Kolmogorov's Theorem

Travis Westura

May 11, 2015

Abstract

This paper discusses a proof of the Kolmogorov Theorem on the conservation of invariant tori. We follow the approach given by Hubbard and Ilyashenko in [8]. Their proof is influenced by the one given by Bennettin, Galgani, Giorgilli, and Strelcyn in [3], which itself resembles Kolmogorov's original argument.

1 Introduction

Andrey Kolmogorov announced an important theorem in the International Congress of Mathematicians in 1954. However, he never wrote down a proof of this theorem. Vladimir Arnold give a proof in 1963, and Jürgen Moser proved a related result in 1962. The resulting theory is called KAM Theory, named after the three of them.

Arnold recalls in [2] how Kolmogorov had been considering the problem since he was a child. He was influenced by reading *The Flammarion Book of Astronomy*, where [4] is a more recent edition than the one Kolmogorov would have read. After Stalin died Kolmogorov entered a period of great mathematical productivity. This theorem was born out of this period.

While Kolmogorov never wrote down a proof, many others did, or at least attempted to. As already noted, Arnold wrote a proof in 1963 and Moser proved a related result in 1962. But the proof is quite difficult, as Arnold's original argument was over 60 pages and quite difficult to understand. The argument with which we concern ourselves is given by Hubbard and Ilyshenko in [8]. This proof, while still containing many messy, technical details, is much shorter. We will discuss the motivation behind the main ideas and key concepts of this proof.

Before moving on, let's take a look at our main goal:

Theorem 1.1 (Kolmogorov's Theorem)**.** *Let $\rho, \gamma > 0$ be given, and let $h(\boldsymbol{q}, \boldsymbol{p}) = h_0(\boldsymbol{p}) + h_1(\boldsymbol{q}, \boldsymbol{p})$ be a Hamiltonian, with $h_0, h_1 \in \mathcal{A}_\rho$ and $\|h\|_\rho \leq 1$. Suppose the Taylor polynomial of h_0 is*

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

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

$$H(\boldsymbol{Q}, \boldsymbol{P}) = A + \omega \boldsymbol{P} + R(\boldsymbol{Q}, \boldsymbol{P}), \tag{1.1}$$

with $R(\boldsymbol{Q}, \boldsymbol{P}) \in O(|\boldsymbol{P}|^2)$. In particular the torus $\boldsymbol{P} = 0$ is invariant under the flow $\nabla_\sigma H$, and on this torus the flow ϕ_H is linear with direction ω.

We will return to this statement after building up some intuition for it. For now, let's just note that the proof involves solving equation (1.1) for the symplectic diffeomorphism Φ.

In this paper we will follow the proof given by Hubbard and Ilyshenko given in [8]. We will begin by examining a toy example of the solar system, to which we will apply the theorem. This example will help us in understanding the statement of the theorem and the tools that we will need in the proof. We will discuss the basics of Hamiltonian mechanics and the notion of irrational vectors while examining this example. Then we will begin examining the technique of the proof and examine the role played by analytic functions. Finally we will put all of this together and explain the main ideas of the proof, leaving out some of the messy, technical details.

2 A Motivating Example

To start understanding the theorem, let's consider an example to which we would like to apply it. While staring up at the night sky, one might be driven to wonder, "Why do the planets orbit around the sun and not just crash into the sun or fly off by themselves in their own directions?" Kolmogorov's Theorem gives us a method to attempt to answer this question.

Let's start by discussing a simplified model of our solar system. In this model we will assume that planets have zero mass. This assumption is reasonable, as the masses of the planets are extremely small compared to that of the sun. Thus we may, at least for the moment, consider these masses to be negligible.

A system with n bodies, each with a mass m_i and a position $\boldsymbol{x}_i$ satisfies Netwon's second law: $\boldsymbol{F} = m\boldsymbol{a}$. For each i, we have

$$m_i\ddot{\boldsymbol{x}} = \sum_{j\neq i} \mathrm{G}\, m_i m_j \frac{\boldsymbol{x}_j - \boldsymbol{x}_i}{|\boldsymbol{x}_j - \boldsymbol{x}_i|^3}.$$

Here, $\mathrm{G} \approx 6.62 \cdot 10^{-11} m^3/(kg\, s^2)$ is the universal gravitational constant. On the left hand side we have each planet's mass and its acceleration. On the right hand side we have the force that acts on this mass. Note that the force is inversely proportional to the square of the distances between the forces. The top of the fraction contains the vector $\boldsymbol{x}_j - \boldsymbol{x}_i$ that has length $|\boldsymbol{x}_j - \boldsymbol{x}_i|$. To cancel out this length, we divide by $|\boldsymbol{x}_j - \boldsymbol{x}_i|^3$. This gives us the desired inverse proportionality to the square of the distance.

In our model of the solar system we take the sun to be given by the index 0 and the planets given by indices from 1 to 8 or 9, depending on the reader's opinions of Pluto. Now we will examine what happens as the masses of the planets, the m_j for $j = 1, \dots, n$, tend to zero.

Rewriting the previous equations, we have

$$\ddot{\boldsymbol{x}}_0 = \mathrm{G}\sum_{j=1}^{n} m_j \frac{\boldsymbol{x}_j - \boldsymbol{x}_0}{|\boldsymbol{x}_j = \boldsymbol{x}_i|^3},$$
$$\ddot{\boldsymbol{x}}_i = \mathrm{G}m_0 \frac{\boldsymbol{x}_0 - \boldsymbol{x}_i}{|\boldsymbol{x}_0 - \boldsymbol{x}_i|^3} + \mathrm{G}\sum_{\substack{j=1,\dots,n\\ j\neq i}} m_j \frac{\boldsymbol{x}_j - \boldsymbol{x}_i}{|\boldsymbol{x}_j - \boldsymbol{x}_i|^3}.$$

Now we keep the mass of the sun m_0 constant but let the masses of the planets go to zero. These equations then become

$$\ddot{\boldsymbol{x}}_0 = 0,$$
$$\ddot{\boldsymbol{x}}_j = \mathrm{G}\,\frac{\boldsymbol{x}_0 - \boldsymbol{x}_i}{|\boldsymbol{x}_0 - \boldsymbol{x}_i|^3}.$$

We can even simplify these equations a bit further. Since $\ddot{\boldsymbol{x}}_0 = 0$, the body with mass m_0 travels in a straight line with constant speed. Thus we can work in a heliocentric system of coordinates with the sun at the center of our solar system with position $\boldsymbol{x}_0 = 0$. We then have

$$\ddot{\boldsymbol{x}}_0 = 0,$$
$$\ddot{\boldsymbol{x}}_j = -\mathrm{G}\, m_0 \frac{\boldsymbol{x}_i}{|\boldsymbol{x}_i|^3}.$$

In particular we can note that this system is stable. Over time, it will never diverge far from its present state.

