DISCRETE AND CONTINUOUS
DYNAMICAL SYSTEMS
Volume **4**, Number **4**, October **2003** pp. 1–20

A PROOF OF KOLMOGOROV'S THEOREM

Dedicated to Prof. Vishik on the occasion of his 80th birthday

John Hubbard and Yulij Ilyashenko

Department of Mathematics
Malott Hall, Cornell University
Ithaca, NY 14853

Abstract. In this paper we will give a proof of Kolmogorov's theorem on the conservation of invariant tori. This proof is close to the one given by Bennettin, Galgani, Giorgilli and Strelcyn in [2]; we follow the outline of their proof, but carry out the steps somewhat differently in several places. In particular, the use of balls rather than polydiscs simplifies several arguments and improves the estimates.

Acknowlegements. We have benefitted from conversations with Adrien Douady and Pierre Lochak. Jerôme Droniou, Nicolas Bédaride, Camille Bièche and Bastien Albriet were students in a course devoted to this material and provided many helpful comments.

In 1954, Kolmogorov [4] announced his theorem on the conservation of invariant tori when an integrable hamiltonian system is perturbed. He never published a proof of this result; Arnold [1] did provide a proof, and at about the same time Moser [4] proved a closely related result, so the whole field has come to be called KAM theory. Over the years many improvements and variants have appeared. This paper does not aim at generality, but instead at providing an easy and short proof of the weakest form of the theorem. Indeed, Bennettin, Galgani, Giorgilli and Strelcyn have provided such a proof in [2], and the present paper is mainly a further simplification of their proof.

1. A crash course in hamiltonian mechanics.

All the results presented in this section are standard; we have collected their proofs in the appendix.

If (X, σ) is a symplectic manifold, then any function H on X has a symplectic gradient $\nabla_\sigma H$, which is the unique vector field such that

$$\sigma(\xi, \nabla_\sigma H) = dH(\xi)$$

for any vector field ξ. We can then consider the Hamiltonian differential equation

$$\dot{x} = (\nabla_\sigma H)(x). \tag{1.1}$$

Example 1.2. If $X = \mathbb{R}^{2n}$, with coordinates $(\mathbf{q}, \mathbf{p}) = (q_1, \ldots, q_n, p_1, \ldots, p_n)$ and $\sigma = \sum_i dp_i \wedge dq_i$, then Equation 1.1 becomes the famous *Hamiltonian equations of motion*

$$\begin{aligned} \dot{q}_i &= \frac{\partial H}{\partial p_i} \\ \dot{p}_i &= -\frac{\partial H}{\partial q_i} \end{aligned} \tag{1.3}$$

AMS 2000 Subject Classification: Primary 70H08, 70H07,70H05.
Key words and phrases:Hamiltonian mechanics, Invariant tori, KAM theory

The vector field $\nabla_\sigma f$ has a flow which we will denote by ϕ_f^t. It has two key properties:

- ϕ_f^t preserves f, i.e., $f \circ \phi_f^t = f$;
- ϕ_f^t preserves σ, i.e., $(\phi_f^t)^* \sigma = \sigma$.

The central construction in Kolmogorov's theorem is a symplectic diffeomorphism, which will be constructed as a composition of hamiltonian flows. In the process, we will need to compute the Taylor polynomial of functions of the form

$$t \mapsto g \circ \phi_f^t.$$

The natural ways to approach this is via the Poisson bracket. The Lie bracket $[\nabla_\sigma f, \nabla_\sigma g]$ on vector fields is of course well defined and symplectic, and we might wonder whether it is the symplectic gradient of some function. This is the case.

Define the Poisson bracket of functions on X by

$$\{f, g\} = \sigma(\nabla_\sigma g, \nabla_\sigma f) = df(\nabla_\sigma g) = -dg(\nabla_\sigma f). \tag{1.4}$$

Then this does correspond to the Lie bracket:

$$\nabla_\sigma \{f, g\} = [\nabla_\sigma f, \nabla_\sigma g].$$

We will say that functions f, g commute if $\{f, g\} = 0$. This certainly implies that their flows commute, in fact the flows commute if and only if the Poisson bracket is constant.

In the "standard case" of Example 1.2, the Poisson bracket is computed by the formula

$$\{f, g\} = \sum_{i=1}^n \left(\frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} \right). \tag{1.5}$$

Of more immediate relevance, it allows us to write Taylor polynomials:

$$f \circ \phi_g^t = f + t\{f, g\} + \frac{t^2}{2}\{\{f, g\}, g\} + \frac{t^3}{3!}\{\{\{f, g\}, g\}, g\} + \dots. \tag{1.6}$$

Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. We will define an *integrable system* to be a system where $X = (\mathbb{T})^n \times \mathbb{R}^n$, with variables $(\mathbf{q} \in \mathbb{T}^n, \mathbf{p} \in \mathbb{R}^n)$ and symplectic form $\sum_i dp_i \wedge dq_i$ as above, and whose Hamiltonian function $H(\mathbf{p})$ depends only on $\mathbf{p}$.

It is easy to integrate the equation (1.3) in this case: the solution with initial value $(\mathbf{q}_0, \mathbf{p}_0)$ is simply

$$\begin{aligned} \mathbf{q}(t) &= \mathbf{q}_0 + t\frac{\partial H}{\partial \mathbf{p}}(\mathbf{p}_0) = \mathbf{q}_0 + t\,\omega(\mathbf{p}_0) \\ \mathbf{p}(t) &= \mathbf{p}_0 \end{aligned}$$

In particular, each coordinate $p_1, \dots, p_n$ is conserved, and the motion is a linear motion on the torus $\mathbb{T}^n \times \{\mathbf{p}_0\}$.

A famous theorem of Liouville (Theorem A 6.1 of the Appendix) asserts that this situation occurs "anytime" you have a mechanical system with n degrees of freedom and n commuting conservation laws. More precisely, if X is a symplectic manifold of dimension $2n$, and $f_1, \dots, f_n$ are n commuting functions such that $F = (f_1, \dots, f_n)$ is a submersion, and if $F^{-1}(0)$ is compact, then $F^{-1}(0)$ is a torus, and there are coordinates $(\mathbf{q} \in \mathbb{T}^n, \mathbf{p} \in \mathbb{R}^n)$ on a neighborhood of this torus such that $\sigma = \sum dp_i \wedge dq_i$, and the hamiltonian is p_1.

2. An informal statement and proof of Kolmogorov's theorem.

Suppose we perturb an integrable hamiltonian $h_0(\mathbf{p})$ by a small $h_1(\mathbf{q}, \mathbf{p})$, so that the functions $p_1, \dots, p_n$ are no longer conserved. Are there still any tori invariant under the hamiltonian flow of $h = h_0 + h_1$?

Let us ask specifically if the torus corresponding to $\mathbf{p} = 0$ is preserved. Set $\omega = \partial h_0/\partial \mathbf{p}$. Then Kolmogorov's theorem asserts that if

- $h_1(\mathbf{q}, \mathbf{p})$ is sufficiently small, (2.1)
- $\omega_0 = \omega(\mathbf{0})$ is sufficiently irrational, and (2.2)
- $\omega(\mathbf{p})$ varies sufficiently fast at $\mathbf{p} = \mathbf{0}$, (2.3)

then there exists a symplectic diffeomorphism $\Phi : (\mathbf{P}, \mathbf{Q}) \to (\mathbf{p}, \mathbf{q})$ close to the identity such that if $H = h \circ \Phi$, then

$$H(\mathbf{Q}, \mathbf{P}) = a + \omega_0 \cdot \mathbf{P} + R(\mathbf{Q}, \mathbf{P}) \quad \text{with} \quad R(\mathbf{Q}, \mathbf{P}) \in O(|\mathbf{P}|^2). \tag{2.4}$$

In particular, the motion

$$\mathbf{Q}(t) = \mathbf{Q}(0) + t\omega_0, \quad \mathbf{P}(t) = 0$$

is a solution of the hamiltonian equation, which is conjugate to the linear flow with direction ω_0, so that the invariant torus $\mathbf{p} = 0$ is preserved by the perturbation.

Strategy of the proof. The equation 2.4 is an equation for a diffeomorphism Φ, which we need to solve. Moreover, the solution should be symplectic, adding the equation $\Phi^*\sigma = \sigma$. As usual when solving non-linear equations, we will use a variant of Newton's method, approximating Φ by a sequence of maps Φ_i each computed from the previous by solving an appropriately linearized equation.

In practice, this will mean writing

$$\Phi_i = \phi_i \circ \phi_{i-1} \circ \cdots \circ \phi_1, \tag{2.5}$$

where each ϕ_i is the time one hamiltonian flow ϕ_{g_i} for some "hamiltonian function" g_i, which is the unknown for which we will solve. This has two important advantages:

• the unknown g_i is a function, and functions are simpler than diffeomorphisms;

• the corresponding mapping ϕ_{g_i} is automatically a diffeomorphism, and it is automatically symplectic.

Thus the proof is by induction: at the ith stage we will have constructed a Hamiltonian $\tilde{h} = \Phi_i^* h$, which we will develop as a Taylor polynomial with respect to $\mathbf{p}$, with coefficients that are Fourier series with respect to $\mathbf{q}$. More precisely, we will write $\tilde{h} = \tilde{h}_0 + \tilde{h}_1$, where

– $\tilde{h}_1$ consists of the constant and linear terms with respect to $\mathbf{p}$, except for the constant terms with respect to $\mathbf{q}$, and

–h_0 is everything else.

We will require that $\tilde{h}_1$ be "of order ϵ_i," whatever that means. We wish to solve a linear equation for a function g such that $\phi_g^* \tilde{h}$ is "better" than $\tilde{h}$. Ideally we would like the troublesome part $(\phi_g^* \tilde{h})_1$ to be of order $\epsilon_{i+1} \sim \epsilon_i^2$; this is the superconvergence of the standard Newton's method. But our Newton's method is not quite the standard one, and we won't do that well; but we will achieve convergence.

