目录

研究探讨
Notes on Optimal Transport: Computations and Applications 陈钇帆 1
On Certain Square-free Triples 谭泽睿 11
McKay Correspondence 吴志翔 17
From Belyi’s Theorem to Dessins d’enfants 赵瑞屾 30

专题介绍
A Brief Introduction of KAM Theory 王森淼 41
Morse Theory And Its Applications 金儒鸿 53
Basics of Iwasawa Theory 张志宇 62

数学小品
Trace Class Operators 李林骏 70
Modules and λ-Matrix 何通木 75
Notes on Optimal Transport: Computations and Applications

陈钇帆 *

Abstract

Optimal transport (OT) has witnessed massive applications in a lot of fields such as statistical learning and image processing. The theory introduced provides a natural distance to compare probability measures or histograms, named the Wasserstein metric. This note gives a short introduction of OT and surveys some of the recent progress in the computations and applications of Wasserstein metric. Corresponding literature and their main ideas are also presented.

1 Background

First we give a short introduction of the optimal transport problem and its several equivalent forms. To simplify the presentation we only deal with probability measures in $\mathbb{R}^n$ and assume they are absolutely continuous with respect to the Lebesgue measure, i.e. their density functions exist. For more detailed discussions we refer to [1].

Consider two probability measures $\mu$ and $\nu$ in $\mathbb{R}^n$ with density functions $\rho_0$ and $\rho_1$, respectively. OT aims to find an “optimal” way to compare the two measures, or in economical terms, to seek the lowest cost for rearranging $\rho_0$ to $\rho_1$. The original problem proposed by French engineer Gaspard Monge writes

$$\inf_{T} \int_{\mathbb{R}^n} c(x, T(x)) \rho_0(x) dx \quad \text{s.t.} \quad T\#\mu = \nu. \tag{1}$$

where $c$ is a cost function that measures the distance between $x$ and $T(x)$, and $\#$ means “push-forward”, that is, $T\#\mu(A) := \mu(T^{-1}(A))$ for any measurable $A$. This requirement implies the mass is conserved during the whole rearranging process, i.e.

$$\int_A \rho_1(y) dy = \int_{T(x) \in A} \rho_0(x) dx.$$

* 数 41
The Monge problem (1) is nonlinear and in section 2.2 we will see the solution is related to the famous Monge-Ampère equation (13). The nonlinear structure renders it very difficult to deal with, for example it is not obvious that the minimum exists. From the mathematical modeling view, Monge assumes the mass \( \rho_0(x) \) at \( x \) could only be transferred to the location \( T(x) \), that is to say, no mass is split. This might not be the case in real world since sometimes we would like to cut up a large object and arrange its small parts separately. To address the issue, Kantorovish took the splitting of mass into consideration and proposed the Kantorovish problem (2). In this formulation, the dimension of the transport plan \( \gamma \) doubles and a simple linear programming problem is obtained:

\[
\min_{\gamma(x,y)} \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x,y)\gamma(x,y)\,dx\,dy
\]

\[
s.t. \quad \int \gamma(x,y)\,dy = \rho_0(x)
\]

\[
\int \gamma(x,y)\,dx = \rho_1(y)
\]

The theory of linear programming is very mature and the existence and uniqueness of Kantorovish problem is easily established. Moreover, using the duality theory in convex analysis we are able to dig more information about this structure. The result is, under mild conditions the optimal \( \gamma \) must be “one to one” and its support gives a map \( T \). In this sense the optimal transport will not split the mass, and the existence of solution in (1) could be established. So lifting the dimension works to handle the Monge problem and the two problems (1),(2) are equivalent.

Precisely, the dual of (2) has the form

\[
\sup_{\varphi, \phi \in C(\mathbb{R}^n)} \int_{\mathbb{R}^n} \varphi(x)\rho_0(x)\,dx + \int_{\mathbb{R}^n} \phi(y)\rho_1(y)\,dy \quad s.t. \quad \varphi(x) + \phi(y) \leq c(x,y),
\]

and \( \phi \) is called the Kantorivish potential. From the duality theory we know that it can be interpreted as the Fréchet gradient of the optimal cost with respect to \( \rho_0 \). This formulation is important because it is equivalent to

\[
\sup_{\varphi \in C(\mathbb{R}^n)} \int_{\mathbb{R}^n} \varphi(x)\rho_0(x)\,dx + \int_{\mathbb{R}^n} \varphi^\ast(y)\rho_1(y)\,dy,
\]

where \( \varphi^\ast(y) := \inf_x c(x,y) - \varphi(x) \), similar as the Legendre transformation. Note that the dimension of the problem reduces back to \( n \) in (4), which is very useful in computation.

Usually the cost function \( c(x,y) = L(x - y) \) and \( L(x - y) = \|x - y\|_p^p \), the \( L^p \) norm. Suppose the optimal value of problem (1),(2),(3) or (4) is \( V_p \), and then the \( L^p \) Wasserstein metric is defined as \( W_p(\mu,\nu) = V_p^{1/p} \). It is a distance on the probability measure space \( \mathcal{M}(\mathbb{R}^n) \). Numerous literature has investigated its properties and finds its fruitful connections to convex and functional analysis, Riemann geometry, PDEs and fluid dynamics, etc.
The fluid dynamics view of OT renders it a new dynamical formulation, and this view is now receiving increasing attentions in both theoretical and applicable sides. By adding a time variable it aims to find the “geodesic” \( \rho(x,t) \) between \( \rho_0 \) and \( \rho_1 \). The feasible \( \rho(x,t) \) will satisfy a continuity equation and it leads to a control problem:

\[
\min_v \int_0^1 \int_{\mathbb{R}^n} L(v)\rho(t,x)dxdt \\
\text{s.t. } \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0 \\
\rho(0,x) = \rho_0, \rho(1,x) = \rho_1
\]

where \( v \) is the flux vector field of the transportation. When \( L(v) = \|v\|^2 \) this formulation makes \( \mathcal{M}(\mathbb{R}^n) \) a Riemannian manifold endowed with the Wasserstein metric, see Otto Calculus in [2] for details.

2 Computation

In this section, we turn to the computational OT. Computation is important to promote a theory to go further and will in return motivate many new research topics. For OT, these four equivalent formulations (1), (2), (4) and (5) can create a lot of methods to compute the Wasserstein metric. This note mainly covers some progress in the computational \( L^1 \) and \( L^2 \) Wasserstein metric. The former has a homogeneous degree one cost function, rendering it very efficient to deal with in numerics. The latter enjoys a fruitful geometric and PDEs view but its computation is more challenging.

2.1 Homogeneous degree one cost function

When \( c(x,y) = \|x - y\|_1 \), (4) has an equivalent form:

\[
\max_{\varphi \in \text{Lip}_1} \int_{\mathbb{R}^n} \varphi(x)(\rho_0(x) - \rho_1(x))dx
\]

where \( \varphi \in \text{Lip}_1 \) means \( |\varphi(x) - \varphi(y)| \leq \|x - y\| \). Consider its discrete version in \( \mathbb{R}^1 \) (higher dimension is similar):

\[
\max_{\varphi} \sum_{i} \varphi_i(\rho_0^{(i)} - \rho_1^{(i)})\Delta x \\
\text{s.t. } \varphi_{i+1} - \varphi_i \leq \Delta x \quad \text{and} \quad \varphi_i - \varphi_{i+1} \leq \Delta x
\]

This is a linear programming problem and could be easily solved with high efficiency by the popular first order algorithms, such as ADMM, primal-dual splitting, etc. An algorithmic example will be [3] in full waveform inversion. Note that if we just use (2) for computation
the dimension will double, but (6) handles it well. In section 3.2 we will see that in deep learning, people use a deep neuron network to approximate the feasible set \( \{ \varphi \in \text{Lip}_1 \} \) and compute a Wasserstein-like quantity, which leads to the popular WGAN [4].

The method above could be generalized to unbalanced OT, which is often the case in WGAN or KR norm-based full waveform inversion since energy might be not equal for two physical objects. However, using the formulation (5) seems to become a trend now for its fruitful structures. The result is, when \( L \) is homogeneous degree one, such as \( L(v) = \|v\|_1 \) or the Manhattan distance \( L(v) = \|v\|_2 \), the dynamic flow problem can be reformulated as a static flow problem, by using a simple Jensen Inequality trick:

\[
\min_{m} \int_{\mathbb{R}^n} L(m(x))dx \\
\text{s.t. } \nabla \cdot m(x) + \rho_1(x) - \rho_0(x) = 0
\]

This problem has a “compressed sensing” structure, i.e. when \( L(v) = \|v\|_1 \) it has the form

\[
\min \|x\|_1 \\
\text{s.t. } Ax = b
\]

and computational tools in compressed sensing could be applied to this setting. Interested reader could refer to [5] for details.

### 2.2 Quadratic cost function

When \( c(x, y) = \|x - y\|_2^2 \) we obtain the quadratic Wasserstein metric

\[
W_2^2(\mu, \nu) = \inf_{T#\mu=\nu} \int_{\mathbb{R}^n} \|x - T(x)\|_2^2 \rho_0(x)dx.
\]

In one dimension the optimal \( T \) has a explicit formula \( T = G^{-1}(F) \), i.e.

\[
W_2^2(\mu, \nu) = \int_{\mathbb{R}} |F^{-1}(x) - G^{-1}(x)|^2dx,
\]

where \( F, G \) are accumulative distribution functions of \( \rho_0 \) and \( \rho_1 \), respectively. In this sense \( W_2^2 \) metric is similar to the \( H^{-1} \) Sobolev norm in one dimension.

In higher dimension, however, explicit formula does not exist. To the best of my knowledge there are two popular ways to compute it: one is to solve the related Monge-Ampére equation, and another is to add entropic regularizations.

Consider the Monge-Ampére equation in space \( X \):

\[
\begin{cases}
\det(D^2u(x)) = \rho_0(x)/\rho_1(\nabla u(x)), \ x \in X \\
u \text{ is convex}
\end{cases}
\]

4
If we have solved (12) (the boundary condition needs considerations), then the squared quadratic Wasserstein is given by

$$W_2^2(\mu, \nu) = \int_X \|x - \nabla u(x)\|_2^2 \rho_0(x) \, dx$$

Numerically solving (12) is not easy. First, it would only have a viscosity solution, and simple finite difference method based on Taylor expansion will not converge. Second, without convexity constraints the solution of (12) might not be unique, which give further restrictions on the approximate solution. In [6] the authors tackle the first issue by using a refined version of Barles and Souganidis framework [7], which says that approximation schemes will converge to the viscosity solution if they are consistent, stable, and nearly monotone. A monotone scheme will preserve maximum principle and leads to stability and convergence. For the convexity constraints, the authors enforce it into the approximation scheme. For example in two dimension, we define the “determinant of the Hessian of a convexity function $u$” as:

$$\det^+(D^2u) = \min_{\{v_1, v_2\} \in V} \left\{ \max\{u_{v_1, v_1}, 0\} \times \max\{u_{v_2, v_2}, 0\} \right\},$$

where $V$ is the set of all orthonormal basis for $\mathbb{R}^2$. Then by substituting $\det^+$ for $\det$ in (12) the convexity constraints could be removed, because at this time a solution must be convex. The $\det^+$ operator is discretized monotonically by computing the minimum over finitely many directions $\{v_1, v_2\}$. See the paper [6] for details. This method can guarantee its correctness even in singular cases, but might need huge computational cost in high dimensional problems.

The second method, namely by adding regularizations, will make the transport map smoother and the optimization problem strongly convex. By adding entropic regularizations [8] it can be transformed to a problem which involves projections of some input Gibbs density on an intersection of convex sets according to the Kullback-Leibler divergence. Iterative Bregman projections or Dykstra’s Algorithm solves this problem very fast and efficiently. Recently there are also attempts [9] to add fisher information regularizations (5) and establishes its connection to Schrödinger bridge problem. As a result, the presence of regularizations enriches the problem structures and makes the computation easier. This idea scales well in high dimension and can be applied to big data problems.

### 3 Application

In this section we discuss some applications of OT. Since it is a natural distance on the probability space $\mathcal{M}(\mathbb{R}^n)$, it can be used to measure the difference between distributions. So it finds massive applications in statistics or other probabilistic models. As we can see, the Wasserstein metric potentially illustrates the difference between measures better than the classical Kullback-Leibler divergence or $L^2$ distance in certain cases, and its fruitful structures make it well-suited to model real world problems.
3.1 Seismology

A central step in seismic exploration is the estimation of basic geophysical properties. Full waveform inversion is a successful procedure for determining structures of the earth from surface measurements. The method is based on minimization of the misfit between recorded and synthetic signals, with respect to some parameters influencing the propagation of seismic waves, such as wave velocities and earthquake locations. Since in some cases signals could be viewed as signed density functions and thus Wasserstein metric can be used to measure the misfit.

The quadratic Wasserstein metric has several good theoretical properties for this task [11], such as the convexity and insensitivity to noise. We take the earthquake location problem in our paper [12] as an example. First we give a simple introduction of its background.

Given a seismic velocity structure below (figure 1), there is an earthquake happened at location $\xi_T$ and time $\tau_T$, both of which are unknown.

$$c(x, z) = \begin{cases} 
5.2 + 0.05z + 0.2\sin(\pi x/25), & 0km \leq z \leq 20km, \\
6.8 + 0.2\sin(\pi x/25), & 20km < z \leq 40km.
\end{cases}$$

The receivers on the ground observed many seismic signals $g_i, 1 \leq i \leq r$, for example in figure 1:

![Figure 1: red triangles indicate the receivers; total number: r](image)

The goal is to utilize these signals information to obtain $\xi_T, \tau_T$. It is an inverse problem and the forward model is that for any $\xi, \tau$, we could compute the synthetic signals using PDEs (for simplicity denoted by an operator $\mathcal{L}$):

$$f_i = \mathcal{L}_i(\xi, \tau, c) \quad 1 \leq i \leq r.$$ 

So under this setting our optimization problem is

$$(\xi_T, \tau_T) = \arg\min_{\xi, \tau} \sum_{i=1}^{r} d(f_i, g_i)$$
where $d$ measures the difference between $f_i$ and $g_i$. Traditional choice of $d$ is the $L^2$ distance, and in our paper we choose

$$d(f, g) = W_2^2 \left( \frac{f^2}{\langle f^2 \rangle}, \frac{g^2}{\langle g^2 \rangle} \right)$$

where $\langle \cdot \rangle$ is the integral operator. As is proved in [11], this metric has many good properties. For example if we consider the shift of signals in figure 3, then the behavior of $W_2^2$ and $L^2$ are very different. The results are shown in figure 4.

Since we need to find the real time $\xi_T$, the time shift is a common phenomenon. If $L^2$ is adopted, from the figure we can see that it leads to many local minimizers in the optimization landscape, and also it cannot measure the difference well because the metric does not change when the shift exceeds a small constant. However, $W_2^2$ has a nice quadratic structure and performs better. There are also many other reasons to choose $W_2^2$ and the transformation $f^2/\langle f^2 \rangle$. Interested readers can refer to [12] for details. Roughly speaking, the quadratic
Wasserstein metric refines the optimization landscape, makes the global minimization easier, and handles noisy occasions well.

### 3.2 Learning problem

Statistical learning often involves characterizations and operations of density functions. The objective function to be optimized often writes \( d(\rho(\theta), \rho_0) \), where \( \theta \in \mathbb{R}^n \) is the optimization variable and \( \rho(\theta), \rho_0 \) are synthetic and true densities, respectively. \( d \) represents a distance which illustrates the similarity between \( \rho(\theta) \) and \( \rho_0 \). We often use the empirical distribution

\[
\rho_0^\text{em} := \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}
\]

to approximate the true \( \rho_0 \). As an example if we set \( d \) to be the Kullback-Leibler divergence then it leads to the famous Maximum Likelihood Estimate Method. Kullback-Leibler divergence enjoys fisher geometry, which could be utilized to accelerate its computation. We are now also working on using the Wasserstein metric to perform statistical learning and developing the corresponding geometric accelerating method. Hope to report it in subsequent papers.

In deep learning, the Generative Adversarial Networks (GAN) has been a very hot topic. The \( L^1 \) Wasserstein metric is adopted in the popular WGAN [4], which could handle the vanishing gradient problem well. As you can see in figure 4, \( W_2^2 \) establishes a quadratic loss (for \( W_1 \), linear), but \( L^2 \) or KL divergence will have gradient zero when the point deviates from the origin.

In GAN, rather than using a family of explicitly parametrized density functions, it fixes a original distribution \( \mu \) and uses a family of functions to push-forward it, i.e. \( \rho(\theta) = f_\theta \# \mu \). By setting \( d \) to be \( W_1 \) and using the formulation (6), the optimization problem writes

\[
\min_{\theta} \max_{\xi} \int_{\mathbb{R}^n} \varphi(\xi)(f_\theta \# \mu(x) - \rho_0^\text{em}(x))dx
\]
where the Kantorivish potential $\varphi$ is parametrized by $\xi$. WGAN uses two deep neuron networks to represent $\theta$ and $\xi$, respectively. The paper [10] also discusses the geometric view of this model.

4 Conclusion

Optimal transport has been a promising topic in pure and applied mathematics. Its tremendous applications in a lot of subjects such as data science and image processing render it a hot research area. This note surveys the basics of OT and as a beginner, I am looking forward to receiving your comments, suggestions or potential discussions.

References


On Certain Square-free Triples

谭泽睿*

Abstract

Square-free numbers are those numbers without square factors other than 1, for example 1, 6, 77 are all square-free but 9, 12 are not square-free. It is widely-known that square-free numbers distribute well among all nature numbers and have a positive density 6/π². An interesting question is to find the number, or density, of square-free pairs (n, n + s), or even triple square-frees (n, n + s, n + t). Let 0 < s < t be two positive integers, x > 0 a real number. By using elementary methods, we obtain the following two asymptotic formulas:

\[
\sum_{n \leq x} \mu^2(n) \mu^2(n + s) = \frac{6}{\pi^2} \prod_p \left(1 - \frac{1}{p^2 - 1}\right) \prod_{p^2 \mid s} \left(1 + \frac{1}{p^2 - 2}\right) x + O\left(x^{3/4}\right),
\]

\[
\sum_{n \leq x} \mu^2(n) \mu^2(n + s) \mu^2(n + t) = \frac{6}{\pi^2} \prod_p \left(1 - \frac{2}{p^2 - 1}\right) \prod_{p^2 \mid s} \left(1 + \frac{1}{p^2 - 3}\right) \prod_{p^2 \mid [t, t-s]} \left(1 + \frac{1}{p^2 - 3}\right) \prod_{p^2 \mid (s, t)} \left(1 - \frac{1}{(p^2 - 2)^2}\right) x + O\left(x^{7/8}\right)
\]

where [a, b] denotes the least common multiple of a, b, (a, b) denotes the greatest common divisor of a, b, and the big O constants depend only on s, t.

1 Evaluation of Q

Let \( \mu(n) \) be the usual Möbius function. First we calculate

\[
Q(x; k, q^2) := \sum_{n \leq x, n \equiv k (q^2)} \mu^2(n)
\]

by using

\[
\mu^2(n) = \sum_{d^2 \mid n} \mu(d).
\]

* 数 62
We have
\[ Q(x; k, q^2) = \sum_{n \leq x, n \equiv k(q^2)} \mu^2(n) \]
\[ = \sum_{n \leq x, n \equiv k(q^2)} \sum_{d^2 | n} \mu(d) \]
\[ = \sum_{d^2 \leq x} \mu(d) \sum_{n \equiv k(q^2), d^2 | n} 1 \]
where finding the number of integers satisfying \( n \equiv k(q^2), d^2 | n \leq x \) is equivalent to finding the number of solutions of the indefinite equation \( d^2 y - q^2 z = k \). When \((d, q)^2 | k\), that number is
\[ \frac{x}{[d, q]^2} + O(1). \]

Therefore we have
\[ Q(x; k, q^2) = \sum_{d^2 \leq x, (d, q)^2 | k} \mu(d) \left( \frac{x}{[d, q]^2} + O(1) \right) \]
\[ = \frac{1}{q^2} \sum_{(d, q) = D(q, D^2)|k} \mu(D) \sum_{d = Dd_1, (d_1, qD^{-1}) = 1, d_1^2 \leq xD^{-2}} \mu^2(Dd_1) \mu(d_1) \frac{x}{d_1^2} + O(\sqrt{x}) \]
\[ = \frac{x}{q^2} \sum_{D^2(q^2, k)} \mu(D) \left( \frac{6}{\pi^2} \prod_{p | q} \left( 1 - \frac{1}{p^2} \right)^{-1} + O \left( \frac{D}{\sqrt{x}} \right) \right) + O(\sqrt{x}) \]
\[ = \frac{6x}{\pi^2 q^2} \prod_{p | q} \left( 1 + \frac{1}{p^2 - 1} \right) \sum_{D^2(q^2, k)} \mu(D) + O \left( \frac{D}{\sqrt{x}} \right) + O(\sqrt{x}) \]
\[ = \frac{6\mu^2(k, q^2)}{\pi^2 q^2} \prod_{p | q} \left( 1 + \frac{1}{p^2 - 1} \right) x + O(\sqrt{x}) \]
where \( \mu^2(a, b) \) denotes \( \mu^2((a, b)) \).

2 Evaluation of \( S_1 \)

We now calculate the following sum
\[ S_1(x; k, q^2, s) := \sum_{n \leq x, n \equiv k(q^2)} \mu^2(n) \mu^2(n + s). \]
Note that by letting $q = 1$ the sum becomes the first asymptotic formula in the abstract.

$$S_1(x; k, q^2, s) := \sum_{n \leq x, n \equiv k(\ell^2)} \mu^2(n) \mu^2(n + s)$$

$$= \sum_{d^2 \leq x+s} \mu(d) \sum_{\mu^2(n) = 1, n \equiv k(\ell^2), n \equiv -s(d^2), n \leq x} 1.$$

We now need to figure out those $n \leq x$ satisfying

$$n \equiv k(q^2), n \equiv -s(d^2).$$

By rewriting it in the form of an indefinite equation

$$n = k + q^2y = -s + d^2z$$

namely

$$k + s = d^2z - q^2y$$

it is easy to see that a necessary condition for the existence of a solution is $(d, q)^2 | k + s$. Now the Chinese Remainder Theorem guarantees a solution, say

$$n = k_0 + [d, q]^2Y.$$

Because of the condition $\mu^2(n) = 1$, we will also need that

$$\mu^2((k_0, [d, q]^2)) = 1$$

we see that

$$\mu^2((k_0, [d, q]^2)) = \mu^2((k_0, d^2), (k_0, q^2)) = \mu^2(k_0, d^2) \mu^2(k_0, q^2) = \mu^2(s, d^2) \mu^2(k, q^2).$$

So the number of solutions can therefore be written as

$$\mu^2(s, d^2) \mu^2(k, q^2) Q(x; k_0, [d, q]^2) = \frac{6\mu^2(k, q^2) \mu^2(s, d^2)}{\pi^2[d, q]^2} \prod_{p | [d, q]} \left( 1 + \frac{1}{p^2 - 1} \right) x + O(\sqrt{x})$$

provided $(d, q)^2 = D^2 | k + s$. 

13
Inserting the above formula, we have

\[ S_1 = \sum_{D^2 \mid (q^2, s+k)} \sum_{d \leq \sqrt{\tau}(q,d) = D} \mu(d) \frac{6 \mu^2(s,d^2) \mu^2(k,q^2)}{\pi^2[d,q]^2} \prod_{p \mid [d,q]} \left( 1 + \frac{1}{p^2 - 1} \right) + O \left( x^{3/4} \right) \]

\[ = \frac{6 \mu^2(k, q^2)}{\pi^2 q^2} \sum_{D^2 \mid (q^2, s+k)} \sum_{d=1}^{D_1 \leq x^{1/4}} \mu(Dd_1) \frac{\mu^2(s, D^2d_1^2)}{d_1^2} \prod_{p \mid d_1q} \left( 1 + \frac{1}{p^2 - 1} \right) + O \left( x^{3/4} \right) \]

\[ = \frac{6 \mu^2(k, q^2)}{\pi^2 q^2} \prod_{p \mid q} \left( 1 + \frac{1}{p^2 - 1} \right) \sum_{D^2 \mid (q^2, s+k)} \mu(D) \sum_{d_1 \leq \frac{x^{1/4}}{p}, (d_1,q)=1} \frac{\mu(d_1) \mu^2(s, D^2d_1^2)}{d_1^2} \prod_{p \mid d_1q} \left( 1 + \frac{1}{p^2 - 1} \right) \]

\[ \prod_{p \mid [s,q^2]} \left( 1 + \frac{1}{p^2 - 2} \right) \sum_{D^2 \mid (q^2, s+k)} \mu(D) \mu^2(s, D^2)x = O \left( x^{3/4} \right) \]

\[ = \frac{6 \mu^2(k, q^2) \mu^2(s + k, q^2)}{\pi^2 q^2} \prod_{p \mid q} \left( 1 + \frac{1}{p^2 - 1} \right) \prod_{p} \left( 1 - \frac{1}{p^2 - 1} \right) \prod_{p \mid [s,q^2]} \left( 1 + \frac{1}{p^2 - 2} \right) x + O \left( x^{3/4} \right) \]

\[ = \frac{6 \mu^2(k, q^2)}{\pi^2 q^2} \prod_{p \mid [s,q^2]} \left( 1 + \frac{1}{p^2 - 2} \right) \prod_{p} \left( 1 - \frac{1}{p^2 - 1} \right) \prod_{p \mid [s,q^2]} \left( 1 + \frac{1}{p^2 - 2} \right) x + O \left( x^{3/4} \right). \]
3 Evaluation of $S$

Now, we are going to calculate the following sum, inserting $S_1$, letting $\alpha = 1/8$ below, we have

$$S = \sum_{n \leq x} \mu^2(n)\mu^2(n + s)\mu^2(n + t)$$

$$= \sum_{n \leq x} \mu^2(n)\mu^2(n + s) \sum_{d \mid n + t} \mu(d)$$

$$= \sum_{d^2 \leq x + t} \mu(d) \sum_{n \equiv -t(d^2) \mod x} \mu^2(n)\mu^2(n + s)$$

$$= \sum_{d^2 \leq x} \mu(d)S_1(x; -t, d^2, s) + O \left( x \sum_{d > x^0} \frac{1}{d^2} \right)$$

$$= \frac{6x}{\pi^2} \sum_{d^2 \leq x} \mu(d)\mu^2(t, d^2)\mu^2(t - s, d^2) \prod_{p \mid d} \left( 1 + \frac{1}{p^2 - 1} \right) \prod_p \left( 1 - \frac{1}{p^2 - 1} \right) \prod_{p \mid d} \left( 1 + \frac{1}{p^2 - 1} \right) \prod_{p \mid d} \left( 1 + \frac{1}{p^2 - 2} \right) \prod_{p \mid d} \left( 1 - \frac{1}{p^2 - 1} \right) + O \left( x^{3/4+\alpha} + x^{-\alpha} \right)$$

$$= \frac{6x}{\pi^2} \prod_p \left( 1 - \frac{1}{p^2 - 1} \right) \prod_{p^2 \mid s} \left( 1 + \frac{1}{p^2 - 2} \right) \prod_{p^2 \mid d} \left( 1 - \frac{1}{p^2 - 1} \right) + O \left( x^{7/8} \right)$$

$$= \frac{6x}{\pi^2} \prod_p \left( 1 - \frac{1}{p^2 - 1} \right) \prod_{p^2 \mid s} \left( 1 + \frac{1}{p^2 - 2} \right) \prod_{p^2 \mid t-s} \left( 1 - \frac{1}{p^2 - 1} \right) + O \left( x^{7/8} \right)$$

where

$$f(d) := \prod_{p \mid d} \left( 1 + \frac{1}{p^2 - 1} \right) \prod_{p \mid d} \left( 1 + \frac{1}{p^2 - 2} \right) \prod_{p \mid d} \left( 1 - \frac{1}{p^2 - 1} \right) \prod_{p \mid d} \left( 1 - \frac{1}{p^2 - 2} \right) \prod_{p \mid d} \left( 1 - \frac{1}{p^2 - 1} \right).$$

Note that $f(d)$ is a multiplicative function. Consider $f(p)$ in two different cases,

$$1 - \frac{f(p)}{p^2} = 1 - \frac{1}{p^2 - 1}, \quad p^2 \nmid s$$

$$= 1 - \frac{1}{p^2 - 1}, \quad p^2 \mid s.$$
We have that
\[
S = \frac{6}{\pi^2} \prod_p \left( 1 - \frac{1}{p^2 - 1} \right) \prod_{p^2 \mid s} \left( 1 - \frac{1}{p^2 - 2} \right) \prod_{p^2 \nmid s} \left( 1 - \frac{1}{p^2 - 1} \right) \prod_{p^2 \mid [t, t-s]} \left( 1 - \frac{1}{p^2 - 2} \right)^{-1} \prod_{p^2 \nmid [t, t-s], p^2 \mid s} \left( 1 - \frac{1}{p^2 - 1} \right)^{-1} x + O\left( x^{7/8} \right).
\]

Upon further simplification we obtain the desired result
\[
S = \frac{6}{\pi^2} \prod_p \left( 1 - \frac{2}{p^2 - 1} \right) \prod_{p^2 \mid s} \left( 1 + \frac{1}{p^2 - 3} \right) \prod_{p^2 \mid [t, t-s]} \left( 1 + \frac{1}{p^2 - 3} \right) \prod_{p^2 \mid (s,t)} \left( 1 - \frac{1}{(p^2 - 2)^2} \right) x
+ O\left( x^{7/8} \right).
\]

\[\text{(1)}\]

\[\text{(2)}\]

4 Generalization

Let
\[
s := \{0, s_1, s_2, \ldots, s_{m-1}\}.
\]

We will call the set \( s \) admissible if for any prime \( p \), there is \( s \in s \) such that \( s \neq 0(p^2) \), otherwise we call it inadmissible. For a prime \( p \), we define the function
\[
v_p(s) := \max_{s_0 \in s} \sum_{s_0 = s_{(p^2), s \in s - \{s_0\}}} 1.
\]

In this way, the definition of admissibility can be reformulated as \( |s| - v_p(s) < p^2 \).

