proof. In characteristic p, if $g^{p^k} = 1$, then $g^{p^k} - 1 = (g-1)^{p^k} = 0$. Therefore in (b) p-elements are always unipotent.

Let g be a unipotent element with, say, $(g-1)^m = 0$. Every $v \in V$ is contained in a g-invariant subspace of dimension m_v at most m, namely $U_v = \sum_{i=0}^{m-1} Kv^{(g-1)^i} = \sum_{i=0}^{m-1} Kv^{g^i}$.

Consider a v with $v^g \neq v$; that is, U_v does not have dimension 1. For an appropriate basis, g is represented on U_v by a lower unitriangular matrix. Choose $i > j$ with $g_{ij} = a \neq 0$ and, subject to that, with $i-j$ minimal. Then $(g^k)_{ij} = ka$, for all k. In characteristic 0 this is never 0; so g has infinite order on U_v, giving (a).

In characteristic p, we have $(g^p)_{ij} = 0$; and this is true for all i', j' with $i' - j' \leq i - j$. Continuing in this manner, we see that $g^{p^{m_v - 1}}$ is trivial on U_v. That is, for $l = \max_{v \in V}(m_v) \leq m$, the endomorphism $g^{p^{l-1}}$ is trivial on every U_v, and so is trivial on V. In particular $g^{p^{m-1}} = 1$. This gives (b).

If $g \in \mathrm{GL}_n(K)$ then as above there is a basis in which g is represented by a unitriangular matrix. And certainly $l \leq n$, so (c) is complete. □

When $\dim_K(V)$ is finite, the unitriangular subgroups of $\mathrm{GL}_K(V)$ are nilpotent. See Problem (6.36).

3.2.5 Sylow's First Theorem (third time)

(3.37). Theorem. (Sylow's First Theorem) *If the finite group G has order $|G| = p^a m$ with p prime, $a \in \mathbb{N}$, and $\gcd(p, m) = 1$, then G contains subgroups of order p^a and index m.*

Proof. Consider G acting faithfully on a module $\mathbb{F}_p^n$ (for instance, the permutation module $V = \mathbb{F}_p G$). In this action, it permutes the set Ω of all nonzero vectors of V. Here $|\Omega| = p^n - 1$ is not a multiple of p. Therefore some orbit of G on Ω has order prime to p. If the orbit has size greater than 1, then by induction on order a point stabilizer contains a Sylow p-subgroup of G. If the orbit has size 1, say v, then G that acts on the nonzero vectors of the smaller space $V/\mathbb{F}_p v$. By induction on dimension the group induced by G on this quotient space has a Sylow p-subgroup $\bar{P}$. The kernel of that action is a normal Sylow p-group, and the preimage P of $\bar{P}$ in G is then a Sylow p-subgroup of G. □

(3.38). Corollary. $\mathrm{U}_n(p^a)$ *is a Sylow p-subgroup of* $\mathrm{GL}_n(p^a)$.

Proof. This can of course be proven by a calculation of the order of $\mathrm{GL}_n(p^a)$. Also the argument used above can be used to show first that every p-subgroup fixes a 1-space and indeed a vector, and then induction takes over. □

3.2.6 Problems

(3.39). Problem. *Let V be the K-permutation module for the finite group $G =$ $\mathrm{Sym}(\Omega)$; that is, $V = \bigoplus_{\omega \in \Omega} K\, e_\omega$, with $e_\omega^g = e_{\omega^g}$, for all $g \in \mathrm{Sym}(\Omega)$ and $\omega \in \Omega$.*

Let

$$U = \{ \sum_{\omega\in\Omega} a_\omega e_\omega \mid \sum_{\omega\in\Omega} a_\omega = 0,\, a_\omega \in K \}$$

and

$$Z = \{ \sum_{\omega\in\Omega} a_\omega e_\omega \mid a_\omega = a \in K,\ \textit{for all } \omega \}\,.$$

(a) *Prove that U and Z are G-submodules of V.*

(b) *Prove that 0, Z, U, and V are the only G-submodules of V.*

(3.40). Problem. *Consider* $\mathrm{Mat}_n(K)$*, the space of $n \times n$ matrices from K, as a module for the group $G = \mathrm{GL}_n(K) \times \mathrm{GL}_n(K)$, where, for $M \in \mathrm{Mat}_n(K)$ and $(A,B) \in G$, we have*

$$M^{(A,B)} = A^\top M B\,.$$

Prove that $\mathrm{Mat}_n(K)$ *is irreducible for G.*

(3.41). Problem. *Let G and H be finite groups, and let*

$$X \colon G \longrightarrow \mathrm{GL}_K(V) \quad \textit{and} \quad Y \colon H \longrightarrow \mathrm{GL}_K(W)$$

be K-representations of G and H.

(a) *Prove that $X \oplus Y \colon G \times H \longrightarrow \mathrm{GL}_K(V \oplus W)$ given by*

$$(v,w)^{X\oplus Y(g,h)} = (v^{X(g)}, w^{Y(h)})$$

is a representation of $G \times H$.

(b) *Prove that $X \otimes Y \colon G \times H \longrightarrow \mathrm{GL}_K(V \otimes_K W)$ (tensor product) given by*

$$(v \otimes w)^{X\otimes Y(g,h)} = v^{X(g)} \otimes w^{Y(h)}$$

is a representation of $G \times H$.

(c) *Prove that if $X \otimes Y$ is irreducible then X and Y are irreducible.*

Remark. *In the special case $G = H$ the group G itself has a natural embedding on the diagonal of $G \times H = G \times G$, namely $\delta \colon G \longrightarrow G \times G$ given by $\delta(g) = (g,g)$. The composition of this with the representations $X \oplus Y$ and $X \otimes Y$ (restricted to the image of δ) gives new K-representations of G as the "sum" and "product" of the two original representations. With this in mind, the collection of all K-representations of G can naturally be given structure of a ring, indeed a K-algebra since the effect of scalar multiplication is easy to see. (We do want to factor by equivalence).*

The source of this nice additional structure is the innocent diagonal mapping δ. The natural abstract setting is that of Hopf algebras. *These are associative K-algebras A that in addition to having the usual multiplication $\mu \colon A \otimes_K A \longrightarrow A$ also have a well-behaved* comultiplication *$\delta \colon A \longrightarrow A \otimes_K A$. The previous paragraph is then about the special (and motivating) example where A is the group algebra KG.*

(3.42). Problem. *Prove that an element g of* $\mathrm{End}_K(V)$ *with $(g-1)^m = 0$ for some m must belong to* $\mathrm{GL}_K(V)$.

Chapter 4

Finiteness and Reduction

Many mathematical arguments follow the path of breaking a large problem into a number of smaller problems. Particular examples are the three "unique factorization" results presented in Section 1.1. The questions then are:

(i) What do we mean by "smaller"?

(ii) How do we achieve the "break"?

For us, smallness will usually be gauged by a finiteness condition—a property that the object under study shares with finite objects. Once a suitable finiteness condition is imposed, then reduction via something resembling induction becomes available.

4.1 Finiteness: Sylow's First Theorem, one last time

Of course, the basic finiteness condition for a group is the requirement that the group be finite. As already discussed, the classification of finite simple groups Theorem (1.3) proceeds by induction, studying an arbitrary finite simple group G all of whose proper simple sections are on the theorem's list and ultimately proving that G then is also on the list.

As an example of how such reductions go, we have a last[1] proof of Sylow's First Theorem.

(4.1). Proposition. *The First Sylow Theorem is valid in all finite groups if and only if it is valid in all finite simple groups.*

Proof. One direction is clear. Now assume that the First Sylow Theorem holds in all finite simple groups. Let G be an arbitrary finite group and p a

[1] I promise!

prime. We prove that G satisfies the First Sylow Theorem for the prime p by induction on $|G|$.

If G is simple then we are done by hypothesis. Therefore we may assume that G contains a normal subgroup N, not equal to G or to 1. By induction N and G/N both have Sylow p-subgroups. Let P be a Sylow p-subgroup of N. By the Second Sylow Theorem, all Sylow p-subgroups of N are conjugate, and by the Frattini argument $G = \mathrm{N}_G(P)N$.

By the Second Isomorphism Theorem $G/N \simeq \mathrm{N}_G(P)/\mathrm{N}_N(P)$. In particular, $[G{:}\mathrm{N}_G(P)]$ is not a multiple of p, and a Sylow p-subgroup of $\mathrm{N}_G(P)$ is a Sylow p-subgroup of G. Especially a Sylow p-subgroup of G/N is isomorphic to one of $\mathrm{N}_G(P)/\mathrm{N}_N(P)$.

As P is Sylow in N, the index $[\mathrm{N}_N(P){:}P]$ is not a multiple of p. By the Third Isomorphism Theorem

$$\mathrm{N}_G(P)/P \Big/ \mathrm{N}_N(P)/P \simeq \mathrm{N}_G(P)/\mathrm{N}_N(P)\,.$$

Therefore a preimage of a Sylow p-subgroup of $\mathrm{N}_G(P)/\mathrm{N}_N(P)$ in $\mathrm{N}_G(P)/P$ is a Sylow p-subgroup of $\mathrm{N}_G(P)/P$. In turn a preimage of that in $\mathrm{N}_G(P)$ is a Sylow p-subgroup of $\mathrm{N}_G(P)$ and so of G. □

(4.2). PROPOSITION. *For the prime p, assume:*

> *Every finite simple group that has order a multiple of p but is not a p-group has a permutation representation with no fixed points and of degree not a multiple of p.*

Then the First Sylow Theorem holds for the prime p holds.

PROOF. The proof is by induction on the order of the arbitrary finite group G. If G itself is a p-group, then there is nothing to prove.

By the previous proposition, we need only consider finite simple G that is not a p-group. By the assumption, it has a faithful permutation representation with no fixed points and of degree not a multiple of p. But then it has in this an orbit, not of length 1 but of length not a multiple of p. As G is simple, it is faithful on this orbit. For ω a point in this orbit, G_ω is a proper subgroup of G. By induction, it contains a Sylow p-subgroup which is then a Sylow p-subgroup of G. □

(4.3). PROPOSITION. *Let p be a prime. If G is a finite simple group of order $p^a m$ with $a \in \mathbb{Z}^+$ and $1 \neq m$ coprime to p, then G has a permutation representation with no fixed points and of degree not a multiple of p.*

PROOF. As $m \neq 1$, the group G is nonabelian simple by Proposition (2.18). We give three constructions of an appropriate G-space Ω:

(1) (Class equation proof) Let $\Omega = G \setminus \{1\}$ with conjugation action. Any orbits of length 1 would correspond to elements of the center of nonabelian simple G.

(2) (Wielandt's permutation proof) Let $\Omega = \binom{G}{p^a}$ with translation action. As $m \neq 1$, all orbits of G on Ω are nontrivial. The following binomial coefficient calculation shows that $|\Omega|$ is not a multiple of p:

$$|\Omega| = \binom{p^a m}{p^a} =$$
$$\frac{(p^a m)(p^a m-1)\cdots(p^a m-p)\cdots(p^a m-p^{a-1})\cdots\cdots(p^a m-p^a+1)}{(p^a)(p^a-1)\cdots(p^a-p)\cdots(p^a m-p^{a-1})\cdots\cdots(1)}\,.$$

(3) (Unipotent subgroup proof) Let $\Omega = V \setminus \{0\}$ where V is an $\mathbb{F}_p$-space that is a faithful G-module and, subject to that, has minimal degree. (The permutation module $\mathbb{F}_p G$ proves that such a V exists.). If there is an orbit of length 1, then G has abelian hence trivial action on the 1-space it spans. But then the quotient of V by that 1-space would give a representation of smaller degree that is faithful as G is nonabelian simple. □

(4.4). Theorem. (Sylow's First Theorem) *If the finite group G has order $|G| = p^a m$ with p prime, $a \in \mathbb{N}$, and* $\gcd(p,m)=1$, *then G contains subgroups of order p^a and index m.*

Proof. Proposition (4.3) combines with Proposition (4.2) to give the result. □

4.2 Finite generation and countability

In many applications the groups studied are not necessarily finite but do have a finite description. The fundamental group of a surface is often finitely generated or even finitely presented. Here the group G is *finitely generated* if there is a finite subset X of G with $G = \langle X \rangle$, and it is additionally *finitely presented* if there is some finite set of relations R such that $G \simeq \langle\, X \mid R \,\rangle$; that is, G is isomorphic to the *free group* $\mathrm{F}(X)$ on X modulo its normal subgroup $\langle R^{\mathrm{F}(X)} \rangle$.

There are many important results in this area; see [Rob82, §14.1]. Here we are content with one that is of great importance both theoretically and computationally.

(4.5). Theorem. *Let G be a group and H a subgroup of finite index in G.*

(a) (Schreier) *If G is finitely generated, then H is finitely generated.*

(b) (Reidermeister-Schreier) *If G is finitely presented, then H is finitely presented.*

Proof. (a) Let G be generated by the finite set X, and let $Y = \{\, g_i \mid i \in I \,\}$ be a finite set of coset representatives for H in $G = \biguplus_{i \in I} H g_i$. For ease, we assume $1_G = g_0 \in Y$.

For each $g \in G$ let the coset representative $\bar{g} \in Y$ be given by $g \in H\bar{g}$. We claim that H is generated by the finite set of Schreier generators

$$W = \{\ (yx)\overline{yx}^{-1} \mid x \in X \cup X^{-1},\, y \in Y\,\}\,.$$

Indeed every element g of G can be written, for some $n \in \mathbb{Z}^+$, as $y_1 \prod_{i=1}^n x_i$ with $x_i \in X \cup X^{-1}$ and $y_1 \in Y$. (Recall that $1_G = g_0 \in Y$.) Then

$$\begin{aligned} g = y_1 \prod_{i=1}^n x_i &= (y_1x_1)(\overline{y_1x_1}^{-1}\overline{y_1x_1}) \prod_{i=2}^n x_i \\ &= ((y_1x_1)\overline{y_1x_1}^{-1})(\overline{y_1x_1}) \prod_{i=2}^n x_i \\ &= w_1 \left(y_2 \prod_{i=2}^n x_i \right) \end{aligned}$$

with $w_1 = (y_1x_1)\overline{y_1x_1}^{-1} \in W \subseteq H$ and $y_2 = \overline{y_1x_1} \in Y$. Therefore by induction on n, every element g of G can be rewritten as a product $\prod_{i=1}^n w_i \, y_{n+1}$ of elements w_i in W and a final member y_{n+1} of Y. If g happens to belong to H, then $\prod_{i=1}^n w_i \in H$ forces $y_{n+1} = g_0 = 1_G$, hence $g \in \langle W \rangle$. Therefore $H = \langle W \rangle$.

(b) (Sketch). If R is a set of relations defining G with respect to the generating set X, then

$$S = \{\, yry^{-1} \mid y \in Y,\, r \in R \,\}$$

is a set of relations defining H with respect to the generating set W, where, as above, the Reidermeister rewriting process allows us to rewrite the relations of S as words in the generating set W of H. The set S is finite if R is finite.

The difficulty that must be resolved before we have a complete proof is that, while by (a) the subset W (actually a bijective preimage) within $\mathrm{F}(X)$ certainly generates the full preimage of H, it might not do so freely. Schreier solved this by proving that, given a careful initial choice of the set Y of representatives, a free generating set results from deleting all elements of W that are the identity in $\mathrm{F}(X)$. □

Every finitely generated group is countable. Indeed every countably generated group is countable. Countability is thus a weaker finiteness property than finite generation. It still can be useful. In particular, a countable group can be expressed as the union of an ascending chain of its subgroups. Suppose $G = \{\, g_i \mid i \in \mathbb{Z}^+ \,\}$ is a countable group. For each i set $G_i = \langle g_1, \dots, g_i \rangle$. Then $G = \bigcup_{i \geq 1} G_i$ with

$$1 = G_0 \leq G_1 \leq G_2 \leq \cdots \leq G_i \leq \cdots$$

This may allow us to check certain properties of G more easily in one of the finitely generated subgroups G_i.

(4.6). Lemma. *The nontrivial group G is simple if and only if for every pair of elements $g, h \in G$ with $g \neq 1$, there is a finite subset $X(h,g)$ of G with $h \in \langle\, g^x, (g^{-1})^x \mid x \in X(h,g) \,\rangle$.*

Proof. If G is simple, then $G = \langle g^G \rangle$ and h is a product of a finite number of conjugates of g or its inverse.

Conversely, assume that for each h and $g \neq 1$ the set $X(h, g)$ exists. Let N be a nontrivial normal subgroup of G and choose $1 \neq n \in N$. Then for every h in G every element of $\{ n^x, (n^{-1})^x \mid x \in X(h, n) \}$ is in N, hence $h \in N$. That is, $N = G$ and G is simple. □

Of course for a given g, h there can be many appropriate finite subsets $X(h, g)$.

(4.7). Theorem. (P. Hall) *Let G be a simple group. Then every countable subset of G is contained in a countable simple subgroup of G.*

Proof. Let S be a countable subset of G. Then define

$$\mathcal{S}(S) = \langle\, S, X(h, g) \mid 1 \neq g, h \in S \,\rangle\,,$$

where $X(h, g)$ is a set as in the lemma. As S is countable and each $X(h, g)$ is finite, $\mathcal{S}(S)$ is also countable. Set $\mathcal{S}^0(S) = \langle S \rangle$ and recursively define $\mathcal{S}^i(S) = \mathcal{S}(\mathcal{S}^{i-1}(S))$ for integral $i \geq 2$. Then

$$S \subseteq \mathcal{S}^0(S) \leq \mathcal{S}^1(S) \cdots \leq \mathcal{S}^i(S) \leq \cdots$$

with S and each subgroup in the sequence countable.

Then $\mathcal{S}^\infty = \cup_{i \geq 1} \mathcal{S}^i(S)$ is a countable subgroup of G containing S, and by the lemma it is simple. □

4.3 Limits and finite approximation

Above we used the fact that every countable set is an ascending union of finite subsets hence every countable group is an ascending union of finitely generated subgroups. But every set is the union of its finite subsets, and so every group is the union of its finitely generated subgroups. Can we make these trivial observations useful?

The set of subgroups $\Gamma = \{ G_i \mid i \in I \}$ is a *local system* in G if

(i) $G_i \leq G$ for each $i \in I$, and $G = \bigcup_{i \in I} G_i$

(ii) for each $i, j \in I$, there is a $k \in I$ with $\langle G_i, G_j \rangle \leq G_k$.

In this case we say that G is the *directed limit* of its local system Γ.[2]

We may also call this the *finitely directed limit*, since an easy induction shows that for any finite subset J of I there is a k with $\langle\, G_j \mid j \in J \,\rangle \leq G_k$. Correspondingly, G is the *countably directed limit* of the local system Γ provided for every countable subset J of I there is always a $k \in I$ with $\langle\, G_j \mid j \in J \,\rangle \leq G_k$.

(4.8). Proposition.

[2] This is actually the *internal directed limit.* There is a corresponding more general *external directed limit*, which we do not define.

(a) *Every group is the directed limit of its finitely generated subgroups.*

(b) *Every group is the countably directed limit of its countable subgroups.*

PROOF. If G_i is generated by X_i, then $\langle\, G_i \mid i \in I \,\rangle$ is generated by $\{\, X_i \mid i \in I \,\}$. □

Phillip Hall's Theorem (4.7) can now be refined to say

(4.9). THEOREM. (P. HALL) *Every simple group is the countably directed limit of its countable simple subgroups.*

PROOF. Let $\{\, G_i \mid i \in I \,\}$ be the set of all countable simple subgroups of simple G. By Theorem (4.7) we have $G = \bigcup_{i\in I} G_i$. Let J be a countable subset of I. Then $\bigcup_{j\in J} G_j$ remains countable, so by Theorem (4.7) again there is a $k \in I$ with $\langle\, G_j \mid j \in J \,\rangle \le G_k$. □

Usually groups have many local systems of finitely generated subgroups. For any group property $\mathcal{X}$, the group G is *locally-$\mathcal{X}$* provided it has a local system, all of whose members enjoy the property $\mathcal{X}$. Especially a group that has a local system of finite subgroups is called a *locally finite group*.

Every finite group is locally finite, of course, but there are others. In particular, every torsion abelian group is locally finite, and we will soon encounter additional interesting infinite examples in Theorem (4.15).

(4.10). THEOREM. *Let G be a group. Then the following are equivalent:*

(1) *G is locally finite.*

(2) *Every finite subset of G generates a finite subgroup.*

(3) *Every finite subset of G is contained in a finite subgroup.* □

(4.11). THEOREM. (SCHMIDT'S THEOREM) *Let N be normal in the group G. Then G is locally finite if and only if G/N and N are locally finite.*

PROOF. If G is locally finite then so are its subgroup N and its image G/N. Now assume that normal N and $\bar{G} = G/N$ are locally finite. Let X be a finite subset of G, and set $H = \langle X\rangle$.

As $\bar{X}$ is a finite subset of locally finite $\bar{G}$, we have $\langle \bar{X}\rangle = \bar{H}$ finite. That is, HN/N is finite. By the Second Isomorphism Theorem, $H/H\cap N$ is finite, and especially $H\cap N$ has finite index in finitely generated H. By Schreier's Theorem (4.5)(a), the subgroup $H\cap N$ is also finitely generated. Within locally finite N this tells us that $H\cap N$ is finite. As both $H/H\cap N$ and $H\cap N$ are finite, H itself is finite, as required. □

Locally finite group theory admits many of the techniques of finite group theory, since the presence of the local system says that the group can be well approximated by its finite subgroups.

Dually, at times we may be able to approximate a group by its finite quotients. For any group property $\mathcal{X}$, the group G is *residually-$\mathcal{X}$* provided that for every nonidentity element g of G there is a normal subgroup N_g with $g \notin N_g$ for which the quotient G/N_g has property $\mathcal{X}$. Of special interest here is the class of groups that are *residually finite*. This does not imply that $\langle g \rangle$ is finite or intersects G/N_g trivially. Indeed $\mathbb{Z}$ is residually finite but has no subgroups of finite order. More generally:

(4.12). Theorem. (Schreier) *Every free group is residually finite.*

Proof. In the free group $\mathrm{F}(X)$ let

$$g = x_1 x_2 \cdots x_n$$

be a nonidentity element, written as a word in the various x and x^{-1} for $x \in X$, the only restriction being that we never have $\{x_i, x_{i+1}\} = \{x, x^{-1}\}$. (That is, $x_1 x_2 \cdots x_n$ is a reduced word in $\mathrm{F}(X)$.)

We define a map $x_i \mapsto \pi_i$ from $\mathrm{F}(X)$ to $\mathrm{Sym}(n+1)$ choosing each π_i to extend the mapping $i \mapsto i+1$. Notice that for different i, j we may have $\pi_i = \pi_j$, but our restriction on the x_i implies that nevertheless the choices can be made consistently. Having done this, we have a map $\pi \colon \mathrm{F}(X) \longrightarrow \mathrm{Sym}(n+1)$ in which $\pi(g) \neq 1_{\mathrm{F}(X)}$ since $1^{\pi(g)} = n+1$. Thus $\ker \pi = N_g$ is the desired finite index normal subgroup of $\mathrm{F}(X)$ with $g \notin N_g$. □

For residually finite groups, again a simple induction shows that for every finite subset S of G with $1_G \notin S$, there is a normal subgroup N_S of G with $S \cap N_S = \emptyset$.

(4.13). Theorem. *Let G be a group. Then the following are equivalent:*

(1) *G is residually finite.*

(2) *G is a subdirect product of finite groups.*

Proof. If G is residually finite, then the Chinese Remainder Theorem tells us that G is a subdirect product of the various finite groups G/N_g.

If G is a subdirect product of finite groups, then every element g of G projects nontrivially onto at least one of the finite quotients. Projection onto that coordinate then gives a homomorphism of G onto a finite group for which g is not in the kernel. □

Especially free groups are subdirect products of finite groups.

For an arbitrary group G, its *profinite topology* declares all subgroups of finite index index in G to be a base of open neighborhoods of the identity (with additional open sets being provided by the cosets of these subgroups). This topology is then Hausdorff precisely when G is residually finite. Many interesting results about finite p-groups as a class arise from study of the larger class of *pro-p groups*: those groups that are residually finite p-groups. Again free groups are examples, for all p, although this is harder to prove.

4.4 Representational finiteness

For a linear representation $\varphi\colon G \longrightarrow \mathrm{GL}_K(V)$, finiteness conditions may impose restrictions on the division ring K or on the K-space V.

The most common finiteness restriction made on a division ring K is that it have finite dimension over its center. Indeed later on we shall only consider division rings equal to their centers—fields. Even there additional restrictions may be of help. For instance, by Jordan canonical form a faithful $\mathbb{C}$-representation of a finite group can be realized over a finite Galois extension of the rationals. We will not spend much time on such matters, although later we will see that mileage can be gained by requiring K itself to be finite.

We have already noted that every group can be faithfully represented as a linear group. On the other hand many but not all infinite groups can be faithfully represented on finite dimensional vectors spaces. We have seen in Clifford's Theorem (3.33) and in Problem (6.36) that finite dimensionality of a representation is a useful finiteness condition. Indeed finite dimensionality of V is such an important hypothesis that, unlike us (see page 39), many (for instance, Robinson [Rob82]) reserve the term *linear group* for those groups that can be faithfully represented on a finite dimensional space. A weaker, but still useful, condition is that of *local linearity*, where G has a local system of subgroups, each of which has a faithful representation of finite dimension. This includes all locally finite groups and many other groups as well.

We now introduce groups that might be termed "internally locally linear" as they have faithful possibly infinite dimensional representations within which every finitely generated subgroup acts faithfully on some finite dimensional subspace.

If $g \in \mathrm{Sym}(\Omega)$, then the *support* of g on Ω, written $\mathrm{Supp}_\Omega(g)$, is the set $\{\, \omega \in \Omega \mid \omega^g \neq \omega \,\}$. The identity certainly has finite support. The inverse of an element of finite support also has finite support, as does the product of two elements of finite support. Therefore the elements of finite support,

$$\mathrm{FSym}(\Omega) \quad = \quad \{\, g \in \mathrm{Sym}(\Omega) \mid |\,\mathrm{Supp}_\Omega(g)| < \infty \,\},$$

form a subgroup of $\mathrm{Sym}(\Omega)$ called the *finitary symmetric group*. Of course, if Ω itself is finite, then $\mathrm{FSym}(\Omega) = \mathrm{Sym}(\Omega)$. If Ω is infinite then this is not the case, and $\mathrm{FSym}(\Omega)$ consists of the "nearly trivial" permutations.

(4.14). THEOREM. $\mathrm{FSym}(\Omega) \trianglelefteq \mathrm{Sym}(\Omega)$.

PROOF. By Lemma (3.5), conjugacy preserves the cardinality of the support of a permutation. □

If N is any nontrivial normal subgroup of $\mathrm{Sym}(\Omega)$ with $1 \neq n \in N$ and $1 \neq g \in \mathrm{FSym}(\Omega)$, then Lemma (3.5) also proves $[n,g] \in N \cap \mathrm{FSym}(\Omega)$. This observation is the beginning of the classification of all normal subgroups of $\mathrm{Sym}(\Omega)$. These include the normal subgroup $\mathrm{FSym}_\alpha(\Omega)$ consisting of all permutations with support of cardinality less than α, for each infinite cardinal α less than or equal to $|\Omega|$.

A fundamental result is

(4.15). THEOREM. *For every* Ω*, the group* $\mathrm{FSym}(\Omega)$ *is locally finite.*

PROOF. For the finite subset $\{g_1, \ldots, g_n\}$ of $\mathrm{FSym}(\Omega)$, let $H = \langle g_1, \ldots, g_n \rangle$ and $\Delta = \bigcup_{i=1}^n \mathrm{Supp}_\Omega(g_i)$, a finite subset of Ω which is H-invariant. We may identify the finite group $\mathrm{Sym}(\Delta)$ with the pointwise stabilizer of $\Omega \setminus \Delta$ in $\mathrm{FSym}(\Omega)$, and then the map $H \longrightarrow \mathrm{Sym}(\Delta) \leq \mathrm{FSym}(\Omega)$ is injective.

Thus every finite subset of G is contained in a finite subgroup, and G is locally finite by Theorem (4.10). □

The linear transformation $g \in \mathrm{GL}_D(V)$ is *finitary* provided the dimension of its commutator subspace $[V, g] = V(g-1)$ is finite.

Let $\mathrm{FGL}_D(V)$ be the set of all finitary elements of $\mathrm{GL}_D(V)$.

(4.16). LEMMA.

(a) $g \in \mathrm{GL}_D(V)$ *is finitary if and only if its fixed point space* $\mathrm{C}_V(g)$ *has finite codimension in* V*.*

(b) *For finitary* g*, the codimension of* $\mathrm{C}_V(g)$ *is equal to the dimension of* $[V, g]$*.*

(c) $\mathrm{FGL}_D(V)$ *is a normal subgroup of* $\mathrm{GL}_D(V)$*.*

PROOF. The subspace $[V, g]$ is the image of the endomorphism $g-1$ while its kernel $\ker(g-1)$ is equal to $\mathrm{C}_V(g)$. This gives the first two parts immediately.

If $\mathrm{C}_V(g)$ and $\mathrm{C}_V(h)$ both have finite codimension, then so does $\mathrm{C}_V(\langle g, h \rangle) = \mathrm{C}_V(g) \cap \mathrm{C}_V(h)$. Thus $\mathrm{FGL}_D(V)$ is a subgroup. It is normal as $[V, g]^h = [V, g^h]$. □

The corresponding group of "nearly trivial" linear transformations is

$$\mathrm{FGL}_K(V) = \{\, g \in \mathrm{GL}_K(V) \mid \dim_K(V^{g-1}) < \infty \,\}.$$

The group (as it is) $\mathrm{FGL}_K(V)$ is the *finitary linear group* and is normal in $\mathrm{GL}_K(V)$. In matrix terms, this group can be thought of as those invertible linear transformations g for which, with a suitable choice of basis, the matrix $g-1$ only has a finite number of nonzero rows. A group and representation are *stably linear* if they satisfy the stronger condition of having a fixed basis for which each $g-1$ has only a finite number of nonzero rows and nonzero columns—that is, a finite number of nonzero entries.

Finitary linear groups need not be locally finite, but groups that are both locally finite and finitary linear have a beautiful and well-understood structure. These include $\mathrm{FSym}(\Omega)$.

(4.17). THEOREM. $\mathrm{FSym}(\Omega)$ *is finitary linear over every* K*.*

PROOF. Consider $g \in \mathrm{FSym}(\Omega)$ acting on the permutation module $K\Omega$. For $\omega \in \Omega$, we have $e_\omega^{g-1} = 0$ if and only if ω is fixed by g. That is, the only elements of the K-basis $\{\, e_\omega \mid \omega \in \Omega \,\}$ that have nontrivial image under $g-1$ are those with ω in the support of g. For elements g of $\mathrm{FSym}(\Omega)$ this support is finite, therefore V^{g-1} has a finite spanning set and $g \in \mathrm{FGL}(K)K\Omega$. □

4.5 Chain conditions

Chain conditions illustrate both "finiteness" and "reduction" in that the groups considered exhibit finiteness properties precisely because they admit some type of reduction.

In any partially ordered set $(I\ \succeq)$, the *descending chain condition* states that any chain

$$a_1 \succeq a_2 \succeq a_3 \succeq \cdots$$

stabilizes at some point: there is an $N \in \mathbb{Z}^+$ with $a_i = a_j$, for all $i, j \geq N$. Equivalently we have the *minimal condition*: every nonempty subset of I contains at least one minimal element. In the opposite partially ordered set $(I, \preceq)$ where $a_1 \preceq a_2$ if and only if $a_1 \succeq a_2$ in $(I\ \succeq)$, these become the *ascending chain condition* and *maximal condition.*

For us, there are group theoretic and representation theoretic versions of each. Clifford's Theorem (3.33) is valid for all G-modules with the minimal condition on submodules.

For a given group G, there are several posets that may be considered: those of all subgroups of G, all normal subgroups of G, and all subnormal subgroups of G.

The Jordan-Hölder Theorem (1.2) addresses *composition series* for the group G—those chains of subnormal subgroups

$$G = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_k = 1$$

with each quotient G_{i-1}/G_i simple—stating that the multiset of simple factors is uniquely determined up to isomorphism. If instead we consider the poset of normal subgroups of a group, a chain that cannot be refined is a *chief series.* In this context there is also the appropriate Jordan-Hölder Theorem. (See [Rob82, Theorem 3.1.4].) The corresponding factors are *chief factors*, and the next result gives their structure in many situations of interest (for instance when G is finite).

A *minimal normal subgroup* of G is a subgroup N with $1 \neq N \trianglelefteq G$ and such that, whenever $W \trianglelefteq G$ with $W \leq N$, either $1 = W$ or $W = N$. For finite groups these must exist by order arguments, but in general they might not. The group $(\mathbb{Z}, +)$ has no minimal normal subgroups.

(4.18). Theorem. *Let G be a group that satisfies the minimal condition on subnormal subgroups and N minimal normal in G. Then there is an index set I and a set of subgroups $\{\, S_i \mid i \in I \,\}$ such that*

(a) *each S_i is simple and minimal normal in N;*

(b) *for each $i, j \in I$, there is a $g_{ij} \in G$ with $S_i^{g_{ij}} = S_j$;*

(c) $N = \langle\, S_i \mid i \in I \,\rangle = \bigoplus_{i \in I} S_i$.

Proof. This theorem and its proof should be compared with Clifford's Theorem (3.33), the proposition on completely reducible modules that precedes it, and their proofs.

Let S be a minimal normal subgroup of N. Then $\langle\, S^g \mid g \in G\,\rangle$ is clearly nontrivial, normal in G, and contained in N; so $N = \langle\, S^g \mid g \in G\,\rangle$. Choose a set of conjugates $\{\, S_i = S^{g_i} \mid i \in I\,\}$ that generates N and is minimal subject to this. That is, $N = \langle\, S_i \mid i \in I\,\rangle$ and $N > S_J = \langle\, S_j \mid j \in J\,\rangle$, for all $J \subsetneq I$. (Such a set exists by Zorn's Lemma.)

As the S_j are conjugates of S, they too are minimal normal in N; and each S_J is normal in N. Set $g_{ij} = g_i^{-1}g_j$. We now have (b) and half of (a).

For each i, the subgroup $S_i \cap S_{I\setminus i}$ is normal in N but is not equal to S_i by minimality of I. Since S_i is minimal normal in N, we find $S_i \cap S_{I\setminus i} = 1$, giving (c) by Theorem (2.34).

Finally, if $T \ntrianglelefteq S_i$, then

$$N_N(T) \geq \langle S_i, S_{I\setminus i}\rangle = N\,,$$

by (c). As S_i is minimal normal in N, this forces $T = 1$. Therefore S_i is simple, completing (a) and the theorem. □

A characteristically simple group G is minimal normal in the split extension of G by $\mathrm{Aut}(G)$. In particular in appropriate situations (for instance for finite G) the theorem describes all characteristically simple groups. (Compare Problem (2.46).)

4.6 Problems

(4.19). Problem. *Prove that a group with a local system of simple subgroups is simple.*

(4.20). Problem. *Recall that a group G is quasisimple if it is perfect $G = G'$ and $G/\mathrm{Z}(G)$ is simple.*

(a) *Prove that the nontrivial group G is quasisimple if and only if for every pair of elements g and h in G with $g \notin \mathrm{Z}(G)$ it is possible to write h as the product of a finite number of conjugates of g and g^{-1}.*

(b) *Prove that a group with a local system of quasisimple subgroups is quasisimple.*

(4.21). Problem. *Let G be a simple and locally finite group. Prove that for every finite subgroup H of G, there are finite subgroups F and N of G with*

(i) *$F \trianglerighteq N$;*

(ii) *F/N simple;*

(iii) *$H \leq F$ and $H \cap N = 1$.*

Remark. *This important observation, due to Kegel, says that every finite subgroup H of the locally finite simple group G can be "covered" by a finite simple section F/N of G. This is a "finite" version of P. Hall's "countable" result Theorem (4.7).*

Hint: *Consider $\mathcal{S}^2(H)$.*

(4.22). PROBLEM. *Let F be a division ring. Prove that F is a locally finite division ring if and only if F is an algebraic extension of the ground field $\mathbb{F}_p$, for some prime p. In particular, a locally finite division ring is a countable field. (Here, by a* locally finite division ring*, we mean a division ring in which every finite subset is contained in some finite sub-division ring.)*

HINT: *You may use Wedderburn's theorem that all finite division rings are fields.*

(4.23). PROBLEM.

(a) *Prove that if G is a permutation group on Ω with all orbits finite, then $G/\ker_\Omega(G)$ is residually finite.*

(b) *An FC*-group *is a group in which all conjugacy classes are finite. Abelian groups are examples; for instance $\mathbb{Z}$ is an FC-group that is residually finite, but $(\mathbb{Q}, +)$ is an FC-group that is not residually finite. (Indeed $(\mathbb{Q}, +)$ has no subgroups of finite index.) Prove that if G is a FC-group, then $G/Z(G)$ is residually finite.*

(4.24). PROBLEM.

(a) *Let $g \in \mathrm{Sym}(\Omega)$ have the orbit Δ on Ω and set $W = \oplus_{\delta\in\Delta} K e_\delta$, a $\langle g\rangle$-invariant subspace in the action on the permutation module $K\Omega$. Prove that W^{g-1} has K-dimension $|\Delta| - 1$.*

(b) *Prove* $\mathrm{FSym}(\Omega) = \mathrm{Sym}(\Omega) \cap \mathrm{FGL}_K(K\Omega)$.

Chapter 5

Symmetric and alternating groups

5.1 Transpositions

In $\mathrm{Sym}(\Omega)$ and its subgroup $\mathrm{FSym}(\Omega)$ a *transposition* is a 2-*cycle* (a, b) for distinct $a, b \in \Omega$.

(5.1). Lemma.

(a) *The transpositions form a conjugacy class of elements of order* 2 *in* $\mathrm{Sym}(\Omega)$ *and* $\mathrm{FSym}(\Omega)$.

(b) *If* $g = (a, b)$ *and* $h = (c, d)$*, then, respectively,*

$$|gh| = 1, 2, 3 \quad \textit{as} \quad |\{a, b\} \cap \{c, d\}| = 2, 0, 1 \ .$$

In the last case $\langle (a, b), (b, c) \rangle = \mathrm{Sym}(\{a, b, c\}) \leq \mathrm{FSym}(\Omega)$.

(c) *We have*

$$\mathrm{C}_{\mathrm{Sym}(\Omega)}(\,(a, b)\,) = \langle (a, b) \rangle \times \mathrm{Sym}(\Omega \setminus \{a, b\})$$

and

$$\mathrm{C}_{\mathrm{FSym}(\Omega)}(\,(a, b)\,) = \langle (a, b) \rangle \times \mathrm{FSym}(\Omega \setminus \{a, b\})\,. \qquad \square$$

Here for each subset Δ of Ω, we have identified $\mathrm{Sym}(\Delta)$ with the subgroup of $\mathrm{Sym}(\Omega)$ that fixes $\Omega \setminus \Delta$ pointwise (and so for the corresponding finitary subgroups as well). This is standard, and we will continue to make this identification without comment.

The elements $(1, 2)(1, 3) = (1, 2, 3)$ of Lemma (5.1)(b) are the 3-*cycles* of $\mathrm{Sym}(\Omega)$ while the products $(1, 2) \cdot (3, 4) = (1, 2)(3, 4)$ are the 2^2-*elements*.

(5.2). Theorem.

(a) *The subgroup of* $\mathrm{Sym}(\Omega)$ *generated by the class of transpositions is* $\mathrm{FSym}(\Omega)$.

(b) *If G is a subgroup of* $\mathrm{FSym}(\Omega)$ *that is generated by transpositions, then, for the partition* $\Omega = \biguplus_{i\in I} \Omega_i$ *of* Ω *into distinct* G*-orbits* Ω_i*, we have* $G = \bigoplus_{i\in I} \mathrm{FSym}(\Omega_i)$*.*

Proof. Each transposition is finitary, so the group they all generate is contained in the finitary symmetric group. For (a) it remains to observe that

$$(1,2,3,\dots,m) = (1,2)(1,3)\cdots(1,m)\,.$$

Now let $G = \langle X\rangle \leq \mathrm{FSym}(\Omega)$ with X a set of transpositions, and let Y be the set of all transpositions that are in G.

Define a graph with vertex set Ω and $a \sim b$ if and only if (a,b) is one of the transpositions in Y. Let the connected components of Ω in this graph be Ω_i for $i \in I$. We claim that $G = \bigoplus_{i\in I} \mathrm{FSym}(\Omega_i)$. The Ω_i are the orbits of G on Ω, so G is contained in this direct sum. It remains to prove that G contains each $\mathrm{FSym}(\Omega_i)$.

Let distinct a, b be in the connected component Ω_i, and select

$$a = a_0 \sim a_1 \sim \cdots \sim a_m = b\,,$$

an Ω_i-path from a to b, chosen to be as short as possible.

If $m > 1$, then we have

$$a = a_0 \sim a_1 \sim a_2 \sim \cdots \sim a_m = b\,.$$

But then, with $f = (a_0, a_1)$ and $g = (a_1, a_2)$, both transpositions of G and Y, we have $f^g = (a_0, a_2)$ in G and hence in Y. But then the path

$$a = a_0 \sim a_2 \sim \cdots \sim a_m = b\,,$$

is shorter than the original one. This is a contradiction, and so $m = 1$. That is, $(a_0, a_m) = (a, b)$ is a transposition of Y and G, and this holds for arbitrary a, b from Ω_i. Therefore all transpositions with support from Ω_i belong to Y and G, hence by (a) the subgroup $\mathrm{FSym}(\Omega_i)$ is in G. □

5.2 The Weyl group $\mathrm{W}(A_k)$

For $k \in \mathbb{Z}^+$ consider the group

$$\mathrm{W}(A_k) = \langle\, a_1\,, \dots, a_k \mid a_i^2 = 1\,,\ (a_i a_{i+1})^3 = 1\,,\ (a_i a_j)^2 = 1 \text{ for } |i-j| > 1\,\rangle\,,$$

the *Coxeter group* of *type* A_k. The W stands for Weyl. This groups is also the *Weyl group* of type A_k, but that involves different definitions.

We first discuss some elementary consequences of these relations. Under the map sending every a_i to $-1 \in \{\pm 1\}$ all relations are satisfied, so there is a homomorphism onto a group of order 2. In particular all of the elements a_i have order exactly two in $\mathrm{W}(A_k)$ rather than one.

The relation $(a_ia_j)^2 = 1$ states $a_ia_ja_ia_j = 1$. As all a_i have order 2, this becomes

$$a_ia_j = a_ja_i\,;$$

that is, if $|i-j| > 1$ then the elements a_i and a_j commute in $\mathrm{W}(A_k)$. Similarly $(a_ia_{i+1})^3 = 1$ reads as $a_ia_{i+1}a_ia_{i+1}a_ia_{i+1} = 1$ or

$$a_ia_{i+1}a_i = a_{i+1}a_ia_{i+1}\,,$$

the so-called braid relation; see Problem (5.23).

(5.3). Theorem. *The map $f\colon a_i \mapsto (i, i+1)$, for $1 \le i \le k$, extends to an isomorphism φ of* $\mathrm{W}(A_k)$ *and* $\mathrm{Sym}(k+1)$*.*

Proof. By Lemma (5.1) the given transpositions in $\mathrm{Sym}(k+1)$ satisfy the corresponding relations of $\mathrm{W}(A_k)$, and by Theorem (5.2) these transpositions generate all of $\mathrm{Sym}(k+1)$. Therefore the map f extends at least to a surjective homomorphism φ from $\mathrm{W}(A_k)$ onto $\mathrm{Sym}(k+1)$.

Set $W_k = \mathrm{W}(A_k)$. We claim:

(i) For $B = \langle a_1, \dots, a_{k-1}\rangle \le W_k$ and $n = a_k$, we have $W_k = B \cup BnB$.

(ii) $|W_k| \le (k+1)!\,.$

Once we have claim (ii) we will be done, since by the previous paragraph we already have $|W_k| \ge (k+1)! = |\mathrm{Sym}(k+1)|$.

Our proof of the claims will be by induction on k. For $k = 1$ we have $B = 1$ and

$$W_1 = \{1\} \cup \{1n1\} = \{1, a_1\} \simeq Z_2 \simeq \mathrm{Sym}(1+1)\,.$$

We now assume $k \ge 2$.

The image of B under φ is $\mathrm{Sym}(k)$. Therefore by induction $B \simeq \mathrm{Sym}(k)$ and also its subgroup $C = \langle a_1, \dots, a_{k-2}\rangle$ is isomorphic to $\mathrm{Sym}(k-1)$. Furthermore, by induction (or direct calculation) $B = C \cup CmC$ for $m = a_{k-1}$.

By the relations for $W_k = \mathrm{W}(A_k)$,

$$nmn = a_ka_{k-1}a_k = a_{k-1}a_ka_{k-1} = mnm$$

and

$$[C, n] = [\langle a_1, \dots, a_{k-2}\rangle, a_k] = 1\,.$$

Under part (i) of the claim, as $B = \langle a_1, \dots, a_{k-1}\rangle$ and $1n1 = a_k$, the subset $B \cup BnB$ contains the generating set for W_k; so it is enough to prove that $B \cup BnB$ is a subgroup. Clearly $1 \in B$; $B^{-1} = B$; $(BnB)^{-1} = Bn^{-1}B = BnB$; $B.B = B$; $B.BnB = BnB$; and $BnB.B = B$; so the only thing that must be verified at length is $BnB.BnB \subseteq B \cup BnB$. Indeed

$$\begin{aligned} BnB.BnB &= BnBnB \\ &= Bn(C \cup CmC)nB \\ &= B(nCn \cup nCmCn)B \\ &= B(C \cup CnmnC)B \\ &= B(C \cup CmnmC)B \\ &= B \cup BmnmB \\ &= B \cup BnB\,, \end{aligned}$$

as desired.

For (ii) we have $|B| = k!$, and so the list of all triples (b_1, n, b_2) with $b_1, b_2 \in B$ has length $(k!)^2$. For $c \in C$, the two list members (b_1, n, b_2) and (b_1c^{-1}, n, cb_2) satisfy $b_1nb_2 = b_1c^{-1}ncb_2$ as c commutes with n. Thus

$$|BnB| \leq (k!)^2/|C| = (k!)^2/(k-1)! = k \cdot k!\,.$$

Therefore by (i)

$$|W_k| = |B \cup BnB| \leq k! + k \cdot k! = (k+1)k! = (k+1)!\,,$$

completing our proof of claim (ii) and so of the theorem. □

5.3 The alternating group and simplicity

The *alternating group* $\mathrm{Alt}(\Omega)$ is the subgroup of $\mathrm{FSym}(\Omega)$ consisting of all finitary permutations that can be written as a product of an even number of transpositions.

(5.4). Theorem. *For $|\Omega| \geq 2$, the alternating group $\mathrm{Alt}(\Omega)$ is a normal subgroup of index 2 in $\mathrm{FSym}(\Omega)$. Indeed $\mathrm{Alt}(\Omega) = \mathrm{FSym}(\Omega)'$ is the unique subgroup of index 2 in $\mathrm{FSym}(\Omega)$.*

Proof. First consider the case $|\Omega| = n$, finite. Then by Theorem (5.3) we have

$$\mathrm{FSym}(\Omega) = \mathrm{Sym}(\Omega) \simeq \mathrm{Sym}(n) \simeq \mathrm{W}(A_{n-1})\,.$$

Thus, as noted before, there is a surjective homomorphism from $\mathrm{W}(A_{n-1})$ to $\{\pm 1\}$ given by $a_i \mapsto -1$, and this provides a surjective homomorphism

$$\mathrm{sgn}\colon \mathrm{Sym}(n) \longrightarrow Z_2\,.$$

Each a_i of $\mathrm{W}(A_{n-1})$ maps to a transposition of $\mathrm{Sym}(n)$, so the kernel of sgn has index 2 in $\mathrm{Sym}(n)$ and consists of all products of an even number of transpositions. That is, $\ker \mathrm{sgn} = \mathrm{Alt}(n)$ is normal of index 2.

Now consider arbitrary Ω. As Ω is the directed limit of its finite subsets Δ, $\mathrm{FSym}(\Omega)$ is the directed limit of its finite subgroups $\mathrm{FSym}(\Delta)$, as in Theorem

(4.15). Each embedding of $\mathrm{FSym}(\Delta)$ into $\mathrm{FSym}(\Omega)$ takes the transpositions of $\mathrm{FSym}(\Delta)$ to transpositions of $\mathrm{FSym}(\Omega)$. Therefore the various sign homomorphisms sgn_Δ combine to give a uniform sign homomorphism sgn_Ω on $\mathrm{FSym}(\Omega)$ with kernel $\mathrm{Alt}(\Omega)$.

