A First Encounter with Classical Groups

Adrien Deloro

Şirince '14 Summer School

Contents

List of Lectures

Introduction

> *La filosofia è scritta in questo grandissimo libro che continuamente ci sta aperto innanzi a gli occhi (io dico l'universo), ma non si può intendere se prima non s'impara a intender la lingua, e conoscer i caratteri, ne' quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi, ed altre figure geometriche, senza i quali mezi è impossibile a intenderne umanamente parola; senza questi è un aggirarsi vanamente per un oscuro laberinto.*
>
> Galileo, Il Saggiatore, Capitolo 6.

> *Philosophy is written in yon grand book which permanently lies open before our eyes (I mean the universe) but cannot be understood lest one learns to understand its language, and know the characters in which it is written. It is written in mathematical language, and the characters are triangles, circles, and other geometric figures, without which means it is manly impossible to understand a word; without those one wanders in vain an obscure labyrinth.*

To which we should add the following: *the book of geometry is written in group-theoretic language.* The latter was clearly put by Felix Klein in his famous text (1872). Less than seventy years later, Weyl published a book on the "classical groups": the topic was already classical. But nowadays the phrase "classical groups" refers to certain matrix groups of fundamental importance. Let us mention two major results of modern mathematics.

Classification of the finite simple groups
Let G be a finite simple group. Then G is isomorphic to one of the following:

- the cyclic group C_p for some prime p;
- the alternating group A_n for some integer $n \geqslant 5$;
- a group of Lie type ("most" of them are in fact classical groups);
- one of the 26 known "sporadic" groups.

Classification of the finite-dimensional simple Lie algebras
Let $\mathfrak{g}$ be a finite-dimensional, simple Lie algebra over the complex numbers. Then $\mathfrak{g}$ is isomorphic to one of the following:

- a classical Lie algebra: A_n, B_n, C_n, or D_n;

- an exceptional Lie algebra: E_6, E_7, E_8, F_4, or G_2.

We shall discuss neither of these theorems. Understanding them would go extremely deep into mathematics. But both are of major significance and both rely on objects called "classical".

What does "classical" mean then? It should not be opposed to baroque or romantic. It means something like "the usual suspects": the same structure behind some groups keeps arising over and over again. This structure is as one should expect of geometric nature.

Sometimes classical groups are discussed only over the field of complex numbers; we shall try and keep things more general, as the topic over finite fields is of extreme interest as well.

Prerequisites

In order to follow this class, one needs to know:

- fields: essentially the definition, and the fact that for every prime power $q = p^n$ there exists a unique field of order q;
- groups (subgroups, normal subgroups, ...) and group actions (stabilisers, orbits, ...);
- vector spaces, linear maps;
- matrices, elementary row and column operations.

For the second week, one also needs:

- bilinear forms (scalar products, isometries, ...).

Recommended Reading

[1] E. Artin. *Geometric algebra*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1988. Reprint of the 1957 original, A Wiley-Interscience Publication
A classical book on classical groups. Very elegant, but parts of the exposition would be different now.

[2] Donald E. Taylor. *The geometry of the classical groups*, volume 9 of *Sigma Series in Pure Mathematics*. Heldermann Verlag, Berlin, 1992.
If you have special interest in finite fields.

[3] Larry C. Grove. *Classical groups and geometric algebra*, volume 39 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002.
Clear, precise, concise: highly recommended.

[4] Roger W. Carter. *Simple groups of Lie type*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1989. Reprint of the 1972 original, A Wiley-Interscience Publication
Harder. Read it only once you master the others.

Week 1: The General and Special Linear Groups

Lecture 1 (Introduction; the General Linear Group)

1.1 The general and special linear groups

1.1.1 The general linear group

Let $\mathbb{F}$ be a field and V be a vector space over $\mathbb{F}$ of dimension n.

Definition 1.1.1. Let $\mathrm{GL}(V)$ be the *general linear group* of all invertible linear transformations from V into itself.

As we know linear transformations can be represented by matrices.

Definition 1.1.2. Let $\mathrm{GL}(n, \mathbb{F})$ be the group of all invertible $n \times n$ matrices over $\mathbb{F}$.

Proposition 1.1.3. *If V is a vector space over $\mathbb{F}$ of dimension n, then* $\mathrm{GL}(V) \simeq \mathrm{GL}(n, \mathbb{F})$.

Proof. Let $e_1, \ldots, e_n$ be a basis for V and $f \in \mathrm{GL}(V)$. Then:

$$f(e_j) = \sum_{i=1}^{n} a_{i,j} e_i$$

for some coefficients $a_{i,j} \in \mathbb{F}$. The map:

$$\begin{array}{ccc} \mathrm{GL}(V) & \to & \mathrm{GL}(n, \mathbb{F}) \\ f & \mapsto & (a_{i,j}) \end{array}$$

is an isomorphism by the very definition of matrix multiplication. □

Remark 1.1.4. There is no *canonical* isomorphism: the isomorphism given in the proof depends on the choice of a basis. In particular there are many so. *Never forget this.* In general reducing a geometric question to a matrix computation hides the interest and meaning of a problem. We shall see an example of such irrelevant hiding at once.

The center of $GL(V)$

Proposition 1.1.5. $Z(\mathrm{GL}(V)) = \{\lambda \,\mathrm{Id} : \lambda \in \mathbb{F}^\times\}$.

Proof 1. For this proof we suppose that the characteristic is not 2.

By the isomorphism above, this is equivalent to showing $Z(\mathrm{GL}(n,\mathbb{F})) = \{\lambda I_n : \lambda \in \mathbb{F}^\times\}$. One inclusion is clear: if $\lambda \in \mathbb{F}^*$, then $\lambda I_n \in Z(\mathrm{GL}(n,\mathbb{F}))$. For the converse, fix $M = (m_{i,j}) \in Z(\mathrm{GL}(n,\mathbb{F}))$; we aim at showing that M has the form λI_n.

Recall that the elementary matrix $E_{i,j}(i \neq j)$ is the matrix which has zeros everywhere except for a 1 in position (i,j). Also recall that $E_{i,j} \cdot E_{k,\ell} = \delta_{j,k} E_{i,\ell}$, where $\delta_{j,k}$ is 1 if $j = k$ and 0 otherwise.

We want to say that M commutes with $E_{i,j}$. Since by assumption $M \in Z(\mathrm{GL}(n,\mathbb{F}))$, it commutes to any matrix in $\mathrm{GL}(n,\mathbb{F})$ but there is a slight catch as $E_{i,j} \notin \mathrm{GL}(n,\mathbb{F})$. Now for all $i,j \in \{1,\dots,n\}$, $I_n + E_{i,j} \in \mathrm{GL}(n,\mathbb{F})$ (when $i = j$ this requires the characteristic to be not 2), so by assumption $M \cdot (I_n + E_{i,j}) = M + M \cdot E_{i,j}$ must equal $(I_n + E_{i,j}) \cdot M = M + E_{i,j} \cdot M$, so $M \cdot E_{i,j} = E_{i,j} \cdot M$.

So let us compute:

$$(M \cdot E_{i,j})_{k,\ell} = \left(\sum_{p,q} m_{p,q} E_{p,q} E_{i,j}\right)_{k,\ell} = \left(\sum_{p} m_{p,i} E_{p,j}\right)_{k,\ell} = m_{k,i}\delta_{j,\ell}$$

and similarly, $(E_{i,j} \cdot M)_{k,\ell} = m_{j,\ell}\delta_{i,k}$. They must be equal: therefore for any i,j and k,ℓ in $\{1,\dots,n\}$, one has $m_{k,i}\delta_{j,\ell} = m_{j,\ell}\delta_{i,k}$.

So take $j = \ell$ and $i \neq k$. Then $m_{k,i} = m_{k,i}\delta_{j,\ell} = m_{j,j}\delta_{i,k} = 0$; this means that M is a diagonal matrix. Now take $j = \ell$ and $i = k$. Then $m_{i,i} = m_{k,i}\delta_{j,\ell} = m_{j,j}\delta_{i,k} = m_{j,j}$. So M has the form λI_n. □

This argument is rather meaningless; it is tedious, one has a hard time getting the subscripts right, and the reader is not more clever at the end. There is worse: we didn't prove it when $\mathbb{F}$ has characteristic 2.

Exercise 1.1.6. *Fix Proof 1 in characteristic* 2.

In short, with matrices one can *prove* things but one can't really *explain* them. So we now give another proof.

Proof 2. Clearly any $\lambda \,\mathrm{Id}$, with $\lambda \in \mathbb{F}^\times$, is in $Z(\mathrm{GL}(V))$. Let $f \in Z(\mathrm{GL}(V))$.

Suppose that there is $v \in V$ such that $(v, f(v))$ is linearly independent. Then there exists $g \in \mathrm{GL}(V)$ such that $g(v) = v$ and $g(f(v)) = v + f(v)$ (this is possible since $(v, v + f(v))$ is another linearly independent family). Now $f(g(v)) = f(v) \neq g(f(v))$, so $f \circ g \neq g \circ f$: a contradiction.

This shows that for any $v \in V$, one has $f(v) \in \langle v \rangle$, the linear span of v. Hence there is $\lambda_v \in \mathbb{F}$ such that $f(v) = \lambda_v v$. A priori λ_v depends on v; if we show that it doesn't we are finished. As a matter of fact λ_v remains constant on $\langle v \rangle$, so we now do it for two linearly independent vectors. So let (v,w) be linearly independent (if $\dim V = 1$ then we are done). Then:

$$\begin{aligned} f(v) + f(w) &= \lambda_v v + \lambda_w w \\ = f(v+w) &= \lambda_{v+w}(v+w) = \lambda_{v+w} v + \lambda_{v+w} w \end{aligned}$$

Since (v,w) is linearly independent, this shows $\lambda_v = \lambda_{v+w} = \lambda_w$. Hence λ_v does not depend on v: for any $v \in V$, $f(v) = \lambda v$ for a common $\lambda \in \mathbb{F}$. By invertibility, $\lambda \neq 0$. □

This is much easier, and our problems with characteristic 2 disappeared.

The order of GL(n, q)

We shall have some interest in finite fields throughout. Assume that $\mathbb{F}$ is finite, say $|\mathbb{F}| = q = p^k$ where p is a prime and $k \geqslant 1$. In this case, by uniqueness up to isomorphism of the field of order q, we simply write $\mathrm{GL}(n, q)$.

Lemma 1.1.7. $|\mathrm{GL}(n, q)| = q^{\frac{n(n-1)}{2}} \prod_{i=1}^{n} (q^i - 1)$.

Here again there will be two proofs, the first being more combinatorial and the second more group-theoretic.

Proof 1. We shall compute the order of $\mathrm{GL}(n, q)$. It is the number of invertible, $n \times n$, matrices with coefficients in $\mathbb{F}_q$.

- The first row of such a matrix can be any of the $q^n - 1$ non-zeros row vectors over $\mathbb{F}$, so there are $q^n - 1$ choices.
- The second row of the matrix must be independent of the first, so there are $q^n - q$ choices left.
- The third row of the matrix must be independent from the first two, so we must remove the span of the two first rows. This span has order q^2, so there are $q^n - q^2$ choices left.
- And so on.

It follows that:

$$\begin{aligned} |\mathrm{GL}(n, q)| &= (q^n - 1)(q^n - q) \dots (q^n - q^{n-1}) \\ &= q^{1+2+\dots+(n-1)} \prod_{i=1}^{n} (q^i - 1) \\ &= q^{\frac{n(n-1)}{2}} \prod_{i=1}^{n} (q^i - 1) \end{aligned}$$

□

We now give another proof of this fact using the language of group actions. This needs a few classical definitions.

Definition 1.1.8. The action of G on Ω is *faithful* if the only $g \in G$ such that $\forall \alpha \in \Omega \; g \cdot \alpha = \alpha$ is $g = 1$ (this means that $\bigcap_{\alpha \in \Omega} \mathrm{Stab}_G(\alpha) = \{1\}$).

Definition 1.1.9. If there is $\alpha \in \Omega$ such that $\Omega = \mathrm{orb}^G(\alpha)$, then we say that G acts *transitively* on Ω. (In that case, this holds of any $\alpha \in \Omega$.)

Example 1.1.10. The action of $\mathrm{GL}(V)$ on $V \backslash \{0\}$ is faithful and transitive.

Proof 2. Since the action of $G = \mathrm{GL}(n, q)$ on $\mathbb{F}^n \backslash \{0\}$ is transitive, one has $|G| = |\mathbb{F}^n \backslash \{0\}| \cdot |\mathrm{Stab}_G(\alpha)|$ for any $\alpha \in \mathbb{F}^n \backslash \{0\}$. So let us compute the stabiliser of the first vector e_1 of the canonical basis of $\mathbb{F}^n$:

$$\mathrm{Stab}_G(e_1) = \left\{ \begin{pmatrix} 1 & * & \dots & * \\ 0 & & & \\ \vdots & & N & \\ 0 & & & \end{pmatrix} : N \in \mathrm{GL}(n-1, q) \right\}$$

The line of $*$ gives q^{n-1} possibilities. Hence $|\operatorname{Stab}_G(e_1)| = q^{n-1} \cdot |\mathrm{GL}(n-1,q)|$, meaning that $|\mathrm{GL}(n,q)| = (q^n-1)q^{n-1}|\mathrm{GL}(n-1,q)|$, which leads to the same formula. □

End of Lecture 1.

Lecture 2 (The Special Linear Group and Transvections Matrices)

1.1.2 $\mathrm{SL}(V)$: matrix study

Recall that the determinant map det is multiplicative from the set of $n \times n$ matrices over $\mathbb{F}$ to $\mathbb{F}$, and that $\mathrm{GL}(n,\mathbb{F}) = \{M : \det M \neq 0\}$. Hence $\det : \mathrm{GL}(n,\mathbb{F}) \to \mathbb{F}^*$ is a group homomorphism. In particular, if $f \in \mathrm{GL}(V)$ and $\mathcal{B}, \mathcal{B}'$ are two bases, then $\mathrm{Mat}_{\mathcal{B}}(f)$ and $\mathrm{Mat}_{\mathcal{B}'}(f)$ are conjugate so by multiplicativity they have the same determinant. Hence $\det f$ is a well-defined number. By multiplicativity again, $\det(g \circ f) = \det g \cdot \det f$.