3 Hamiltonian Mechanics

We will need to consider Hamiltonian mechanics in our study of the KAM Theorem. Hamiltonian mechanics is a reformulation of Newtonian mechanics. The theories produce the same results, but often one of them is a more convenient formulation for some problems than is the other.

A Hamiltonian mechanical system is given by an even-dimensional manifold, the phase space, a symplectic structure on this manifold, and a function H. This function is called a Hamiltonian function.

Let (X, σ) be a symplectic manifold. That is, let X be a differentiable manifold and σ be a nowhere vanishing 2-form such that $d\sigma = 0$. A function H on X has a *symplectic gradient*, denoted $\nabla_\sigma H$. This gradient is defined as the unique vector field $\nabla_\sigma H$ such that for any vector field ξ, we have

$$\sigma(\xi, \nabla_\sigma H) = dH(\xi). \tag{3.1}$$

We will be considering a *Hamiltonian differential equation* that makes use of this symplectic gradient:

$$\dot{\boldsymbol{x}} = (\nabla_\sigma H)(\boldsymbol{x}). \tag{3.2}$$

Before moving on, let's compare the symplectic gradient and the gradient ∇f of a function $f : \mathbb{R}^n \to \mathbb{R}$. This gradient is not an element of $\mathbb{R}$, but rather is the vector

$$\nabla f = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix}.$$

In the symplectic gradient, we used the symplectic form σ to give a geometric structure to our underlying space. Similarly, we can define the gradient by using the scalar product $\langle \cdot, \cdot \rangle$ to provide a geometric structure. In comparison to equation (3.1), the gradient ∇f is the unique vector field such that

$$df(\xi) = \langle \xi, \nabla f \rangle.$$

Further, just as we considered a Hamiltonian differential equation in equation (3.2), we can consider a gradient differential equation:

$$\dot{\boldsymbol{x}} = \nabla f(\boldsymbol{x}).$$

The gradient and Hamiltonian differential equations are fundamentally different. In the gradient equation, our function f increases along solutions. Thus these solutions may never return to their starting points. The gradient differential equation prohibits any type of recurrence from occurring. In contrast, the Hamiltonian differential equation not just allows but practically imposes recurrent behavior. This difference means we can observe much more interesting recurrent behavior when using Hamilton's equation than we could with the gradient equation.

As an example of a Hamiltonian system, let's consider our toy model of the solar system from section 2. We take our manifold X to be $\mathbb{R}^{2n}$ with coordinates $(q_1, \dots, q_n, p_1, \dots, p_n)$. Our symplectic form will be given by

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

The Hamiltonian differential equation then becomes the *Hamiltonian equations of motion*:

$$\begin{aligned} \dot{q}_i &= \frac{\partial H}{\partial p_i}, \\ \dot{p}_i &= -\frac{\partial H}{\partial q_i}. \end{aligned} \tag{3.3}$$

Now we consider the case of a single body of zero mass. It is sufficient to consider the single body system, since the planets' having zero mass means that they do not affect each other in our toy system. For $\boldsymbol{x} \in \mathbb{R}^2$, the equation

$$\ddot{\boldsymbol{x}} = -\frac{\boldsymbol{x}}{|\boldsymbol{x}|^3}$$

is Hamilton's equation for the manifold $X = \mathbb{R}^2 \times \mathbb{R}^2$ with points $(\boldsymbol{q}, \boldsymbol{p})$, standard symplectic form

$$\sigma = dp_1 \wedge dq_1 + dp_2 \wedge dq_2,$$

and Hamiltonian

$$H(\boldsymbol{q}, \boldsymbol{p}) = \frac{1}{2}(p_1^2 + p_2^2) - \frac{1}{\sqrt{q_1^2 + q_2^2}}.$$

We write $\boldsymbol{p} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$ and $\boldsymbol{q} = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}$. Now we have

$$\begin{aligned} &(dp_1 \wedge dq_1 + dp_2 \wedge dq_2)\left(\left(\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}\right), \left(\begin{pmatrix} p1 \\ p2 \end{pmatrix}, -\frac{1}{(q_1^2 + q_2^2)^{\frac{3}{2}}}\begin{pmatrix} q1 \\ q2 \end{pmatrix}\right)\right) \\ &= \eta_1 p_1 + \frac{\xi_1 q_1}{(q_1^2 + q_2^2)^{\frac{3}{2}}} + \eta_2 p_2 + \frac{\xi_2 q_2}{(q_1^2 + q_2^2)^{\frac{3}{2}}} \\ &= \left[DH\left(\begin{pmatrix} q1 \\ q1 \end{pmatrix}, \begin{pmatrix} p1 \\ p2 \end{pmatrix}\right)\right]\left(\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}\right). \end{aligned}$$

Following from this computation, we see that

$$\nabla_\sigma H = \left(\boldsymbol{p}, -\frac{1}{(q_1^2 + q_2^2)^{\frac{3}{2}}}\boldsymbol{q}\right).$$

The Hamiltonian differential equation $\boldsymbol{x} = \nabla_\sigma H(\boldsymbol{x})$ is then given by

$$\begin{aligned} q_1' &= p_1 \\ p_1' &= -\frac{q_1}{(q_1^2 + q_2^2)^{\frac{3}{2}}}, \\ q_2' &= p_2 \\ p_2' &= -\frac{q_2}{(q_1^2 + q_2^2)^{\frac{3}{2}}}. \end{aligned}$$

From here, we can recover the desired $\ddot{\boldsymbol{x}} = -\dfrac{\boldsymbol{x}}{|\boldsymbol{x}|^3}$.

