Math 55a - Honors Abstract Algebra

Taughy by Yum-Tong Siu
Notes by Dongryul Kim

Fall 2015

The course was taught by Professor Yum-Tong Siu. We met twice a week, on Tuesdays and Thursdays from 2:30 to 4:00. At the first lecture there were over 30 people, but at the end of the add-drop period, the class consisted of 11 students. There was an in-class midterm exam and a short take-home final. The course assistants were Calvin Deng and Vikram Sundar.

Contents

1 September 3, 2015

1.1 Overview

My name is Siu([see-you]), and there are no textbooks for this course. The website for this course is `http://math.harvard.edu/~siu/math55a`. There will be no clear division between abstract algebra and analysis.

These are the things I will tell you during the lectures.

- Motivation, background, and history for the material
- Techniques, methods, ideas, and structures
- "Rigorous" presentation

I will emphasize the last one, but it is useless to only know rigorous things.

There will be weekly problem sets, and we encourage discussions. And of course, you need to write the solutions down in you own words.

The actual level of difficulty will depend on the feedback I get from your assignments.

1.2 Things we will cover

We focus on solving equations in a number system. There are two kinds of equations:

- polynomial equations - This is algebra, and will be the A part
- differential equations - This is real and complex analysis and will be covered in the B part

We start with Peano's five axioms, and from this, we can define $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$. You can choose what number system you would like work in, and this is why number systems are important. For instance, the fundamental theorem of algebra holds in $\mathbb{C}$, but does not hold in $\mathbb{R}$ or $\mathbb{Q}$.

Historically, the whole algebra came from solving polynomial equations. There are symmetry involved in solving equations. For instance, if

$$(x-a_1)\cdots(x-a_n) = x^n - \sigma_1 x_{n-1} + \sigma_2 x_{n-1} - \cdots,$$

we get $\sigma_1 = a_1 + \cdots + a_n$, $\sigma_2 = a_1a_2 + \cdots + a_{n-1}a_n$. The coefficients have symmetry between the a_is. So basically solving a polynomial equation is bringing all-symmetry down to no-symmetry. This is basically what Galois did, but by going down steps of partial symmetry.

1.3 Peano's axioms

I want to start from:

Russell's paradox. *Consider the set A of all sets which do not belong to themselves. Is $A \in A$?*

In any way, you get a contradiction. The problem start from the fact that you should know all sets before defining the set of all set, which is absurd. To put it another way, it is circular. To overcome it, you may use the theory of types (which I will not go into explaining). The point is that you need to be very rigorous when doing axiomatic stuff.

Peano's five axioms.

1. *(Non-emptiness) There exists* $1 \in \mathbb{N}$.
2. *(Successor map) There exists a map* $' : \mathbb{N} \to \mathbb{N}$ *called the successor map which sends* $x \mapsto x'$.
3. *(Special element)* $x' \neq 1$ *for any* $x \in \mathbb{N}$.
4. *(Infectivity of successor map)* $x' = y'$ *implies* $x = y$.
5. *(Induction axiom) If a subset* $A \subset \mathbb{N}$ *contains* 1, *and* $x' \in A$ *for any* $x \in A$, *then* $A = \mathbb{N}$.

I'll give two examples of proof using the axioms as a warm-up.

Proposition 1.1. *There exist no fixed point for the successor map.*

Proof. We make use of the induction axiom. Let

$$A = \{x \in \mathbb{N} : x \neq x'\}.$$

First, $1 \in A$ since $1 \neq x'$ for any $x \in A$. Next we need to prove that $x \in A$ implies $x' \in A$. If $x \in A$, by definition $x \neq x'$. And this implies $x' \neq x''$ because of injectivity of the successor map. Lastly, using the induction axiom, we get $A = \mathbb{N}$. □

Proposition 1.2. *The image of the successor map is* $\mathbb{N} \setminus \{1\}$.

Proof. To use the induction axiom, we add 1 to the set we are interested in. Let

$$A = \{1\} \cup \{x' : x \in A\}.$$

$1 \in A$ is clear. Also, if $x \in A$, then $x \in \mathbb{N}$ and thus $x' \in A$. □

The amazing thing about Peano's axioms is that you can get addition from them. Let us define addition and multiplication.

Because we have two variables, we first fix the first variable.

Definition 1.3 (Addition)**.** Define $x + 1 = x'$. Suppose we have defined $x + y$. Then we define $x + y' = (x + y)'$. By induction axiom, we have defined $x + y$ for all x and y.

Theorem 1.4. *Addition is associative.*

Proof. Let us prove $(x+y)+z = x+(y+z)$. First fix x and y. Let

$$A_{x,y} = \{z \in \mathbb{N} : (x+y)+z = x+(y+z)\}.$$

First $1 \in A_{x,y}$ since

$$(x+y)+1 = (x+y)' = x+(y') = x+(y+1).$$

Also if $z \in A_{x,y}$, then

$$(x+y)+z' = ((x+y)+z)' = (x+(y+z))' = x+(y+z)' = x+(y+z').$$

Thus $z \in A$ implies $z' \in A$, and it follows that $A_{x,y} = \mathbb{N}$. □

Theorem 1.5. *Addition is commutative.*

Proof. We want to prove $x+y = y+x$. Fix $y \in \mathbb{N}$. Let

$$A_y = \{x \in A : x+y = y+x\}.$$

The first thing we need to prove is $1 \in A_y$, which is $1+y = y+1$. We use another induction inside this induction.

Let

$$B = \{y \in \mathbb{N} : 1+y = y+1\}.$$

Obviously $1 \in B$, and $y \in B$ implies

$$1+y' = (1+y)' = (y+1)' = y+(1+1) = (y+1)+1 = y'+1,$$

which in turn, implies $y' \in B$. Thus $1+y = y+1$ for all $y \in \mathbb{N}$.

Now suppose that $x \in A_y$. For x', we have

$$y+x' = (y+x)' = (x+y)' = x+y' = x+(1+y) = (x+1)+y = x'+y.$$

Thus $A_y = \mathbb{N}$. □

Now let us define multiplication.

Definition 1.6 (Multiplication)**.** Let $x \cdot 1 = x$ and $x \cdot y' = x \cdot y + x$. This defines multiplication in general because of the induction axiom.

Theorem 1.7. *For any $x, y, z \in \mathbb{N}$, we have the following:*
(a) $x \cdot y = y \cdot x$
(b) $(x \cdot y) \cdot z = x \cdot (y \cdot z)$
(c) $x \cdot (y+z) = x \cdot y + x \cdot z$

Proof. Homework. □

1.4 Rational numbers

Now we begin to handle division. We construct the set $\mathbb{Q}_+$ of all positive fractions (or positive rational numbers). But first we need the concept of equivalence relations, because we need to say $\frac{a}{b} = \frac{c}{d}$ if $ad = bc$.

Definition 1.8. Let X be a set. A **relation** in X is a subset $\mathcal{R} \subset X \times X$. We use the notation $a \sim b$ to mean $(a, b) \in \mathcal{R}$. The relation $\mathcal{R}$ (also denoted by $\sim$) is an **equivalence relation** if

(Reflexivity) $x \sim x$ for all $x \in X$,
(Symmetry) $x \sim y$ if and only if $y \sim x$,
(Transitivity) $x \sim y$ and $y \sim z$ imply $x \sim z$.

Theorem 1.9 (Decomposition statement)**.** *An equivalence relation divides up X into a disjoint union of subsets.*

Proof. For $x \in X$, let $X_x = \{y \in X : y \sim x\}$, known as the equivalence class which contains x. It is clear that

$$X = \bigcup_{x \in X} X_x.$$

We also need to show that what we have is a disjoint union in the following sense:

$$X_x \cap X_y \neq \emptyset \text{ implies } X_x = X_y.$$

Because of symmetry, it is sufficient to show $X_x \subset X_y$. By assumption there exists an element $z \in X_x \cap X_y$, and we get $z \sim x$ and $z \sim y$. Take any $u \in X_x$. Because $u \sim x$, $x \sim z$ and $z \sim y$, we have $u \sim y$. This shows $u \in X_y$. □

Now we finally define $\mathbb{Q}_+$, the set of positive rational numbers.

Definition 1.10. Introduce $\sim$ in $X = \mathbb{N} \times \mathbb{N}$ such that $(a, b) = (c, d)$ if and only if $ad = bc$. We call the equivalence classes $\mathbb{Q}_+$.

You can check that it actually is a equivalence relation.

Next class, we will define $\mathbb{R}_+$ by Dedekind cuts. We have to go into the realm of analysis to define the reals, because we need the mean-value property. For instance, let me sketch a proof of the fundamental theorem of algebra.

Let

$$P(z) = z^n + \sum_{j=0}^{n-1} a_j z^j$$

be a monic polynomial with complex variables and no roots. Let $f(z) = 1/P(z)$. Then by certain facts in complex analysis,

$$f(c) = \frac{1}{2\pi} \int_{\theta=0}^{2\pi} f(c + re^{i\theta}) d\theta$$

and

$$|f(c)| \leq \frac{1}{2\pi} \int_{\theta=0}^{2\pi} |f(c + re^{i\theta})| d\theta.$$

Sending $r \to \infty$, we get a contradiction.

2 September 8, 2015

The CAs would like to get the solutions for the problem sets typed in LaTeX.

Last time we studied the five Peano axiom, and defined $\mathbb{N}$, addition, multiplication, and $\mathbb{Q}_+$. Now we will introduce $\mathbb{R}_+$, and lastly $\mathbb{C}$.

2.1 Non-rigorous proof of the fundamental theorem of algebra

We do we need Dedekind cuts to make sure every polynomial equation is solvable? Last class, I said that the crucial thing was the process of averaging. Let me prove the fundamental theorem of algebra in more detail, but not in a rigorous way.

Theorem 2.1. *For any*

$$P(z) = z^n + \sum_{j=0}^{n-1} a_j z^j$$

with $a_j \in \mathbb{C}$ and $n \geq 1$, there exists a $z_0 \in \mathbb{C}$ such that $P(z_0) = 0$.

Assume $P(z) \neq 0$ for any $z \in \mathbb{C}$. Then $f(z) = 1/P(z)$ is well-defined on $\mathbb{C}$. Obviously $|f(z)| \to 0$ as $|z| \to \infty$ since $n \geq 1$. The limit of the difference quotient exists for $f(z) = 1/P(z)$ in the setting of complex numbers.

For a real-valued function $g(x)$ of a real variable $x \in \mathbb{R}$, the difference quotient (or quotient of difference) is defined as

$$\lim_{x \to x_0} \frac{g(x) - g(x_0)}{x - x_0} = g'(x_0).$$

We have not defined what a limit is, but let us just assume that we know this.

In the complex numbers, differentiation is defined similarly. A complex-valued function $f(z)$ of a complex variable $z \in \mathbb{C}$. We say that the complex derivative exists if

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} = f'(z_0).$$

This looks similar to the differentiablity of real functions, but it is much stronger in a sense that there are many ways to approach a single point. If $z : \mathbb{R} \to \mathbb{C}$ is a curve passing a point z_0 at time t_0, we have

$$\frac{g(t) - g(t_0)}{t - t_0} = \frac{f(z(t)) - f(z(t_0))}{z(t) - z(t_0)} \cdot \frac{z(t) - z(t_0)}{t - t_0}$$

where $g(t) = f(z(t))$. Then we have

$$g'(t_0) = f'(z_0) \left(\frac{dz}{dt} \right)_{t=t_0}.$$

Because $f'(z_0)$ is independent of the curve, we get a degree of freedom.

As we have set $f(z) = 1/P(z)$, we have

$$f'(z_0) = -\frac{nz_0^{n-1} + \sum_{j=1}^{n-1} a_j j z_0^{j-1}}{P(z_0)^2}.$$

There exists a derivative of f under the assumption that P has no zeros, although we have not proved it yet.

Now the mean value property states that

$$f(z_0) = \text{average of } f(z) \text{ at the circle centered at } z_0 \text{ of radius } r > 0.$$

This can be deduced from the chain rule.

Proof of the mean value property. Analytically it can be written down as

$$f(z_0) = \frac{1}{2\pi} \int_{\theta=0}^{2\pi} f(z_0 + re^{i\theta}) d\theta.$$

I have not defined $e^{i\theta}$ yet, but $e^{i\theta} = \cos\theta + i\sin\theta$. We consider the map

$$r \mapsto \frac{1}{2\pi} \int_{\theta=0}^{2\pi} f(z_0 + re^{i\theta}) d\theta.$$

If we prove that the derivative is always 0, and that the limit when $r \to 0^+$ is $f(z_0)$, we have proven the formula. The latter is immediate since $f(z_0 + re^{i\theta}) \to f(z_0)$ as $r \to 0^+$. Note that this is possible because any curve in the complex plane can can be retracted to a point. So we prove the former. We will apply the chain rule to two different curve; the line going through the origin, and the circle.

First, we have

$$\frac{d}{dr} \int_{\theta=0}^{2\pi} f(z_0 + re^{i\theta}) d\theta = \int_{\theta=0}^{2\pi} \left(\frac{\partial}{\partial r} f(z_0 + re^{i\theta}) \right) d\theta.$$

Looking in the radial direction, we obtain

$$\begin{aligned} \frac{\partial}{\partial r} f(z_0 + re^{i\theta})\Big|_{r=r_0} &= \lim_{r \to r_0} \frac{f(z_0 + re^{i\theta}) - f(z_0 + r_0 e^{i\theta})}{r - r_0} \\ &= \lim_{r \to r_0} \frac{f(z_0 + re^{i\theta}) - f(z_0 + r_0 e^{i\theta})}{(z_0 + re^{i\theta}) - (z_0 + r_0 e^{i\theta})} \cdot \frac{(z_0 + re^{i\theta}) - (z_0 + r_0 e^{i\theta})}{r - r_0} \\ &= f'(z_0 + r_0 e^{i\theta}) e^{i\theta}, \end{aligned}$$

and thus

$$\int_{\theta=0}^{2\pi} \left(\frac{\partial}{\partial r} f(z_0 + re^{i\theta} \right) d\theta = \int_{\theta=0}^{2\pi} e^{i\theta} f'(z_0 + re^{i\theta}) d\theta.$$

To calculate $f'(z_0 + re^{i\theta})$, we do the same thing over again. Looking at the circle, we get

$$\begin{aligned}\frac{\partial}{\partial\theta} f(z_0 + re^{i\theta})\Big|_{\theta=\theta_0} &= \lim_{\theta\to\theta_0} \frac{f(z_0 + re^{i\theta}) - f(z_0 + re^{i\theta_0})}{\theta - \theta_0} \\ &= \lim_{\theta\to\theta_0} \frac{f(z_0 + re^{i\theta}) - f(z_0 + re^{i\theta_0})}{(z_0 + re^{i\theta}) - (z_0 + re^{i\theta_0})} \cdot \frac{(z_0 + re^{i\theta}) - (z_0 + re^{i\theta_0})}{\theta - \theta_0} \\ &= f'(z_0 + re^{i\theta_0}) rie^{i\theta}\end{aligned}$$

Therefore

$$\begin{aligned}\int_{\theta=0}^{2\pi} e^{i\theta} f'(z_0 + re^{i\theta}) d\theta &= \int_{\theta=0}^{2\pi} \frac{1}{ri} \left(\frac{\partial}{\partial\theta} f(z_0 + re^{i\theta}) \right) d\theta \\ &= \frac{1}{ri} f(z_0 + re^{i\theta})\Big|_{\theta=0}^{2\pi} = 0.\end{aligned}$$

□

Proof of the fundamental theorem of algbera. Take any $z_0 \in \mathbb{C}$. Since

$$f(z_0) = \frac{1}{2\pi} \int_{\theta=0}^{2\pi} f(z_0 + re^{i\theta}) d\theta$$

and the $f(z_0 + re^{i\theta})$ goes to 0 as $r \to \infty$, we get

$$|f(z_0)| \le \frac{1}{2\pi} \int_{\theta=0}^{2\pi} |f(z_0 + re^{i\theta})| d\theta = 0$$

which contradicts $f(z_0) = 1/P(z_0) \neq 0$. □

As you can see, analysis is needed to prove a theorem in algebra.

We needed two things; first is the notion of averaging which is same as integrals, and the two-dimensional situation which makes it possible to consider multiple directions.

2.2 Order relations

Back to rigorous presentations. Let us define upper bounds, and the least upper bound. But first we need to define what $x < y$ or $x \le y$ means.

Definition 2.2. Let $x, y \in \mathbb{N}$. We say that $x > y$ if and only if there exists a $u \in \mathbb{N}$ such that $x = y + u$. Let $x < y$ if and only if $y > x$.

Theorem 2.3 (Trichotomy)**.** *For any $x, y \in \mathbb{N}$, precisely one of the following three statements holds.*

$$x = y, \qquad x > y, \qquad x < y$$

The key point in the proof is that there are no fixed points in the addition operation. In other words, for any fixed $x \in \mathbb{N}$, we have $y \neq x + y$ for any $y \in \mathbb{N}$.

Proof. First, fix x and let

$$A_x = \{y \in \mathbb{N} : y \neq x + y\}.$$

$1 \in A_x$ since 1 is not a successor. And $y \neq x + y$ impies $y' \neq x + y'$ by the injectivity of the successor function. Thus $A_x = \mathbb{N}$, and addition has no fixed points.

Now let us go back to the trichotomy. We need to prove two things, namely exclusivity and inclusivity. Actually exclusivity is immediately verified by the fact that addition has no fixed proof. For instance, if $x > y$ and $y > x$, we get $x = y + a$ and $y = x + b$, and thus $x = x + (a + b)$.

For inclusivity, we fix $x \in \mathbb{N}$ and let

$$B_x = \{y \in \mathbb{N} : \text{either } x = y \text{ or } x > y \text{ or } x < y\}.$$

One can prove $1 \in B_x$ by dividing into cases $x = 1$ and $x \neq 1$. If $x = 1$, then $1 = 1$. If $x \neq 1$, there exists a u for which $x = u' = 1 + u$. Then $x > 1$.

Assume $y \in A_x$. Now we have three cases. If $y = x$, then $y' = x + 1 > x$. If $y < x$, there exists a u for which $x = y + u$. If $u = 1$, we have $x = y'$, and if $u \neq 1$, we have $x = y + u = y + v' = y' + v$ for some $v \in \mathbb{N}$. If $y > x$, there exists a u for which $y = x + u$, and then $y' = x + u'$. Therefore $B_x = \mathbb{N}$. □

As we have introduced ordering in $\mathbb{N}$, we can extend this to $\mathbb{Q}_+$.

Definition 2.4. We say

$$\frac{a}{b} > \frac{c}{d}$$

if and only if $ad > bc$.

You can check the ordering is well-defined.

2.3 Dedekind cuts

Definition 2.5. An **upper bound** of a subset $A \subset Q_+$ is a number U such that $x \leq U$ for any $x \in A$. A **least upper bound** means an upper bound lub such that $lub \leq U$ for any upper bound U.

We want to make every set which has an upper bound admits a least upper bound. Dedekind used some good logic to make this true.

To define $\sqrt{2}$, you look at the set

$$\left\{\frac{a}{b} \in \mathbb{Q}_+ : \frac{a^2}{b^2} < 2\right\}.$$

This set doesn't have a least upper bound, but we want it to exist. So you just throw the number in. It doesn't cost money. You just consider any set of the form

$$\left\{\frac{a}{b} \in \mathbb{Q}_+ : \frac{a}{b} < \xi\right\}$$

as a real number.

Definition 2.6. A **Dedekind cut** is a proper subset ξ of $\mathbb{Q}_+$ such that

1. (containing all numbers less than some non-member) For any $x \in \xi$ and $y \in \mathbb{Q}_+ \setminus \xi$, we have $x < y$.
2. (containing no upper bound) There does not exist an $x \in \xi$ such that $x \geq y$ for all $y \in \xi$.

Definition 2.7. The (positive) **real numbers** is defined as

$$\mathbb{R}_+ = \{\text{all Dedekind cuts}\}.$$

We can embed $\mathbb{Q}_+$ into $\mathbb{R}_+$ according to the map

$$r \in \mathbb{Q}_+ \mapsto \xi_r = \{s \in \mathbb{Q}_+ : s < r\}.$$

We can also easily define ordering, addition, and multiplication on the real numbers.

Definition 2.8. Let ξ and η be two distinct Dedekind cuts. Define $\xi > \eta$ if and only if $\xi \supset \eta$, and $\xi < \eta$ if and only if $\xi \subset a$. Also, define

$$\xi + \eta = \{x + y : x \in \xi, y \in \eta\}$$

and

$$\xi \cdot \eta = \{xy : x \in \xi, y \in \eta\}.$$

Now we can define $\mathbb{Q}$ as

$$\mathbb{Q} = (-\mathbb{Q}_+) \cup \{0\} \cup \mathbb{Q}_+$$

and also

$$\mathbb{R} = (-\mathbb{R}_+) \cup \{0\} \cup \mathbb{R}_+.$$

You can define addition, multiplication, ordering on these sets, but I am not going to do this, because I do not want to write a whole book.

Definition 2.9. The **complex numbers** is defined as the product $\mathbb{C} = \mathbb{R} \times \mathbb{R}$. The operations on the set are given as

$$(a, b)(c, d) = (ac - bd, ad + bc),$$

$$(a, b) + (c, d) = (a + c, b + d).$$

Letting $i = (0, 1)$, we get the notation we are used to.

In the first class, I said that we will be studying polynomial equations. There are two kinds of things we want to do.

- Single polynomial of a single variable - This is mainly Galois theory.
- System of line equations in several variables - We will be doing this to do Stokes' theorem.

Next time, we will discuss how to solve a polynomial equation with one variable.

3 September 10, 2015

Solving quadratic equations is easy. Given a equation $ax^2 + bx + c = 0$, we can solve it by "completing the squares".

$$a\Big(x + \frac{b}{2a}\Big)^2 = \frac{b^2 - 4ac}{4a},$$

and now if we take roots, we get the solution.

3.1 Scipione del Ferro's solution of the cubic equation

A general method of solving the cubic equation was first discovered by del Ferro. Let $F(X) = ax^3 + bx^2 + cx + d$. Imitating the quadratic case, one can translate the variable x by letting $x = t + \alpha$. For a good α, one can eliminate the second degree term and obtain

$$t^3 + pt + q = 0.$$

But this does not solve the equation.

So we try some other translation. Let $t = u + v$. Then

$$t^3 + pt + q = (u^3 + v^3) + (u + v)(3uv + p) + q.$$

Note that this is a polynomial of degree 3 over u. But we don't want to just see this as a polynomial over u, because it destroys the symmetry between u and v. Instead, we set $3uv + p = 0$. Then it is the same as

$$\begin{cases} u^3 + v^3 + q = 0 \\ 3uv + p = 0 \end{cases}$$

Note that $3uv + p = 0$ is the artificial relation, and $u^3 + v^3 + q = 0$ is the original equation. Cubing the second equation, we get $u^3v^3 = -p^3/27$, and then we get a quadratic equation

$$X^2 + qX - \frac{p^3}{27}$$

whose zeroes are u^3 and v^3. Then you get three solutions for each variable, and plugging each of the solutions, you finally get three solution pairs. This quadratic polynomial is called the resolvent.

3.2 Lagrange's idea

Lagrange saw this solution of del Ferro's and realized that actually what del Ferro had done was same as this.

Let $\epsilon = (-1 + \sqrt{3})i$ be the cubic root of unity. The main trick is just setting

$$x_1 = u + v, \quad x_2 = \epsilon u + \epsilon^2 v, \quad x_3 = \epsilon^2 u + \epsilon v.$$

Generally this is not possible since there are two variables and three equations, but because $x_1 + x_2 + x_3 = 0$, we can do this. If we solve the system of linear equations, we get

$$\begin{cases} u = \frac{1}{3}(x_1 + \epsilon x_2 + \epsilon^2 x_3) \\ v = \frac{1}{3}(x_1 + \epsilon^2 x_2 + \epsilon x_3) \end{cases}.$$

This is called the Lagrange's resolvent.

Solving an equation is basically given the elementary symmetric polynomials

$$\sigma_1 = x_1 + \cdots + x_n, \quad \sigma_2 = \sum_{i<j} x_i x_j, \quad \ldots, \quad \sigma_n = x_1 \cdots x_n,$$

describing each $x_1, \ldots, x_n$ in terms of these polynomials using radicals and rationals. When $x_1, \ldots, x_n$ are permuted, note that the symmetric polynomials are not changed. So solving the equation is the same as bring the whole symmetry to no symmetry in this sense.

Lagrange started to observe what happens to u and v when x_1, x_2, x_3 are permuted. Using a "ladder diagram", you see that all permutations are generated by (12) and (23). The permutation (12) acts on u and v as

$$u \mapsto \frac{1}{3}(x_2 + \epsilon x_1 + \epsilon^2 x_3) = \frac{\epsilon}{3}(x_1 + \epsilon^2 x_2 + \epsilon x_3) = \epsilon v, \quad v \mapsto \epsilon u.$$

Also, (23) acts as

$$u \mapsto v, \quad v \mapsto u.$$

To get rid of the ϵ, we consider the cube of u and v. Then we see that $u^3 + v^3$ and $u^3 v^3$ are both symmetric functions in terms of x_1, x_2, x_3. Hence we get a equation with lower degree.

Lagrange applied this idea to quartic equations. Quartic roots of 1 are $1, i, -1, -i$. Hence according to what previously did, we need to look at

$$\begin{aligned} &x_1 + x_2 + x_3 + x_4 \\ &x_1 + ix_2 - x_3 - ix_4 \\ &x_1 - x_2 + x_3 - x_4 \\ &x_1 - ix_2 - x_3 + ix_4. \end{aligned}$$

But just considering the third term $x_1 - x_2 + x_3 - x_4$, there are three possible outcomes when a permutation acts on $\{x_1, x_2, x_3, x_4\}$. So Lagrange just let

$$\begin{cases} y_0 = \frac{1}{2}(x_1 + x_2 + x_3 + x_4) \\ y_1 = \frac{1}{2}(x_1 - x_2 + x_3 - x_4) \\ y_2 = \frac{1}{2}(x_1 + x_2 - x_3 - x_4) \\ y_3 = \frac{1}{2}(x_1 - x_2 - x_3 + x_4). \end{cases}$$

Because any permutation acts by changing y_i to $\pm y_j$, the symmetric polynomials of y_1^2, y_2^2, y_3^2 are symmetric respect to x_1, x_2, x_3, x_4. Then you can calculate y_1, y_2, y_3 using the cubic formula, and subsequently, x_1, x_2, x_3, x_4.

Actually the quartic formula was first discovered by Ferrari. But this is not relevant with our topic, so I will go over it quickly. Starting with the equation $x^4 + ax^3 + bx^2 + cx + d = 0$, we change it to

$$x^2(x^2 + ax) = -bx^2 - cx - d,$$
$$\Big(x\Big(x + \frac{a}{2}\Big)\Big)^2 = \frac{1}{4}a^2x^2 - bx^2 - cx - d,$$
$$\Big(x^2 + \frac{a}{2}x\Big)^2 = \frac{1}{4}a^2x^2 - bx^2 - cx - d.$$

In the cubic formula, we introduced a generic translation $t = u + v$ and imposed an additional condition. We do this again. Translating $(x^2 + \frac{a}{2}x)$, we get

$$\Big(x^2 + \frac{a}{2}x + \frac{1}{2}y\Big)^2 = \Big(\frac{1}{4}a^2 - b + y\Big)x^2 + \Big(-cx + \frac{1}{2}ay\Big)x + \Big(-d + \frac{1}{4}y^2\Big).$$

Ferrari wanted to make the right-hand side a square of a polynomial, or in other words, make its discriminant zero. This condition in terms of y is a cubic equation. So it is possible to calculate y, and thus x by solving the corresponding quadratic equation.

3.3 Schematics for solving a polynomial equation

As I have said, solving a polynomial equation is performing on the coefficients of the polynomial equations (or symmetric functions $\sigma_1, \ldots, \sigma_n$) the operations of the form of rational functions and roots(radicals). Actually the roots is what destroys the symmetry, because you need to choose what roots you will use. We can drawing the schematic as:

$$\begin{array}{c} \sigma_1, \sigma_2, \ldots, \sigma_n \\ \downarrow \text{root-taking} \\ \tau_1^{(1)}, \tau_2^{(1)}, \ldots, \tau_{n_1}^{(1)} \\ \downarrow \\ \vdots \\ \downarrow \\ \tau_1^{(l)}, \tau_2^{(l)}, \ldots, \tau_{n_l}^{(l)} \\ \downarrow \\ x_1, \ldots, x_n \end{array}$$

Each "layer" actually represents the field of functions which share some specific symmetry. For instance the first layer is $\mathbb{C}(\sigma_1, \ldots, \sigma_n)$ which is the set of rational symmetric functions. In each step, we take roots to extend the set of functions.