The group $\mathrm{FSym}(\Omega)$ is generated by its transposition class by Theorem (5.2)(a), so its derived quotient has order at most 2 by Proposition (2.17). The subgroup $\mathrm{Alt}(\Omega)$ has index 2 in $\mathrm{FSym}(\Omega)$. If $N \neq \mathrm{Alt}(\Omega)$ also had index 2, then $N \cap \mathrm{Alt}(\Omega)$ would be normal with an abelian quotient of order 4. The contradiction proves that $\mathrm{Alt}(\Omega) = \mathrm{FSym}(\Omega)'$ is the unique subgroup of index 2 in $\mathrm{FSym}(\Omega)$. □

(5.5). Corollary. *For $n \geq 2$, the group $\mathrm{Alt}(n) = \mathrm{Alt}(\{1, 2, \dots, n\})$ has order $n!/2$.* □

The homomorphism

$$\mathrm{sgn}\colon \mathrm{FSym}(\Omega) \longrightarrow Z_2 .$$

of the theorem, which takes each transposition to $-1 \in \{\pm 1\} \simeq Z_2$, is called the *sign* homomorphism. The elements of its kernel $\mathrm{Alt}(\Omega)$ are the *even permutations* of Ω while those of $\mathrm{FSym}(\Omega)$ that are not in $\mathrm{Alt}(\Omega)$ are the *odd permutations.*

(5.6). Proposition.

(a) *$\mathrm{Alt}(\Omega)$ is generated by its 3-cycles and its 2^2-elements.*

(b) *If $|\Omega| > 4$, then $\mathrm{Alt}(\Omega)$ is generated by its conjugacy class of 3-cycles.*

(c) *If $|\Omega| > 4$, then $\mathrm{Alt}(\Omega)$ is generated by its conjugacy class of 2^2-elements.*

Proof. (a) If $g = \prod_{i=1}^{2m} t_i$ then $g = \prod_{j=1}^{m} (t_{2j-1} t_{2j})$, so this follows from Lemma (5.1).

(b) By Lemma (3.5) the 3-cycles form a single conjugacy class in $\mathrm{FSym}(\Omega)$. As the odd permutation $(4, 5)$ centralizes $(1, 2, 3)$, this is also a single class in $\mathrm{Alt}(\Omega)$ provided $|\Omega| > 4$. Then $(1, 3, 2)(3, 2, 4) = (1, 2)(3, 4)$, so (a) gives the result.

(c) By Lemma (3.5), in $\mathrm{Sym}(\Omega)$ the 2^2-elements form a single conjugacy class. As the odd permutation $(1, 2)$ centralizes $(1, 2)(3, 4)$, the collection of 2^2-elements remains a single class in $\mathrm{Alt}(\Omega)$. Furthermore $(1, 2)(4, 5) \cdot (1, 3)(4, 5) = (1, 2, 3)$, so again (a) gives the result. □

Of course, for $n = 3, 4$, the group $\mathrm{Alt}(n)$ is still generated by its 3-cycles, but a 3-cycle is not conjugate to its inverse in these two groups.

(5.7). Proposition. *For $n \leq 4$, the group $\mathrm{Sym}(n)$ is solvable.*[1] □

(5.8). Theorem. *If $|\Omega| > 4$, then $\mathrm{Alt}(\Omega)$ is nonabelian and simple.*

[1] As Galois knew!

Proof. Let $1 \neq n \in N \trianglelefteq \mathrm{Alt}(\Omega)$ and $a, b, c \in \Omega$ with $a^n = b \neq a$ and $b^n = c$. Choose $d \notin \{a, b, c\}$ so that for $h = (a, b, d) \in \mathrm{Aut}(\Omega)$ we have $h^n = (b, c, e)$ where $e = d^n$. Then

$$g = [h, n] = h^{-1}h^n = (a, d, b)(b, c, e) \in N \cap \mathrm{Alt}(\{a, b, c, d, e\}) .$$

As $d^g = c \neq d$, we have $g \neq 1$. In particular $\mathrm{Alt}(\Omega)$ is not abelian. The element g must have cycle type one of 3, 2^2, or 5. In the first two cases we find $N = \mathrm{Alt}(\Omega)$ by Proposition (5.6), therefore we may assume that a, b, c, d, e are distinct and so $g = (a, d, c, e, b)$.

Now let $k = (a, d, c)$. Then

$$[g, k] = g^{-1}g^k = (b, e, c, d, a)(d, c, a, e, b) = (a, d, e) \in N ,$$

and again $N = \mathrm{Alt}(\Omega)$ by Proposition (5.6). □

(5.9). Corollary. *If $|\Omega| > 4$, then* $\mathrm{Alt}(\Omega)$ *is the unique minimal normal subgroup in* $\mathrm{Sym}(\Omega)$.

Proof. A permutation of Ω that maps every 3-subset to itself must then map every point to itself (as the intersection of 3-subsets, using $|\Omega| > 3$). Therefore by Lemma (3.5) and Proposition (5.6)(b) we have $\mathrm{C}_{\mathrm{Sym}(\Omega)}(\mathrm{Alt}(\Omega)) = 1$.

For $1 \neq n \in N \trianglelefteq \mathrm{Sym}(\Omega)$ choose a $g \in \mathrm{Alt}(\Omega)$ that does not commute with n. Then

$$1 \neq [g, n] = g^{-1}(n^{-1}gn) = (g^{-1}n^{-1}g)n \in \mathrm{Alt}(\Omega) \cap N .$$

Therefore $\mathrm{Alt}(\Omega) \cap N$ is a nontrivial normal subgroup of simple $\mathrm{Alt}(\Omega)$, hence $\mathrm{Alt}(\Omega) \leq N$. □

5.4 Geometry and automorphisms

This section is focused on proof of:

(5.10). Theorem. *For $|\Omega| \neq 2, 6$ we have*

$$\mathrm{Aut}(\mathrm{Sym}(\Omega)) = \mathrm{Aut}(\mathrm{FSym}(\Omega)) = \mathrm{Sym}(\Omega) .$$

Additionally $\mathrm{Aut}(\mathrm{Sym}(2)) = 1$ *and* $\mathrm{Sym}(6)$ *has index at most* 2 *in* $\mathrm{Aut}(\mathrm{Sym}(6))$.

The proof of the theorem follows a standard and important model: we define a geometric object upon which the group acts; we then reconstruct the geometry within the group; we finally identify the group's automorphisms within the geometry's automorphisms. The process is not always precise (as in the present case for $|\Omega| = 6$),[2] but in practice this allows us to locate a large portion of the

[2] Indeed, one of the most compelling aspects of finite group theory and geometry is the way in which a small number of anomalous or sporadic examples force themselves upon us in the midst of a general result.

automorphism group of the group within the more manageable automorphism group of the geometry. For instance, here the geometry is that of the underlying set for the symmetric group as structured by its set of unordered pairs.

The general model is our motivation for a somewhat grandiose name:

(5.11). Theorem. (Fundamental Theorem of Set Geometry) *Let* $|\Omega| \geq 5$. *Consider the graph* $K(\Omega, 2)$ *whose vertex set is the set* $\binom{\Omega}{2}$ *with* $\{i,j\}$ *adjacent to* $\{k,l\}$ *precisely when* $|\{i,j\} \cap \{k,l\}| = 1$. *Then* $\mathrm{Aut}(K(\Omega,2)) = \mathrm{Sym}(\Omega)$ *with the natural action.*

Proof. As $|\Omega| \geq 3$ the group $\mathrm{Sym}(\Omega)$ acts naturally and faithfully on $K(\Omega, 2)$.

The maximal cliques (maximal complete subgraphs) of $K(\Omega, 2)$ are of two distinct types:

$$T_{a,b,c} = \{\{a,b\},\{b,c\},\{a,c\}\} \text{ for } \{a,b,c\} \in \binom{\Omega}{3}$$

of cardinality 3 and

$$C_a = \{\, \{a,w\} \mid w \in \Omega \setminus \{a\} \,\} \text{ for } a \in \Omega\,.$$

of cardinality $|\Omega| - 1 \gneq 3$. Automorphisms of the graph must take maximal cliques to maximal cliques. Therefore for any $g \in \mathrm{Aut}(K(\Omega,2))$ and each $a \in \Omega$, there is a unique $a^g \in \Omega$ with $C_a^g = C_{a^g}$. This gives a natural action of the automorphism group on Ω and a surjective homomorphism $\mathrm{Aut}(K(\Omega,2)) \longrightarrow \mathrm{Sym}(\Omega)$.

The kernel of this homomorphism fixes globally each clique C_a. But then, for each pair $a, b \in \Omega$, the kernel must fix $\{a,b\} = C_a \cap C_b$. Therefore every pair $\{a,b\} \in \binom{\Omega}{2}$ is fixed, and the kernel is trivial. We conclude that $\mathrm{Aut}(K(\Omega,2))$ and $\mathrm{Sym}(\Omega)$ are isomorphic. □

(5.12). Proposition. *Let* T *be the conjugacy class of transpositions in* $\mathrm{Sym}(\Omega)$ *with* $|\Omega| \geq 3$. *The noncommuting graph* Γ *of* T *is the graph whose vertices are the members of* T *with* a *and* b *adjacent when* a *and* b *do not commute.*

(a) $\Gamma \simeq K(\Omega, 2)$.

(b) *The subgroup of* $\mathrm{Aut}(\mathrm{Sym}(\Omega))$ *that stabilizes the class* T *of transpositions is equal to* $\mathrm{Sym}(\Omega)$.

(c) *The subgroup of* $\mathrm{Aut}(\mathrm{FSym}(\Omega))$ *that stabilizes the class* T *of transpositions is equal to* $\mathrm{Sym}(\Omega)$.

Proof. We have $T = \{\, (a,b) \mid \{a,b\} \in \binom{\Omega}{2} \,\}$. By Lemma (5.1) the two transpositions (i,j) and (k,l) do not commute if and only if $|\{i,j\} \cap \{k,l\}| = 1$. Therefore Γ is isomorphic to $K(\Omega, 2)$, as in (a).

In (b) and (c) let G be, respectively, the stabilizer in $\mathrm{Aut}(\mathrm{FSym}(\Omega))$ and $\mathrm{Aut}(\mathrm{Sym}(\Omega))$ of the class T. By Lemma (3.5) we have $\mathrm{C}_{\mathrm{Sym}(\Omega)}(\mathrm{FSym}(\Omega)) = 1$, and we may identify $\mathrm{Sym}(\Omega)$ with its image in G.

By (a) and Theorem (5.11), if $\Omega \geq 5$ then G induces automorphisms of Γ and $K(\Omega, 2)$; and we have a surjective homomorphism $\varphi\colon G \longrightarrow \mathrm{Sym}(\Omega)$ with natural action. The kernel of this action fixes all vertices of Γ, the class of transpositions. This remains true for $|\Omega| = 3$, as in Sym(3) the point stabilizers are the transpositions plus the identity, and also for $\Omega = 4$, as in Sym(4) the point stabilizers are the Sylow 3-normalizers, isomorphic to Sym(3).

Thus the kernel of φ acts trivially on $\mathrm{FSym}(\Omega)$, the group generated by all transpositions (a, b) (by Theorem (5.2)). In particular, we are done in (b), where $G \leq \mathrm{Aut}(\mathrm{FSym}(\Omega))$.

Now suppose g is an element of $G \leq \mathrm{Aut}(\mathrm{Sym}(\Omega))$ that is in the kernel of φ and so acts trivially on $\mathrm{FSym}(\Omega)$. For arbitrary $s \in \mathrm{Sym}(\Omega) \leq G$, the commutator $[s, g] = s^{-1}s^g$ is in $\mathrm{Sym}(\Omega)$ but remains in the kernel of φ. As already noted $\mathrm{C}_{\mathrm{Sym}(\Omega)}(\mathrm{FSym}(\Omega)) = 1$, so we conclude $[s, g] = 1$. Therefore g fixes all elements s of $\mathrm{Sym}(\Omega)$, and again $\ker\varphi$ is trivial. This completes (c). □

(5.13). PROPOSITION. *If $d \in \mathrm{Sym}(\Omega)$ is an element of order 2 having at least k orbits of length 2 on Ω, then there is a normal elementary abelian 2-subgroup of order at least 2^k that is normal in the centralizer of d in both $\mathrm{FSym}(\Omega)$ and $\mathrm{Sym}(\Omega)$.*

PROOF. The centralizer of the element $(a_1, a_2) \cdots (a_{2k-1}, a_{2k}) \cdots$ has the normal subgroup $\langle (a_1, a_2), \ldots, (a_{2k-1}, a_{2k}), \ldots \rangle$ both in $\mathrm{FSym}(\Omega)$ and in $\mathrm{Sym}(\Omega)$. □

(5.14). LEMMA. *For $|\Omega| \geq 2$ with $|\Omega| \neq 6$, the conjugacy class of transpositions is stabilized by $\mathrm{Aut}(\mathrm{FSym}(\Omega))$ and by $\mathrm{Aut}(\mathrm{Sym}(\Omega))$.*

PROOF. By Lemma (5.1) the centralizers of (a, b) in $\mathrm{FSym}(\Omega)$ and $\mathrm{Sym}(\Omega)$ are, respectively, $\langle (a, b)\rangle \times \mathrm{Sym}(\Omega \setminus \{a, b\})$ and $\langle (a, b)\rangle \times \mathrm{FSym}(\Omega \setminus \{a, b\})$. Thus, by Theorem (5.8) and Corollary (5.9), for $|\Omega| \geq 7$ the largest normal 2-subgroup of the centralizer of a transposition in $\mathrm{FSym}(\Omega)$ has order 2. On the other hand, for any other element of order 2 in $\mathrm{FSym}(\Omega)$, the corresponding centralizer has a normal 2-subgroup of larger order by the previous proposition. Therefore all automorphisms of these groups cannot take a transposition to an element of order 2 with support of size greater than 2 and must take transpositions to transpositions.

For $|\Omega| \in \{2, 3, 4, 5\}$ the transpositions are the only elements of order 2 in $\mathrm{Sym}(\Omega)$ but not in $\mathrm{Alt}(\Omega) = \mathrm{Sym}(\Omega)'$, so again automorphisms must take transpositions to transpositions. □

PROOF OF THEOREM (5.10).

Certainly $\mathrm{Aut}(\mathrm{Sym}(2)) = 1$. Also an automorphism of Sym(6) must fix Alt(6) (say, by Corollary (5.9)). But $\mathrm{Sym}(6) \setminus \mathrm{Alt}(6)$ has two conjugacy classes of involutions—the transpositions and those of cycle type 2^3. Therefore the index of Sym(6) in Aut(Sym(6)) is at most 2 by Proposition (5.12).

We may now assume $|\Omega| \neq 2, 6$. Thus by Lemma (5.14) both $\mathrm{Aut}(\mathrm{Sym}(\Omega))$ and $\mathrm{Aut}(\mathrm{FSym}(\Omega))$ stabilize the class of transpositions. By Proposition (5.12) again, $\mathrm{Sym}(\Omega)$ is the full automorphism group of $\mathrm{Sym}(\Omega)$ and $\mathrm{FSym}(\Omega)$. □

5.5 Problems

(5.15). Problem. *Let $n \in \mathbb{Z}^+$. A subgroup of* $\mathrm{Sym}(n)$ *is* $(n-2)$*-transitive if and only if it contains* $\mathrm{Alt}(n)$.

Remark. $\mathrm{Sym}(6)$ *contains a* 3*-transitive subgroup that does not contain* $\mathrm{Alt}(6)$*; see Problem (5.20) below.*

(5.16). Problem. (Thompson Transfer) *Assume that the finite group G has a Sylow* 2*-subgroup S containing subgroup T of index* 2 *in S and an element s of order* 2 *in S with $s^G \cap T = \emptyset$. Prove that G has a normal subgroup of index* 2 *that does not contain s.*

Hint: *Consider s in the permutation action of G on the cosets of T.*

(5.17). Problem. *Prove that the finite group G with a cyclic Sylow* 2*-subgroup S has a normal subgroup N of odd order with $G = N \rtimes S$.*

Hint: *Consider the action of S in the right regular representation of G.*

(5.18). Problem. *Prove that if G is a subgroup of* $\mathrm{Alt}(\Omega)$ *that is generated by* 3*-cycles, then, for the partition $\Omega = \biguplus_{i\in I} \Omega_i$ of Ω into distinct G-orbits Ω_i, we have $G = \bigoplus_{i\in I} \mathrm{Alt}(\Omega_i)$.*

(5.19). Problem. *Prove that for $|\Omega| \geq 4$ but $|\Omega| \neq 6$ we have*

$$\mathrm{Aut}(\mathrm{Alt}(\Omega)) = \mathrm{Sym}(\Omega)\,.$$

Also $\mathrm{Sym}(6)$ *has index at most* 2 *in* $\mathrm{Aut}(\mathrm{Alt}(6))$.

(5.20). Problem.

(a) *Let G be* $\mathrm{Sym}(5)$ *or* $\mathrm{Alt}(5)$*. Prove that the normalizer of a Sylow* 5*-subgroup has index* 6 *in G.*

(b) *Prove that* $\mathrm{Sym}(6)$ *contains a* 3*-transitive subgroup that does not contain* $\mathrm{Alt}(6)$.

(c) *Prove that* $[\mathrm{Aut}(\mathrm{Sym}(6)){:}\mathrm{Sym}(6)] \geq 2$ *and* $[\mathrm{Aut}(\mathrm{Alt}(6)){:}\mathrm{Sym}(6)] \geq 2$*. (Hence we have equality by Theorem (5.10) and the previous problem.)*

(5.21). Problem. *Let $G = \langle D \rangle$ be generated by the conjugacy class $D = d^G$ of elements of order* 2 *with the property that:*

$$\textit{for all } d, e \in D \textit{ we have } |de| \in \{1,2,3\}\,.$$

Such a conjugacy class is called a class of 3-transpositions *in the* 3-transposition group *G. This concept and terminology, due to* B. Fischer*, arise because a basic example is the class of transpositions in* $\mathrm{FSym}(\Omega)$*; see Lemma (5.1).*

Consider the special case in which we additionally have:

> **(Sym(5))** For $a, b, c, d \in D$ with connected diagram, the subgroup $\langle a, b, c, d\rangle$ is isomorphic to $\mathrm{Sym}(k)$ for $k \leq 5$ with a, b, c, d all being mapped to transpositions.

Assume also that there is a subgroup H of G with $H = \langle H \cap D\rangle \simeq$ $\mathrm{Sym}(5)$ *(the elements of $D \cap H$ acting as the transpositions of* $\mathrm{Sym}(5)$*). Prove that there is a set Ω with $G \simeq$* $\mathrm{FSym}(\Omega)$*, the elements of D being mapped to the transpositions of* $\mathrm{FSym}(\Omega)$.

Hint: *Consider the maximal subsets C of D with the property that $e, f \in C$ implies $(ef)^3 = 1$.*

(5.22). PROBLEM. *Let G_0 be the symmetric group* $\mathrm{Sym}(5)$*, and let G_1 be the symmetric group* $\mathrm{Sym}(G_0) \simeq \mathrm{Sym}(120)$*. Identify G_0 with the subgroup of* $\mathrm{Sym}(G_0)$ *that is the right regular representation of G_0. Continue in this fashion: identify each G_i with its regular representation as a subgroup of $G_{i+1} = \mathrm{Sym}(G_i)$. Set $G = \bigcup_{i \in \mathbb{N}} G_i$, a group that is the ascending union of the G_i.*[3]

(a) *Prove that G is locally finite and simple.* HINT: *Consider Problem (4.19).*

(b) *Prove that if H_1 and H_2 are isomorphic finite subgroups of G, then there is an element $s \in G$ with $H_1^s = H_2$.* HINT: *Consider Problem (3.24).*

REMARK. *The group G is Phillip Hall's* universal locally finite group.

(5.23). PROBLEM. *Consider the* braid group

$$\mathrm{B}_k = \langle\, \alpha_1\,, \ldots, \alpha_k \mid \alpha_i\alpha_{i+1}\alpha_i = \alpha_{i+1}\alpha_i\alpha_{i+1}\,,\ \alpha_i\alpha_j = \alpha_j\alpha_i \textit{ for } |i-j| > 1 \,\rangle\,.$$

(a) *Prove that the map $\alpha_i \mapsto a_i$ extends to a homomorphism of the braid group onto* $\mathrm{W}(A_k) \simeq \mathrm{Sym}(k+1)$.

(b) *Prove that in* B_k*, every generator α_i has infinite order and* $\mathrm{B}_k/\mathrm{B}_k' \simeq \mathbb{Z}$.

[3] This should be formalized using directed limits, but we hope it is clear what is intended.

Chapter 6

Matrices

6.1 Elementary matrices and operations

We start by considering the group $\mathrm{GL}_n(R)$ of all invertible matrices with entries from the ring R with identity.

Within $\mathrm{GL}_n(R)$ there are three types of *elementary matrices*:

(i) *Elementary permutations* $\pi_{(i,j)}$. These are just the permutation matrices corresponding to the transpositions (i, j) of the symmetric group;

(ii) *Elementary diagonal matrices* $h_j(u)$. These are the diagonal matrices in which all diagonal entries are 1 except for the (j, j)-entry which is u, a unit in R;

(iii) *Elementary transvections* $\mathrm{t}_{i,j}(a)$. These are the matrices $I + ae_{i,j}$ with $a \in R$ and $i \neq j$, where $e_{i,j}$ is a matrix unit—all its entries are 0 except for a 1 in the (i, j)-position.

The matrix group generated by the elementary permutation matrices $\pi_{(i,j)}$ is the group of all $n \times n$ permutation matrices—the image of $\mathrm{Sym}(n)$ under π—and we often identify this subgroup with $\mathrm{Sym}(n)$ (so we may write (i, j) in place of $\pi_{(i,j)}$ and in general w for π_w.)

The group generated by all the elementary diagonal matrices $h_j(U)$ is the group $\mathrm{H}_n(R)$ of all invertible diagonal matrices, and the elementary permutation and diagonal matrices together generate the group of monomial matrices.

The group $\mathrm{E}_n(R)$ is $\langle\, \mathrm{t}_{i,j}(a) \mid i \neq j\,, a \in R\,\rangle$, generated by all elementary transvections, while the group generated by all three types of elementary matrices is $\mathrm{GE}_n(D)$.

As $e_{i,j}e_{k,l} = \delta_{j,k}e_{i,l}$ we immediately have

(6.1). Lemma.

(a) $\mathrm{t}_{i,j}(a)\,\mathrm{t}_{i,j}(b) = \mathrm{t}_{i,j}(a + b)$ *and especially* $\mathrm{t}_{i,j}(a)^{-1} = \mathrm{t}_{i,j}(-a)$.

(b) $[\mathrm{t}_{i,j}(a), \mathrm{t}_{k,l}(b)] = \mathrm{t}_{i,l}(ab)^{\delta_{j,k}}$, *for* $i \neq l$. □

Let $\mathrm{U}_n^+(R)$ be the subgroup of upper *unitriangular matrices* and $\mathrm{U}_n^-(R)$ be the corresponding subgroup of lower unitriangular matrices. We will usually write $\mathrm{U}_n(R)$ in place of $\mathrm{U}_n^+(R)$.

(6.2). PROPOSITION.

(a) $\mathrm{U}_n(R) = \mathrm{U}_n^+(R) = \langle\, \mathrm{t}_{i,j}(a) \mid a \in R, i < j\,\rangle$.

(b) $\mathrm{U}_n^-(R) = \langle\, \mathrm{t}_{i,j}(a) \mid a \in R, i > j\,\rangle$.

PROOF. (a) The proof is algorithmic (which is to say, by induction on n). Let A be upper unitriangular, so that $a_{j,j} = 1$ for all j and $a_{i,j} = 0$ for all $i > j$. Set $A = A^{(1)}$. Then let $t_1 = \prod_{j=2}^n \mathrm{t}_{1,j}(-a_{1,j})$ and $A^{(2)} = A^{(1)}t_1$. The matrix $A^{(2)}$ is upper unitriangular and additionally $a_{1,1}^{(1)} = 1$ is the only nonzero entry in its first row. We proceed to construct from each $A^{(i)}$ an new matrix $A^{(i+1)} = A^{(i)}t_i$ where $t_i = \prod_{j=i+1}^n \mathrm{t}_{i,j}(-a_{i,j}^{(i)})$, the matrix $A^{(i+1)}$ being upper unitriangular and having all nondiagonal entries 0 in its first i rows.

In particular $A^{(n)}$ is the identity matrix. Therefore $A \prod_{i=1}^{n-1} t_i = I$ and $A = (\prod_{i=1}^{n-1} t_i)^{-1} \in \langle\, \mathrm{t}_{i,j}(a) \mid a \in R, i < j\,\rangle$, as desired.

A similar argument gives (b). □

(6.3). PROPOSITION. $\mathrm{U}_n(R)$ *is nilpotent of class* $n - 1$.

PROOF. For each $1 \leq i \leq n$, let $U_k = \langle\, \mathrm{t}_{i,j}(a) \mid j - i \geq k\,\rangle$. In particular $U_n = 1$ and $U_1 = U$ by the previous proposition. By Lemma (6.1)(b) we always have $[U_k, U_l] \leq U_{k+l}$. Therefore

$$1 = U_n \leq U_{n-1} \leq \cdots \leq U_k \leq \cdots \leq U_1 = U$$

is an ascending central series for U. Indeed the same lemma tells us that $U_{n-1} = \mathrm{Z}(U)$; and, continuing in this fashion (by induction), we actually have the upper central series for U. Its length is $n-1$, so U has class $n-1$. (See also Problem (6.36).) □

There are many proofs for Proposition (6.2), some perhaps more elegant than the one given here. But we have chosen a proof that shows its relation to elementary matrix operations. We have shown that by a succession of elementary column operations (corresponding to right multiplication by $\mathrm{t}_{i,j}(b_{i,j})$ for $i < j$) an upper unitriangular can be reduced to the identity. This means that the original matrix is the inverse of the product of the corresponding elementary transvections $\mathrm{t}_{i,j}(a)$.

Let us discuss *elementary operations* on rows and columns in terms of multiplication by elementary matrices. There are three types of elementary row operations that can be made on the matrix A:

(i) Left multiplication by $\pi_{(i,j)}$ switches rows i and j of A;

(ii) Left multiplication by $h_j(u)$ modifies row j of A by multiplying all its entries on the left by u;

(iii) Left multiplication by $\mathrm{t}_{i,j}(a)$ adds a times row j of to row i of A.

Similarly for elementary column operations:

(i) Right multiplication by $\pi_{(i,j)}$ switches columns i and j of A;

(ii) Right multiplication by $h_j(u)$ modifies column j of A by multiplying all its entries on the right by u;

(iii) Right multiplication by $\mathrm{t}_{i,j}(a)$ adds column i times a to column j of A.

(6.4). Lemma. (Whitehead Lemma)
Let R be a ring with identity. For u, v units in R and $a \in R$:

(a) $\begin{pmatrix} 0 & -u^{-1} \\ u & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix} \begin{pmatrix} 1 & -u^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix}$;

(b) $\begin{pmatrix} u & 0 \\ 0 & u^{-1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -u^{-1} \\ u & 0 \end{pmatrix}$;

(c) $\begin{pmatrix} [u,v] & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} (vu)^{-1} & 0 \\ 0 & vu \end{pmatrix} \begin{pmatrix} u & 0 \\ 0 & u^{-1} \end{pmatrix} \begin{pmatrix} v & 0 \\ 0 & v^{-1} \end{pmatrix}$;

(d) $\begin{pmatrix} u & 0 \\ a & v \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} v & a \\ 0 & u \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$;

(e) $\begin{pmatrix} 1 & v^{-1}au \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} v^{-1} & 0 \\ 0 & u^{-1} \end{pmatrix} \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v & 0 \\ 0 & u \end{pmatrix}$;

(f) $\begin{pmatrix} 1 & -a + v^{-1}au \\ 0 & 1 \end{pmatrix} = \left[\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} v & 0 \\ 0 & u \end{pmatrix} \right]$. □

A primary lesson the Whitehead Lemma (6.4) teaches is that almost all elementary operations can be accomplished using elementary transvections: the monomial operation $\pi_{(i,j)}h_i(-1)$ and the diagonal operations $h_i(u)h_j(u^{-1})$ and $h_i([u,v])$ can all be realized by a succession of elementary transvections (using Lemma (6.4)(a,b,c)); as matrices they belong to $\mathrm{E}_n(R)$. More specifically

(6.5). Proposition. *Let U be the group of units of the ring R and $\mathrm{GL}_1(R) = \{ h_1(u) \mid u \in U \} \leq \mathrm{GL}_n(R)$. Then $\mathrm{E}_n(R) \trianglelefteq \mathrm{GE}_n(R) = \mathrm{E}_n(R)\mathrm{GL}_1(R)$ with $\mathrm{GE}_n(R)/\mathrm{E}_n(R)$ a quotient of U/U'.*

Proof. By the Whitehead Lemma (6.4)(d,e) the group $\mathrm{GL}_1(R)$ normalizes $\mathrm{E}_n(R)$.

We have

$$\begin{aligned} \pi_{(i,j)} &= \pi_{(i,j)} \left(h_i(-1)h_i(-1)\right) \left(h_1(-1)h_1(-1)\right) \\ &= \left(\pi_{(i,j)}h_i(-1)\right) \left(h_i(-1)h_1(-1)\right) h_1(-1) \in \mathrm{E}_n(R)\mathrm{GL}_1(R) \end{aligned}$$

and

$$\begin{aligned} h_i(u) &= h_i(u) \left(h_1(u^{-1})h_1(u)\right) \\ &= \left(h_i(u)h_1(u^{-1})\right) h_1(u) \in \mathrm{E}_n(R)\mathrm{GL}_1(R)\,. \end{aligned}$$

Therefore $\mathrm{E}_n(R) \trianglelefteq \mathrm{E}_n(R)\mathrm{GL}_1(R) = \mathrm{GE}_n(R)$. Furthermore

$$\mathrm{GE}_n(R)/\,\mathrm{E}_n(R) = \mathrm{E}_n(R)\mathrm{GL}_1(R)/\,\mathrm{E}_n(R) \simeq \mathrm{GL}_1(R)/\,\mathrm{E}_n(R) \cap \mathrm{GL}_1(R)$$

with $\mathrm{GL}_1(R) \simeq U$ and

$$\mathrm{GL}_1(R)' = \{\, h_1([u,v]) \mid u,v \in U\,\} \le \mathrm{E}_n(R) \cap \mathrm{GL}_1(R)\,. \qquad \square$$

6.2 Bruhat decomposition

Gaussian elimination over the field F is designed to move, via a sequence of elementary row operations, from an arbitrary matrix A to one P that is in *row echelon form*:

> the nonzero rows of P are at its bottom, and, for each nonzero row i and minimal j with $p_{i,j} \neq 0$, all other entries $p_{k,l}$ with $i \le k$ and $l \le j$ are equal to zero.

This is a canonical form result in the sense that Gaussian elimination seeks a relatively simple representative for the orbits of $\mathrm{GL}_n(F)$ in its left action on the set of matrices with n rows and entries from F.

In this section we address a similar problem. We consider square matrices $\mathrm{Mat}_n(D)$ with entries from the division ring D and three different equivalence relations on this set:

(i) Row equivalence, only allowing as elementary row operations the addition of a multiple of a row to a row higher in the matrix.

(ii) Column equivalence, only allowing as elementary column operations the addition of a multiple of a column to a column to its right in the matrix.

(iii) Row and column equivalence, only allowing as elementary operations the addition of a multiple of a row to a row higher in the matrix and the addition of a multiple of a column to a column to its right in the matrix.

That is, in view of Proposition (6.2), we look at the orbits of $\mathrm{Mat}_n(D)$ under, respectively, left multiplication by elements of $\mathrm{U}_n(D)$, right multiplication by elements of $\mathrm{U}_n(D)$, and left and right multiplication by elements of $\mathrm{U}_n(D)$.

The special nonzero p_{ij} of echelon form are sometimes called the *pivots*. With that in mind we say that a *row pivot* of the nonzero row vector $\vec{v}$ is its first nonzero element. More precisely, the *row pivot location* is the smallest j with $\vec{v}_j \neq 0$, the corresponding *row pivot value* then being $\vec{v}_j$. We say that a matrix is in *row pivot form* if every column contains at most one row pivot.

We define a *column pivot* similarly. The *column pivot location* of a nonzero column vector is the largest row index for which that column has a nonzero entry, the corresponding *column pivot value*. A matrix is then in *column pivot form* if every row contains at most one column pivot.

A *partial monomial matrix* is a square matrix that has at most one nonzero entry in each row and each column. It is additionally a *monomial matrix* if it has exactly one nonzero entry in each row and each column.[1] Every partial monomial matrix can be written as the product of a permutation matrix and a diagonal matrix, with this factorization unique if the matrix is monomial. Equally well, every partial monomial matrix can be written as the product of a diagonal matrix and a permutation matrix, again uniquely when the matrix is monomial.

(6.6). LEMMA. *For the partial monomial matrix $P \in \mathrm{Mat}_n(D)$, the following are equivalent:*

(1) *P is invertible.*

(2) *P has no zero row.*

(3) *P has no zero column.*

(4) *P is monomial.* □

If a matrix is in row or column pivot form, then the corresponding pivot locations and values form a partial monomial matrix which we call the *pivot matrix.*

(6.7). PROPOSITION. *Let $P \in \mathrm{Mat}_n(D)$ be a partial monomial matrix.*

(a) $P\,\mathrm{U}_n(D) = \{\, M \mid M$ *is in row pivot form with pivot matrix* $P\,\}$.

(b) $\mathrm{U}_n(D)P = \{\, M \mid M$ *is in column pivot form with pivot matrix* $P\,\}$.

PROOF. The first part actually includes Proposition (6.2)(a) for division rings, and the proof by elementary column operations is essentially the same. The second part is similar, using elementary row operations. □

(6.8). LEMMA. *If the matrix $M \in \mathrm{Mat}_n(D)$ is simultaneously in row pivot form and in column pivot form, then the row pivot locations are the same as the column pivot locations.*

PROOF. The proof is by induction on n with nothing to prove when $n = 1$ or M is a 0-matrix.

The leftmost nonzero column of M has exactly one nonzero entry, as otherwise it would contain two row pivots and M would not be in row pivot form. Delete from M that column and the row containing that pivot.

Certainly the remaining matrix is still in row pivot form. Furthermore, the only column pivot in the deleted row must have been that of the deleted column, since M was in column pivot form. Therefore the row and column pivots of the new matrix are those it inherits from M, and we are done by induction. □

(6.9). THEOREM. *For $A \in \mathrm{Mat}_n(D)$ there are $U_1, U_2 \in \mathrm{U}_n(D)$ and a partial monomial matrix P with $A = U_1PU_2$. Furthermore this determines P uniquely.*

PROOF. We first prove existence of such a factorization. The proof makes use of a "Gaussian elimination" style algorithm:

[1] If this were over a ring R, we would require further that these entries be units of R.

```
Initialize. $j = 1$; $A^{(1)} = A$; label all locations in $A^{(1)}$ open.

Step $j$. If $j = n+1$, halt.
If there is no $i$ with $(i,j)$ open and $a^{(j)}_{i,j} \neq 0$,
then
    $\rho(j) = 0\,;\ t_j = 1\,;\ A^{(j+1)} = A^{(j)}\,;$
    the closed locations of $A^{(j+1)}$ are the closed locations of $A^{(j)}\,;$
    $j \longrightarrow j+1$ and continue to the next step;
else
    $\rho(j) =$ the largest $i$ with $(i,j)$ open and $a^{(j)}_{i,j} \neq 0\,;$
    $t_j = \prod_{i=1}^{\rho(j)-1} \mathrm{t}_{i,\rho(j)}(-a^{(j)}_{i,j}(a^{(j)}_{\rho(j),j})^{-1})\,;$
    $A^{(j+1)} = t_j A^{(j)}\,;$
    the closed locations of $A^{(j+1)}$ are the closed locations of $A^{(j)}$
    and the locations $(\rho(j),k)$ for $k \geq j\,;$
    $j \longrightarrow j+1$ and continue to the next step.
```

In words, we scan the columns of A from left to right. At each new column, we choose the lowest nonzero entry that is not in a row from which we have already chosen an entry. We then use elementary rows operations to zero out all entries in the column above the chosen one.

> *Claim:* All nonzero entries in columns $1, \dots, j-1$ of $A^{(j)}$ are in closed locations.
>
> Proof. The proof is by induction on j with nothing to prove for $j = 1$. Assume the claim is true for j.
>
> By its construction at Step j, the only nonzero entries in column j of $A^{(j+1)}$ are in closed locations. Additionally, since $a^{(j)}_{\rho(j),j}$ was open, all $a^{(j)}_{\rho(j),k}$ for $k < j$ were open as well, hence 0 by induction. But then the various multiplications $\mathrm{t}_{i,\rho(j)}(*)$ and their product t_j leave columns $1, \dots, j-1$ unchanged from $A^{(j)}$ to $A^{(j+1)}$. In particular all locations in these columns that were nonzero, hence closed, in $A^{(j)}$ remain nonzero and closed in $A^{(j+1)}$.
>
> This completes the proof of the claim.

In particular, all nonzero entries in $B = A^{(n+1)}$ are in closed locations. Therefore in a given row i, the only possible nonzero entries are $b_{i,\rho^{-1}(i)}$ and those to its right.

We conclude that B is in row pivot form with pivots $b_{\rho(j),j}$ for those j with $\rho(j) \neq 0$. Set $V = \prod_{j=1}^{n} t_j \in \mathrm{U}_n(D)$ so that $VA = B$. By Proposition (6.7) there is a $U \in \mathrm{U}_n(D)$ with $VA = B = PU$ for a partial monomial matrix P. That is, $A = U_1 P U_2$ with $U_1 = V^{-1}$ and $U_2 = U$ both in $\mathrm{U}_n(D)$, as desired.

Suppose $U_1 P U_2 = A = W_1 Q W_2$ with Q partial monomial and $W_1, W_2 \in \mathrm{U}_n(D)$. Then for $X_1 = W_1^{-1} U_1$ and $X_2 = W_2 U_2^{-1}$, both in $\mathrm{U}_n(D)$, we have $X_1 P = C = Q X_2$. By Proposition (6.7) again, the matrix C is in both row pivot and column pivot form with respective pivot matrices P and Q. By Lemma (6.8) we have $P = Q$. That is, factorizations $A = U_1PU_2$, with $U_1, U_2 \in \mathrm{U}_n(D)$ and P partial monomial, determine P uniquely. □

The matrices U_1 and U_2 of the theorem are not in general unique (think of the case $A = I$), but see Problem (6.39) below.

The P of the theorem is then the *partial monomial form* $\mathrm{pmon}(A)$ of the matrix A. When A and P are invertible (as in Lemma (6.6)) we also write $\mathrm{mon}(A)$ for P, the *monomial form* of A. The matrix $\mathrm{pmon}(A)$ can always be written as H_1J and JH_2 with H_1 and H_2 uniquely determined diagonal matrices (the same, up to a permutation of the diagonal entries) and a permutation matrix J. When A is invertible, J is uniquely determined as the permutation matrix π_σ for $\sigma = \rho^{-1}$, where ρ is the permutation constructed during the proof of the theorem.

(6.10). THEOREM. *Let $A \in \mathrm{Mat}_n(D)$. The following are equivalent:*

(1) $A \in \mathrm{GL}_n(D)$.

(2) $\mathrm{pmon}(A) \in \mathrm{GL}_n(D)$.

(3) *A is left invertible.*

(4) $\mathrm{RS}(A) = {}_DD^{1,n}$.

(5) *A is right invertible.*

(6) $\mathrm{CS}(A) = D^{n,1}_D$.

PROOF. Clearly (1) implies (3) and (5). Also (3) and (4) are equivalent as $\mathrm{RS}(A) = {}_DD^{1,n}$ if and only if there exists an X with $XA = I$, since the rows of X give the coefficients needed to write the canonical basis elements for ${}_DD^{1,n}$ as linear combinations of the rows of A. Similarly (5) and (6) are equivalent.

Let $P = \mathrm{pmon}(A)$ with $A = U_1PU_2$ for $U_1, U_2 \in \mathrm{U}_n(D)$. If A is invertible, then $P^{-1} = U_2A^{-1}U_1$. If P is invertible, then $A^{-1} = U_2^{-1}P^{-1}U_1^{-1}$. Therefore (1) and (2) are equivalent.

It remains to prove that (3) and (5) imply (2). Suppose

$$X_1QX_2 \cdot Y_1RY_2 = I\,,$$

with $X_1, X_2, Y_1, Y_2 \in \mathrm{U}_n(D)$ and Q, R partial monomial matrices. Then $QWR = Z$, with $W = X_2Y_1$ and $Z = X_1^{-1}Y_2^{-1}$ both invertible. Therefore partial monomial Q has no zero row and so by Lemma (6.6) is invertible, and similarly partial monomial R has no zero column and so is invertible. □

(6.11). COROLLARY. *Let $A, B \in \mathrm{GL}_n(D)$. Then AB is invertible if and only if A and B are invertible.*

PROOF. If A and B are invertible then $(AB)^{-1} = B^{-1}A^{-1}$. If X is the inverse of AB, then XA is a left inverse for B and BX is a right inverse for A. □

The following technical lemma is the basis for our verification of the Bruhat Decomposition in the next theorem and of the Dieudonné determinant in the next section.

(6.12). LEMMA. *Let $W \in \mathrm{Sym}(n)$, $U \in \mathrm{U}_n(D)$, and $J = (j, j+1) \in \mathrm{Sym}(n)$ (where we identify $\mathrm{Sym}(n)$ with the permutation matrices of $\mathrm{GL}_n(D)$). Then*

$$\mathrm{mon}(WUJ) = \begin{cases} WJ & \textit{or} \\ Wh_j(b)h_{j+1}(-b^{-1}) & \textit{for some nonzero } b \in D\,. \end{cases}$$

PROOF. Let $k^W = j$ and $l^W = j+1$.

The matrix $A = WU$ is in row pivot form with $\mathrm{mon}(A) = W$. Then in $AJ = WUJ$ we have switched columns j and $j+1$ and left all else unchanged. In particular a_{i,i^W} remains a row pivot as long as (i, i^W) is not (k, j) or $(l, j+1)$. Therefore we expect pivots for rows $\{k, l\}$ ultimately to be in columns $\{j, j+1\}$. That is, appropriate elementary operations should reveal $\mathrm{mon}(AJ)$ as WH or WJH for some $H = h_j(b)h_{j+1}(c)$. The precise verification remains.

First suppose that $k < l$ so that in rows k and l of A we have

$$A_{\{k,l\},*} = \left(\begin{array}{ccc|cc|ccc} 0 & \cdots & 0 & 1 & a & * & \cdots & * \\ 0 & \cdots & 0 & 0 & 1 & * & \cdots & * \end{array}\right)$$

with

$$W_{\{k,l\},*} = \left(\begin{array}{ccc|cc|ccc} 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \end{array}\right).$$

Here we have separated the columns j and $j+1$ from the others.

Therefore

$$(AJ)_{\{k,l\},*} = \left(\begin{array}{ccc|cc|ccc} 0 & \cdots & 0 & a & 1 & * & \cdots & * \\ 0 & \cdots & 0 & 1 & 0 & * & \cdots & * \end{array}\right),$$

and for $U_1 = \mathrm{t}_{k,l}(-a)$

$$(U_1AJ)_{\{k,l\},*} = \left(\begin{array}{ccc|cc|ccc} 0 & \cdots & 0 & 0 & 1 & * & \cdots & * \\ 0 & \cdots & 0 & 1 & 0 & * & \cdots & * \end{array}\right).$$

Thus $U_1(WUJ) = U_1AJ = WJU_2$, for some $U_2 \in \mathrm{U}_n(D)$. That is,

$$\mathrm{mon}(WUJ) = \mathrm{mon}(U_1WUJ) = \mathrm{mon}(WJU_2) = WJ\,,$$

one of the stated conclusions.

Next we suppose $l < k$ so that

$$W_{\{l,k\},*} = \left(\begin{array}{ccc|cc|ccc} 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 \end{array}\right)$$

with

$$A_{\{l,k\},*} = \left(\begin{array}{ccc|cc|ccc} 0 & \cdots & 0 & 0 & 1 & * & \cdots & * \\ 0 & \cdots & 0 & 1 & b & * & \cdots & * \end{array}\right)$$

and

$$(AJ)_{\{l,k\},*} = \left(\begin{array}{ccc|cc|ccc} 0 & \cdots & 0 & 1 & 0 & * & \cdots & * \\ 0 & \cdots & 0 & b & 1 & * & \cdots & * \end{array}\right).$$

If b happens to be equal to 0, then $AJ = WJU_2$ for some $U_2 \in \mathrm{U}_n(D)$, so that $\mathrm{mon}(WUJ) = \mathrm{mon}(AJ) = WJ$ again. Therefore it remains to consider the case $b \neq 0$. Let $U_1 = \mathrm{t}_{l,k}(-b^{-1})$ so that

$$\begin{aligned}(U_1AJ)_{\{l,k\},*} &= \left(\begin{array}{ccc|cc|ccc} 0 & \cdots & 0 & 0 & -b^{-1} & * & \cdots & * \\ 0 & \cdots & 0 & b & 1 & * & \cdots & * \end{array}\right) \\ &= \left(\begin{array}{ccc|cc|ccc} 0 & \cdots & 0 & 0 & 1 & * & \cdots & * \\ 0 & \cdots & 0 & 1 & -b & * & \cdots & * \end{array}\right) h_j(b)h_{j+1}(-b^{-1})\,.\end{aligned}$$

As the group of diagonal matrices normalizes $\mathrm{U}_n(D)$, this gives

$$U_1AJ = (WU_2)h_j(b)h_{j+1}(-b^{-1}) = Wh_j(b)h_{j+1}(-b^{-1})U_3\,,$$

for $U_2, U_3 \in \mathrm{U}_n(D)$. Therefore

$$\mathrm{mon}(WUJ) = \mathrm{mon}(AJ) = \mathrm{mon}(U_1AJ) = Wh_j(b)h_{j+1}(-b^{-1})\,,$$

the second of our two possibilities. □

(6.13). Theorem. (Bruhat decomposition)

Let D be a division ring. Set $G = \mathrm{GL}_n(D)$, $U = \mathrm{U}_n(D)$, and $H = \mathrm{H}_n(D)$. Next let $B = HU = UH$. Finally let N be the subgroup of all monomial matrices in G so that its subset $S = \{\, s_j = (j, j+i) \mid 1 \leq j \leq n-1 \,\}$ generates $W = \mathrm{Sym}(n)$, which we identify with the subgroup of all permutation matrices in G.

(a) (**BN1**) *$G = BNB$ and $H = B \cap N \trianglelefteq N$.*

(b) (**BN2**) *$N/H \simeq W = \langle S\rangle$ and $s^2 = 1 \neq s$ for each $s \in S$.*

(c) (**BN3**) *For $w \in W$ and $s \in S$ we have $BwB.BsB \subseteq BwB \cup BwsB$.*

(d) (**BN4**) *For each $s \in S$ we have $sBs \neq B$.*

Proof.

(a) Certainly $H = B \cap N \trianglelefteq N$. As $B = HU = UH$, also $G = UNU = BNB$ by Theorem (6.9).

(b) We have seen this before. For instance, it is a consequence of Theorem (5.2)(b).

(c) By Lemma (6.12)

$$wUs \subseteq UwHU \cup UwsU\,.$$

As $B = HU = UH$,

$$BwUsB \subseteq BwB \cup BwsB\,.$$

Also $H \trianglelefteq N$, so $Bw \geq Hw = wH$ and $BwU = BwB$. Therefore

$$BwB.BsB = BwBsB \subseteq BwB \cup BwsB\,.$$

(d) We have $sHs = H$ but $sUs \neq U$. For instance $\mathrm{t}_{j,j+1}(a)^{(j,j+1)} = \mathrm{t}_{j+1,j}(a)$. That is,

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}. \qquad \square$$

6.3 The Dieudonné determinant

For the division ring D, let $D' = [D, D]$ be the derived subgroup of the multiplicative subgroup of D and let $\bar{D}$ be the commutative monoid whose elements are the orbits $\bar{d} = dD'$ for $d \in D$ with multiplication given by

$$\bar{d}\bar{e} = (dD')(eD') = (de)D' = \overline{de}$$

That is, $\bar{D}$ is the abelian group D/D' extended to a monoid by adjoining the element $\bar{0}$ and declaring $\bar{0}\bar{d} = \bar{0} = \bar{0}\bar{d}$ for all $\bar{d}$. As D is a division ring, $\bar{D}$ has no nonzero zero divisors.

