UNIVERSITÀ DI PISA
DIPARTIMENTO DI MATEMATICA

Dottorato di ricerca in matematica

AN INTRODUCTION TO SMALL DIVISORS PROBLEMS

by

Stefano Marmi

Preface

The material treated in this book was brought together for a PhD course I taught at the University of Pisa in the spring of 1999. It is intended to be an introduction to small divisors problems. The book is divided in two parts. In the first one I discuss in some detail the theory of linearization of germs of analytic diffeomorphisms of one complex variable. This is a part of the theory where many complete results are known. The second part is more informal. It deals with Nash–Moser's implicit function theorem in Fréchet spaces and Kolmogorov–Arnol'd–Moser theory. Many results (and even some statements) are just briefly sketched but I always refer the reader to a choice of the huge original literature on the subject.

I am particularly fond of the topics described in the first part, especially because of their interplay with complex analysis and number theory. The second part is also fascinating both because of its generality and because it leads to applications to Hamiltonian systems. Both are the object of major active research.

These lectures contain many problems (some of which may challenge the reader) : they should be considered as an essential part of the text. The proof of many useful and important facts is left as an exercise.

I hope that the reader will find these notes a useful introduction to the subject. However the reason of the long list of references at the end of these notes is my belief that the best way to learn a subject is to study directly the papers of those who invented it : Poincaré, Siegel, Kolmogorov, Arnol'd, Moser, Herman, Yoccoz, etc.

I am very grateful to Mariano Giaquinta for his invitation to give this series of lectures. I also wish to thank Carlo Carminati, whose enthusiasm is also at the origin of this project, and whose remarks have been essential in correcting some mistakes.

Udine, December 8, 1999.

Stefano Marmi

Table of Contents

Part I. One–Dimensional Small Divisors. Yoccoz's Theorems

1. Germs of Analytic Diffeomorphisms. Linearization

A *dynamical system* is the action of a group (or a semigroup) on some space. In looking for the simplest cases we are led to ask for the lowest possible dimension of the ambient space together with the highest possible regularity of the action. A remarkably rich but elementary situation is obtained considering the group of germs of holomorphic local diffeomorphisms of $\mathbb{C}$ which leave the point $z = 0$ fixed. In what follows we will omit the symbol $\circ$ for the composition of two germs (unless some confusion may be possible).

Let $\mathbb{C}[[z]]$ denote the ring of formal power series and $\mathbb{C}\{z\}$ denote the ring of convergent power series.

Let G denote the group of germs of holomorphic diffeomorphisms of $(\mathbb{C}, 0)$ and let $\hat{G}$ denote the group of formal germs of holomorphic diffeomorphisms of $(\mathbb{C}, 0)$: $G = \{f \in z\mathbb{C}\{z\}\,, f'(0) \neq 0\}$, $\hat{G} = \{\hat{f} \in z\mathbb{C}[[z]]\,, \hat{f}_1 \neq 0\}$. One has the trivial fibrations

$$\begin{array}{ccc} G = \cup_{\lambda\in\mathbb{C}^*} G_\lambda & \qquad & \hat{G} = \cup_{\lambda\in\mathbb{C}^*} \hat{G}_\lambda \\ \Big\downarrow{\scriptstyle \pi} & & \Big\downarrow{\scriptstyle \hat{\pi}} \\ \mathbb{C}^* & & \mathbb{C}^* \end{array} \tag{1.1}$$

where

$$\hat{G}_\lambda = \{\hat{f}(z) = \sum_{n=1}^{\infty} \hat{f}_n z^n \in \mathbb{C}[[z]]\,,\ \hat{f}_1 = \lambda\}\,, \tag{1.2}$$

$$G_\lambda = \{f(z) = \sum_{n=1}^{\infty} f_n z^n \in \mathbb{C}\{z\}\,,\ f_1 = \lambda\}\,. \tag{1.3}$$

1.1 Conjugation, Symmetries

Let $\mathrm{Ad}_g f$ denote the adjoint action of g on f : $\mathrm{Ad}_g f = g^{-1} f g$.

Definition 1.1 *Let $f \in G$ (resp. $\hat{f} \in \hat{G}$). We say that a germ g (resp. a formal germ $\hat{g} \in \hat{G}$) is* equivalent *or* conjugate *to f (resp. $\hat{f}$) if it belongs to the orbit of f (resp. $\hat{f}$) under the adjoint action of G_1 (resp. $\hat{G}_1$) :*

$$
\begin{aligned}
f \sim g &\iff \exists h \in G_1 \ : \ g = h^{-1} f h \ , \\
\hat{f} \sim \hat{g} &\iff \exists \hat{h} \in \hat{G}_1 \ : \ \hat{g} = \hat{h}^{-1} \hat{f} \hat{h} \ .
\end{aligned}
$$

The set of germs equivalent to f obviously forms an equivalence class, the *orbit* of f under the adjoint action of G_1 :

$$
[f] = \mathrm{Ad}_{G_1} f = \{g \in G \, , \ \exists h \in G_1 \ : g = \mathrm{Ad}_h f = h^{-1} f h\} \ .
$$

The same holds in the formal case.