Write $\phi_g^*\tilde{h}$ as a Taylor polynomial with respect to g:

$$\phi_g^*\tilde{h} = \tilde{h} + \{g, \tilde{h}\} + o(|g|) = \tilde{h}_0 + \tilde{h}_1 + \{g, \tilde{h}_0\} + \{g, \tilde{h}_1\} + o(|g|).$$

Our objective is to eliminate the terms which are not $O(|\mathbf{p}|)^2$ except those that are constant with respect to $\mathbf{q}$. To apply Newton's method in the standard way, we would need to solve the linear equation

$$\tilde{h}_1 + \{g, \tilde{h}_0\} + \{g, \tilde{h}_1\} \in o(|\mathbf{p}|)$$

But we won't quite do this; we will consider $\{g, \tilde{h}_1\}$ as quadratically small, since g and h_1 are both small. Of course, we can decide to treat anything we want as small; the question is whether the inequalities which come out at the end justify this view. Thus the linear equation we will solve will be

$$\tilde{h}_1 + \{g, \tilde{h}_0\} \in o(|\mathbf{p}|). \tag{2.6}$$

The equation 2.6 is a system of "diophantine partial differential equations;" we will study such equations in Section 5. In the mean time, even to make the statement precise, we need to say exactly what "sufficiently" means in the statements 2.1, 2.2 and 2.3.

3. Norms. One of Kolmogorov's key insights is that using real analytic functions and the associated sup-norms over regions in $\mathbb{C}^n$ substantially simplifies the proofs. Usually, convergence criteria for Newton's method require bounds on the second derivatives (for an elementary treatment of Newton's method, see [3]); the Cauchy inequalities of course give such bounds in terms of the sup-norms for analytic functions, and at heart that is why these norms simplify the proof so much.

If $X \subset \mathbb{C}^k$ is a compact subset, we will use the corresponding script letter $\mathcal{X}$ to denote the Banach algebra of continuous functions on X, analytic in the interior, with the sup-norm

$$\|f\|_X = \sup_{\mathbf{x} \in X} |f(\mathbf{x})|.$$

We will consistently endow $\mathbb{C}^n$ with the Euclidean norm, denoted simply by an absolute value sign. The regions we will be interested in are

$$\begin{aligned} B_\rho &= \{\mathbf{p} \in \mathbb{C}^n \mid |\mathbf{p}| \le \rho\} \\ C_\rho &= \{\mathbf{q} \in \mathbb{C}^n/\mathbb{Z}^n \mid |\operatorname{Im}(\mathbf{q})| \le \rho\} \\ A_\rho &= C_\rho \times B_\rho = \{(\mathbf{q}, \mathbf{p}) \in \mathbb{C}^n/\mathbb{Z}^n \times \mathbb{C}^n \mid |\mathbf{p}| \le \rho,\ |\operatorname{Im}(\mathbf{q})| \le \rho\}. \end{aligned}$$

and the corrresponding Banach algebras $\mathcal{B}_\rho, \mathcal{C}_\rho, \mathcal{A}_\rho$. In these spaces, the sup-norm will be denoted $\|f\|_\rho$. Elements of $\mathcal{B}_\rho$ can be developed in power series, and elements of $\mathcal{C}_\rho$ can be developed in Fourier series

$$f(\mathbf{z}) = \sum_{\mathbf{k} \in \mathbb{Z}^n} f_{\mathbf{k}} e^{2\pi i\, \mathbf{k} \cdot \mathbf{z}}$$

which we will use in an essential way when solving *diophantine partial differential equations.*

For vector-valued functions $\mathbf{f} : X \to \mathbb{C}^n$ (i.e., $\mathbf{f} \in \mathcal{X}^n$), we will use the Euclidean norm in the range, i.e,

$$\|\mathbf{f}\|_X = \sup_X |\mathbf{f}|.$$

This applies also to covectors (i.e., line-matrices) and for matrices we will use the operator-norm associated to the Euclidean norm on the domain and range, still denoted by an absolute value sign.

4. Cauchy's inequalities on balls. Just as in 1 dimension, we can bound derivatives of analytic functions on balls in terms of the values of the function itself.

Proposition 4.3. *If $f \in \mathcal{B}_\rho$, then we have*

$$\|Df\|_{\rho-\delta} \le \frac{1}{\delta}\|f\|_\rho \quad \text{and} \quad \|D^2 f\|_{\rho-\delta} \le \frac{4}{\delta^2}\|f\|_\rho. \tag{4.4}$$

Proof. Take $\mathbf{z} \in B_{\rho-\delta}$, and $\mathbf{u} \in \mathbb{C}^n$. Since the ball of radius δ around $\mathbf{z}$ is contained in B_ρ, the function

$$g : t \mapsto f(\mathbf{z} + t\delta\mathbf{u})$$

is defined on the unit disc , so that the standard Cauchy inequality says that

$$\delta\ |(Df(\mathbf{z}))\mathbf{u}| = |g'(0)| \le \|g\|_1 \le \|f\|_\rho.$$

For the second derivative estimate, apply the argument above twice:

$$|D^2 f(\mathbf{z})(\mathbf{u},\mathbf{v})| \le \frac{2}{\delta}\|Df(\mathbf{z})(\mathbf{u})\|_{\rho-\delta/2}|\mathbf{v}| \le \frac{4}{\delta^2}\|f\|_\rho|\mathbf{u}||\mathbf{v}|.$$

□

The case $\delta = \rho$ bounds derivatives of functions at the center of balls.

Corollary 4.5. *If $f \in \mathcal{B}_\rho$, then*

$$|Df(\mathbf{0})| \le \frac{1}{\rho}\|f\|_\rho \quad \text{and} \quad |D^2 f(\mathbf{0})| \le \frac{4}{\rho^2}\|f\|_\rho.$$

5. Diophantine conditions. The notion of "sufficiently irrational" is absolutely key to Kolmogorov's proof. To motivate it, let us start with diophantine conditions on numbers.

A number θ is called *diophantine of exponent* d if there exists a constant C such that for all coprime integers p, q, we have

$$\left|\theta - \frac{p}{q}\right| > \frac{\gamma}{|q|^d}. \tag{5.1}$$

It is clearly a stronger requirement to be diophantine with a smaller exponent. In fact, since for any irrational θ there exists an arbitrarily large q and p prime to q so that

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{\sqrt{5}|q|^2},$$

we see that no number is diophantine of any exponent smaller than 2. The numbers which are diophantine of exponent 2 are exactly the numbers whose continued fraction have bounded entries; these form a set of measure 0.

Proposition 5.2. *For any $\epsilon > 0$, the set of diophantine numbers of exponent $2+\epsilon$ is of full measure.*

Proof. We will consider our "numbers" to be in $\mathbb{R}/\mathbb{Z}$. For any integer $q \geq 1$, there are at most q elements of $\mathbb{Q}/\mathbb{Z}$ which, in reduced form, have denominator q, and so for any constant γ, the set of numbers $\theta \in \mathbb{R}/\mathbb{Z}$ with

$$\left|\theta - \frac{p}{q}\right| < \frac{\gamma}{|q|^{2+\epsilon}},$$

has total length at most $2\gamma/q^{1+\epsilon}$. Summing this over all q, we see that the set of numbers θ for which there exists q such that

$$\left|\theta - \frac{p}{q}\right| < \frac{\gamma}{|q|^{2+\epsilon}},$$

has total length $< 2\gamma m(\epsilon)$, where $m(\epsilon)$ is the sum of the convergent series

$$m(\epsilon) = \sum_{q=1}^{\infty} 1/q^{1+\epsilon}$$

and in particular, the intersection of these sets as $\gamma \to 0$ has measure 0. But the set of diophantine numbers of exponent $2+\epsilon$ is precisely the complement of this set. $\square$

We will now consider "diophantine" vectors in $\mathbb{R}^n$. Our diophantine condition, the simplest in this setting that corresponds to a set of vectors of full measure, is the set Ω_γ of $\omega \in \mathbb{R}^n$ such that

$$|\mathbf{k} \cdot \omega| > \frac{\gamma}{|\mathbf{k}|^n}$$

for all $\mathbf{k} \in \mathbb{Z}^n - \{0\}$.

Proposition 5.3. *The union*

$$\Omega = \bigcup_{\gamma > 0} \Omega_\gamma$$

is of full measure.

Proof. This is very similar to Proposition 5.2 and reduces to the case $\epsilon = 1$ when $n = 2$. The region $S_{\mathbf{k},\gamma}$ where

$$|\mathbf{k} \cdot \omega| \leq \frac{\gamma}{|\mathbf{k}|^n}$$

is a slab around the hyperplane orthogonal to $\mathbf{k}$, of thickness $2\gamma/|\mathbf{k}|^{n+1}$. The part within the unit cube Q then has measure $\leq M\gamma/|\mathbf{k}|^{n+1}$, where M is the universal constant giving the maximal $(n-1)$-dimensional measure of the intersection of a hyperplane and Q. As above, the sum

$$\sum_{\mathbf{k} \in \mathbb{Z}^n - \{0\}} \frac{1}{|\mathbf{k}|^{n+1}}$$

is finite, so $\operatorname{vol} \cup_{\mathbf{k} \in \mathbb{Z}^n - \{0\}} (S_{\mathbf{k},\gamma} \cap Q) \leq (\mathrm{cst})\gamma$, thus

$$\bigcap_{\gamma > 0} \bigcup_{\mathbf{k} \in \mathbb{Z}^n - \{0\}} S_{\mathbf{k},\gamma} \cap Q$$

has measure 0. Our set Ω is the complement. $\square$

6. A precise statement of Kolmogorov's theorem.

We are finally in a position to state Kolmogorov's theorem exactly.

