15 Lecture 15. Siegel and KAM

There is almost no hope to solve celestial n ($n > 2$) body problem analytically in closed forms, the only remaining analytic hope is perturbative. Perturbative approaches generally have a grave difficulty called the small divisor problem. If the unperturbed periodic orbit is sufficiently 'irrational' (the Diophantine condition), this difficulty may be overcome. How to do this systematically is the key issue and is the core of its solution is proposed by Kolmogorov.

Here, (1) we consider perturbation around a center: when are the majority of invariant periodic orbits survive? [Poincaré-Siegel problem], then go to (2) the Kolmogorov-Arnold-Moser (KAM) theorem, asserting the survival of numerous invariant tori under perturbation. (1) is given with full demonstration details to taste the strategy used in (2). (2) is complicated, so only an outline of the demonstration is given.

15.1 Perturbative solution of motion

We know what we mean by solving the motion of a Hamiltonian system. It is to devise a smooth canonical transformation that 'linearizes the equation of motion (see **11.1**). What happens if the system is not solvable in the sense we already discussed? If the nonlinear part of the dynamics is small (only a perturbation), a natural idea is to construct a transformation near identity to get rid of this nonlinearity. This is the basic idea of perturbation. Especially because the results due to Poincare and others (**14.7**, **14.8**) that (almost) dashed the hope of solving celestial mechanics by quadrature, we desperately need perturbative approaches.

If we try to implement this strategy, we almost immediately encounter a grave difficulty called the problem of small divisors. Thus, first let us taste the difficulty and how to overcome it with a 'simple'[153] complex map problem: Find $x \to y$ that can linearize the original ODE in terms of x that has small nonlinear term g (that is $dg/dx|_{x=0} = 0$):

$$\dot{x} = Ax + g(x) \;\Rightarrow\; y = Ay, \tag{15.1}$$

where A is a diagonal matrix (for simplicity). This is the famous Poincaré-Siegel linearization problem (answers are **15.8** and **15.12**).

[153] The map problem is simple, but the solution is anything but simple as you will see.

15.2 Formal linearization around singular point

Consider an ode (if needed let us complexify it 4.7)

$$\dot{x} = f(x), \tag{15.2}$$

where $x \in \mathbb{R}^n$ and f is real analytic with $f(0) = 0$. The linear term is assumed to be diagonalized as Ax, where $A = [\alpha_1, \cdots, \alpha_n]$.

Let us consider the formal power series transformation

$$x = u(y) = y + u^2(y) + \cdots + u^k(y) + \cdots, \tag{15.3}$$

where u^k is a degree k homogeneous polynomial of y. (15.2) reads componentwisely as

$$\dot{x}_j = \alpha_j x_j + \sum_{|k|\geq 2} f_j^k x^k. \tag{15.4}$$

We have used the Hadamard notation.[154] (15.3) reads in this notation

$$x_j = y_j + \sum_{|k|\geq 2} u_j^k y^k. \tag{15.5}$$

15.3 Formal transformation to the second order

If we solve (15.3) of (15.5) for y we obtain

$$y_j = x_j - \sum_{|k|=2} u_j^k x^k + \cdots. \tag{15.6}$$

Here the higher order terms are very complicated and are suppressed. The original ODE in terms of y can be obtained by differentiating this as[155]

$$\dot{y}_j \overset{(15.6)}{=} \dot{x}_j - \sum_{|k|=2} u_j^k (k \cdot \alpha) x^k + \cdots \tag{15.7}$$

---

[154] ⟪**Hadamard notation**⟫ For $x = (x_1, \cdots, x_n)$ and $\alpha = (\alpha_1, \cdots, \alpha_n)$, $x^\alpha = \prod_{j=1}^n x_j^{\alpha_j}$. Also we write for $k = (k_1, \cdots, k_n) \in \mathbb{N}^n$, $k! = \prod_{j=1}^n k_j!$. Very often $|k| = \sum_j k_j$. If $f : \mathbb{R}^n \to \mathbb{R}$ and differentiable, the multivariate Taylor expansion reads

$$f(x+y) = \sum_{k\in\mathbb{N}^n} \frac{1}{k!} y^k \frac{d^k}{dx^k} f(x).$$

[155] $dx^k = d(\sum_j x_j^{k_j}) = \sum_j k_j dx_j (x_j^{k_j - 1}) = \sum_j \alpha_j x_j k_j (x_j^{k_j-1}) dt = \sum_j \alpha_j k_j (x_j^{k_j}) dt = (a \cdot k) x^k dt$. More formally, $dx^k = (k \cdot dx) x^{k-1} = (Ak) x^k$.

$$\overset{(15.4)}{=} \sum_j \alpha_j x_j + \sum_{|k|=2} f_j^k x^k - \sum_{|k|=2} u_j^k (k\cdot\alpha) x^k + \cdots \tag{15.8}$$

$$\overset{(15.5)}{=} \sum_j \alpha_j \left(y_j + \sum_{|k|=2} u_j^k y^k \right) + \sum_{|k|=2} f_j^k y^k - \sum_{|k|=2} u_j^k (k\cdot\alpha) y^k + \cdots \tag{15.9}$$

$$= \sum_j \alpha_j y_j + \sum_{|k|=2} \left[\alpha_j u_j^k + \ f_j^k - u_j^k (k\cdot\alpha) \right] y^k + \cdots \tag{15.10}$$

$$= \sum_j \alpha_j y_j + \sum_{|k|=2} \left\{ f_j^k + \left[\alpha_j - (k\cdot\alpha) \right] u_j^k \right\} y^k + \cdots . \tag{15.11}$$

Therefore, if we can set

$$u_j^k = f_j^k / (k\cdot\alpha - \alpha_j), \tag{15.12}$$

the second order terms have been eliminated. For this to be possible we need a non-resonance condition for $|k| = 2$

$$\alpha_j - k\cdot\alpha \neq 0. \tag{15.13}$$

15.4 Non-resonance condition

As we will see the non-resonance condition is crucial, so let us state the condition clearly:

