Problems and Topics in Linear Algebra

Rui Xiong

Contents

Last updated: September 21, 2024

1 Simultaneously Triangularity

We will investigate when a family of matrices can be diagonalized/triangulated simultaneously.

Definition 1.1 To be exact, let $\{A_i\}_{i\in I}$ be a family of $n\times n$ complex matrices. We say that they can be **simultaneously diagonalized (SD)** if there exists a $P \in \mathrm{GL}_n(\mathbb{C})$ such that every PA_iP^{-1} is a diagonal matrix.

Similarly, we say that they can be **simultaneously triangulated (ST)** by changing the above condition by requiring each PA_iP^{-1} to be an upper triangular matrix.

The first reduction can be made on the number of matrices. Note that A_i's spans a finite-dimensional subspace of matrix algebra, which allow us to reduce the problem to finite many of matrices.

Remark 1.2 Geometrically, SD is equivalent to the existence of a decomposition

$$V_1 \oplus \cdots \oplus V_n = \mathbb{C}^n$$

with each V_j a one-dimensional subspace which is A_i-invariant for all i.

Similarly, ST is equivalent to the existence of a flag

$$0 = V_0 \subseteq V_1 \subseteq \cdots \subseteq V_{n-1} \subseteq V_n = \mathbb{C}^n$$

with each V_j an j-dimensional subspace which is A_i-invariant for all i. In particular, any non-zero element $v \in V_1$ is a **common eigenvector** of all A_i's.

From this point of view, a single matrix is definitely able to be triangulated since we can always find an eigen-subspace and proceed by induction on the quotient space.

A motivating statement is the coming theorem. Actually, all of the statements in this section can be viewed as a generalization of this argument.

Theorem 1.3 Assume A and B commutes, then they can be ST.

— Proof. Let V be the eigen-subspace of A for some eigenvalue λ. Then V is B-invariant, since $A(Bv) = B(Av) = \lambda Bv$ for $v \in V$. Therefore, we can pick a B-eigenvector v inside V which is the desired **common eigenvector**. Now by making induction on $\mathbb{C}^n/\operatorname{span}(v)$, we see the existence of the flag. □

▶ **Problem 1.4** Assume A and B are diagonalizable matrices commuting each other, show that they can be SD. ◀ **P7**

▶ **Problem 1.5** For a commutative family $\{A_i\}$, show that they can be ST.

▶ **Problem 1.6** Show that A_i's can be SD if and only if they are all diagonalizable and commute each other.

It is natural to ask for an equivalent condition for ST, which does exist. Before starting it as a theorem, we shall make attempt on the necessary conditions for two matrices A and B able to be ST. Note that if A and B can be ST, then

$$[A, B] = AB - BA$$

is nilpotent. Actually, this is not extremely far from the equivalent condition.

Theorem 1.7 ([1]) The family $\{A_i\}$ can be ST if and only if

$$[A_i, A_j]f(A)$$

is nilpotent for any non-commutative polynomial f in A_i's.

— *Proof.* The author dislikes the elementary proofs, so a proof using the representation of associative algebra will be given here.

It is clear the condition holds when it is possible to be ST. Assuming the condition holds, we will show the converse. Consider the algebra $\mathcal{A}$ generated by $\{A_i\}$. We will view $V = \mathbb{C}^n$ as representation of $\mathcal{A}$. The condition stated here tells that $[A_i, A_j]$ lies in the radical of $\mathcal{A}$. In other words, $\mathcal{A}$ is basic, i.e. all simple modules are one-dimensional. The condition for ST is equivalent to a $\mathcal{A}$-flag which is ensured by Jordan–Hölder theorem. □

Good as this theorem looks, Theorem 1.7 only serves as a theoretic statement in practice. Frankly speaking, only the generalization of Theorem 1.3 would work in face of any less abstract cases.

▶ **Problem 1.8([2])** Assume $\operatorname{rank}[A, B] = 1$, show that A and B can be ST. ◀ **P7**

The rest of the section will be devoted to explaining the solubility of Lie algebra and groups. Actually, the condition stated in 1.7 can be viewed as the associative-algebra-theoretic notation for solvable algebras which are already named by "basic algebras".

Recall a (concrete) **Lie algebra** is a subspace of matrix algebra which is closed under the Lie bracket $[\cdot, \cdot]$. For the theorems of Lie algebras and its group theoretic analogy, see [3].

Definition 1.9 Let $\mathfrak{g}$ be a Lie algebra, we say it is **solvable** if $\mathfrak{g}^{(N)} = 0$ for some N where $\mathfrak{g}^{(0)} = \mathfrak{g}$ and $\mathfrak{g}^{(i+1)} = [\mathfrak{g}^{(i)}, \mathfrak{g}^{(i)}]$ the space of elements $[x, y]$ for $x, y \in \mathfrak{g}^{(i)}$.

The first non-commutative solvable lie algebra is

$$\mathfrak{b} = \left\{ \begin{pmatrix} a & b \\ & -a \end{pmatrix} : a, b \in \mathbb{C} \right\}.$$

Denoting $H = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$ and $E = \begin{pmatrix} 0 & 1 \\ & 0 \end{pmatrix}$, we have $[H, E] = 2E$ which forms all the defining relations. Due to the defining form of $\mathfrak{b}$, it is already ST. The following problem is to ask no matter how $\mathfrak{b}$ is embedded (or more generally, represented) it can always be ST.

▶ **Problem 1.10** For two matrices H and E with relation $[H, E] = 2E$, show that they can be ST. ◀ **P7**

Lie's Theorem 1.11 A solvable Lie algebra $\mathfrak{g}$ can be ST.

— *Proof.* We can find a codimensional one ideal $\mathfrak{a}$ of $\mathfrak{g}$ by doing the same thing in the commutative Lie algebra $\mathfrak{g}/\mathfrak{g}^{(1)}$. By induction, we can find a linear functional $\lambda : \mathfrak{a} \to \mathbb{C}$ such that

$$V = \{v \in \mathbb{C}^n : {}^{\forall a \in A,} av = \lambda(a)v\} \neq 0.$$

We will show that this space is invariant under $\mathfrak{g}$. Pick any $x \in \mathfrak{g} \setminus \mathfrak{a}$. For any $a \in \mathfrak{a}$ and $v \in V$,

$$a \cdot xv = xav + [a, x]v = \lambda(a)xv + \lambda([a, x])v, \tag{$*$}$$

since $[a, x] \in \mathfrak{a}$. This does not tell the invariance immediately, but it tells a flagged version that

$$a \cdot xv \subseteq \operatorname{span}(xv, v).$$

In particular, by an easy induction, we get

$$a \cdot \operatorname{span}(x^r v, \cdots, v) \subseteq \operatorname{span}(x^r v, \cdots, v).$$

We see all $a \in \mathfrak{a}$ fix a common partial flag $\operatorname{span}(x^r v, \cdots, v)$ (deleting the repetitive spaces if necessary). Denote the biggest member of this flag by

$$V' = \operatorname{span}(v, xv, x^2 v, \ldots).$$

By the above discussion, under the matrix presenting over V', $\mathfrak{a}$ can be ST with all diagonal elements $\lambda(a)$ for any $a \in \mathfrak{a}$. Note that V' is also x-invariant, thus the trace of a bracket $[a, x]$ of two operators over V' is zero. On the other hand, $[a, x] \in \mathfrak{a}$, we thus get

$$\dim V' \cdot \lambda([a, x]) = 0.$$

Luckily, we are working over $\mathbb{C}$ where the above condition implies that $\lambda([a, x]) = 0$. Substituting back to $(*)$, we see the space V is x-invariant (thus $\mathfrak{g}$-invariant). Finally, argue just as Theorem 1.3. □

In Lie theory, there is another important theorem on ST concerning nilpotent matrices.

Engel's Theorem 1.12 For a Lie algebra $\mathfrak{g}$, if all its elements are nilpotent, then $\mathfrak{g}$ can be ST.

— *Proof.* It suffices to show the existence of a vector $v \neq 0$ such that for any $x \in \mathfrak{g}$,

$$x \cdot v = 0.$$

For any proper subalgebra $\mathfrak{h}$ of $\mathfrak{g}$, there is an induced adjoint action of $\mathfrak{h}$ on $\mathfrak{g}/\mathfrak{h}$. Note that

$$\operatorname{ad} x = \text{left multiplication by } x - \text{right multiplication by } x.$$

If x is nilpotent, the adjoint action of x can be written as a sum of two commutative nilpotent operators as above, and in particular, it is still nilpotent. So $\mathfrak{h}$ acts on $\mathfrak{g}/\mathfrak{h}$ nilpotently. By applying induction hypothesis on $\mathfrak{g}/\mathfrak{h}$, there exists an $g \in \mathfrak{g} \setminus \mathfrak{h}$ such that

$$[\mathfrak{h}, g] \subseteq \mathfrak{h}.$$

In particular, $\mathfrak{h} + \operatorname{span}(g)$ is a bigger subalgebra. If we pick $\mathfrak{h}$ to be maximal, then $\mathfrak{h}$ must be an ideal. By induction,

$$\mathfrak{h}^{\perp} = \bigcap_{x \in \mathfrak{h}} \ker x \neq 0.$$

For any $v \in \mathfrak{h}^{\perp}$, $x \in \mathfrak{h}$ and $g \in \mathfrak{g}$, we have

$$x \cdot gv = gxv + [x, g]v = 0.$$

This tells us that $\mathfrak{h}^{\perp}$ is $\mathfrak{g}$-invariant, so we can conclude the theorem by induction. □

Now we turn to the group theoretic analogy of the above two theorems. We will work with subgroups of $\mathrm{GL}_n(\mathbb{C})$.

Definition 1.13 Let G be a group, we say it is **solvable** if $G^{(N)} = 0$ for some N where $G^{(0)} = G$ and $G^{(i+1)} = [G^{(i)}, G^{(i)}]$ the subgroup generated by elements $xyx^{-1}y^{-1}$ for $x, y \in G^{(i)}$.

Theorem 1.14 A connected solvable subgroup G of $\mathrm{GL}_n(\mathbb{C})$ can be ST.

— *Proof.* By induction, we can find a character χ of $G^{(1)}$ such that

$$V = \{v \in \mathbb{C}^n : {}^{\forall x \in G^{(1)}}, xv = \chi(x)v\}$$

is nonzero. Note that $\chi \mapsto \chi(g^{-1} \cdot g)$ permutes all such characters, since

$$xv = \chi(x)v \iff xgv = g(g^{-1}xg)v = \chi(g^{-1}xg)gv.$$

But there are finite of them, so it acts trivially thanks to the connectivity. Note that for any $x \in V$ and $x \in G^{(1)}$, we have

$$x(gv) = g(g^{-1}xg)v = g\chi(g^{-1}xg)v = \chi(x)gv.$$

The rests of the proof are the same as before. □

From the geometric point of view, the proof essentially shows the existence of a fixed point G over the projective space $\mathbb{C}P^{n-1}$ — the space of all one-dimensional subspaces. Actually, there is a theorem generalize in this direction but in terms of algebraic groups, see [1].

Theorem 1.15 Assume a connected solvable algebraic group G acts on a complete variety X, then X has a fixed point.

— *Proof.* We can replace X by the fixed loci of G'. Note that for any $x \in X$, the stabilizer G_x is normal since it is contained in G'. In particular, G/G_x is a connected **affine** algebraic group which isomorphic to the orbit of x (under the reduced structure). We can pick the orbit of minimal dimension which must be closed (thus complete). But the only connected affine variety which is complete is a single point. □

There is a funny analogy over finite fields. Philosophically, we shall think p-group as a characteristic-p analogy of unipotent groups, a special case of solvable groups. For example $\begin{pmatrix} 1 & \mathbb{F}_p & \mathbb{F}_p \\ & 1 & \mathbb{F}_0 \\ & & 1 \end{pmatrix}$ is a non-commutative p-group.

▶ **Problem 1.16** Let $G \subseteq \mathrm{GL}_n(\mathbb{F}_p)$ be a p-group. Show that G can be ST over $\mathbb{F}_p$. ◀ **P8**

Kolchin's Theorem 1.17 A subgroup G of $\mathrm{GL}_n(\mathbb{C})$ can be ST if all its elements are unipotent.

— *Proof.* It suffices to show the existence of a nonzero fixed vector. We can view $V = \mathbb{C}^n$ as a representation of G. Firstly, we can assume V is simple. Then by Burnside's theorem, G spans $\mathrm{End}(V)$, the space of matrices. For any $g \in G$, $g - 1$ is nilpotent by definition. Note that for any $g' \in G$,

$$\mathrm{tr}((g-1)g') = \mathrm{tr}(gg') - \mathrm{tr}(g') = 0.$$

As a result, for any matrices x,

$$\operatorname{tr}((g-1)x) = 0.$$

By taking in elementary matrices, we see it says $g - 1 = 0$. □

References

[1] M. A. Drazin, J. W.Dungey, K. W. Gruenberg, Some theorems on commutative matrices, J. London Math. Soc. 26 (1951), 221-228.

[2] R. M. Guralnick, A note on pairs of matrices with rank one commutator, Linear and Multilinear Algebra 8 (1979/80), no. 2, 97-99.

[3] J. P. Serre, Lie algebras and Lie group, 1964 Lectures given at Harvard University. Springer series of Lecture Notes in Mathematics. 978-3-540-55008-2. (1992)

[4] J. E. Humphreys, Linear Algebraic Groups. Springer seires of Graduate Texts in Mathematics. 978-1-4684-9445-7. (1975)

Hints

1.4 By our discussion above, we need to show B can be diagonalized on a B-invariant subspace V. This is clear since B is diagonalizable if and only if its minimal polynomial has not multiple roots.

If one wants, this argument can be done by using **Krull–Schmidt theorem**.

1.8 Show that $\ker A$ or $\operatorname{im} A$ is B-invariant.

Firstly, for $v \in \ker A$, we have $A(Bv) = BAv + [A, B]v = [A, B]v$. If $\ker A$ is B-invariant, then the argument of Theorem 1.3 still works. So we assume the converse, then $\operatorname{im}[A, B]$ is spanned by ABv for some $v \in \ker A$. So $\operatorname{im} A$ is B-invariant. Actually, $BAu = ABu - [A, B]u = A(Bu - xBv)$ for some $x \in \mathbb{C}$. By replacing A by $A - \lambda$ for any eigenvalue, we can prove by induction.

1.10 Actually,

$$HEx = E(H + 2)x$$

This tells that if $Hv = \lambda v$, then $H(Ev) = (\lambda + 2)Hv$. We pick an eigenvalue λ of H such that $\lambda + 2$ is no longer one. Then $Ev = 0$, for any eigenvector v of H belonging to λ. This is the desired common eigenvector.

1.16 We have

$$\#(\mathbb{F}_p^n \setminus 0) \equiv \{\text{fixed points}\} \mod p,$$

since all stabilizers are proper subgroups of G.

2 AB and BA

We will investigate the relation between AB and BA for two matrices in the following diagram

$$\mathbb{C}^n \underset{A}{\overset{B}{\rightleftarrows}} \mathbb{C}^m.$$

The main purpose of this section is to find the exact characterization of two matrices P and Q such that $P = BA$ and $Q = AB$, see Theorem 2.7.

Actually, this is equivalent to the problem of finding all indecomposable modules for a two-cyclic quiver and without extra efforts, the main arguments can be moved for any cyclic quiver. Actually, the material of this section comes from the book in this direct [1]. It turns out that it relates to the representation theory of Hecke algebras (of Type A) at the roots of unity from the geometric side. This provides one of the most successful examples of geometric methods in representation theory.

Firstly, we present here two basic problems which warm us up about the relation of AB and BA.

▶ **Exercise 2.1** Show that

$$x^m \det(x\mathbf{1}_n - AB) = x^n \det(x\mathbf{1}_m - BA).$$

◀ **P12**

▶ **Exercise 2.2** Denote f the minimal polynomial for AB, then the minimal polynomial for BA is $f(x)$, $xf(x)$ or $\frac{f(x)}{x}$. ◀ **P12**

The following problem shows that if the Jordan type of AB is fixed, then BA has some constraints.

▶ **Problem 2.3** If AB is diagonalizable, then so is $BABA$. ◀ **P12**

▶ **Exercise 2.4** If $ABAB$ is diagonalizable, then so is $BABABA$.

Aemma 2.5 We can decompose

$$\mathbb{C}^n \underset{A}{\overset{B}{\rightleftarrows}} \mathbb{C}^m = U_\circ \underset{A}{\overset{B}{\rightleftarrows}} V_\circ \oplus U_\bullet \underset{A}{\overset{B}{\rightleftarrows}} V_\bullet$$

such that over BA is nilpotent over $U_\circ$ and invertible over $U_\bullet$, and AB is nilpotent over $V_\circ$ and invertible over $V_\bullet$.

— *Proof.* This is a special case of the Fitting lemma. For the readers' convenience, we will present the proof in this special case. Actually,

$$U_\circ = \bigcup_N \ker(AB)^N, \quad U_\bullet = \bigcap_N \operatorname{im}(AB)^N,$$
$$V_\circ = \bigcup_N \ker(BA)^N, \quad V_\bullet = \bigcap_N \operatorname{im}(BA)^N.$$

Then they are generalized eigenspace of the eigenvalue 0 and the direct sum of the rest of generalized eigenspaces for BA and AB subspaces respectively. It suffices to show $B(U_x) \subseteq V_x$ for $x = \circ, \bullet$ and vice versa. But this is clear. □

The above lemma reduces the problem to two parts — nilpotent case and the invertible case. For the latter case, i.e. when A and B are both invertible, it is kind of trivial since it is equivalent to say that, $AB = A(BA)A^{-1}$ similar to BA. In general, the Jordan blocks of AB other than that belonging to 0 are the same as those of BA. The most crucial part is the nilpotent case.

Recall that by the theory of Jordan canonical form any nilpotent matrix A takes the following form

$$\begin{aligned}
&0 \xleftarrow{A} \bullet \xleftarrow{A} \bullet \xleftarrow{A} \bullet \xleftarrow{A} \bullet\\
&0 \xleftarrow{A} \bullet \xleftarrow{A} \bullet \xleftarrow{A} \bullet \xleftarrow{A} \bullet\\
&0 \xleftarrow{A} \bullet \xleftarrow{A} \bullet\\
&0 \xleftarrow{A} \bullet,
\end{aligned}$$

with all $\bullet$'s form a set of basis. Actually, for the case we care about, it is reasonable to guess the same picture but alternating $\xleftarrow{A}$ and $\xleftarrow{B}$. To be exact, we call any one of two spaces

$$\begin{aligned}
&0 \xleftarrow{A} \bullet \xleftarrow{B} \bullet \xleftarrow{A} \bullet \xleftarrow{B} \cdots \longleftarrow \bullet\\
&0 \xleftarrow{B} \bullet \xleftarrow{A} \bullet \xleftarrow{B} \bullet \xleftarrow{A} \cdots \longleftarrow \bullet
\end{aligned}$$

a **strip**.

Theorem 2.6 If AB and BA are both nilpotent, then

$$\mathbb{C}^n \underset{A}{\overset{B}{\rightleftarrows}} \mathbb{C}^m$$

can be decomposed into a direct sum of strips.

— *Proof.* The proof is similar to the proof of Jordan's canonical form. We denote $V = \mathbb{C}^n \oplus \mathbb{C}^m$, and T be the sum $A + B$. We have a filtration

$$0 \subseteq \ker T \subseteq \ker T^2 \subseteq \cdots \subseteq V.$$

Note that each

$$\ker T^r = \ker(\overset{r}{\cdots} AB) \oplus \ker(\overset{r}{\cdots} BA)$$

is $\mathbb{Z}/2$-graded. Moreover, we have induced maps

$$\frac{\ker T}{0} \xleftarrow{\overline{T}} \frac{\ker T^2}{\ker T} \xleftarrow{\overline{T}} \cdots \xleftarrow{\overline{T}} \frac{\ker T^N}{\ker T^{N-1}} \xleftarrow{\overline{T}} \cdots .$$

Note that all of them are injective. By identifying them with their image, they form a flag of

$$\ker T = \ker B \oplus \ker A.$$

Note that every $\frac{\ker T^r}{\ker T^{r-1}}$ is still $\mathbb{Z}/2$-graded and $\overline{T}$ has degree $\bar{1} \in \mathbb{Z}/2$. In particular, the flag is compatible with the grading, i.e. it is a direct sum of two flags in $\ker A$ and $\ker B$. By picking a set of the basis for this flag, and picking a lifting, we see that V decomposes into strips. □

In particular, thanks to the theorem above, the nilpotent case reduces to a combinatorial problem.

Theorem 2.7 For two matrices P and Q, there exists A and B such that $P = BA$ and $Q = AB$ if and only if

(i) the Jordan blocks belonging to nonzero eigenvalues of P and Q are the same, and

(ii) there exists a pairing of sizes of Jordan blocks belonging to eigenvalue 0 for P and Q such that in each pair their difference is at most 1.

For example, if AB is diagonalizable (i.e. all Jordan blocks are of size 1), the Jordan blocks of BA belonging to 0 can be predicted to have size at most 2.

For example, if the minimal size of the Jordan block of AB (resp. BA) belonging to 0 is r (resp. r'), then $|r - r'| \leq 1$. For other eigenvalues, the minimal sizes must be the same. This is exactly Exercise 2.2.

Finally, we left readers to think about what happened for more matrices.

References

[1] Ariki S. Representations of quantum algebras and combinatorics of Young tableaux[M]. American Mathematical Soc., 2002.

Hints

2.1 $\det \begin{pmatrix} x\mathbf{1}_n - AB & A \\ & x\mathbf{1}_m \end{pmatrix} = \det \begin{pmatrix} x\mathbf{1}_n & A \\ xB & x\mathbf{1}_m \end{pmatrix} = \det \begin{pmatrix} x\mathbf{1}_n & A \\ & x\mathbf{1}_m - BA \end{pmatrix}$.

2.2 It is clear $xf(x)$ is a polynomail for BA since $Af(BA)B = ABf(AB) = 0$. So the minimal polynomial g for BA divides $xf(x)$. Conversely, $f(x)$ divides $xg(x)$ which leads to the assertion.

2.3 Note that the minimal polynomial BA has a multiple root only if 0 has multiplicity 2 by Exercise 2.2. But in this case $BABA$ is multiplicity free.

3 Estimation of Eigenvalues

In this section, we will discuss how to estimate eigenvalues. To be exact, for an $n \times n$ complex matrix $A = (a_{ij})$, can we predict the possible places on the complex plane where its eigenvalues appear? Let $\lambda_1, \dots, \lambda_n$ be all the eigenvalues of A. The first theorem that comes to analysts' minds is probably the formula of spectral radius. Recall the **spectral radius** of a matrix A is the maximum of norms of eigenvalues of A.

▶ **Spectral radius 3.1** For any matrix A, its spectral radius is

$$\lim_{n\to\infty} \|A^n\|^{1/n}$$

for any matrix norm $\|\cdot\|$. ◀ **P16**

But in this section, we would like to give a couple of theorems on the estimation of eigenvalues more directly by matrix indices. The first affirmative relation is

$$\mathrm{tr} = \sum a_{ii} = \sum \lambda_i.$$

Thus it indicates that in the estimation the **diagonal elements** should contribute mostly. Here is a theorem on general matrices.

Gerŝgorin's disk Theorem 3.2 For any eigenvalue λ of A, we have

$$|\lambda - a_{ii}| \le |a_{i1}| + \cdots + \widehat{|a_{ii}|} + \cdots + |a_{in}|$$

for some i.

— *Proof.* Assume $Ax = \lambda x$, for $x = (x_1, \dots, x_n)^{\mathsf{t}} \neq 0$. That is, for any i,

$$a_{i1}x_1 + \cdots + a_{in}x_n = \lambda x_i.$$

Pick i such that $|x_i|$ is maximal. We get

$$(\lambda - a_{ii})x_i = a_{i1}x_1 + \cdots + \widehat{a_{ii}x_i} + \cdots + a_{in}x_n$$

Thus we have

$$\begin{aligned} |\lambda - a_{ii}| \cdot |x_i| &\le |a_{i1}| \cdot |x_1| + \cdots + \widehat{|a_{ii}| \cdot |x_i|} + \cdots + |a_{in}| \cdot |x_n| \\ &\le \big(|a_{i1}| + \cdots + \widehat{|a_{ii}|} + \cdots + |a_{in}|\big) \cdot |x_i| \end{aligned}$$

Since $|x_i| \neq 0$, we can conclude the assertion. □

▶ **Problem 3.3** We call a matrix $A = (a_{ij})$ to be **diagonally dominant** if

$$a_{ii} > |a_{i1}| + \cdots + \widehat{|a_{ii}|} + \cdots + |a_{in}|.$$

Show that every eigenvalue of any diagonally dominant matrix has a positive real part.

The problem turns out to be interesting if we restrict ourselves to the case of **hermitian matrices**. From now on, we assume A to be an Hermitian matrix, i.e. $A^{\mathrm{h}} = A$. Note that in this case, both diagonal elements and eigenvalues are real numbers so we can assume that

$$a_{11} \geq a_{22} \geq \cdots \geq a_{nn},$$
$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

Before stating the general theorem, let us see a couple of illustrating examples.

Example 3.4 Let $\left(\begin{smallmatrix} a & z \\ \bar{z} & b \end{smallmatrix}\right)$ be a hermitian matrix with $a \geq b$ with two eigenvalues $\lambda_1 \geq \lambda_2$. Then

$$\lambda_1 + \lambda_2 = a + b,$$
$$\lambda_1 \lambda_2 = ab - |z|^2.$$

So

$$(a - \lambda_1)(b - \lambda_1) = ab - \lambda_1(a + b - \lambda_1) = |z|^2 \geq 0.$$

Since $\lambda_1 \geq \lambda_2$ so that $\lambda_1 \geq \frac{a+b}{2}$, the above condition is equivalent to $\lambda \geq a$.

Example 3.5 The diagonal elements stand between the minimal eigenvalue and the maximal eigenvalue. That is,

$$\lambda_n \leq a_{nn} \leq \cdots \leq a_{11} \leq \lambda_1.$$

Actually, we can assume $A = U^{\mathrm{h}} D U$ for unitary matrix U with $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Assume that

$$U = (u_1, \ldots, u_n),$$

then

$$a_{ii} = u_i^{\mathrm{h}} D u_i \leq \lambda_1 u_i^{\mathrm{h}} u_i = \lambda_1.$$

Similar argument shows that $a_{ii} \geq \lambda_n$.