We define the following sum \( S(x; s) \) related to \( s \)
\[
S(x; s) := \sum_{n \leq x} \prod_{s \in s} \mu^2(n + s).
\]

It is obvious that when \( s \) is inadmissible, the above sum vanishes. When \( s \) is admissible, intuitively, we suggest the following asymptotic formula for \( S(x; s) \)
\[
S(x; s) = \prod_p \left( 1 - \frac{|s| - v_p(s)}{p^2} \right) x + O\left( x^{1-2-|s|} \right).
\]

By using the exactly same method in this paper, we can prove this formula by induction without much difficulty.
1 Kleinian Singularities

Let $G$ be a finite subgroup of $\text{SL}(2, \mathbb{C})$. We can assume $G$ is a subgroup of $\text{SU}(2)$. Then $G$ acts on $\mathbb{A}^2$ by $g(x, y) = (ax + by, cx + dy)$, where $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The origin is the unique point with non-trivial stabilizer and we get an isolated singularity of surface $\mathbb{A}^2/G = \text{spec}(\mathbb{C}[x, y]^G)$. Let $\mathbb{H}_1$ be the group of quaternions of norm 1. Then

$$\mathbb{H}_1 \cong \text{SU}(2), a + bi + cj + dk \mapsto \begin{pmatrix} a + bi & c + di \\ -c + di & a - bi \end{pmatrix}.$$ 

Let $\mathbb{H}_1 \cong \text{SU}(2)$ acts on pure quaternions, i.e quaternions of form $q = bi + cj + dk$ and identify the space with $\mathbb{R}^3$. We get an exact sequence

$$0 \rightarrow \{ \pm 1 \} \rightarrow \text{SU}(2) \rightarrow \text{SO}(3) \rightarrow 0.$$ 

Hence to classify finite subgroups of $\text{SL}(2)$, it suffices to classify subgroups of $\text{SO}(3)$. In fact, they are symmetry groups of the Platonic solids or dihedrals or rotations around an axis. Lifted to $\text{SU}(2)$, the so-called binary polyhedral group([8]) is classified below:

- $A_n$: cyclic group of order $n$
- $D_n$: binary dihegral of order $2n$
- $E_6$: binary tetrahedral group of order 24
- $E_7$: binary octahedral group of order 48
- $E_8$: binary icosahedral group of order 120

*数 43
Here A, D, E denote the type of the simply laced affine Dykin diagrams, they are Mckay diagrams of the groups, which will be defined in next section. The fact is that the Klein singularity of $\mathbb{A}^2/G$ are of the same type of $G$, i.e. the intersection matrix of the exception divisor of the minial resolution of the singularity is the Cartan matrix of the same type. The goal of the article is to explain the correspondence.

**Remark 1.1.** The surface $\text{spec}(\mathbb{C}[x, y]^G)$ is studied as early as Klein or du Val([2]). It turns out that $\mathbb{C}[x, y]^G$ is finitely generated and they are hypersurfaces in $\mathbb{A}^3$ defined by:

- $A_n: XY - Z^n$
- $D_n: Z^2 + X(Y + X^n)$
- $E_6: Z^2 + X^4 + Y^3$
- $E_7: Z^2 + X(Y^3 + X^2)$
- $E_8: X^2 + Y^3 + Z^5$

For example, if $G = \langle a \rangle$ is the cyclic group of order $n$ and $ax = \mu_n x, ay = \mu_n^{-1} y$, where $\mu_n$ is a $n$-th root of unity. Then $\mathbb{C}[x, y]^G$ is generated by polynomials $X = x^n, Y = y^n, Z = xy$.

## 2 McKay graph

Let $G$ be a finite subgroup of SU(2) and $\rho_0$ the natural 2-dimensional representation of $G$ as a subgroup of SU(2). Then define the Mckay graph as follows. The vertices are irreducible representations $\rho_i$ of $G$ labeled by $\delta_i$ the dimensions of $\rho_i, i = 0, 1, \cdots m$. Let $m_{ij} = \text{dim}(\text{Hom}(\rho_i \otimes \rho_0, \rho_j))$ and draw an arrow from $i$ to $j$ labeled by $m_{ij}$ if $m_{ij} \neq 0$. Then erase the arrow ends if they go into both directions and erase the label if it is equal to 1.

**Theorem 2.1. (J. McKay)** The Mckay graphs are:

- **cyclic group of order $n$:**

```
  \begin{array}{c}
  1 \\
  1 \\
  \cdots \\
  1 \\
  \end{array}
```

- **binary dihegral of order $2n$:**

```
  \begin{array}{c}
  1 \\
  1 \\
  \cdots \\
  1 \\
  \end{array}
```

- **binary tetrahedral group:**

```
  \begin{array}{c}
  1 \\
  1 \\
  \cdots \\
  1 \\
  \end{array}
```
In all cases, trival representation of $G$ corresponds to a vertex labeled 1.

They are affine Dykin diagrams of type $A_n, D_n, E_6, E_7, E_8$. Let $e_i$ be simple roots. Let $\delta = (d_0, \cdots, d_m)$. Then the Cartan matrix is given by $\langle e_i, e_i \rangle = -2, \langle e_i, e_j \rangle = 1$ if $i$ and $j$ are connected or 0 if $i$ and $j$ are not. Since A,D,E are simple laced, the affine Cartan matrix are semi-definite negative with radical spanned by $\delta = \sum d_i e_i$. We can check these fact case by case. However, there is a system way given by T.Springer, see [8].

3 Quiver varieties

We will identify $\mathbb{A}^2/G$ and its minimal resolution with certain quiver variety of the Mckay graph. We first review the definition of quivers.

3.1 Quivers

Definition 3.1. A quiver is a finite graph $Q = (Q_0, Q_1)$, where $Q_0$ is an ordered set of vertices and $Q_1$ a set of arrows.

A linear representation $\rho$ of a quiver $Q$ is a map which assigns each vertex $v_i \in Q_0$ an finite-dimensional vector space $V_i = \rho(v_i)$ over an algebraic closed field $k$ and each arrow $a \in Q_1$ from $v_i$ to $v_j$ an linear morphism $\rho(a) : E_i \rightarrow E_j$.

In the remaining part of the article, we fix $k = \mathbb{C}$ and any module or representation is finite dimensional.

Given a quiver $Q$, we have the path algebra $kQ$. A path of $Q$ is a sequence of arrows $a_n, \cdots, a_1$ such that the head of $a_{i+1}$ is the tail of $a_i$ for any $i$. Each $v_i \in Q_0$ is considered as a path $e_i$. Then $kQ$ is the $k$-vector space with paths in $Q$ form a basis. The multiplication of two path $a_1 \cdots a_n$ and $b_m \cdots b_1$ is $b_m \cdots b_1 a_n \cdots a_1$ if it is a path or 0 if it is not. We see $e_i$ are idempotents of $KQ$ and $1 = \sum e_i, e_i e_j = 0$ if $i \neq j$.

It is well-known that a representation of $Q$ is the same as a $kQ$-module. For a $kQ$-module $V$, the representaion $\rho_V$ assign each vertex $v_i$ the vector space $V_i = e_i V$ and assign any arrow $a$ from $v_i$ to $v_j$ the linear map $V_i \rightarrow V_j : x \mapsto ax$.

Let $d_i = \dim V_i, d = (d_0, d_1, \cdots, d_m)$ be the dimension vector of a representation $V$. $\forall a \in Q_1$, let $h(a)$ be the head of $a$ and $t(a)$ be the tail of $a$. Choose a basis for each $V_i$, then $\rho(a) \in$
the matrix of size $d_h(a) \times d_t(a)$. Hence a representation with dimension vector $d$ can be mapped to a point of $\text{Rep}(Q, d) = \prod_{a \in Q_1} M_{d_h(a), d_t(a)}$. Let $GL(d) = \prod_{v \in Q_0} GL(d_v, k)$. $GL(d)$ acts on $\text{Rep}(Q, d)$ by simultaneous conjugation, which is just the change of bases of vector spaces. Then equivalent classes of representations of quiver of dimension $d$ are mapped bijectively to orbits of the action of $GL(d)$ on $\text{Rep}(Q, d)$.

### 3.2 Quotients

We consider $\text{Rep}(Q, d)$ as an affine variety with the action of $GL(d)$. Then the orbits are not always closed and the naive quotient is not well-behaved. But Geometric Invariant Theory (GIT) will give us a way to parameterize some orbits.

Suppose $G$ is a reductive algebraic group acting on $X = \text{spec } A$ and $\chi : G \to \mathbb{C}^*$ is an one-dimensional character of $G$. Define

$$A^G(\chi) = \bigoplus_{n=0}^{\infty} A^G(\chi)_n,$$

where $A^G(\chi)_n = \{ \phi \in A : g^*\phi = \chi(g)^n \phi, \forall g \in G \}$.

In general, $A^G(\chi)$ is finitely generated over $\mathbb{C}$ and we define

$$X/\!\!/\chi G := \text{Proj} A^G(\chi).$$

In our case, a character of $G = GL(d)$ is a vector $\theta = (\theta_0, \theta_1, \cdots, \theta_m) \in \mathbb{Z}^m$. $\chi_\theta((g_0, g_1, \cdots, g_m)) = \prod_{i=1}^{m} \det(g_i)^{\theta_i}$. Set

$$R_\theta(Q, d) := \text{Rep}(Q, d)/\!\!/\theta GL(d)$$

be the moduli space of representations. Since the center of $G$ act identically, we always assume $d \cdot \theta = 0$, otherwise $R_\theta(Q, d) = \emptyset$.

In GIT quotients, different stable orbits correspond to different points while two semistable orbits may correspond to same point if their closures intersect([3]). In our case, a representation $\rho \in \text{Rep}(Q, d)$ is said to be $\theta$-semistable if the closure of its orbit does not contain the zero vector or, equivalently, if there exists non-constant homogeneous semiinvariant $f(\rho) \neq 0$. A representation $\rho$ is stable if its orbit is closed in the set of semistable points and its stabilizer is a finite group. We have a criterien given by A.King([5]):

**Theorem 3.2.** A representation $\rho \in \text{Rep}(Q, d)$ is semistable if and only if any subrepresentation $\rho'$ of $\rho$ with dimension vector $d'$ satisfies $d' \cdot \theta \geq 0$, it is stable if and only if the equality is strict for any proper subrepresentation.

### 3.3 Preprojective algebra

Let $Q = (Q_0, Q_1)$ be a quiver. $\tilde{Q}$ is obtained from $Q$ by doubling the number of arrows: $\forall a \in Q_1$, add $a^*$ to be the arrow such that $t(a^*) = h(a), h(a^*) = t(a)$. Define the deformed
preprojective algebra

\[ \Pi(Q) := k\tilde{Q}/(\sum_{a \in Q_1} (a^*a - aa^*)). \]

A representation of \( \Pi(Q) \) can be considered as a representation of \( k\tilde{Q} \) satisfying that

\[ \sum_{a \in Q_1, \ 	ext{h}(a) = v_i} \rho(a)\rho(a^*) - \sum_{a \in Q_1, \ 	ext{t}(a) = v_i} \rho(a^*)\rho(a) = 0, \forall v_i \in Q_0. \]

Then \( \text{Rep}(\Pi(Q), d) \) is a \( \text{GL}(d) \) invariant closed subvariety of \( \text{Rep}(\tilde{Q}, d) \). \( \Pi(Q) \) is similar with \( kQ \) and in future, we will use \( \Pi(Q) \) other than \( kQ \). The reason will be explicit in next section.

**Example 3.3.** Let \( Q \) be the McKay graph of \( G \). Order the vertices and consider it as a quiver. A \( \Pi(Q) \)-module \( V \) is called \( i \)-generated if it is generated by an element of \( e_iV \). We assume \( d_V = \delta \). Recall \( \delta \) is the dimension vector of dimensions of all irreducible representations of \( G \) and \( \delta_0 = 1 \), where \( \rho_0 \) is the trivial representation. We can choose \( \theta = (\theta_0, \theta_1, \cdots, \theta_m) \) such that \( \theta_j > 0, \forall j \neq 0 \) and \( \delta \cdot \theta = 0 \). If \( V \) is 0-generated, then any proper submodule has dimension vector \( \delta' \) such that \( \delta'_i = 0 \). Hence \( \delta' \cdot \theta > 0 \). We conclude that any representation of \( \Pi(Q) \) with dimension vector \( d = \delta \) is \( \theta \)-semistable if and only if it is 0-generated. And all semistable points are stable. Any representation is 0-semistable.

From now on, we fix \( \theta \) as in the example. We have the map

\[ \pi : R_\theta(\Pi(Q), \delta) \to R_0(\Pi(Q), \delta) = \text{Rep}(\Pi(Q), \delta)/\text{GL}(\delta) \]

which map \( \theta \)-semistable, i.e. 0-generated representations to their images in \( \text{Rep}(\Pi(Q), \delta)/\text{GL}(\delta) \). To explain McKay correspondence, we will show \( \pi \) is a minimal resolution of Kleinian singularity.

4 Quiver resolution of Kleinian singularity

4.1 Morita equivalence

An orbit \( G \cdot x_0 \) in \( \mathbb{A}^2/G \) which is not the origin is a closed subscheme \( Z \) of \( \mathbb{A}^2 \). \( H^0(\mathcal{O}_Z) \cong k^n \) and is \( G \)-invariant, where \( n = |G| \). Since \( G \) acts \( G \cdot x_0 \) freely and transitively and by the reason that as a finite subgroup of \( \text{SL}(2) \), \( k[G] \cong k[G]^* \) as \( G \)-module, \( H^0(\mathcal{O}_Z) \) is a regular representation of \( G \). Hence such orbit corresponds to a \( G \)-invariant ideal \( I \) of \( k[x, y] \) such that \( k[x, y]/I \) is isomorphic to \( k[G] \) as \( G \)-module. \( R = k[x, y] \) is a \( G \)-module and \( k[x, y]/I \) is a \( R \)-module. This motivates us to consider the skew group algebra \( R \# G \). \( R \# G \) consists of linear combination \( \sum_{g \in G} r_g g \), where \( r_g \in R \). The multiplication is defined by

\[ (rg)(r'g') = r^g r' gg', \]
where \( g_r = g_r \). Then a \( R\#G \) module is the same as a \( R \) module with \( G \) action which is compatible with the \( R \) action. It is easy to verify that

**Lemma 4.1.** \( Z(R\#G) = R^G \), where \( Z(R\#G) \) is the center of \( R\#G \).

The following theorem corresponds Kleinian surfaces to quivers.

**Theorem 4.2.** (W.Crowley-Bovey, M.Holland,[4]) \( \Pi(Q) \) is Morita equivalent to \( k[x, y]\#G \).

Before the proof of the theorem, we recall that two ring \( S \) and \( T \) are called Morita equivalent if there is an additive equivalence of the categories \( S\text{Mod} \) and \( T\text{Mod} \) of left modules. Any equivalent of categories \( F : S\text{Mod} \rightarrow T\text{Mod} \) and \( G : T\text{Mod} \rightarrow S\text{Mod} \) is given by a bimodule \( U \) such that \( F(M) = U \otimes_S M \) and \( G(N) = V \otimes_T N \), where \( V = \text{Hom}_T(U, T) \).

**Example 4.3.** Let \( e \in S \) be an idempotent. \( T = eSe \) is a \( k \)-algebra with the unity \( e \). Suppose that \( SeS = S \). Then \( S \) is Morita equivalent to \( eSe \) given by the bimodule \( eSe \).

Proof of Theorem 4.1 We should choose idempotents \( f_0, f_1, \cdots, f_m \in k[G] \) such that \( f_i f_j = 0 \) if \( i \neq j \) and \( 1 \in k[G]/f[k[G]], \) where \( f = f_0 + \cdots + f_m \). \( k[G] \) is a product of Matrix algebra \( M_{i,j} \). In fact we can take \( f_i = E_{i,1} \in M_{i,j} \). Let \( R_i, 0 \leq i \leq m \) be irreducible representations of \( G \). Then \( R_i \cong k[G]f_i \) as \( G \)-module. \( 1 \in k[G]/f[k[G]]. \) Hence \( 1 \in (R\#G)f(R\#G) \). By the above example, \( R\#G \) is Morita equivalent to \( fR\#Gf \). We claim \( fR\#Gf \) is isomorphic to \( \Pi(Q) \).

We first show that \( fk(x, y)\#Gf \) is isomorphic to \( k\tilde{Q} \), where \( k(x, y) \) is the noncommutative polynomials ring and \( k[x, y] = k(x, y)/(xy - yx) \). Let \( N = k[x, y]_1 \), the standard representation of \( G \). It is a \( k[G] \)-bimodule, where \( G \) acts on the left diagonally and on the right acts on \( k[G] \) by right multiplication. It can be proved that \( k(x, y)\#G \) is isomorphic to the tensor algebra of the bimodule \( N\#k[G]\).

\[
\text{Hom}_{k[G]}(R_i, N \otimes R_j) \cong \text{Hom}_{k[G]}(f_i, N \otimes k[G]f_j) = fiN \otimes k[G]fj
\]

In McKay quiver \( k\tilde{Q} \), each arrow is labeled \( \text{dim}(\text{Hom}_{k[G]}(R_i, N \otimes R_j)) \). We define \( \alpha : k\tilde{Q} \rightarrow fk(x, y)\#Gf \) by sending \( e_i \) to \( f_i \) and \( a \in \tilde{Q}_1 \) with \( h(a) = j, t(a) = i \) to \( \phi(a) \in fiN \otimes k[G]fj \cong \text{Hom}_{k[G]}(R_i, N \otimes R_j) \). By the definition of the path algebra, \( \alpha \) is an isomorphic of rings.

Now \( k[x, y] = k(x, y)/(xy - yx) \) and \( \Pi(Q) = k\tilde{Q}/(\sum_{a \in \tilde{Q}_1}(a^*a - aa^*)) \). To show \( fR\#Gf \cong \Pi(Q) \), we should only check that

\[
\sum_{a \in \tilde{Q}_1, h(a) = i} (1 \otimes \phi_a)\phi_{a^*} - \sum_{a \in \tilde{Q}_1, t(a) = i} (1 \otimes \phi_{a^*})\phi_a = -\delta_i f_i(xy - yx)
\]

for any \( i \). This is checked case by case for different finite subgroups of \( \text{SL}(2) \).

Let \( V \) be a \( R\#G \)-module, then under Morita equivalence, it is sent to \( fV \) as a \( fR\#Gf \) \( \cong \Pi(Q) \)-module. Note that \( f_i V \cong \text{Hom}_{k[G]}(R_i, V) \), where \( R_i \) are irreducible representations. Since \( f_i \) is send to \( e_i \in \Pi(Q) \), \( \text{dim}(e_i(fV)) = \text{dim}\text{Hom}_{k[G]}(R_i, V) \). Hence

**Lemma 4.4.** \( \Pi(Q) \)-modules with dimension vector \( \delta \) is Morita equivalent to \( k[x, y]\#G \)-modules which are isomorphic to \( k[G] \) as \( G \)-modules.
4.2 Resolution

We begin to study \( \pi : R_\delta(\Pi(Q), \delta) \to R_\delta(\Pi(Q), \delta) = \text{Rep}(\Pi(Q), \delta) / \text{GL}(\delta) \). By previous sections, points on \( R((Q); \delta) \) are \( \Pi(Q) \)-modules which are 0-generated with dimension vector \( \delta \), they are Morita equivalent to \( R\#G \)-modules which are isomorphic to \( k[G] \) as \( G \)-module and is \( R\#Gf_0 \) generated. Note that \( f_0 = \frac{1}{n} \sum_{g \in G} g \). Hence they are \( R\#Gf_0 \cong R \) generated, which means they are isomorphic to quotients of \( R = k[x, y] \). So points on \( R_\delta(\Pi(Q), \delta) \) are mapped bijectively to ideals

\[ \{ I \text{ ideal of } k[x, y]/I \cong k[G] \text{ as } G\text{-module} \}. \]

For any \( R\#G \)-module \( V \), let \( gr^V \) be the direct sum of the decomposition factors of \( V \), which is well-defined by the Jordan-Holder Theorem. Then \( gr^V \) is semisimple and is a \( R\#G/Rad(R\#G) \)-module, where \( Rad(R\#G) \) is the Jacobson radical.

**Lemma 4.5.** Two \( R\#G \)-module \( V, W \in \text{Rep}(\Pi(Q), \delta) \) are mapped to same point in \( R_0(\Pi(Q), \delta) = \text{Rep}(\Pi(Q), \delta) / \text{GL}(\delta) \) if and only if \( gr^V \cong gr^W \)

Proof Since we only consider finite dimensional modules, trace functions is well-defined. Then similar to representation of finite groups, \( gr^V \) is a \( R\#G/\text{Rad}(R\#G) \)-module, where \( \text{Rad}(R\#G) \) is the Jacobson radical.

Since \( gr^V \) is a direct sum of simple modules, it suffices to classify all simple \( R\#G \)-modules. Recall \( Z(R\#G) = R^G \). For any simple module \( V \), \( Z(R\#G) \) act as scalars, which defines a morphism from \( R^G \) to \( k \). So any simple module is correspond to a maxial ideal \( m_V \) of \( k[x, y]^G \).

If \( m_V \) corresponds to the orbit of origin. In this case \( (x, y) \) is the only maximal ideal of \( k[x, y] \) which is over \( m_V \). We know that \( m_V \subset k[x, y]^G \) acts as 0 on \( V \). Hence \( V \) is a \( (k[x, y]/m_V)^G \)-module. Elements in \( (x, y) \) are nilpotent in \( k[x, y]^G \). It is an easy exercise that \( (x, y)^G \) lies in the Jacobson radical of \( (k[x, y]/m_V)^G \), hence acts 0 on the simple module. Hence \( V \) is a simple \( (k[x, y]/(x, y))^G \) \( G \)-module. Thus \( V \) is an irreducible representation \( R \) of \( G \). It corresponds to simple \( \Pi(Q) \)-module \( S_0 \), which have dimension vector \( e_i \). (we use \( e_i \) to denote the dimension vector which is 1 on \( v_i \) and 0 on other vertices).

If \( m_V \) corresponds to an orbit \( Z = G \cdot x_0 \) of length \( |G| \). There are \( |G| \) maximal ideal of \( k[x, y] \) over \( m_V \). \( k^2 \to k^2/G \) is étale outside the origin. Hence \( k[x, y]/m_V \) is already reduced and isomorphic to \( H^0(\mathcal{O}_Z) \). We can identify the ring \( H^0(\mathcal{O}_Z) \) with \( k[G^\ast] \) by defining \( \phi(g) = f(gx_0) \). Then \( V \) is a \( k[G^\ast] \#G \)-module. Let \( \{ g^\ast | g \in G \} \) be the dual basis of \( k[G^\ast] \). We have

\[
k[G]^\ast \#G \to \text{End}(k[G]) \quad (4.1)
\]
\[
g_0g_1 \mapsto (g_1^{-1}g_0)^\ast \otimes g_0 \quad (4.2)
\]

23
is an isomorphism of rings. But there are only one isomorphic class of simple module of
the matrix ring \(\text{End}(k[G])\), which is \(k[G]\). Hence \(V\) is uniquely determined by \(m_V\) and is
isomorphic to \(k[G]\) as \(G\)-module. It is Morita equivalent to a simple \(\Pi(Q)\)-module with
dimension vector \(\delta\). Different orbits correspond to different simple modules.