Definition 1.1.11. Let $\mathrm{SL}(V) = \ker\det = \{f \in \mathrm{GL}(V) : \det f = 1\}$ be the *special linear group*. In matrix form, let $\mathrm{SL}(n,\mathbb{F}) = \ker\det = \{M \in \mathrm{GL}(n,\mathbb{F}) : \det M = 1\}$.

Geometric interpretation: $\mathrm{SL}(V)$ is the group of "volume"-preserving linear transformations (in $\mathbb{R}^3$, the determinant of a basis equals the volume of the parallelepiped it determines).

If $\mathbb{F}$ is finite with q elements, we simply write $\mathrm{SL}(n,q)$.

Remark 1.1.12. Since the det map is onto $\mathbb{F}^*$, a group with $q-1$ elements, one finds $[\mathrm{GL}(n,q) : \mathrm{SL}(n,q)] = q-1$. Combining with the order of $\mathrm{GL}(n,q)$, we deduce $|\mathrm{SL}(n,q)| = q^{\frac{n(n-1)}{2}} \prod_{i=2}^{n}(q^i-1)$.

Generation by transvection matrices

Proposition 1.1.13. $\mathrm{SL}(n,\mathbb{F})$ *is generated by the matrices of the form* $I_n + \lambda E_{i,j}$ *(*$i \neq j$*,* $\lambda \in \mathbb{F}^*$*).*

Proof. This is Gauß' algorithm for computing determinants. Let $M \in \mathrm{SL}(n,\mathbb{F})$. We call *transvection matrix* every matrix of the form $I_n + \lambda E_{i,j}$ (with $i \neq j$); the reason will be clear later. In any case, transvection matrices are in $\mathrm{SL}(n,\mathbb{F})$.

Observe that multiplying by $E_{i,j}$ on the left takes the j^{th} row to the i^{th} row (and the rest of the matrix is erased). So a row operation $R_i \leftarrow R_i + \lambda R_j$ corresponds to multiplication on the left by $I_n + \lambda E_{i,j}$. Similarly, a column operation $C_j \leftarrow C_j + \lambda C_i$ corresponds to multiplication by $I_n + \lambda E_{i,j}$ on the right. These are the only operations we may use (a row swapping matrix has determinant -1; a non-trivial, row multiplication matrix has determinant $\neq 1$).

If we apply Gauß' algorithm naively we will find the following: there are transvection matrices $T_1, \ldots, T_t$ and $S_1, \ldots, S_s$ such that $T_t \ldots T_1 M S_1 \ldots S_s$ is diagonal. The proof is not finished then: one needs to prove that diagonal matrices with determinant 1 are generated by transvection matrices (which amounts to assuming that M is diagonal). We do not want to do so.

When we start the algorithm, we look if there is $i \geqslant 2$ with $m_{1,i} \neq 0$.

- If so, then $L_1 \leftarrow L_1 + \frac{1-m_{1,1}}{m_{i,1}} L_i$ will bring 1 in position $(1,1)$.

- If there is none, then only $m_{1,1}$ is non-zero in the first column. Be careful that perhaps $m_{1,1} \neq 1$. But observe that:

$$(I_n + E_{i,j})(I_n - E_{j,i})(I_n + E_{i,j}) = I_n + E_{i,j} - E_{j,i} - E_{i,i} - E_{j,j}$$

$$= \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & -1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

  Let $W_{i,j}$ denote the latter matrix. Multiplication on the left by $W_{i,j}$ performs the transformation: $L_i \leftarrow L_j, L_j \leftarrow -L_i$. So up to multiplying by $W_{1,2}$ we may assume that $m_{2,1} \neq 0$, and we are back to the first case.

  Note that this part of the proof involves expressing row-swapping (with one sign change of course, because of the determinant) in terms of transvection matrices!

In any case we see that applying this modified version of the algorithm we can bring M to the block matrix:

$$\begin{pmatrix} 1 & \\ & N \end{pmatrix}$$

so inductively, to the diagonal form:

$$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & d \end{pmatrix}$$

using *only transvection matrices*. Since $d = \det M = 1$, it follows that M is a product of transvection matrices. □

I hope that this proof is difficult to understand. It seems to hold by miracle, and its meaning is carefully hidden. We should unfold it and try and find the geometric interpretation of transvection matrices.

Exercise 1.1.14. *Write* $-I_2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ *as a product of transvections.*

1.2 Transvections

1.2.1 Transvections: Geometric aspects

Consider a transvection matrix

$$T = I_n + \lambda E_{i,j} = \begin{pmatrix} 1 & & & \\ & \ddots & \lambda & \\ & & 1 & \end{pmatrix}$$

where the λ is in position (i,j) with $i \neq j$. What is its meaning?

Observe that T is "close" to I_n; more specifically $T - I_n = \lambda E_{i,j}$ which has rank 1. Moreover, the image of $T - I_n$ is generated by e_j, which is in the kernel of $T - I_n$ as $i \neq j$.

Definition 1.2.1. Let V be a vector space over $\mathbb{F}$. A *transvection* of V is a linear map $\tau : V \to V$ such that:

- $\mathrm{rk}(\tau - \mathrm{Id}) = 1$;
- $(\tau - \mathrm{Id})^2 = 0$.

Remark 1.2.2. Depending on the context, Id is regarded as a transvection or not (in the above definition, it was not the case).

Remark 1.2.3. If τ is a transvection of V, then there is a basis $\mathcal{B}$ of V in which $\mathrm{Mat}_{\mathcal{B}}(\tau)$ is a transvection matrix $I_n + \lambda E_{i,j}$.

Take a basis v_1 of $\mathrm{im}(\tau - \mathrm{Id})$; since $\mathrm{im}(\tau - \mathrm{Id}) \leqslant \ker(\tau - \mathrm{Id})$ which has dimension $n-1$, extend v_1 to a basis $(v_1, \dots, v_{n-1})$ of $\ker(\tau - \mathrm{Id})$, and then to a basis $\mathcal{B} = (v_1, \dots, v_n)$ of V. We claim that $\mathrm{Mat}_{\mathcal{B}}(\tau)$ has the desired form.

Indeed, for $i \leqslant n-1$, one has $\tau(v_i) = \mathrm{Id}(v_i) = v_i$, but also $\tau(v_n) = \mathrm{Id}(v_n) + (\tau(v_n) - \mathrm{Id}(v_n)) = v_n + \lambda v_1$. Hence:

$$\mathrm{Mat}_{\mathcal{B}}(\tau) = \begin{pmatrix} 1 & & \lambda \\ & \ddots & \\ & & 1 \end{pmatrix} = I_n + \lambda E_{1,n}$$

Of course the description still has an algebraic flavor. We need more tools in order to reach the geometric form.

Definition 1.2.4. A *linear form* is a linear map $\varphi : V \to \mathbb{F}$.

The set V^* of linear forms is a vector space over $\mathbb{F}$ for the laws:

$$(\varphi_1 + \varphi_2)(v) := \varphi_1(v) + \varphi_2(v)$$
$$(a\varphi)(v) := a\varphi(v)$$

V^* is called the *dual space* of V.

Proposition 1.2.5. *If V has dimension n then so does V^*.*

Proof. Let $\mathcal{B} = (e_1, \dots, e_n)$ be a basis of V. For each $i = 1 \dots n$, let e_i^* be defined by:

$$e_i^* \left(\sum_j a_j e_j \right) = a_i$$

Then e_i^* is a linear form, that is, an element of V^*. We claim that $\mathcal{B}^* = (e_1^*, \dots, e_n^*)$ is a basis of V^* (called the dual basis to $\mathcal{B}$).

First suppose that $\sum_i \lambda_i e_i^* = 0$ as linear forms. Then applying to e_j, one finds $\lambda_j = 0$. It follows that $\mathcal{B}^*$ is linearly independent.

Now let $\varphi \in V^*$ be any linear form. Let $\lambda_i = \varphi(e_i)$. Then by construction, φ and $\sum \lambda_i e_i^*$ agree on a basis of V. By linearity they agree everywhere, so they are equal as linear forms. Hence $\varphi = \sum \lambda_i e_i^*$ lies in the span of $\mathcal{B}^*$. □

Remark 1.2.6. If $\dim V = \infty$ then $\mathcal{B}^*$ remains linearly independent, but no longer generates V^*. Find a counter-example.

Definition 1.2.7. A *hyperplane* of V is the kernel of a non-zero linear form.

Lemma 1.2.8. *A hyperplane has codimension* 1 *(this simply means that with* $\dim V = n$*, one has* $\dim H = n-1$*). Conversely, if* H *is a subspace with codimension* 1*, then there exists a non-zero linear form* φ *with* $H = \ker\varphi$*; moreover, all such* φ *are collinear.*

Proof. First suppose $H = \ker\varphi$ with $\varphi \neq 0$. Then $\varphi : V \to \mathbb{F}$ is non-zero, so it is onto; therefore $\dim H = \dim\ker\varphi = \dim V - \dim\operatorname{im}\varphi = n-1$.

Suppose conversely that H has codimension 1. Let $(h_1, \dots, h_{n-1})$ be a basis of H, and extend it to a basis $(h_1, \dots, h_{n-1}, a)$ of V. Define a linear form as follows: $\varphi(h_i) = 0$ and $\varphi(a) = 1$. This defines a non-zero linear form with $H \leqslant \ker\varphi$; by dimension equality, $H = \ker\varphi$.

We finally show that such φ is unique up to collineation in V^*. So suppose $H = \ker\varphi = \ker\psi$; take the same basis $(h_1, \dots, h_{n-1}, a)$ as before and let $\lambda = \psi(a)$. Then on every element b of the basis, $\psi(b) = \lambda\varphi(b)$, so $\psi = \lambda\varphi$ as linear forms. □

We can now give a better description of transvections.

Lemma 1.2.9. *Let* $\tau \neq \mathrm{Id}$ *be a transvection of* V*. Then there exist a linear form* φ *and a vector* $a \in \ker\varphi$ *such that: for all* $v \in V$*,* $\tau(v) = v + \varphi(v)a$*.*

Proof. Let $H = \ker(\tau - \mathrm{Id})$; then $\dim H = n-1$ so H is a hyperplane of V. There exists $\varphi \in V^*$ with $H = \ker\varphi$. Let $v_0 \in V \backslash H$ and $a = \frac{1}{\varphi(v_0)}(\tau(v_0) - v_0)$. We contend that the pair (φ, a) meets the requirements.

Observe that $(\tau - \mathrm{Id})(v_0) \in \ker(\tau - \mathrm{Id}) = H$ so $a \in H = \ker\varphi$. Moreover $\tau(a) = a$. Now let $v \in V$. Since $v_0 \notin H$, one has $\langle H, v_0\rangle = V$: there exist $h \in H$ and $\lambda \in \mathbb{F}$ with $v = h + \lambda v_0$. Then $\varphi(v) = \lambda\varphi(v_0)$, and:

$$\tau(v) = \tau(h) + \lambda\tau(v_0) = h + \lambda(v_0 + \varphi(v_0)a) = v + \varphi(v)a$$

as desired. □

It should be checked that $v \mapsto v + \varphi(v)a$ does define a transvection (using $a \in \ker\varphi$). As a consequence, for every transvection t there is a basis in which the matrix of t is a transvection matrix $I_n + \lambda E_{i,j}$. It follows that every transvection is in $\mathrm{SL}(V)$.

Notation 1.2.10. For $\varphi \in V^*$ and $a \in \ker\varphi$ let $t_{\varphi,a}(v) = v + \varphi(v)a$.

One may wonder to which extent the pair (φ, a) is unique. For instance, if we start with the transvection whose matrix (in some fixed basis $(v_1, \dots, v_n)$) is $I_n + \lambda E_{i,j}$, one could have $\varphi = \lambda e_i^*$, $a = e_j$; one could also have $\varphi = e_i^*$, $a = \lambda e_j$; as a matter of fact one could have $\varphi = \mu e_i^*$, $a = \frac{\lambda}{\mu} e_j$ for any $\mu \neq 0$.

Lemma 1.2.11. $t_{\varphi,a} = t_{\psi,b}$ *iff there is* $\lambda \in \mathbb{F}^*$ *with* $\varphi = \lambda\psi$ *and* $a = \lambda^{-1}b$*.*

Proof. One implication is obvious. Conversely suppose $t_{\varphi,a} = t_{\psi,b}$. Then $\ker\varphi = \ker\psi$ is a hyperplane, so by collinearity in Lemma 1.2.8 there is $\lambda \neq 0$ with $\varphi = \lambda\psi$. Now on any $v_0 \notin \ker\varphi = \ker\psi$, $t_{\varphi,a}(v_0) = v_0 + \varphi(v_0)a = v_0 + \lambda\psi(v_0)a = t_{\psi,b}(v_0) = v_0 + \psi(v_0)b$. So dividing by $\psi(v_0) \neq 0$, one has $b = \lambda a$. □

Remark 1.2.12. The appropriate setting would be the *tensor product* of V and V^*, a topic we shall not discuss.

End of Lecture 2.

Lecture 3 (Generation by transvections)

1.2.2 Generation by transvections

Last time we saw that transvection matrices generate $\mathrm{SL}(n, \mathbb{F})$ (Proposition 1.1.13). We prove the geometric form of this statement.

Proposition 1.2.13. *The transvections generate* $\mathrm{SL}(V)$.

Proof. We have observed that the transvections are elements of $\mathrm{SL}(V)$. Let $f \in \mathrm{SL}(V)$ be any map. We shall write f as a product of transvections. The beginner should think that we shall adapt the matrix proof of Proposition 1.1.13, that is argue by induction. Having a 1 in position $(1,1)$ amounts to finding a vector fixed by f.

So let us first reduce to the case where f has a (non-trivial) fixed point.

- Suppose that there exists $v_0 \in V$ with $f(v_0) \notin \langle v_0 \rangle$. Consider $a = v_0 - f(v_0)$; then a and $f(v_0)$ are linearly independent, so there exists a linear form φ with $\varphi(a) = 0$ and $\varphi(f(v_0)) = 1$. Then $t_{\varphi,a}(f(v_0)) = f(v_0) + \varphi(f(v_0))a = v_0$, so $f \circ t_{\varphi,a}(f(v_0)) = f(v_0)$ and $f t_{\varphi,a}$ has a fixed point.
- If there is no such v_0 then for all $v_0 \in V$, $f(v_0) \in \langle v_0 \rangle$; we saw in the second proof of Proposition 1.1.5 that f is a vector dilation $\lambda \,\mathrm{Id}$. The product with *any* transvection will no longer be such: we are back to the first case.