Let us represent the process of solving a quadratic equation in this way.

$$\begin{array}{c} \mathbb{C}(\sigma_1, \sigma_2) \\ \downarrow \text{root-taking} \\ \mathbb{C}(x_1, x_2) = \mathbb{C}(\tau_1^{(1)}, \tau_2^{(1)}) \end{array}$$

Writing down the symmetry of each layer in terms of groups(you can just think this as a set of permutations for now), this is

$$\{1\} = G_1 \subset G_0 = S_2,$$

where S_n is the set of permutations on $\{1, 2, \ldots, n\}$ and 1 is the identity permutation.

The cubic equation has two steps.

$$\{1\} = G_2 \subset G_1 \subset G_0 = S_3$$

where $G_1 = \{1, (123), (132)\}$ is the alternating group. It can be drawn as

$$\begin{array}{c} \mathbb{C}(\sigma_1, \sigma_2, \sigma_3) \\ \downarrow \\ \mathbb{C}(\tau_1^{(1)}, \tau_2^{(1)}, \tau_3^{(1)}, \tau_4^{(1)}) \\ \downarrow \\ \mathbb{C}(x_1, x_2, x_3, x_4) \end{array}$$

where

$$\tau_1^{(1)} = \sigma_1, \quad \tau_2^{(1)} = y_2 y_3, \quad \tau_3^{(1)} = y_2^3, \quad \tau_4^{(1)} = y_3^3.$$

The schematic for solving the quartic equation can be drawn as

$$\{1\} \subset K_4 \subset A_4 \subset S_4$$

where A_4 is the alternating group and $K_4 = \{1, (12)(34), (13)(24), (14)(23)\}$ is the Klein four-group. This diagram is not the solution itself; it is more of a reverse engineering kind of thing that shows us how complete symmetry was brought down to no symmetry in each of the cases.

4 September 15, 2015

4.1 More on solving polynomial equations

Last class, I have explained the schematics of solving a polynomial. Galois was interested in the possibility of such a scheme.

$$\begin{array}{c} \sigma_1, \dots, \sigma_n \\ \downarrow \\ \tau_1^{(1)}, \dots, \tau_{n_1}^{(1)} \\ \downarrow \\ \vdots \\ \downarrow \\ x_1, \dots, x_n \end{array}$$

As I has said, the process of taking rationals is not very important, because it does not destroy the symmetry. For instance, σ_1/σ_2 still possess whole symmetry. The important thing is roots, because it involves taking the root. In quadratic equations, the symmetry is destroyed when we consider $\sqrt{\sigma_1^2 - 4\sigma_2}$. This can be either $x_1 - x_2$ or $x_2 - x_1$, but because we don't really know what is what, we can go down to partial (or no) symmetry.

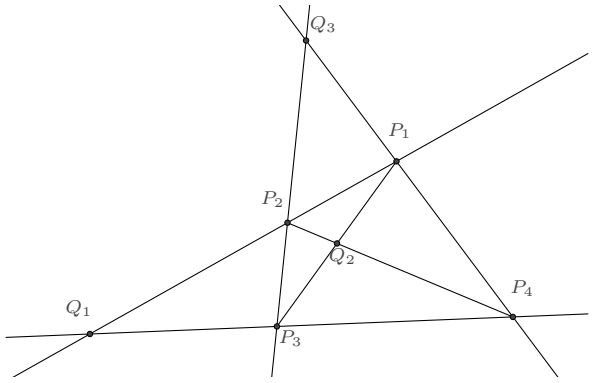
When we go down a step

$$\begin{array}{c} \mathbb{C}(\sigma_1, \dots, \sigma_n) = \mathbb{C}(\tau_1^{(0)}, \dots, \tau_{m_0}^{(0)}) \\ \downarrow \\ \mathbb{C}(\tau_1^{(1)}, \dots, \tau_{m_1}^{(1)}) \\ \downarrow \\ \vdots \\ \downarrow \\ \mathbb{C}(x_1, \dots, x_n) \end{array}$$

it is not clear what means by taking 'roots' of a polynomial. So instead, we consider it to be $(\tau_j^{(1)})^{\kappa_{1,j}} \in \mathbb{C}(\sigma_1, \dots, \sigma_n)$, that is, some power of an element in the lower step is in the upper step. Galois' theory states that in each step, $\mathbb{C}(\tau_1^{(j)}, \dots, \tau_{n_j}^{(j)})$ is equal to $\mathbb{C}(x_1, \dots, x_n)^{G_j}$ for some G_j, which is the subset of $\mathbb{C}(x_1, \dots, x_n)$ consisting of elements which are invariant under the action of G_j. So, basically, this schematics is equivalent to a 'tower' of groups

$$\{1\} \subset G_l \subset \cdots \subset G_0 = S_n.$$

There is a geometrical interpretation of the solution of quartic equation. Consider four points P_1, P_2, P_3, P_4 on the plane, and $Q_1 = P_1P_2 \cap P_3P_4$, $Q_2 = P_1P_3 \cap P_2P_4$, $Q_3 = P_1P_4 \cap P_2P_3$.



Because the set $\{Q_1, Q_2, Q_3\}$ is only permuted by the permutation of $\{P_1, P_2, P_3, P_4\}$, we can represent the elementary symmetric polynomials of Q_1, Q_2, Q_3 in terms of elementary symmetric polynomials of P_1, P_2, P_3, P_4. But it is not so simple as it looks, because the formula for Q_1, Q_2, Q_3 involves complex conjugates. What you need to do is just write down the formula, and take the part which does not involve any complex conjugates. Then because permutations does not change where conjugates are, you get a polynomial with no conjugates, and does not change after permutation.

4.2 Basic linear algebra

When we talked about groups, they were finite groups which lay in a symmetric group. In Lagrange's resolvent, y_i's were represented by linear combinations of x_i's.

These are the things we are going to do now.

- Solution of a system of linear equations
- Change of variables as a matrix multiplication
- Inverse of a matrix
- Determinant of a matrix
- Cramer's rule and the adjoint matrix

We are actually doing determinants to do higher dimensional analysis.

When we have a curve, we calculate the length of the curve by projecting it to an axis, and then adding up the lengths. In other word, it is

$$\int \sqrt{dx^2 + dy^2} = \int \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx.$$

In calculating higher dimensional objects, such as the area of a surface, we do the same thing with an higher dimensional analogue of Pythagoras theorem.

We will do some review. A system of linear equations

$$\begin{cases} y_1 = a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ y_m = a_{m1}x_1 + \cdots + a_{mn}x_n \end{cases}$$

can be represented by

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

or

$$\vec{y} = \sum_{j=1}^{n} x_j \vec{A_j}$$

where A_j is the jth column of the matrix.

Gauss came up with a procedure to solve the equation. Using the following elementary row operations, we can make the matrix in to a row echelon form.

- Multiply a row by a nonzero number.
- Switch two rows.
- replace the ith row by adding a constant times the jth row.

Everyone knows this. The important observation is that an elementary row operation E applied to A to get A' is the same as applying the operation to I_m to get I'_m and left multiply A to get A'.

Why is this? The jth column of the $n \times m$ matrix A is

$$\begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} = a_{1j}\vec{e_1} + \cdots + a_{mj}\vec{e_m}.$$

Looking at each vector separately, applying a row operation is actually manipulating the coefficient of the vector expansion correspondingly. Thus it is same as left-multiplying a matrix.

Now consider the equation

$$A\vec{x} = \vec{b}$$

where $\vec{b}$ is a column m-vector, and $\vec{x}$ is a column n-vector to solve as the coefficients of the n column m-vectors A. We look at the augmented matrix

$$(A|\vec{b})$$

and apply k elementary row operations. Let those elementary row operations operated on the identity matrix be $E_1, E_2, \ldots, E_k$. Then after the operations, we will get

$$(E_k \cdots E_1 A | E_k \ldots E_1 \vec{b}).$$

Using the Gauss elimination, there exists a wise choice of operations which makes $E_k \cdots E_1 A = A'$ a row echelon matrix. The system is solvable if and only if when the last t rows of A' are identically zero the last t entries of $E_1 \cdots E_1 \vec{b}$ are zero.

Applying this theory to square matrices, we get the following theorem.

Theorem 4.1. *Let A be a square matrix. Then the followings are equivalent.*
(a) Elementary row operation recedes A to I.
(b) A is a product of elementary row operations.
(c) The inverse A^{-1} exists.
(d) $A\vec{x} = 0$ implies $\vec{x} = 0$.

4.3 Determinant of a matrix

Definition 4.2. We define the **determinant** of a matrix inductively as the following. For a 1×1 matrix, the determinant is same as the only entry. For a $n \times n$ matrix, denote A_{ij} by the matrix obtained by throwing the ith row and the jth column of A away. Then define

$$\det A = \sum_{j=1}^{n} (-1)^{j-1} a_{j1} \det A_{j1}.$$

This is characterized by the following properties.

- The function det is a polynomial of the entries.
- Multiplying on row by a leaves the determinant multiplied by a.
- Switching two row changes the sign of the determinant.
- Adding a scalar multiple of a row to another row leaves the determinant unchanged.

You can check each of these by induction. In fact, these properties characterizes the determinant, because you can use the elementary operations to reduce any matrix to either a I or a matrix with a zero row. We can also prove $\det(AB) = \det A \det B$ from this argument.

5 September 17, 2015

The next topic we want to discuss is a preparation for analysis. Stokes' theorem is the higher dimensional version of the fundamental theorem of calculus. We need the notion of exterior algebra to do this. For two vectors $\vec{u}, \vec{v} \in \mathbb{R}^3$, there is the inner product(=dot product) $\vec{u} \cdot \vec{v}$. And there is also the vector product(=cross product) $\vec{u} \times \vec{v}$. Exterior algebra is the higher dimensional version of the cross product. This will be used to calculate the surface, volume, etc, of an object.

5.1 Review of basic matrix theory

The main ingredient of this theory is the elementary row operations, which is motivated by the Gauss elimination.

There were three kinds of elementary row operations. The first one is $E_{(i)\leftrightarrow(j)}$ which is exchanging the ith row and the jth row. (These are not standard notations.) The second one is $E_{(i)\to c(i)}$ which is multiplying the ith row by a nonzero number c. The third one is $E_{(i)\to(i)+c(j)}$ which is adding c times the jth row to the ith row. Here, the entries of the matrix can be either $\mathbb{Q}, \mathbb{R}, \mathbb{C}$, or other fields.

We can reduce a $m \times n$ matrix A to a row echelon form A'.

Definition 5.1. A **row with a pivot** means a row, not identically zero, where the first (from the left) entry is 1, called the **pivot**.

Definition 5.2. A $m \times n$ matrix A is in **row echelon form** if for some $0 \leq r \leq m$

- the first r rows are rows with a pivot,
- the last $m - r$ rows are identically zero,
- the position of the pivot is strictly to the right of a pivot on the preceding row,
- and all entries above a pivot are zero.

We can make a matrix into a row echelon form using the following procedure.

1. Locate the first not identically zero column.
2. $E_{(i)\leftrightarrow(j)}$ (if needed) to make the number at the first row.
3. $E_{(i)\to c(i)}$ (if needed) to make the number 1.
4. $E_{(j)\to(i)+c(j)}$ to make all entries under the 1 zero.

This originally was developed to solve equations and determine the solvable ones, of form

$$A\vec{x} = \vec{b}$$

where A is a $m \times n$ matrix, $\vec{x}$ is a column n-vector, and $\vec{b}$ is a column m-vector.

Elementary row operations is equivalent to left multiplication by the result of applying the operation to the identity matrix I_m of order m.

Apply elementary row operation matrices $E_1, E_2, \ldots, E_k$ on A to make A a row echelon form. Then

$$E_k E_{k-1} \cdots E_2 E_1 A\vec{x} = E_k \cdots E_1 \vec{b}.$$

Letting $E_k \cdots E_1 A = A'$ and $E_k \ldots E_1 \vec{b} = \vec{b'}$, we have

$$A'\vec{x} = \vec{b'}.$$

We can now determine solvability. Obviously, the last $m - r$ components (from the top) of $\vec{b'}$ should vanish for a solution to exist. And actually this condition is sufficient, because you can arbitrarily prescribe values for x_j for the values where there is no pivot in the jth column.

We can translate this into matrix algebra. If we let

$$S = (0|I_{m-r})E_k E_{k-1} \cdots E_2 E_1$$

then the equation $A\vec{x} = \vec{b}$ is solvable if and only if $S\vec{b} = 0$. This matrix is called the **compatibility matrix**.

The name if this matrix actually come from partial differential equations. A system of equations

$$\begin{cases} \frac{\partial u}{\partial x} = P(x, y) \\ \frac{\partial v}{\partial x} = Q(x, y) \end{cases}$$

is solvable only if

$$\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x} = 0.$$

This kind of thing is called the compatibility.

Our second goal is determining when the compatibility is not necessary; i.e., when the equation is solvable (and also unique) for all $\vec{b}$. Let A be a $m \times n$ matrix, where m is the number of equations, and n is the number of variables.

If $m < n$, the solution is never unique. There are not enough pivots, and there is always a column free with pivots. Because we can set the variable to any number, there are at least two solutions to $A\vec{x} = \vec{0}$. If $m > n$, compatibility comes in.

Therefore, $m = n$ should be true. Then the row echelon form of A should be the identity matrix I_n. This means that $E_k \cdots E_1 A = I_n$, and then we have $A^{-1} = E_1^{-1} \cdots E_k^{-1}$. Thus A is invertible. If A is invertible, the solution for $A\vec{x} = \vec{b}$ can be easily described as $\vec{x} = A^{-1}\vec{b}$.

This also gives an algorithm for inverting a matrix. If we make the matrix $(A|I)$ into a row echelon form, then the result will be $(I|A^{-1})$.

5.2 Determinants again

Last time, we defined

$$\det A = \sum_{j=1}^{n} (-1)^{j-1} a_{j1} \det A_{j1}.$$

Consider the effect on the determinant by the elementary row operations. We get

$$\begin{aligned} \det(E_{(i)\leftrightarrow(j)} A) &= -\det A \\ \det(E_{(i)\to c(i)} A) &= c \det A \\ \det(E_{(i)\to(i)+c(j)} A) &= \det A \end{aligned}$$

In the first operation, the sign always changes, because two switch the ith row and the jth row, you need to move the ith row down $j-i$ steps, and then move the jth row up $j-i-1$ steps. The sum is always odd. Note that

$$\det E_{(i)\leftrightarrow(j)} = -1, \quad \det E_{(i)\to c(i)} = c, \quad \det E_{(i)\to(i)+c(j)} = 1.$$

If naturally follows that $\det(EA) = \det E \det A$. Then we get the following.

Theorem 5.3.

$$\det(AB) = \det A \cdot \det B.$$

Proof. First suppose that A is invertible. Then there exist elementary matrices $E_1, \ldots, E_k$ such that $E_k \cdots E_1 A = I$. Then $A = E_1^{-1} \cdots E_k^{-1}$ and

$$\det A = \det E_1^{-1} \cdots \det E_k^{-1}.$$

Then

$$\det(AB) = \det(E_1^{-1} \cdots E_k^{-1} B) = \det A \det B.$$

If $A$is not invertible, then the row reduced echelon form has a zero row. Then $\det A = 0$, and because AB is also not invertible, $\det AB = 0 = \det A \det B$. □

Note that we have also proved in this proof that a matrix A is invertible if and only if $\det A \neq 0$.

Inductively we can obtain the formula to compute the determinant from elementary row operation matrix.

$$\det A = \sum_{\rho \in S_n} \operatorname{sgn}(\rho) a_{\rho(1)1} a_{\rho(2)2} \cdots a_{\rho(n)n},$$

where sgn is the signature, the number of exchanges to make ρ.

5.3 Cramer's rule

Definition 5.4. The **adjugate matrix** $\operatorname{adj} A$ is defined as

$$(\operatorname{adj} A)_{ij} = (-1)^{i+j} \det A_{ji},$$

where A_{ji} is minor matrix.

Actually the ajugate was originally called the adjoint matrix, but the adjoint matrix is used to call the transpose. So we use the non-word 'adjugate'.

Theorem 5.5 (Cramer's rule)**.**

$$A^{-1} = \frac{1}{\det A}(\operatorname{adj} A).$$

Proof. Let us look at the (i,k)th entry of $(\operatorname{adj} A)A$. It is

$$\sum_{j=1}^{n} (\operatorname{adj} A)_{ij} a_{jk} = \sum_{j=1}^{n} (-1)^{i+j} \det A_{ji} \cdot a_{jk}.$$

If $i = k$, the result is $\det A$, by definition. If $i \neq k$, the the result is the determinant of the matrix obtained by replacing the ith column with the kth column. After one elementary column operation, we can make the ith column all zero. Then the value is zero. Thus we get the desired formula. □

Note that if $A\vec{x} = \vec{b}$, then

$$\vec{x} = \frac{1}{\det A}(\operatorname{adj} A)\vec{b}$$

and each entry of $(\operatorname{adj} A)\vec{b}$ is a determinant of a matrix with some column of A replaced by $\vec{b}$.

6 September 22, 2015

Today, we are going to do matrix theory in a coordinate-free manner. It is necessary to go to exterior algebra, which is motivated by Stokes' theorem, which is a generalization of the fundamental theorem of calculus.

6.1 Groups, rings, and fields

Definition 6.1. A **group** is a set G along with the map $G \times G \to G$ which satisfies the following.

(i) Associativity: $(xy)z = x(yz)$.

(ii) Identity: There exists $1 \in G$ such that $1 \cdot x = x \cdot 1 = x$.

(iii) Inverse: There is a $x^{-1} \in G$ such that $x^{-1} \cdot x = x \cdot x^{-1} = 1$.

Example 6.2. *$G = \{$invertible matrices$\}$ and $G = S_n = \{$permutations of $\{1, \dots, n\}\}$ are groups.*

Definition 6.3. A **ring** is a set R with two maps $+ : R \times R \to R$ and $\cdot : R \times R \to R$, where R with addition is a commutative group with identity 0, and the associativity of multiplication and both right and left distributivity holds.

Definition 6.4. A ring D with 1 is called a **division ring** if for any $x \neq 0$, the multiplicative inverse x^{-1} exists.

Definition 6.5. A **field** is a commutative division ring.

Example 6.6. *$F = \mathbb{Q}, \mathbb{R}, \mathbb{C}$ are all fields.*

Definition 6.7. A **module** over a ring R is a set M with addition $M \times M \to M$ with scalar multiplication $R \times M \to M$ where M with addition is an abelian group and the (mixed) associative law $a(bx) = (ab)x$ and the distributive laws $a(x + y) = ax + ay$ and $(a + b)x = ax + bx$ holds.

Definition 6.8. A **vector space** over a field F is a F-module V such that $1 \cdot x = x$.

We are only interested in vector spaces. A vector space has scalar multiplication and addition.

6.2 Vector spaces

Definition 6.9. A set $\{x_1, \dots, x_m, \dots\}$ in a vector space V is called **spanning** if every $v \in V$ is an F-linear combination of a finite number of elements form the set.

$$v = \sum_{j \in \mathcal{J}} a_j x_j$$

Definition 6.10. A vectors space V is **finite dimensional** if there exists a finite spanning subset.

For now, we will only consider finite dimensional matrices.

Definition 6.11. For $v_1, \ldots, v_m \in V$, they are called **linearly independent** if $\sum_{j=1}^{m} a_j v_j = 0$ for some $a_j \in F$ implies $a_j = 0$ for all j.

Definition 6.12. A spanning linearly independent set is called a **basis** of V.

Note that $\{e_1, \ldots, e_m\}$ is a basis if and only if every $v \in V$ can be uniquely expressed as

$$v = \sum_{j=1}^{m} a_j e_j.$$

Proposition 6.13. *Every spanning set contains a subset which is a basis.*

Proof. Take away dependent elements. If $\{v_1, \ldots, v_m\}$ is the spanning set, and it is linearly dependent, then there is $a_j \in J$ such that $\sum_{j=1}^{m} a_j v_j = 0$ where a_j are not all 0. If $a_m \neq 0$, then $v_m = -\frac{1}{a_m} \sum_{j=1}^{m-1} a_j v_j$ can be spanned by other elements. Thus it may be taken away with conserving the spanning property. □

Using this technique of removing linearly dependent elements according to some prescribed preference order, we can also prove the following.

Proposition 6.14. *The cardinality of any linearly independent set is less than the cardinality of any spanning set. Also, any linearly independent set can be expanded to a basis.*

Proof. If $v_1, \ldots, v_m$ are linearly dependent, you remove v_j wit the largest j according to the earlier argument.

(a) Let $\{v_1, \ldots, v_m\}$ be a spanning set, and let $\{u_1, \ldots, u_n\}$ be a linearly independent set. We will substitute $v_1, \ldots, v_m$ by $u_1, \ldots, u_n$ one at a time.

Add u_n to v_ks to make $u_n, v_1, \ldots, v_m$, and because $v_1, \ldots, v_m$ span u_n, we can remove one element of v_k. Without loss of generality, let it be v_m. Then we get a new spanning set $\{u_n, v_1, \ldots, v_{m-1}\}$. Then you add u_{n-1}, and pull another element out. If $n > m$, then you would need to remove u_k at some time. This means that some u_is are linearly dependent. Therefore, we arrive at a contradiction and $n \leq m$.

(b) Start from any linearly independent set $u_1, \ldots, u_n$. And let $v_1, \ldots, v_m$ be a spanning set. (This is possible since it is a finite vector space.) We do the same thing on $\{u_1, \ldots, u_n, v_1, \ldots, v_m\}$ and remove the elements. Since we cannot remove u_is, the resulting set would be a basis which contains $u_1, \ldots, u_n$. □

Corollary 6.15. *The cardinality of any two bases are the same.*

Definition 6.16. The **dimension** of a vector space V is defined as the cardinality of any basis.

6.3 Linear maps and the dual space

Definition 6.17. A **F-linear map** from a vector space V over F to another vector space W over F is a map $T : V \to W$ such that

$$T(au + bv) = aT(u) + bT(v)$$

for any $a, b \in F$ and $u, v \in V$.

Given a basis $e_1, \ldots, e_n$ of V and $f_1, \ldots, f_m$ of W, a F-linear map $T : V \to W$ such that

$$T(e_j) = \sum_{k=1}^{m} a_{kj} f_k,$$

it can be expressed as a matrix (a_{kj}) with columns

$$(T(e_1), \ldots, T(e_n)).$$

Definition 6.18. The **dual** V^* of a vector space V is defined as

$$V^* = \{F\text{-linear maps from } V \text{ to } W\}$$

when F is regarded as a 1-dimensional vector space over F. (Any nonzero element of F is a basis.) The addition is defined for any $\phi, \psi \in V^*$ as

$$(\phi + \psi)(v) = \phi(v) + \psi(v),$$

and the scalar multiplication is defined as

$$(a\phi)(v) = a(\phi(v)).$$

The dimension $\dim V^* = \dim V$, because if $e_1, \ldots, e_n$ is a basis for V over F, then $e_j^* : V \to F$ such that $e_j^*(e_k) = \delta_{jk}$ form a basis of V^*.

The trivial conclusion is that V is the double dual of itself. That is, $V^{**} = V$. We can define the $\Phi : V \to (V^*)^*$ as

$$\Phi(v)(f) = f(v).$$

This is injective, and because the dimension are the same, this is a isomorphism.

In the homework, there was a problem about the Pythagorean theorem for parallelepiped. We can do this coordinate-free. But let us first introduce the length abstractly via inner product. We confine F to either $\mathbb{R}$ or $\mathbb{C}$.

Definition 6.19. Let V be a vector space over C. An **inner product** $(-,-)_V$ means a $V \times V \to \mathbb{C}$ such that:

(i) $u \mapsto (u, v)$ is $\mathbb{C}$-linear for fixed $v \in V$.

(ii) $(u, v) = \overline{(v, u)}$. (conjugate symmetric, or Hermitian symmetric)

(iii) $(u, u) \geq 0$ for all u, and equality hold if and only if $u = 0$.

Proposition 6.20. *A vector space over $\mathbb{C}$ with an inner product $(-,-)_V$ is conjugate self-dual.*

Proof. This is because for a fixed $u \in V$, we can define a $\mathbb{C}$-linear map from V into $\mathbb{C}$ (or a functional on V) by

$$v \mapsto (v,u)_V.$$

But (v,u) is conjugate $\mathbb{C}$-linear in u. □

This is important, especially in partial differential equations, and is know as the Riesz representation theorem.

Definition 6.21. Given a vector space V over $\mathbb{C}$, the complex **conjugate** $\bar{V}$ is a vector space defined as follows.

1. $(\bar{V},+)$ is same as $(V,+)$.
2. For $a \in \mathbb{C}$, $v \in \bar{V}$, define the scalar product of a and v as $\bar{a}v$.

Then the map $\Phi : V \to \bar{V}$ such that $v \to v$ is not linear, but complex conjugate $\mathbb{C}$-linear.

Definition 6.22. Suppose we have finite dimensional $\mathbb{C}$-vector spaces V and W with inner products $(-,-)_V$ and $(-,-)_W$. Given a $\mathbb{C}$-linear map $T : V \to W$, the **adjoint** T^* of T is a $\mathbb{C}$-linear map $W^* \to V^*$ such that

$$(Tv,w)_W = (v,T^*w)_V.$$

There is a conjugate $\mathbb{C}$-linear map $\Psi_V : V \to V^*$ defined via the inner product as

$$(\Psi_V v)(u) = (u,v)_V.$$

That is, $\Psi_V(av) = \bar{a}\Psi_V(v)$. There is the same for W, Ψ_W. Then we get a commutative diagram

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ \downarrow{\scriptstyle \Psi_V} & & \downarrow{\scriptstyle \Psi_W} \\ V^* & \xleftarrow[T^*]{} & W^* \end{array}$$

6.4 Tensor products

Let us define tensor products of V^* and W^*. The reason we use the dual is because of the the historical context.

Definition 6.23. Let V, W be a vector spaces over F. For an element f of V^* and g of W^*, define $f \otimes g$ as an F-linear map from $V \times W \to F$ as

$$(f \otimes g)(v,w) = f(v)g(w).$$

Here, F-linear means that the function is linear respect to each variable v and w.

Definition 6.24. The **tensor product** $V^* \otimes W^*$ of V^* and W^* is defined as the set of all F-linear maps from $V \times W \to F$.

So, for two $f \in V^*$ and $g \in W^*$, the tensor product $f \otimes g \in V^* \otimes W$.

The dimension of the tensor product is

$$\dim_F(V^* \otimes W^*) = (\dim_F V^*)(\dim_F W^*).$$

This is because if we take a basis $\{f_1, \ldots, f_m\}$ and $\{g_1, \ldots, g_n\}$, then the set $\{f_j \otimes g_k\}_{1 \le j \le m, 1 \le k \le n}$ is a basis of $V^* \otimes W^*$.

This was all motivated to calculate the area of a parallelepiped. It started from analysis, but afterwards people wanted to get rid of analysis, and think differential forms as linear functionals.

We can also define the tensor product of many vector spaces

$$V_1^* \otimes V_2^* \otimes \cdots \otimes V_k^* = \otimes^k V$$

This is the set of multilinear functionals $f : V \times \cdots \times V \to F$.

The exterior algebra $\bigwedge^k V$ is the set of all alternating multilinear maps f from $V \times \cdots \times V \to F$ such that

$$f(v_1, \ldots, v_k) = (\operatorname{sgn} \sigma) f(v_\sigma(1), \ldots, v_{\sigma(k)}).$$

7 September 24, 2015

We will talk in more detail about tensor products and exterior products. We defined it for dual spaces, and it is because this is the historic order. I will do exterior differential forms and Riemannian metric tensors for examples. Tensor products will be used in handling polynomials of higher degree in many variables.

7.1 More explanation on tensor products

We started with two dual vectors spaces V^* and W^*. The tensor product $V^* \otimes W^*$ is the set of functions

$$f : V^* \times W^* \to F$$

which is linear in each variable. That is,

$$f(a_1 v_1 + a_2 v_2, w) = a_1 f(v_1, w) + a_2 f(v_2, w)$$
$$f(v, b_1 w_1 + b_2 w_2) = b_1 f(v, w_1) + b_2 f(v, w_2)$$

for each variable. If $\dim_F V = n$ with basis $e_1, \ldots, e_n$ and $\dim_F W = m$ with basis $\hat{e}_1, \ldots, \hat{e}_m$, then $f \in V^* \otimes W^*$ is determined by $f(e_j, \hat{e}_k)$ for $1 \le j \le n$ and $1 \le k \le m$. So $\dim_F V^* \otimes W^* = nm$.

For an element $\alpha \in V^*$ and $\beta \in W^*$, we can define

$$(\alpha \otimes \beta)(v, w) = \alpha(v)\beta(w).$$

Such $\alpha \otimes \beta$ is called decomposable. Not all elements are decomposable; some sum

$$\sum_{l=1}^{N} \alpha_l \otimes \beta_l \in V^* \otimes W^*$$

may be not decomposable.