Let \(M\) be a \(R\#G\)-module corresponds to a module in \(\text{Rep}(\Pi(Q), \delta)\). Then by our classi-
fication, either \(M \cong \text{gr}M \cong V\), where \(V\) is a simple module corresponds to a point in
\((\mathbb{A}^2/G)\setminus\{0\}\), or \(\text{gr}M \cong \oplus_{i=0}^m R_i^\delta\) since \(\text{gr}M \cong k[G]\) as \(G\)-module. Define a map from \(\mathbb{A}^2/G\)
to \(\text{Rep}(\Pi(Q), \delta)\)/\(\text{GL}(\delta)\): for any orbit \(G \cdot x_0\), we define an \(R\)-module structure on \(k[G]::
f \cdot g = f(gx_0)g, \forall f \in k[x, y]\). We see the map coincides with the correspondence given in
the classification above, is well-defined and is a bijection. So we can identify \(\mathbb{A}^2/G\) with \(\text{Rep}(\Pi(Q), \delta)\)/\(\text{GL}(\delta)\). Recall
\[
\pi : R \theta(\Pi(Q), \delta) \rightarrow \mathbb{A}^2/G = \text{Rep}(\Pi(Q), \delta)\)/\(\text{GL}(\delta)\)
\]
We see \(\pi^{-1}(V)\) must isomorphic to \(V\) if \(V\) is a simple module. Hence \(\pi\) is 1-1 outside the
singularity and
\[
E := \pi^{-1}(0) = \{V \in \text{Rep}(\Pi(Q), \delta)|V\text{ is 0-generated, } \text{gr}V \cong \oplus_{i=0}^m S_i^g\} \]
is the exceptional divisor.

### 4.3 The main theorem

For a \(\Pi(Q)\)-module \(V\), define the socle \(\text{soc}(V)\) to be the direct sum of simple submodules
of \(V\). Then \(\text{soc}(V)\) is semisimple. For any simple module \(S\), let \([V : S] := \dim \text{Hom}(S, V)\).
Recall \(S_i, i = 0, 1, \cdots, m\) are simple modules with dimension vector \(e_i\). If \(V\) is 0-generated
with dimension vector \(\delta\), then \([V : S_0] = 0\). Otherwise \(V/S_0\) is 0-generated. While \(\delta_0 = 1, e_0(V/S_0)\) has dimension 0, which is impossible. For \(i = 1, \cdots, m\), let
\[
E_i = \{V \in \pi^{-1}(0)|[V, S_i] \neq 0\}.
\]
Then \(E = \cup_{i \neq 0} E_i\).

**Theorem 4.6.** (W. Crawley-Boevey, [4]) \(E_i \cap E_j \neq \emptyset\) if and only if \(i\) and \(j\) are adjacent
in the Mckay graph. If \(E_i \cap E_j \neq \emptyset, E_i \cap E_j\) consists a unique module with socle \(S_i \oplus S_j\).
Moreover, \(E_i\) is a closed subset of \(\pi^{-1}(0)\) isomorphic to \(\mathbb{P}^1\).

Here we use \(0, 1, \cdots, m\) to present vertices. The theorem shows projective lines \(E_1, \cdots, E_m\)
intersect each other in the way of the Cartan matrix of the same type. This is our desired
explanation.

**Remark 4.7.** The main theorem is an explanation. The quivers don’t tell us self-intersection
numbers of \(E_i\) and we have not proved that \(\pi\) is a minimal resolution yet. These will be done
in next section by identifying quiver varieties with Hilbert schemes.
The proof depends on the following lemma from representation theory, which connects Π(Q)-modules with root systems.

**Lemma 4.8.** Let M be a t-generated Π(Q)-module with dimension vector d such that d_i = 1. Then d is a root. Conversely, for any real root α, there exists a unique t-generated Π(Q)-module with dimension vector α.

**Remark 4.9.** In A,D,E root systems, equip the space spanned by roots the symmetric product: \( \langle e_i, e_i \rangle = 2, \langle e_i, e_j \rangle = -1 \) if i and j are adjacent and \( \langle e_i, e_j \rangle = 0 \) if i and j are not adjacent given by Cartan matrix. Then real roots are those roots with positive length (they are all of length 2), and imaginary roots lie in the 1-dimensional radical spanned by \( \delta = \sum_{i=0}^{m} \delta_i e_i ([?] ) \). There are infinite 0-generated simple Π(Q)-module with dimension vector δ!

The proof of the lemma and the theorem are based on Ringel’s formula([4]): let M, N be two Π(Q)-modules with dimension vector \( d_M, d_N \), then

\[
\dim \text{Hom}(M, N) + \dim \text{Hom}(N, M) - \dim \text{Ext}^1(M, N) = \langle d_M, d_N \rangle
\]

Proof of the theorem If \( V \in E_i \cap E_j \), then \( V/(S_i \oplus S_j) \) is 0-generated with dimension vector \( \delta - e_i - e_j \). Then by the lemma, \( \delta - e_i - e_j \) is a root. This can happen if and only if \( i \neq j \) and i and j are adjacent.

If i and j are adjacent, \( \delta - e_i - e_j \) is a real root. By the lemma, there exists a unique 0-generated module M with dimension vector \( \delta - e_i - e_j \). By Ringel’s formula

\[
\dim \text{Hom}(V, S_i) + \dim \text{Hom}(S_i, V) - \dim \text{Ext}^1(V, S_i) = \langle \delta - e_i - e_j, e_i \rangle = -1.
\]

Notice that since V are 0-generated, \( \text{Hom}(V, S_i) = 0 \). \( \delta - 2e_i - e_j \) is not a root, hence \( \text{Hom}(S_i, V) = 0 \). We get \( \dim \text{Ext}^1(V, S_i) = 1 \). Similarly \( \dim \text{Ext}^1(V, S_j) = 1 \). We have the universal extension (the pullback of the two nontrivial extensions of V)

\[
0 \to S_i \oplus S_j \to W \to V \to 0.
\]

Then \( \text{soc}(W) = S_i \oplus S_j \) and \( W \in E_i \cap E_j \). Notice that non-split of extensions is equivalent to 0-generated, hence the uniqueness of such W up to isomorphisms is guaranteed by the uniqueness of the universal extensions.

We now show for \( i \neq 0 \), \( E_i \) is isomorphic to \( \mathbb{P}^1 \). Let L be the unique 0-generated Π(Q)-module with dimension vector \( \delta + e_i \). Since \( \delta + 2e_i \) is not a root, \( \text{Ext}^1(L, S_i) = 0 \). By Ringel’s formula, \( \dim \text{Hom}(S_i, L) = 2 \). For any \( f \in \text{Hom}(S_i, L) \), \( V_f = L/f(S_i) \) is a 0-generated module with dimension vector \( \delta \). Conversely, for any module V ∈ \( E_i \), by Ringel’s formula and previous arguments, \( \dim \text{Ext}^1(V, S_i) = \dim \text{Hom}(S_i, V) = 1 \), so there is a nontrivial extension

\[
0 \xrightarrow{f} S_i \to L \to V \to 0.
\]

Notice \( \dim \text{End}(L) = 1 \) since \( \dim e_0L = 1 \). So the map between \( \text{Hom}(S_i, L) \) and modules V with dimension vector \( \delta \) such that \( \text{Hom}(S_i, V) \neq 0 \) descends to a bijection between \( \mathbb{P}^1 = \mathbb{P} (\text{Hom}(S_i, L)) \) and \( E_i \). For the proof that it is a morphism between varieties, see [4] or [8].
Proof of the lemma We use induction on $|d| = \sum d_i$. Assume $M$ is $t$-generated with dimension vector $d$. There are two cases. If $\langle d, e_i \rangle > 0$ for some $i \neq t$. Then by Ringel’s formula $k := \dim \text{Hom}(S, M) = \langle d, e_i \rangle + \dim \text{Ext}^1(S, M) \geq \langle d, e_i \rangle$. By induction, $d - ke_i$ is a root, thus $s_i(d - ke_i) = d - ke_i - \langle d - ke_i, e_i \rangle e_i = d + (k - \langle d, e_i \rangle)e_i$ is also a root. By basic properties of root systems, $d$ is a root. If $\langle d, e_i \rangle \leq 0, \forall i \neq t$. Since $\langle \delta, e_i \rangle = 0$ and $\delta = \sum \delta e_i$, we get $\langle d, e_i \rangle \geq 0$. If $\langle d, e_i \rangle = 0$, $d$ lies in the radical and is an imaginary root. If $\langle d, e_i \rangle > 0$, $\langle d, e_i \rangle = 2 + \sum_{i \in I(t)} d_i e_i = 2 - \sum_{i \in I(t)} d_i$, where $I(t)$ is the simple roots adjacent to $t$ such that $d_i \neq 0$. Hence $I(t) = \{i\}$ for some $i$ and $d_i = 1$. Assume $a$ is the root such that $h(a) = i, t(a) = t$. Then $\rho(a)\rho(a^*) = 0$. Since the module is $t$-generated and $d_i = d_i = 1$, we see $\rho(a^*) = 0$. Let $M'$ be the submodule generated by an element in $e_iM$. Then $M'$ is an $i$-generated module with dimension vector $d - e_i$. By induction, $d - e_i$ is a root. Hence $s_{d - e_i}e_i = e_i - \langle d - e_i, e_i \rangle (d - e_i) = d$ is a root.

The existence is proved similarly by induction on $|d|$. Assume $d$ is a real root. If $\langle d, e_i \rangle > 0$ for some $i \neq t$. $d' = s_i(d) = d - \langle d, e_i \rangle e_i$ is a real root. By induction, there exists a unique $t$-generated module $M'$ with dimension vector $d'$. Then by Ringel’s formula $\dim \text{Ext}^1(S, M') = \langle d, e_i \rangle$. This give rise to a unique universal extension $0 \to S_i^{\langle d, e_i \rangle} \to M \to M' \to 0$.

If $\langle d, e_i \rangle \leq 0, \forall i \neq t$. Then using the same argument, we see there is a unique $i$ adjacent to $t$ such that $d_i \neq 0$. By induction, there is a unique $i$-generated module $M'$ with dimention vector $d - e_i$. We can construct the desired module $M$ by adding an 1-dimensional space at $t$ and let $\rho(a^*) = 0, \rho(a) \neq 0$.

\[\-boxed{}\]

## 5 Resolution of Hilbert schemes

In this section, we will show $\pi : R_0(P(Q), \delta) \to R_0(P(Q), \delta) = \mathbb{A}^2/G$ is a minimal resolution. We need the punctual punctual Hilbert schemes of $\mathbb{A}^2$, which is the moduli space of 0-dimensional subschemes.

Define a functor $\mathcal{H}_{X,n}$ from $k$-schemes to sets for a quasi-projective variety $X$ over $k$

$$\mathcal{H}_{X,n}(S) = \{\text{closed subschemes of } Z_S \subset X_S, (p_S)_*(\mathcal{O}_{Z_S}) \text{ is locally free of rank } n\}.$$

**Theorem 5.1** (A. Grothendieck). $\mathcal{H}_{X,n}$ is represented by a scheme of finite type over $k$.

The scheme is denoted by $X^{[n]}$ and is called the punctual Hilbert schemes. Then $X^{[n]}(\text{spec}(k)) = \{Z| \text{ closed subschemes of } X \text{ of dimension } 0, \dim_k H^0(Z, \mathcal{O}_Z) = n\}$ is the set of closed points. Let $X^{(n)}$ be quotient of $X^n$ by the action of symmetric group $S_n$. Then $X^{(n)}$ is just the moduli space of $n$ points. There is a natural map $c : X^{[n]} \to X^{(n)}$
called the cycle map/Hilbert-Chow morphism. If $S = \text{spec}(k)$, $c(Z) = \sum_{x \in \text{supp}(O_Z)} \dim(\kappa(x))x$

Let $v = (v_1, \cdots, v_n)$ be a partition of $n$, i.e. $n = v_1 + \cdots + v_k$, $v_1 \leq v_2 \cdots \leq v_k$. Let $X^{(n)}_i \subset X^{(n)}$ consists of $n$ points which is $\sum_{i=1}^{k} v_i x_i$ for distinct $k$ points $x_1, \cdots, x_k$. Let $(1^n)$ be the part

ition $n = 1 + \cdots + 1$. Then $X^{(n)}_{1^n}$ are subsets of distinct $n$ points. Define $X^{[n]}_{1^n} = c^{-1}X^{(n)}_{1^n}$.

Then $c$ is a bijection restricted to $X^{[n]}_{1^n}$.

**Theorem 5.2** (J. Forgarty). Assume $X$ is a quasiprojective smooth surface. Then $X^{[n]}_{1^n}$ is a smooth connected variety of dimension $2n$ and $c : X^{[n]} \to X^{(n)}$ is a resolution of singularity.

We only prove the theorem for Hilbert schemes of $\mathbb{A}^2([6])$. Let $\tilde{H}$ be the set of triples $(B_1, B_2, i) \in (\text{End}(k^n), \text{End}(k^n), \text{Hom}(k, k^n))$ such that (1) $[B_1, B_2] = 0$ and (2) (stability) there exists no proper subspace $S \subset k^n$ such that $B_2(S) \subset S(\alpha = 1, 2)$ and $\text{im}r \subset S$. Define an action of $\text{GL}(n)$ on $\tilde{H}$ by $g(B_1, B_2, i) = (gB_1g^{-1}, gB_2g^{-1}, gi)$. The $(\mathbb{A}^2)^{[n]} = H := \tilde{H}/\text{GL}(n)$. In fact, a point in $(\mathbb{A}^2)^{[n]}$ is an ideal $I \subset k[x, y]$ such that $\dim(k[x, y]/I) = n$. Then we can let the linear space $V = k[x, y]/I$ and $B_1, B_2$ be the linear transformation of multiplication by $x, y$ and $i(1) = 1 \in k[x, y]$. By the conditions on triples we can construct the converse map, hence we have $(\mathbb{A}^2)^{[n]} = H$. It can be shown that the map $\text{End}(k^n) \times \text{Hom}(k, k^n) \to \text{End}(k^n), (B_1, B_2, i) \mapsto [B_1, B_2]$ has constant rank. Thus $H$ is smooth. $\text{GL}(n)$ acts freely on $\tilde{H}$. Hence $(\mathbb{A}^2)^{[n]}$ is a nonsingular variety and it can be shown that $H$ represents $\mathcal{H}_{\mathbb{A}^2, n}$.

**Example 5.3.** Since $[B_1, B_2] = 0$, the two matrix can be made into upper triangular simultaneously as

$$B_1 = \begin{pmatrix}
\lambda_1 & \cdots & * \\
& \ddots & \\
\lambda_n & & \\
\end{pmatrix},
B_2 = \begin{pmatrix}
\mu_1 & \cdots & * \\
& \ddots & \\
& & \mu_n \\
\end{pmatrix}$$

Then the cycle map is given by $c(B_1, B_2, i) = \{(\lambda_1, \mu_1), \cdots, (\lambda_n, \mu_n)\}$.

Let $G$ be a finite subgroup of $\text{SU}(2)$ and $|G| = n$. The orbit which is not the origin as a reduced closed subscheme is then a point on $(\mathbb{A}^2)^{[n]}$.

**Definition 5.4.** The $G$-Hilbert scheme of $\mathbb{A}^2$ is the irreducible component $G$-$\text{Hilb}(\mathbb{A}^2)$ of $(\mathbb{A}^2)^{[n]}$ that contains a $G$-orbit of length $|G|$. 

**Proposition 5.5.** $G$ acts identically on $G$-$\text{Hilb}(\mathbb{A}^2)$ and $\forall Z \in G$-$\text{Hilb}(\mathbb{A}^2), H^0(\mathcal{O}_Z)$ is a regular representation of $G$.

**Proposition 5.6.** $G$-$\text{Hilb}(\mathbb{A}^2)$ coincide with the set of points $Z \in ((\mathbb{A}^2)^{[n]})^G$ such that $H^0(\mathcal{O}_Z)$ is a regular representation of $G$.

**Remark 5.7.** The proposition is a conjecture in general case([8]).
Recall that

\[ R_0(\Pi(Q), \delta) = \{ I \text{ ideal of } k[x, y]|k[x, y]/I \cong k[G] \text{ as } G\text{-module}\}. \]

Hence by the proposition, we have

\[ R_0(\Pi(Q), \delta) = G-Hilb(\mathbb{A}^2) \]

If a finite group \( G \) acts on a smooth algebraic variety \( X \) over \( k \), then \( X^G \) is smooth. Hence \( G-Hilb(\mathbb{A}^2) \) is smooth. It is easy to see that \( ((\mathbb{A}^2)^{(n)})^G \cong \mathbb{A}/G \). Restrict the cycle map to the \( G \) invariant subvarieties, we get a resolution of \( \mathbb{A}/G \). We conclude that

\[
\begin{array}{ccc}
R_0(\Pi(Q), \delta) & \longrightarrow & G-Hilb(\mathbb{A}^2) \subset (\mathbb{A}^2)^{(n)} \\
\downarrow \pi & & \downarrow \iota \\
R_0(\Pi(Q), \delta) & \longrightarrow & \mathbb{A}^2/G \subset (\mathbb{A}^2)^{(n)}
\end{array}
\]

commutes.

It remains to show that the resolution is minimal. Recall a symplectic structure on a smooth variety \( X \) is a section \( \omega \) of \( \Omega^2_{X/k} \) which is nondegenerated everywhere. If \( X \) is of dimension \( 2n \), then \( \wedge^n\omega \) is a nowhere vanishing section of \( \Omega^2_{X/k} \). Hence the canonical divisor \( K_X = 0 \). \( \mathbb{A}^2 \) has the natural symplectic structure.

**Theorem 5.8.** Suppose a nonsingular surface \( X \) has a symplectic structure. Then \( X^{[n]} \) has a symplectic structure.

**Lemma 5.9.** Let \( G \) be a finite group acting symplectically on \((X, \omega)\). Then \( X^G \) is a symplectic subvariety.

The proofs can be found in [8]. \( SL(2) \) acts symplectically on \( \mathbb{A}^2 \). Hence \( G-Hilb(\mathbb{A}^2) \) is a symplectic subvariety.

**Corollary 5.10.** The restriction of the cycle map to \( G-Hilb(\mathbb{A}^2) \) is a minimal resolution of \( \mathbb{A}^2/G \).

Proof Since \( G-Hilb(\mathbb{A}^2) \) is symplectic, the canonical divisor \( K \) is trivial. Let \( D \) be an irreducible component of the exceptional divisor. Then \( K|_D = 0 \). We have the adjunction formula \( K_D = (K_S + D)|_D = D_D \). Hence \( D \cdot D = k_D \cdot D = \deg(K_D) = 2g - 2 \). Since \( D \cdot D < 0 \), we must have \( g = 0 \). Hence \( D \) is isomorphic to \( \mathbb{P}^1 \) with self-intersection number \(-2\). Thus it is a minimal resolution. 

\( \square \)
References


From Belyi’s theorem to dessins d’enfants

赵瑞屾*

Abstract

In this paper, firstly we introduce Belyi’s theorem and give a concrete (effective) proof for the ‘only if’ side. After that, we give two important applications. Finally, we introduce the conception of dessins d’enfants, a combinatorial subject that is closely related with Belyi’s theorem and many other fields.

Belyi’s theorem

As we all know, the category of compact Riemann surface is equivalent to the category of smooth projective curve over \( \text{spec}(C) \). Then we wonder when will the curve \( X \) be defined over a number field \( K \)? For a curve \( X \) over \( \text{spec}(C) \), it is said to be defined over a subfield \( K \) of \( C \), if and only if there exists a curve \( Y \) over \( \text{spec}(K) \), and under the base change \( \text{spec}(C) \rightarrow \text{spec}(K) \), \( \tilde{Y} \) is isomorphic to \( X \) over \( \text{spec}(C) \):

\[
\begin{array}{ccc}
X & \xrightarrow{f} & \tilde{Y} \\
\downarrow & & \downarrow \\
Y & \rightarrow & \text{spec}(K)
\end{array}
\]

(f is an isomorphism over \( \text{spec}(C) \))

For example, the projective curve \( X \) in \( P^2(C) \) defined by \( x^2 + y^2 + z^2 = 0 \) is obviously defined over \( Q \). For another curve \( Z \) in \( P^2(C) \) defined by \( x^2 + y^2 + \pi z^2 = 0 \), although its defining coefficients has transcend number, but \( Z \) is isomorphic to \( X \) over \( C \), thus \( Z \) is also defined over \( Q \).

*数 41
We concern about the case that $K$ is a number field due to consideration in arithmetic. Notice that a curve $X$ can always be described by finite number of equations, therefore it is equivalent to fixed the field $K$ to be $\mathbb{Q}$, when a curve is defined over $\mathbb{Q}$, it can always be descent to be defined over a number field (for example, we take all coefficient in its defining equations, then due to the total number of coefficient is finite and each one is algebraic over $\mathbb{Q}$, they will be contained in a finite extension of $\mathbb{Q}$).

This question is solved by Belyi and to our surprise, the answer looks quite simple:

**Belyi theorem:** A compact Riemann surface $X$ is defined over $\mathbb{Q}$ if and only if there exists a meromorphic function on $X$, $f : X \to \mathbb{P}^1$ such that $f$ only has three ramification values $0, 1, \infty$. In such case, the function $f$ can be chosen to be defined over $\text{spec}(\mathbb{Q})$.

For the 'if' side, it is solved quite early before Belyi. It is an interesting example of descent method. We omit its proof because we won’t use this side in this paper and the method is classical. Here are some good references for interested readers. The standard way is using weil’s criterion (Weil’s paper, *The field of definition of a variety*). If you are only concerning about the base change $\text{spec}(\mathbb{C}) \to \text{spec}(\mathbb{Q})$, we refer to a theorem in *Galois groups and fundamental groups* (Tamás Szamuely, chapter 4, this theorem is also useful latter, it can help us compute etale fundamental group of curves over $\text{spec}(\mathbb{Q})$).

For the 'only if' side, we will give a concrete construction so that our method is effective (this point is important in latter application).

Before going into the proof, we’d like to make some remark. The most astonishing thing is that the proof looks very simple as we will see! But the conclusion is strong. In fact if $X$ is not $\mathbb{P}^1$, 3 is the least number of ramification values: Just notice that a map $f$ between compact Riemann surfaces, $X \to Y$, is determined by unramify cover $f^{-1}(U) \to U$ ($U$ is $Y$ dropping out ramification values of $f$), and for $\mathbb{P}^1$, when it drops zero or one point, the remaining part is still simply-connected; if it drops two points, the remaining part is $U = \mathbb{C} - 0$ (complex plane without origin), each cover is isomorphic to $U \to U$, $z \to z^n$. And moreover in the second part, we will give a simple and combinatorial description (dessins d’enfants).

We will divide the proof into two steps, each step is proceeding by using the dilation formula: For map $f$ between smooth projective curves $X \to Y$, we denote ramification values of $f$ by $\text{Ram}(f)$, then for another map $g : Y \to Z$, we have $\text{Ram}(gf) = g(\text{Ram}(f)) \cup \text{Ram}(g)(1)$.

**First step:**

The first step is to construct a map $f : X \to \mathbb{P}^1(\mathbb{Q})$ over $\text{spec}(\mathbb{Q})$ such that $\text{Ram}(f) \subseteq \mathbb{Q} \cup \{\infty\}$. (Here we denote $[x:1]$ by $x$ simply ($x$ lies in $\mathbb{Q}$), denote $[1:0]$ by $\infty$).

By definition, $X$ can be expressed by $V(I)$ in $\mathbb{P}^n(\mathbb{Q})$, where $I$ is a homogeneous ideal of $\mathbb{Q}[x_0, \ldots, x_n]$.

Choose the standard affine open cover for the projective space, $\{D^+(x_i)\}$, at least one open set has nonempty intersection with $X$. Take out this intersection, we get an affine
open subset $U$ for $X$, and $U$ is also a closed sub-variety of $A^n_\mathbb{Q}$. Combine with the projection map $A^n_\mathbb{Q} \to A^n_\mathbb{Q}$, the closed immersion $U \to A^n_\mathbb{Q}$, the open immersion $A^n_\mathbb{Q} \to P^1(\mathbb{Q})$, we get a map $U \to P^1(\mathbb{Q})$, and it can be extended into $X \to P^1(\mathbb{Q})$ explicitly (just notice that $X$ is irreducible smooth curve, and $X$ is covered by $D^+(x_i)$, for each pair $D^+(x_i), D^+(x_j)$, they have explicit coordinate transform in the intersection part, use this formula to extend this map to $X$), thus we construct a nontrivial map $f:X \to P^1(\mathbb{Q})$ over $\text{spec}(\mathbb{Q})$ explicitly.

Then, $\text{Ram}(f)$ is a finite subset in $\mathbb{Q} \cup \{\infty\}$. Now we only need to work on $P^1(\mathbb{Q})(X$ plays no role in the following process!). Our method is to construct a series of rational polynomials $P^1(\mathbb{Q}) \to P^1(\mathbb{Q})$, repeat the dilation formula (1) to reach our goal.

Due to polynomial (except constant) will always send $\infty$ to $\infty$, send points like $[x:1]$ to $[y:1]$, we can ignore $\infty$ in the following, can the problem becomes to make $\mathbb{Q} \cap \text{Ram}(f)$ (denote this by $S$) into a subset of $\mathbb{Q}$.

If $S$ is already a subset of $\mathbb{Q}$, we’re done, otherwise consider the monic rational polynomial $g$ with the least degree such that $g(a)=0$ for any irrational $a$ in $S$. ($g$ can be constructed explicitly, for each $a$, let $p_a$ denote its annihilation polynomial over $\mathbb{Q}$, the $g$ is their product without repeating).

Composite with $g$, we replace $f$ by $g(f)$, and use dilation formula (1) to analyze change of $S$.

This composition change rational numbers into rational numbers, and it send original part of irrational numbers of $S$ into $0$. If $S$ has no irrational numbers, we’re done, if not, the possible irrational number can only come from $g(Z(g'))(Z(g')$ denote zeros of $g'$). Then repeat the above process, to find new $g_2$, and continue the same thing. Notice that $g'$ will send each irrational number in $S$ into $0$, thus by definition, $\text{deg}(g_2) \leq \text{deg}(g') < \text{deg}(g)$ (use elementary Galois theory). Thus after each dilation process, the degree of new ‘$g’ will lower than the degree of previous degree of $g$, then after finite steps, we will reach our goal, $S$ has no irrational numbers.

Second step:
Like step 1, we can still ignore $\infty$ in the dilation process, and the problem becomes we have $f$ such that $S = \text{Ram}(f) \cap \mathbb{Q} \subseteq \mathbb{Q}$, we need to adjust it to make $\text{Ram}(f) \subseteq \{0, 1\}$.

The idea is similar as in step 1.

It doesn’t matter that we assume $\{0, 1\} \subseteq S$ (We can use $\mathbb{Q}$-linear form to make any rational number become 0 or 1, if $S$ has only one or two elements, we have finished our proof). Denote other elements by $\{x_1, \ldots, x_m\}$. And we can apply mobius transform $z \to -z, z \to \frac{1}{z}$ if necessary to make $0 < x < 1$ (the second transform is not polynomial, but it interchange $0$ and $\infty$, so it will cause no trouble). Thus we can assume $x = \frac{1}{r+t}(l$ and $r$ are positive integers).