Now let's again consider the Hamiltonian Differential Equation equation (3.2). The vector field $\nabla_\sigma f$ has a flow that we denote by ϕ_f^t. This flow has two properties:

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

Flows will be central to our proof of Kolmogorov's Theorem. We will be constructing a symplectic diffeomorphism and will need the flows of Hamiltonian functions to do so.

Further, in our proof we will be constructing Taylor polynomials of functions of the form $t \mapsto g \circ \phi_f^t$. To perform this construction, we will use the Poisson bracket.

Definition 3.1 (Poisson Bracket)**.** Let f and g be two function on X. The Poisson bracket, denoted $\{f, g\}$, is defined by

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

We say that two functions f and g commute if the Poisson bracket is zero:

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

This implies that their flows also commute:

$$\phi_f(s) \circ \phi_g(t) = \phi_g(t) \circ \phi_f(s).$$

The Poisson bracket allows to write Taylor polynomials in the form

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

The Poisson bracket is related to the Lie bracket: the Lie bracket is the symplectic gradient of the Poisson bracket.

Proposition 3.2 (Relation of Poisson and Lie Brackets)**.** *For any two functions f and g on a symplectic manifold (X, σ), we have*

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

Proof. This proof will use the Jacobi identity for the Poisson bracket:

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

Given a function h, we can compute

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

QED

As an example of a Poisson bracket, let's consider again the Hamiltonian equations of motion (equation (3.3)). The Poisson bracket is calculated by

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

Next we discuss what it means for a system to be totally integrable. We will be denoting the torus by $\mathbb{T} = \mathbb{R}/\mathbb{Z}$.

Definition 3.3 (Totally Integrable System)**.** A *totally integrable system* is a symplectic manifold $X = \mathbb{T}^n \times \mathbb{R}^n$ with variables $(\boldsymbol{q} \in \mathbb{T}^n, \boldsymbol{p} \in \mathbb{R}^n)$, symplectic form $\sum_i dp_i \wedge dq_i$, and Hamiltonian function $H(\boldsymbol{p})$ depending only on $\boldsymbol{p}$.

The first example is to integrate the Hamiltonian equations of motion (equation (3.3)). If we take $(\boldsymbol{q}_0, \boldsymbol{p}_0)$ to be the initial value and let $\omega = \dfrac{\partial H}{\partial \boldsymbol{p}}$, then the solution is

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

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

Let's take a moment to consider the importance of this example. Indeed, this is the example that comes up whenever you are dealing with a mechanical system that has n degrees of freedom and n commuting conservation laws. To make this notion more precise, we discuss a theorem of Liouville.

Theorem 3.4 (Liouville's Theorem)**.** *Let (X, σ) be a symplectic manifold of dimension $2n$. Let*

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

be C^∞ functions such that pairwise, the Poisson brackets vanish, that is, $\{f_i, f_j\} = 0$. Suppose further that the set X_0 of equation

$$f_1 = \cdots = f_n = 0$$

is compact, and that the 1-forms df_i for $i = 1, \dots, n$ are linearly independent at all points of X_0. Then one can choose coordinates $\boldsymbol{q}, \boldsymbol{p}$ on a neighborhood X' of X_0 that make the Hamiltonian system (X', σ, f_1) isomorphic to a neighborhood of the 0 section in a standard completely integrable system.

In the statement of this theorem, (X, σ, H) is 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 depends only on the variables $p_1, \dots, p_n$.

Liouville's Theorem can be applied to our toy solar system example. We consider the case of a single planet. The system has two degrees of freedom, so we need two conservation laws. The first one is the function

$$H\left(\begin{pmatrix} q_1 \\ q_2 \end{pmatrix}, \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}\right) = \frac{1}{2}(p_1^2 + p_2^2) - \frac{1}{(q_1^2 + q_2^2)^{\frac{1}{2}}}.$$

For the second conservation law we take the function

$$M = q_1 p_2 - q_2 p_1.$$

Then we have that we have a conservation law by computing

$$\begin{aligned} M' &= q_1' p_2 + q_1 p_2' - (q_2' p_1 + q_2 p_1') \\ &= q_1' p_2 + q_1 p_2' - q_2' p_1 - q_2 p_1' \\ &= p_1 p_2 + q_1 \frac{q_2}{(q_1^2 + q_2^2)^{\frac{3}{2}}} - p_2 p_1 - q_2 \frac{q_1}{(q_1^2 + q_2^2)^{\frac{3}{2}}} = 0. \end{aligned}$$

Finally we check that the conservation functions H and M commute, that is, that their Poisson bracket is zero.

$$\begin{aligned}\{H, M\} &= \frac{\partial H}{\partial q_1}\frac{\partial M}{\partial p_1} - \frac{\partial H}{\partial p_1}\frac{\partial M}{\partial q_1} + \frac{\partial H}{\partial q_2}\frac{\partial M}{\partial p_2} - \frac{\partial H}{\partial p_2}\frac{\partial M}{\partial q_2} \\ &= q_1\frac{q_2}{(q_2^2+q_2^2)^{\frac{3}{2}}} - p_1p_2 + q_2\frac{p_1}{(q_1^2+q_2^2)^{\frac{3}{2}}}.\end{aligned}$$

4 Irrationality

In our statement of Kolmogorov's Theorem, we included the hypothesis that $\omega \in \Omega_\gamma$. We now define this notation and begin to explain its importance. The set Ω_γ consists of vectors that are "sufficiently irrational," a notion that we need to make more precise.