Following Dieudonné, we shall define a map

$$\mathrm{Ddet}\colon \mathrm{Mat}_n(D) \longrightarrow \bar{D}\,,$$

the *Dieudonné determinant.* In this section we will prove:

(6.14). Theorem. *The map* Ddet *is a surjective multiplicative homomorphism from* $\mathrm{Mat}_n(D)$ *to* $\bar{D}$ *with* $\mathrm{Ddet}(A) \neq \bar{0}$ *if and only if* $A \in \mathrm{GL}_n(D)$*. If* D *is a field, then* $\mathrm{Ddet} = \det$*, the usual determinant.*

We define Ddet in three stages:

(i) For the diagonal matrix H with diagonal entries h_i, for $1 \leq i \leq n$, we set $\mathrm{Ddet}(H) = (\prod_{i=1}^n h_i)D' = \prod_{i=1}^n \bar{h}_i$.

(ii) For the partial monomial matrix $N = PH$ with P a permutation matrix and H a diagonal matrix, we set $\mathrm{Ddet}(N) = \mathrm{sgn}(P)\,\mathrm{Ddet}(H)$.

(iii) For arbitrary $A \in \mathrm{Mat}_n(D)$, we set $\mathrm{Ddet}(A) = \mathrm{Ddet}(\mathrm{pmon}(A))$.

This is well-defined by Theorem (6.9) with justification only needed for $\mathrm{Ddet}(N)$. In that case, the factorization $N = PH$ is unique as long as N is invertible. For noninvertible N there can be more than one factorization $P_1H_1 = N = P_2H_2$; however $H_1 = H_2$ contains at least one diagonal entry 0, so $\mathrm{Ddet}(N) = \mathrm{sgn}(P_1)\,\mathrm{Ddet}(H_1) = \mathrm{sgn}(P_2)\,\mathrm{Ddet}(H_2) = \bar{0}$, independent of the signs of P_1 and P_2.

We instead could have used any factorization $N = KP$ of partial monomial N, since in that case $K = PHP^{-1} = H^{P^{-1}}$ is a diagonal matrix with the same diagonal entries as H, only permuted using P. In particular

$$\mathrm{Ddet}(K) = \prod_{j=1}^n k_jD' = \prod_{j=1}^n h_{j^P}D' = \prod_{i=1}^n h_iD' = \mathrm{Ddet}(H)\,.$$

Several parts of Theorem (6.14) are immediate. Purely as a map Ddet is surjective since the diagonal matrix H with $h_{1,1} = d$ and $h_{i,i} = 1$ for $i \geq 2$ has $\mathrm{Ddet}(H) = \bar{d}$. Over fields D the determinant of a unitriangular matrix is 1; so

$$\det(A) = \det(\mathrm{pmon}(A)) = \mathrm{Ddet}(\mathrm{pmon}(A)) = \mathrm{Ddet}(A)$$

always, and Ddet recovers the usual determinant.

Invertibility is easy to verify.

(6.15). LEMMA. *The matrix $A \in \mathrm{Mat}_n(D)$ is noninvertible if and only if $\mathrm{Ddet}(A) = \bar{0}$. In particular, $\mathrm{Ddet}(AB) = \bar{0}$ if and only if $\mathrm{Ddet}(A) = \bar{0}$ or $\mathrm{Ddet}(B) = \bar{0}$.*

PROOF. By Theorem (6.10) the noninvertible matrices A are precisely those with $\mathrm{pmon}(A) = PH$ for P a permutation and H a diagonal matrix having at least one 0 on its diagonal. That is, A is noninvertible if and only if $\mathrm{Ddet}(A) = \bar{0}$. The rest follows from Corollary (6.11). □

To complete our proof of the theorem, we are reduced to showing that Ddet is a multiplicative homomorphism. By the lemma we need only consider $\mathrm{GL}_n(D)$. As a starting point, we handle some easy cases.

(6.16). LEMMA.

(a) *If $U \in \mathrm{U}_n(D)$, then* $\mathrm{Ddet}(AU) = \mathrm{Ddet}(UA) = \mathrm{Ddet}(A)$.

(b) *If $H \in \mathrm{H}_n(D)$, then* $\mathrm{Ddet}(AH) = \mathrm{Ddet}(HA) = \mathrm{Ddet}(A)\,\mathrm{Ddet}(H)$.

PROOF.

(a) This is immediate as $\mathrm{pmon}(AU) = \mathrm{pmon}(UA) = \mathrm{pmon}(A)$.

(b) Let $A = U_1PKU_2$ with $U_1, U_2 \in \mathrm{U}_n(D)$, P a permutation matrix, and K diagonal. As $\mathrm{H}_n(D)$ normalizes $\mathrm{U}_n(D)$, we have $AH = U_1PKU_2H = U_1PKHU_3$ for some $U_3 \in \mathrm{U}_n(D)$; so $\mathrm{pmon}(AH) = PKH$ and

$$\begin{aligned}\mathrm{Ddet}(AH) = \mathrm{Ddet}(PKH) &= \mathrm{sgn}(P)\prod_{i=1}^{n} k_ih_iD' \\ &= \mathrm{sgn}(P)\prod_{i=1}^{n} k_iD'\prod_{i=1}^{n} h_iD' = \mathrm{Ddet}(A)\,\mathrm{Ddet}(H)\,.\end{aligned}$$

Similarly $HA = HU_1PKU_2 = U_4HPKU_2$ so that $\mathrm{pmon}(HA) = HPK = PH^PK$ and

$$\begin{aligned}\mathrm{Ddet}(HA) = \mathrm{Ddet}(PH^PK) &= \mathrm{Ddet}(P)\,\mathrm{Ddet}(H^P)\,\mathrm{Ddet}(K) \\ &= \mathrm{Ddet}(P)\,\mathrm{Ddet}(H)\,\mathrm{Ddet}(K) = \mathrm{Ddet}(A)\,\mathrm{Ddet}(H)\end{aligned}$$

by the previous calculation (used several times). □

(6.17). THEOREM. *We have* $\mathrm{Ddet}(A)\,\mathrm{Ddet}(B) = \mathrm{Ddet}(AB)$ *for all* $A, B \in \mathrm{GL}_n(D)$*. In particular the map* $\mathrm{Ddet}\colon \mathrm{GL}_n(D) \longrightarrow D/D'$ *is a surjective homomorphism from* $\mathrm{GL}_n(D)$ *onto the abelian group* D/D'*.*

PROOF. By Theorem (6.9) we may write $A = U_1HWU_2$ and $B = V_1TKV_2$ for $U_1, U_2, V_1, V_2 \in \mathrm{U}_n(D)$, $H, K \in \mathrm{H}_n(D)$, and $W, T \in \mathrm{Sym}(n)$ (identified with permutation matrices).

Write $T = \prod_{i=1}^{m} J_i$, where each J_i is one of the transpositions $(j, j+1)$ with $1 \leq j \leq n-1$. In particular $\mathrm{Ddet}(T) = \mathrm{sgn}(T) = (-1)^m$. Our proof is by induction on m, the case $m = 0$ being contained in the previous lemma. By that lemma, we also have:

$$\begin{aligned}
\mathrm{Ddet}(A) &= \mathrm{Ddet}(U_1HWU_2) = \mathrm{Ddet}(H)\mathrm{sgn}(W)\\
\mathrm{Ddet}(B) &= \mathrm{Ddet}(V_1TKV_2) = \mathrm{sgn}(T)\,\mathrm{Ddet}(K)\\
\mathrm{Ddet}(AB) &= \mathrm{Ddet}(U_1HWU_2V_1TKV_2) = \mathrm{Ddet}(H)\,\mathrm{Ddet}(WUT)\,\mathrm{Ddet}(K)
\end{aligned}$$

for $U = U_2V_1 \in \mathrm{U}_n(D)$. Therefore to prove $\mathrm{Ddet}(A)\,\mathrm{Ddet}(B) = \mathrm{Ddet}(AB)$ we must verify

$$\mathrm{Ddet}(WUT) = \mathrm{sgn}(W)\mathrm{sgn}(T)\,.$$

Assume $m \geq 1$ and write $WUT = WUJ\prod_{i=2}^{m} J_i$ for $J = J_1 = (j, j+1)$. Then by Lemma (6.12) we have

$$WUJ = U_3WJU_4 \quad \text{or} \quad WUJ = U_3Wh_j(b)h_{j+1}(-b^{-1})U_4$$

for some $U_3, U_4 \in \mathrm{U}_n(D)$ and some $0 \neq b \in D$.

In these two cases the previous lemma gives, respectively,

$$\mathrm{Ddet}(WUJ) = \mathrm{Ddet}(U_3WJU_4) = \mathrm{Ddet}(WJ) = -\mathrm{sgn}(W)$$

and

$$\begin{aligned}
\mathrm{Ddet}(WUJ) &= \mathrm{Ddet}(U_3Wh_j(b)h_{j+1}(-b^{-1})U_4)\\
&= \mathrm{Ddet}(W)\bar{b}(\overline{-b^{-1}}) = -\mathrm{sgn}(W)\,.
\end{aligned}$$

That is, in both cases WUJ is a matrix C with $\mathrm{Ddet}(C) = -\mathrm{sgn}(W)$.

Now by induction

$$\begin{aligned}
\mathrm{Ddet}(WUT) &= \mathrm{Ddet}(WU\prod_{i=1}^{m} J_i) = \mathrm{Ddet}(WU(J\prod_{i=2}^{m} J_i))\\
&= \mathrm{Ddet}(C\prod_{i=2}^{m} J_i) = \mathrm{Ddet}(C)\,\mathrm{Ddet}(\prod_{i=2}^{m} J_i)\\
&= -\mathrm{sgn}(W)(-1)^{m-1} = \mathrm{sgn}(W)\mathrm{sgn}(T)\,,
\end{aligned}$$

as desired. □

We let $\mathrm{SL}_n(D)$ denote the kernel of the Dieudonné determinant, the *special linear group*.[2]

[2] This notation and terminology are not used uniformly in the literature; see Hahn and O'Meara [HaO89, p. 84].

6.4 (B,N)-pairs

The pair of subgroups B and N of the group G is a (B,N)-*pair* provided:

(**BN1**) *$\langle B,N\rangle = G$ and $H = B\cap N \trianglelefteq N$.*

(**BN2**) *$N/H = W = \langle S\rangle$ and $s^2 = 1 \neq s$ for each $s\in S$.*

(**BN3**) *For $w\in W$ and $s\in S$ we have $BwB.BsB \subseteq BwB\cup BwsB$.*

(**BN4**) *For each $s\in S$ we have $sBs\neq B$.*

This is an important unifying concept (due to Tits), since many of the classical and Lie type groups possess a (B,N)-pair. We already saw in Theorem (6.13) that $G = \mathrm{GL}_n(D)$ is a group with a (B,N)-pair consisting of B, the upper triangular subgroup, and N, the monomial subgroup. We shall find below in Theorem (6.30) that its normal subgroup $\mathrm{SL}_n(D)$ does as well.

The subgroup B of G is a *Borel subgroup* of G. The subgroup H is a *Cartan subgroup*, and the quotient $W = N/H$ is the *Weyl group* of G. These terms are usually extended to include any G-conjugates of B and H. The subgroup N does not seem to have a common name. Note that the generating subset S of W is part of the data required to define a (B,N)-pair. When we need to emphasize the specific generating set S being considered, we may write P_S in place of G; see Lemma (6.19) and the remarks that follow it.

The (**BN**) axioms, as presented above, vary somewhat from the properties seen under Theorem (6.13). The most significant variance is that the group W is not required to be a subgroup of G but is instead defined to be the section N/H, where by (**BN1**) the subgroup H, now defined as $B\cap N$, is normal in N. This also means that we have abused notation under (**BN3**) and (**BN4**) by writing elements w,s of W as though they belong to G. In both places this is unambiguous since the elements of W are actually cosets of the normal subgroup H in N whereas the subgroup B contains H. The distinction is important. In particular in Theorem (6.30), where we see that the Dieudonné kernel $\mathrm{SL}_n(D)$ inherits a (B,N)-pair B_0, N_0 from $\mathrm{GL}_n(D)$, it is not always the case that the extension of H_0 by W splits within N_0; see Problem (6.40).

As we shall see, the axiom (**BN3**) is very powerful. The double coset product $BxB.ByB = BxByB$ is always a union of double cosets, one of them clearly being $BxyB$. We saw in Problem (2.39) that the product is always this one coset if and only if the subgroup B is normal. Axiom (**BN3**) then says that a significant portion of the time, the double coset product includes at most one further coset.

Axiom (**BN3**) appears to say more about right multiplication by members of S than left. However when we invert (**BN3**) and set $u = w^{-1}$ we find an equivalent and lefthanded version:

($\mathbf{BN3^{-1}}$) *For $u\in W$ and $s\in S$ we have $BsB.BuB \subseteq BuB\cup BsuB$.*

Although (**BN4**) is usually presented in the form given here, there are at least two alternative and equivalent formulations (given (**BN2**) and (**BN3**)) that are of interest:

(**BN4′**) *For each $s \in S$ we have $s^{-1}Bs \neq B$.*

(**BN4″**) *For each $s \in S$ we have $BsB.BsB = B \cup BsB$.*

Both of these state emphatically that B is not normal in G. Were B to be normal, then we would always have $sBs = s^{-1}Bs = B$ and $BsB.BsB = B$ with $G/N = BN/B \simeq N/B \cap N = W$.

(6.18). Lemma. *Let B and N form a (B,N)-pair in G.*

(a) *If $K = \ker_G(B)$, the core of B in G, then B/K and NK/K form a (B,N)-pair in G/K.*

(b) *For arbitrary K, the subgroups $K \times B$ and $K \times N$ form a (B,N)-pair in $K \times G$.* □

The last change from Theorem (6.13) to here is in (**BN1**) where $G = BNB$ has been weakened to $G = \langle B, N\rangle$; however:

(6.19). Lemma. *Let B and N form a (B,N)-pair in G.*

(a) *$G = BNB$.*

(b) *For the subset T of S set $W_T = \langle T\rangle \leq W$ and then let N_T be the preimage of W_T in N and $P_T = BN_TB$. Then P_T is a subgroup of G within which B and N_T form a (B,N)-pair.*

Proof. The first part follows from the second, since $P_S = BN_SB = BNB$ is then a subgroup of G that contains the generating set $\{B, N\}$.

The double coset union BN_TB is nonempty as it contains B; it is closed under multiplication by (**BN3**) and induction; and it is closed under inverses as $BtB = (BtB)^{-1}$ for all $t \in T$. Therefore P_T is a subgroup, and all P_T contain $B = P_\emptyset$. The (B,N)-pair axioms for P_T follow directly from those for G. □

The subgroups P_T of the lemma are the *parabolic subgroups* containing B, with the conjugates P_T^g being the parabolic subgroups containing the Borel subgroup B^g. In Theorem (6.26) below we shall find various properties of the parabolic subgroups including their characterization as the only subgroups of G containing a Borel subgroup. Observe that $B = P_\emptyset$ and $G = P_S$ (as promised above).

Throughout the rest of this section we will consider a specific nontrivial group G with (B,N)-pair as given above.

In the Weyl group W generated by the elements of S, we define the *length* $\ell_S(w)$ (sometimes just $\ell(w)$) of the element w to be the minimal m with $w = \prod_{i=1}^m s_i$ for $s_i \in S$. For instance $\ell_S(1) = 0$ and $\ell_S(s_i) = 1$. The length function has already made a cameo appearance in our proof of Theorem (6.17).

(6.20). Lemma. *If $w, u \in W$ with $BwB = BuB$ then $w = u$.*

PROOF. Assume $\ell(w) \leq \ell(u)$. We induct on $\ell(w)$. If this is 0, then $w = 1_W$ and $B = BwB = BuB$. A coset representative n_u then belongs to $B \cap N = H$, so also $u = 1_W$.

Now assume $\ell(w) \geq 1$ and write $w = vs$ with $\ell(v) = \ell(w) - 1$ and $s \in S$. Then

$$BvB \subseteq BwBsB = BuBsB \subseteq BuB \cup BusB$$

by (**BN3**). That is, the double coset BvB is either BuB or $BusB$. By induction either $v = u$ or $v = us$. We cannot have $v = u$ as

$$\ell(v) = \ell(w) - 1 < \ell(u)\,.$$

Therefore $v = us$, which is to say $w = vs = (us)s = u$, as desired. □

(6.21). PROPOSITION. *Let $w \in W$ and $s \in S$.*

(a) *If $\ell(ws) \geq \ell(w)$ then $BwB.BsB = BwsB$.*

(b) *If $\ell(ws) \leq \ell(w)$ then $BwB.BsB = BwB \cup BwsB$.*

PROOF. Let $w = \prod_{i=1}^m s_i$ with $m = \ell(w)$ and all s_i in S.

(a) We use induction on m, the case $m = 0$ being clear. Assume $m \geq 1$, then set $r = s_1$ and $u = \prod_{i=2}^m s_i$ so that $w = ru$ with $\ell(u) = m - 1$.

We cannot have $\ell(us) < \ell(u) = m - 1$ as then we would also have

$$\ell(ws) = \ell(rus) \leq m - 1 < m = \ell(w)\,,$$

which is not the case by hypothesis. Therefore $\ell(us) \geq \ell(u) = m - 1$, so by induction $BuBsB = BusB$ hence $uBs \subseteq BusB$. Thus by ($\mathbf{BN3^{-1}}$)

$$BwBsB = Br(uBs)B \subseteq BrBusB \subseteq BrusB \cup BusB = BwsB \cup BusB\,.$$

On the other hand, by (**BN3**)

$$BwBsB \subseteq BwB \cup BwsB\,.$$

Suppose $BwBsB \cap BwB \neq \emptyset$ (that is, $BwBsB \supseteq BwB$). Then we must have $BwB = BwsB$ or $BwB = BusB$. Lemma (6.20) tells us that $w = ws$ or $w = us$, the first clearly false as $s \neq 1$ by (**BN2**). On the other hand, if $w = us$, then $ws = u$, and

$$\ell(ws) = \ell(u) = m - 1 < m = \ell(w)\,,$$

against hypothesis. We conclude that $BwBsB \cap BwB = \emptyset$ and so $BwBsB = BwsB$.

(b) Here $m = 0$ is not possible. We now set $s = s_m$ and $v = \prod_{i=1}^{m-1} s_i$ so that $w = vs$ with $\ell(v) = m - 1$.

By ($\mathbf{BN4''}$) $BsBsB \cap BsB \neq \emptyset$, hence $sBs \cap BsB \neq \emptyset$ and $vsBs \cap vBsB \neq \emptyset$; so indeed

$$BvsBsB \cap BvBsB = BwBsB \cap BvBsB \neq \emptyset\,.$$

As $w = vs$,

$$\ell(vs) = \ell(w) = m > m - 1 = \ell(v)\,;$$

so by (a), $BvBsB = BvsB = BwB$ meets $BwBsB$ nontrivially, hence is contained in $BwBsB$. □

(6.22). Corollary. *For $w \in W$ and $s \in S$, we have $\ell(ws) = \ell(w) \pm 1$.*

Proof. As $(ws)s = w$, also $\ell(w) - 1 \leq \ell(ws) \leq \ell(w) + 1$. By the Proposition we cannot have $\ell(w) = \ell(ws)$. □

Recall that s is a coset of H in G.

(6.23). Corollary. *If $\ell(ws) < \ell(w)$ then $s \subseteq Bw^{-1}BwB$.*

Proof. We have $wBs \cap BwB \neq \emptyset$, so $Bs \cap w^{-1}BwB \neq \emptyset$ hence $s \subseteq Bw^{-1}BwB$. □

(6.24). Proposition.
Let $w = \prod_{i=1}^m s_i$ with $m = \ell_S(w)$, and set $T = \{\, s_i \mid 1 \leq i \leq m \,\}$. (Note that we may have $|T| < m$.) Then $\langle B, B^w \rangle = \langle B, w \rangle = P_T$.

Proof. Set $s = s_m$ and $v = \prod_{i=1}^{m-1} s_i$ so that $w = vs$ and $ws = v$ with $\ell(v) = m - 1$.

By induction $s_1, \ldots, s_{m-1} \subseteq \langle B, B^v \rangle$, so

$$P_T = \langle B, s_1, \ldots, s_{m-1}, s_m \rangle \leq \langle B, B^v, s \rangle\,.$$

As $\ell(ws) = \ell(v) = m - 1 < m = \ell(w)$, by Corollary (6.23)

$$s \subseteq Bw^{-1}BwB \subseteq \langle B, B^w \rangle\,.$$

Therefore as $v = ws$

$$P_T \leq \langle B, B^v, s \rangle \leq \langle B, B^w \rangle \leq \langle B, w \rangle \leq P_T\,,$$

and we have equality throughout. □

(6.25). Corollary. *If $w \in W$ with $w = \prod_{i=1}^{\ell_S(w)} s_i$ and $w \subseteq P_T$ then $s_i \in T$ for all i. In particular $\ell_T(w) = \ell_S(w)$, and for $s \in S$ we have $s \subseteq P_T$ if and only if $s \in T$.*

Proof. As $T \subseteq S$ we must have $\ell_T(w) \leq \ell_S(w)$. Write $w = \prod_{i=1}^{\ell_S(w)} s_i$ with $S_0 = \{\, s_i \mid 1 \leq i \leq \ell_S(w) \,\} \subseteq S$ and $w = \prod_{j=1}^{\ell_T(w)} t_j$ for $T_0 = \{\, t_j \mid 1 \leq j \leq \ell_T(w) \,\} \subseteq T$. Then by the proposition applied within $G = P_S$, we have

$$\langle B, w \rangle = P_{S_0} = \bigcup_{h \in W_{S_0}} BhB\,.$$

On the other hand, the parabolic subgroup P_T also has a (B, N)-pair by Lemma (6.19)(b), and applied within P_T the proposition gives

$$\langle B, w \rangle = P_{T_0} = \bigcup_{k \in (W_T)_{T_0}} BkB = \bigcup_{k \in W_{T_0}} BkB\,.$$

In particular, for every $s_i \in S_0$, we have

$$Bs_iB \subseteq \bigcup_{k \in W_{T_0}} BkB\,.$$

By Lemma (6.20) we must have $s_i \in W_{T_0}$ for all i.

Now write $s_i = \prod_{l=1}^{\ell_T(s_i)} r_l$ with $r_l \in T$. Then by the proposition again (in P_T and in G)

$$Br_lB \subseteq \langle B, r_1, \dots \rangle = \langle B, s_i \rangle = B \cup Bs_iB\,.$$

Lemma (6.20) also applies again, now telling us that $\ell_T(s_i) = 1$ and $s_i = r_1 \in T$.

We conclude that any minimal S-factorization of $w \in P_T$ is in fact a minimal T-factorization. Especially $\ell_T(w) = \ell_S(w)$. □

(6.26). THEOREM.

(a) *If $B \leq P \leq G$, then there is a $T \subseteq S$ with $P = P_T$.*

(b) *The maps $T \mapsto W_T$ and $T \mapsto P_T$ give isomorphisms of the lattice of subsets of S with, respectively, the lattice of subgroups of W generated by subsets of S and the lattice of subgroups P of G with $B \leq P \leq G$.*

PROOF. (a) Let $B \leq P \leq G$ so that $P = \bigcup_{u \in U} BuB$ for some subset U of W. As P is a subgroup of G, we have $U^{-1} = U$ and $1 \in U$. Furthermore, for all $u, v \in U$ with have $BuvB \subseteq BuB.BvB \subseteq P$; so U is a subgroup of W.

For each $u \in U$ write $u = \prod_{i=1}^{\ell(u)} s_{u,i}$. Then set $S_u = \{\, s_{u,i} \mid 1 \leq i \leq \ell(u) \,\}$ and $T = \bigcup_{u \in U} S_u$. From Proposition (6.24) we learn

$$P_T \geq P = \bigcup_{u \in U} BuB = \langle\, B, u \mid u \in U \,\rangle = \langle\, P_{S_u} \mid u \in U \,\rangle = P_T\,.$$

That is, $P = P_T$ and $U = W_T$.

(b) By Lemma (6.19) and Corollary (6.25) the map $W_T \mapsto P_T = BW_TB$ is a lattice isomorphism. By definition $P_{R \cup T} = \langle P_R, P_T \rangle$, and certainly $P_{R \cap T} \leq P_R \cap P_T$. But if $w \in P_R \cap P_T$, then $w \in P_R$ and $w \in P_T$, again by Corollary (6.25). Thus we have in turn a lattice isomorphism $T \mapsto P_T$ of the set of subsets of S with the set of parabolic subgroups containing B. □

(6.27). THEOREM. *Assume that S is indecomposable; that is, it is not possible to write S as the disjoint union of two nonempty subsets S_1 and S_2 for which $(s_1s_2)^2 = 1$ for all $s_1 \in S_1$ and $s_2 \in S_2$. If $X \trianglelefteq G$ then either $X \leq \ker_G(B)$ or $G = BX$.*

PROOF. By Theorem (6.26) the subgroup BX is P_T for some $T \subseteq S$. Here $T \supseteq \{\, t \in S \mid BtB \cap X \neq \emptyset \,\}$ by Corollary (6.25). Indeed we have equality since $P_T = BX = BXB$ implies that every $t \in T$ is represented in N_T by some $n_t \in X$. We have $T = \emptyset$ if and only if $X \leq \ker_G(B)$. Assume this is not the case.

Let $t \in T$ and $s \in S$ with $(st)^2 \neq 1$. By Corollary (6.25) $\ell_{\{s,t\}}(w) = \ell_S(w)$ for all w in the nonabelian dihedral group $P_{\{s,t\}} = \langle s, t \rangle$. Especially

$$3 = \ell_S(sts) > \ell_S(st) > \ell_S(s) = 1 \,.$$

Let n_s be a representative for s in $N = N_S$. As X is normal in G we have $n_s^{-1} n_t n_s \in X$ hence $B n_s^{-1} n_t n_s B = BstsB \leq P_T$. Thus $sts \subseteq P_T$ and $\ell(sts) = 3$, so Corollary (6.25) give $s \in T$. Indecomposability of S now forces $T = S$, hence $BX = P_T = P_S = G$, as claimed. □

A frequent application of this is the following:

(6.28). Corollary. *Assume that S is indecomposable. If G is perfect and B is solvable, then $G/\ker_G(B)$ is simple.*

Proof. By the theorem, any normal X not contained in $\ker_G(B)$ has $G = BX$. Then $G/X = BX/X \simeq B/B \cap X$, which is solvable. As G is perfect, we conclude $B/B \cap X = 1$ hence $G = X$. Therefore $G/\ker_G(B)$ is simple. □

6.5 Simplicity

Let $\mathrm{R}_n(D)$ be the subgroup of nonzero scalar matrices in $\mathrm{GL}_n(D)$. Next $\mathrm{Z}_n(D)$ consists of those nonzero scalar matrices whose diagonal entries are from the center of D. The *projective special linear group* $\mathrm{PSL}_n(D)$ over the division ring D is then $\mathrm{SL}_n(D)/\,\mathrm{Z}_n(D) \cap \mathrm{SL}_n(D)$.

In this section we prove

(6.29). Theorem. *Let D be a division ring and $n \geq 2$ such that $(n, |D|) \neq (2,2), (2,3)$. Then all proper normal subgroups of $\mathrm{SL}_n(D)$ are contained in $\mathrm{Z}_n(D)$. Especially $\mathrm{PSL}_n(D)$ is simple provided $(n, |D|) \neq (2,2), (2,3)$.*

The two excluded cases are genuine exceptions with $\mathrm{PSL}_2(\mathbb{F}_2) \simeq \mathrm{Sym}(3)$ and $\mathrm{PSL}_2(\mathbb{F}_3) \simeq \mathrm{Alt}(4)$; see Corollary (7.10).

(6.30). Theorem. *Let D be a division ring and $n \geq 2$. Set $G_0 = \mathrm{SL}_n(D)$, $U = \mathrm{U}_n(D)$, and $H_0 = \mathrm{H}_n(D) \cap G_0$. Next let $B_0 = H_0 U = U H_0$ and N_0 be the subgroup of all monomial matrices having Dieudonné determinant $\bar{1}$. Finally set $S = \{\, s_j = (j, j+i) \mid 1 \leq j \leq n-1 \,\} \subseteq \mathrm{Sym}(n) = W$, which we identify with the quotient N_0/H_0. Then B_0 and N_0 form a (B,N)-pair in the group G_0 with the same Weyl group $W = \langle S \rangle \simeq \mathrm{Sym}(n)$.*

Proof. This should be compared with the proof of Theorem (6.13).

Certainly $H_0 = B_0 \cap N_0 \trianglelefteq N_0$. By Theorem (6.9) $\mathrm{GL}_n(D) = UNU$. Since each element u of U has $\mathrm{Ddet}(u) = \bar{1}$, those matrices with Dieudonné determinant $\bar{1}$ are precisely those of $UN_0U = B_0N_0B_0$, so we have (**BN1**).

Again (**BN2**) is a known property from Theorem (5.2)(b) for the symmetric group. Within $\mathrm{GL}_n(D)$ we have $N = N_0H$, so that coset representatives for elements of W can always be chosen from N_0. To prove (**BN4**) of Theorem (6.13) we noted that $s_jUs_j \neq U$, and with a suitable choice of coset representative for s_j in N_0 (say $(j, j+1)h_j(-1)$) this remains true in G_0.

It remains to check (**BN3**). By Lemma (6.12)

$$wUs \subseteq UwHU \cup UwsU$$

in $\mathrm{GL}_n(D)$. Multiplying by H_0 we get

$$wB_0s \subseteq B_0wHU \cup B_0wsB_0\,,$$

at which point we can choose our coset representatives for w, s, and ws all within N_0. As everything else then has Dieudonné determinant $\bar{1}$, this becomes

$$wB_0s \subseteq B_0wH_0U \cup B_0wsB_0\,,$$

hence

$$B_0wB_0.B_0sB_0 = B_0wB_0sB_0 \subseteq B_0wB_0 \cup B_0wsB_0\,,$$

as desired. □

It must be emphasized that the Weyl group $\mathrm{Sym}(n)$ is given as a section of $\mathrm{SL}_n(D)$, not a subgroup. The transpositions $(j, j+1)$ are not in $\mathrm{SL}_n(D)$. The coset representatives $(j, j+1)h_j(-1)$ generate a supplement to $H_0 = \mathrm{H}_n(D) \cap \mathrm{SL}_n(D)$ in N_0, but the extension does not always split (see Problem (6.40)).

(6.31). Lemma. *For $n \geq 2$ and B_0 as in Theorem (6.30) the core $\ker_{\mathrm{SL}_n(D)}(B_0)$ is equal to $\mathrm{Z}_n(D) \cap \mathrm{SL}_n(D)$.*

Proof. The Whitehead Lemma (6.4)(d) shows that any subgroup of B_0 (indeed of B) that is normalized by $\mathrm{SL}_n(D)$ must consist of scalar matrices. If the scalar u is not central, then Lemma (6.4)(f) with $u = v$ and a an element of D not commuting with u tells us that the matrix uI does not belong to a normal subgroup of diagonal matrices. On the other hand, the subgroup $\mathrm{Z}_n(D)$ is indeed normal in $\mathrm{GL}_n(D)$. □

(6.32). Proposition. *For $n \geq 2$, $\mathrm{SL}_n(D)$ is the normal closure of $\mathrm{U}_n(D)$ in $\mathrm{SL}_n(D)$.*

Proof. By the Whitehead Lemma (6.4)(d) the subgroup $\mathrm{U}_n^-(D)$ of lower unitriangular matrices is in the normal closure of $U = \mathrm{U}_n^+(D) = \mathrm{U}_n(D)$ in $\mathrm{SL}_n(D)$. Therefore by Lemma (6.4)(a,b,c) the normal closure also contains generators for N_0, the subgroup of monomial matrices with Dieudonné determinant $\bar{1}$. Theorem (6.30) and (**BN1**) then tell us that the normal closure contains $\langle U, N_0\rangle = \langle B_0, N_0\rangle = \mathrm{SL}_n(D)$, as desired. □

(6.33). Proposition. *For $n \geq 2$, $\mathrm{SL}_n(D) = \mathrm{SL}_n(D)' = \mathrm{GL}_n(D)'$ provided $(n, |D|) \neq (2,2), (2,3)$.*

Proof. For $n \geq 3$, this is immediate by the previous proposition, Lemma (6.1)(b), and Proposition (6.5). For $n = 2$, let b be arbitrary in D and choose

$u \in D$ that commutes with b but $u \notin \{0, 1, -1\}$. If b is not in $\mathrm{Z}(D)$, the center of the division ring, then we can choose $u = b$. If b is in the center then everything commutes with b, and we only need an element u of D that is not 0, 1, or -1. As long as $|D| > 3$, such u exist.

Then, for $a = b(u^2 - 1)^{-1}$, by the Whitehead Lemma (6.4)(f)

$$\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -a + uau \\ 0 & 1 \end{pmatrix} = \left[\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} u^{-1} & 0 \\ 0 & u \end{pmatrix} \right]$$

Therefore $\mathrm{U}_2(D) \leq \mathrm{SL}_2(D)'$ provided $|D| > 3$. The previous proposition then shows that $\mathrm{SL}_2(D) = \mathrm{SL}_2(D)'$. □

Proof of Theorem (6.29).

Let $G_0 = \mathrm{SL}_n(D)$, and let X be a normal subgroup of G_0 not contained in $\ker_{G_0}(B_0) \leq \mathrm{Z}_n(D)$ (using Lemma (6.31)).

By Theorem (6.30) the subgroups B_0 and N_0 form a (B, N)-pair in G_0, so by Theorem (6.27) we have $G_0 = B_0 X$. As $U \trianglelefteq B_0$, we then find $UX \trianglelefteq G_0$. But G_0 is the normal closure of U within G_0 by Proposition (6.32). Therefore $G_0 = UX$, and $G_0/X = UX/X \simeq U/U \cap X$, a nilpotent hence solvable group by Proposition (6.3). By Proposition (6.33) the group $G_0 = G_0'$ is perfect provided $(n, |D|) \neq (2, 2), (2, 3)$. In those cases we must have G_0/X trivial hence $G_0 = X$. That is, any normal subgroup of G_0 not contained in $\ker_{G_0}(B_0)$ is all of G_0. In particular $G_0/\ker_{G_0}(B_0)$ is simple. □

6.6 Problems

(6.34). Problem. *Let a and b be noncommuting elements of the division ring D.*

(a) *Prove that $\begin{pmatrix} 1 & a \\ b & ab \end{pmatrix}$ is invertible in $\mathrm{Mat}_2(D)$.*

(b) *Prove that the transpose $\begin{pmatrix} 1 & b \\ a & ab \end{pmatrix}$ is not invertible in $\mathrm{Mat}_2(D)$ but is invertible in $\mathrm{Mat}_2(D^{\mathrm{op}})$.* ***Where do we define D^{op}?***

(6.35). Problem. *Let $G = \mathrm{GL}_2(R)$, the group of 2×2 invertible matrices with entries from the ring with identity R. Set $B = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \,\middle|\, a, d \in U(R) \right\}$ and $n = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. We have seen that for division rings R we have $G = B \,\dot{\cup}\, BnB$. Prove that $G \neq B \,\dot{\cup}\, BnB$ when R is not a division ring.*

(6.36). Problem. *Let R be a ring with identity. Let $G \leq \mathrm{U}_n(R)$ be block upper unitriangular with k blocks. This problem outlines a proof that G is nilpotent of class at most $k - 1$. In particular $\mathrm{U}_n(R)$ itself has class $n - 1$.*

Suppose the blocks have dimensions, respectively, $d_1, d_2, \ldots, d_k$, so that $\sum_{i=1}^{k} d_k = n$. Set $d_0 = 0$. The "corners" of G are then the positions

$$\begin{aligned} \mathcal{C} \quad = \quad & \big\{ (d_1, d_1 + 1), (d_1 + d_2, d_1 + d_2 + 1), \ldots, \\ & \quad (d_1 + \cdots + d_i, d_1 + \cdots + d_i + 1), \ldots, \\ & \quad (d_1 + \cdots + d_{k-1}, d_1 + \cdots + d_{k-1} + 1) \big\} \end{aligned}$$

Let U be the full block diagonal group with these corners. In a sequence of steps we show that U is nilpotent of class $k-1$. Verify the following.

(a) *By Gaussian elimination U is generated by the elementary subgroups*

$$E_{ij} = \{\, I + \alpha\, e_{ij} \mid \alpha \in R \,\}$$

it contains. (Here e_{ij} is the standard matrix unit.) That is

$$U = \langle\, E_{ij} \mid i \leq a \text{ and } b \leq j\,,\ \text{for some } (a,b) \in \mathcal{C} \,\rangle\,.$$

(b) *For each corner $(a,b) \in \mathcal{C}$,*

$$U_{ab} = \langle\, E_{ij} \mid a \leq i \text{ and } j \leq b \,\rangle$$

is abelian and normal in U.

(c) *By the previous two parts, $U = \langle\, U_{ab} \mid (a,b) \in \mathcal{C} \,\rangle$ is nilpotent of class at most $|\mathcal{C}|$; and so $U \geq G$, nilpotent of class at most $k-1$, as claimed.* Hint: *Problem (2.45).*

(d) *Set*

$$I_i = [d_0 + \cdots + d_{i-1}, d_0 + \cdots + d_i + 1]$$

and

$$E^{ab} = \langle\, E_{i,j} \mid i \in I_a,\, j \in I_b \,\rangle \leq U\,,$$

for $1 \leq a < b \leq k$. Then $[E^{ab}, E^{bc}] = E^{ac}$, and so

$$[E^{1,2},\, E^{2,3},\, E^{3,4},\, \cdots,\, E^{k-2,k-1},\, E^{k-1,k}] = E^{1,k}\,.$$

In particular, U is not nilpotent of class less than $k-1$; so it is nilpotent of class exactly $k-1$.

(6.37). Problem. *Let R be a ring with identity. For $n \leq m$, we consider the group $\mathrm{GL}_n(R)$ as embedded in the upper lefthand corner of $\mathrm{GL}_m(R)$:*

$$\mathrm{GL}_n(R) \simeq \begin{pmatrix} \mathrm{GL}_n(R) & 0 \\ 0 & I_{m-n} \end{pmatrix} \leq \mathrm{GL}_m(R)\,.$$

We then let the stable linear group *$\mathrm{GL}(R)$ be the directed limit of the various $\mathrm{GL}_n(R)$ for $n = 1, 2, 3, \ldots.$ (so $\mathrm{GL}(R)$ can be thought of as $\mathbb{N} \times \mathbb{N}$ invertible matrices, each of which differs from the identity matrix only in some finite dimensional upper lefthand corner.)*

Always $\mathrm{E}_n(R) \leq \mathrm{GL}_n(R)$, and we let $\mathrm{E}(R)$ be the corresponding directed limit subgroup of $\mathrm{GL}(R)$. Prove:

(a) $\mathrm{E}_n(R) = \mathrm{E}_n(R)'$ *for $n \geq 3$;*

(b) $\mathrm{GL}_n(R)' \leq \mathrm{E}_{2n}(R)$ *for all n;*

(c) (The Whitehead Lemma) $\mathrm{GL}(R)' = \mathrm{E}(R) = \mathrm{E}(R)'$.

Remark. *Along the spectrum of things referred to as the Whitehead Lemma this is probably at the top while our Lemma (6.4) is probably at the bottom. But Whitehead's insight here was that the simple calculations of that earlier lemma readily lead to this important result.*

The abelian group quotient $\mathrm{GL}(R)/\,\mathrm{E}(R)$ is the Whitehead group *of R and is denoted $\mathrm{K}_1(R)$. It is of central interest in the field of algebraic* K*-theory.*

(6.38). Problem. *In the group* $\mathrm{E}(R)$ *of Problem (6.37) let* $\mathrm{U}^+(R)$ *be the subgroup that is the directed limit of the upper unitriangular subgroups* $\mathrm{U}_n^+(R)$*. Prove that* $\mathrm{U}^+(R)$ *is a locally nilpotent group with trivial center.*

(6.39). Problem. *Let* D *be a division ring and* $A \in \mathrm{GL}_n(D)$ *with* $\mathrm{mon}(A) = HP$ *for diagonal* $H \in \mathrm{H}_n(D)$ *and permutation matrix* P*. Set*

$$\mathrm{U}_n(D)_P = \langle\, \mathrm{t}_{i,j}(a) \mid a \in D\,, i < j\,,\, i^{P^{-1}} > j^{P^{-1}}\,\rangle\,.$$

Prove that $A = U_1HPU_2$ *with* $U_1 \in \mathrm{U}_n(D)$*,* $H \in \mathrm{H}_n(D)$*,* $P \in \mathrm{Sym}(n)$*, and* $U_2 \in \mathrm{U}_n(D)_P$ *all uniquely determined.*

Hint: *The permutation matrix* P *is* $\pi_{\rho^{-1}}$*, where* ρ *is the permutation found while carrying through the algorithm of our proof for Theorem (6.9). That is,* $P^{-1} = \pi_\rho$*; so the condition* $i^{P^{-1}} > j^{P^{-1}}$ *is equivalent to* $\rho(i) > \rho(j)$*.*

(6.40). Problem. *Consider the group* N_0 *of monomial matrices of (Dieudonné) determinant* 1 *in* $\mathrm{GL}_{2m}(\mathbb{F}_3)$*, and let its normal subgroup of diagonal matrices of determinant* 1 *be* $H_0 \simeq Z_2^{2m-1}$*. The quotient is then* $N_0/H_0 \simeq \mathrm{Sym}(2m)$*. Prove that this extension is nonsplit.*

Hint: *In each of* N_0 *and the split extension* $H_0 \rtimes \mathrm{Sym}(2m)$ *(with the same action), let* A *be the preimage of the subgroup* $\langle(1,2)\rangle \times \mathrm{Alt}(3,\ldots,2m)$ *of* $\mathrm{Sym}(2m)$*, and let* S *be the preimage of the subgroup* $\langle(1,2)\rangle$*. Thus the normal subgroup* H_0 *has index* 2 *in* S *which is itself normal in* A*. Consider the action of* A *on the elements of order* 2 *in the coset* $S \setminus H_0$*.*

(6.41). Problem. *Let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{Mat}_2(D)$*. Calculate the Dieudonné determinant of* A*.*

(6.42). Problem. *It* B *and* N *form a* (B,N)*-pair in the group* G*, prove that* $\mathrm{Z}(G)$ *is contained in the core of* B*.*

Chapter 7

Linear transformations

We often consider a matrix ring or group in terms of action on its natural module. Rings and groups of linear transformations provide the appropriate level of generality.

Many of the arguments are reminiscent of earlier arguments about the symmetric and alternating groups but are usually somewhat more difficult.

7.1 The dual space

For V a D-vector space, the *dual space* V^* is $\mathrm{Hom}_D(V,D)$. Under pointwise action it is an abelian group. Indeed, assuming V is a left D-space, the dual V^* is a *right* D-vector space with operations given by

$$v(\lambda+\mu) = v\lambda + v\mu \quad \text{and} \quad v(\lambda.k) = (v\lambda)k\,,$$

for all $\lambda,\mu \in V^*$, $v \in V$, and $k \in D$. The dual of a right D-space is in turn a left D-space.

(7.1). LEMMA. *Let V be a D-space with basis $\{\, v_i \mid i \in I\,\}$. For $\lambda \in V^*$, set $\lambda_i = v\lambda$. Then, for each $v = \sum_{i\in I} a_i v_i \in V$, we have $v\lambda = \sum_{i\in I} a_i\lambda_i$. Indeed, the map $\lambda \mapsto (\lambda_i)_{i\in I}$ is an isomorphism of V^* and $\prod_{i\in I} D$ as* right *D-spaces.*

PROOF. We have

$$v\lambda = \left(\sum_{i\in I} a_i v_i\right)\lambda = \sum_{i\in I} a_i(v_i\lambda) = \sum_{i\in I} a_i\lambda_i\,.$$

The map is directly checked to be a right D-space injection. Conversely, for any $(\lambda_i)_{i\in I} \in \prod_{i\in I} D$,

$$\left(\sum_{i\in I} a_i v_i\right)\lambda = \sum_{i\in I} a_i\lambda_i$$

defines a member λ of V^*, since all but a finite number of the a_i equal 0. □

(7.2). Corollary. *If V is finite dimensional, then*

$$\dim(V_D^*) = \dim_D(V_D^*) = \dim_D({}_DV) = \dim({}_DV)\,.$$

In any event $|V^*| = |D|^{\dim_D(V)}$. □

While V is a direct product of $|I|$ copies of D, the dual V^* is a cartesian product of $|I|$ copies of D. For infinite I, we therefore expect the dual to bigger than the original space; and this is indeed the case. Especially, if V is finite dimensional as left D-space, then V^* has the same dimension as right D-space.

(7.3). Proposition. *Let V be infinite dimensional over the division ring D.*

(a) $|V| = \max(\dim_D(V), |D|)$.

(b) *If D is a field, then* $\dim_D(V^*) = |D|^{\dim_D(V)} > \dim_D(V)$.

(c) $\dim_D(V^*) > \dim_D(V)$.

(d) $\dim_D(V^*) = |D|^{\dim_D(V)} > \dim_D(V)$.

Proof.

□

On the other hand, for finite $|I|$ the dimensions of V and V^* are equal. Nevertheless there is no canonical isomorphism of V and V^*, a fact most easily appreciated by remembering that V is a left D-space while V^* is a right D-space.

With each basis $\mathcal{B} = \{\, v_i \mid i \in I \,\}$ of V, we can associate a nice subset $\mathcal{B}^* = \{\, v_i^* \mid i \in I \,\}$ of V^*, given by

$$v_i.v_i^* = 1 \quad \text{and} \quad v_j.v_i^* = 0 \text{ for } i \neq j\,.$$

The dual set $\mathcal{B}^*$ is always linearly independent. In particular, for finite I, it is a basis of V^*, the *dual basis* to $\mathcal{B}$.

In the proof of Proposition (7.3), we have used the fact that there is little difference in properties between right D-spaces and left D-spaces. In particular, if W is a right D-space, then we may equally well consider its dual $\mathrm{Hom}_D(W, D) = W^*$ (which perhaps should be ${}^*W$). Similar properties hold but with right and left reversed, so W^* is now naturally a *left* D-space. It is then natural to consider the *double dual* of V, the left D-space $V^{**} = (V^*)^*$. Here there is a canonical embedding.

(7.4). Lemma. *V is isomorphic to its image in V^{**} under the map*

$$v \longrightarrow v^{**} \quad \textit{where} \quad v^{**}(\lambda) = v\lambda\,,$$

for all $v \in V$ and $\lambda \in V^*$. □

7.2 Matrix representation

(7.5). Proposition. $\mathrm{Mat}_n(R) \simeq \mathrm{Mat}_n(R^{\mathrm{op}})^{\mathrm{op}}$ *via transpose.*

Proof.

□

(7.6). Corollary. $\mathrm{GL}_n(R) \simeq \mathrm{GL}_n(R^{\mathrm{op}})$ *via transpose-inverse.*

The automorphism group of the division ring D acts as a group of automorphisms on $\mathrm{GL}_n(D)$ via

$$A \mapsto A^\alpha \text{ with } (A^\alpha)_{i,j} = (A_{i,j})^\alpha\,,$$

for each $\alpha \in \mathrm{Aut}(D)$. This gives the semidirect product

$$\Gamma\mathrm{L}_n(D) = \mathrm{Aut}(D) \ltimes \mathrm{GL}_n(D)\,.$$

(7.7). Theorem. *Let V be a D-space of dimension n.*

(a) $\mathrm{End}_D(V) \simeq \mathrm{Mat}_n(D)$.

(b) $\Gamma\mathrm{L}_D(V) \simeq \Gamma\mathrm{L}_n(D)$.

(c) $\mathrm{GL}_D(V) \simeq \mathrm{GL}_n(D)$.

Proof. As $\mathrm{GL}_n(D)$ is the group of units in $\mathrm{Mat}_n(D)$ and a normal subgroup of $\Gamma\mathrm{L}_n(D)$, the last part is a consequence of either one of the preceding parts.

□

(7.8). Proposition. *For $0 \neq r \in D$*

(a) $r \mapsto \mathrm{R}_r = (1, rI) \in \mathrm{GL}_n(D)$ *is an isomorphic embedding of $D^\times$ as the normal subgroup* $\mathrm{R}_n(D)$ *of* $\mathrm{GL}_n(D)$.

(b) $r \mapsto \mathrm{L}(r) = (r^{-1}, rI) \in \Gamma\mathrm{L}_n(D)$ *is an isomorphic embedding of $D^\times$ as a normal subgroup* $\mathrm{L}_n(D)$ *of* $\Gamma\mathrm{L}_n(D)$.

(c) $\mathrm{L}_n(D) \cap \mathrm{R}_n(D) = \mathrm{Z}_n(D)$ *consists of the nonzero central scalars. It is the center of* $\Gamma\mathrm{L}_n(D)$ *and the centralizer of* $\mathrm{SL}_n(D)$ *in* $\Gamma\mathrm{L}_n(D)$.