For $\alpha \in \mathbb{C}^n$ If the following set $\Gamma(\alpha) = \emptyset$, we say α is non-resonant:

$$\Gamma(\alpha) = \{ j \in \{1, 2, \cdots, n\},\ k \in \mathbb{N}^n \,|\, \alpha_j - k\cdot\alpha = 0 \}. \tag{15.14}$$

15.5 Formal transformation to order *k*

Replacing '2' with K in (15.5) as

$$x_j = y_j + \sum_{|k|=K} u_j^k y^k + \cdots , \tag{15.15}$$

we can repeat the computation in **15.3** to get the following instead of (15.11)

$$\dot{y}_j = \sum_j \alpha_j y_j + \sum_{2\le|k|<K} f_j^k y_j^k + \sum_{|k|=K} \left[f_j^k + (\alpha_j - (k\cdot\alpha))\, u_j^k \right] y^k + \cdots \tag{15.16}$$

Therefore, under the non-resonance condition (15.13), we can eliminate the order K terms.

Notice that in (15.15) $\cdots$ terms are not needed. Thus we can get the following obvious theorem:

15.6 Finite order elimination of nonlinear terms

For any $M \in \mathbb{N}^+$ we can make a polynomial transformation of order M

$$x = u(y) = y + u^2(y) + \cdots + u^M(y) \tag{15.17}$$

such that the original ODE (15.2) with the non-resonance condition **15.4** can be transformed into the following form:

$$\dot{y}_j = \alpha_j y_j + \sum_{|k|>M} g_j^k y^k. \tag{15.18}$$

15.7 What happens if $|\boldsymbol{k}|$ is very large?

If $|k|$ is very large, then $\alpha_j - k \cdot \alpha$ can be very close to zero (imagine a lattice and then try to draw a line not hitting any lattice point).

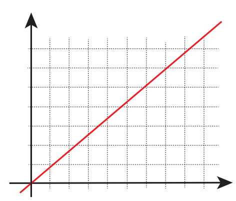

Figure 15.1: Suppose $\alpha_1 > 0$ and $\alpha_2 < 0$. $\alpha_1 k_1 + \alpha_2 k_2 = 0$ (red line) can be very close to a positive integer lattice points for large $|k|$ even if $\alpha_1/\alpha_2 \notin \mathbb{Q}$. In this case Poincaré's condition is trivially violated.

Thus, non-resonance alone cannot guarantee that u_k^j is of manageable size. Poincaré proved that if the convex hull of $\{\alpha_1, \cdots, \alpha_n\}$ does not contain the origin on the complex plane, and if α satisfies the non-resonance condition, then we can iterate the procedure in **15.6** indefinitely. That is,

15.8 Poincaré's lemma

For an ODE with the linear portion diagonalized and the nonlinear part denoted as

g:

$$\dot{x} = Ax + g(x), \tag{15.19}$$

if the convex hull of the eigenvalues do not contain the origin, and the non-resonance condition is satisfied, then there is an analytic transformation $x = y + u(y)$ with $du/dy|_{y=0} = 0$ such that

$$\dot{y} = Ay. \tag{15.20}$$

Its proof is through an 'honest' construction of the majorizing series that is convergent (tedious but straightforward).

15.9 Significance of the convex hull condition

Let $\alpha_1, \cdots, \alpha_n \in \mathbb{C}$ and its convex hull $[\alpha_1, \cdots, \alpha_n]$ does not contain the origin. Then, for any α in this convex hull there is $\delta > 0$ such that $|k \cdot \alpha| \geq \delta |k|$ for any $k \in \mathbb{N}^n$. Thus, what we worried in **15.7** never happens.

This is easy to see. Consider a unit vector $k/|k|$. Then $|k \cdot \alpha|/|k|$ is the length of the projection of the vector $(\alpha_1, \cdots, \alpha_n)$ onto $k/|k|$. Thus, this cannot be smaller than the distance between $[\alpha_1, \cdots, \alpha_n]$ and the origin.

15.10 What happens if Poincare condition fails?

Siegel demonstrated the following theorem:

Theorem [Siegel] For

$$\dot{x} = Ax + f(x), \tag{15.21}$$

where (as above) $A = [\alpha_1, \cdots, \alpha_n]$ is diagonal and f is analytic and $f'(0) = 0$. If there is $\gamma > 0$ such that for any $k \in \mathbb{Z}^n$ ($|k| \neq 0$)

$$|k \cdot \alpha| > \gamma |k|^{-n}, \tag{15.22}$$

then (15.21) is analytically transformed to $y = Ay$.

Notice (see **15.11**) that $\{\alpha_1, \cdots, a_n\}$ satisfying (15.22) is with full measure (exception is measure zero) wrt the usual Lebesgue measure m on $\mathbb{R}^{2n}$ when we identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$.

15.11 Diophantine approximation

Lemma. For almost all $\alpha \in \mathbb{C}^n$[156] we can take a positive number γ such that for all

[156] This means: regarding α as a $2n$-real vector, it is almost sure with respect to the Lebesgue measure m on $\mathbb{R}^{2n}$.

$k \in \mathbb{Z}^n$ ($|k| \neq 0$)

$$|k \cdot \alpha| > \gamma/|k|^n. \tag{15.23}$$

[Demo]
Choose an arbitrary $R > 0$. Consider the totality Σ_γ of α for which (15.23) does not hold (i.e., (15.23) is untrue for any choice of $\gamma > 0$) and $|\alpha_i| < R$ for $i \in \{1, \cdots, n\}$. To estimate $m(\Sigma_\gamma)$ we consider the condition for α:

$$|k \cdot \alpha| \leq \gamma/|k|^n \tag{15.24}$$

for each shell of k: $K \leq |k| < K + 1$; it is an $n-1$-sphere of radius K with shell thickness 1. There are $\sim (2K)^{n-1}$ vectors. For each such k, the projection of α must be smaller than $\sim 1/K^{n+1}$ according to (15.24), so the α must be in the slab of thickness $\sim 2\gamma/K^{n+1}$ perpendicular to k and the remaining $n-1$ dimensions are with span $2R$. Thus, the total number of α satisfying (15.24) for $|k| \sim K$ is bounded by

$$\sim (K)^{n-1} \times 2\gamma/K^{n+1} \times (2R)^{n-1} = C(R)\gamma/K^2, \tag{15.25}$$

where $C(R)$ is an R dependent constant. Thus, we have an upper bound

$$m(\Sigma_\gamma) \leq C(R)\gamma \sum_K K^{-2}. \tag{15.26}$$

Therefore, the infimum of this wrt γ is zero. Since R is arbitrary, we have shown the lemma.