Now we consider the polynomial: $p(z) = \frac{(l+r)^{l+r}}{l^{r+r}}z^{l}(1-z)^{r}$.
Notice that its possible ramification points are only 0,1 or \( x \). And it send \( \{0, 1, x\} \) into \( \{0, 1\} \), thus composite with this polynomial, we can reduce the size of \( S \) at least by 1, if \( S \) has only \( \{0, 1\} \), we’re done, if not, repeat the above process, due to the original size of \( S \) is finite, after finite steps, we will reach our goal.

**Remark:**

We refer to *Introduction to compact Riemann surfaces and dessins d’enfants* for concrete calculation. This book is very elementary (even freshman can read it happily). But it makes some small mistakes. (For example, it assumes that compact Riemann surface is planar curve, this is not right)

**Application 1**

An amazing application is that abc conjecture implies Mordell conjecture. In fact, we will show effective abc conjecture implies effective Mordell conjecture.

Both two conjectures can be defined over any number field, but we shall concern only on \( \mathbb{Q} \) to make things simple. (In fact, the proof can be translated into number field version easily, but we need more technical terms to define these two conjectures on that case)

Give a nonzero rational number \( a \) (\( a \) is neither 0 nor 1), we define:

\[
\begin{align*}
N_0(a) &= \prod_{V_p(a) > 0} p, \quad N_1(a) = \prod_{V_p(a-1) > 0} p, \quad N_{\infty}(a) = \prod_{V_p(a) < 0} p, \quad N(a) = N_0(a)N_1(a)N_{\infty}(a),
\end{align*}
\]

\( H(a)=H([a:1]) \) (\( V_p \) is the valuation by prime \( p \), the latter height is standard height function on \( \mathbb{P}^1(\mathbb{Q}) \), \( H([a:b])=\max\{||a||, ||b||\} \) (\( a,b \) is coprime, such pair for a point in the projective plane is unique up to sign )).

The abc conjecture says that for any \( \epsilon > 0 \), there exists a constant \( C_\epsilon \), such that for any coprime positive integers \( (a,b,c)(a+b=c) \), we have \( N(c) \geq C_\epsilon c^{1-\epsilon} \). In our setting, it can be restated as for any rational number \( x \) (neither 0 nor 1), \( N(x) \geq C\epsilon H(x)^{1-\epsilon} \).

Mordell’s conjecture is easier to state, it says that A smooth projective curve \( C \) over a number field with genus \( g \) larger than 1, \( C \) has finite many rational points. This is first proved by Faltings (so it is also called Faltings’ theorem). He use deep theory in algebraic geometry. Our method (under abc conjecture) is much simpler than that. However, we still need lots of height theory in this short proof.

An interesting fact is that abc conjecture in function field is very easy while there are no analogue of Belyi theorem in function field.

The key role played by Belyi theorem is the following estimate:

Choose Belyi map \( f:C \rightarrow \mathbb{P}^1(\bar{\mathbb{Q}}) \), denote its degree by \( d \), the cardinality of \( f^{-1}\{0, 1, \infty\} \) by \( m \), then apply **Riemann – Hurwitz theorem:**

\[
2g-2 = d*(\#\{\text{ramification points}\}) = -2d + \sum_{x \in C} (e(x)-1) \quad \text{(e(x) is ramification index at x)}
\]

\[
\Rightarrow d > m(\#) \quad \text{(e(x) is ramification index at x)}
\]

Therefore \( d > m(\#) \). This is the key point, we can’t get this estimate without Belyi
In the following, we need to use some standard theorems in height theory, we refer to Silverman’s *Diophantine geometry: an introduction* and Serre’s *Lectures on the Mordell-Weil Theorem*.

Let $D$ denote the divisor $f^*(0)$, $D = \sum_{Q \in V(f)} V_Q(Q)$, let $D_0 = \sum_{Q \in V(f)} (Q)$ (this divisor counts zero of $f$ without multiplicity). Then $\text{deg}(D) = d, \text{deg}(D_0) = d_0$.

For any rational point $P$ in $C$ that $f(P) \in P^1(\bar{Q}) - \{0,1,\infty\}$, we will bound its height using abc conjecture.

The reason for considering divisor $D_0$ is this: consider the height function attached to the divisor $h_{D_0}$, use knowledge from local height theory, we get $\log N_0(f(P)) \leq h_{D_0}(P) + O(1)$.

Then notice that $D$ is ample divisor, consider divisor $F = d_0 \ast D - d \ast D_0$ (which has degree 0), for smooth projective variety over number field, we will have the following estimate $h_F(P) \leq c \sqrt{h_D(P)} + 1 + O(1)$ ($c$ is constant, $P$ is $\bar{Q}$ point).

Then we have $\log N_0(f(P)) \leq h_{D_0}(P) + O(1) \leq \frac{d}{d_0} h_D(P) + O(\sqrt{h_D(P)}) \leq \frac{d}{d_0} h(f(P)) + O(\sqrt{h(f(P)))}$ (1) ($h$ is log($H$), height function in projective space in log form).

Composite with the isomorphism $P^1(\bar{Q}) \longrightarrow P^1(Q)$, we get other Belyi map $f_2, f_3$, repeat the above process, we get divisor $D_1 = \sum_{V_Q(f-1)>0} (Q), \text{deg}(D_1) = d_1, D_\infty = \sum_{V_Q(f)<0} (Q), \text{deg}(D_\infty) = d_\infty$, and:

\[
\log N_1(f(P)) \leq \frac{d_1}{d} \log h(f(P)) + O(\sqrt{h(f(P)))},
\]

\[
\log N_\infty(f(P)) \leq \frac{d_\infty}{d} \log h(f(P)) + O(\sqrt{h(f(P)))},
\]

Sum them together: $\log N(f(P)) \leq \frac{d}{d} \log h(f(P)) + O(\sqrt{h(f(P)))}$, apply abc conjecture and estimate (*):

\[
(1 - \epsilon - \frac{d}{d}) \log h(f(P))) \leq O(\sqrt{h(f(P)))}, \text{ thus } h(f(P)) \text{ is bounded!}
\]

Use height theory, $h(f(P))$ is the height function attached to divisor $D$, and points with bounded height is finite. Thus we prove the Mordell conjecture. Moreover, the above inequalities can be compute explicitly (we have construct the Belyi map explicitly, the error term $O(1)$ can be computed explicitly too), the only unknown thing is the constant $C_\epsilon$ in abc conjecture, if we can make it explicitly, then we can solve the Mordell conjecture effectively. Thus effective abc conjecture implies effective Mordell conjecture! Faltings shows that the number of rational points is finite, but the explicit bound heights of these rational points is unknown, a proof that can give bound of heights among these points will be called effective proof, the effective Mordell conjecture is a famous open problem in arithmetic geometry (Shinichi Mochizuki claims to prove abc conjecture, but I don’t know his proof and whether it is effective abc conjecture). For more discussion about effective Mordell conjecture, we refer to *Diophantine Geometry, An introduction*.

For other similar application, we refer to *Heights in diophantine geometry*, chapter 12, it
shows that abc conjecture can imply Roth’s approximation theorem (also a famous theorem in Diophantine Geometry), Belyi’s theorem also appears in its proof.

**Application 2**

For the second example, we’d like to talk about the out-Galois action, and embed $\text{Gal}(\overline{Q}/Q)$ into $\text{Out}(\widehat{F}_2)$. (Here $\widehat{G}$ means the profinite completion of $G$, $\widehat{G} = \lim \frac{G}{N}$ (N run over normal subgroup in G with finite index)).

For this application and the second part of this article, we need to consider Galois theory in the sense of Grothendieck. He made this theory more powerful and gave a deeper description. Roughly speaking, in a modern way, Galois theory is about equivalence between some different categories.

For example, we know the following categories are equivalent:

1. compact Riemann surface;
2. smooth projective curve over $\text{spec}(C)$.

Also fix a compact Riemann surface $X$, and a finite subset $S \subseteq X$, the three following categories are equivalent:

3. $(Y,f)$, $Y$ is compact Riemann surface over $X$, $f: Y \to X$, such that ramification values of $f$ lies in $S$;
4. $(W,h)$, $W$ is a connected topological manifold over $U = X - S$, $h: W \to U$ is a finite cover;
5. finite set with continuous $G$ ($G = \pi_1(U)$) action, and the action is transitive.

Category (5) is remarkable because it is closely related with Galois theory for field extension. In fact, give Galois field extension $K \to L$, let $G$ denote its Galois group. Then the classical Galois correspondence is in fact an equivalence between the following categories:

5. finite set with continuous $G$ action, and the action is transitive;
6. subextension $F$, $K \to F \to L$.

To make algebraic geometry involve into this picture, we need to develop fundamental group theory in algebraic geometry. Grothendieck develop etale fundamental group theory. Etale fundamental group unify geometry and arithmetic (For example, the etale fundamental group of $\text{spec}(k)$ is exactly $\text{Gal}(k_s/k)$ ($k_s$ the separated closure inside its algebraic closure)).

Now let’s come back to out Galois action. From now on, the following fundamental group will always refer to etale fundamental group. Let $U$ be the curve $P^1(Q) - \{0, 1, \infty\}$ (defined over $\text{spec}(Q)$ and denote $U$ to be its base change under $\text{spec}(\overline{Q}) \to \text{spec}(Q)$, then we have an exact sequence:

1. $\to \pi_1(U) \to \pi_1(\overline{U}) \to \text{Gal}(\overline{Q}/Q) \to 1.$ (2)

For an exact sequence $1 \to A \to B \to C \to 1$, we have a natural map, $C \to \text{Out}(A)$:

For any $c$ in $C$, choose $b$ in $B$ represents $c$, then due to $A$ is normal, the inner automor-
phism: $x \mapsto bxb^{-1}$ defines automorphism of $A$, if we modulo $\text{Inn}(A) \langle \text{Out}(A) = \text{Aut}(A)/\text{Inn}(A) \rangle$, this will define a group map $C \rightarrow \text{Out}(A)$.

Apply above idea to exact sequence (2), we get the so-called out Galois action (notice that each group is profinite group, thus has topology on it, we can also shrink the algebraic $\text{Out}(A)$ by $\text{Out}_c(A) = \text{Aut}_c(A)/\text{Inn}(A)$ (require the group map to be continuous)).

For more details about etale fundamental group and the above exact sequence, we recommend Lei Fu’s book *etale cohomology theory* strongly (this book gives a clean introduction to etale fundamental group in chapter 3).

Concerning about the out Galois action defined by exact sequence (2), the wonderful application of Belyi theorem is that we can show this map is injection! Thus we embed the absolute Galois group over $\mathbb{Q}$ into an outer group of an etale fundamental group. Moreover, we can shrink the latter group. Grothendieck has many fascinating idea about this. We refer to his *Esquisse d’un programme*. Later Drinfeld construct the Grothendieck-Teichmuller group, and the Galois group can be embedded into this group. Some people even conjecture that these two groups are the same (but this remains to be proved). Drinfeld employ deep knowledge about quantum algebra and so on, this is amazing (There are no obvious relation between $\text{Gal}(\overline{\mathbb{Q}})$ and quantum group after all), but I know little about this huge field. Thus we can only refer to Drinfeld’s paper *On quasi-triangular quasi-Hopf algebras and a group closely connected with $\text{Gal}(\overline{\mathbb{Q}})$*, and Ihara’s detailed proof, *On the embedding of $\text{Gal}(\overline{\mathbb{Q}})$ into $\widehat{GT}$*. Geometric Galois Actions has some other discussion.

Now let’s show the map $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{Out}(\mathbb{F}_2)$ (the etale fundamental group of $\overline{U}$) is injection.

The essential point is that we should let the Galois group acts on other things, such as curve over $\text{spec}(\overline{\mathbb{Q}})$ and their function fields:

Notice that the Galois group $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ gives a right action on $\text{spec}(\overline{\mathbb{Q}})$, for an element $g$, it defines an automorphism $\bar{g}$ of $\text{spec}(\overline{\mathbb{Q}})$, then through the base change defined by it, $g$ becomes a functor, transform $X$ (scheme over $\text{spec}(\overline{\mathbb{Q}})$) to $\bar{g}(X)$, thus the Galois group acts on curves defined over $\text{spec}(\overline{\mathbb{Q}})$ (this is left action).

$$
\begin{align*}
\bar{g}(X) & \rightarrow \text{spec}(\overline{\mathbb{Q}}) \\
\downarrow & \\
X & \rightarrow \text{spec}(\overline{\mathbb{Q}})
\end{align*}
$$

More precisely, if $X$ is given by concrete defining equations, the $\bar{g}$ action is just acting on the coefficients. For example, if $X$ is an elliptic curve in Legendre form, $y^2z = x(x - z)(x - \sqrt{2}z)$, under the conjugate action in the Galois group (send $a + bi$ to $a - bi$), it becomes $\bar{X}$,
defined by \(y^2z = x(x - z)(x + \sqrt{2}z)\)

Restrict on the category of smooth projective curve over \(\text{spec}(\mathbb{Q})\), this Galois action is faithful (The set is isomorphic classes in this category), which means that for any element \(g\) (\(g\) is not identity), there exists a smooth projective curve \(X\), \(\bar{g}(X)\) is not isomorphic to \(X\) (over \(\text{spec}(\mathbb{Q})\)).

For this point, We use some knowledge from elliptic curves (for more details, we recommend the excellent textbook, silverman’s *The arithmetic of elliptic curves* ). For an elliptic curve \(X\) over \(\mathbb{Q}\), we can associate it to an invariant \(j(X)\) (defined by a rational function of the coefficient, thus its value lies in \(\mathbb{Q}\)), the isomorphic class of \(X\) is determined by \(j(X)\) (\(X\) is isomorphic to \(Y\) if and only if \(j(X) = j(Y)\)). Moreover the j-map is surjection (any \(x\) in \(\mathbb{Q}\), there exists an elliptic curve \(X\), such that \(j(X) = x\)). From the discussion above, we know that for an element \(g\) in \(\text{Gal}(\mathbb{Q})\), \(j(\bar{g}(X)) = g(j(X))\). Then for any \(g\) not identity, there exists \(x\) in \(\mathbb{Q}\), \(g(x) \neq x\), choose the elliptic curve \(X\) with j-invariant \(x\), then \(\bar{g}(X)\) is not isomorphic to \(X\).

Now we can show the map \(\text{Gal}(\mathbb{Q}) \rightarrow \text{Out}(\tilde{F}_2)\) is injection:

Keep the idea that the Galois action on smooth projective curves is faithful, we shall argue this in function field version.

Write the exact sequence in field version (field \(L\) is the direct limit of all finite extension arising from an etale cover of \(\tilde{U}\), see Lei Fu’s book *etale cohomology theory*):

\[
1 \rightarrow \text{Gal}(L/\bar{Q}(x)) \rightarrow \text{Gal}(L/Q(x)) \rightarrow \text{Gal}(\bar{Q}(x)/Q(x)) \rightarrow 1.
\]

Now let’s translate the Galois action on curves(discussed above) into field version. Because the canonical isomorphism: \(\text{Gal}(\bar{Q}/Q) \rightarrow \text{Gal}(\bar{Q}(x)/Q(x))\), an element \(g\) in the first Galois group defines a field map \(\bar{Q}(x) \rightarrow \bar{Q}(x)(\text{fix } Q(x))\). Then extend it to an automorphism \(\bar{g}: L \rightarrow L\):

\[
\begin{array}{c}
Q(x) \xrightarrow{j} \bar{Q}(x) \\
\downarrow \text{Id} \quad \downarrow g \\
\bar{Q}(x) \xrightarrow{\bar{g}} \bar{Q}(x)
\end{array}
\]

\[
\begin{array}{c}
\bar{Q}(x) \rightarrow \bar{Q}(x) \\
\downarrow \bar{g} \\
\bar{Q}(L)
\end{array}
\]

Then for \(\bar{Q}(x) \rightarrow K \rightarrow L\), \(\bar{g}\) transform it into another subextension \((\bar{Q})^{(x)} \rightarrow \bar{g}(K) \rightarrow L\). Now for a smooth projective curve \(X\) with a Belyi map \(f:X \rightarrow P^1(\mathbb{Q})\)(which means that ramification values lie in \(\{0, 1, \infty\}\)), it will define an etale cover for \(\tilde{U}\): \(f^{-1}(\tilde{U}) \rightarrow \tilde{U}\), the function field of \(U\) is \(Q(x)\), then the function field \(K\) of the first one is a subextension \((\bar{Q})^{(x)} \rightarrow K \rightarrow L\), and two kinds Galois action(on curves and on its function field) are compatible: \(\text{Fun}(\bar{g}(f^{-1}(\tilde{U}))) = \bar{g}(\text{Fun}(f^{-1})(\tilde{U}))\) (the field is identified up to field isomorphism over \(\bar{Q}(x)\)).
If the element $g(g \neq 1)$ gives an inner automorphism, thus it acts on field isomorphism class of subfield extension trivially (here two subfield extension $M$ and $N$, we call them isomorphic if and only they are isomorphic to each other over $Q(x)$). However, we have shown that the Galois action on elliptic curves is faith, thus $g$ must be 1 (for a smooth projective curve, its isomorphic class is determined by isomorphic function field class), this is contradiction.

**Dessins d’enfants**

As we say above, Galois theory is in fact equivalence between some categories, especially we focus on equivalence among category (3),(4),(5). We fix the compact Riemann surface to be $Y = P^1$, let $S=\{0,1,\infty\}$, thus $U$ is just the complex plane $C$ without $\{0,1\}$.

Now consider a Belyi pair $(X,f)$ ($X$ is a compact Riemann surface over $Y$, with $f:X \to Y$ being Belyi map, with ramification value in $S$)(this is just an object in category(3)). Grothendieck’s idea to give a combinatorial description:

Denote the inverse image of 0 by white points on $X$ (each point in $f^{-1}(0)$, we colour it white), denote the inverse image of 1 by black points on $X$, and connects them use $f^{-1}(0,1)$ (the inverse image of interval $(0,1)$). Then we embed a bipartite graph (a graph with points being or black, and each edge connects points with different colour) into the compact Riemann surface $X$.

For example, the map in $P^1$, $z \mapsto z^3$ is drawing like this:

\[
\begin{array}{c}
\bullet \\
\circ \\
\bullet
\end{array}
\]

Above discussion is **dessins d’enfants**:

A dessins d’enfants is a pair $(X,D)$, where $X$ is an oriented compact topological surface, $D$ is a finite connected bipartite graph embedding in $X$, and $X - D$ is finite union of topological discs (called faces of $D$).

**Remark** From the definition, we know that it is just a CW decomposition of $X$ in topology. However, it has much more information than merely CW decomposition. For each $(X,D)$, it has monodromy action: Denote the finite set of edges by $E$, according to the orientation, for each white point $A$, enumerate its neighbor edge $\{j_0, j_1, \ldots, j_k\}$ by anti-clock order (we may identity a small neighbor of $A$ with disk in the plate (keep the orientation), $j_{i+1}$ is the edge follow $j_i$ if we count the edge around $A$ in anti-clock direction), do this for all white points, then we divide $E$ into disjoint ‘cycles’, we define action $M_0$ by this cycle decomposition (so $M_0$ rotate edges around white point positively), similarly use black points, we define $M_1$ action, then $E$ is transitive under the subgroup $< M_0, M_1 >$ (inside the permutation group of $E$). Recall our motivation for dessins d’enfants, for each pair $(X,D)$.
produced by Belyi map \( (X,f) \), we can see it coincides with monodromy group in topology \((f^{-1}(U))\) is finite cover of \( U \), while fundamental group of \( U \) is \( F_2 \), if we choose 0.5 as the base point, two anti-clock circle around 0 and 1, with radius 0.5, intersect at 0.5, will become the generators of \( F_2 \), its monodromy action is exactly what we defined.

Category (3) is equivalent to the category of dessins d’enfants (call it category (7)) (the map between \((X,D)\) and \((Y,H)\) is a continuous map from \( X \) to \( Y \) that send \( D \) to \( H \), send points to the same colour, and keep the action of \( M_0, M_1 \)). The proof is some topological argument (Each dessins d’enfants determine a finite cover over \( U \), thus a Belyi map), we refer to Introduction to compact Riemann surfaces and dessins d’enfants, chapter 4.

The most amazing thing is that object in category (7) looks so simple! For example, any tree is bipartite graph and we can embed it into complex plane (thus embed into \( P^1 \)), then we get a Belyi pair easily (it has only one face). Grothendieck name them as dessins d’enfants, which is in French, it means ‘children’s drawing’ (even children can draw it for fun). This category doesn’t involve complex structure at first glance.

Belyi’s theorem says that a curve can lie in category (3) if and only if it can be defined over \( \text{spec}(Q) \). Thus the equivalence connects arithmetic with topology and combinatorics. For example, the absolute Galois group can act on dessins d’enfants now. Then how to classify the Galois orbit? Can we distinguish Galois orbits of dessins d’enfants by topological or combinatorial invariant? These questions have motivated many study, lots of invariant have been constructed, but we still don’t know the answer.

Finally, we want to mention some example and give references for this researching field. For example, we can study the polynomial \( f: P^1 \rightarrow P^1 \), and consider when it will be a Belyi map. It is Belyi map if and only if it has at most two critical values (restrict on complex plane, we don’t need to consider the point \( \infty \)). Such polynomial is called Shabat polynomial. An interesting class among them is chebyshev polynomials.

For more discussion about these fields, we refer to graph on surfaces and their applications. Besides that, Geometric Galois Action 1,2 and The Grothendieck theory of dessins d’enfants are also remarkable. Grothendieck’s draft Esquisse d’un programme is invaluable, it will give you more wonderful idea. You will see the above discussion is only the simplest example in his great program.

Reference


A Brief Introduction of KAM Theory

王森淼*

Introduction

The purpose of my report is to give a brief introduction of the classical KAM theorem. I'd like to sketch the ideas and pick the key points rather than giving a detailed proof. KAM theory can be applied to solve abundant problems related to quasiperiodic motions. In this report, I only introduce the applications of KAM theory in nearly integrable Hamiltonian systems, but the techniques and methods used in the proof can be extended into other systems.

Before we go into the main theorem, let’s pave the way to make the statement of the theorem clear and comprehensible.

Action-angle coordinates

We introduce action-angle coordinates which is a set of canonical coordinates useful in solving integrable systems, especially useful when we want to analyze the periodicity or obtain the frequencies of oscillatory or rotational motion without solving the Hamilton’s equations. It defines an invariant torus where the action constant defines the surface of the torus and the angle variables parametrize the coordinates on the torus.

Integrable system

Moreover, recall that a Hamiltonian system is a dynamical system governed by Hamilton’s equations whose form is given by

\[
\begin{align*}
\dot{p}_i &= -\frac{\partial H}{\partial q_i} \\
\dot{q}_i &= \frac{\partial H}{\partial p_i}
\end{align*}
\]

where \(H(p,q)\) denotes the Hamiltonian. (For simplicity, I assume that all Hamiltonians \(H\) is time independent in this report.) In general, these 2\(N\) equations are hard to solve. But if

*数 52
the system is integrable, the Hamiltonian function can be obtained easily. It has the form:

\[ H(I, \phi) = h(I), \]

where \((I, \phi)\) are action-angle coordinates.

Note that the Hamilton’s equations is unchanged if we transform the coordinates canonically.

\[
\dot{I}_i = -\frac{\partial H}{\partial \phi_i}, \\
\dot{\phi}_i = \frac{\partial H}{\partial I_i}
\]

(2)

Hence, we find that

\[
I(t) = I^*, \\
\phi(t) = \phi^* + \omega(I)t
\]

(3)

\(\omega(t)\) is called the frequency map. Hence, every solution curve is a trajectory winding around the invariant torus (Kronecker torus) with constant frequencies,

\[
\omega(I^*) = (\omega_1(I^*), \ldots, \omega_N(I^*))
\]

(4)

Such tori are also called **Kronecker tori**.

**Nearly Integrable Hamiltonian system**

Now suppose we already have an integrable Hamiltonian function \(h(I)\), and then make small pertubations. So the new Hamiltonian function takes the form:

\[ H(I, \phi) = h(I) + f(I, \phi) \]

(5)

**Note:** In order to avoid complicated details in the proof of KAM theorem, we will assume some analytic properties of \(H(I, \phi)\), or equivalently, of \(f(I, \phi)\). Furthermore, we will assume that the \(H\) can be extended to a small subset \(A_{\sigma, \rho} = \{(I, \phi) \in C^N \times C^N : ||I - I^*|| < \rho, |Im(\phi_j)| < \sigma\}\) (\(||\cdot||\) denotes the \(l^1\) norm), and \(f\) is analytic on \(A_{\sigma, \rho}\), \(\|f\|_{\sigma, \rho} = \sup_{A_{\sigma, \rho}} |f(I, \phi)|\).

**Resonant and Nonresonant**

We can easily observe that we can classify the flows on Kronecker tori by the arithmetical properties of \(\omega\).

1. - The frequencies \(\omega\) are nonresonant:

\[
\langle k, \omega \rangle \neq 0 \text{ for all } 0 \neq k \in \mathbb{Z}^N.
\]

then on the Kronecker torus, each flow is dense and ergodic.
2. - The frequencies $\omega$ are resonant:

$$\langle k, \omega \rangle = 0 \text{ for some } 0 \neq k \in \mathbb{Z}^N.$$  

However, there comes a heated debate on whether these tori are stable or not. Or say it another way, do the trajectories of this perturbed Hamiltonian system still lie on invariant tori? The first result goes back to Poincare. He found that the resonant tori disintegrate when pertubated arbitrarily in a nondegenerate system. I will explain later what is a nondegenarate system. Only finitely many tori survive such a perturbation. It means that a dense set of tori is destroyed. So it seems that there is little chance for other tori to survive. For example, one planet exerts a periodic force on the motion of of a second, and if the orbital periods of the two are commensurate, this can lead to resonance and instability. But in 1954, Kolmogorov went in the opposite direction. He proved that most tori (strongly resonant tori) survived!

In this sense, we can say that nonresonant solutions are more stable than resonant ones. In fact, KAM theorem just states that the invariant tori are stable under small pertubations if we have certain conditions.

**Definition 1.** We say a vector $\omega \in \mathbb{R}^N$ satisfies $(L, \gamma)$-diophantine condition if 

$$|\langle \omega, n \rangle| = |\sum_{j=1}^{N} \omega_j n_j| \geq \frac{L}{|n|^\gamma} \text{ for all } n \in \mathbb{Z}^N - \{0\}$$

Maybe it’s not a regular definition, but I use it in this report in order to imply that it has a number theory background.

**Remark 2.** For every $\gamma > N$ and $L > 0$, almost every frequency $\omega$ satisfies $(L, \gamma)$-diophantine condition. We can use Dirichlet pigeon method to prove it.