Let's first consider the definition of an irrational number. If a real number θ is irrational, then for all pairs if integers p and q, with q positive, we have the following

$$\left|\theta - \frac{p}{q}\right| \neq 0.$$

This equation tells us simply that there does not exist and rational number $\frac{p}{q}$ that equals our irrational number θ. Here, the "not equal to zero" part of the equation will be stressed, as an expression similar to the one on the left hand side will later appear as the denominator of a fraction (see section 8). As dividing by zero can be rather troublesome, we wish to avoid it. This condition of irrationality is the tool we use to do so: if θ is irrational, then the left side will not be zero, so we can divide by it without any problems.

Our condition of "sufficiently irrational" will mean that $\left|\theta - \frac{p}{q}\right|$ is "sufficiently nonzero," or since we are using an absolute value, "sufficiently big." However, as we learn in our introductory courses in real analysis, the rationals are dense in the reals, and every real number, specifically every irrational number θ, may be approximated arbitrarily closely by the rationals. More precisely, given any real $\epsilon > 0$, there exists a rational number $\frac{p}{q}$ such that $\left|\theta - \frac{p}{q}\right| < \epsilon$. Thus trying to coerce $\left|\theta - \frac{p}{q}\right|$ to be big is quite impossible.

Unsatisfied with our answer, let's instead consider a different question. Instead of wanting $\left|\theta - \frac{p}{q}\right|$ to be "big,", we ask that it is small *only if the denominator is big.* This is the beginning of the theory of Diophantine approximation.

The numbers that we seek will satisfy the following definition.

Definition 4.1 (Diophantine Number of Exponent d)**.** A number θ is *Diophantine of exponent d* if there exists a constant $\gamma > 0$ such that for all coprime integers p and q we have

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

From this definition we see that it is a stronger requirement for a number to be Diophantine of a smaller exponent. For all irrational numbers θ there exist arbitrarily large q and p prime to q such 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. And the number that are Diophantine of exponent exactly 2 are precisely the numbers whose continued fractions have bounded entries. These numbers form a set of measure zero.

But what about exponents greater than 2, that is, of the form $2 + \epsilon$ for $\epsilon > 0$? In the sense of Lebesgue measure, these numbers are quite abundant, as for any $\epsilon > 0$ they form a set of full measure.

Proposition 4.2 (Diophantine Numbers with Full Measure)**.** *For all $\epsilon > 0$, the set of Diophantine numbers of exponent $2+\epsilon$ is of full measure.*

Proof. We consider numbers in $\mathbb{R}/\mathbb{Z}$. Given any positive integer q, there are at most q elements of $\mathbb{Q}/\mathbb{Z}$ that, in reduced form, have denominator q. Hence for any constant γ, we consider the set

$$\left\{\theta \in \mathbb{R}/\mathbb{Z} : \left|\theta - \frac{p}{q}\right| < \frac{\gamma}{|q|^{2+\epsilon}}\right\}.$$

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

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

has length strictly less than

$$2\gamma \sum_{q=1}^{\infty} \frac{1}{q^{1+\epsilon}}.$$

Take the intersection over all these sets as $\gamma \to 0$ and note that this intersection has measure 0. But this set is the complement of the set of Diophantine numbers of exponent $2+\epsilon$. Hence the claim holds. QED

Having this definition for numbers, we want to extend "irrationality" to vectors. As an example, we can consider the solar system with the vector $\omega = (\omega_1, \ldots, \omega_n)$, where each ω_i represents the frequency of the ith planet's orbit. Our statement of Kolmogorov's theorem will require that such a vector be irrational according to the following definition.

Definition 4.3 (Diophantine Vector)**.** Let $\omega = (\omega_1, \ldots, \omega_n)$. We say ω is Diophantine if there exists $\gamma > 0$ such that for all vectors with integer coefficients $\boldsymbol{k} = (k_1, \ldots, k_n)$, we have

$$|k_1\omega_1 + \cdots + k_n\omega_n| \geq \frac{\gamma}{(k_1^2 + \cdots + k_n^2)^{\frac{n}{2}}}.$$

Let Ω_γ^n be the subset of such $\omega \in \mathbb{R}^n$.

We could rewrite the condition in the definition as

$$\boldsymbol{k} \cdot \omega \geq \frac{\gamma}{|\boldsymbol{k}|^n}. \tag{4.1}$$

Also, we will often drop then n from out notation and write simply Ω_γ instead of Ω_γ^n when it is clear that we are working in $\mathbb{R}^n$.

Again, we want to examine how common it is for such vectors to occur. We do not want just exceptional motions to be preserved, but rather we wish that most motions are preserved, and that we should not need to look hard to find such vectors. In the case of numbers, we have that satisfying answer that Diophantine numbers have full measure. We establish an analogous result for vectors, following a similar proof to the one we have just seen.

Proposition 4.4 (Diophantine Vectors are of Full Measure)**.** *The union over $\gamma > 0$ of sets of Diophantine vectors with constant γ,*

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

is of full measure.

Proof. Consider the region $S_{\boldsymbol{k},\gamma}$, in which

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

is a region around the hyperplane orthogonal to $\boldsymbol{k}$ and with thickness $\dfrac{2\gamma}{|\boldsymbol{k}|^{n+1}}$. Denote the unit cube by Q. The part of $S_{\boldsymbol{k},\gamma}$ within Q has measure at most $\dfrac{\mathrm{M}\gamma}{|\boldsymbol{k}|^{n+1}}$, where M denotes the constant giving the maximal $(n-1)$-dimensional measure of the intersection of Q with a hyperplane. Now consider the sum

$$\sum_{k \in \mathbb{Z}^n \setminus \{0\}} \frac{1}{|\boldsymbol{k}|^{n+1}}.$$

This sum is finite, so the volume of

$$\bigcup_{\boldsymbol{k} \in \mathbb{Z}^n \setminus \{0\}} S_{\boldsymbol{k},\gamma} \cap Q$$

is bounded by some constant times γ. As before, we now consider the intersection of these sets as $\gamma \to 0$:

$$\bigcap_{\gamma > 0} \bigcup_{\boldsymbol{k} \in \mathbb{Z}^n \setminus \{0\}} S_{\boldsymbol{k},\gamma} \cap Q.$$

This intersection has measure 0, and this set is the complement of our desired set Ω. Thus Ω has full measure.