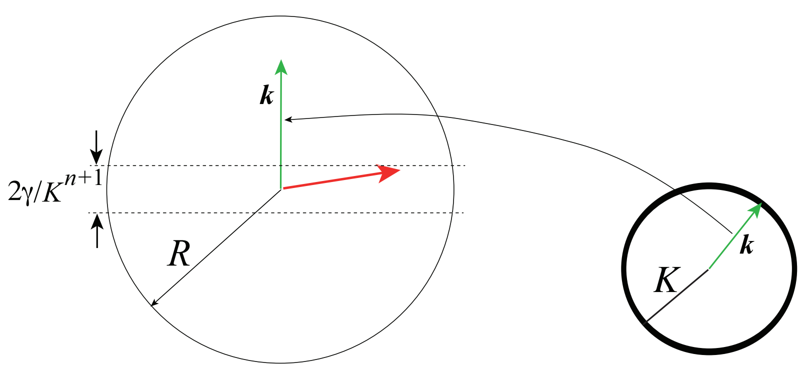

Figure 15.2: (15.25) illustrated

15.12 Siegel's stability theorem for conformal maps

We should continue to study the ODE considered by Poincaré, but here, we go to a simpler problem with the same difficulty.[157]

Let $S : \mathbb{C} \to \mathbb{C}$ be defined as $z \to f(z)$:

$$f(z) = \lambda z + g(z), \tag{15.27}$$

where $\lambda \neq 0$, $g(0) = 0$ which is holomorphic in $z < r$

We say the origin is stable for the system S, if and only if for any neighborhood of the origin U, there is a neighborhood of the origin $V \subset U$ such that $S^n V \subset U$ for all $n \in \mathbb{Z}$.

Siegel proved:

Theorem: If $|\lambda| = 1$ and $|\lambda^q - 1|^{-1} < C_0 q^2$ for $C_0 > 0$ and for any $q \in \mathbb{N}^+$ (the Diophantine condition), then S is stable.

15.13 Strategy to demonstrate Siegel's theorem

The strategy of the proof here (due to Moser) following Kolmogorov's idea is as follows:
First we show
Lemma: S is stable iff
(1) $|\lambda| = 1$,
(2) There is a holomorphy u in a neighborhood of the origin such that $\lim_{|z|\to 0} u(z)/z \to 1$ (asymptotically an identity) and

$$u(\lambda\zeta) = f(u(\zeta)). \tag{15.28}$$

That is, the variable change due to $z = u(\zeta)$ converts the original into $\lambda\zeta$:

$$\begin{array}{ccc} z & \xrightarrow{S} & f(z) \\ u\uparrow & & u\uparrow \\ \zeta & \longrightarrow & \lambda\zeta. \end{array} \tag{15.29}$$

Thus, the theorem requires to solve (15.28). Instead of solving this, that is,

$$u(\lambda\zeta) = \lambda u(\zeta) + g(u(\zeta)) \tag{15.30}$$

at once, we solve a partially linearized version:

$$u_1(\lambda\zeta) = \lambda u_1(\zeta) + g(\zeta). \tag{15.31}$$

[157] Polya said: if you cannot understand a problem, there must be a simpler problem you cannot understand. Find it.

This u_1 cannot render the original equation to $\lambda\zeta_1$ even if $z = u_1(\zeta_1)$; actually,

$$u_1(\lambda\zeta_1 + \Phi^{(1)}(\zeta_1)) = f(\lambda u_1(\zeta_1)). \tag{15.32}$$

That is,

$$\begin{array}{ccc} z & \xrightarrow{\quad S \quad} & f(z) \\ {\scriptstyle u_1}\Big\uparrow & & {\scriptstyle u_1}\Big\uparrow \\ \zeta_1 & \longrightarrow & \lambda\zeta_1 + \Phi^{(1)}(\zeta_1). \end{array} \tag{15.33}$$

Rewrite the original equation in terms of the new variable $\zeta_1 = u_1(z)$. The nonlinear term is reduced. That is, if $g < \varepsilon$, then, roughly speaking, $|\Phi^{(1)}| < \varepsilon^2$. Repeating this procedure rapidly decreases the size of the nonlinear term.

$$\begin{array}{ccc} z & \xrightarrow{\quad S \quad} & f(z) \\ {\scriptstyle u_1}\Big\uparrow & & {\scriptstyle u_1}\Big\uparrow \\ \zeta_1 & \xrightarrow{S_1} & \lambda\zeta_1 + \Phi^{(1)}(\zeta_1) \\ {\scriptstyle u_2}\Big\uparrow & & {\scriptstyle u_2}\Big\uparrow \\ \zeta_2 & \xrightarrow{S_2} & \lambda\zeta_2 + \Phi^{(2)}(\zeta_2) \\ {\scriptstyle u_3}\Big\uparrow & & {\scriptstyle u_3}\Big\uparrow \\ \cdots & & \cdots. \end{array} \tag{15.34}$$

Thus, $u = \cdots \circ u_n \circ u_{n-1} \circ \cdots \circ u_2 \circ u_1$.

15.14 Preparatory lemma

$S : \mathbb{C} \to \mathbb{C}$ is stable iff for $\forall U$ neighborhood of $z = 0$ there is a simply connected neighborhood V $(\subset U)$ of origin such that $SV = V$.