Schur–Horn's Theorem 3.6 For an hermitian matrix A,

$$a_{11} + \cdots + a_{ii} \leq \lambda_1 + \cdots + \lambda_i$$

for any $1 \leq i \leq n - 1$ and

$$a_{11} + \cdots + a_{nn} \leq \lambda_1 + \cdots + \lambda_n.$$

Moreover, if the above conditions hold then there exists such a matrix.

A clear proof using symplectic geometry can be found in [1, Example 5.5.4] as a shadow of the **Atiyah–Guillemin–Sternberg Theorem**.

On the other hand, one would think about what is the geometric notation for a "diagonal element". For any unit vector x, i.e. a vector with $x^{\mathrm{h}}x = 1$, $x^{\mathrm{h}}Ax$ is a diagonal element of A up to a unitary transformation. This observation leads to the following theorem.

▶ **Rayleigh–Ritz Theorem 3.7** For an hermitian matrix A,

$$\lambda_1 = \max_{x\neq 0} \frac{x^{\mathrm{h}}Ax}{x^{\mathrm{h}}x}, \qquad \lambda_n = \min_{x\neq 0} \frac{x^{\mathrm{h}}Ax}{x^{\mathrm{h}}x}$$

If we want to go further — to see the second minimal eigenvalue λ_{n-1}, in principle, we can work over the orthogonal complement V of the eigensubspace of λ_n. Luckily, we have a good observation that this space V maximize

$$\min_{x\in V\setminus 0} \frac{x^{\mathrm{h}}Ax}{x^{\mathrm{h}}x},$$

among all $(n-1)$-dimensional subspaces. Actually, this can be seen immediately by assuming $A = D$ to be diagonal. To be exact, denoting $U = \operatorname{span}(\mathbf{e}_{n-1}, \mathbf{e}_n)$. For general V,

$$\min_{x\in V\setminus 0} \frac{x^{\mathrm{h}}Dx}{x^{\mathrm{h}}x} \leq \min_{x\in (V\cap U)\setminus 0} \frac{x^{\mathrm{h}}Dx}{x^{\mathrm{h}}x} = \lambda_{n-1}.$$

Here we use the fact that $\dim V = n-1$ to ensure that $V \cap U \neq 0$.

▶ **Courant–Fischer Theorem 3.8** For an hermitian matrix A,

$$\lambda_i = \max_{\dim V = i} \min_{x\in V\setminus 0} \frac{x^{\mathrm{h}}Ax}{x^{\mathrm{h}}x} = \min_{\dim V = n+1-i} \max_{x\in V\setminus 0} \frac{x^{\mathrm{h}}Ax}{x^{\mathrm{h}}x}.$$

▶ **Interlacing 3.9** Let $A = \left(\begin{smallmatrix} B & x \\ x^{\mathrm{h}} & c \end{smallmatrix}\right)$ be a hermitian matrix with B a submatrix of size one less than A. Show that

$$\begin{array}{ccccccccccccc} \lambda_1 & & \lambda_2 & & \cdots & & & & \lambda_{n-1} & & \lambda_n \\ \geq & \geqslant & \geq & \geqslant & & \cdots & & \geqslant & \geq & \geqslant & \\ & \mu_1 & & \mu_2 & & & \cdots & & & \mu_{n-1} & \end{array}$$

with $\lambda_1 \geq \cdots \geq \lambda_n$ eigenvalues of A, and $\mu_1 \geq \cdots \geq \mu_{n-1}$ eigenvalues of B. ◀ P16

▶ **Weyl Theorem 3.10** We have

$$\lambda_{i+j-1}(A+B) \leq \lambda_i(A) + \lambda_j(B) \leq \lambda_{i+j-n}(A+B)$$

for hermitian matrices A, B with $\lambda_1(\cdot) \geq \cdots \geq \lambda_n(\cdot)$ equipped with the obvious meaning and $\lambda_{<1} = \infty$ and $\lambda_{>n} = -\infty$. ◀ P17

▶ **Problem 3.11** Let $B = A + xx^{\mathrm{h}}$ be an Hermitian matrix, a rank-one positive modification. Show that

$$\begin{array}{ccccccccc} & \lambda_1 & & \lambda_2 & & \cdots & & \lambda_n & \\ \mu_1 & \geq & \mu_2 & \geq & \cdots & \geq & \mu_n & \geq & \end{array}$$

with $\lambda_1 \geq \cdots \geq \lambda_n$ eigenvalues of A, and $\mu_1 \geq \cdots \geq \mu_n$ eigenvalues of B.

We lastly remark that there is a description for the exact conditions $(\lambda_1 \geq \cdots)$, $(\mu_1 \geq \cdots)$ and $(\nu_1 \geq \cdots)$ should satisfy such that they are eigenvalues of hermitian matrices A, B and $A+B$. This is known as **Horn's conjecture** which is proved by Schubert calculus and symplectic geometry. A good survey in this direction is [2].

References

[1] D. Mcduff, D. Salaon, Introduction to Symplectic Topology. Third Edition. Oxford Express.

[2] Knutson, Allen, Tao, Terence. Honeycombs and sums of Hermitian matrices. Notices Amer. Math. Soc. 48. (2000).

Hints

3.1 Since all the norms are equivalent over finite-dimensional spaces, we can pick $\|\cdot\|$ such that $\|Ax\| \leq \|A\| \cdot \|x\|$ (for example $\|A\| = \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}$). Assume $Ax = \lambda x$, then we have

$$|\lambda|^n \|x\| = \|A^n x\| \leq \|A^n\| \cdot \|x\|.$$

This shows $|\lambda| \leq \|A^n\|^{1/n}$. On the other hand, denoting d the spectral radius, we see that $A_1 = A/(d+\epsilon)$ satisfies $\|A_1^n\| \to 0$ (for example, computing with Jordan canonical form). In particular, for $n \gg 0$,

$$\frac{\|A^n\|}{(d+\epsilon)^n} < 1,$$

which shows $\|A^n\|^{1/n} < d + \epsilon$.

3.9 It suffices to show $\lambda_1 \geq \mu_1$ since the rest follows from the result for $-A = \left(\begin{smallmatrix} -B & -x \\ -x^{\mathrm{h}} & -c \end{smallmatrix}\right)$. This is a direct application of Theorem 3.8.

3.10 By Theorem 3.8,

$$\begin{aligned}
\lambda_i(A)+\lambda_j(B) &= \max_{\substack{\dim V=i\\ \dim U=j}} \min_{\substack{x\in V\setminus 0\\ y\in U\setminus 0}} \left(\frac{x^{\mathtt{h}}Ax}{x^{\mathtt{h}}x}+\frac{y^{\mathtt{h}}By}{y^{\mathtt{h}}y}\right)\\
&\le \max_{\substack{\dim V=i\\ \dim U=j}} \min_{x=y\in(U\cap V)\setminus 0} \left(\frac{x^{\mathtt{h}}Ax}{x^{\mathtt{h}}x}+\frac{y^{\mathtt{h}}By}{y^{\mathtt{h}}y}\right)\\
&\le \max_{\dim(U\cap V)\ge i+j-n} \min_{x\in(U\cap V)\setminus 0} \left(\frac{x^{\mathtt{h}}(A+B)x}{x^{\mathtt{h}}x}\right)\\
&\le \max_{d\ge i+j-n} \lambda_d(A+B)=\lambda_{i+j-n}(A+B).
\end{aligned}$$

4 Convexity

In this section, we will work with a linear space V over $\mathbb{R}$ and use its analytic properties. Our main concern is convex sets. Recall that a subset $\mathcal{C}$ in V is called **convex** if

$$\forall x, y \in \mathcal{C}, \quad \{tx + (1-t)y\}_{0 \le t \le 1} \subseteq \mathcal{C}.$$

We will show the separation theorem 4.3 stating that any two disjoint convex sets can be separated by a hyperplane. This statement is intuitively clear, but the proof is far from easy.

Definition 4.1 For each convex set containing 0, we can define the associated **Minkowski functional** $m_{\mathcal{C}} : V \to \mathbb{R}$ by setting

$$m_{\mathcal{C}}(x) = \inf \left\{ \lambda > 0 : \tfrac{x}{\lambda} \in C \right\}.$$

For instance, for the disk (solid sphere) of $\mathbb{R}^n$, the associated Minkowski functional is nothing but the standard Euclidean norm.

Aemma 4.2 For any convex set $\mathcal{C}$ containing 0, the Minkowski functional $m = m_{\mathcal{C}}$ satisfies

(i) $m(\lambda x) = \lambda m(x)$ for any $\lambda \ge 0$;
(ii) $m(x + y) \le m(x) + m(y)$.

— *Proof.* The first property is clear. For the second, we see

$$\frac{x}{m(x) + \epsilon} \quad \text{and} \quad \frac{y}{m(y) + \epsilon} \in \mathcal{C}.$$

Then

$$\frac{x + y}{m(x) + m(y) + 2\epsilon} \in \mathcal{C}.$$

We get $m(x + y) \le m(x) + m(y)$. □

Separation Theorem 4.3 For a convex set $\mathcal{C}$, and a point $x \notin \mathcal{C}$, there exists a linear function ϕ over V such that

$$\phi(\mathcal{C}) \le \phi(x).$$

— *Proof.* Firstly, we can assume $\mathcal{C}$ contains an interior point. The reason is the following. Let V' be the affine space spanned by $\mathcal{C}$. By translation, we can

assume V' is a subspace. Assume V' is defined by $\{\phi_i = 0\}$ for some linear functions ϕ_i.

(i) If $x \notin V'$, then there is one ϕ_i such that $\phi_i(x) \neq 0$, so $\pm\phi_i$ works.

(ii) If $x \in V'$, then $\mathcal{C}$ contains an interior point as a subset of V'. If we can find ϕ' over V' separating $\mathcal{C}$ and x, then any extension of ϕ' to V works.

By translation, we can assume $\mathcal{C}$ containing 0 as an interior point. Denote $m = m_{\mathcal{C}}$. Note that

$$m(\mathcal{C}) \leq 1 \leq m(x).$$

Denote ϕ the linear function over the space spanned by x given by $\phi(\lambda x) = \lambda$. Thus, it suffices to extend ϕ to V such that

$$\phi(v) \leq m(v)$$

for any $v \in V$. Note that this is already true over $\operatorname{span}(x)$, that is,

$$\phi(\lambda x) = \lambda \leq \begin{cases} \lambda m(x) = m(\lambda x), & \lambda > 0, \\ 0 \leq m(\lambda x), & \lambda \leq 0. \end{cases}$$

To extend ϕ, inductively, it suffices to extend it by one dimension. Say, assume ϕ is already defined over V_1, then for any $y \notin V_1$, to extend ϕ to y, we need to ensure that for any $\lambda \neq 0$ and $v \in V_1$

$$\lambda\phi(y) + \phi(v) \leq m(\lambda y + v).$$

It suffices to ensure when $\lambda = \pm 1$ by (i) of Lemma 4.2. So we get

$$-m(-y + v) + \phi(v) \leq \phi(y) \leq m(y + v) - \phi(v).$$

In particular, to show the existence of the extension, we need the supremum of the left-hand side not greater than the infimum of the right-hand side. In other words, it suffices to show for any $v, v' \in V'$,

$$-m(-y + v) + \phi(v) \leq m(y + v') - \phi(v').$$

This is true since

$$m(y + v') + m(-y + v) \geq m(v + v') \geq \phi(v) + \phi(v').$$

The proof is complete. □

Actually, the above separation theorem still works in Banach spaces using **Hahn–Banach extension theorem** in which case we should assume the existence of interior points, see [1].

▶ **ℙProblem 4.4** Show that for two disjoint convex subset $\mathcal{C}_1$ and $\mathcal{C}_2$, there exists a linear function ϕ on V such that

$$\phi(\mathcal{C}_1) \leq \phi(\mathcal{C}_2).$$

◀ **P21**

▶ **ℙProblem 4.5(Cancellation)** Let P and Q be two compact convex subsets, we define **Minkowski sum** and **Minkowski difference**

$$P + Q = \{p + q : p \in P, q \in Q\}$$
$$P - Q = \{x : x + Q \subseteq P\}$$

Show that $(P + Q) - Q = P$. ◀ **P21**

▶ **𝔽arkas Lemma 4.6** For a real matrix A and real vector b, then

Either	$Ax < b$ has a solution $x \geq \mathbf{0}$,
or	$y^{\mathsf{t}}A \geq \mathbf{0}$ has a nonzero solution $y \geq \mathbf{0}$ with $y^h b \leq 0$.

Here $(x_1, \ldots, x_n)^{\mathsf{t}} \geq (x'_1, \ldots, x'_n)^{\mathsf{t}}$ means $x_i \geq x'_i$ for all i. ◀ **P21**

▶ **ℙProblem 4.7** For a symmetric semi-positively definite matrix A, show that there exists a nonzero $x \geq 0$ such that $Ax \geq 0$. ◀ **P21**

▶ **𝔼Exercise 4.8** Assume $v_1, \ldots, v_d \in V$ such that for any **non-negative** number $x_1, \ldots, x_d \geq 0$,

$$x_1 v_1 + \cdots + x_d v_d = 0 \iff x_1 = \cdots = x_d = 0.$$

Show that there exists a linear function ϕ such that $\phi(v_i) > 0$ for all i.

𝒟efinition 4.9 We call $\mathcal{C}$ is a **polyhedron** if it is intersection of finite many hyperplanes, i.e.

$$\mathcal{C} = \bigcap \{\phi_i \geq b_i\}$$

for finite many linear functions ϕ_i's and real numbers b_i's. If all $b_i = 0$, then $\mathcal{C}$ is known as a **cone**. A bounded polyhedron is called a **polytope**.

𝒟efinition 4.10 For a subset $S \subseteq V$, we denote the **convex hull** hull(S) the set of element $x \in V$ able to be written as

$$x = \sum_{v \in S} t_v v \quad \text{(finite sum)}, \qquad \sum_{v \in S} t_v = 1.$$

We denote the **conic hull** $\operatorname{cone}(S)$ the set of element $x \in V$ able to be written

$$x = \sum_{v \in S} t_v v \quad \text{(finite sum)}, \qquad {}^{\forall v \in S,}\ t_v \geq 0.$$

For two sets S and T, we denote their **(Minkowski) sum**

$$S + T = \{s + t : s \in S, t \in T\}.$$

Theorem 4.11 A subset $\mathcal{C}$ is a polytope (resp., cone) if and only if it is a convex (resp., conic) hull of a finite set. A subset $\mathcal{C}$ is a polyhedron if and only if it can be written as a Minkowski sum of a polytope and a polyhedral.

We refer [2] for the proof.

Hints

4.4 Consider $\mathcal{C}_1 - \mathcal{C}_2$.

4.5 Assume $x + Q \subseteq P + Q$. If $x \notin P$, we can separate x and P by a hyperplane, say $\phi(x) \leq \phi(P)$. Note that $\phi(Q)$ is a closed interval, it is impossible to have

$$\phi(x + Q) = \phi(x) + \phi(Q) \subseteq \phi(P) + \phi(Q) = \phi(P + Q).$$

4.6 The first condition says that $\{Ax\}_{x>\mathbf{0}}$ intersects $\{t : t < b\}$. Otherwise, they can be separated by $\{t : y^{\mathsf{h}} t = 0\}$ (after translation) for some nonzero y. That is $y^{\mathsf{t}} Ax \geq 0$ for all $x \geq 0$ and $y^{\mathsf{t}} t \leq 0$ for all $t < b$. This is equivalent to the second condition.

4.7 Otherwise, $\{Ax : x \geq 0\} \setminus 0$ is disjoint from $\{x \geq 0\}$. So we can find a nonzero y such that $y^{\mathsf{t}} Ax \leq 0$ and $y^{\mathsf{t}} x \geq 0$ for all $x \geq 0$. This shows $y \geq 0$, but then $y^{\mathsf{t}} Ay \leq 0$, a contradiction.

References

[1] Rudin, Walter. Functional analysis. Internat. Ser. Pure Appl. Math (1991).

[2] Ziegler, Günter. Lectures on Polytopes. Springer seires of Graduate Texts in Mathematics. 10.1007/978-1-4613-8431-1. (1994)

5 Perturbation Method

In this section, we will discuss the perturbation method making it possible to reduce problems to "regular" cases. A good example is the following proof of **Hamilton–Cayley theorem**.

Proof 5.1 Let A be a matrix with characteristic polynomial $\chi_A(x) = \det(x\mathbf{1} - A)$. Firstly, if A is diagonalizable, then it is clear that $\chi_A(A) = 0$. Secondly, $\chi_A(A)$ is a polynomial in entries of A, and it vanishes when A is diagonalizable. Thus $\chi_A(A) = 0$ for any A, by the theorem below.

Theorem 5.2 The subset of matrices A with different eigenvalues is a non-empty Zariski open (thus dense) subset.

— *Proof.* Actually, the subset is given by $\{f \neq 0\}$ with f the **resultant** of the characteristic polynomial χ_A. □

In the above example, the fact that $\chi_A(A)$ is continuous in its indices is vital in the argument. Thus it is unfair not to discuss the problem of **continuous/smooth dependence**. Fortunately, we are studying linear **algebra** where most of the values are polynomial-dependent. But, there still rest some non-continuous concepts.

Recall a function f is called **lower semi-continuous** if

$$\forall \epsilon > 0, \exists \text{ a neighborhood } U \text{ of } x_0 \text{ such that}$$
$$\text{for any } x \in U, \text{ we have } f(x) > f(x_0) - \epsilon.$$

Equivalently,

$$\forall x \in \mathbb{R}, f^{-1}\big((x, \infty)\big) \text{ is open,}$$
$$\forall x \in \mathbb{R}, f^{-1}\big((-\infty, x]\big) \text{ is closed.}$$

Intuitively, it says, among all the way tending to x, the value $f(x)$ coincides with the lowest limit value. Similarly, we define **upper semi-continuous**.

Theorem 5.3 The function rank is a lower semi-continuous function over matrices spaces (under Zariski topology and thus in usual topology).

— *Proof.* We need to show the subset of matrix A with rank $\leq r$ is closed. Note that it suffices to show when r is an integer. This is true since it is defined by all minors of size $r + 1$. □

▶ **Problem 5.4** Show that $(V, U) \longmapsto \dim(V+U)$ is an upper semi-continuous function over the space of pair of subspaces of given dimensions (say, the product of two Grassmannians). ◀ **P25**

▶ **Problem 5.5** For a nilpotent matrix A, we denote $\lambda'(A) = \lambda'_1 \geq \cdots$ with $\lambda'_i = \dim \ker A^{i-1} - \dim \ker A^i$. Show that λ' is lower semi-continuous in the following sense.

$$\lambda' \geq \mu' \iff {}^{\forall i,}\ \lambda'_1 + \cdots + \lambda'_i \geq \mu'_1 + \cdots + \mu'_i.$$

For example,

◀ **P25**

Now, let us turn to the continuity of eigenvalues. Note that this problem is essentially the continuous dependence of the roots of a polynomial on its coefficients. Before going deep, we first remark that the roots are **NOT** a continuous function in its coefficients. For example, for $f(x) = x^2 + bx + c$, its two roots

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - c}}{2}$$

are continuous over $\mathbb{R}$ but can never be continuously extended to $\mathbb{C}$.

To be exact, consider a morphism

$$\pi : R_{\mathbb{C}} = \mathbb{C}^n \longrightarrow \mathbb{C}^n = P_{\mathbb{C}}, \qquad (z_1, \ldots, z_n) \longmapsto (e_1(-z), \ldots, e_n(-z))$$

where e_i is the elementary symmetric polynomial. We will think $R_{\mathbb{C}}$ as the space of roots, and $P_{\mathbb{C}}$ as the space of monic polynomials of degree n. Note that generically, π is $n!$ to 1, but not all point is the case.

Theorem 5.6 Let $f(z) = z^n + \sum a_i z^i$ be a monic polynomial of degree n with distinct roots $\lambda_1, \ldots, \lambda_m$ with multiplicity m_i. Then for any $\epsilon > 0$, there exists $\delta > 0$ such that for any monic polynomial $g = z^n + \sum b_i z^n$ of degree n such that $\|f - g\| < \delta$ (any given norm $\|\cdot\|$), there are exactly m_i many roots in the disk $\{z : |z - \lambda_i| < \epsilon\}$.

— *Proof.* Now we consider the subset of $(z_1, \ldots, z_n)$ such that there are exactly m_i many of them in the disk $\{z : |z - \lambda_i| < \epsilon\}$. Note that this is an open subset. Now the theorem follows from the fact that π is a full-rank holomorphic function which is in particular open. This also follows from **Rouché's Theorem**. □

Aemma 5.7 For a monic polynomial $f = x^n + \sum a_i x^i$, for any $\delta > 0$ sufficiently small such that the δ-disks centered at roots of f are disjoint,

we can find $\delta > 0$ such that whenever $\max|a_i - b_i| < \delta$, the polynomial $g = x^n + \sum b_i x^i$ has the same number of roots as that of f in each δ-disks centered at roots of f.

Proof. Firstly, we pick δ' to be the minimum of $|f(x)|$ for x on the boundary of those disks. Then whenever $|g(z) - f(z)| < \delta' < |f(z)|$ on the boundary of each disk, g and f share the same number of roots inside the disk by **Rouché theorem**.

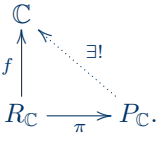
Denote C the union of the boundaries. Secondly, we claim that there exists $\delta > 0$ such that $\sum |a_i - b_i| < \delta$ implies $|g(z) - g(z)| < \delta'$ for all $z \in C$ where $g = x^n + \sum b_i x^i$. This follows from the following observation. Other then $(a_i) \mapsto \|(a_i)\| = \max|a_i|$, we can define a new norm

$$(a_i) \longmapsto \|(a_i)\|' = \max|a_i| + \min_{z \in C} \left|z^n + \sum a_i z^i\right|$$

which is equivalent to $\|\cdot\|$. As a result, such δ exists. □

Theorem 5.8 Let $\mathfrak{F}(X)$ be the space of continuous function, holomorphic function, or polynomial function over $X = P_{\mathbb{C}}$ or $R_{\mathbb{C}}$.

Then any symmetric $f \in \mathfrak{F}(R_{\mathbb{C}})$ can be uniquely written as $\tilde{f} \circ \pi$ for $\tilde{f} \in \mathfrak{F}(P_{\mathbb{C}})$.



In other words, $\mathfrak{F}(R_{\mathbb{C}})^{\mathcal{S}_n} = \mathfrak{F}(P_{\mathbb{C}})$.

— *Proof.* When $\mathfrak{F}$ is the space of polynomial functions, this is exactly the fundamental theorem of symmetric polynomials — every symmetric polynomial is a unique polynomial in e_i's.

For the continuous world, we consider the quotient space $R_{\mathbb{C}}/\mathcal{S}_n$ with the natural quotient topology. Note that it induces $\hat{\pi} : R_{\mathbb{C}}/\mathcal{S}_n \to P_{\mathbb{C}}$ which is a continuous bijection. Actually, π is open by Lemma 5.7.

Note that when $\mathfrak{F}$ is the space of continuous functions, $R_{\mathbb{C}}/\mathcal{S}_n$ satisfies the above universal property. Thus it suffices to show π induces an isomorphism between $R_{\mathbb{C}}/\mathcal{S}_n$ and $P_{\mathbb{C}}$. This follows from the fact that π is open since it is a full-rank holomorphic morphism.

When it comes to the case of holomorphic functions. We equip $R_{\mathbb{C}}/\mathcal{S}_n$ with complex structure with coordinates induced by $e_1, \ldots, e_n \in \mathfrak{F}(R_{\mathbb{C}})^{\mathcal{S}_n}$. Then $\hat{\pi} : R_{\mathbb{C}}/\mathcal{S}_n \to P_{\mathbb{C}}$ is a morphism of complex manifold which is bijective. Then by **Rouché's theorem**, $\hat{\pi}$ is actually an isomorphism which tells $\mathfrak{F}(R_{\mathbb{C}})^{\mathcal{S}_n} = \mathfrak{F}(P_{\mathbb{C}})$. □

In particular, a symmetric and continuous statement on the roots is continuously dependent on its coefficients. Another useful conclusion on the distribution of roots is the following.

Recall that the (first) Gerŝgorin Theorem 3.2 shows that the eigenvalues are contained in a union of disks. The above theorem is used to deduce the **second Gerŝgorin Theorem** predicting the number in each connected component.

▶ **Second Gerŝgorin Theorem 5.9** Let $A = (a_{ij})$ be a matrix. Consider the union of disks in Theorem 3.2. In each connected component consisting of d many disks, there are exactly d many eigenvalues (counting with multiplicities). ◀ **P25**

More specifically, a symmetric and continuous statement on the eigenvalues is continuously dependent on matrices entries. For example, **spectral radius** is a continuous function.

▶ **Problem 5.10** Assume A commute with C, show that

$$\det \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) = \det(AD - CB).$$

◀ **P25**

Hints

5.4 Actually, by picking base, this map is the rank of two sets of vectors which is semi-continuous.

5.5 This is simply because $\lambda_1'(A)+\cdots+\lambda_i(A) = \dim \ker A^{i-1} = n - \operatorname{rand} A^i$ with n the size of A.

5.9 Consider $\begin{pmatrix} a_{11} & ta_{12} & \cdots & ta_{1n} \\ ta_{21} & a_{22} & \cdots & ta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ta_{n1} & ta_{n2} & \cdots & a_{nn} \end{pmatrix}$.

5.10 Assume firstly A is invertible.

$$\det \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) = \det \left(\begin{smallmatrix} A & B \\ 0 & D - CA^{-1}B \end{smallmatrix}\right) = \det A \det(D - CA^{-1}B).$$

6 Non-negative Matrices

In this section, we will discuss **non-negative matrices**, that is, matrices with non-negative indices. We will use the notation $\geq$ (or $\leq$) between matrices or vectors if each index is the case. If furthermore, they are not equal, we denote $>$ (or $<$). On the other hand, we say a matrix or vector is **positive** (or **negative**), if all of its indices are the case. Note that

$$\begin{aligned}(x_1,\dots,x_n) \geq 0 &\iff \text{all } x_i \geq 0,\\ (x_1,\dots,x_n) > 0 &\iff \text{all } x_i \geq 0 \text{ and some } x_i > 0,\\ (x_1,\dots,x_n) \text{ is positive} &\iff \text{all } x_i > 0.\end{aligned}$$

The most fundamental theorem in the study of non-negative matrices is the **Frobenius theorem**. There are a lot of variations of it which we will discuss after proving the most fundamental one.

Theorem 6.1 (Frobenius) Positive square matrix A admits a positive eigenvector belonging to a positive eigenvalue.