Note that this proof is very similar to that of proposition 4.2, and reduces to the case $\epsilon = 1$ if we take $n = 2$. QED

We can now understand the condition $\omega \in \Omega_\gamma$ from our statement of Kolmogorov's Therorem. This requirement means that the vector ω must be "suitably irrational," and such vectors are rather "common" in the sense of Lebesgue.

5 KAM and the Solar System

With these preliminaries out of the way, we are going to discuss Kolmogorov's Theorem in the context of our solar system example.

We view the solar as system as an n dimensional torus, where n is the number of planets. This gives us a geometric way to describe the motions of the planets. Suppose that the planets have initial positions given by $\boldsymbol{a} = (a_1, \ldots, a_n)$. Then a trajectory with frequency vector $\omega = (\omega_1, \ldots, \omega_n)$ is at the point $\boldsymbol{a} + t\omega$ at time t.

Define a *linear flow on* $(\mathbb{R}/\mathbb{Z})^n$ *in the direction* ω to be the motion given by

$$t \mapsto \boldsymbol{a} + t\omega = (a_1 + t\omega_1, \ldots, a_n + t\omega_n).$$

Such a trajectory is dense on the torus if and only if ω is irrational according to the Diophantine condition from definition 4.3.

Kolmogorov's Theorem tells us about when motions of a system are preserved, so let's examine what it means for two motions to be the same as each other. Consider a motion $\boldsymbol{x}(t)$ to be a motion of the perturbed system and a motion $\boldsymbol{x_1}(t)$ to be a motion of the unperturbed system that is dense on a torus T_1. For the motions to be the same, we mean that $\boldsymbol{x}(t)$ is dense on the corresponding torus T and that it fills in T in the same way that $\boldsymbol{x}_1$ fills in T_1. That is, there exists a homeomorphism $\Phi : T \to T_1$ such that

$$\Phi(\boldsymbol{x}(t)) = \boldsymbol{x}_1(t).$$

Using this notation, let's see what Kolmogorov's theorem says for our solar system.

Theorem 5.1 (Kolmogorov's Theorem Applied to the Solar System)**.** *Let $\boldsymbol{x}_1(t)$ be a motion of the zero-masses system with Diophantine frequency vector. Then there exists $\epsilon > 0$ such that, if the planets are given masses $m_i < \epsilon$, there exists a trajectory $\boldsymbol{x}(t)$ of the perturbed system dense on a torus T and a homeomorphism $\Phi : T \to T_1$ such that*

$$\Phi(\boldsymbol{x}(t)) = \boldsymbol{x}_1(t).$$

The set of such trajectories is of positive measure in the set of all trajectories. The probability of being on such a trajectory tends to 1 as ϵ tends to zero.

We would like that our solar system should fit the hypotheses of this theorem. But alas, the periods of the orbits of Jupiter and Saturn at in a $5:2$ ratio. Thus we have trouble with the important hypothesis of the irrationality of the planets' frequency vector. However, there are refinements of the theorem that state that there may still exist stable motions where the ratios are rational. We will not pursue these refinements here.

6 Analytic Functions

As this is a paper for a complex analysis class, we begin this section by noting that the complex analysis is located here.

Take another look at the statement of the theorem. We will concern ourselves with equation (1.1). Here we have $H = h \circ \Phi$, where Φ is a symplectic diffeomorphism. Our proof will involve solving for this diffeomorphism Φ.

We need a tool for solving equations, and the tool that we use is similar to Newton's Method. So let's take a moment first to consider what goes on to using Newton's Method.

Newton's Method can be used to solve an equation $f(\boldsymbol{x}) = 0$. The process involves choosing an initial guess $\boldsymbol{x}_0$. Then successive points $\boldsymbol{x}_i$ are defined by

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - [Df(\boldsymbol{x}_i)]^{-1} f(\boldsymbol{x}_i).$$

These points $\boldsymbol{x}_{i+1}$ will, under "good" conditions, converge to a solution to the equation.

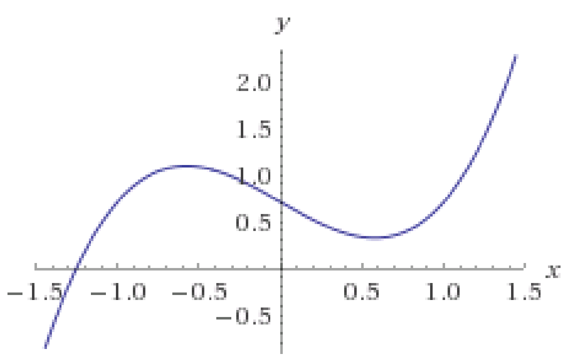

Figure 1: $f(x) = x^3 - x + \dfrac{\sqrt{2}}{2}$

To see why some additional conditions are necessary, consider the function $f(x) = x^3 - x + \dfrac{\sqrt{2}}{2}$. Let our initial guess be $x_0 = 0$. Note that this function has derivative $f'(x) = 3x^2 - 1$. Thus plugging in our initial guess gives us

$$x_1 = x_0 - \frac{1}{f'(x_0)} f(x_0) = \frac{\sqrt{2}}{2}.$$

But then we encounter a problem once we use x_1 to determine x_2.

$$x_2 = x_1 - \frac{1}{f'(x_1)} f(x_1) = 0.$$

This brings us back to our initial guess! So now Netwon's method will oscillate between 0 and $\dfrac{\sqrt{2}}{2}$ and will not converge to a solution.

The additional conditions that we need for the convergence of Newton's Method are given in a theorem by Kantorovitch. In order to guarantee convergence, we need a bound on the second derivative of the function f. For completeness, let's state this theorem.