Theorem 6.1. *Let $\rho, \gamma > 0$ be given, and let $h(\mathbf{q}, \mathbf{p}) = h_0(\mathbf{p}) + h_1(\mathbf{q}, \mathbf{p})$ be a hamiltonian, with $h_0, h_1 \in \mathcal{A}_\rho$ and $\|h\|_\rho \le 1$.*

Suppose the Taylor polynomial of h_0 is

$$h_0(\mathbf{p}) = a + \omega\ \mathbf{p} + \frac{1}{2}\mathbf{p} \cdot C\mathbf{p} + o(|\mathbf{p}|^2),$$

with $\omega \in \Omega_\gamma$ and C is symmetric and invertible.

Then for any $\rho_* < \rho$, there exists $\epsilon > 0$, which depends on C and γ, but not on the remainder term in $o(|\mathbf{p}|^2)$, such that if $\|h_1\|_\rho < \epsilon$, there exists a symplectic mapping $\Phi : A_{\rho_*} \to A_\rho$ such that if we set $(\mathbf{q}, \mathbf{p}) = \Phi(\mathbf{Q}, \mathbf{P})$ and $H = h \circ \Phi$, we have

$$H(\mathbf{Q}, \mathbf{P}) = A + \omega\ \mathbf{P} + R(\mathbf{Q}, \mathbf{P})$$

with $R(\mathbf{Q}, \mathbf{P}) \in O(|\mathbf{P}|^2)$.

In particular, the torus $\mathbf{P} = \mathbf{0}$ is invariant under the flow of $\nabla_\sigma H$, and on this torus the flow is linear with direction ω.

7. Partial differential equations and small divisors.

Let $g \in \mathcal{C}_\rho$, i.e., a function of just $\mathbf{q} \in \mathbb{C}^n/\mathbb{Z}^n$. A key role will be played by the partial differential equation

$$Df(\omega) = \sum_{i=1}^{n} \omega_i \frac{\partial f}{\partial q_i} = g, \tag{7.1}$$

which is to be solved for $f \in \mathcal{C}_{\rho'}$, for an appropriate $\rho' < \rho$.

This equation is easy to solve in formal Fourier series: if

$$f(\mathbf{q}) = \sum_{\mathbf{k} \in \mathbb{Z}^n} f_{\mathbf{k}} e^{2\pi i\ \mathbf{k} \cdot \mathbf{q}}, \quad g(\mathbf{q}) = \sum_{\mathbf{k} \in \mathbb{Z}^n} g_{\mathbf{k}} e^{2\pi i\ \mathbf{k} \cdot \mathbf{q}},$$

then the unique solution to the problem is

$$f_{\mathbf{k}} = \frac{1}{2\pi i\ (\mathbf{k} \cdot \omega)} g_{\mathbf{k}}. \tag{7.2}$$

One thing we see immediately is that for there to be a solution we must have $g_{\mathbf{0}} = 0$, and that $f_{\mathbf{0}}$ is then arbitrary; both these properties will be important when we come to solving such equations in Section 10.

It is clear from this formula that the convergence of the series for F depends crucially on the diophantine properties of ω. If $g \in \mathcal{C}_\rho$ and $\omega \in \Omega$, then the Fourier series for f still converges on the interior of C_ρ, but we can no longer guarantee that f is bounded on C_ρ. But it is bounded on $C_{\rho'}$ for any $\rho' < \rho$, of course, and we need to choose ρ' so as not to lose too much on the radius, and not to lose too much on the norm either.

Proposition 7.3. *If $g \in \mathcal{C}_\rho$ and $\omega \in \Omega_\gamma$, then for all δ with $0 < \delta < \rho$ we have*

$$\|f\|_{\rho-\delta} \le \frac{\kappa_n}{\gamma \delta^{2n}} \|g\|_\rho \quad \textit{and} \quad \|Df\|_{\rho-\delta} \le \frac{\kappa_n}{\gamma \delta^{2n+1}} \|g\|_\rho,$$

where κ_n is a constant which depends only on n.

Proof. For every $\mathbf{y} \in \mathbb{R}^n$ with $|\mathbf{y}| \leq \rho$, the function $\mathbf{q} \mapsto g(\mathbf{q} - i\mathbf{y})$ is a continuous periodic function of $\mathbf{q}$ of period 1, which can be written

$$g(\mathbf{q} - i\mathbf{y}) = \sum_{\mathbf{k}\in\mathbb{Z}^n} g_{\mathbf{k}} e^{2\pi i\mathbf{k}\cdot(\mathbf{q}-i\mathbf{y})} = \sum_{\mathbf{k}\in\mathbb{Z}^n} \big(g_{\mathbf{k}} e^{2\pi\mathbf{k}\cdot\mathbf{y}}\big) e^{2\pi i\mathbf{k}\cdot\mathbf{q}}$$

so Parseval's theorem says

$$\|g\|_\rho^2 \geq \int_{\mathbb{T}^n} |g(\mathbf{q} - i\mathbf{y})|^2 |d^n\mathbf{q}| = \sum_{\mathbf{k}\in\mathbb{Z}^n} |g_{\mathbf{k}}|^2 e^{4\pi\mathbf{k}\cdot\mathbf{y}}.$$

This is true for every $\mathbf{y}$ with $|\mathbf{y}| \leq \rho$, in particular for $\mathbf{y} = \rho\frac{\mathbf{k}}{|\mathbf{k}|}$ for any $\mathbf{k} \in \mathbb{Z}^n - \{0\}$. Since the series above is a series of positive numbers, this gives

$$\|g\|_\rho^2 \geq |g_{\mathbf{k}}|^2 e^{4\pi\rho|\mathbf{k}|}. \tag{7.4}$$

Next, we see what $\omega \in \Omega_\gamma$, together with our formal expression 7.2 for $f_{\mathbf{k}}$ gives.

Lemma 7.5. *We have*

$$|f_{\mathbf{k}}| \leq \frac{1}{2\pi\gamma}\|g\|_\rho |\mathbf{k}|^n e^{-2\pi|\mathbf{k}|\rho}.$$

Proof. This follows immediately from 7.4 and 7.2. □ Lemma 7.5

Now we need to go back to the sup-norm.

Lemma 7.6. *We have*

$$\|f\|_{\rho-\delta} \leq \frac{\kappa_n}{(2\pi\delta)^{2n}} \frac{\|g\|_\rho}{2\pi\gamma}.$$

Proof. We have $f(\mathbf{q}) = \sum f_{\mathbf{k}} e^{2\pi i(\mathbf{k}\cdot\mathbf{q})}$. Thus when $|\mathbf{q}| \leq \rho - \delta$, we have

$$\begin{aligned}\Big|\sum f_{\mathbf{k}} e^{2\pi i(\mathbf{k}\cdot\mathbf{q})}\Big| &\leq \sum_{\mathbf{k}\in\mathbb{Z}^n} \frac{|\mathbf{k}|^n}{2\pi\gamma}\|g\|_\rho e^{-2\pi\rho|\mathbf{k}|} e^{2\pi|\mathbf{k}|(\rho-\delta)} \\ &= \frac{\|g\|_\rho}{2\pi\gamma} \sum_{\mathbf{k}\in\mathbb{Z}^n} |\mathbf{k}|^n e^{-2\pi|\mathbf{k}|\delta}.\end{aligned}$$

The sum can be rewritten

$$\sum_{\mathbf{k}\in\mathbb{Z}^n} |\mathbf{k}|^n e^{-2\pi|\mathbf{k}|\delta} = \frac{1}{(2\pi\delta)^{2n}} \left[(2\pi\delta)^n \sum_{\mathbf{k}'\in 2\pi\delta\mathbb{Z}^n} |\mathbf{k}'|^n e^{-|\mathbf{k}'|}\right],$$

and the term in brackets is a continuous function of $\delta > 0$. It is a Riemann sum, so as $\delta \to 0$ it approaches the convergent integral

$$\int_{\mathbb{R}^n} |\mathbf{x}|^n e^{-|\mathbf{x}|} |d^n\mathbf{x}|.$$

So there exists a constant κ_n' depending only on n such that for $\delta \le 1$, we have

$$\sum_{\mathbf{k}\in\mathbb{Z}^n} |\mathbf{k}|^n e^{-2\pi|\mathbf{k}|\delta} \le \frac{\kappa_n'}{(2\pi\delta)^{2n}}.$$

□ Lemma 7.6

The case of Df is similar, using $Df(\mathbf{q})(\mathbf{u}) = 2\pi i \sum (\mathbf{k}\cdot\mathbf{u}) f_{\mathbf{k}} e^{2\pi i(\mathbf{k}\cdot\mathbf{q})}$. One finds

$$|Df(\mathbf{q})(\mathbf{u})| \le \frac{|\mathbf{u}|\|g\|_\rho \kappa_n''}{\gamma(2\pi\delta)^{n+1}}$$

where κ_n'' exists because the integral

$$\int_{\mathbb{R}^n} |\mathbf{x}|^{n+1} e^{-|\mathbf{x}|} |d^n\mathbf{x}|$$

is convergent. Finally, set

$$\kappa_n = \max\left\{ \frac{\kappa_n'}{(2\pi)^{2n}}, \frac{\kappa_n''}{(2\pi)^{2n+1}} \right\}.$$

□ Proposition 7.3

8. The main iterative step. The next proposition constructs a symplectic change of variables $\phi(\mathbf{q}_1, \mathbf{p}_1) = (\mathbf{q}, \mathbf{p})$, such that $\tilde{h} = h \circ \phi$ is "better" than h: "closer" to the form 2.4. We will construct using Proposition 8.1 a succession of changes of variables $\phi_k(\mathbf{q}_k, \mathbf{p}_k) = (\mathbf{q}_{k-1}, \mathbf{p}_{k-1})$, and hamiltonians $h_k = h_{k-1} \circ \phi_k$; the solution to our problem will be $\Phi = \lim_{k\to\infty} \phi_k \circ \cdots \circ \phi_1$. Of course, the convergence of this sequence is the real issue in the proof.