7.2 Wedge products and some differential geometry

Let me explain how people arrived at this, in an analytic perspective. Ricci first introduced this thing. Let's look at $\mathbb{R}^n$, and a function F on a neighborhood of 0 in $\mathbb{R}$. People know how do differentiate; partial differentiation in one direction. Let $\vec{v} \in V = \{\text{vector of } \mathbb{R}^n \text{ at } 0\}$. The we can differentiate in the direction of $\vec{v}$;

$$\nabla_{\vec{v}} F = \sum_{j=1}^{n} \left(\frac{\partial F}{\partial x_j} \right) v_j$$

where $\vec{v} = (v_1, \ldots, v_n)$. This is actually the $dF \in V^*$, because given a vector in v, we have a number $\nabla_{\vec{v}} F$. If F is the coordinate function $F = x_j$, then $dx_j \in V^*$. Then $dx_j \otimes dx_k \in V^* \otimes V^*$ makes sense, where

$$dx_j \otimes dx_k : V \times V \to \mathbb{R}.$$

This is known as the tensor.

There is also the Riemannian metric tensor. Suppose that you have a way of measuring a vector. It may not be the Euclidean length, because it is not interesting. This is the Riemannian metric tensor, and is written as

$$\sum_{i,j=1}^{n} g_{ij}(P)(dx_i \otimes dx_j) \in V^* \otimes V^*$$

where $g_{ij}(P) \in \mathbb{R}$ and $g_{ij}(P) = g_{ji}(P)$. In the standard Euclidean metric, $g_{ij}(P) = \delta_{ij}$. In this case, the length of the vector will be

$$\|\vec{v}\| = \sqrt{\Big(\sum_{i,j=1}^{n} g_{ij}(P)(dx_j \otimes dx_k)\Big)(\vec{v},\vec{v})}.$$

The exterior product is only defined only for $V = W$. We write $V^* \bigwedge V^* = \bigwedge^2 V^*$.

Definition 7.1. The **wedge product** $V^* \wedge V^*$ is a subset of $V^* \otimes V^*$. A map f is in $V^* \otimes V^*$ if and only if it is skew-symmetric (also called alternating), i.e., $f(v_1, v_2) = -f(v_2, v_1)$.

Definition 7.2. The wedge product

$$\bigwedge^k V^* = V^* \bigwedge V^* \bigwedge \ldots \bigwedge V^* \subset V^* \otimes \cdots \otimes V^* = \otimes^k V^*$$

is the set of functions which is F-linear in each variable and alternating, i.e.,

$$f(v_{\sigma(1)}, v_{\sigma(2)}, \ldots, v_{\sigma(k)}) = \operatorname{sgn}(\sigma) \cdot f(v_1, \ldots, v_k)$$

for any permutation σ.

The function $f \wedge g$ is the is the skew-symmetrization of $f \otimes g$. That is,

$$(f \wedge g)(v_1, v_2) = (f \otimes g)(v_1, v_2) - (g \otimes f)(v_1, v_2).$$

In the case when V is the tangent space, then

$$\begin{aligned}(dx_1 \wedge dx_2)(v_1, v_2) &= (dx_1 \otimes dx_2)(v_1, v_2) - (dx_2 \otimes dx_1)(v_1, v_2) \\ &= (dx_1)(v_1)(dx_2)(v_2) - (dx_2)(v_1)(dx_1)(v_2) \\ &= \begin{vmatrix} \nabla_{v_1} x_1 & \nabla_{v_2} x_1 \\ \nabla_{v_1} x_2 & \nabla_{v_2} x_2 \end{vmatrix}.\end{aligned}$$

Suppose that we have a 2-dimensional surface M in $\mathbb{R}^n$ which is parametrized by s and t, and a point P on M. Let

$$V = T_{M,P} = \text{plane of all tangent vectors to } M \text{ at } P.$$

Then we can restrict the form dx_j as $dx_j\big|_{T_{M,P}} \in (T_{M,P})^*$. Then

$$dx_j \wedge dx_k \in (T_{M,P})^* \bigwedge (T_{M,P})^*$$

and we can calculate the volume as

$$\int_M \sqrt{\sum_{1 \le j < k \le n} |(dx_j \wedge dx_k)(\vec{v}_s, \vec{v}_t)|^2} ds\, dt.$$

As we have done in the problem sets, this is something like the higher dimensional Pythagorean theorem.

7.3 Polarization of a polynomial

Consider a polynomial in many variables of general degree. It will look like

$$F(x_1, \ldots, x_n) = \gamma_0 + \sum_{j=1}^{n} \gamma_j x_j + \sum_{j,k=1}^{n} \gamma_{jk} x_j x_k + \cdots .$$

We divide the variables to two parts $x_1, \ldots, x_n, y_1, \ldots, y_m$, and write it as

$$\begin{aligned} G(x_1, \ldots, x_n, y_1, \ldots, y_m) = a_0 &+ \left(\sum_{j=1}^{n} a_j x_j + \sum_{k=1}^{n} b_k y_k \right) \\ &+ \left(\sum_{j,k=1}^{n} a_{jk} x_j x_k + \sum_{j=1}^{n} \sum_{k=1}^{m} b_{jk} x_j x_k + \sum_{j,k=1}^{m} c_{jk} y_j y_k \right) + \cdots . \end{aligned}$$

If the two a_{jk}s and c_{jk}s are zero, then the polynomial is bilinear in each parts. Then the second degree middle term can be thought as a function in

$$(\mathbb{C}^n)^* \otimes (\mathbb{C}^m)^*.$$

The thing we did by assuming things zero is not actually a special thing, because even in

$$F(x_1, \ldots, x_n) = \gamma_0 + T_1 + T_2 + \cdots ,$$

we can view T_2 as a element $\tilde{T_2} = (\mathbb{C}^n)^* \otimes (\mathbb{C}^n)^*$ where

$$T_2(x_1, \ldots, x_n) = \tilde{T_2}(x_1, \ldots, x_n; x_1, \ldots, x_n).$$

This technique is called polarization. It is useful, because it is looking a function of many variables regard the variables as coordinates in a vector space.

7.4 Binet-Cauchy formula from wedge products

I deliberately assigned the Binet-Cauchy formula as a homework. It is a formula for the determinant of a product of non-square matrices. This follows very naturally from wedge products.

Let $T : V \to W$ be a map, and consider the map

$$T^{(k)} : V \times \cdots \times V \to W \times \cdots \times W$$

which takes each element to its map. Now I can to do the alternation of this and make the map

$$\bigwedge^k T : \bigwedge^k T \to \bigwedge^k W.$$

Note that I am skipping $V^{**} = V$.

Now let

$$V \xrightarrow{T} W \xrightarrow{S} U$$

where $\dim V = \dim U = k$ and $\dim W = n$ and $n > k$. We can extend this to

$$\bigwedge^k V \xrightarrow{\bigwedge^k T} \bigwedge^k W \xrightarrow{\bigwedge^k S} \bigwedge^k U$$

Note that the dimension of $\bigwedge^k T$ and $\bigwedge^k U$ is 1 while the dimension of $\bigwedge^k W$ is $\binom{n}{k}$. If we let $e_1, \ldots, e_k$ be the basis of V and $\hat{e}_1, \ldots, \hat{e}_k$ be the basis of W, then one can check

$$\bigwedge^k (S \cdot T)(e_1 \wedge \cdots \wedge e_k) = \det(ST)(\hat{e}_1 \wedge \cdots \wedge \hat{e}_k).$$

This is because each space is a spaces of dimension 1. Also, in each space, $\bigwedge^k T$ and $\bigwedge^k U$ are represented by $\binom{n}{k}$ vectors. Each of them exactly corresponds to the choice of k rows or columns.

Also the Pythagorean theorem for the volume of a parallelepiped is the relation between inner products of V and inner products of $\bigwedge^k V$. Let $\dim_{\mathbb{R}} V = n$. Note that the inner product $(\cdot, \cdot)_V$ on V is the same as specifying an orthonormal basis $e_1, \ldots, e_n$. Then $e_i \cdot e_j = (e_i, e_j)_V = \delta_{ij}$. The question is "Is there a naturally induced inner product on $\bigwedge^k V$?" There is, because we can just naturally say that $e_{j_1}, \ldots, e_{j_k}$ where $1 \le j_1 < \cdots < j_k \le n$. But then does it depend on the choice of another orthonormal basis?

Let $u_1, \ldots, u_n$ be another choice of the orthonormal basis of V. Presumably, the $u_{j_1} \wedge \cdots \wedge u_{j_k}$ for $1 \le j_1 < \cdots < j_k \le n$ should be orthonormal with respect to the inner product defined by $e_1, \ldots, e_n$. We need to check that

$$(u_{i_1} \wedge \cdots \wedge u_{i_k}, u_{j_1} \wedge \cdots \wedge u_{j_k}) = \delta_{i_1 j_1} \cdots \delta_{i_k j_k}.$$

The left hand side is precisely the determinant

$$\det \begin{pmatrix} u_{i_1} \cdot u_{j_1} & \cdots & u_{i_1} \cdot u_{j_k} \\ \vdots & \ddots & \vdots \\ u_{i_k} \cdot u_{j_1} & \cdots & u_{i_k} \cdot u_{j_k} \end{pmatrix} = \begin{pmatrix} u_{i_1} \\ \vdots \\ u_{i_k} \end{pmatrix} \begin{pmatrix} v_{j_1} & \cdots & v_{j_k} \end{pmatrix}.$$

Then we use can use the Binet-Cauchy formula and get a sum of some products of the determinant $l_1, \ldots, l_k$th column matrix and the $l_1, \ldots, l_k$th row matrix. This is the wedge product

$$(u_{j_1} \wedge \cdots \wedge u_{j_k}, e_{l_1} \wedge \cdots \wedge e_{l_k}).$$

Then we have

$$(u_{i_1} \wedge \cdots \wedge u_{i_k}, u_{j_1} \wedge \cdots \wedge u_{j_k})_{\wedge^k V} = \sum_{1 \le l_1 < \cdots < l_k \le n} (u_{i_1} \wedge \cdots \wedge u_{i_k}, e_{l_1} \wedge \cdots \wedge e_{l_k})_{\wedge^k V} \cdot (e_{l_1} \wedge \cdots \wedge e_{l_k}, u_{j_1} \wedge \cdots \wedge u_{j_k})_{\wedge^k V}$$

What we get after all this is the Lefschetz decomposition theorem.

8 September 29, 2015

There will be a in-class midterm exam on November 3, Tuesday. It will mostly cover solving equations; polynomial equations in one variable, and many linear equations in many variables.

8.1 Historical background

Let me explain the historical starting point of tensors as higher dimensional arrays. There is the vector $\vec{v} = (v_1, \dots, v_n)$ which is an 1-dimensional array, and is used in physics as force, acceleration, etc. Then there is the matrix, which is a 2-dimensional array $A = (a_{ij})$. In physics, this is used in describing forces exerted on many particles, such as stress tensors or strain tensors. Then there is the k-dimensional array. After a while, people realized that it can be defined abstractly, and without choosing a basis.

Let $V_1, \dots, V_p$ be finite dimensional vector spaces over a field F. We defined the $V_1^* \otimes \cdots \otimes V_p^*$ as the set of linear functionals

$$\{f \mid f : V_1 \times \cdots \times V_p \to F\}.$$

If $\dim_F V_i = n_i$ and the basis for V_i is $e_1^{(i)}, \dots, e_{n_i}^{(i)}$, there is the dual basis $(e_*^{(i)})^1, \dots, (e_*^{(i)})^{n_i}$. Then every function f can be written as

$$f = \sum_{1 \le i_t \le n_t} f_{i_1 \cdots i_p} (e_*^{(1)})^{i_1} \otimes (e_*^{(2)})^{i_2} \otimes \cdots \otimes (e_*^{(p)})^{i_p}$$

where $f_{i_1 \cdots i_n}$ is some element of F. Then f can be viewed as the p-dimensional array $(f_{i_1 \dots i_p})_{1 \le i_t \le n_t}$.

Moreover, if $T \in V_1^* \otimes \cdots \otimes V_p^* \otimes W_1 \otimes \cdots \otimes W_q$ then it is a multilinear map $V_1 \times \cdots \times V_p \times W_1^* \times \dots W_q^* \to F$. Then T can be represented as $T = (T_{i_1 \cdots i_q}^{j_1 \cdots j_q})$. Then p is called the the covariant rank, and q is called the contravariant rank. The reason we write some indices on the top, and some on the bottom is because we want to distinguish between the dual and the not dual.

8.2 Evaluation tensor and the contraction map

Let V be a vector space over F. An element $f \in V \otimes V^*$ is a linear functional $f : V^* \times V \to F$. Consider the **evaluation tensor**, which is

$$f(v^*, u) = v^*(u)$$

for $v^* \in V^*$ and $u \in V$. This has covariant rank 1 and contravariant rank 1. Let us give this a name Eval_V, and represent it in terms of basis.

Let $e_1, \dots, e_n$ be a basis of V, and $e_*^1, \dots, e_*^n$ be the dual basis. Then

$$\text{Eval}_V(e_*^j, e_k) = e_*^j(e_k) = \delta_k^j.$$

Then the array is (δ_k^j). Note that we wrote j on top and k on the bottom. If we write it as δ_{jk}, it is not a tensor.

Before introducing the contraction operator, let us consider the dual of tensor product. The product $V_1^* \otimes \cdots \otimes V_p^*$ is the set of multilinear functionals $V_1 \times \cdots \times V_p \to F$. What is the dual of this vector space? We can show that it is $V_1 \otimes \cdots \otimes V_p$. Let $f \in V_1^* \otimes \cdots \otimes V_p^*$. Then for a decomposable element $v_1 \otimes \cdots \otimes v_p \in V_1 \times \cdots \times V_p$, we can define

$$(v_1 \otimes v_p)(f) = f(v_1, \ldots, v_p).$$

Then we have shown that $V_1 \times \ldots V_p \subset (V_1^* \otimes \cdots \otimes V_p^*)^*$, and because $V_1 \times \cdots \times V_p$ generates $V_1 \otimes \cdots \otimes V_p$. After calculating the dimension, you can show that

$$(V_1^* \otimes \cdots \otimes V_p^*)^* = V_1 \otimes \cdots \otimes V_p.$$

Rephrasing it, we can write this as

$$V_1^* \otimes \cdots \otimes V_p^* = \mathrm{Hom}_F(V_1 \otimes \cdots \otimes V_p, F).$$

In a general setting, you can replace F by a tensor product. Then

$$V_1^* \otimes \cdots \otimes V_p^* \otimes W_1 \otimes \cdots \otimes W_q = \mathrm{Hom}_F(V_1 \otimes \cdots \otimes V_p, W_1 \otimes \cdots \otimes W_q).$$

Specifically,

$$V_1^* \otimes \cdots \otimes V_p^* \otimes W^* \otimes W = \mathrm{Hom}_F(V_1 \otimes \cdots \otimes V_p, W^* \otimes W).$$

But we have the evaluation map $\mathrm{Eval}_W : W^* \otimes W \to F$, and composing it with the $\mathrm{Hom}_F(V_1 \otimes \cdots \otimes V_p, W^* \otimes W)$, we get a map

$$V_1^* \otimes \cdots \otimes V_p^* \otimes W^* \otimes W \to \mathrm{Hom}_F(V_1 \otimes \cdots \otimes V_p, F) = V_1^* \otimes \cdots \otimes V_p^*.$$

This looks complicated, but it is more simpler in terms of bases. In terms of bases, this map is actually $(T_k^{i_1 \ldots i_p j}) \mapsto (T^{i_1 \cdots i_p} \delta_k^j)$. This is called the **contraction map**.

If there is a inner product, we don't need the dual space. If V is a vector space over $\mathbb{R}$, then the inner product gives an isomorphism $V \cong V^*$ by $u \mapsto (v \mapsto (v, u)_V)$. It is more complicated if V is a vector space over $\mathbb{C}$. The same map gives us an isomorphism $V^* \cong \bar{V}$, where $\bar{V}$ is the complex conjugate vector space.

Now back to contraction maps. If W is a vector space over $\mathbb{R}$, then we can define the contraction map

$$V_1 \otimes \cdots \otimes V_p \otimes W \otimes W \to V_1 \otimes \cdots \otimes V_p,$$

because we have $W \cong W^*$. In terms of basis, if $u_1, \ldots, u_m$ is a basis for W and $(u_j, u_k)_W = g_{jk} \in \mathbb{R}$, then

$$(T^{i_1 \cdots i_p jk}) \mapsto (T^{i_1 \cdots i_p})$$

where

$$T^{i_1\cdots i_p} = \sum_{j,k} T^{i_1\cdots i_p jk} g_{jk}.$$

In the complex cases, we can similarly define

$$V_1 \otimes \cdots \otimes V_p \otimes \bar{W} \otimes W \to V_1 \otimes \cdots \otimes V_p.$$

8.3 Exterior product of two different vector spaces

Recall that if V is a vector space over F, the exterior product $V^* \bigwedge V^*$ is defined as the set of elements of $V^* \otimes V^*$ which is skew symmetric. It doesn't make sense if the two vector spaces are different.

But there is a need for this. In complex analysis, we have $dz^1, \ldots, dz^n$, and sometimes we have $d\bar{z^1}, \ldots d\bar{z^n}$. But because $z^1 \neq \bar{z^1}$, sometimes exterior product of different vector spaces are needed.

It can be done by embedding V and W into $V \oplus W$. This is generally impossible, but in this case, we can specify an element, namely 0. When there is a linear map $V \otimes W \to F$, we can extend it to a multilinear map $(V \oplus W) \otimes (V \oplus W) \to F$, and then this is an element of $(V \oplus W)^* \otimes (V \oplus W)^*$. Then

$$V^* \otimes W^* \subset (V \oplus W)^* \otimes (V \oplus W)^*,$$

and inside this, we can define $V^* \bigwedge W^*$.

There was a homework problem about this. If V is a vector space over $\mathbb{C}$, we consider as a vector space over $\mathbb{R}$. Then

$$V \otimes_{\mathbb{R}} \mathbb{C} = V \otimes \bar{V}.$$

Then we have

$$V \bigwedge \bar{V} \subset (V \oplus \bar{V}) \bigwedge (V \otimes \bar{V}) = (V \otimes_{\mathbb{R}} \mathbb{C}) \bigwedge (V \otimes_{\mathbb{R}} \mathbb{C}).$$

Generally,

$$\bigwedge^k (V \otimes_{\mathbb{R}} \mathbb{C}) = \bigwedge^k (V \oplus \bar{V}) \cong \bigoplus_{p+q=k} (\bigwedge^p V) \otimes (\bigwedge^q \bar{V}).$$

This is known as the **Hodge decomposition**, and let me explain.

8.4 Hodge decomposition

For instance, let $k = 2$, and consider $\bigwedge^2 (V^* \oplus \bar{V^*})$, and consider an element f. Then f is a multilinear map

$$f : (V \oplus \bar{V}) \times (V \oplus \bar{V}) \to \mathbb{C}.$$

Then $f(v_1 \oplus \bar{v_2}, w_1 \oplus \bar{w_2}) \in \mathbb{C}$. Then you can break up f into four parts $f = f_1 + f_2 + f_3 + f_4$ by linearity. The four pieces will be complex function

defined on $V \times V, \bar{V} \times V, V \times \bar{V}, \bar{V} \times \bar{V}$. When we impose the skew symmetry condition on the pieces, the first and fourth pieces are just skew symmetry. So it is just $\bigwedge^2 V$ and $\bigwedge^2 \bar{V}$. The second and third pieces have some kind of relation, and if we view it as the third being coming out from the second one, we can write it as $V \otimes \bar{V}$. However, if we view it as a whole, we can write it down as $V \wedge \bar{V}$. In fact, these two are the same thing. So

$$\bigwedge^2 (V \oplus \bar{V}) = \bigwedge^2 V \otimes (V \otimes \bar{V}) \otimes \bigwedge^2 \bar{V} = \bigwedge^2 V \otimes (V \bigwedge \bar{V}) \otimes \bigwedge^2 \bar{V}.$$

Generally, we have

$$\bigwedge^k (V \oplus \bar{V}) = \bigoplus_{p+q=k} (\bigwedge^p V) \otimes (\bigwedge^q \bar{V}) = \bigoplus_{p+q=k} (\bigwedge^p V) \bigwedge (\bigwedge^q \bar{V}).$$

9 October 1, 2015

We started out with determinants, and introduced exterior products and tensor products, which are array or entries. There were various techniques; multi linearity, duality, alternation, contraction, complex structure for real vector spaces, and ultimately, Lefschetz theorem. The Lefschetz theorem breaks down the product into simpler building blocks, by contraction with inner product. It only works in $\mathbb{C}$-vector spaces with the inner product.

9.1 Philosophy of the Lefschetz theorem

Let me explain more about the Lefschetz theorem. We consider the vector space V over $\mathbb{C}$. Then $(\bigwedge^p V)\bigwedge(\bigwedge^q \bar{V}) \subset \bigwedge^{p+q}(V \oplus \bar{V})$. The contraction sends (p,q) of $\binom{n}{p}\binom{n}{q}$ dimension to $(p-1,q-1)$ of $\binom{n}{p-1}\binom{n}{q-1}$ dimension. On the other hand, for the Lefschetz operator does the exterior product and sends $(\bigwedge^p V^*)\bigwedge(\bigwedge^q \bar{V}^*)$ to $(\bigwedge^{p+1} V^*)\bigwedge(\bigwedge^{q+1} \bar{V}^*)$. This sends the dimension $\binom{n}{p}\binom{n}{q}$ to $\binom{n}{p+1}\binom{m}{q+1}$.

Let $e_1, \dots, e_n$ be a $\mathbb{C}$-basis of V and let $(\cdot,\cdot)_V$ be a Hermitian inner product. Let $g_{j\bar{k}} = (e_j, e_k)_V$. Then the tensor by which we product in the Lefschetz operator is $\sum g_{j\bar{k}} e_*^j \otimes \bar{e}_*^k = \sum g_{j\bar{k}} e_*^j \wedge \bar{e}_*^k$.

Because people are finding wedge products to hard, I am going to give some time to digest, and delay the proof of the Lefschetz theorem. Meanwhile, I will do something else, which will be used in the proof.

9.2 Hodge star operator

Let V be a vector space over $\mathbb{R}$ with a inner product $(\cdot,\cdot)_V$. Let $e_1, \dots, e_n$ be an orthonormal basis, and the dimension of $\bigwedge^k V$ and the dimension of $\bigwedge^{n-k} V$ is the same, because the basis of $\bigwedge^k V$ is $e_{j_1} \wedge \cdots \wedge e^{j_k}$ for $j_1 < \cdots < j_k$ and the basis of $\bigwedge^{n-k} V$ is $e_{i_1} \wedge \cdots \wedge e_{i_{n-k}}$ for $i_1 < \cdots < i_{n-k}$. Then for a pair, we can assign a number as

$$(e_{j_1}\wedge\cdots\wedge e_{j_k}, e_{i_1}\wedge\cdots\wedge e_{i_{n-k}}) \mapsto (e_{j_1}\wedge\cdots\wedge e_{j_k})\wedge(e_{i_1}\wedge\cdots\wedge e_{i_{n-k}}) = \operatorname{sgn}(\pi) e_1\wedge\cdots\wedge e_n.$$

Then we get a map $(\bigwedge^k V)\bigwedge(\bigwedge^{n-k} V) \to \mathbb{R}$. This map is independent of the basis, up to orientation. The star operator $* : \bigwedge^k V \to \bigwedge^{n-k} V$ as a composite of the paring and the use of the inner product. Because the paring induces an isomorphism $(\bigwedge^k V)^* = \bigwedge^{n-k} V$, and because there is a inner product on $\bigwedge^k V$, we get a isomorphism $* : \bigwedge^k V \to \bigwedge^{n-k} V$.

What does it mean to be independent up to orientation? It means that we need a choice of an element θ which has unit length in $\bigwedge^k V$. Then we can express the definition of the star operator as

$$v \wedge (*u) = (v,u)_{\bigwedge^k V}\theta$$

for $v, u \in \bigwedge^k$.

9.3 Normal form of a matrix

The idea of the normal form is building up from simpler things, just like the Lefschetz theorem. Because tensor is far more complex than the matrix, for the time being, we just consider matrices, or 2-tensors.

We use the technique of elementary row operations and column operations. But we do it on a ring. The difference is that we can't divide an element, and make the value of the pivot to 1. So we choose the 'smallest' element on the pivot column. If the ring is $\mathbb{Z}$, we choose the one with smallest absolute value, and if the ring is $F[\lambda]$, we choose the one with smallest degree. The we use the Euclidean algorithm to make the elements of the column smaller. Then eventually, we will end up with a unique nonzero element. We do this also for the column. Then we end up with

$$PAQ = \begin{pmatrix} d_1 & & & & \\ & d_2 & & & \\ & & \ddots & & \\ & & & d_s & \\ & & & & O \end{pmatrix}$$

with $d_1 \mid d_2 \mid \cdots \mid d_s$, where P and Q are invertible matrices.

Let F be a field and let V be a vector space over F. Let $T : V \to V$ be a linear map. Consider the $F[\lambda] \oplus \cdots \oplus F[\lambda]$. This has a additive structure, and scalar multiplication. This kind of thing is called a free $F[\lambda]$-module of rank n, and is denoted as $F[\lambda]^n$. Anyways, we can give a $F[\lambda]$-module structure over V by

$$(f(\lambda), v) \mapsto f(T) \cdot v.$$

Then the module structure is linked to the normal form of the linear map T.

10 October 6, 2015

We did elementary row and column operation over a ring last class. There are two applications.

- Structure of a finitely generated abelian groups ($R = \mathbb{Z}$)
- Normal form of a matrix ($R = F[\lambda]$)

The important thing is the you need Euclidean algorithm. The "size" of an element is defined as the absolute value in the ring of integers, and the degree in the ring of polynomials. If $\text{size}\, b \leq \text{size}\, a$, then there is a r such that $\text{size}\, r < \text{size}\, b$ and $a = qb + r$. You can successively apply this algorithm to get the greatest common divisor $\gcd(a, b) = c$ such that $c = pa + qb$ for some $p, q \in R$.

For a matrix A, consider the element with smallest size

$$s_{\min} = \min\{\text{size}(a_{ij})\}_{1 \leq i \leq m, 1 \leq j \leq n}.$$

Using row and column exchanges, we can move this to the top left place. Then using Euclid's algorithm, we can either make some element smaller than $s_{\min}$, or make them all zero. If some element becomes smaller than $s_{\min}$, we replace this as the $s_{\min}$ element. Otherwise, we get a first we get a smaller $(m-1) \times (n-1)$ matrix. Since elementary row operations and column operations can be regarded as multiplying matrices, we get

$$PAQ = \begin{pmatrix} d_1 & & & & \\ & \ddots & & & \\ & & d_s & & \\ & & & 0 & \\ & & & & \ddots \end{pmatrix}.$$

10.1 $F[\lambda]$-module structure of a vector space

Let $R = F[\lambda]$ and T be an $n \times n$ matrix with coefficients in a field F. Then we can regard T as $T : V \to V$, or $T \in \text{Hom}_F(V, V)$. Here,

$$V = \bigoplus_{k=1}^{n} Fe_k$$

which means that $e_1, \ldots, e_n$ is a basis. If $T = (a_{i,j})_{1 \leq i,j \leq n}$, then

$$Te_j = \sum_{i=1}^{n} a_{ij} e_i,$$

where e_j is the column vector with one 1 in the jth place.

V is a vector space over F, which is also a module over F. We want to make V a module over the ring $F[\lambda]$. The scalar multiplication is defined as

$$(f(\lambda), v) \mapsto f(T)v$$

where $f(\lambda)$ is a polynomial in $F[\lambda]$ and $v \in V$. What does $f(T)$ mean? If $f(\lambda) = \sum_{j=1}^{m} c_j \bigwedge^j$ with $c_j \in F$, then $f(T) = \sum_{j=0}^{m} c_j T^j$ so that

$$f(T)v = \sum_{j=0}^{m} c_j (T^j v).$$

This is a more complicated structure, because the map T is incorporated in the structure. If we know the structure of this module, we know the map T.

Now we need to come up with a matrix, because we want to apply the elementary row and column operations. We introduce the notion of a free module.