— *Proof.* Consider the subset $\mathcal{S}$ of $x \in \mathbb{R}^n$ with $x > 0$ and $|x| = 1$. Note that $\mathcal{S}$ is homoeomorphic to a closed disk. Then the map $x \mapsto \frac{Ax}{|Ax|}$ is well-defined over $\mathcal{S}$ since all indices of A are positive. By the **Brouwer fixed point theorem**, this map has a fixed point, i.e. there exists $x \in \mathcal{S}$ such that $Ax = |Ax| \cdot x$. Note that since A is positive and $x \geq 0$, so Ax is positive, thus x is actually positive. □

The readers are welcomed to refer [1, Chatper 3] for the application of this theorem in category theory.

▶ **Problem 6.2** Under the assumption of theorem 6.1, show that there is exactly one **simple** eigenvalue admits a positive eigenvector. ◀ **P28**

▶ **Problem 6.3** Show that if $A \geq 0$, then A admits an eigenvector $v \geq 0$ belonging to $\lambda \geq 0$. ◀ **P28**

▶ **Exercise 6.4** Computing the eigenvalues of the matrix $N = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$.

▶ **Problem 6.5** If $A \geq 0$ with the biggest real eigenvalue $\lambda \geq 0$, then any eigenvalue μ satisfying $|\mu| \leq \lambda$. ◀ **P29**

Now, let us restrict ourselves to the case of symmetric matrices. Assume A is symmetric with λ the maximal eigenvalue. Equivalently, in this case,

$$\lambda = \inf\left\{\lambda : -A + \lambda \mathbf{1} \text{ is positively definite}\right\}$$

Note that a positive definite matrix must have positive diagonal elements. In particular, the problem reduces to C-matrices in the following sense.

Definition 6.6 We call a symmetric matrix C a **C-matrix** if the entries of C are all non-negative except diagonal elements all positive.

Remark 6.7 Note that geometrically, a matrix $C = (c_{ij})$ is positively definite if there exists a set of basis $\{v_1, \dots, v_n\}$ such that

$$c_{ij} = \langle v_i, v_j \rangle$$

under a given inner product $\langle \cdot, \cdot \rangle$. We can assume that technically the diagonal entries are all 1's and all v_i's are the units so that

$$c_{ij} = \cos \theta_{ij}$$

with θ_{ij} the angel between v_i and v_j. So the statement about positive definiteness is equivalent to the existence of certain vectors with promised angles pairwise. Now, being a positively definite C-matrix is the case when all angles are obtuse.

Example 6.8 For three $\theta_1, \theta_2, \theta_3 \in [\pi/2, \pi]$,

$$\begin{pmatrix} 1 & \cos\theta_1 & \cos\theta_2 \\ \cos\theta_1 & 1 & \cos\theta_3 \\ \cos\theta_2 & \cos\theta_3 & 1 \end{pmatrix}$$

is positive definite if and only if $\theta_1 + \theta_2 + \theta_3 < 2\pi$.

Theorem 6.9 A C-matrix C is positively definite if and only if $Cx > 0$ for some positive x.

— *Proof.* We already established the existence of $x \geq 0$ such that $Cx \geq 0$ by Problem 4.7. Denote $I = \{i : x_i = 0\}$. Then by assumption, for $i \in I$

$$(Cx)_i = c_{ii}x_i + \sum_{j \neq i} c_{ij}x_j = \sum_{j \neq i} c_{ij}x_j \geq 0$$

This happens if and only if $c_{ij} = 0$ for all $j \notin I$. In other words, by reordering the index, we can assume $C = \begin{pmatrix} C_1 & \\ & C_2 \end{pmatrix}$, and $x = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$ with x_1 positive. By induction, there is a positive x_2 with $C_2 x_2 > 0$. So finally, $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ serves.

Conversely, we can assume without loss of generality that $x = (1, \dots, 1)^{\mathsf{t}}$. Then

$$\sum c_{ii}v_i^2 + \sum_{i \neq j} c_{ij}v_i v_j = \sum_i \Big(\sum_j c_{ij}\Big)v_i^2 - \sum_{i \neq j} c_{ij}(v_i - v_j)^2,$$

which is positive definite. □

▶ **Problem 6.10** Let C be an invertible C-matrix, then the following statements are equivalent

(1) C is positively definite;
(2) $P^{\mathsf{t}}CP = \mathbf{1}$ for some upper triangular matrix $P \geq 0$;
(3) $C^{-1} > 0$. ◀ **P29**

▶ **Exercise 6.11** A C-matrix C is positively semi-definite if and only if $Cx \geq 0$ for some positive x.

In particular if C is positively semi-definite but not definite if and only if $Cx = 0$ for some positive x.

▶ **Problem 6.12** For two C-matrices $C = (c_{ij})$ and $D = (d_{ij})$, if $c_{ii} = d_{ii}$ and for $i \neq j$,

$$|c_{ij}| \leq |d_{ij}|. \qquad (*)$$

If D is positively semi-definite, and at least one $(*)$ is strict, then C is positive definite. ◀ **P29**

▶ **Problem 6.13** For a C-matrix $C = \left(\begin{smallmatrix} C_1 & X \\ X^{\mathsf{t}} & C_2 \end{smallmatrix}\right)$, if C is positively semi-definite and $X \neq 0$, then C_1 is positively definite.

References

[1] Etingof, P., Gelaki, S., Nikshych, D., and Ostrik, V. (2016). Tensor categories (Vol. 205). American Mathematical Soc..

Hints

6.2 Note that A^{t} has the same eigenvalue as A, assume $x^{\mathsf{t}}A = \lambda x^{\mathsf{t}}$ and $Ay = \mu y$ for x, y, then $\lambda x^{\mathsf{t}}y = x^{\mathsf{t}}Ay = \mu x^{\mathsf{t}}y$. Note that $x^{\mathsf{t}}y \neq 0$. Assume $Ax = \lambda x$ for a positive x and positive λ. Assume $Av = \lambda v$ for some v. If v is not a scalar of x, then we can choose t such that $v' = v + tx > 0$ with at least one zero indices. Now $Av' = \lambda v'$ is positive, which is a contradiction. We also need to show that $Av - \lambda v = x$ has no solution. We can pick t such that $v' = v + tx < 0$ with at least one zero indices. Then $Av' = x + \lambda v'$ invokes a contradiction.

6.3 Just repeat the proof of Frobenius theorem 6.1. If $x \mapsto \frac{Ax}{|Ax|}$ is not well-defined, i.e. $Ax = 0$ for some x, then 0 is an eigenvalue.

6.5 Firstly assume A is positive. Assume $x^{\mathrm{t}}A = x^{\mathrm{t}}\lambda$. We denote a new norm $|v|_1 = x^{\mathrm{t}}v' := \sum x_i|v_i|$ where $v'_i = |v_i|$ for each i. We find

$$|Av| = \sum x_i\big|A_{ij}v_j\big| \le \sum x_iA_{ij}\big|v_j\big| = \lambda\sum x_j|v_j| = \lambda|v|.$$

If $Av = \mu v$, then $|Av| = |\mu|\cdot|v|$. For general A, we consider $A + tN$ for N the matrices with all indices 1's.

6.10 (1) $\Rightarrow$ (2). We can assume without loss of generality that $c_{11} = 1$. Now $C = \left(\begin{smallmatrix} 1 & \\ x^{\mathrm{t}} & \mathbf{1}\end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & -x^{\mathrm{t}} \\ -x & C'\end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & x^{\mathrm{t}} \\ & \mathbf{1}\end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & \\ & C'-xx^{\mathrm{t}}\end{smallmatrix}\right)$. Note that when $i \ne j$,

$$(C' - xx^{\mathrm{t}})_{ij} = c_{i+1,j+1} - c_{i+1,1}c_{1,j+1} \le 0.$$

Since C is positively definite, thus so is $C' - xx^{\mathrm{t}}$. In particular, $C' - xx^{\mathrm{t}}$ is still a C-matrix.

The implication (2) $\Rightarrow$ (3) is clear. Actually, $C^{-1} = PP^{\mathrm{t}}$. For (3) $\Rightarrow$ (1), it follows easily from Theorem 6.9.

6.12 Assume $Dx \ge 0$ for some positive x, note that $Cx - Dx > 0$ and is strict if one of $(*)$ is strict.

7 ADE Classifications

In this section, we will discuss the famous phenomenon called **ADE classification**. It naturally appears in the study of integer quadratic form. More precisely, we are asking what conditions should $c_{ij} \in \mathbb{Z}_{\leq 0}$ satisfy such that

$$q(x) = \sum_{i=1}^{n} x_i^2 + \sum_{1 \leq i < j \leq n} c_{ij} x_i x_j$$

is positive definite. Let us denote the corresponding **Cartan matrix**

$$C = \begin{pmatrix} 2 & c_{12} & \cdots & c_{1n} \\ c_{12} & 2 & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1n} & c_{2n} & \cdots & 2 \end{pmatrix}$$

It is clear that q is positive definite if and only if C is. Let us denote the associated **Dynkin diagram** $\mathcal{D}$ be the graph with vertices $1, \ldots, n$, and there are $|c_{ij}|$ many edges between i and j. We denote

$i \quad j$	i —— j	$i \overset{\lvert c_{ij} \rvert}{\text{——}} j$
$c_{ij} = 0$	$c_{ij} = -1$	$\lvert c_{ij} \rvert > 1$

Note that the order of vertices does not essentially matter, so we will denote $\circ$ by a vertex rather than i.

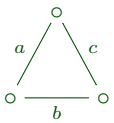
Example 7.1 For example, for $n = 2$, $c \leq 0$

$$x^2 + cxy + y^2 \text{ is } \begin{cases} \text{positively definite} & |c| = 0, 1 \\ \text{positively semi-definite} & |c| = 2 \\ \text{indefinite} & |c| > 2 \end{cases}$$

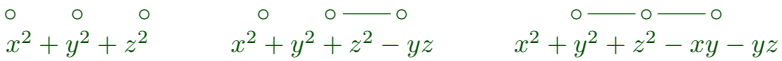
The Cartan matrix is $\left(\begin{smallmatrix} 2 & c \\ c & 2 \end{smallmatrix}\right)$. Their Dynkin diagrams are

$\circ \quad \circ$	$\circ$ —— $\circ$	$\circ \overset{2}{\text{——}} \circ$	$\circ \overset{3}{\text{——}} \circ$
$x^2 + y^2$	$x^2 - xy + y^2$	$x^2 - 2xy + y^2$	$x^2 - 3xy + y^2$

Example 7.2 For $n = 3$, $a, b, c \leq 0$, denote

$\mathcal{D} =$ 

Note that if any one of a, b, c are more than 3, then the quadratic form is indefinite, i.e. $\mathcal{D}$ does not contain $\circ \overset{3}{\text{—}} \circ$ as a subgraph. By Example 6.8, the quadratic form with the following diagram is positively definite (up to a permutation of vertices)



$x^2 + y^2 + z^2$ $\qquad$ $x^2 + y^2 + z^2 - yz$ $\qquad$ $x^2 + y^2 + z^2 - xy - yz$

The following diagrams correspond to positive semi-definite but not definite forms

$x^2 + y^2 + z^2 - 2yz$ $\qquad$ $x^2 + y^2 + z^2 - xy - yz - xz$

Definition 7.3 Denote the following **affine ADE diagrams**

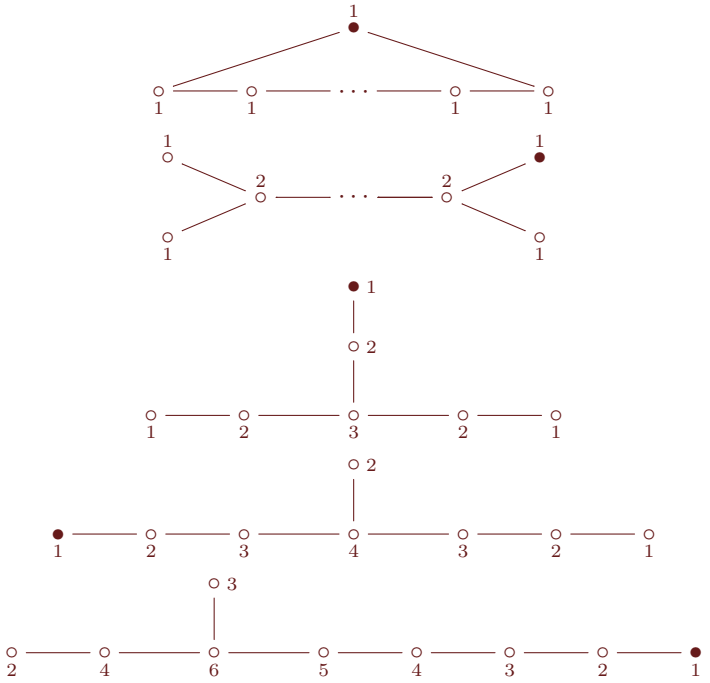
$\widetilde{\mathbb{A}}_n (n \geq 1)$

$\widetilde{\mathbb{D}}_n (n \geq 4)$

$\widetilde{\mathbb{E}}_6$

$\widetilde{\mathbb{E}}_7$

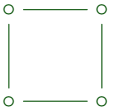
$\widetilde{\mathbb{E}}_8$



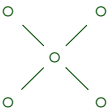
Note that the subscript is one less than the number of vertices. Note that $\widetilde{A}_1$ is $\circ \overset{2}{—} \bullet$.

▶ **Problem 7.4** The corresponding quadratic forms of affine ADE diagrams are positively semi-definite but not definite. ◀ **P34**

Example 7.5 For example,



$$\tfrac{1}{2}\left[(x_1-x_2)^2+(x_2-x_3)^2+(x_3-x_4)^2+(x_4-x_5)^2\right]$$



$$\tfrac{1}{4}\left[(2x_1-x_0)^2+(2x_2-x_0)^2+(2x_3-x_0)^2+(2x_4-x_0)^2\right]$$

Definition 7.6 Denote the following **(finite) ADE diagrams**

$\mathbb{A}_n (n \geq 1)$

$\mathbb{D}_n (n \geq 4)$

$\mathbb{E}_6$

$\mathbb{E}_7$

$\mathbb{E}_7$

where the subscript stands for the number of vertices. Note that, they are all possible proper connected subgraphs of affine ADE diagrams.

▶ **Problem 7.7** The corresponding quadratic forms of affine ADE diagrams are positive definite. ◀ **P34**

▶ **Problem 7.8** Denote the quadratic form

$$q(x) = \sum_{i=1}^{n} x_i^2 + \sum_{1 \le i < j \le n} c_{ij} x_i x_j,$$

for $c_{ij} \in \mathbb{Z}_{\le 0}$. Then

(1) $q(x)$ is positively semi-definite if and only if each connected component of the corresponding Dynkin diagram is an affine or finite ADE diagram.

(2) $q(x)$ is positive definite if and only if each connected component of the corresponding Dynkin diagram is a finite ADE diagram. ◀ **P34**

We shall remark on a couple of generalizations of ADE classification as follows. In the study of Lie theory, see [1], we only need to assume C to be symmetrizable integer matrices where the vertices are of different "sizes". The resulting diagrams are known as **affine or finite Dynkin diagrams**. In the study of Coxeter groups (and regular polytopes), see [2], we assume $c_{ij} = 2\cos\frac{\pi}{d}$ for some $d \ge 2$. The resulting diagrams are known as **affine or finite Coxeter diagrams**. Other than the above cases, ADE classification appears in some seeming-non-relative cases. For example, there is a one-to-one correspondence between finite subgroups of SL_2 and ADE diagrams known as **McKay correspondence**, see [3, Chapter 8]. In the representation theory of associative algebra, a connected quiver is of finite representation type if and only if the underlying graph is a finite ADE diagram, known as **Gabriel theorem**, see [1, Chapter 4].

References

[1] Serre J P. Complex semisimple Lie algebras[M]. Springer Science & Business Media, 2000.

[2] Humphreys J E. Reflection groups and Coxeter groups[M]. Cambridge university press, 1990.

[3] Kirillov Jr A. Quiver representations and quiver varieties[M]. American Mathematical Soc., 2016.

[4] Benson D J. Representations and cohomology: Volume 1, basic representation theory of finite groups and associative algebras[M]. Cambridge university press, 1998.

Hints

7.4 Use Exercise 6.11. The labeled number is the required positive vector.

7.7 By Problem 6.12 and Problem 6.13.

7.8 Firstly, any $q(x)$ whose Dynkin diagram is connected and contains an affine Dynkin diagram properly cannot be positively semi-definite by Problem 6.12 and Problem 6.13. Note that $\widetilde{A}_n$'s exclude graphs with a cycle, $\widetilde{D}_n$'s conclude graphs with more than 4 branches. Finally, $\widetilde{E}_n$'s restrict the graphs with 3 branches.

8 Krull–Schmidt Theorem

In this section, we will discuss **Krull–Schmidt theorem** and how it is reflected in terms of linear algebra. It can be stated in the language of module theory as follows.

Theorem 8.1 (Krull–Schmidt Theorem) Assume we have a module isomorphism

$$X = M_1 \oplus \cdots \oplus M_m \cong N_1 \oplus \cdots \oplus N_n,$$

with all M_i and N_i and indecomposable (unable to write it as a direct sum of smaller submodules). If X satisfies Artinian and Noetherian conditions, then $n = m$, and $M_i \cong N_{\sigma(i)}$ for a permutation σ.

But in this section, we will give another form in terms of linear algebra. Before this, we give a couple of usual exercises which can be regarded as a shadow of Krull–Schmidt's theorem.

▶ **Problem 8.2** Let $\mathbb{F} \subseteq \mathbb{C}$ be a subfield (for example $\mathbb{R}$). If two $\mathbb{F}$-matrices A and B are similar over $\mathbb{C}$, then they are similar over $\mathbb{F}$. ◀ P38

Remark 8.3 Problem 8.2 is also true over finite field. This follows from the theory of λ-matrices. This is also a special case of **Noether–Deuring Theorem**, an application of Krull–Schmidt theorem.

▶ **Problem 8.4** If $\begin{pmatrix} A & \\ & A \end{pmatrix}$ is similar to $\begin{pmatrix} B & \\ & B \end{pmatrix}$, show that A is similar to B. ◀ P38

Definition 8.5 Let I be a fixed index set. Let $\mathcal{A} = \{A_i\}_{i\in I}$ and $\mathcal{B} = \{B_i\}_{i\in I}$ be two I-families of matrices of the same size (at least 1). We denote

$$\mathcal{A} \cong \mathcal{B}$$

if there exists a $P \in \mathrm{GL}_n$ such that for each $i \in I$

$$PA_iP^{-1} = B_i,$$

Definition 8.6 For families $\mathcal{A}^{(i)} = \{A_i^{(i)}\}_{i\in I}$, we denote

$$\mathcal{A}^{(1)} \oplus \cdots \oplus \mathcal{A}^{(r)} = \left\{ \begin{pmatrix} A_i^{(1)} & & \\ & \ddots & \\ & & A_i^{(r)} \end{pmatrix} \right\}.$$

We say $\mathcal{A}$ is **indecomposable** if we cannot write $\mathcal{A} \cong \mathcal{A}_1 \oplus \mathcal{A}_2$.

By definition, we can decompose any family $\mathcal{A}$ until all summands are indecomposable, or more precisely, by the finiteness of dimension, this process terminates. In other words,

$$\mathcal{A} \cong \mathcal{A}^{(1)} \oplus \cdots \oplus \mathcal{A}^{(r)}$$

with each $\mathcal{A}^{(i)}$ indecomposable. For example, if the size of matrices of each $\mathcal{A}_i$ is 1, then we usually say $\mathcal{A}$ can be **simultaneously diagonalized**.

Example 8.7 For the case I is a single point, then $\mathcal{A} = \{A\}$ is indocomposable if and only if A is similar to a Frobenius block. In particular, if everything is over $\mathbb{C}$, the condition is to require A similar to a Jordan block.

Remark 8.8 Geometrically, a decomposition

$$\mathcal{A} = \mathcal{A}^{(1)} \oplus \cdots \oplus \mathcal{A}^{(r)}$$

corresponds to a direct sum

$$V = V_1 \oplus \cdots \oplus V_r$$

with each V_j subspace invariant under $\mathcal{A}$. Moreover, the matrix of $A_i \in \mathcal{A}$ over V_j is $A_i^{(j)}$.

Lemma 8.9 Let $\mathcal{A}$ be indecomposable. If a matrix P commutes all matrices of $\mathcal{A}$, then P is either invertible or nilpotent.

— *Proof.* This is also an application of Fitting Lemma. Denote

$$V_\circ = \bigcup \ker P^i \quad \text{and} \quad V_\bullet = \bigcap \operatorname{im} P^i.$$

In other words, $V_\circ$ is the generalized eigenspace of 0 and $V_\bullet$ is the direct sum of the rest of generalized eigenspaces. We have $\mathbb{C}^n \cong V_\circ \oplus V_\bullet$. It is very easy to see from the definition that V_x is $\mathcal{A}$-invariant for $x = \circ, \bullet$. Since we assume $\mathcal{A}$ is indecomposable, we have $V_\bullet = \mathbb{C}^n$ or $V_\circ = \mathbb{C}^n$. Correspondently P is invertible or nilpotent. □

Actually, from the proof, if we work over $\mathbb{C}$, we can conclude that P has the same eigenvalues.

We will use the following elementary exercise.

▶ **Exercise 8.10** Assume $P_1, \ldots, P_d$ satisfy the condition in Lemma 8.9. If

$$P = P_1 + \cdots + P_d$$

is invertible, then one of P_i is invertible. ◀ P38

Theorem 8.11 Assume

$$\mathcal{A}^{(1)} \oplus \cdots \oplus \mathcal{A}^{(a)} \cong \mathcal{B}^{(1)} \oplus \cdots \oplus \mathcal{B}^{(b)},$$

with each $\mathcal{A}^{(i)}$ and $\mathcal{B}^{(i)}$ indecomposable. Then $a = b$ and $\mathcal{A}_i \cong \mathcal{B}_{\sigma(i)}$ for a permutation σ.

— *Proof.* By definition of $\cong$, there exists $P = (P_{ij})$ with its inverse $Q = (Q_{ij})$ such that

$$\begin{array}{ll} P_{ij}A^{(j)}_* = B^{(i)}P_{ij} & A^{(j)}_* Q_{ji} = Q_{ij}B^{(i)} \\ \sum_j P_{ij}Q_{jk} = \delta_{ik}\mathbf{1} & \sum_j Q_{ij}P_{jk} = \delta_{ik}\mathbf{1}. \end{array}$$

Note that

$$P_{11}Q_{11}, \ldots, P_{1a}Q_{a1}$$

commute with all matrices of $\mathcal{B}^{(1)}$ with sum to be $\mathbf{1}$. Thus one of them is invertible. By reordering, we can assume $P_{11}Q_{11}$ to be invertible. Note that $P_{11}A^{(1)}_* = B^{(1)}_* P_{11}$ which implies $\mathcal{A}^{(1)} \cong \mathcal{B}^{(1)}$. By conjugating by $\left(\begin{smallmatrix} P_{11}^{-1} & \\ & \mathbf{1} \end{smallmatrix}\right)$, we can assume $P = \left(\begin{smallmatrix} \mathbf{1} & X \\ Y & P' \end{smallmatrix}\right)$ and $A^{(1)}_* = B^{(1)}_*$. Denote

$$\begin{aligned} \mathcal{A}' = \mathcal{A}^{(2)} \oplus \cdots \oplus \mathcal{A}^{(a)} &= \{A'_* = \operatorname{diag}(A^{(2)}_*, \ldots A^{(a)}_*)\}_{*\in I}, \\ \mathcal{B}' = \mathcal{B}^{(2)} \oplus \cdots \oplus \mathcal{B}^{(b)} &= \{B'_* = \operatorname{diag}(B^{(2)}_*, \ldots B^{(b)}_*)\}_{*\in I}. \end{aligned}$$

Note that

$$(P' - YX)A'_* = B'_*P' - YB^{(1)}_* X = B'_*P' - YA^{(1)}_* X = B'_*(P' - YX).$$

Since $\left(\begin{smallmatrix} \mathbf{1} & X \\ Y & P' \end{smallmatrix}\right)\left(\begin{smallmatrix} \mathbf{1} & -X \\ & \mathbf{1} \end{smallmatrix}\right) = \left(\begin{smallmatrix} \mathbf{1} & \\ Y & P'-YX \end{smallmatrix}\right)$, the matrix $P' - YX$ is always invertible. We can conclude that $\mathcal{A}' \cong \mathcal{B}'$ and our theorem follows from induction. □

▶ **Problem 8.12** For two matrices A and B, recall that we call them **unitarily similar** if there exists a unitary matrix U such that $UAU^{\mathrm{h}} = B$.

Show that A is unitarily similar to B if and only if A and A^{h} similar to B and B^{h} simultaneously. ◀ P38

▶ **Problem 8.13** If $\left(\begin{smallmatrix} A & \\ & A \end{smallmatrix}\right)$ is unitarily similar to $\left(\begin{smallmatrix} B & \\ & B \end{smallmatrix}\right)$, show that A is unitarily similar to B.

▶ **Problem 8.14** For two real matrices A and B, recall that we call them **orthogonally similar** if there exists an orthogonal matrix U such that $UAU^{\mathrm{t}} = B$.

For two real matrices A and B, show that A is orthogonally similar to B if and only if A is unitarily similar to B.

Hints

8.2 Assume $A = PBP^{-1}$. Let $\mathbb{E}$ be a finite field extension of $\mathbb{F}$ containing entries of P. We pick a set of basis B for $\mathbb{E}/\mathbb{F}$. Assume $P = \sum_{\beta\in B} P_\beta \beta$. Now $AP = PB$ implies $AP_\beta = P_\beta B$. Now consider the polynomial in $\#B$ many variables $\{z_\beta\}_{\beta\in B}$,

$$P\big((z_\beta)_{\beta\in A}\big) = \sum_{\beta\in B} P_\beta z_\beta.$$

Now we have $AP(z_\beta) = P(z_\beta)B$. It suffices to find some choice of $z_\beta \in \mathbb{F}$ such that $P(z_\beta)$ is invertible. Since $\mathbb{F}$ is an infinite field, this is equivalent to requiring $\det(P(z_\beta)) \neq 0$ as a polynomial. Being zero does not depend on the base field. Thus this is clear since $\det(P(\beta)) = \det(P) \neq 0$.

8.4 By the previous problem, we can do so over $\mathbb{C}$ which we can use Jordan canonical forms. Note that Jordan blocks of $\begin{pmatrix} A & \\ & A \end{pmatrix}$ are just obtained by doubling those for A.