The construction is very similar to doing one step of Newton's method: we will write the non-linear equation saying that the new hamiltonian is of the form 2.4, linearize the equation and solve it.

If we want to prove the existence of a root of an equation $\mathbf{f}(\mathbf{x}) = 0$ in $\mathbb{R}^n$ using Newton's method, it is enough to show that the Newton map is contracting. This is misleading since ignores the fact that Newton's method *superconverges*; which is why Newton's method is an essential tool of numerical analysis. Here the situation is different: the improved hamiltonian h_k is only defined on some A_{ρ_k} where $\rho_{k+1} = \rho_k - \delta_k$, and the superconvergence is essential to guarantee that the "error" decreases faster than the domain.

Thus our proof will depend on recursive inequalities: "hard analysis" at its hardest. We will have to keep precise track of all the constants, to be sure that the conclusion of the previous step really is the hypothesis of the next step.

In the statement below, we describe our initial h using numbers

- ρ, describing the domain of h,
- γ, describing how irrational the desired flow is,
- m, describing how fast the angle of the flow is changing, and
- ϵ, describing how big the perturbation to be overcome is.

We will also give ourselves numbers $\rho_* < \rho$, $m_* < m$, targets below which we don't want the new ρ and the new m to fall.

For a function $f \in \mathcal{C}_\rho$, we set $\overline{f} = \int_{\mathbb{R}^n/\mathbb{Z}^n} f(\mathbf{q})|d^n\mathbf{q}|$.

Proposition 8.1. *Let the numbers $\rho,\ \rho_*,\ \gamma,\ m,\ m_*,\ \epsilon$ all be in $(0,1)$ and satisfy $\rho_* < \rho,\ m_* < m$. Let $h \in \mathcal{A}_\rho$ satisfy $\|h\|_\rho \le 1$. We will write $h(\mathbf{q},\mathbf{p}) = h_0(\mathbf{q},\mathbf{p}) + h_1(\mathbf{p},\mathbf{q})$, where*

$$\begin{aligned} h_0(\mathbf{q},\mathbf{p}) &= a + \omega\ \mathbf{p} + \frac{1}{2}\mathbf{p}\cdot C(\mathbf{q})\mathbf{p} + R(\mathbf{q},\mathbf{p}) \\ &\qquad \text{with } a \in \mathbb{R},\ \omega \in \Omega_\gamma, \text{ and } R(\mathbf{q},\mathbf{p}) \in O(|\mathbf{p}|^3) \\ h_1(\mathbf{q},\mathbf{p}) &= A(\mathbf{q}) + B(\mathbf{q})\ \mathbf{p} \text{ with } \overline{A} = 0. \end{aligned}$$

Suppose we have the following inequalities:

$$\|A\|_\rho < \epsilon, \quad \|B\|_\rho < \epsilon, \qquad \text{and} \tag{8.2}$$

$$m|\mathbf{v}| < |\overline{C}\mathbf{v}|, \quad \|C\mathbf{v}\|_\rho < \frac{|\mathbf{v}|}{m} \tag{8.3}$$

Choose δ so that $\rho - 3\delta > \rho_*$, and suppose that ϵ is so small that

$$m - \frac{2\eta}{\rho_*^2} > m_*, \quad \text{where} \quad \eta = \frac{10\kappa_n^2}{\gamma^2}\frac{\epsilon}{m^3\delta^{4n+3}}. \tag{8.4}$$

Then there exists a change of variables $\phi : A_{\rho-3\delta} \to A_\rho$, with

$$\|\phi - id\|_{\rho-3\delta} \le \delta/2,$$

such that if we denote all the quantities associated to ϕ^*h by $\tilde{\omega}, \tilde{A}, \tilde{B}, \tilde{C}, \tilde{R}$, then $\tilde{\omega} = \omega$, and

$$\tilde{\rho} = \rho - 3\delta > \rho_*, \quad \tilde{m} = m - \frac{2\eta}{\rho_*^2} > m_*, \quad \tilde{\epsilon} = \frac{\eta^2}{2\rho_*}. \tag{8.5}$$

9. Ending the proof using 8.1. Let us set

$$\epsilon_k = \frac{\epsilon_0}{2^{2\tau k}}, \quad \text{where } \tau = 4n+3, \tag{9.1}$$

$$\delta_k = \frac{\epsilon_0^{\frac{1}{2\tau}}}{2^k}K, \quad \text{where } K = 2\left(\frac{1}{2\rho_*}\right)^{\frac{1}{2\tau}}\left(\frac{10\kappa_n^2}{\gamma^2 m_*^3}\right)^{\frac{1}{\tau}}, \tag{9.2}$$

$$m_{k+1} = m_k - \frac{\sqrt{\epsilon_0}}{2^{k\tau}}L, \quad \text{where } L = \frac{2\sqrt{2}}{2^\tau \rho_*^{3/2}}, \tag{9.3}$$

and ϵ_0 is chosen sufficiently small so that

$$\rho - 3\sum_{k=1}^{\infty}\delta_k > \rho_* \quad \text{and} \quad \lim_{k\to\infty} m_k \ge m_*. \tag{9.4}$$

To justify these equations, define

$$\eta_k = \frac{10\kappa_n^2}{\gamma^2}\frac{\epsilon_k}{m_k^3\delta_k^\tau} \quad \text{and} \quad \eta_{k,*} = \frac{10\kappa_n^2}{\gamma^2}\frac{\epsilon_k}{m_*^3\delta_k^\tau}.$$

The formula 9.1 for the epsilon's is simply parachuted. The formula 9.2 for the delta's comes from solving the equation

$$\epsilon_{k+1} = \frac{\eta_{k,*}^2}{2\rho_*} = \frac{1}{2\rho_*}\left(\frac{10\kappa_n^2}{\gamma^2}\frac{\epsilon_k}{m_*^3\delta_k^\tau}\right)^2$$

for δ_k. The formula 9.3 for the m's comes from setting

$$m_{k+1} = m_k - \frac{2\eta_{k,*}}{\rho_*^2} = m_k - \frac{2\sqrt{2\epsilon_{k+1}\rho_*}}{\rho_*^2}.$$

The second of the requirements 9.4 guarantees that $m_k > m_*$ and thus that $\eta_{k,*} > \eta_k$ for all k.

Now to prove Theorem 6.1 using Proposition 8.1. First set $\rho_0 = \rho$ and $\rho_{i+1} = \rho_i - 3\delta_i$, and recursively define a sequence of hamiltonian functions $f_i : A_{\rho_i} \to \mathbb{R}$ and changes of variables $\phi_i : A_{\rho_{i+1}} \to A_{\rho_i}$ as follows. Set $f_0 = h$, and suppose that for all $i \leq k$ we have defined f_i and ϕ_i, satisfying the hypotheses 8.2 and 8.3 with $\epsilon = \epsilon_i$ and $m = m_i$.

With $\delta = \delta_k$, the requirement 8.4 is satisfied:

$$m_k - \frac{2\eta_k}{\rho_*^2} > m_k - \frac{2\eta_{k,*}}{\rho_*^2} = m_{k+1} > m_*,$$

so there exists a symplectic mapping $\phi_{k+1} : A_{\rho_{k+1}} \to A_{\rho_k}$ such that

$$\|\phi_{k+1} - \mathrm{id}\|_{\rho_{k+1}} \leq \frac{\delta_k}{2},$$

and that if we set $f_{k+1} = \phi^* f_k$, and denote all the corresponding quantities with tilde's, then

$$\tilde{m} = m - \frac{2\eta_k}{\rho_*^2} > m - \frac{2\eta_{k,*}}{\rho_*^2} = m_k > m_*, \quad \tilde{\epsilon} = \frac{\eta_k^2}{2\rho_*} < \frac{\eta_{k,*}^2}{2\rho_*} = \epsilon_{k+1}.$$

Now define

$$\psi_k = \phi_k \circ \phi_{k-1} \circ \cdots \circ \phi_1.$$

Clearly ψ_k maps $A_{\rho_k} \to A_\rho$. In particular, all ψ_k are defined in A_{ρ_*}. Moreover, the sequence of ψ_k converges to a symplectic mapping $\Psi : A_{\rho_*} \to A_\rho$, since

$$\|\psi_{k+1} - \psi_k\|_{\rho_*} \leq \|\phi_{k+1} \circ \psi_k - \psi_k\|_{\rho_*} \leq \|\phi_{k+1} - \mathrm{id}\|_{\rho_*} \leq \frac{\delta_k}{2},$$

and thus the sequence (ψ_k) converges.

Now consider the decomposition of $H = \Psi^* h \in \mathcal{A}_{\rho_*}$ as $H = H_0 + H_1$, where H_1 consists of the constant and linear terms with respecto to $\mathbf{P}$, except for those which are constant with respect to $\mathbf{Q}$. We have $H = \lim_{k\to\infty} f_k$, and in particular $H_1 = \lim_{k\to\infty} f_{k,1}$. Thus

$$\|H_1\|_{\rho_*} = \lim_{k\to\infty} \|f_{1,k}\|_{\rho_*} \leq \lim_{k\to\infty} \epsilon_k = 0,$$

so $H = H_0$. $\square$

10. Proving Proposition 8.1. In this section we will prove Proposition 8.1. This is quite lengthy, so we first present the strategy.

Strategy. The idea is to write the required symplectic diffeomorphism ϕ as the time one hamiltonian flow ϕ_g, for some function g which is the unknown for which we will solve. Develop everything in Taylor polynomial with respect to $\mathbf{p}$, and isolate the terms that are linear with respect to g.