[Demo] $\Leftarrow$ Trivial.
$\Rightarrow$ According to the definition of stability $\exists W \ni S^n W \subset U$ for $\forall n \in \mathbb{Z}$. Define

$$\tilde{V} = \cup_{n\in\mathbb{Z}} S^n W. \tag{15.35}$$

Take a connected component V of $\tilde{V}$ containing $z = 0$. Then, any curve C connecting any point in V and the origin inside V satisfies $S^n C \subset V$ (i.e., different connected components do not map between them by S). Therefore, $SV = V$.

If V is not simply connected, repeat the above procedure all over starting from a smaller U. Suppose the resultant V is not simply connected even if U is very small. This, however, contradicts the assumption that S is holomorphic around $z = 0$, because there must be a neighborhood on which S looks very close to a rigid rotation (= linearized version). Thus, we can get the desired V.

15.15 Poof of Lemma in 15.13

Our key lemma already mentioned in **15.13** is:
Lemma: S is stable iff
(1) $|\lambda| = 1$
(2) There is a holomorphy u in a neighborhood of the origin such that $\lim_{|z|\to 0} u(z)/z \to 1$ (asymptotically an identity) and

$$u(\lambda\zeta) = f(u(\zeta)). \tag{15.36}$$

[Demo]
$\Leftarrow$ is obvious. Take a small disk D and $u(D) = V$.
$\Rightarrow$ According to the preparatory lemma **15.14**, stability implies the existence of $V \not\in SV = V \ni 0$, where V is simply connected (i.e., topologically equivalent to a disk). Then the Riemann mapping theorem guarantees the existence of a conformal map u from V to a disk D of radius ρ such that $u(0) = 0$ and

$$u(\zeta) = \zeta + b_2\zeta^2 + \cdots. \tag{15.37}$$

$u^{-1} \circ S \circ u : D \to D$ must be a rigid rotation of D, so there is μ ($|\mu| = 1$) and

$$u^{-1} \circ S \circ u(\zeta) = \mu\zeta. \tag{15.38}$$

$\mu = \lambda$, so $|\lambda| = 1$.
Now, we can start the strategy outlined in **15.13**.

15.16 If g is small, $u_1(z) - z$ must be small

If g, which satisfies $g(0) = g'(0) = 0$, in (15.27) is small in a neighborhood of the origin, u_1 constructed as an approximation to u should not be very far from the identity. This is shown in the Lemma below.

More precisely,
Lemma: Let g be holomorphic on a disk $|\zeta| < r$ and $|g| < \varepsilon$ with $g(0) = g'(0) = 0$. Define u_1 as the solution to

$$u_1(\lambda\zeta) = \lambda u(\zeta) + g(u_1(\zeta)). \tag{15.39}$$

Then, u_1 is holomorphic in $|\zeta| < r$ and on $|\zeta| < r(1-\theta)$ ($\theta \in (0,1)$)

$$|u_1(\zeta) - \zeta| < 2C_0\varepsilon/\theta^3, \tag{15.40}$$

where C_0 is a positive constant.

> Technically, to show that u_n are all close to identity, we must show $\Phi^{(n)}$ are all small. It is more convenient to show that u_1 solving (15.39) should not be very far from the identity, if g', instead of g, is small. This is given as a corollary **15.17**.

[Demo of Lemma]
Since g is holomorphic in $|\zeta| < r$, we can expand it as

$$g(\zeta) = \sum_{k\geq 2} g_k \zeta^k \tag{15.41}$$

with $|g_k| < \varepsilon/r^k$ (the Cauchy bound). Let us formally expand u_1 as

$$u_1 = \zeta + \sum_{k\geq 2} u_k \zeta^k. \tag{15.42}$$

Putting this into (15.39), we obtain

$$\sum_{k\geq 2} u_k[(\lambda\zeta)^k - \lambda\zeta^k] = \sum_{k\geq 2} g_k \zeta^k. \tag{15.43}$$

Thus, we must solve

$$[\lambda^k - \lambda]u_k = g_k. \tag{15.44}$$

That is (notice that we potentially have a small denominator difficulty, which is overcome by the Diophantine condition),

$$u_k = g_k/[\lambda^k - \lambda]. \tag{15.45}$$

We can show that u_1 is well defined and its deviation from the identity is bounded by ε by an explicit calculation as follows.

$$|u_1(\zeta) - \zeta| = \left|\sum_{k\geq 2}[\lambda^k - \lambda]^{-1} g_k \zeta^k\right| \leq \sum_{k\geq 2} |\lambda^{k-1} - 1|^{-1}|g_k||\zeta|^k \tag{15.46}$$

$$\leq \sum_{k\geq 2} C_0(k-1)^2 \varepsilon r^{-k}|\zeta|^k < \sum_{k\geq 2} C_0(k-1)^2\varepsilon(1-\theta)^k \tag{15.47}$$

where we have used the Cauchy bound for g_k and the Diophantine condition. Notice that

$$\sum_{k\geq 2}(k-1)^2 x^2 \sim \frac{d^2}{dx^2}\sum x^k \sim \frac{d^2}{dx^2}\frac{1}{1-x} \sim (1-x)^{-3}, \tag{15.48}$$

so (set $x = 1 - \theta$ to use this identity)

$$|u_1(\zeta) - \zeta| < \varepsilon \alpha C_0/\theta^3, \tag{15.49}$$

where α is a positive constant (which can be chosen as 2 with a bit more detailed calculation).

15.17 If g' is small, $u_1(z) - z$ must be small

If g, which satisfies $g(0) = g'(0) = 0$, in (15.27) and if its derivative g' is small in a neighborhood of the origin, u_1 constructed as an approximation to u should not be very far from the identity.