### Canonical transformations and generating functions

Suppose there is a canonical transformation $(I, \phi) = \Phi(\dot{I}, \dot{\phi})$ i.e. $(\dot{I}, \dot{\phi}) = \Phi^{-1}(I, \phi)$, it preserves the symplectic form, therefore preserves the form of Hamilton’s equations. Then there exists $S_1(I, \phi)$ such that

$$Id\phi - \dot{I}d\dot{\phi} = dS_1$$  

We need the condition:

$$\det \frac{\partial^2 S_1}{\partial I \partial \phi} |_{(I, \phi)} \neq 0$$

We want to define a new generating function:

$$Id\phi + \dot{\phi}d\dot{I} = d(\langle \dot{I}, \dot{\phi} \rangle + S_1) = \Sigma(\dot{I}, \phi)$$

And the condition:

$$\det \frac{\partial^2 \Sigma}{\partial I \partial \phi} |_{(I, \phi)} \neq 0$$
\[ \Sigma(\tilde{I}, \phi) = \langle \tilde{I}, \tilde{\phi} \rangle + S_1(I, \phi) \]  

(10)

\[ I = \frac{\partial \Sigma}{\partial \phi} \]

\[ \tilde{\phi} = \frac{\partial \Sigma}{\partial I} \]  

(11)

are equivalent to \( \Phi(I, \phi) = (\tilde{I}, \tilde{\phi}) \).

\textbf{Note:} If \( \Phi \) is identity transformation, then \( \Sigma(\tilde{I}, \phi) = \langle \tilde{I}, \phi \rangle \).

\textbf{Remark 3.} There are also many other ways of generating canonical transformations.

\textbf{Now} it’s time to state classical KAM theorem explicitly.

\section*{Main theorem}

\textbf{Theorem 4} (KAM). Suppose that \( \omega(I^*) = \omega^* \) satisfies \((L, \gamma)\)-diophantine condition, and the Hessain matrix \( \frac{\partial^2 h}{\partial I^2} \) is invertible at \( I^* \) (also on some neighborhood). Then there exists \( \epsilon_0 > 0 \) such that if \( \| f \|_{\sigma, \rho} < \epsilon_0 \), the Hamiltonian system has a non-resonant solution with frequencies \( \omega(I^*) \). It means that there is a new pair of action-angle coordinates such that in terms of these variables the system is integrable.

\textbf{Remark 5.} In fact, we can weaken the assumption without destroy the validity of the KAM theorem. We can suppose the Hamiltonian is only finitely differentiable, but the proof need to be modified.

The proof of the theorem can be divided into two crucial parts whose ideas can be applied in other KAM problems. The two basic ideas are: (1) Linearize the problem and solve the linearized problem. Then we get an approximate solution. (2) Improve the approximate solution by using the solution of the linearized problem as the basis of a Newton’s method argument. KAM induction lemma will be powerful.

\section*{Step1: Linearize the Hamilton’s equations}

A naive idea is to use new variables \((\tilde{I}, \tilde{\phi})\) such that in terms of the new variables, the equations become integrable. However, not any change of variables (diffeomorphism) is permitted, they must be also canonical transformations. To find a proper transformation, generating function method is a pretty good method.

\[ h(\frac{\partial \Sigma}{\partial \phi}) + f(\frac{\partial \Sigma}{\partial \phi}) = \tilde{h}(\tilde{I}). \]  

(12)
Since $H$ is close to $h$ for $f$ small, we also hope that the canonical transformation $\Phi$ is close to the identity transformation. Knowing that the generating function for the identity transformation is the inner product, we will look for the generating functions of the form

$$\Sigma = \langle \tilde{I}, \phi \rangle + S$$  \hfill (13)

We substitute it into

$$h(\tilde{I} + \frac{\partial S}{\partial \phi}) + f(\tilde{I} + \frac{\partial S}{\partial \phi}, \phi) = \tilde{h}(\tilde{I})$$  \hfill (14)

Expand it, and omit the terms of higher orders. Note that

$$\frac{\partial h}{\partial \tilde{I}} = \omega$$  \hfill (15)

Then we obtain the linearized equation:

$$\langle \omega, \frac{\partial S}{\partial \phi} \rangle + f(\tilde{I}, \phi) = \tilde{h}(\tilde{I}) - h(\tilde{I})$$  \hfill (16)

We expand $f(\tilde{I}, \phi)$ and $S(\tilde{I}, \phi)$ into Fourier series with respect to $\tilde{I}$,

$$f(\tilde{I}, \phi) = \sum_{n \in \mathbb{Z}^N} \hat{f}(\tilde{I}, n) e^{i2\pi \langle n, \phi \rangle}$$

$$S(\tilde{I}, \phi) = \sum_{n \in \mathbb{Z}^N} \hat{S}(\tilde{I}, n) e^{i2\pi \langle n, \phi \rangle}$$  \hfill (17)

Substitute them into (16), then we get

$$S(\tilde{I}, \phi) = \frac{i}{2\pi} \sum_{n \in \mathbb{Z}^N \setminus \{ 0 \}} \frac{\hat{f}(\tilde{I}, n) e^{i2\pi \langle n, \phi \rangle}}{\langle \omega(\tilde{I}), n \rangle}$$  \hfill (18)

But in fact, $S$ satisfies

$$\langle \omega, \frac{\partial S}{\partial \phi} \rangle + f(\tilde{I}, \phi) = 0$$  \hfill (19)

We have to estimate these two equations then. There is a question: whether these series convergent or not? Many people including Poincare believed that they diverged. Nonetheless, Kolmogorov, Arnold and Moser show that for "most" $\tilde{I}$, these series converge. However, having $S$ defined only on the complement of a dense set of points would be a problem. To deal with it, we can use the technique – truncating the sum.

Notice that $f$ is an analytic function in $A_{\sigma, \rho}$, so $\hat{f}(\tilde{I}, \phi)$ are decaying exponentially fast, then the error can be controlled. Define
\[ f^<(\tilde{I}, \phi) = \sum_{|n|<M} \frac{\hat{f}(\tilde{I}, n)e^{i2\pi(n, \phi)}}{\langle \omega(\tilde{I}), n \rangle} \]

\[ S^<(\tilde{I}, \phi) = \frac{i}{2\pi} \sum_{n \in \mathbb{Z}\setminus\{0\}} \frac{\hat{f}(\tilde{I}, n)e^{i2\pi(n, \phi)}}{\langle \omega(\tilde{I}), n \rangle} \]

\[ \Sigma^< = \langle \tilde{I}, \phi \rangle + S^< \]

It solves

\[ \langle \omega, \frac{\partial S^<}{\partial \phi} \rangle + f^<(\tilde{I}, \phi) = 0 \]  \hspace{1cm} (21)

If we choose \( M \) properly, \( S^< \) will be analytic, and its norm can be controlled. The nondegeneracy will be used here. We show that the generating function is well-defined and analytic.

First, define \( \Omega \), such that

\[ \max\{\sup \| \frac{\partial^2 h}{\partial I^2} \|, \sup \| (\frac{\partial^2 h}{\partial I^2})^{-1} \|, 1\} < \Omega \]

Now we know that \( \omega(I^*) = \omega^* \) and \( |\langle \omega^*, n \rangle| > \frac{L}{|n|} \). Recall that \( \frac{\partial h}{\partial I} = \omega \), we see that \( |\langle \omega(I), n \rangle - \langle \omega^*, n \rangle| \leq \Omega |n| \rho \), then \( |\langle \omega^*, n \rangle| \geq \frac{L}{2|n|} \) if \( \rho \) is small enough (need to be digested).

If we notice the fact that \( |\hat{f}(\tilde{I}, n)| \leq \|f\|_{\sigma, \rho}e^{-2\pi|n|} \), by Cauchy’s theorem, we find

\[ \|S^<\|_{\sigma-\delta, \rho} \leq C(N, \sigma, \rho; \delta)\|f\|_{\sigma, \rho} = C(\delta)\|f\|_{\sigma, \rho} \]  \hspace{1cm} (22)

**Remark 6.** We sacrifice the size of the domain in the action variables to preserve the analyticity of \( S^< \) and get an upper bound of it.

Now define

\[ I = \frac{\partial \Sigma^<}{\partial \phi} = \tilde{I} + \frac{\partial S^<}{\partial \phi} \]

\[ \tilde{\phi} = \frac{\partial \Sigma^<}{\partial \tilde{I}} = \phi + \frac{\partial S^<}{\partial \tilde{I}} \]  \hspace{1cm} (23)

then by the analyticity of \( S^< \) and inverse function theorem, we get an analytic and invertible canonical trasformation

\[ (I, \phi) = \Phi(\tilde{I}, \tilde{\phi}) \]  \hspace{1cm} (24)

**Remark 7.** Moreover, define \( \tilde{H}(\tilde{I}, \tilde{\phi}) = H \circ \Phi(\tilde{I}, \tilde{\phi}) = \tilde{h}(\tilde{I}) + \hat{f}(\tilde{I}, \tilde{\phi}) \). One of the advantages of this form of \( H \) is that if \( (\tilde{I}(t), \tilde{\phi}(t)) \) is the solution of \( \tilde{H} \), then \( \Phi(\tilde{I}(t), \tilde{\phi}(t)) \) is the solution of \( H \).
Step 2: The Newton Step

The basic idea is to repeat the step above, i.e. to define $\Psi_n = \Phi_0 \circ \Phi_1 \circ \cdots \circ \Psi_n$. Obviously, we can’t get the proper canonical transformation by iterating $\Phi$ in a finite number of steps. Besides, we need to give an explicit upper bound of $S$ and $\Phi$.

What’s more, there are two essential inductive constants $n$ and $M_n$ which must be defined carefully. $n$ (and $\delta_n$) controls the size of the domain of action variables. $M_n$ determines how we truncate the sum.

\[
\hat{h}(\tilde{I}) + \hat{f}(\tilde{I}, \tilde{\phi}) = H(\tilde{I}, \tilde{\phi}) = H(\tilde{I}, \frac{\partial S^{<}}{\partial \phi}, \phi) = h(\tilde{I} + \frac{\partial S^{<}}{\partial \phi}, \phi) + f(\tilde{I} + \frac{\partial S^{<}}{\partial \phi}, \phi)
\]

\[
= h(\tilde{I} + \frac{\partial S^{<}}{\partial \phi}) + f(\tilde{I} + \frac{\partial S^{<}}{\partial \phi}, \phi)
\]

\[
= h(\tilde{I}) + \{\omega(\tilde{I}), \frac{\partial S^{<}}{\partial \phi}\} + f(\tilde{I}, \phi) + \text{47} + \text{25}
\]

Define

\[
\hat{h}(\tilde{I}) = h(\tilde{I}) + \text{47} + \text{47} + \text{47} + \text{47}
\]

\[
\text{47} = \text{average} \{\text{47}\}
\]

\[
\text{47} = \text{average} \{\text{47}\}
\]

\[
\text{47} = \text{average} f(\tilde{I}, \phi(\tilde{I}, \tilde{\phi}))
\]

Notice that

\[
\langle \omega \tilde{I}, \frac{\partial S^{<}}{\partial \phi}\rangle + f(\tilde{I}, \phi) = f^{\geq}(\tilde{I}, \phi) = \sum_{|n| \geq M} \widehat{f}(\tilde{I}, n) e^{2\pi i (\phi, n)}
\]

We have the estimate

\[
\|f^{\geq}\| \leq \sum_{|n| \geq M} \|f\|_{\sigma, \rho} e^{-2\pi |n|} \leq C(\delta) \|f\|_{\sigma, \rho}^{2}
\]

if $M$ is large enough.

Finally, we get that

\[
\|h - \hat{h}\| \leq C(\delta) \|f\|_{\sigma, \rho}^{2}
\]

\[
\|\hat{f}\| \leq C(\delta) \|f\|_{\sigma, \rho}^{2}
\]

Before we go into the induction step, we need to make every inductive constant, variables and functions clear.
• \( \delta_n = \frac{\sigma}{36(1+n^2)} \), represent the reduction of the size of the domain of angle-variables.

• \( \sigma_0 = \sigma \), and \( \sigma_{n+1} = \sigma_n - 4\delta_n \), control the size of the domain of angle-variables.

• \( \rho_0 \leq \rho \), and \( \rho_{n+1} = \rho_n / 8 \), control the size of action-variables. \( \rho_0 \) is to be defined.

• \( \epsilon_0 = \| f \|_{\sigma, \rho} \), and \( \epsilon_n = \epsilon_0^{\frac{3}{4}} \), control the change of \( h_n \) and the norm of \( f_n \).

• \( M_n = \frac{|\log \epsilon_n|}{\pi \delta_n} \), control where we cut the sum of Fourier series.

• \( H_{n+1} = H_n \circ \Phi_n = h_{n+1} + f_{n+1} \)

• \( \omega_n(\tilde{I}) = \frac{\partial h_n}{\partial \phi}(\tilde{I}) \)

• \( A_{\sigma_n, \rho_n} = \{(I, \phi) \in C^n \times C^N : \| I - I_n \| < \rho_n, |Im(\phi_j)| < \sigma_n \} \)

• \( I_n \) is chosen s.t. \( \omega_n(I_n) = \omega^* \).

• \( \Omega_n = \max(1, \sup \| \frac{\partial^2 h_n}{\partial I^2} \|, \sup \| (\frac{\partial^2 h_n}{\partial I^2})^{-1} \|) \)

\( \Phi_n \) is the canonical transformation whose generating function is

\[
S_n^{<}(\tilde{I}, \phi) = \frac{i}{2\pi} \sum_{n \in \mathbb{Z}^N \setminus \{0\}} \tilde{f}(\tilde{I}, n) e^{2\pi i \langle n, \phi \rangle} \langle \omega(\tilde{I}), n \rangle \tag{30}
\]

It solves

\[
\langle \omega, \frac{\partial S_n^{<}}{\partial \phi} \rangle + f_n^{<}(\tilde{I}, \phi) = 0 \tag{31}
\]

\[
f_n^{<}(\tilde{I}, \phi) = \sum_{|n| \in M_n} \tilde{f}(\tilde{I}, n) e^{2\pi i \langle n, \phi \rangle} \tag{32}
\]

Then we can introduce the Induction Lemma.
Theorem 8 (KAM Induction Lemma). There exists a positive constant $c_1$ such that if

$$\epsilon_0 < 2^{-c_1 N(\gamma + 1)} \sigma^8 N(4 \gamma + 1) \rho_0^8 L^{16} \Omega N^{8}$$

then

- $\| S_n \|_{\sigma_n - \delta_n, \rho_n} \leq C(\delta_n) \epsilon_n$
- $\| f_n \|_{\sigma_n, \rho_n} \leq \epsilon_n$
- $\| h_{n+1} - h_n \|_{\sigma_{n+1}, \rho_{n+1}} \leq \epsilon_{n+1}$
- $\| I_{n+1} - I_n \| < \rho_n / 8$
- $\Phi_n$ is defined and analytic on $A_{\sigma_n - 3 \delta_n, \rho_n / 4}(I_n)$ and maps this set into $A_{\sigma_n - 2 \delta_n, \rho_n / 2}$.

The proof of this lemma is horrible, and I don’t want to discuss it in detail. I only point out that the nondegeneracy of $h$ is crucial in the proof. But before proving this lemma, I will show how the KAM Induction theorem follows from it. On the condition that $f$ is sufficiently small, all the hypothesis of Induction lemma will be satisfied.

Begin by defining $\Psi_n = \Phi_0 \circ \Phi_1 \cdots \Phi_n$. By the Induction lemma, $\Psi_n : A_{\sigma_n - 3 \delta_n, \rho_n / 4}(I_n) \to A_{\sigma_0, \rho_0}$, and $H_n = H_0 \circ \Psi_{n-1}$. In particular, if $(I^n(t), \phi^n(t))$ is a solution of Hamilton’s equations with Hamiltonian $H_n$, then $\Psi_{n-1}(I^n(t), \phi^n(t))$ is a solution of Hamilton’s equations with Hamiltonian $H_0$.

Consider the equations of motion of $H_n$:

$$\dot{i} = -\frac{\partial f_n}{\partial \phi}$$
$$\dot{\phi} = \omega_n(I) + \frac{\partial f_n}{\partial \rho}$$

Since $\| \frac{\partial f_n}{\partial \phi} \|_{\sigma_n, \rho / 2} \leq 2 \epsilon_n N / \rho_n$, and $\| \frac{\partial f_n}{\partial \rho} \| \leq \epsilon_n N / \delta_n$, the trajectory with initial conditions $(I_n, \phi_0)$, with remain in $A_{\sigma_n - 3 \delta_n, \rho_n / 4}(I_n)$ for all times $|t| \leq T_n = 2^n$, by our hypothesis on $\epsilon_0$, and the definition of induction constants. Furthermore, if $(I^n(t), \phi^n(t))$ is the solution with these initial conditions, we have

$$\max(\sup_{|t| \leq T_n} |I^n(t) - I_n|, \sup_{|t| \leq T_n} |\phi^n(t) - \phi_0 - \omega^* t|) \leq 2^{2n+2} \Omega_n N / \rho_n \delta_n.$$ (34)

Note that the inductive estimates on $I_n$ imply that there exists $I^\infty$ with $I_n \to I^\infty$, we see that for $t$ in any compact subset of the real line, $(I^n(t), \phi^n(t)) \to (I^\infty, \omega^* t + \phi_0)$. Using the
inductive bounds on the canonical transformation we can control $\|\Psi_n(I, \phi) - (I, \phi)\|_{\sigma_{n+1}, \rho_{n+1}}$ and $\|\Psi_n(I, \phi) - \Psi_{n-1}(I, \phi)\|_{\sigma_{n+1}, \rho_{n+1}}$.

Using the definition of the inductive constants, we get from the sum over $n$ of this last expression’s upper boundary converges and hence there exists

$$(I^*(t), \phi^*(t)) = \lim_n \Psi_n(I^\infty, \omega^* t + \phi_0)$$

which is a non-resonant function with frequency $\omega^*$. Similarly, for $t$ in any compact subset,

$$\lim_n |\Psi_n(I^n(t), \phi^n(t)) - \Psi_n(I^\infty, \omega^* t + \phi_0)| = 0 \quad (36)$$

Combining these two remarks, find that $\lim \Psi_n(I^n(t), \phi^n(t)) = (I^*(t), \phi^*(t))$, so $(I^*(t), \phi^*(t))$ is a non-resonant solution of Hamilton’s equations for the system with Hamiltonian $\mathcal{H}_0$ as claimed.

Remark 9. From the point of view of applications of this theory, it is often convenient to know not just what happens to a single trajectory, but rather the behavior of whole sets of trajectories. Suppose we have a bounded set $V \subset \mathbb{R}^N$ in which the nondegenerate conditions is true. For every $\delta > 0$, there exists a set $P \subset V \times T^N$, such that the Lebesgue measure of $P$ is less than $\delta$, and when the initial conditions $(I_0, \phi_0)$ is in the remainder, the trajectory is nonresonant. Moreover, those tori that are destroyed by perturbation become invariant Cantor sets.

With this remark, we can control a system by control the pertubations on it instead of control it explicitly.

Remark 10. In this proof (and the original paper of Arnold), the construction of induction steps and the domains are processed at the same time. But J. Poschel modified the method by making the two processes independent of each other. It’s more convenient.

Remark 11. The non-resonance conditions and nondegenerate conditions are very crucial (not only in the KAM theory of the nearly integrable system). However, these conditions of the KAM theorem become increasingly difficult to satisfy for systems with more degrees of freedom. As the number of dimensions of the system increases, the volume occupied by the tori decreases. And if the frequency is resonant, we can cut off the dimension of the tori, these tori are also stable.

Remark 12. In some dynamic systems, the number of the action variables and the angle variables are not the same. We can also find invariant (stable) tori.

Remark 13. When the smooth curves disintegrate or the pertubation is not that small, KAM theory is not powerful. Then we move to Aubry-Mather theory.
Some applications

There is a simpler problem which can be solved through the KAM method.

Consider orientation preserving diffeomorphisms of the circle. For brevity, we lift it to the real line:

\[ \phi : R \to R \]  

(37)

The simplest diffeomorphism is a rotation \( \phi(x) = R_\alpha(x) = x + \alpha \), and then we make a small perturbation: \( \phi(x) = x + \alpha + \eta(x) \). The question here is changed. We are also asking whether the rotation is stable (preserve some properties), i.e. analytically diffeomorphic to pure rotation.

We first introduce an important characteristic of circle diffeomorphisms, the rotation number:

**Definition 14.** The rotation number of \( \phi \) is

\[ \rho(\phi) = \lim_{n \to \infty} \frac{\phi^{(n)}(x) - x}{n}. \]

**Remark 15.** The definition makes sense because the right hand side of the equation exists and doesn’t rely on \( x \).

Obviously, we can assume that the perturbation is done on \( R_\rho \).

Secondly, like we have discussed before, we can also define diophantine condition of a real number.

**Definition 16.** The real number \( \rho \) satisfies \((K, \nu)\)-diophantine condition, if there exist positive numbers \( K \) and \( \nu \) s.t. \( |\rho - m/n| > K/|n|^\nu \), for all pairs of integers \((m, n)\).

**Remark 17.** For every \( \nu > 2 \) and \( K > 0 \), almost every irrational number \( \rho \) satisfies \((K, \nu)\)-diophantine condition. We can also use Dirichlet pigeon method to prove it.

**Theorem 18** (Arnold’s Theorem). Suppose that \( \rho \) satisfies \((K, \nu)\)-diophantine condition, and the rotation number of the diffeomorphism \( \phi(x) = x + \rho + \eta(x) \) is just \( \rho \). Then there exists \( \epsilon > 0 \), when \( \|\eta\|_\sigma = \sup_{|\text{Im}(z)| < \sigma} |\eta(z)| < \epsilon \), there exists an analytic and invertible change of variable \( H(x) \) which conjugates \( \phi \) to \( R_\rho \).

The proof is parallel to the proof of classical KAM theorem in nearly integrable systems. And we also have a corollary to describe the behavior of a "bundle" of diffeomorphisms.

**Theorem 19.** Consider the family of diffeomorphisms:

\[ \phi_{\alpha, \epsilon}(x) = x + \alpha + \epsilon \eta(x). \]

(38)

For every \( \delta > 0 \), there exists \( \epsilon_0 > 0 \) s.t. if \( \epsilon < |\epsilon_0| \), there exists a set \( A(\epsilon) \subset [0, 1] \) such that the Lebesgue measure of \( A(\epsilon) \) is less than \( \delta \) and for \( \alpha \) in the remainder, \( \phi_{\epsilon, \alpha} \) is analytically conjugate to a rotation of the circle.
一则 Lang 的趣事

如我们所知，Lang 是一位多产的作者，但是他的写作风格并不是读者友好型的。

Mordell 爵士在他为 Lang 的“丢番图几何”写的书评中对 Lang 的话“一个人写一篇高等的专著只是为了自己而写，因为他想用永恒方式的表达自己对数学中某些美的事物的看法，并不是出于为了易于为人接受的考虑。”表示不满。

Lang 在他对 Mordell 的“丢番图方程”一书写的书评中，做出了如下的回应：“Mordell 在他的书评中引用了我写的微积分教材的前言中的一段话，但有一些省略。我的原文是这样说的：“一个人写一篇高等的专著只是为了自己而写，因为他想用永恒方式的表达自己对数学中某些美的事物的看法，并不是出于为了易于为人接受的考虑，这就像作曲家按照乐谱演奏他的交响乐。”我这段话要强调的是写书的过程和音乐演奏的相似性。”
Morse Theory And Its Applications

金儒鸿

Abstract

Morse theory is a powerful tool to analyze the topology of a manifold by studying differentiable functions on that manifold. It enables one to find CW structures and handle decompositions on manifolds and to obtain substantial information about their homology. In the report, here are main conclusion of morse theory but without any prove of them since it is kind of complex. Besides, Morse Inequality is proved in detail and some interesting application such as Reeb sphere theorem, assuring the CW structure of $\mathbb{C}P^n$ and Poincaré-Hopf theorem, which gives directly solution of Hairy Ball Theorem.

1 Basic definitions

In the following statement, the words “smooth” and “differentiable” will be used interchangeably to mean differentiable of class $C^\infty$.

Definition 1.1 (Critical point). Let $f$ be a smooth real valued function on a manifold $M$. A point $p \in M$ is called a critical point of $f$ if the induced map $f_* : TM_p \to TR_{f(p)}$ is zero. If we choose a local coordinate system $(x^1, \cdots, x^n)$ in a neighborhood of $U$ of $p$ this means that
\[
\frac{\partial f}{\partial x^1}(p) = \cdots = \frac{\partial f}{\partial x^n}(p) = 0
\]
The real number $f(p)$ is called a critical point of $f$

Definition 1.2 (Hessian and Index). Let $f$ be a smooth real valued function on a manifold $M$. A critical point $p$ is called non-degenerate if and only if the matrix
\[
\left( \frac{\partial^2 f}{\partial x^i \partial x^j}(p) \right)
\]
is non-singular. This matrix induces a bilinear map $f_* : TM_p \times TM_p \to \mathbb{R}$, called the Hessian of $f$ at $p$. Its index is defined to be the maximal dimension of $s$ subspace of $TM_p$ on which $f_*$ is negative definite.
Definition 1.3. Let $f$ be a smooth real valued function on a manifold $M$. Define

$$M^a = \{ x \in M : f(x) \leq a \}$$

2 Morse Theory

Now let's turn to the description of Morse Theory, which consists of several lemmas and theorems and here will not give all the proofs of them.

2.1 Some Lemmas

At first, I shall illustrate some lemmas needed in the proof of main theorem of Morse Theory.

Lemma 2.1. Let $f$ be a $C^\infty$ function in a convex neighborhood $V$ of $0$ in $\mathbb{R}^n$, with $f(0) = 0$. Then

$$f(x_1, \cdots, x_n) = \sum_{i=1}^{n} x_i g_i(x_1, \cdots, x_n)$$

for some suitable $C^\infty$ functions $g_i$ defined in $V$, with $g_i(0) = \frac{\partial f}{\partial x_i}(0)$.

Therefore we can let $g_i(x_1, \cdots, x_n) = \int_0^1 \frac{\partial f}{\partial x_i}(tx_1, \cdots, tx_n) dx_i dt$

Lemma 2.2 (Lemma of Morse). Let $p$ be a non-degenerate critical point for $f$. Then there is a local coordinate system $(y^1, \cdots, y^n)$ in a neighborhood $U$ of $p$ with $y^i(p) = 0$ for all $i$ and such that the identity

$$f = f(p) - (y^1)^2 - \cdots - (y^\lambda)^2 + (y^{\lambda+1})^2 + \cdots + (y^n)^2$$

holds throughout $U$, where $\lambda$ is the index of $f$ at $p$.

Lemma 2.3. A smooth vector field on $M$ which vanishes outside of a compact set $K \subset M$ generates a unique 1-paramater group of diffeomorphism.

Remark. Here are all lemmas using in the proof of Morse Theory. First two lemmas show some local properties of a special point in manifold, which give rise to a way of attaching cells. And the third lemma gives a way to formula homotopy map.
2.2 Main Theorem of Morse Theory

Now let’s turn to main part of Morse Theory. Here we ignore the proof of Theorem 2.5 since it is too complex. And the proof can be found in [1].