Theorem 6.1 (Kantorovitch). *Let E and F be Banach spaces, U be an open subset of E, and $f : U \to F$ be C^1. Suppose that $\boldsymbol{x}_0 \in U$ is a point where $[Df(\boldsymbol{x}_0)] : E \to F$ is an isomorphism. Set $h_0 = -[Df(\boldsymbol{x}_0)]^{-1}(f(\boldsymbol{x}_0))$, $\boldsymbol{x}_1 = \boldsymbol{x}_0 + h_0$, and define the ball $U_0 = B_{\|h_0\|}(\boldsymbol{x}_1)$. Then if the following hold:*

1. $U_0 \subset U$
2. $\|[Df(\boldsymbol{y}_1)] - [Df(\boldsymbol{y}_2]\| \leq M$ *for all* $\boldsymbol{y}_1, \boldsymbol{y}_2 \in U_0$
3. $\|f(\boldsymbol{x})\| \left\|[Df(\boldsymbol{x}_0)]^{-1}\right\|^2 M \leq \dfrac{1}{2}$

then the equation $f(\boldsymbol{x}) = 0$ has a unique solution in U_0, and Newton's method starting from $\boldsymbol{x}_0$ is defined for all i and converges to this solution. If moreover we have

$$\|f(\boldsymbol{x})\| \left\|[Df(\boldsymbol{x}_0)]^{-1}\right\|^2 M = k < \frac{1}{2},$$

Then the method is superconvergent. *That is, if we put*

$$\begin{aligned} C &= \frac{1-k}{2(1-2k)} \left\|[Df(\boldsymbol{x}_0)]^{-1}\right\| M, \\ \boldsymbol{x}_{i+1} &= \boldsymbol{x}_i + h_i m \\ h_{i+1} &= -[Df(\boldsymbol{x}_i)]^{-1} f(\boldsymbol{x}_i), \end{aligned}$$

then $\|h_{i+1}\| \leq C \|h_i\|^2$.

Note that superconvergence corresponds to doubling the number of correct digits at each iteration. Geometric convergence differs from superconvergence, as it adds only a fixed number of significant figures at each iteration. The interested reader can follow a proof given in [6] and fill in some minor details to apply it to this more general case.

In solving for the diffeomorphism Φ in the proof of Kolmogorov's theorem, we will be using a similar technique to Newton's method. Where with Newton's method we needed a bound on the second derivative, we will need some type of analogous condition for our iterative process. We need some way to measure "size," that is, we need to choose a norm for our functions.

First let's recall the definition of a Banach algebra.

Definition 6.2 (Banach Algebra). Let k be $\mathbb{R}$ or $\mathbb{C}$. A *normed algebra* over k is an algebra $\mathcal{A}$ over k with a sub-multiplicative norm $\|\cdot\|$. That is, for all $x, y \in \mathcal{A}$, we have

$$\|xy\| \leq \|x\| \|y\| .$$

If $\mathcal{A}$ is a Banach space, then it is called a Banach algebra.

Let $X \subseteq \mathbb{C}^k$ be compact, and let the caligraphic letter $\mathcal{X}$ be the Banach algebra of continuous functions on X that are analytic in the interior and have the sup norm

$$\|f\|_X = \sum_{\boldsymbol{x} \in X} |f(\boldsymbol{x})|.$$

Here we use the normal absolute value as the standard Euclidean norm on $\mathbb{C}^n$.

We consider three regions:

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

Denote by $\mathcal{B}_\rho$, $\mathcal{C}_\rho$, and $\mathcal{A}_\rho$ the corresponding Banach algebras. We can expand elements of $\mathcal{B}_\rho$ as power series, and elements of $\mathcal{C}_\rho$ as Fourier series:

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

This Fourier series expansion will be important in section 8.

Let's take a look at how we can bound derivatives of analytic functions, analogous to how we use the second derivative to bound the derivative of functions in Kantorovitch's Theorem. We will use Cauchy's Inequalities on Balls.

Theorem 6.3 (Cauchy's Inequalities on Balls)**.** *If $f \in \mathcal{B}_\rho$, then*

$$
\begin{aligned}
\|Df\|_{\rho-\delta} &\leq \frac{1}{\delta} \|f\|_\rho , \\
\left\|D^2 f\right\|_{\rho-\delta} &\leq \frac{4}{\delta^2} \|f\|_\rho .
\end{aligned}
$$

As a corollary, the case $\delta = \rho$ bounds the derivatives at the center of balls:

$$
\begin{aligned}
|\, Df(0)| &\leq \frac{1}{\rho} \|f\|_\rho , \\
|\, D^2 f(0)| &\leq \frac{4}{\rho^4} \|f\|_\rho .
\end{aligned}
$$

Proof. Take $\boldsymbol{z} \in B_{\rho-\delta}$ and $\boldsymbol{u} \in \mathbb{C}^n$. Since $B_\delta(\boldsymbol{z}) \subseteq B_\rho$, the function

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

is defined on the unit disc. The normal Cauchy inequality implies that

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

Applying the argument twice gives the result for the second derivative:

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

QED

7 Main Idea of the Proof

Let's take another look at the statement of Kolmogorov's theorem and recall equation (1.1), where we have $H = h \circ \Phi$:

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

This equation is an equation for a diffeomorphism Φ, and the proof involves solving for this diffeomorphism. We would like to use Newton's method to do this solving, but that in itself is not quite sufficient for out purposes. However, we can still do something with a similar flavor.

The proof uses an iterative process in which we obtain the diffeomorphism Φ as a limit of Φ_i, where

$$\Phi_i = \phi_i \circ \phi_{i-1} \circ \cdots \circ \phi_1.$$

Here, ϕ_i denotes the Hamiltonian flow ϕ_{g_i} for a Hamiltonian function g_i. This g_i is the unknown for which we solve.

Let's take a moment to pause and reflect on what was just done. Instead of solving for a diffeomorphism Φ, at each iteration we are solving for a Hamiltonian function g_i. This simplification is rather important, as a diffeomorphism can be a rather complicated object. The g_i are much easier to reason about. Furthermore, the map ϕ_{g_i} is itself a symplectic diffeomorphism.