By formula 1.6, the pullback $\phi_g^* h$ has the Taylor expansion

$$\phi_g^* h = h + \{g, h\} + O(|g|^2) = h_0 + h_1 + \{g, h_0\} + \{g, h_1\} + O(|g|^2).$$

In the subsection below, we will solve the linearized equation

$$h_1 + \{g, h_0\} \in O(|\mathbf{p}|^2) \tag{10.1}$$

and in particular we will see that the solution is small approximately of the same order as h_1. The cumbersome terms of $\phi_g^* h$, i.e., those which prevent the torus $\mathbf{p} = 0$ from being invariant with linear flow, are the terms $\{g, h_1\} + O(|g|^2)$, and we see that these are now approximately quadratic with respect to h_1. Hence at the beginning of the next step the cumbersome terms are quadratic with respect to the previous ones. This provides the superconvergence of the successive coordinate changes.

Thus, we write

$$g = \lambda\ \mathbf{q} + X(\mathbf{q}) + \sum_{i=1}^{n} Y_i(\mathbf{q}) p_i;$$

there is no sense in developing further, since only the linear terms of g with respect to $\mathbf{p}$ can contribute to the linear terms of $\phi_g^* h$.

Remark. Notice that because of the term $\lambda \cdot \mathbf{q}$, the function g is defined on $\mathbb{R}^n \times \mathbb{R}^n$, not $\mathbb{T}^n \times \mathbb{R}^n$, though the flow of $\nabla_\sigma g$ is perfectly well defined on $\mathbb{T}^n \times \mathbb{R}^n$. This is clearly necessary. For instance, set $n = 1$, and consider the family of hamiltonians

$$h_s(p) = \frac{1}{2}(p + s)^2,$$

whose associated motion is to rotate at speed $p_0 + s$ in the circle $p = p_0$. Think of $\mathbb{T} \times \mathbb{R}$ as a vertical cylinder. Evidently vertical translation by s transforms h_s into h_0. But vertical translation by s is the flow at time s of the "hamiltonian" $g(q, p) = -q$, which is only defined on $\mathbb{R} \times \mathbb{R}$, not on $\mathbb{T} \times \mathbb{R}$, since the q on the right is a real variable, not an angular variable periodic of period 1. The freedom to choose λ will be essential for our purposes. This also explains why the term $\lambda \cdot \mathbf{q}$ cannot be incorporated in $X(\mathbf{q})$.

Computing g using diophantine PDE's. We are planning to make $h_0 + h_1 + \{g, h_0\}$ be of the form $a + \omega_0 \cdot \mathbf{P} + R(\mathbf{Q}, \mathbf{P})$ with $R(\mathbf{Q}, \mathbf{P}) \in O(|\mathbf{P}|^2)$ as in Equation 2.4. The function h_0 is already of that form, so we need to bring $h_1 + \{g, h_0\}$ to that form. Expanding out using 1.5, we find

$$\begin{aligned}(h_1 + \{g, h_0\})(\mathbf{q}, \mathbf{p}) &= \omega \cdot \lambda + A(\mathbf{q}) + DX(\mathbf{q})(\omega) \\ &\quad + \Big(B(\mathbf{q}) + \big(\lambda + DX(\mathbf{q})\big)C(\mathbf{q}) + \omega DY(\mathbf{q})\Big) \cdot \mathbf{p} + O(|\mathbf{p}|^2).\end{aligned}$$

To solve Equation 10.1, we need to solve the system of linear equations for X and Y

$$DX(\mathbf{q})(\omega) = -A(\mathbf{q}) \tag{10.2}$$
$$DY(\mathbf{q})(\omega) = -B(\mathbf{q}) - \big(\lambda + DX(\mathbf{q})\big)C(\mathbf{q}). \tag{10.3}$$

The first equation can be rewritten

$$\sum_i \omega_i \frac{\partial X}{\partial q_i} = -A(\mathbf{q}), \tag{10.5}$$

and the second is really n different equations (an equation for a line-matrix):

$$\sum_i \omega_i \frac{\partial Y_j}{\partial q_i} = -B_j(\mathbf{q}) - \sum_i C_{i,j}(\mathbf{q}) \left(\lambda_i + \frac{\partial X}{\partial q_i}\right). \tag{10.6}$$

so all these equations are of the form studied in 7.1. We will solve 10.2 first, then 10.3.

Recall that $\overline{A} = 0$, so we can find a unique X with $\overline{X} = 0$ and satisfying the estimates

$$\|X\|_{\rho-\delta} \le \frac{\kappa_n \epsilon}{\gamma \delta^{2n}} \quad \text{and} \quad \|DX\|_{\rho-\delta} \le \frac{\kappa_n \epsilon}{\gamma \delta^{2n+1}} \tag{10.8}$$

for all δ satisfying $0 < \delta < \rho$.

Now for the second lot. We can't solve that unless we make the average of the right hand side 0, which we accomplish by setting

$$\lambda = -\overline{C}^{-1}\overline{(B + (DX)C)}.$$

We have

$$\|B - (DX)C\|_{\rho-\delta} \le \epsilon + \frac{\kappa_n \epsilon}{m\gamma\delta^{2n+1}} \le \frac{2\kappa_n \epsilon}{m\gamma\delta^{2n+1}}, \tag{10.9}$$

which gives

$$|\lambda| \le \frac{2\kappa_n \epsilon}{m^2\gamma\rho^{2n+1}} \le \frac{2\kappa_n \epsilon}{m^2\gamma\delta^{2n+1}}. \tag{10.10}$$

This gives

$$\begin{aligned} \|B - (\lambda + DX)C\|_{\rho-\delta} &\le \|B - (DX)C\|_{\rho-\delta} + \|\lambda C\|_{\rho-\delta} \\ &\le \frac{2\kappa_n \epsilon}{m\gamma\delta^{2n+1}} + \frac{2\kappa_n \epsilon}{m^3\gamma\delta^{2n+1}} \le \frac{4\kappa_n \epsilon}{m^3\gamma\delta^{2n+1}}. \end{aligned} \tag{10.11}$$

Now applying 7.3 again, we can find $Y \in \mathcal{C}^n_{\rho-2\delta}$ with $\overline{Y} = 0$ satisfying the estimates

$$\|Y\|_{\rho-2\delta} \le \frac{4\kappa_n^2 \epsilon}{m^3\gamma^2\delta^{4n+1}} \quad \text{and} \quad \|DY\|_{\rho-2\delta} \le \frac{4\kappa_n^2 \epsilon}{m^3\gamma^2\delta^{4n+2}}. \tag{10.12}$$

Bounding the vector field $\nabla_\sigma g$. We now have our hamiltonian g, together with the estimate

$$\begin{aligned} \|\nabla_\sigma g\|_{\rho-2\delta} &\le \left(\left\| \frac{\partial g}{\partial \mathbf{q}} \right\|^2_{\rho-2\delta} + \left\| \frac{\partial g}{\partial \mathbf{p}} \right\|^2_{\rho-2\delta} \right)^{1/2} \\ &\le \left(\|Y\|^2_{\rho-2\delta} + \|\lambda + DX + (DY)\mathbf{p}\|^2_{\rho-2\delta} \right)^{1/2} \\ &\le \left(\left(\frac{4\kappa_n^2 \epsilon}{m^3\gamma\delta^{4n+1}} \right)^2 + \left(\frac{2\kappa_n \epsilon}{m^3\gamma\delta^{2n+1}} + \frac{\kappa_n \epsilon}{\gamma\delta^{2n+1}} + \frac{4\kappa_n^2 \epsilon}{m^3\gamma^2\delta^{4n+2}} \right)^2 \right)^{1/2} \\ &\le \frac{10\kappa_n^2 \epsilon}{m^3\gamma^2\delta^{4n+2}} = \eta\delta \le \frac{\delta}{2}. \end{aligned}$$

Bounding the new hamiltonian. The time one flow map $\phi_g : A_{\rho-3\delta} \to A_\rho$ is well defined, in fact the image lies in $A_{\rho-5\delta/2}$. Of course, we get immediately

$$|\phi_g - id\|_{\tilde\rho} \le \frac{\delta}{2} \quad \text{and} \quad \|\phi_g^* h\|_{\rho-3\delta} \le \|h\|_\rho \le 1.$$

In order to prove 8.5, we need to estimate $\phi^* h - h_0$. We have the identity

$$\phi^* h = h_0 + \overbrace{h_1 + \{g, h_0\}}^{-\omega\cdot\lambda + O(|\mathbf{p}|^2)} + \big[\{g, h_1\} + \phi^* h - h - \{g, h\}\big] = h_0 - \omega\cdot\lambda + O(|\mathbf{p}|^2) + \hat h_1,$$

where by definition $\hat h_1$ is the expression in brackets. This function $\hat h_1$ is almost but not quite $\tilde h_1$; in particular, it contributes all the troublesome terms $\tilde A + \tilde B\mathbf{p}$; more specifically,

$$\tilde A(\mathbf{q}) = \hat h_1(\mathbf{q}, \mathbf{0}) - \overline{\hat h_1(\mathbf{q}, \mathbf{0})} \quad \text{and} \quad \tilde B(\mathbf{q}) = \frac{\partial \hat h_1}{\partial \mathbf{p}}(\mathbf{q}, \mathbf{0}).$$

So we need to bound $\hat h_1$. The case of $\{g, h_1\}$ is staightforward; we are bounding a derivative after restricting by 3δ, so we get

$$\|\{g, h_1\}\|_{\rho-3\delta} = \|dh_1(\nabla_\sigma g)\|_{\rho-3\delta} \le \frac{\epsilon}{3\delta}\|\nabla_\sigma g\|_{\rho-2\delta} \le \frac{\epsilon\eta}{3}$$

Now for the other term of $\hat h_1$:

$$\begin{aligned}\|\phi_g^* h - h - \{g, h\}\|_{\rho-3\delta} &\le \frac{1}{2}\sup_{0\le t\le 1}\frac{d^2}{dt^2}|(\phi_g^t)^* h| \le \frac{1}{2}\|\{\{h, g\}, g\}\|_{\rho-5\delta/2} \\ &\le \frac{1}{2}\|D^2 h(\nabla_\sigma g, \nabla_\sigma g)\| \le \frac{4}{2}\left(\frac{2}{5\delta}\right)^2 (\eta\delta)^2 < \frac{\eta^2}{3}.\end{aligned}$$

Using $\epsilon < \eta/2$ (it is really much smaller than that), we get

$$\|\hat h\|_{\tilde\rho} < \frac{\eta^2}{6} + \frac{\eta^2}{3} = \frac{\eta^2}{2}$$

where $\tilde\rho = \rho - 3\delta$.