Definition 10.1. A **free module** over R of rank n is

$$R \otimes \cdots \otimes R = \{(r_1, \ldots, r_n) : r_1, \ldots, r_n \in R\}.$$

Let me write $\hat{e}_j$ for the column vector with 1 in the jth position in the free module $F[\lambda]^{\oplus n}$. Then we can make a map

$$\Phi : F[\lambda]^{\oplus n} \to V$$

such that $\hat{e}_j \to e_j$. This map is over the ring $F[\lambda]$, which means that it is linear in the sense of scalar multiplication of a polynomial. That is,

$$\Phi\left(\sum_{j=1}^{n} f_j(\lambda)\hat{e}_j\right) = \sum_{j=1}^{n} f_j(T)e_j.$$

Then we get an so-called exact sequence:

$$0 \longrightarrow \operatorname{Ker}\Phi \hookrightarrow F[\lambda]^{\oplus n} \xrightarrow{\Phi} V \longrightarrow 0$$

We claim that

Claim. *As an $F[\lambda]$-module, we have the isomorphism*

$$\operatorname{Ker}\Phi \cong F[\lambda]^{\oplus n}.$$

Using this claim, we can change the diagram to:

$$0 \longrightarrow F[\lambda]^{\oplus n} \hookrightarrow F[\lambda]^{\oplus n} \xrightarrow{\Phi} V \longrightarrow 0$$

This second arrow is then a $n \times n$ matrix, which contains the information of T.

10.2 Kernel of the map induced by T

Now we prove the claim. By definition,

$$Te_j = \sum_{i=1}^{n} a_{ij} e_i \in V.$$

In $F[\lambda]^{\oplus n}$ we have

$$\lambda \hat{e}_j - \sum_{i=1}^{n} a_{ij} \hat{e}_i \in \operatorname{Ker} \Phi$$

because the image is zero. Denote this element by $\hat{f}_j$. Then it suffices to prove that $\hat{f}_1, \ldots, \hat{f}_n$ make $\operatorname{Ker} \Phi$ a free $F[\lambda]$-module of rank n, or in other words,

$$\operatorname{Ker} \Phi = \oplus_{j=1}^{n} F[\lambda] \hat{f}_j.$$

We need to prove two things; spanning and independence. First let us prove spanning. Take any element $\sum_{j=1}^{n} g_j(\lambda) \hat{e}_j \in \operatorname{Ker} \Phi$. Note that

$$\lambda \hat{e}_j = \hat{f}_j + \sum_{j=1}^{n} a_{ij} \hat{e}_i \in \sum_{i=1} F[\lambda] \hat{f}_i + \sum_{i=1}^{n} F \hat{e}_i \subset F[\lambda]^{\oplus n}.$$

We want to get rid of the $\sum_{i=1}^{n} F \hat{e}_i$ part.

Now we prove that

$$\textstyle\bigwedge^{\ell} \hat{e}_j \in \sum_{i=1}^{n} F[\lambda] \hat{f}_i + \sum_{i=1}^{n} F \hat{e}_i.$$

This is done by induction on ℓ. If $\bigwedge^{\ell} \hat{e}_j$ is in that module, than

$$\textstyle\bigwedge^{\ell+1} \hat{e}_j \in \sum_{i=1}^{n} F[\lambda] \hat{f}_i + \sum_{i=1}^{n} F(\lambda \hat{e}_i)$$

which, in turn, is then in $\sum_{i=1}^{n} F[\lambda] \hat{f}_i + \sum_{i=1}^{F} \hat{e}_i$.

Any

$$\sum_{j=1}^{n} g_j(\lambda) \hat{e}_j \in \sum_{i=1}^{n} F[\lambda] \hat{f}_i + \sum_{i=1}^{n} F \hat{e}_i$$

can be represented as

$$\sum_{j=1}^{n} g_j(\lambda) \hat{e}_j = \sum_{j=1}^{n} h_j(\lambda) \hat{f}_j + \sum_{j=1}^{n} b_j \hat{e}_j.$$

Since the left hand side g is in the kernel of Φ, we can apply the map Φ. Then we get

$$0 = \sum_{j=1}^{n} b_j \Phi(\hat{e}_j) = \sum_{j=1} b_j e_j$$

and thus $b_j = 0$. Therefore, $\sum_{j=1}^{n} g_j(\lambda) \hat{e}_j$ is generated by $\hat{f}_j$s.

We now prove linear independence. Suppose that

$$\sum_{j=1}^{n} h_j(\lambda) \hat{f}_j = 0.$$

Then

$$\sum_{j=1}^{n} h_j(\lambda)\Big(\lambda\hat{e}_j - \sum_{i=1}^{n} a_{ij}\hat{e}_i\Big) = 0$$

and it follows that

$$\sum_{j=1}^{n}\left(h_j(\lambda)\lambda - \sum_{i=1}^{n} a_{ji}h_i(\lambda)\right)\hat{e}_j = 0.$$

Then for each h,

$$h_j(\lambda)\lambda = \sum_{i=1}^{n} a_{ji}h_i(\lambda).$$

If you look at the degree of the h_is, you get a contradiction, because the h_j with the maximal degree times λ cannot be expressed as a linear combination of other polynomials.

10.3 Decomposition of the module structure on V

Finally, we get a $n \times n$ matrix with entries in $F[\lambda]$. Because we actually know what the basis is, we can actually specify what the matrix looks like. Because we are changing the basis from $\hat{f}_j$ to $\hat{e}_j$, and

$$\hat{f}_j = \lambda\hat{e}_j - \sum a_{ij}\hat{e}_i,$$

the jth column is the matrix is

$$\begin{pmatrix} -a_{1j} \\ -a_{2j} \\ \vdots \\ \lambda - a_{jj} \\ \vdots \\ -a_{nj} \end{pmatrix}.$$

Then we get

$$0 \longrightarrow F[\lambda]^{\oplus n} \xrightarrow{\lambda I_n - A} F[\lambda]^{\oplus n} \xrightarrow{\Phi} V \longrightarrow 0$$

We can apply the elementary row and column operations over $F[\lambda]$ to get

$$P(\lambda I_n - A)Q = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & O & \\ & & 1 & & & & \\ & & & d_1(\lambda) & & & \\ & & & & d_2(\lambda) & & \\ & O & & & & \ddots & \\ & & & & & & d_s(\lambda) \end{pmatrix}$$

where $d_j \in F[\lambda]$ and $d_1(\lambda) \mid d_2(\lambda) \mid \cdots \mid d_s(\lambda)$. Note that there are no zeros on the diagonal, because the determinant is nonzero. Also, V is the cokernel of the map $\lambda I_n - A$. The each $d_1(\lambda), \ldots, d_s(\lambda)$ is called the **invariant factor**.

The matrix P is the change of basis in the third $F[\lambda]^{\oplus n}$, and Q is the change of basis in the second $F[\lambda]^{\oplus n}$.

The lots of 1 on the diagonal does not contribute anything to the cokernel. The entry $d_j(\lambda)$ contribute

$$F[\lambda]/d_j(\lambda)F[\lambda] = F \oplus F\lambda \oplus \cdots \oplus F\bigwedge^{\deg d_j - 1}$$

to the cokernel. But because V is of dimension n, we have

$$\deg d_1 + \cdots + \deg d_s = n.$$

Now we have decomposed V into parts which are invariant under T. But we can further reduce things by interpolation techniques (Chinese remainder theorem). That is, if $d(\lambda) = g(\lambda)h(\lambda)$ with g and h relatively prime, then

$$F[\lambda]/d(\lambda)F[\lambda] \cong F[\lambda]/g(\lambda)F[\lambda] \oplus F[\lambda]/h(\lambda)F[\lambda].$$

This is because they have the same dimension and the map is surjective. To show that it is surjective, we need to find f_1 and f_2 such that

$$\begin{cases} f_1(\lambda) \equiv 1 \bmod g(\lambda) \\ f_1(\lambda) \equiv 0 \bmod h(\lambda) \end{cases} \quad \text{and} \quad \begin{cases} f_2(\lambda) \equiv 0 \bmod g(\lambda) \\ f_2(\lambda) \equiv 1 \bmod h(\lambda) \end{cases}$$

These exist because there are q_1 and q_2 such that $1 = g_1 g + g_2 h$.

In the special case $F = \mathbb{C}$, we can use the fundamental theorem of algebra to write

$$F[\lambda]/d(\lambda)F[\lambda] = \bigoplus_{j=1}^{t} F[\lambda]/(\lambda - \gamma_j)^{h_j} F[\lambda].$$

Then each of V is a direct summand of

$$V = \bigoplus_{j} F[\lambda]/(\lambda - r_j)^{k_j} F[\lambda].$$

This gives rise to the Jordan normal form.

We will focus on each $F[\lambda]/(\lambda - r)^k F[\lambda]$. Because this is the cokernel, the element $1 \bmod (\lambda - r)^k$ goes to some $v \in V$. Then λ will go to $Tv \in V$, and likewise, $\lambda k - 1$ will go to $T^{k-1}v \in V$. Then, $\bigwedge^k = \bigwedge^k - (\lambda - r)^k$ will then go to $T^k v - (T - r)^k v$.

11 October 8, 2015

Now we continue our discussion of a normal form of a matrix. The technique is using elementary row and column operations over a ring. The main reason we are doing only for $\mathbb{Z}$ and $F[\lambda]$ is because we can use Euclid's algorithm.

11.1 Review of the decomposition of V as a $F[\lambda]$-module

We started we a linear map $T : V \to V$, and make V a $F[\lambda]$-module by defining $\lambda \cdot v = Tv$. If $e_1, \dots, e_n$ is a basis of V over F, we constructed a map $F[\lambda]^{\oplus n} \to V$ by $\hat{e}_1, \dots, \hat{e}_n \mapsto e_1, \dots, e_n$ and got the exact sequence

$$0 \longrightarrow \operatorname{Ker}\Phi \longrightarrow F[\lambda]^{\oplus n} \xrightarrow{\Phi} V \longrightarrow 0.$$

If we let $\hat{f}_j = \lambda \hat{e}_j - \sum_{i=1}^n a_{ij}\hat{e}_j$, then $\hat{f}_1, \dots, \hat{f}_n$ was the basis for the $F[\lambda]$-module of $\operatorname{Ker}\Phi$. We proved spanning by calculating the error, and independence by considering the maximal degree. The map $\operatorname{Ker}\Phi \to F[\lambda]^{\oplus n}$ was actually $\lambda I_n - T$, and with the elementary operations, made it into a diagonal matrix with entries $1, \dots, 1, d_1(\lambda), \dots, d_s(\lambda)$. Also, V was the cokernel of $\lambda I_n - T$. If $P(\lambda I_n - T)Q$ is the diagonal matrix, then P is changing the basis $\hat{e}_1, \dots, \hat{e}_n$ and actually replaces $\hat{e}_j$ by the jth column of $P = (p_{ij}(\lambda))$. In other words, $\hat{e}_j$ is replaced by

$$\begin{pmatrix} p_{1j}(\lambda) \\ \vdots \\ p_{nj}(\lambda) \end{pmatrix} = p_{1j}(\lambda)\hat{e}_1 + \cdots + p_{nj}(\lambda)\hat{e}_j.$$

This new basis is the good basis, and this means that

$$\{p_{1j}(T)e_1 + \cdots + p_{nj}(T)e_n\}_{j=1}^n$$

is a basis of V, because P is an invertible basis with polynomial entries.

Now when we take the cokernel, the many 1s in the diagonal matrix have zero contribution, and we obtain that V is isomorphic as a $F[\lambda]$-module to

$$\bigoplus_{j=1}^{s} (F[\lambda]/d_j(\lambda)F[\lambda])\hat{e}_{n-s+j}.$$

We want to decompose further. Assume that $F = \mathbb{C}$ (so as to apply the fundamental theorem of algebra to get roots of polynomials.) Decompose $F[\lambda]/d_j(\lambda)F[\lambda]$ by the Chinese Remainder Theorem.

11.2 Chinese remainder theorem

For $\mathbb{Z}$, consider $n_1, \dots, n_k \in \mathbb{Z}_+$ such that they are pairwise relatively prime. Then there exist integers $q_1, \dots, q_k$ such that

$$1 = \sum_{j=1}^{k} q_j n_1 \cdots n_{j-1} n_{j+1} \cdots n_k.$$

If we let $a_j = q_j n_1 \cdots n_{j-1} n_{j+1} \cdots n_k$, we have

$$\begin{cases} a_j \equiv 1 \pmod{n_j} \\ a_j \equiv 0 \pmod{j_k} \quad (k \neq j). \end{cases}$$

Then we see that given b_j, there exists an a such that $a \equiv b_j \bmod n_j$ for all j, namely $a = \sum_j b_j a_j$.

Ditto for $F[\lambda]$. Let $g_1(\lambda), \ldots, g_k(\lambda)$ relatively prime, and we can do the same thing

$$1 = \sum_{j=1}^{k} q_j(\lambda) g_1(\lambda) \cdots g_{j-1}(\lambda) g_{j+1}(\lambda) \cdots g_k(\lambda).$$

Then again,

$$\begin{cases} a_j(\lambda) \equiv 1 \pmod{g_j(\lambda)} \\ a_j(\lambda) \equiv 0 \pmod{g_k(\lambda)} \quad (k \neq j). \end{cases}$$

It then follows that given $b_j(\lambda) \bmod g_j(\lambda)$, there exists a unique $a(\lambda)$ such that $a(\lambda) \equiv b_j(\lambda) \bmod g_j(\lambda)$, defined by $a \equiv \sum_{j=1}^{k} a_j b_j \bmod g_1 \cdots g_k$.

Now factor $d_j(\lambda) = g_{1,j}(\lambda) \cdots g_{k_j,j}(\lambda)$. Then we have

$$F[\lambda]/d_j(\lambda)F[\lambda] \cong \bigoplus_{l=1}^{k_j} F[\lambda]/g_{l,j}(\lambda)F[\lambda].$$

If we want to get the inverse of this decomposition, we need to write

$$1 = \sum_{l=1}^{k_j} q_{l,j}(\lambda) g_{1,j}(\lambda) \cdots g_{l-1,j}(\lambda) g_{l+1,j}(\lambda) \cdots g_{k_j,j}(\lambda).$$

Then the inverse map is

$$b_1, \ldots, b_{k_j} \mapsto \sum_{l=1}^{k_j} q_{l,j} g_{1,j}(\lambda) \cdots g_{l-1,j}(\lambda) g_{l+1,j}(\lambda) \cdots g_{k_j,j}(\lambda) b_l(\lambda).$$

We can phrase it differently. We have an embedding

$$F[\lambda]/g_{l,j}(\lambda)F[\lambda] \hookrightarrow F[\lambda]/d_j(\lambda)F[\lambda]$$

which is the multiplication by $\hat{q}_{l,j} = q_{l,j} g_{1,j}(\lambda) \cdots g_{l-1,j}(\lambda) g_{l+1,j}(\lambda) \cdots g_{k_j,j}(\lambda)$.

We can extend the identity

$$F[\lambda]^{\oplus n}/(\lambda I_n - T)F[\lambda]^{\oplus n} = \bigoplus_{j=1}^{s} (F[\lambda]/d_j(\lambda)F[\lambda])\hat{e}_{n-s+j},$$

using the decomposition we just made to write

$$F[\lambda]^{\oplus n}/(\lambda I_n - T)F[\lambda]^{\oplus n} = \bigoplus_{j=1}^{s} \left(\bigoplus_{l=1}^{k_j} \hat{q}_{l,j}(\lambda)(F[\lambda]/g_{l,j}(\lambda)F[\lambda]) \right).$$

11.3 Jordan normal form

Now let us get back to V. Using the map Φ, we see that

$$V = \bigoplus_{j=1}^{s} \left(\bigoplus_{l=1}^{k_j} (\hat{q}_{l,j} F[\lambda])(T) e_{n-s+j} \right).$$

Because this is a direct summand, T maps to each part $\hat{q}_{l,j}(T)e_{n-s+j}$ to itself. This is a single vector, and let us denote $\tilde{e}_{l,j} = \hat{q}_{l,j}(T)e_{n-s+j}$.

Each of $g_{l,j}(\lambda) = (\lambda - \lambda_{l,j})^{s_{l,j}}$ should be a power of a linear polynomial. Note that

$$(\hat{q}_{l,j} F[\lambda]) e_{n-s+j}) = F[T]\tilde{e}_{l,j} \subset V.$$

as an $F[\lambda]$-module. (This subspace is not necessarily dimension 1 over F.) Because $\hat{q}_{l,j}(\lambda)$ was missing only $g_{l,j}(\lambda)$, when we multiply it, the element $\tilde{e}_{l,j}$ becomes zero. That means that

$$(T - \lambda_{l,j})^{s_{l,j}} \tilde{e}_{l,j} = 0$$

in V. Then the F-basis of $F[T]\tilde{e}_{l,j}$ is

$$\tilde{e}_{l,j}, (T - \lambda_{l,j})\tilde{e}_{l,j}, \ldots, (T - \lambda_{l,j})^{s_{l,j}-1}\tilde{e}_{l,j}.$$

This is because any polynomial $h(\lambda)$ can be written uniquely as

$$h(\lambda) = c_0 + c_1(\lambda - \lambda_{l,j}) + c_2(\lambda - \lambda_{l,j})^2 + \cdots + c_{s_{l,j}-1}(\lambda - \lambda_{l,j})^{s_{l,j}-1} \pmod{(\lambda - \lambda_{l,j})^{s_{l,j}}}.$$

What is the matrix representing this basis? It is

$$\begin{pmatrix} \lambda_{l,j} & 1 & & & \\ & \lambda_{l,j} & 1 & & \\ & & \lambda_{l,j} & & \\ & & & \ddots & \\ & & & & \lambda_{l,j} \end{pmatrix}.$$

This is because T maps $\tilde{e}_{l,j}$ to

$$T\tilde{e}_{l,j} = \lambda_{l,j}\tilde{e}_{l,j} + (T - \lambda_{l,j})\tilde{e}_{l,j}$$

and

$$T(T - \lambda_{l,j})^k \tilde{e}_{l,j} = \lambda_{l,j}(T - \lambda_{l,j})^k \tilde{e}_{l,j} + (T - \lambda_{l,j})^{k+1}\tilde{e}_{i,j}.$$

Now because these are direct decompositions, we can do this for every part and choose a basis of V so that T is represented by the matrix

$$\begin{pmatrix} J_{1,1} & & & \\ & J_{1,2} & & \\ & & \ddots & \\ & & & J_{s,k_s} \end{pmatrix}.$$

This is the **Jordan normal form**. Note that the sum of the size of the blocks is $\sum_{j=1}^{s}\sum_{l=1}^{k_j}\deg g_{l,j} = \sum_{j=1}^{s}\deg d_j = n$.

Let me give some names now. Consider the matrix $\lambda I_n - T$. Then $d_j(\lambda)$ is the same as

gcd of the det of all $(n-s+j)$-minors/gcd of the def of all $(n-s+j-1)$-minors.

Suppose that the $\lambda_1, \ldots, \lambda_n$ are all distinct. Then we see that the invariant factor is just $(\lambda - \lambda_1)\cdots(\lambda - \lambda_n)$.

Definition 11.1. The last invariant factor $d_s(\lambda)$ is called the **minimal polynomial**. This is because d_s is the minimal polynomial such that $d_s(T)\cdot v = 0$ for all $v \in V$.

Definition 11.2. A nonzero vector $v \in V$ is an **eigenvector** with **eigenvalue** λ for $T : V \to V$ if $(T-\lambda)v = 0$. The **eigenspace** E_λ is the set

$$\{\text{eigenvectors for eigenvalue } \lambda\} \cup \{0\}.$$

Definition 11.3. A **generalized eigenvector** v (of rank m) of T with eigenvalue λ for a $T : V \to V$ is a nonzero vector such that

$$(T-\lambda)^m v = 0 \quad \text{and} \quad (T-\lambda)^{m-1} v \neq 0.$$

Likewise, the **generalized eigenspace** is

$$\tilde{E}_\lambda\{v : v \text{ is a generalized eigenvector}\} \cup \{0\}.$$

Now you might feel home, so I will again begin talking about tensor and wedge products. You will be comfortable with the complex structure of a vector space V over $\mathbb{R}$. When we give a complex structure, you have to give some kind of i. So, you need to give a map J such that $J : V \to V$ such that $J^2 = -I$.

One bad thing is that we want to say things about the eigenvalues or eigenvectors, but because V is over $\mathbb{R}$, we cannot use the fundamental theorem of algebra. So we extend J to the map $V \otimes \mathbb{C} \to V \otimes \mathbb{C}$ between tensor products. Then J becomes a $\mathbb{C}$-linear map, and then the eigenvalues of J will be i and $-i$, because $J^2 = -1$. If we let $P_1 = \frac{1}{2}(1 - iJ)$ and $P_2 = \frac{1}{2}(1 + iJ)$, the eigenspace of i is

$$E_i = \operatorname{Im} P_1 = \operatorname{Im} \frac{1}{2}(1 - iJ).$$

12 October 13, 2015

For a matrix $T : V \to V$ over the field F, we used the technique of replacing the action of T on V by an $F[\lambda]$-module structure for V. We used the elementary row and column operations over the ring $F[\lambda]$ on the characteristic matrix $\lambda I_n - T$. Note that this can be done over any field F, while F has to be algebraically closed to further decompose it to a Jordan normal form.

12.1 Justifying complex multiplication on real vector spaces

We go back to the complex structure J of a real vector space $V/\mathbb{R}$. The map $J : V \to J$ is an element $J \in \mathrm{Hom}_{\mathbb{R}}(V, V)$ such that $J^2 = -1$. We want to look at the eigenspace and eigenvalues. (The reason I come back to this is because it is a bridge between real analysis and complex analysis. This is very important.) If v is an eigenvector, $Jv = \lambda v$ and hence $J^2 v = \bigwedge^2 v = -v$, so $\lambda = \pm i$. But because this is not real, we need to extend it to a vector space over $\mathbb{C}$.

The intuitive idea is to multiply a vector from V by i by brute force. For instance, if the basis of V over $\mathbb{R}$ is $e_1, \ldots, e_n$, we just write ie_j. But we have to justify this. We do it by regarding $V = (V^*)^*$. For an element $v \in \mathrm{Hom}_{\mathbb{R}}(V^*, \mathbb{R}) = V$, we can regard it as a $\mathbb{R}$-linear map $V^* \to R$. Then we can compose it with the embedding $\mathbb{R} \to \mathbb{C}$, and then v can be regarded as an element of $\mathrm{Hom}_{\mathbb{R}}(V^*, \mathbb{C}) = V \otimes_{\mathbb{R}} \mathbb{C}$.

More generally, let F be a field, and E be an extension field of F. Then for any vector space W over F, we have

$$W \otimes_F E = \mathrm{Hom}_F(W^*, E) = \mathrm{Hom}_F(\mathrm{Hom}_F(W, F), E).$$

This is called a basis change.

Back to the complex structure. For a linear map $J : V \to V$, we consider the pullback $J^* : V^* \to V^*$ of J. This is defined by

$$f \circ J = J^*(f).$$

Note that the pullback of the pullback $(J^*)^*$ is the same as the original map J. For this, we consider the diagram

$$\begin{array}{ccccc} \mathrm{Hom}(V^*, \mathbb{R}) & \xrightarrow{v} & \mathbb{R} & \longrightarrow & \mathbb{C} \\ {\scriptstyle J^*}\uparrow & \nearrow_{Jv} & & & \\ \mathrm{Hom}(V^*, \mathbb{R}) & & & & \end{array}$$

and compose the map with the embedding with $\mathbb{R} \to \mathbb{C}$. Then we get an extension $J : \mathrm{Hom}_{\mathbb{R}}(V^*, \mathbb{C}) \to \mathrm{Hom}_{\mathbb{R}}(V^*, \mathbb{C})$, and a $\mathbb{C}$-linear map $J \otimes \mathrm{id}_{\mathbb{C}} : V \otimes_{\mathbb{R}} \mathbb{C} \to V \otimes_{\mathbb{R}} \mathbb{C}$. This is just a justification, and you don't have to worry about this when doing real things.

We now decompose the space $V \otimes_{\mathbb{R}} \mathbb{C}$ into eigenspaces by Chinese remainder theorem. There are two eigenvalue i and $-i$ for (the $\mathbb{C}$-linear extension of) J.

Then

$$Id = 1 = \frac{1}{2}(1 - iJ) + \frac{1}{2}(1 + iJ).$$

Then as in the homework assignment, the two spaces $\frac{1}{2}(1 - iJ)(V \otimes_{\mathbb{R}} \mathbb{C})$ is the eigenspace for the eigenvalue i in the $\mathbb{C}$-vectorspace $V \otimes \mathbb{C}$. The two maps $\pi_1 = \frac{1}{2}(1-iJ)$ and $\pi_2 = \frac{1}{2}(1+iJ)$ are projection maps; $\pi_1+\pi_2 = 1$ and $\pi_1^2 = \pi_1$, $\pi_2^2 = \pi_2$.

Before we go back, I want to talk about normalizing constants. We had the identity $1 = \frac{1}{2}(1-iJ)+\frac{1}{2}(1+iJ)$. But there is the constant $1/2$ which is a kind of nuisance. Some authors just don't write the constant explicitly, and some authors do. Similar things happens for wedge products too. When we want to calculate for instance

$$\left(\sum_{i_1<\cdots<i_p} a_{i_1\cdots i_p}(e_{i_1} \wedge \cdots \wedge e_{i_p}) \right) \wedge \left(\sum_{j_1<\cdots<j_q} a_{j_1\cdots j_q}(e_{j_1} \wedge \cdots \wedge e_{j_q}) \right)$$

we need to change the order. If we want to avoid lots of $(-1)^n$, we used the alternating convention, requiring the $a_{i_1\cdots i_p}$ alternating $i_1, \ldots, i_p$. Then

$$\sum_{i_1<\cdots<i_p} (e_{i_1} \wedge \cdots \wedge e_{i_p}) = \frac{1}{p!} \sum_{i_1,\ldots,i_p} a_{i_1\cdots i_p}(e_{i_1} \wedge \cdots \wedge e_{i_p}).$$

This also arises when looking at double integrals.

12.2 Field extensions

Let me move on. We will look at the logical foundation of Galois theory. For an equation $ax^2 + bx + c = 0$, we can solve it as

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

But what does this $\pm\sqrt{\ }$ mean? When working in complex numbers, it is not clear what it exactly is.

We work in the field $F = \mathbb{Q}(a, b, c)$ which consists of all rational functions of the three independent variables a, b, c. What we want to do is to solve the equation $ax^2+bx+c = 0$ in the field $F = \mathbb{Q}(a, b, c)$. Of course, it is not solvable, because the solution involves square roots. So wee need to justify this radical.

How do we justify solving $x^2 - a = 0$ over the field $F = \mathbb{Q}(a)$. We want to construct an extension field E of F such that in E the equation $x^2 - a = 0$ can be solved. Because the rule is that you can only do addition, subtraction, division, multiplication and taking radicals, we are solving only the equations of form $x^n - k$.

The idea is that $F[x]/(x^2 - a)F[x]$ is a field. In general, if $p(x)$ is an irreducible element of $F[x]$, then $F[x]/p(x)F[x]$ is a field. That is, for any $f(x) \in F[x]$ such that $f(x) \not\equiv 0 \bmod p(x)$, then by Euclid's algorithm, we can write

$$1 = g(x)f(x) + h(x)p(x)$$

and $g(x)$ becomes an inverse of $f(x)$.

Then $E = F[x]/(x^2 - a)F[x]$ is an extension of F, and the image x^* of x of $F[x]$ in the quotient $F[x]/(x^2 - a)F[x]$ satisfies $(x^*)^2 - a = 0$. This x^* then is the justification of $\sqrt{a}$. Because the other root is automatically $-x^*$, we don't really need to distinguish between $\sqrt{a}$ and $-\sqrt{a}$, and it doesn't really make sense to distinguish between them.

In the 1820s, people proposed the problem to find a formula, but people first needed to define what a formula means. For a polynomial equation

$$x^n - \sigma_1 x^{n-1} + \sigma_2 x^{n-2} + \cdots + (-1)^n \sigma_n = 0$$

where $\sigma_1, \ldots, \sigma_n$ are independent variables. The field we are working on is $F_0 = F = \mathbb{C}(\sigma_1, \ldots, \sigma_n)$. Now for some $d_0 \geq 2$ and $a_0 \in F_0$, we extend the field $F_1 = F_0[x]/(x_{d_0} - a_0)F_0[x]$. This is one step of the formula. For instance, if $n = 2$, we choose $a_0 = \sigma_1^2 - 4\sigma_2$ and $d_0 = 2$. Next you choose some $d_1 \geq 2$ and $a_1 \in F_1$. You do this until you get some F_ℓ such that the solution $x_1, \ldots, x_n$ are the elements of F_ℓ.