So we may assume that f has a fixed vector v_1 with $f(v_1) = v_1$. Extend v_1 to a basis $\mathcal{B} = (v_1, \ldots, v_n)$ of V and write $H = \langle v_2, \ldots, v_n \rangle$. We hope to do induction. The problem is that f does not necessarily fix H setwise; the matrix of f in $\mathcal{B}$ could have the following form:

$$\begin{pmatrix} 1 & * & \ldots & * \\ 0 & & & \\ \vdots & & A & \\ 0 & & & \end{pmatrix}$$

So we need to take care of the "$*$" line. We shall inductively fill it with zeros (this is again very much in the spirit of the Gauß algorithm).

Let us prove that we may assume that $f(v_2), \ldots, f(v_n)$ all lie in H. By induction on i.

- Since $V = \langle v_1 \rangle \oplus H$ there are $\lambda_2 \in \mathbb{F}$ and $h_2 \in H$ with $f(v_2) = \lambda_2 v_1 + h_2$. Let τ_2 be a transvection fixing v_1 and mapping v_2 to $v_2 - \lambda_2 v_1$. Observe how $f \circ \tau_2(v_1) = f(v_1) = v_1$ and $f \circ \tau_2(v_2) = f(v_2 - \lambda_2 v_1) = h_2$. So we may assume that f maps v_2 into H.
- Suppose the result holds of $i-1$. We wish to do the same for v_i. The problem is that we should not touch the value of f on $v_1, \ldots, v_{i-1}$.

Write $f(v_i) = \lambda_i v_1 + h_i$. Let K be a hyperplane containing $v_1, \dots, v_{i-1}$ but not v_i. Let φ be a linear form which is zero on K but with $\varphi(v_i) = -\lambda_i$. Clearly $t_{\varphi, v_1} = \mathrm{Id} + \varphi v_1$ fixes $v_1, \dots, v_{i-1}$ and maps v_i to $v_i - \lambda_i v_1$. So $f \circ t_{\varphi, v_1}$ fixes v_1, maps $v_2, \dots, v_{i-1}$ to H, and maps v_i to $h_i \in H$.

At $i = n$ we get that f fixes v_1 and maps H to H. So f restricts to an element $g \in \mathrm{GL}(H)$; since $f(v_1) = 1$ and the determinant is 1, $g \in \mathrm{SL}(H)$. We apply induction there; g is a product of transvections of H. These naturally extend (by identity on v_1) to transvections of V. So f is a product of transvections of V. □

The underlying principle in the previous proof is that given independent vectors $v_1, \dots, v_i$, there exists a transvection fixing each v_k for $k < n$ and mapping v_i to any vector not in $\langle v_1, \dots, v_{i-1} \rangle$. This is better explained in the language of group actions.

Lemma 1.2.14. *Let $i \leqslant n-1$ and X_i be the set of* linearly independent *families $(v_1, \dots, v_i)$ of vectors. Then* $\mathrm{SL}(V)$ *is transitive on X_i.*

Proof. This is a part of the argument above. □

Remark 1.2.15.

- It is *not* true that $\mathrm{SL}(V)$ is transitive on the set of *all* families $(v_1, \dots, v_i)$ (that is, on V^i). It is *not* true that $\mathrm{SL}(V)$ is transitive on the set of families of n distinct vectors.

  The reason is that an element in $\mathrm{SL}(V)$ cannot map a linearly independent family to a linearly dependent one.

- It is *not* true that $\mathrm{SL}(V)$ is transitive on X_n (defined as one imagine, and which is the set of bases of V), because $\mathrm{SL}(V)$ must preserve the determinant of a basis.

  However, adjusting a coefficient, something can be done with $i = n$. This requires going to the projective space.

Before, we introduce some terminology.

Definition 1.2.16. Suppose $|\Omega| \geqslant 2$. We say that G acts *doubly transitively* on Ω if for any pairs (α_1, β_1) and (α_2, β_2) in $\Omega \times \Omega$ satisfying $\alpha_1 \neq \beta_1$ and $\alpha_2 \neq \beta_2$, there is $g \in G$ such that $g \cdot \alpha_1 = \alpha_2$ and $g \cdot \beta_1 = \beta_2$.

Exercise 1.2.17. *G is doubly transitive on Ω iff G is transitive on Ω and for all $\alpha \in \Omega$, G_α is transitive on $\Omega \backslash \{\alpha\}$.*

Definition 1.2.18. G acts *n-transitively* on Ω if for any tuples $(\alpha_1, \dots, \alpha_n)$ and $(\beta_1, \dots, \beta_n)$ of pairwise distinct elements of Ω, there is $g \in G$ such that $g \cdot \alpha_i = \beta_i$ for all $i = 1, \dots, n$.

Exercise 1.2.19. *State and prove an equivalent definition in terms of stabilisers.*

1.2.3 Projective action and transitivity

Recall that for $v \in V\backslash\{0\}$, $\langle v\rangle = \mathbb{F}v$ denotes the line through the origin spanned by v.

Definition 1.2.20. The *projective space* associated to V, denoted $\mathbb{P}_{n-1}(V)$ or simply $\mathbb{P}(V)$, is the set of all distinct lines $\langle v\rangle$. We then call $\langle v\rangle$ a *projective point*.

Be very careful that when viewed as a subset of V, $\langle v\rangle$ is of course a line; but when we view it as an element of $\mathbb{P}(V)$, it is now a *point*.

Exercise 1.2.21. *Let $\sim$ be the equivalence relation on $V\backslash\{0\}$: $v_1 \sim v_2$ iff there is $\lambda \in \mathbb{F}^*$ with $v_2 = \lambda v_1$. Construct a natural bijection between $\mathbb{P}_{n-1}(V)$ and $V\backslash\{0\}/\sim$.*

Lemma 1.2.22. *If $|\mathbb{F}| = q$ and $\dim V = n$ then $|\mathbb{P}(V)| = \frac{q^n-1}{q-1}$.*

Proof. V has exactly $q^n - 1$ non-zero vectors, and each line has exactly $q - 1$ points. □

Lemma 1.2.23. *The formula $g\cdot\langle v\rangle = \langle g\cdot v\rangle$ for $(g, \langle v\rangle) \in \mathrm{GL}(V)\times\mathbb{P}(V)$ defines an action of $\mathrm{GL}(V)$ on $\mathbb{P}(V)$. The kernel is $Z(\mathrm{GL}(V)) = \{\lambda\,\mathrm{Id} : \lambda \in \mathbb{F}^*\} \simeq \mathbb{F}^*$.*

Proof. If $\langle v_1\rangle = \langle v_2\rangle$ then v_1 and v_2 are collinear in V, say $v_2 = \lambda v_1$. Then $g \in \mathrm{GL}(V)$ will map v_2 to $\lambda g(v_1)$, so $\langle g \cdot v_2\rangle = \langle g \cdot v_1\rangle$, and the action is well-defined.

It is clear that a vector dilation $\lambda\,\mathrm{Id}$ will stabilize every projective point. Conversely, if g is in the kernel of the action, then for any $v \in V\backslash\{0\}$, $g \cdot v \in \langle v\rangle$. We saw in the second proof of Proposition 1.1.5 that g is of the form $\lambda\,\mathrm{Id}$. □

All this strongly suggests to consider the following quotients.

Definition 1.2.24.

- Let $\mathrm{PGL}(V) = \mathrm{GL}(V)/Z(\mathrm{GL}(V))$ be the *projective general linear group*.
- Let $\mathrm{PSL}(V) = \mathrm{SL}(V)/Z(\mathrm{SL}(V))$ be the *projective special linear group*.

(One defines $\mathrm{PGL}(n, \mathbb{F})$, $\mathrm{PGL}(n, q)$, and so on, as one easily imagines.)

By construction, the action of $\mathrm{PGL}(V)$ on $\mathbb{P}(V)$ is now faithful. Similarly, $\mathrm{PSL}(V)$ acts faitfhully on $\mathbb{P}(V)$.

Lemma 1.2.25. *Let Y_i be the set $\{(\langle v_1\rangle, \ldots, \langle v_i\rangle) : (v_1, \ldots, v_i) \in V^i$ is linearly independent$\}$. Then $\mathrm{PSL}(V)$ acts transitively on Y_i for all $i \leqslant \dim V$.*

Proof. It suffices to do it for $i = n$. So take any basis $v_1, \ldots, v_n$; we claim that we can map $(\langle e_1\rangle, \ldots, \langle e_n\rangle)$ to $(\langle v_1\rangle, \ldots, \langle v_n\rangle)$. Let $\lambda \in \mathbb{F}^*$ be yet undetermined. Let f_λ map e_i to v_i for $i \leqslant n-1$ and map e_n to λv_n. Then by n-linearity of the determinant function, $\det f_\lambda = \lambda \det f_1$. But since f_1 sends a basis to another basis, it has non-zero determinant. So with $\lambda = \frac{1}{\det f_1}$, one finds $\det f_\lambda = 1$, that is $f_\lambda \in \mathrm{SL}(V)$.

The image of f_λ in $\mathrm{PSL}(V)$ sends $\langle e_i\rangle$ to $\langle v_i\rangle$ for all $i = 1 \ldots n$. □

Here again, one should *not* think that $\mathrm{PSL}(V)$ is n-transitive on $\mathbb{P}(V)$, because linearly independent families can be sent only to linearly independent families. Howevere there is an interesting corollary which will be extremely useful later.

Corollary 1.2.26. $\mathrm{PSL}(V)$ *is* 2*-transitive on* $\mathbb{P}(V)$*.*

Proof. The definition of 2-transitivity (1.2.16 above) was that any pair of *distinct* elements $\langle v_1\rangle \neq \langle v_2\rangle$ can be sent to any other pair of *distinct* elements. But $\langle v_1\rangle \neq \langle v_2\rangle$ is equivalent to (v_1, v_2) being linearly independent. So this is just a special case of the previous lemma with $i = 2$. □

Lemma 1.2.27.

- $|\mathrm{PGL}(n,q)| = |\mathrm{SL}(n,q)| = \frac{1}{q-1}|\mathrm{GL}(n,q)| = q^{\frac{n(n-1)}{2}} \prod_{i=2}^{n}(q^i - 1)$
- *Let* $d = (n, q-1)$ *(standard notation for gcd). Then* $|\mathrm{PSL}(n,q)| = \frac{1}{d}|\mathrm{SL}(n,q)| = \frac{1}{d} q^{\frac{n(n-1)}{2}} \prod_{i=2}^{n}(q^i - 1)$

Proof.

- Since $Z(\mathrm{GL}(V)) = \mathbb{F}^* \,\mathrm{Id}$, one sees that $|\mathrm{PGL}(n,q)| = \frac{1}{q-1}|\mathrm{GL}(n,q)|$.
- $Z(\mathrm{SL}(V)) = \{\lambda \,\mathrm{Id} : \lambda^n = 1\}$. Recall from field theory that $\mathbb{F}_q^* \simeq C_{q-1}$, the cyclic group with $q-1$ elements. So the equation $\lambda^n = 1$ has exactly $(n, q-1) = d$ solutions in $\mathbb{F}_q$, meaning that $|\mathrm{PSL}(n,q)| = \frac{1}{d}|\mathrm{SL}(n,q)|$. □

Remark 1.2.28. Although $|\mathrm{PGL}(n,q)| = |\mathrm{SL}(n,q)|$, these groups are not in general isomorphic. Suppose $(n,q) \neq 1$ and take $\lambda \in \mathbb{F}\backslash\{0,1\}$ such that $\lambda^n = 1$. Then $\lambda I_n \in \mathrm{SL}(n,q)$, and actually $\lambda I_n \in Z(\mathrm{SL}(n,q))$. But $\mathrm{PGL}(n,q)$ is always centerless. Here is why.

Let $f \in \mathrm{GL}(V)$ map to a central element of $\mathrm{PGL}(V)$. This means that for any $g \in \mathrm{GL}(V)$, the commutator $[f,g] = fgf^{-1}g^{-1}$ lies in $Z(\mathrm{GL}(V))$, i.e. for any $g \in \mathrm{GL}(V)$ there is $\lambda_g \in \mathbb{F}^*$ with $fgf^{-1}g^{-1} = \lambda_g \,\mathrm{Id}$. Suppose that $f \notin Z(\mathrm{GL}(V))$. Then there is $v \in V$ such that $(v, f(v))$ is linearly independent. Let $g \in \mathrm{GL}(V)$ map v to $v + f(v)$ and fix $f(v)$ (this is possible). Then:

$$fgf^{-1}g^{-1}(f(v)) = fgf^{-1}(f(v)) = fg(v) = f(v+f(v)) = f(v)+f^2(v) = \lambda_g f(v)$$

So $(f(v), f^2(v))$ is not linearly independent, and neither is $(v, f(v))$: a contradiction.

End of Lecture 3.

Lecture 4 (Study of $\mathrm{SL}(V)$)

1.2.4 Study of $\mathrm{SL}(V)$

Conjugacy of transvections

Since $\mathrm{SL}(V)$ is a group it acts on itself by conjugation. We are now interested in what it does to the transvection $t_{\varphi,a}$. But in order to write the answer it is relevant to understand how $\mathrm{GL}(V)$ can act on V^*. We begin with this.

Definition 1.2.29. The *dual action* of $\mathrm{GL}(V)$ on V^* is defined by: $g * \varphi = \varphi \circ g^{-1}$ for $(g, \varphi) \in \mathrm{GL}(V) \times V^*$, i.e. $(g * \varphi)(v) := \varphi(g^{-1}v)$.

It is the case indeed that $\varphi \circ g^{-1}$ is a linear form, and $(gh) * \varphi = \varphi \circ (gh)^{-1} = \varphi \circ h^{-1} \circ g^{-1} = g * (h * \varphi)$. From now on, we simply write $g\varphi$ for $g * \varphi$. There is no risk of confusion since "$g \circ \varphi$" (attempted composition) is meaningless.