This now gives

$$\|\tilde A\|_{\tilde\rho} = \|\hat h_1(\mathbf{q}, \mathbf{0}) - \overline{\hat h_1(\mathbf{q}, \mathbf{0})}\|_{\tilde\rho} \le \eta^2.$$

Similarly, $\|\tilde B\|_{\tilde\rho}$ is bounded (using the Cauchy estimates of Corollary 4.5; note that we are estimating the derivative of $\hat h_1$ at the center of a ball of radius $\tilde\rho > \rho_*$):

$$\|\tilde B\|_{\tilde\rho} = \|[D\hat h](\mathbf{q}, \mathbf{0})\| \le \frac{\eta^2}{2\rho_*}.$$

Bounding the $\tilde C$ below and above. Next, we attack $h - \tilde h = h - h\circ\phi$. Again this is a form of Taylor's theorem, using Equation (1.6):

$$\begin{aligned}\|h - \tilde h\|_{\tilde\rho} &\le \sup_{0\le t\le 1}\left|\frac{d}{dt}(h\circ\phi_g^t)\right| \le \|\{g, h\}\|_{\rho-5\delta/2} \\ &\le \|Dh(\nabla_\sigma g)\|_{\rho-2\delta} \le \frac{1}{2\delta}\|\nabla_\sigma g\|_{\rho-2\delta} \le \frac{\eta\delta}{2\delta} = \frac{\eta}{2}.\end{aligned}$$

This now allows us to estimate $\tilde{C}(\mathbf{q}) - C(\mathbf{q}) = D^2(\tilde{h} - h)(\mathbf{q}, \mathbf{0})$. Again we are evaluating the second derivative of a function in the center of a ball, and can apply Corollary 4.5, to find

$$\|\tilde{C}(\mathbf{q}) - C(\mathbf{q})\| \leq \frac{4}{\tilde{\rho}^2}\frac{\eta}{2} \leq \frac{2\eta}{\rho_*^2}.$$

This yields

$$|\overline{\tilde{C}}| \geq |\overline{C}| - |\overline{(\tilde{C} - C)}| \geq m - \frac{2\eta}{\rho_*^2},$$

and

$$|\tilde{C}| \leq |C| + |\tilde{C} - C| \leq \frac{1}{m} + \frac{2\eta}{\rho_*^2} \leq \frac{1}{m - \frac{2\eta}{\rho_*^2}},$$

using that if $0 < b < a < 1$ (in our case, $0 < 2\eta/\rho_*^2 < m < 1$), then

$$\frac{1}{a} + b \leq \frac{1}{a - b}.$$

Thus we find that we can take

$$\tilde{\rho} = \rho - 3\delta, \;\; \tilde{\epsilon} = \frac{\eta^2}{2\rho_*}, \quad \text{and} \quad \tilde{m} = m - \frac{2\eta}{\rho_*^2}. \quad \square$$

Appendix: Filling in the crash course

A 1. The symplectic structure of the cotangent bundle. Almost all symplectic manifolds which come up in practice are cotangent bundles of some other manifold: such cotangent bundles have a natural symplectic structure. In fact, they carry a cononical 1-form ω defined as follows. Let $\pi : T^*M \to M$ be the canonical projection, and let $\alpha \in T^*_{\mathbf{x}}M$ be a point of T^*M with $\pi(\alpha) = \mathbf{x}$. Then if $\xi \in T_\alpha(T^*M)$ (a space which is rather hard to think about), we define

$$\omega(\xi) = \alpha([D\pi(\alpha)]\xi).$$

Let us bring this definition down to earth. Let $q_1, \dots, q_n$ be local coordinates on a subset $U \subset M$, so that $\mathbf{q} : U \to V$ is a diffeomorphism, where V is an open subset of $\mathbb{R}^n$. Any point $\alpha \in T^*M$ can be written $(\mathbf{q}, \sum_i p_i \, dq_i)$, and $q_1, \dots, q_n, p_1, \dots, p_n$ are local coordinates on T^*U, i.e., together define a diffeomorphism $(\mathbf{q}, \mathbf{p}) : T^*U \to V \times \mathbb{R}^n = T^*V$. The coordinates p_i are called the *conjugates* of the coordinates q_i, and are clearly the coefficients of the dual basis of the $\partial/\partial q_i$; together the q_i, p_i are called *canonical coordinates*. With respect to canonical coordinates, we have $\omega = \sum_i p_i \, dq_i$.

The form ω is not closed: we define $\sigma = d\omega$. Clearly in canonical coordinates we have

$$\sigma = \sum dp_i \wedge dq_i.$$

Then (T^*M, σ) is a symplectic manifold.

A 2. Lagrangian submanifolds. Later in section A 6, we will need Lagrangian submanifolds. If (X, σ) is a symplectic manifold of dimension $2n$, then an n-dimensional submanifold is called *Lagrangian* if $\sigma|_Y = 0$. To understand the main example we use the following notation. If ϕ is a 1-form on a manifold M, then ϕ is a section of $\pi : T^*M \to M$; we will write the section as $\tilde{\phi} : M \to T^*M$; of course, ϕ and $\tilde{\phi}$ are just two ways of thinking of the same thing. Still, the following lemma shows why the notation might be helpful.

Lemma A 2.1. *If ϕ is a 1-form on a manifold M, then $\tilde{\phi}^*\omega = \phi$.*

Think about it this way: $\tilde{\phi}^*\omega$ is a 1-form on M which depends only on ϕ. What else could it possibly be?

Proof. We have

$$\tilde{\phi}^*(\xi) = \omega(D\tilde{\phi}(\xi)) = \phi(\xi).$$

□

Proposition A 2.2. *Let M be a manifold, and ϕ be a 1-form on M. Then the image of $\tilde{\phi}$ is a Lagrangian submanifold of T^*M if and only if $d\phi = 0$.*

Proof. This is just A 2.1 and the naturality of the exterior derivative: Saying that the image of $\tilde{\phi}$ is Lagrangian is precisely saying that $\tilde{\phi}^*\sigma = 0$, but

$$\tilde{\phi}^*\sigma = \tilde{\phi}^*d\omega = d(\tilde{\phi}^*\omega) = d\phi.$$

□

A 3. Poisson and Lie brackets. We want to check that for any functions f, g on a symplectic manifold, we have

$$\nabla_\sigma\{f,g\} = [\nabla_\sigma f, \nabla_\sigma g].$$

This requires as a preliminary the Jacobi identity for Poisson brackets.

Proposition A 3.1. *If f, g, h are functions on a symplectic manifold, then*

$$\{\{f,g\},h\} + \{\{g,h\},f\} + \{\{h,f\},g\} = 0.$$

Proof. This is a fairly complicated computation, which uses $d\sigma = 0$ in an essential way. It is easiest to use the intrinsic formula for the exterior derivative of a k-form ϕ:

$$\begin{aligned} d\phi(\xi_1,\dots,\xi_{k+1}) = &\sum_{i=1}^{n+1}(-1)^{i+1} d\phi\big(\xi_1,\dots,\hat{\xi}_i,\dots,\xi_{n+1}\big)(\xi_i) \\ &+ \sum_{1\le i<j\le n+1}(-1)^{i+j}\phi\big([\xi_i,\xi_j],\xi_2,\dots,\hat{\xi}_i,\dots,\hat{\xi}_j,\dots,\xi_{n+1}\big). \end{aligned}$$

Thus

$$\begin{aligned} 0 &= d\sigma(\nabla_\sigma f, \nabla_\sigma g, \nabla_\sigma h) \\ &= d\big(\sigma(\nabla_\sigma g, \nabla_\sigma h)\big)\nabla_\sigma f - d\big(\sigma(\nabla_\sigma f, \nabla_\sigma h)\big)\nabla_\sigma g + d\big(\sigma(\nabla_\sigma f, \nabla_\sigma g)\big)\nabla_\sigma h \\ &\qquad - \sigma\big([\nabla_\sigma f, \nabla_\sigma g], \nabla_\sigma h\big) + \sigma\big([\nabla_\sigma f, \nabla_\sigma h], \nabla_\sigma g\big) - \sigma\big([\nabla_\sigma g, \nabla_\sigma h], \nabla_\sigma f\big) \\ &= d\big(\{h,g\}\big)\nabla_\sigma f - d\big(\{h,f\}\big)\nabla_\sigma g + d\big(\{g,f\}\big)\nabla_\sigma h \\ &\qquad - d\big(dh(\nabla_\sigma g)\big)\nabla_\sigma f + d\big(dh(\nabla_\sigma f)\big)\nabla_\sigma g \\ &\qquad\qquad + d\big(dg(\nabla_\sigma h)\big)\nabla_\sigma f - d\big(dg(\nabla_\sigma f)\big)\nabla_\sigma h \\ &\qquad\qquad\qquad - d\big(df(\nabla_\sigma h)\big)\nabla_\sigma g + d\big(df(\nabla_\sigma g)\big)\nabla_\sigma h \\ &= \big\{\{h,g\},f\big\} - \big\{\{h,f\},g\big\} + \big\{\{g,f\},h\big\} \\ &\qquad - \big\{\{h,g\},f\big\} + \big\{\{h,f\},g\big\} \\ &\qquad\qquad + \big\{\{g,h\},f\big\} - \big\{\{g,f\},h\big\} \\ &\qquad\qquad\qquad - \big\{\{f,h\},g\big\} + \big\{\{f,g\},h\big\} \\ &= \{\{f,g\},h\} + \{\{g,h\},f\} + \{\{h,f\},g\}. \qquad \square \end{aligned}$$

The relation of the Poisson bracket and the Lie bracket is now straightforward.