12.3 The rise of Galois theory

Let $x_1, \ldots, x_n$ be independent variables over F. We need to find a chain of extensions

$$\begin{array}{c} F_\ell = F(x_1, \ldots, x_n) = F_{\ell-1}[x]/(x^{d_{\ell-1}} - a_{\ell-1})F_{\ell-1}[x] \\ \cup \\ F_{\ell-1} = F_{\ell-2}[x]/(x^{d_{\ell-2}} - a_{\ell-2})F_{\ell-2}[x] \\ \cup \\ \vdots \\ F_1 = F_0[x]/(x^{d_0} - a_0)F_0[x] \\ \cup \\ F_0 = F(\sigma_1, \ldots, \sigma_n) \end{array}$$

Lagrange first observed that each step is actually the stabilizer of some group. The chain of groups $1 \subset G_1 \subset G_2 \subset \cdots \subset G_\ell = S_n$ then corresponds to

$$\begin{array}{c} F_\ell = F(x_1, \ldots, x_n) \\ \cup \\ F_{\ell-1} = F_\ell G_1 \\ \cup \\ \vdots \\ \cup \\ F_0 = F_\ell^{G_\ell}. \end{array}$$

For each step the group G_{j-1} should be a normal subgroup of G_j, and the quotient group G_j/G_{j-1} should be a cyclic group. I will explain this later.

Starting with an irreducible polynomial $p(x) \in F[x]$ (generalizing $x^2 - a$), we extend F to $E = F[x]/(p(x)F[x])$. Then E is actually a F-vector space. We define the **degree of extension** as the dimension $\dim_F E = [E : F]$.

But when we extend, are the "other" roots of $p(x)$ inside E of not? Generally they are not all in E, and you can easily write down an example. In resolving this problem, the key idea is applying Euclid's algorithm. Denote by x^* the image of $x \in F[x]$ in $F[x]/(p(x)F[x])$. Then after extending to E, the $p(x) \in E[x]$ will contain the factor $x - x^*$, because $p(x^*) = 0$. Write

$$p(x) = (x - x^*)p_1(x) = (x - x^*)g_1(x) \cdots g_k(x)$$

for some irreducible polynomials $g_1, \ldots, g_k \in E[x]$. Then we extend the field E respect to g_1, and we strictly increase the number of roots of $p(x)$. This means that after a finite number of step, we have an extension field $\tilde{F} \supset F$ such that $[\tilde{F} : F] < \infty$ and $p(x)$ can be regarded as and element of $\tilde{F}[x]$ which completely factor in to linear factors

$$p(x) = \alpha(x - x_1) \cdots (x - x_n)$$

with $x_1, \ldots, x_n \in \tilde{F}$.

Now we might have extended more than needed. So we just consider the field $F(x_1, \ldots, x_n)$, which is the set of all elements of $\tilde{F}$ which can be expressed as rational functions of the elements of $x_1, \ldots, x_n$ of $\tilde{F}$ with coefficients in F. This has a name

Definition 12.1. $F(x_1, \ldots, x_n)$ is called a **splitting field** for $p(x) \in F[x]$. (Note that p need not be irreducible.)

Actually the dimension of the splitting field $[F(x_1, \ldots, x_n) : F]$ is unique, and it is related the automorphism group $\mathrm{Aut}_F(F(x_1, \ldots, x_n))$ which is the group of automorphisms which fixes the field F. This is the contribution of Galois.

13 October 15, 2015

We want to look at Galois theory. The question is how to solve polynomial equation over single variable. The main point of Galois theory is the relation between intermediate fields and subgroup of partial symmetry. There is a one-one correspondence, and this is the key of Galois theory.

13.1 Galois theory

We start with elementary symmetric functions $F_0 = F(\sigma_1, \dots, \sigma_n)$ and make it down to $F_\ell = F(x_1, \dots, x_n)$. Starting with an equation

$$x^n - \sigma_1 x^{n-1} + \cdots + (-1)^n \sigma_n,$$

we want to end up with the solution in F_ℓ, and the whole point is expressing $x_1, \dots, x_n$ with $\sigma_1, \dots, \sigma_n$ using rational functions and roots. We want

$$\begin{array}{c} F_0 = F(\sigma_1, \dots, \sigma_n) \\ \cap \\ F_1 = F_0(\sqrt[d_0]{a_0}) \\ \cap \\ F_2 = F_1(\sqrt[d_1]{a_1}) \\ \cap \\ \vdots \\ \cap \\ F_\ell = F(x_1, \dots, x_n) \end{array}$$

But this is not Galois theory.

The field F_0 can be written as $F_0 = F_\ell^{S_n = G_0}$, where this means the set which is fixed by S_0. Actually eat of F_k can be written as $F_k = F_\ell^{G_k}$ with the tower

$$G_0 \supset G_1 \supset \cdots \supset G_\ell.$$

This is the key of Galois theory. It show that the chain of fields is finite.

You start out with a small field F_0 and extend it to F_ℓ. But there is a rule that in each step, it should be extended to a splitting field.

We have to define the automorphism group.

Definition 13.1. An **automorphism group** $\mathrm{Aut}_F E$ of a field E over F is the set of bijections $f : E \to E$ such that the restriction of f to F is the identity.

Let $\tilde{F}$ be an intermediate field between E and F.

$$\begin{array}{c} E \\ \cup \\ \tilde{F} \\ \cup \\ F \end{array}$$

We want to show that there is a one-one correspondence between intermediate fields and the automorphism groups. Let $G = \mathrm{Aut}_F E$ and let

$$H = \{g \in G : g \text{ fixes every elements in } F\}.$$

We can go back to the field by

$$E^H = \{\xi \in E : g(\xi) = \xi \text{ for all } g \in H\}.$$

We can do it the other way. Starting with a subgroup $H \subset G$, we can think of a intermediate field $\tilde{F} = E^H$ and back to groups by $\mathrm{Aut}_{\tilde{F}}(E)$.

What I presented is a way of going from intermediate fields to automorphism groups, and another way of going from automorphism groups to intermediate fields by considering the set of elements fixed by the group. The question is whether these two procedures are inverses of each other. In other words, is

$$E^H = \tilde{F} \quad \text{and} \quad \mathrm{Aut}_{\tilde{F}}(E) = H \text{ in the two situations?}$$

These are true if F is a splitting extension of E.

Let us look at the process of making a splitting field. For a field F, we take a polynomial $f(x)$ with coefficients in F. Then we construct it as $F[x]/f(x)F[x]$. Then the root of $f(x)$ in E is ξ, which is the image of x. If $F \subset \tilde{F} \subset E$ is an intermediate field, then if we want to construct E from $\tilde{F}$, we can just take the exactly same polynomial $f(x)$. Since $f \in F[x]$, it is also inside $\tilde{F}[x]$. Then we see that E is also a splitting field of $\tilde{F}$.

But is $\tilde{F}$ a spitting field of F? Yes it is, if and only if $\tilde{F} = E^H$ for a certain H. We call subgroups with this certain property normal.

13.2 Normal groups and solvability

Definition 13.2. A subgroup H of a group G is called **normal**, if $ghg^{-1} \in H$ for any $g \in G$ and $h \in H$.

This means that if we consider the tower

$$\{1\} = G_\ell \subset G_{\ell-1} \subset \cdots \subset G_1 \subset G_0 = S_n,$$

which is the solution to solving equations, G_j should be normal in G_{j-1}. But because we allow only taking roots, the quotient G_j/G_{j-1} should be cyclic. Note that any abelian group can be written as a direct sum of cyclic groups. So we can loosen the condition G_j/G_{j-1} being cyclic to G_j/G_{j-1} being abelian.

Also, the condition $H \subset G$ is normal and G/H is abelian, is actually equivalent to H containing the **commutator subgroup**, which is the group generated by

$$\{ghg^{-1}h^{-1} : g, h \in G\}.$$

We then can just consider the tower

$$G \supset \mathrm{comm}(G) \supset \mathrm{comm}(\mathrm{comm}(G)) \supset \cdots \supset \{1\}.$$

We can introduce a definition.

Definition 13.3. A finite group G is called **solvable** if and only if some

$$\operatorname{comm}(\cdots(\operatorname{comm}(G))\cdots) = \{1\}$$

Now if we assume Galois theory, the solvability of a equation of degree n boils down to the solvability of S_n.

Let me introduce one theory about the solvability of groups.

Theorem 13.4 (Feit-Thompson)**.** *Every group of odd order is solvable.*

But this doesn't really help when finding out whether S_n is solvable. Because S_n has a lots of 2s, we need to take out all the trouble makers. If $|S_n| = n! = 2^k \cdot (\text{odd})$, then we would want a $H \subset S_n$ with $|H| = 2^k$ which is normal. If we

This was actually done by Sylow.

Theorem 13.5 (Sylow)**.** *If $|G| = p^m \cdot (p\text{-free product})$, then G contains a subgroup of order p^m.*

But I will not go into all this. Maybe I will assign as a homework problem.

13.3 Bounding theorems for Galois extensions

Now let us go back to Galois theory. There were two operations: forming the automorphism group over a field, and forming the fixed field. We will prove that they are inverses by bounding the dimension of the intermediate field.

One direction is bonding the "size" of the field.

Theorem 13.6 (Artin)**.** *Let E be an field, and let $G \subset \operatorname{Aut} E$ be a finite subgroup. If we consider the field $F = E^G$, then $[E : F] \leq |G|$.*

Proof. Let $m = |G|$, suppose that $n > m$ and let $x_1, \ldots, x_n$ be any elements in E. It suffices to prove that there exists a nonzero solution to the equation

$$a_1 x_1 + \cdots + a_n x_n = 0$$

such that $a_1, \ldots, a_n \in F$. Then it will show that $n > [E : F]$ since any $x_1, \ldots, x_n$ are linearly independent. Because $n > [E : F]$ for any $n > m$, it will follow that $[E : F] \leq m = |G|$.

Let $G = \{g_j\}_{j=1}^n$, and consider the system of equations

$$a_1 g_j(x_1) + \cdots + a_n g_j(x_n) = 0.$$

Since there are more variables than equations, there should be a nonzero solution $(a_1, \ldots, a_n) \in E^n$. But since it is not a solution in F, we need to bring it down to F.

We now have a nonzero solution $a_1, \ldots, a_n \in E$. Consider the solution $a_1, \ldots, a_n \in E$ with the maximal number of elements which are not zero. Since we can multiply any nonzero element of E to all $a_1, \ldots, a_n$ simultaneously, we may assume without loss of generality $a_1 = 1$. Suppose that there is some a_j

which is not in F, say $a_2 \notin F$. Then there should be some $g_k \in G$ such that $a_2 \neq g_k(a_2)$. Now for any $1 \leq j \leq m$, we have

$$g_k(a_1)g_j(x_1) + g_k(a_2)g_j(x_2) + \cdots + g_k(a_n)g_j(x_n) = 0$$

since we can apply the whole automorphism g_k to the system of equations. On the other hand we have

$$a_1 g_j(x_1) + a_2 g_j(x_2) + \cdots + a_n g_j(x_n) = 0,$$

which is the original system of equations. Subtracting the two equations, we get a new solution

$$(g_k(a_1) - a_1, g_k(a_2) - a_2, \ldots, g_k(a_n) - a_n).$$

This solution is nonzero because $g_k(a_2) \neq a_2$, and there is at least one more zero because $g_k(a_1) - a_1 = 0$. Therefore, we get a contradiction, and it follows that $a_1, \ldots, a_n \in F$. □

The other direction is:

Theorem 13.7. *Given a field F and a polynomial $f(x) \in F[x]$, construct a splitting field E of $f(x)$ over F. Then $[E : F] = |\mathrm{Aut}_F E|$.*

Proof. The idea is to count the degree of freedom in the remainder of Euclid's algorithm. If $g(x) \in F[x]$ is irreducible, then $\deg g$ is the number of embeddings of $F(\xi)$ into a splitting field E, where ξ is the image of x in $F[x]/g(x)F[x]$. This is because $f(x)$ can be decomposed into linear factors.

Now $\deg g = [F(\xi) : F]$, and thus $[F(\xi) : F]$ is the number of embeddings of $F(\xi)$ into E. Since g is irreducible, the minimal polynomial of ξ should be exactly g. This means that actually $F(\xi)$ is just E, and thus $[F(\xi) : F] = |\mathrm{Aut}_F F(\xi)|$.

But g may not be irreducible. Suppose that $g(x) = g_1(x) \cdots$. Consider a root ξ_1 such that $g_1(\xi_1) = 0$. Then the field $F(\xi_1)$ factors in $F[\xi]$ to $g(x) = (x - \xi_1)h_1(x) \cdots$. Let ξ_2 be the root of h_1 and consider $F(\xi_1)(\xi_2)$.

Then $E = F(\xi_1)(\xi_2) \cdots (\xi_k)$. For each k, the number of embedding

$$F(\xi_1) \cdots (\xi_k) \hookrightarrow E$$

which fixes $F(\xi_1) \cdots (\xi_{k-1})$ will be $[F(\xi_1, \ldots, \xi_k) : F(\xi_1, \ldots, \xi_{k-1})]$. Then the number of embeddings $E \hookrightarrow E$ fixing F will be the product

$$\prod_{k=1}^{n} [F(\xi_1, \ldots, \xi_k) : F(\xi_1, \ldots, \xi_{k-1})] = [F(\xi_1, \ldots, \xi_k) : F] = [E : F].$$

□

14 October 20, 2015

We want to solve a polynomial equation, and the only known ones are quadratic, cubic, and quartic cases. If you start out $F_0 = F(\sigma_1, \dots, \sigma_n)$, and extend it to $F_1 = F_0(\sqrt[d_0]{a_0})$, $F_2 = F_1(\sqrt[d_1]{a_1})$, and so forth, until $F_q = F(x_1, \dots, x_n)$, we get a complicated solution. Lagrange first related it to fixed fields of subgroups of S_n.

The whole game is the correspondence between intermediate fields and subgroups. If you look at intermediate fields, it looks like there are infinitely many choices, but if you look at subgroups, it is much more easier. Historically people tried to to things in intermediate fields and got stuck. Galois came and said that you can look at subgroups instead.

Generally, let E be a splitting field of a field for a polynomial $f(x)$ with coefficients in F. This means that the polynomial $f(x)$ factors into linear factors, i.e.,

$$f(x) = a(x - r_1) \cdots (x - r_n)$$

where $r_1, \dots, r_n \in E$. Also, because E should be the minimal field, we have $E = F(r_1, \dots, r_n)$. These two conditions are the definition. The reason for making this notion is to justify logically the formulas.

14.1 Separability of a polynomial

One more important thing is that all roots of a irreducible polynomial in $F[x]$ should be distinct. Later, it will be used in counting elements in the subgroup of automorphism group. For a field with characteristic 0, that is, $1+1+\cdots+1 \neq 0$, the roots will be always distinct. This is because for any irreducible f, the polynomial f and f' cannot share a common factor. If the characteristic is zero, problems can occur. For instance, let $F = (\mathbb{Z}/p\mathbb{Z})(t)$, and consider the polynomial $X^p - t \in F[X]$. In the splitting field, if factors into

$$X^p - t = (X - \sqrt[p]{t})^p.$$

We want to say that a polynomial has distinct roots. But there might be two or more non-isomorphic splitting fields, and one might have same roots while the other have distinct roots. So we need the following theorem.

Theorem 14.1. *Let $f(x) \in F[x]$, and let $F \hookrightarrow E$ be a splitting field, and suppose that you have another splitting field $F \hookrightarrow \tilde{E}$. Then there is an isomorphism $E \cong \tilde{E}$.*

Proof. We prove this by induction. Let $f(x) = g_1(x) \cdots g_k(x)$ and $x^* \in E$ be a root of $f(x)$. In $\tilde{E}$, the polynomial $f(x)$ will split into

$$f(x) = (x - r_1) \cdots (x - r_n),$$

and we should map x^* to some r_i. Suppose that $g_1(x^*) = 0$. In $\tilde{E}$ there should be some r_1 such that $g_1(r_1) = 0$. Then we construct

$$F[x]/g_1(x)F[x] \to F(r_1) \subset \tilde{E}$$

and the elements $1, x^*, (x^*)^2, \ldots, (x^*)^{\deg g_1 - 1}$ will be linearly independent in the left side, while $1, r_1, r_1^2, \ldots, r_1^{\deg g_1 - 1}$ will be linearly independent in the right side. So we construct a isomorphism $F[x]/g_1(x)F[x] \to F(r_1)$ by sending $x^* \mapsto r_1$. Now if we let $\tilde{F} = F[x]/g_1(x)F[x] = F(r_1)$, the fields E and $\tilde{E}$ are both splitting fields of $\tilde{F}$. Then we can inductively construct the isomorphism. □

Definition 14.2. A polynomial $f(x) \in F[x]$ is called **separable** if all its foots (in a splitting field) are distinct.

14.2 The second counting argument

There are two counting proposition. One is Artin's theorem. The other one is this. This bounds the size of the group by the degree.

Theorem 14.3. *Let E be a splitting field for a separable polynomial $f(x)$ with coefficients in F. Then $[E : F] = |\mathrm{Aut}_F(E)|$.*

Proof. Again, we use induction on $n-k$, where $n = \deg f$ and $f(x) = g_1(x) \cdots g_k(x)$. We prove the following more general formulation.

> *Suppose $\tilde{E}$ is another extension field of F which contains a splitting field of F. Then the number of embeddings of E into $\tilde{E}$ is equal to $[E : F]$.*

Let $\tilde{F} = F[x]/g_1(x)F[x]$. The the number of embedding of $\tilde{F}$ into $\tilde{E}$ over F is $\deg g_1 = m$, because an embedding is uniquely determined by the image of x^*. (Note that the fact that g_1 has distinct roots is used here.) Now fix any embedding φ_j. Then

$$\tilde{F} \to \varphi_j(\tilde{F}) \subset \tilde{E}.$$

We identify the two field so that both E and $\tilde{E}$ are extension fields of $\tilde{F}$. Then in the new field $\tilde{F}$, the polynomial $f(x)$ splits into

$$f(x) = (x - x^*)\tilde{g}_1(x)g_2(x) \cdots g_k(x).$$

Because this has smaller $n - k$, we can use the induction hypothesis.

Then the number of embeddings $E \hookrightarrow \tilde{E}$ over $\tilde{F}(\varphi_j(\tilde{F}))$ is $[E : \tilde{F}]$. Then the number of all embeddings $E \hookrightarrow \tilde{E}$ over F is

$$[E : \tilde{F}][\tilde{F} : F] = [E : F].$$

This shows that $[E : F] = |\mathrm{Aut}_F(E)|$. □

14.3 Galois extension

There are three equivalent characteristics of a splitting field.

(i) E is a splitting field of a polynomial with coefficients in F.

(ii) For some finite subgroup of G of $\mathrm{Aut}(E)$, the field F is the fixed field E^G.

(iii) E is a finite extension field of F, which is both normal and separable.

Definition 14.4. Let E be a extension field of F. The field E is **normal** over F if the minimum (monic) polynomial for any element of E with coefficient in F splits in E. The field E is **separable** over F if the minimum polynomial is separable, i.e., all roots are distinct. If E is both normal and separable over F, then it is called **Galois** over F.

We will prove this next class.

15 October 22, 2015

You start out with a field F and a splitting field for a separable polynomial. (The theory started out with no condition on separability.) The main point is that the splitting field is unique, before induction. The main trick is that you can generally construct $F[x]/f(x)F[x]$. After we have that it is unique, we can say something about separability. You need separability because you want to count things. The first thing is the dimension of E over F, and the second thing is the number of automorphisms of E over F. Then you have $[E : F] = |\mathrm{Aut}_F(E)|$. You want to make a correspondence between intermediate fields and subgroups, and one inclusion is obvious. The other inclusion requires these countings.

15.1 Three equivalent definitions of Galois extensions

We pick up from what we left of last time.

Theorem 15.1. *The three statements are equivalent.*

(i) E is a splitting field of a separable *polynomial $f(x)$ over F.*[1]

(ii) For a finite group $G \subset \mathrm{Aut}\, E$, the field F is the fixed field E^G.

(iii) E is a separable, normal finite extension of F.

Proof of (i) $\Rightarrow$ (ii). We have $f(x)$, and we need the group G. Because we have $f(x)$, we have $\mathrm{Aut}_F(E)$. We use this group as G. Of course, we have to check. Let $F' = E^G$, and we want to show that $F = F'$. By definition, $F \subset F'$, because all automorphisms of G fixes F.

Now we use the counting. The second counting tells us that $[E : F] = \mathrm{Aut}_F(E)$, and because $F \subset F' \subset E$, the field E is a splitting field over F'. So again, $[E : F'] = \mathrm{Aut}_{F'}(E)$. But let us look at $\mathrm{Aut}_{F'}(E)$. This is in $\mathrm{Aut}_F(E)$, because F' is bigger than F. On the other hand, every element of $\mathrm{Aut}_F(E)$ fixes F', so it follows that $\mathrm{Aut}_{F'}(E) = \mathrm{Aut}_F(E)$. Hence

$$[E : F'] = |\mathrm{Aut}_{F'}(E)| = |\mathrm{Aut}_F(E)| = [E : F].$$

By the tensor product thing, we have $[E : F'][F' : F] = [E : F]$. But because $[E : F'] = [E : F]$, we see that $[F' : F] = 1$. Therefore, $F' = F$. □

Proof of (ii) $\Rightarrow$ (iii). Given a finite group G inside $\mathrm{Aut}(E)$, we want to show that E it is separable and normal over $F = E^G$. Let $\xi \in E \setminus F$, and we want to show that the minimal polynomial $f_\xi(x)$ has all roots in E, and that they are distinct. The trick is to produce another polynomial by Artin's technique and compare to $f_\xi(x)$.

The group G acts on ξ to produce the orbit $\{\gamma\xi\}$ for $\gamma \in G$. And we can consider the stabilizer subgroup

$$G_\xi = \{\gamma \in G : \gamma\xi = \xi\}.$$

[1] No irreducibility is used, because we want E to be also a splitting field for f over K for any $F \subset K \subset E$.

Then G is the disjoint union

$$G = \bigcup_{1 \le j \le m} \gamma_j G_\xi$$

for some $\gamma_1, \dots, \gamma_m \in G$, and moreover the orbit is

$$\{\gamma\xi\} = \{\gamma_1\xi, \dots, \gamma_m\xi\}.$$

We now form the polynomial $g_\xi(x)$ from the elementary symmetric functions of elements in the orbit G_ξ, i.e.,

$$g_\xi(x) = (x - \gamma_1\xi) \cdots (x - \gamma_m\xi) \in F[x].$$

Then $g_\xi(x)$ is separable, and al the roots are in E. All the elements in the orbit are in $f_\xi(x)$ because $f_\xi \in F[x]$, and hence $g_\xi(x)$ divides $f_\xi(x)$. But because $f_\xi(x)$ is irreducible, we see that $f_\xi(x) = ag_\xi(x)$ for some $a \in F \setminus \{0\}$. This shows that the extension is normal and separable.

Lastly, Artin's theorem shows that the extension is finite. This finishes the proof. □

Proof of (iii) ⇒ (i). Let $\xi_1, \dots, \xi_\ell$ be a basis, so that $E = F\xi_1 + \cdots + F\xi_\ell$. Consider g_{ξ_k} be the minimal polynomial of ξ_k. Let

$$f(x) = g_{\xi_1}(x) \cdots g_{\xi_k}(x),$$

and delete all duplicates. Then we get a separable polynomial. The splitting field of F over $f(x)$ shall be E. □

15.2 Some comments about normality

Let E/F be a Galois extension. By this, I mean that E is separable, normal, and finite over F. If we have a intermediate $F \subset K \subset E$, we know that if E is Galois of F, then E is also Galois over K. But is K Galois over E?

You want to imitate the argument in (ii) ⇒ (iii) to show that K is normal over F, but the big problem is that the orbit of some element in K might get out of K. Then you would ask, "Is $\gamma(K) \subset K$ for all $\gamma \in G = \mathrm{Aut}_F(E) = \mathrm{Gal}(E/F)$?" This condition is called the invariance of K under G.

Proposition 15.2. *The extension K/F is Galois if and only if $G = \mathrm{Aut}_F(E)$ maps K to K. And this is true if and only if $\mathrm{Aut}_K(E)$ is normal in G.*

We shall prove this claim. This is part of the Fundamental theorem of Galois theory.

15.3 Fundamental theorem of Galois theory

We are now ready to state the theorem

Theorem 15.3 (Fundamental theorem of Galois theory)**.** *Let E/F be a Galois extension, i.e., E is a splitting field of some separable polynomial with coefficients in F. Let $G = \mathrm{Aut}_F(E)$. Let*

$$\mathcal{H} = \{H : H \text{ is a subgroup of } G\},$$

and let

$$\mathcal{K} = \{K : K \text{ is a field with } E \subset K \subset F\}.$$

Then there is a correspondence between $\mathcal{H}$ and $\mathcal{K}$, which is natural in the following sense: for $H \leftrightarrow K$, we have

$$H = \mathrm{Aut}_K(E), \quad K = E^H.$$

Moreover, $[E : K] = |K|$ and $|G|/|H| = [K : F]$. Furthermore, H is normal G if and only if K/F is Galois.

Proof of the first direction of correspondence (group $\to$ field $\to$ group).
Let E be a field, and G be a finite subgroup of $\mathrm{Aut}(E)$. Let F be the fixed field $F = E^G$. We want to prove that $G = \mathrm{Aut}_F(E)$. Note that I have removed the base field to simplify the notation.

First, there is the trivial inclusion $G \subset \mathrm{Aut}_F(E)$, because by definition G fixes F.

Now by Artin's theorem counting, we have $[E : F] \leq |G|$. Since F is a fixed field of some subgroup, we see that E is a splitting field of some polynomial over F. Hence we can use the second counting to get $[E : F] = |\mathrm{Aut}_F(E)|$. Hence

$$|\mathrm{Aut}_F(E)| = [E : F] \leq |G|.$$

Together with the trivial inclusion, we see that $G = \mathrm{Aut}_F(E)$. □

Proof of the second direction of the correspondence (field $\to$ group $\to$ field).
We start with the Galois extension E/F, and let $F \subset K \subset E$. Let $G = \mathrm{Aut}_K(E)$ and $L = E^G$. We want to show that $K \subset L$. Again, we have the trivial inclusion $K \subset L$.

We have two Galois extensions E/K and E/L. The second one is Galois since $K \subset L \subset E$. Using the second counting, we see that

$$[E : L] = |\mathrm{Aut}_L(E)|, \quad [E : K] = |\mathrm{Aut}_K(E)|.$$

But because $L = E^G$ and $G = \mathrm{Aut}_K(E)$, from the first part of correspondence we have, we know that $\mathrm{Aut}_L(E) = G = \mathrm{Aut}_K(E)$. Hence $[E : L] = [E : K]$, and therefore we get $L = K$. □

Now we need to show the last sentence of the theorem; that H is normal in G if and only if K/F is Galois. We will do this next time.

16 October 27, 2015

There will be four questions in the midterm: two questions in linear algebra, and two questions in Galois theory.

16.1 Wrapping up Galois theory

If E/F is Galois, and $F \subset K \subset E$ is an intermediate subfield, then E/K is automatically Galois. But the question is when K/F is also Galois.

Theorem 16.1. *An intermediate field K is Galois over F if and only if $\mathrm{Aut}_K(E)$ is normal in $\mathrm{Aut}_F(E)$.*

Proof. By Galois theory, the field K is the fixed field of $\mathrm{Aut}_K(E) = H$. First assume that K is Galois over F. Then we need to show that for any $\gamma \in G$, the conjugate $\gamma H \gamma^{-1}$ is in H. Consider any $x \in K$. Because K is an Galois extension of F, the map $\gamma^{-1}x$ should be in K.[2] Then H fixes K, so then γ turns it back to x. So $\gamma H \gamma^{-1} x = x$. Hence H is normal in G.

(Siu skipped the other direction.) □

16.2 Solvability of the polynomial with degree n

So let us get back the the solvability of a general degree n polynomial. Since $F(x_1, \dots, x_n)$ is Galois over $F(\sigma_1, \dots, \sigma_n)$, we know that they are all fixed fields of some groups.

$$\begin{array}{c} F_0 = F(\sigma_1, \dots, \sigma_n) = F(x_1, \dots, x_n)^{S_n = G_0} \\ \cap \\ F_1 = F(x_1, \dots, x_n)^{G_1} \\ \cap \\ \vdots \\ \cap \\ F_j = F(x_1, \dots, x_n)^{G_j} \\ \cap \\ F_{j+1} = F(x_1, \dots, x_n)^{G_{j+1}} \\ \cap \\ \vdots \\ \cap \\ F(x_1, \dots, x_n)^{G_q} \end{array}$$

In each step, F_{j+1} is an extension field of F_j by adjoint the root of $X^{d_j} - a_j$. For instance, if we let $F_{j+1} = F_j(\xi_j)$, then $a_j = \xi_j^{d_j} \in F_j$. That is, $F_{j+1} = F_j[X]/(X^{d_j} - a_j)F_j[X]$. Assume that F contains all d_jth roots of unity. Then

[2] This is Artin's technique. You look at the orbit of x under G, which is $\bigcup_{\gamma_k \in G} \gamma_k x$. Then $\prod(X - \gamma_k x)$ is the minimal polynomial of x.

all roots of $X^{d_j} - a_j$ is in F_{j+1} and hence F_{j+1} is just the splitting field. This means that F_{j+1} should be Galois over F_j. We can even claim the following.