**Theorem 2.4.** Let $f$ be a smooth real valued function on a manifold $M$. Let $a < b$ and suppose that the set $f^{-1}[a, b]$, consisting of all $p \in M$ with $a \leq f(p) \leq b$, is compact, and contains no critical points of $f$. Then $M^a$ is diffeomorphic to $M^b$. Furthermore, $M^a$ is a deformation retract of $M^b$, so that the inclusion map $M^a \rightarrow M^b$ is a homotopy equivalence.

**Theorem 2.5.** Let $f : M \rightarrow \mathbb{R}$ be a smooth function, and let $p$ be a non-degenerate critical point with index $\lambda$. Setting $f(p) = c$, suppose that $f^{-1}[c - \epsilon, c + \epsilon]$ is compact, and contains no critical point of $f$ other than $p$, for some $\epsilon > 0$. Then, for all sufficiently small $\epsilon$, the set $M^{c+\epsilon}$ has the homotopy type of $M^{c-\epsilon}$ with a $\lambda$–cell attached.

**Lemma 2.6.** As for a compact manifold $M$, there exists Morse function $f : M \rightarrow \mathbb{R}$ which maps different critical points to different values.

> Just consider suitable “bump functions” to modify a Morse-function near a critical point. Since the compactness of manifold, there are only finite many critial points for each critical values. Thus we can modify the Morse-function in the local position of each critical points.

From this lemma, we can obtain following **Theorem 2.7** by combining **Theorem 2.4** and **Theorem 2.5**.

**Theorem 2.7.** If $f$ is a differentiable function on a manifold $M$ with no degenerate critical points, and if each $M^a$ is compact, then $M$ has the homotopy type of a CW-complex, with one cell of dimension $\lambda$ for each critical point of index $\lambda$.

2.3 Morse inequality

In this subsection, the original treatment of this subject by Morse is introduced.

**Definition 2.8.** Let $S$ be a function from the space $\Omega = \{(X, Y) | X, Y$ is topological space $\}$ to the integers. $S$ is subadditive if whenever $X \supset Y \supset Z$ we have $S(X, Z) \leq S(X, Y) + S(Y, Z)$. If equality holds, $S$ is called additive.

**Lemma 2.9.** Let $S$ be subadditive and let $X_0 \subset \cdots \subset X_n$. Then $S(X_n, X_0) \leq \sum_{i=1}^{n} S(X_i, X_{i-1})$. If $S$ is additive then equality holds.

\[
S(X_n, X_0) \leq S(X_n, X_{n-1}) + S(X_{n-1}, X_0) \\
\leq S(X_n, X_{n-1}) + S(X_{n-1}, X_{n-2}) + S(X_{n-2}, X_0) \\
\leq \cdots \\
\leq \sum_{i=1}^{n} S(X_i, X_{i-1})
\]
Lemma 2.10. Let $R_\lambda(X,Y)$ denotes $\lambda$ th Betti number of $(X,Y)$, which equal to rank of $H_\lambda(X,Y)$. Then the function $S_\lambda$ is subadditive, where

$$S_\lambda(X,Y) = R_\lambda(X,Y) - R_{\lambda-1}(X,Y) + \cdots + (-1)^\lambda R_0(X,Y)$$

Consider long exact sequence

\[ \cdots \xrightarrow{\partial_{n+1}} A_n \xrightarrow{i_n} B_n \xrightarrow{j_n} C_n \xrightarrow{} \cdots \]

\[ \xrightarrow{\partial_n} A_{n-1} \xrightarrow{i_{n-1}} B_{n-1} \xrightarrow{j_{n-1}} C_{n-1} \]

\[ \cdots \xrightarrow{\partial_1} A_0 \xrightarrow{i_0} B_0 \xrightarrow{j_0} C_0 \xrightarrow{} 0 \]

We can get another long exact sequence:

\[ \cdots \rightarrow \ker(i_n) \rightarrow A_n \rightarrow \text{Im}(i_n) \rightarrow B_n \]

\[ \rightarrow \text{Im}(j_n) \rightarrow C_n \rightarrow \text{Im}(\partial_n) \rightarrow A_{n-1} \]

\[ \rightarrow \text{Im}(i_{n-1}) \rightarrow B_n \rightarrow \text{Im}(j_{n-1}) \rightarrow \cdots \]

and a short exact sequence:

\[ 0 \rightarrow \ker(i_n) \rightarrow A_n \rightarrow \text{Im}(i_n) \rightarrow 0 \]

Then we obtain

\[ \frac{A_n}{\ker(i_n)} \cong \text{Im}(i_n) \]

Thus $\text{rank}(A_n) = \text{rank}(\ker(i_n)) + \text{rank}(\text{Im}(i_n))$. The same equalities hold for $B_n$ and $C_n$.

Then we obtain that

\[ \text{rank}(\partial_{n+1}) = \text{rank}(A_n) - \text{rank}(\text{Im}(i_n)) \]

\[ = \text{rank}(A_n) - \text{rank}(B_n) + \text{rank}(\text{Im}(j_n)) \]

\[ = \cdots \]

\[ = \text{rank}(A_n) - \text{rank}(B_n) + \text{rank}(c_n) - \text{rank}(A_n) + \cdots \]

\[ = S_n(A_n) - S_n(B_n) + S_n(C_n) \geq 0 \]

Let $A_n = H(Y,Z), B_n = H(X,Z), C_n = (X,Y)$ for $X,Y,Z \in \Omega$. Then lemma follows.

Theorem 2.11. If $C_\lambda$ denotes the number of critical points of index $\lambda$ on the compact manifold $M$ then

\[ R_\lambda(M) \leq C_\lambda, \text{ and} \]

\[ S_\lambda(M) \leq C_\lambda - C_{\lambda-1} + \cdots + (-1)^\lambda C_\lambda \]
As for a compact manifold $M$, $f$ is a differentiable function on $M$ with isolated, non-degenerate, critical points. Let $a_1 \cdots < a_k$ be such that $M^{a_i}$ contains exactly $i$ critical points, and $M^{a_k} = M$. Then

$$H_n(M^{a_i}, M^{a_{i-1}}) = H_n(M^{a_{i-1}} \cup e^{\lambda_i}, M^{a_{i-1}}) = H_n(e^{\lambda_i}, \partial e^{\lambda_i}) = \tilde{H}_n(S^{\lambda_i}) = \begin{cases} \mathbb{Z} & n = \lambda_i, \\ 0 & n \neq \lambda_i \end{cases}$$

Thus according to the sequence $\emptyset = M^{a_0} \subset M^{a_1} \subset \cdots \subset M^{a_k} = M$ and Lemma 2.9, we have

$$R_\lambda(M) = R_\lambda(M, \emptyset) \leq \sum_{i=1}^k R_\lambda(M^{a_i}, M^{a_{i-1}}) = C_\lambda$$

and

$$S_\lambda(M) \leq \sum_{i=1}^k S_\lambda(M^{a_i}, M^{a_{i-1}}) = \sum_{j=1}^\lambda \sum_{i=1}^k (-1)^{\lambda-j} R_j(M^{a_i}, M^{a_{i-1}}) = \sum_{j=1}^\lambda \sum_{i=1}^k (-1)^{\lambda-j} C_j = C_\lambda - C_{\lambda-1} + \cdots + (-1)^\lambda C_\lambda$$

**Remark.** From Morse Inequality, we have

$$R_\lambda(M) - R_{\lambda-1}(M) + \cdots + (-1)^\lambda R_0(M) \leq C_\lambda - C_{\lambda-1} + \cdots + (-1)^\lambda C_\lambda$$

$$R_{\lambda+1}(M) - R_{\lambda}(M) + \cdots + (-1)^{\lambda+1} R_0(M) \leq C_{\lambda+1} - C_\lambda + \cdots + (-1)^{\lambda+1} C_\lambda$$

Thus if $C_{\lambda+1} = 0$, then $R_{\lambda+1} = 0$ and **Morse Equality**

$$R_\lambda(M) - R_{\lambda-1}(M) + \cdots + (-1)^\lambda R_0(M) = C_\lambda - C_{\lambda-1} + \cdots + (-1)^\lambda C_\lambda$$

57
3 Examples and Applications

3.1 Reeb sphere theorem

**Theorem 3.1.** If $M$ is a compact manifold and $f$ is a differentiable function on $M$ with only two critical points, both of which are non-degenerate, then $M$ is homeomorphic to a sphere.

Since $M$ is compact, $f$ must attain its minimum and maximum. Let $f(p) = a$ and $f(q) = b$, where $a$ is its minimum and $b$ is its maximum. Since $M$ is boundaryless, these points must be critical points of $f$. Besides, by the hypothesis, $p$ and $q$ are the only two critical points and thus we have $f^{-1}(a) = p$ and $f^{-1}(q)$.

Now we choose a value $\epsilon$ satisfies $a < a + \epsilon < b - \epsilon < b$. By **Theorem 2.4**, we see that $M^{a+\epsilon}$ is diffeomorphic to $M^{b-\epsilon}$. Besides, From **Lemma of Morse** we see that when $\epsilon$ is small enough, $M^{a+\epsilon}$ and $f^{-1}(b - \epsilon, b]$ are homeomorphic to two closed $n-$cells.

Hence, we obtain some homeomorphisms. If we note $D^n_+$ as upper semi-sphere and $D^n_-$ as lower semi-sphere then from previous paragraph, there are three homeomorphisms as follows.

\[
\begin{align*}
  f_1 : M^{a+\epsilon} &\rightarrow D^n_-
  
  f_2 : M^{a+\epsilon} &\rightarrow M^{b-\epsilon}
  
  f_3 : f^{-1}(b - \epsilon, b] &\rightarrow D^n_+
\end{align*}
\]

Finally, one can combine these functions and obtain a homeomorphism $h$ from $M$ to a sphere.

\[ h = (f_2^{-1} \circ f_1) \amalg f_3 : M \rightarrow S^n \]

\[ \square \]

**Remark.** It’s very interesting to find that this special function can uniquely determine the shape of a $n-$manifold and can only in homeomorphism sence. Since their differentiable structure can be totally different such as a 7-sphere.\[3\] However, we can see in the case of $n = 2$, it is really a diffeomorphism.\[2\]

When it comes to the case that there are only three critical points, it will have a homotopy type of $\frac{n}{2}$-sphere with an $n-$cell attached, or a Eells-Kuiper manifold.\[4\]

3.2 An application to complex projective plane

I still remember when I deal with the complex projective plane for the last homework, I find it really difficult to figure out its CW decomposition. However, will usage of Morse Theory, the CW decomposition becomes more trival than before. Let’s see how Morse Theory is used here.

**Theorem 3.2.** $\mathbb{C}P^n$ has the homotopy type of a CW-complex of the form

\[ e^0 \cup e^2 \cup \cdots \cup e^{2n} \]
Consider $\mathbb{C}P^n$ as equivalence classes of $(n+1)$-tuples $(z_0, \cdots, z_n)$ of complex numbers, with $\sum \lvert z_j \rvert^2 = 1$. Denote the equivalence class of $(z_0, z_1, \cdots, z_n)$ by $(z_0; z_1; \cdots; z_n)$. Define $f : \mathbb{C}P^n \to \mathbb{R}$ as follows:

$$f(z_0, z_1, \cdots, z_n) = \sum_{j=1}^{n} c_j |z_j|^2$$

where $c_0, \cdots, c_n$ are distinct real constants. It’s easy to see this definition is well-defined since different representation of $(z_0, z_1, \cdots, z_n)$ only differ from a multiplication of a complex number whose norm equal to 1.

Choose local coordinate system $(\psi_k, U_k)$ where $U_k = (z_0; z_1; \cdots; z_n)$ with $0 \neq z_k \in \mathbb{R}$ and set $\lvert z_k \rvert \frac{z_i}{z_k} = x_{k,j} + y_{k,j}$. $(i = 0, \cdots, n)$.

Let $\psi_k : U_k \to \mathbb{R}^{2n}$

$$(z_0; z_1; \cdots; z_n) \to (x_{k,0}, y_{k,0}, \cdots, x_{k,n}, y_{k,n})$$

with $x_{k,0}^2 + y_{k,0}^2 + \cdots + x_{k,n}^2 + y_{k,n}^2 = 1$ and $x_{k,k}^2 + y_{k,k}^2 \neq 0$. Then function $f$ can be represented as follows:

$$f(z_0; \cdots; z_n) = \sum_{j=0}^{n} c_j (x_{k,j}^2 + y_{k,j}^2)$$

$$= c_k + \sum_{j \neq k} (c_j - c_k)(x_{k,j}^2 + y_{k,j}^2)$$

Thus we can see that $f'$ equal to 0 if and only if $x_{k,j} = y_{k,j} = 0$ for $j \neq k$ and $z_k = 1$. It leads to that the only critical point in $U_k$ is $P_k = (0, \cdots, 0, 1, 0, \cdots, 0)$ in which 1 is in the $k$ position.

Besides, Since $f(z_0; \cdots; z_n) = c_k + \sum_{j \neq k} (c_j - c_k)(x_{k,j}^2 + y_{k,j}^2)$, point $P_k$ must have index $\lambda_k$ = twice of the number of $j$ with $c_j < c_k$. Thus $\mathbb{C}P^n$ has CW-decomposition

$$e^0 \cup e^2 \cup \cdots \cup e^{2n}$$

Then one can easy calculate its homology group.

\[\square\]

### 3.3 Poincaré-Hopf theorem

**Definition 3.3** (index of vector field). Let $M$ be a differentiable manifold of dimension $n$, and $v$ is a vector field on $M$. Suppose that $p$ is an isolated zero of $v$, and fix some local coordinates near $p$. Pick a closed ball $D$ centered at $p$, so that $p$ is the only zero of $v$ in $D$. Then we define the index of $v$ at $p$, $\text{index}_p(v)$, to be the degree of map $u : \partial D \to S^{n-1}$ from the boundary of $D$ to the $(n-1)$-sphere given by $u(z) = \frac{v(z)}{\lvert v(z) \rvert}$. Then we define the index of vector field $\text{ind}(v)$

$$\text{ind}(v) = \sum_p \text{index}_p(v)$$

where $\sum$ takes over all critical points $p$ in $M$. 

59
Definition 3.4 (Transversality). Let \( f : M \to N \) be a \( C^1 \) map and \( A \subset N \) a submanifold. If \( K \subset M \) we say \( f \) is transverse to \( A \) along \( K \), that is, whenever \( x \in K \) and \( f(x) = y \in A \), the tangent space \( N_y \) is spanned by \( A_y \) and the image \( T_x f(M_x) \).

Definition 3.5 (intersection number). Let \( W \) be an oriented manifold of dimension \( m + n \) and \( N \subset W \) a closed oriented submanifold of dimension \( n \). Let \( M \) be a compact oriented \( m \)-manifold. Suppose \( \partial M = \partial N = \emptyset \).

Let \( f : M \to W \) be a \( C^\infty \) map transverse to \( N \) along \( f^{-1}(N) \). A point \( x \in f^{-1}(N) \) has positive or negative type according as the composite linear isomorphism

\[
M_x \xrightarrow{T_f} W_y \to W_y/N_y, \quad y = f(x)
\]

preserves or reverses orientation; we write \( \#_x(f, N) = 1 \) or \(-1\), respectively. The intersection number of \((f, N)\) is the integer

\[
\#(f, N) = \sum \#_x(f, N),
\]

summed over all \( x \in f^{-1}(N) \).

Proposition 3.6. If \( f, g : M \to W \) are homotopic \( C^\infty \) maps transverse to \( N \) then \( \#(f, N) = \#(g, N) \).

Remark. According to this proposition, we can define any map \( g : M \to W \) its \( \#(g, N) = \#(f, N) \), where \( f \) is a \( C^\infty \) map transverse to \( N \) and homotopic to \( g \). Hence we can define self-intersection number \( \#(N; W) = \#(f, N) \). if \( M = N \) and \( f \) is a inclusion map.

In order to give a proof of Poincaré-Hopf theorem, we need a theorem without giving proof.

Theorem 3.7. Let \( \sigma \) be a smooth vector field with finite many zeros on a closed manifold \( X \). Then the sum of all the indices of \( \sigma \) at these points is equal to the self-intersection number \( \#(\Delta_X; X \times X) \), where \( \Delta_X \) is the diagonal of the space \( X \times X \), i.e,

\[
\text{Ind}(\sigma) = \#(\Delta_X; X \times X)
\]

Theorem 3.8 (Poincaré-Hopf theorem). The Euler characteristic of a closed manifold \( M \) is equal to the index of any vector field on \( M \) with finitely many zeros.

At first, suppose \( f \) is a morse function on \( M \) and \( p \) is a critical point of \( M \) with index \( \lambda \). Consider vector field \( \text{grad}(f) \) on \( M \). Then \( p \) is also the point where vector field vanishes. From Morse Lemma we see that in a neighborhood of \( p \), \( f \) can be given:

\[
f = f(p) - \sum_{i=1}^{\lambda} x_i^2 + \sum_{i=\lambda+1}^{n} x_i^2
\]

60
Hence the vector field around $p$ in suitable coordinate system is

$$-2x_1, \cdots, -2x_\lambda, 2x_{\lambda+1}, \cdots, 2x_n$$

Thus index of $p$ is simply $(-1)^\lambda$.

Now we consider all the critical points in $M$. The index of vector field $\nabla f$ can be given by $\sum_{i=0}^{n} (-1)^i C_i$, where $C_i$ is the number of critical points whose index is $i$.

Thus by Morse Equality, we have $\chi(M) = \text{ind}(\nabla f)$. Hence $\chi(M)$ equals to the self-intersection number of the manifold $M$. By Theorem 3.7, we obtain Poincaré-Hopf theorem.

**Remark.** By Poincaré-Hopf theorem, we can easily prove Hairy Ball Theorem, which states there is no nowhere vanished vector fields on 2-dimension sphere.

### 4 Conclusion

Here is another interesting application of Morse Theory, Classification of Compact Manifold[5], which classifies all compact manifold in the sense of diffeomorphism. However, it is two long and I can’t fully understand it by now. But I still want to write it down here for its beauty.

After some reading, I think Morse Theory can really help us make how manifold forms steps by steps clear and give some deep understanding of differential topology.

### References


1 Introduction

To begin with, recall the famous claim by Fermat:

**Conjecture 1.1.** If \( n \geq 3 \), then \( x^n + y^n = z^n \) has no integer solution with \( xyz \neq 0 \).

Fermat proved this conjecture in the case \( n = 4 \) by the method of infinite descent. Therefore it’s sufficient to deal with the case \( n = p \) is a odd prime (Without otherwise mention, we always assume \( p \) is odd from now on). As \( \mathbb{Z}[\zeta_p] \) is a Dedekind domain and \( z^p = x^p + y^p = \prod_{\zeta_p} (x + \zeta_p y) \), Kummer noticed that if \( p \) does not divide the class number of \( \mathbb{Q}(\zeta_p) \) then there is a contradiction under the condition \( p \nmid xyz \). This lead to Kummer’s study of such primes which are called regular, and Kummer proved:

**Theorem 1.2.** Let \( p \) be a odd prime, then \( p \) is regular if and only if \( p \) does not divide any of the numerators of \( B_i \), \( \forall i = 2, 4, \ldots, p - 3 \). (\( B_n \) are Bernoulli numbers such that \( \sum_{i=0}^{\infty} B_n \frac{x^n}{n!} \).

**Example 1.** \( B_{12} = \frac{-691}{2730} \), so \( p = 691 \) is not regular.

Bernoulli numbers have many applications such as computing the sum of a fixed power of natural numbers, and they are connected with Riemann zeta function by \( \zeta(-n) = (-1)^n \frac{B_{n+1}}{n+1} \), \( \forall n \geq 1 \). Therefore it’s natural that they have deep arithmetical properties, and Kummer proved:

**Theorem 1.3.** If \( n \equiv m \not\equiv -1 \pmod{p-1} \), then \( B_{n+1} \equiv B_{m+1} \pmod{p} \)

Let \( C = Cl_{Q(\zeta_p)}[p^\infty] \) be the \( p \)-part of idea class group with the action of \( G = \text{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q}) \), then \( C = \bigoplus_{n=1}^{p-1} C^n \) where \( G \) acts on \( C^n \) by \(-n\)-th power of the Teichmuller character \( w : G \to \mathbb{Z}_p^\times \). Above two theorems have more precise generalizations:

**Theorem 1.4.** (Herbrand, Ribet) If \( n \) is odd, then \( C^n \neq 0 \Leftrightarrow p | \zeta(-n) \)
Theorem 1.5. (Kubota, Leopoldt) For every $i$, there exist a unique continuous function $L_p(s, \omega^i) : \mathbb{Z}_p \to \mathbb{Q}_p$ such that

$$L_p(-n, \omega^i) = \zeta(-n)(1 - p^n), \forall n \equiv i - 1 \pmod{p - 1}$$

In conclusion, we see there is a deep connection between $p$-part of ideal class group and $p$-adic L function which is the main part of Iwasawa theory, and all theorems above are easy consequences of the main conjecture in Iwasawa theory.

The idea behind Iwasawa theory is to study algebraic objects such as ideal class groups by placing them in a $p$-adic tower. For instance, given $F$ there is a $\mathbb{Z}_p$-extension $F_\infty$ i.e $\Gamma := Gal(F_\infty/F) \cong \mathbb{Z}_p$. The Iwasawa module $X$ is defined as inverse limit of the $p$-parts of the ideal class groups $Cl_{F_n}[p^n]$ of the intermediate fields $F_n$ in the $\mathbb{Z}_p$-extension $F_\infty$. Note that $X$ is a module over the Iwasawa algebra $\Lambda := \mathbb{Z}_p[[\Gamma]] \cong \mathbb{Z}_p[[T]]$ which is the ring of power series over $\mathbb{Z}_p$. The Iwasawa algebra $\Lambda$ is a 2-dimensional regular local ring. Thus the structure theory of finitely generated modules over $\Lambda$ is good and we have the classification up to pseudo-isomorphism. More precisely we have:

**Theorem 1.6.** If $M$ is a finitely generated $\Lambda$-module, then there exists a homomorphism with finite kernel and cokernel (i.e a quasi-isomorphism):

$$M \to \Lambda^r \oplus \bigoplus_{i=1}^s \Lambda/(p^{m_i}) \oplus \bigoplus_{j=1}^t \Lambda/(f_j(T)^{n_j})$$

where $f_j \equiv x^{deg f_j} \pmod{p}$ is irreducible.

Through such method, Iwasawa proved:

**Theorem 1.7.** Let $F$ be a number field with a $\mathbb{Z}_p$-extension $F_\infty$, then there exist integers $\lambda, \mu, \nu$ (called Iwasawa invariants) such that $\# Cl_{F_n}[p^n] = p^{\lambda n + \mu p^{\nu} + \nu}, \forall n \gg 0$.

**Proof.** This is just a sketch of the proof. The main point is that Fitting ideals are invariant under pseudo-isomorphism so we can still define invariants. One can show $X = \lim Cl_{F_n}[p^n]$ is a f.g torsion $\Lambda$-module using class field theory as $Gal(L_n/F_n) \cong Cl_{F_n}[p^n]$ where $L_n$ is the $p$-Hilbert class field of $F_n$, so $X = Gal(L_\infty/F_\infty)$ with action of $\Lambda$ by conjugation. Then we can define $char(X) = \prod p^{m_i} \prod f_j^{n_j}, \lambda = deg(char(X)), \mu = ord_p(char(X))$ and finish the proof by computation. For instance, we know $X/((T + 1)^p - 1) \cong Cl_{F_1}[p^n]$ if there is only one prime in $O_F$ over $p$ which is totally ramified in $F_\infty/F$. \qed

Determining such invariants are difficult, but there are some partial results:

**Theorem 1.8.** If $p$ does not divide the class number of $F$ and there is only one prime in $O_F$ over $p$, then $p$ does not divide the class number of $F_n, \forall n = 1, 2, \ldots$.

**Theorem 1.9.** Assume $p$ splits completely in $F$ and $F_\infty/F$ is the cyclotomic $\mathbb{Z}_p$-extension, then $\lambda \geq r_2(F) = \# \text{complex places of } F$. 63
We also have Iwasawa’s conjecture, which is only proved in the case $F/\mathbb{Q}$ is abelian and is called Ferrero-Washington theorem:

**Conjecture 1.10.** (Iwasawa) If $F_\infty/F$ is the cyclotomic $\mathbb{Z}_p$-extension, then $\mu = 0$.

Now we come to the main conjecture of Iwasawa theory. Assume $F = \mathbb{Q}(\zeta_p)$, then $G = Gal(\mathbb{Q}(\zeta_p)/\mathbb{Q})$ acts on $X$ so we have $X = \oplus_{n=1}^{p-1} X^n$. We can find a $p$-adic L function $\hat{L}_p(T, w^{n+1}) \in \Lambda$ such that $\hat{L}_p((1 + p)^s - 1, w^{n+1}) = L_p(s, w^{n+1}), \forall s \in \mathbb{Z}_p$. Now we can state the main conjecture which is proved by Mazur and Wiles in 1984:

**Theorem 1.11.** If $n$ is odd and $n \not\equiv -1(\mod p - 1)$, then $(\text{char}(X^n)) = (\hat{L}_p(T, w^{n+1}))$ as ideals in $\Lambda$.

The proof relies on the arithmetic of modular form, and another later proof uses Euler system of cyclotomic units.

Besides, one can consider Iwasawa theory for elliptic curves, where the main conjecture shows a connection between Selmer group and special values of $p$-adic L function. Using modular symbols, one can get information of the special values hence the information of the arithmetic of elliptic curves. For example, Coates and Wiles showed that $r_{an} = 0 \Rightarrow r_{alg} = 0$ for elliptic curves over $\mathbb{Q}$ with complex multiplication, which is one of the early results on BSD conjecture.

## 2 Classical Iwasawa theory

We provide some details in the above argument before further discussion. Let $A$ be a noetherian integrally closed domain, $h_1(A)$ be the set of height 1 prime ideals of $A$ and $K = \text{Frac}(A)$.

**Definition 2.1.** A finitely generated $A$-module $M$ is called pseudo-null if $M_P = 0, \forall P \in h_1(A)$; A morphism between $A$-modules is called a pseudo-isomorphism if its kernel and cokernel are pseudo-null (We use $\sim$ to denote a pseudo-isomorphism).

**Proposition 2.2.** Let $M$ be a finitely generated $A$-module and $T(M)$ be its torsion part. Then we have

1. $M \sim T(M) \oplus M/T(M)$.
2. There exists unique $P_i \subseteq h_1(A), n_i \in \mathbb{N}$ such that $T(M) \sim \bigoplus_i A/P_i^{n_i}$.