More precisely,
Lemma: Let g be holomorphic on a disk $|\zeta| < r$ and $|g'| < \varepsilon$ with $g(0) = g'(0) = 0$. Define u_1 as the solution to

$$u_1(\lambda\zeta) = \lambda u(\zeta) + g(u_1(\zeta)). \tag{15.50}$$

Then, u_1 is holomorphic in $|\zeta| < r$ and in $|\zeta| < r(1-\theta)$ ($\theta \in (0,1)$)

$$|u_1(\zeta) - \zeta| < 2C_0\varepsilon r/\theta^3, \tag{15.51}$$

where C_0 is a positive constant.
[Demo]
This should not be hard to guess, since $|g|$ must be of order $r \times |g'| < r\varepsilon$. A formal demonstration may be as follows. From (15.50) we get

$$\lambda u_1'(\lambda\zeta) = \lambda u_1'(\zeta) + g'(\zeta). \tag{15.52}$$

Let $v = \zeta u_1'(\zeta)$. We have

$$\lambda v(\lambda\zeta) = \lambda v(\zeta) + \zeta g'(\zeta). \tag{15.53}$$

Since $|\zeta g'(\zeta)| < \varepsilon|\zeta|$ in $|\zeta| < r$, the lemma in **15.16** tells us that in $|\zeta| < r(1-\theta)$

$$|v(\zeta) - \zeta| < 2C_0|\zeta|/\theta^3. \tag{15.54}$$

That is,

$$|u_1' - 1| < 2C_0\varepsilon/\theta^3. \tag{15.55}$$

Thus,

$$|u_1 - \zeta| = \left| \int_0^\zeta (u_1' - 1) d\zeta \right| < 2C_0 r\varepsilon/\theta^3. \tag{15.56}$$

15.18 (15.34) is a commutative diagram

To construct u we must be able to show that all the 'squares' in (15.34) must commute. That is, for example, u_1 and u_1^{-1} must be well defined in a neighborhood of the origin. We have already shown that u_1 is well defined in $|\zeta| < r(1-\theta)$.

To demonstrate u_1^{-1} is well-defined, we show that $u_1(\zeta)$ is topologically the same as ζ. To show this we use Rouche's theorem:[158]

Lemma. Choose ε small enough to satisfy for $\theta \in (0, 1/4)$

$$2C_0\varepsilon < \theta^4. \tag{15.57}$$

Then, u_1^{-1} is well defined in $|z| < r(1-2\theta)$ and

$$u_1^{-1}(\{z \,|\, |z| < r(1-2\theta)\}) \subset \{\zeta \,|\, |\zeta| < r(1-\theta)\}. \tag{15.58}$$

[Demo]

(15.57) and (15.56) tells us that for $v(\zeta) = u_1(\zeta) - \zeta$

$$|v(\zeta)| \le r\theta. \tag{15.59}$$

If $|z| < r(1-2\theta)$ and $|\zeta| < r(1-\theta)$, then this implies that

$$|v(\zeta)| \le |\zeta| - |z| \le |\zeta - z|. \tag{15.60}$$

Since $|\lambda| = 1$, this implies that the equations $\lambda\zeta + v(\zeta) = u_1(\zeta) = z$ and $\zeta = z$ have the same zeros. Thus, u_1^{-1} is well defined in $|z| < r(1-2\theta)$ and its image is in $|\zeta| < r(1-\theta)$.

15.19 S_1 is well defined close to the origin

We can show

Lemma: $S_1(\zeta) = u_1^{-1} \circ f \circ u_1 = \lambda\zeta + \Phi^{(1)}(\zeta)$ in (15.34) is well defined in $|\zeta| < r(1-4\theta)$ if $\varepsilon < \theta$.

[Demo]

Since $|f| \le |\zeta| + |g|$ and $|g| < r\varepsilon$, if $|g'| < \varepsilon$

$$|f| \le |\zeta| + |g| \le |\zeta| + r\varepsilon \le |\zeta| + r\theta, \tag{15.61}$$

because we assume $\varepsilon < \theta$.

The image of f must be in the domain of u_1^{-1}, i.e., $|z| < r(1-2\theta)$. Therefore,

[158] ⟪**Rouche's theorem**⟫ Let D be a simply connected open set. Suppose f and g are non-constant holomorphic functions on $[D]$ (closure), and $|f| > |g|$ on ∂D. Then, the number of zeros of f and $f+g$ agree in D.

the domain of f must be restricted to $|z| < r(1-3\theta)$. Then, (15.58) implies that the domain of u_1 must be restricted to $|\zeta| < r(1-4\theta)$.

15.20 $\Phi^{(1)}$ is small and we can iterate the above argument indefinitely

If we can show that $\Phi^{(1)}$ is sufficiently small, we can repeat the above argument for u_2, and ad infinitum. We can show (we wish to use an analogue of **15.17** to show that u_2 is close to identity)

Lemma: $|{\Phi^{(1)}}'(\zeta)| < C_1\varepsilon^2/\theta^4$, where $C_1 < 3C_0$ in $|\zeta| > r(1-5\theta)$ with $0 < \varepsilon < \theta < 1/5$.

[Demo]

$$u_1(\lambda\zeta + \Phi^{(1)}) = f(u_1) = \lambda u_1(\zeta) + g(u_1), \tag{15.62}$$

so (recall $v = u_1 - \zeta$ and $v(\lambda\zeta) = \lambda v(\zeta) + g(\lambda\zeta)$)

$$\lambda\zeta + \Phi^{(1)} + v(\lambda\zeta + \Phi^{(1)}) = \lambda\zeta + \lambda v(\zeta) + g(\lambda\zeta + v(\zeta)). \tag{15.63}$$

Therefore,

$$\Phi^{(1)} = \lambda v(\zeta) - v(\lambda\zeta + \Phi^{(1)}) + g(\lambda\zeta + v(\zeta)), \tag{15.64}$$

but $\lambda v(\zeta) = v(\lambda\zeta) - g(\lambda\zeta)$ gives

$$\Phi^{(1)} = v(\lambda\zeta) - v(\lambda\zeta + \Phi^{(1)}) + g(\lambda\zeta + v(\zeta)) - g(\lambda\zeta). \tag{15.65}$$