Proof. (a) By the Whitehead Lemma (6.4)(f), the Cartan subgroup of diagonal matrices $\mathrm{H}_n(D) = B \cap N$ is normalized by the upper triangular Borel subgroup B, an observation central to much of the previous chapter. By Lemma (6.4)(d), the scalar diagonal matrices are then normalized by the Symmetric subgroup $S = \mathrm{Sym}(n)$ and so by all of $\mathrm{GL}_n(D) = \langle B, N\rangle = \langle B, S\rangle$.

(b) This could be verified directly as in the previous part, but it is more enlightening to realize that in the action of $\Gamma\mathrm{L}_n(D)$ on the left D-space of row vectors (as described in the previous result), the element (r^{-1}, rI) acts as scalar multiplication by r. Thus the group $\{\, (r^{-1}, rI) \mid 0 \neq r \in D \,\}$ is the kernel of the action of $\Gamma\mathrm{L}_n(D)$ on the set of 1-spaces of D^n. (That is, the projective space of D^n; see the next chapter.)

(c) Again in terms of action on the set of row vectors D^n, $\mathrm{R}(r)$ is right scalar multiplication by r while $\mathrm{L}(r)$ is left scalar multiplication. A left scalar can be equal to a right scalar on all row vectors if and only if the scalars are identical and central in D □

Just as $\mathrm{PSL}_n(D)$ was defined to be $\mathrm{SL}_n(D)$ modulo its normal scalar subgroups, so $\mathrm{P\Gamma L}_n(D)$ is $\Gamma\mathrm{L}_n(D)$ modulo the group of scalars $\mathrm{L}_n(D)$ and $\mathrm{PGL}_n(D)$ is the image of its subgroup $\mathrm{GL}_n(D)$.

7.3 The finite linear groups

(7.9). Theorem. *Let $q = p^a$ for p a prime and a a positive integer.*

(a) $|\mathrm{GL}_n(q)| = N_{n,q} = \prod_{i=0}^{n-1}(q^n - q^i) = q^{\binom{n}{2}} \prod_{i=1}^{n}(q^i - 1)$.

(b) $|\mathrm{SL}_n(q)| = |\mathrm{PGL}_n(q)| = N_{n,q}/q - 1$.

(c) $|\mathrm{PSL}_n(q)| = |\mathrm{SL}_n(q)|/\gcd(n, q-1)$.

(d) $|\,\Gamma\mathrm{L}_n(q)| = a|\mathrm{GL}_n(q)|$.

Proof. The group $\mathrm{GL}_n(q)$ is regular on the set of (ordered) bases of $\mathbb{F}_q^n$, essentially by definition. But the number of such ordered bases is $N_{n,q}$. The rest of the theorem follows directly. □

(7.10). Corollary.

(a) $\mathrm{PSL}_2(2) = \mathrm{GL}_2(2) \simeq \mathrm{Sym}(3)$.

(b) $\mathrm{PGL}_2(3) \simeq \mathrm{Sym}(4)$.

(c) $\mathrm{PSL}_2(3) \simeq \mathrm{Alt}(4)$.

(d) $\mathrm{PSL}_2(4) \simeq \mathrm{Alt}(5)$.

(e) $\mathrm{PSL}_2(5) \simeq \mathrm{Alt}(5)$.

Proof. $\mathrm{PGL}_2(q)$ acts faithfully on the projective line, which consists of $q+1$ points. Its subgroup $\mathrm{PSL}_2(q)$ is generated by its Sylow p-subgroups (elementary transvections), where $q = p^a$. In the last two parts $\mathrm{PSL}_2(q)$ is simple (by Theorem (6.29)) of order 60. For the last part, see also Problem (5.20). □

7.4 Transvections and elements of small degree

The finitary symmetric group is generated by its 2-cycles; and, in a sense, the alternating group is defined as the group generated by all 3-cycles. The classical groups also have special generating elements of small degree.

The element $t \in \mathrm{GL}_D(V)$ is a *ℓ-root element* if it satisfies:

(i) $\dim_D([V,t]) \leq \ell$;

(ii) t is unipotent.

The second condition says that the endomorphism $t-1$ is nilpotent: $(t-1)^k = 0$, for some $k \in \mathbb{Z}^+$. Indeed, as the series $V > V(t-1) > V(t-1)^2 \cdots$ must strictly descend until 0 is reached, we must have $k \leq \ell + 1$. Of course, every nilpotent t is an ℓ-root element for sufficiently large ℓ. We focus on the cases $\ell = 1$ and $\ell = 2$.

A *transvection* is a 1-root element on V. That is, it is invertible with $V(t-1)$ of dimension 1 and $V(t-1)^2 = 0$.

(7.11). LEMMA. *Every transvection has the form*

$$t(\lambda, x) \colon v \longrightarrow v + (v\lambda)x$$

for nonzero $\lambda \in V^*$ and $x \in V$ with $x\lambda = 0$.

PROOF.

□

The *center* of $t(\lambda, x)$ is the 1-space $\langle x \rangle$, and its *axis* is the hyperplane $\ker \lambda = \ker\langle \lambda \rangle$. (We allow the abuse of notation $t(0, x) = 1 = t(\lambda, 0)$, but we do not consider the identity to be a transvection.)

(7.12). PROPOSITION. *For $\lambda, \lambda_1, \lambda_2 \in V^*$, $x, x_1, x_2 \in V$, and $d \in D$:*

(a) $t(\lambda d, x) = t(\lambda, dx)$. *Indeed if the transvection $t(\lambda_1, x_1)$ is equal to $t(\lambda_2, x_2)$, then there is an $e \in D$ with $\lambda_1 e = \lambda_2$ and $e^{-1}x_1 = x_2$.*

(b) $t(\lambda_1, x)t(\lambda_2, x) = t(\lambda_1 + \lambda_2, x)$.

(c) $t(\lambda, x_1)t(\lambda, x_2) = t(\lambda, x_1 + x_2)$.

(d) *For $g = [\gamma, g] \in \Gamma\mathrm{L}_D(V)$, we have $g^{-1}t(\lambda, x)g = t(g^{-1}\lambda, xg)$.*

PROOF. (*a*) comes from the previous lemma.

For (*b*) and (*c*) consider $t_1 = t(\lambda_1, x_1)$, $t_2 = t(\lambda_2, x_2)$ and $s = t_1t_2$. We have

$$(*) \qquad v.s = v.t_1t_2 = v + (v\lambda_1)x_1 + (v\lambda_2)x_2 + (v\lambda_1)(x_1\lambda_2)x_2 \,.$$

In (*b*) and (*c*) we always have $x_1\lambda_2 = 0$, giving the results.

For (*d*), let $[\tau, h] \in \Gamma\mathrm{L}(V)$ with associated automorphism τ of D. Then $[\tau, h]$ acts (semilinear on the left) on $\mu \in V^*$ via $v(h\mu) = ((vh)\mu)^{\tau^{-1}}$ by Lemma (9.10) below. Thus, for a $[\gamma, g]$ associated with, we have[1]

$$\begin{aligned} v(g^{-1}t(\lambda, x)g) &= ((vg^{-1})t(\lambda, x))g \\ &= (vg^{-1} + (((vg^{-1})\lambda)x))g \\ &= vg^{-1}g + ((v(g^{-1}\lambda))^{\gamma^{-1}}x)g \\ &= v + (v(g^{-1}\lambda))xg \\ &= v(t(g^{-1}\lambda, xg))\,. \qquad \square \end{aligned}$$

Let V be a D-spaces with subspaces U of V and W of V^*. We let $\mathrm{T}(W, U)$ be the subgroup of $\mathrm{GL}_D(V)$ generated by all the $\mathrm{t}(\lambda, v)$ with $\lambda \in W$, $v \in V$,

[1] "It can be easily checked . . . "

and $u\lambda = 0$. In particular if W is a 1-space of V^* and U a 1-space of V with $UW = 0$, then $\mathrm{T}(W, U)$ is a *transvection subgroup* of $\mathrm{GL}_D(V)$. We will abuse this notation somewhat by writing $\mathrm{T}(\lambda, u)$ for the transvection subgroup $\mathrm{T}(W, U)$ when λ spans the 1-space W and u spans the 1-space U.

(7.13). Proposition.

(a) *If W is a 1-space of V^* and U a 1-space of V, then $\mathrm{T}(W, U) = \{\, \mathrm{t}(\lambda, u) \mid \lambda \in W,\, u \in U \,\}$ is a subgroup of $\mathrm{GL}_D(V)$ that is isomorphic to $(D, +)$. if $x\lambda = 0$. In this second case, $T(W, p)$ is called a* transvection subgroup.

(b) *$T(W) = \{1\} \cup \{\, t(\mu, U) \,|\, \langle\mu\rangle = W\,, \langle U\rangle \in \ker W \,\}$ is a subgroup of $SL(V)$ isomorphic to $(D, +)^r$.*

(c) *$T(p) = \{1\} \cup \{\, t(\mu, U) \,|\, p \in \ker\langle\mu\rangle\,, \langle U\rangle = p \,\}$ is a subgroup of $SL(V)$ isomorphic to $(D, +)^r$.*

(d) *$T(W) \cap T(p) = T(W, p)$.*

(e) *If $n \geq 3$, then $T(W) \neq T(p)$.*

(f) *The transvection subgroups $\mathrm{T}(W, U)$ for $UW = 0$ are all conjugate in the group $\mathrm{T}(V^*, V)$. $SL(V)$; the subgroups $T(W)$ are all conjugate in $SL(V)$; the subgroups $T(p)$ are all conjugate in $SL(V)$.*

Proof. The first part follows immediately from the previous proposition.

We wish to prove $\mathrm{T}(\lambda, x)$ and $\mathrm{T}(\gamma, y)$ are conjugate. First assume that $\langle x\rangle \neq \langle y\rangle$, and choose a basis $\{x, y, z_1, z_2, \ldots\}$ for V. The hyperplane $\langle x - y, z_1, z_2, \ldots\rangle$ is $\ker\eta$, for some $\eta \in V^*$ with $\alpha = y\eta \neq 0$. Then, for $t = \mathrm{t}(\eta, \alpha^{-1}(x - y))$, we have

$$y.t = y + (y\eta)\alpha^{-1}(x - y) = y + \alpha(\alpha^{-1}(x - y)) = x$$

Especially $\mathrm{T}(\lambda, x)^t = \mathrm{T}(\delta, y)$, for some $\delta \in V^*$.

So it is enough to prove $\mathrm{T}(\lambda, y)$ and $\mathrm{T}(\gamma, y)$ conjugate in $\mathrm{T}(V^*, V)$. We may assume λ and γ span different 1-spaces of V^*. Let $\{y = y_0, y_1, \ldots\}$ be a basis of $K = \ker\lambda \cap \ker\gamma$, and choose u and v with $\ker\lambda = \langle u, K\rangle$ and $\ker\gamma = \langle v, K\rangle$. For $\mu \in V^*$ with $\ker\mu = \langle v - u, K\rangle$ and $\beta = v\mu \neq 0$, set $s = t(\mu, \beta^{-1}(v - u))$. Then

$$u.s = u + (u\mu)\beta^{-1}(v - u) = u + \beta\beta^{-1}(v - u) = v\,.$$

Therefore $\mathrm{T}(\lambda, y) = \mathrm{T}(\gamma, y)^s$, as desired. □

(7.14). Lemma.

(a) *If $t_1 = t(\lambda_1, x_1)$ and $t_2 = t(\lambda_2, x_2)$ with $\langle\lambda_1\rangle \neq \langle\lambda_2\rangle$ and $\langle x_1\rangle \neq \langle x_2\rangle$, then for $s = t_1t_2$, we have $V(s - 1) = \langle x_1, x_2\rangle$. In particular, s is not a transvection.*

(b) *Let T be a subgroup of $\mathrm{GL}_D(V)$ with $T^\# = T\backslash\{1\}$ completely composed of transvections. Then either there is a 1-space W in V^* with $T \leq \mathrm{T}(W, V)$ or there is a 1-space U in V with $T \leq \mathrm{T}(V^*, U)$.*

Proof. In equation $(*)$ from the proof of Proposition (7.12), if we first choose $v \in \ker \lambda_1 \setminus \ker \lambda_2$ then we find $x_2 \in V(s-1)$. Next with $v \notin \ker \lambda_1$ we also have $x_1 \in V(s-1)$.

The second part follows directly from the first. □

The element $s \in \mathrm{GL}_D(V)$ is a 2-*root element* if and only if there are $u, v \in V$ with $V(s-1) \leq \langle u, v\rangle$ and s unipotent (that is, $(s-1)^k = 0$, for some k). Equivalently, there is a series of D-spaces $0 \neq \langle x\rangle \leq U < V$ with $\dim V/U = 1$ and such that $[V, s] \leq U$, $[U, s] \leq \langle x\rangle$, and $[x, s] = 0$.

(7.15). Proposition. *Let $0 \neq \langle x\rangle \leq U < V$ be a series of D-spaces with $\dim V/U = 1$. Suppose $s \in \mathrm{GL}_D(V)$ stabilizes this series. (That is, $[V, s] \leq U$, $[U, s] \leq \langle x\rangle$, and $[x, s] = 0$.)*

(a) *Then*

$$s = \mathrm{r}(\lambda, x, \mu, y) \colon v \mapsto v + v\lambda.y + v\mu.x$$

for $y \in V$ and $\mu, \lambda \in V^\star$ with $x \in \ker \lambda = U$, $y\lambda = 0$, and $x\mu = 0$. Conversely, any such s stabilizes the given series.

(b) $\mathrm{r}(\lambda, x, \mu, y)\,\mathrm{r}(\lambda, x, \eta, z) = \mathrm{r}(\lambda, x, \mu+\eta, y+z+y\eta.x) = \mathrm{r}(\lambda, x, \mu+\eta+\lambda.y\eta, y+z)$.

(c) $\mathrm{r}(\lambda, x, \mu, y)^{-1} = \mathrm{r}(\lambda, x, -\mu, -y + y\mu.x)$.

(d) $[\mathrm{r}(\lambda, x, \mu, y), \mathrm{r}(\lambda, x, \eta, z)] = \mathrm{r}(\lambda, x, 0, (y\eta - z\mu).x)$.

(e) *For $g \in \mathrm{GL}_D(V)$, $g^{-1}\,\mathrm{r}(\lambda, x, \mu, y)g = \mathrm{r}(g^{-1}\lambda, xg, g^{-1}\mu, yg)$.*

(f) *For $a, b \in D^\times$, $\mathrm{r}(\lambda a, bx, \mu, y) = \mathrm{r}(\lambda, x, \mu b, ay)$.*

(g) *For $a \in D$, $\mathrm{r}(\lambda, x, \mu, y + ax) = \mathrm{r}(\lambda, x, \mu + \lambda a, y)$. Indeed $\mathrm{r}(\lambda, x, \mu, y) = \mathrm{r}(\lambda, x, \mu', y')$ if and only if there is an $a \in D$ with and $\mu' - \mu = \lambda a$ and $y' - y = -ax$.*

(h) *The set $\mathrm{R}(\lambda, x) = \{\mathrm{r}(\lambda, x, \mu, y) \mid y\lambda = 0, x\mu = 0\}$ is a subgroup of $\mathrm{GL}_D(V)$, normal in the stabilizer of $\langle x\rangle$ and U. $\mathrm{R}(\lambda, x)$ has class 2 with $\mathrm{R}(\lambda, x)' \leq \mathrm{Z}(\mathrm{R}(\lambda, x)) = \mathrm{T}(\lambda, x)$, the transvection subgroup with axis $\ker \lambda$ and center $\langle x\rangle$. If $\dim_D(V) = 2$, then $\mathrm{R}(\lambda, x) = \mathrm{Z}(\mathrm{R}(\lambda, x)) = \mathrm{T}(\lambda, x)$ is abelian. If $\dim_D(V) \geq 3$, then $\mathrm{R}(\lambda, x)' = \mathrm{Z}(\mathrm{R}(\lambda, x)) = \mathrm{T}(\lambda, x)$.*

(i) $\mathrm{R}(\lambda, x) = \mathrm{T}(\lambda, \cdot)\,\mathrm{T}(\cdot, x)$, *where $\mathrm{T}(\lambda, \cdot) = \{\mathrm{r}(\lambda, x, 0, y) \mid y\lambda = 0\}$ is the subgroup of all transvections with axis $U = \ker \lambda$ and $\mathrm{T}(\cdot, x) = \{\mathrm{r}(\lambda, x, \mu, 0) \mid x\mu = 0\}$ is the subgroup of all transvections with center $\langle x\rangle$. Both are abelian and normal in $\mathrm{R}(\lambda, x)$, and their intersection is $\mathrm{T}(\lambda, x)$.*

(j) *Let $W = U/\langle x\rangle$ with the image of u being denoted $\bar{u}$. For $\eta \in V^*$ with $x\eta = 0$, let $\bar{\eta}$ be the member of $W^\star$ induced by η. The map $\mathrm{r}(\lambda, x, \mu, y) \mapsto (\bar{\mu}, \bar{y})$ is a well-defined surjective homomorphism from $\mathrm{R}(\lambda, x)$ onto $W^* \oplus W$ with kernel $\mathrm{T}(\lambda, x)$.*

(k) *If $|D| = q$ and $\dim_D(V) = n$, then $|\,\mathrm{R}(\lambda, x)| = q^{2n-3}$ with center $\mathrm{T}(\lambda, x)$ of order q.*

PROOF.

□

7.5 Problems

(7.16). PROBLEM. *This problem will (among other things) show again that the sign homomorphism on finite* $\mathrm{Sym}(\Omega)$ *(and hence on arbitrary* $\mathrm{FSym}(\Omega)$*) is well-defined. Let F be a field and V the* $\mathrm{Sym}(\Omega)$ *permutation module $F\Omega$ with basis* $\{\, e_\omega \mid \omega \in \Omega \,\}$ *and action given by* $e_\omega^g = e_{\omega^g}$*. Let $n \geq 2$ and set* $\Omega = \{1, 2, \ldots, n\}$ *so that* $\mathrm{Sym}(\Omega) = \mathrm{Sym}(n)$*.*

(a) *Recall that for* $g \in \mathrm{GL}_F(V)$*, we have defined* $[V, g] = V(g-1)$*. Prove:*

(i) *The subspace W of V is g-invariant with g trivial on V/W if and only if* $W \geq [V, g]$*.*

(ii) $[V, gh] \leq [V, g] + [V, h]$*.*

(b) *For* $g \in \mathrm{Sym}(n)$*, let $\ell(g)$ be the smallest number of transpositions with product g. Prove that* $\dim_D[V, g] \leq \ell(g)$*.*

(c) *Prove that* $\dim_D[V, g] = \ell(g) = n - n_k = |\mathrm{Supp}(g)| - c_k$ *where n_k is the number of cycles in g (including cycles of length 1) and c_k is the number of cycles in g of length greater than 1. (*HINT: *First prove that if g is a k-cycle, then* $\dim_D[V, g] = \ell(g) = k - 1$*.)*

(d) *For t a transposition, prove that* $\ell(gt) = \ell(g) \pm 1$*.*

(e) *Prove that* $\mathrm{sgn}\colon g \mapsto (-1)^{\ell(g)} = (-1)^{\dim_k[V,g]}$ *is a homomorphism from* $\mathrm{Sym}(n)$ *onto the multiplicative group* ± 1*. (This is the sign homomorphism.)*

(7.17). PROBLEM. *Let* $(I, \leq)$ *a totally ordered set. For the field E, let* $\{\, e_i \mid i \in I \,\}$ *be the canonical basis of the I-tuple space* $E^{(I,\leq)}$*, defined to be* $\oplus_{i \in I} E e_i$*. The upper triangular group* $\mathrm{FU}_E(E^{(I,\leq)}) \leq \mathrm{FGL}_E(E^{(I,\leq)})$ *then consists of all linear transformations g given by* $e_i^g = e_i + \sum_{i<j} \alpha_{ij} e_j$ *for all i, where only a finite number of the α_{ij} are nonzero. For instance, in Problems (6.36) and (6.38), we discussed* $\mathrm{U}_n^+(E) = \mathrm{FU}_E(E^{([1,n],\leq)})$*. and* $\mathrm{U}^+(E) = \mathrm{FU}_E(E^{(\mathbb{Z},\leq)})$*.*

(a) *For infinite I, prove that* $\mathrm{FU}_E(E^{(I,\leq)})$ *is locally nilpotent with trivial center.*

(b) *Prove that* $\mathrm{FU}_E(E^{(\mathbb{Q},\leq)})$ *is a perfect locally nilpotent group with trivial center.*

(c) *If E is a locally finite field, prove that* $\mathrm{FU}_E(E^{(I,\leq)})$ *is locally finite.*

REMARK. *The group* $\mathrm{FU}_E(E^{(\mathbb{Q},\leq)})$ *is in fact characteristically simple and is called a* McLain group*.*

(7.18). PROBLEM. *Prove* $\mathrm{PSL}_2(9) \simeq \mathrm{Alt}(6)$*.* HINT: *???*

Chapter 8

Projective Spaces

8.1 Several versions of projective spaces

8.1.1 Lattices

For a vector space V over the division ring D, the *projective space* $\mathbb{P}V$ is the lattice of subspaces of V (excluding $\{0\}$ and V). The *rank* of $\mathbb{P}V$ is one less than the D-dimension of V.

For any vector subspace W of V, the projective space $\mathbb{P}W$ is naturally contained in $\mathbb{P}V$. Somewhat abusing terminology, we refer to both W and $\mathbb{P}W$ as subspaces (members, elements) of $\mathbb{P}V$ of rank $d-1$, where $d = \dim_D(W)$. Lemma (8.1) below allows this abuse. (The other deleted subspace $\{0\}$ is the unique element of rank -1.) The subspaces of $\mathbb{P}V$ of rank 0 are *projective points*; those of rank 1 are *projective lines*, and those of rank 2 are *projective planes*, usually abbreviated to *points*, *lines*, and *planes*.[1] If W has codimension 1 in V, then $\mathbb{P}W$ is a *hyperplane* of $\mathbb{P}V$ (just as W is a hyperplane of V).

8.1.2 Incidence geometry

Two members u, w of $\mathbb{P}V$ are incident, written $u \sim w$, if one contains the other. (That is, subspaces U and W are incident if either $U \leq W$ or $U \geq W$.) In particular, two members of the same rank are incident if and only if they are equal.

There is a great deal of redundant information in the incidence relations of the projective space. For $w \in \mathbb{P}V$ and $0 \leq i \leq \text{rank}(V)$, let

$$\mathbb{P}_i V_w = \{\, p \in \mathbb{P}_i V \,|\, p \sim w \,\}$$

[1] We avoid a common vector space terminology that identifies vectors as points, 1-spaces or 1-flats as lines, and 2-spaces or 2-flats as planes.

the *shadow* of w in $\mathbb{P}_iV$. Once we realize that (b) follows directly from (a), we have

(8.1). Lemma.

(a) *For $w, v \in \mathbb{P}V$, $\mathbb{P}_iV_w = \mathbb{P}_iV_v$ if and only if $w = v$.*

(b) *For $0 \le i \le \mathrm{rank}(V)$ and $w, v \in \mathbb{P}_iV$, $\mathbb{P}_iV_w = \mathbb{P}_iV_v$ if and only if $w = v$.*

(c) *A subset P of $\mathbb{P}_0V$ is equal to $\mathbb{P}_0V_w$, for some subspace w, if and only if, for every projective line $\ell \in \mathbb{P}_1V$, we have either $|\mathbb{P}_0V_\ell \cap P| \le 1$ or $\mathbb{P}_0V_\ell \subseteq P$.* □

Part (a) of the lemma allows us, without much confusion, to abuse notation by identifying a subspace with the set of projective points contained within it. Then (b) says that P is a subspace if and only if, for all lines ℓ, either $|\ell \cap P| \le 1$ or $\ell \subseteq P$; the subspaces are exactly the line-closed subsets of $\mathbb{P}_0V$.

For any subset P of $\mathbb{P}V$, the *subspace generated* by P, denoted $\langle P\rangle$, is the intersection of all subspaces containing P. So $\langle P\rangle$ is the smallest subspace that contains P. As mentioned above, we typically identity $\langle P\rangle$ with its shadow in $\mathbb{P}_0V$.

It is an easy consequence of Lemma (8.1) that every automorphism of the incidence system $\Pi(V) = (\mathbb{P}_0V, \mathbb{P}_1V)$ extends uniquely to a lattice automorphism of the projective space $\mathbb{P}V$ and, conversely, any lattice automorphism of $\mathbb{P}V$ restricts to an automorphism of $\Pi(V)$. For this reason, the automorphisms of $\mathbb{P}V$ are usually called *collineations* and the full lattice automorphism group of $\mathbb{P}V$ is called the *collineation group* of $\mathbb{P}V$, denoted $\mathrm{Coll}(\mathbb{P}V)$.

Veblen-Young
Moufang-Hall
projectivities, collineations

8.1.3 Buildings and Chamber systems

dualities, polarities
Tits

8.2 The Fundamental Theorem of Projective Geometry

Recall that the morphisms in Vec are *semilinear maps.* Thus $\Sigma = [\sigma, s]$ is a semilinear map from ${}_DV$ to ${}_EW$ provided:

(i) σ is a homomorphism of D into E;

(ii) s is an additive homomorphism from $(V, +)$ to $(W, +)$;

(iii) for all $a \in D$ and $v \in V$ we have $(av)^\Sigma = a^\sigma v^s$.

Especially a semilinear map $[1, s]$ is a linear transformation from ${}_DV$ to ${}_DW$. Since D is a division ring, σ always realizes an isomorphism of D with a subdivision ring of E.

(8.2). Theorem. (Fundamental Theorem of Projective Geometry) *Let D and E be division rings with V a D-space of dimension at least 3 and W an E-space. Let $\mathcal{S}\colon \mathbb{P}_0V \longrightarrow \mathbb{P}_0W$ be a map on projective points with the "small rank" property that:*

$$(\mathbf{SmRk})\ \textit{if } p, q, r \in \mathbb{P}_0V,\ \textit{then } \operatorname{rank}\langle p, q, r\rangle = \operatorname{rank}\langle p^{\mathcal{S}}, q^{\mathcal{S}}, r^{\mathcal{S}}\rangle.$$

Then there is a semilinear transformation $\Sigma = [\sigma, s]\colon V \longrightarrow W$ with D^σ a subfield of E isomorphic to D (via σ) and $\langle v\rangle^{\mathcal{S}} = \langle v^s\rangle$, for all vectors $v \in V$.

(8.3). Lemma. *Let $p, q \in \mathbb{P}_0V$.*

(a) *$\mathcal{S}\colon \mathbb{P}_0V \longrightarrow \mathbb{P}_0W$ is injective.*

(b) *For $t \in \mathbb{P}_0W$ we have $t \in \left[p^{\mathcal{S}}, q^{\mathcal{S}}\right] \cap \operatorname{im}(\mathcal{S})$ if and only if there is an $r \in \langle p, q\rangle$ with $r^{\mathcal{S}} = t$.*

Proof.

□

(8.4). Corollary.

(a) *$\mathcal{S}\colon \mathbb{P}_1V \longrightarrow \mathbb{P}_1W$ given by $\langle p, q\rangle^{\mathcal{S}} = \left[p^{\mathcal{S}}, q^{\mathcal{S}}\right]$ is injective.*

(b) *If $\langle p, q\rangle$, and $\langle r, s\rangle$ are lines of $\mathbb{P}_1V$ with $t = \langle p, q\rangle \cap \langle r, s\rangle$, then $t^{\mathcal{S}} = \langle p, q\rangle^{\mathcal{S}} \cap \langle r, s\rangle^{\mathcal{S}}$.*

Proof.

□

We further extend our definition of the map $\mathcal{S}$ by defining $\langle u, v\rangle^{\mathcal{S}}$ to be $\langle\langle u\rangle, \langle v\rangle\rangle^{\mathcal{S}}$ for $u, v \in V$.

Proof of the Fundamental Theorem of Projective Geometry (8.2).

The proof is accomplished in a series of steps. We will use Property (**SmRk**), the lemma, and the corollary often and usually without reference.

Let $\{x_0\} \cup \{\, x_i \mid i \in \mathcal{I}\,\}$ be a D-basis for V. Choose $x_0' \in W$ with $\langle x_0\rangle^{\mathcal{S}} = [x_0']$ and then for each $i \in \mathcal{I}$ choose $x_i' \in W$ such that:

(1) $\langle x_i\rangle^{\mathcal{S}} = [x_i']$ and

(2) $\langle x_0 + x_i\rangle^{\mathcal{S}} = [x_0' + x_i']$.

By Property (**SmRk**), any subset of $\{x_0'\} \cup \{\, x_i' \mid i \in \mathcal{I}\,\}$ of size up to three is linearly independent in W since its preimage in V is. This is not necessarily the case for subsets of size greater than three.

Step (i). *For each $i \in \mathcal{I}$ and $d \in D$, define $d^{(i)}$ by $\langle x_0 + dx_i \rangle^{\mathcal{S}} = \left[x'_0 + d^{(i)} x'_1\right]$. Then for all $i, j \in \mathcal{I}$ and all $d \in D$ we have $d^{(i)} = d^{(j)}$.*

Proof. We first note that $d \mapsto d^{(i)}$ is well-defined. As $\langle x_0 + dx_i \rangle^{\mathcal{S}} \in [x'_0, x'_1]$, this only fails if $\langle x_0 + dx_i \rangle^{\mathcal{S}} = [x'_i]$; but in that case $\langle x_0, x_i \rangle^{\mathcal{S}} = \langle x_0 + dx_i, x_i \rangle^{\mathcal{S}} = [x'_i]$, which is not the case (by Property (**SmRk**)).

The claim of the step is obvious for $i = j$ and true by choice (see above, in particular (2)) for $d = 0, 1$. Now assume $i \neq j$ and $0 \neq d \neq 1$.

Consider the three distinct lines

$$\langle x_i, x_j \rangle \ , \ \langle x_0 + x_i, x_0 + x_j \rangle \ , \ \langle x_0 + dx_i, x_0 + dx_j \rangle \ ,$$

all intersecting in the common point $\langle x_i - x_j \rangle = \langle dx_i - dx_j \rangle$.

Step (ii). *Consider the map $\sigma \colon D \longrightarrow E$ given by $d^\sigma = d^{(i)}$, for any $i \in \mathcal{I}$. Then σ is a well-defined injection $\sigma \colon D \longrightarrow E$ with $0^\sigma = 0$ and $1^\sigma = 1$.*

Proof.

□

By (i) the map σ is well-defined.

Step (iii). *For all finite $I \subseteq \mathcal{I}$ and all $d_i \in D$ for $i \in I$, we have $\left\langle x_0 + \sum_{i \in I} d_i x_i \right\rangle^{\mathcal{S}} = \left[x'_0 + \sum_{i \in I} d_i^\sigma x'_i\right]$.*

Proof.

□

Step (iv). *For all finite $I \subseteq \mathcal{I}$ and all $d_i \in D$ for $i \in I$, we have $\left\langle \sum_{i \in I} d_i x_i \right\rangle^{\mathcal{S}} = \left[\sum_{i \in I} d_i^\sigma x'_i\right]$.*

Proof.

□

Step (v). *For all $d, e \in D$ we have $(de)^\sigma = d^\sigma e^\sigma$.*

Proof.

□

Step (vi). *For all $d, e \in D$ we have $(d + e)^\sigma = d^\sigma + e^\sigma$.*

Proof.

□

Step (vii). *For all finite $H \subseteq \{0\} \cup \mathcal{I}$ and all $d_i \in D$ for $i \in H$, we have $\left\langle \sum_{i \in H} d_i x_i \right\rangle^{\mathcal{S}} = \left[\sum_{i \in H} d_i^\sigma x'_i\right]$.*

Proof.

□

Define the map $s\colon (V,+) \longrightarrow (W,+)$ by $\left(\sum_{i\in I} d_i x_i\right)^s = \sum_{i\in I} d_i^\sigma x_i'$.
repn of semilinear map

STEP (viii). *$\mathcal{S}$ is induced by the semilinear map $\Sigma = [\sigma, s]$: for all $\langle v\rangle \in \mathbb{P}_0 V$ we have $\langle v\rangle^{\mathcal{S}} = [v^s]$.*

PROOF.

□

This completes our proof of the Fundamental Theorem of Projective Geometry (8.2). □

Chapter 9

Pairings, Isometries, and Automorphisms

The classical groups are linear groups that are isomorphism (isometry) groups of forms defined on the underlying space. The underlying concept is that of pairings of spaces.

9.1 Pairings

As before, D is a division ring. We let $V = {}_DV$, a left D-space, and $W = W_D$, a right D-space. A *pairing* of V and W is a bilinear map $m\colon V \times W \longrightarrow D$. That is, for all $u, v \in V$, $w, y \in W$, and $a, b \in D$:

(i) $m(u + v, w) = m(u, w) + m(v, w)$;

(ii) $m(u, w + y) = m(u, w) + m(u, y)$;

(iii) $m(av, wb) = am(v, w)b$.

The motivating example is the *canonical pairing* m^{can} of V with its dual $W = V^*$, where

$$m^{\mathrm{can}}(v, \lambda) = v\lambda\,,$$

for all $v \in V$ and $\lambda \in V^*$. If instead we start with a right D-space W, then the canonical pairing is $m^{\mathrm{can}}\colon W^* \times W \longrightarrow D$ given by $m^{\mathrm{can}}(\mu, w) = \mu w$.

Let U be a subspace of V and Y a subspace of W. Then

$$U^\perp = \{\, w \in W \mid m(u, w) = 0, \text{ for all } u \in U \,\} \text{ and}$$

$$^\perp Y = \{\, v \in V \mid m(v, y) = 0, \text{ for all } y \in Y \,\}\,.$$

The *right radical* of the pairing m is $V^\perp$ and its *left radical* is $^\perp W$. The pairing m is *nondegenerate* if both its radicals are 0: $V^\perp = 0$ and $^\perp W = 0$. If $U \leq V$

and $Y \leq W$ with $m|_{U \times Y}$ identically 0, then we call the pair (U, Y) *totally isotropic.*

(9.1). Lemma.

(a) *For a pairing $m\colon V \times W \longrightarrow D$, the map $\rho^m\colon w \mapsto m(\cdot, w)$ is a D-homomorphism of W into V^* and the map $\lambda^m\colon v \mapsto m(v, \cdot)$ is a D-homomorphism of V into W^*. Here $\ker \rho^m = V^\perp$ and $\ker \lambda^m = {}^\perp W$.*

(b) *The pairing $m\colon V \times W \longrightarrow D$ is nondegenerate if and only if the map $\rho^m\colon w \mapsto m(\cdot, w)$ is an injection of W into V^* and the map $\lambda^m\colon v \mapsto m(v, \cdot)$ is an injection of V into W^*.*

Proof. For all $a, b \in D$ and $u, v \in V$

$$\begin{aligned}(au + bv)\rho^m_w &= m(au + bv, w) = m(au, w) + m(bv, w) \\ &= am(u, w) + bm(v, w) = a(u\rho^m_w) + b(w\rho^m_w)\,,\end{aligned}$$

so $\rho^m_w \in V^*$. Furthermore for $x \in W$

$$v\rho^m_{wa+xb} = m(v, wa + xb) = m(v, w)a + m(v, x)b = v(\rho^m_w a + v\rho^m_x b)\,,$$

and ρ is D-linear. It kernel is

$$\{\, w \in W \mid v\rho^m_w = m(v, w) = 0\,, \text{ for all } v \in V \,\} = V^\perp\,.$$

This gives the first part of (a), and the rest of that part follows similarly (or by applying the first part to the opposite pairing; see page 113 below).

Part (b) then follows directly. □

(9.2). Corollary. *For the pairing $m\colon V \times W \longrightarrow D$, set $V^0 = V/{}^\perp W$ and $W^0 = W/V^\perp$. Then $m^0\colon V^0 \times W^0 \longrightarrow D$ given by $m^0(v + {}^\perp W, w + V^\perp) = m(v, w)$ is a well-defined nondegenerate pairing.* □

(9.3). Lemma. *Let $m\colon V \times W \longrightarrow D$ be a nondegenerate pairing. Let finite dimensional $U \leq V$ and finite dimensional $Y \leq W$.*

(a) *The codimension of $U^\perp$ in W equals the dimension of U, and ${}^\perp(U^\perp) = U$.*

(b) *The codimension of ${}^\perp Y$ in V equals the dimension of Y, and $({}^\perp Y)^\perp = Y$.*

(c) *$m|_{U \times Y}$ is nondegenerate if and only if $\dim_D(U) = \dim_D(Y)$, $V = U \oplus {}^\perp Y$, and $W = Y \oplus U^\perp$.*

Proof. (a) Let $\dim_D(U) = d$.

When we apply Lemma (9.1) to the restriction of m to $U \times W$ we learn that $U^\perp$ is the kernel of the map $\rho\colon W \mapsto U^*$. Thus the codimension e of $U^\perp$ in W is at most $\dim_D(V^*) = \dim_D(V) = d$. Write $W = \left(\bigoplus_{i=1}^e w_i D\right) \oplus U^\perp$. Then by nondegeneracy

$$0 = {}^\perp W = \Big(\bigcap_{i=1}^e {}^\perp w_i\Big) \cap {}^\perp(U^\perp)\,.$$

In particular

$$\dim_D({}^\perp(U^\perp)) \le \operatorname{codim}_D\Big(\bigcap_{i=1}^{e} {}^\perp w_i\Big) \le e \le d\,.$$

But always ${}^\perp(U^\perp) \ge U$ of dimension d. Therefore

$$\dim_D(U) = d = e = \operatorname{codim}_D(U^\perp)$$

and ${}^\perp(U^\perp) = U$.

(b) is similar.

(c) Again by Lemma (9.1) there are injections $U \longrightarrow Y^*$ and $Y \longrightarrow U^*$ so $\dim_D(U) \le \dim_D(Y^*) = \dim_D(Y)$ and $\dim_D(Y) \le \dim_D(U^*) = \dim_D(U)$, hence $\dim_D(U) = \dim_D(Y)$. We have just seen that the codimension of $U^\perp$ is equal to the dimension of U and so also to the dimension of Y. As $m|_{U\times Y}$ is nondegenerate, $Y \cap U^\perp = \{0\}$. Therefore $W = Y \oplus U^\perp$ and similarly $V = U \oplus {}^\perp Y$. □

(9.4). Lemma. *Let $m\colon V \times W \longrightarrow D$ be a nondegenerate pairing. Let finite dimensional $U_0 \le V$ and finite dimensional $Y_0 \le W$. Then there are U and Y with $U_0 \le U \le V$, $Y_0 \le Y \le W$, $m|_{U\times Y}$ nondegenerate, and* $\dim_D(U) = \dim_D(Y) \le \dim_D(U_0) + \dim_D(Y_0)$.

Proof. Let $\dim_D(U_0) = k$ and $\dim_D(Y_0) = l$.

Let Y_1 be a complement to $(U_0 \cap {}^\perp Y_0)^\perp$ in W, so Y_1 has dimension $d = \dim(U_0 \cap {}^\perp Y_0) \le k$ by Lemma (9.3). Similarly let U_1 be a complement to ${}^\perp(Y_0 \cup U_0^\perp)$ in V of dimension $e = \dim(Y_0 \cap U_0^\perp) \le l$. We then set

$$U = U_0 \oplus U_1 \quad \text{and} \quad Y = Y_0 \oplus Y_1\,,$$

both of dimension at most $k + l$. We claim that $m|_{U\times Y}$ is nondegenerate.

Let $u = u_0 + u_1 \in U$ be a nonzero element with $u_0 \in U_0$ and $u_1 \in U_1$. We will find a $y \in Y$ with $m(u, y) \ne 0$, and so demonstrate $U^\perp \cap Y = 0$. If $u_1 \ne 0$ then there is a $y \in U_0^\perp \cap Y_0$ with

$$0 \ne m(u_1, y) = m(u_0, y) + m(u_1, y) = m(u, y)\,,$$

so we may assume $u = u_0 \in U_0$.

If $u_0 \notin U_0 \cap {}^\perp Y_0$, then there is a $y \in Y_0$ with $m(u, y) = m(u_0, y) \ne 0$. Finally if $0 \ne u_0 \in U_0 \cap {}^\perp Y_0$, then by nondegeneracy of m there is a $y \in Y_1$ with $m(u, y) \ne 0$.

We conclude that $U^\perp \cap Y = 0$ and similarly $U \cap {}^\perp Y = 0$, thus $m|_{U\times Y}$ is nondegenerate. In particular $\dim U = \dim Y \le k + l = \dim U_0 + \dim Y_0$. □

A particular consequence of Lemma (9.3) is that for the finite dimensional space $V = U$ there is an essentially unique nondegenerate pairing, the canonical one m^{can} with $W = Y = V^*$ (see Lemma (9.7) for a precise statement). This is not the case for infinite dimensional V. As in Section 7.1, to each basis $\mathcal{B}$ of V, we associate the dual subset $\mathcal{B}^* = \{\, \lambda_y \mid y \in \mathcal{B} \,\}$ of V^* given by $x\lambda_y = \delta_{x,y}$ for

$x, y \in \mathcal{B}$. Then $\mathcal{B}^*$ is linearly independent, and we let $V^{\mathcal{B}}$ be the subspace of V^* with basis $\mathcal{B}^*$. The restriction of the canonical pairing, $m^{\mathcal{B}} = m^{\mathrm{can}}|_{V \times V^{\mathcal{B}}}$, is a nondegenerate pairing of V and $V^{\mathcal{B}}$. For finite dimensional V, the space $V^{\mathcal{B}}$ is all of V^* (by Corollary (7.2)) and so $m^{\mathcal{B}} = m^{\mathrm{can}}$. For $\dim_D(V)$ infinite,

$$\dim_D(V^*) > \dim_D(V) = \dim_D(V^{\mathcal{B}})$$

by Proposition (7.3)(c); so the two nondegenerate pairings m^{can} and $m^{\mathcal{B}}$ of V are different in an essential way.

In general, we say that a pairing $m \colon V \times W \longrightarrow D$ is *hyperbolic* provided it admits dual bases: there are a basis $\mathcal{V} = \{ v_i \mid i \in I \}$ for V and basis $\mathcal{W} = \{ w_i \mid i \in I \}$ for W such that

$$m(v_i, w_j) = \delta_{i,j}, \quad \text{for all } i, j \in I\,.$$

Thus each $m^{\mathcal{B}}$ is hyperbolic, essentially by definition. Clearly a necessary condition for the pairing to be hyperbolic is that it be nondegenerate with $\dim V = |I| = \dim W$. This is not, in general, sufficient, although it is in in certain cases. We have remarked above that all nondegenerate pairings in finite dimension are hyperbolic. More surprising is that this remains true in countable dimension. The proof is a nice example of a “back-and-forth” argument, here combined with a Gram-Schmidt style calculation.

(9.5). THEOREM. *If ${}_DV$ and W_D both have countable dimension, then every nondegenerate pairing $m \colon V \times W \longrightarrow D$ is hyperbolic.*

PROOF. Choose the bases $\mathcal{U} = \mathcal{U}_0 = \{u_1, u_2, \dots\}$ for V and $\mathcal{X} = \mathcal{X}_0 = \{x_1, x_2, \dots\}$ for W. We construct bases $\mathcal{V} = \{v_1, v_2, \dots\}$ for V and $\mathcal{W} = \{w_1, w_2, \dots\}$ for W. At Step n we replace some $u_i \in \mathcal{U}$ by the vector $v_n \in \mathcal{V}$ and some $x_j \in \mathcal{X}$ by $w_n \in \mathcal{W}$. We do this in such a way that:

(i) $\mathcal{U}_n = \{v_n\} \cup \mathcal{U}_{n-1} \setminus \{u_i\}$ is still a basis for V and $\mathcal{X}_n = \{w_n\} \cup \mathcal{X}_{n-1} \setminus \{x_j\}$ is still a basis for W;

(ii) $m(v_a, w_b) = \delta_{a,b}$ for $a, b \leq n$;

(iii) all u_i of $\mathcal{U}$ and x_j of $\mathcal{X}$ are eventually replaced.

The result of this process is then a pair of dual bases $\mathcal{V} = \mathcal{U}_\infty$ for V and $\mathcal{W} = \mathcal{X}_\infty$ for W that reveal m as hyperbolic.

We describe Step n precisely. If n is odd, let i be the smallest index i with u_i not yet replaced. Set

$$v_n = u_i - \sum_{k=1}^{n-1} m(u_i, w_k) v_k\,,$$

and let $V_n = \langle v_1, \dots, v_n \rangle$ of dimension n. By Lemma (9.4) every element of V_n^* can be induced by some element of W. Let j be minimal subject to the condition that the linear functional $\rho_j \colon V_n \longrightarrow D$, given by $\rho_j(v) = m(v, x_j)$, is not induced by any element of $W_{n-1} = \langle w_1, \ldots, w_{n-1}\rangle$. Set

$$w_n = x_j - \sum_{k=1}^{n-1} w_k m(v_k, x_j)\,.$$

If instead n is even, we first replace x_j and only then replace u_i. That is, we first choose j to be the smallest index for which x_j has not already been replaced. We then define w_n according to the formula given above. Next we set $W_n = \langle w_1, \ldots, w_n\rangle$ and choose u_i with i minimal subject to the linear functional $\lambda_i \colon W_n \longrightarrow D$, given by $\lambda_i(w) = m(u_i, w)$, not being induced by any element of V_{n-1}. (Again Lemma (9.4) guarantees that such an i exists.) The element v_n is then defined as above.

Whether n is even or odd, these choices of v_n and w_n certainly give (i). Also Step $n-1$ provides us with (ii) for $a, b < n$.

For $b < n$

$$\begin{aligned} m(v_n, w_b) =& m(u_i - \sum_{k=1}^{n-1} m(u_i, w_k) v_k, w_b) \\ =& m(u_i, w_b) - \sum_{k=1}^{n-1} m(u_i, w_k) m(v_k, w_b) = 0\ . \end{aligned}$$

Similarly, for $a < n$

$$\begin{aligned} m(v_a, w_n) =& m(v_a, x_j - \sum_{k=1}^{n-1} w_k m(v_k, x_j)) \\ =& m(v_a, x_j) - \sum_{k=1}^{n-1} m(v_a, w_k) m(v_k, x_j) = 0\ . \end{aligned}$$

Depending upon whether n is odd or even, our choice of j or i guarantees that $m(v_n, w_n) = d$ is nonzero, so to complete (ii) at Step n we only need to replace one of v_n or w_n with its multiple by the scalar d^{-1}.

Finally, the element u_i will be replaced by the i^{th} odd step if not earlier, while x_j will have been replaced by the j^{th} even step. Thus both u_k and x_k are replaced by the time we have completed Step $2k$, giving (iii) and the theorem. □

(9.6). PROPOSITION. *For spaces ${}_D V$ and W_D there exists a nondegenerate pairing $m \colon V \times W \longrightarrow D$ if and only if* $\dim W \le \dim V^*$ *and* $\dim V \le \dim W^*$.

PROOF. Necessity follows from Lemma (9.1)(b).

For the other direction, we may assume $\dim V \le \dim W$. Choose a basis $\mathcal{B}$ of V, and let $V^{\mathcal{B}}$ be the subspace of the same dimension in V^* that was constructed above. By hypothesis there is a vector space injection $\theta \colon W \longrightarrow V^*$ with $V^{\mathcal{B}} \le W^\theta$. Define $m \colon V \times W \longrightarrow D$ by $m(v, w) = v(w^\theta)$. Then $V^\perp = 0$ as θ is injective and ${}^\perp W = 0$ as $m^{\mathcal{B}} \colon V \times V^{\mathcal{B}} \longrightarrow D$ is nondegenerate. □

9.2 Isometry groups of pairings

Given two pairings $m\colon {}_DV \times W_D \longrightarrow D$ and $n\colon {}_EU \times X_E \longrightarrow E$, we wish again to formalize the feeling that there is no essential difference between the two. Here this should mean that there are semilinear isomorphisms $s = [\sigma, s]$ from V to U and $t = [t, \sigma]$ from W to X with $n(vs, tw) = m(v, w)^\sigma$ always:

$$\begin{tikzcd}
{}_DV \arrow[d, "s"] & \times & W_D \arrow[d, "t"] \arrow[r, "m"] & D \arrow[d, "\sigma"] \\
{}_EU & \times & X_E \arrow[r, "n"] & E
\end{tikzcd}$$

Conversely, given (semi)isomorphisms $[\sigma, s]$ from V to U and $[t, \sigma]$ from W to X we can construct from the pairing $m\colon {}_DV \times W_D \longrightarrow D$ a new pairing $n\colon {}_EU \times X_E \longrightarrow E$ given by

$$n(u, x) = m(us^{-1}, t^{-1}w)^\sigma \,.$$

The triple (s, t, σ) is then an *equivalence* of the pairings (V, W, m) and (U, X, n). Equivalent pairs of spaces and forms are said to be *isometric*, although it might be better to say they are *semi-isometric*, since equivalences are induced by semilinear maps. We reserve the term *isometry* for situations where $D = E$ and $\sigma = 1$.