Proposition A 3.2. *For any functions f, g on a symplectic manifold, we have*

$$\nabla_\sigma\{f,g\} = [\nabla_\sigma f, \nabla_\sigma g].$$

Given a vector field ξ and a function h, the function $dh(\xi)$ is just a partial derivative. Sometimes we find it easier to think in those terms, and write $dh(\xi) = \partial_\xi h$.

Proof. Given some third function h, we need to compute:

$$\begin{aligned} dh([\nabla_\sigma f, \nabla_\sigma g]) &= \partial_{[\nabla_\sigma f, \nabla_\sigma g]} h = \partial_{\nabla_\sigma f}(\partial_{\nabla_\sigma g} h) - \partial_{\nabla_\sigma g}(\partial_{\nabla_\sigma f} h) \\ &= \partial_{\nabla_\sigma f}\{h,g\} - \partial_{\nabla_\sigma g}\{hf\} = d\{h,g\}(\nabla_\sigma f) - d\{h,f\}(\nabla_\sigma g) \\ &= \{\{h,g\},f\} - \{\{h,f\},g\} = \{\{h,g\},f\} + \{\{f,h\},g\} \\ &= \{\{f,g\},h\} = dh(\nabla_\sigma\{f,g\}). \end{aligned}$$

The second equality is the definition of the Lie bracket, the next to last is the Jacobi identity, all the others are the equivalent forms of the Poisson bracket. $\square$

A 4.1.

Let $\phi_f^t : M \to M$ be the flow at time t of the vector-field $\nabla_\sigma f$.

Proposition A 4.1. *We have* $(\phi_f^t)^*\sigma = \sigma$.

Remark. It is perfectly possible to prove this by differential calculus, using the Lie derivative $L_{\nabla_\sigma f}\sigma$; we find the integral form closer to the intuition.

Proof. Clearly if a 2-form ψ on M has integral $\int_S \psi = 0$ for all embedded closed discs $S \subset M$, then $\psi = 0$. Let S be such an embedded disc, set $I = [0, a]$ and consider the map $F : S \times I \to M$ given by $F(x,t) = \phi_f^t(x)$. By Stokes theorem we have

$$\int_{\partial(S\times I)} F^*\sigma = \int_{S\times I} d(F^*\sigma) = \int_{S\times I} F^*(d\sigma) = 0.$$

Thus

$$\int_{S\times\{0\}} F^*\sigma - \int_{S\times\{a\}} F^*\sigma + \int_{(\partial S)\times I} F^*\sigma = \int_S \sigma - \int_{\phi_f^a S} \sigma + \int_{(\partial S)\times I} F^*\sigma = 0.$$

We need to see that the last term vanishes. Let $\gamma : J \to M$ be a parametrization of the simple closed curve ∂S, which gives the boundary orientation. We can rewrite

$$\begin{aligned}\int_{(\partial S)\times I} F^*\sigma &= \int_0^a \left(\int_J \sigma((\phi_f^t \circ \gamma)'(s), \nabla_\sigma f) ds\right) dt \\ &= \int_0^a \left(\int_J df(\phi_f^t \circ \gamma)'(s) ds\right) dt = \int_0^a (0)\, dt = 0.\end{aligned}$$

This last is because we are integrating df around a closed curve.

Thus

$$0 = \int_S \sigma - \int_{\phi_f^a S} \sigma = \int_S (\sigma - (\phi_f^a)^*\sigma),$$

and so $\sigma = (\phi_f^a)^*\sigma$ for all a. $\square$

A 5. Darboux's theorem. Riemannian manifolds have lots of *local geometry*. A piece of a sphere is not locally isometric to a piece of a plane or a piece of a hyperboloid. All the various *curvatures* (Ricci, sectional, total, ...) are particular local invariants of the geometry. It comes as a surprise at first that symplectic forms have *no local invariants*: all symplectic manifolds of dimension $2n$ are locally symplectomorphic. This result, due to Darboux, is not strictly necessary for our purposes, but it is conceptually important, and helps to justify computations in local coordinates. The proof given is essentially that in [2].

Theorem A 5.1. *Let* (X, σ) *be a symplectic manifold of dimension* $2n$, *and* $\mathbf{x} \in X$ *a point in* X. *Then there exists an open neighborhood* $U \subset X$ *of* $\mathbf{x}$ *and a diffeomorphism* $\Psi : U \to \mathbb{R}^{2n}$ *with* $\Psi(\mathbf{x}) = \mathbf{0}$ *such that* $\Psi^*(\sum_i dq_i \wedge dp_i) = \sigma$.

Proof. Choose any function q_1 on a neighborhood U_1 of $\mathbf{x}$ with $dq_1(\mathbf{x}) \neq 0$. Next, define a function p_1 as follows: choose a smooth hypersurface $M \subset U_1$ through $\mathbf{x}$ so that $T_\mathbf{x}X = T_\mathbf{x}M \oplus \mathbb{R}\nabla_\sigma q_1$. Consider the map $\Phi : M \times \mathbb{R} \to X$ given by $(\mathbf{m}, t) \mapsto \phi_{q_1}^t(\mathbf{m})$. This is a local diffeomorphism at $\mathbf{x} \times \{0\}$ by the inverse function theorem; let $\Psi : U_2 \to M \times \mathbb{R}$ be the inverse, defined on an appropriate neighborhood $U_2 \subset U_1$ of $\mathbf{x}$. Now define $p_1 : U_2 \to \mathbb{R}$ to be the composition $pr_2 \circ \Psi$. Note that

$$\{p_1, q_1\} = dp_1(\nabla_\sigma q_1) = 1,$$

in particular $[\nabla_\sigma p_1, \nabla_\sigma q_1] = 0$, so the flows of $\nabla_\sigma p_1$ and $\nabla_\sigma q_1$ commute.

If $n = 1$ we are done; otherwise the other coordinate functions are constructed by induction. Consider the subset $X \subset U_2$ given by $q_1 = p_1 = 0$. By the implicit function theorem, a neighborhood $X' \subset X$ is a manifold of dimension $2(n-1)$, and by choosing X' sufficiently small we may assume that the restriction of σ to X' is non-degenerate. Indeed, it is enough to show that the restriction of σ to $T_\mathbf{x}X$ is non-degenerate. Set $\mathbf{w}_1 = \nabla_\sigma q_1,\ \mathbf{w}_1 = \nabla_\sigma p_1$, and choose any basis $\mathbf{w}_3, \dots, \mathbf{w}_{2n}$ of $T_\mathbf{x}X'$. Then

$$\sigma(\mathbf{w}_i, \nabla_\sigma p_1) = dp_1(\mathbf{w}_i) = 0, \quad \sigma(\mathbf{w}_i, \nabla_\sigma q_1) = dq_1(\mathbf{w}_i) = 0 \quad \text{for all } i \geq 3,$$

and it follows that if we set A to be the matrix $A = (\sigma(\mathbf{w}_i, \mathbf{w}_j))_{1 \leq i,j \leq 2n}$ and A', A'' the principal minors formed of the first two lines and columns, and the last $2n-2$ respectively, then $0 \neq \det A = \det A' \det A''$, and in particular $\det A'' \neq 0$.

By induction we may assume that there is a neighborhood V of $\mathbf{x}$ in X' and coordinates

$$\tilde{q}_2, \tilde{p}_2, \dots, \tilde{q}_n, \tilde{p}_n : V \to \mathbb{R}$$

such that $\sigma|_V = \sum_{i=2}^n d\tilde{q}_i \wedge d\tilde{p}_i$.

Consider the mapping $\Phi' : V \times \mathbb{R}^2$ given by

$$\Phi'(\mathbf{y}, s, t) = \phi^s_{\nabla_\sigma q_1}\left(\phi^t_{\nabla_\sigma p_1}(\mathbf{y})\right) = \phi^t_{\nabla_\sigma p_1}\left(\phi^s_{\nabla_\sigma q_1}(\mathbf{y})\right).$$

Again by the inverse function theorem Φ' is a local diffeomorphism, so there exists a neighborhood U_3 of $\mathbf{x}$ in X and an inverse $\Psi' : U_3 \to V \times \mathbb{R}^2$ which is a diffeomorphism onto its image. Now set

$$q_i = \tilde{q}_i \circ pr_1 \circ \Psi' \quad \text{and} \quad p_i = \tilde{p}_i \circ pr_1 \circ \Psi'.$$

This gives us our local coordinates: we still need to show that $\sigma = \sum_{i=1}^n dq_i \wedge dp_i$. Since $\{f, g\} = \sigma(\nabla_\sigma f, \nabla_\sigma g)$, this is equivalent to showing that $\{q_i, p_j\} = \delta_{i,j}$, and that all other Poisson brackets are 0.

There are several cases to consider. First, if $i, j \geq 2$, it is enough to verify the condition on V, since the functions p_i, q_i are by definition invariant under the flow $\phi^t_{\nabla_\sigma p_1} \circ \phi^s_{\nabla_\sigma q_1}$, and this flow is hamiltonian. On V, the Poisson brackets are what is required by the inductive hypothesis.