Using the mean value theorem

$$|v(\lambda\zeta) - v(\lambda\zeta + \Phi^{(1)})| \le \sup|v'| \sup|\Phi^{(1)}| \le \frac{2\varepsilon C_0}{\theta^3} \sup|\Phi^{(1)}| < \theta \sup|\Phi^{(1)}| < \frac{1}{5} \sup|\Phi^{(1)})|. \tag{15.66}$$

From this and (15.65)

$$|\Phi^{(1)}| \le \frac{1}{5} \sup|\Phi^{(1)}| + |g(\lambda\zeta + v(\zeta)) - g(\lambda\zeta)| \tag{15.67}$$

or

$$\frac{4}{5} \sup|\Phi^{(1)}| \le \sup|g(\lambda\zeta + v(\zeta)) - g(\lambda\zeta)|. \tag{15.68}$$

Again we use the mean-value theorem

$$\frac{4}{5} \sup|\Phi^{(1)}| \le \sup|g'| \sup|v(\zeta)| < \varepsilon \frac{2C_0\varepsilon}{\theta^3} r. \tag{15.69}$$

Thus we get for $|\zeta| < r(1-4\theta)$

$$\sup|\Phi^{(1)}| < \frac{5C_0\varepsilon^2}{2\theta^3} r. \tag{15.70}$$

Its derivative may be estimated following Cauchy in a somewhat smaller domain $|\zeta| < r(1-5\theta)$ as

$$|\Phi^{(1)\prime}(\zeta)| < C_1\varepsilon^2/\theta^4. \tag{15.71}$$

15.21 Summary of a recursion step

Thus we have completely constructed the single step for the recursion scheme ((15.33) realized) [Please ignore strange arrowheads at the ends of the formulas]

$$\begin{array}{ccc} z & \xrightarrow{S} & \lambda z + g(z) \\ {\scriptstyle u_1}\uparrow & & {\scriptstyle u_1}\uparrow \\ \zeta_1 & \xrightarrow{S_1} & \lambda\zeta_1 + \Phi^{(1)}(\zeta_1) \end{array} \qquad \begin{array}{c} |g'(z)| < \varepsilon \text{ in } |z| < r \\ \\ |\Phi^{(1)}(\zeta)| < C_1\varepsilon^2/\theta^4 \text{ in } |\zeta| < r(1-5\theta). \end{array} \tag{15.72}$$

or more generally

$$\begin{array}{ccc} \zeta_n & \xrightarrow{S_n} & \lambda\zeta_n + \Phi^{(n)}(\zeta_n) \\ {\scriptstyle u_{n+1}}\uparrow & & {\scriptstyle u_{n+1}}\uparrow \\ \zeta_1 & \xrightarrow{S_{n+1}} & \lambda\zeta_{n+1} + \Phi^{(n+1)}(\zeta_{n+1}) \end{array} \qquad \begin{array}{c} |\Phi^{(n)\prime}(\zeta_n)| < \varepsilon_n \text{ in } |\zeta_n| < r_n \\ \\ |\Phi^{(n+1)}(\zeta_{n+1})| < C_1\varepsilon_n^2/\theta_n^4 = \varepsilon_{n+1} \text{ in } |\zeta_{n+1}| < r_n(1-5\theta_n) = r_{n+1}. \end{array} \tag{15.73}$$

15.22 Convergence of the recursive transformations

For the recursive transformations to converge we must at least demand $\varepsilon_n \to 0$ and $r_n \to r_\infty > 0$. Thus, $\sum \theta_n$ must be convergent. Moreover,

$$U_n = u_n \circ u_{n-1} \circ \cdots \circ u_1 \tag{15.74}$$

must be uniformly convergent in $|\zeta| < r_\infty$. Or equivalently

$$U_n' = u_n' u_{n-1}' \cdots u_1' \tag{15.75}$$

must be uniformly convergent. Since

$$|U_n'| \leq \prod_{k=1}^{n} (1 + |\Phi^{(n)\prime}|), \tag{15.76}$$

if

$$\sum |\Phi^{(n)\prime}| \leq \sum \varepsilon_n \tag{15.77}$$

converges, we are done!

Thus the requirements are

$$\sum \varepsilon_n < \infty, \tag{15.78}$$

$$\sum \theta_n < \infty \tag{15.79}$$

with $0 < \varepsilon_n < \theta_n < 1/5$. This is satisfied if we choose $\theta_n = (1/5)c^{-n}$ for $c > 1$. Indeed,

$$\varepsilon_{n+1} = \frac{C_1}{\theta_n^4}\varepsilon_n^2 = \frac{C_1}{5}c^{4n}\varepsilon_n^2 < C_2^{n+1}\varepsilon_n^2 \tag{15.80}$$

Define $\varepsilon_n = C^{n+2}\varepsilon_n$. Then $\varepsilon'_{n+1} < (\varepsilon'_n)^2$, so choose $\varepsilon'_0 \leq 1$ and all the requirements for θ_n are met.

15.23 Setup of the simplest version of KAM theorem

Let $H(\theta, I, \varepsilon)$ be a near integrable Hamiltonian; for $\varepsilon = 0$

$$H_0(I) = H(\theta, I, 0). \tag{15.81}$$

That is, I is invariant, and determines an invariant T^n as asserted by the Liouville-Arnold theorem. We wish to see whether the invariant T^n with $I = I_0$ survives the perturbation. Since we may add any constant to I, we choose $I_0 = 0$. We consider I close to $I_0 = 0$: $B = \{I \mid |I| < b\}$ for some small $b > 0$. We assume $H(\theta, I, \varepsilon)$ is holomorphic on $\mathcal{M} \times (-\varepsilon_0, \varepsilon_0)$. We write the zeroth and the first order terms as

$$H(\theta, I, 0) = E + \omega \cdot I + Q(I), \tag{15.82}$$

where $Q = O[I^2]$ in the $I \to 0$ limit.