Claim. *Assume that d_j is prime. The extension from F_j to F_{j+1} is taking the d_jth root if and only if $\mathrm{Gal}_{F_j}(F_{j+1})$ is cyclic of order d_j.*

Proof. There are two directions. We have assumed that the ground field F contains the dth root of unity. Then if you have a root ξ of $X^d - a$ then you get all the roots by the form $\omega^k \xi$. Let γ_k be the automorphisms sending $X \mapsto \omega^k X$. Then $\gamma_k \circ \gamma_l$ maps X to $\omega^{k+l} X$. So this shows that the Galois group is cyclic.

The other direction uses Lagrange resolvent. We convert the Galois group action to multiplication by root of unity. Let η be a generator. Then the all the automorphisms are $1, \eta, \ldots, \eta^{d-1}$. Take one element $\xi \in E \setminus F$. Then $F(\xi) = E$, because $[E : F]$ is prime and $[F(\xi) : E] > 1$. Then $\xi, \eta(\xi), \eta^2(\xi), \ldots, \eta^{d-1}(\xi)$ shall be the basis of F over E. Define

$$\zeta_\ell = \sum_{j=1}^{d-1} (\omega^\ell)^j \eta^j(\xi)$$

for $0 \le \ell \le d-1$. This is the same thing we did for the cubic formula. Then

$$\eta(\zeta_\ell) = \sum_{j=1}^{d-1} (\omega^\ell)^j \eta^{j+1}(\xi) = \omega^{-\ell} \zeta_\ell.$$

Then $(\zeta_\ell)^d$ is invariant under η, and hence $(\zeta_\ell)^d$ is in the field F_j. □

So solvability of a polynomial of degree n is equivalent to the existence of the tower of groups

$$\{1\} = G_q \subset \cdots \subset G_2 \subset G_1 \subset G_0 = S_n$$

such that G_{j+1} is normal in G_j and G_j/G_{j+1} is the cyclic group of prime order.

There is an easier formulation using abelian groups. If H is a finite abelian group, then it can be represented as

$$H = \bigoplus_j (\mathbb{Z}/p_j^{\ell_j}\mathbb{Z}).$$

Then you can easily find a subgroup $H' \subset H$ such that H/H' is cyclic of prime order. This shows that you can replace the condition "G_j/G_{j+1} is the cyclic group of prime order" with "G_j/G_{j+1} is an abelian group."

You can further shorten the tower by doing the following. You can consider the minimal $G_{j+1} \subset G_j$ such that G_j/G_{j+1} is abelian. This minimal subgroup G_{j+1} is actually the commutator subgroup of G_j, which is the group generated by elements of form $aba^{-1}b^{-1}$ with $a, b \in G$. So such a tower exists if and only if the commutator of the commutator of the commutator of the $\cdots$ of the commutator is $\{1\}$.

Suppose that $n \ge 5$. First, if you consider the commutator subgroup of S_n, it is A_n.

Theorem 16.2. *For any $n \geq 5$, the group A_n has no proper normal subgroup except for $\{1\}$.*

Proof. Suppose that there is a proper normal subgroup N of A_n. Consider the "least disturbing permutation" inside N, that is the permutation with most fixed point (except for the identity). We will first show that this is $\sigma = (123)$. Apply the inner automorphism of A_n on σ, and then we get all 3-cycles. Then N should contain all permutations generated by 3-cycles. But using "ladder diagrams," we see that every even permutation is generated by 3-cycles. This contradicts our assumption that N is proper in A_n.

Now all we are left with is proving that σ indeed is a 3-cycle. It is obvious that σ is a composition of distinct cycles. Then σ should either have a cycle with at least 3 elements, or have at least 2 disjoint transposition. We claim that if σ is not a 3-cycle, then there is a less disturbing permutation. It can be done by observing a thing like $\tau\sigma\tau^{-1}\sigma^{-1}$ for like $\tau = (345)$. I will check this later. □

16.3 Digression: Primitive element theorem

There is a recipe for computing the Galois group. Suppose that E/F is a Galois extension and let $[E : F] = n$. Also for convenience, assume that F has characteristic zero. We want to reduce it into a simple extension. This means that one step is enough to construct F. That is, there exists a $\xi \in E$ such that $E = F(\xi)$. Then for the minimal polynomial $f_\xi(x)$ of ξ, the field $E = F[X]/f_\xi(X)F[X]$. This kind of element is called primitive.

The idea is roughly the following. Suppose that $E = F(\xi_1, \ldots, \xi_k)$. Then there is a sufficiently "generic" choice of $a_1, \ldots, a_k$ such that $\xi = \sum_{j=1}^{k} a_j\xi_j$ is a primitive element. This is possible, because F is infinite.

Now you can compute the Galois group from this using this fact. Besides ξ, there exists other roots, and let them by $\xi = \xi_1, \ldots, \xi_n$. We want $\xi_j = h_j(\xi)$ to be true for some $h_j(X) \in F[X]$. Let $G = \mathrm{Gal}(E/F)$ and then we will have $|G| = n = [E : F]$. Let $G = \{\gamma_1, \ldots, \gamma_n\}$, where $\gamma_1 = \mathrm{id}_E$ and $\gamma_j(\xi) = \xi_j$. Then γ_j corresponds to the polynomial h_j. That is, you can compute the Galois group by composing the polynomial h_j modulo $f(x)$.

17 October 29, 2015

We continue on our discussion of the insolvability of polynomial of degree $n \geq 5$. Using Galois theory, we showed that a solution corresponds to a tower of groups

$$\{1\} = G_q \subset \cdots \subset G_{j+1} \subset G_j \subset \cdots \subset G_1 \subset G_0 = S_n$$

where G_{j+1} is normal in G_j and G_j/G_{j+1} is cyclic of order d_j. The the question reduces to whether S_n is solvable or not. In retrospect, we haven't used a lot of tricks. We used some estimates, and the trick of extending fields.

17.1 Insolvability of S_n

We were showing that

Theorem 17.1. *A_n is simple, i.e., has no proper normal subgroup other than 1, for $n \geq 5$.*

Proof. Suppose the contrary, and assume that there is a normal subgroup $\{1\} \neq N \neq A_n$. Note that A_n is generated by 3-cycles. This is because every cycle is a composition of a even number of transposition, and $(13)(24) = (234) \cdot (123)$.

If N contains a 3-cycle, then it contains all 3-cycles. This is because if ρ is a 3-cycle then $\sigma\rho\sigma^{-1}$ can be any other 3-cycle for $\sigma \in A_n$. Then N would just be A_n. Therefore it suffices to show that N contains a 3-cycle.

Consider the $1 \neq \sigma \in N$ with the most number of fixed points. Obviously, σ cannot be a 4-cycle since a 4-cycle is not in A_n. This means that σ will have the form of either

$$(123\cdots)\cdots \quad \text{or} \quad (12)(34)\cdots.$$

Let $\tau = (345)$. Let us look at the permutation $\tau\sigma\tau^{-1}\sigma^{-1}$. If you write it out, you will see that it is not the identity, and that it has an additional fixed point. So σ should be the 3-cycle, and then we arrive at a contradiction. □

Then people started to look at subclass of quintic polynomials which are solvable. But it is not simple, and I don't want to get into this topic.

17.2 Galois group of $x^{p+1} - sx - t$

This was problem 8 in the problem set.

Problem 8. *Let F_0 be a field of characteristics p (where p is an odd prime) and $F = F_0(s,t)$ where s and t are two independent indeterminates over F_0. Let $f(x) \in F(x)$ be the polynomial $x^{p+1} - sx - t$ with coefficients in F. Show that the Galois group of the polynomial $f(x)$ over F is isomorphic to the group of all linear fractional transformations*

$$x \mapsto \frac{ax+b}{cx+d} \quad (\text{where } a,b,c,d \in \mathbb{Z}/p\mathbb{Z} \text{ with } ad - bc \neq 0)$$

on $\mathbb{Z}/p\mathbb{Z}$.

Consider the field $F = (\mathbb{Z}/p\mathbb{Z})(s,t)$. Our goal is showing that the Galois group, which acts on F, is isomorphic to some group which acts on $\mathbb{Z}/p\mathbb{Z}$. So we need to bring F down to $\mathbb{Z}/p\mathbb{Z}$. One important trick is that if you want to show that $\alpha \in \mathbb{Z}/p\mathbb{Z}$ then you can alternatively show $\alpha^{p-1} = 1$.

Let me try to explain the geometry.

- There is the Euclidean geometry. In this geometry, lengths are fixed, so there is only rigid motion, and there are only the translations $x \mapsto x + \ell$. For any two points p and q, the length $p - q$ is invariant. The frame of reference is defined by only 1 point.
- There is also the affine geometry. You can translate, but also rescale. So the translations look like $x \mapsto ax + b$. In this case, the ratio $(q-p)/(r-p)$ is invariant. The frame of reference is defined by 2 points.
- In projective geometry, the maps are factional translations. This is called projective geometry, because it is analogous to projective a line to another line from a light source. The cross ratio

$$\frac{q-s}{r-s}\Big/\frac{q-p}{r-p} = \frac{(q-s)(r-p)}{(q-p)(r-s)}$$

  is preserved in this case. The frame of reference is defined by 3 points in this case.

You can actually show that preserving the cross ratio is equivalent to the transformation being a linear fractional transformation. That is,

$$\frac{x-p}{x-q}\Big/\frac{r-p}{r-q} = \frac{\tilde{x}-\tilde{p}}{\tilde{x}-\tilde{q}}\Big/\frac{\tilde{r}-\tilde{p}}{\tilde{r}-\tilde{q}}$$

is equivalent to

$$\tilde{x} = \frac{ax+b}{cx+d}$$

for some a, b, c, d.

Let γ be a automorphism of the splitting field. Because an extension E of F is also a vector space over F with degree $n = [E : F]$, then you can consider γ as a matrix. This gives a injection

$$G \hookrightarrow GL_n(F).$$

But this looses a lot of information, because it does not contain any information about multiplication. So there is another embedding

$$G \hookrightarrow S_n.$$

We do a similar thing to embed it into the group of linear fractional transformations.

Let α, β, γ be three roots of $x^{p+1} - sx - t = 0$. We want to mirror the action of the Galois group onto $\mathbb{Z}/p\mathbb{Z}$. We use the cross-ration as some kind of

coordinate. We already have some frame α, β, γ and a automorphism σ changes the frame to $\sigma(\alpha), \sigma(\beta), \sigma(\gamma)$. The coordinates change by the map

$$\frac{x-\alpha}{x-\beta}\Big/\frac{\gamma-\alpha}{\gamma-\beta} \mapsto \frac{\sigma(x)-\sigma(\alpha)}{\sigma(x)-\sigma(\beta)}\Big/\frac{\sigma(\gamma)-\sigma(\alpha)}{\sigma(\gamma)-\sigma(\beta)}.$$

That is, the σ acts on the coordinates in the same manner.

But actually, the coordinates are in $\mathbb{Z}/p\mathbb{Z}$ for roots x. This can be proved by computing. If $\gamma^{p+1}-s\gamma-t=0$ and $\alpha^{p+1}-s\alpha-t=0$ then $\gamma^{p+1}-\alpha^{p+1}=s(\gamma-\alpha)$. Likewise, we have $\gamma^{p+1}-\beta p+1=s(\gamma-\beta)$. Then

$$\alpha(\gamma^{p-1}+\gamma^{p-2}\alpha+\cdots+\gamma\alpha^{p-2}+\alpha^{p-1}) = \beta(\gamma^{p-1}+\gamma^{p-2}\beta+\cdots+\gamma\beta^{p-2}+\beta^{p-1}).$$

Then because we are working in characteristic p, we have

$$\alpha(\gamma-\alpha)^{p-1} = \beta(\gamma-\beta)^{p-1}$$

and

$$\left(\frac{\gamma-\alpha}{\gamma-\beta}\right)^{p-1} = \frac{\beta}{\alpha}.$$

This shows that the $(p-1)$th power of the cross ratio is always 1, and hence the cross ratio is always inside $\mathbb{Z}/p\mathbb{Z}$. This brings down F to $\mathbb{Z}/p\mathbb{Z}$.

Why care about the polynomial $x^{p+1}-sx-t$? It started before Abel and Galois. When solving the equation $x^n-\sigma_1 x^{n-1}+\cdots+(-1)^n\sigma_n=0$, you can get rid of one term by translating $x \mapsto x+a$. But this introduces only one degree of freedom. So people started to do other translations, and somehow got down to x^5-sx-t. This is why we care about such things.

17.3 Constructing a regular polygon

You are allowed to use only an unmarked straightedge and compass. We want to construct a regular n-gon, that is, all the vertices. Then getting to a point by only a ruler and a compass is actually considering the tower of fields, where only square roots are allowed. Let $z=e^{2\pi i/17}$. You want to solve the equation $z^{17}=1$ using only square roots. How will you be able to do this?

This is an alternative formulation. Let $F_0=\mathbb{Q}$, and let $F_1=F_0[X]/(X^2+a_0X+b_0)F_0[X]$. Then let $F_2=F_1[X]/(X^2+a_1X+b_1)F_1[X]$. We keep constructing extension fields until z is in F_q. The problem is constructing the intermediate fields.

Gauss's idea is to start from $z^17=1$. Then we have

$$z^{16}+z^{15}+\cdots+z+1=0.$$

And then you break $z^{16}+\cdots+z$ into two parts of eight terms such that the product is inside $\mathbb{Q}$. His idea is breaking up into

$$z^{3^0}+z^{3^2}+\cdots+z^{3^{14}} \quad \text{and} \quad z^{3^1}+\cdots+z^{3^{15}}.$$

Then you break up each into two parts, and then do the similar things. I will finish doing this next time.

18 November 3, 2015

There was an in-class midterm this day. There were 4 problems and we were supposed to solve them in one and a half hour.

18.1 Midterm

This was the hardest problem in the exam.

Problem[3]**.** *(Wronskian and Linear Dependency)* Let F be a field of characteristic zero and let x be an indeterminate. Define the F-linear operator $D : F(x) \to F(x)$ by first defining D as an F-linear map form $F[x]$ to $F[x]$ which sends

$$P(x) = \sum_{j=0}^{n} a_j x^j \quad \text{with} \quad a_0, \ldots, a_n \in F$$

to

$$(DP)(x) = \sum_{j=1}^{n} j a_j x^{j-1}$$

and then defining

$$D\left(\frac{P}{Q}\right) = \frac{(DP)Q - (DQ)P}{Q^2}$$

for $P(x), Q(x) \in F[x]$ with $Q(x)$ being nonzero element of $F[x]$. *Assume as known* the well-definedness of $D : F(x) \to F(x)$ described above and *assume also as known* its derivation property that

$$D(f_1 \cdots f_n) = \sum_{j=1}^{n} f_1 \cdots f_{j-1} (Df_j) f_{j+1} \cdots f_n$$

for $f_1, \ldots, f_n \in F(x)$.
(b) Let $n \geq 2$ be an integer. Let

$$\mathbf{f} = (f_1, \ldots, f_n) \in F(x)^{\oplus n}$$

and

$$D^k \mathbf{f} = (D^k f_1, \ldots, D^k f_n) \in F(x)^{\oplus n}$$

for $1 \leq k \leq n-1$. Show that the set $f_1, \ldots, f_n$ in $F(x)$ is F-linearly dependent if and only if

$$\mathbf{f} \wedge D\mathbf{f} \wedge \cdots \wedge D^{n-1}\mathbf{f}$$

is the zero element of $\bigwedge^n (F(x)^{\oplus n})$, where the exterior product $\bigwedge^n (F(x)^{\oplus n})$ is taken with $F(x)^{\oplus n}$ regarded as a vector space of dimension n over the field $F(x)$.

[3]This was Problem 1 in the exam. Part (a) was more or less same as the case $n = 2$.

19 November 5, 2015

Today I will continue the discussion on tower of fields.

19.1 Gauss's straightedge-and-compass construction of a regular polygon of 17 sides

The main point of the construction is to look at $z = e^{2\pi i}17$. The rules are to first take 0 and 1. and rational functions of the points, and square roots. That is, we need to find a field extension

$$\begin{array}{c} F_q = F_{q-1}(a_{q-1}) \\ \cup \\ \vdots \\ \cup \\ F_j = F_{j-1}(a_{j-1}) \\ \cup \\ \vdots \\ \cup \\ F_1 = F_0(a_0) \\ \cup \\ F_0 = \mathbb{Q} \end{array}$$

such that $z \in F_q$.

Because $z^17 = 1$ and $z \neq 1$, we have $1 + z + \cdots + z^{16} = 0$. That is,

$$z + z^2 + \cdots + z^{16} = -1.$$

We break this into two pieces x_1 and x_2 such that $x_1 \cdot x_2$ is computable in $\mathbb{Q}$. Then x_1 and x_2 will be the roots of a quadratic polynomials and hence will be in a degree 2 extension. Next break $x_1 = y_1 + y_2$ and $x_2 = y_3 + y_4$ and etcetera.

Gauss realized that if you break it up into $z + z^3 + \cdots + z^{15}$ and $z^2 + \cdots + z^{16}$ it doesn't work. So he did it in a multiplicative way to

$$z^{3^0} + z^{3^2} + \cdots + z^{3^14} \quad \text{and} \quad z^{3^1} + z^{3^3} + \cdots + z^{3^15}$$

by observing that $\{3^\ell\}$ is exactly $(\mathbb{Z}/17\mathbb{Z})^*$.

We will first do it by brute force. It is not that bad. We have

$$x_1 = \sum_{k=0}^{3} \left(z^{3^{2k}} + z^{-3^{2k}}\right) \quad \text{and} \quad x_2 = \sum_{k=0}^{3} \left(z^{3^{2k+1}} + z^{-3^{2k+1}}\right).$$

Checking the numbers, we see that

$$x_1 = Z_1 + Z_8 + Z_4 + Z_2 \quad \text{and} \quad x_2 = Z_3 + Z_7 + Z_5 + Z_6$$

where $Z_j = z^j + z^{-j}$. Because $Z_j Z_k = Z_{j+k} + Z_{|j-k|}$, we can multiply the two things and get

$$\begin{aligned} x_1 x_2 &= (Z_1 + Z_8 + Z_4 + Z_2)(Z_3 + Z_7 + Z_5 + Z_6) \\ &= (Z_4 + Z_2) + (Z_6 + Z_5) + (Z_7 + Z_1) + (Z_5 + Z_1) \\ &\quad + (Z_8 + Z_6) + (Z_2 + Z_1) + (Z_6 + Z_3) + (Z_8 + Z_5) \\ &\quad + (Z_6 + Z_4) + (Z_4 + Z_3) + (Z_8 + Z_1) + (Z_7 + Z_3) \\ &\quad + (Z_7 + Z_5) + (Z_3 + Z_2) + (Z_7 + Z_2) + (Z_8 + Z_4) \\ &= -4 \end{aligned}$$

and you can see that every term appears four times. Hence x_1 and x_2 are the roots of $X^2 + X - 4 = 0$ and then $X = (-1 + \sqrt{17})/2$.

We now break x_1 to $y_1 + y_2$. We let

$$y_1 = \sum_{j=0}^{1} Z_{3^{4j}} \quad \text{and} \quad y_2 = \sum_{j=0}^{1} Z_{3^{2(2j+1)}}.$$

Then actually $y_1 = Z_1 + Z_4$ and $y_2 = Z_8 + Z_2$. The sum is something we already know. The product is

$$\begin{aligned} y_1 y_2 &= (Z_1 + Z_4)(Z_8 + Z_2) \\ &= (Z_8 + Z_7) + (Z_5 + Z_4) + (Z_3 + Z_1) + (Z_6 + Z_2) = -1. \end{aligned}$$

We see that y_1 and y_2 are the roots of $X^2 - \frac{-1+\sqrt{17}}{2} X - 1 = 0$.

Likewise, we can let $y_3 = Z_3 + Z_5$ and $y_4 = Z_7 + Z_6$ and get

$$y_3 y_4 = Z_7 + Z_4 + Z_5 + Z_2 + Z_8 + Z_3 + Z_6 + Z_1 = -1.$$

Now $y_1 = Z_1 + Z_4$ and $Z_1 Z_4 = Z_5 + Z_3 = y_3$. This shows that Z_1 is computable, and then $Z_1 = z + z^{-1}$ and thus z is computable.

In general, for this method to work, the prime p should be of the form $p = 2^n + 1$. If $n = k\ell$ where ℓ is odd, then one can factor

$$p = 2^{k\ell} + 1 = (2^k + 1)(\cdots).$$

So ℓ should be just 1 and thus $p = 2^{2^k} + 1$.

We can do this with less brute force. We have

$$1 \equiv 3 - 2 \equiv 3^1 - 3^{14} \equiv 3^1 + 3^6 \pmod{17}.$$

The claim is that every number is covered in the form $3^{\text{odd}} + 3^{\text{even}}$ exactly four times. This can be checking that 1 is covered exactly four times and multiplying 3^k to see that 3^k is covered exactly four times.

So how do you actually add, subtract, multiply, divide complex numbers with straightedge and compass? Adding and subtracting is just constructing parallelograms; multiplying and dividing is just drawing similar triangles. Lastly, to get the square root, you can just consider the circle with diameter $1 + a$ and draw a perpendicular line to obtain the length $\sqrt{a}$.

19.2 Lefschetz decomposition

This is extremely important in algebraic geometry and differential geometry. But the argument is purely linear algebraic.

We start out be a finite-dimensional $\mathbb{C}$-vector spaces with a Hermitian inner product. The vector space V can be regarded as an $\mathbb{R}$-vector space together with the (almost) complex structure J.

The complex structure of a real vector space is a $\mathbb{R}$-linear map $J : V \to V$ such that $J^2 = -1$. Then the scalar multiplication is defined by $(i, v) \mapsto J(v)$. We can then look at the eigenspace of J for eigenvalues i and $-i$ when we go from V to $V \otimes_{\mathbb{R}} \mathbb{C}$. A decomposition

$$V \otimes_{\mathbb{R}} \mathbb{C} = \frac{1}{2}(I - iJ)(V \otimes_{\mathbb{R}} \mathbb{C}) + \frac{1}{2}(1 + iJ)(V \otimes_{\mathbb{R}} \mathbb{C})$$

into the eigenspaces follows. The maps $\frac{1}{2}(1 + iJ)$ and $\frac{1}{2}(1 - iJ)$ are projection maps.

The space $V \otimes_{\mathbb{R}} \mathbb{C}$ has two complex structures. The multiplication by $\sqrt{-1}$ can be interpreted as a $J \otimes_{\mathbb{R}} (1_{\mathbb{C}})$ and $(1_V) \otimes_{\mathbb{R}} i$. This is not surprising, because it is the tensor product. But the surprising thing is that the two complex structures are equal on the eigenspace

$$\frac{1}{2}(I - iJ)(V \otimes_{\mathbb{R}} \mathbb{C}).$$

So in the literature, people just write

$$V \otimes_{\mathbb{R}} \mathbb{C} = V \oplus \bar{V}.$$

There was a Hermitian inner product. Let $v \in V$ and consider the norm $\|v\|$. Then the square of the norm can be written as $\|v\| = (v, v)$ for some inner product if and only if the parallelogram law

$$\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2$$

holds. A Hermitian inner product also satisfies $(iu, v) = i(u, v)$ and $\overline{(u, v)} = (v, u)$.

At first, V was a $\mathbb{C}$-vector space. We apply the forgetful functor to forget the $\mathbb{C}$-vector space structure and consider it as a $\mathbb{R}$-vector space, but keep the complex structure J. Then by making it to $V \otimes_{\mathbb{R}} \mathbb{C}$ again, we can now consider eigenspaces. If $(\cdot, \cdot)$ was a Hermitian inner product on V, then we can extend $(\cdot, \cdot)$ by $\mathbb{C}$-bilinearity to $V \otimes_{\mathbb{R}} \mathbb{C}$.

Now we have

$$\bigwedge^*(V \otimes_{\mathbb{R}} \mathbb{C}) = \bigwedge^*(V \oplus \bar{V})$$

where $\bigwedge^* = \bigoplus_p \bigwedge^p$. The wedge product $V \bigwedge \bar{V}$ is contained in the space $\bigwedge^2(V \oplus \bar{V})$.

20 November 10, 2015

We were looking at Lefschetz decomposition. Normally it is not part of the abstract algebra curriculum, but it is important in algebraic geometry and complex analysis in higher dimension.

20.1 Setting of the Lefschetz decomposition

We set V as a vector space over $\mathbb{R}$ with a complex structure $J \in \mathrm{Hom}_{\mathbb{R}}(V, V)$ with $J^2 = -1$. This is not that much, and the important thing is the inner product $(\cdot, \cdot)_V : V \times V \to \mathbb{R}$ which is *compatible with* J. Under this situation, we have a matrix, which is nilpotent, on a finite dimensional vector space, which is the exterior algebra V. And we ask for a normal form for the matrix, which is compatible with both the bidegrees of the exterior algebra of $V \otimes \mathbb{C}$. Let me explain what this means.

If you have a vector space V over F, and another W over F, we can think of the tensor product $V \otimes_F W$. If $\Phi : V \to V$ and $\Psi : W \to W$ are both F-linear, then the map $\Phi \otimes_F \Psi : V \otimes_F W \to V \otimes_F W$ defined by the commutative diagram

$$\begin{array}{ccc} V^* & \xrightarrow{f} & W \\ {\scriptstyle \Phi^*}\uparrow & & \downarrow{\scriptstyle \Psi} \\ V^* & \xrightarrow{(\Phi\otimes\Psi)(f)} & W \end{array}$$

or alternatively,

$$(\Phi \otimes \Psi)(v \otimes w) = \Phi(v)\Psi(w).$$

Now if we have $V \otimes_{\mathbb{R}} \mathbb{C}$, there are two complex structures J_V defined on V and $J_{\mathbb{C}}$ defined on $\mathbb{C}$, which is multiplication by i. Now we can extend the map J_V on V to $J_V \otimes \mathrm{Id}_{\mathbb{C}} : V \otimes \mathbb{C} \to V \otimes \mathbb{C}$. Then we will have

$$(J_V \otimes \mathrm{Id}_{\mathbb{C}})^2 = -\mathrm{Id}_{V\otimes\mathbb{C}}$$

and likewise, we have another complex structure

$$(\mathrm{Id}_V \otimes J_{\mathbb{C}})^2 = -\mathrm{Id}_{V\otimes\mathbb{C}}\,.$$

Also, if we let $\pi^+ = \frac{1}{2}(1 - iJ)$ and $\pi^- = \frac{1}{2}(1 + iJ)$ we have the decomposition

$$V \otimes \mathbb{C} = \mathrm{Im}\,\pi^+ \oplus \mathrm{Im}\,\pi^-$$

where π^+ and π^- are projections. The spaces are the eigenspaces of J_V. On the space $\mathrm{Im}\,\pi^+$, the two complex structures $J_V \otimes \mathrm{Id}_{\mathbb{C}}$ and $\mathrm{Id}_V \otimes J_{\mathbb{C}}$ agree. Because there is a injection $V \hookrightarrow V \otimes \mathbb{C}$ and a projection $\pi^+ : V \otimes \mathbb{C} \to \mathrm{Im}\,\pi^+$, we get a bijection $V \to \mathrm{Im}\,\pi^+$. Because, the complex structures agree, this can be seen as a isomorphism of complex vector spaces. So we write $\pi^+ \simeq V$. Likewise, we can write $\pi^- \simeq \bar{V}$. This is why people write just

$$V \otimes \mathbb{C} = V \oplus \bar{V}.$$

Now from this we get

$$\bigwedge^k(V \otimes \mathbb{C}) = \bigoplus_{p+q=k} ((\bigwedge^p V)\bigwedge(\bigwedge^q V)).$$

We define the **exterior algebra** as

$$\bigwedge^*(V \otimes \mathbb{C}) = \bigoplus_{k=0}^{2n} \bigwedge^k(V \otimes \mathbb{C}) = \bigoplus_{0\le p,q\le n} (\bigwedge^p V)\bigwedge(\bigwedge^q \bar{V}).$$

20.2 Inner product on the complexified vector space

Given an inner product $(\cdot,\cdot)_V : V \times V \to \mathbb{R}$, we can extend it by $\mathbb{C}$-bilinearity to

$$(\cdot,\cdot)_{V\otimes\mathbb{C}} : (V \otimes \mathbb{C}) \times (V \otimes \mathbb{C}) \to \mathbb{C}.$$

Definition 20.1. An inner product $(\cdot,\cdot)_V$ is said to be **compatible** with J if

$$(Ju, Jv)_V = (u, v)_V.$$

The extended inner product $(\cdot,\cdot)_{V\otimes\mathbb{C}}$ can be broken up into 4 pieces by the decomposition

$$(V \otimes \mathbb{C}) \times (V \otimes \mathbb{C}) = (V \times V) \oplus (V \times \bar{V}) \oplus (\bar{V} \times V) \oplus (\bar{V} \times \bar{V})$$

to $(\cdot,\cdot)_{V,V}$, $(\cdot,\cdot)_{V,\bar{V}}$, $(\cdot,\cdot)_{\bar{V},V}$, and $(\cdot,\cdot)_{\bar{V},\bar{V}}$.