Exercise 1.2.30. *Prove that* $\ker g(\varphi) = g(\ker \varphi)$.

Exercise 1.2.31. *Let V have basis $\mathcal{B} = (e_1, \ldots, e_n)$ and let $\mathcal{B}^* = (e_1^*, \ldots, e_n^*)$ be the dual basis of V^*, which was discussed in the proof of Proposition 1.2.5. Show that the matrix representing the action of g on V^* in $\mathcal{B}^*$ is the inverse transpose of* $\mathrm{Mat}_{\mathcal{B}}(g)$.

Lemma 1.2.32. *If $g \in \mathrm{GL}(V)$ then $gt_{\varphi,a}g^{-1} = t_{g\varphi,ga}$.*

Proof. Let $v \in V$ be any vector. Since g is onto, write $v = g(x)$. Then:

$$\begin{aligned} gt_{\varphi,a}g^{-1}(v) &= gt_{\varphi,a}(x) \\ &= g(x + \varphi(x)a) \\ &= v + \varphi(x)g(a) \\ &= v + \varphi(g^{-1}v)ga \\ &= v + (g * \varphi)(v) \cdot ga \\ &= t_{g\varphi,ga}(v) \end{aligned}$$

□

Remark 1.2.33. Here again, the natural setting for Lemma 1.2.32 would be the tensor product of V^* and V.

Hence when we conjugate a transvection we get another transvection. This yields one more question: are transvections conjugate?

Lemma 1.2.34.

- *Transvections are conjugate in* $\mathrm{GL}(V)$.
- *If $n \geqslant 3$ then transvections are conjugate in* $\mathrm{SL}(V)$.

Proof. Let $t_{\varphi,a}$ and $t_{\psi,b}$ be transvections. Write $H = \ker \varphi$ and $K = \ker \psi$. Let $a', b' \in V$ be such that $\varphi(a') = \psi(b') = 1$.

Since $a \in H$, there exists a basis $(a_1 = a, \ldots, a_{n-1})$ of H extending a. Then $(a, a_2, \ldots, a_{n-1}, a')$ is a basis of V. And similarly, there exists a basis $(b, b_2, \ldots, b_{n-1}, b')$ of V where the $(n-1)$ first vectors span K.

Let $g \in \mathrm{GL}(V)$ send $(a, a_2, \ldots, a_{n-1}, a')$ to $(b, b_2, \ldots, b_{n-1}, b')$. Then $g(H) = K$. In particular,

$$\ker(g \cdot \varphi) = \{v \in V : \varphi(g^{-1}(v)) = 0\} = \{v \in V : g^{-1}(v) \in H\} = g(H) = K$$

We know from Lemma 1.2.8 that there is $\lambda \in \mathbb{F}^*$ with $g \cdot \varphi = \lambda\psi$. But let us apply this to b':

$$\lambda = \lambda\psi(b') = (g \cdot \varphi)(b') = \varphi(g^{-1}(ga')) = \varphi(a') = 1$$

Hence $g \cdot \varphi = \psi$, and by construction $g \cdot a = b$. It follows from Lemma 1.2.32 that:

$$gt_{\varphi,a}g^{-1} = t_{g\varphi,ga} = t_{\psi,b}$$

If $n \geqslant 3$, then playing with $(b_2, \ldots, b_{n-1})$ (which is not the empty tuple), one can choose $g \in \mathrm{SL}(V)$. □

Remark 1.2.35. If $n = 2$ then the conjugacy classes of transvections in $\mathrm{SL}(V)$ are in bijection with the cosets of $\{a^2 : a \in \mathbb{F}^*\}$ in $\mathbb{F}^*$.

We know that every transvection matrix gives rise to a transvection map, and that for every transvection map τ there is a basis in which the matrix of τ is of the form $I_2 + \lambda E_{1,2}$. In particular, every transvection matrix is conjugate to a transvection of the form $I_2 + \lambda E_{1,2}$, and it suffices to determine when two such are conjugate in $\mathrm{SL}(V)$.

Now:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & \lambda \\ & 1 \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} 1 - \lambda ac & \lambda a^2 \\ -\lambda c^2 & 1 + \lambda ac \end{pmatrix}$$

which has the form $I_2 + \mu E_{1,2}$ iff $c = 0$, in which case we find $I_2 + \lambda a^2 E_{1,2}$. So we can conjugate $I_2 + \lambda E_{1,2}$ only to transvections having λ in the same class modulo the squares of $\mathbb{F}^\times$.

Here is a more geometric argument. Every transvection is conjugate (in $\mathrm{SL}(V)$) to a transvection with fixed hyperplane. So it will suffice to study the set $\{ft_{\varphi,a}f^{-1} : f \in \mathrm{SL}(V), \ker(f\varphi) = \ker(\varphi)\}$. But $f\varphi$ and φ have the same kernel iff there is $\lambda \in \mathbb{F}^*$ with $f\varphi = \lambda\varphi$, in which case $fa = \lambda a$. Then $ft_{\varphi,a}f^{-1} = t_{\lambda\varphi,\lambda a} = t_{\varphi,\lambda^2 a}$. So the only conjugates of the desired form $t_{\varphi,b}$ satisfy $b \in (\mathbb{F}^\times)^2 a$.

The remark shows in particular that transvections are conjugate in $\mathrm{SL}_2(\mathbb{C})$ (since every element there is a square).

Lemma 1.2.36. *Write $t_{0,a} = t_{\varphi,0} = \mathrm{Id}$ (which is consistent). Then:*

- $t_{\varphi,a}t_{\psi,a} = t_{\varphi+\psi,a}$
- $t_{\varphi,a}t_{\varphi,b} = t_{\varphi,a+b}$
- $t_{\varphi,a}^{-1} = t_{-\varphi,a} = t_{\varphi,-a}$.

Proof. All obvious. □

As a consequence, and although transvections do *not* form a group, both $\{t_{\varphi,a_0} : \varphi \in V^*\}$ (for fixed a_0) and $\{t_{\varphi_0,a} : a \in V\}$ (for fixed φ_0) are subgroups of $\mathrm{SL}(V)$.

The commutator subgroup of $\mathrm{SL}(V)$

Corollary 1.2.37. *If $n \geqslant 3$ or $|\mathbb{F}| \geqslant 4$, then $(\mathrm{SL}(V))' = \mathrm{SL}(V)$.*

Proof. We know from Proposition 1.2.13 (or even Proposition 1.1.13) that $\mathrm{SL}(V)$ is generated by transvections. Moreover, even if $n = 2$, then transvections are $\mathrm{GL}(V)$-conjugate by Lemma 1.2.34. Since the commutator subgroup is not only normal but also characteristic as a subgroup, it suffices to find one transvection in $(\mathrm{SL}(V))'$.

Take a hyperplane H, $a \in H$, and $\varphi \in V^*$ with $\ker\varphi = H$. Let $b \notin H$, so that $V = H \oplus \langle b \rangle$.

- First suppose $n \geqslant 3$. Then there is $f \in \mathrm{SL}(V)$ with $f(\varphi) = \varphi$ but $f(a) \neq a$: say for instance that f acts as the identity on $\langle b \rangle$ and fixes H setwise but

moves a (this is possible as the dimension is $\geqslant 3$). Hence by Lemmas 1.2.32 and 1.2.36:

$$ft_{\varphi,a}f^{-1}t_{\varphi,a}^{-1} = t_{f(\varphi),f(a)}t_{\varphi,-a} = t_{\varphi,f(a)-a}$$

Since $f(a) - a \neq 0$, we thus have a transvection in $(\mathrm{SL}(V))'$.

- Now suppose $n = 2$. Then the former construction fails: $\dim H = 1$ so $H = \langle a \rangle$ and $f(a) = \lambda a$. If we want f to fix b, then using $\det f = 1$ we see that $\lambda = 1$ and $f = \mathrm{Id}$. We can remedy this as follows.

  Take $\lambda \in \mathbb{F}^*$ with $\lambda^2 \neq 1$: this is possible by assumption on $|\mathbb{F}|$ in this case. Let f map a to λa and b to $\frac{1}{\lambda}b$. (Note that $f \neq \mathrm{Id}$ since $\lambda^{-1} \neq \lambda$ in $\mathbb{F}$.) Then $(f \cdot \varphi)(a) = \varphi(\lambda^{-1}(a)) = 0$ and $(f \cdot \varphi)(b) = \lambda\varphi(b)$, so $f \cdot \varphi = \lambda\varphi$. Hence:

  $$ft_{\varphi,a}f^{-1}t_{\varphi,a}^{-1} = t_{\lambda\varphi,\lambda a}t_{\varphi,-a} = t_{\varphi,\lambda^2 a - a}$$

  which is a non-trivial transvection since $\lambda^2 \neq 1$. Here again, $(\mathrm{SL}(V))'$ contains a non-trivial transvection, so $(\mathrm{SL}(V))' = \mathrm{SL}(V)$. □

Remark 1.2.38. In case you did not follow how to retrieve a transvection in the derived subgroup, you may want to view this in matrix form. Consider the transvection matrices $I + E_{1,2}$ and $I + E_{1,3}$. Then:

$$\begin{aligned}
&(I + E_{1,2})(I + E_{2,3})(I + E_{1,2})^{-1}(I + E_{2,3})^{-1} \\
&= (I + E_{1,2} + E_{2,3} + E_{1,3})(I - E_{1,2} - E_{2,3} + E_{1,3}) \\
&= I - E_{1,2} - E_{2,3} + E_{1,3} + E_{1,2} - E_{1,3} + E_{2,3} + E_{1,3} \\
&= I + E_{1,3}
\end{aligned}$$

which is a transvection matrix, and an element of $(\mathrm{SL}(n, \mathbb{F}))'$. This works only if $n \geqslant 3$.

Exercise 1.2.39. *Write a matrix computation yielding a transvection if $n = 2$ and $|\mathbb{F}| \geqslant 4$.*

Remark 1.2.40. If $\mathbb{F} = \mathbb{F}_2$, then $\mathrm{SL}_2(\mathbb{F}_2)$ has order 6, so it is solvable, and certainly $\mathrm{SL}_2(\mathbb{F}_2)' < \mathrm{SL}_2(\mathbb{F}_2)$. If $\mathbb{F} = \mathbb{F}_3$, then $\mathrm{SL}_2(\mathbb{F}_3)$ has order 24, and the same argument works.

It is an exercise to find the actual isomorphism type of $\mathrm{SL}_2(\mathbb{F}_2)$ and $\mathrm{SL}_2(\mathbb{F}_3)$.

As an application of Corollary 1.2.37 we give another proof of $Z(\mathrm{PGL}(V)) = 1$ (which was already discussed in Remark 1.2.28).

Proof. Let $f \in \mathrm{GL}(V)$ map to an element of $Z(\mathrm{PGL}(V))$; then for any $g \in \mathrm{GL}(V)$, $[f, g] \in Z(\mathrm{GL}(V))$ so there is $\lambda_g \in \mathbb{F}^*$ with $[f, g] = \lambda_g \,\mathrm{Id}$. Consider the function $\Lambda : \mathrm{GL}(V) \to \mathbb{F}^*$ with $\Lambda(g) = \lambda_g$. Observe that for $g, h \in \mathrm{GL}(V)$:

$$\lambda_{gh}\,\mathrm{Id} = [f, gh] = [f, g] \cdot g[f, h]g^{-1} = \lambda_g\,\mathrm{Id} \cdot \lambda_h\,\mathrm{Id}$$

so the map Λ is a group homomorphism. Since the values are in an abelian group, $(\mathrm{GL}(V))' \leqslant \ker \Lambda$. Now by Corollary 1.2.37, $\mathrm{SL}(V) \leqslant (\mathrm{GL}(V))'$, so Λ is constant on $\mathrm{SL}(V)$.

This means that for any $g \in \mathrm{SL}(V)$, $[f, g] = \mathrm{Id}$. We then go back to the (second) proof of Proposition 1.1.5, and notice that g there constructed could be taken in $\mathrm{SL}(V)$. So the argument gives $f \in Z(\mathrm{GL}(V))$, and therefore $Z(\mathrm{PGL}(V)) = 1$. □

Lecture 5 (The Simplicity of PSL(V))

1.3 Simplicity of PSL(V)

1.3.1 Iwasawa's criterion for simplicity

Today we introduce more abstract group-theoretic tools.

Definition 1.3.1. Suppose that G is transitive on Ω. A *block of imprimitivity* for G in Ω is a subset $\Delta \subseteq \Omega$ with $|\Delta| \geqslant 2$, $\Delta \neq \Omega$, such that for all $g \in G$, either $g \cdot \Delta = \Delta$ or $g \cdot \Delta \cap \Delta = \varnothing$.

The group G (more precisely, the action) is *primitive* if there are no blocks of imprimitivity.

Exercise 1.3.2. *Prove that G is primitive iff there is no equivalence relation on Ω compatible with the action of G, apart from equality and the trivial relation (where all points are equivalent).*

Proposition 1.3.3. *If G is doubly transitive on Ω, then G is primitive.*

Proof. Suppose that $\Delta \subseteq \Omega$ satisfies $\Delta \neq \Omega$ and $|\Delta| \geqslant 2$. Let $\alpha \neq \beta$ be *distinct* elements of Δ, and let $\gamma \in \Omega \backslash \Delta$.

By double transitivity, there is $g \in G$ such that $g \cdot \alpha = \alpha$ and $g \cdot \beta = \gamma$. Then $\alpha \in g \cdot \Delta \cap \Delta$ which is therefore non-empty. But $\gamma \in g \cdot \Delta \backslash \Delta$, so $g \cdot \Delta \neq \Delta$.

It follows that Δ is not a block of imprimitivity. □

Proposition 1.3.4. *Suppose that G acts primitively on Ω, $N \trianglelefteqslant G$, and N is* not *contained in the kernel of the action. Then N is transitive.*

Proof. Since N is not in the kernel, there is $\alpha \in \Omega$ which is not stabilized by all of N. Let $\Delta = N \cdot \alpha$; by construction, $|\Delta| \geqslant 2$.