We assume

(1) ω satisfies a Diophantine condition **15.11**

$$|k \cdot \omega| \geq \frac{\alpha}{|k|^\tau} \tag{15.83}$$

with $\tau > n - 1$.

(2) $H(\theta, I, 0)$ is non-degenerate in the sense that $\det(\partial^2 Q/\partial I_i \partial I_j) \neq 0$.

15.24 KAM theorem

With the setup **15.23** there is a holomorphic canonical transformation $\phi : T^n \times$

$B^* \times (-\varepsilon_*, \varepsilon_*) \to \mathcal{M}$, where $B^* \subset B$ and $0 < \varepsilon_* < \varepsilon_0$, such that the canonical transformation $(\theta, I) \to (\theta', I')$ transforms $H(\theta, I, \varepsilon)$ as

$$K(\theta', I'.\varepsilon) = E_*(\varepsilon) + \omega \cdot I' + Q_*(\theta', I', \varepsilon), \tag{15.84}$$

where $Q_* = O[I'^2]$ in the $I' \to 0$ limit. Here, ϕ is an identity for $\varepsilon = 0$.

This implies that the canonical equation of motion reads

$$\frac{d\theta'}{dt} = \frac{\partial K}{\partial I'} = \omega + \frac{\partial Q_*}{\partial I'}, \tag{15.85}$$

$$\frac{dI'}{dt} = -\frac{\partial K}{\partial \theta'} = -\frac{\partial Q_*}{\partial \theta'}. \tag{15.86}$$

Notice that the derivatives of Q_* for $I' = 0$ vanish because $Q_* = O[I'^2]$. This means that the 'holomorphically deformed' $I = 0$ torus survives and the motion on it is just the original (almost) periodic flows. This torus is called a KAM torus.

15.25 Strategy of proof of KAM

The idea is just as explained for Siegel's theorem. Let the perturbed Hamiltonian reads

$$H = H_0 + \varepsilon P - 0, \tag{15.87}$$

where $H_0 = E_0 + \omega I + Q(I)$ with $Q = O[I^2]$.

We construct a canonical transformation $\phi_1 : (\theta, I) \to (\theta', I')$ such that

$$H_1 = H \circ \phi_1 = K_1 + \varepsilon^2 P_1, \tag{15.88}$$

where

$$K_1 = E_1(\varepsilon) + \omega I' + Q_1(\theta', I', \varepsilon), \tag{15.89}$$

with $Q_1 = O[I'^2]$.

This is accomplished by removing the θ-dependent constant portion and the first order in I of P_0 of order ε choosing ϕ_1.

If we repeat the same strategy, we get

$$H_2 = H_1 \circ \phi_2 = K_2 + \varepsilon^4 P_2. \tag{15.90}$$

Of course we must show that $\phi_n \circ \cdots \circ \phi_2 \circ \phi_1$ converges.

15.26 Canonical transformation that can remove $O[\varepsilon]$ terms

We introduce $(\theta, I) \to (\theta', I')$ with the generator $G(\theta, I')$

$$dG = Id\theta + \theta' dI' + (K - H)dt, \tag{15.91}$$

but $G(\theta, I')$ is time independent, so $K = H(I(\theta', I'), \theta(\theta', I'))$.

We wish to remove I'-independent θ' dependence (to keep I' invariant) and the $O[I']$ terms that cannot be written as $\omega \cdot I'$ to order ε. We assume the following form:

$$I = \frac{\partial G}{\partial \theta} = I' + \varepsilon \beta(\theta, I'), \tag{15.92}$$

$$\theta' = \frac{\partial G}{\partial I'} = \theta + \varepsilon a(\theta). \tag{15.93}$$

If ε is sufficiently small, we can invert the second equation as

$$\theta = \varphi(\theta', \varepsilon). \tag{15.94}$$

Thus we may set (b is a constant vector)

$$G = \theta \cdot I' + \varepsilon[\theta \cdot b + s(\theta) + a(\theta) \cdot I'] \tag{15.95}$$

We see that

$$\beta(\theta, I') = b + \frac{\partial}{\partial \theta} s(\theta) + \frac{\partial a}{\partial \theta} I'. \tag{15.96}$$

Notice that we can choose s and a integration over T^n to vanish (by subtracting appropriate constants tat we can ignore from G).

15.27 Transformation of Hamiltonian

Let us write

$$H(\theta, I, \varepsilon) = H(\theta, I, 0) + \varepsilon P(\theta, I, \varepsilon) = E + \omega \cdot I + Q(I) + \varepsilon P(\theta, I, \varepsilon). \tag{15.97}$$

First, we change $I \to I'$:

$$H(\theta, I, \varepsilon) = H(\theta, I' + \varepsilon \beta(\theta, I'), \varepsilon) \tag{15.98}$$

$$= E + \omega \cdot I' + Q(\theta, I') + \varepsilon \left[\omega \cdot \beta + \frac{\partial Q}{\partial I'} \beta + P(\theta, I', 0) \right] + \varepsilon^2 P'(\theta, I', \varepsilon), \tag{15.99}$$

where P' denotes all the remaining terms.

Since we do not care for $O[I'^2]$, but we wish to rewrite the rest of $O[\varepsilon]$ as constant. Let us expand $O[\varepsilon]$ in powers of I':

$$\omega \cdot \beta + \frac{\partial Q}{\partial I'}\beta + P(\theta, I', 0) = \omega \cdot \left(b + \frac{\partial s}{\partial \theta} + \frac{\partial a}{\partial \theta} I'\right) + \frac{\partial Q}{\partial I'}\left(b + \frac{\partial s}{\partial \theta} + \frac{\partial a}{\partial \theta} I'\right) + P(\theta, I', 0) \tag{15.100}$$

$$= \omega \cdot (b + \frac{\partial s}{\partial \theta}) + P(\theta, 0, 0) \tag{15.101}$$

$$+ \left[\omega \cdot \frac{\partial a}{\partial \theta} + (b + \frac{\partial s}{\partial \theta})\frac{\partial^2 Q}{\partial I' I'} + \frac{\partial P}{\partial I'}\right] I' + O[I'^2] \tag{15.102}$$