8.10 If $d = 1$, this is automatically true, so we assume $d \geq 2$. Note that it suffices to show when $P = \mathbf{1}$, since we can replace each P_i by $P^{-1}P_i$. To be exact, P_i is invertible if and only if so is $P^{-1}P_i$. Thus P_i is nilpotent if and only if so is $P^{-1}P_i$ by Lemma 8.9. In this case, if P_1 is not invertible, $P_2 + \cdots + P_d = \mathbf{1} - P_1$ is invertible. Thus the assertion follows from induction.

8.12 One direction is clear. For the converse, assume $PAP^{-1} = B$ and $PA^{\mathrm{h}}P^{-1} = B^{\mathrm{h}}$, i.e. $(P^{\mathrm{h}})^{-1}AP^{\mathrm{h}} = B$. We thus have $P^{\mathrm{h}}PA = AP^{\mathrm{h}}P$. Assume $P = UQ$ for a unitary matrix U and a positive definite matrix Q by polar decomposition. So we have $Q^2A = AQ^2$. Then $QA = AQ$, since Q^2 is a polynomial in Q. Now $B = PAP^{-1} = UAU^{-1}$.

9 Matrices Decompositions

In this section, we will make an attempt to summarize the matrices decompositions.

Gauss decomposition 9.1 For any invertible matrix A, we can write $A = PSQ$ with P lower triangular, Q upper triangular, and S a permutation matrix.

— *Proof.* Note that if $a \neq 0$,

$$\begin{pmatrix} 1 & \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & -b \\ & a \end{pmatrix}\begin{pmatrix} 1/a & \\ & 1/a \end{pmatrix} = \begin{pmatrix} 1 & \\ & ad-bc \end{pmatrix}.$$

Thus, if one entry of A is non-zero, we can without any loss of generality to assume that this entry is 1 and any entries lower than or right to it is zero. In particular, each column or row has exactly one 1, i.e. a permutation matrix. □

We see that for most (a dense set many) invertible matrices can be written as a product of a lower triangular matrix and an upper one.

▶ **ℙroblem 9.2(Bruhat decomposition)** For any invertible matrix A, we can write $A = PSQ$ with P and Q both upper triangular, and S a permutation matrix. ◀ **P42**

▶ **ℙroblem 9.3** Show that any matrix A can be written as $A = PSQ$ with P and Q upper triangular, and S a partial permutation matrix (i.e. at most one 1 in each column and row).

We remark that the Bruhat decomposition can be generalized to any reductive algebraic groups, see [1] where triangular matrices are generalized to be an element of a Borel subgroup, and permutation matrices are nothing but Weyl group elements.

▶ **ℙroblem 9.4** Any positively definite matrix A can be decompose into PP^{t} for an upper triangular matrix P.

Jordan decomposition 9.5 Any matrix A can be uniquely decomposed into $A = D + N$ with D diagonalizable, N nilpotent and $DN = ND$. Moreover, D and N are both polynomials in A.

— *Proof.* Firstly, let us show this over $\mathbb{C}$.

Over each root space V, we define the action of D to be the scalar of the corresponding eigenvalue and $N = A - D$. It is clear that D and N are polynomials in A over V. By the Chinese remainder theorem, they are still polynomials in A globally. This shows the existence of decomposition.

Assume we have any other decomposition $A = D' + N'$. Since we have constructed D and N which are polynomials in A, we see D', N', D and N commute each other. In particular, $D - D'$ is still diagonalizable and $N - N'$ is still nilpotent. But we require $D - D' = N' - N$, which forces both sides to be zero. This is the proof of uniqueness.

For a general subfield $\mathbb{F}$ of $\mathbb{C}$, we pick a Galois extension $\mathbb{E}$ the entries of D and N belong to. By the uniqueness, D and N are both invariant under any Galois group action which shows D and N are both $\mathbb{F}$-matrices. □

▶ **Jordan decomosition 9.6** Assume A is invertible, show that $A = DU$ with D diagonalizable, U unipotent ($\mathbf{1}$ + (nilpotent)) and $DU = UD$. Moreover, D and N are both polynomials in A. ◀ P42

QR decomposition 9.7 Any invertible real matrix A can be decomposed in to $A = QR$ with Q orthogonal and R upper triangular. Moreover, this decomposition is uniquely determined if we require the diagonal elements of R to be positive.

— *Proof.* This is a record of Gram–Schmidt orthogonalization. To be exact, assume $A = (a_1, \dots, a_n)$ for column vectors a_i's. Gram–Schmidt orthogonalization tells that there is a set of unit orthogonal basis $v_1, \dots, v_n$ such that for all i,

$$v_i \in \operatorname{span}(a_1, \dots, a_i).$$

In other words,

$$(a_1, \dots, a_n)R' = (v_1, \dots, v_n)$$

for an upper triangular matrix R'. Note that since $\operatorname{diag}(\pm 1, \dots, \pm 1)$ is orthogonal, we can always assume the diagonal entries R to be positive. This shows the existence.

Assume $A = QR = Q'R'$ for another such decomposition. We have $Q^{-1}Q' = R(R')^{-1}$ which is both orthogonal and upper triangular. But such matrices only be $\operatorname{diag}(\pm 1, \dots, \pm 1)$. Thus, if we require the signs over diagonal, both sides is $\mathbf{1}$. □

▶ **Problem 9.8** Any invertible complex matrix A can be decomposed in to $A = QR$ with Q unitary and R upper triangular. Moreover, this decomposition is uniquely determined if we require the diagonal elements of R to be positive.

This shows that we have a decomposition

$$\mathrm{GL}_n(\mathbb{R}) \cong O(n) \times \begin{pmatrix} \mathbb{R}_{>0} & \cdots & \mathbb{R} \\ & \ddots & \vdots \\ & & \mathbb{R}_{>0} \end{pmatrix} \quad \text{(as topological space)},$$

$$\mathrm{GL}_n(\mathbb{C}) \cong U(n) \times \begin{pmatrix} \mathbb{R}_{>0} & \cdots & \mathbb{C} \\ & \ddots & \vdots \\ & & \mathbb{R}_{>0} \end{pmatrix} \quad \text{(as topological space)},$$

which in particular shows that $\mathrm{GL}_n(\mathbb{R})$ and $O(n)$ (resp., $\mathrm{GL}_n(\mathbb{C})$ and $U(n)$) are of the same homotopy type. This can also be generalized to complex groups, known as **Iwasawa decomposition**.

▶ **Problem 9.9** Show that any matrix A (not necessarily invertible) can be decomposed into $A = QR$ with Q orthogonal and R upper triangular. ◀ P42

▶ **Problem 9.10** For a matrix $A = (a_{ij})$, show that

$$\det A \leq \prod_{i=1}^{n} \sqrt{|a_{i1}|^2 + \cdots + |a_{in}^2|}.$$

Singular Value Decomposition, SVD 9.11 For any real matrices A, we can decompose $A = P\Sigma Q$ with P and Q orthogonal and Σ diagonal with non-negative entries. Moreover, it is clear that the diagonal entries of D are square roots of eigenvalues of AA^{t} which are known as the **singular values** of A.

— *Proof.* Note that AA^{t} is positively definite. We can find an orthogonal P such that

$$AA^{\mathsf{t}} = P\Sigma^2 P^{\mathsf{t}}$$

for Σ a diagonal matrix whose diagonal entries are the square roots of eigenvalues of AA^{t}. If Σ is invertible, equivalently A is invertible, it suffices to show $Q = \Sigma^{-1}P^{\mathsf{t}}A$ is orthogonal. This is clear by observing that

$$QQ^{\mathsf{t}} = (\Sigma^{-1}P^{\mathsf{t}}A)(A^{\mathsf{t}}P\Sigma^{-1}) = \mathbf{1}.$$

As a result, we get the decomposition $A = P\Sigma Q$.

In the general case, we pick a sequence A_i of invertible matrices tending to A. Assume $A_i = P_i\Sigma_i Q_i$ the decomposition as above. Since the space of orthogonal matrices is compact, we can find a convergent sub-sequence for (P_i, Q_i). By replacing the sequence by the sub-sequence, we can assume (P_i, Q_i) has a limit (P, Q). Then $\Sigma = P^{\mathsf{t}}AQ^{\mathsf{t}}$ is the limit of Σ_i which is non-negative and diagonal. Hence we get the decomposition $A = P\Sigma Q$. □

▶ **Problem 9.12** For any complex matrices A, we can decompose $A = P\Sigma Q$ with P and Q unitary and Σ diagonal with non-negative entries.

▶ **Polar decomposition 9.13** For any real/complex matrices A, we can decompose $A = PO$ with P an orthogonal/unitary matrix and O an positively semi-definite matrix. ◀ P42

References

[1] J. E. Humphreys, Linear Algebraic Groups. Springer series of Graduate Texts in Mathematics. 978-1-4684-9445-7. (1975)

Hints

9.2 Denote $w_0 = \begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix}$, Note that $w_0 \begin{pmatrix} * & \cdots & * \\ & \ddots & \vdots \\ & & * \end{pmatrix} w_0 = \begin{pmatrix} * & & \\ \vdots & \ddots & \\ * & \cdots & * \end{pmatrix}$.

9.6 Assume $A = N + D = D(1 + D^{-1}N)$. Note that D^{-1} is a polynomial in D thus in A.

9.9 If A is not invertible, then we pick a sequence A_i of invertible matrices tending to A. Assume $A_i = Q_i R_i$ the QR decomposition. Since the set of orthogonal matrices is compact, we can find a convergent sub-sequence. By replacing the sequence with the sub-sequence, we can assume Q_i has a certain limit Q. Then $Q^{\mathsf{t}}A$ is the limit of R_i which is also upper triangular.

9.13 Assume $A = P\Sigma Q$, then $A = PQ(Q^{\mathsf{t}}\Sigma Q)$.

10 Three Subspaces

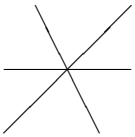
We have a well-known formula stating the relation of dimensions of intersection and sum, say for two subspaces X and Y, we have

$$\dim(X) + \dim(Y) = \dim(X \cap Y) + \dim(X + Y).$$

It was warned in a lot of contexts that the analogy is no longer true for three subspaces. That is, in general, for three subspaces X, Y, and Z

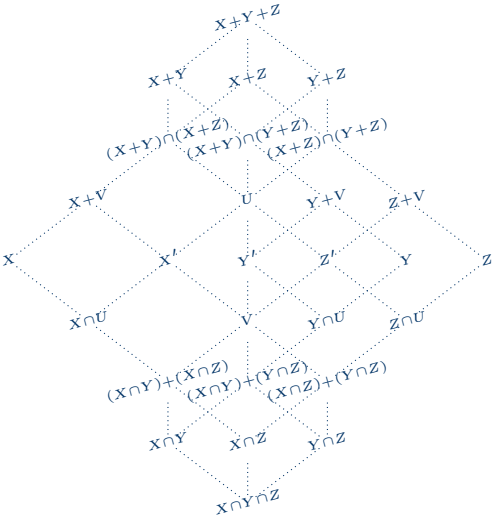
$$\begin{aligned}\dim(X+Y+Z) \neq\ & \dim(X) + \dim(Y) + \dim(Z) \\ & - \dim(X \cap Y) - \dim(Y \cap Z) - \dim(X \cap Z) \\ & + \dim(X \cap Y \cap Z).\end{aligned}$$

For example, consider three different lines (going through 0) in a plane.



Then $\dim(X + Y + Z) = 2$, but the right-hand-side is $3 + 0$. In fact, to state a correct formula for three spaces, we need to do a lot.

Theorem 10.1 For three subspaces X ,Y, and Z, there are 28 many possible different subspaces obtained by summing and intersecting among them.



$U=(X+Y)\cap(Y+Z)\cap(Z+X)$
$V=(X\cap Y)+(Y\cap Z)+(Z\cap X)$
$X'=V+X\cap U$
$Y'=V+Y\cap U$
$Z'=V+Z\cap U$

This diagram is known as the free modular lattice generated by 3 elements, see [1] for more information.

▶ **Problem 10.2** For three subspaces X, Y and Z, show that

$$\dim X+\dim Y+\dim Z = \dim(X+Y+Z)+\dim(X\cap Y\cap Z)+\frac{1}{2}(\dim U+\dim V),$$

where

$$U = (X+Y)\cap(Y+Z)\cap(Z+X)$$
$$V = (X\cap Y)+(Y\cap Z)+(Z\cap X)$$

are as in Theorem 10.1. ◀ P47

For example, if we have a base $\mathcal{B}$ such that X, Y and Z are spanned by $\mathcal{B}_X = \mathcal{B}\cap X$, $\mathcal{B}_Y = \mathcal{B}\cap Y$ and $\mathcal{B}_Z \cap Z$ respectively. Then $U = V$ is nothing but the space spanned by

$$(\mathcal{B}_X\cup\mathcal{B}_Y)\cap(\mathcal{B}_Y\cup\mathcal{B}_Z)\cap(\mathcal{B}_Z\cup\mathcal{B}_X) = (\mathcal{B}_X\cap\mathcal{B}_Y)\cup(\mathcal{B}_Y\cap\mathcal{B}_Z)\cup(\mathcal{B}_Z\cap\mathcal{B}_X).$$

For another example, consider three different lines in a plane as before. The subspace U is the whole 2-dimensional subspace, and V is zero. In particular,

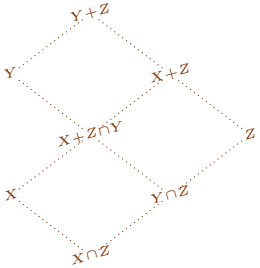
$$1+1+1 = 2+0+\frac{2}{2}.$$

We will see just after one more example below that these two examples above actually include all possible cases. In particular, knowing Theorem 10.5, problem 10.2 follows from those two examples.

▶ **Modular property 10.3** For three subsapces X, Y, Z, if $X\subseteq Y$, then $X+(Z\cap Y) = (X+Z)\cap Y$. Thus, it makes no confusion to write $X+Z\cap Y$, and

$$(X+Y)\cap(Y+Z)\cap(Z+X) = (X\cap Y)+(Y\cap Z)+(Z\cap Z) = X+Z\cap Y.$$

▶ **Problem 10.4** For three subspaces X, Y and Z with $X\subseteq Y$, show that there are 8 many possibly different subspaces obtained by sum and intersection among them.



In particular,

$$\dim X + \dim Y + \dim Z = \dim(Y + Z) + \dim(X \cap Z) + \dim(X + Z \cap Y).$$

In general, any two chains of subspaces admits finitely many possible subspaces obtained by sum and intersection. Moreover, the diagram, in the most general case, is the same as the lattice of Young diagrams inside a rectangle, which is a part of the Young lattice.

Now, we are in the position to state our main theorem on the structure of three subspaces.

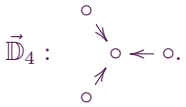
Theorem 10.5 For three subspaces X, Y and Z, we can find a base

$$\{v_1, \dots, v_d\} \cup \{x_1, \dots, x_r\} \cup \{y_1, \dots, y_r\}$$

such that

(i) X is spanned by $\{v_i \in X\} \cup \{x_i\}$;
(ii) Y is spanned by $\{v_i \in Y\} \cup \{y_i\}$;
(iii) Z is spanned by $\{v_i \in Z\} \cup \{x_i + y_i\}$.

— *Proof.* This can be proved by the quiver representation of



From the representation theory of quiver, the full list of indecomposable representations of the above quiver are

$$\underbrace{\begin{smallmatrix}1\\0\end{smallmatrix}00 \quad \begin{smallmatrix}0\\1\end{smallmatrix}00 \quad \begin{smallmatrix}0\\0\end{smallmatrix}01}_{\text{(I)}} \qquad \underbrace{\begin{smallmatrix}0\\0\end{smallmatrix}10 \quad \begin{smallmatrix}1\\0\end{smallmatrix}10 \quad \begin{smallmatrix}0\\1\end{smallmatrix}10 \quad \begin{smallmatrix}0\\0\end{smallmatrix}11 \quad \begin{smallmatrix}1\\1\end{smallmatrix}10 \quad \begin{smallmatrix}0\\1\end{smallmatrix}11 \quad \begin{smallmatrix}1\\0\end{smallmatrix}11 \quad \begin{smallmatrix}1\\1\end{smallmatrix}11}_{\text{(II)}} \qquad \underbrace{\begin{smallmatrix}1\\1\end{smallmatrix}21}_{\text{(III)}}$$

Now, we view three inclusions as a $\vec{\mathbb{D}}_4$-representation. It decomposes into indecomposable modules. Note that (I) will not appear since the maps are not injective.

The direct summand of (II) tells that there exists a base

$$\{v_1, \dots, v_d\}$$

such that X, Y or Z is spanned by the basis contained respectively.

The direct summand of (III) tells that there exists a base

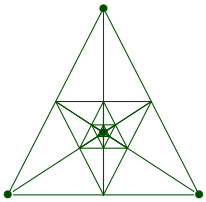
$$\{x_1, \dots, x_r\} \cup \{y_1, \dots, y_r\}$$

such that X is spanned by $\{x_i\}$, Y is spanned by $\{y_i\}$ and Z is spanned by $\{x_i + y_i\}$.

Combining the two parts together, we get the assertion in the theorem. □

Remark 10.6 We remark that the story about arbitrary subspaces as what we discussed in this section stops at the number 3. It is easy to construct four subspaces with infinitely many different subspaces from sum and intersection. For example,

$$\{x = y = z\}, \quad \{x = y = 0\}, \quad \{y = z = 0\}, \quad \{z = x = 0\}.$$



There would not able to get a "finite type" classification as in Theorem 10.5. Since, the space of four spaces of dimension $d_1, \cdots, d_4$ in an n-dimensional space has dimension

$$n(n - d_1) + \cdots + n(n - d_4)$$

which would achieve $\frac{n^2}{4} + \cdots + \frac{n^2}{4} = n^2$ for some n and $d_1, \ldots, d_4$. If there are only finitely many possibilities of structures, then there are only finite many GL_n-orbits, which in particular indicates the dimension of it is not more than $n^2 - 1$ (the action of scalar matrices in GL_n are trivial), a contradiction. We refer [2] for the more detailed classification.

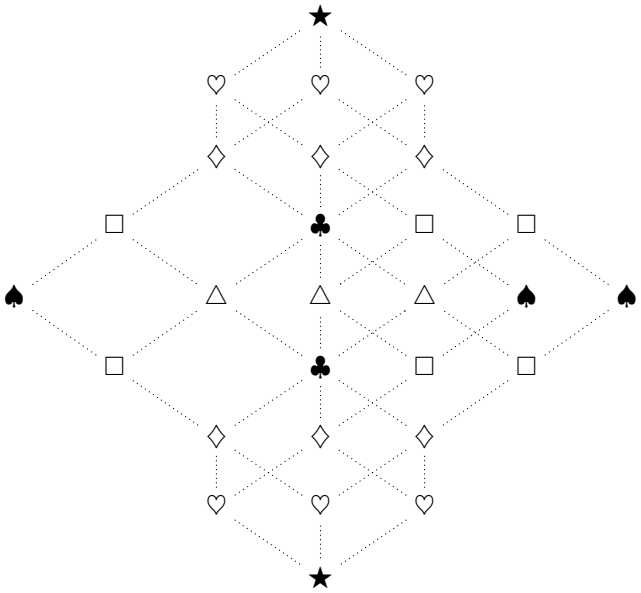
References

[1] George Grätzer. Lattice theory: foundation. Birkhäuser/Springer Basel AG, Basel, 2011.

[2] Magyar P, Weyman J, Zelevinsky A. Multiple flag varieties of finite type[J]. Advances in Mathematics, 1999, 141(1): 97-118.

Hints

10.2 Denote



We have

$$\left.\begin{aligned} \bigstar - \heartsuit + \diamondsuit - \clubsuit = 0 \\ 2\heartsuit = 3\bigstar + \diamondsuit \end{aligned}\right\} \Longrightarrow 2\heartsuit = \clubsuit + 2\bigstar \left.\begin{aligned} \\ \spadesuit = 2\heartsuit \end{aligned}\right\} \Longrightarrow \spadesuit = \clubsuit + 2\bigstar.$$

11 Substituting Matrices into Functions

For a square matrix A, we are able to evaluate it for any polynomial $f(x) = \sum a_i x^i$, i.e.

$$f(A) = \sum a_i A^i.$$

In this section, we will discuss to what extent we are allowed to substitute a matrix into a function f. We will restrict to $\mathbb{R}$ or $\mathbb{C}$ in order to motivate two ways of generalization of it to functional analysis, see Remark 11.8. We will lastly study the special case when f is exponential or logarithm.

A typical example is that it makes sense to denote $f(N)$ for a nilpotent matrix N and $f(x) = \frac{1}{1-x}$. The observation is

$$f(x) = \frac{1}{1-x} = 1 + x + x^2 + \cdots$$

in some (for example, topological) sense.

▶ **Problem 11.1** For a nilpotent matrix N, show that

$$(1-N)^{-1} = 1 + N + N^2 + \cdots.$$

Note that the sum is finite.

Aemma 11.2 For a finite set $\{\lambda_1, \dots, \lambda_n\} \subseteq \mathbb{R}$ or $\mathbb{C}$ and any function f which is n_i-differential at λ_i, there exists a polynomial p such that

$$f(z) = p(z) + o\big((z-\lambda)^{n_i}\big) \quad \text{when} \quad z \to \lambda_i.$$

— *Proof.* This lemma can be viewed as a generalization of Lagrangian interpolation with multiplicities. For any $0 \le r \le n_1$, since $(z-\lambda_1)^{n_1+1-r}$ is relative prime to $\prod_{i\neq 1}(z-\lambda_i)^{n_i+1}$, we can write

$$f(x)(z-\lambda_1)^{n_1+1-r} + g(x)\prod_{i\neq 1}(z-\lambda_i)^{n_i+1} = 1.$$

Denote

$$\chi_1^{(r)}(z) = g(z)\prod_{i\neq 1}(z-\lambda_i)^{n_i+1}(z-\lambda_i)^r.$$

We see for $i \neq 1$,

$$\chi_1^{(r)}(z) = o\big((z-z_i)^{n_i}\big) \quad \text{when} \quad z \to \lambda_i,$$

and

$$\chi_1^{(r)}(z) = (z-\lambda_1)^r + o\big((z-\lambda_1)^{n_1}\big) \text{when} \quad z \to \lambda_1.$$

We can similarly construct $\chi_i^{(r)}$. As a result,

$$p(z) = \sum_i \sum_{r=0}^{n_i} \frac{f_i^{(r)}(\lambda_i)}{r!} \chi_i^{(r)}(z)$$

serves. □

Construction 11.3 Let $f : \Omega \to \mathbb{C}$ for an open subset Ω of $\mathbb{C}$ (resp., $\mathbb{R}$) and A be an $n \times n$ complex (resp., real) matrix. Assume that

(1) all eigenvalues of A are contained in Ω;

(2) f is $(n(\lambda)-1)$-differentiable at eigenvalue λ of A, where $n(\lambda)$ is the multiplicity of $(z-\lambda)$ of minimal polynomial of A.

Then we define

$$f(A) = p(A)$$

where $p(z)$ is any polynomial such that

$$f(z) = p(z) + o\big((z-\lambda)^{n(\lambda)-1}\big) \quad \text{when} \quad z \to \lambda$$

for any eigenvalues λ of A.

▶ **Problem 11.4** Show this construction is well-defined, i.e. $p(A)$ does not depend on the choice of $p(z)$. ◀ P55

▶ **Spectral Mapping Theorem 11.5** Show that the eigenvalues of $f(A)$ is the image of eigenvalues of A under f.

Example 11.6 For example, we can evaluate $f(D)$ when D is diagonalizable. Assume

$$D = P\,\mathrm{diag}(d_1, \ldots, d_n)P^{-1}.$$

Then

$$f(D) = P\,\mathrm{diag}(f(d_1), \ldots, f(d_n))P^{-1}.$$

Example 11.7 For example, we can evaluate $f(J)$ when J is a Jordan block

$$J = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & \ddots & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}_{n\times n},$$

where $f(z)$ is differentiable $(n-1)$-differentiable. It is not hard to see

$$f(J) = \begin{pmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2} & \cdots & \cdots \\ & \ddots & \ddots & \ddots & \vdots \\ & & f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2} \\ & & & f(\lambda) & f'(\lambda) \\ & & & & \lambda \end{pmatrix}_{n\times n}$$

Remark 11.8 We remark that there are two great generalizations in functional analysis of what we discussed in the section. Both of them require more restrictions on functions or operators.

- we can substitute **bounded** linear operators A over a Banach space into any **analytic function** f defined over an open subset of $\mathbb{C}$ containing the spectrum of A. The definition uses Cauchy integral

$$f(A) = \oint \frac{f(z)dz}{z-A},$$

  where the integral is over any curve around the spectrum of A. See [1, §3.30].

- We can substitute **normal** bounded linear operators A over a Hilbert space into any **continuous** function $f : \mathbb{C} \to \mathbb{C}$. The definition is

$$f(A) = \int f(z)dE(z)$$

  where E is an operator-valued measure over the spectrum of A. See [1, §12.24].

Both of them reduce to what we defined above when the underlying vector space is of finite dimension.

▶ **ℙroblem 11.9** Show that a matrix A is normal if and only if its conjugate transposition A^h is a polynomial of A. ◀ **P55**

▶ **ℙroblem 11.10** For an ∞-differentiable function f, show that

$$d\big(f(A)\big) = f'(A)dA.$$

◀ **P55**

As an application, we will use the construction in this section to solve the matrix equation

$$f(X) = A,$$

where f is function and A is a given **complex** matrix. The following theorem is necessary to establish.

Theorem 11.11 For two infinitely differentiable functions f and g,

$$g(f(A)) = (g \circ f)(A)$$

holds whenever both sides make sense.

— *Proof.* Assume

$$\begin{aligned} f(z) &= p(z) + o\big((z-\lambda)^n\big) \quad \text{when} \quad z \to \lambda \\ g(w) &= q(w) + o\big((w - f(\lambda))^n\big) \quad \text{when} \quad z \to f(\lambda) \end{aligned}$$

for any eigenvalues λ of A. We will show that

$$g(f(z)) = q(p(z)) + o\big((z-\lambda)^n\big) \quad \text{when} \quad z \to \lambda.$$

Actually, by picking a non-constant p if necessary, we have

$$\begin{aligned} \left|\frac{g(f(z)) - q(p(z))}{(z-\lambda)^n}\right| &\le \left|\frac{g(p(z)) - q(p(z))}{(z-\lambda)^n}\right| + \left|\frac{g(f(z)) - g(p(z))}{(z-\lambda)^n}\right| \\ &= \left|\frac{g(p(z)) - q(p(z))}{(p(z)-p(\lambda))^n}\right| \cdot \left|\frac{p(z)-p(\lambda)}{z-\lambda}\right|^n \\ &\quad + \text{constant} \cdot \frac{o\big((z-\lambda)^n\big)}{(z-\lambda)^n} \end{aligned}$$

which tends to zero. □

▶ **Problem 11.12** For any r, if A is nonsingular, then there exists a matrix B with $B^n = A$.