We focus on the action of N on Ω. Then $\Delta = N \cdot \alpha = \mathrm{orb}^N(\alpha)$ is an N-orbit. Now for all $g \in G$, then $g \cdot \Delta = g \cdot (N \cdot \alpha) = N \cdot (g \cdot \alpha) = \mathrm{orb}^N(g \cdot \alpha)$ is also an N-orbit (perhaps the same). Since N-orbits form a partition of Ω, either $\Delta = g \cdot \Delta$ or $\Delta \cap g \cdot \Delta = \varnothing$, and this holds of any $g \in G$.

But the action of G on Ω is primitive, and $|\Delta| \geqslant 2$. It follows $\Delta = \Omega$, which exactly means that N is transitive on Ω. □

Proposition 1.3.5. *Suppose that G acts on Ω, and $H \leqslant G$ is transitive on Ω. If $\alpha \in \Omega$, then $G = H \cdot G_\alpha$.*

Proof. Let $g \in G$. By transitivity of H, there is $h \in H$ with $h \cdot \alpha = g \cdot \alpha$. Hence $h^{-1}g \in \mathrm{Stab}_G(\alpha)$, and $g \in h \,\mathrm{Stab}_G(\alpha) \subseteq HG_\alpha$. □

Proposition 1.3.6. *Suppose that G acts transitively on a finite set Ω. If $|\Omega| > 1$ and $\alpha \in \Omega$, then G_α is maximal (as a proper subgroup) iff G is primitive (in its action on Ω).*

Proof.

- Suppose G_α is *not* maximal. So there exists a subgroup H with $G_\alpha < H < G$. Let $\Delta = \{h \cdot \alpha : h \in H\}$. Since $H > G_\alpha$, one has $|\Delta| \geqslant 2$.

  We claim that $\Delta \neq \Omega$. Otherwise H is transitive on Ω, which implies by Proposition 1.3.5: $G = HG_\alpha \subseteq H \cdot H = H$, a contradiction.

  We now prove that Δ is a block of imprimitivity. Let $g \in G$. If $\Delta \cap g \cdot \Delta \neq \varnothing$, there are $h_1, h_2 \in H$ such that $h_1 \cdot \alpha = g \cdot (h_2 \cdot \alpha) = gh_2 \cdot \alpha$. Hence $h_1^{-1} g h_2 \in G_\alpha \leqslant H$, and it follows $g \in H$. But in that case, since Δ is an H-orbit, $g \cdot \Delta = \Delta$.

  This proves that Δ is a block of imprimitivity; hence G is not primitive.

- Suppose conversely that G is not primitive, and let $\Delta \subseteq \Omega$ be a block of imprimitivity.

  Remember that G was assumed to be transitive. In particular, all stabilisers are conjugate in G; if we want to prove that G_α is maximal, we may without loss of generality suppose $\alpha \in \Delta$. Let $H = \{g \in G : g \cdot \Delta = \Delta\}$, the *setwise* stabiliser of Δ. Then H is a subgroup of G.

  Suppose $g \in G_\alpha$. Then $g \cdot \alpha = \alpha \in g \cdot \Delta \cap \Delta$, so by imprimitivity, $g \cdot \Delta = \Delta$. This means $G_\alpha \leqslant H$.

  We claim that H acts transitively on Δ. Let $\beta, \gamma \in \Delta$. By transitivity *of* G, there is $g \in G$ with $g \cdot \beta = \gamma$. In particular, $\gamma \in g \cdot \Delta \cap \Delta$, so by imprimitivity $g \cdot \Delta = \Delta$, exactly meaning $g \in H$. Hence there is $g \in H$ sending β to γ: H is transitive on Δ.

  Remember that $G_\alpha \leqslant H$; this simply means that $H_\alpha = G_\alpha$, and by transitivity on Δ, we get $|\Delta| = [H : G_\alpha] > 1$.

  Finally, by transitivity of G on Ω, $[G : G_\alpha] = |\Omega| > |\Delta| = [H : G_\alpha]$, so $G > H > G_\alpha$, and G_α is not maximal in G. □

Notice that only the very end of the second half of the argument did require finiteness of Ω.

Theorem 1.3.7 (Iwasawa)**.** *Suppose G acts faithfully and primitively on Ω and $G' = G$. Let $\alpha \in \Omega$. Suppose there is an abelian normal subgroup K of G_α such that $G = \langle K^g : g \in G \rangle$. Then G is simple.*

Proof. Let $N \trianglelefteqslant G$ be a normal subgroup; suppose $N \neq 1$. Then by Proposition 1.3.4, N is transitive on Ω, and it follows by Proposition 1.3.5: $G = NG_\alpha$. Since K is normal in G_α, $KN \trianglelefteqslant NG_\alpha = G$.

Hence for all $g \in G$, $K^g \leqslant (KN)^g = KN$. Therefore $G = \langle K^g : g \in G \rangle \leqslant KN$.

Now by the classical isomorphism theorems, $G/N \simeq KN/N \simeq K/(K \cap N)$ is abelian. In particular, $(G/N)' = 1$ which means $G' \leqslant N$. But $G' = G$ by assumption, so $G = N$, which proves simplicity of G. □

1.3.2 Simplicity of $\mathrm{PSL}(V)$

Notation 1.3.8. For $u \in V \backslash \{0\}$ let $K_u = \{t_{\varphi,u} : \varphi \in V^* : \varphi(u) = 0\}$.

Theorem 1.3.9.

- *K_u is an abelian, normal subgroup of* $\mathrm{Stab}_{\mathrm{SL}(V)}(u)$*;*

- $\langle K_u^f : f \in \mathrm{SL}(V)\rangle = \mathrm{SL}(V)$.

Proof. First observe that since $u \in \ker\varphi$, the transvection $t_{\varphi,u}$ fixes u. So K_u is a subset of the stabiliser. Now $t_{\varphi,u}t_{\psi,u} = t_{\varphi+\psi,u}$ by Lemma 1.2.36, so K_u is an abelian subgroup of $\mathrm{Stab}_{\mathrm{SL}(V)}(u)$. Moreover, if $f \in \mathrm{Stab}(u)$, then $ft_{\varphi,u}f^{-1} = t_{f\cdot\varphi,u}$ where $(f\cdot\varphi)(u) = \varphi(f^{-1}(u)) = 0$. This shows that K_u is a normal subgroup of $\mathrm{Stab}\, u$.

We move to the generation part. Be careful, when reading Lemma 1.2.34 again, that transvections are not all $\mathrm{SL}(V)$-conjugate (this can fail if $n = 2$). But by transitivity of $\mathrm{SL}(V)$ on $V\backslash\{0\}$, every transvection is $\mathrm{SL}(V)$-conjugate to one of K_u. As a consequence, the group suggested contains all transvections. It then equals $\mathrm{SL}(V)$ by Proposition 1.2.13. □

Theorem 1.3.10. $\mathrm{PSL}(n,q)$ *is simple except for* $\mathrm{PSL}(2,2)$ *and* $\mathrm{PSL}(2,3)$.

Proof. This is an application of Iwasawa's criterion, Theorem 1.3.7. As we know, $\mathrm{PSL}(V)$ acts faithfully and 2-transitively on $\mathbb{P}(V)$, in particular it is faithful and primitive by Proposition 1.3.3.

By Corollary 1.2.37, $(\mathrm{SL}(V))' = \mathrm{SL}(V)$ so $(\mathrm{PSL}(V))' = \mathrm{PSL}(V)$, except for $\mathrm{PSL}(2,2)$ and $\mathrm{PSL}(2,3)$, which are the cases we excluded by assumption.

For $L \in \mathbb{P}(V)$ a projective point (see Definition 1.2.20 again if necessary), let $u \in V$ be such that $L = \langle u\rangle$. Let $X_L = \overline{K_u}$. This is well-defined; clearly X_L is an abelian normal subgroup of $\mathrm{Stab}\,\mathrm{PSL}(V)(L)$ and $\mathrm{PSL}(V) = \langle fX_Lf^{-1} : f \in \mathrm{PSL}(V)\rangle$. This is what we need for Theorem 1.3.7. □

End of Lecture 5.

Lecture 6 (The finite case)

1.3.3 The Finite Case

Lemma 1.3.11. $\mathrm{PSL}(2,2) \simeq S_3$ *and* $\mathrm{PSL}(2,3) \simeq A_4$.

Proof. First note that if $\mathbb{F} = \mathbb{F}_2$, then $\mathrm{GL}(n,2) = \mathrm{SL}(n,2)$, and $Z(\mathrm{GL}(n,2)) = \{\mathrm{Id}\}$, hence $\mathrm{PSL}(n,2) \simeq \mathrm{GL}(n,2)$. In particular $\mathrm{PSL}(2,2) \simeq \mathrm{GL}(2,2)$ has 6 elements. It is easily seen that $\mathrm{PSL}(2,2)$ is non-abelian, whence isomorphic to S_3.

The $n = 3$ case is hardly more subtle. Observe that $|\mathrm{PSL}(2,3)| = 12$, and finish as an exercise. □

Remark 1.3.12. $\mathrm{PSL}(2,q)$ acts faithfully on $\mathbb{P}(V)$ (where $\dim V = 2$). Now $|\mathbb{P}(V)| = q+1$, so $\mathrm{PSL}(2,q)$ embeds into S_{q+1}.

1. If $q = 2$, then $\mathrm{PSL}(2,2)$ embeds into S_3; one finds $\mathrm{PSL}(2,2) \simeq S_3$.
2. If $q = 3$, then $\mathrm{PSL}(2,3)$ embeds into S_4; one finds $\mathrm{PSL}(2,3) \simeq A_4$.
3. If $q = 4$, then $\mathrm{PSL}(2,4)$ embeds into S_5; one finds $\mathrm{PSL}(2,4) \simeq S_5$.
4. There is a unique simple group of order 60. Hence $\mathrm{PSL}(2,5) \simeq \mathrm{PSL}(2,4) \simeq S_5$.
5. There is a unique simple group of order 168. Hence $\mathrm{PSL}(2,7) \simeq \mathrm{PSL}(3,2)$.

6. $|A_8| = |\mathrm{PSL}(3,4)| = |\mathrm{PSL}(4,2)| = 2^6 \cdot 3^2 \cdot 5 \cdot 7$. Although it is the case that $\mathrm{PSL}(4,2) \simeq A_8$, one has $\mathrm{PSL}(3,4) \not\simeq \mathrm{PSL}(4,2)$.

   Let indeed
   $$U_1 = \begin{pmatrix} 1 & * & * \\ & 1 & * \\ & & 1 \end{pmatrix}$$
   and
   $$U_2 = \begin{pmatrix} 1 & * & * & * \\ & 1 & * & * \\ & & 1 & * \\ & & & 1 \end{pmatrix}$$
   It can be shown that U_1 is a Sylow 2-subgroup of $\mathrm{PSL}(3,4)$, and U_2 a Sylow 2-subgroup of $\mathrm{PSL}(4,2)$.

   It is an exercise that for the matrix group:
   $$U = \begin{pmatrix} 1 & & * \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$
   one has:
   $$Z(U) = \begin{pmatrix} 1 & & * \\ & \overset{0}{\ddots} & \\ & & 1 \end{pmatrix}$$
   It follows that $|Z(U_1)| = 4$ whereas $|Z(U_2)| = 2$. In particular, U_1 and U_2 are not isomorphic; neither are $\mathrm{PSL}(3,4)$ and $\mathrm{PSL}(4,2)$.

7. One has $\mathrm{PSL}(2,9) \simeq A_6$.

8. One also has $\mathrm{PSL}(4,2) \simeq A_8$.

There are no other isomorphisms between $\mathrm{PSL}(n,q)$ and A_m.

End of Lecture 6.

Week 2: The Symplectic Group

Lecture 7 (Introduction; Sesquilinear Forms)

Clasical groups are groups associated to some familiar ("classical") linear geometries. A linear geometry consists in a vector space plus some extra structure.

Example 2.0.13. Let V be a vector space over a field $\mathbb{F}$.

1. We simply consider V (with no extra structure). The group associated to this geometry is $\mathrm{GL}(V) = \{f : V \to V \text{ a linear bijection}\}$, the general linear group.

2. We now consider V with the notion of the determinant of a basis. The group associated to this structure is $\mathrm{SL}(V) = \{f \in \mathrm{GL}(V) : \det(f) = 1\}$, the special linear group.

3. Consider the real vector space $\mathbb{R}^3$ with the usual Euclidean distance. The group associated to this structure is $O(\mathbb{R}^3) = \{f \in \mathrm{GL}(\mathbb{R}^3) : \forall x \in \mathbb{R}^3, \|f(x)\| = \|x\|\}$, the orthogonal group of $\mathbb{R}^3$.

4. One could consider $\mathbb{R}^3$ with the distance and the determinant. The resulting group is $\mathrm{SO}(\mathbb{R}^3) = \{f \in O(\mathbb{R}^3) : \det f = 1\} = \{f \in \mathrm{SL}(\mathbb{R}^3) : f \text{ preserves distances}\}$.

These are only examples, and one sees that one needs a general setting for the notion of "extra structure". We shall explain this during the first two lectures, dedicated to studying *sequilinear forms*; these lectures are completely independent from Week 1. Then we shall study an example, the geometry of so-called *symplectic forms*.

2.1 Sesquilinear Forms

We wish to add some structure of metric type on our vector space V. Of course one has first in mind scalar products, but we must be more general. First, because even in the metric case one must go a little beyond: over $\mathbb{F} = \mathbb{C}$, the usual dot product of $\mathbb{C}^n$ is *not* bilinear. Also, because of some other geometries which arise naturally. To cover all cases our setting is that of sequilinear forms.

2.1.1 A classification result

Definition 2.1.1. Let σ be an automorphism of $\mathbb{F}$. A *sesquilinear* form on V with respect to σ is a map $\beta : V \times V \to \mathbb{F}$ such that $\forall x, y, z \in V, \forall \lambda \in \mathbb{F}$:

- $\beta(x + \lambda y, z) = \beta(x, z) + \sigma(\lambda)\beta(y, z)$ (left semi-linearity);
- $\beta(x, y + \lambda z) = \beta(x, y) + \lambda\beta(x, z)$ (right linearity).

When $\sigma = \mathrm{Id}$, a sesquilinear form is simply called a bilinear form.

Remark 2.1.2. From the Latin, "one and a half": because β is linear (on the right) plus half-linear (semi-linear, on the left).