15.28 Removing $\boldsymbol{\theta}$-dependence from energy

The I' independent term in (15.102) is

$$\omega \cdot b + D_\omega s(\theta) + P(\theta, 0, 0), \tag{15.103}$$

where

$$D_\omega = \omega \cdot \frac{\partial}{\partial \theta}. \tag{15.104}$$

Therefore, to remove the θ dependence, we must choose s so that

$$\omega \cdot b + D_\omega s(\theta) + P(\theta, 0, 0) = \omega \cdot b + P_0 \tag{15.105}$$

where P_0 is the average value of $P(\theta, 0, 0)$. Thus, we must solve

$$D_\omega s(\theta) = -(P(\theta, 0, 0) - P_0), \tag{15.106}$$

or we must show the well-definedness of D_ω^{-1}. Here, we encounter the small devisor problem.

As shown below **15.31-15.33**, since the average of $P(\theta, 0, 0) - P_0$ over T^n vanishes, s is well-defined.

15.29 Removing the $\boldsymbol{O[I']}$ term

We wish to eliminate

$$D_\omega a + (b + \frac{\partial s}{\partial \theta})\frac{\partial^2 Q}{\partial I' I'} + \frac{\partial P}{\partial I'}. \tag{15.107}$$

Since we wish to determine a the average of the terms other than $D + w\alpha$ over T^n must vanish. First, we choose b to satisfy this condition. This is possible due to the non-degeneracy condition (Hess $\neq 0$). Then, a can be computed just as s.

15.30 Full transformation

The remaining task is to replace θ in $H(\theta, I' + \varepsilon\beta(\theta, I'), \varepsilon)$ with θ' (see (15.94)).

15.31 Well-definedness of D_ω^{-1}: formal solution

We wish to solve

$$D_\omega u = \omega \cdot \frac{\partial u}{\partial \theta} = f. \tag{15.108}$$

To solve this formally, u and f are Fourier expanded as

$$u(\theta) = \sum_{k \in \mathbb{R}^n} u_k e^{ik\theta}, \quad f(\theta) = \sum_{k \in \mathbb{R}^n} f_k e^{ik\theta}, \tag{15.109}$$

where

$$u_k = \frac{1}{(2\pi)^n} \int_{T^n} d\theta\, u(\theta) e^{-ik\theta}, \quad f_k = \frac{1}{(2\pi)^n} \int_{T^n} d\theta\, f(\theta) e^{-ik\theta}. \tag{15.110}$$

Therefore, (15.108) gives

$$ik\omega u_k = f_k \tag{15.111}$$

For this to be solvable $f_0 = 0$ (i.e., f averaged over T^n is zero. Therefore, the general solution to (15.108) is given be

$$u(\theta) = u_0 + \sum_{k \in \mathbb{R}^n \setminus \{o\}} \frac{f_k}{i\omega k} e^{ik\theta}. \tag{15.112}$$

Thus, if the average of u over T^n vanishes (i.e., $u_0 = 0$), the solution to (15.108) is formally unique.

15.32 Well-definedness of D_ω^{-1}: convergence of formal solution

We assume that the real analytic function on T^n is analytic in a 'strip' T_ξ^n containing the real axis:

$$T_\xi^n = \{\theta \in \mathbb{C}^n \,|\, |\mathrm{Im}\theta_j| < \xi\}/(2\pi\mathbb{Z})^n \tag{15.113}$$

If f is analytic on T_ξ^n, then

$$|f_k| \leq \|f\| e^{-|k}\xi, \tag{15.114}$$

where $\|f\|$ is the max of $|f$ on T^n_ξ. Its demonstration is in **15.33**.

Now, we assume the Diophantine condition

$$|k\omega| \geq \alpha|k|^\tau. \tag{15.115}$$

Then,

$$|u_k| = \left|\frac{f_k}{i\omega k}\right| \leq \alpha^{-1}\|f\|e^{-|k|\xi}|k|^\tau. \tag{15.116}$$

Using this estimate the maximal convergence of u is shown.

15.33 Cauchy estimate of $|f_k|$

$$f_k = \frac{1}{(2\pi)^n}\int_{T^n} d\theta\, f(\theta)e^{-ik\theta} \tag{15.117}$$

The integration path may be shifted to the imaginary direction so that jth component is shifted by $i\xi\text{sgn}(k_j)$ ($\theta_j \to \theta_j + i\xi\text{sgn}(k_j)$). Then, $ik\theta \to ik\theta - \xi|k|$ (we stick to the Hadamard notation).

$$\begin{aligned} f_k &= \frac{1}{(2\pi)^n}\int_{T^n} d\theta\, f(\theta)e^{-ik\theta} && (15.118)\\ &= \frac{1}{(2\pi)^n}\int_{T^n} d\theta\, f(\theta_j + i\xi\text{sgn}(k_j))e^{-ik_j(\theta_j+i\xi\text{sgn}(k_j))} && (15.119)\\ &= \frac{1}{(2\pi)^n}\int_{T^n} d\theta\, f(\theta_j + i\xi\text{sgn}(k_j))e^{-ik\theta-\xi|k|}. && (15.120)\end{aligned}$$

Therefore,

$$\begin{aligned} |f_k| &= e^{-\xi|k|}\left|\frac{1}{(2\pi)^n}\int_{T^n} d\theta\, f(\theta_j + i\xi\text{sgn}(k_j))e^{-ik\theta}\right| && (15.121)\\ &\leq e^{-\xi|k|}\frac{1}{(2\pi)^n}\int_{T^n} d\theta\, \left|f(\theta_j + i\xi\text{sgn}(k_j))e^{-ik\theta}\right| && (15.122)\\ &\leq e^{-\xi|k|}\frac{1}{(2\pi)^n}\int_{T^n} d\theta\, \|f\| = \|f\|e^{-|k|\xi}. && (15.123)\end{aligned}$$