▶ **Problem 11.13** Show that there is no matrix A such that $A^2 = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$.

Theorem 11.14 For a infinitely differentiable function f, the equation

$$f(X) = A$$

has a solution if for any eigenvalue a of A, there exists an x with $f(x) = a$ such that $f'(x) \neq 0$.

Proof. From our construction, it suffices to find an infinitely differentiable function g defined near eigenvalues of A such that $f \circ g$ is identity. So the problem finally reduces to local calculus. By implication theorem, we can find a differential g near a such that

$$g'(z) = \frac{1}{f'(g(z))}.$$

By induction, $g^{(n)}(z)$ is infinitely differentiable. As a result, $X = g(A)$ is a solution. □

▶ **Problem 11.15** Assume we have two ∞-differentiable functions f and g, show that

$$d\big(gf(A)\big) = g'(f(A))f'(A)dA.$$

Next, we will study the matrix exponentials. That is, we will substitute $e^A = \exp A$ for a square matrix A. Recall that we have the following expansion

$$e^z = \exp z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots$$

Note that we can take derivative item-wise for the above series, so that

$$e^A = \exp A = 1 + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots$$

under the index-wise limit. This can also be obtained by introducing a norm of matrices. We will list some properties of matrix exponentials by exercise.

▶ **Problem 11.16** For any $t_1, t_2 \in \mathbb{C}$, we have

$$\exp(t_1 A)\exp(t_2 A) = \exp((t_1 + t_2)A).$$

In particular, e^A is always invertible, and its inverse is e^{-A}.

▶ **Problem 11.17** Any invertible matrix can be written as e^A for some complex matrix A, i.e. the map

$$\exp : \mathfrak{gl}_n(\mathbb{C}) \longrightarrow \mathrm{GL}_n(\mathbb{C})$$

is surjective. Recall $\mathfrak{gl}_n(\mathbb{C})$ is another notation for matrix space $\mathbb{M}_n(\mathbb{C})$. ◀ P55

▶ **Problem 11.18** We have

$$\frac{d}{dt}e^{tA} = e^{tA}A.$$

In particular, $\frac{d}{dt}e^{tA}\big|_{t=0} = A$. Actually, by theory of ordinary differential equation, $X(t) = e^{tA}$ is the unique solution for

$$\frac{dX}{dt} = XA,$$

such that $X(0) = \mathbf{1}_n$. ◀ P55

▶ **Problem 11.19** Solve differential equation

$$\frac{dX}{dt} = AX.$$

▶ **Problem 11.20** We have

$$\det e^A = e^{\operatorname{tr} A}.$$

◀ P55

Next theorem explains the significance of **Lie bracket**

$$[A, B] = AB - BA.$$

To explain the notation, we need a well-defined **logarithm**. We firstly remark that the logarithm is NOT a well-defined function over $\mathbb{C} \setminus 0$, thus it makes no sense to denote $\log A$ in general. But log can be defined over the open disk $D = \{z \in \mathbb{C} : |z - 1| < 1\}$. Moreover, we have

$$\log(1 + z) = z - \frac{z^2}{2} + \frac{z^3}{3} + \cdots.$$

In this case, it is well-defined to define $\log A$ if the eigenvalues of A are sufficiently close to 1.

Theorem 11.21 When $|t| > 0$ is sufficiently small, the eigenvalues of $e^{tA}e^{sB}$ all lie in D, and

$$\log(e^{tA}e^{tB}) = t(A + B) + \frac{t^2}{2}[A, B] + o(t^2).$$

Proof. Note that the function sending the maximum distance of eigenvalues from 1 is a continuous function (Theorem 5.8). Thus when $|t|$ is sufficiently

small, all eigenvalues of $e^{tA}e^{tB}$ will be sufficiently close to 1. Now,

$$\begin{aligned}\log(e^{tA}e^{tB}) &= (e^{tA}e^{tB}-1) - \tfrac{(e^{tA}e^{tB}-1)^2}{2} + o(t^2)\\ &= \big((1+tA+\tfrac{t^2}{2}A^2)(1+tB+\tfrac{t^2}{2}B^2)-1\big) - \tfrac{\big((1+tA)(1+tB)-1\big)^2}{2} + o(t^2)\\ &= t(A+B) + t^2AB + \tfrac{t^2}{2}(A^2+B^2) - t^2\tfrac{(A+B)^2}{2} + o(t^2)\\ &= t(A+B) + \frac{t^2}{2}[A,B] + o(t^2).\end{aligned}$$

□

Remark 11.22 Actually, if we expand more, we will find

$$\log(e^{tA}e^{tB}) = t(A+B) + \frac{t^2}{2}[A,B] + \frac{t^3}{12}[A,[A,B]] - \frac{t^3}{12}[B,[A,B]] + o(t^3).$$

The famous **Baker–Campbell–Hausdorff theorem** tells that the coefficients are sum of iterated Lie bracket of A and B. In particular, the multiplication structure of $\mathrm{GL}_n(\mathbb{C})$ is recorded by the Lie bracket over $\mathfrak{gl}_n(\mathbb{C})$. We refer [1] for an elementary introduction to this topic.

▶ **Problem 11.23** Show that actually

$$\log(e^{tA}e^{sB}) = tA + sB + \frac{ts}{2}[A,B] + o(|s|^2+|t|^2).$$

▶ **Problem 11.24** Show that

$$\lim_{t\to 0}\frac{e^{tA}e^{tB}e^{-tB}e^{-tA}-1}{t^2} = [A,B].$$

Another explanation of the Lie bracket is the following.

▶ **Problem 11.25** Show that

$$\frac{d}{dt}e^{tA}Xe^{-tA}\Big|_{t=0} = [A,X].$$

◀ P55

References

[1] W. Rudin, 1991. Functional analysis, Mcgrawhill. Inc, New York, 45.

[2] B. C. Hall, Lie groups, Lie algebras, and representations. Springer, New York, NY.

Hints

11.4 It suffices to show when

$$p(z) = o\big((z-\lambda)^{n(\lambda_i)-1}\big) \quad \text{when} \quad z \to \lambda_i$$

then $p(A) = 0$. Actually, $\prod (z-\lambda)^{n(\lambda_i)}$ divides p.

11.9 Since A can be unitarily diagonalized when A is normal, the result follows directly from Lagrangian interpolation.

11.10 It suffices to show when f is a monomial. Thus everything reduces to Leibniz rule $d(AB) = A(dB) + (dA)B$.

11.17 Since $\frac{d}{dx}e^x$ does not vanish.

11.18 $de^{tA} = e^{tA}d(tA) = e^{tA}Adt$.

11.20 This is obviously true for the diagonal matrix. The general case follows from the Jordan decomposition or perturbation argument.

11.25 $e^{tA}Xe^{-tA} = (1+tA)X(1-tA) + o(t) = t(AX - XA) + o(t)$.

12 Topology of Classical Groups

In this section, we will list the basic topological properties of classical Groups. Here we list some of them.

$$\begin{aligned}
\mathrm{GL}_n(\mathbb{R}) &= \{\text{invertible real matrices}\},\\
\mathrm{GL}_n(\mathbb{C}) &= \{\text{invertible complex matrices}\},\\
\mathrm{SL}_n(\mathbb{R}) &= \{A \in \mathrm{GL}_n(\mathbb{R}) : \det A = 1\},\\
\mathrm{SL}_n(\mathbb{C}) &= \{A \in \mathrm{GL}_n(\mathbb{C}) : \det A = 1\},\\
\mathrm{O}(n) &= \{A \in \mathrm{GL}_n(\mathbb{R}) : A \cdot A^{\mathtt{t}} = A^{\mathtt{t}} \cdot A = \mathbf{1}_n\},\\
\mathrm{U}(n) &= \{A \in \mathrm{GL}_n(\mathbb{C}) : A \cdot A^{\mathtt{h}} = A^{\mathtt{h}} \cdot A = \mathbf{1}_n\},\\
\mathrm{SO}(n) &= \{A \in \mathrm{O}(n) : \det A = 1\},\\
\mathrm{SU}(n) &= \{A \in \mathrm{U}(n) : \det A = 1\},
\end{aligned}$$

We do not mention symplectic groups, but the computation is similar.

Proposition 12.1 The group $\mathrm{GL}_n(\mathbb{R})$ has two connected components,

$$\mathrm{GL}_n^{\pm}(\mathbb{R}) = \{A \in \mathrm{GL}_n(\mathbb{R}) : \pm \det A > 0\}.$$

— *Proof.* It suffices to show $\mathrm{GL}_n^{+}(\mathbb{R})$ is path connected. Let us denote the path-connected component of identity $\mathbf{1}_n$ by P_n. It suffices to show $P_n = \mathrm{GL}_n^{+}(\mathbb{R})$. We first list some basic properties of P_n.

(1) If $A_1 \in P_n$ and $A_2 \in P_m$, then $\begin{pmatrix} A_1 & \\ & A_2 \end{pmatrix} \in P_{m+n}$.

(2) If $A, B \in P_n$, then $A^{-1}, AB \in P_n$, i.e. P_n is a subgroup.

(3) If $A \in P_n$, then $QAQ^{-1} \in P_n$ for any $Q \in \mathrm{GL}_n(\mathbb{R})$, i.e. P_n is a normal subgroup.

Note that any orthogonal matrix A with $\det A = 1$ is contained in P_n. To be exact, by (1) and (3), it suffices to check

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in P_2,$$

which is trivial. On the other hand, upper triangular matrices with positive diagonal entries, i.e.

$$A = \begin{pmatrix} \mathbb{R}_{>0} & \cdots & \mathbb{R} \\ & \ddots & \vdots \\ & & \mathbb{R}_{>0} \end{pmatrix} \in P_n.$$

By (2), the general case follows from QR decomposition Theorem 9.7. □

▶ **Problem 12.2** Show that $\mathrm{O}(n)$ has two connected components, $\mathrm{SO}(n)$ and $\mathrm{SO}^{-}(n) = \{A \in \mathrm{O}(n) : \det A = -1\}$.

▶ **Problem 12.3** The following groups are connected

$$\mathrm{SL}_n(\mathbb{R}), \qquad \mathrm{SL}_n(\mathbb{C}), \qquad \mathrm{GL}_n(\mathbb{C}), \qquad \mathrm{U}(n), \qquad \mathrm{SO}(n), \qquad \mathrm{U}(n).$$

We remark that, using the surjectivity of exponential map $\exp : \mathfrak{gl}_n(\mathbb{C}) \to \mathrm{GL}_n(\mathbb{C})$, we also conclude the connectedness of $\mathrm{GL}_n(\mathbb{C})$, see Problem 11.17.

The above discussion about connected components of G is essentially the computation of $\pi_0(G, 1)$ in terms of topology. Our next step is to understand $\pi_1(G, 1)$ the **fundamental group**. Before doing computation, we present here a famous result saying $\pi_1(G, 1)$ is always abelian.

Proposition 12.4 For any topological group G, its fundamental group $\pi_1(G, 1)$ is always abelian.

— *Proof.* Denote the usual product of $\pi_1(G, 1)$ by $\circ$. The product $G \times G \to G$ induces another group structure over $\pi_1(G, 1)$ which we denoted by $\diamond$. Note that $\diamond$ is $\circ$-homomorphism, i.e.

$$(x \circ y) \diamond (z \circ w) = (x \diamond z) \circ (y \diamond w) =: \begin{bmatrix} x \circ y \\ \diamond \quad \diamond \\ z \circ w \end{bmatrix}.$$

Note that the constant path 1 is the identity not only for $\diamond$ but also for $\circ$. Thus we can apply the **Eckmann–Hilton argument**

$$x \circ y = \begin{bmatrix} x \circ y \\ \diamond \quad \diamond \\ 1 \circ 1 \end{bmatrix} = \begin{bmatrix} 1 \circ y \\ \diamond \quad \diamond \\ x \circ 1 \end{bmatrix} = \begin{bmatrix} y \circ 1 \\ \diamond \quad \diamond \\ 1 \circ x \end{bmatrix} = \begin{bmatrix} y \circ x \\ \diamond \quad \diamond \\ 1 \circ 1 \end{bmatrix} = y \circ x.$$

This shows commutativity. □

▶ **Problem 12.5** Show that $\circ = \diamond$. ◀ P60

The understanding of fundamental groups of mentioned groups starts from the case of small n. For example, it is easy to see

$$\mathrm{U}(1) = \mathrm{SO}(2) = S^1 = \text{ the unit circle in } \mathbb{R}^2.$$

Thus

$$\pi_1(\mathrm{U}(1)) = \pi_1(\mathrm{SO}(2)) = \mathbb{Z}.$$

The next lemma concerns a little bit big n, which turns out to be important when computing the general cases.

Aemma 12.6 We have

$$\mathrm{SU}(2) \cong S^3 = \text{ 3-dimensional sphere in } \mathbb{R}^4$$

$$\mathrm{SO}(3) \cong \mathbb{R}P^3 = \text{ 3-dimensional real projective space}$$

— Proof. Note that $\mathrm{SU}(2)$ is homeomorphic to the space of unitary vectors in $\mathbb{C}^2$, i.e. the unit ball of $\mathbb{R}^4$.

Note that any $A \in \mathrm{SO}(3)$ is a rotation. Denote

$$\pi : S^2 \times [0, \pi] \longrightarrow \mathrm{SO}(3)$$

by sending (u, θ) to the rotation of angle θ with axis $u\mathbb{R}$. Note that all possible collision of (u, θ) is

(1) for any $u, v \in S^2$, $R(u, 0) = R(v, 0)$;
(2) for any $u \in S^2$, $R(u, \pi) = R(-u, \pi)$.

By (1), we can glue

$$(u, 0) \sim (v, 0) \qquad \text{for all } u, v \in S^2,$$

the resulting quotient space is a solid ball D^3 with θ parametrized by radium. By (2), we can glue

$$(u, \pi) \sim (u, -\pi) \qquad \text{for all } u \in S^2,$$

the resulting quotient space is $\mathbb{R}P^3$ by definition. By above discussion, π induces a continuous bijection

$$\hat{\pi} : \mathbb{R}P^3 \longrightarrow \mathrm{SO}(3),$$

which must be a homeomorphism since both of them are Hausdorff compact. □

Remark 12.7 Actually, one can show that $\mathrm{SU}(2)/\{\pm 1\} \cong \mathrm{SO}(3)$, see [1].

As a result,

$$\pi_1(\mathrm{SU}(2)) = \pi_1(S^3) = 0, \qquad \pi_1(\mathrm{SO}(3)) = \pi_1(\mathbb{R}P^3) = \mathbb{Z}/2\mathbb{Z}.$$

In general, we can compute fundamental groups for all mentioned groups.

Theorem 12.8 We have

$$\pi_1(\mathrm{GL}_n(\mathbb{C})) = \pi_1(\mathrm{U}(n)) = \mathbb{Z},$$
$$\pi_1(\mathrm{SL}_n(\mathbb{C})) = \pi_1(\mathrm{SU}(n)) = 0,$$
$$\pi_1(\mathrm{GL}_n^+(\mathbb{R})) = \pi_1(\mathrm{SL}_n(\mathbb{R})) = \pi_1(\mathrm{SO}(n)) = \begin{cases} 0, & n = 1, \\ \mathbb{Z}, & n = 2, \\ \mathbb{Z}/2, & n \geq 3. \end{cases}$$

— *Proof.* By QR decomposition, the groups in each row are homotopy equivalent, thus it suffices to compute any one of them. We first show the last case as an example. We did the computation for $n = 2, 3$ above, and it is trivial when $n = 1$, so we assume $n > 3$. The natural identification $\mathrm{SO}(n-1) \subseteq \mathrm{SO}(n)$ gives a long exact sequence

$$\pi_2(\tfrac{\mathrm{SO}(n)}{\mathrm{SO}(n-1)}) \to \pi_1(\mathrm{SO}(n-1)) \to \pi_1(\mathrm{SO}(n)) \to \pi_1(\tfrac{\mathrm{SO}(n)}{\mathrm{SO}(n-1)}) \to 0.$$

We remark that $\mathrm{SO}(n)/\mathrm{SO}(n-1)$ is nothing but the $(n-1)$-dimensional sphere S^{n-1} in $\mathbb{R}^n$. To be exact, $\mathrm{SO}(n)$ acts transitively on S^{n-1} with the stablizer of $(0, \dots, 0, 1)$ isomorphic to $\mathrm{SO}(n-1)$. As a result, when $n > 3$, $\pi_2(S^{n-1}) = \pi_1(S^{n-1}) = 0$. This shows $\pi_1(\mathrm{SO}(n)) = \pi_1(\mathrm{SO}(n-1))$ for $n > 3$.

By a similar manner, we can show

$$\pi_1(\mathrm{SU}(n)) = \pi_1(\mathrm{SU}(n-1)), \qquad \pi_1(\mathrm{U}(n)) = \pi_1(\mathrm{U}(n-1))$$

for $2n > 3$ i.e. $n \geq 2$. □

▶ **Problem 12.9** Show $\pi_1(\mathrm{GL}_n(\mathbb{C})) = \mathbb{Z}$ from $\pi_1(\mathrm{SL}_n(\mathbb{C})) = 0$ directly. ◀ P60

▶ **Problem 12.10** Compute the universal cover of $\mathrm{GL}_n(\mathbb{C})$. ◀ P60

Remark 12.11 The universal covering of $\mathrm{SO}(n)$ is known as the spin group. To construct them, we need **Clifford algebra**. We refer readers to [2, §I.6]. The fundamental groups of a compact group can be computed in terms of root systems, see [2, §V.7].

References

[1] B. C. Hall, Lie groups, Lie algebras, and representations. Springer, New York, NY.

[2] T. Bröcker, T.Tom Dieck, Representations of compact Lie groups (Vol. 98). Springer Science & Business Media.

Hints

12.5 $x \circ y = \begin{bmatrix} x \circ y \\ \diamond \quad \diamond \\ 1 \circ 1 \end{bmatrix} = \begin{bmatrix} x \circ 1 \\ \diamond \quad \diamond \\ 1 \circ y \end{bmatrix} = \begin{bmatrix} 1 \circ x \\ \diamond \quad \diamond \\ 1 \circ y \end{bmatrix} = x \diamond y.$

12.9 Note that $\mathrm{GL}_n(\mathbb{C})/\mathrm{SL}_n(\mathbb{C}) = \mathbb{C}^\times$. We have a long exact sequence

$$\pi_1(\mathrm{SL}_n(\mathbb{C})) \to \pi_1(\mathrm{GL}_n(\mathbb{C})) \to \pi_1(\mathbb{C}^\times) \to 0.$$

Note that $\mathbb{C}^\times$ is homotopy equivalent to circle S^1, thus $\pi_1(\mathrm{GL}_n(\mathbb{C})) \cong \mathbb{Z}$.

12.10 It is $\mathbb{C} \times \mathrm{SL}_n(\mathbb{C})$, with covering map be $(z, A) \longmapsto e^{z/n} A$.

13 Counting Subspaces

In this section, we will compute the number of subspaces over a finite field $\mathbb{F}_{\mathbf{q}}$ where $\mathbf{q}$ is a power of a prime. There are two methods correspondent to addition principle and multiplication principle. The first method is to decompose (stratify) the space into disjoint pieces. The second method is to find a surjective map to with preimage (fibre) of each point isomorphic. As a result, our method works over any field except counting at the end.

We first introduce some notations.

Definition 13.1 We define **quantum number**

$$[n] = \frac{\mathbf{q}^n - 1}{\mathbf{q} - 1} = 1 + \mathbf{q} + \cdots + \mathbf{q}^{n-1}.$$

We define **quantum factorials**, and **quantum binomial coefficients**

$$[n]! = [n] \cdots [1], \qquad \begin{bmatrix} n \\ k \end{bmatrix} = \frac{[n]!}{[k]![n-k]!}.$$

The main object is Grassmannian, the set of k-subspaces of $\mathbb{F}_{\mathbf{q}}^n$.

Definition 13.2 For $0 \le k \le n$, denote **Grassmannian**

$$\mathrm{Gr}(k, n) = \{V \subseteq \mathbb{F}_q^n : \dim V = k\}.$$

Note that $\mathrm{Gr}(1, n) = \mathbb{P}^{n-1}$ is the projective space. We also denote **complete flag variety**

$$\mathrm{Fl}(n) = \{0 \subseteq V_1 \subseteq \cdots \subseteq V_{n-1} = \mathbb{F}_{\mathbf{q}}^n : \dim V_i = i\}.$$

▶ **Problem 13.3** Show that $\#\mathbb{P}^{n-1} = [n]$. ◀ P65

Theorem 13.4 We have

$$\#\,\mathrm{Fl}(n) = [n]!, \qquad \#\,\mathrm{Gr}(k, n) = \begin{bmatrix} n \\ k \end{bmatrix}.$$

— *Proof.* We have a natural map

$$\pi : \mathrm{Fl}(n) \longrightarrow \mathrm{Gr}(1, n) = \mathbb{P}^{n-1}$$

by sending a flag $(V_\bullet)$ to V_1. Note that π is surjective. For each $V \in \mathrm{Gr}(1, n)$, the fibre $\pi^{-1}(V)$ is in bijection to the set of complete flag of $\mathbb{F}_{\mathbf{q}}^n/V$, i.e. $\mathrm{Fl}(n-1)$ up to an isomorphism. As a result,

$$\#\,\mathrm{Fl}(n) = \#\mathbb{P}^{n-1} \times \#\,\mathrm{Fl}(n-1) = [n] \cdot \#\,\mathrm{Fl}(n-1).$$

By induction, we see $\#\operatorname{Fl}(n) = [n]!$.

Next, let us consider

$$\pi_k : \operatorname{Fl}(n) \longrightarrow \operatorname{Gr}(k, n)$$

by sending a flag $(V_\bullet)$ to V_k. Note that π_k is surjective with each fiber isomorphic to $\operatorname{Fl}(k) \times \operatorname{Fl}(n-k)$. Thus

$$[k]! \cdot [n-k]! \times \#\operatorname{Gr}(k, n) = [n]!.$$

This is exactly what we asserted. □

▶ **Problem 13.5** For $\mathbf{k} = (0 < k_1 < \cdots < k_m < n)$, compute the number of **partial flag variety**

$$\operatorname{Fl}(\mathbf{k}, n) = \{0 \subseteq V_1 \subseteq \cdots \subseteq V_{n-1} = \mathbb{F}_{\mathbf{q}}^n : \dim V_i = k_i\}$$

equals to so-called **quantum multinomial coefficient**

$$\frac{[n]!}{[k_1]![k_2 - k_1]! \cdots [n - k_m]!}.$$

Our next problem is to compute the number of subspaces (U, V) with the following dimension condition

$$\dim \begin{bmatrix} & U+V & \\ U & & V \\ & U \cap V & \end{bmatrix} = \begin{bmatrix} & t & \\ u & & v \\ & s & \end{bmatrix}.$$

Note that $u + v = t + s$. Let us denote this set by $\operatorname{Gr}(\begin{smallmatrix} & t & \\ u & & v \\ & s & \end{smallmatrix}, n)$.

Example 13.6 For example, let us consider the case

$$\begin{bmatrix} & t & \\ u & & v \\ & s & \end{bmatrix} = \begin{bmatrix} & k{+}1 & \\ k & & 1 \\ & 0 & \end{bmatrix}.$$

In this case, for each line V, we need to avoid those U containing V. In other words, for a fixed V, the choice of U is $\operatorname{Gr}(k, n) \setminus X$ with $X \cong \#\operatorname{Gr}(k-1, n-1)$. As a result,

$$\#\operatorname{Gr}\left(\begin{smallmatrix} & k+1 & \\ k & & 1 \\ & 0 & \end{smallmatrix}, n\right) = [n] \cdot \left(\begin{bmatrix} n \\ k \end{bmatrix} - \begin{bmatrix} n-1 \\ k-1 \end{bmatrix} \right).$$

▶ **Problem 13.7** Show that

$$\#\operatorname{Gr}(u\begin{smallmatrix}t\\s\end{smallmatrix}v,n)=\begin{bmatrix}n\\t\end{bmatrix}\#\operatorname{Gr}(u\begin{smallmatrix}t\\s\end{smallmatrix}v,t)=\begin{bmatrix}n\\t\end{bmatrix}\begin{bmatrix}t\\s\end{bmatrix}\#\operatorname{Gr}(u\text{-}s\begin{smallmatrix}t-s\\0\end{smallmatrix}v\text{-}s,t-s).$$

▶ **Problem 13.8** Prove that

$$\#\operatorname{GL}_n(\mathbb{F}_{\mathbf{q}})=(\mathbf{q}^n-1)(\mathbf{q}^n-\mathbf{q})\cdots(\mathbf{q}^n-\mathbf{q}^{n-1})=(\mathbf{q}-1)^n\mathbf{q}^{n(n-1)/2}[n]!.$$

◀ P65

▶ **Problem 13.9** Prove that

$$\#\operatorname{Gr}(k\begin{smallmatrix}n\\0\end{smallmatrix}n\text{-}k,n)=\frac{\#\operatorname{GL}_n(\mathbb{F}_{\mathbf{q}})}{\#(\operatorname{GL}_k(\mathbb{F}_{\mathbf{q}})\times\operatorname{GL}_{n-k}(\mathbb{F}_{\mathbf{q}}))}=\mathbf{q}^{k(n-k)}\begin{bmatrix}n\\k\end{bmatrix}.$$

In other words, we are counting all possible direct sum decomposition of the given dimensions for $\mathbb{F}_{\mathbf{q}}^n$.