Example 2.1.3. An inner product on V is a bilinear form (satisfying some extra properties).

Suppose that V has finite dimension n and let $\mathcal{B} = (e_1, \ldots, e_n)$ be a basis. Let M be the matrix $(\beta(e_i, e_j))$. For two vectors x, y with coordinates $X = \mathrm{Col}_{\mathcal{B}}(x)$ (a column vector), $Y = \mathrm{Col}_{\mathcal{B}}(y)$ respectively, one has:

$$\beta(x, y) = X^{t\sigma} M Y$$

where the line X^t is the transpose of the column X and $X^{t\sigma}$ is obtained from X^t by applying σ to all its coefficients.

Exercise 2.1.4. *Suppose that $\mathcal{B}' = (e'_1, \ldots, e'_n)$ is another basis and $M' = (\beta(e'_i, e'_j))$. Let P denote the basis change matrix. Show that $M' = P^{t\sigma} M P$.*

Definition 2.1.5. A sesquilinear form β is *non-degenerate* if: the only vector $v \in V$ such that $\beta(u, v) = 0$ for all $u \in V$ is the vector $v = 0$.

This is equivalent to: the only vector $u \in V$ with $\beta(u, v) = 0$ for all $v \in V$ is the vector $u = 0$.

Remark 2.1.6. Equivalence is not entirely trivial as we have made no assumption on symmetry/skew-symmetry/etc. But it can easily be seen in matrix form: β is non-degenerate iff $(\beta(e_i, e_j))$ has determinant non-zero, in any basis.

Our definitions so far were introduced precisely to encompass the following three fundamental cases.

Definition 2.1.7. The sesquilinear form β is called:

1. *symmetric* if β is bilinear and $\beta(x, y) = \beta(y, x)$ for all $x, y \in V$;
2. *Hermite-symmetric* if $\sigma^2 = \mathrm{Id} \neq \sigma$ and $\beta(x, y) = \sigma(\beta(y, x))$ for all $x, y \in V$.
3. *alternating* if β is bilinear and $\beta(x, x) = 0$ for all $x \in V$;

Remark 2.1.8. Observe that if β is alternating then $\beta(x, y) = -\beta(y, x)$ (skew-symmetry). In characteristic $\neq 2$ the converse is easily proved. However in characteristic 2 skew-symmetry and symmetry coincide.

The three special cases (symmetric, Hermite-symmetric, alternating) enjoy the following common property.

Definition 2.1.9. The sesquilinear form β is *reflexive* if $\beta(u,v) = 0$ implies $\beta(v,u) = 0$.

Definition 2.1.10. Let β be a reflexive form. Two vectors u, v are *orthogonal* if $\beta(u,v) = 0$; one writes $u \perp v$.

(We avoid to speak about non-orthogonality for non-reflexive forms; this would be too counter-intuitive).

Theorem 2.1.11. *If β is a non-degenerate reflexive sequilinear form, then up to a scalar multiple β is of one of the above types (symmetric, alternating, Hermite-symmetric).*

Proof. For every $x \in V$, the function $\varphi_x : y \mapsto \beta(x,y)$ is a linear form. The function $y \mapsto \beta(y,x)$ is not necessarily, because of sesquilinearity. But the other function $\psi_x : y \mapsto \sigma^{-1}(\beta(y,x))$ is. Observe that $\ker \varphi_x = \ker \psi_x$ by reflexivity. As a consequence, there is a scalar $\lambda_x \in \mathbb{F}$ such that $\psi_x = \lambda_x \varphi_x$.

It is an easy exercise that λ_x actually does not depend on x. Hence, there is $\lambda \in \mathbb{F}$ such that for all $x \in V$, $\psi_x = \lambda \varphi_x$. Write $\mu = \sigma(\lambda)$. Then for all $(x,y) \in V^2$:

$$\beta(y,x) = \sigma(\psi_x(y)) = \sigma(\lambda \varphi_x(y)) = \mu\sigma(\beta(x,y)) \qquad (*)$$

By non-degeneracy, there is a pair (x,y) with $\beta(x,y) \neq 0$. So up to multiplying we may suppose $\beta(x,y) = 1$ and write the following computation where $a \in \mathbb{F}$:

$$a = \beta(x,ay) = \mu\sigma(\beta(ay,x)) = \mu\sigma(\mu\sigma(\beta(x,ay))) = \mu\sigma(\mu)\sigma^2(a)$$

With $a = 1$ one finds $\mu\sigma(\mu) = 1$. So there remains $\sigma^2(a) = a$ for any $a \in \mathbb{F}$, that is $\sigma^2 = \mathrm{Id}$.

If $\sigma \neq \mathrm{Id}$ then it is a fact from field theory that μ can be written $\frac{\nu}{\sigma(\nu)}$ for some $\nu \in \mathbb{F}^*$. Hence $(*)$ rewrites $\frac{1}{\nu}\beta(y,x) = \sigma\left(\frac{1}{\nu}\beta(x,y)\right)$ for all $x, y \in V$: up to multiplication, β is hermitian.

We now suppose $\sigma = \mathrm{Id}$. Hence $\beta(y,x) = \mu\beta(x,y)$ where $\mu^2 = 1$. In characteristic $\neq 2$ this solves into $\mu = \pm 1$ and we find a symmetric or alternating form, respectively. In characteristic 2 this solves into $\mu = 1$, and $\beta(y,x) = \beta(x,y)$, so β is symmetric (but could fail to be alternating). □

The three geometries one can study are defined as follows. Our focus will be on the symplectic case.

Definition 2.1.12. A *symplectic form* is an alternating bilinear form. A *symplectic space* is a finite-dimensional space equipped with a symplectic form.

Sometimes one implicitely requires non-degeneracy. We shall not.

Definition 2.1.13. A *unitary form* is a Hermite-symmetric sesquilinear form. Similarly for a *unitary space*.

In characteristic $\neq 2$, an orthogonal form will simply be a symmetric bilinear form. The same definition in characteristic 2 leads to nothing classifiable. One must speak about quadratic forms instead. The topic is remarkably more subtle.

End of Lecture 7.

Lecture 8 (Geometric aspects: Orthogonality and the radical)

2.1.2 Orthogonality and the radical

We shall start with a few common features of the various geometries one may wish to describe: symplectic, unitary, and orthogonal in characteristic $\neq 2$. So we fix a reflexive sesquilinear form β; *we do not assume non-degeneracy.*

Definition 2.1.14. Let $X \subseteq V$ be any subset. The *orthogonal* of X is $X^\perp = \{u \in V : \forall x \in X, \beta(u,x) = 0\}$.

Observe that $X^\perp$ is a subspace of V and $X^\perp = \langle X \rangle^\perp$.

Definition 2.1.15. A subspace $W \leqslant V$ is *(totally) isotropic* if $W \leqslant W^\perp$. A vector $v \in V$ is isotropic if its span $\langle v \rangle$ is.

Example 2.1.16.

- For a symplectic form, every vector (and therefore, every line) is isotropic. This is confusing at first but one gets used to it.
- If $\beta((x_1,y_1,z_1,t_1),(x_2,y_2,z_2,t_2)) = x_1x_2 + y_1y_2 + z_1z_2 - t_1t_2$ is the symmetric form of special relativity, then any vector with $t = \sqrt{x^2+y^2+z^2}$ (photon) is isotropic.

Lemma 2.1.17. *If $U \leqslant V$ is a subspace then $U \leqslant U^{\perp\perp}$. If β is non-degenerate then equality holds.*

Proof. The inclusion is clear. For the converse, we suppose β non-degenerate and show $\operatorname{codim} U^\perp = \dim U$.

We shall do it only when $\sigma = \mathrm{Id}$, that is, when β is bilinear. Generalising to the Hermitian case is not hard, but not fascinating either (just use semi-linear functions instead of linear ones).

For $x \in V$ let $\varphi_x : U \to \mathbb{F}$ map $y \in U$ to $\beta(x,y)$. Then φ_x is linear, so $\varphi_x \in U^*$. This defines a map $\Phi : V \to U^*$ which is linear. But $\ker \Phi = U^\perp$, and as a consequence $V/U^\perp \hookrightarrow U^*$.

Now consider for $y \in U$ the function $\psi_y : V \to \mathbb{F}$ which maps $x \in V$ to $\beta(x,y)$. This function is linear. Since $y \in U$, one has $U^\perp \leqslant \ker \psi_y$. So ψ_y goes to the quotient and induces $\bar{\psi}_y : V/U^\perp \to \mathbb{F}$, that is $\overline{\psi_y} \in (V/U^\perp)^*$. Here again the map $\overline{\Psi} : U \to (V/U^\perp)^*$ which sends y to $\overline{\psi_y}$ is easily proved to be linear. But $\ker \overline{\Psi} = 0$, so $U \hookrightarrow (V/U^\perp)^*$.

Let us sum this up and consider dimensions. On the one hand, $\operatorname{codim} U^\perp \leqslant \dim U^* = \dim U$. On the other hand, $\dim U \leqslant \dim (V/U^\perp)^* = \dim V/U^\perp = \operatorname{codim} U^\perp$. Equality follows.

Hence $\operatorname{codim} U^\perp = \dim U$, and $\dim U^{\perp\perp} = \operatorname{codim} U^\perp = \dim U$. □

Too abstract? Here is an alternative proof.

Alternative proof of Lemma 2.1.17. By induction on $\dim U$. First note that this is trivial by non-degeneracy if $\dim U = 0$.

Suppose $\dim U = 1$. Then $U = \langle x \rangle$ for some vector $x \neq 0$. But then, $\varphi_x : V \to \mathbb{F}$ which maps y to $\beta(x,y)$ is a (non-trival, by non-degeneracy) linear form on V, so $\operatorname{codim} U^\perp = \operatorname{codim} \ker \varphi_x = 1$: the result is proved for $\dim U = 1$.

We now proceed by induction. Suppose the result is known for subspaces of dimension $< \dim U = m$. By non-degeneracy there is $x \in V \backslash U^\perp$. Then $x^\perp$ is a hyperplane. It does not contain U as otherwise $U \leqslant x^\perp$ and $x \in \langle x \rangle \leqslant x^{\perp\perp} \leqslant$

$U^\perp$, against its definition. Hence $W = U \cap x^\perp$ is a *proper* subspace of U; it actually is a hyperplane of U.

By induction, $\operatorname{codim} W^\perp = \dim W$. Now let $u \in U \backslash W$. Then $U = W + \langle u \rangle$, and $U^\perp = W^\perp \cap u^\perp$. By the case $m = 1$, we know that $u^\perp$ is a hyperplane of V. If $W^\perp \leqslant u^\perp$, then $u \in u^{\perp\perp} \leqslant W^{\perp\perp} = W$ as we know from induction. So $W^\perp \cap u^\perp$ is a hyperplane of $W^\perp$.

As a consequence, $\dim U^\perp = \dim W^\perp - 1 = n - \operatorname{codim} W^\perp - 1 = n - (\dim W + 1) = n - \dim U = \operatorname{codim} U$. □

Notation 2.1.18. We write $V = U_1 \oplus U_2$ if $V = U_1 \oplus U_2$ and in addition, $U_1 \perp U_2$, i.e. $U_1 \leqslant U_2^\perp$ (equivalently, $U_2 \leqslant U_1^\perp$).

Remark 2.1.19. Notice that we did *not* prove that $V = U \oplus U^\perp$; in the symplectic case, every vector is orthogonal to itself.

Definition 2.1.20. The *radical* of V is the subspace $\operatorname{Rad}(V) = V^\perp$.

By definition, β is non-degenerate iff $\operatorname{Rad} V = 0$.

Lemma 2.1.21. *Let β be a reflexive σ-sesquilinear form on V. Then β induces on $\overline{V} = V/\operatorname{Rad} V$ a reflexive σ-sesquilinear form, and $\operatorname{Rad} \overline{V} = 0$.*

Proof. Let $\overline{\beta}(\bar{x}, \bar{y}) = \beta(x, y)$ for $(x, y) \in V^2$. This is well-defined by definition of the radical. Suppose that $\overline{v} \in \operatorname{Rad} \overline{V}$ and take a representative v of $\overline{v}$. Then for all $x \in V$, $\beta(x, v) = 0$, so $v \in \operatorname{Rad}(V)$, meaning that $\overline{v} = 0$. □

We thus know how to construct non-degenerate spaces (at least if we avoid the orthogonal case in even characteristic). Can we avoid the quotient and find a direct sum?

Lemma 2.1.22. *Every space (V, β) (β a reflexive form) is a sum $\operatorname{Rad}(V) \oplus W$ where W is non-degenerate.*

Proof. Let W be *any* linear complement of $\operatorname{Rad}(V)$. Clearly $\operatorname{Rad}(V) \perp W$. Now β induces on W a form of the same type as on V. If $x \in W$ is such that $\beta(x, y) = 0$ for all $y \in W$, then $\beta(x, y + r) = 0$ for all $(y, r) \in W \times \operatorname{Rad}(V)$. Hence $\beta(x, z) = 0$ for all $z \in V$, whence $x \in \operatorname{Rad}(V)$. But $x \in W$, so $x = 0$: W is non-degenerate. □

Remark 2.1.23. Unfortunately this is not true of orthogonal spaces in characteristic 2, as restricting quadratic forms can get tricky. We shall avoid the topic anyway.

Lemma 2.1.24. *Assume that β is non-degenerate and $U \leqslant V$. Then:*

1. $\operatorname{Rad}(U^\perp) = \operatorname{Rad} U$.
2. *U is non-degenerate iff $V = U \oplus U^\perp$.*
3. *If U is totally isotropic (see Definition 2.1.15) then $\dim U \leqslant \frac{n}{2}$.*

Proof.

1. By Lemma 2.1.17, $U^{\perp\perp} = U$. Hence $\operatorname{Rad}(U^\perp) = U^\perp \cap U^{\perp\perp} = U^\perp \cap U = \operatorname{Rad}(U)$.

2. Suppose that U is non-degenerate. Then $U \cap U^\perp = 0$. But since $\dim U + \dim U^\perp = \dim V$, one has $V = U \oplus U^\perp$. Conversely if $V = U \oplus U^\perp$ then $U \cap U^\perp = 0$: U is non-degenerate.