As a direct consequence of Lemmas (9.1) and (9.3) we have:

(9.7). LEMMA. *Let $m\colon V \times W \longrightarrow D$ be a nondegenerate pairing. Then there are subspaces W^ρ of V^* and V^λ of W^* with (V, W, m) isometric to (V, W^ρ, m^ρ) and $(V^\lambda, W, m^\lambda)$, where m^ρ and m^λ are the restrictions to W^ρ and V^λ of the canonical pairings $(V, V^*, m^{\mathrm{can}})$ and $(W^*, W, m^{\mathrm{can}})$.*

In particular, if either V or W has finite dimension, then m is isometric to the canonical pairings $(V, V^*, m^{\mathrm{can}})$ and $(W^*, W, m^{\mathrm{can}})$ and is hyperbolic. □

The group $\mathrm{GL}({}_DV) \times \mathrm{GL}(W_D)$ acts on $V \times W$ on the right. For $g \in \mathrm{GL}_D(V)$ and $h \in \mathrm{GL}_D(W)$ the element $f = (g, h) \in \mathrm{GL}({}_DV) \times \mathrm{GL}(W_D)$ acts according to

$$(v, w).f = (v, w).(g, h) = (vg, hw)\,,$$

for all $(v, w) \in V \times W$. We may also write $v.f$ for $v.g$ and $f.w$ for $h.w$. The notation $\mathrm{GL}(W_D)$ for $\mathrm{GL}_D(W)$ reminds us that $\mathrm{GL}_D(W)$ normally acts on W on the left since W is a right D-space. We have

$$(v, w)(g_1, h_1)(g_2, h_2) = (v.g_1, h_1.w)(g_2, h_2) = (v.g_1g_2, h_2h_1.w)\,.$$

Thus $(g_1, h_1)(g_2, h_2) = (g_1g_2, h_2h_1)$ in $\mathrm{GL}({}_DV) \times \mathrm{GL}(W_D)$.

An *isometry* (rather than self-isometry) of the pairing $m\colon V \times W \longrightarrow D$ is then a self-equivalence—an element

$$(g, h) \in \mathrm{GL}({}_DV) \times \mathrm{GL}(W_D)$$

with

$$m(v,w) = m(v.g, h.w)\,,$$

for all $(v,w) \in V \times W$. This can be viewed as the stabilizer of m in the (right) action of $\mathrm{GL}({}_DV) \times \mathrm{GL}(W_D)$ on the abelian group of pairings $\mathrm{Pair}_D(V,W)$ ($\simeq \mathrm{Hom}_D(V,W^*)$) given by $m \mapsto n = m^{(g,h)}$. The subgroup of $\mathrm{GL}({}_DV) \times \mathrm{GL}(W_D)$ consisting of all isometries of m (with the above product) will be denoted $\mathrm{GL}_D(V,W,m)$.

More generally, a (self-)*semi-isometry* (s,t,σ) of (V,W,m) is a pair of semilinear maps $[\sigma,s] \in \Gamma\mathrm{L}_D(V)$ and $[t,\sigma] \in \Gamma\mathrm{L}_D(W)$ satisfying

$$m(v,w) = m(v.s, t.w)^{\sigma^{-1}}\,.$$

The group of all semi-isometries of m is then $\Gamma\mathrm{L}_D(V,W,m)$ and has $\mathrm{GL}_D(V,W,m)$ as its normal subgroup of all semi-isometries with $\sigma = 1$.

(9.8). LEMMA. *Let $m\colon V \times W \longrightarrow D$ be a nondegenerate pairing, and $(g,h) \in \mathrm{GL}_D(V,W,m)$. Then $g = 1$ if and only if $h = 1$.*

PROOF. Suppose $(1,h) \in \mathrm{GL}_D(V,W,m)$. Then, for all $v \in V$ and $w \in W$, we have

$$0 = m(v,w) - m(v,hw) = m(v,(1-h)w)\,.$$

That is, $(1-h)W \in V^\perp$. As m is nondegenerate, this gives $(1-h)w = 0$ for all $w \in W$. Therefore $hw = w$ and $h = 1$.

Similarly, if $(g,1) \in \mathrm{GL}_D(V,W,m)$, then all $(1-g)v \in {}^\perp W = 0$ and $g = 1$. □

(9.9). COROLLARY. *Let $m\colon V \times W \longrightarrow D$ be a nondegenerate pairing, and $G = \mathrm{GL}_D(V,W,m)$ its isometry group.*

(a) *The restriction map $(g,h) \mapsto g$ is an injection of G into $\mathrm{GL}_D(V)$.*

(b) *The restriction map $(g,h) \mapsto h^{-1}$ is an injection of G into $\mathrm{GL}_D(W)$.*

PROOF. Given that $(g_1,h_1)(g_2,h_2) = (g_1g_2, h_2h_1)$ in G, this is immediate from Lemma (9.8) . □

(9.10). LEMMA.

(a) *Each element $g \in \mathrm{GL}_D(V)$ has a natural linear action on V^* given by*

$$v(g\mu) = (vg)\mu\,,$$

for all $v \in V$ and $\mu \in V^*$. In particular, for all $g \in \mathrm{GL}_D(V)$ we have $(g,g^{-1}) \in \mathrm{GL}_D(V,V^*,m^{\mathrm{can}})$.

(b) *Each element $[\sigma,s] \in \Gamma\mathrm{L}_D(V)$ has a natural σ^{-1}-semilinear action on V^* given by*

$$v(s\mu) = ((vs)\mu)^{\sigma^{-1}} \ : \qquad \begin{tikzcd} {}_DV \arrow[r, "{[\sigma,s]}"] \arrow[d, "s\mu"] & {}_DV \arrow[d, "\mu"] \\ D \arrow[r, "\sigma"] & D \end{tikzcd}$$

for all $v \in V$ and $\mu \in V^*$. In particular, for all $[\tau, t] \in \Gamma\mathrm{L}_D(V)$ we have $(t, t^{-1}, \tau) \in \Gamma\mathrm{L}_D(V, V^*, m^{\mathrm{can}})$.

Proof. [1] We need only prove (b). We first check that we have an action—that $s\mu \in V^*$. Let $u, v \in V$, $\mu \in V^*$, and $a \in D$:

$$\begin{aligned}(u+v)s\mu &= (((u+v)s)\mu)^{\sigma^{-1}} = ((us+vs)\mu)^{\sigma^{-1}}\\ &= ((us)\mu + (vs)\mu)^{\sigma^{-1}} = ((us)\mu)^{\sigma^{-1}} + ((vs)\mu)^{\sigma^{-1}}\\ &= u(s\mu) + v(s\mu)\end{aligned}$$

and

$$\begin{aligned}(av)(s\mu) &= (((av)s)\mu)^{\sigma^{-1}} = ((a^{\sigma}(vs))\mu)^{\sigma^{-1}}\\ &= (a^{\sigma}((vs)\mu))^{\sigma^{-1}} = a((vs)\mu)^{\sigma^{-1}}\\ &= a(v(s\mu))\,.\end{aligned}$$

Then we check that the action is σ^{-1}-semilinear. Let $v \in V$, $\mu, \lambda \in V^*$, and $a \in D$:

$$\begin{aligned}v(s(\mu+\lambda)) &= ((vs)(\mu+\lambda))^{\sigma^{-1}} = ((vs)\mu + (vs)\lambda)^{\sigma^{-1}}\\ &= ((vs)\mu)^{\sigma^{-1}} + ((vs)\lambda)^{\sigma^{-1}} = v(s\mu) + v(s\lambda)\\ &= v(s\mu + s\lambda)\end{aligned}$$

and

$$\begin{aligned}v(s(\mu a)) &= ((vs)(\mu a))^{\sigma^{-1}} = (((vs)\mu)a)^{\sigma^{-1}}\\ &= ((vs)\mu)^{\sigma^{-1}} a^{\sigma^{-1}} = (v(s\mu))a^{\sigma^{-1}}\\ &= v((s\mu)a^{\sigma^{-1}})\,. \qquad \square\end{aligned}$$

The previous two results immediately give

(9.11). Corollary.

(a) $\mathrm{GL}_D(V) = \mathrm{GL}_D(V, V^*, m^{\mathrm{can}})|_V$.

(b) $\Gamma\mathrm{L}_D(V) = \Gamma\mathrm{L}_D(V, V^*, m^{\mathrm{can}})|_V$. □

However (semi)linear transformations of V need not extend to (semi)isometries of nondegenerate pairings. Using Lemmas (9.7) and (9.10), we can say exactly when extension is possible.

(9.12). Proposition. *Let $W \le V^*$ with $m^{\mathrm{can}} \colon V \times W \longrightarrow D$ nondegenerate. For $[\sigma, s] \in \Gamma\mathrm{L}_D(V)$ there is a $[t, \sigma] \in \Gamma\mathrm{L}_D(W)$ with $(s, t, \sigma) \in \Gamma\mathrm{L}_D(V, W, m^{\mathrm{can}}|_{V\times W})$ if and only if $s(W) = W$ in the action described under Lemma (9.10). In this case $t = s^{-1}|_W$ gives the unique element (s, t, σ) of $\Gamma\mathrm{L}_D(V, W, m^{\mathrm{can}}|_{V\times W})$ extending $[\sigma, s]$.*

[1] "It can be easily checked . . ."

Proof. If $s(W)=W$ then $(s, s^{-1}|_W, \sigma) \in \Gamma\mathrm{L}_D(V, W, m^{\mathrm{can}}|_{V\times W})$ by Lemma (9.10)(b).

Now write m for $m^{\mathrm{can}}|_{V\times W}$ and assume $(s,t,\sigma) \in \Gamma\mathrm{L}_D(V,W,m)$. For fixed $w \in W$

$$m(v,w)=0 \iff m(vs,tw)=0\,.$$

Therefore the hyperplane ${}^{\perp}(tw)$ of V is equal to $({}^{\perp}w)s$. On the other hand, since $(s,s^{-1},\sigma) \in \Gamma\mathrm{L}_D(V,V^*,m^{\mathrm{can}})$ we similarly have ${}^{\perp}(s^{-1}w)=({}^{\perp}w)s$. Both m^{can} and its restriction to $V\times W$ are nondegenerate, so we find an equality of 1-spaces in V^*:

$$s^{-1}(wD)=(s^{-1}w)D=(({}^{\perp}w)s)^{\perp}=(tw)D \in W\,.$$

That is, s^{-1} takes 1-spaces of W to 1-spaces of W. This then is also true for s, and we get $s(W)=W$, as desired.

If $(s,s^{-1}|_W,\sigma)$ and (s,t,σ) are in $\Gamma\mathrm{L}_D(V,W,m^{\mathrm{can}}|_{V\times W})$, then

$$(s,s^{-1}|_W,\sigma)(s,t,\sigma)^{-1}=(1,(s^{-1}|_W)t^{-1},1) \in \mathrm{GL}_D(V,W,m^{\mathrm{can}}|_{V\times W})\,.$$

By Lemma (9.8) we have $(s^{-1}|_W)t^{-1}=1$ and $s^{-1}|_W=t$. □

9.3 Opposites

If $m\colon V\times W \longrightarrow D$ is a pairing, then there is a natural associated *opposite pairing* $m^{\mathrm{op}}\colon {}_{D^{\mathrm{op}}}W \times V_{D^{\mathrm{op}}} \longrightarrow D^{\mathrm{op}}$ given by

$$m^{\mathrm{op}}(w,v)=m(v,w)\,,$$

for all $v\in V$ and $w \in W$. Although the underlying sets are unchanged, it might be clearer to write this as

$$m^{\mathrm{op}}(w^{\mathrm{op}},v^{\mathrm{op}})=m(v,w)\,.$$

Each s that is σ-semilinear on V gives rise to the σ^{-1}-semilinear s^{op} on V^{op} given by

$$s^{\mathrm{op}}v^{\mathrm{op}}=(vs^{-1})^{\mathrm{op}}\,,$$

and actions on W and W^{op} are related similarly. (Compare with Corollary (7.6).)

We then easily have:

(9.13). Theorem. *The map* $(s,t,\sigma)\mapsto((t^{\mathrm{op}})^{-1},(s^{\mathrm{op}})^{-1},\sigma)$ *is an isomorphism of the groups* $\Gamma\mathrm{L}_D(V,W,m)$ *and* $\Gamma\mathrm{L}_{D^{\mathrm{op}}}(W^{\mathrm{op}},V^{\mathrm{op}},m^{\mathrm{op}})$ *which restricts, for* $\sigma=1$*, to an isomorphism of* $\mathrm{GL}_D(V,W,m)$ *and* $\mathrm{GL}_{D^{\mathrm{op}}}(W^{\mathrm{op}},V^{\mathrm{op}},m^{\mathrm{op}})$. □

If m is nondegenerate, then by Lemma (9.7) and Proposition (9.12), the elements of $\Gamma\mathrm{L}(V,W,m)$ all have the form (s,s^{-1}). With this and our naming convention in mind, the isomorphisms of the theorem take a more striking form:

(9.14). Corollary. *Let the pairing* $m\colon V\times W\longrightarrow D$ *be nondegenerate.*

(a) $\Gamma\mathrm{L}_D(V,W,m)=\Gamma\mathrm{L}_{D^{\mathrm{op}}}(W^{\mathrm{op}},V^{\mathrm{op}},m^{\mathrm{op}})$.

(b) $\mathrm{GL}_D(V,W,m)=\mathrm{GL}_{D^{\mathrm{op}}}(W^{\mathrm{op}},V^{\mathrm{op}},m^{\mathrm{op}})$. □

9.4 Finitary isometry groups

We start with an important general result.

(9.15). LEMMA. *Let $m\colon V \times W \longrightarrow D$ be a nondegenerate pairing, and let $(g,h) \in \mathrm{GL}_D(V,W,m)$.*

(a) *$C_W(h) = V(g-1)^\perp$ and $C_V(g) = {}^\perp(h-1)W$.*

(b) *$g \in \mathrm{FGL}_D(V)$ if and only if $h \in \mathrm{FGL}_D(W)$. In this case $\deg_V(g) = \deg_W(h)$.*

PROOF. For all $v \in V$ and fixed w,

$$\begin{aligned} m(v(g-1),w) &= m(vg,w) - m(v,w) \\ &= m(vg,w) - m(vg,hw) \\ &= m(vg,(1-h)w)\,. \end{aligned}$$

Therefore $w \in V(g-1)^\perp$ if and only if $w \in C_W(h)$. This gives (a).

For (b) assume that $\deg_V(g)$ is finite. Then

$$\begin{aligned} \deg_V(g) &= \dim_D(V(g-1)) = \operatorname{codim}_D(V(g-1)^\perp) \\ &= \operatorname{codim}_D(C_W(h)) = \dim_D((h-1)W) \\ &= \deg_W(h)\,, \end{aligned}$$

as desired. □

The *finitary general linear group* $\mathrm{FGL}_D(V,W,m)$ consists of those elements $(g,h) \in \mathrm{GL}_D(V,W,m)$ with $g \in \mathrm{FGL}_D(V)$ and $h \in \mathrm{FGL}_D(W)$. By Lemma (9.15)(c) it is enough to require one of these and the degree of the element (g,h) in this action is well-defined. As in Corollary (9.11) we have $\mathrm{FGL}_D(V) = \mathrm{FGL}_D(V,V^*,m^{\mathrm{can}})|_V$.

(9.16). PROPOSITION. *If $m\colon U \times Y \longrightarrow D$ is a nondegenerate pairing with U and Y finite dimensional, then*

$$\mathrm{GL}_D(U,Y,m) \simeq \mathrm{GL}_D(U,Y,m)|_U = \mathrm{GL}({}_DU) = \mathrm{GL}_D(U)$$

and

$$\mathrm{GL}_D(U,Y,m) \simeq \mathrm{GL}_D(U,Y,m)|_Y = \mathrm{GL}(Y_D) = \mathrm{GL}_D(Y)\,.$$

PROOF. We have

$$\begin{aligned} \mathrm{GL}_D(U,Y,m) &\simeq \mathrm{GL}_D(U,Y,m)|_U && \text{by Corollary (9.9)} \\ &\simeq \mathrm{GL}_D(U,Y^\rho,m^{\mathrm{can}}|_{U\times Y'})|_U && \text{by Lemma (9.7)} \\ &= \mathrm{GL}_D(U,U^*,m^{\mathrm{can}})|_U && \text{by Corollary (7.2)} \\ &= \mathrm{GL}_D(U) && \text{by Corollary (9.11).} \end{aligned}$$

Similarly $\mathrm{GL}_D(U,Y,m) \simeq \mathrm{GL}_D(Y)$. □

By Theorem (7.7) there is an isomorphism of $\mathrm{GL}_n(D)$ and $\mathrm{GL}_D(U)$ for $n = \dim_D(U)$. Different isomorphisms are related by a change of basis. In particular, via any such isomorphism the Dieudonné determinant is well-defined on $\mathrm{GL}_D(U)$ and so also on $\mathrm{GL}_D(U,Y,m)$ (using the proposition). Let $\mathrm{SL}_D(U)$ and $\mathrm{SL}_D(U,Y,m)$ be the images in $\mathrm{GL}_D(U)$ and $\mathrm{GL}_D(U,Y,m)$ of $\mathrm{SL}_n(D)$, the kernel of the Dieudonné determinant. Especially by Theorem (6.29) we have $\mathrm{SL}_D(U,Y,m)$ quasisimple and equal to $\mathrm{GL}_D(U,Y,m)'$ provided $(n,|D|) \neq (2,2),(2,3)$.

(9.17). PROPOSITION. *For the nondegenerate pairing $m\colon V \times W \longrightarrow D$, the group $\mathrm{FGL}_D(V,W,m)$ is the directed limit of its subgroups*

$$G_{U,Y} \simeq \mathrm{GL}_D(U,Y,m|_{U\times Y}) \simeq \mathrm{GL}_D(U)$$

for U finite dimensional in V, Y finite dimensional in W, and $m|_{U\times Y}$ nondegenerate. Here the element of $(g,h) \in G_{U,Y}$ corresponding to $(g_0,h_0) \in \mathrm{GL}_D(U,Y,m|_{U\times Y})$ is defined to act on $V = U \oplus {}^{\perp}Y$ according to

$$g|_U = g_0 \quad \textit{and} \quad {}^{\perp}Y(g-1) = 0$$

and on $W = Y \oplus U^{\perp}$ via

$$h|_Y = h_0 \quad \textit{and} \quad (h-1)U^{\perp} = 0\,.$$

PROOF. By Lemma (9.3)(c), the groups $G_{U,Y}$ with g and h acting as described are subgroups of $\mathrm{FGL}_D(V,W,m)$ and are isomorphic to $\mathrm{GL}_D(U)$ by the previous proposition.

Each $(g,h) \in \mathrm{FGL}_D(V,W,m)$ is in some $G_{U,Y}$ by Lemma (9.4) with $U_0 = V(g-1)$ and $Y_0 = (h-1)W$.

Furthermore, if U_1, U_2, Y_1, Y_2 are finite dimensional with both $m|_{U_1\times Y_1}$ and $m|_{U_2\times Y_2}$ nondegenerate, then a second application of Lemma (9.4), now with $U_0 = U_1 + U_2$ and $Y_0 = Y_1 + Y_2$, provides a finite dimensional and nondegenerate pairing $m|_{U\times Y}$ with $\langle G_{U_1,Y_1}, G_{U_2,Y_2}\rangle \leq G_{U,Y}$. Therefore the set of all $G_{U,Y}$ is indeed directed with $\mathrm{FGL}_D(V,W,m)$ as its directed limit. □

(9.18). THEOREM. *Let $m\colon V \times W \longrightarrow D$ be a nondegenerate pairing. For $(\dim_D(U),|D|) \neq (2,2),(2,3)$ let $S_{U,Y}$ be the derived group of the group $G_{U,Y}$ defined in Proposition (9.17).*

The group $\mathrm{SL}_D(V,W,m) = \mathrm{FGL}_D(V,W,m)'$ is the directed limit of its subgroups

$$S_{U,Y} \simeq \mathrm{SL}_D(U,Y,m|_{U\times Y}) \simeq \mathrm{SL}_D(U)$$

for U finite dimensional in V, Y finite dimensional in W, and $m|_{U\times Y}$ nondegenerate.

This group $\mathrm{SL}_D(V,W,m)$ is quasisimple if V and W have finite dimension (with $(\dim_D(V),|D|) \neq (2,2),(2,3)$) and simple if V and W have infinite dimension.

Proof. By Proposition (6.33) we have $G'_{U,Y} = S_{U,Y} \simeq \mathrm{SL}_D(U)$.

By the proposition, for each $g, h \in \mathrm{FGL}(V, W, m)$ there is a a $G_{U,Y}$ containing both g and h hence $[g, h] \in S_{U,Y}$.

Furthermore, for finite dimensional nondegenerate $m\colon U_1 \times Y_1$ and $m\colon U_2 \times Y_2$, by Lemma (9.4) there is a finite dimensional and nondegenerate $m\colon U \times Y$ with

$$\langle G_{U_1,Y_1}, G_{U_2,Y_2}\rangle \leq G_{U,Y}$$

hence

$$\langle S_{U_1,Y_1}, S_{U_2,Y_2}\rangle = \langle G'_{U_1,Y_1}, G'_{U_2,Y_2}\rangle \leq G'_{U,Y} = S_{U,Y}\,.$$

Therefore the subgroup of $\mathrm{FGL}_D(V, W, m)$ generated by all commutators is the directed limit of the various quasisimple subgroups $S_{U,Y}$. By Problem (4.20) the groups itself must then be quasisimple.

If V has finite dimension, then $\mathrm{SL}_D(V, W, m) = S_{V,W}$ is quasisimple. Assume V has infinite dimension, and let $z \in \mathrm{Z}(\mathrm{SL}_D(V, W, m))$. Then there is some $S_{U,Y}$ with $z \in S_{U,Y}$, hence z scalar is on U (and Y). Thus either $z = 1$ of $\deg_V(z) = \deg_U(z) = \dim_D(U)$. Since V has infinite dimension, we may find finite dimensional U' and Y' with $U < U', Y < Y$ and $m|_{U'\times Y'}$ nondegenerate. But then z is in the center $S_{U',Y'}$, having degree strictly less than $\dim_D(U')$. We conclude that $z = 1$, and so $\mathrm{SL}_D(V, W, m)$ is simple. □

Here we have defined the *finitary special linear group* $\mathrm{SL}_D(V, W, m)$ to be the derived group of the finitary general linear group $\mathrm{FGL}_D(V, W, m)$. Instead we could have observed, using Proposition (9.17), that the Dieudonné determinant has a well-defined and unique extension from the finite dimensional to the finitary groups $\mathrm{FGL}_D(V, W, m)$ (with m nondegenerate) and that $\mathrm{SL}_D(V, W, m)$ is the corresponding kernel. Thus the finitary special linear groups SL relate to the finitary general linear groups FGL in the same way that the alternating groups Alt relate to the finitary symmetric groups FSym.

(9.19). Theorem. *Let $\mathcal{B}$ be the canonical basis of the D-vector space $V = D^{\mathbb{N}}$. Then the finitary linear group $\mathrm{FGL}_D(V, V^{\mathcal{B}}, m^{\mathcal{B}})$ is isomorphic to the stable linear group $\mathrm{GL}(D)$ of Problem (6.37), and that isomorphism restricts to an isomorphism of $\mathrm{SL}_D(V, V^{\mathcal{B}}, m^{\mathcal{B}}) = \mathrm{FGL}_D(V, V^{\mathcal{B}}, m^{\mathcal{B}})'$ with the elementary stable linear group $\mathrm{E}(D) = \mathrm{GL}(D)'$, which is simple.*

Proof. The dual bases $\mathcal{B}$ of V and $\mathcal{B}^*$ of $V^{\mathcal{B}}$ can be used to represent the group $\mathrm{FGL}_D(V, V^{\mathcal{B}}, m^{\mathcal{B}})$ by infinite matrices that differ from the identity only in a finite dimensional upper-lefthand corner, as in Problem (6.37). That is, if V_n is the span in V of the first n vectors of the basis $\mathcal{B}$ and $V_n^{\mathcal{B}}$ the subspace of V^* spanned by the corresponding initial segment of $\mathcal{B}^*$, then $\mathrm{FGL}_D(V, V^{\mathcal{B}}, m^{\mathcal{B}})$ is the ascending directed limit of the groups $\mathrm{GL}_D(V_n)$ ($\simeq \mathrm{GL}_D(V_n, V_n^{\mathcal{B}}, m^{\mathcal{B}}|_{V_n\times V_n^{\mathcal{B}}})$) with respect to the natural embedding.

Simplicity then comes from the theorem. □

(9.20). Corollary. *For $m\colon V \times W \longrightarrow D$ nondegenerate and infinite dimensional, the unique minimal normal subgroup of $\mathrm{GL}_D(V, W, m)$ is $\mathrm{SL}_D(V, W, m)$.*

Proof. Compare Corollary (5.9).

For $1 \neq n \in N \trianglelefteq \mathrm{GL}_D(V,W,m)$ choose a $g \in \mathrm{SL}_D(V,W,m)$ that does not commute with n. Then

$$1 \neq [g,n] = g^{-1}(n^{-1}gn) = (g^{-1}n^{-1}g)n \in \mathrm{SL}_D(V,W,m) \cap N\,.$$

Therefore $\mathrm{SL}_D(V,W,m) \cap N$ is a nontrivial normal subgroup of the simple group $\mathrm{SL}_D(V,W,m)$, hence $\mathrm{SL}_D(V,W,m) \leq N$. □

9.5 Transvections and elations

In a classical group $\mathrm{GL}_D(V,W,m)$, the element $t = (g,h)$ is an *ℓ-root element* (for $\ell = 1$ or 2) provided:

(a) $\ell = \dim_D(V(t-1))$ $(= \dim_D(V(g-1)) = \dim_D((h-1)W))$;

(b) $(t-1)^2 = 0$ (that is, $V(g-1)^2 = 0$ and $(h-1)^2W = 0$);

(c) the restriction of f to $V(g-1) \times (h-1)W$ is trivial.

Given an ℓ-root element t, the associated *ℓ-root subgroup* is the subgroup consisting of the identity and all ℓ-root elements t_0 with $V(t-1) = V(t_0-1)$ and $(t-1)W = (t_0-1)W$.

A $\mathrm{GL}_D(V,W,m)$ conjugate of an ℓ-root element is an ℓ-root element, and a $\mathrm{GL}_D(V,W,m)$ conjugate of an ℓ-root subgroup is an ℓ-root subgroup.

An element $t \in \mathrm{GL}_D(V,V^*,m^{\mathrm{can}})$ with $\deg Vt = 1$ and $(t-1)^2 = 0$ is a *transvection*. Every transvection t is has the form $\mathrm{t}(\lambda,v)$, with action on $x \in V$ given by

$$x.\,\mathrm{t}(\lambda,v) = x + x\lambda.v\,,$$

for some $v \in V$ and $\lambda \in V^*$ with $v.\lambda = 0$. The 1-space $\langle v\rangle = Dv \leq V$ is called the *center* of $\mathrm{t}(\lambda,v)$ while the 1-space $\langle\lambda\rangle \leq V^*$ is its *axis*. (Although the identity is not a transvection, the notation $\mathrm{t}(0,v) = 1 = \mathrm{t}(\lambda,0)$ is convenient.) By Lemma (9.15) a transvection on V also acts as a transvection on V^*. The action on V^* is given by

$$\mathrm{t}(\lambda,v).\mu = \mu + \lambda.v\mu\,.$$

By Lemma (9.15) a transvection on V also acts as a transvection on V^*. The action on V^* is given by

$$\mathrm{t}(\lambda,v).\mu = \mu + \lambda.v\mu\,.$$

(9.21). Theorem.

(a) *For $m\colon V \times W \longrightarrow D$ a nondegenerate pairing, we have*

$$\mathrm{SL}_D(V,W,m) \simeq \mathrm{T}_D(W^\rho,V) \simeq \mathrm{T}_D(W,V^\lambda),$$

the isomorphisms given by restriction.

(b) *Let V have infinite dimension and W be a subspace of V^* with ${}^\perp W = 0$. Then $m = m^{\mathrm{can}}|_{V\times W}$ is nondegenerate. Furthermore*

$$\mathrm{SL}_D(V, W, m) \simeq \mathrm{T}(W, V)$$

is simple. Especially

$$\mathrm{SL}_D(V, V^*, m^{\mathrm{can}}) \simeq \mathrm{SL}_D(V) = \mathrm{T}(V^*, V)$$

is simple.

Proof. (a) For $(\dim_D(V), |D|) = (2, 2)$ this is clear, and in finite dimensions the result is immediate from Proposition (9.16).

In general, there are many ways of seeing the isomorphisms, but perhaps the most elegant is to observe that all three groups are (isomorphic to) the directed limit of the subgroups $S_{U,Y}$ of Theorem (9.18). Simplicity then follows from that theorem as well.

(b) As $W \leq V^*$ we have $V^\perp = 0$, so m is nondegenerate. Then (a) and Theorem (9.18) apply. □

An *elation* e is the image in $\mathrm{PSL}_D(V)$ of a transvection of $\mathrm{SL}_D(V)$. The image $\bar{\mathrm{T}}(\varphi, x)$ of the transvection subgroup $\mathrm{T}(\varphi, x)$ is an *elation subgroup.*

(9.22). Lemma.

(a) *Every elation has a unique transvection preimage.*

(b) *Set $\bar{\mathrm{T}}(\varphi) = \bar{\mathrm{T}}(\varphi, V)$ and $\bar{\mathrm{T}}(v) = \bar{\mathrm{T}}(V^*, v)$ for each nonzero $\varphi \in V^*$ and $v \in V$. If $\bar{\mathrm{T}}(a) \cap \bar{\mathrm{T}}(b) \neq 1$ for nonzero $a, b \in V^* \cup V$ with $\langle a\rangle \neq \langle b\rangle$, then there are $\varphi \in V^*$ and $v \in V$ with $\{a, b\} = \{\varphi, v\}$, $v \in \ker\varphi$, and $\bar{\mathrm{T}}(\varphi, x) = \bar{\mathrm{T}}(\varphi) \cap \bar{\mathrm{T}}(v)$ isomorphic to $(D, +)$. For $\dim V \geq 3$, $\bar{\mathrm{T}}(\varphi) \neq \bar{\mathrm{T}}(v)$.*

(c) *Every abelian subgroup that is maximal subject to containing only the identity and elations is either $\bar{\mathrm{T}}(\varphi)$, for some $\varphi \in V^*$, or is $\bar{\mathrm{T}}(v)$, for some $v \in V$.*

Proof.

□

9.6 Rigidity and automorphisms

$\mathrm{Aut}_0(\mathrm{PSL}_D(V))$ will denote that subgroup of $\mathrm{Aut}(\mathrm{PSL}_D(V))$ composed of automorphisms that take elation subgroups to elation subgroups.

(9.23). Theorem. *Let $n \geq 3$.*

(a) $\mathrm{Aut}_0(\mathrm{PSL}_n(D))$ *has* $\mathrm{P\Gamma L}_D(V)$ *as a normal subgroup of index at most* 2.

(b) $\mathrm{Aut}_0(\mathrm{PSL}_n(D)) \neq \mathrm{P\Gamma L}_n(D)$ *if and only if D is isomorphic to D^{op}.*

Proof.

□

(9.24). Corollary. *For $n \geq 3$,*

$$\mathrm{Aut}(\mathrm{PSL}_n(q)) = \mathrm{P\Gamma L}_n(q)\langle\tau\rangle\,.$$

Proof. Let q be a power of the prime p. An automorphism of finite $\mathrm{PSL}_n(q)$ must take Sylow p-subgroups to Sylow p-subgroups. But the center of a Sylow p-subgroup of $\mathrm{PSL}_n(q)$ is an elation subgroup. □

9.7 Problems

(9.25). Problem. *The matrix "inverse-transpose" map $g \mapsto (g^{-1})^\top$ induces an automorphism of $\mathrm{PSL}_n(F)$, for any field F. For $n \geq 3$ this automorphism is not induced by any semilinear map.*

Find a semilinear $\Sigma = (\sigma, S) \in \Gamma\mathrm{L}_2(F)$ with $g^\Sigma = (g^{-1})^\top$ on $\mathrm{PSL}_2(F)$.

Remark. *The next problem implies that this automorphism must be semilinear when $n = 2$, but the direct calculation is more elementary.*

(9.26). Problem. *Let F be a field. This problem approaches*

Theorem (9.26)(a). $\mathrm{Aut}(\mathrm{PSL}_2(F)) = \mathrm{P\Gamma L}_2(F)$.

We actually prove the slightly easier

Theorem (9.26)(b). $\mathrm{Aut}_0(\mathrm{PSL}_2(F)) = \mathrm{P\Gamma L}_2(F)$.

As in Corollary (9.24) we get the important:

Corollary (9.26)(c). For finite fields F, we have $\mathrm{Aut}(\mathrm{PSL}_2(F)) = \mathrm{P\Gamma L}_2(F)$.

The proofs are presented through a sequence of parts.

Let $V = F^2$, the two dimensional F-space of row vectors, admitting the group $\Gamma\mathrm{L}_2(F) = \mathrm{Aut}(F) \ltimes \mathrm{GL}_2(F)$ acting via

$$(a,b)^{(\sigma,S)} = (a^\sigma, b^\sigma)S\,.$$

Thus $\mathrm{P\Gamma L}_2(F)$ (and its various subgroups) acts on the associated projective line $\mathbb{P}V = \mathbb{P}F^2$.

For the vectors $(a,b) \in V$ and 1-spaces $\langle(a,b)\rangle$, define the transvections

$$t_{\langle\vec{(0,1)}\rangle}(d) = \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix}$$

and

$$t_{\langle\vec{(1,a)}\rangle}(d) = \begin{pmatrix} 1-ad & -a^2d \\ d & 1+ad \end{pmatrix}$$

for each $d \in F$; so in particular

$$t_{\langle\vec{(1,0)}\rangle}(d) = \begin{pmatrix} 1 & 0 \\ d & 1 \end{pmatrix}.$$

Also define the transvection subgroup

$$T_{\langle\vec{v}\rangle} = \{\, t_{\langle\vec{v}\rangle}(d) \mid d \in F \,\}\,,$$

for each 1*-space* $\langle\vec{v}\rangle$ *in* V.

(a) (*i*) *For fixed* $\langle\vec{v}\rangle$, *prove that* $T_{\langle\vec{v}\rangle}$ *is a subgroup of* $\mathrm{SL}_2(F)$ *and that* $d \mapsto t_{\langle\vec{v}\rangle}(d)$ *is an isomorphism of the additive group* $(F, +)$ *with* $T_{\langle\vec{v}\rangle}$.

(*ii*) *For each* $t \in T_{\langle\vec{v}\rangle}$ *with* $t \neq 1$, *prove that* $V(t-1) = \langle\vec{v}\rangle$ *and* $V(t-1)^2 = \vec{0}$. *Conversely, show that any* $t \in \mathrm{GL}_2(F)$ *that has these two properties is* $t_{\langle\vec{v}\rangle}(d)$, *for some non-zero* $d \in F$.

*(*REMARK. *Once you have proved this characterization of transvections and transvection subgroups, in the rest of the problem you should not need to do much calculation with the complicated matrices above.)*

(*iii*) *For each* $g \in \Gamma\mathrm{L}_2(F)$, *prove that* $g^{-1}T_{\langle\vec{v}\rangle}g = T_{\langle\vec{v}g\rangle}$. *(Make sure you get the correct inverse for* g. *You may want to use the characterization of* (*ii*) *for this.)*

(*iv*) *Prove that* $g \in \Gamma\mathrm{L}_2(F)$ *normalizes* $T_{\langle\vec{v}\rangle}$ *if and only if* $\langle\vec{v}\rangle g = \langle\vec{v}\rangle$.

(*v*) *If finite* $|F| = p^k$, *for some prime* p, *prove that the set of the various* $T_{\langle\vec{v}\rangle}$ *is precisely the set of Sylow* p*-subgroups of* $\mathrm{GL}_2(F)$.

Each transvection subgroup $T_{\langle\vec{v}\rangle}$ *meets the group* $\mathrm{Z}_2(F)$ *of scalars trivially and contains all transvections of the subgroup* $T_{\langle\vec{v}\rangle} \times \mathrm{Z}_2(F)$ *(as*

$$\begin{pmatrix} r-1 & 0 \\ d & r-1 \end{pmatrix}$$

has rank 2 *for central* $r \neq 1$*). Thus the image* $\bar{T}_{\langle\vec{v}\rangle}$ *of* $T_{\langle\vec{v}\rangle}$ *in the quotient group* $\mathrm{PSL}_2(F)$ *inherits the properties of part* (a). *The subgroups* $\bar{T}_{\langle\vec{v}\rangle}$ *are the* elation *subgroups of* $\mathrm{PSL}_2(F)$.

$\mathrm{Aut}_0(\mathrm{PSL}_2(F))$ *is defined to be the subgroup of* $\mathrm{Aut}(\mathrm{PSL}_2(F))$ *that takes elation subgroups to elation subgroups. That is, for each* r *in* $\mathrm{Aut}_0(\mathrm{PSL}_2(F))$ *and each nonzero* $\vec{v} \in V$, *we have* $\bar{T}^r_{\langle\vec{v}\rangle} = \bar{T}_{\langle\vec{w}\rangle}$, *for some* $\vec{w}$. *By part* (a)(*iii*), $\mathrm{P\Gamma L}_2(F)$ *is at least contained in* $\mathrm{Aut}_0(\mathrm{PSL}_2(F))$.

(b) *Show that if* $|F|$ *is finite, then* $\mathrm{Aut}_0(\mathrm{PSL}_2(F)) = \mathrm{Aut}(\mathrm{PSL}_2(F))$.

We now commence with the proof of Theorem (9.26)(b). Set $G = \mathrm{PSL}_2(F)$, *and choose an arbitrary* $S \in \mathrm{Aut}_0(G)$.

We want to show that there is a semilinear map Σ *such that* $g^S = g^\Sigma$, *for each* $g \in \mathrm{PSL}_2(F)$, *for then* $\Sigma \mapsto S$ *describes a homomorphism of* $\Gamma\mathrm{L}_2(F)$ *onto* $\mathrm{Aut}_0(\mathrm{PSL}_2(F))$ *with kernel the group of scalar maps* $\mathrm{Z}_2(F)$, *proving Theorem 2.*

There are vectors $\vec{x}, \vec{y}, \vec{z} \in V$ *with* $\bar{T}^S_{\langle(1,0)\rangle} = \bar{T}_{\langle\vec{x}\rangle}$, $\bar{T}^S_{\langle(0,1)\rangle} = \bar{T}_{\langle\vec{y}\rangle}$, *and* $\bar{T}^S_{\langle(1,1)\rangle} = \bar{T}_{\langle\vec{z}\rangle}$. *Set* $\vec{x}_0 = \vec{x}$. *Then we can find a scalar* $e \in F$ *with so that* $\vec{z} = \vec{x}_0 + \vec{x}_1$ *upon setting* $\vec{x}_1 = e\vec{y}$. *Hence*

$$\bar{T}^S_{\langle(1,0)\rangle} = \bar{T}_{\langle\vec{x}_0\rangle}, \quad \bar{T}^S_{\langle(0,1)\rangle} = \bar{T}_{\langle\vec{x}_1\rangle}, \text{ and } \quad \bar{T}^S_{\langle(1,1)\rangle} = \bar{T}_{\langle\vec{x}_0+\vec{x}_1\rangle}.$$

Next, for every $a \in F$, *there is a uniquely determined* $a^\sigma \in F$ *with*

$$\bar{T}^S_{\langle(1,a)\rangle} = \bar{T}_{\langle\vec{x}_0+a^\sigma\vec{x}_1\rangle}\,.$$

Notice that $0^\sigma = 0$ *since* $\bar{T}^S_{\langle(1,0)\rangle} = \bar{T}_{\langle\vec{x}_0\rangle}$, *and* $1^\sigma = 1$ *since* $\bar{T}^S_{\langle(1,1)\rangle} = \bar{T}_{\langle\vec{x}_0+\vec{x}_1\rangle}$.

(c) *Prove that for all* $a, b \in F$:

(*i*) $(a+b)^\sigma = a^\sigma + b^\sigma$;

(*ii*) $(ab^2)^\sigma = a^\sigma(b^2)^\sigma$;

(*iii*) $(a^\sigma)^{-1} = (a^{-1})^\sigma$;

*(*HINT: *For (i) let $g \in \bar{T}_{\langle(0,1)\rangle}$ be represented by $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$. Therefore $g^S \in \bar{T}^S_{\langle(0,1)\rangle} = \bar{T}_{\langle \vec{x}_1 \rangle}$. Thus there is a $b' \in F$ for which g^S is represented by the transvection t with*

$$t\colon \vec{x}_0 \mapsto \vec{x}_0 + b'\vec{x}_1 \qquad t\colon \vec{x}_1 \mapsto \vec{x}_1\,.$$

Now calculate both sides of

$$(g^{-1}\bar{T}_{\langle(1,a)\rangle}g)^S = (g^{-1})^S\,\bar{T}^S_{\langle(1,a)\rangle}\,g^S$$

to conclude that, for all $a \in F$, $(a+b)^\sigma = a^\sigma + b'$. Using this, complete (i).

For (ii) and (iii) consider the action of elements represented by $\begin{pmatrix} b^{-1} & 0 \\ 0 & b \end{pmatrix}$ and $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ as well as their images under S.)

(d) *Prove that σ is in fact an automorphism of F.*

*(*REMARK. *More generally, any bijection $\sigma\colon F \longrightarrow F$ with* (i), (ii), *and* (iii) *of* (c) *is an automorphism of F.)*

(e) *Prove that $\Sigma\colon (a,b) \mapsto a^\sigma\vec{x}_0 + b^\sigma\vec{x}_1$ is a semilinear map on V with $\bar{T}^\Sigma_{\langle\vec{v}\rangle} = \bar{T}^S_{\langle\vec{v}\rangle}$, for all 1-spaces $\langle\vec{v}\rangle$ of V.*

(f) *Prove that $g^\Sigma = g^S$, for all $g \in G$.*

*(*HINT: *Let $I \in \mathrm{Aut}_0(G)$ be the automorphism $\bar{\Sigma}S^{-1}$, where $\bar{\Sigma}$ is the image of Σ in* $\mathrm{P\Gamma L}_2(F)$. *Then I fixes each transvection subgroup of G. For arbitrary $g \in G$, calculate $(g^{-1}\bar{T}_{\langle v\rangle}g)^I = (g^{-1})^I\bar{T}^I_{\langle v\rangle}g^I$ to prove that $\bar{T}^g_{\langle v\rangle} = \bar{T}^{g^I}_{\langle v\rangle}$, for all 1-spaces $\langle\vec{v}\rangle$. Therefore g^Ig^{-1} fixes all 1-spaces $\langle\vec{v}\rangle$ and so is the identity element of $G = \mathrm{PSL}_2(F)$.)*

Part (f) *completes the proof of Theorem (9.26)(b) and together with part* (b) *completes the proof of Corollary (9.26)(c).*

Chapter 10

Sesquilinear and Pseudoquadratic Forms

10.1 Sesquilinear forms

Of special interest in the previous chapter were those pairings $m\colon {}_DV \times W_D \longrightarrow D$ that gave rise to isomorphisms of the two projective spaces $\mathbb{P}_DV$ and $\mathbb{P}W_D$. This happens if and only if $\mathbb{P}_DV$ and $\mathbb{P}_{D^{\mathrm{op}}}W^{\mathrm{op}}$ are isomorphic. In this situation the Fundamental Theorem of Projective Geometry (8.2) tells us that the isomorphism is induced by a semilinear isomorphism $[\tau, t]$ from ${}_DV$ to ${}_{D^{\mathrm{op}}}W^{\mathrm{op}}$. This consists of an abelian group isomorphism $t\colon V \longrightarrow W^{\mathrm{op}}\,(= W)$ and a division ring isomorphism $\tau\colon D \longrightarrow D^{\mathrm{op}}$ (associated with the anti-isomorphism $\sigma = \tau\mathrm{op}\colon D \longrightarrow D$), such that

$$(ax + by)^t = (ax)^t + (by)^t = a^\tau x^t + b^\tau y^t\,,$$

for all $a, b \in D$ and $x, y \in V$.

Beginning with the pairing m we can now define a map $f\colon V \times V \longrightarrow D$ by

$$f(x, y) = m(x, y^{t\mathrm{op}})\,.$$

We have

$$\begin{aligned} f(x_1 + x_2, y) &= m(x_1 + x_2, y^{t\mathrm{op}}) \\ &= m(x_1, y^{t\mathrm{op}}) + m(x_2, y^{t\mathrm{op}}) \\ &= f(x_1, y) + f(x_2, y) \end{aligned}$$

and

$$\begin{aligned} f(x, y_1 + y_2) &= m(x, (y_1 + y_2)^{t\mathrm{op}}) \\ &= m(x, y_1^{t\mathrm{op}} + y_2^{t\mathrm{op}}) \\ &= m(x, y_1^{t\mathrm{op}}) + m(x, y_2^{t\mathrm{op}}) \\ &= f(x, y_1) + f(x, y_2)\,; \end{aligned}$$

so f is biadditive. Additionally

$$\begin{aligned} f(ax,by) &= m(ax,(by)^{t\mathrm{op}}) \\ &= m(ax,(b^\tau y^t)^{\mathrm{op}}) \\ &= m(ax,y^{t\mathrm{op}}b^{\tau\mathrm{op}}) \\ &= a\,m(x,y^{t\mathrm{op}})b^{\tau\mathrm{op}} \\ &= a\,f(x,y)b^\sigma\,. \end{aligned}$$

In general, a biadditive map $f\colon V\times V\longrightarrow D$ that satisfies

$$f(ax,by)=af(x,y)b^\sigma\,,$$

for a fixed anti-automorphism σ of D, all $a,b\in D$, and $x,y\in V$, is called a *σ-sesquilinear form.* In the special case where the anti-automorphism σ is the identity map, the division ring D must be a field and f is a *bilinear form.*

Above we have seen that every pairing of V with a space isomorphic to its opposite yields such a form. We will next see that every σ-sesquilinear form comes from a pairing of V and an appropriate opposite.

Given an anti-isomorphism $\sigma\colon D\longrightarrow D$ and a left D-space V, we provide the additive group $(V,+)$ with the structure of a right D-space by defining, for all $a,b\in D$ and $v\in V$,

$$v.a=a^{\sigma^{-1}}.v\quad\text{or, equivalently,}\quad b.v=v.b^\sigma\,.$$

That is, we recast the right D^{op}-space V^{op} as a right D-space using the isomorphism $\tau=\sigma\mathrm{op}$ of D with D^{op}. We use the notation V^σ to denote V when viewed as a right D-space via the anti-isomorphism σ. The subset U of V is a subspace of the left D-space V if and only if it is a subspace of the right D-space V^σ.

Given a σ-sesquilinear form $f\colon V\times V\longrightarrow D$ consider the map $m\colon V\times V^\sigma\longrightarrow D$, given by

$$m(x,y)=f(x,y)\,,$$

for all $x,y\in V$.[1] Since the addition in V^σ is identical to that in ${}_DV$, the function m is biadditive. We also have, for all $a,b\in D$ and $x,y\in V$,

$$\begin{aligned} m(ax,yb) &= f(ax,yb) \\ &= f(ax,b^{\sigma^{-1}}y) \\ &= a\,f(x,y)\,(b^{\sigma^{-1}})^\sigma \\ &= a\,m(x,y)\,b\,. \end{aligned}$$

[1] Strictly speaking, this definition is unnecessary. The underlying sets V and V^σ are identical, and the maps m and f on the cartesian square of this set are also identical. That is, $m=f$. But it is best to hold the distinction; see, for instance, page 141.

Therefore, as promised above, the sesquilinear form f can be viewed as arising from the pairing m of V with V^σ. The sesquilinear form f is *nondegenerate* provided that the associated pairing m is nondegenerate.

There are division rings D that are not isomorphic to their opposites, but as long as we have this necessary condition, every D-space admits nondegenerate σ-sesquilinear forms. The following comes from easy calculation.

(10.1). Lemma. *Let σ be an anti-automorphism of the division ring D.*

(a) *Let $\mathcal{X} = \{ x_i \mid i \in I \}$ be a basis of V and $g_{i,j} \in D$ for $i, j \in I$. Then*

$$g\Big(\sum_{i\in I} a_i x_i, \sum_{j \in I} b_j x_j\Big) = \sum_{i,j\in I} a_i g_{i,j} b_j^\sigma$$

defines a σ-sesquilinear form on V.

(b) *The set $\mathrm{Sesq}^\sigma(V)$ of σ-sesquilinear forms on the D-space V is an abelian group under pointwise addition: $(f+g)(x,y) = f(x,y) + g(x,y)$.* □

The matrix of coefficients, $G = (g_{i,j})_{i,j}$, is the *Gram matrix* of the form g, and the process of passing from the values on the basis elements to the values on arbitrary elements, as in Lemma (10.1)(a), is called *sesquilinearization* (or just *linearization* if $\sigma = 1$).