Example 13.10 Alternatively, it is easy to see

$$\#\operatorname{Gr}(k\begin{smallmatrix}n\\0\end{smallmatrix}n\text{-}k,n)=\begin{bmatrix}n\\k\end{bmatrix}\times\#X$$

where X is the set $U\in\operatorname{Gr}(n-k,n)$ such that $\mathbb{F}_{\mathbf{q}}^n=U\oplus V$ for a **given** V of dimension k. For each $U_0\in X$, we claim

$$X\cong\operatorname{Hom}(U_0,V).$$

To be exact, any $U\in X$ is the **graph** of some linear map $\varphi:U_0\to V$, i.e.

$$U=\{u\oplus\varphi(u):u\in U_0\}\subseteq U_0\oplus V=\mathbb{F}_{\mathbf{q}}^n.$$

That is, for each $u\in U_0$, $u+V$ intersects U exactly once. As a result,

$$\#\operatorname{Gr}(k\begin{smallmatrix}n\\0\end{smallmatrix}n-k,n)=\mathbf{q}^{k(n-k)}\begin{bmatrix}n\\k\end{bmatrix}.$$

Theorem 13.11 We have

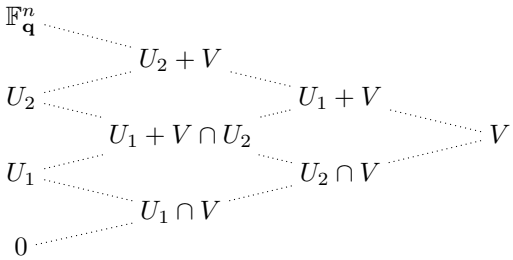
$$\#\operatorname{Gr}(u\begin{smallmatrix}t\\s\end{smallmatrix}v,n)=\mathbf{q}^{(u-s)(v-s)}\frac{[n]!}{[s]![u-s]![v-s]![n-t]!}.$$

Corollary 13.12 For a given u-dimensional space U, the number of v-dimensional spaces V with $\dim(V \cap U) = d$ is $\mathbf{q}^{(u-d)(v-d)} \begin{bmatrix} u \\ d \end{bmatrix} \begin{bmatrix} n-u \\ v-d \end{bmatrix}$. Actually, the fiber of

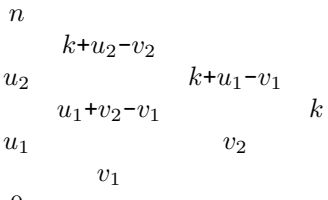
$$V \longmapsto (U \cap V, \tfrac{U+V}{U}) \in \mathrm{Gr}(d, u) \times \mathrm{Gr}(v-d, n-u)$$

has cardinality $\mathbf{q}^{(u-d)(v-d)}$.

Now, assume we have a 2-step flag $U_1 \subseteq U_2$, we can form the lattice for any subspace V



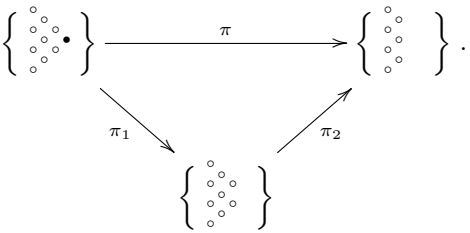
Assume $U_1 \subseteq U_2$ is given, of dimensions $u_1 \leq u_2$. We can compute the number of subspaces V such that $U_1 \cap V \subseteq U_2 \cap V$ have dimension $v_1 \leq v_2$. We can first predict the dimensions of all subspaces in the diagram, say,



Then we set

$$V \longmapsto (U_1 \cap V, \tfrac{U_1+V\cap U_2}{U_1}, \tfrac{U_2+V}{U_2}) \in \mathrm{Gr}(v_1, u_1) \times \mathrm{Gr}(v_2-v_1, u_2-u_1) \times \mathrm{Gr}(k-v_2, n-u_2).$$

The diagram is



We see the computation of fiber π_1 and π_2 can be reduced to Corollary 13.12. As a result, the final answer is

$$\mathbf{q}^{(u_1-v_1)(v_2-v_1)+(u_2-v_1)(k-v_2)} \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} \begin{bmatrix} u_2-u_1 \\ v_2-v_1 \end{bmatrix} \begin{bmatrix} n-u_2 \\ k-v_2 \end{bmatrix}.$$

In general, for any chain of subspaces, we can do a similar computation.

▶ **Problem 13.13** Let us counting the number of nilpotent matrices over a finite field. Let

$$N(n) = \{x \in \mathbb{M}_n(\mathbb{F}_\mathbf{q}) : x^n = 0\}.$$

Prove that

$$\begin{aligned} \#\mathbb{M}_n(\mathbb{F}_\mathbf{q}) &= \sum_{k=0}^{n} \#\operatorname{Gr}(k\,{}_0^{\,n}\text{-}k) \cdot \#\operatorname{GL}_{n-k}(\mathbb{F}_\mathbf{q}) \cdot N(k) \\ &= \sum_{k=0}^{n} \frac{\#\operatorname{GL}_n(\mathbb{F}_\mathbf{q})}{\#\operatorname{GL}_k(\mathbb{F}_\mathbf{q})} N(k) = N(n) + \sum_{k=1}^{n-1} \frac{\#\operatorname{GL}_n(\mathbb{F}_\mathbf{q})}{\#\operatorname{GL}_k(\mathbb{F}_\mathbf{q})} N(k) \\ &= N(n) + \frac{\#\operatorname{GL}_n(\mathbb{F}_\mathbf{q})}{\#\operatorname{GL}_{n-1}(\mathbb{F}_\mathbf{q})} \sum_{k=1}^{n-1} \frac{\#\operatorname{GL}_{n-1}(\mathbb{F}_\mathbf{q})}{\#\operatorname{GL}_k(\mathbb{F}_\mathbf{q})} N(k) \\ &= N(n) + \frac{\#\operatorname{GL}_n(\mathbb{F}_\mathbf{q})}{\#\operatorname{GL}_{n-1}(\mathbb{F}_\mathbf{q})} \#\mathbb{M}_{n-1}(\mathbb{F}_\mathbf{q}). \end{aligned}$$

As a result, $N(n) = \mathbf{q}^{n(n-1)}$.

Hints

13.3 Actually, we can decompose $\mathbb{P}^{n-1} = \mathbb{A}^0 \sqcup \cdots \sqcup \mathbb{A}^{n-1}$ where each $\mathbb{A}^i \cong \mathbb{F}_\mathbf{q}^i$ whose cardinaty is $\mathbf{q}^i$. Alternative, we see $\mathbb{P}^{n-1}$ is obtained from $\mathbb{F}_\mathbf{q}^n \backslash 0$ quotient by a free action of $\mathbb{F}_\mathbf{q}^\times$.

13.8 It is equivalent to count the number of linearly independent vectors of cardinality n. The number of linearly independent vectors of cardinality $k \le k$ is $(\mathbf{q}^n - 1)(\mathbf{q}^n - \mathbf{q}) \cdots (\mathbf{q}^n - \mathbf{q}^{k-1})$.

14 Commutating Matrices

We will describe all matrices commuting with a given matrix. We will only deal with nilpotent matrices and left the general cases to readers. See Problem 14.10.

We will use the following diagram (of length r)

$$0 \xleftarrow{A} \bullet \xleftarrow{A} \bullet \xleftarrow{A} \cdots \xleftarrow{A} \bullet \xleftarrow{A} \bullet$$

to represent a nilpotent matrix A, i.e. Jordan block of size r. We call the vector corresponding to the rightmost dot the **generator** of A, or it generates A. The theory of Jordan canonical forms tells that any nilpotent matrices are the direct sum of Jordan blocks. For a nilpotent matrix A, if A decomposes to Jordan blocks of sizes $\lambda_1 \geq \lambda_2 \geq \cdots$, we call λ is the **type** of A. For example, the following diagram

$$\begin{aligned}
&0 \xleftarrow{A} \bullet \xleftarrow{A} \bullet \xleftarrow{A} \bullet \xleftarrow{A} \bullet \\
&0 \xleftarrow{A} \bullet \xleftarrow{A} \bullet \xleftarrow{A} \bullet \xleftarrow{A} \bullet \\
&0 \xleftarrow{A} \bullet \xleftarrow{A} \bullet \\
&0 \xleftarrow{A} \bullet,
\end{aligned}$$

represents a nilpotent matrix of type $(4, 4, 2, 1)$.

Definition 14.1 For a partition λ, we denote

$$\begin{aligned}
G_\lambda &= \{X \in \mathrm{GL}(V) : XA = AX\}, \\
\mathfrak{g}_\lambda &= \{T \in \mathrm{End}(V) : TA = AT\},
\end{aligned}$$

for a nilpotent transform A over V of type λ, i.e. the automorphism group of V preserves A.

▶ **Problem 14.2** Let $A = J$ be single Jordan block (belonging to 0) of size r. Show that

$$\begin{aligned}
G_\lambda &= \{a_0\mathbf{1} + a_1 J + \cdots + a_{r-1}J^{r-1} : a_0 \neq 0\}, \\
\mathfrak{g}_\lambda &= \{a_0\mathbf{1} + a_1 J + \cdots + a_{r-1}J^{r-1}.
\end{aligned}$$

Or, in terms of matrix

$$G_\lambda = \left\{ \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_{r-1} \\ & \ddots & \ddots & \ddots & \vdots \\ & & a_0 & a_1 & a_2 \\ & & & a_0 & a_1 \\ & & & & a_0 \end{pmatrix} : a_0 \neq 0 \right\}.$$

In particular, $\dim G_\lambda = r$.

▶ **Problem 14.3** Note that type of $A = \begin{pmatrix} 0 & & \\ & 0 & 1 \\ & & 0 \end{pmatrix}$ is $(2,1)$. Prove

$$G_{(2,1)} = \left\{ \begin{pmatrix} a & 0 & b \\ c & d & e \\ 0 & 0 & d \end{pmatrix} : ad \neq 0 \right\} \qquad \mathfrak{g}_{(2,1)} = \left\{ \begin{pmatrix} a & 0 & b \\ c & d & e \\ 0 & 0 & d \end{pmatrix} \right\}.$$

In particular, $\dim G_{(2,1)} = 5$.

Let us fix a Jordan canonical form of A. For two coordinate subspaces $M \subseteq N$, we denote $N \ominus M$ to be the complement coordinate subspace. For example,

$$V_i' = \ker x^i \ominus x(\ker x^{i+1})$$

is space spanned by generators of Jordan block of size i. Note that $\dim V_i'$ is exactly the number of i's appearing in λ.

Example 14.4 Assume we are given a

$$\varphi = (\varphi_i) \in \prod \operatorname{Hom}(V_i', \ker A^i \ominus V_i').$$

Then φ defines $T_\varphi : V \to V$ by sending

$$A^j v \longmapsto A^j v + A^j \varphi(v)$$

for $v \in V_i'$. It is clear that $T_\varphi - 1$ commutes with A and is nilpotent. Actually any $X \in G_\lambda$ such that the submatrix for

$$V_1' \oplus V_2' \oplus \cdots \oplus V_i' \subseteq V$$

is identity admits a unique φ such that $X = T_\varphi$. We denote $\operatorname{rad} G_\lambda$ the subgroup of all transforms obtained in this way.

Example 14.5 Assume we are given

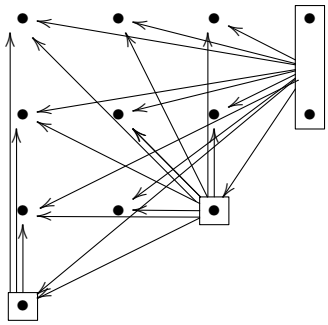
$$\sigma = (\sigma_i) \in \prod \operatorname{GL}(V_i')$$

Each σ defines $g_\sigma : V \to V$ by sending

$$A^j v \longrightarrow A^j \sigma_i(v)$$

for $v \in V_i'$. It is clear that g_σ commutes with A and is invertible. We denote L_λ the subgroup of all transforms obtained in this way.

We use the following diagram to illustrate them.



morphisms for arrows $\leftarrow \in \operatorname{rad} G_\lambda$

automorphisms inside frames $\square \in L_\lambda$

Theorem 14.6 We have a split short exact sequence

$$0 \longrightarrow \operatorname{rad} G_\lambda \longrightarrow G_\lambda \longrightarrow L_\lambda \longrightarrow 0.$$

— *Proof.* Note that $\ker A^i \ominus V_i'$ is A-invariant, thus the operator sending T to the submatrix of V_i is well-defined. By our discussion above, this is clear. □

▶ **Problem 14.7** Find the description for $\mathfrak{g}_\lambda = \{T \in \operatorname{End}(V) : TA = AT\}$. Actually, $\mathfrak{g}_\lambda$ is the Lie algebra of G_λ.

We are going to compute the dimension of G_λ.

To establish this theorem, we need more notations concerning partitions. Let us denote λ' the conjugation of λ, i.e. transpose. That is, λ_i' is the number of •'s in the i-th column. For example $(4,4,3,1)' = (4,3,3,2)$

Note that $\lambda_i' - \lambda_{i+1}'$ is the number of i's appearing in λ. A slightly less trivial computation is

▶ **Problem 14.8** Show that

$$n(\lambda) := \sum (i-1)\lambda_i = \sum \binom{\lambda_i'}{2}$$

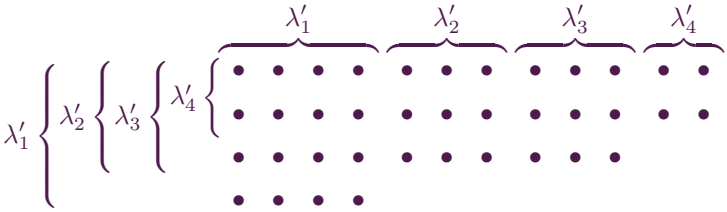
In other words, $\sum \lambda_i'^2 = |\lambda| + 2n(\lambda)$. ◀ P71

Theorem 14.9 We have $\dim G_\lambda = \dim \mathfrak{g}_\lambda = |\lambda| + 2n(\lambda)$.

— *Proof.* By Theorem 14.6,

$$\begin{aligned}\dim G_\lambda &= \dim \operatorname{rad} G_\lambda + \dim L_\lambda \\ &= \sum (\lambda'_i - \lambda'_{i+1})(\lambda'_1 + \cdots + \lambda'_{i-1} + \lambda'_{i+1}) + \sum (\lambda'_i - \lambda'_{i+1})^2 \\ &= \sum (\lambda'_i - \lambda'_{i+1})(\lambda'_1 + \cdots + \lambda'_i) \\ &= \sum \lambda_i'^2\end{aligned}$$

where the last equality uses summation by parts. For example,



Thus the result follows from above problem. □

See also [1, Lemma 2.8] or [2, II (1.6)].

▶ **Problem 14.10** Denote

$$\begin{aligned}G_A &= \{T \in \mathrm{GL}(V) : TA = AT\}, \\ \mathfrak{g}_A &= \{T \in \mathrm{End}(V) : TA = AT\},\end{aligned}$$

for any linear transform A. Assume A decomposes into $\operatorname{diag}(A_1, \ldots, A_d)$ with A_i belonging to different eigenvalues σ_i. Show that

$$\begin{aligned}G_A &= \operatorname{diag}(G_{A_1}, \ldots, G_{A_d}), \\ \mathfrak{g}_A &= \operatorname{diag}(\mathfrak{g}_{A_1}, \ldots, \mathfrak{g}_{A_d}).\end{aligned}$$

In particular,

$$\dim \mathfrak{g}_A = \sum |\lambda_i| + 2n(\lambda_i),$$

where λ_i is the type, i.e. the sizes of Jordan blocks belonging to σ_i.

▶ **Problem 14.11** Assume we have Jordan decomposition $A = D + N$ with D diagonalizable and N nilpotent. Show that

$$G_A = G_D \cap G_N, \qquad \mathfrak{g}_A = \mathfrak{g}_D \cap \mathfrak{g}_N.$$

▶ **Problem 14.12** Assume the coefficients of A are all in $\mathbb{F}$ a subfield of $\mathbb{C}$. We can form

$$\mathfrak{g}_A(\mathbb{F}) = \{T \in \mathbb{M}_n(\mathbb{F}) : TA = AT\}.$$

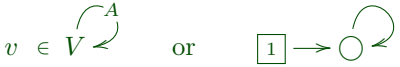
Show that

$$\dim_{\mathbb{F}} \mathfrak{g}_A(\mathbb{F}) = \dim_{\mathbb{C}} \mathfrak{g}_A(\mathbb{C}).$$

◀ P71

▶ **Problem 14.13** Show that any matrix B commutes with all elements of $\mathfrak{g}_A$ is a polynomial of A. ◀ P71

Remark 14.14 We can classify the pairs (A, x) for $A \in \mathbb{M}_n(\mathbb{C})$ and $x \in \mathbb{C}^n$. Geometrically, this problem is equivalent to the classification of the following data



Similarly, we will only deal with the case when A is nilpotent. Let A be a nilpotent matrix of type λ. We pick a Jordan canonical form of A, say,

$$V = \bigoplus_i \bigoplus_{j=1}^{m_i} (\mathbb{C}\mathbf{e}_i^{(j)} \oplus \cdots \oplus \mathbb{C}A^{i-1}\mathbf{e}_i^{(j)}),$$

with each $\mathbf{e}_i^{(j)} \in \ker A^i$, and m_i the number of i's appearing in λ. For any $v \in V$, there exists a $g \in G_\lambda$ such that

$$gv = A^{b_1}\mathbf{e}_{a_1+b_1}^{(1)} + \cdots + A^{b_d}\mathbf{e}_{a_d+b_d}^{(1)},$$

for two partitions

$$a : a_1 > a_2 > \cdots > a_d > 0, \qquad b : b_1 > b_2 > \cdots > b_d \geq 0,$$

such that $a_1 + b_1, \ldots, a_d + b_d$ represent different parts of λ. We illustrate this with an example.

$$\begin{array}{c}1\,2\,0\,0\\3\,4\,5\,0\\6\,7\\8\end{array} \xmapsto{L_\lambda} \begin{array}{c}?\,?\,1\,0\\?\,?\,0\,0\\6\,7\\8\end{array} \xmapsto{\operatorname{rad} G_\lambda} \begin{array}{c}0\,0\,1\,0\\0\,0\,0\,0\\0\,7\\8\end{array} \xmapsto{L_\lambda} \begin{array}{c}0\,0\,1\,0\\0\,0\,0\,0\\0\,1\\8\end{array} \xmapsto{\operatorname{rad} G_\lambda} \begin{array}{c}0\,0\,1\,0\\0\,0\,0\,0\\0\,1\\0\end{array}$$

Thus $a = (3, 2)$ and $b = (1, 0)$. Moreover, it is not hard to see two partitions (a, b) are uniquely determined. See [3, Proposition 2.5]. We left it to the reader to figure out the general case when A is not necessarily nilpotent.

References

[1] O. Schiffmann, Lectures on Hall algebras, arXiv:0611617.

[2] I. G. Macdonald, Symmetric functions and Hall polynomials. Oxford university press.

[3] Springer T A. The exotic nilcone of a symplectic group[J]. Journal of Algebra, 2009, 321(11): 3550-3562.

Hints

14.8 Since $n(\lambda)$ is the sum of numbers obtained by attaching 0 to each $\bullet$ in the top row, and 1 to each $\bullet$ in the second row, and so on.

14.12 Actually, $TA = AT$ is a linear equation in entries of T.

14.13 This follows from the direct computation.

15 Subspaces Avoidence

In this section, we will discuss the phenomenon of subspaces avoidance. The simplest version states that over an infinite field $\mathbb{F}$ a finite-dimensional vector space V cannot be covered by finite many proper subspaces.

If $\mathbb{F} = \mathbb{R}$, this assertation follows from a measure theory argument. To be exact, any proper subspaces of $\mathbb{R}^n$ have measure zero under the standard Lebesgue measure. Therefore the union of at most countable proper subspaces is also of measure zero. Similarly, this also follows a Baire category argument, see [1]. Note that the same is true for $\mathbb{C}$ since we can view a complex vector space as a real one with dimension doubled.

In the general case, we can make an attempt for the union of two subspaces. For two proper subspaces U_1 and U_2 of V. If $U_1 \subseteq U_2$ or $U_2 \subseteq U_1$, then the union is a single proper subspace, thus cannot cover V. Otherwise, we can pick $a \in U_1 \setminus U_2$ and $b \in U_2 \setminus U_1$, in this case, it is easy to show $a + b$ is not in $U_1 \cup U_2$. This can be generalized to more subspaces.

Theorem 15.1 Let $\mathbb{F}$ be an infinite field. Any vector space (not necessarily finite-dimensional) cannot be covered by finite many proper subspaces.

— *Proof.* Assume $U_1, \dots, U_n$ be subspaces of V. This is clear when $n = 1$, so we assume $n \geq 2$. By induction, we can assume each U_i is not contained in the union of U_j for $j = i$. In other words, we can find $x_i \in U_i$ but not in U_j for $j \neq i$. Note that

$$\#\left\{ x_1 + \lambda x_2 : \lambda \in \mathbb{F}^\times \right\} = \infty.$$

Assume $U_1 \cup \cdots \cup U_n = V$, then there must be one U_i contains at least two elements, i.e.

$$v_1 = x_1 + \lambda_1 x_2 \in U_i, \qquad v_2 = x_1 + \lambda_2 x_2 \in U_i$$

for $\lambda_1 \neq \lambda_2$, which implies

$$x_2 = \frac{v_1 - v_2}{\lambda_1 - \lambda_2} \in U_i, \qquad x_1 = v_1 - \lambda_1 x_2 \in U_i.$$

This contradicts our choice. □

Theorem 15.2 Let V be a finite-dimensional vector space over an infinite field $\mathbb{F}$. If

$$V = \bigcup_{i \in I} U_i$$

for $\{U_i\}_{i \in I}$ a family of proper subspaces, then $|I| \geq |\mathbb{F}|$.

— *Proof.* We can assume $V = \mathbb{F}^n$. We can construct the following curve

$$\gamma : \mathbb{F} \longrightarrow \mathbb{F}^n, \qquad t \longmapsto (1, t, t^2, \dots, t^{n-1}).$$

The main feature of this curve is, any n different points on γ are not contained in any proper subspace. That is, for different $t_1, \dots, t_n$,

$$\begin{pmatrix} 1 \\ t_1 \\ \vdots \\ t_1^{n-1} \end{pmatrix}, \begin{pmatrix} 1 \\ t_2 \\ \vdots \\ t_2^{n-1} \end{pmatrix}, \cdots, \begin{pmatrix} 1 \\ t_n \\ \vdots \\ t_n^{n-1} \end{pmatrix}$$

are linear independent. This is clear by Vandemonde's determinant. As a result, any U_i contains at most $n-1$ many points on γ. In particular, since $|I| = \infty$,

$$|I| = (n-1)|I| \geq |\mathbb{F}|.$$

This finishes the proof. □

▶ **Problem 15.3** Give a counterexample when $\dim V = \infty$. ◀ P74

▶ **Problem 15.4** Let $\mathbb{F}_q$ be a finite field of q elements. Show that we need all $q+1$ one-dimensional subspaces to cover $\mathbb{F}_q^2$.

▶ **Problem 15.5** Let $f(x_1, \dots, x_n)$ be a polynomial and $\mathcal{S} \subseteq \mathbb{F}$ be a finite set. If $\deg f < |\mathcal{S}|$, show that we can always find $x_1, \dots, x_n \in \mathcal{S}$ such that $f(x_1, \dots, x_n) \neq 0$. ◀ P74

Theorem 15.6 Let V be a finite-dimensional vector space over a finite field $\mathbb{F}$. If

$$V = \bigcup_{i \in I} U_i$$

for $\{U_i\}_{i \in I}$ a family of proper subspaces, then $|I| \geq |\mathbb{F}| + 1$.

— *Proof.* We assume $V = \mathbb{F}^n$. By deleting the repeating subspaces, we can assume I to be finite. Without loss of generality, we assume every W_i to be a hyperplane, say $U_i = \{f_i = 0\}$. The condition says

$$f(x_1, \dots, x_n) = \prod f_i(x_1, \dots, x_n) = 0 \tag{$*$}$$

for all $x \in V$. Note that f is homogeneous of degree $|I|$. By the problem above, we see $|I| > |\mathbb{F}|$. Assume $|I| = |\mathbb{F}|$. Now since 0 is contained in each U_i, we have

$$\#\left(\bigcup U_i\right) < |I| \cdot |\mathbb{F}^{n-1}| = |\mathbb{F}^n| = |V|,$$

which is absurd. As a result, $|I| \geq |\mathbb{F}| + 1$. □

Remark 15.7 Lastly, we remark a vast extension of Problem 15.5, combinatorial nullstellensatz by N.Alon [2].

> Let $\mathbb{F}$ be an arbitrary field, and let $f = f(x_1, \ldots, x_n)$ be a polynomial in $\mathbb{F}[x_1, \ldots, x_n]$. Suppose the degree $\deg(f)$ of f is $\sum_{i=1}^{n} t_i$, where each t_i is a nonnegative integer, and suppose the coefficient of $\prod_{i=1}^{n} x^{t_i}$ in f is nonzero. Then, if $\mathcal{S}_1, \ldots, \mathcal{S}_n$ are subsets of $\mathbb{F}$ with $|\mathcal{S}_i| > t_i$, there are $s_1 \in \mathcal{S}_1$, $s_2 \in \mathcal{S}_2$, $\ldots$, $s_n \in \mathcal{S}_n$ so that
>
> $$f(s_1, \ldots, s_n) \neq 0.$$

References

[1] J C.Oxtoby, A Survey of the Analogies between Topological and Measure Spaces[J]. SPRINGERVERLAG, New York Heidelberg Berlin, 1971.

[2] N.Alon, Combinatorial nullstellensatz[J]. Combinatorics, Probability and Computing, 1999, 8(1-2): 7-29.

Hints

15.3 Consider $\bigoplus_{i=1}^{\infty} \mathbb{R}$, it is the union of $\bigoplus_{i=1}^{n} \mathbb{R}$ for all n.

15.5 Apply the induction hypothesis on the coefficient of the leading term of f with respect to x_1.

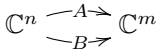
16 Classification of $A + \lambda B$

In this section, we will classify the equivalence class of pair of complex matrices where (A, B) is similar to (A', B') if there exists invertible P and Q such that

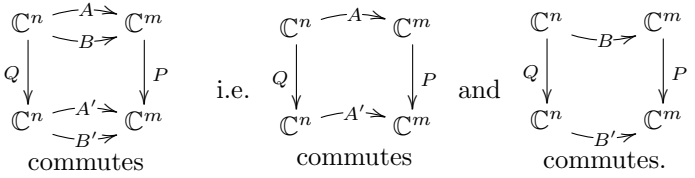
$$PA = A'Q, \qquad PB = B'Q.$$

Note that we are actually classifying indecomposable representation of **Kronecker quiver**

More precisely, if we draw (A, B) as



then (A, B) is equivalent to (A', B') if and only if the following



▶ **Exercise 16.1** Assume A is invertible, then (A, B) are similar to (I, J) where J is a Jordan canonical form.