3. If U is totally isotropic then $U \leqslant U^\perp$, so $\dim U \leqslant \dim U^\perp = \operatorname{codim} U = n - \dim U$ and $2 \dim U \leqslant n$. □

Definition 2.1.25. A *linear isometry* of spaces between (V_1, β_1) and (V_2, β_2) is a bijective linear transformation $f : V_1 \to V_2$ such that for all $x, y \in V_1$, one has $\beta_2(f(x), f(y)) = \beta_1(x, y)$.

Theorem 2.1.26 (Witt)**.** *Let (V, β) be an orthogonal, symplectic or unitary space. Let V_1, V_2 be subspaces of V. Suppose that $f : V_1 \to V_2$ is an isometry. Then there is an isometry $F : V \to V$ extending f.*

We shall not prove this theorem in general, but deal with it in special cases. The general proof is not difficult at all, but a little tedious, especially if one wishes to encompass the orthogonal geometry in characteristic 2 as well.

Remark 2.1.27.

1. A corollary (not used) is that any two maximal totally isotropic subspaces of V have the same dimension. This dimension is called the *Witt index* of the form β.

   Now if U is a totally isotropic subspace and β is non-degenerate, then $U \leqslant U^\perp$, so by Lemma 2.1.24 the Witt index is at most $\frac{1}{2} \dim V$.

2. Another corollary (not used either) is that if V is non-degenerate then the group of isometries acts transitively on the set of totally isotropic subspaces of fixed dimension.

End of Lecture 8.

Lecture 9 (Symplectic spaces and symplectic bases)

2.2 Linear Symplectic Geometry

Let β be a symplectic form on V, that is a non-degenerate alternating bilinear form ($\beta(x, x) = 0$ for all $x \in V$). (Recall that β is then reflexive, and skew-symmetric. Be careful that subspaces of V could be degenerate; for instance, any line is degenerate as we know.)

Example 2.2.1. On $\mathbb{F}^2$ define

$$\beta\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) = x_1 y_2 - x_2 y_1$$

More generally, on $\mathbb{F}^{2n}$ (and using line notation)

$$\beta((x_1, \ldots, x_{2n}), (y_1, \ldots, y_{2n})) = x_1 y_{n+1} + \cdots + x_n y_{2n} - x_{n+1} y_1 - \cdots - x_{2n} y_n$$

As we shall see, this example is typical: there is a unique symplectic space of dimension $2n$ (Proposition 2.2.6 below).

Definition 2.2.2. Let $\mathrm{Sp}(V) = \{f \in \mathrm{GL}(V) : \forall (u,v) \in V^2, \beta(f(u), f(v)) = \beta(u,v)\}$ be the *symplectic group*.

Let $\mathcal{B} = (e_1, \ldots, e_n)$ be a basis for V and J be the matrix of β in $\mathcal{B}$, that is $J = (\beta(e_i, e_j))_{i,j}$. Let $f \in \mathrm{GL}(V)$ and $A = \mathrm{Mat}_{\mathcal{B}}(f)$ be the matrix of f with respect to $\mathcal{B}$. Then $f \in \mathrm{Sp}(V)$ iff $J = A^t J A$.

Remark 2.2.3. Although it is obvious that every symplectic transformation has determinant ± 1, it so happens that $+1$ is the only value, as we shall prove later (Corollary 2.3.4). Also, it is obvious that $\lambda \mathrm{Id}$ is symplectic iff $\lambda = \pm 1$. But it is not obvious that $Z(\mathrm{Sp}(V))$ consists only of transformations of the form $\lambda \mathrm{Id}$ (Corollary 2.2.11).

2.2.1 Symplectic spaces

The "atomic" blocks in vector spaces are lines. What are the atomic blocks in symplectic spaces?

Definition 2.2.4. A pair of vectors (u, v) is called a *hyperbolic pair* if both u and v are isotropic, and $\beta(u, v) = 1$. In that case $\langle u, v \rangle$ is called a *hyperbolic plane*.

Remark 2.2.5.

- The definition is actually general, for any sesquilinear form. In the symplectic case, all vectors are isotropic, so it suffices to check that $\beta(u, v) = 1$.
- A hyperbolic plane is non-degenerate.

Proposition 2.2.6. *Let (V, β) be a symplectic space. Then V has a basis $(e_1, f_1, \ldots, e_m, f_m)$ such that the $P_i = \langle e_i, f_i \rangle$ are hyperbolic planes pairwise orthogonal.*

In particular $\dim V$ is even, and for every even integer $n = 2m$, there is up to isometry a unique (non-degenerate) symplectic space of dimension n.

Proof. Recall that since β is symplectic, $\beta(v, v) = 0$ for all $v \in V$. Let $e_1 \in V \backslash \{0\}$ and choose $f_1 \in V$ so that $\beta(e_1, f_1) \neq 0$; this is possible as otherwise β is degenerate. Replacing f_1 by $\beta(e_1, f_1)^{-1} f_1$, we may assume $\beta(e_1, f_1) = 1$: hence (e_1, f_1) is a hyperbolic pair (see Definition 2.2.4).

By non-degeneracy of V and of $\langle e_1, f_1 \rangle$ (and by Lemma 2.1.24), one has $V = \langle e_1, f_1 \rangle \oplus \langle e_1, f_1 \rangle^{\perp}$. Continuing in this way, choose a hyperbolic pair (e_2, f_2) in $\langle e_1, f_1 \rangle^{\perp}$ and so on. Hence:

$$V = \langle e_1, f_1 \rangle \oplus \langle e_2, f_2 \rangle \oplus \cdots \oplus \langle e_m, f_m \rangle$$

Uniqueness up to linear isometry is now clear. □

Definition 2.2.7. Any such basis $(e_1, f_1, e_2, f_2, \ldots, e_m, f_m)$ (pay attention to the ordering) is called a *symplectic basis* of V.

Consider the other ordering of the same basis: $e_1, \ldots, e_n, f_1, \ldots, f_n$. Then in this basis β has matrix:

$$J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$$

Observe that $M = \langle e_1, \ldots, e_n \rangle$ is totally isotropic and as $\dim M = n$, M is maximal as such.

Proposition 2.2.8. $|\mathrm{Sp}(2m,q)| = q^{m^2}\prod_{i=1}^{m}(q^{2i}-1)$.

Proof. By definition, $\mathrm{Sp}(V)$ acts faithfully and transitively on the set of (ordered) symplectic bases of V. So $|\mathrm{Sp}(V)|$ equals the number of such bases, which we now count.

Let $e_1 \in V\backslash\{0\}$; there are $q^{2m}-1$ possibilities. Now e_1 being fixed, the equation $\beta(e_1,f_1)=1$ has solution set an *affine hyperplane* (that is, a translate of a vector hyperplane), so there are exactly q^{2m-1} possible values for f_1.

It follows that there are $(q^{2m}-1)q^{2m-1}$ distinct hyperbolic pairs. Write $V = \langle e_1,f_1\rangle \oplus W$ with $W = \langle e_1,f_1\rangle^{\perp}$. Now W is a symplectic space of dimension $2n-2$. Hence $|\mathrm{Sp}(V)| = (q^{2m}-1)q^{2m-1}|\mathrm{Sp}(W)|$. By induction,

$$|\mathrm{Sp}(V)| = \prod_{i=1}^{m}(q^{2i}-1)q^{2i-1} = q^{m^2}\prod_{i=1}^{m}(q^{2i}-1) \qquad \square$$

Observe in particular that for $n=1$, $|\mathrm{Sp}(2,\mathbb{F})| = |\mathrm{SL}(2,\mathbb{F})|$. This actually is an isomorphism.

Theorem 2.2.9. $\mathrm{Sp}(2,\mathbb{F}) \simeq \mathrm{SL}(2,\mathbb{F})$.

Proof. Equip $V=\mathbb{F}^2$ with the form:

$$\begin{array}{rccc} \det: & V\times V & \to & \mathbb{F} \\ & (x,y) & \mapsto & \det(x,y) \end{array}$$

As we know, det is bilinear; the group of det is by definition $\mathrm{SL}(2,\mathbb{F})$. But det is alternating, so it is a symplectic form. It follows that $\mathrm{Sp}(2,\mathbb{F}) = \mathrm{SL}(2,\mathbb{F})$. $\square$

End of Lecture 9.

Lecture 10 (Witt's theorem; symplectic transvections)

2.2.2 Witt's Theorem for symplectic spaces

Theorem 2.2.10 (the symplectic version of Witt's Theorem)**.** *Let (V,β) be a symplectic space. Suppose that $V_1, V_2 \leqslant V$ are subspaces and that there exists a* partial *linear isometry $f: V_1 \to V_2$. Then there is a* global *linear isometry $g: V \to V$ (that is, $g \in \mathrm{Sp}(V)$) extending f.*

Proof. By descending induction on $\dim V_1 = \dim V_2$. The claim trivially holds when $V_1 = V$. Inside V let V_1 be a maximal counterexample: the property will hold of subspaces of larger dimension. There are two cases.

- If $\mathrm{Rad}(V_1) = 0$ then $\mathrm{Rad}(V_2) = 0$ as well since they are isometric. As $\mathrm{Rad}(V_i^{\perp}) = \mathrm{Rad}(V_i)$ by Lemma 2.1.24, the spaces $V_i^{\perp}$ are non-degenerate as well. Since $\dim V_1 = \dim V_2$, one has $\dim V_1^{\perp} = \dim V_2^{\perp}$. It follows from Proposition 2.2.6 and the uniqueness of the symplectic space of dimension $2k$ that $V_1^{\perp}$ and $V_2^{\perp}$ are isometric, by some f'.

  But since V_i is non-degenerate, by Lemma 2.1.24 we know that $V = V_i \oplus V_i^{\perp}$. So the map $g = f \oplus f'$ (understand that it acts as f on V_1 and as f' on $V_1^{\perp}$) is an isometry of V extending f.

  This case was rather straightforward: we did not even use induction.

- Now suppose that $R = \mathrm{Rad}(V_1) \neq 0$ (by isometry, $\mathrm{Rad}(V_2) \neq 0$ as well). The idea is to extend f one dimension up to some $g : V_1 \oplus L_1 \to V_2 \oplus L_2$ for cleverly chosen lines $L_1 = \langle t_1 \rangle, L_2 = \langle t_2 \rangle$. We want to do it isometrically, so we must carefully control the values of $\beta(t_i, V_i)$.

  Write $V_1 = R \oplus W$ as in Lemma 2.1.22. Let $\langle r_1, \dots, r_d \rangle$ be a basis of R. Let $S = \langle r_2, \dots, r_d \rangle$, so that $R = \langle r_1 \rangle \oplus S$.

  Since $\langle r_1 \rangle^{\perp\perp} = \langle r_1 \rangle \not\subseteq W = W^{\perp\perp}$, one has $W^\perp \not\subseteq r_1^\perp$. This means that $\beta(r_1, W^\perp) \neq 0$. So we find $t \in W^\perp$ with $\beta(r_1, t) = 1$. If $\beta(r_2, t) \neq 0$ then replacing r_2 with $r_2 - \lambda r_1 \in R$ we may actually assume $\beta(r_2, t) = 0$, and so on. Hence there is $t \in W^\perp \cap S^\perp$ such that $\beta(r_1, t) = 1$. Observe that $t \notin V_1$ as otherwise $t \in R^\perp$, which is not the case.

  So starting with a basis of $\mathrm{Rad}(V_1)$ extended to a basis $\mathcal{B}_1$ of V_1, we found a vector $t_1 = t \notin V_1$ orthogonal to every vector in $\mathcal{B}_1$ except to the first.

  Since V_2 is isometric to V_1, we take $\mathcal{B}_2 = f(\mathcal{B}_1)$ and the same argument yields a vector $t_2 \notin V_2$ orthogonal to every vector in $\mathcal{B}_2$ except to the first.

  We now map t_1 to t_2. This defines an isometry $g : V_1 \oplus \langle t_1 \rangle \to V_2 \oplus \langle t_2 \rangle$. By reverse induction, g (and hence f) extends to an isometry of V.

In either case we are done. □

Here is a Corollary to Witt's theorem.

Corollary 2.2.11. $Z(\mathrm{Sp}(V)) = \{\pm \mathrm{Id}\}$.

Proof. Let $f \in Z(\mathrm{Sp}(V))$. We shall simply adapt the second proof of Proposition 1.1.5 in a symplectic way. It actually suffices to show: $\forall v \in V, f(v) \in \langle v \rangle$. This will prove that $f = \lambda \mathrm{Id}$. But such a map is symplectic iff $\lambda = \pm 1$ and we will be done.

So suppose that there is $v \in V$ such that $(v, f(v))$ is linearly independent. Notice that $(v, v + f(v))$ is another linearly independent family; moreover, $\beta(v, v + f(v)) = \beta(v, f(v))$. So there is a partial linear isometry mapping $(v, f(v))$ to $(v, v + f(v))$. By Witt's Theorem, it extends to a symplectic function $g \in \mathrm{Sp}(V)$ with $g(v) = v$ and $g(f(v)) = v + f(v)$. Then $g \circ f(v) = v + f(v) \neq f(v) = f \circ g(v)$, a contradiction. □

2.3 Group-Theoretic Analysis

2.3.1 Symplectic Transvections

Transvections were an essential tool when we studied $\mathrm{SL}(V)$. We try to do the same.

Proposition 2.3.1. *A transvection $t_{\varphi,a}$ is in $\mathrm{Sp}(V)$ iff $\ker \varphi = a^\perp$. In that case there is $\lambda \in \mathbb{F}$ with $t_{\varphi,a}(x) = x + \lambda \beta(x, a)a$.*

Proof. A priori, the condition on $\tau = t_{\varphi,a}$ to be symplectic is that, for all $x, y \in V$:

$$\beta(x, y) = \beta(\tau(x), \tau(y)) = \beta(x, y) + \varphi(y)\beta(x, a) + \varphi(x)\beta(a, y) + \beta(a, a)$$

So this is equivalent to:

$$\varphi(y)\beta(x,a) + \varphi(x)\beta(a,y) = 0$$

Suppose this holds and let $b \in V$ be such that $\beta(a,b) \neq 0$. Substituting $y = b$ one has $\varphi(b)\beta(x,a) + \varphi(x)\beta(a,b) = 0$, whence $\varphi(x) = -\frac{\varphi(b)}{\beta(a,b)}\beta(x,a) = \lambda\beta(x,a)$ where λ is independent on x.