The first part of the lemma actually demonstrates the second part concretely by showing $\mathrm{Sesq}^\sigma(V)$ to be isomorphic as abelian group to $\mathrm{Mat}_I(D)$; if F is the Gram matrix for f and G is the Gram matrix for g, then $F+G$ is the Gram matrix for $f+g$.

When we write the vectors $\sum_{i\in I} a_i x_i$ and $\sum_{j\in I} b_j x_j$ in the basis $\mathcal{X}$ as I-tuples $a = (\dots, a_i, \dots)$ and $b = (\dots, b_j, \dots)$, then we have the matrix representation of the form:

$$g(a,b) = a\,G\,b^{\sigma\top}\,.$$

Of course the Gram matrix G of g is actually $G_\mathcal{X}$, since it depends upon the choice of basis $\mathcal{X}$. If $\mathcal{Y}$ is a second basis and A is the $I \times I$ base change matrix that takes vectors written in the basis $\mathcal{Y}$ to their corresponding representation in the basis $\mathcal{X}$, then $G_\mathcal{Y} = AG_\mathcal{X}A^{\sigma\top}$.

(10.2). Lemma. *If $f\colon V \times V \longrightarrow D$ is a σ-sesquilinear form, then $\lambda^f : v \mapsto f(v,\cdot)$ is a D-homomorphisms from V to $(V^\sigma)^*$ and $\rho^f : w \mapsto f(\cdot, w)$ is D-homomorphism from V to $(V^*)^{\sigma^{-1}}$.* □

Proof. We apply Lemma (9.1) to the pairing $m\colon V \times V^\sigma$ given by $m(v,w) = f(v,w)$ and find that D-linear λ^f goes from V to $(V^\sigma)^*$ while ρ goes from V^σ to V^*. But then ρ^f can also be thought of as going from V to $(V^*)^{\sigma^{-1}}$. □

The σ-sesquilinear form f is *nondegenerate* if both ρ^f and λ^f are injective.

(10.3). Corollary. *If f is a σ-sesquilinear form on finite dimensional D-space V, then the following are equivalent:*

(1) *f is nondegenerate;*

(2) *λ^f is injective;*

(3) *λ^f is surjective;*

(4) *ρ^f is injective;*

(5) *ρ^f is surjective;*

(6) *λ^f is an isomorphism;*

(7) *ρ^f is an isomorphism.*

PROOF. V and V^σ have the same dimension. When finite, this is also equal to the dimensions of V^* and $(V^\sigma)^*$ by Corollary (7.2). The result then follows from Lemma (9.7). □

(10.4). COROLLARY. *The σ-sesquilinear form f on the finite dimensional D-space V is nondegenerate if and only if its Gram matrix is invertible if and only if its Gram matrix has nonzero Dieudonné determinant.*

PROOF. If F is the Gram matrix for f, then the vector v is in ${}^{\perp}V$ if and only if $vF = 0$ (in matrix representation). Similarly $v \in V^{\perp}$ if and only if $Fv^{\sigma\top} = 0$. Therefore the corollary follows from the previous corollary, Theorem (6.10), and Lemma (6.15). □

(10.5). LEMMA.

(a) *If $g\colon V \times V \longrightarrow D$ is a σ-sesquilinear form and $0 \neq c \in D$, then $f(x, y) = g(x, y)c$ is a σc-sesquilinear form.*

(b) *Let f and g be nondegenerate sesquilinear forms on V with $\dim_D(V) \geq 2$. The following are equivalent:*

  (i) *for all $x, y \in V$ we have*

$$f(x, y) = 0 \iff g(x, y) = 0\,;$$

  (ii) *there is a constant $0 \neq c \in D$ with $f(x, y) = g(x, y)c$ for all $x, y \in V$.*

PROOF. (a) Clearly f is biadditive. Also

$$f(ax, by) = g(ax, by)c = ag(x, y)b^\sigma c = a(g(x, y)c)c^{-1}b^\sigma c = af(x, y)b^{\sigma c}\,.$$

(b) Certainly (ii) implies (i). Now assume that $f(x, y) = 0$ if and only if $g(x, y) = 0$.

Let f_y be the linear functional from V to D given by $x.f_y = f(x, y)$ and similarly let g_y be the linear functional given by $x.g_y = g(x, y)$. Hypothesis (i) then says that, for all $y \in V$, the functionals f_y and g_y have the same kernel. That is, there is a nonzero constant $c_y \in D$ with $f_y = g_y c_y$. It remains to show that c_y is independent of y.

We have

$$\begin{aligned} & & f_{x+y} - f_x - f_y &= 0\,, \\ &\text{so} & g_{x+y}c_{x+y} - g_x c_x - g_y c_y &= 0\,, \\ &\text{and} & g_{x+y} - g_x - g_y &= 0\,, \\ &\text{so} & g_{x+y}c_{x+y} - g_x c_{x+y} - g_y c_{x+y} &= 0\,, \end{aligned}$$

hence

$$g_x(c_{x+y} - c_x) - g_y(c_{x+y} - c_y) = 0\,.$$

For linearly independent x and y, the functionals g_x and g_y are linearly independent by nondegeneracy, so $c_x = c_{x+y} = c_y$. If x and y are dependent, then by hypothesis there is a z independent of both, hence $c_x = c_z = c_y$. □

If g is a σ-sesquilinear form, then the σc-sesquilinear form $f = gc$ is said to be *proportional* to the form g.

10.2 Hermitian forms

The σ-sesquilinear form $f\colon V \times V \longrightarrow D$ of the first section comes from a duality automorphism of the pairing graph for m, whose bipartition $\mathbb{P}_D V \cup \mathbb{P} W_D$, we now view as $\mathbb{P}V \cup \mathbb{P}V^\sigma$. Although such automorphisms switch the two parts of the graph, they may not actually have order 2.

When the duality automorphism does have order 2, then we say it is a *polarity automorphism* and the geometric map exchanging the 1-spaces of $\mathbb{P}V$ and those of $\mathbb{P}V^\sigma \leq \mathbb{P}V^*$ is called a *polarity.* Once we have identified $\mathbb{P}V$ and $\mathbb{P}V^\sigma$ via a polarity automorphism we have the nice property

$$m(u,v) = 0 \quad \text{if and only if} \quad m(v,u) = 0\,.$$

That is, for every subset of S of V we have ${}^\perp S = S^\perp$, so we may dispense with this particular right-left distinction.

In the language of sesquilinear forms, the form $f\colon V \times V \longrightarrow D$ is *reflexive* if

$$f(u,v) = 0 \quad \text{if and only if} \quad f(v,u) = 0\,,$$

for all $u, v \in V$. We write $S^\perp$ for the subspace $\{\, v \in V \mid f(v,s) = 0\,,\ \text{for all } s \in S\,\}$ and say that V and f are *nondegenerate* provided its *radical* $\mathrm{Rad}(V,f) = \mathrm{Rad}(V) = V^\perp$ is equal to $\{0\}$. The reflexive form f restricts to a reflexive form on each subspace U of V, and U is a *nondegenerate subspace* provided its radical under this restriction is 0; that is, $U \cap U^\perp = 0$.

These definitions are entirely consistent with their previous use for pairings. When we specialize the earlier Lemma (9.3) to the present situation we find:

(10.6). Lemma. *Let h be a nondegenerate reflexive σ-sesquilinear form on V, and let U be a finite dimensional subspace of V.*

(a) *The codimension of $U^\perp$ in V is equal to the dimension of U, and $U^{\perp\perp} = U$.*

(b) *The restriction of h to U is nondegenerate if and only if $V = U \oplus U^\perp$.* □

There is also a current version of Lemma (9.4).

(10.7). LEMMA. *Let h be a nondegenerate reflexive σ-sesquilinear form on V. If S is a finite dimensional subspace of V, there is a subspace T of dimension at most $2\dim_D(S)$ that contains S and has the restriction of h to T nondegenerate.*

PROOF. We follow the proof of Lemma (9.4). For $S\,(= U_0 = Y_0)$ let $X\,(= U_1 = Y_1)$ be a complement in V to $(S \cap S^\perp)^\perp$. Then $T = S \oplus X\,(= U_0 \oplus U_1 = Y_0 \oplus Y_1)$ has the desired properties. □

(10.8). LEMMA. *Let h be a nondegenerate reflexive σ-sesquilinear form on V with $\dim_D(V) \geq 2$. Then there is a $0 \neq k \in D$ with*

$$h(x,y) = kh(y,x)^\sigma$$

and

$$k^\sigma = k^{-1} \quad \textit{and} \quad a^{\sigma^2} = k^{-1}ak \textit{ for all } a \in D\,.$$

PROOF. Let $g\colon V \times V \longrightarrow D$ be given by $g(x,y) = h(y,x)^{\sigma^{-1}}$. Then g is biadditive and

$$\begin{aligned} g(ax,by) = h(by,ax)^{\sigma^{-1}} &= (bh(y,x)a^\sigma)^{\sigma^{-1}} \\ &= ah(y,x)^{\sigma^{-1}}b^{\sigma^{-1}} = ag(x,y)b^{\sigma^{-1}}\,, \end{aligned}$$

so g is σ^{-1}-sesquilinear.

As h is reflexive, $h(x,y) = 0$ if and only if $g(x,y) = 0$. Therefore by Lemma (10.5) there is a nonzero $c \in D$ with $h(x,y) = g(x,y)c = f(y,x)^{\sigma^{-1}}c$ hence $kh(x,y)^\sigma = h(y,x)$ for $k = (c^{-1})^\sigma$. Next

$$h(x,y) = kh(y,x)^\sigma = k(kh(x,y)^\sigma)^\sigma = kh(x,y)^{\sigma^2}k^\sigma\,.$$

As h is nondegenerate, there are x, y with $h(x,y) = 1$, which leads to $1 = kk^\sigma$ and $k^\sigma = k^{-1}$. This in turn yields $h(ax,y) = kh(ax,y)^{\sigma^2}k^{-1}$ or $k^{-1}ak = a^{\sigma^2}$, so the automorphism σ^2 of D is conjugation by k. □

Sesquilinear forms in dimension 1 are always reflexive. A σ-sesquilinear form h with

$$h(x,y) = kh(y,x)^\sigma\,,$$

as in the lemma, is clearly reflexive. Such a form is called (σ,k)-*hermitian* (and even a (σ,k)-*form* at times). We call the pair (V,h) a *hermitian space*.[2]

When $k = 1$ the form h is said to be σ-*hermitian* and when $k = -1$ it is σ-*skew-hermitian*. As σ^2 is conjugation by k, in the σ-hermitian and σ-skew-hermitian cases we must have $\sigma^2 = 1$. The hermitian and skew-hermitian

[2] Of course, this is imprecise. The data that goes into such a space is (D,σ,k,V,h), but we rarely record this so specifically. Instead we rely on the context for clarity.

terminology is usually reserved for the case $\sigma \neq 1$, the cases $(\sigma, k) = (1, 1)$ and $(\sigma, k) = (1, -1)$ then giving, respectively, *symmetric bilinear forms* and *alternating bilinear forms.* This is problematic in characteristic 2 where $1 = -1$, so we shall avoid the alternating terminology. Instead we call any bilinear form h with $h(x, x) = 0$, for all x, a *symplectic form* or *null form.* If the characteristic is not 2, this is equivalent to h being alternating.

(10.9). PROPOSITION. *Let σ be an anti-automorphism of D such that the automorphism σ^2 is conjugation by the element k of D with $k^\sigma = k^{-1}$. For each form $f \in \mathrm{Sesq}^\sigma(V)$, consider $f^\pi \colon V \times V \longrightarrow D$ given by*

$$f^\pi(x, y) = kf(y, x)^\sigma \,.$$

(a) *The map $\pi \colon f \mapsto f^\pi$ is an automorphism of the abelian group $\mathrm{Sesq}^\sigma(V)$ with $\pi^2 = 1$.*

(b) *In $\mathrm{End}_{\mathsf{AbGrp}}(\mathrm{Sesq}^\sigma(V))$ we have $\mathrm{im}(1 + \epsilon\pi) \leq \ker(1 - \epsilon\pi)$ for $\epsilon = \pm 1$.*

PROOF. We have already seen in Lemma (10.1) that $\mathrm{Sesq}^\sigma(V)$ is an abelian group. If f is one of the forms belonging to it, then f^π is biadditive and

$$\begin{aligned} f^\pi(ax, by) &= kf(by, ax)^\sigma = k(bf(y, x)a^\sigma)^\sigma \\ &= ka^{\sigma^2} f(y, x)^\sigma b^\sigma = kk^{-1}akf(y, x)^\sigma b^\sigma \\ &= af^\pi(x, y)b^\sigma \,. \end{aligned}$$

Thus π is a map from $\mathrm{Sesq}^\sigma(V)$ to itself and is certainly additive. Furthermore

$$\begin{aligned} f^{\pi^2}(x, y) &= k(kf(x, y)^\sigma)^\sigma = kf(x, y)^{\sigma^2} k^\sigma \\ &= kk^{-1}f(x, y)kk^\sigma = f(x, y) \,, \end{aligned}$$

so π is an automorphism of order 2, completing (a). As an endomorphism $0 = 1 - \pi^2 = (1 - \pi)(1 + \pi)$, so (b) follows directly. □

With the notation of the proposition, we set

$$\mathrm{Herm}^{(\sigma,\epsilon k)}(V) = \ker(1 - \epsilon\pi) = \{\, f \in \mathrm{Sesq}^\sigma(V) \mid f(x, y) = \epsilon kf(y, x)^\sigma \,\} \,,$$

the group of $(\sigma, \epsilon k)$-*hermitian* forms on V, and

$$\begin{aligned} &\mathrm{THerm}^{(\sigma,\epsilon k)}(V) = \mathrm{im}(1 + \epsilon\pi) \\ &\quad = \{\, h \in \mathrm{Sesq}^\sigma(V) \mid h(x, y) = f(x, y) + \epsilon kf(y, x)^\sigma \,,\ \text{some } f \in \mathrm{Sesq}^\sigma(V) \,\} \,, \end{aligned}$$

the group of $(\sigma, \epsilon k)$-*trace-hermitian* forms on V.

Consider the special case $V = D$. Here $\mathrm{Sesq}^\sigma(D)$ is naturally isomorphic to $(D, +)$, since the Gram matrix for each form has degree 1. We define subgroups of $(D, +)$:

$$D_{(\sigma,\epsilon k)} = \mathrm{THerm}^{(\sigma,\epsilon k)}(D) = \{\, a + \epsilon ka^\sigma \mid a \in D \,\} \,,$$

the $(\sigma, \epsilon k)$-*traces* in D, and

$$D^{(\sigma,\epsilon k)} = \mathrm{Herm}^{(\sigma,\epsilon k)}(D) = \{\, a \in D \mid a = \epsilon k a^\sigma \,\}\,,$$

the $(\sigma, \epsilon k)$-*symmetric elements* in D.

(10.10). Lemma. *If F is a field of characteristic 2 and σ is an automorphism of F of order 2, then there is an element t of F with $t + t^\sigma = 1$.*

Proof. For $s \neq s^\sigma$, set $t = s(s + s^\sigma)^{-1}$. □

(10.11). Proposition. *Let σ be an anti-automorphism of D such that the automorphism σ^2 is conjugation by the element k of D with $k^\sigma = k^{-1}$.*

(a) $\mathrm{THerm}^{(\sigma,\epsilon k)}(V) \le \mathrm{Herm}^{(\sigma,\epsilon k)}(V)$ *and especially* $D_{(\sigma,\epsilon k)} \le D^{(\sigma,\epsilon k)}$.

(b) *If* $\mathrm{char}(D) \neq 2$ *then*

$$\begin{aligned}\mathrm{Sesq}^\sigma(D) &= \mathrm{THerm}^{(\sigma,k)}(V) \oplus \mathrm{THerm}^{(\sigma,-k)}(V)\\ &= \mathrm{Herm}^{(\sigma,k)}(V) \oplus \mathrm{Herm}^{(\sigma,-k)}(V)\,.\end{aligned}$$

(c) *If* $\mathrm{char}(D) \neq 2$ *or* $\sigma|_{\mathrm{Z}(D)} \neq 1$ *then* $\mathrm{THerm}^{(\sigma,\epsilon k)}(V) = \mathrm{Herm}^{(\sigma,\epsilon k)}(V)$ *and especially* $D_{(\sigma,\epsilon k)} = D^{(\sigma,\epsilon k)}$.

Proof. (a) This is just the second part of the previous result, rewritten in the appropriate notation.

(b) For $f \in \mathrm{Sesq}^\sigma(V)$ we have

$$2f = f(1+\pi) + f(1-\pi) \in \mathrm{THerm}^{(\sigma,k)}(V) + \mathrm{THerm}^{(\sigma,-k)}(V)\,,$$

and also for $h \in \mathrm{Herm}^{(\sigma,k)}(V) \cap \mathrm{Herm}^{(\sigma,-k)}(V)$ we have

$$2h = h + h = kh^\sigma + (-k)h^\sigma = 0\,.$$

(c) In characteristic other than 2 this follows from the previous part, but we can handle both cases at once. We have already noted $\mathrm{THerm}^{(\sigma,\epsilon k)}(V) \le \mathrm{Herm}^{(\sigma,\epsilon k)}(V)$, so for arbitrary $h \in \mathrm{Herm}^{(\sigma,\epsilon k)}(V)$ we must find a $g \in \mathrm{Sesq}^\sigma(V)$ with $h(x,y) = g(x,y) + kg(y,x)^\sigma$.

Choose a t in the subfield $\mathrm{Z}(D)$ with $t + t^\sigma = 1$. If D has characteristic other than 2, then $t = 2^{-1}$ has this property. If D has characteristic 2, such a t is guaranteed by the previous lemma. Let $g = ht$. Then

$$\begin{aligned}g(x,y) + \epsilon k g(y,x)^\sigma &= h(x,y)t + \epsilon k(h(y,x)t)^\sigma\\ &= h(x,y)t + \epsilon k t^\sigma h(y,x)^\sigma\\ &= h(x,y)t + \epsilon k h(y,x)^\sigma t^\sigma\\ &= h(x,y)t + h(x,y)t^\sigma\\ &= h(x,y)\end{aligned}$$

as desired. □

A $(\sigma, \epsilon k)$-hermitian form h is *trace-valued* if, for all $x \in V$, we have $h(x,x) \in D_{(\sigma,\epsilon k)}$. Certainly all trace-hermitian forms are trace-valued. If the characteristic is not 2, then the converse is true for the trivial reason that all (σ, k)-hermitian forms are trace-hermitian forms, as we saw in Proposition (10.11)(c). But the converse is valid without restriction.

(10.12). THEOREM. *Every trace-valued $(\sigma, \epsilon k)$-hermitian form h on V is a $(\sigma, \epsilon k)$-trace-hermitian form.*

Indeed, for $\{ x_i \mid i \in I \}$ a D-basis of V indexed by the well-ordered set $(I, <)$, let f be the σ-sesquilinear form with Gram matrix given by:

$$\begin{aligned} f(x_i, x_j) &= h(x_i, x_j) && \textit{for } i < j\,; \\ f(x_i, x_i) &= q_i && \textit{for } i = j\,; \\ f(x_i, x_j) &= 0 && \textit{for } j < i\,, \end{aligned}$$

where each element $h(x_i, x_i)$ is the $(\sigma, \epsilon k)$-trace $q_i + \epsilon k q_i^\sigma$. Then the set of all σ-sesquilinear forms mapped to h by the endomorphism $1 + \epsilon\pi$ of Proposition (10.9) is $f + \mathrm{Herm}^{(\sigma,-\epsilon k)}(V)$.

PROOF. If F is the given Gram matrix for f, then by construction $F + \epsilon k (F^\sigma)^\top$ is the Gram matrix for h; so h is trace-hermitian, being $f(1 + \epsilon\pi)$. By definition the kernel of $1 + \epsilon\pi$ is the subspace $\mathrm{Herm}^{(\sigma,-\epsilon k)}(V)$. □

From the next section on, we shall primarily be concerned with trace-valued, hence trace-hermitian, forms. This is because all forms that are not trace-valued give rise rigidly to trace-valued forms.

(10.13). PROPOSITION. *Let h be a (σ, k)-hermitian form on the vector space V over the division ring D of characteristic 2. Let V_0 be the set of all vectors $x \in V$ with $h(x,x) \in D_{(\sigma,k)}$. Then V_0 is the unique maximal trace-valued subspace of V, and the value map $v \longrightarrow h(v,v)$ induces an additive injection of the elementary abelian 2-group V/V_0 into $D^{(\sigma,k)}/D_{(\sigma,k)}$.*

PROOF. Always $h(x,x) \in D^{(\sigma,k)}$. If $x \in V_0$ so that $h(x,x) = c \in D_{(\sigma,k)}$, then $c = a + ka^\sigma$, for some a. For all $b \in D$

$$\begin{aligned} h(bx, bx) &= bh(x,x)b^\sigma = bcb^\sigma = b(a + ka^\sigma)b^\sigma \\ &= bab^\sigma + bka^\sigma b^\sigma = bab^\sigma + kk^{-1}bka^\sigma b^\sigma \\ &= bab^\sigma + kb^{\sigma^2}a^\sigma b^\sigma = bab^\sigma + k(bab^\sigma)^\sigma \in D_{(\sigma,k)}\,. \end{aligned}$$

Also

$$\begin{aligned} h(x+y, x+y) =& h(x,x) + h(y,y) + h(x,y) + h(y,x) \\ =& h(x,x) + h(y,y) + (h(x,y) + kh(x,y)^\sigma) \\ \in& \ h(x,x) + h(y,y) + D_{(\sigma,k)}\,. \qquad \square \end{aligned}$$

The form $f = gc$ is said to be *proportional* to g. Clearly g is reflexive if and only if the proportional form gc is. The algebraic version of this is recorded in the next lemma.

(10.14). LEMMA. *Let σ be an anti-automorphism of D such that the automorphism σ^2 is conjugation by the element k of D with $k^\sigma = k^{-1}$. Further let $0 \neq c \in D$.*

(a) *Set $k' = k(c^\sigma)^{-1}c$. Then $ka^\sigma c = k'(ac)^{\sigma c}$ for all $a \in D$.*

(b) $\mathrm{Herm}^{(\sigma,k)}(V)c = \mathrm{Herm}^{(\sigma c,k')}(V)$*, and especially* $D^{(\sigma,k)}c = D^{(\sigma c,k')}$.

(c) $\mathrm{THerm}^{(\sigma,k)}(V)c = \mathrm{THerm}^{(\sigma c,k')}(V)$*, and especially* $D_{(\sigma,k)}c = D_{(\sigma c,k')}$.

PROOF. (a) Indeed

$$\begin{aligned} ka^\sigma c &= k((c^\sigma)^{-1}cc^{-1}(c^\sigma))a^\sigma c \\ &= (k(c^\sigma)^{-1}c)(c^{-1}(ac)^\sigma c) \\ &= (k(c^\sigma)^{-1}c)(ac)^{\sigma c} \\ &= k'(ac)^{\sigma c}\,. \end{aligned}$$

(b) For $h \in \mathrm{Herm}^{(\sigma,k)}(V)$ the form hc is σc-sesquilinear by Lemma (10.5). As h is (σ,k)-hermitian,

$$h(x,y)c = kh(y,x)^\sigma c = k'(h(y,x)c)^{\sigma c}$$

by (a). Thus multiplication by c gives a map from $\mathrm{Herm}^{(\sigma,k)}(V)$ to $\mathrm{Herm}^{(\sigma c,k')}(V)$, and its inverse is multiplication by c^{-1}.

(c) Similarly, if $f(x,y) + kf(y,x)^\sigma \in \mathrm{THerm}^{(\sigma,k)}(V)$, for σ-sesquilinear f, then fc is σc-sesquilinear (again by Lemma (10.5)) and

$$\begin{aligned} (f(x,y) + kf(y,x)^\sigma)c &= f(x,y)c + kf(y,x)^\sigma c \\ &= f(x,y)c + k'(f(y,x)c)^{\sigma c} \in \mathrm{THerm}^{(\sigma c,k')}(V)\,. \end{aligned} \qquad \square$$

(10.15). THEOREM. (BESTIARY OF HERMITIAN FORMS) *Let h be a (σ,k)-hermitian form on the D-space V with h not identically 0. Set $D_h = \{\, h(x,x) \mid x \in V\,\}$. Then, up to proportionality, we have one of:*

(1) SYMPLECTIC CASE: *$\sigma = 1$; $k = -1$; D is a field; $D_h = \{0\} = D_{(\sigma,k)}$; and h is a symplectic bilinear form.*

(2) ORTHOGONAL CASE: *$\sigma = 1$; $k = 1$; D is a field; $D_h \neq \{0\}$; $D_{(\sigma,k)} = 2D$; and h is a symmetric bilinear form.*

(3) UNITARY CASE: *$\sigma^2 = 1 \neq \sigma$; $D_h \neq \{0\} \neq D_{(\sigma,k)}$; k can be taken to be either 1 or -1; and h is a hermitian or skew-hermitian form (both possible, subject to proportionality).*

Proof.

(a) *Case $D_{(\sigma,k)} = \{0\}$.*

We have $1+k\,1^\sigma = 0$, so $k = -1$. Thus $a+(-1)a^\sigma = 0$ for all a, and $a = a^\sigma$. That is, $\sigma = 1$ and, as it is an anti-automorphism of D, D must be a field. Furthermore $D^{(\sigma,k)} = \{\, a \mid a = ka^\sigma = -a \,\}$, which is 0 unless $\mathrm{char}(D) = 2$ where it is all of D.

(b) *Case $D_h = \{0\}$.*

As h is not identically 0 there are $x, y \in V$ with $h(x, y) = b \neq 0$. For all $a \in D$,

$$\begin{aligned} a + ka^\sigma &= h(ab^{-1}x, y) + kh(ab^{-1}x, y)^\sigma \\ &= h(ab^{-1}x, y) + h(y, ab^{-1}x) \\ &= h(ab^{-1}x + y, ab^{-1}x + y) - h(ab^{-1}x, ab^{-1}x) - h(y, y) \\ &= 0 - 0 - 0 = 0\,. \end{aligned}$$

That is, $D_{(\sigma,k)} = 0$. By (i) we are in the Symplectic Case (1).

(c) *Case $D_h \neq \{0\}$ but $D_{(\sigma,k)} = \{0\}$*

As $D^{(\sigma,k)} \supseteq D_h \neq \{0\}$ we have $D_{(\sigma,k)} = \{0\} \neq D^{(\sigma,k)}$. As we saw under (i), this can only happen in characteristic 2. Thus, again by (i), $\sigma = 1$, D must be a field, and $k = -1 = 1$. We are in the Orthogonal Case (2).

(d) *Case $D_h \neq \{0\} \neq D_{(\sigma,k)}$.*

The arguments to this point imply that a nonzero (σ, k)-hermitian form with $\sigma \neq 1$ can only occur under this case.

Choose $a \in D$ with $0 \neq a + ka^\sigma \in D_{(\sigma,k)}$. Set $s = (a + ka^\sigma)^{-1}$. Then by Lemma (10.14) the form $g = hs$ is $(\sigma s, k')$-hermitian for $k' = k(s^\sigma)^{-1}s$. But

$$\begin{aligned} k(s^\sigma)^{-1}s &= k(((a + ka^\sigma)^{-1})^\sigma)^{-1}(a + ka^\sigma)^{-1} \\ &= k((a + ka^\sigma))^\sigma)(a + ka^\sigma)^{-1} \\ &= k(a^\sigma + ka^{\sigma^2}k^\sigma)(a + ka^\sigma)^{-1} \\ &= (ka^\sigma + kk^{-1}akk^\sigma)(a + ka^\sigma)^{-1} = 1\,. \end{aligned}$$

Therefore g is $(\sigma s, 1)$-hermitian. As $(\sigma s)^2$ is conjugation by 1, we have $(\sigma s)^2 = 1$. Also $D_g = (D_h)s \neq \{0\}$.

(i) $\sigma s = 1$.

As the anti-automorphism σs is 1, again D is a field. Also $D_{(\sigma s,1)} = \{\, a + 1a^{\sigma s} \mid a \in D \,\} = 2D$. After replacing h with g and σ with σs, we are again in the Orthogonal Case (2).

(ii) $\sigma s \neq 1$.

The remark at the beginning of Case (iv) tells us that the $(\sigma s, 1)$-hermitian form g also comes under this case. Replacing h with g and σ with σs, we are in the Unitary Case (3) for $k = 1$.

As $\sigma s \neq 1$ there is an $a \in D$ with $a \neq a^{\sigma s}$, hence $t = a - a^{\sigma s} \neq 0$. Then the form $gt = hst$ is (by Lemma (10.14) again) $(\sigma st, k'')$-hermitian with

$$\begin{aligned} k'' &= k'(t^{\sigma s})^{-1}t \\ &= 1((a - a^{\sigma s})^{\sigma s})^{-1}t \\ &= (a^{\sigma s} - a^{(\sigma s)^2})^{-1}t = -t^{-1}t = -1\,. \end{aligned}$$

The only inner anti-automorphism of D is 1. As $\sigma s \neq 1$, it must be a noninner anti-automorphism. In particular $\sigma s \neq t^{-1}$ and $\sigma st \neq 1$. Thus after replacing h with gt and σ with σst, we remain under Case (iv) and so are again in the Unitary Case (3) for $k = -1$. □

10.3 Pseudoquadratic forms

Throughout this section σ will be an anti-automorphism of the division ring D for which the automorphism σ^2 is conjugation by the element k of D with $k^\sigma = k^{-1}$. It is also true that σ^2 is conjugation by $-k$ and that $(-k)^\sigma = (-k)^{-1}$; so, unlike the previous section, we will write k rather than ϵk with $\epsilon = \pm 1$. It is, however, worth remembering that everything to follow would remain valid and consistent with the roles of k and $-k$ reversed.

Proposition (10.11) tells us that much of the time we have

$$\mathrm{Sesq}^\sigma(V)/\,\mathrm{Herm}^{(\sigma,-k)}(V) \simeq \mathrm{Herm}^{(\sigma,k)}(V)\,.$$

Pseudoquadratic forms, as introduced by Tits, provide a more general version of this. A second motivation comes from a direct calculation:

(10.16). LEMMA. *Let $f\colon V \times V \longrightarrow D$ be a (σ, k)-hermitian form. Define the value map $v\colon V \longrightarrow D$ by $v(x) = f(x, x)$. Then we have, for all $x, y \in V$, $a \in D$:*

(a) $v(x + y) - v(x) - v(y) = f(x, y) + kf(x, y)^\sigma$ *;*

(b) $v(ax) = av(x)a^\sigma$ *;*

(c) $v(x) = kv(x)^\sigma$ *.* □

Let Λ be an additive subgroup of $(D, +)$ satisfying:

(i) $D_{(\sigma,-k)} \leq \Lambda \leq D^{(\sigma,-k)}$;

(ii) $a\Lambda a^\sigma = \Lambda$ for all $0 \neq a \in D$.

Such a Λ is called a (σ, k)-*form parameter* or *form parameter* in D. We already have several examples.

(10.17). Lemma.

(a) $D_{(\sigma,-k)}$ *is a* (σ,k)*-form parameter.*

(b) $D^{(\sigma,-k)}$ *is a* (σ,k)*-form parameter.*

Proof. Both

$$D_{(\sigma,-k)} = \{\, a - ka^\sigma \mid a \in D\,\}$$

and

$$D^{(\sigma,-k)} = \{\, a \in D \mid a = -ka^\sigma\,\}$$

are subgroups of $(D,+)$. By Proposition (10.13) $bD_{(\sigma,-k)}b^\sigma = D_{(\sigma,-k)}$ for all $b \in D$. Also when $a = -ka^\sigma$,

$$bab^\sigma = b(-ka^\sigma)b^\sigma = -k(k^{-1}bk)a^\sigma b^\sigma = -kb^{\sigma^2}a^\sigma b^\sigma = -k(bab^\sigma)^\sigma\,. \qquad \square$$

(10.18). Lemma. *Let* Λ *be a* (σ,k)*-form parameter. Then under the scalar action*

$$a \cdot (x + D_{(\sigma,-k)}) = axa^\sigma + D_{(\sigma,-k)}\,,\ \textit{for } a \in D\,,$$

the quotient group $\Lambda/D_{(\sigma,-k)}$ *is a* D*-space.*

Proof. The quotient is an abelian group, but we must show that it is a unital D-module under the given action. Certainly

$$1 \cdot (x + D_{(\sigma,-k)}) = x + D_{(\sigma,-k)}$$

and

$$\begin{aligned}(ab) \cdot (x + D_{(\sigma,-k)}) &= (ab)x(ab)^\sigma + D_{(\sigma,-k)}\\ &= a(bxb^\sigma)a^\sigma + D_{(\sigma,-k)}\\ &= a \cdot (b \cdot (x + D_{(\sigma,-k)}))\,.\end{aligned}$$

Finally

$$\begin{aligned}(a+b) \cdot (x + D_{(\sigma,-k)}) &= (a+b)x(a+b)^\sigma + D_{(\sigma,-k)}\\ &= axa^\sigma + bxb^\sigma + axb^\sigma + bxa^\sigma + D_{(\sigma,-k)}\\ &= a \cdot (x + D_{(\sigma,-k)})) + b \cdot (x + D_{(\sigma,-k)}))\,,\end{aligned}$$

since, as $x \in \Lambda \le D^{(\sigma,-k)}$,

$$bxa^\sigma = -bkx^\sigma a^\sigma = -kk^{-1}bkx^\sigma a^\sigma = -kb^{\sigma^2}x^\sigma a^\sigma = -k(axb^\sigma)^\sigma\,. \qquad \square$$

Let Λ be a (σ,k)-form parameter in D. The pair (q,h) is a Λ-*pseudoquadratic form* on V, provided q is a map from V to the abelian group D/Λ and $h\colon V \times V \longrightarrow D$ is a (σ,k)-hermitian form on V that together satisfy:

(i) $q(x+y) - q(x) - q(y) = h(x,y) + \Lambda$;

(ii) $q(ax) = aq(x)a^\sigma + \Lambda$;

(iii) $h(x,x) = q(x) + kq(x)^\sigma$;

for all $x, y \in V$ and $a \in D$. The triple (V, q, h) is then a *pseudoquadratic space*.[3] The map q is the associated *quadratic part*, just as the hermitian form h is the associated *hermitian part*. The set of all Λ-pseudoquadratic forms on V is $\mathrm{PQuad}_\Lambda^{(\sigma,k)}(V)$, again an abelian group under pointwise addition.

As we have seen in Proposition (10.11)(b), we often have $D = D_{(\sigma,k)} \oplus D_{(\sigma,-k)}$. Therefore the quotients $D/D_{(\sigma,-k)}$ and D/Λ provide generalizations of the group of (σ,k)-traces $D_{(\sigma,k)}$.

The condition (iii) contains an abuse of notation, since $h(x,x) \in D$ while $q(x) \in D/\Lambda$. Let q_0 be a coset representative for $q(x)$. Then, for all $\lambda \in \Lambda$,

$$(q_0 + \lambda) + k(q_0 + \lambda)^\sigma = (q_0 + kq_0^\sigma) + (\lambda + k\lambda^\sigma) = (q_0 + kq_0^\sigma) + 0 = q_0 + kq_0^\sigma\,,$$

as $\Lambda \le D^{(\sigma,-k)}$. Thus the righthand side of (iii) gives a well-defined element of D, and the abuse is small.

We write $\mathrm{Herm}_\Lambda^{(\sigma,k)}(V)$ for the subgroup of all (σ,k)-hermitian forms with $h(x,x) \in \Lambda$, for all $x \in V$, so that

$$\mathrm{Herm}_{D_{(\sigma,k)}}^{(\sigma,k)}(V) \le \mathrm{Herm}_\Lambda^{(\sigma,k)}(V) \le \mathrm{Herm}_{D^{(\sigma,k)}}^{(\sigma,k)}(V) = \mathrm{Herm}^{(\sigma,k)}(V)\,,$$

the final equality being a consequence of Lemma (10.16)(c).

(10.19). Theorem. *Let Λ be a (σ,k)-form parameter on D. The map $f \mapsto (q,h)$ given by*

$$q(x) = f(x,x) + \Lambda \quad \textit{and} \quad h(x,y) = f(x,y) + kf(y,x)^\sigma$$

gives an isomorphism

$$\mathrm{Sesq}^\sigma(V)/\,\mathrm{Herm}_\Lambda^{(\sigma,-k)}(V) \simeq \mathrm{PQuad}_\Lambda^{(\sigma,k)}(V)\,.$$

Proof. The map as described is additive, and $h \in \mathrm{Herm}^{(\sigma,k)}(V)$ by Proposition (10.11)(a). The things that must be verified are:
(a) The image is in $\mathrm{PQuad}_\Lambda^{(\sigma,k)}(V)$.
(b) The kernel is $\mathrm{Herm}_\Lambda^{(\sigma,-k)}(V)$.
(c) The map is surjective.

(a) *Image.* We must verify that the given q and h satisfy the requirements (i), (ii), and (iii) for a pseudoquadratic form.

(i) For $x, y \in V$

$$\begin{aligned} q(x+y) - q(x) - q(y) &= f(x+y, x+y) - f(x,x) - f(y,y) + \Lambda \\ &= f(x,y) + f(y,x) + \Lambda \\ &= f(x,y) + f(y,x) + (kf(y,x)^\sigma - kf(y,x)^\sigma) + \Lambda \\ &= f(x,y) + kf(y,x)^\sigma + (f(y,x) - kf(y,x)^\sigma) + \Lambda \\ &= f(x,y) + kf(y,x)^\sigma + \Lambda \end{aligned}$$

as $f(y,x) - kf(y,x)^\sigma \in D_{(\sigma,-k)} \le \Lambda$.

[3] Again, this should really be the tuple $(D, \sigma, k, V, h, \Lambda, q)$.

(ii) $q(ax) = f(ax, ax) + \Lambda = af(x, x)a^\sigma + \Lambda = aq(x)a^\sigma + \Lambda$.

(iii) For arbitrary $a \in D$ we calculate $q((a + 1)x) = q(ax + 1)$ in two ways—using (i) and using (ii).

$$\begin{aligned} q((a+1)x) &= q(ax+x) \\ (a+1)q(x)(a+1)^\sigma &= q(ax) + q(x) + h(ax, x) + \Lambda \\ aq(x)a^\sigma + q(x) + aq(x) + q(x)a^\sigma &= aq(x)a^\sigma + q(x) + ah(x, x) + \Lambda \\ aq(x) + q(x)a^\sigma &= ah(x, x) + \Lambda\,. \end{aligned}$$

Again as $D_{(\sigma,-k)} \leq \Lambda$

$$\begin{aligned} aq(x) + q(x)a^\sigma - (q(x)a^\sigma - k(q(x)a^\sigma)^\sigma) &= ah(x, x) + \Lambda \\ aq(x) + ka^{\sigma^2}q(x)^\sigma = aq(x) + kk^{-1}akq(x)^\sigma &= ah(x, x) + \Lambda \\ aq(x) + akq(x)^\sigma &= ah(x, x) + \Lambda\,. \end{aligned}$$

That is, for all $a \in D$,

$$a\,(q(x) + kq(x)^\sigma - h(x, x)) \in \Lambda\,.$$

If $\Lambda \neq D$ and $q(x) + kq(x)^\sigma - h(x, x)$ is not identically 0, this gives a contradiction.

Therefore we have (iii) except possibly when $\Lambda = D$. In that case as $\Lambda \leq D^{(\sigma,-k)}$ we also have $D^{(\sigma,-k)} = \{\, a \in D \mid a = -ka^\sigma \,\} = D$. With $a = 1$ we find $1 = -k1^\sigma$, so $k = -1$. Then always $a = -ka^\sigma = -(-1)a^\sigma = a^\sigma$, so $\sigma = 1$. Thus $h(x, x) = f(x, x) + kf(x, x)^\sigma = f(x, x) - f(x, x) = 0$, and so we again have (iii):

$$h(x, x) = 0 = q(x) - q(x) = q(x) + kq(x)^\sigma\,.$$

(b) *Kernel.* The kernel consists of those f for which $q(x) = \Lambda$ and $h(x, y) = 0$ for all $x, y \in V$. Therefore

$$0 = h(x, y) = f(x, y) + kf(y, x)^\sigma\,,$$

which is to say $f(x, y) = -kf(y, x)^\sigma$ and so $f \in \mathrm{Herm}^{(\sigma,-k)}(V)$. Furthermore $f(x, x) + \Lambda = q(x) + \Lambda = \Lambda$. Thus $f \in \mathrm{Herm}_\Lambda^{(\sigma,-k)}(V)$.

(c) *Surjective.* Let (q, h) be a Λ-pseudoquadratic form. For $\{\, x_i \mid i \in I \,\}$ a D-basis of V indexed by the totally ordered set $(I, <)$, let the Gram matrix for a form f be:

$$\begin{aligned} f(x_i, x_j) &= h(x_i, x_j) && \text{for } i < j\,; \\ f(x_i, x_j) &= q_i && \text{for } i = j\,; \\ f(x_i, x_j) &= 0 && \text{for } j < i\,, \end{aligned}$$

where each q_i is a representative in D for the coset $q(x_i)$. As in Theorem (10.12) we have $h(x, y) = f(x, y) + kf(y, x)^\sigma$ for all $x, y \in V$.

We must check $q(x) = f(x,x) + \Lambda$ for all $x, y \in V$. Let $x = \sum_{i\in I} a_i x_i \in V$. By (i) (and induction) and (ii) for the pseudoquadratic form (q,h):

$$\begin{aligned} q(x) &= q\Big(\sum_{i\in I} a_i x_i\Big) + \Lambda \\ &= \sum_{i\in I} q(a_i x_i) + \sum_{i<j} h(a_i x_i, a_j x_j) + \Lambda \\ &= \sum_{i\in I} a_i q(x_i) a_i^\sigma + \sum_{i<j\in I} a_i h(x_i, x_j) a_j^\sigma + \Lambda\,. \end{aligned}$$

On the other hand

$$\begin{aligned} f(x,x) &= f\Big(\sum_{i\in I} a_i x_i, \sum_{j\in I} a_j x_j\Big) \\ &= \sum_{i,j\in I} f(a_i x_i, a_j x_j) = \sum_{i,j\in I} a_i f(x_i, x_j) a_j^\sigma \\ &= \sum_{i\in I} a_i q_i a_i^\sigma + \sum_{i<j\in I} a_i h(x_i, x_j) a_j^\sigma\,. \end{aligned}$$

As $q_i \in q(x_i)$ for all i, we get $q(x) = f(x,x) + \Lambda$ for all $x \in V$, as desired.

With (a), (b), and (c) all complete, we have the theorem. □

Two observations embedded in the proof of the theorem are of more general interest.

(10.20). Lemma.

(a) *If (q,h) is a Λ-pseudoquadratic form on V with $\Lambda \neq D$, then the hermitian form h is uniquely determined by q.*

(b) *If (q,h) is a Λ-pseudoquadratic form on V, then h is a (σ,k)-trace-hermitian form. Conversely, if h is a trace-valued (σ,k)-hermitian form on V, then, for every (σ,k)-form parameter Λ in D, there is a quadratic part $q\colon V \longrightarrow D/\Lambda$ such that (q,h) is a Λ-pseudoquadratic form.*

Proof. (a) Suppose (q,h_1) and (q,h_2) are both Λ-pseudoquadratic forms. In particular always

$$h_1(x,y) + \Lambda = q(x+y) - q(x) - q(y) = h_2(x,y) + \Lambda\,.$$

Thus $h_0 = h_1 - h_2$ is a (σ,k)-hermitian form, for which $h_0(x,y)$ always belongs to Λ. If some $h_0(x_0,y_0) \neq 0$, then $h_0(ax_0,y_0) = ah(x_0,y_0)$ runs through D, hence $D \leq \Lambda \leq D$.

Therefore whenever $\Lambda \neq D$ the form h_0 must be identically 0, and $h_1 = h_2$.

(b) The condition $h(x,x) = q(x) + kq(x)^\sigma$ shows that the hermitian part of a pseudoquadratic form must be trace-valued and so, by Theorem (10.12), is trace-hermitian. On the other hand, the previous theorem tells us that for any trace-hermitian form $h(x,y) = f(x,y) + kf(y,x)^\sigma$ and any form parameter Λ, the map $q(x) = h(x,x) + \Lambda$ extends h to a pseudoquadratic form (q,h). □

(10.21). Lemma. *Let $0 \neq c \in D$. If Λ is a (σ, k)-form parameter and (q, h) is a Λ-pseudoquadratic form, then Λc is a $(\sigma c, k')$-form parameter for $k' = k(c^\sigma)^{-1}c$ and (qc, hc) is a Λc-pseudoquadratic form.*

Proof. From Lemma (10.14)

$$D_{(\sigma c, -k')} = D_{(\sigma, -k)}c \leq \Lambda c \leq D^{(\sigma, -k)}c = D^{(\sigma c, -k')}$$

for the additive subgroup Λc of $(D, +)$. Furthermore

$$a(\Lambda c)a^{\sigma c} = a(\Lambda c)c^{-1}a^\sigma c = a(\Lambda cc^{-1})a^\sigma c = (a\Lambda a^\sigma)c = \Lambda c\,,$$

for all $0 \neq a \in D$. Therefore Λc is a $(\sigma c, k')$-form parameter.

Following Theorem (10.19), let f be a σ-sesquilinear form, from which the Λ-pseudoquadratic form (q, h) is given by

$$q(x) = f(x, x) + \Lambda \quad \text{and} \quad h(x, y) = f(x, y) + kf(y, x)^\sigma\,.$$

Clearly $q(x)c = f(x, x)c + \Lambda c$. By Lemma (10.14) again, hc is a $(\sigma c, k')$-hermitian form completing the Λc-pseudoquadratic form (qc, hc). □

As before, the pseudoquadratic form (qc, hc) is said to be *proportional* to (q, h).

(10.22). Theorem. (Bestiary of Pseudoquadratic Forms) *Let Λ be a (σ, k)-form parameter on the division ring D, and let (q, h) be a Λ-pseudoquadratic form on V. Then, up to proportionality, we have one of:*

(1) Symplectic Case: *$\Lambda = D$; $k = -1$; $\sigma = 1$; D is a field; q is identically 0; and the symplectic bilinear form h has $h(x, x) = 0$ for all $x \in V$.*

(2) Generic Orthogonal Case: *$\Lambda = 0$; $k = 1$; $\sigma = 1$; D is a field; and there are $q\colon V \longrightarrow D$ and symmetric bilinear form $h\colon V \times V \longrightarrow D$ satisfying:*

  (i) *$q(x + y) - q(x) - q(y) = h(x, y)$,*

  (ii) *$q(ax) = a^2 q(x)$,*

  for all $x, y \in V$ and $a \in D$.

(3) Generic Unitary Case: *$1 \in \Lambda = D^{(\sigma,1)} < D$; $k = -1$; $\sigma^2 = 1 \neq \sigma$; and h is a σ-skew-hermitian form.*

(4) Exceptional Orthogonal Case: *$1 \in \Lambda < D^{(\sigma,1)} = D$; $k = \pm 1$; $\sigma = 1$; D is a field;* $\mathrm{char}(D) = 2$*; and the symplectic bilinear form h has $h(x, x) = 0$ for all $x \in V$.*

(5) Exceptional Unitary Case: *$1 \in \Lambda < D^{(\sigma,1)} < D$; $k = \pm 1$; $\sigma^2 = 1 \neq \sigma$;* $\mathrm{char}(D) = 2$*; D is not a field; and h is a σ-hermitian form.*

Proof.

(a) *Case* $\Lambda = D$.

Certainly $q\colon V \longrightarrow D/\Lambda = 0$ is identically 0. Thus, for all $x \in V$, we have $h(x,x) = q(x) + kq(x)^\sigma = 0$.

As $D = \Lambda \leq D^{(\sigma,-k)}$, we have $a = -ka^\sigma$ for all $a \in D$. In particular $1 = -k1^\sigma$, so $k = -1$. Then $a = -(-1)a^\sigma = a^\sigma$ gives $\sigma = 1$. Especially D is a field. Thus we are in the Symplectic Case (1).

(b) *Case* $\Lambda = \{0\}$.

Here $D_{(\sigma,-k)} \leq \Lambda = \{0\}$; so $a - ka^\sigma = 0$ and $a = ka^\sigma$, for all $a \in D$. Especially $1 = k1^\sigma$, and $k = 1$. Then $a = ka^\sigma = a^\sigma$, and $\sigma = 1$. Again D is a field.