**Proof.** (1) If $\text{Supp}(M) \cap h_1(A) = \emptyset$ then $M$ is pseudo null and there is nothing to prove. Otherwise, we localize $A$ by $S = A - \bigcup_{P \in \text{Supp}(M) \cap h_1(A)} P$ and get a PID (as $S^{-1}A$ is a Dedekind domain with finitely many prime ideals). So $S^{-1}T(M)$ is a direct summand of $S^{-1}M$ hence there exists $f_0 : M \to T(M)$ and $s_0 \in S$ such that $\frac{f_0}{s_0}$ is the projection from $S^{-1}T(M)$
to $S^{-1}M$ (i.e $f_0|_{S^{-1}TM} = id$). Thus there exists $s_1 \in S, f_1 = s_1 f_0$ such that $f_1|_{TM} = s_1 s_0 id$. By check the stalk at $P \in \text{Supp}(M) \cap h_1(A)$, one see $f = (f_1, pr) M \to T(M) \oplus M/T(M)$ is a pseudo-isomorphism.

(2) localizing $A$ by $S$ and using structure theorem of finitely generated modules over a PID, the proof is similar to (1).

The torsion-free parts are more subtle. Write $M^+ = \text{Hom}_A(M, A)$ for any finitely generated torsion-free $A$-module $M$, then there is a natural map $M \to M^{++}$ which is an injective pseudo-isomorphism (as it’s an isomorphism if $A$ is a DVR) and $M^{++}$ is reflexive ($N$ is called reflexive if the natural map $N \to N^{++}$ is an isomorphism). So WLOG we can assume $M$ is reflexive. Notice that

**Proposition 2.3.** If $A$ is a 2-dimensional regular local ring, then a finitely generated reflexive module $M$ must be free.

**Proof.** $M$ is reflexive, hence torsion free. Choose a regular system of parameters $p_1, p_2$ for $A$ so $A/p_1$ is regular of dimension 1 hence a DVR. We have $M/p_1 = M^{++}/p_1 = \text{Hom}_A(M^+, A) \otimes A/p_1 \to \text{Hom}_A(M^+, A/p_1)$ and $\text{Hom}_A(M^+, A/p_1)$ is torsion free as $A/p_1$-module, so $M/p_1$ is free over $A/p_1$. Now choose a minimal resolution $\phi : A^r \to M$ (i.e a lifting of basis of $M/mAM$), then $\phi$ is surjective and $p_1 \ker\phi = \ker\phi$ as $\phi \mod p_1$ is an isomorphism. By Nakayama, we know $\ker\phi = 0$ so $M$ is free.

**Theorem 2.4.** (Structure Theorem) Let $A$ be a 2-dimensional regular local ring and $M$ be a finitely generated $A$-module, then there exists unique $P_i \subseteq h_1(A), n_i \in \mathbb{N}, r \in \mathbb{N}$ such that

$$M \cong A^r \oplus \bigoplus_i A/P_i^{n_i}$$

where $r = \dim_K M \otimes_A K, \{P_i\} = \text{Supp} M \cap h_1(A)$.

Let $\Lambda \cong \mathbb{Z}_p[[T]]$ be the Iwasawa algebra and its modules are called Iwasawa modules.

**Theorem 2.5.** $\Lambda$ is a 2-dimensional regular local ring with maximal ideal $m = (p, T)$, and $h_1(\Lambda) = \{(p)\} \cup \{(f)\} | f \in \mathbb{Z}_p[T] \text{ irreducible}, f \equiv T^{\deg f} \mod p$.

**Proof.** $A$ is 2-dimensional local ring and $\dim m/m^2 = 2$ so it’s regular hence a UFD. One can easily prove the division lemma for $O[[T]]$ where $O$ is a complete DVR and deduce the Weierstrass Preparation Theorem, so every principle ideal is contained in $(p)$ or some $(f)$ in the theorem. Finally, every height 1 prime ideal is principle in a UFD.

**Remark 2.6.** By Weierstrass Preparation Theorem, we know that for any $p \not| g \in \mathbb{Z}_p[[T]]$, $\mathbb{Z}_p[[T]]/(g) \cong \mathbb{Z}_p[[T]]/(f)$ for some $f \in \mathbb{Z}_p[T]$ such that $f \equiv T^{\deg f} \mod p$. In fact $f$ is the characteristic polynomial of the linear transformation $T : \mathbb{Z}_p[[T]]/(g) \to \mathbb{Z}_p[[T]]/(g)$. 

65
Remark 2.7. We should define \( \Lambda := Z_p[\Gamma] = \lim Z_p[[Z/pnZ]] = \lim Z_p[[T]]/((1 + T)^n - 1) \), where \( \Gamma \cong Z_p \) with a topological generator \( \gamma \) and the morphism \( \hat{Z}_p[[Z/p^nZ]] \cong Z_p[[T]]/((1 + T)^n - 1) \) is given by \( \gamma \mapsto 1 + T \). But \( Z_p[[T]] \cong Z_p[[\Gamma]]T \mapsto \gamma - 1 \) so there is no problem if a topological generator is fixed. Write \( w_n(T) = (1 + T)^n - 1 \), it’s important to keep in mind we should compute \( \# X/w_nX \) rather than \( \# X/T^nX \) after such translation.

As \( V(\text{Ann}M) = \text{Supp}M \) we find that a \( \Lambda \)-module is pseudo-null if and only if it’s finite as a set, so above argument finishes the proof of Theorem 1.6 (The uniqueness requires the theory of Fitting ideals and it’s omitted). As finite set will only contribute a constant to \( \# M/w_nM, n \gg 0 \), while \( \Lambda/p^k \) contributes \( p^{kn} \) to \( \# M/w_nM, n \gg 0 \). Besides we have:

Lemma 2.8. If \( V = \Lambda/(f) \) where \( f \in Z_p[T], f \equiv T^{\text{deg}f} \mod p \) and assume that \( \# V/w_nV \) are finite for every \( n \), then \( \# V/w_nV = p^{\text{deg}f} n + c, \forall n \gg 0 \) where \( c \) is a constant.

Proof. We know \( T^{\text{deg}f} = pQ(T) \mod f \) for some \( Q \in Z_p[T] \), so \( (1 + T)^n - 1 = 1 + p \cdot Q_n \mod f \cdot p^n \geq \text{deg} f \), where \( Q_n \in Z_p[T] \). Thus \( (1 + T)^n - 1 = (1 + pQ_n)^n \) is \( 1 + p^2 \cdot \text{poly} \mod f \), and we find that \( n^{-1} + (1 + T)^{n+1} + \ldots + (1 + T)^{n+1(p-1)} \) acts on \( V \) by \( p \cdot \text{unit} \). Therefore \( w_{n+2}V = pw_{n+1}V \).

\[ \# V/w_{n+2}V = \# V/pV \# V/pw_nV = p^{\text{deg}f} \# V/w_{n+1}V, \forall n \gg 0 \]

so the result follows. \( \square \)

In short, we have shown that

**Theorem 2.9.** Let \( M \) be a finitely generated torsion \( \Lambda \)-module such that \( M/w_nM \) is finite for all \( n \). Then there exist integers \( \lambda \geq 0, \mu \geq 0 \) and \( \nu \) such that \( \# M/w_nM = p^{\lambda n + p^\lambda + \nu}, \forall n \gg 0 \)

To finish the proof of Theorem 1.7, we just need to show \( X = \text{Gal}(L_\infty/F_\infty) \) is a finitely generated torsion module. Using Nakayama’s lemma and that \( X \) is a compact \( \Lambda \)-module, one can show \( X \) is finitely generated under the assumption there is only one prime in \( O_F \) over \( p \) which is totally ramified in \( F_\infty/F \) (which can be removed by passing from \( F \) to \( F_n \) for some \( n \)), and the torsion part follows from Nakayama’s lemma and finiteness of \( X/w_nX \).

To be more precise, we should include basics of \( Z_p \)-extensions:

**Proposition 2.10.** Let \( F \) be a number field with a \( Z_p \)-extension \( F_\infty/F \), then \( F_\infty/F \) is unramified outside \( p \), and every ramified prime ideals (they exist and must be over \( p \)) will be totally ramified in some \( F_\infty/F_n \).

Proof. By class field theory, there is a surjective map from \( U_F \) to the inertia group \( I_v \) for every place \( v \) in \( F \). If \( p \nmid v \) then this map must be zero (maps from pro-\( l \) group to pro-\( p \) group are zero), hence \( F_\infty/F \) is unramified outside \( p \). Besides, \( F_\infty/F \) must be ramified at some place as the ideal class group is finite. As there are only finite many ramified prime ideals, we can choose \( n \gg 0 \) such that \( \text{Gal}(F_\infty/F_n) \) is contained in intersection of those inertia groups. \( \square \)
Note that every number field $F$ has at least one $\mathbb{Z}_p$-extension: $\mathbb{Q}$ has only one $\mathbb{Z}_p$-extension $\mathbb{Q}_\infty/\mathbb{Q}$ by Kronecker-Weber theorem, thus $F\mathbb{Q}_\infty/F$ is a $\mathbb{Z}_p$-extension for $F$ which is called the cyclotomic $\mathbb{Z}_p$-extension of $F$. More precisely, Let $E = O_F^\times, E_1 = \{x \in E | x \equiv 1 \mod P \text{ for } P \mid \mathbb{P} \text{ in } F\}$, then $E_1$ (global) can be diagonally embedded in $U_1 = \prod_{P \mid \mathbb{P}} \mathbb{U}_1P$ where $U_1, P = \{x \in F_P | x \equiv 1 + P\mathbb{O}_P \}$ (local units congruent to 1 mod $P$). Let $E_1$ be the closure of $E_1$ in $U_1$, then $\delta_p(F) := \text{rank}Z_1E_1 - \text{rank}Z_2E_1 \geq 0$. By class field theory one can prove that:

**Proposition 2.11.** Let $F_1$ be the composite of all $\mathbb{Z}_p$-extension of $F$ which lies in the maximal unramified outside $p$ abelian extension of $F$, then we have $\text{Gal}(F_1/F) \cong \mathbb{Z}^{1+r_2(F)+\delta_p(F)}_{p}$ where $r_2(F) = \# \text{complex places of } F$. In particular, there are at least $1 + r_2(F)$ independent $\mathbb{Z}_p$-extensions over $F$.

So $\mathbb{Z}_p$-extensions are not rare. The famous Leopoldt conjecture concerns about $\delta_p(F)$:

**Conjecture 2.12.** $\delta_p(F) = 0$.

Iwasawa theory is also used in the study of Leopoldt conjecture (Iwasawa’s original paper about Iwasawa theory proved the weak Leopoldt conjecture i.e $\delta_p$ is bounded along cyclotomic $\mathbb{Z}_p$-extension), which is only proved in some special cases such as the case that $F$ is abelian over $\mathbb{Q}$.

Last but not least, let $F$ be any abelian number field with Galois group $G$, by Kronecker-Weber theorem there is an integer $m$ (conductor of $F$) such that $F \subseteq \mathbb{Q}(\zeta_m)$ and a natural map $a \mapsto \sigma_a : (\mathbb{Z}/m\mathbb{Z})^\times \rightarrow G$. The Stickelberger element of $F$ is defined as $\theta_F = -\frac{1}{m} \sum_{a=1}^{m} a \cdot \sigma_a^{-1} \in \mathbb{Q}[G]$, and the Stickelberger ideal is $I = \mathbb{Z}[G] \cap \theta_F \mathbb{Z}[G]$. We have:

**Theorem 2.13.** (Stickelberger) The Stickelberger ideal $I$ annihilates the class group of $F$.

The proof is based on Gauss sums. Now assume $F = \mathbb{Q}(\zeta_p)$ with the standard cyclotomic $\mathbb{Z}_p$-extension and $G = \text{Gal}(F_\infty/F) \cong \Delta \times \Gamma$ where $\Delta \cong (\mathbb{Z}/p\mathbb{Z})^\times$ with Teichmuller character $w : G \rightarrow \Delta \rightarrow \mathbb{Z}_p^\times$. For $i \in \mathbb{Z}/(p-1)\mathbb{Z}$, we can define $\epsilon_i = \frac{1}{p-1} \sum_{a=1}^{p-1} w^i(a) \sigma_a \in \mathbb{Z}_p[\Delta]$ to split the Stickelberger element $\theta_n = \theta_{F_n}$ into $\theta_n^i = \epsilon_i \cdot \theta_n$.

**Lemma 2.14.** If $i \neq 1, \text{then } \theta_n = (\theta_n^i) \in \mathbb{Z}_p[\text{Gal}(F_n/F)] = \Lambda$, and $\theta_i = 0$ for $i \neq 0$ even.

$\theta_i$ can be computed explicitly and is related to generalized Bernoulli number, and $p$-adic L function $L_p(s, w^i)$ can be defined using special values at $\theta_1 - j$. Let $X$ be the inverse limit of $p$-parts of the class groups of $F_n$ and $X_i = \epsilon_i X$ which are finitely generated $\Lambda$-modules, then the main conjecture states that the characteristic ideal $\text{char}(X_i)$ is generated by $\theta_i$ for all odd $3 \leq i \leq p - 2$. 

67
3 Iwasawa theory for elliptic curves

Besides the standard cyclotomic $\mathbb{Z}_p$-extension for any number field, for an imaginary quadratic field one can also introduce another $\mathbb{Z}_p$-extension called anti-cyclotomic $\mathbb{Z}_p$-extension. Using the theory of elliptic curves with complex multiplication, which reveals some connections between Iwasawa theory and elliptic curves. Besides, unit groups and elliptic curves are both one dimensional commutative algebraic groups over number rings and the Tate–Shafarevich group is similar to the ideal class group, so the classical Iwasawa theory which involves study of unit group may be generalized.

First of all, we should review basics of elliptic curves:

**Proposition 3.1.** Let $E$ be an elliptic curve over $\mathbb{Q}$ (i.e. a smooth projective curve over $\mathbb{Q}$ of genus one with a distinguished $\mathbb{Q}$-rational point), then

1. (Mordell-Weil) $E(\mathbb{Q})$ is a finitely generated abelian group.
2. $E$ has a global minimal model over $\mathbb{Q}$. Under such model, define $a_l = l + 1 - \#(E(\mathbb{F}_l))$ for every prime $l$, then the $L$ function $L(E/\mathbb{Q}, s) = \prod_l (1 - a_l l^{-s} + c(l) l^{1-2s})^{-1}$

has an analytic continuation to the entire complex plane where $c(l) = 1$ if $E$ has good reduction at $l$, and 0 otherwise.

Let $F$ be a number field, two important groups are involved in the study of elliptic curves. For $n > 1$ consider the exact sequence

$$0 \to E[n](\overline{\mathbb{Q}}) \to E(\overline{\mathbb{Q}}) \to E(F) \to 0$$

Take cohomology gives

$$0 \to E(F)/nE(F) \to H^1(F, E[n])$$

One problem is that $H^1(\mathbb{Q}, E[n])$ is too large, but we can get similar local exact sequence for every place of $F$ and define the Selmer group

$$Sel_n(E/F) = \{ c \in H^1(F, E[n]) | res_v c \in Im(E(F_v)/nE(F_v) \to H^1(F_v, E[n])) \text{ for all } v \}$$

and the Tate-Shafarevich group

$$\Sha(E/F) = \bigcap_v Ker(H^1(F, E) \to H^1(K, E_v))$$

then we have the exact sequence

$$0 \to E(F)/nE(F) \to Sel_n(E/F) \to \Sha(E/F)[n] \to 0$$

Note that Selmer group is finite and Tate-Shafarevich group is torsion. Let $r_{alg} := \text{rank}_Z E(\mathbb{Q})$ and $r_{an} := \text{ord}_{s=1} L(E/\mathbb{Q}, s)$ then we have the famous BSD conjecture:
Conjecture 3.2. (BSD conjecture) (1) $r_{\text{alg}} = r_{\text{an}}$.
(2) $\text{III}(E/Q)$ is finite.
(3) $$\frac{1}{r!} \frac{d^r}{ds^r} L(E/Q, s)|_{s=1} = \#\text{III}(E/Q) \cdot \frac{R_{E/Q}}{\#E(Q)_{\text{tor}}} \cdot \left( \prod_l c_l \right) \cdot \Omega_E^+$$
where $r = r_{\text{alg}} = r_{\text{an}}$, $R_{E/Q}$ is the volume of $E(Q)_{\text{free}}$ in $E(Q) \otimes \mathbb{R}$ under the canonical height, $c_l$ is the local Taganawa number of $E$ at prime $l$ and $\Omega_E^+ := \int_{E(C)^+} w_E$ is the real period.

While it is still open, controls for Selmer groups can give information for the rank of $E$ and the size of $\text{III}(E/Q)$ and Iwasawa theory for elliptic curves concerns about the growth of Selmer groups. Let $\mathcal{S}(E/F) := \lim_{\to} \text{Sel}_{p^n}(E/F)$, from above exact sequence we get:

Proposition 3.3. There is a short exact sequence:
$$0 \to E(F) \otimes_{\mathbb{Z}} \mathbb{Q}_p/\mathbb{Z}_p \to \mathcal{S}(E/F) \to \text{III}(E/F)[p^\infty] \to 0$$

It happens that the behaviour of the Pontryagin dual of this group in a $\mathbb{Z}_p$-extension of $F$ is well understood. Let $X = \lim_{\to} \text{Hom}(\mathcal{S}(E(F_n)), \mathbb{Q}_p/\mathbb{Z}_p)$ where the maps are induced by the natural inclusions $E(F_n) \subseteq E(F_{n+1})$. Then we have:

Theorem 3.4. (1) $X$ is a finitely generated $\Lambda$-module.
(2) (Kato, [Kat04]) If $E/Q$ has good reduction at $p$, $F/Q$ is abelian with the standard cyclotomic extension, then $X$ is torsion $\Lambda$-module.

Mazur conjectured that if $E/F$ has good reduction at all primes above $p$ and $F_{\infty}/F$ is the standard cyclotomic extension then $X$ is torsion, which is still open. If $X$ is torsion and assume the Tate-Shafarevich group of $E/F$ has finite $p$-part, we can get similiar growth formula for Selmer groups and Tate-Shafarevich groups.

References


Trace Class Operators

李林骏*

Abstract

The intension of this article is to provide an invitation to the nuclear operators material which only assumes its readers are acquainted with the undergraduate functional analysis. In this short essay, we first state some basic propositions on the nuclear operators by using elementary methods. Then we compare the p-class operators ideals ($I_p$) with the Lebesgue spaces $L_p$.

Let $\mathcal{H}$ be a separable Hilbert space, $B(\mathcal{H})$ is the operators on $\mathcal{H}$, $K$ is the compact operators ideal of $\mathcal{H}$. Recall that any self adjoint compact operator can be diagonalized with respect to its spectra, and its eigenvalues converge to 0. Recall the polar decomposition $T = U|T|$, where $|T| = (T^*T)^{1/2}$, $||T|| = ||T||$, and $U$ is partial iso between $\text{Ran}(|T|)$ and $\text{Ran}(T)$.

**Definition 1.** For an element $a \in K$, define $\text{Trace}(a) = \sum_i \langle ae_i, e_i \rangle$ for some orthonormal base (ONB) $\{e_i\}^\infty_1$. If $\text{Trace}(|a|) < \infty$, call $a$ is of Class 1, and define $|a|_1 = \text{Trace}(|a|)$.

**Lemma 1.**

$$|a|_1 = \inf_{\{e_i\}^\infty_1} \sum_i ||ae_i|| = \sum_{\lambda_i \in \text{sp}(|a|)} \lambda_i$$

*Proof.* Firstly, $\text{Trace}(a)$ is well defined, provided $|a|_1 < \infty$.

Let $\{\xi_i\}, \{e_i\}$ be two ONBs, correlated by $e_i = \sum_j c_{ij} \xi_j$. Then

$$\sum_i \langle ae_i, e_i \rangle = \sum_i \langle a \sum_j c_{ij} \xi_j, \sum_j c_{ij} \xi_j \rangle = \sum_{i,j,k} \langle ac_{ij} \xi_j, c_{ik} \xi_k \rangle$$

$$= \sum_{j,k} \sum_i c_{ij} c_{ik} \langle a \xi_j, \xi_k \rangle = \sum_j \langle a \xi_j, \xi_j \rangle$$

Secondly, $|a|$ is a compact operator, thus there is an ONB $\{\xi_i\}^\infty_1$, s.t. $|a| \xi_i = \lambda_i \xi_i$. Now,

$$\text{Trace}(|a|) = \sum_i \langle |a| \xi_i, \xi_i \rangle = \sum_i \lambda_i = \sum_{\lambda_i \in \text{sp}(|a|)} \lambda_i$$

*数 42
Finally, suppose \( \{\xi_i\} \) is picked above, and \( e_i = \sum_j c_{ij} \xi_j \). Then

\[
\sum_i ||ae_i|| = \sum_i \left( \sum_j |c_{ij} \lambda_j|^2 \right)^{1/2}
\]

Let

\[
\varphi(t) = \sum_i \left( \sum_j |c_{ij}|^2 |\lambda_j|^t \right)^{1/t}
\]

then \( \varphi \) is increasing. Hence,

\[
\varphi(t) \geq \varphi(1) = \sum_j |\lambda_j| = |a|_1
\]

i.e. \( \sum_i ||ae_i|| \geq |a|_1 \). Select \( \{e_i\} \) to be \( \{\xi_i\} \), the equation holds. \( \square \)

**Proposition 1.** For \( T \in B(H) \), \( Ta \) and \( aT \) are of Class 1, provided \( a \) is of Class 1.

**Proof.**

\[
|Ta|_1 \leq \sum_i \||Ta||\xi_i|| = \sum_i \|u^* Ta\xi_i\| \leq \sum_i \|u^* T\| \|a\xi_i\| \leq \|T\| \sum_i \|a\xi_i\|
\]

where \( Ta = U|Ta| \) is the polar decomposition.

Now pick \( \{\xi_i\}_1^\infty \) an eigenvectors base of \( a \),

\[
\sum_i \langle u^* aT \xi_i, \xi_i \rangle = \sum_i \langle u^* \sum_j T_{ij} \lambda_j \xi_j, \xi_i \rangle = \sum_j \sum_i T_{ij} \langle u^* \xi_j, \xi_i \rangle \lambda_j
\]

\[
\left| \sum_i T_{ij} \langle u^* \xi_j, \xi_i \rangle \right| = \left| \langle u^* \xi_j, T^* \xi_j \rangle \right| \leq \|T\|
\]

Thus \( |Ta|_1 \leq \|T\| |a|_1, \|aT\|_1 \leq \|T\| |a|_1 \) \( \square \)

**Lemma 2.** \( Trace(Ta) = Trace(aT) \).

**Definition 2.** Generally, for \( p \geq 1 \), define Class \( p \) norm:

\[
|a|_p = \left( \sum_i \lambda_i^p \right)^{1/p} = \left( Trace(|a|^p) \right)^{1/p}
\]

where \( \{\lambda_i\}_1^\infty = sp(|a|) \) counted as multiplicity.
Theorem 1. $I_p$: $p$ class is an ideal of $K$ which is self-adjoint and $|a|_p$ is a complete norm for $I_p$ and $|a|_p \geq ||a||$.

Proof. we only prove $a \in I_p \Rightarrow a^* \in I_p$. And $\| \cdot \|_p$ is complete, the others are similar to the case of $p = 1$. Since $a = U|a|$, $a^* = |a|U^*$, we have $a \in I_p \Rightarrow |a| \in I_p \Rightarrow a^* \in I_p$.

If $|a_n - a_m|_p \to 0(n, m \to \infty)$, then $a = \lim_n a_n$ in norm topology. Then

$$|a_n - a_m| = ((a_n - a_m)^*(a_n - a_m))^{1/2} \to 0$$

$$\sum_{i=1}^{N} \langle |a_n - a_m| |p e_i, e_i \rangle \to 0$$

since $a_n, a_m$ norm bounded and $\forall i, \langle |a_n - a_m| |p e_i, e_i \rangle \to 0$.

If $\forall m, n > M$,

$$\sum_{i=1}^{\infty} \langle |a_n - a_m| |p e_i, e_i \rangle < \epsilon$$

then

$$\sum_{i=1}^{N} \langle |a_n - a_m| |p e_i, e_i \rangle < \epsilon$$

let $n \to \infty$,

$$\sum_{i=1}^{N} \langle |a - a_m| |p e_i, e_i \rangle \leq \epsilon$$

let $N \to \infty$ we get $|a - a_m|_p \leq \epsilon^{1/p}$.

For any $T \in B(\mathcal{H}), T = U_1 + U_2 + U_3 + U_4, U_i$ is unitary, and $U_i a = U_i(U|a|), aU_i = UU_i(U_i^{-1}|a|U_i)$. Thus $a \in I_p \Rightarrow aT, Ta \in I_p$. □

Proposition 2. $I_p = \{ S \in K : S(x) = \sum_{k=1}^{\infty} T_k \langle x, x_k \rangle y_k, x_k, y_k \text{ are ONBs}, \{ T_k \}_1^\infty \in L^p \}$.

Proposition 3. $a \in I_1$ iff $\exists b, c \in I_2, a = bc$.

Proof. $\Rightarrow$: Obvious.

$\Leftarrow$: Set $|bc| = ubc$, then $|bc|_1 = \text{Trace}(ubc) \leq |c|_2 |b^* u^*|_2 \leq 4|c|_2 |b^*|_2 < \infty$. □

Now, we have a natural question analogy to $L^p$ spaces: what is the dual of $I_p(1 \leq p < \infty)$?

And since when $p \to \infty, (\text{Trace}(|s|^p))^{1/p} \to ||S||$. Thus it is reasonable to define $I_\infty = B(\mathcal{H})$.

We now proof $(I_1)^* = I_\infty$.

Definition 3. $\phi_h(x) = \text{Trace}(xh)$ for $x \in B(\mathcal{H}), h \in I_1$.

Theorem 2. $\forall \phi \in I_1^*, \exists x \in B(\mathcal{H}), \text{s.t., } \phi(x) = \phi_h(x)$ and $||\phi|| = ||x||$.  

72
Proof. Let \( f_{ij} : e_i \to e_j \in B(\mathcal{H}) \), \( X = (x_{ij})_{i,j} \) (matrix representation), where \( \langle Xe_i, e_j \rangle = x_{ij} = \phi_{f_{ji}}(x) = \phi(f_{ji}). \) \( \forall \{a_i\}_1^n, \)

\[
\left| \varphi(\sum_{i=1}^n a_i f_{ij}) \right| \leq \left\| \sum_{i=1}^n a_i f_{ij} \right\|_1 = \left( \sum_{i=1}^n |a_i|^2 \right)^{1/2}
\]

Thus \( \sum |f_{ij}|^2 < \infty. \)

For any \( l \in \langle e_1, \ldots, e_n \rangle, \) let \( S : X(l) : \to ||x(l)|| \) and \( S = 0 \) on \( X(l) \). Then use \( S_n \xrightarrow{\|\|} S \) where \( S_n \) are finite matrices and \( \varphi(S_n) = \text{Trace}(S_n x) \to \frac{||X(l)||}{||l||}. \)

(More explicitly, \( S = U|S| \), let \( \widetilde{S}_n \) be first \( n \times n \) matrix of \( |S| \), \( S_n = U \widetilde{S}_n \), then \( |S_n|_1 \leq |\widetilde{S}_n|_1 \), but \( \widetilde{S}_n \xrightarrow{\|\|} |S| \Rightarrow S_n \xrightarrow{\|\|} S \), at last use the first \( n \) rows of \( U \), then \( U_n \widetilde{S}_n \) finite and converge to \( S \).