Definition 1.2 *A germ $g \in G$ is a* symmetry *of $f \in G$ if $g \in Cent\,(f)$, i.e. if $Ad_g f = f$. We will denote by $\widehat{Cent}\,(\hat{f})$ the formal analogue of $Cent\,(f)$.*

Exercise 1.3 Let $f \in G_\lambda$ (resp. $\hat{f} \in \hat{G}_\lambda$) and assume $g \sim f$ (resp. $\hat{g} \sim \hat{f}$), i.e. $f = h^{-1} g h$ for some $h \in G_1$. Then show that
(1) $g \in G_\lambda$ (resp. $\hat{g} \in \hat{G}_\lambda$) thus $f_1 = f'(0) = \lambda$ is invariant under conjugation.
(2) $\mathrm{Cent}\,(f)$ is conjugated to $\mathrm{Cent}\,(g)$, i.e. $\mathrm{Cent}\,(f) = h^{-1} \mathrm{Cent}\,(g) h$;
(3) $f^{\mathbb{Z}} = \{f^n \, , \ n \in \mathbb{Z}\} \subset \mathrm{Cent}\,(f)$.

1.2 Linearization

Let R_λ denote the germ $R_\lambda(z) = \lambda z$. This is the simplest element of G_λ. It is easy to check that, if λ is not a root of unity, its centralizer is $\mathrm{Cent}\,(R_\lambda) = \{R_\mu \, , \ \mu \in \mathbb{C}^*\}$.

Exercise 1.4 Let $f \in G_\lambda$ and assume that λ is not a root of unity. The morphism

$$
\begin{aligned}
\mu \ : \mathrm{Cent}\,(f) &\to \mathbb{C}^* \\
g &\mapsto \mu(g) := g_1 = g'(0)
\end{aligned}
$$

is injective. [Hint : this is equivalent to showing that $g \in G_1$, $g \in \mathrm{Cent}\,(f) \Rightarrow g = \mathrm{id}$. On the other hand if $g \in G_\mu$ and $g \in \mathrm{Cent}\,(f)$ one can recursively determine the power series coefficients of g : one has

$$(\lambda^n - \lambda) g_n = (\mu^n - \mu) f_n + \sum_{j=2}^{n-1} f_j \sum_{n_1 + \ldots n_j = n} g_{n_1} \cdots g_{n_j} - \sum_{j=2}^{n-1} g_j \sum_{n_1 + \ldots n_j = n} f_{n_1} \cdots f_{n_j} \ ,$$

for all $n \geq 2$.]

Definition 1.5 *A germ $f \in G_\lambda$ is* linearizable *if there exists $h_f \in G_1$ (a linearization of f) such that $h_f^{-1} f h_f = R_\lambda$, i.e. f is conjugate to (its linear part) R_λ. f is* formally linearizable *if there exists $\hat{h}_f \in \hat{G}_1$ such that $\hat{h}_f^{-1} f \hat{h}_f = R_\lambda$ (note that in this case this is a functional equation in the ring $\mathbb{C}[[z]]$ of formal power series).*

From Exercise 1.4 it follows that when λ is not a root of unity the linearization (if it exists) is unique : if h_1 and h_2 are two linearizations of the same $f \in G_\lambda$ then $h_1 h_2^{-1} \in \ker \mu$.

Our first result on the existence of linearizations will concern the case when λ is a root of unity.

Proposition 1.6 *Assume λ is a primitive root of unity of order q. A germ $f \in G_\lambda$ is linearizable if and only if $f^q = \mathrm{id}$. The same holds for a formal germ $\hat{f} \in \hat{G}_\lambda$.*

Proof. Assume that f is linearizable. Then $z = \lambda^q z = (h_f^{-1} \circ f \circ h_f)^q(z) = (h_f^{-1} \circ f^q \circ h_f)(z)$ from which one gets $f^q(z) = (h_f \circ \mathrm{id} \circ h_f^{-1})(z) = z$.

Conversely if $f^q = \mathrm{id}$ then defining $h_f^{-1} := \frac{1}{q} \sum_{j=0}^{q-1} \lambda^{-j} f^j$ one immediately checks that $h_f^{-1} \in G_1$ if $f \in G_\lambda$ (resp. $h_f^{-1} \in \hat{G}_1$ if $f \in \hat{G}_\lambda$) and $h_f^{-1} \circ f \circ h_f = R_\lambda$. □

1.3 Formal Conjugacy Classes

In the formal case, all conjugacy classes of germs whose linear part is a root of unity are well known :