It is clear in this case that $\ker\varphi = a^\perp$. Conversely if $\ker\varphi = a^\perp = \ker\beta(\cdot,a)$, then there is λ such that $\varphi(x) = \lambda\beta(x,a)$, and τ is symplectic. □

Notation 2.3.2. For $a \in V$ and $\lambda \in \mathbb{F}$ let $\tau_{\lambda,a}(v) = v + \lambda\beta(v,a)a$.

End of Lecture 10.

Lecture 11 (Generation and consequences)

Theorem 2.3.3. *The symplectic transvections generate* $\mathrm{Sp}(V)$.

Proof. Since the result is known for $\mathrm{SL}(2,\mathbb{F}) \simeq \mathrm{Sp}(2,\mathbb{F})$, we may suppose that the dimension is $\geqslant 4$. Let T be the subgroup generated by the symplectic transvections.

- We first show that T is transitive on $V\backslash\{0\}$. Let $v_1, v_2 \in V\backslash\{0\}$.
  If $\beta(v_1,v_2) \neq 0$, let $\lambda = \frac{1}{\beta(v_1,v_2)}$ and $a = v_1 - v_2$. Then:

$$\begin{aligned} t_{\lambda,a}(v_1) &= v_1 + \lambda\beta(v_1,a)a \\ &= v_1 + \frac{\beta(v_1, v_1 - v_2)}{\beta(v_1,v_2)}(v_1 - v_2) \\ &= v_1 - (v_1 - v_2) \\ &= v_2 \end{aligned}$$

  So there is a transvection sending v_1 to v_2.

  Now suppose that $v_1 \neq v_2$ and $\beta(v_1,v_2) = 0$. Since V is not the union of two proper subspaces, there is $w \notin v_1^\perp \cup v_2^\perp$. By the previous case there are symplectic transvections t_1 and t_2 with $t_1(v_1) = w$ and $t_2(w) = v_2$.

  This proves transitivity on non-zero vectors.

- We now claim that T is transitive on the set of hyperbolic pairs.

  Let (u_1,v_1) and (u_2,v_2) be two hyperbolic pairs. Since T is transitive on the non-zero vectors, we may assume that $u_1 = u_2$, and write it simply u. By assumption, $\beta(u,v_1) = \beta(u,v_2) = 1$, so $u \perp v_1 - v_2$.

  If $\beta(v_1,v_2) \neq 0$, then consider the symplectic transvection:

$$t(v) = v + \frac{\beta(v, v_1 - v_2)}{\beta(v_1,v_2)}(v_1 - v_2)$$

  Observe that $t(u) = u$ and $t(v_1) = v_2$. In this case, we are done.

  Now suppose that $\beta(v_1,v_2) = 0$. Then $(u, u+v_1)$ is a hyperbolic pair satisfying $\beta(v_1, u+v_1) \neq 0$ and $\beta(u+v_1,v_2) \neq 0$. Thus there exist symplectic transvections:

$$\begin{aligned} t_1 &: (u,v_1) \mapsto (u, u+v_1) \\ t_2 &: (u,u+v_1) \mapsto (u,v_2) \end{aligned}$$

  Hence T is transitive on the set of hyperbolic pairs.

- Now suppose that $f \in \mathrm{Sp}(V)$ and (u, v) is a hyperbolic pair. Then $(f(u), f(v))$ is another hyperbolic pair. So for some $t \in T$, we have $t(f(u)) = u$ and $t(f(v)) = v$. This means that $t \circ f$ acts as the identity on $L = \langle u, v \rangle$; as it is an isometry it fixes $L^\perp$ (setwise).

  By induction, the restriction of $t \circ f$ to $L^\perp$ is a product of symplectic transvections $t_1, \ldots, t_k$ of $L^\perp$. But L is non-degenerate, so $V = L \oplus L^\perp$, and each t_i extends to a global transvection $t_i' = \mathrm{Id}_L + t_i$ of V. It then follows that $f = t^{-1} t_1' \ldots t_k' \in T$. □

Corollary 2.3.4. $\mathrm{Sp}(V) \leqslant \mathrm{SL}(V)$.

Proof. Each transvection has determinant 1. □

Corollary 2.3.5. $\mathrm{Sp}(2n, \mathbb{F})' = \mathrm{Sp}(2n, \mathbb{F})$, *except for* $\mathrm{Sp}(2, 2)$, $\mathrm{Sp}(2, 3)$, *and* $\mathrm{Sp}(4, 2)$.

Proof. By induction on the dimension. We first deal with the inductive step. Suppose that $\mathrm{Sp}(2n, \mathbb{F})' = \mathrm{Sp}(2n, \mathbb{F})$, and prove it for $\mathrm{Sp}(2n+2, \mathbb{F})$. By Theorem 2.3.3 it suffices to show that every symplectic transvection is in the commutator subgroup. So let $t = t_{\lambda, a} \in \mathrm{Sp}(2n + 2, \mathbb{F})$; we shall prove $t \in (\mathrm{Sp}(2n + 2, \mathbb{F}))'$. Let P be a hyperbolic plane in $a^\perp$; bear in mind $V = P \oplus P^\perp$.

The restriction $s = t_{|P^\perp}$ of t to $P^\perp$ belongs to $\mathrm{Sp}(P^\perp)$, and by induction, $\mathrm{Sp}(P^\perp)' = \mathrm{Sp}(P^\perp)$. Hence $s \in \mathrm{Sp}(P^\perp)'$. Since t acts as the identity on $P \leqslant a^\perp = \ker \beta(a, \cdot)$, it follows that $t = \mathrm{Id}_P + s$; therefore $t \in \mathrm{Sp}(2n + 2, \mathbb{F})'$ (easily seen blockwise).

Of course the previous argument works only if we could get induction started. But we know from Theorem 2.2.9 that $\mathrm{Sp}(2, \mathbb{F}) \simeq \mathrm{SL}(2, \mathbb{F})$ for any field, and from Corollary 1.2.37 from last week that $\mathrm{SL}(2, \mathbb{F})' = \mathrm{SL}(2, \mathbb{F})$ as soon as $|\mathbb{F}| \geqslant 4$. So for $\mathbb{F} \neq \mathbb{F}_2, \mathbb{F}_3$ there is no problem: the induction starts already at $2n = 2$.

When $\mathbb{F} = \mathbb{F}_2$ we are supposed to start induction at $2n = 6$; when $\mathbb{F} = \mathbb{F}_3$ we are supposed to start induction at $2n = 4$. So to complete the proof, we need to check that $\mathrm{Sp}(4, 3)' = \mathrm{Sp}(4, 3)$ and $\mathrm{Sp}(6, 2)' = \mathrm{Sp}(6, 2)$. This is a little tedious but not difficult. □

End of Lecture 11.

Lecture 12 (Simplicity)

2.3.2 Transitivity

Definition 2.3.6. Let $\mathrm{PSp}(V) = \mathrm{Sp}(V)/Z(\mathrm{Sp}(V)) = \mathrm{Sp}(V)/\pm \mathrm{Id}$ be the *projective symplectic group*.

Theorem 2.3.7. $\mathrm{PSp}(V)$ *is transitive on* $\mathbb{P}(V)$. *If* $\dim V = 2$, *then* $\mathrm{PSp}(V)$ *is* 2*-transitive. If* $\dim V \geqslant 4$ *and* $P \in \mathbb{P}(V)$, *then* $\mathrm{PSp}(V)_P$ *has exactly three orbits on* $\mathbb{P}(V)$.

Proof. For clarity, write $G = \mathrm{PSp}(V)$. Two lines are isotropic, so they certainly are isometric. Hence by Witt's Theorem (Theorem 2.2.10), G is transitive on the points of $\mathbb{P}(V)$. Let $P \in \mathbb{P}(V)$. Of course $\{P\}$ is an orbit under the stabiliser G_P.

Now if $Q_1, Q_2 \in P^\perp$ are *both* distinct from P, then by Witt's Theorem again, Q_1 can be sent to Q_2 fixing P: so they are in the same orbit. Note that this cannot happen in dimension 2 (as $P^\perp = \{P\}$ in $\mathbb{P}(V)$).

Finally let $Q \notin P^\perp$. Then $\langle P, Q\rangle$ is a hyperbolic plane. But always by Witt's Theorem, G acts transitively on the set of hyperbolic planes. Hence G_P acts transitively on $\mathbb{P}(V)\backslash P^\perp$. We have three orbits under G_P:

$$\{P\}, P^\perp\backslash\{P\}, \mathbb{P}(V)\backslash P^\perp$$

The one in the middle is actually empty when $n = 2$. □

As a consequence $\mathrm{PSp}(V)$ is *not* 2-transitive in general. It is however primitive.

Theorem 2.3.8. *The action of* $\mathrm{PSp}(V)$ *on the points of* $\mathbb{P}(V)$ *is primitive.*

Proof. If $\dim V = 2$, then $\mathrm{PSp}(V)$ is 2-transitive (alternatively, $\mathrm{PSp}(V) = \mathrm{PSL}(V)$), so we are done. Therefore assume that $\dim V \geqslant 4$. Let Δ be a block of imprimitivity with $|\Delta| > 1$. Let $L \neq M \in \Delta$.

- First suppose $M \in L^\perp$ (that is, $L \perp M$). Then for any $M' \in L^\perp$, by Witt's Theorem there is $g \in G$ fixing L and mapping M to M'. Now $L \in g\cdot\Delta \cap \Delta$, so $g \cdot \Delta = \Delta$ and $M' \in \Delta$. Hence $L^\perp \subseteq \Delta$.

  Let $N \notin L^\perp$ and $P \in (L+N)^\perp$: this is possible as $\dim V \geqslant 4$. Since $P \perp L$ one has $P \in \Delta$. But applying the previous argument with P instead of L one finds $N \in P^\perp \subseteq \Delta$. So $\mathbb{P}(V)\backslash L^\perp \subseteq \Delta$ as well and $\Delta = \mathbb{P}(V)$.

- Now suppose $M \notin L^\perp$. Then $\langle L, M\rangle$ is a hyperbolic line and can be sent to any other hyperbolic line. In particular $\mathbb{P}(V)\backslash\{L^\perp\} \subseteq \Delta$.

  Let $N \in L^\perp$. Since V is not a union of two proper subspaces, $\mathbb{P}(V) \neq L^\perp \cup N^\perp$: let $P \notin L^\perp \cup N^\perp$. Hence $P \notin L^\perp$ and $P \in \Delta$. But $N \notin P^\perp$, so applying the same argument with P instead of L, one has $N \in \Delta$.

In either case $\Delta = \mathbb{P}(V)$: $\mathrm{PSp}(V)$ is primitive. □

2.3.3 Simplicity of $\mathrm{PSp}(V)$

Recall that for the special linear group we had in Notation 1.3.8 introduced for $u \in V\backslash\{0\}$ the abelian group $K_u = \{t_{\varphi,u} : \varphi \in V^* : \varphi(u) = 0\}$. Of course we must now restrict to symplectic transvections.

Notation 2.3.9. Let $K_{u,u^\perp} = \{t_{\varphi,u} : \varphi \in V^* : \ker\varphi = u^\perp\}$.

Lemma 2.3.10. $K_{u,u^\perp} = K_u \cap \mathrm{Sp}(V)$ *is an abelian, normal subgroup of the stabiliser* $\mathrm{Stab}_{\mathrm{Sp}(V)}(u)$ *whose* $\mathrm{Sp}(V)$*-conjugates generate* $\mathrm{Sp}(V)$*.*

Proof. Bear in mind Theorem 1.3.9 and the fact that $\mathrm{Sp}(V) \leqslant \mathrm{SL}(V)$.

Recall that a transvection $t_{\varphi,u}$ (where by definition of a transvection $\varphi(u) = 0$) is in $\mathrm{Sp}(V)$ iff $\ker\varphi = u^\perp$. So $K_{u,u^\perp} = K_u \cap \mathrm{Sp}(V)$; it is thus abelian. But $\mathrm{Stab}_{\mathrm{Sp}(V)}(u) = \mathrm{Stab}_{\mathrm{SL}(V)}(u) \cap \mathrm{Sp}(V)$, so $K_{u,u^\perp}$ is normal in $\mathrm{Stab}_{\mathrm{Sp}(V)}(u)$.

Now any symplectic transvection is of the form $t_{\psi,v}$ with $\ker\psi = v^\perp$. By transitivity of $\mathrm{Sp}(V)$ on the set of non-zero vectors, there is $f \in \mathrm{Sp}(V)$ with $f(u) = v$. Hence $ft_{\varphi,u}f^{-1} = t_{f\varphi,v} \in K_u^f \cap \mathrm{Sp}(V) = K_{u,u^\perp}$. □

Theorem 2.3.11. *The symplectic groups* $\mathrm{PSp}(2n, \mathbb{F})$ *are all simple, except for* $\mathrm{PSp}(2,2)$*,* $\mathrm{PSp}(2,3)$*, and* $\mathrm{PSp}(4,2)$*.*

Proof. We shall use Iwasawa's criterion, Theorem 1.3.7.

- $\mathrm{PSp}(V)$ acts primitively on the points of $\mathbb{P}(V)$ (Theorem 2.3.8).
- $\mathrm{PSp}(V) = \mathrm{PSp}(V)'$ except for $\mathrm{PSp}(2,2)$, $\mathrm{PSp}(2,3)$, and $\mathrm{PSp}(4,2)$ (Corollary 2.3.5).
- The image $L = \overline{K_{u,u^\perp}}$ of $K_{u,u^\perp} \trianglelefteqslant \mathrm{Sp}(V)_u$ is an abelian normal subgroup of $\mathrm{PSp}(V)_u$ and $\mathrm{PSp}(V) = \langle L^g : g \in \mathrm{PSp}(V) \rangle$ by Lemma 2.3.10.

The consequence is immediate. □

Remark 2.3.12.

- $\mathrm{Sp}(2,2) = \mathrm{SL}(2,2)$ and $\mathrm{Sp}(2,3) = \mathrm{SL}(2,3)$ as we know (Theorem 2.2.9).
- $\mathrm{Sp}(4,2) \simeq S_6$.

End of Lecture 12.