Before preceding, let us consider the translation of this problem in terms of λ-matrices. Recall a λ-matrix is nothing but an element in $\mathbb{M}_n(\mathbb{C})[\lambda] = \mathbb{M}_n(\mathbb{C}[\lambda])$.

▶ **Exercise 16.2** Show that a λ-matrix $P(\lambda)$ is invertible if and only if $\det P(\lambda)$ is a nonzero number. Equivalently, $P(\lambda)$ is invertible for any complex number λ.

▶ **Problem 16.3** Assume A is invertible Show that two pairs (A, B) is similar to (A', B') if and only if there exists invertible λ-matrices $P(\lambda)$ and $Q(\lambda)$ such that

$$P(\lambda)(A + \lambda B) = (A' + \lambda B')Q(\lambda).$$

◀ P80

By replacing A by $A + \lambda_0 B$ for some λ_0, we actually find the canonical form for pairs (A, B) with the polynomial $\det(A + \lambda B)$ nonzero. The details are left to readers. Now, we will assume $\det(A + \lambda B) = 0$ (including the case (A, B) are not square matrices).

Kronecker Theorem 16.4 If $\det(A+\lambda B) = 0$, then (A, B) is equivalent to

$$\left(\begin{pmatrix} R & \\ & A' \end{pmatrix}, \begin{pmatrix} L & \\ & B' \end{pmatrix}\right) \quad \text{or} \quad \left(\begin{pmatrix} R^{\mathrm{t}} & \\ & A' \end{pmatrix}, \begin{pmatrix} L^{\mathrm{t}} & \\ & B' \end{pmatrix}\right)$$

where

$$R = \begin{pmatrix} 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 1 \end{pmatrix}, \qquad L = \begin{pmatrix} 1 & & & \\ & \ddots & & 0 \\ & & 1 & \vdots \\ & & 0 & 0 \end{pmatrix}.$$

Since $\det(A+\lambda B) = 0$ or $A+\lambda B$ is not even a square matrix, by replacing $A + \lambda B$ by its transpose, we can assume there is a nonzero null λ-vector for $(A - \lambda B)$ (here we use $A - \lambda B$ to simplify the signs). Precisely, there exists

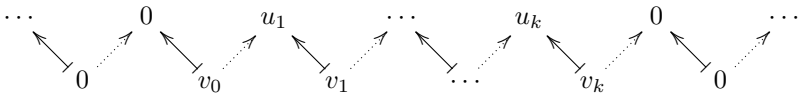
$$v(\lambda) = v_0 + v_1\lambda + \cdots + v_k\lambda^k \in \mathbb{C}^n[\lambda]$$

such that

$$(A - \lambda B)v(\lambda) = 0.$$

We assume the degree k of $v(\lambda)$ is as small as possible.

By expanding the equation, we get



where $A : \longmapsto$ and $B : \dashrightarrow$. Conversely, the diagram of the shape above provides a null λ-vector.

We shall denote vectors by

$$\begin{bmatrix} \cdots & 0 & u_1 & u_2 & \cdots & u_k & 0 & \cdots \\ \cdots & 0 & v_0 & v_1 & \cdots & \cdots & v_k & 0 & \cdots \end{bmatrix}.$$

Claim Vectors $u_1, \ldots, u_k$ are linearly independent.

— *Proof.* Let $\lambda_1 u_1 + \cdots + \lambda_k u_k = 0$ be linear relation. Then we consider

$$\begin{aligned}
& \lambda_1 \begin{bmatrix} \cdots & 0 & u_1 & u_2 & \cdots & u_k & 0 & \cdots \\ \cdots & 0 & v_0 & v_1 & \cdots & \cdots & v_k & 0 & \cdots \end{bmatrix} \\
+ \; & \lambda_2 \begin{bmatrix} \cdots & u_1 & u_2 & \cdots & u_k & 0 & 0 & \cdots \\ \cdots & v_0 & v_1 & \cdots & \cdots & v_k & 0 & 0 & \cdots \end{bmatrix} \\
+ \; & \cdots \\
+ \; & \lambda_k \begin{bmatrix} \cdots & u_{k-1} & u_k & 0 & 0 & \cdots & 0 & \cdots \\ \cdots & v_{k-2} & v_{k-1} & v_k & 0 & \cdots & \cdots & 0 & \cdots \end{bmatrix} \\
= \; & \begin{bmatrix} \cdots & * & 0 & * & \cdots & * & 0 & \cdots \\ \cdots & * & * & * & \cdots & \cdots & * & 0 & \cdots \end{bmatrix}
\end{aligned}$$

We get a null vector

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

By our choice, all $*$ vanish, and thus all $\lambda_i = 0$. The proof is complete. □

▶ **Exercise 16.5** Show that $v_0, v_1, \ldots, v_k$ are linearly independent.

Now we find the matrices of (A, B) restricting to

$$V = \operatorname{span}(v_i) \underset{B}{\overset{A}{\rightrightarrows}} \operatorname{span}(u_i) = U$$

is nothing but (R, L). Now, (A, B) is equivalent to

$$\left(\begin{pmatrix} R & C \\ & A_1 \end{pmatrix}, \begin{pmatrix} L & D \\ & B_1 \end{pmatrix} \right)$$

for some matrices C and D. Now we assume

$$(A, B) = \left(\begin{pmatrix} R & C \\ & A_1 \end{pmatrix}, \begin{pmatrix} L & D \\ & B_1 \end{pmatrix} \right)$$

I claim there exist matrices X and Y such that

$$\begin{pmatrix} 1 & X \\ & 1 \end{pmatrix} \begin{pmatrix} R & C \\ & A_1 \end{pmatrix} = \begin{pmatrix} R & \\ & A_1 \end{pmatrix} \begin{pmatrix} 1 & Y \\ & 1 \end{pmatrix}$$
$$\begin{pmatrix} 1 & X \\ & 1 \end{pmatrix} \begin{pmatrix} L & D \\ & B_1 \end{pmatrix} = \begin{pmatrix} L & \\ & B_1 \end{pmatrix} \begin{pmatrix} 1 & Y \\ & 1 \end{pmatrix}$$

That is, the following matrix equation

$$\begin{cases} RY - XA_1 = C \\ -LY + XB_1 = -D \end{cases}$$

is solvable. To see this, we need to show the linear map

$$(X, Y) \longmapsto (RY - XA_1, -LY + XB_1)$$

is surjective.

▶ **Exercise 16.6** Check that the adjoint of

$$\mathbb{M}_{a\times b}(\mathbb{C}) \longrightarrow \mathbb{M}_{c\times d}(\mathbb{C}), \qquad A \longmapsto PAQ$$

is

$$\mathbb{M}_{d\times c}(\mathbb{C}) \longrightarrow \mathbb{M}_{b\times a}(\mathbb{C}), \qquad B \longmapsto QBP.$$

In particular, the matrix equation $PXQ = A$ has a solution for any A if and only if $QYP = 0$ has only zero solution.

Equivalently, the adjoint

$$(M, N) \longmapsto (MR - NL, A_1M - B_1N)$$

is injective.

Claim The equation

$$\begin{cases} MR - ML = 0 \\ A_1M - B_1N = 0 \end{cases}$$

has only zero solution.

— *Proof.* Let us denote $M = (x_0, \dots, x_{k-1})$, the equation tells that we have the following diagram

where $A_1 : \Longmapsto$ and $B_1 : = = \Rightarrow$. Firstly, we will lift the chain to

where $A : \longmapsto$ and $B : \dashrightarrow$. We require it is a lift in the following sense The last $n - k - 1$ component of $\tilde{x}_i$ is x_i, i.e. $\tilde{x}_i = \binom{*}{x_i}$. Then the last $m - k$ component of $\tilde{y}_i$ is y_i, i.e. $\tilde{t}_i = \binom{*}{t_i}$. We include $i = 0$ by assuming $y_0 = 0$.

Note that

$$B\begin{pmatrix} 0 \\ x_{k-1} \end{pmatrix} = Dx_{k-1} + B_1x_{k-1} = Dx_{k-1} \in \operatorname{span}(v_i)$$

So we can find $-Lx'_{k-1} = B\binom{0}{x_{k-1}}$. We put $\tilde{x}_{k-1} = \binom{x'_{k-1}}{x_{k-1}}$, we have $B\tilde{x}_{k-1} = 0$. Let us denote $\tilde{y}_{k-1} = A\tilde{x}_{k-1}$.

Note that

$$B\begin{pmatrix} 0 \\ x_{k-2} \end{pmatrix} - \tilde{y}_{k-1} \in \operatorname{span}(v_i).$$

So we can find $-Lx'_{k-2} = B\binom{0}{x_{k-2}} - \tilde{y}_{k-1}$. We put $\tilde{x}_{k-2} = \binom{x'_{k-2}}{x_{k-2}}$, we have

$$B\tilde{x}_{k-2} = A\tilde{x}_{k-1}.$$

Let us denote $\tilde{y}_{k-2} = A\tilde{x}_{k-2}$.

We can extend this process to get a chain claimed above to get

$$\begin{bmatrix} \cdots & \tilde{y}_0 & \tilde{y}_1 & \tilde{y}_2 & \cdots & \tilde{y}_k & 0 & \cdots \\ \cdots & \cdots & \tilde{x}_0 & \tilde{x}_1 & \cdots & \cdots & \tilde{x}_k & 0 & \cdots \end{bmatrix}.$$

Now assume $\tilde{y_0} = \lambda_1 u_1 + \cdots + \lambda_k u_k$. We then consider

$$\begin{aligned} & \lambda_1 \begin{bmatrix} \cdots & 0 & u_1 & u_2 & \cdots & u_k & 0 & \cdots \\ \cdots & 0 & v_0 & v_1 & \cdots & \cdots & v_k & 0 & \cdots \end{bmatrix} \\ + \quad & \lambda_2 \begin{bmatrix} \cdots & u_1 & u_2 & \cdots & u_k & 0 & 0 & \cdots \\ \cdots & v_0 & v_1 & \cdots & \cdots & v_k & 0 & 0 & \cdots \end{bmatrix} \\ + \quad & \cdots \\ + \quad & \lambda_k \begin{bmatrix} \cdots & u_{k-1} & u_k & 0 & 0 & \cdots & 0 & \cdots \\ \cdots & v_{k-2} & v_{k-1} & v_k & 0 & \cdots & \cdots & 0 & \cdots \end{bmatrix} \\ - \quad & \begin{bmatrix} \cdots & \cdots & \tilde{y}_0 & 0 & \cdots & \tilde{y}_k & 0 & \cdots \\ \cdots & \cdots & \cdots & \tilde{x}_0 & \cdots & \cdots & \tilde{x}_k & 0 & \cdots \end{bmatrix} \\ = \quad & \begin{bmatrix} \cdots & * & 0 & * & \cdots & * & 0 & \cdots \\ \cdots & * & * & * & \cdots & \cdots & * & 0 & \cdots \end{bmatrix} \end{aligned}$$

We get a null vector

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

By our choice, all $*$ should vanish, but it implies $\tilde{x}_i \in \operatorname{span}(u_i)$ and thus $x_i = 0$. The proof is complete. □

Now, we finished the proof of the Kronecker theorem.

Remark 16.7 The quiver

is known as **Kronecker quiver**. It helps to

- understand of (modular) representation of Klein four-group, see [1];
- set a derived equivalence between coherent sheaves over $\mathbb{P}^1$, see [2].

References

[1] Benson D J. Representations and cohomology: Volume 1, basic representation theory of finite groups and associative algebras[M]. Cambridge university press, 1998.

[2] A.A. Beilinson, Coherent sheaves on $\mathbb{P}^n$ and problems of linear algebras, Func. Anal. Appl. 12 (1978), pp. 214-216.

Hints

16.3 The "only if" part is trivial. Now we assume

$$P(\lambda)(A + \lambda B) = (A' + \lambda B')Q(\lambda).$$

Note that

$$\det(A + \lambda B) = \det(A' + \lambda B) \quad \text{up to some nonzero complex number.}$$

Thus A' is also invertible. Pick any $N > \deg p$ and $N > \deg q$. Then we can write by Euclidean algorithm

$$\begin{aligned} P(\lambda) &= (A' + \lambda B')U(\lambda) + P_0\lambda^n \\ Q(\lambda) &= V(\lambda)(A + \lambda B) + Q_0\lambda^n \end{aligned}$$

where $\deg U(\lambda) < N$ and $\deg V(\lambda) < N$. Now we have

$$(A' + \lambda B')\Big(U(\lambda) - V(\lambda)\Big)(A + \lambda B) = -\Big(P_0(A + \lambda B) - (A + \lambda B)Q_0\Big)\lambda^N.$$

Take $\lambda = 0$, we see

$$A'\big(U(0) - V(0)\big)A = 0.$$

Since we assume A and A' to be invertible, $\frac{1}{\lambda}(U(\lambda) - V(\lambda))$ is a polynomial of degree $< N - 1$. Substituting and canceling λ both side, we can continuous our process to conclude that $U(\lambda) = V(\lambda)$. This leads to

$$P_0(A + \lambda B) = (A + \lambda B)Q_0.$$

By comparing the coefficients, it is equivalent to say (A, B) is equivalent to (A', B').

17 Quadratic Forms

In this section, we are going to present the basic theory of quadratic forms over an arbitrary field $\mathbb{F}$ of characteristic $\neq 2$. We recommend [1] for a short introduction. A **quadratic form** is a homogeneous polynomial of degree 2. In general, we can write

$$q(x) = \sum_{ij} a_{ij} x_i x_j$$

for a symmetric matrix (a_{ij}). There is an associated symmetric bilinear form

$$B_q(x, y) := \sum_{ij} a_{ij} x_i y_j = \frac{1}{2}(q(x+y) - q(x) - q(y)).$$

Instead of considering quadratic forms, we shall consider **quadratic space**, i.e. a vector space V equipped with some symmetric bilinear form $q \in \mathsf{S}^2 V^*$. For example, square sum

$$q(x) = x_1^2 + \cdots + x_n^2$$

is a quadratic form, and it equips the space $\mathbb{F}^n$ the structure of quadratic space.

Example 17.1 (Diagonalized form) For any $a \in \mathbb{F}$, we have a one-dimensional quadratic form $q = \langle a \rangle$ defined by

$$q(x) = ax^2.$$

More generally, for $a_1, \ldots, a_n \in \mathbb{F}$, the quadratic form $q = \langle a_1, \ldots, a_n \rangle$ given by

$$q(x) = a_1 x_1^2 + \cdots + a_n x_n^2.$$

Example 17.2 (Hyperbolic plane) We denote the **hyperbolic plane** by

$$\mathbb{H} = \langle 1, -1 \rangle .$$

The corresponding quadratic form is

$$q(x, y) = x^2 - y^2 = (x - y)(x + y).$$

By changing of variable, it is equivalent to

$$q'(x, y) = xy.$$

We say a quadratic form is **hyperbolic** if it is an orthogonal sum of copies of hyperbolic planes.

Definition 17.3 Let (V, q) be a quadratic space. For any subspace U, we define the **orthogonal**

$$U^\perp = \{v \in V : \forall u \in W,\ B_q(v, u) = 0\}.$$

Note that it is possible to have $q(x) = 0$ for some x, thus it is possible to have $(\mathbb{F}x) \subseteq (\mathbb{F}x)^\perp$.

Definition 17.4 (Regular) For a quadratic space (V, q), we denote the **radical**

$$\operatorname{rad} q = \{v \in V : B_q(v, -) = 0\} = V^\perp.$$

We say q is **regular** or **nondegenerate** if $\operatorname{rad} q = 0$. We say q is **totally isotropic** or **trivial** if $\operatorname{rad} q = V$, i.e. $q = 0$.

▶ **Problem 17.5** Show that up to isomorphism, we have a unique decomposition

$$q = q_{\mathrm{r}} \oplus q_{\mathrm{t}} \qquad \text{(as quadratic space)}$$

where q_{r} is regular, and q_{t} is trivial. ◀ **P85**

Thus the problem of classifying quadratic forms reduces to the case when q is regular.

▶ **Exercise 17.6** Let (V, q) be a regular quadratic space. For any subspace u, we have

$$\dim U^\perp + \dim U = \dim V.$$

Moreover, if $(U, q|_U)$ is regular, we have

$$V = U \oplus U^\perp \qquad \text{(as quadratic space)}.$$

Particularly, $(U^\perp, q|_{U^\perp})$ is still regular.

We can apply this exercise to one-dimensional subspace $\mathbb{F}v$. Note that the restriction of q to $\mathbb{F}v$ is $\langle q(v)\rangle$. Thus if $a_1 = q(v) \neq 0$, we have a decomposition

$$q = \langle a_1\rangle \oplus q_1 \qquad \text{as quadratic space.}$$

By induction, we obtain the following.

Theorem 17.7 Any (regular) quadratic forms can be diagonalized, i.e. equivalent to $\langle a_1, \dots, a_n\rangle$ for some $a_i \in \mathbb{F}$ ($a_i \in \mathbb{F}^\times$).

Example 17.8 Over $\mathbb{C}$, and quadratic forms are of the form

$$\langle 1, \dots, 1, 0, \dots, 0 \rangle$$

where the number of 1's is the rank.

Example 17.9 Over $\mathbb{R}$, and quadratic forms are of the form

$$\langle 1, \dots, 1, -1, \dots, -1, 0, \dots, 0 \rangle$$

where the number of 1's (resp., -1's) is the **positive (resp., nagative) index of inertia**.

Definition 17.10 Let (V, q) be a regular quadratic space. A vector $v \neq 0$ is said to be **isotropic** if $q(v) = 0$. The space (V, q) is said to be **isotropic** if it admits an isotropic vector. Otherwise, (V, q) is said to be **anisotropic**.

Example 17.11 All regular one-dimensional quadratic forms $\langle a \rangle$ for $a \in \mathbb{F}^\times$ are anisotropic.

Example 17.12 Hyperbolic plane $\mathbb{H}$ is isotropic. Since $q(x, y) = xy$, any nonzero vector over x-axis or y-axis is isotropic.

Example 17.13 Over $\mathbb{R}$, any positively (negatively) definite quadratic form is anisotropic.

Example 17.14 Over $\mathbb{C}$, only one-dimensional quadratic forms can be anisotropic.

► **Problem 17.15** Over $\mathbb{Q}$, for two integers p and q, show that if p is not a square mod q, then

$$\langle 1, -p, -q, \ pq \ \rangle$$

is anisotropic. ◄ **P85**

Aemma 17.16 Assume q is regular but isotropic, then we have

$$q = q_{\mathtt{h}} \oplus q_{\mathtt{a}} \qquad \text{(as quadratic spaces)}$$

where $q_{\mathtt{h}}$ is hyperbolic and $q_{\mathtt{a}}$ is anisotropic.

— *Proof.* If q is anisotropic, then we are done since we can take $q_{\mathtt{h}} = 0$. Now, assume q isotropic, i.e. there exists x such that $q(x) = 0$. Since q is regular

$$\mathbb{F}x \subseteq (\mathbb{F}x)^\perp \neq V$$

i.e. there exists $y \notin \mathbb{F}x$ such that $B_q(x,y) \neq 0$. Note that

$$\begin{aligned} B_q(x, y+\lambda x) &= B_q(x,y) \\ B_q(y+\lambda x, y+\lambda x) &= q(y) + 2\lambda B_q(x,y) \end{aligned}$$

By replacing y by $y+\lambda x$ for suitable λ, we can assume $q(y) = 0$. Now, the restriction of q over $\mathbb{F}x \oplus \mathbb{F}y$ is hyperbolic. Thus we have

$$q = \mathbb{H} \oplus q_1 \qquad \text{(as quadratic spaces)}.$$

The proof follows from induction. □

Now we are going to show the decomposition in the Lemma 17.16 is unique. It follows from the following Witt cancelation theorem.

Witt cancelation 17.17 Assume

$$\langle a \rangle \oplus q \cong \langle a \rangle \oplus p \qquad \text{(as quadratic spaces)}$$

then

$$p \cong q \qquad \text{(as quadratic spaces)}.$$

Aemma 17.18 Let (V,q) be a quadratic space. For two vectors u and v such that $q(u) = q(v) \neq 0$, there exists

$$A \in O(q) = \{A \in \mathrm{GL}(V) : q(Ax) = q(x)\}$$

such that $Au = v$.

— *Proof.* We should first try the reflection by $\delta := u - v$.

Case A. If $q(\delta) \neq 0$, the reflection is defined to be

$$R_\delta : x \longmapsto x - 2\frac{B_q(x,\delta)}{q(\delta)}\delta.$$

It is easy to see $R_\delta \in O(q)$, and $R_\delta(u) = v$.

Case B. If $q(\delta) = 0$, then we should try to reflect u to some place such that we can reduce to Case A. The condition tells

$$q(u) = q(v) = B_q(u,v) \neq 0$$

Then we can compute $R_v u = u - 2v$. Note that

$$q\big((u-2v)-v\big) = q(x)(1-6+9) \neq 0$$

So we reduce to Case A. □

— *Proof of 17.17.* Since the decomposition into totally isotropic and regular part is unique, we can assume $a \neq 0$.

Geometrically, it says there exists u and v such that

$$q(u) = q(v) = a \neq 0 \quad \text{and} \quad \begin{cases} q = (\mathbb{F}u)^\perp \\ p = (\mathbb{F}u)^\perp \end{cases}$$

By Lemma above, we can move u to v without changing the quadratic form, say by A. As a result $q \cong p$ by A. □

Witt decomposition 17.19 For any quadratic form q, we have a unique decomposition

$$q = q_{\mathtt{t}} \oplus q_{\mathtt{h}} \oplus q_{\mathtt{a}} \qquad \text{(as quadratic spaces)}$$

where $q_{\mathtt{t}}$ is trivial, $q_{\mathtt{h}}$ is hyperbolic and $q_{\mathtt{a}}$ is annisotropic.

▶ **Sylvester's law of inertia 17.20** The **positive index of inertia** and **nagative index of inertia** are invariants of quadratic forms over $\mathbb{R}$.

References

[1] Serre J P. A course in arithmetic [M]. Springer Science & Business Media, 2012.

Hints

17.5 $\left(\begin{smallmatrix} 1 & \\ -C^{\mathtt{t}}A^{-1} & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} A & C \\ C^{\mathtt{t}} & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & -A^{-1}C \\ & 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} A & \\ & 0 \end{smallmatrix}\right)$

17.15 That is, the equation

$$x^2 + pqy^2 = pz^2 + qw^2$$

has no nonzero integer solution. Assume we have a nonzero solution (x, y, z, w), we have

$$x^2 \equiv pz^2 \bmod q.$$

By assumption, it implies $q|x$ and $q|z$. Say $x = qx_0$ and $z = qz_0$, so (w, z_0, y, x_0) is another solution. Note that the norm of the solution would infinitely decrease which is a contradiction.

18 Symplectic Spaces

A **symplectic space** is a vector space V equipped with a nondegenerate anti-symmetric form $\omega \in \Lambda^2 V^*$. Here ω is **nondegenerate** if the pairing

$$V \times V \longmapsto \mathbb{R}, \qquad (x, y) \longmapsto \omega(x, y)$$

induces an isomorphism

$$V \xrightarrow{\sim} V^* \qquad x \longmapsto \omega(x, -).$$

Correspondently, by picking a basis, the measure matrix $\big(\omega(\mathbf{e}_i, \mathbf{e}_j)\big)$ is anti-symmetric and of full rank.

The following problem would explain the reason we should care about symplectic spaces.

▶ **Birkhoff–von Neumann 18.1** For a bilinear form $f \in V^* \otimes V^*$, if

$$f(x, y) = 0 \iff f(y, x) = 0$$

then f is symmetric or anti-symmetric. Geometrically, it says the left annihilator coincides with the right annihilator if and only if f is symmetric or anti-symmetric. ◀ **P92**

Example 18.2 For any vector space X, the space

$$X \oplus X^*$$

has a natural symplectic form ω defined by

$$\omega\big(x + \phi, y + \psi\big) = -\phi(y) + \psi(x).$$

By picking basis and dual basis for X and X^*, the matrix is $\begin{pmatrix} & -I \\ I & \end{pmatrix}$.

Remark 18.3 There would be a more geometric method to define this symplectic form. Picking a basis of X, we get an isomorphism

$$(q_i) : X \longmapsto \mathbb{R}^n,$$

it extends to an isomorphism by considering the dual basis

$$(q_i, p_i) : X \oplus X^* \longmapsto \mathbb{R}^{2n}.$$

Let us denote the **Lagrangian form**

$$\lambda = \sum_i p_i dq_i \in \Omega^1(T^* X)$$

to be the universal 1-form. It is universal in the following sense,

for any $\alpha \in \Omega^1(X)$, viewed as a section $\alpha : X \to T^*X$, we have $\alpha^*(\lambda) = \alpha$.

$$\begin{array}{ccc} \alpha & \xleftarrow{\alpha^*} & \lambda \\ X & \xrightarrow[\alpha]{} & T^*X \end{array}$$

We define

$$\omega = d\lambda = \sum dp_i \wedge dq_i \in \Omega^2(T^*X).$$

For real vector spaces, nondegenerate symmetric forms (i.e. regular quadratic forms, or full-rank symmetric matrices) are classified by indices of inertia. However, there is only one symplectic space up to isomorphism.

Theorem 18.4 Any symplectic space is isomorphic to $X \oplus X^*$ for some vector space X.

— *Proof.* Pick any nonzero $x \in V$. Since ω is nondegenerate, we have

$$\mathbb{R}x \subseteq (\mathbb{R}x)^\perp \neq V.$$

So there exists $y \notin \mathbb{F}x$ such that

$$\omega(x, y) \neq 0.$$

We can without loss of generality to assume $\omega(x, y) = 1$. So the matrix over $\mathbb{F}x \oplus \mathbb{F}y$ is $\left(\begin{smallmatrix} & 1 \\ -1 & \end{smallmatrix}\right)$. We then have

$$V \cong (\mathbb{F}x \oplus \mathbb{F}y) \oplus (\mathbb{F}x \oplus \mathbb{F}y)^\perp \qquad \text{(orthogonal)}$$

Thus the theorem follows from induction. □

Let us denote the group of **symplectic transform**

$$\mathrm{Sp}(V) = \big\{A \in \mathrm{GL}(V) : \omega(A-, A-) = \omega(x, y)\big\}.$$

In terms of matrices, it tells

$$A^{\mathsf{t}} M_\omega A = M_\omega$$

for measure matrices M_ω. Taking the determinant on both sides, we get $\det A = \pm 1$. But actually, only $\det A = 1$ is possible.