We have already computed $\{q_1, p_1\} = 1$, and of course $\{q_1, q_1\} = \{p_1, p_1\} = 0$. So we need to show that for $i > 1$ we have $\{p_1, p_i\} = \{p_1, q_i\} = \{q_1, p_i\} = \{q_1, q_i\} = 0$. Again, since p_i and q_i are invariant under the flow $\phi^t_{\nabla_\sigma p_1} \circ \phi^s_{\nabla_\sigma q_1}$, we have

$$\{p_i, q_1\} = dp_i(\nabla_\sigma q_1) = 0, \quad \{q_i, q_1\} = dq_i(\nabla_\sigma q_1) = 0 \quad \text{for } i = 2, \dots, n.$$

But since $\phi^t_{\nabla_\sigma p_1} \circ \phi^s_{\nabla_\sigma q_1} = \phi^s_{\nabla_\sigma q_1} \circ \phi^t_{\nabla_\sigma p_1}$, we also have

$$\{p_i, p_1\} = dp_i(\nabla_\sigma p_1) = 0, \quad \{q_i, p_1\} = dq_i(\nabla_\sigma p_1) = 0 \quad \text{for } i = 2, \dots, n.$$

□

A 6. Liouville's theorem.

A Hamiltonian system (X, σ, H) will be called a *standard completely integrable system* if $X = T^*(\mathbb{T}^n)$ is the cotangent bundle of the torus with its canonical symplectic structure, and the Hamiltonian H depends only on the variables $p_1, \dots, p_n$ conjugate to the canonical variabls $q_1, \dots, q_n \in \mathbb{R}/\mathbb{Z}$ on $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$.

It is then very easy to integrate the equations of motion

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}} \stackrel{\text{def}}{=} \omega(\mathbf{p}), \quad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}} = 0.$$

We find

$$\mathbf{p}(t) = \mathbf{p}(0) = \mathbf{p}_0, \quad \text{and} \quad \mathbf{q}(t) = \mathbf{q}(0) + t\omega(\mathbf{p}_0).$$

This is actually not quite so simple as the formula makes it look: the formula for $\mathbf{q}(t)$ corresponds to linear motion on the torus $\mathbb{T}^n$, such as an irrational flow when $n = 2$; the trajectory may be periodic, or dense in $\mathbb{T}^n$, or dense in a subtorus, if all, or none, or some of the ratios $\omega_i(\mathbf{p})/\omega_j(\mathbf{p})$ are rational; just how they fill up the torus depends in a delicate way on the diophantine properties of these ratios.

The object of this section is to prove the following theorem.

Theorem A 6.1. *Let (M, σ) be a symplectic manifold of dimension $2n$, and let*

$$f_1, \dots, f_n : M \to \mathbb{R}$$

be C^∞ functions such that the Poisson brackets $\{f_i, f_j\}$ all vanish. Suppose that the set M_0 of equation $f_1 = f_2 = \cdots = f_n = 0$ is compact, and that the 1-forms $df_i,\ i = 1, \dots, n$ are linearly independent at all points of M_0. Then it is possible to choose coordinates $\mathbf{q}, \mathbf{p}$ on a neighborhood M' of M_0 which make the Hamiltonian system (M', σ, f_1) isomorphic to a neighborhood of the 0 section in a standard completely integrable system as above.

Remark. The hypothesis and the implicit function theorem imply that M_0 is a compact n-dimensional manifold; part of the proof is to show that this manifold is in fact diffeomorphic to a torus.

Proof. Note that M_0 is a Lagrangian submanifold; therefore it is possible to choose a Lagrangian submanifold Z intersecting M_0 at a point $\mathbf{m}_0$, such that $T_{\mathbf{m}_0}M = T_{\mathbf{m}_0}M_0 \oplus T_{\mathbf{m}_0}Z$. By the inverse function theorem, we may assume that restrictions of the f_i to Z give a diffeomorphism of Z onto an open neighborhood U of $\mathbf{0}$ in $\mathbb{R}^n$.

We can now define a mapping $\Phi : Z \times \mathbb{R}^n \to M$ by the formula

$$\Phi(z, \mathbf{t}) = \phi_{t_1 f_1 + \cdots + t_n f_n}(1, z) = \phi_{f_1}^{t_1}(\phi_{f_2}^{t_2}(\dots, \phi_{f_n}^{t_n}(z)\dots)). \qquad (A\ 6.2)$$

The mapping Φ is an $\mathbb{R}^n$-action in the sense that $\Phi(z, \mathbf{s} + \mathbf{t}) = \Phi(\Phi(z, \mathbf{s}), \mathbf{t}))$, since the vector-fields $\nabla_\sigma(f_i)$ commute.

The domain $Z \times \mathbb{R}^n$ can be thought of as T^*Z, since Z has explicit coordinates, i.e., one can think of $(z, \mathbf{t}) \in Z \times \mathbb{R}^n$ as $(z, \sum t_i df_i) \in T^*Z$. Thus $Z \times \mathbb{R}^n$ carries the canonical symplectic structure of a cotangent bundle, which we will denote σ_Z. The key point of the proof is the following lemma.

Lemma A 6.3. *The mapping Φ is (almost) symplectic, i.e., we have $\Phi^*\sigma = -\sigma_Z$.*

Proof. First, observe that is enough to prove this on Z, since σ is invariant under the hamiltonian flow, and so is σ_Z. Further, Z and M_z are Lagrangian for both σ and σ_Z. So to verify Lemma A 6.3, it is enough to show that

$$\Phi^*\sigma\left(\frac{\partial}{\partial z_i},\frac{\partial}{\partial t_j}\right)=\sigma_Z\left(\frac{\partial}{\partial z_i},\frac{\partial}{\partial t_j}\right)$$

for all $1\le i,j\le n$.

We have

$$\Phi^*\sigma\left(\frac{\partial}{\partial z_i},\frac{\partial}{\partial t_j}\right)=\sigma\left(\frac{\partial}{\partial z_i},\nabla_\sigma f_j\right)=df_j\left(\frac{\partial}{\partial z_i}\right)=\delta_{i,j}.$$

On the other hand

$$\begin{aligned}\sigma_Z(\frac{\partial}{\partial z_i},\frac{\partial}{\partial t_j})&=\sum_{k=1}^n(dt_k\wedge dz_k)\left(\frac{\partial}{\partial z_i},\frac{\partial}{\partial t_j}\right)\\&=\sum_{k=1}^n\left(dt_k\left(\frac{\partial}{\partial z_i}\right)dz_k\left(\frac{\partial}{\partial t_j}\right)-dz_k\left(\frac{\partial}{\partial z_i}\right)dt_k\left(\nabla_\sigma f_j\right)\right)=-\delta_{i,j}.\end{aligned}$$

□

Define $\Lambda=\Phi^{-1}(Z)$ and $\Lambda_z=\Phi^{-1}(\{z\})$. For each z the subset $\Lambda_z\subset\mathbb{R}^n$ is a discrete subgroup of the additive group $\mathbb{R}^n$. It is a subgroup because Φ is a group action: if $\mathbf{s},\mathbf{t}\in\Lambda_z$ so that if $\Phi(z,\mathbf{s})=\Phi(z,\mathbf{t})=z$, then

$$\Phi(z,\mathbf{s}+\mathbf{t})=\Phi(\Phi(z,\mathbf{s}),\mathbf{t})=\Phi(z,\mathbf{t})=z$$

so that $\mathbf{s}+\mathbf{t}\in\Lambda_z$. It is discrete because Φ is a diffeomorphism on a neighborhood of Z and the flow it describes is transversal to Z.

Moreover, Φ induces a homeomorphism between the compact space M_0 and $\mathbb{R}^n/\Lambda_{\mathbf{m}_0}$, so $\Lambda_{\mathbf{m}_0}$ is a lattice, i.e., a discrete subgroup of $\mathbb{R}^n$ isomorphic to $\mathbb{Z}^n$; it follows that M_0 is homeomorphic to a torus. Moreover, by shrinking Z if necessary we may assume that Λ_z is still a lattice for all $z\in Z$. Thus we see that the inclusion

$$\begin{array}{ccc}\Lambda & \hookrightarrow & Z\times\mathbb{R}^n=T^*Z\\ & \searrow\quad\swarrow & \\ & Z & \end{array}$$

makes Λ into a bundle of lattices over Z, which we may take to be trivial by taking Z smaller yet.

We can then choose sections $\phi_1,\dots,\phi_n$ of Λ such that for each $z\in Z'$, the elements $\phi_1(z),\dots,\phi_n(z)$ of Λ_z form a basis. These ϕ_i are also sections of T^*Z, i.e., 1-forms on Z, and their images are Lagrangian submanifolds of T^*Z, so they are closed forms on Z. Taking Z smaller yet if necessary, we can set $\phi_i=dp_i$ for appropriate functions p_i on Z; moreover, the p_i are coordinates on Z, since their derivatives form a basis; by further shrinking Z if necessary, the functions p_i define a diffeomorphism $\mathbf{p}:Z\to V$ for an appropriate neighborhood V of 0 in $\mathbb{R}^n$. We now have our coordinates: the map

$$\Psi:\mathbb{R}^n\times V\to M\quad\text{given by}\quad\Psi(\mathbf{p},\mathbf{q})=\Phi(\mathbf{p},\sum q_i\phi_i).$$

There is not much to prove: clearly $\sum dq_i \wedge dp_i = \sigma_Z$ (these are canonical coordinates), so $\Phi^*\sigma = \sum dp_i \wedge dq_i$. $\square$

Bibliography

[2] V. Arnold, *Mathematical methods of classical mechanics*, Springer Verlag, 1985.
[2] G. Benettin, L. Galgani, A. Giorgilli, J.-M. Strelcyn, *A proof of Kolmogorov's theorem on invariant tori using canonical transformations defined by the Lie method*, Il Nuevo Cimento 79B, 2 (1984), 201-223
[3] B. Hubbard and J. Hubbard, *Vector Calculus, Linear Algebra and Differential forms*, 2nd ed., Prentice Hall, Upper Saddle river, New Jersey 07458, 2002
[4] W. Rudin, *Function theory on balls*, Springer Verlag, NY (1980)