We have \( \varphi(S_n) \to \varphi(S) \), and \( S_n \) are all of range \( \langle l \rangle \Rightarrow |S|_1 \geq \frac{||X(l)||}{||l||} \) for any finite \( l \). Now, \( X \) is a bounded operator on \( (H) \).

And \( \varphi(S) = \text{Trace}(S X) \) for \( S \) finite matrix. As before, use finite \( S_n \xrightarrow{\|\|} S' \) for \( S' \in I_1 \), then \( \varphi(S') = \lim_n \varphi(S_n) = \lim_n \text{Trace}(S_n X) = \text{Trace}(S' X). \)

Theorem 3. \( I_1^* \simeq B(\mathcal{H}) \) and the ultraweakly continuous functionals on \( B(\mathcal{H}) \) are \( I_1 \).

In fact, the weak-*topology on \( B(\mathcal{H}) \) coincides with its ultraweakly topology. Since on ball of \( B(\mathcal{H}) \), ultraweak topology coincides with weakly topology which is compact and metrizable, thus \( (I_1,|\cdot|) \) is separable provided \( \mathcal{H} \) is separable.

Remark: For a Banach space \( X \), the unit ball of its dual is metrizable under the *-topology iff \( X \) is separable.

Proposition 4. When \( p = 2 \), \( I_2 \) is a Hilbert space with \( [a,b] := \text{Trace}(b^* a) \) called Schmidt operators.

Lemma 3. Let \( U \) be unitary, \( T = V|T| \), then \( UT = (UV)|T| \) and \( TU = V(U^*|T|U) \) are polar decompositions. Thus \( |UT|_p = |TU|_p = |T|_p. \)

Lemma 4. Any operator is a linear combination of 4 unitary operators. Thus \( |Tx|_p, |xT|_p \leq 2||T|| \cdot |x|_p. \)

(hint: for self-adjoint \( a, ||a|| < 1, a = ((a + \sqrt{1-a^2})i) + (a - \sqrt{1-a^2})/2). \)

Proposition 5. \( |Ta|_p \leq ||T|| |a|_p. \)

In fact, when \( 1 \leq p \leq 2, \)

\[
|a|_p = \inf_{\{\xi_i\}} \left( \sum_i ||a \xi_i||^p \right)^{1/p}
\]

73
and for $p > 2$,

$$|a|_p = \sup_{\{\xi_i\}} \left( \sum_i ||a\xi_i||^p \right)^{1/p}$$

**Proof.** As in the proof of Lemma 1, $e_i = \sum_j c_{ij}\xi_j$ with $\{\xi_j\}_{i=1}^\infty$ satisfies $|a|_i = \lambda_i\xi_i$. Now

$$\left( \sum_i ||ae_i||^p \right)^{1/p} = \left( \sum_i \left( \sum_j |c_{ij}\lambda_j|^2 \right)^{p/2} \right)^{1/p}$$

$$\geq \left( \sum_i \sum_j |c_{ij}|^2|\lambda_j|^p \right)^{1/p} = \left( \sum_j |\lambda_j|^p \right)^{1/p} = |a|_p$$

when $1 \leq p \leq 2$ by Holder Inequality. Hence

$$|Ta|_p = \inf_{\{e_i\}} \left( \sum_i ||Te_i||^p \right)^{1/p} \leq \inf_{\{e_i\}} ||T|| \left( \sum_i ||ae_i||^p \right)^{1/p} = ||T|| |a|_p$$

for $1 \leq p \leq 2$. If $p > 2$, the inequality reverses, thus $|a|_p = \sup_{\{\xi_i\}} (\sum_i ||a\xi_i||^p)^{1/p}$. □

**Theorem 4.** $I_p$ is a Banach space (algebra), which is an ideal in $K$. Furthermore, for $S \in I_p, T \in I_q, 1/p + 1/q = 1, |\text{Trace}(S^*T)| \leq |S|_p |T|_q$. The mapping $h \mapsto \phi_h : \phi_h(t) = \text{Trace}(ht)$ gives a homomorphism between $I_p$ and $I_q^*$, when $1 < p < \infty$.

**Proof.** We first proof the easier one: $I_p \cong I_q^*$. Firstly, $|\text{Trace}(ht)| \leq ||h||_p ||t||_q$, i.e. $||\phi_h|| \leq ||h||_p$, or $\Phi : h \mapsto \phi_h$ is continuous.

Now we prove $\Phi$ is a surjection. Let $\varphi \in I_q^*$, then $I_1 \xrightarrow{id} I_q \xrightarrow{\varphi} C$ gives $\varphi = \varphi C_1$, which is $\phi_x$ by Theorem 2.

Let $x = |x|$ be its polar decomposition. For any $0 \leq x_n \leq |x|^{p'-1}$ which commutes with $|x|$, we have

$$\text{Trace} \left( x_n^{q'/2} \right) \leq \text{Trace}(x_n |x|) \leq ||\varphi|| \cdot 4|x_n|_{p'}$$

or $(\text{Trace}(x_n^{q'/2}))^{1/q'} \leq 4||\varphi||$ provided $x_n \in I_{q'=p'}$. If we pick $x_n, x_n^q \xrightarrow{\text{strong}} |x|^p$ increasingly, then $\sum_i (x_n^q e_i, e_i) \rightarrow \sum_i (|x|^p e_i, e_i) (n \rightarrow \infty)$ by increasing convergence. Thus $|x|_p \leq 4||\varphi||$, i.e. $x \in I_p$. □

Remark: In fact $||x||/2 \leq ||\phi_x|| \leq ||x||$ from the proof above. We picked $x_n$ s.t. $x_n^q \xrightarrow{\text{strong}} |x|^p$ increasingly. This is provided since Finite Rank Operator $S = B(\mathcal{H})$ and we have Kaplansky density theorem. This guarantees $x_n^q \in I_1$. 74
Modules and $\lambda$-Matrix

何通木*

Abstract

We use the classification of homomorphisms of finitely generated free modules over PID to prove the validity of $\lambda$-matrix method, which serves to find Jordan basis of a given linear transformation. We only need some basic notions of modules, and linear algebra.

1 Introduction

This is an old problem to simplify a linear homomorphism between finite-dimensional vector spaces. One of the best result is Jordan form. It’s important since each Jordan block reveal subtle phenomenon in linear dynamical systems (we decompose the whole complex phenomenon into special cases). Precisely, each matrix $A$ over $\mathbb{C}$ is linear conjugate to its Jordan form:

$$P^{-1}AP = \begin{pmatrix} J_1 & & \\ & J_2 & \\ & & \ddots \\ & & & J_g \end{pmatrix}$$

where $J_i = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}$

A critical question rises that how we find the transition matrix $P$, or equivalently how we find the new basis in the vector space $V$. Let’s focus on one block $A(v_1, \ldots, v_k) = (v_1, \ldots, v_k) \begin{pmatrix} \lambda & 1 & & \\ & \cdots & 1 \\ & & \cdots & 1 \\ & & & \lambda \end{pmatrix}$. $v_1$ is the eigenvector which can be found by solving $Av_1 = \lambda v_1$.

Then $v_2$ satisfies the linear equations: $Av_2 = \lambda v_2 + v_1$. One wishes this recursive process would go on, but this procedure might get stuck. For example, the eigenvectors $v_{11}, v_{12}, \ldots$ of some eigenvalue $\lambda$ you chose (a basis of $Ker(\lambda I - A)$) might not be in the subspace $(\lambda I - A)(Ker(\lambda I - A)^2)$. However, we still have a classical method stated in theorem 4.1.

*数
The idea of resolution will be extremely useful. Roughly speaking, suppose we have a surjective homomorphism \( q \) from a free module \( F \) onto \( V \), which records the information of \( A \). We take the kernel of \( q \), then the information about \( A \) is recorded in the embedding \( i \) of free modules from \( \ker(q) \) into \( F \), which can be diagonalized by looking for a basis of \( \ker(q) \) and a basis of \( F \). Hence we know the structure of \( i \), so is the quotient \( F/\ker(q) \cong V \). (The details of diagonalization are contained in theorem 2.6)

\[
\begin{array}{c}
0 \rightarrow \ker(q) \rightarrow F \quad q \rightarrow V \rightarrow 0
\end{array}
\]

In fact, we regard \( V \) as \( \mathbb{C}[\lambda] \)-module with action \( \lambda v = Av \) (denote \( R = \mathbb{C}[\lambda] \)). We choose \( F = \mathbb{C}[\lambda] \mathbb{R}^n \) where \( n = \dim V \). Fix a basis \( e_1, e_2, \ldots, e_n \) of \( V \), let \( q(x_1, x_2, \ldots, x_n) = x_1e_1 + \cdots + x_ne_n \).

We should notice that we cannot allow the kernel to remain such abstract, or else we can’t do any calculation! Observe that the torsion part of the module action of \( \mathbb{C}[\lambda] \) onto \( V \) is totally induced by \( \lambda e_j = \sum a_{ij}e_i \). So we guess we have such a short exact sequence:

\[
0 \rightarrow \mathbb{R}^n \xrightarrow{\lambda I - A} \mathbb{R}^n \xrightarrow{q} V \rightarrow 0
\]

where \( \lambda I - A \) is the matrix multiplication by the \( \lambda \)-matrix \( \lambda I - A \).

\( \lambda \)-matrix method is based on this observation: Each matrix representation \( A \) of linear transformation \( \phi \) corresponds to such a short exact sequence. The precise statements are in the lemmas 3.1, 3.2 and concluded in remark 3.4. And corollary 3.3 is the main tool of \( \lambda \)-matrix method. And we finally use \( \lambda \)-matrices to find Jordan basis, that’s theorem 4.2.

This note will prove the validity of \( \lambda \)-matrix method in the view of classifying homomorphisms of free modules. The preliminaries beyond linear algebra are only the basic notion of modules and some elementary algebraic properties of PID. For the conciseness of this note, we tacitly approve the existence and uniqueness of Jordan forms, which in fact is easy to be verified by using diagonalization theorem 2.6.

2 Diagonalization of Module Homomorphism

Basic knowledge about modules can be found in Hungerford[1] or Jacobson[2], we leave out the proof.

**Definition 2.1.** A ring is called PID, if it is a principle ideal domain, i.e. a domain whose ideal is generated by one element. For example, \( \mathbb{C}[\lambda] \), the complex polynomial ring with one variable, is a PID.

**Definition 2.2.** A module \( M \) over a ring \( R \) is an abelian group with left ring action. It’s called free, if \( M \) has a set of generators, any finite subset of them is \( R \)-linearly independent. The cardinality of this set is called the rank of \( M \).
Definition 2.3. A short sequence of modules is called exact, $0 \to M \overset{i}{\to} N \overset{q}{\to} V \to 0$ if $\text{Ker}(q) = \text{Im}(i)$ and $i$ is injective, $q$ is surjective.

Proposition 2.4. A PID ring $R$ satisfies unique factorization. The greatest common divisor $w$ of two nonzero elements $a, b \in R$ is a linear combination of $a, b$, that is the Bézout identity: $as + bt = w$.

Proposition 2.5. The Jordan form of a matrix over $\mathbb{C}$ exists, and is uniquely determined by its invariant factors, or equivalently elementary factors.

Theorem 2.6. Let $\varphi : M \to N$ is a homomorphism between finitely generated free modules $M, N$ over PID $R$. Then there is a basis $y_1, \ldots, y_m$ in $M$ and a basis $x_1, \ldots, x_n$ in $N$, such that $\varphi(y_i) = d_i x_i$ where $d_1 | d_2 | \cdots | d_k$, $d_{k+1} = \cdots = d_m = 0$, $d_i \in R$.

Proof. We choose basis in $M, N$ arbitrarily, say $y_1, \ldots, y_m$ and $x_1, \ldots, x_n$. Let $\varphi$ have the matrix representation:

$$\varphi(y_1, \ldots, y_m) = (x_1, \ldots, x_n) A$$

where $A = (a_{ij})$, $\varphi(y_j) = \sum a_{ij}x_i$. According to the transition formula:

$$\varphi(y_1, \ldots, y_m)Q = (x_1, \ldots, x_n)P \cdot P^{-1} AQ$$

we only need to diagonalize the matrix $A$ by left and right multiplication. We have three types of basic transformations:

1. Exchange rows (columns): $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} b & d \\ a & c \end{pmatrix}$;

2. Sum rows (columns): $\begin{pmatrix} 1 & 0 \\ r & 1 \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} a + ra & c + rc \\ b + ra & d + rc \end{pmatrix}$;

3. Division of rows (columns): $\begin{pmatrix} s & t \\ \frac{-b}{w} & \frac{a}{w} \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} l & cs + dt \\ 0 & \frac{ad - bc}{w} \end{pmatrix}$ where $as + bt = w$ is the Bézout identity in PID $R$ ($w$ is the greatest common divisor of $a, b$);

Consider an integer value $L = \min\{l(a_{ij})\}$, where $l(a)(a \in R)$ is the number of its prime divisors, that is, $a = p_1 \cdots p_{l(a)}$, $l(a) = 0$ if $a$ is unit and set $l(0) = +\infty$. We take induction on the scale of $A$ and also on this value. (The case $L = +\infty$ is trivial)

If $L = 0$, we assume $a_{11}$ is a unit by exchanging rows and columns. We use $a_{11}$ to eliminate other elements which is on the same row or column with it. Hence we have:

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & \vdots & & A_{n-1,m-1} \\ 0 & & & \end{pmatrix}$$

77
by induction on scale, we are done.

If \( L > 0 \), we assume \( a_{11} \) has the minimal value of \( l \) by exchanging rows and columns. If there is another element on the same row or column with it which can’t be divided by it, say \( a_{21} \). Then we use the division of row to generate their greatest common divisor, so that \( L \) decrease, and done by induction. If there is no such element, we still use \( a_{11} \) to eliminate other elements which is on the same row or column with it. By induction on scale, we have:

\[
\begin{pmatrix}
  a_{11} & 0 & 0 & \cdots & 0 & 0 \\
  0 & b_2 & 0 & \cdots & 0 & 0 \\
  0 & 0 & b_3 & \cdots & 0 & 0 \\
  \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
  0 & 0 & 0 & \cdots & b_k & 0 \\
  0 & 0 & 0 & \cdots & 0 & \ddots \\
\end{pmatrix}
\]

where \( b_2 | b_3 | \cdots | b_k \). If \( a_{11} \nmid b_2 \) we add the second column onto the first column, and use once division of rows, then \( L \) decrease. Or else \( a_{11} \mid b_2 \), we are done.

\[\square\]

Remark 2.7. \( d_1, \ldots, d_k \) are uniquely determined by \( \varphi \). They are called invariant factors.

### 3 \( \lambda \)-Matrix Method

Given a \( n \)-dimensional complex vector space \( V \), and a linear transformation \( \phi : V \rightarrow V \). Let \( R = \mathbb{C}[\lambda] \). Fix a basis \( e_1, \ldots, e_n \) of \( V \), under which the representation matrix of \( \phi \) is \( A = (a_{ij}) \). \( V \) is viewed as a \( R \)-module by action \( \lambda v = \phi(v) \), particularly \( \lambda e_j = \sum a_{ij} e_i \).

**Lemma 3.1.** \( 0 \rightarrow R^n \xrightarrow{\varphi} R^n \xrightarrow{q} V \rightarrow 0 \) is exact, where \( q((x_1, x_2, \ldots, x_n)^T) = x_1 e_1 + \cdots + x_n e_n \), \( \varphi((y_1, y_2, \ldots, y_n)^T) = (\lambda - A)(y_1, y_2, \ldots, y_n)^T \).

**Proof.** It’s obvious that \( q \) is surjective. If \( \varphi((y_1, y_2, \ldots, y_n)^T) = 0 \), suppose \( y_1 \) has the maximal degree regarded as polynomial. Since \( (\lambda - a_{11})y_1 - \sum a_{ij} y_j = 0 \), so we must have \( (y_1, y_2, \ldots, y_n)^T = 0 \). Hence \( \varphi \) is injective.

Since \( q \circ \varphi((0, 1, \ldots, 0)^T) = q((-a_{11}, \ldots, \lambda - a_{jj}, \ldots, -a_{nj})^T) = 0 \), \( q \circ \varphi = 0 \). If \( q((x_1, x_2, \ldots, x_n)^T) = 0 \), we may subtract some elements in \( Im(\varphi) \) from it so that \( x_i \in \mathbb{C} \). Hence \( q((x_1, x_2, \ldots, x_n)^T) = 0 \) indicates \( x_i = 0 \). We have \( Ker(q) = Im(\varphi) \).

According to the diagonalization theorem 2.6, we may find a basis \( y_1, y_2, \ldots, y_n \) in the first \( R^n \) and a basis \( x_1, x_2, \ldots, x_n \) in the middle \( R^n \), so that \( \varphi(y_i) = d_i x_i \) where \( d_1 | d_2 | \cdots | d_k, d_{k+1} = \cdots = d_n = 0, d_i \in R \). Since we have isomorphism

\[ q : R^n/Im(\varphi) \rightarrow V \]

78
The images of generators $x_1, \ldots, x_k$ in $R^n/\text{Im}(\phi)$ are the generators of cyclic subspaces of invariant factors $d_1, \ldots, d_k$. Thus we derive the elementary factors from invariant factors, and we know what the Jordan form of $\phi$ is. We'll show how we find the transition matrix in the following.

The opposite of lemma 3.1 is

**Lemma 3.2.** Let $M,N$ be free $R$-modules of rank $n$, $0 \to M \xrightarrow{\phi} N \xrightarrow{\nu} V \to 0$ is exact. If $\phi$ has matrix representation $\lambda I - A$ under the basis $y_1, y_2, \ldots, y_n$ in $M$ and a basis $x_1, x_2, \ldots, x_n$ in $N$, then $q(x_1), \ldots, q(x_n)$ form a basis of $V$ under which the linear transformation $\phi$ on $V$ has the matrix representation $A$.

**Proof.** Since $q$ is surjective, $q(x_1), \ldots, q(x_n)$ generate $V$ as $R$-module, that is for any $v \in V$, there are $f_1, \ldots, f_n \in R$ such that $v = \sum f_i q(x_i)$. Notice that the exactness indicates $q(\lambda x - \sum a_{ij} x_i) = 0$, that is $\lambda q(x_j) - \sum a_{ij} q(x_i) = 0$. Hence $v$ is indeed a linear combination of $q(x_1), \ldots, q(x_n)$, which indicates $q(x_1), \ldots, q(x_n)$ form a basis of $V$ as vector space. $\lambda q(x_j) - \sum a_{ij} q(x_i) = 0$ tells us the matrix of $\phi$ is $A$. □

**Corollary 3.3.** $\lambda I - A, \lambda I - B$ are equivalent $\iff$ $A, B$ are conjugate. Moreover, if $P(\lambda)^{-1}(\lambda I - A)Q(\lambda) = \lambda I - B$, then $P(A)^{-1}AP(A) = B$, where $P(\lambda) = P_0 + \lambda P_1 + \cdots + \lambda^p P_p$, $P(A) = P_0 + AP_1 + \cdots + A^p P_p (P_i \in M_n(C))$.

**Proof.** "$\Leftarrow$" is trivially guaranteed by lemma 3.1 and "$\Rightarrow$" is guaranteed by lemma 3.2. Suppose $\lambda I - A$ corresponds to a basis $x_1^A, x_2^A, \ldots, x_n^A$ of $N$, and $\lambda I - B$ corresponds to a basis $x_1^B, x_2^B, \ldots, x_n^B$ of $N$. The transition formula (2) shows:

$$(x_1^B, x_2^B, \ldots, x_n^B) = (x_1^A, x_2^A, \ldots, x_n^A)P(\lambda)$$

Then by computation:

$$(q(x_1^B), q(x_2^B), \ldots, q(x_n^B)) = \sum_{t=0}^p \lambda^t(q(x_1^A), q(x_2^A), \ldots, q(x_n^A))P_t$$

$$= \sum_{t=0}^p (q(x_1^A), q(x_2^A), \ldots, q(x_n^A))A^t P_t$$

$$=(q(x_1^A), q(x_2^A), \ldots, q(x_n^A))P(\lambda)$$ □

**Remark 3.4.** We can summarize those results in a commutative diagram:

$$
\begin{array}{c}
0 \longrightarrow R^n \xrightarrow{\phi} R^n \xrightarrow{q} C^n \longrightarrow 0 \\
Q(\lambda) \downarrow \quad P(\lambda) \downarrow \quad P(A) \downarrow \\
0 \longrightarrow R^n \xrightarrow{\nu} R^n \xrightarrow{q} C^n \longrightarrow 0
\end{array}
$$
where \( q((0, \ldots, 1, \ldots, 0)^T) = (0, \ldots, 1, \ldots, 0)^T \).

## 4 Find Jordan Basis

**Theorem 4.1** (Naive method). The following algorithm computes the Jordan basis:

1. Given a matrix \( A \), find all of its eigenvalues: solve \( \det(\lambda I - A) = 0 \).
2. For each eigenvalue \( \lambda \), write down \((\lambda I - A), (\lambda I - A)^2, \ldots \) and simultaneously solve \((\lambda I - A)x = 0, (\lambda I - A)^2x = 0, \ldots \), until the dimensions of solutions stay constant, i.e. suppose \( \text{rank}(\lambda I - A)^{k-1} > \text{rank}(\lambda I - A)^k = \text{rank}(\lambda I - A)^{k+1} = \cdots \). Denote the basis of \( \text{Ker}(\lambda I - A)^i \) by \( v_{i1}, \ldots, v_{im_i} \).
3. Solve this equation:
   \[
   x^T \left( (\lambda I - A)^{k-1} \right) v_{(k-1)1} \ v_{(k-1)2} \ \ldots \ v_{(k-1)n_{k-1}} = 0
   \]
   denote its solution basis by \( w_{k1}, \ldots, w_{km_k} \). This is the basis of \( \text{Ker}(\lambda I - A)^k \cap (\text{Ker}(\lambda I - A)^{k-1})^\perp \cong \text{Ker}(\lambda I - A)^k / \text{Ker}(\lambda I - A)^{k-1} \). Write \( w_{kj} = (\lambda I - A)^i w_{kj} \). They are the generators of \( k \times k \)-Jordan blocks.
4. Suppose we’ve found \( w_{k1}, \ldots, w_{km_k}, \ldots, w_{(i+1)1}, \ldots, w_{(i+1)m_{i+1}} \). Solve this equation:
   \[
   x^T \left( (\lambda I - A)^{i-1} \right) v_{(i-1)1} \ \ldots \ v_{(i-1)n_{i-1}} \ w_{k1}^{i-1} \ \ldots \ w_{(i+1)m_{i+1}}^i = 0
   \]
   denote its solution basis by \( w_{i1}, \ldots, w_{im_i} \). They are the generators of \( i \times i \)-Jordan blocks.
5. \( \bigcup_{\lambda} \{ w_{ij}^l \mid 1 \leq i \leq k, 1 \leq j \leq m_i, 0 \leq l \leq i - 1 \} \) are the Jordan basis for eigenvalue \( \lambda \).

**Proof.** Only need to show \( \{ w_{ij}^l \} \) form a basis for \( \text{Ker}(\lambda I - A)^n \). According to step 4, we see that \( \{ w_{(i+1)j}^l \mid 1 \leq j \leq m_{i+1}, 0 \leq l \leq k - i \} \) forms a basis of \( \text{Ker}(\lambda I - A)^i \cap (\text{Ker}(\lambda I - A)^{i-1})^\perp \).

Since \( \text{Ker}(\lambda I - A)^n = V_k = V_{k-1} \oplus (V_k \cap V_{k-1}^\perp) = V_{k-2} \oplus (V_{k-1} \cap V_{k-2}^\perp) \oplus (V_k \cap V_{k-1}^\perp) = \cdots = \bigoplus_{i=1}^k V_i \cap V_{i-1}^\perp \)

\[ \square \]

**Theorem 4.2** (\( \lambda \)-matrix method). The following algorithm computes the transition matrix:
1. Given a matrix $A$, extend it to a $n \times 2n$-matrix $(\lambda I - A \ I)$. Do the three types of row or column basic transformation on $\lambda I - A$ (if it’s row-type, then do the same on the right). Hence we get

\[
\begin{pmatrix}
d_1 & & \\
& \ddots & \\
& & d_k \\
& & P(\lambda)^{-1} \\
0 & & \cdots
\end{pmatrix}
\]

2. Use these invariant factors $d_1, \ldots, d_k$ to write down the Jordan form $J$ of $A$. Consider $(\lambda I - J \ I)$, and do the same thing as above, we get:

\[
\begin{pmatrix}
d_1 & \cdots & \\
& \ddots & \\
& & d_k \\
& & \bar{P}(\lambda)^{-1}
\end{pmatrix}
\]

3. Consider $\begin{pmatrix} P(\lambda)^{-1} & \bar{P}(\lambda)^{-1} \end{pmatrix}$, and do the same thing as above, we get:

\[
\begin{pmatrix}
I & P(\lambda)\bar{P}(\lambda)^{-1}
\end{pmatrix}
\]

4. Write $\bar{P}(\lambda) = P(\lambda)\bar{P}(\lambda)^{-1}$, then $\bar{P}(A)^{-1}A\bar{P}(A) = J$.

Proof. Suppose $P(\lambda)^{-1}(\lambda I - A)Q(\lambda)$ is the diagonal matrix. By the diagonalization theorem 2.6, we know the step 1 computes $P(\lambda)^{-1}$. For the same reason, step 2 computes $\bar{P}(\lambda)^{-1}$, where $\bar{P}(\lambda)^{-1}(\lambda I - J)\bar{Q}(\lambda) = P(\lambda)^{-1}(\lambda I - A)Q(\lambda)$. So $\lambda I - J = (P(\lambda)\bar{P}(\lambda)^{-1})^{-1}(\lambda I - A)Q(\lambda)\bar{Q}(\lambda)^{-1}$, and the corollary 3.3 tells us $\bar{P}(A)^{-1}A\bar{P}(A) = J$. \hfill \Box

References


征稿启事

《荷思》是清华大学数学系学生自主创办的数学学术刊物，面向各个院系中对数学感兴趣的本科生及研究生。

本刊欢迎全校师生任何与数学有关的投稿，无论是长篇的论述，还是精彩的小品，抑或学习/教学的心得、习题的妙解。在原作者的允许下，推荐他人的作品也同样欢迎。

为了编辑方便，建议投稿者提供电子版，并且采用 TeX 推荐) 或者 Word排版。来稿注明作者，联系方式。

投稿请寄THUmathgooglegroups.com

欢迎访问荷思网站http://math.tsinghua.me/hesi/

《荷思》编辑部
2017-07
但它具有某些永恒的性质。

我们所做的事可能是渺小的，