As $\Lambda = \{0\}$, the conditions for a pseudoquadratic form with $q\colon V \longrightarrow D/\Lambda = D$ give us a symmetric bilinear form h defined by

$$\text{(i)} \quad q(x+y) - q(x) - q(y) = h(x,y)$$

with the quadratic form q satisfying additionally

$$\text{(ii)} \quad q(ax) = aq(x)a^\sigma = a^2 q(x)\,.$$

The condition

$$\text{(iii)} \quad h(x,x) = q(x) + kq(x)^\sigma = 2q(x)$$

is actually a direct consequence of (i) (for $x = y$) and (ii) and need not be repeated. We are in the Generic Orthogonal Case (2).

(c) *Case* $0 \lneq \Lambda \lneq D$.

By passing to a proportional form (see Lemma (10.21)), we may assume that $1 \in \Lambda$. As $\Lambda \leq D^{(\sigma,-k)}$, this gives $1 = -k1^\sigma$ hence $k = -1$. But σ^2 is conjugation by k, so we also have $\sigma^2 = 1$. Furthermore $D^{(\sigma,-k)} = \{\, a \in D \mid a = -ka^\sigma = a^\sigma \,\}$ is the additive subgroup of elements fixed by σ.

(i) $\sigma = 1$. Again D is a field. Here

$$D^{(\sigma,-k)} = D \gneq \Lambda \geq D_{(\sigma,-k)} = \{\, a - (-1)a^1 = 2a \mid a \in D \,\} = 2D\,,$$

which can only happen in characteristic 2. In particular always $h(x,x) = q(x) + kq(x)^\sigma = 2q(x) = 0$, so the bilinear form h is symplectic. We are in the Exceptional Orthogonal Case (4).

(ii) $\sigma \neq 1 = \sigma^2$. We divide this into two subcases, depending upon whether or not Λ is proper in the fixed point additive subgroup $D^{(\sigma,1)}$, itself proper in D. When $\Lambda = D^{(\sigma,1)}$ we are in the Generic Unitary Case (3). When $\Lambda \lneq D^{(\sigma,1)}$ we find $D_{(\sigma,-k)} \leq \Lambda \lneq D^{(\sigma,-k)}$, but by Proposition (10.11)(c) this can only happen when the characteristic is 2 and $\sigma|_{\mathrm{Z}(D)}$ is trivial. As $\sigma \neq 1$, this second restriction forces $D \neq \mathrm{Z}(D)$ so that D is not a field. We finish with the Exceptional Unitary Case (5). □

The Generic Orthogonal Case (10.22)(2) gives the standard definition for a function $q\colon V \longrightarrow D$ to be a *quadratic form* of the space V over the field D.

10.4 Isometries and adjoints

As we have seen above, for instance in Lemma (10.20), often the pseudoquadratic form (q, h) is uniquely determined by its quadratic part $q \colon V \longrightarrow D$. In that case the isometries of the form are naturally those $g \in \mathrm{GL}_D(V)$ with $q(xg) = q(x)$, for all $x \in V$. Then we have

$$\begin{aligned} h(xg, yg) + \Lambda &= q((x+y)g) - q(xg) - q(yg) \\ &= q(x+y) - q(x) - q(y) = h(x, y) + \Lambda\,, \end{aligned}$$

so in this setting the isometry condition inherited by the associated hermitian part h is $h(xg, yg) = h(x, y)$ for all $x, y \in V$.

Accordingly, an *isometry* of the σ-sesquilinear form $f \colon V \times V \longrightarrow D$ is defined to be an element $g \in \mathrm{GL}_D(V)$ with

$$f(xg, yg) = f(x, y)\,, \quad \text{for all } x, y \in V\,.$$

The full *isometry group* of the sesquilinear form f on the D-space V is then $\mathrm{I}_D(V, f)$ or $\mathrm{I}(V, f)$.

Then, an *isometry* of an arbitrary pseudoquadratic form (q, h) is a $g \in \mathrm{GL}_D(V)$ with

$$q(xg) = q(x) \text{ and } h(xg, yg) = h(x, y)\,, \quad \text{for all } x, y \in V\,.$$

The corresponding full *isometry group* of the pseudoquadratic form (q, h) on the D-space V is then $\mathrm{I}_D(V, q, h)$ or $\mathrm{I}(V, q, h)$.

Tits introduced pseudoquadratic forms in part to inject some uniformity into the treatment of the classical groups. The next result suggests the background for that desire. The generic classical groups (including the symplectic groups) can all be viewed as the isometry groups of a single form. But sometimes its a bilinear form, sometimes a quadratic form, and sometimes a hermitian form:

Given an anti-automorphism σ, to each $g \in \mathrm{GL}_D(V)$ we can associate $g^\sigma \in \mathrm{GL}_D(V^\sigma)$, acting on the left:

$$g^\sigma.v = v.g \quad \text{or, equivalently,} \quad v.g^{\sigma^{-1}} = g.v\,.$$

For a basis $\{\, e_i \mid i \in I \,\}$ of V, if we have $e_i.g = \sum_{j\in I} g_{ij} e_j$, then $g^\sigma.e_i = \sum_{j\in I} e_j.g_{ij}^\sigma$; so the matrix representing g^σ in this basis is the transpose-σ-conjugate of that representing g. Therefore g is an isometry of the σ-sesquilinear form f precisely when $(g, g^\sigma) \in \mathrm{GL}_D(V, V^\sigma, m)$, where $m \colon V \times V^\sigma$ is the pairing given by $m(x, y) = f(x, y)$.

Here it is important to remember the distinction between the seemingly identical m and f. Not every isometry (g, h) of the pairing m gives an isometry g of the sesquilinear form f; we must additionally require $h = g^\sigma$. In view of Proposition (9.12) this provides us with the familiar requirement $gg^\sigma = 1$.

The fact that (in at least the finite dimensional case) every $g \in \mathrm{GL}_D(V)$ extends to a $(g, h) \in \mathrm{GL}_D(V, V^\sigma, m)$ still plays a role in the study of σ-sesquilinear forms.

(10.23). LEMMA. *Let f be a nondegenerate σ-sesquilinear form on the finite dimensional D-space V. Then for every $g \in \mathrm{GL}_D(V)$ there is a unique $g^\circ \in \mathrm{GL}_D(V)$ with*

$$f(xg, y) = f(x, yg^\circ)\,, \quad \textit{for all } x, y \in V\,.$$

PROOF. As V is finite dimensional, every $g \in \mathrm{GL}_D(V)$ extends to an isometry (g, h) of the associated pairing $m \colon V \times V^\sigma \longrightarrow D$ by Proposition (9.16). For all $x, y \in V$,

$$\begin{aligned} f(xg, y) &= m(xg, h(h^{-1}y)) \\ &= m(x, h^{-1}y) \\ &= f(x, y((h^{-1})^{\sigma^{-1}}))\,. \end{aligned}$$

Thus $g^\circ = (h^{-1})^{\sigma^{-1}}$ has the desired property. Uniqueness follows from Lemma (9.8). □

The linear transformation g° is the *adjoint* of g (with respect to f).

10.5 Extension and reduction

10.5.1 Perpendicular direct sums

(10.24). PROPOSITION. *Let h be a σ-hermitian form of the space $V = A \oplus B$ with $h(a, b) = 0$ for all $a \in A$ and $b \in B$. Then for $u, v \in V$ with $u = a + b$ and $v = c + d$ where $a, c \in A$ and $b, d \in B$, we have $h(u, v) = h(a, c) + h(b, d)$.*

Additionally if (q, h) is a pseudoquadratic form on V, then $q(a+b) = q(a) + q(b)$.

PROOF. $h(u, v) = h(a + b, c + d) = h(a, c) + h(a, d) + h(b, c) + h(b, d) = h(a, c) + h(b, d)$ and $q(a + b) = q(a) + q(b) + h(a, b) = q(a) + q(b)$. □

(10.25). PROPOSITION. *Let h_A be a σ-hermitian form on the D-space A and h_A be a σ-hermitian form on the D-space B. Then on $V = A \oplus B$ we have the σ-hermitian form $h = h_A + h_B$ given by*

$$h((a, b), (c, d)) = h_A(a, c) + h_B(b, d)\,.$$

Additionally if (q_A, h_A) is a pseudoquadratic form on A and (q_B, h_B) is a pseudoquadratic form on B, both for the (σ, k)-form parameter Λ in D, then (q, h) is a pseudoquadratic form on V for Λ with h as above and $q = q_A + q_B$ given by

$$q((a, b)) = q_A(a) + q_B(b)\,. \qquad \square$$

In the situation described by these two propositions, the space V is the *perpendicular direct sum* of the spaces A and B, written $A \perp B$. The first proposition presents the internal version of the perpendicular direct sum and the second the external version.

(10.26). Corollary. *If $g_A \in \mathrm{I}_D(A, h_A)$, respectively $\mathrm{I}_D(A, q_A, h_A)$, and $g_B \in \mathrm{I}_D(B, h_B)$, respectively $\mathrm{I}_D(B, q_B, h_B)$, then $g = (g_A, g_B) \in \mathrm{I}_D(A \perp B, h)$, respectively $\mathrm{I}_D(A \perp B, q, h)$, with the action of g given by*

$$(a, b)g = (a, b)(g_A, g_B) = (ag_A, bg_B)\,. \qquad \Box$$

10.5.2 Radicals

As already discussed, the *radical* of the hermitian space (V, h) is

$$\mathrm{Rad}(V, h) = V^\perp = \{\, v \in V \mid h(v, x) = 0\,, \text{for all } x \in V \,\}\,.$$

The form and space are *nondegenerate* if this radical is $\{0\}$.

As in Corollary (9.2) the form h induces a well defined form h^0 on $V^0 = V/\mathrm{Rad}(V, h)$ given by

$$h^0(x + V^\perp, y + V^\perp) = h(x, y)\,.$$

Indeed we could take V^0 to be any complement to $\mathrm{Rad}(V, h)$ in V. We then would have $V = \mathrm{Rad}(V, h) \perp V^0$. As h restricted to $\mathrm{Rad}(V, h)$ is trivial, we can combine this with Corollary (10.26) to find an injection

$$\mathrm{GL}_D(\mathrm{Rad}(V, h)) \times \mathrm{I}_D(V^0, h^0) \longrightarrow \mathrm{I}_D(V, h)\,.$$

More can be said:

(10.27). Theorem. *In the situation described above there is a split short exact sequence*

$$1 \longrightarrow \mathrm{Hom}_D(V^0, V^\perp) \longrightarrow \mathrm{I}_D(V, h) \longrightarrow \mathrm{GL}_D(\mathrm{Rad}(V, h)) \times \mathrm{I}_D(V^0, h^0) \longrightarrow 1\,.$$

Proof. We have already described the surjection and splitting, so we now need to find the kernel—trivial both on the quotient $V/V^\perp$ and on $V^\perp$ These are precisely the elements which, for each $v = v_0 + r$ of V, with $v_0 \in V^0$ and $r \in V^\perp$ have $v \longrightarrow v_0 + r + v_0\varphi$ for some $\varphi \in \mathrm{Hom}_D(V^0, V^\perp)$. $\Box$

A consequence is that in discussing the isometry groups of hermitian forms we are largely reduced to questions about the isometry groups of nondegenerate forms.

A similar reduction for isometry groups of pseudoquadratic forms is not quite so elementary. For the pseudoquadratic form (q, h) on V the *singular radical* $\mathrm{SRad}(V, q, h) = \{\, v \in \mathrm{Rad}(V, h) \mid q(v) \in \Lambda \,\}$ may be a proper subspace of $\mathrm{Rad}(V, h)$. The form (q, h) is *nonsingular*[4] provided $\mathrm{SRad}(V, q, h) = \{0\}$.

(10.28). Proposition. *For the Λ-pseudoquadratic space (V, q, h), set $V^1 = V/\mathrm{SRad}(V, q, h)$. The maps $q^1 \colon V^1 \longrightarrow D/\Lambda$ and $h^1 \colon V^1 \times V^1 \longrightarrow D$ given by*

$$q^1(x + \mathrm{SRad}(V, q, h)) = q(x)$$

[4] This is unhappy terminology. Other terms are used for this elsewhere, and this term is used to mean other things elsewhere. We avoid it for the most part.

and

$$h^1(x + \mathrm{SRad}(V,q,h), y + \mathrm{SRad}(V,q,h)) = h(x,y)$$

provide a Λ-pseudoquadratic form (q^1, h^1) on V^1 with $\mathrm{SRad}(V^1, q^1, h^1) = \{0\}$. □

The space (V^1, q^1, h^1), while nonsingular, may be degenerate. The dimension of the space

$$\mathrm{Rad}(V^1, h^1) = \mathrm{Rad}(V,h)/\mathrm{SRad}(V,q,h)$$

is the *defect* of the form (q,h), and the form is *defective* if it has positive defect. In the defect 0 case the quadratic part q naturally induces, as before, a Λ-pseudoquadratic form q^0 on V^0, yielding the nondegenerate Λ-pseudoquadratic space $(V^0, q^0, h^0) = (V^1, q^1, h^1)$. In positive defect, the appropriate induced form q^0 may not exist. The previous proposition provides one partial remedy for this, and the next another.

(10.29). Proposition. *Let (V,q,h) be Λ-pseudoquadratic space over D, and let Λ^0 be the additive subgroup of $(D,+)$ generated by Λ and $\{\, q(x) \mid x \in$ $\mathrm{Rad}(V,h)\,\}$. Then Λ^0 is a form parameter in D, and the map $q^0 \colon V \longrightarrow D/\Lambda^0$ given by*

$$q^0(x + \mathrm{Rad}(V,h)) = q(x) + \Lambda^0$$

provides a nondegenerate Λ^0-pseudoquadratic form (q^0, h^0) on V^0.

Proof.

□

We can at least be happy in that our two partial remedies lead to the same group.

(10.30). Theorem. *With the same notation as above we have*

$$\mathrm{I}_D(V^0, q^0, h^0) \simeq \mathrm{I}_D(V^1, q^1, h^1)\,.$$

Proof.

□

10.6 Singular and hyperbolic spaces

Let Λ be a (σ, k)-form parameter in D and (q,h) be a Λ-pseudoquadratic form on the D-space V. The nonzero vector v is a *singular vector* if $q(v) = \Lambda$ (which might more suggestively be written $q(v) = 0$). If v is singular, then so is the 1-space it spans.

The space and form are *singular* if there is a singular vector not in the radical. Of course, there might be no such vectors, in which case the space and form are *asingular*. What we are calling singular and asingular are elsewhere called *isotropic* and *anisotropic*. For us, an *isotropic vector* is a nonzero vector with $h(x,x) = 0$; so a singular vector is a special type of isotropic vector. A space is *isotropic* when it contains a nonradical isotropic vector. Thus a singular space is isotropic, but the converse need not be the case. Similarly an anisotropic space is asingular, but the converse may not hold.

A subspace U of V is *totally isotropic* provided h is identically 0 on U. Similarly, U is *totally singular* if both q and h are identically 0 on U; it is totally isotropic and all of its nonzero vectors are (Λ-)singular.

From now on, almost all our work on the classical groups and geometries will be concerned with the singular case. That is all because of the following important lemma.

(10.31). LEMMA. *Let (V, q, h) be a pseudoquadratic space of dimension 2, and let x a singular vector of V that is not in the radical $\mathrm{Rad}(V, h)$.*

(a) *V is nondegenerate, and there is a singular vector y with $V = \langle x, y\rangle$ and $h(x, y) = 1$.*

(b) *For $a \in D$, we have $q(ax + y) = a$. In particular, there are $1 + |\Lambda|$ singular 1-spaces in V, being those spanned by x and by $ax + y$ for $a \in \Lambda$.*

PROOF.

□

In this case V is a *hyperbolic 2-space* and the ordered pair of vectors (x, y) is a *hyperbolic pair.*

(10.32). COROLLARY. *Let (V, q, h) be a singular pseudoquadratic space. The map $q \colon V \longrightarrow D/\Lambda$ is surjective, and the value map $v \colon V \longrightarrow D_{(\sigma,k)}$ given by $v(x) = h(x, x)$ is surjective.*

PROOF. By the lemma every singular space contains a hyperbolic 2-space H, and the conclusions already hold within H. □

If the pseudoquadratic space (V, q, h) is isometric to a direct sum of hyperbolic 2-spaces, then it is called a *hyperbolic space.* The hyperbolic spaces are of fundamental importance. The union of hyperbolic pairs for these summands is then a *hyperbolic basis* for V.

(10.33). PROPOSITION. *Let the hyperbolic space (V, q, h) be the perpendicular direct sum of the hyperbolic 2-spaces V_i, for $i \in I$, with corresponding hyperbolic pairs (x_i, y_i). Let X be the span of the x_i and Y the span of the y_i. Then X and Y are both maximal totally singular and $V = X \oplus Y$.*

Conversely, if finite or countable dimensional V is nondegenerate and equal to $X \oplus Y$ where X and Y are maximal totally singular subspaces, then V is hyperbolic.

PROOF.

□

In finite dimension, this property has often been used as the definition of a hyperbolic space.

(10.34). THEOREM.

(a) *Let the hyperbolic space (V, q, h) be equal to the sum $V = X \oplus Y$, with X and Y maximal totally singular, and let $m\colon X \times Y^\sigma \longrightarrow D$ be the corresponding hyperbolic pairing given by $h(x, y) = m(x, y)$. Then the global stabilizer of the two subspaces X and Y in $\mathrm{I}_D(V, q, h)$ is $\mathrm{GL}_D(X, Y^\sigma, m) \simeq \mathrm{GL}_D(X)$.*

(b) *The hyperbolic 2-space (V, q, h) is unique up to isometry. Its isometry group is transitive on hyperbolic pairs and, especially, is 3-transitive on its singular 1-spaces.*

PROOF.

□

10.7 Problems

(10.35). PROBLEM. *Let Λ be a (σ, k)-form parameter that is proper in the division ring D; that is, $\Lambda \lneqq D$. Let q be a map from V to the abelian group D/Λ and $h\colon V \times V \longrightarrow D$ a (σ, k)-hermitian form on V that together satisfy, for all $x, y \in V$ and $a \in D$,*

(i) $q(x + y) - q(x) - q(y) = h(x, y) + \Lambda$;

(ii) $q(ax) = aq(x)a^\sigma + \Lambda$.

Prove that, for all $x \in V$, we have $h(x, x) = q(x) + kq(x)^\sigma$ (well-defined). Thus (q, h) is a pseudoquadratic form on V. That is, show that in the case $\Lambda \lneqq D$, the third defining condition for a pseudoquadratic form (from page 136) is a consequence of the first two conditions.

(10.36). PROBLEM. *Let $(p, j) = (p^+, j^+)$ be a nondegenerate Λ-pseudoquadratic form on the space $U = U^+$. Let U^- be a second copy of U, now equipped with the Λ-pseudoquadratic form $(-p, -j) = (p^-, j^-)$ Prove that the Λ-pseudoquadratic form $(q, h) = (p^+ + p^-, j^+ + j^-)$ is hyperbolic on $V = U^+ \perp U^-$.*

(10.37). PROBLEM. *Let Λ be a (σ, k)-form parameter in the division ring D, and suppose $m\colon {}_DX \times Y_D \longrightarrow D$ is a nondegenerate pairing. Set $V = X \oplus Y^{\sigma^{-1}}$ and define on the D-space V the pair of maps*

$$q((x, y)) = m(x, y) + \Lambda \quad \textit{and} \quad h((x_1, y_1), (x_2, y_2)) = m(x_1, y_2) + km(x_2, y_1)^\sigma .$$

(a) *Prove that (V, q, h) is a Λ-pseudoquadratic space and X and Y are maximal totally singular subspaces of V.*

(b) *Prove that if m is a hyperbolic pairing, then (V, q, h) is a hyperbolic space.*

(10.38). PROBLEM. *This problem indicates how, if we extend definitions to include modules over rings, the general linear group $\mathrm{GL}_D(V)$, for a division ring D, can be realized as the full isometry group of a hermitian form defined on a suitable module.*

Let D^+ be the ring $D \oplus D^{\mathrm{op}}$ and set $V^+ = V \oplus V^*$, where we view V^* as a left D^{op}-space. The group V^+ is then a unital left D^+-module with scalar multiplication given by

$$(a, b) \star (x, \lambda) = (ax, b \cdot \lambda) = (ax, \lambda b) .$$

Define a map $h\colon V^+ \times V^+ \longrightarrow D^+$ by

$$h((x, \lambda), (y, \mu)) = (x\mu, y\lambda) .$$

(a) *Prove that h is a σ-hermitian form on V^+ with respect to the anti-automorphism $\sigma\colon D^+ \longrightarrow D^+$ given by $(a,b)^\sigma = (b,a)$.*

(b) *Prove that*

$$\begin{aligned} \mathrm{GL}_D(V) &\simeq \mathrm{I}_{D^+}(V^+, h) \\ &= \{\, g \in \mathrm{GL}_{D^+}(V^+) \mid h(x,y) = h(xg, yg) \text{ for all } x, y \in V^+ \,\}. \end{aligned}$$

HINT: *What are the idempotents of D^+?*

REMARKS. *There are more general versions of this that involve form parameters on rings; see [HaO89]. In that context this problem is related to the previous problem: if σ is an order 2 anti-automorphism of D—an isomorphism of D and D^{op}—then D^+ is $D \oplus D^\sigma$ and D sits on the diagonal of D^+ as $\{\, (d, d^\sigma) \mid d \in D \,\}$.*

Chapter 11

Isometry Groups

11.1 The Wall form of an isometry

Let (V, q, h) be a nondegenerate Λ-pseudoquadratic form on the D-space V for the (σ, k)-form parameter Λ. For each subspace U of V, the space $(U, q|_U, h|_U)$ is also a Λ-pseudoquadratic space, but it may be degenerate. G.E. Wall [Wal63] observed that in certain important cases a second nondegenerate form on U can be of great help.

(11.1). PROPOSITION. *Let (V, q, h) be a Λ-pseudoquadratic form on the D-space V for the (σ, k)-form parameter Λ, and let $g \in \mathrm{I}_D(V, q, h)$. Set $\gamma = g - 1$ and $U = [V, g] = V(g-1) = V\gamma$, and assume that U meets the radical of (V, q, h) trivially. Then the map $h_g \colon U \times U \longrightarrow D$ given by*

$$h_g(x\gamma, y\gamma) = h(x, y\gamma) = -h(x\gamma, yg)$$

is a well-defined nondegenerate σ-sesquilinear form on U.

PROOF. We first observe that, for all $x, y \in V$,

$$\begin{aligned} h(x, y\gamma) &= h(x, y(g-1)) = h(x, yg) - h(x, y) \\ &= h(x, yg) - h(xg, yg) = h(x - xg, yg) \\ &= h(-x\gamma, yg) = -h(x\gamma, yg)\,. \end{aligned}$$

As $h_g(x\gamma, y\gamma) = h(x, y\gamma)$ it is well-defined in its second coordinate, and as $h_g(x\gamma, y\gamma) = h(-x\gamma, yg)$ it is well-defined in its first coordiante; it is well-defined. As γ is a D-linear transformation, h_g is σ-sesquilinear, although unlikely to be reflexive/hermitian. If $h_g(x\gamma, y\gamma) = 0$ for all $x\gamma \in U$, then $h(x, y\gamma) = 0$ for all $x \in V$. That is, $y\gamma \in U \cap V^\perp = \{0\}$ by hypothesis. Similarly if $h_g(x\gamma, y\gamma) = h(-x\gamma, yg) = 0$ for all $y\gamma \in U$ then $x\gamma \in U \cap V^\perp = \{0\}$. We conclude that h_g is nondegenerate on U. □

The form h_g is the *Wall form* for the isometry g.

(11.2). PROPOSITION. *Continue the notation of Proposition (11.1). Also, for $x, y \in V$ set $u = x\gamma$ and $v = y\gamma$.*

(a) $h_g(u,v) + kh_g(v,u)^\sigma = -h(u,v)$;

(b) $h_g(u,u) + \Lambda = -q(u)$;

(c) $h_g(ug,v) = -kh_g(v,u)^\sigma$;

(d) $h_g(u,v) = h_g(ug,vg)$.

PROOF. (a)

$$\begin{aligned} h_g(u,v) + kh_g(v,u)^\sigma &= h(x,v) - kh(v,xg)^\sigma = h(x,v) - k(kh(xg,v)^\sigma)^\sigma \\ &= h(x,v) - kh(xg,v)^{\sigma^2}k^\sigma = h(x,v) - kk^{-1}h(xg,v)kk^\sigma \\ &= h(x,v) - h(xg,v) = -h(u,v)\,. \end{aligned}$$

(b) As $q(x) - kq(x)^\sigma \in D_{(\sigma,-k)} \leq \Lambda$,

$$\begin{aligned} h_g(u,u) + \Lambda &= h(x, x(g-1)) + \Lambda = h(x,xg) - h(x,x) + \Lambda \\ &= h(x,xg) - (q(x) + kq(x)^\sigma) + \Lambda \\ &= -h(-x,xg) - q(-x) + (-q(xg) + q(xg)) - kq(x)^\sigma + \Lambda \\ &= (-h(-x,xg) - q(-x) - q(xg)) + (q(xg) - kq(x)^\sigma) + \Lambda \\ &= -q(xg - x) + (q(x) - kq(x)^\sigma) + \Lambda = -q(u)\,. \end{aligned}$$

(c) By (a)

$$\begin{aligned} h_g(ug,v) &= h_g(u(g-1) + u, v) = h_g(u(g-1),v) + h_g(u,v) \\ &= h(u,v) + h_g(u,v) = -kh_g(v,u)^\sigma\,. \end{aligned}$$

(d) We use (c) twice:

$$\begin{aligned} h_g(ug,vg) &= -kh_g(vg,u)^\sigma = k(k(h_g(u,v))^\sigma)^\sigma \\ &= k(h_g(u,v))^{\sigma^2}k^\sigma = kk^{-1}h_g(u,v)kk^\sigma = h_g(u,v)\,. \qquad \square \end{aligned}$$

The astonishing thing is that the propositions have a strong converse.

(11.3). THEOREM. *Let (V,q,h) be a Λ-pseudoquadratic form on the D-space V for the (σ,k)-form parameter Λ. Let U be a finite dimensional subspace of V that intersects the radical trivially. Assume that nondegenerate $f \in \mathrm{Sesq}^\sigma(U)$ satisfies, for all $u,v \in U$,*

$$f(u,v) + kf(v,u)^\sigma = -h(u,v) \quad \textit{and} \quad f(u,u) + \Lambda = -q(u)\,.$$

Then, for each $x \in V$, there is a unique $x\gamma \in U$ with $f(x\gamma,v) = h(x,v)$ for all $v \in U$. Additionally $g = \mathrm{w}(U,f) = \mathrm{Id}_V + \gamma$ is the unique isometry g of (V,q,h) with $U = V(g-1)$ and $f = h_g$.

Proof. By Lemma (10.2) for each $v \in U$

$$\lambda^f_v = f(v, \cdot) \in (U^\sigma)^*$$

and for each $x \in V$

$$\lambda^h_x = h(x, \cdot)|_U \in (U^\sigma)^* \,.$$

By Corollary (10.3), as U is finite dimensional and f is nondegenerate, for every $x \in V$ there is a unique $x\gamma \in U$ with

$$\lambda^f_{x\gamma} = \lambda^h_x \,.$$

Conversely (again by Corollary (10.3)) as U is finite dimensional and meets the radical of h trivially, for every $v \in U$ the linear functional λ^f_v is induced by some λ^h_x.

Therefore the map $\gamma \colon V \longrightarrow U$ is well-defined and surjective. By Lemma (10.2) the maps λ^h and λ^f are D-linear, so γ is as well:

$$\lambda^f_{(ax+by)\gamma} = \lambda^h_{ax+by} = a\lambda^h_x + b\lambda^h_y = a\lambda^f_{x\gamma} + b\lambda^f_{y\gamma} = \lambda^f_{a(x\gamma)+b(y\gamma)} \,.$$

Set $g = \mathrm{Id}_V + \gamma = 1 + \gamma \ \in \mathrm{End}_D(V)$.

The above equality gives

$$f(x\gamma, y\gamma) = \lambda^f_{x\gamma}(y\gamma) = \lambda^h_x(y\gamma) = h(x, y\gamma)\,,$$

for all $x, y \in V$. In particular, if g is an isometry then $f = h_g$.

For all $x, y \in V$

$$\begin{aligned} h(xg, yg) &= h(x(1+\gamma), y(1+\gamma)) \\ &= h(x,y) + h(x\gamma, y\gamma) + h(x, y\gamma) + h(x\gamma, y) \\ &= h(x,y) + h(x\gamma, y\gamma) + h(x, y\gamma) + kh(y, x\gamma)^\sigma \\ &= h(x,y) + h(x\gamma, y\gamma) + f(x\gamma, y\gamma) + kf(y\gamma, x\gamma)^\sigma \\ &= h(x,y) + h(x\gamma, y\gamma) - h(x\gamma, y\gamma) = h(x,y) \end{aligned}$$

and

$$\begin{aligned} q(xg) &= q(x(1+\gamma)) = q(x + x\gamma) \\ &= q(x) + q(x\gamma) + h(x, x\gamma) + \Lambda \\ &= q(x) + q(x\gamma) + f(x\gamma, x\gamma) + \Lambda \\ &= q(x) + q(x\gamma) - q(x\gamma) + \Lambda = q(x)\,. \end{aligned}$$

The first calculation implies that $\ker g \leq V^\perp$. For $v \in \ker g$ we then have

$$-v = -v + vg = v\gamma \leq U \cap V^\perp = \{0\}\,.$$

Thus g is an isometry with $U = V(g-1)$ and, as anticipated, $f = h_g$.

If $a = 1 + \alpha$ is a second such isometry, then for $x, z \in V$ with $x\gamma = z\alpha$ we have, for all $v \in U$,

$$h(x, v) = h_g(x\gamma, v) = f(x\gamma, v) = f(z\alpha, v) = h_a(z\alpha, v) = h(z, v)\,.$$

Therefore $x - z \in U^\perp$. Thus, for all $v \in U$

$$0 = h(x - z, v) = h_g((x - z)\gamma, v)\,.$$

Since h_g is nondegenerate, $0 = (x - z)\gamma = x\gamma - z\gamma$; so $z\alpha = x\gamma = z\gamma$. That is, $\alpha = \gamma$ and $a = 1 + \alpha = 1 + \gamma = g$. □

The isometry g constructed in the Theorem is a *Wall isometry* and will be denoted $\mathrm{w}(U, f)$.

While the two propositions are valid for arbitrary isometries, the construction of the theorem is only given for finitary isometries. That is because in two places in the proof we have appealed to Corollary (10.3), valid only in finite dimensions. We have effectively used the fact that a nondegenerate pairing of finite dimensional spaces is isometric to the canonical hyperbolic pairing. For infinite dimensional spaces this is not in general true.

(11.4). PROPOSITION. *Let $g \in \mathrm{I}_D(V, q, h)$ with $U = [V, g]$ of finite dimension and meeting* $\mathrm{Rad}(V, h)$ *trivially. Set $f = h_g$ so that $g = \mathrm{w}(U, f)$, as in the previous theorem.*

Suppose $U = A \oplus B$ with $f(A, B) = 0$. (That is, $f(a, b) = 0$, for all $a \in A$ and $b \in B$.) Then A and B are nondegenerate under the appropriate restrictions of f, and with $a = \mathrm{w}(A, f|_A)$ and $b = \mathrm{w}(B, f|_B)$ we have $g = ba$.

PROOF.

□

(11.5). PROPOSITION. *Let $a, b \in \mathrm{I}_D(q, h)$ such that $a = \mathrm{w}(A, f_A)$ and $b = \mathrm{w}(B, f_B)$ with $A \cap B = \{0\}$. Set $U = A \oplus B$ and $ba = g \in \mathrm{I}_D(q, h)$. Then there is a nondegenerate $f \in \mathrm{Sesq}^\sigma(U)$ with $f(A, B) = 0$, $f_A = f|_A$, $f_B = f|_B$, and $g = \mathrm{w}(U, f)$.*

PROOF.

$$\mathrm{Gram}(f) = \begin{pmatrix} \mathrm{Gram}(f_B) & -h(b, a) \\ 0 & \mathrm{Gram}(f_A) \end{pmatrix}. \qquad \square$$

(11.6). COROLLARY. *$a = \mathrm{w}(A, f_A)$ and $b = \mathrm{w}(B, f_B)$ with $A \cap B = \{0\}$ commute if and only if in $U = [V, ba]$ we have $U = A \perp B$.* □

11.2 Isometries of small degree and root isometries

If (V, q, h) is a singular Λ-pseudoquadratic space for the (σ, k)-form parameter Λ in D, then it contains hyperbolic 2-spaces by Lemma (10.31) and so, by

Theorem (10.34), it possesses nontrivial isometries g of degree $\dim_D([V,g])$ at most 2.

The identity is the only isometry of degree 0. It is possible to describe all isometries g of relatively small degree by adopting an *ad hoc* approach. For instance for g of degree 1 with $[V,g] = \langle v\rangle$, we must have $C_V(g) = \ker\lambda$ for some $\lambda \in V^*$; so g has the general form $x \mapsto x + (x\lambda)av$ for some $a \in D$. The possibilities for λ and a can then be analyzed.

We take a different approach. For a given finite dimensional U, Wall's Theorem (11.3) allows us to construct all isometries g of (V,q,h) with $[V,g] = U$ as Wall isometries $\mathrm{w}(U,f)$ by considering all σ-sesquilinear forms on U. In practice there are many forms, but if the dimension of U, which is to say the degree of g, is small and the restriction of (q,h) to U reasonably simple, it can become manageable.

According to Theorem (10.19) the candidates for f consist of the coset $f_0 + \mathrm{Herm}_\Lambda^{(\sigma,-k)}(U)$, where f_0 has upper triangular Gram matrix

$$\begin{aligned}
f_0(x_i,x_j) &= -h(x_i,x_j) && \text{for } i<j\,,\\
f_0(x_i,x_i) &= -q_i && \text{for } i=j\,,\\
f_0(x_i,x_j) &= 0 && \text{for } j<i\,,
\end{aligned}$$

with $\{\, x_i \mid 1 \le i \le d\,\}$ a basis of U and the q_i representatives of $q(x_i)$. The only issue is which elements of the coset are nondegenerate. For instance, if any of the x_i are singular then for $q_i = 0$ the form f_0 itself is degenerate.

(11.7). Theorem. *Let $g \in \mathrm{I}_D(V,q,h)$ with $[V,g] = \langle v\rangle$. Then there is a $0 \neq b \in q(v)$ with $g = \mathrm{w}(v,b)$ given by*

$$\mathrm{w}(v,b)\colon x \mapsto x - h(x,v)b^{-1}v\,.$$

Correspondingly, for any nonzero $b \in q(v)$, the linear transformation $\mathrm{w}(v,b)$ is an isometry.

Proof.

□

Here we have abused terminology by writing $\mathrm{w}(v,b)$ for the degree 1 Wall isometry $\mathrm{w}(Dv, f(v,v) = -b)$.

If $h(v,v) = 0$ then $\mathrm{w}(v,b) - 1$ has rank 1 and $(\mathrm{w}(v,b)-1)^2 = 0$. Therefore $\mathrm{w}(v,b)$ is a transvection. If $h(v,v) \neq 0$, then $\mathrm{w}(v,b)$ is a *generalized reflection*, as standard reflections arise a special case.

If $q(v) = \Lambda$ (so that $h(v,v) = 0$) the transvection $\mathrm{w}(v,b)$ is a *root transvection.* If $q(v) \neq \Lambda$, then $\mathrm{w}(v,b)$ is a *symmetry.* (Note that in characteristic 2 it is possible for a symmetry to be a transvection.)

(11.8). Theorem. (Cartan's Generation Theorem) *Let (V,q,h) be a nondegenerate Λ-pseudoquadratic form for the (σ,k)-form parameter and the finite dimensional D-space V. Assume additionally that (V,q,h) is asingular. Then $\mathrm{I}_D(V,q,h)$ is generated by its symmetries. Indeed the isometry g of degree d is the product of exactly d symmetries.*

Proof.

□

Recall from Proposition (7.15) that, for $0 \neq x \in V$ and $0 \neq \lambda \in V^*$, the group $\mathrm{R}(\lambda, u)$ consists of the 2-root elements determined by λ and x: those elements s that stabilize the series $0 \neq \langle x\rangle \leq \ker\lambda < V$. (That is, $[V, s] \leq \ker\lambda$, $[U, s] \leq \langle x\rangle$, and $[x, s] = 0$.)

(11.9). Theorem.

(a) *We have* $r \in \mathrm{R}_x = \mathrm{R}(h(\cdot, x), x) \cap \mathrm{I}_D(V, q, h)$ *if and only if*

$$r = \mathrm{r}(u, v, b)\colon z \mapsto z + h(z, ku)v - h(z, v)u - h(z, ku)bu\,,$$

for u singular, $v \in u^\perp$, and $b \in q(v)$.

(b) *We have a singular transvection* $t \in \mathrm{T}_x = \mathrm{R}_x \cap \mathrm{T}(h(\cdot, x), x)$ *if and only if*

$$t = \mathrm{r}(u, 0, b)\colon z \mapsto z - h(z, ku)bu\,,$$

$b \in \Lambda$.

(c) *For* $g \in \mathrm{I}_D(V, q, h)$ *we have* $\mathrm{r}(u, v, b)^g = \mathrm{r}(ug, vg, b)$.

(d) $\mathrm{r}(u, v, b)\,\mathrm{r}(u, w, c) = \mathrm{r}(u, v + w, b + c + h(v, w))$.

(e) $\mathrm{r}(u, v, b)^{-1} = \mathrm{r}(u, -v, -b + h(v, v))$.

(f) $[\mathrm{r}(u, yv, b), \mathrm{r}(u, w, c)] = \mathrm{r}(u, 0, h(v, w) - kh(v, w)^\sigma)$.

(g) *For* $a \in D^\times$, $\mathrm{r}(a^\sigma u, v, b) = \mathrm{r}(u, av, a^\sigma ba)$.

(h) *For* $a \in D$, $\mathrm{r}(u, au + v, b) = \mathrm{r}(u, v, b - a + ka^\sigma)$.

Proof.

□

Note that $\langle u_1\rangle = \langle u_2\rangle$ implies $\mathrm{R}_{u_1} = \mathrm{R}_{u_2}$.

11.3 Witt's Theorem

(11.10). Theorem. (Witt's Theorem) *Let (V, q, h) be a nondegenerate Λ-pseudoquadratic form for the (σ, k)-form parameter Λ in D. Let W be a finite dimensional subspace of V and $s\colon W \longrightarrow V$ be a partial isometry of (V, q, h), in that s is injective and, for all $u, v \in W$, we have*

$$h(u, v) = h(us, vs) \quad \text{and} \quad q(v) = q(vs)\,.$$

Then there is an isometry $g \in \mathrm{I}_D(V, q, h)$ with $s = g|_W$.

Proof. The proof is by induction on $d = \dim_D(W)$, the result being true trivially for $d = 0$.

Let $W = \langle w\rangle \oplus W_1$. By induction there is a $g_1 \in \mathrm{I}_D(V, q, h)$ with $s|_{W_1} = g_1|_{W_1}$. Set $t = sg_1^{-1}$. Then for all $v \in W_1$ we have $vt = vsg_1^{-1} = v$. If t can be extended to an isometry g_2 of V, then s can be extended to the isometry g_2g_1. Therefore we may replace s with t. That is, without loss of generality we may assume $s|_{W_1}$ is the identity.

Set $v = -[w, s] = w(1 - s) = w - ws$ hence $ws = w - v$. For every $x \in W_1$ we have

$$h(x, v) = h(x, w) - h(x, ws) = h(x, w) - h(xs, ws) = h(x, w) - h(x, w) = 0\,.$$

That is, $v \in W_1^{\perp}$ and $v^{\perp} \geq W_1$. We also have

$$\begin{aligned} q(v) &= q(w - ws) = q(w) + q(-ws) + h(w, -ws) + \Lambda \\ &= q(w) + q(w) + (kq(w)^{\sigma} - kq(w)^{\sigma}) - h(w, ws) + \Lambda \\ &= (q(w) + kq(w)^{\sigma}) + (q(w) - kq(w)^{\sigma}) - h(w, ws) + \Lambda \\ &= h(w, w) - h(w, ws) + \Lambda = h(w, v) + \Lambda\,. \end{aligned}$$

Especially $h(w, v) \in q(v)$. We consider in turn the two cases $0 \neq h(w, v)$ and $0 = h(w, v)$.

First assume $0 \neq h(w, v) = r \in q(v)$. Then $g = \mathrm{w}(v, r) \in \mathrm{I}_D(V, q, h)$ and

$$\begin{aligned} wg &= w\,\mathrm{w}(v, r) = w - h(w, v)r^{-1}v \\ &= w - h(w, v)h(w, v)^{-1}v = w - v = ws\,. \end{aligned}$$

Also for all $w_1 \in W_1 \leq v^{\perp}$ we find $w_1 g = w_1\,\mathrm{w}(v, r) = w_1$, so that $g|_W = s$, as desired.

We may now assume $0 = h(w, v) \in q(v)$. In particular, v is singular and $h(v, v) = 0$. As $w \notin W_1 = (W_1^{\perp})^{\perp}$ also $v - w = ws \notin W_1 s = W_1 = (W_1^{\perp})^{\perp}$. Thus $W_1^{\perp}$ is spanned by its hyperplane complement $W_1^{\perp} \setminus W_1^{\perp} \cap (v - w)^{\perp}$ and this complement is not contained in $w^{\perp}$. Therefore we may choose a $z \in W_1^{\perp}$ with $h(v, z) \neq h(w, z) \neq 0$. Additionally choose a representative $r \in q(z)$.

Let U be the 2-space $\langle v, z\rangle$, and on U let f be the the σ-sesquilinear form with Gram matrix

$$G = \begin{pmatrix} f(v, v) & f(v, z) \\ f(z, v) & f(z, z) \end{pmatrix} = \begin{pmatrix} 0 & -h(v, z) \\ 0 & -r \end{pmatrix} + \begin{pmatrix} 0 & h(w, z) \\ -kh(w, z)^{\sigma} & 0 \end{pmatrix}.$$

Since v is singular, the first summand has the correct upper triangular shape, as in Theorem (10.12). The second summand is the Gram matrix of a form in $\mathrm{Herm}_{\Lambda}^{(\sigma,-k)}(U)$. As $h(v, z) \neq h(w, z) \neq 0$, the two off-diagonal entries of G are nonzero; hence the form f is nondegenerate.

Therefore

$$G + k(G^{\top})^{\sigma} = \begin{pmatrix} -h(v, v) & -h(v, z) \\ -h(z, v) & -h(z, z) \end{pmatrix};$$

and the nondegenerate $f \in \mathrm{Sesq}^{\sigma}(U)$ satisfies, for all $u, y \in U$,

$$f(u, y) + kf(y, u)^{\sigma} = -h(u, y) \quad \text{and} \quad f(u, u) + \Lambda = -q(u)\,.$$

Theorem (11.3) then provides an isometry $a = \mathrm{w}(U, f)$.

Consider the action of $a = 1 + \alpha$, as given in that theorem: for every $x \in V$ the vector $x\alpha$ is characterized by

$$h(x, y) = f(x\alpha, y) \text{ for all } y \in U\,.$$

Equivalently

$$h(x, v) = f(x\alpha, v) \quad \text{and} \quad h(x, z) = f(x\alpha, z)\,.$$

In particular, as $v, z \in W_1^\perp$, for every $x \in W_1$ we must have $x\alpha = 0$. For $x = ws$ we claim $x\alpha = v$. Indeed

$$h(ws, v) = h(w - v, v) = h(w, v) - h(v, v) = 0 - 0 = 0 = f(v, v)\,,$$

as we are in the case $h(w, v) = 0$ and v is singular, and

$$h(ws, z) = h(w - v, z) = h(w, z) - h(v, z) = f(v, z)\,,$$

as we find in the Gram matrix G.

Therfore, for all $w_1 \in W_1$, we have $w_1 a = w_1(1 + \alpha) = w_1$; and

$$(ws)a = ws(1 + \alpha) = ws + v = (w - v) + v = w\,.$$

That is, with $g = a^{-1} \in \mathrm{I}_D(V, q, h)$, we have $g|_W = s$, as desired. □

(11.11). Corollary. (Witt Cancellation) *Let (V, q, h) be a nondegenerate Λ-pseudoquadratic form for the (σ, k)-form parameter Λ in D. Suppose that the finite dimensional subspaces W and U are isometric. Then $W^\perp$ and $U^\perp$ are also isometric.*

Proof. The assumption is that there is a partial isometry $s\colon W \longrightarrow U$. By Witt's Theorem, there is an isometry of (V, q, h) extending s to all of V. But then $(W^\perp)g = U^\perp$, and $W^\perp$ and $U^\perp$ are isometric. □

The name of the corollary comes from its frequent application in the special case where W and U are nondegenerate themselves. The corollary then has the form of a cancellation result:

(11.12). Corollary. (Witt Cancellation) *If W and U are isometric and nondegenerate $W \perp A = V = U \perp B$ then A and B are isometric.* □

11.4 Canonical forms

A subspace of a totally singular subspace is totally singular. A nondegenerate subspace of a hyperbolic space is hyperbolic, and a subspace of an asingular space is asingular. In an arbitrary pseudoquadratic space Zorn's Lemma guarantees maximal such subspaces, and those of finite dimension are of particular interest.

(11.13). Theorem. (Witt Decomposition) *Let (V, q, h) be a nondegenerate Λ-pseudoquadratic form for the (σ, k)-form parameter Λ in D. Assume that some maximal totally singular subspace M of V has finite dimension.*

(a) *All maximal singular subspaces are equivalent to M under $\mathrm{I}_D(V, q, h)$. Especially, they all have the same finite dimension.*

(b) *For every maximal singular subspace N there is a second maximal totally singular subspace P with $H = N \oplus P$ maximal hyperbolic in V. All maximal hyperbolic subspaces are equivalent to H under $\mathrm{I}_D(V, q, h)$.*

(c) *$H^\perp$ is a nondegenerate asingular subspace of V. If J is a hyperbolic subspace of V with $J^\perp$ asingular, then J is equivalent to H and $J^\perp$ is equivalent to $H^\perp$ under $\mathrm{I}_D(V, q, h)$.*

Proof.

□

The common dimension of finite dimensional maximal totally singular spaces is the *Witt index*. Finite dimensional spaces (V, q, h) of course have a well-defined Witt index. As we have seen in Problem (10.37) this need not be the case for infinite dimension spaces; there can be maximal totally singular subspaces of different infinite dimension.

(11.14). Corollary. *A finite dimensional nondegenerate symplectic space is hyperbolic. In particular, it has even dimension.*

Proof.

□

In particular, a nondegenerate symplectic form on a finite dimensional space is uniquely determined up to isometry by the field and its dimension, which must be even. In this case we write $\mathrm{I}_F(V, h) = \mathrm{Sp}_F(V, h) = \mathrm{Sp}_{2m}(F)$ for the *symplectic group* of the symplectic pseudoquadratic space $(V, h) = (V, q, h)$ in which the quadratic part q is trivial since $\Lambda = D$ is a field.

As a consequence of the corollary, symplectic spaces and groups are, in a sense, the most accessible of the classical geometries and groups. They in fact exhibit most of the behaviour of all the classical groups, often in a form that is simplified but nontrivial. That makes them ideal to keep in mind (or study) as the basic models for classical groups.

11.5 Finite isometry groups

We consider nondegenerate pseudoquadratic forms on over finite fields and in finite dimension. We shall discover that the field and dimension alone come close to determining the form uniquely and so also the isometry group.

The cases from Theorem (10.22) to consider are:

(i) Symplectic Case: $\Lambda = D$; $k = -1$; $\sigma = 1$; D is a field; q is identically 0; and the symplectic bilinear form h has $h(x, x) = 0$ for all $x \in V$.

(ii) Generic Orthogonal Case: $\Lambda = 0$; $k = 1$; $\sigma = 1$; D is a field; and there are $q \colon V \longrightarrow D$ and symmetric bilinear form $h \colon V \times V \longrightarrow D$ satisfying:

  (i) $q(x+y) - q(x) - q(y) = h(x,y)$,
  (ii) $q(ax) = a^2 q(x)$,

  for all $x, y \in V$ and $a \in D$.

(iii) Generic Unitary Case: $1 \in \Lambda = D^{(\sigma,1)} < D$; $k = -1$; $\sigma^2 = 1 \neq \sigma$; and h is a σ-skew-hermitian form.

For a finite division ring D, the two exceptional cases in Theorem (10.22) do not occur. For the Exceptional Unitary Case, the division ring D is not a field. But Wedderburn's Theorem [Hll59, Theorem 20.6.1] tells us that all finite division rings are fields, so this case does not arise here.

In the Exceptional Orthogonal Case the field D has characteristic 2, $\sigma = 1$, and $1 \in \Lambda < D^{(\sigma,1)} = D$. As $1 \in \Lambda$, for each $a \in D$ we have $a1a^\sigma = a^2 \in \Lambda$. That is, $D^2 \leq \Lambda$. But here D is a finite field of characteristic 2 hence perfect: $D^2 = D$. The contradiction shows that this case does not occur either.