If we have the compatibility condition, we will have

$$(\tilde{u}, \tilde{v})_{V,V} = (J\tilde{u}, J\tilde{v})_{V,V} = -(\tilde{u}, \tilde{v})_{V,V}.$$

Then we have $(\tilde{u}, \tilde{v})_{V,V}$ for any $\tilde{u}, \tilde{v}$, and hence $(\cdot,\cdot)_{V,V} \equiv 0$. Likewise, we will have $(\cdot,\cdot)_{\bar{V},\bar{V}} \equiv 0$. Moreover, by symmetry of the inner product, we have $(\tilde{u}, \tilde{v})_{V,\bar{V}} = (\tilde{v}, \tilde{u})_{\bar{V},V}$. Then we can define a Hermitian inner product by

$$(w_1, w_2) = (w_1, \bar{w}_2)_{V,\bar{V}}$$

for $w_1, w_2 \in V$. I will assign verifying that it is a Hermitian as a homework assignment.

Definition 20.2. An orthonormal basis for a real vector space V is said to be **compatible** with J if it is of the form

$$\xi_1, J\xi_1, \xi_2, J\xi_2, \ldots, \xi_n, J\xi_n,$$

where $\dim_{\mathbb{R}} V = 2n$.

There always exists an orthonormal basis. We can actually construct it inductively. First pick any ξ_1 of unit length, and automatically we have $J\xi_1$. Then because

$$(\xi_1, J\xi_1)_V = (J\xi_1, J(J\xi_1))_V = (J\xi_1, -\xi_1)_V = -(\xi_1, J\xi_1)_V,$$

we see that ξ is perpendicular with $J\xi_1$. You then pick ξ_2, and etcetera.

If $\xi_1, J\xi_1, \dots, \xi_n, J\xi_n$ is a basis, we can easily construct a basis of $V \otimes \mathbb{C}$ from it. If we project it, we get

$$\xi_1 - iJ\xi_1, \xi_2 - iJ\xi_2, \dots, \xi_n - iJ\xi_n$$

in $\operatorname{Im} \pi^+$, and one can check that it is a orthogonal basis of $\operatorname{Im} \pi^+$ by calculating

$$(\xi_k - iJ\xi_k, \xi_j + iJ\xi_j)_{V \otimes \mathbb{C}}.$$

So we write

$$e^j = \xi_j - iJ\xi_j, \quad \bar{e}^j = \xi_j + iJ\xi_j.$$

Let

$$\omega = \frac{i}{2} \sum_{j=1}^{n} e^j \wedge \bar{e}^j \in V \bigwedge \bar{V}.$$

Why is there a factor $i/2$? This is because everything started from differential geometry. If $z_j = x_j + \sqrt{-1} y_j$ then in the $V = \oplus_{j=1}^n \mathbb{C} dz_j$ we have

$$\frac{i}{2}(dz_j \wedge d\bar{z}_j) = \frac{i}{2}((dx_j + i dy_j) \bigwedge (dx_j - i dy_j)) = dx_j \wedge dy_j.$$

20.3 Lefschetz operator and Hodge star operator

Back to our discussion, the multiplication by ω gives a $\mathbb{C}$-linear map

$$L : \bigwedge\nolimits^*(V \otimes \mathbb{C}) \to \bigwedge\nolimits^*(V \otimes \mathbb{C}).$$

This is the Lefschetz operator.[4] This is clearly nilpotent because if you wedge ω many times it goes to zero.

Let us try to actually try the decomposition. We have this

$$L = \text{Lefschetz operation}$$

which is the (exterior) multiplication by ω. The exterior algebra is

$$\bigwedge\nolimits^*(V \otimes \mathbb{C}) = \bigoplus_{0 \le p,q \le n} (\bigwedge\nolimits^p V) \bigwedge (\bigwedge\nolimits^q \bar{V}),$$

and we will just call this component (p, q). Because L is nilpotent, we see that all the eigenvalues of L are 0. Also, L sends $(p, q) \to (p + 1, q + 1)$.

There are two tools we can use. The first one is the contraction operator Λ. This operator, which we have defined many weeks before, sends $(p+1, q+1) \to (p, q)$. The useful thing about this contraction is that the commutator is

$$[L, \Lambda] = c.$$

[4] The motivation for this whole operator is from algebraic geometry. If we have a variety but is too complicated, we can cut it with a hyperplane. This is same as introducing a new linear equation. The square of this linear equation can be roughly regarded as ω.

The other tool is the Hodge star operator for a real vector space V with an inner product, assumed that there is a choice of orientation. If $e_1, \dots, e_m$ is a orthonormal basis of V, then $\bigwedge^m V \simeq \mathbb{R}$. The orientation is a element in this space with unit length, for instance $e_1 \wedge \cdots \wedge e_m$. Then the Hodge start operator $* : \bigwedge^p V \to \bigwedge^{m-p} V$ is defined by

$$(\varphi, \psi)(*1) = \varphi \wedge (*\psi),$$

where $*1$ is the orientation.

Now let us consider the whole thing in the context of our setting. If $\dim_{\mathbb{R}} V = 2n = m$ then the star operator $* : \bigwedge^k V \to \bigwedge^{2n-k} V$ is defined using the orientation

$$(\tfrac{i}{2} e^1 \wedge \bar{e}^1) \wedge \cdots \wedge (\tfrac{i}{2} e^n \wedge \bar{e}^n) = (\xi_1 \wedge J\xi_1) \wedge \cdots \wedge (\xi_n \wedge J\xi_n) \in \bigwedge^{2n} V.$$

By $\mathbb{C}$-linearity, we can extend this to $* : \bigwedge^k (V \otimes \mathbb{C}) \to \bigwedge^{2n-k} (V \otimes \mathbb{C})$ and then $*$ will send

$$* : (\textstyle\bigwedge^p V) \bigwedge (\bigwedge^q \bar{V}) \to (\bigwedge^{n-q} V) \bigwedge (\bigwedge^p \bar{V}).$$

The contraction Λ is conjugate to L (up to some normalizing factor by the Hodge star operator. That is,

$$\Lambda = (\text{const}) * L *^{-1}.$$

It follows from this fact that the (generalized) eigenspaces for L are the same as those for Λ. This makes things much easier, because Λ is simpler. The elements in the kernel $\operatorname{Ker} \Lambda$ are called the **primitive elements**.

20.4 Statement of the Lefschetz decomposition

Now we finally get to the Lefschetz decomposition. The space

$$\bigwedge^* (V \otimes \mathbb{C}) = \bigoplus_{k=0}^{2n} \bigwedge^k (V \otimes \mathbb{C})$$

has midpoint n. First thing to observe is that if $k > n$ then there is no primitive element in $\bigwedge^k (V \otimes \mathbb{C})$. This is because among the indices of the $\bigwedge^p V$ part an the indices of the $\bigwedge^q \bar{V}$ part there should be some index in common.

Let $m = (k-n)^+ = \max\{k-n, 0\}$. Then any $\varphi \in \bigwedge^k (V \otimes \mathbb{C})$ can be decomposed into

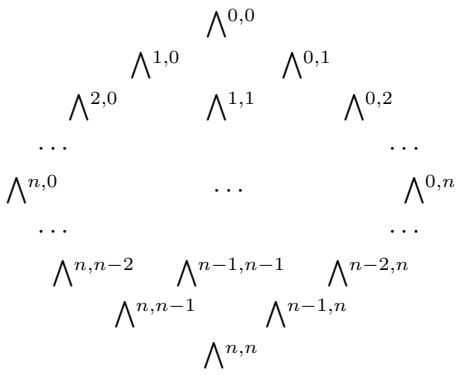
$$\varphi = \sum_{\ell \geq m} L^\ell \varphi_\ell$$

where $\varphi_\ell \in \bigwedge^{k-2\ell} (V \otimes \mathbb{C})$ is primitive. This is the statement of the theorem.

The exterior algebra breaks up into

$$\bigwedge^* (V \otimes \mathbb{C}) = \bigoplus_{k=0}^{2n} \bigwedge^k (V \otimes \mathbb{C}) = \bigoplus_{0 \leq p,q \leq n} \bigwedge^{p,q}.$$

Then we have the **Hodge diamond**:



This is just a visualization of what the operators do on the space. The star operator reflects the spaces with respect to the horizontal axes. Complex conjugation flips the diaper with respect to the vertical axes. The operator L moves things down, and Λ moves things up.

21 November 12, 2015

Before I start I ought to tell you about the final. It will be a take-home final during the reading period.

21.1 Overview of Lefschetz decomposition

So we are doing this Lefschetz decomposition, and I posted the notes in great detail. The setting, as I explained, is you start out with a $2n$-dimensional real vector space V. There are two additional structures: the almost complex structure $J : V \to V$ for which $J^2 = -1$, and an inner product $(\cdot,\cdot)_V : V \times V \to \mathbb{R}$. The inner product should be J-invariant, i.e., $(u, v)_V = (Ju, Jv)_V$ for any $u, v \in V$. The Lefschetz decomposition is about the normal form for the linear transformation defined by the inner product which is compatible with J.

In algebraic geometry, one of the important tools is the Lefschetz theorem. There are two part: the theory of harmonic forms, and the linear algebra part. But the hard part is the linear algebra, and we are doing this part.

The inner product is a kind of an operator. If you give two vectors u and v, the inner product gives a scalar. Now the inner product

$$(u, v) \mapsto (u, Jv)_V$$

is skew-symmetric, and therefore it can be considered as an element of $\bigwedge^2 V$. The multiplication with this element is the Lefschetz operator, and because it shifts the dimension, it is considered as a map $\bigwedge^* V \to \bigwedge^* V$.

Now we can extend this to $L : \bigwedge^*(V \otimes \mathbb{C}) \to \bigwedge^*(V \otimes \mathbb{C})$. If $m = 2n$ then we would have $L^{m+1} = 0$. So all eigenvalues would all be 0. We can also think of generalized eigenspaces, by considering $\ker L^\nu$.

We also have another operation $* : \bigwedge^* V \to \bigwedge^* V$ defined by

$$(\phi, \psi)(*1) = \phi \wedge (*\psi).$$

This sends $* : \bigwedge^p V \to \bigwedge^{2n-p} V$. This identifies $\psi \in \bigwedge^p V$ with $*\psi \in \bigwedge^{2n-p} V$. Suppose that we are using the orientation $\alpha_1 \wedge \cdots \wedge \alpha_{2n}$, where $\alpha_1, \ldots, \alpha_{2n}$ is an orthonormal basis. Then $*$ will send

$$\alpha_1 \wedge \cdots \wedge \alpha^p \mapsto \alpha_{p+1} \wedge \cdots \wedge \alpha_{2n}.$$

Now we have the construction operator Λ, and it is conjugate to L under $*$ by

$$\Lambda = (\text{const}) *^{-1} L *.$$

So understanding the normal form of L is same as understanding the normal form of Λ. But people like Λ better, so we call the elements of kernel of Λ the primitive elements. The Lefschetz decomposition decomposes the base space $\bigwedge^*(V \otimes \mathbb{C})$ according to the eigenspaces of Λ.

21.2 Notations and basic formulas

We are going to use the orientation

$$(\xi_1 \wedge J\xi^1) \wedge \cdots \wedge (\xi^n J\xi^n) = (\tfrac{\sqrt{-1}}{2} e^1 \wedge \bar{e}^1) \wedge \cdots \wedge (\tfrac{\sqrt{-1}}{2} e^n \wedge \bar{e}^n)$$

for $*$, where $e^j = \xi^j + \sqrt{-1} J\xi^j$. Let $A = (\alpha_1, \ldots, \alpha_a)$ and $B = (\beta_1, \ldots, \beta_b)$ and $M = (\mu_1, \ldots, \mu_m)$ be disjoint ordered subsets of $\{1, 2, \ldots, n\}$, we denote

$$e^A = e^{\alpha_1} \wedge \cdots \wedge e^{\alpha_a}, \quad \omega^M = e^{\mu_1} \wedge \bar{e}^{\mu_1} \wedge \cdots \wedge e^{\mu_m} \wedge \bar{e}^{\mu_m}.$$

So for instance,

$$e^A \wedge \bar{e}^B \wedge \omega^M$$

is in $\bigwedge^{p,q}$, where $p = m + a$ and $q = m + b$. This notation is good for Λ and L, but is bad for $*$.

To compute the $*$, we let A_p and A_{n-p} be a partition of $\{1, 2, \ldots, n\}$, and likewise let B_q and B_{n-q} also be a partition. For the basis elements, if we have $\varphi = e^{A_p} \wedge \bar{e}^{B_q}$, we will get

$$*\varphi = C_{p,q} e^{B_{n-q}} \wedge \bar{e}^{A_{n-p}}.$$

The difficult thing is determining the constant $C_{p,q}$. From the definition we have

$$(\varphi, \varphi)(*1) = \varphi \wedge \overline{*\varphi}.$$

Because we have

$$(e^j, e^j)_{\text{Herm}} = (e^j, \bar{e}^j)_V = (\xi^j + \sqrt{-1} J\xi^j, \xi^j - \sqrt{-1}\xi^j) = 2,$$

we have, for the left hand side,

$$(\varphi, \varphi)(*1) = 2^{p+q} (\tfrac{i}{2}) (e^1 \wedge \bar{e}^1) \wedge \cdots \wedge (e^n \wedge \bar{e}^n).$$

Then for the right hand side, we have

$$\begin{aligned} \varphi \wedge \overline{*\varphi} &= e^{A_p} \wedge \bar{e}^{B_q} \wedge \overline{C_{p,q} e^{B_{n-q}} \wedge \bar{e}^{A_{n-p}}} \\ &= \bar{C}_{p,q} (-1)^{n(n-p)} \operatorname{sgn} \begin{pmatrix} A_p & A_{n-p} \\ B_q & B_{n-q} \end{pmatrix} (-1)^{n(n-1)/2} e^1 \wedge \bar{e}^1 \wedge \cdots \wedge e^n \wedge \bar{e}^n. \end{aligned}$$

So after comparing, we get

$$C_{p,q} = \frac{i^n}{2^{n-(p+q)}} (-1)^{\frac{n(n-1)}{2} + np} \operatorname{sgn} \begin{pmatrix} A_p & A_{n-p} \\ B_q & B_{n-q} \end{pmatrix}.$$

Consequently we have

$$*(e^{A_p} \wedge \bar{e}^{B_q}) = \frac{i^n}{2^{n-(p+q)}} (-1)^{\frac{n(n-1)}{2} + np} \operatorname{sgn} \begin{pmatrix} A_p & A_{n-p} \\ B_q & B_{n-q} \end{pmatrix} e^{B_{n-q}} \wedge \bar{e}^{A_{n-p}}.$$

Now let us apply $*$ to the habitat for L and Λ. We will have

$$*(e^A \wedge \bar{e}^B \wedge \omega^M)$$
$$= (-1)^{\frac{m(m-1)}{2}+\frac{n(n-1)}{2}+n(a+m)+ab+ma+sb+\frac{s(s-1)}{2}} \frac{i^n}{2^{n-(a+b+2m)}}(e^A \wedge \bar{e}^B \wedge \omega^S),$$

where A, B, M, S form a partition of $\{1, 2, \ldots, n\}$. The horrible horrible sign comes from switching the various things around.[5]

21.3 Relations between L, Λ, and $*$

Let us check that $L* = *\Lambda$. We apply both things to $e^A \wedge \bar{e}^B \wedge \omega^M$ and check that the results agree. We have

$$L * (e^A \wedge \bar{e}^B \wedge \omega^M) = (*(e^A \wedge \bar{e}^B \wedge \omega^M)) \wedge \left(\frac{i}{2}\sum_{j=1}^{n} e^j \wedge \bar{e}^j\right)$$
$$= \sum_{j=1}^{m} (\text{const}) e^A \wedge \bar{e}^B \wedge \omega^S \wedge \omega^{\mu_j}$$

where $M = \{\mu_1, \ldots, \mu_m\}$. On the other hand, we have

$$\Lambda(e^A \wedge \bar{e}^B \wedge \omega^M) = \sum_{j=1}^{m} (e^A \wedge \bar{e}^B \wedge \omega^{M-\{\mu_j\}})$$

and hence

$$*(\Lambda(e^A \wedge \bar{e}^B \wedge \omega^M)) = \sum_{j=1}^{m} *(e^A \wedge \bar{e}^B \wedge \omega^{M-\{\mu_j\}})$$
$$= \sum_{j=1}^{m} (\text{const})_{\mu_j} (e^A \wedge \bar{e}^B \wedge \omega^S \wedge \omega^{\mu_j}).$$

One can write down the constants and check that they agree .

So we have $\Lambda = (\text{const}) *^{-1} L*$. Because you have done it before in the problem set, I skip the commutator part. We have

$$[\Lambda, L] = n - (p + q)$$

if it acts on $\bigwedge^{p,q}$.

21.4 Commutator of powers of Λ and L

Denote by Π_k the projection map of $\bigwedge^*(V \otimes \mathbb{C})$ to $\bigwedge^k(V \otimes \mathbb{C})$. Then the commutator of Λ and L can be written as

$$[\Lambda, L] = \sum_{k=0}^{2n} (n - k)\Pi_k.$$

[5] Actually I am not so sure I've written down things right.

We now look at the commutator of Λ and L^r. Because Λ and L do not commute, there is a discrepancy, and it accumulates telescopically as

$$\begin{aligned}[\Lambda, L^r] &= \sum_{\ell=0}^{r-1} L^{r-\ell-1}[\Lambda, L]L^\ell = \sum_{\ell=0}^{r-1} L^{r-\ell-1}\left(\sum_{k=0}^{2n}(n-k)\Pi_k\right)L^\ell \\ &= \sum_{\ell=0}^{r-1} L^{r-\ell-1}\left(\sum_{k=0}^{2n} L^\ell(n-k)\Pi_{k-2\ell}\right) = L^{r-1}\sum_{\ell=0}^{r-1}\sum_{k=0}^{2n}(n-k)\Pi_{k-2\ell}.\end{aligned}$$

Note that we have $\Pi_k L^\ell = L^\ell \Pi_{k-2\ell}$ because Π does nothing except projecting, and L shifts the degree. Therefore we get

$$[\Lambda, L^r] = \sum_{k=0}^{2n} r(n-k-r+1)L^{r-1}\Pi_k.$$

Now we look at how $\Lambda^s L^r$ acts on a primitive element φ in $\bigwedge^k(V\otimes\mathbb{C})$. From what we have already, we see that

$$\begin{aligned}\Lambda^s L^r\varphi &= \Lambda^{s-1}\Lambda L^r\varphi = \Lambda^{s-1}(\Lambda L^r - L^r\Lambda)\varphi \\ &= \Lambda^{s-1}[\Lambda, L^r]\varphi = r(n-k-r+1)\Lambda^{s-1}L^{r-1}\varphi.\end{aligned}$$

Then if $r \geq s$, we will have

$$\Lambda^s L^r\varphi = r(r-1)\cdots(r-s+1)(n-k-r+1)(n-k-r+2)\cdots(n-k-r+s)L^{r-s}\varphi.$$

One conclusion we can draw from this formula is:

Proposition 21.1. *There is no primitive element strictly below the middle row of the Hodge diamond.*

Proof. Suppose that there is a primitive element $\varphi \in \bigwedge^k(V\otimes\mathbb{C})$ with $k > n$. Then letting $s = r = n+1$, we get

$$\Lambda^{n+1}L^{n+1}\varphi = (n+1)!(-k)(-k+1)\cdots(-k+n)\varphi.$$

Because there is not enough room, we have $L^{n+1}\varphi = 0$ and hence the right hand side is zero. But the right hand side is nonzero. □

Proposition 21.2. *For any $\varphi \in \bigwedge^k(V\otimes\mathbb{C})$, we have*

$$\varphi = \sum_{\ell\geq m} L^\ell\varphi_\ell$$

where $\varphi_\ell \in \bigwedge^{k-2\ell}(V\otimes\mathbb{C})$ are primitive elements and $m = \max(k-n, 0)$.

This is the Lefschetz decomposition, and unfortunately, we cannot finish it today.

22 November 17, 2015

The Lefschetz decomposition is an important example of a normal form of a matrix. The setting is a vector space V over $\mathbb{R}$ with dimension $2n$. There are two additional structures: a complex structure $J : V \to V$ such that $J^2 = -1$, and a J-compatible inner product $(\cdot,\cdot)_V$ for V. With these structures, we are going to consider a matrix (a $\mathbb{R}$-linear transformation) on $\bigwedge^* V = \bigoplus_{j=0}^{2n} \bigwedge^j V$. If we go to $V \otimes_{\mathbb{R}} \mathbb{C}$ over $\mathbb{C}$, we can decompose

$$V \otimes_{\mathbb{R}} \mathbb{C} = V \oplus \bar{V}.$$

We can also consider an element of $\bigwedge^2 V$ which maps

$$u^*, v^* \in V^* \mapsto (Ju^*, v^*) \in \mathbb{R}.$$

This is in $\mathrm{Hom}_{\mathbb{R}}(V^* \times V^*, \mathbb{R})$ and because it is alternating. The Lefschetz operator L is exterior product by this element. In fact, it is

$$\omega = \frac{\sqrt{-1}}{2} \sum_{j=1}^{n} e^j \wedge \bar{e}^j$$

where $\xi^1, J\xi^1, \cdots, \xi^n, J\xi^n$ is an orthonormal basis, which is J-compatible, and $e^j = \xi^j - \sqrt{-1} J\xi^j$. We also showed last time that $\Lambda = *^{-1} L *$.[6] The eigenspace of Λ, which is just the kernel $\ker \Lambda$ is called the primitive elements.

22.1 Proof of the Lefschetz decomposition

Now the Lefschetz decomposition states that

Theorem 22.1. *Given $\varphi \in \bigwedge^k (V \otimes_{\mathbb{R}} \mathbb{C}) \subset \bigwedge^* (V \otimes_{\mathbb{R}} \mathbb{C})$, there is a unique decomposition*

$$\varphi = \sum_{\ell \geq m} L^\ell \varphi_\ell$$

where $\varphi_\ell \in \bigwedge^{k-2\ell}(V \otimes \mathbb{C})$ and $m = \max(0, k-n)$.

The reason we start from m is because we want to get the Lefschetz isomorphism between $\bigwedge^{p,q}$ and $\bigwedge^{n-q,n-p}$.

Let us prove this now. We have two tools: $\Lambda = *^{-1} L *$ and $[\Lambda, L] \sum (n-k)\Pi_k$. From this, we obtained

$$[\Lambda, L^r] = \sum_{k=0}^{2n} r(n-k-r+1) L^{r-1} \Pi_k$$

and

$$\Lambda^s L^r \varphi = r(r-)\cdots(r-s+1)(n-k-r+1)(n-k-r+2)\cdots(r-k-r+s)\varphi$$

for primitive $\varphi \in \bigwedge^k (V \otimes \mathbb{C})$ and $r \geq s$.

[6] Siu changed the notation some time. The normal 'contraction' we were looking at is now $\tilde{\Lambda}$ and Λ is defined by $\Lambda = \frac{2}{i}\tilde{\Lambda}$.

Proof. We first prove existence. Let $\varphi \in \bigwedge^k(V \otimes \mathbb{C})$. We want to find a decomposition.

Case 1. $k \leq n$

Consider the minimal $r \geq 0$ such that $\Lambda^t \varphi = 0$. If $r = 0$ or $r = 1$, it is trivial. We use induction on r. Suppose that we already have the decomposition for r, and consider an element φ such that

$$\Lambda(\Lambda^r \varphi) = 0$$

and then it follows that $\Lambda^r \varphi$ is primitive in $\bigwedge^{k-2r}(V \otimes \mathbb{C})$. Using some identity we found, we see that

$$\Lambda^r L^r(\Lambda^r) = A(\Lambda^r \varphi)$$

for some $A \neq 0$. The we see that

$$\Lambda^r(\varphi - \frac{1}{A} L^r \Lambda^r \varphi) = 0$$

and hence by induction we are done.

Case 2. $k > n$

Let $m = k - n \geq 1$. Let r be the smallest integer at least m such that $\Lambda^r \varphi = 0$. This is essentially the same as the first case, and using the smear induction, we can show that there are primitive elements $\varphi_\ell \in \bigwedge^{k-2\ell}(V \otimes \mathbb{C})$ such that

$$\Lambda^m \left(\varphi - \sum_{\ell=m}^{r-1} L^\ell \varphi_\ell \right) = 0.$$

Let $\varphi' = \varphi - \sum_{\ell=m}^{r-1} L^\ell \varphi_\ell$. This is in $\bigwedge^k(V \otimes \mathbb{C})$ and $\Lambda^m \varphi' = 0$. Using the star operator, we get a $*\varphi' \in \bigwedge^{2n-k}(V \otimes \mathbb{C})$. Now we can decompose

$$\varphi' = \sum_{\ell=0}^{r'-1} L^\ell \varphi'_\ell.$$

Because $\Lambda^m \varphi' = 0$, we have

$$0 = L^m(*\varphi') = \sum_{\ell=0}^{r'-1} L^{m+\ell} \varphi'_\ell$$

and by uniqueness of Lefschetz decomposition, we have $\varphi'_\ell = 0$. This shows that $*\varphi' = 0$ and hence $\varphi' = 0$. So $\varphi = \sum_{\ell=m}^{r-1} L^\ell \varphi_\ell$ is a Lefschetz decomposition.

We now prove uniqueness. It suffices to show that

$$\sum_{\ell \geq m} L^\ell \varphi_\ell = 0$$

implies $\varphi_\ell = 0$ for each ℓ. Let s be the largest ℓ such that $\varphi_\ell \neq 0$. We see that

$$\Lambda^s L^s \varphi_s = (\text{some nonzero constant}) \varphi_s.$$

If we apply Λ^s to the assumption, we get

$$0 = \sum_{\ell \geq m} \Lambda^s L^\ell \varphi_\ell.$$

But if $\ell < s$ the term $\Lambda^s L^\ell \varphi_\ell$ vanishes, and if $\ell = s$, we get a nonzero constant times φ_s. This contradicts our assumption that $\varphi_s \neq 0$. □

22.2 Prelude to our next topic

Now we finished Lefschetz decomposition, and I want to move to another topic, namely spin system. This is used in quantum mechanics and is really useful. I want to look it from the point of view of composing 2 rotations in $\mathbb{R}^3$. Rotations in $\mathbb{R}^2$ is just a multiplication by a complex number, and is simple, but rotations in $\mathbb{R}^3$ is complicated. The key in the theory is that rotation is two reflections involving $1/2$ of the angle.

Let me try two explain this in two dimensions. Say we want to rotate P by angle θ. This is easy in dimension 2, but in higher dimensions it is not easy. The right way to look at things is considering two lines differing by $\theta/2$, and reflecting P by one line and then the other line.

In three dimensions, this turns into choosing two planes by which we perform our reflections. But there is a certain degree of freedom, because as long as both planes contain the axis, we can rotate the plane around. So if you have two rotations, we can make the second reflection of the first rotation and the first reflection of the second rotation agree, by taking the reflection plane to be the plane containing both the first axis and the second axis. Then composition is reduced to two reflections.

23 November 19, 2015

There is the quaternions developed by Hamilton, from the viewpoint of composition of space rotations. Then people started to think about higher dimension analogues, for instance octonions. However, there are no "good" hypercomplex number systems in higher settings. This is connected with Clifford algebras and independent vector fields on spheres of higher dimension. These were established by the works of Hurwitz, Radon, and Eckmann in the 1920s.

Let us look what happens what when we extend $\mathbb{R}$ to $\mathbb{C}$. Basically, $\mathbb{C}$ is a vector space over $\mathbb{R}$, but with some multiplication. We can do this further.

$$\mathbb{R} \longrightarrow \mathbb{C} \longrightarrow \mathbb{H} \longrightarrow \mathbb{O}$$

But when we extend $\mathbb{C}$ to $\mathbb{H}$, we lose commutativity, and when we extend to $\mathbb{O}$ we lose associativity. But still we have involution, and a multiplicative absolute value. If we go further, we even lose this. The absolute value means that for indeterminates $x_1, \dots, x_n$ and $y_1, \dots, y_n$, there is a rule of multiplication

$$z_j = \sum_{k,l=1}^{n} a_{jul} x_k y_l$$

such that

$$(x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) = z_1^2 + \cdots + z_n^2.$$

23.1 Rotations of $\mathbb{R}^3$

Rotations of $\mathbb{R}^3$ fix a point, which is the axis. We can decompose any rotation into two reflections by two planes containing the axis and making angle $\theta/2$. Suppose we have two rotations, we can make the plane containing the first axis and the second axis, the first plane for the second rotation and the second plane for the first rotation.