At the ith iteration we have a Hamiltonian $\tilde{h} = \Phi_i^* h$. We expand $\tilde{h}$ up to order 2 in $\boldsymbol{p}$, and the coefficients are Fourier series in $\boldsymbol{q}$. We write $\tilde{h} = \tilde{h}_0 + \tilde{h}_1$, where

- $\tilde{h}_1$ has the terms constant or linear in $\boldsymbol{p}$, except the constant in $\boldsymbol{q}$,
- $\tilde{h}_0$ is everything else.

We would like to eliminate $\tilde{h}_1$, be we won't be able to do so by solving a linear equation. Instead we will solve a linear equation for a function g such that $\phi_g^*\tilde{h}$ is "better" than $\tilde{h}$, in some reasonable sense of "better."

We expand $\phi_g^*\tilde{h}$ to first order in g. This gives us

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

We want to eliminate the terms, other than the term constant in $\boldsymbol{q}$, that are not $O(|\boldsymbol{p}|)^2$. Applying the standard Newton's method would necessitate solving the equation

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

But we will do something a bit different.

We can assume that anything we want is small as long as the choice is justified by the resulting inequalities. So suppose that $\{g, \tilde{h}_1\}$ is of order 2 since g and $\tilde{h}_1$ are both small. Then the linear equation that we need to solve is

$$\tilde{h}_1 + \{g, \tilde{h}_0\} \in o(|\boldsymbol{p}|). \tag{7.1}$$

Thus we need to solve this *Diophantine differential equation*. Our ability to solve it relies on the Diophantine vectors that we saw previously in section 4.

8 Diophantine Differential Equations

Let $g \in \mathcal{C}_\rho$. We are going to be solving linear equations of the form

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

with $f \in \mathcal{C}_{\rho'}$ for some $\rho' < \rho$. Recall, as was mentioned in section 6, that since $f \in \mathcal{C}_{\rho'}$ and $g \in \mathcal{C}_\rho$, we can write f and g as Fourier Series.

$$\begin{aligned}f(\boldsymbol{q}) &= \sum_{\boldsymbol{k}\in\mathbb{Z}^n} f_{\boldsymbol{k}} \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k}\cdot\boldsymbol{q}},\\ g(\boldsymbol{q}) &= \sum_{\boldsymbol{k}\in\mathbb{Z}^n} g_{\boldsymbol{k}} \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k}\cdot\boldsymbol{q}}.\end{aligned}$$

The solution is given by

$$f_{\boldsymbol{k}} = \frac{1}{2\pi\mathrm{i}(\boldsymbol{k}\cdot\omega)} g_{\boldsymbol{k}}. \tag{8.1}$$

We need g_0 to be zero. Then f_0 is arbitrary, and otherwise the series for f is unique.

Take note of the $(\boldsymbol{k}\cdot\omega)$ in the denominator of equation (8.1) and compare this to equation (4.1). Dividing by zero in equation (8.1) would mean that there is no solution. So we need the Diophantine condition for the vector ω to ensure that we do not divide by zero in equation (8.1). Hence the convergence properties of f depend on the Diophantine properties of ω.

But even though this condition gives us convergence for ρ, we might not have boundedness. So we need some $\rho' < \rho$ where we have boundedess. But we must choose ρ' large enough so that the limit is nonempty. Thus the choice of ρ' must be done rather "carefully."

We will use the following tool for choosing ρ'.

Proposition 8.1. *If $g \in \mathcal{C}_\rho$ and $\epsilon \in \Omega_\gamma$, then for all δ such that $0 < \delta < \rho$, we have the following two inequalities:*

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

Here, κ_n is a constant that depends only on n.

Proof. For $\boldsymbol{y} \in \mathbb{R}^n$ with $|\boldsymbol{y} \le \rho$ (in particular $\boldsymbol{y} = \rho\frac{\boldsymbol{k}}{|\boldsymbol{k}|}$), the function

$$\boldsymbol{q} \mapsto g(\boldsymbol{q} - \mathrm{i}\boldsymbol{y})$$

is continuous and periodic in $\boldsymbol{q}$ of period 1. This function can be written

$$g(\boldsymbol{q} - \mathrm{i}\boldsymbol{y}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^n} g_{\boldsymbol{k}} \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{q}-\mathrm{i}\boldsymbol{y})} = \sum_{\boldsymbol{k}\in\mathbb{Z}^n} \left(g_{\boldsymbol{k}}\mathrm{e}^{2\pi\boldsymbol{k}\cdot\boldsymbol{y}}\right)\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{q}}.$$

By Parseval's theorem, we have

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

Since the series has positive numbers, we have

$$\|g\|_\rho^2 \ge |g_{\boldsymbol{k}}|^2 \mathrm{e}^{4\pi\rho|\boldsymbol{k}|}.$$

Using the Diophantine-ness of ω and the Fourier coefficients $f_{\boldsymbol{k}}$, we have

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

Next we split the proposition into two parts: the first part dealing with the f case and the second part dealing with the Df case.

First we handle the case for f. For $|\boldsymbol{q}| \le \rho - \delta$, we can write

$$\begin{aligned} \left|f_{\boldsymbol{k}}\mathrm{e}^{2\pi\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{q})}\right| &\le \sum_{\boldsymbol{k}\in\mathbb{Z}^n} \frac{|\boldsymbol{k}|^n}{2\pi\gamma} \|g\|_\rho \mathrm{e}^{-2\pi\rho|\boldsymbol{k}|(\rho-\delta)} \\ &= \frac{\|g\|_\rho}{2\pi\gamma} \sum_{\boldsymbol{k}\in\mathbb{Z}^n} |\boldsymbol{k}|^n \mathrm{e}^{-2\pi|\boldsymbol{k}|\delta}. \end{aligned}$$

We then rewrite the sum on the right:

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^n} |\boldsymbol{k}|^n \mathrm{e}^{-2\pi|\boldsymbol{k}|\delta} = \frac{1}{(2\pi\delta)^{2n}} \left((2\pi\delta)^n \sum_{\boldsymbol{k}'\in 2\pi\delta\mathbb{Z}^n} |\boldsymbol{k}'|^n \mathrm{e}^{-|\boldsymbol{k}'|} \right).$$

As $\delta \to 0$, the expression inside the parentheses approaches

$$\int_{\mathbb{R}^n} |\boldsymbol{x}|^n \mathrm{e}^{-|\boldsymbol{x}|} |\, d^n\boldsymbol{x}|.$$

Since the integral is convergent there exist κ_n' such that for $\delta \leq 1$, it holds that

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^n} |\boldsymbol{k}|^n \mathrm{e}^{-2\mathrm{i}|\boldsymbol{k}|\delta} \leq \frac{\kappa_n'}{(2\pi\delta)^{2n}}.$$

Thus we are finished with the f part of the proof.

The Df part of the proof is similar. We use

$$Df(\boldsymbol{q})(\boldsymbol{u}) = 2\pi\mathrm{i} \sum (\boldsymbol{k}\cdot\boldsymbol{u}) f_{\boldsymbol{k}} \mathrm{e}^{2\pi\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{q})}.$$

Then we obtain the inequality

$$|\, D(\boldsymbol{q})(\boldsymbol{u})| \leq \frac{|\boldsymbol{u}|\, \|g\|_\rho \,\kappa_n''}{\gamma(2\pi\gamma)^{n+1}}.$$

To complete the proof of the proposition, take

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

QED

9 Solving Equations

Let's return to our previous equation

$$h_1 + \{g, h_0\} \in O(|\boldsymbol{p}|^2).$$

This equation is the Diophantine differential equation that we had before in equation (7.1), but without tildes and with $o(|\boldsymbol{p}|)$ replaced by $O(|\boldsymbol{p}|^2)$, which is equivalent since our functions are analytic. The unknown is g, which we take to be degree 1 in $\boldsymbol{p}$:

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

Note that only the linear terms of g contribute to the linear terms of $\{g, h_0\}$. The unknowns are λ, X, and Y_i. We can expand X and Y_i as Fourier series, since they are functions of $\boldsymbol{q}$ only.

Let $a \in \mathbb{R}$, $\omega \in \Omega_\gamma$, $R(\boldsymbol{q},\boldsymbol{p}) \in O(|\boldsymbol{p}|^3)$, and $\overline{A} = 0$, where $\overline{A}$ is the average of A on the torus $\boldsymbol{p} = 0$. We can then write

$$\begin{aligned} h_0(\boldsymbol{q},\boldsymbol{p}) &= a + \omega\boldsymbol{p} + \frac{1}{2}\boldsymbol{p}\cdot C(\boldsymbol{q})\boldsymbol{p} + R(\boldsymbol{q},\boldsymbol{p}), \\ h_1(\boldsymbol{q},\boldsymbol{p}) &= A(\boldsymbol{q}) + B(\boldsymbol{q})\boldsymbol{p}. \end{aligned}$$

We then have

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

We need to solve the equations

$$DX(\boldsymbol{q})(\omega) = -A(\boldsymbol{q})$$
$$DY(\boldsymbol{q})(\omega) = -B(\boldsymbol{q}) - \big(\lambda + DX(\boldsymbol{q})\big)C(\boldsymbol{q})$$

for X and Y_i, $i = 1, \dots, n$. These are Diophantine differential equations. Since we have the hypothesis $\overline{A} = 0$, we can solve the first equation and find $X \in \mathcal{C}_{\rho'}$ for all $\rho' < \rho$. We then substitute into the second equation. We can then determine λ since the averages of the right-hand sides in the n are all zero. Then we can colve the second equation and find $Y_i \in \mathcal{C}_{\rho''}$ for all $\rho'' < \rho'$. Iterating the process requires choosing ρ' and ρ'' "carefully."

10 Conclusion

With everything that we have discussed, the statement of Kolmogorov's Theorem should now be very much readable. We have an equation that we solve for a diffeomorphism. This equation is solved iteratively using an analogue of Newton's Method. We used analytic functions in order for this technique to converge properly, analogous to how we would need to bound the second derivative is we were to apply Kantorovitch's Theorem when using Newton's Method. And we need the Diophantine vectors in order to avoid dividing by zero when solving for our Fourier coefficients.

The proof by Hubbard and Ilyshenko in [8] is quite short, and solely by following it without considering the work before it we may miss out on what make it so succinct, and why the simplifications that it makes are so powerful and useful. We can tell that the proof is difficult merely by the fact that Moser's proof, given in involves functions that are 333 times differentiable.

References

[1] V.I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer-Verlag, 1978.

[2] V.I. Arnold. From superpositions to KAM theory. *Vladimir Igorevich Arnold. Selected-60, Moscow: PHASIS*, pages 727–740, 1997.

[3] G. Benettin, L. Galgani, A Giorgilli, and J.-M. Strelcyn. A proof of Kolmogorov's theorem on invariant tori using canonical transformations defined by the lie method. *Il Nuovo Cimento*, 79B, 1984.

[4] C. Flammarion. *The Flammarion Book of Astronomy*. Simon and Schuster, New York, 1964.

[5] É. Ghys. Resonances and small divisors. In E. Charpentier, A. Lesne, and N. Nikolski, editors, *Kolmogorov's Heritage in Mathematics*. Springer, 2007.

[6] B. Hubbard and J. Hubbard. *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach*. Matrix Editions, Ithaca, NY, 2009.

[7] J. Hubbard. The KAM theorem. In E. Charpentier, A. Lesne, and N. Nikolski, editors, *Kolmogorov's Heritage in Mathematics*. Springer, 2007.

[8] J Hubbard and Y. Ilyashenko. A proof of Kolmogorov's theorem. *Discrete and Continuous Dynamical systems*, 4:1–20, 2003.

[9] J. Lee. *Introduction to Smooth Manifolds*. Springer:, 2013.