11.5.1 Finite symplectic groups

As an immediate consequence of Corollary (11.14) we have:

(11.15). Theorem. *Let F be a finite field and V a finite dimensional vector space over F. Then up to isometry there is a unique nondegenerate symplectic form h on V when $\dim_F(V) = 2m$ is even and no nondegenerate symplectic form on V when $\dim_F(V)$ is odd.* □

In this case we often write $\mathrm{Sp}_{2m}(r)$ in place of $\mathrm{Sp}_{2m}(F)$ where $F = \mathbb{F}_r$.

(11.16). Proposition. $\mathrm{Sp}_2(F) \simeq \mathrm{SL}_2(F)$.

11.5.2 Finite unitary groups

As already mentioned, over finite fields only the generic case arises, so we drop that term. Also as σ is an order 2 automorphism of the finite field F, we must have $F = \mathbb{F}_{r^2}$ for some prime power r with σ the *Frobenius automorphism* $\alpha \longrightarrow \alpha^r$.

(11.17). Lemma. *Let (V, q, h) be a nondegenerate Λ-pseudoquadratic form in the generic unitary case over a field D. Then, there is a nonzero $t \in D$ such that, for every $x \in V$ we have $q(x) = th(x,x) + \Lambda$. In particular x is singular if and only if it is isotropic and $\mathrm{I}_D(V, q, h) = \mathrm{I}_D(V, h)$.*

Proof.

□

(11.18). Lemma. *If h is a nondegenerate $(\sigma, \pm 1)$-hermitian form on the 1-space $V = Dv$ with $h(v,v) = a \neq 0$, then $\mathrm{GU}_D(V, q, h) = \{d \in D \mid d^a d^\sigma = 1\}$. In particular $\mathrm{GU}_1(r^2) \simeq Z_{r+1}$.*

Proof.

□

(11.19). Lemma. *If h is a nondegenerate skew-hermitian form on the 2-space V over the finite field $\mathbb{F}_{r^2}$, then V is hyperbolic.*

Proof.

□

(11.20). Corollary. *Let F be a finite field and h a skew-hermitian form on the vector space V over F of dimension n. If n is even, then h is hyperbolic (Witt index $n/2$). If n is odd, then its Witt index $(n-1)/2$ and h is unique up to scaling.*

Proof.

□

In particular the *unitary group*

$$\mathrm{I}_F(V,q,h) = \mathrm{I}_F(V,h) = \mathrm{GU}_F(V,h) = \mathrm{GU}_n(F) = \mathrm{GU}_n(r^2)$$

is uniquely determined up to isomorphism by the dimension n and the field order r^2.

11.5.3 Finite orthogonal groups

(11.21). Lemma. *Let (V,q) be a non-degenerate orthogonal space of dimension 2 over F. Then either*

(a) *V is hyperbolic or*

(b) *for some nonzero constant $a \in F$ and some quadratic extension K of F, (V,q) is isometric to K (as F-space) provided with the quadratic form $q_K(\alpha) = a\alpha\bar{\alpha}$, where the bar denotes Galois conjugation in K over F.*

Proof.

□

(11.22). Corollary. *Let (V,q) be a nondegenerate orthogonal space of dimension 2 over the finite field $\mathbb{F}_r$. Then for every nonzero element b of $\mathbb{F}_r$ there is a $v \in V$ with $q(v) = b$. If (V,q) is not hyperbolic, then for every $0 \neq b \in \mathbb{F}_r$ there are exactly $r+1$ vectors y in V with $q(y) = b$.*

Proof. If V is hyperbolic, this follows from Lemma (10.31). In the asingular case, we have $V = \mathbb{F}_{r^2}$ with quadratic form $q(\alpha) = a\alpha^{1+r}$ for nonzero $a \in \mathbb{F}_q$. The map $\alpha \longrightarrow \alpha^{r+1}$ is a surjective homomorphism from the cyclic multiplicative subgroup of $\mathbb{F}_{r^2}$ of order r^2-1 to that of its subfield $\mathbb{F}_r$. The image thus has order $r-1$ and the kernel order $r+1$. □

(11.23). Lemma. *Let (V,q) be a nondegenerate orthogonal space of dimension 3 over the finite field $\mathbb{F}_r$. Then V contains singular vectors.*

Proof. Let $0 \neq u \in V$ and $v \in V \setminus u^\perp$. Then $U = \langle u, v\rangle$ is nondegenerate of dimension 2 with $V = U \perp W$ for some $W = \mathbb{F}_r w$, where w is nonsingular as V is nondegenerate. By the previous corollary $-q(w) = q(z)$ for some $z \in U$. But then $z + w \neq 0$ and $q(z + w) = q(z) + q(w) + h(z, w) = -q(z) + q(z) = 0$, as desired. □

11.5.4 Orders of finite isometry groups

In this section we prove:

(11.24). Theorem.

(a) $|\mathrm{Sp}_{2k}(r)| = r^{k^2} \prod_{i=1}^{k}(r^{2i} - 1)$.

(b) $|\mathrm{GU}_n(r^2)| = r^{\binom{n}{2}} \prod_{i=1}^{n}(r^i - (-1)^i)$

(c) $|\mathrm{GO}_{2k+1}(r)| = 2\, d^{-1} r^{k^2} \prod_{i=1}^{k}(r^{2i} - 1)$, *where* $d = \gcd(2, r)$.

(d) $|\mathrm{GO}^{\epsilon}_{2k}(r)| = 2\, r^{k(k-1)}(r^k - \epsilon) \prod_{i=1}^{k-1}(r^{2i} - 1)$.

It is convenient to see all of these groups as specializations of a single isometry group $G = \mathrm{Cl}^{[t]}_n(s)$, which acts on $V = V_n = V^{[t]}_n = \mathbb{F}^n_s$ with trivial singular radical. The relevant parameters are:

t	$\mathrm{Cl}^{[t]}_n(s)$	s	m	κ	g
1	$\mathrm{Sp}_{2k}(r)$	r	k	1	1
2	$\mathrm{GU}_{2k}(r^2)$	r^2	k	r^{-1}	1
3	$\mathrm{GU}_{2k+1}(r^2)$	r^2	k	r	$r+1$
4	$\mathrm{GO}_{2k+1}(r)$	r	k	1	$2d^{-1}$
5	$\mathrm{GO}^+_{2k}(r)$	r	k	r^{-1}	1
6	$\mathrm{GO}^-_{2k}(r)$	r	$k-1$	r	$2(r+1)$

The field of definition is $\mathbb{F}_s$. The dimension n of V is $2k$ or $2k+1$, as appropriate, and m is the Witt index of the corresponding form. Set $|\Lambda| = \lambda$. The parameter κ is then λs^{n-2m-1} and first appears in Lemma (11.28) below.

The Witt Decomposition Theorem (11.13) gives us, for $0 \leq i \leq m$,

$$V^{[t]}_n = H_{2i} \perp V^{[t]}_{n-2i}\,,$$

where H_{2i} is a hyperbolic subspace of Witt index i and the space $V^{[t]}_{n-2i}$ has the same type t as V. In particular $A = V^{[t]}_{n-2m}$ is asingular and nonsingular of dimension $n - 2m \in \{0, 1, 2\}$. The parameter $g = g^{[t]}_{n-2m}$ of the table is defined to be the order of the corresponding asingular isometry group $|\,\mathrm{I}_{\mathbb{F}_s}(A)|$. Of course in the hyperbolic cases $t \in \{1, 2, 5\}$ we have $n - 2m = 0$, $A = \{0\}$, and $g = 1$. The remaining values of g are verified in the next lemma.

(11.25). Lemma.

(a) $g^{[3]}_1 = |\mathrm{GU}_1(r^2)| = r + 1$.

(b) $g_1^{[4]} = |\mathrm{GO}_1(r)| = 2\,d^{-1}$ *for* $d = \gcd(2, r)$*,*

(c) $g_2^{[6]} = |\mathrm{GO}_2^-(r)| = 2(r+1)$*.*

Proof. For the first two parts, think in terms of the corresponding 1×1 Gram matrix $G = (g)$ with $g = kg^\sigma \in \mathbb{F}_s$, where $s = r^2$ in the first part and $s = r$ in the second. The isometry group then consists of all $A = (a) \in \mathrm{GL}_1(s)$ with $AGA^{\top\sigma} = G$. That is, $aga^\sigma = g$. Since $\mathbb{F}_s$ is commutative and g is invertible, this amounts to $a^{\sigma+1} = 1$. In the first case, this is $a^{r+1} = 1$, which has $r+1$ solutions in $\mathbb{F}_{r^2}$; and in the second case this is $a^2 = 1$ which has one solution $a = 1$ if $\mathbb{F}_r$ has characteristic 2 and otherwise has the two solutions $a = \pm 1$.

Now consider $g_2^{[6]} = |\mathrm{GO}_2^-(r)|$. For $0 \neq v \in V$, there are exactly $r+1$ vectors $y \in V$ with $0 \neq b = q(v) = q(y)$ by Corollary (11.22). By Witt's Theorem (11.10) $\mathrm{GO}_2^-(r)$ is transitive on these $r+1$ vectors. By Theorem (11.7) the stabilizer of v is a group of order 2, being the identity and the unique degree 1 element $\mathrm{w}(v,b)\colon x \mapsto x - h(x,v)b^{-1}v$. □

Consider a fixed but arbitrary case t. As above, let $G = \mathrm{Cl}_n^{[t]}(s)$ and $V = V_n = V_n^{[t]} = \mathbb{F}_s^n$. Set

$$\begin{aligned}
g_n &= g_n^{[t]} = |\,\mathrm{Cl}_n^{[t]}(s)|\,;\\
i_n &= i_n^{[t]} = |\{\,\langle x\rangle \in V \mid \langle x\rangle \text{ singular}\,\}|\,;\\
j_n &= j_n^{[t]} = (s-1)i_n = |\{\,0 \neq x \in V \mid x \text{ singular}\,\}|\,;\\
h_n &= h_n^{[t]} = |\{\,(x,y) \in V^2 \mid (x,y) \text{ a hyperbolic pair}\,\}|\,.
\end{aligned}$$

(11.26). Lemma. *The group G is transitive on hyperbolic pairs. Thus if V is singular (that is, $m > 0$) then G has order $g_n = h_n g_{n-2}$.*

Proof. If H is the hyperbolic 2-space spanned by the hyperbolic pair (x,y), then Witt Decomposition (11.13) gives

$$V_n^{[t]} = H \perp V_{n-2}^{[t]}\,.$$

Transitivity on hyperbolic pairs is immediate from Witt's Theorem (11.10). The stabilizer of the pair (x,y) in this transitive action then induces the identity on $H = \langle x,y\rangle$ but can induce any isometry of $V_{n-2}^{[t]}$. Therefore

$$g_n = g_n^{[t]} = h_n^{[t]} g_{n-2}^{[t]} = h_n g_{n-2}\,. \qquad \square$$

(11.27). Lemma. *Assume $V = V_n$ is singular: $m > 0$.*

(a) $i_n = 1 + \lambda s^{n-2} + s i_{n-2}$.

(b) $j_n = (s-1)i_n = s - 1 + (s-1)\lambda s^{n-2} + s j_{n-2}$.

(c) $h_n = \lambda s^{n-2} j_n$.

Proof.

□

(11.28). Lemma. *Set $\kappa = \lambda s^{n-2m-1}$. For $m \geq 0$, $j_n = (s^m - 1)(\kappa s^m + 1)$.*

Proof.

□

(11.29). Theorem. $g_n = g\,\kappa^m s^{m^2} \prod_{j=1}^{m} (s^j - 1)(\kappa s^j + 1)$.

Proof. We have:

$$
\begin{aligned}
g_n &= g \prod_{i=0}^{m-1} h_{n-2i} && \text{by Lemma (11.26) and induction} \\
&= g \prod_{i=0}^{m-1} \lambda\, s^{n-2-2i}(s^{m-i} - 1)(\kappa s^{m-i} + 1) && \text{by Lemmas (11.27)(c) and (11.28)} \\
&= g\,\lambda^m s^{m(n-2m-1+2m-1)} \prod_{i=0}^{m-1} (s^{-2})^i \prod_{j=1}^{m} (s^j - 1)(\kappa s^j + 1) && j = m - i \\
&= g\,\kappa^m s^{m(2m-1)} s^{-m(m-1)} \prod_{j=1}^{m} (s^j - 1)(\kappa s^j + 1) && \kappa = \lambda s^{n-2m-1} \\
&= g\,\kappa^m s^{m^2} \prod_{j=1}^{m} (s^j - 1)(\kappa s^j + 1)\,. && \square
\end{aligned}
$$

Proof of Theorem (11.24).

We break the proof into cases—three pairs of like groups.

Case 1. $t \in \{1, 4\}$: $\mathrm{Sp}_{2k}(r)$ and $\mathrm{GO}_{2k+1}(r)$; $s = r$, $m = k$, $\kappa = 1$.

$$
\begin{aligned}
g_n &= g\kappa^m s^{m^2} \prod_{j=1}^{m} (s^j - 1)(\kappa s^j + 1) \\
&= g(1^m) r^{k^2} \prod_{j=1}^{k} (r^j - 1)(1r^j + 1) \\
&= g\, r^{k^2} \prod_{j=1}^{k} (r^{2j} - 1)\,.
\end{aligned}
$$

with $g = 1$ when $t = 1$ and $g = 2d^{-1}$ when $t = 4$.

Case 2. $t \in \{2, 3\}$: $\mathrm{GU}_n(r^2)$; $s = r^2$, $2m + \delta = n$ for $\delta \in \{0, 1\}$, $\kappa = r^{2\delta - 1}$,

$g = (r+1)^{\delta}$.

$$\begin{aligned} g_n &= g\kappa^m s^{m^2} \prod_{j=1}^{m} (s^j - 1)(\kappa s^j + 1) \\ &= (r+1)^{\delta} (r^{2\delta-1})^m r^{2m^2} \prod_{j=1}^{m} (r^{2j} - 1)(r^{2j-1+2\delta} + 1) \\ &= r^{2m^2+2\delta m - m} (r+1)^{\delta} \prod_{i=\delta+1}^{n} (r^i - (-1)^i) \\ &= r^{\binom{n}{2}} \prod_{i=1}^{n} (r^i - (-1)^i)\,. \end{aligned}$$

Case 3. *$t \in \{5,6\}$: $\mathrm{GO}_{2k}^{\epsilon}(r)$; $s = r$, $m = k - \delta$ for $\delta \in \{0,1\}$, $\kappa = r^{2\delta-1}$, $g = (2(r+1))^{\delta}$, $\epsilon = 1 - 2\delta$.*

$$\begin{aligned} g_n &= g\kappa^m s^{m^2} \prod_{j=1}^{m} (s^j - 1)(\kappa s^j + 1) \\ &= (2(r+1))^{\delta} (r^{2\delta-1})^{k-\delta} r^{(k-\delta)^2} \prod_{j=1}^{k-\delta} (r^j - 1)(r^{j+2\delta-1} + 1) \\ &= 2^{\delta} (r+1)^{\delta} r^{k^2-k-\delta^2+\delta} \prod_{j=1}^{k-\delta} (r^j - 1)(r^{j+2\delta-1} + 1) \\ &= 2^{\delta} r^{k(k-1)} (r^k - 1)^{1-\delta} \left(\prod_{i=1}^{k-1} (r^i - 1)(r^i + 1) \right) (r^0 + 1)^{1-\delta} (r^k + 1)^{\delta} \\ &= 2\, r^{k(k-1)} (r^k - \epsilon) \prod_{i=1}^{k-1} (r^{2i} - 1)\,. \qquad \Box \end{aligned}$$

Chapter 12

Root generation

This chapter is primarily devoted to a proof of the following theorem.

(12.1). THEOREM. *Let (V,q,h) be a nondegenerate Λ-pseudoquadratic space of positive Witt index for the (σ,k)-form parameter Λ in the division ring D. Let G be the subgroup $\mathrm{RI}_D(V,q,h)$ of $\mathrm{I}_D(V,q,h)$ generated by its root isometries. Then G is perfect and $G/\mathrm{Z}(G)$ is simple, except in the following cases of $\mathrm{I}_D(V,q,h)$:*

$$\mathrm{Sp}_2(2), \mathrm{Sp}_2(3), \mathrm{Sp}_4(2), \mathrm{GU}_2(2^2), \mathrm{GU}_2(3^2), \mathrm{GU}_3(2^2), \mathrm{GO}_3(3), \mathrm{GO}_2^+(F), \mathrm{GO}_4^+(F)\,.$$

All of these are genuine exceptions.

12.1 Problem children

12.1.1 Kids behaving badly

(12.2). LEMMA.

(a) $|\mathrm{GU}_3(2^2)| = 2^3 3^4$. *Especially* $\mathrm{GU}_3(2^2)$ *is solvable.*

(b) $\mathrm{Sp}_4(2) \simeq \mathrm{Sym}(6)$. *Especially* $\mathrm{Sp}_4(2)$ *is not perfect.*

PROOF. By Theorem (11.24) $|\mathrm{GU}_3(2^2)| = 2^3 3^4$ and $|\mathrm{Sp}_4(2)| = 6!$.

In particular $\mathrm{GU}_3(2^2)$ is solvable by Problem (3.20).

We saw earlier that the action of $\mathrm{Sym}(2m)$ on its natural module gives an embedding of $\mathrm{Sym}(2m)$ into $\mathrm{Sp}_{2m-2}(2)$. The orders match for $m = 3$, so this is an isomorphism in that case. □

12.1.2 Hyperbolic 2-spaces

(12.3). PROPOSITION. *Let (V,q,h) be a Λ-pseudoquadratic hyperbolic 2-space for the (σ,k)-form parameter Λ in the division ring D. Let (x,y) be a hyperbolic pair in V.*

(a) *For $a \in D$, we have $q(ax+y) = a$. In particular, there are $1 + |\Lambda|$ singular 1-spaces in V, being those spanned by x and by $ax+y$ for $a \in \Lambda$. The hyperbolic pairs (x,z) are precisely those with $z = ax+y$ for $a \in \Lambda$.*

(b) *The hyperbolic 2-space (V,q,h) is unique up to isometry. Its isometry group is regular on hyperbolic pairs and, especially, is 2-transitive on its singular 1-spaces. Indeed $\mathrm{RI}(V,q)$ is 2-transitive on its singular 1-spaces.*

(c) *When we write matrices with respect to the hyperbolic basis (x,y), the stabilizer J in $\mathrm{I}(V,q,h)$ of the unordered pair $\langle x\rangle, \langle y\rangle$ of singular 1-spaces is*

$$\left\{ \begin{pmatrix} d & 0 \\ 0 & d^{-\sigma} \end{pmatrix} \mid d \in D^\times \right\} \rtimes \left\langle \begin{pmatrix} 0 & k \\ 1 & 0 \end{pmatrix} \right\rangle$$

(d) *Assume (V,q,h) is generic orthogonal; that is, $\Lambda = 0$ hence $k=1$ and $\sigma = 1$. Then $\mathrm{GO}_2^+(F) = \mathrm{I}(V,q) = \mathrm{I}(V,q,h) = J$ (as above). Every element of $\mathrm{GO}_2^+(F)$ that is not a symmetry is a product of two symmetries, and $\mathrm{RGO}_2^+(F) = \mathrm{RI}(V,q,h) = 1$.*

(e) *Assume $\Lambda \neq \{0\}$ so that (following Theorem (10.22)) the hyperbolic form h can be taken to be σ-skew-hermitian. Then $\mathrm{RI}(V,q,h) = \mathrm{RI}(V,h)$ contains $\mathrm{SL}_2(F_0)$, where F_0 is any subfield of D contained in Λ (an additive subgroup of $D^{(\sigma,1)} = E$, the additive subgroup of D fixed by σ.) In particular, if $\Lambda = E$ then $\mathrm{RI}(V,q) \simeq \mathrm{SL}_2(E)$.*

Furthermore, the kernel of the action of $\mathrm{I}(V,q,h)$ of the set of singular 1-spaces is the subgroup of central scalars.

Proof.

□

(12.4). Proposition. $\mathrm{Sp}_2(F) = \mathrm{SL}_2(F)$.

Proof.

□

12.1.3 Orthogonal groups of small dimension

(12.5). Proposition. *Assume (V,q,h) is generic orthogonal; that is, $\Lambda = 0$ hence $k=1$ and $\sigma=1$. Then $\mathrm{GO}_2^+(F) = \mathrm{I}(V,q) = \mathrm{I}(V,q,h) = J$, as in Proposition (12.3). Every element of $\mathrm{GO}_2^+(F)$ that is not a symmetry is a product of two symmetries, and $\mathrm{RGO}_2^+(F) = \mathrm{RI}(V,q,h) = 1$.*

Proof.

□

(12.6). Proposition. *Let (V,q) be a non-degenerate orthogonal space that is hyperbolic of dimension 4 over the field F.*

(a) *Up to proportionality, we have*

$$q((w,x,y,z)) = wz - xy = \det \begin{pmatrix} w & x \\ y & z \end{pmatrix}.$$

(b) *Identify V with* $\mathrm{Mat}_2(F)$ *provided with the determinant form, as above. Then* $(\mathrm{GL}_2(F) \times \mathrm{GL}_2(F))\langle t \rangle$ *with action given by* $X.(A,B) = A^\top XB$ *and* $X.t = X^\top$ *induces the full group of similarities of the form. The full orthogonal isometry group is induced by those* (A,B) *and* $(A,B)t$ *with* $\det A = \det B^{-1}$. *The kernel of the action consists of the scalar pairs* $(rI, r^{-1}I)$.

(c) $\mathrm{RGO}_4^+(F) \simeq \mathrm{SL}_2(F) * \mathrm{SL}_2(F)$.

PROOF.

□

(12.7). COROLLARY. *Let (V,q) be a non-degenerate orthogonal space of Witt index* 1 *and dimension* 3 *over the field F.*

(a) *Up to proportionality, we have*

$$q((w,y,z)) = wz - y^2 = \det \begin{pmatrix} w & y \\ y & z \end{pmatrix}.$$

(b) $\mathrm{RGO}_3(F) \simeq \mathrm{PSL}_2(F)$ *acting via* $X.A = A^\top XA$, *for* $A \in \mathrm{SL}_2(F)$.

(c) *In* $\mathrm{GO}_3(F)$ *the kernel of the permutation action on the set of singular* 1*-spaces is the central scalar subgroup* $\langle \pm I \rangle$.

(d) $\mathrm{GO}_3(3) \simeq 2 \times \mathrm{Sym}(4)$. *Especially* $\mathrm{GO}_3(3)$ *is solvable.*

PROOF. After we locate $\mathrm{GO}_3(F)$ as a subgroup of $\mathrm{GO}_4^+(F)$, the first two parts follow from the proposition.

By Lemma (11.28) (or a direct calculation) there are four singular 1-spaces in $V = \mathbb{F}_3^3$, and the kernel for this action is the isometry $-I$ of determinant -1. Finally by Theorem (11.24) $|\mathrm{GO}_3(3)| = 2 \cdot 4! = |2 \times \mathrm{Sym}(4)|$, so we have an isomorphism. □

(12.8). LEMMA. *Let (V,q) be a non-degenerate orthogonal space of dimension* 2 *over the field F. Then either*

(a) *V is hyperbolic or*

(b) *for some nonzero constant $a \in F$ and some quadratic extension K of F, (V,q) is isometric to K (as F-space) provided with the quadratic form* $q_K(\alpha) = a\alpha\bar{\alpha}$, *where the bar denotes Galois conjugation in K over F.*

PROOF.

□

(12.9). PROPOSITION. *Let (V,q) be a non-degenerate orthogonal space of Witt index 1 and dimension 4 over the field F.*

(a) *Up to proportionality, we have*

$$q((w,x,y,z)) = wz - (x-\gamma y)(x-\bar{\gamma} y) = \det \begin{pmatrix} w & x-\gamma y \\ x-\bar{\gamma} y & z \end{pmatrix},$$

where $K = F(\gamma)$ is a quadratic extension of F and $\bar{\gamma}$ is the Galois conjugate of γ over F.

(b) $\mathrm{RGO}_4^-(F) = \mathrm{RI}(V,q) \simeq \mathrm{PSL}_2(K)$ *acting via $X.A = \bar{A}^\top XA$ for $A \in \mathrm{SL}_2(K)$.*

PROOF.

□

12.2 Iwasawa

Following Theorem (12.1), throughout we let (V,q,h) be a nondegenerate Λ-pseudoquadratic space of positive Witt index for the (σ,k)-form parameter Λ in the division ring D. Let G be the subgroup $\mathrm{RI}_D(V,q,h)$ of $I = \mathrm{I}_D(V,q,h)$ generated by its root isometries.

For any subset U of V, denote $\Omega \cap U$ by U_Ω.

The (nonempty) set of singular 1-spaces will be denoted Ω. We will prove Theorem (12.1) by applying the Brey-Wilson version of Iwasawa's Lemma (3.15) to the action of G on Ω.

(12.10). LEMMA. *$V = \langle u\rangle + \langle \Omega \setminus u^\perp\rangle$ for each $\langle u\rangle \in \Omega$. Especially $V = \langle \Omega\rangle$. If V is not of type $\mathrm{GO}_2^+(F)$, then $V = [V,G]$.*

PROOF.

□

(12.11). PROPOSITION. *For $\langle u\rangle \in \Omega$, the group R_u is nilpotent of class at most 2 and normal in $G_{\langle u\rangle}$.*

PROOF. This follows immediately from Propositions (7.15) and Theorem (11.9). □

(12.12). PROPOSITION.

(a) R_u *is regular on $\Omega \setminus u^\perp$.*

(b) T_u *is regular on $\emptyset \neq \langle u,x\rangle_\Omega \setminus \langle u\rangle$ for each $x \in V \setminus u^\perp$.*

(c) R_u *is transitive on $\langle u,v\rangle_\Omega \setminus \langle u\rangle$ for each $v \in u_\Omega^\perp$.*

Proof. Parts (b) and (c) are immediate from Theorem (11.9).

(a) For any $z \notin u^\perp$ and any $1 \neq r = \mathrm{r}(u,v,b) \in \mathrm{R}_u$ we have $[z,r] = zr - z \in \langle u,v\rangle \setminus \langle u\rangle$ when $v \neq 0$ and $0 \neq [z,r] = -h(z,ku)bu$ when $v = 0$. Thus the stabilizer of $\langle z\rangle$ in R_u is 1.

In view of (b), it remains to prove transitivity of R_u on the set of hyperbolic 2-spaces $\langle u,x\rangle$. Let $\langle u,z\rangle$ and $\langle u,w\rangle$ be two such spaces with z and w singular and such that $h(z,ku) = 1$. As $u^\perp$ is a hyperplane, within the 2-space $\langle z,w\rangle$ there is an α such that the vector $v = -z+\alpha w$ is in $u^\perp$. Then for $s = \mathrm{r}(u,v,b) \in \mathrm{R}_u$

$$zs \in z + h(z,ku)v + \langle u\rangle = z + v + \langle u\rangle = \alpha w + \langle u\rangle \subseteq \langle u,w\rangle \setminus \langle u\rangle\,.$$

Therefore $\langle u,z\rangle^s = \langle u,w\rangle$, as desired. □

(12.13). Proposition. *Let $H = \mathrm{I}_D(V,q,h)_{u,v}$, for $\{u,v\}$ a hyperbolic pair. Then $H = \mathrm{I}_D(W,q|_W,h|_W)$ for $W = \langle u,v\rangle^\perp$. If $z \mapsto \bar z$ is the projection map from $u^\perp$ onto W, then $\mathrm{r}(u,z,a) \mapsto \bar z$ gives a $\mathbb{Z}H$-module isomorphism between $\mathrm{R}_u\,/\,\mathrm{T}_u$ and W.*

Proof. This is immediate from Theorem (11.9). □

(12.14). Corollary. *Assume V has Witt index bigger than 1 but is not of type $\mathrm{GO}_4^+(F)$. Then for each singular $0 \neq u \in V$*

$$\mathrm{R}_u = \mathrm{T}_u(\mathrm{R}_u \cap G')\,.$$

Proof. Let (u,v) and $H\,(\not\simeq \mathrm{GO}_2^+(F))$ be as in the previous proposition. By Lemma (12.10) we have $[W,H] = W$ hence by the proposition

$$\mathrm{R}_u = \mathrm{T}_u[\mathrm{R}_u,H] \leq \mathrm{T}_u(\mathrm{R}_u \cap G')\,,$$

as desired. □

(12.15). Proposition. *The graph $(\Omega, \sim)$ given by*

$$\langle u\rangle \sim \langle v\rangle \iff h(u,v) \neq 0$$

is connected of diameter at most 2.

Proof.

□

(12.16). Proposition. *The kernel of the action of G on Ω consists of those central scalars of $\mathrm{I}(V,q,h)$ that are contained in G.*

Proof. Let g be in the kernel of this action, and let $\alpha_v \in D$ be determined by $vg = \alpha_v v$, for each singular 1-space $\langle v\rangle \in \Omega$. By Lemma (12.10) the module V is spanned by Ω, it is enough to show that α_v is independent of $\langle v\rangle$. Let (u,v) be a hyperbolic pair. If we are not in the orthogonal case $\Lambda = \{0\}$, then we have $\alpha_u = \alpha_v$ central by Proposition (12.3). In the orthogonal case, as long as we have dimension greater than 2, we still have $\alpha_u = \alpha_v$ central by Propositions (12.6), (12.9), and Corollary (12.7).

Now as Ω is connected under $\sim$, the scalars α_v all equal some fixed scalar for all singular v, as desired. □

(12.17). Lemma. *Assume that (V, q, h) has Witt index greater than 1. For $\langle u\rangle \in \Omega$, the global stabilizers in $\mathrm{I}_D(V, q, h)$ of $u^\perp_\Omega$ and of $\langle u\rangle \cup (\Omega \setminus u^\perp)$ are both equal to $\mathrm{I}_D(V, q, h)_{\langle u\rangle}$.*

Proof. As the Witt index is at least 2, $u^\perp$ is spanned by $u^\perp_\Omega \setminus \langle u\rangle$ by Lemma (12.10). Therefore both stabilizers globally fix $u^\perp$ and so also $\langle u\rangle = (u^\perp)^\perp$. □

(12.18). Proposition.

(a) *Except when (V, q, h) has type $\mathrm{GO}_2^+(F)$, G is transitive on Ω. Especially $G = \langle \mathrm{R}_u^G\rangle$.*

(b) *Except when (V, q, h) has type $\mathrm{GO}_2^+(F)$ or $\mathrm{GO}_4^+(F)$, G is primitive on Ω.*

Proof. By Proposition (12.12) the subgroup R_u is transitive on $\Omega \setminus u^\perp$. In particular, for Witt index 1 the group G is 2-transitive on Ω, hence transitive and primitive except when $G = 1$ and $|\Omega| = 2$; that is, except for type $\mathrm{GO}_2^+(F)$.

Assume the Witt index at least 2. By Witt's Theorem (11.10) the stabilizer of $\langle u\rangle$ in $\mathrm{I}_D(V, q, h)$ is transitive on $u^\perp_\Omega \setminus \langle u\rangle$ and $\Omega \setminus u^\perp$. Therefore by the previous lemma $\mathrm{I}_D(V, q, h)$ preserves no nontrivial block containing $\langle u\rangle$ and so is primitive. Especially G is transitive (completing (a)).

The stabilizer in G of the hyperbolic pair (u, w) and the nondegenerate space $W = \langle u, w\rangle^\perp$ acts as $G \cap \mathrm{I}(W, q|_W, h|_W)$ on the set of singular 2-spaces $\langle u, v\rangle$. By (a) this action is transitive as long as W does not have type $\mathrm{GO}_2^+(F)$, which is to say V does not have type $\mathrm{GO}_4^+(F)$. By Proposition (12.12) the group R_u is transitive on each set $\langle u, v\rangle_\Omega \setminus \langle u\rangle$ for $v \in u^\perp$. Therefore as long as V is not of type $\mathrm{GO}_4^+(F)$ the group $G_{\langle u\rangle}$ is itself transitive on $u^\perp_\Omega \setminus \langle u\rangle$ and $\Omega \setminus u^\perp$. Again, the previous lemma gives primitivity. □

(12.19). Corollary. *Except when (V, q, h) has type $\mathrm{GO}_2^+(F)$, G is transitive on the set of hyperbolic pairs of V.*

Proof. If (u_1, x_1) and (u_2, x_2) are hyperbolic pairs, then there is a $g \in G$ with $(u_1, x_1)g = (u_2, x)$, for some hyperbolic pair (u_2, x). Then by Proposition (12.12) there is an $r \in \mathrm{R}_{u_2}$ with $(u_2, x)r = (u_2, x_2)$, hence $(u_1, x_1)gr = (u_2, x_2)$. □

(12.20). Proposition. *Let (v, u) be a hyperbolic pair in (V, q, h).*

(a) *Assume $\{0, 1\} \subseteq \Lambda$. Then for each $0 \neq s \in \Lambda$, the isometry g given by*

$$v \mapsto s^{-1}v\,, u \mapsto su\,, g|_{\langle v,u\rangle^\perp} = 1$$

belongs to $G = \mathrm{R}\,\mathrm{I}_D(V, q, h)$.

(b) *Assume $\{0, 1, -1\} \subsetneq \Lambda$. Then for every singular $\langle u\rangle \in \Omega$ we have $\mathrm{T}_u \leq G'$.*

Proof. It is enough to prove this for the hyperbolic 2-space $\langle v, u\rangle$ with $k = -1$. Recall that Λ is an additive subgroup of $D^{(\sigma,-k)} = D^{(\sigma,1)}$, the fixed elements of the antiautomorphism σ. For arbitrary nonzero $t \in \Lambda$

$$t^{-1} = t^{-1}tt^{-1} = t^{-1}(t)(t^{-1})^\sigma \in t^{-1}\Lambda(t^{-1})^\sigma = \Lambda\,,$$

by our definition of a form parameter. That is, for $0 \neq t \in \Lambda$ we also have $t^{-1} \in \Lambda$.

We shall write our isometries as matrices with respect to the hyperbolic pair basis (v, u). For $b, c \in \Lambda$ the singular transvection

$$\mathrm{t}_u(b)\colon z \longrightarrow z - h(z, ku)bu = z + h(z, u)bu$$

is represented by the matrix

$$\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$$

while

$$\mathrm{t}_v(c)\colon z \longrightarrow z - h(z, kv)cv = z + h(z, v)cv$$

is represented by

$$\begin{pmatrix} 1 & 0 \\ -c & 1 \end{pmatrix}.$$

(a) As in the Whitehead Lemma (6.4)(a),

$$\begin{pmatrix} 0 & -s^{-1} \\ s & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix} \begin{pmatrix} 1 & -s^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}$$

and then

$$\begin{pmatrix} s^{-1} & 0 \\ 0 & s \end{pmatrix} = \begin{pmatrix} 0 & -s^{-1} \\ s & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Here 1, s, s^{-1}, and $-s^{-1}$ are all in the additive group Λ. Therefore all these calculations go on inside of $G \geq \langle \mathrm{T}_v, \mathrm{T}_u \rangle$.

(b) We repeat the calculation made in the proof of Proposition (6.33), which showed that transvections of $\mathrm{SL}_2(D)$ are in its derived group when $|D| > 3$. For $b \notin \{0, 1, -1\}$ set $a = (b - b^{-1})^{-1}$ and $s = b$. Then in $\mathrm{SL}_2(D)$ we have the Whitehead Lemma (6.4)(f) calculation

$$\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -a + sas \\ 0 & 1 \end{pmatrix} = \left[\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} s^{-1} & 0 \\ 0 & s \end{pmatrix} \right]$$

In our present situation the calculation will prove $\mathrm{t}_u(b) \in G'$ for all $b \in \Lambda \setminus \{0, 1, -1\}$ provided we show that all the matrix entries on the right in this identity are in Λ.

By hypothesis $s = b \in \Lambda$, so by the above $s^{-1} = b^{-1} \in \Lambda$ as well. As Λ is an abelian group $b - b^{-1}$ belongs to Λ, and a second application of the above remark gives $a = (b - b^{-1})^{-1} \in \Lambda$ as desired.

We conclude that all $\mathrm{t}_u(b)$ for all $b \in \Lambda \setminus \{0, 1, -1\}$ belong to G'. But as long as $\{0, 1, -1\} \subsetneq \Lambda$, the abelian group Λ is generated by all such b. Therefore even when $b \in \{1, -1\}$ (and of course when $b = 0$) we have $\mathrm{t}_u(b) \in G'$. □

(12.21). Lemma. $\{0, 1\} \subseteq \Lambda \subseteq \{0, 1, -1\}$ *if and only if we have* $(|D|, |\sigma|) \in \{(2, 1), (3, 1), (4, 2), (9, 2)\}$.

Proof.

□

(12.22). Proposition. *Except when (V,q,h) has type $\mathrm{GO}_2^+(F)$ or one of $\mathrm{Sp}_2(2), \mathrm{Sp}_2(3), \mathrm{GU}_2(2^2), \mathrm{GU}_2(3^2)$, for (V,q,h) a hyperbolic 2-space we have $\mathrm{R}_u = \mathrm{T}_u \leq G'$. Especially G is perfect.*

Proof. This follows immediately from the previous two results. □

(12.23). Proposition. *Except when (V,q,h) has type $\mathrm{GO}_2^+(F)$ or one of $\mathrm{Sp}_2(2), \mathrm{Sp}_2(3), \mathrm{GU}_2(2^2), \mathrm{GU}_2(3^2), \mathrm{GU}_3(2^2), \mathrm{GO}_3(3)$, for (V,q,h) of Witt index 1 we have $\mathrm{R}_u \leq G'$. Especially G is perfect.*

Proof. If V has dimension 2, this comes from the previous result.

First suppose $\Lambda = \{0\}$, but the dimension is not 2. In dimension 3 Corollary (12.7) applies with $\mathrm{GO}_3(3)$ the only exceptional case (since in even characteristic (V,q,h) has even dimension). For $\Lambda = \{0\}$ and V of dimension 4 or larger, for each $\mathrm{r}(u,v,b)$ with $0 \neq b = q(v)$ there is a nondegenerate 4-subspace W with $u,v \in W$. By Proposition (12.9) $\mathrm{r}(u,v,b)$ is in the derived group of $\mathrm{R\,I}(W, q|_W, h|_W)$, a simple group.

If $\{0,1\} \subseteq \Lambda \subseteq \{0,1,-1\}$ then $(|D|,|\sigma|) \in \{(2,1),(3,1),(4,2),(9,2)\}$. With Witt index 1, this leads only to the groups $\mathrm{Sp}_2(2), \mathrm{Sp}_2(3), \mathrm{GU}_2(2^2), \mathrm{GU}_2(3^2)$ (already handled), the new exception $\mathrm{GU}_3(2^2)$, and $\mathrm{GU}_3(3^2)$.

In the final case $\mathrm{GU}_3(3^2)$, the group is in fact perfect. We handle this together with the general case of $\{0,1,-1\} \subsetneq \Lambda$, making use of Proposition (12.20)(a) ... □

(12.24). Proposition. *Except when (V,q,h) has type $\mathrm{GO}_4^+(F)$ or $\mathrm{Sp}_4(2)$, for (V,q,h) of Witt index greater than 1 we have $\mathrm{R}_u \leq G'$. Especially G is perfect.*

Proof. Assume that (V,q,h) of Witt index greater than 1 does not have type $\mathrm{GO}_4^+(F)$. By Corollary (12.14) for each singular $0 \neq u \in V$ we have $\mathrm{R}_u = \mathrm{T}_u(\mathrm{R}_u \cap G')$. Thus it remains to prove $\mathrm{T}_u \leq G'$ when we are not in the excluded cases. This is trivially true when $\Lambda = \{0\}$. and $\mathrm{T}_u = 1$.

Assume $\Lambda \neq 0$. Within V each singular 1-space $\langle u \rangle$ is contained within a hyperbolic 2-space to which Proposition (12.22) can be applied. The singular transvection subgroup T_u is contained within the derived group of the corresponding root-generated subgroup of G except for the possibilities/exceptions $\mathrm{Sp}_2(2)$, $\mathrm{Sp}_2(3)$, $\mathrm{GU}_2(2^2)$, $\mathrm{GU}_3(2^2)$, and $\mathrm{GU}_2(3^2)$.

As the Witt index is greater than 1, in the exceptional cases each hyperbolic 2-space is contained within a hyperbolic 4-space of respective types $\mathrm{Sp}_4(2)$, $\mathrm{Sp}_4(3)$, $\mathrm{GU}_4(2^2)$, $\mathrm{GU}_4(2^2)$, and $\mathrm{GU}_4(3^2)$. In the last four cases, $\mathrm{T}_u \leq \mathrm{R}_u'$ by Theorem (11.9).

This leaves $\mathrm{Sp}_4(2)$ where Lemma (12.2) reveals that the transvections—the 2-cycles—are not in the derived subgroup of $\mathrm{Sym}(6) \simeq \mathrm{Sp}_4(2) = \mathrm{RSp}_4(2)$. For symplectic groups over $\mathbb{F}_2$ we must go one step higher. In $\mathrm{Sp}_6(2)$ the group R_u is an elementary abelian 2-group of order 2^5, easily seen to be isomorphic to the even submodule $[\mathbb{F}_2\{1,2,3,4,5,6\}, \mathrm{Alt}(6)]$ of the natural $\mathbb{F}_2$-permutation module for Sym(6) and $\mathrm{Alt}(6) \leq G'_{\langle u \rangle}/\,\mathrm{R}_u$. In particular T_u of order 2 is again in G', completing the proposition. □

Proof of Theorem (12.1).

Assume we are not in one of the (genuinely) exceptional cases.

By Propositions (12.11) and (12.18)(a) $G = \langle R^G \rangle$ for solvable $R \trianglelefteq G_\omega$ with $\omega \in \Omega$. By Proposition (12.18)(b) G is primitive on Ω. By Propositions (12.23) and (12.24) $G = G'$. Therefore by the Brey-Wilson version of Iwasawa's Lemma (3.15) all normal subgroups of G are contained in $\ker \Omega$, which is $\mathrm{Z}(G)$ by Proposition (12.16). □

12.3 Symplectic groups

An issue not addressed in the previous section is the structure of the quotient $\mathrm{I}_D(V,q,h)/\mathrm{R}\,\mathrm{I}_D(V,q,h)$. The next theorem shows that it always has a relatively restricted structure, and the theorem that follows it solves the problem completely in the symplectic case.

(12.25). Theorem. *If (V,q,h) is not of type $\mathrm{GO}_2^+(F)$, then G is edge-transitive on Ω and $\mathrm{I}_D(V,q,h) = G.J$ for J as in Proposition (12.3). In particular $I/G \simeq J/J \cap G'$.*

Proof. This follows from Corollary (12.19) and the Frattini Argument (3.8). □

(12.26). Theorem. *Except for the genuine exceptions* $\mathrm{Sp}_2(2)$, $\mathrm{Sp}_2(3)$, *and* $\mathrm{Sp}_4(2)$, *the symplectic group $G = \mathrm{Sp}_{2m}(F)$ is perfect with $\mathrm{Z}(G) = \{\pm I\}$ and $G/\mathrm{Z}(G)$ simple.*

Proof. By Theorem (12.25) $G = \mathrm{Sp}_2(F).\mathrm{RSp}_{2m}(F)$. In view of Theorem (12.1), it suffices to prove $\mathrm{Sp}_2(F) = \mathrm{RSp}_2(F)$.

In the symplectic case, $\Lambda = F$ and all 1-spaces are singular. So by Propositions (12.4) and (12.20) we have $\mathrm{Sp}_2(F) = \mathrm{SL}_2(F) = \mathrm{RSp}_2(F)$, and we are done. □

Other Stuff as of: 6 May 2015
fake index: (0)-*dummy index 2015-05-06*

Bibliography

[Art88] E. Artin, "Geometric Algebra," Wiley Classics Library, John Wiley and Sons, Inc., New York, 1988.

[Asc00] M. Aschbacher, "Finite Group Theory," Second edition, Cambridge Studies in Advanced Mathematics, **10**, Cambridge University Press, Cambridge, 2000.

[BuC13] F. Buekenhout and A.M. Cohen, "Diagram Geometry. Related to Classical Groups and Buildings," Ergebnisse der Mathematik und ihrer Grenzgebiete, 3 Folge, **57**, Springer, Heidelberg, 2013.

[Cam92] P.J. Cameron, "Projective and Polar Spaces," QMW Maths Notes, **13**, Queen Mary and Westfield College, School of Mathematical Sciences, London, 1992.

[Cho49] W.L. Chow, *On the geometry of algebraic homogeneous spaces*, Ann. of Math. (Series 2), **50** (1949), 32–67.

[Coh13] A.M. Cohen, "Diagram Geometry," draft, 23 August 2013.

[Dic58] L.E. Dickson, "Linear groups: With an Exposition of the Galois Field Theory, with an Introduction by W. Magnus," Dover Publications, Inc., New York, 1958

[Die48] J. Dieudonné, "Sur les Groupes Classiques," Actualités Sci. Ind., **1040**, Hermann et Cie., Paris, 1948.

[Die51] J. Dieudonné, *Algebraic homogeneous spaces over fields of characteristic two*, Proc. Amer. Math. Soc., **2** (1951), 295–304.

[Die71] J. Dieudonné, "La géométrie des groupes classiques," Troisième édition, Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, **5**, Springer-Verlag, Berlin-New York, 1971.

[Fro90] M. Frolov, *Recherches sur les permutations carrées*, J. Math. Spec., (3) **4** (1890), 8–11.

[Gro02] L.C. Grove, “Classical Groups and Geometric Algebra,” Graduate Studies in Mathematics, **39**, American Mathematical Society, Providence, RI, 2002.

[HaO89] A.J. Hahn and T.O. O’Meara, “The Classical Groups and K-Theory, with a Foreword by J. Dieudonné,” Grundlehren der Mathematischen Wissenschaften, **291**, Springer-Verlag, Berlin, 1989.

[Hll43] M. Hall, Jr., *Projective planes*, Trans. Amer. Math. Soc., **54** (1943), 229–277,

[Hll49] M. Hall, Jr., *Correction to “Projective planes,”* Trans. Amer. Math. Soc., **65** (1949), 473–474.

[Hll59] M. Hall, Jr, “The Theory of Groups,” The Macmillan Co., New York, 1959.

[Hig78] D.G. Higman, “Classical Groups, with an Appendix by D.E. Taylor,” T.H.-Report 78-WSK-04, Technological University Eindhoven, 1978.

[Hil62] D. Hilbert and E.J. Townsend, “The Foundations of Geometry,” Open Court Pub. Co, Chicago, 1962.

[Hum90] J.E. Humphreys, “Reflection Groups and Coxeter Groups,” Cambridge Studies in Advanced Mathematics, **29**, Cambridge University Press, Cambridge, 1990.

[Jac53] N. Jacobson, “Lectures in Abstract Algebra II. Linear Algebra,” Springer Verlag, New York, 1953.

[Jac89] N. Jacobson, “Basic Algebra II,” Second edition, W.H. Freeman and Company, New York, 1989.

[Kan79] W.M. Kantor, ”Classical groups from a nonclassical viewpoint,” Oxford University, Mathematical Institute, Oxford, 1979.

[KMRT98] M.-A. Knus, A. Merkurjev, M. Rost, and J.-P. Tignol, “The Book of Involutions, with a Preface in French by J. Tits,” American Mathematical Society Colloquium Publications, **44**, American Mathematical Society, Providence, RI, 1998.

[OMe77] O.T. O’Meara, *A general isomorphism theory for linear groups*, J. Algebra, **44** (1977), 93–142.

[Par70] B. Pareigis, “Categories and Functors,” Academic Press, New York-London, 1970.

[Pic55] G. Pickert, “Projektive Ebenen,” Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Bercksichtigung der Anwendungsgebiete, **LXXX**, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1955.

[Rob82] D.J.S. Robinson, “A Course in the Theory of Groups,” Graduate Texts in Mathematics, **80**, Springer-Verlag, New York-Heidelberg-Berlin, 1982.

[Shu11] E.E. Shult, “Points and Lines. Characterizing the Classical Geometries,” Universitext, Springer, Heidelberg, 2011.

[Tay92] D.E. Taylor, “The Geometry of the Classical Groups,” Sigma Series in Pure Mathematics, **9**, Heldermann Verlag, Berlin, 1992.

[VeY16] O. Veblen and J.W. Young, “Projective Geometry, Vol. 1, 2,” Ginn and Co., Boston, 1916, 1917.

[Wal63] G.E. Wall, *On the conjugacy classes in the unitary, symplectic and orthogonal groups*, J. Austral. Math. Soc., **3** (1963), 1–62.

[Wil09] R.A. Wilson, “The Finite Simple Groups,” Graduate Texts in Mathematics, **251**, Springer-Verlag London, Ltd., London, 2009.

Index