Then we have three planes:

$$\begin{cases} \text{1st plane of 1st rotation} \\ \text{2nd plane of 1st rotation} = \text{1st plane of 2nd rotation} \\ \text{2nd plane of 2nd rotation} \end{cases}$$

Because two reflections cancel out, we have the representation of the composition of two rotations as the composition of two reflections. If we draw it on the 2-sphere, representing planes by great circles, we have the figure below.

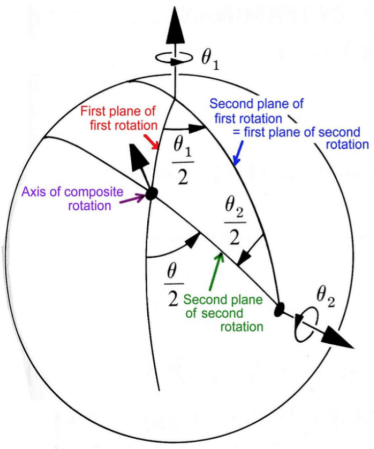


FIGURE 2

The good thing about this representation is that it gives a geometric algorithm determining the axis and the angle of the composite rotation. If we know the angels and the axes, then we can draw an arc joining the two axes, and draw two arcs making a given angle with the original arc, and take the intersection.

Now people wanted to relate "composition of rotations" to "multiplication of 'hypercomplex' numbers." This was motivated by rotations in $\mathbb{R}^2$; they are represented with complex numbers with unit length. That was the job Hamilton set out to solve.

23.2 Representation of rotation by quaternions and $SU(2)$

The complex numbers $\mathbb{C}$ is represented by $a+bi$ for $a, b \in \mathbb{R}$. People then tried $a+ib+jc$ for $a, b, c \in \mathbb{R}$ and $i^2 = j^2 = -1$, but failed. Then Hamilton came along and said that you need one more variable. If we let

$$\vec{i}\vec{j} = \vec{k}, \quad \vec{k}\vec{i} = \vec{j}, \quad \vec{j}\vec{k} = \vec{i}, \quad \vec{i}^2 = \vec{j}^2 = \vec{k}^2 = -1,$$

we have

$$(a + \vec{i}b + \vec{j}c + \vec{d})(a - \vec{i}b - \vec{j}c - \vec{k}d) = a^2 + b^2 + c^2 + d^2.$$

But how is it related to rotations? if we need multiplication to represent rotations, we need some "axis" which is unchanged by the rotation. The key here is to consider the conjugation instead of multiplication. The map

$$\vec{x} \mapsto A\vec{x}A^{-1}$$

leaves 1 unchanged. Since it is an isometry, perpendicularity is preserved, and hence the hyperplane $1^\perp$, which is the set of pure imaginary numbers, is preserved. So it maps

$$\vec{i}b + \vec{j}c + \vec{k}d \mapsto \vec{i}b' + \vec{j}c' + \vec{k}d'$$

which now is a rotation in $\mathbb{R}^3$.

Another breakthrough was identifying the rotation with a complex 2×2 matrix. If we want $a^2 + b^2 + c^2 + d^2$ as the determinant, we would have

$$\begin{aligned}\begin{pmatrix} a+ib & c+id \\ -c+id & a-ib \end{pmatrix} &= a\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} + c\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + d\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \\ &= ae_0 + be_1 + ce_2 + de_3.\end{aligned}$$

The great thing is that $e_0 \mapsto 1$, $e_1 \mapsto \vec{i}$, $e_2 \mapsto \vec{j}$, and $e_3 \mapsto \vec{k}$ is a representation of quaternions of length 1 by

$$SU(2) = \left\{ \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix} : |\alpha|^2 + |\beta|^2 = 1 \right\}.$$

W. Pauli introduced the "infinitesimal" form. We can write a complex number of absolute value 1 as

$$e^{i\theta}, \quad \text{where } \theta \in \mathbb{R}.$$

Likewise, if we have a 2×2 unitary matrix, we can write it as

$$e^{iA}, \quad \text{where } \bar{A}^t = A \text{ is Hermitian.}$$

So the **Pauli matrices** are

$$\sigma_x = -i\vec{e}_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = -i\vec{e}_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = -i\vec{e}_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and we have the relations

$$\sigma_x\sigma_y = i\sigma_z, \quad \sigma_y\sigma_z = i\sigma_x, \quad \sigma_z\sigma_x = i\sigma_y, \quad \sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1.$$

We will talk more about this when we (possibly) do Lie algebras.

Suppose we rotate $\mathbb{R}^3$ around an axis $(\cos\alpha, \cos\beta, \cos\gamma)$ by angle θ. Then the quaternion representing this rotation is explicitly given by

$$R = \cos\frac{\theta}{2} - i\sin\frac{\theta}{2}(\sigma_x\cos\alpha + \sigma_y\cos\beta + \sigma_z\cos\gamma$$

and then this rotations sends a pure imaginary $\vec{x}$ to $R\vec{x}R^{-1}$ in the sense of quaternion multiplication.

23.3 Hypercomplex number systems

The first major breakthrough was make by Hurwitz in 1922. We want to find a a_{jkl} such that

$$\sum_{j=1}^{n} z_j^2 = \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{j=1}^{n} y_j^2\right)$$

for $z_j = \sum_{k,l=1}^{n} a_{jkl}x_k y_l$. This problem is reduced to the Clifford algebra. The

Definition 23.1. The **Clifford algebra** is the algebra generated by $e_0 = 1, e_1, e_2, \ldots, e_k$ with the relations

$$e_j e_k = -e_k e_j \text{ for distinct } j, k \neq 0, \quad e_j^2 = -1.$$

Now because the things of the form $e_{i_1} \cdots e_{i_k}$ form a basis, we see that the dimension should be 2^k.

Example 23.2. *In the case $k = 2$, the Clifford algebra is the quaternions. This is because we can set $e_3 = e_1 e_2$ and everything is the same.*

Hurwitz proved that the problem can be solved only for $n = 1, 2, 4, 8$. Let me explain briefly how he did it. We generalize the situation to

$$\sum_{j=1}^{n} z_j^2 = \left(\sum_{i=1}^{p} x_i^2 \right) \left(\sum_{j=1}^{n} y_j^2 \right).$$

Then we can write z as $z = x_1 A_1 + \cdots + x_p A_p$ for some $n \times n$ matrices $A_1, \ldots, A_p$. Then because we want the length to be preserved, the matrices should satisfy

$$(x_1 A_1^t + \cdots + x_p A^t)(x_1 A_1 + \cdots + x_p A_p) = x_1^2 + \cdots + x_p^2$$

after normalizing the matrices. The we have the system of equations

$$A_k^t A_k = I_n, \quad A_j^t A_k + A_k^t A_j = 0 \quad (k \neq j).$$

This looks like the Clifford algebra, and if we multiply i to everything, we get the Clifford algebra. Then you can replace A_k by $A_p^{-1} A_k$ and make $A_p = I$. Then you can use the anti-commutativity to do things.

Theorem 23.3 (Hurwitz, 1922)**.** *Given any $n = u 2^{4\alpha+\beta}$ such that u is odd and $\beta = 0, 1, 2, 3$, a solution exists if and only if $p \leq 8\alpha + 2^\beta$. If $p = n$, it is true only for $n = 1, 2, 4, 8$.*

If we denote the Clifford algebra by Cliff_k, its dimension $\dim_{\mathbb{R}} \text{Cliff}_k = 2^k$ and the set

$$\{e_{i_1} \cdots e_{i_m} : m \geq 0, i_1 < \cdots < i_m\}$$

is a basis. Then

$$G = \{\pm e_{i_1} \cdots e_{i_m} : m \geq 0, i_1 < \cdots < i_m\}$$

is a group.

Let V be a $\mathbb{R}$-vector space of dimension n. Assume that V is a G-module, and it acts as an isometry. Let $S(V)$ be the unit sphere in V, which will be the same as S^{n-1}.

Theorem 23.4. *For $x \in S^{n-1}$, the set $\{e_1 x, \ldots, e_k x\}$ form a orthonormal frame on $S(V)$.*

This kind of gives an almost complex structure on S^n.

24 November 24, 2015

I will start a new topic, which is the Young diagram. It ties the polynomial equations and linear systems up.

24.1 Decomposing a function into symmetric parts

What are the Young diagrams, and why is it important? First let us look at S_2. This acts on the set of functions $f(x, y)$. We see that there is a decomposition

$$f(x, y) = f_{\text{odd}}(x, y) + f_{\text{even}}(x, y)$$

where

$$f_{\text{odd}}(x, y) = \frac{1}{2}(f(x, y) - f(y, x)),$$
$$f_{\text{even}}(x, y) = \frac{1}{2}(f(x, y) + f(y, x)).$$

Then the question is what happens to more variables? Let see the three variable case. If we have $f(x_1, x_2, x_3)$, we have the analogous symmetric component

$$f_{\text{sum}}(x_1, x_2, x_3) = \frac{1}{3!} \sum_{\sigma \in S_3} f(x_{\sigma(1)}, x_{\sigma(2)}, x_{\sigma(3)})$$

and the alternating component

$$f_{\text{alt}}(x_1, x_2, x_3) = \frac{1}{3!} \sum_{\sigma \in S_3} \text{sgn}(\sigma) f(x_{\sigma(1)}, x_{\sigma(2)}, x_{\sigma(3)}).$$

What is the things in-between? We may guess it as a partially alternating and partially symmetric object. But if $f(x_1, x_2, x_3)$ is symmetric between x_1 and x_2, and alternating in x_2 and x_3, the function should be zero. So it means that this doesn't work.[7] This was observed by Alfred Young and Issai Schur independently.

So instead of trying to make f possess both the symmetry of x_1 and x_2 and the alternation for x_2 and x_3, we do it independently and use the non-commutativity of the two processes.

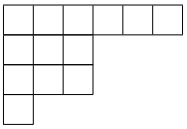
24.2 Young diagrams and Young symmetrizers

In general, consider any element in S_n, and suppose that there are α_k number of k-cycles. Then we see that

$$n = \alpha_1 + 2\alpha_2 + 3\alpha_3 + \cdots + n\alpha_n.$$

[7] This related to the braiding lemma, which is something we will do next semester in proving the fundamental theorem of Riemannian geometry.

We draw empty boxes, so that in the first row, there are $f_1 = \alpha_1 + \cdots + \alpha_n$ boxes, and in the second row, there are $f_2 = \alpha_2 + \cdots + \alpha_n$ boxes, and so forth. Then $f_1 + \cdots + f_n = n$. The diagram consisting of boxes is called a **Young diagram**.



A Young diagram corresponds to a partition of n identical balls into certain groups, so it corresponds to a conjugacy class of S_n. If we put in numbers $1, 2, \ldots, n$ inside the boxes, then we get something more complicated. This is called a **Young tableau**.

Given a Young tableau, we can construct the Young symmetrizer as the following. Consider all permutations σ which only changes elements inside the same rows, and add them up without the $\operatorname{sgn}(\sigma)$ factor. So this will be something like the symmetrizing the f. Next consider all permutations τ which only changes elements inside the same columns, and add them up with the $\operatorname{sgn}(\tau)$ factor. We stop here.

Let us consider the case $n = 3$. There are three Young diagrams. Consider one of the more complex diagrams:

The symmetrizer then shall be

$$\frac{1}{4}\big((f(e_1, e_2, e_3) + f(e_2, e_1, e_3)) - (f(e_3, e_2, e_1) + f(e_3, e_1, e_2))\big).$$

24.3 Representation of a finite group

Consider the group $G = S_n$. Then you can represent G as a matrix, because we can just consider the group algebra $\mathbb{C}[G]$ defined as the vector space generated by the elements of G. Then for any $h \in S_n$, it acts as a linear map on $\mathbb{C}[G]$, because

$$h\left(\sum_{g \in G} a(g)g\right) = \sum_{g \in G} a(g)hg.$$

So if we let $m = |G|$, we get a homomorphism $G \to GL(m, \mathbb{C})$. We also can think about the normal form of the representation, so that the matrices look like block matrices for some good basis choice. Then we get a decomposition into smaller representations.

For example, if you have $G = S_2$, we have a representation $G \to GL(2, \mathbb{C})$ which maps

$$1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (12) \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

This is called a representation.

Definition 24.1. Suppose that G is a finite group. A **representation** of the group G is a homomorphism $\rho : G \to GL(m, F) = GL(V)$, where V is a n-dimensional vector space over the field F.

Definition 24.2. A representation ρ is called **irreducible** if the only subspaces W of V with the property that $g(W) \subset W$ for all $g \in G$ are $W = 0$ and $W = V$.

Suppose that $F = \mathbb{C}$. Then we can introduce an inner product on V which is invariant under the action of any $g \in G$ by averaging the intro product over G. That is, if we already have an inner product $(\cdot, \cdot)_V$, we can let

$$(u, v)_V^* = \frac{1}{|G|} \sum_{g \in G} (gu, gv)_V.$$

Then this is clearly a G-invariant inner product. Suppose that ρ is not irreducible. Then by definition there is a G-invariant subspace W of V where $W \neq 0, V$. Then for the G-invariant inner product we defined, we have

$$V = W \oplus W^\perp$$

and the $W^\perp$ will also be G-invariant. This means that we can break everything down to irreducible representations.

24.4 Results of Schur's theory

These are the main results in Schur's theory.

1. A representation can be identified by a presentation by its character (which is the trace of each matrix). This reduces matrices into scalars.

   Theorem 24.3 (Schur)**.** *Let χ_ρ be the character of an irreducible representation. Then the set $\{\chi_\rho\}$ is an orthonormal basis of character functions of G.*

   This tells you how to calculate the number of irreducible representations.

2. We define the degree of ρ as the order of the matrix for the representation ρ, and let us denote it by m_ρ. Then we have

   $$|G| = \sum m_\rho^2$$

   and

   $$m_\rho \mid |G| \quad \text{for each } \rho.$$

   Also, there is one representation called the regular representation of G, defined by the left multiplication on $\mathbb{C}[G]$. This representation shall be decomposed into many irreducible representations. Each representation ρ occurs exactly m_ρ times in the decomposition.

We will prove this next time, but let me tell you the main tool. Suppose we have two representations $\rho : G \to GL(V)$ and $\tau : G \to GL(W)$. The two representations are equivalent if for any g

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ {\scriptstyle \rho(g)}\downarrow & & \downarrow{\scriptstyle \tau(g)} \\ V & \xrightarrow{T} & W \end{array}$$

commutes. The T here is called the intertwining operator. In general, given any T, possibly not intertwining, we can average it to get an intertwining one. That is, we can let

$$T' = \frac{1}{|G|} \sum_g \tau(g) T \sigma(g^{-1})$$

to get a intertwining T'. Because both σ and τ are irreducible, we see that T' is either invertible, which will mean that σ and τ are equivalent, or $T' = 0$.

25 December 1, 2015

Young diagrams are related to representations of finite groups. Suppose that we are trying to represent S_n. Let $\theta \in S_n$ act on an arrangement

$\alpha(1)$	$\alpha(2)$	$\cdots$	$\alpha(n)$

to yield

$\beta(1)$	$\beta(2)$	$\cdots$	$\beta(n)$

There are two natural ways to make θ act.

i) The first one is change the locations with respect to θ. That is, we let $\theta : 1 \mapsto k$ if $\alpha(1) = \beta(k)$. This means that $\alpha(j) = \beta(\theta(j))$ and hence $\alpha = \beta\theta$ and $\theta = \beta^{-1}\alpha$.

ii) The second one is changing the elements with the same location. That is, we have
$$\theta = \begin{pmatrix} \alpha(1) & \cdots & \alpha(n) \\ \beta(1) & \cdots & \beta(n) \end{pmatrix}.$$
Then $\theta(\alpha(j)) = \beta(j) = \theta = \beta\alpha^{-1}$.

We will be using the first action. As long as we don't confuse one with the other, it doesn't matter.

We work in the group algebra. If we apply the action of some element on f twice as
$$\Big(\sum \tilde{c}(\tilde{g})\tilde{g}\Big)\Big(\sum c(g)g\Big)f(x_1, \ldots, x_n)$$
then it is same as the action of
$$\Big(\sum \tilde{c}(\tilde{g})\tilde{g}\Big)\Big(\sum c(g)g\Big)$$
on f.

25.1 Decomposition of the regular representation

Let G be a finite group. Let G act on $\mathbb{C}[G]$ by left multiplication. The action of any $g \in G$ is clearly an automorphism of $\mathbb{C}[G]$. Hence we get a homomorphism
$$\mathrm{reg}_G : G \to \mathrm{Aut}_{\mathbb{C}}(\mathbb{C}[G])$$
which we will call the **regular representation**. We will decompose this in to invariant summands.

Recall that a representation is a homomorphism $\rho : G \to GL(V_\rho)$ where $\dim_{\mathbb{C}} V_\rho = m_\rho$. This representation is called irreducible if any G-invariant subspace is either 0 or V_ρ.

First note that there is always a G-invariant inner product on V_ρ. Take any Hermitian inner product on V_ρ and let

$$(u, v)_{\text{avg}} = \frac{1}{|G|} \sum_{g \in G} (gu, gv).$$

Then this new inner product will be a G-invariant inner product. If W_ρ is a G-invariant subspace, then the orthogonal part $W_\rho^\perp$ will be a new G-invariant subspace and moreover we have a decomposition

$$V_\rho = W_\rho \oplus W_\rho^\perp.$$

This proves that every non-irreducible representation can be always decomposed to smaller representations.

Two representations $\rho : G \to GL(V_\rho)$ and $\sigma : G \to GL(V_\sigma)$ are called equivalent if $\dim V_\rho = \dim V_\sigma$ and there is an invertible $T : V_\rho \to V_\sigma$ such that

$$T\rho(g)T^{-1} = \sigma(g)$$

for any $g \in G$. This just means that the representation differs only in a change of basis.

Now in order to get a description of an irreducible representation, we need to find some "invariant" which does not change under basis change. Clearly the trace of a matrix is invariant under conjugation. The surprising result of Schur is that if only the trace agrees, then the two irreducible representation are actually equivalent.

Definition 25.1. A **character** of a representation ρ is the function

$$\chi_\rho : g \to \operatorname{tr} \rho(g).$$

The character is a class function. That is, $\chi_\rho(hgh^{-1}) = \chi_\rho(g)$ and hence χ_ρ can be viewed as a function on the conjugacy classes.

There are several results regarding irreducible representations.

1. One of the main result we will prove is that the set of characters of all irreducible representations $\{\chi_\rho\}$ form an orthonormal basis for the vector space of class functions. Here, the inner product on class functions is defined just by
$$(\alpha, \beta) = \frac{1}{|G|} \sum_{g \in G} \alpha(g)\overline{\beta(g)}.$$
If we have the result, we know precisely how many irreducible representations there are. Because the number of irreducible representations is just the dimension of class functions, we se that it is the number of conjugacy classes.

2. The second result is that for any irreducible ρ, the order m_ρ always divides $|G|$. This is a deep result using algebraic integers.

3. The third result is that if we decompose the regular representation to

$$\text{reg}_G = \bigoplus (\text{irred. rep. } \rho)$$

then each ρ occurs precisely m_ρ times in the decomposition. This means that for each matrix $\text{reg}_G(g)$ of size $|G| \times |G|$ can be simultaneously block-diagonalized so that the $\rho_G(g)$ block occurs m_ρ times. If we count the order, we get

$$|G| = \sum_\rho m_\rho^2.$$

Example 25.2. *Consider S_3. There is the trivial representation $\rho_1 : S_3 \to GL(1, \mathbb{C})$ such that $\rho_1(g) = 1$, and there is the alternated representation $\rho_1' : S_3 \to GL(1, \mathbb{C})$ such that $\rho_1'(g) = \text{sgn}(g)$. And there is the ρ_2 using the Young symmetrizer. Then we have one ρ_1, one ρ_1', and two ρ_2 in the regular representations.*

25.2 Intertwining operator and Schur's lemma

Let us start proving things. Most of these things are proved by Schur.

Definition 25.3. Let $\rho : G \to GL(V)$ and $\sigma : G \to GL(W)$ be two irreducible representations. An **intertwining operator** is a map $T : V \to W$ such that the diagram

$$\begin{array}{ccc} V & \xrightarrow{\rho(g)} & V \\ {\scriptstyle T}\downarrow & & \downarrow{\scriptstyle T} \\ W & \xrightarrow{\sigma(g)} & W \end{array}$$

commutes for all $g \in G$.

Note that both $\ker T \subset V$ and $\text{Im}\, T \subset W$ are both G-invariant. Because both are irreducible representations, we see that $\ker T$ is either 0 or V. Likewise, we have $\text{Im}\, T$ is either 0 or W. From this, we see that T is either 0 or invertible.

Suppose that $V = W$. Consider an eigenvalue λ of T. Then $T - \lambda I$ is also an intertwining operator. This cannot be invertible. Thus $T - \lambda I = 0$. Therefore we have the following lemma.

Lemma 25.4 (Schur's lemma)**.** *Suppose that T intertwines two inequivalent irreducible representations. Then we have $T = 0$. Suppose that T intertwines the same irreducible representation. Then we have $T = cI$ for some $c \in \mathbb{C}$.*

We can using the averaging technique to explicitly construct an intertwining operator. Suppose $\rho : G \to GL(V)$ and $\sigma : G \; to GL(W)$ be two representations, and let T be any linear map $T : V \to W$. Then we see that

$$T_{\text{avg}} = \frac{1}{|G|} \sum_{g \in G} \rho(g) T \sigma(g^{-1})$$

is an intertwining operator because

$$\rho(h)T_{\text{avg}}\sigma(h^{-1}) = \frac{1}{|G|}\sum_{g\in G}\rho(h)\rho(g)T\sigma(g^{-1})\sigma(h^{-1}) = T_{\text{avg}}.$$

Then by Schur's lemma, we would have $T_{\text{avg}} = cI$.

Theorem 25.5. *The characters of the irreducible representations form a orthonormal basis in the space of class functions.*

Proof. Start with a map $T_{jk} : V \to W$, which is defined by the matrix with all zeros and only one 1 in the jth row and kth column, where $1 \le j \le m$ and $1 \le k \le n$ where $\dim V = n$ and $\dim W = m$. Then the averaging will be

$$\sum_{g\in G}\sigma(g)T_{jk}\rho(g^{-1}) = 0 \text{ or } cI$$

for some c. Now assume that $j \neq k$, and consider the (j,k)th entry of the matrix. No matter what the matrix is, that entry must be zero. This means that

$$\sum_{g\in G}\sigma(g)_{jj}\rho(g^{-1})_{kk} = 0.$$

Note that $\rho(g)^{|G|} = 1$ and hence every eigenvalue of $\rho(g)$ has absolute value one. Therefore $\chi_\rho(g^{-1}) = \overline{\chi_\rho(g)}$. If ρ and σ are inequivalent, we have

$$\sum_{g\in G}\sigma(g)_{jj}\rho(g^{-1})_{kk} = 0$$

even if $j = k$. If we sum it over j and k, we get

$$\sum_{g\in G}\chi_\sigma(g)\overline{\chi_\rho(g)} = 0.$$

If ρ are σ are equivalent, then $\chi_\rho = \chi_\sigma$. Because the matrix

$$\sum_{g\in G}\rho(g)T_{jj}\rho(g^{-1})$$

is a constant times the identity, but the trace is

$$\sum_{g\in G}\operatorname{tr}(\rho(g)T_{jj}\rho(g^{-1})) = \sum_{g\in G}\operatorname{tr}T_{jj} = |G|,$$

we see that the map is $|G|/m_\rho$ times the identity. So we get

$$\frac{1}{|G|}\sum_{g\in G}\chi_\rho(g)\overline{\chi_\rho(g)} = \frac{1}{|G|}\sum_{g\in G}\Big(\sum\rho(g)_{jj}\Big)\Big(\sum\rho(g^{-1})_{kk}\Big) = 1.$$

This shows that the $\{\chi_\rho\}$ are orthonormal vectors.

Lastly, we should prove that any class function orthogonal to all χ_ρ is identically zero. We will prove this next time. □

26 December 3, 2015

Last time we showed that χ_g are orthogonal and have unit length in the space of all class functions on G. We now actually prove that it is a basis. We will prove that the orthogonal complement is zero, i.e.,

$$\sum_{g\in G} \overline{\varphi(g)}\chi_\rho(g) = 0$$

for all irreducible ρ implies $\varphi \equiv 0$ on G.

Proof. Consider

$$T = T_{\varphi,\rho} = \frac{1}{|G|}\sum_{g\in G} \overline{\varphi(g)}\rho(g).$$

Clearly $\operatorname{tr} T = 0$. Also, note that because

$$\rho(h)T\rho(h^{-1}) = \frac{1}{|G|} = \sum_{g\in G} \overline{\varphi(g)}\rho(h)\rho(g)\rho(h^{-1}) = T$$

since φ is a class function, we see that T is an intertwining operator between ρ and ρ. Therefore T is a constant times the identity. But because T is traceless, we see that $T = 0$.

Now note that this is 0 for all irreducible representation ρ. Because any representation can be decomposed into irreducible representations, we see that for any representation ρ, the T should be zero. In particular, we have

$$\frac{1}{|G|}\sum_{g\in G} \overline{\varphi(g)}\mathrm{reg}_G(g) = 0.$$

If we consider the action of this on $1_G \in \mathbb{C}[G]$, we see that

$$\sum_{g\in G} \overline{\varphi(g)}g = 0 \in \mathbb{C}[G].$$

Therefore $\varphi \equiv 0$. □

This was the most complicated part. We now prove the following.

Theorem 26.1. *Each irreducible representation ρ occurs as many times as m_ρ in the regular representation* reg_G.

Proof. Note that when reg_G is decomposed into irreducible representations and block diagonalized, the character χ_{reg_G} will be a sum of χ_ρs as many times as the number it occurs in the decomposition. Because χ_ρ form an orthonormal basis, we see that that number is just

$$(\chi_{\mathrm{reg}_G}, \chi_\rho).$$

But χ_{reg_G} is just $|G|$ at 1_G and 0 at other points. So we see that

$$(\chi_{\mathrm{reg}_G}, \chi_\rho) = \frac{1}{|G|}\sum_{g\in G} \overline{\chi_{\mathrm{reg}_G}(g)}\chi_\rho(g) = \chi_\rho(1_G) = m_\rho.$$

Therefore ρ occurs m_ρ times. □

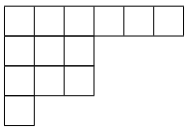
26.1 Representations of S_n

Let us look at representations of S_n. Because S_n is a permutation group, we can consider special permutations by perhaps making blocks and considering only the permutations preserving blocks. Young's idea was representing it as boxes.

Consider a partition $n_1 \geq n_2 \geq \cdots \geq n_\ell$ of n so that

$$n = n_1 + n_2 + \cdots + n_\ell.$$

We represent it we a Young diagram with n_j boxes in the jth row.



If $\tilde{Y} = (\tilde{n}_1, \ldots, \tilde{n}_{\tilde{\ell}})$ is another Young diagram, we can consider the lexicographical order and write $Y \geq \tilde{Y}$ if and only if the first $n_k - \tilde{n}_k$ is positive.

If $\mathbb{C}[G]$ is decomposed into irreducible representations

$$\mathbb{C}[G] = V_1 \oplus V_2 \oplus \cdots \oplus V_\ell,$$

then we can consider the projection $\Pi_1 : \mathbb{C} \to V_1$ and it will commute with any $g \in G$.

$$\begin{array}{ccc} \mathbb{C}[G] & \xrightarrow{\Pi_1} & V_1 \\ \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle g} \\ \mathbb{C}[G] & \xrightarrow{\Pi_1} & V_1 \end{array}$$

Suppose that $T : \mathbb{C}[G] \to \mathbb{C}[G]$ is an G-equivariant map; i.e., that T is a left-multiplication by some $T = \sum_g T_g g$. Then after some calculation one sees that there is a $\tilde{T}$ such that $Ta = a\tilde{T}$ for any $a \in \mathbb{C}[G]$.

A decomposition of $\mathbb{C}[S_n]$ is a set of idempotent maps $e_1, e_2, \ldots, e_\ell$ such that $e_1 + e_2 + \cdots + e_\ell = 1$ and $e_j e_k = 0$. For a Young diagram Y, let e_Y be the element defined by

$$e_Y = \sum_{g \in G} e_{Y,g} g$$

where

$$e_{Y,g} = \begin{cases} \operatorname{sgn} \gamma & \text{if } g = \gamma\rho \text{ for a column-preserving } \gamma \text{ and row-preserving } \rho \\ 0 & \text{otherwise.} \end{cases}$$

8

[8] I gave up taking notes because I was not able to understand.

Index