Theorem 18.5 For any $A \in \mathrm{Sp}(V)$, $\det A = 1$.

In the rest of this section, we will show this fact through three approaches.

Approach A. Recall that $\det A$ has the following description. For a linear transform $A : V \to V$, it induces

$$\Lambda^k A : \Lambda^k V \longrightarrow \Lambda^k V$$

for any k. Since $\Lambda^n V$ is one-dimensional for $n = \dim A$, the map $\Lambda^k A$ is a scalar and the scalar is $\det A$.

▶ **Exercise 18.6** Assume (V, ω) is a symplectic space of dimension $2n$. Show that

$$\omega^n := \underbrace{\omega \wedge \cdots \wedge \omega}_{n} \in \Lambda^{2n} V^*$$

is nonzero.

— *Proof of Theorem 18.5.* Let A^* be the adjoint of A. Now $A^*\omega$ or more precisely $(\Lambda^2 A^*)\omega$ is ω. Thus we have

$$(\Lambda^{2n} A^*)\omega^n = \omega^n$$

So $\det A^* = \det A = 1$. □

Approach B. We are going to introduce a new invariant for antisymmetric matrices.

▶ **Problem 18.7** For an anti-symmetric matrix A with integer coefficients, then $\det A$ is zero or a perfect square. ◀ **P93**

Let x_{ij} for $1 \le i < j \le n$ be $\frac{n(n-1)}{2}$ many variables. We denote $x_{ii} = 0$ and $x_{ji} = -x_{ij}$. Now

$$\det(x_{ij}) \in \mathbb{Z}[x_{ij}].$$

Since we can find invertible matrix P with coefficients in rational field $\mathbb{Q}(x_{ij})$ such that

$$PAP^{\mathsf{t}} = \operatorname{diag}\left(\begin{pmatrix} & 1 \\ -1 & \end{pmatrix}, \ldots, \begin{pmatrix} & 1 \\ -1 & \end{pmatrix}, 0, \cdots\right).$$

So we have $\det A \in \mathbb{Q}(x_{ij})^2$.

Aemma 18.8 Show that

$$\mathbb{Q}(x_{ij})^2 \cap \mathbb{Z}[x_{ij}] = \mathbb{Z}[x_{ij}]^2.$$

— *Proof.* Assume $(p/q)^2 = h$ for some polynomials p, q, h. We assume p, q to be relatively prime. Note that $q^2 | hq^2 = p^2$ which implies q is a unit. This shows p/q is a polynomial. □

Note that this implies that there exists a polynomial $P(x_{ij}) \in \mathbb{Z}[x_{ij}]$ such that $P(x_{ij})^2 = \det(x_{ij})$. Can we find an explicit choice? Recall that for an $n \times n$ matrix (x_{ij}), we have

$$\det(x_{ij}) = \sum_{\sigma \in \mathfrak{S}_n} (-1)^{\ell(w)} x_{1\sigma(1)} \cdots x_{n\sigma(n)}$$

where $\mathfrak{S}_n$ is the group of permutations, and

$$\ell(\sigma) = \#\big\{(i < j) : w(i) < w(j)\big\}$$

is the number of inversions.

Definition 18.9 (Pffafian) For an anti-symmetric matrix $A = (x_{ij})$ or size $2n$, we define

$$\operatorname{pf}(A) = \sum (-1)^{\ell(\pi)} x_{i_1 j_1} \cdots x_{i_n j_n}$$

where the sum goes over all possible pairings π of $\{1, \dots, 2n\}$

$$(i_1 < j_1), \cdots, (i_n < j_n) \qquad \text{(unordered)}$$

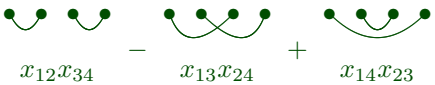
where

$$\ell(\pi) = \#\{(a < b) : i_a < i_b < j_a < j_b\}.$$

Example 18.10 We have $\operatorname{pf}\begin{pmatrix} & x \\ -x & \end{pmatrix} = x$ since there is only one pairing  .

Example 18.11 We have

$$\operatorname{pf}\begin{pmatrix} & x_{12} & x_{13} & x_{14} \\ -x_{12} & & x_{23} & x_{24} \\ -x_{13} & -x_{23} & & x_{34} \\ -x_{14} & -x_{24} & -x_{34} & \end{pmatrix} =$$



Theorem 18.12 We have

$$\operatorname{pf}(A) = \frac{1}{2^n n!} \sum_{w \in \mathfrak{S}_{2n}} (-1)^{\ell(\sigma)} x_{\sigma(1)\sigma(2)} \cdots x_{\sigma(2n-1)\sigma(2n)}.$$

— *Proof.* Let Π be the set of pairings. Note that $\mathfrak{S}_{2n}$ acts on Π by permuting indices. We have a surjective map $\mathfrak{S}_{2n} \to \Pi$ by sending σ to the pairing

$$(\sigma(1), \sigma(2)), \dots, (\sigma(2n-1), \sigma(2n)).$$

Since the stabilizer of this action is $\mathfrak{B}_n = (\mathbb{Z}_2)^n \rtimes \mathfrak{S}_n$, i.e. the group generated by

- transpositions of $2i-1$ and $2i$ and
- permutations of pairs $(1<2),\dots,(2n-1,2n)$.

Moreover, the summand

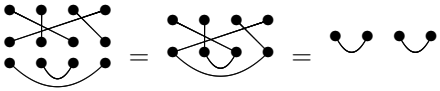
$$(-1)^{\ell(w)}x_{\sigma(1)\sigma(2)}\cdots x_{\sigma(2n-1)\sigma(2n)}$$

does not change under $\mathfrak{B}_n$. Moreover, by a combinatorial argument, we have

$$\ell(\pi)\equiv\ell(\sigma) \mod 2.$$

This finishes the proof. □

Actually, the action can be illustrated as follows



▶ **Problem 18.13** Show that

$$\operatorname{pf}(PAP^{\mathsf{t}})=\det(P)\operatorname{pf}(A).$$

Show that

$$\operatorname{pf}(A)^2=\det(A).$$

— *Proof of Theorem 18.5.* Since

$$A^{\mathsf{t}}M_\omega A=M_\omega$$

we can take Pffafian on both sides to get

$$\det(A)\operatorname{pf}(M_\omega)=\operatorname{pf}(M_\omega).$$

Since $\det(M_\omega)=\operatorname{pf}(M_\omega)^2\neq 0$, we can conclude $\det A=1$. □

Approach C. Note that $\det A$ might equal -1 when A is orthogonal, i.e. preserving a fixed inner product $\langle -,-\rangle$. The simplest example is a reflection with respect to a nonzero vector y

$$x\longmapsto x-2\frac{\langle x,y\rangle}{\langle y,y\rangle}y.$$

Note that this refection fixes the hyperplane $(\mathbb{R}y)^\perp$ and sends y to $-y$. In symplectic space, there is a similar notion for this.

Definition 18.14 (Symplectic transvection) A **symplectic transvection** is a linear transform $A \in \mathrm{Sp}(V)$ such that A fixes a hyperplane.

▶ **Exercise 18.15** For $A \in \mathrm{Sp}(V)$, show that

$$\ker(A-1) = \mathrm{im}(A-1)^{\perp}.$$

◀ **P93**

Let t be a symplectic transvection. Note that $x \mapsto tx - x$ has only one-dimensional, say $\mathbb{R}y$ for some $y \in V$. By the above exercise, the hyperplane fixed by t is $H = (\mathbb{R}y)^{\perp}$. Thus we can assume

$$tx = x + a\omega(x,y)y$$

for some a. It is direct to check $t \in \mathrm{Sp}(V)$ for any $a \in \mathbb{R}$.

Theorem 18.16 Symplectic transvections generate $\mathrm{Sp}(V)$.

Let T be the group generated by symplectic transvections. By induction, it suffices to show the following claim.

Claim Assume x_i, y_i such that $\omega(x_i, y_i) = 1$ for $i = 1, 2$, then there exists $t \in T$ such that $tx_1 = x_2$ and $ty_1 = y_2$.

— *Proof.* We first ensure that $tx_1 = x_2$.

If $\omega(x_1, x_2) \neq 0$, then we can take the symplectic transvection,

$$z \longmapsto z + \frac{\omega(z,\delta)}{\omega(x_1,\delta)}\delta, \qquad \text{where } \delta := x_2 - x_1.$$

Note that it sends x_1 to x_2.

If $\omega(x_1, x_2) = 0$, then we can find $x' \in V$ such that $\omega(x_1, x') \neq 0$ and $\omega(x_2, x') \neq 0$ since

$$(\mathbb{R}x_1)^{\perp} \cup (\mathbb{R}x_2)^{\perp} \neq V.$$

It reduces to the case above.

We can now assume $x := x_1 = x_2$. If $\omega(y_1, y_2) \neq 0$, then we can take the symplectic transvection,

$$z \longmapsto z + \frac{\omega(z,\delta')}{\omega(y_1,\delta')}\delta', \qquad \text{where } \delta' := y_2 - y_1.$$

Note that it sends y_1 to y_2 and fixes x at the same time.

If $\omega(y_1, y_2) = 0$, by applying

$$z \longmapsto z \pm \omega(z,x)x$$

we can replace y_i by $y_i \pm x$ to reduce to the case above since

$$\omega(x, y_i \pm x) = \omega(x, y_i)$$

$$\omega(y_1 - x, y_2 + x) = 2\omega(y_1, x) \neq 0.$$

The proof is complete. □

— *Proof of Theorem 18.5.* By the above theorem, it suffices to show each symplectic transvection has determinant 1. Assume the fixed hyperplane is $H = (\mathbb{R}y)^\perp$, then

$$V/H \xrightarrow{\sim} \mathbb{R}, \qquad x \mapsto \omega(x, y).$$

Since t fixes H, only the induced action of t over V/H contributes determinant. Since t preserves ω and fixes H, thus the action is identity over V/H, so that the determinant is 1. □

Hints

18.1 Note that f is anti-symmetric if $f(x, x) = 0$ for any $x \in V$. Assume f is not anti-symmetric, so there exists x_0 such that $f(x_0, x_0) \neq 0$.

We first show $f(x_0, y) = f(y, x_0)$ for any y. Note that

$$\begin{cases} f(x_0, y + \lambda x_0) = f(x_0, y) + \lambda f(x_0, x_0) \\ f(y + \lambda x_0, x_0) = f(y, x_0) + \lambda f(x_0, x_0) \end{cases}$$

Since $f(x_0, x_0) \neq 0$, we can take suitable λ such that $f(x_0, y + \lambda x_0) = 0$, so by assumption we have $f(y + \lambda x_0, x_0) = 0$. Then $f(x_0, y) = f(y, x_0)$ by above identities.

Next, we show $f(x, y) = f(y, x)$ for any x, y. If $f(x, x_0) = f(x_0, x) \neq 0$, we can apply the same trick as above

$$\begin{cases} f(x, y + \lambda x_0) = f(x_0, y) + \lambda f(x, x_0) \\ f(y + \lambda x, x_0) = f(y, x_0) + \lambda f(x, x_0) \end{cases}$$

to conclude $f(x, y) = f(y, x)$. Similarly, if $f(y, x_0) = f(x_0, y) \neq 0$, then we are done.

Now assume $f(x, x_0) = f(y, x_0) = 0$. Note that

$$\begin{cases} f(x + \mu x_0, y + \lambda x_0) = f(x, y) + \lambda\mu f(x_0, x_0) \\ f(y + \lambda x_0, x + \mu x_0) = f(y, x) + \lambda\mu f(x_0, x_0). \end{cases}$$

By picking $\mu = 1$ and suitable λ, we can conclude $f(x, y) = f(y, x)$. The proof is complete.

18.7 We can find invertible matrix P over $\mathbb{Q}$ such that

$$PAP^{\mathsf{t}} = \operatorname{diag}\left(\begin{pmatrix} & 1 \\ -1 & \end{pmatrix}, \dots, \begin{pmatrix} & 1 \\ -1 & \end{pmatrix}, 0, \cdots\right).$$

So we have $\det A \in \mathbb{Q}^2$. Note that $\mathbb{Q}^2 \cap \mathbb{Z} = \mathbb{Z}^2$.

18.15 For $x \in \ker(A-1)$, i.e. $Ax = x$, we have

$$\omega(x, Ay - y) = \omega(x, Ay) - \omega(x, y) = \omega(A^{-1}x, y) - \omega(x, y) = 0.$$

So $\ker(A-1) \subseteq \operatorname{im}(A-1)^{\perp}$. By dimension reason, they are actually equal.

19 Invariant Theory

In linear algebra, there are different sorts of equivalence relations. The purpose of this section is to study the **(continuous, polynomial) invariants** of the relations. To be exact, for an equivalence relation $\sim$ over some matrix space X, a **(continuous, polynomial) invariant** is a (continuous, polynomial) function $f : X \to \mathbb{R}$ or $\mathbb{C}$ such that if

$$A \sim B \quad \Longrightarrow \quad f(A) = f(B).$$

In terms of set theory, we are actually studying the space of functions over the quotient set $X/\sim$.

$$\mathsf{Fun}(X/\sim, \mathbb{R}) \quad \text{or} \quad \mathsf{Fun}(X/\sim, \mathbb{C}).$$

Continuous invariants and polynomial invariants correspond to quotient space and GIT quotient in topology and algebraic geometry.

Note that the space of invariants forms a ring, i.e. constant maps are invariants and

$$f \text{ and } g \text{ are invariants} \quad \Longrightarrow \quad f+g \text{ and } fg \text{ are invariants.}$$

Similarly, continuous, polynomial invariants form subrings.

Equivalent matrices Recall that two matrices in $\mathbb{M}_{n\times m}(\mathbb{C})$ are **equivalent** if

$$A \sim B \iff \exists P \in \mathrm{GL}_n(\mathbb{C}), Q \in \mathrm{GL}_m(\mathbb{C}), \text{ such that } A = PBQ.$$

Note that if $A \sim B$, then $\mathrm{rank}(A) = \mathrm{rank}(B)$. So rank provides an invariant.

Proposition 19.1 Any invariant of $\sim$ is a function of rank.

— *Proof.* Let f be an invariant. Since $A \sim B$ if and only if $\mathrm{rank}(A) = \mathrm{rank}(B)$, so that $f(A)$ is only determined by $\mathrm{rank}(A)$. So f is actually a function of rank. □

Proposition 19.2 Any continuous invariant of $\sim$ is a constant.

— *Proof.* Let f be a continuous invariant. Note that the space of matrices of full rank is dense, so f is a constant. □

Congruent matrices Recall that two (anti-)symmetric matrices of size n are **congruent** if

$$A \sim B \iff \exists P \in \mathrm{GL}_n(\mathbb{R}), \text{ such that } A = PBP^{\mathrm{t}}.$$

▶ **Exercise 19.3** Show that invariants of $\sim$ over symmetric matrices are functions of the positive and negative indices of inertia.

▶ **Exercise 19.4** Show that invariants of $\sim$ over anti-symmetric matrices are functions of rank.

Left Equivalent matrices Recall that two matrices in $\mathbb{M}_{n\times m}(\mathbb{C})$ are **left equivalent** if

$$A \sim B \iff \exists P \in \mathrm{GL}_n(\mathbb{C}), \text{ such that } A = PB.$$

It is not quite possible to find the full list of invariants. So we turn to continuous and polynomial invariants. As above, to study continuous invariants, we only need to be concerned about the space of matrices of full ranks $\min(n, m)$.

Remark 19.5 There are two ways to understand this relation geometrically.

(1) Let the column vectors of B be $b_1, \dots, b_m$. The left multiplication by $P \in \mathrm{GL}_n(\mathbb{C})$ moves them together inside $\mathbb{C}^n$. So the equivalence classes are nothing but the "configurations" of m-vectors in $\mathbb{C}^n$, say realizable matroids.

(2) Let the row vectors of B is $x_1, \dots, x_n$. If they are linearly independent, then the space $\mathrm{span}(x_1, \dots, x_n)$ is invariant under the left action of $P \in \mathrm{GL}_n(\mathbb{C})$. So the equivalence classes are nothing but the subspaces of dimension n in $\mathbb{C}^m$.

▶ **Problem 19.6** If $n > m$, then any continuous invariants of $\sim$ are constant. ◀ P97

Definition 19.7 Assume $n \leq m$. We call

$$\mathrm{St}(n, m) = \{A \in \mathbb{M}_{n\times m}(\mathbb{C}) : \mathrm{rank}\, A = m\}$$

the **Stiefel varierty**. We denote

$$\mathrm{Gr}(n, m) = \{V \subseteq \mathbb{C}^m : \dim V = n\}$$

the **Grassmannian varierty**.

Theorem 19.8 By the discussion above, when $n \le m$, we have

$$\mathrm{Gr}(n,m) = {}_{\mathrm{GL}_n(\mathbb{C})}\backslash \mathrm{St}(n,m).$$

In particular,

- any continuous invariant of $\sim$ is a continuous function over the Grassmannian variety;
- any polynomial invariants of $\sim$ is a constant (since $\mathrm{Gr}(n,m)$ is a projective variety).

Remark 19.9 From above, the polynomial invariants are not interesting for $\sim$. Actually, instead of considering polynomials f such that

$$f(PA) = f(A) \qquad \text{for all } A,$$

we can consider the **twisted invariant**

$$f(PA) = \det(P)^d f(A) \qquad \text{for all } A.$$

For example, when $d = 1$, any such f is spanned by $\binom{m}{n}$ minors of size $n \times n$. Note that

$$\bigoplus_d \left\{ f : {}^{\forall A}\, f(PA) = (\det P)^d f(A) \right\}$$

is a graded ring. Actually, it defines the projective coordinate of $\mathrm{Gr}(n,m)$ under the Plücker embedding.

Similar matrices Recall that two matrices in $\mathbb{M}_n(\mathbb{C})$ are **similar** if

$$A \sim B \iff \exists P \in \mathrm{GL}_n(\mathbb{C}), \text{ such that } A = PBP^{-1}.$$

Note that $\det(PAP^{-1}) = \det(A)$. So determinant is a polynomial invariant. Similarly, trace is also a polynomial invariant. More generally, we consider

$$\chi_A(t) = \det(t1 - A) = t^n - (\operatorname{tr} A)t^{n-1} + \cdots + (-1)^n \det A.$$

The coefficients are all polynomial invariants.

▶ **Problem 19.10** Show that the coefficients of t^{n-i} in $\chi_A(t)$ is

$$(-1)^i \sum \det(X)$$

with the sum over all principle minors X of size i.

Let $\lambda_1, \dots, \lambda_n$ be eigenvalues of A (counting multiplicity). Then we have

$$\chi_A(t) = \prod (t - \lambda_i).$$

Theorem 19.11 Any continuous invariants of $\sim$ are functions of coefficients of characteristic polynomial.

— *Proof.* Note any invariant f is determined by its values at Jordan canonical forms. Note that

$$\lim_{t \to 0} \begin{pmatrix} \lambda & t \\ & \lambda \end{pmatrix} \to \begin{pmatrix} \lambda & \\ & \lambda \end{pmatrix}$$

where the Jordan canonical forms of $\begin{pmatrix} \lambda & t \\ & \lambda \end{pmatrix}$ are $\begin{pmatrix} \lambda & 1 \\ & \lambda \end{pmatrix}$ whenever $t \neq 0$. If we assume f is continuous, then f is determined by its values at diagonal matrices. Note that two diagonal matrices are similar if and only if the values on the diagonals differ by a permutation. That is, they have the same characteristic polynomial. So any continuous invariants of $\sim$ are functions of coefficients of the characteristic polynomial. □

Theorem 19.12 Any polynomial invariants of $\sim$ are polynomials of coefficients of characteristic polynomial.

— *Proof.* The restriction of a polynomial to the space of diagonal matrices is still a polynomial. Note that any polynomial is represented by $f(x_1, \dots, x_n)$ for $\operatorname{diag}(x_1, \dots, x_n) \in \mathbb{M}_n(\mathbb{C})$. Since f is a symmetric polynomial, it is a polynomial of coefficients of the characteristic polynomial. □

Hints

19.6 Now the rank of B is m, i.e. $b_1, \dots, b_m$ are linearly independent, we can move them to a part of the standard basis.

20 Division algebras

The main question in this section is the requirement on n such that

> for any vector space V of dimension n, there exists linear transformations $A_1, \dots, A_n$ such that for any nonzero vector $v \in V$, $A_1 v, \dots, A_n v$ forms a basis of V. (∗)

Here is some example in low dimensions.

- When $n = 1$, we can take $A_1 = \mathrm{id}$;
- When $n = 2$ and over $\mathbb{R}$ we can take $A_1 = \mathrm{id}$ and $A_2 = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. Actually,

$$\det\left(A_1\tbinom{x}{y}, A_2\tbinom{x}{y}\right) = \det\left(\begin{smallmatrix} x & -y \\ y & x \end{smallmatrix}\right) = x^2 + y^2 = 0 \iff x = y = 0.$$

- When $n = 3$ and over $\mathbb{R}$, we can assume $A_1 = \mathrm{id}$ without loss of generality. Note that as a 3×3 matrix A_2 must admits a real eigenvalue and a real eigenvector v. Thus $v = A_1 v$ and $A_2 v$ cannot be always linearly independent.
- When $n = 4$, insprited from the case $n = 4$, we first notice that

$$\det \begin{pmatrix} x & -y & -z & -w \\ y & x & w & -z \\ z & -w & x & y \\ w & z & -y & x \end{pmatrix} = (x^2 + y^2 + z^2 + w^2)^2.$$

  Thus we take $A_i(x, y, z, w) =$ the i-th colume.

▶ **Problem 20.1** The condition (∗) is not satisfied for $n = 2$ over $\mathbb{C}$. ◀ P100

▶ **Problem 20.2** The condition (∗) is not satisfied for odd $n > 1$ over $\mathbb{R}$.

▶ **Problem 20.3** The condition (∗) is not satisfied for any $n > 1$ over $\mathbb{C}$.

▶ **Problem 20.4** Recall that we can identify

$$\mathbb{C} \xrightarrow{\subset} \mathbb{M}_2(\mathbb{R}), \qquad x + yi \longmapsto \left(\begin{smallmatrix} x & -y \\ y & x \end{smallmatrix}\right).$$

Using this trick to prove the determinant above is $(x^2 + y^2 + z^2 + w^2)^2$. ◀ P100

As reader may observe, these determinants are related to **quaternion algebra**. Recall that

$$\mathbb{H} = \mathbb{R} \oplus \mathbb{R}\mathbf{i} \oplus \mathbb{R}\mathbf{j} \oplus \mathbb{R}\mathbf{k}$$

The relations are

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1$$
$$\mathbf{ij} = \mathbf{k},\ \mathbf{jk} = \mathbf{i},\ \mathbf{ki} = \mathbf{j},$$
$$\mathbf{ji} = -\mathbf{k},\ \mathbf{kj} = -\mathbf{i},\ \mathbf{ik} = -\mathbf{j}.$$

Let us denote

$$\overline{x + y\mathbf{i} + z\mathbf{j} + w\mathbf{k}} = x - y\mathbf{i} - z\mathbf{j} - w\mathbf{k}.$$

We have

$$(x + y\mathbf{i} + z\mathbf{j} + w\mathbf{k})(\overline{x + y\mathbf{i} + z\mathbf{j} + w\mathbf{k}}) = x^2 + y^2 + z^2 + w^2.$$

Example 20.5 Now let us restate the construction for $n = 4$ over $\mathbb{R}$. We take $V = \mathbb{H}$, and the linear transformations are $A_1 = \mathrm{id}$, $A_2(z) = \mathbf{i}z$, $A_3(z) = \mathbf{j}z$, $A_4(z) = \mathbf{k}z$. For any nonzero z, if $\lambda_1 A_1(z) + \cdots + \lambda_4 A_4(z) = 0$ for some real numbers $\lambda_1, \ldots, \lambda_4$, i.e.

$$(\lambda_1 + \lambda_2\mathbf{i} + \lambda_3\mathbf{j} + \lambda_4\mathbf{k})z = 0$$

We can multiply $\overline{z}$ and cancel $\overline{z}z \neq 0$ to conclude $\lambda_1 + \lambda_2\mathbf{i} + \lambda_3\mathbf{j} + \lambda_4\mathbf{k} = 0$. Thus $\lambda_1 = \cdots = \lambda_4 = 0$

▶ **Problem 20.6** If $\mathbb{F}$ admits a field extension of degree d, then $(*)$ is possible for $n = d$.

Example 20.7 Thank to the existence of **octonion algebra** $\mathbb{O}$, we can get a construction for $n = 8$ over $\mathbb{R}$. Note that $\mathbb{O}$ is no longer associative, but it still satisfies $(xy)\overline{y} = x(y\overline{y})$. So all the steps are the same as above.

Theorem 20.8 Over $\mathbb{R}$, the condition $(*)$ is possible if and only if when $n = 1, 2, 4, 8$.

— *Proof.* Without loss of generality, we can take $A_1 = \mathrm{id}$. Let us consider the sphere $S^{n-1} \subset \mathbb{R}^n$. For each $v \in S^{n-1}$, we can project $A_2 v, \ldots, A_n v$ to the tangent hyperplane at v (i.e. $\mathrm{span}(v)^\perp$). This defines a trivialization of the tangent bundle. By Adams theorem, S^{n-1} has trivial tangent bundle (parallelizable) if and only if $n = 1, 2, 4, 8$. Since we have constructed the answer for $n = 1, 2, 4, 8$, this concludes the theorem. □

Actually, the maximal number r of matrices $A_1, \ldots, A_r$ such that $A_1 v, \ldots, A_r v$ are linearly independent for any nonzero v is known as the **Hurwitz–Radon number**.

Hints

20.1 Note that $\det \left(\begin{smallmatrix} x & ax+by \\ y & cx+dy \end{smallmatrix}\right) = 0$ cannot have only one solution over $\mathbb{C}$.

20.4 $\det \left(\begin{smallmatrix} x+yi & -z+wi \\ z+wi & x-yi \end{smallmatrix}\right) = x^2+y^2+z^2+w^2$. Note that we have the following commutative diagram

$$\begin{array}{ccc} \mathbb{M}_n(\mathbb{C}) & \xrightarrow{\subset} & \mathbb{M}_{2n}(\mathbb{R}) \\ \Big\downarrow{\scriptstyle \det} & & \Big\downarrow{\scriptstyle \det} \\ \mathbb{C} & \xrightarrow{|\cdot|^2} & \mathbb{R}. \end{array}$$