Proposition 1.7 *Let λ be a primitive root of unity of order q. Let $\hat{f} \in \hat{G}_\lambda$ and assume that $\hat{f}^q \neq id$. Then there exists a unique integer $n \geq 1$ and two complex numbers $a, b \in \mathbb{C}$, $a \neq 0$, such that $\hat{f}$ is formally conjugated to*

$$P_{n,a,b,\lambda}(z) = \lambda z(1 + az^{nq} + a^2 b z^{2nq}) \ .$$

Exercise 1.8 Prove Proposition 1.7. Note that if one allows to conjugate also with homoteties then $\hat{f}$ is formally conjugated to $P_{n,c,\lambda}(z) = \lambda z(1+z^{nq}+cz^{2nq})$. [Hint : the idea of the proof is to iterate conjugations by polynomials $\varphi_j(z) = z + \beta_j z^j$ with $j \geq 2$ and suitably chosen β_j. See also [Ar3], [Be].]

But in the formal case everything is very simple :

Proposition 1.9 *Assume that λ is not a root of unity. Then $\hat{G}_\lambda$ is a conjugacy class and $\hat{G}_1$ acts freely and transitively on $\hat{G}_\lambda$.*

Proof. To see that any $f \in \hat{G}_\lambda$ is conjugate to R_λ we look for $\hat{h}_f \in \hat{G}_1$ such that $\hat{f}\hat{h}_f = \hat{h}_f R_\lambda$. We develop and solve this functional equation by recurrence : we get, for $n \geq 2$ (denoting $\hat{h}_f(z) = \sum_{n=1}^{\infty} \hat{h}_n z^n$, $\hat{h}_1 = 1$)

$$\hat{h}_n = \frac{1}{\lambda^n - \lambda} \sum_{j=2}^{n} f_j \sum_{n_1+\ldots+n_j=n} \hat{h}_{n_1} \cdots \hat{h}_{n_j} \ . \tag{1.4}$$

The action of $\hat{G}_1$ on $\hat{G}_\lambda$ is free. This follows from the fact that the only germ tangent to the identity belonging to the centralizer of $\hat{f}$ is the identity (see Exercise 1.4). Transitivity of the action is trivial : given two formal germs $\hat{f}_1$ and $\hat{f}_2$ both in $\hat{G}_\lambda$ there exist two formal linearizations $\hat{h}_1$ and $\hat{h}_2$ and clearly $\hat{f}_1 = \mathrm{Ad}_{\hat{h}_2\hat{h}_1^{-1}}(\hat{f}_2)$. □

Collecting propositions 1.6, 1.7 and 1.9 together we have a complete classification of the conjugacy classes of $\hat{G}$:

(I) if λ is not a root of unity then $\hat{G}_\lambda$ is a conjugacy class ;

(II) if $\lambda = e^{2\pi i p/q}$, $q \geq 1$, $(p,q) = 1$ then the conjugacy classes in $\hat{G}_\lambda$ are $[R_\lambda]$ and $\{[P_{n,a,b,\lambda}]\}_{a\in\mathbb{C}^*,\, b\in\mathbb{C}\,,\, n\geq 1}$.

1.4 Koenigs–Poincaré Theorem

In the holomorphic case the problem of a *complete* classification of the conjugacy classes is still open and, as Yoccoz showed, perhaps unreasonable. The first important result in the holomorphic case is the Koenigs–Poincaré Theorem :

Theorem 1.10 (Koenigs–Poincaré) *If $|\lambda| \neq 1$ then G_λ is a conjugacy class, i.e. all $f \in G_\lambda$ are linearizable.*

Proof. Since f is holomorphic around $z = 0$ there exists $c_1 > 1$ and $r \in (0,1)$ such that $|f_j| \leq c_1 r^{1-j}$ for all $j \geq 2$. Since $|\lambda| \neq 1$ there exists $c_2 > 1$ such that $|\lambda^n - \lambda|^{-1} \leq c_2$ for all $n \geq 2$.

Let $(\sigma_n)_{n\geq 1}$ be the following recursively defined sequence :

$$\sigma_1 = 1 \ , \quad \sigma_n = \sum_{j=2}^{n} \sum_{n_1+\ldots+n_j=n} \sigma_{n_1} \cdots \sigma_{n_j} \ . \tag{1.5}$$

The generating function $\sigma(z) = \sum_{n=1}^{\infty} \sigma_n z^n$ satisfies the functional equation

$$\sigma(z) = z + \frac{\sigma(z)^2}{1-\sigma(z)} \ , \tag{1.6}$$

thus $\sigma(z) = \frac{1+z-\sqrt{1-6z+z^2}}{4}$ is analytic in the disk $|z| < 3 - 2\sqrt{2}$ and bounded and continuous on its closure. By Cauchy's estimate one has $\sigma_n \leq c_3 (3-2\sqrt{2})^{1-n}$ for some $c_3 > 0$.

Since λ is not a root of unity, f is formally linearizable and the power series coefficients of its formal linearization $\hat{h}_f$ satisfy (1.4). By induction one can check that $|\hat{h}_n| \leq (c_1 c_2 r^{-1})^{n-1} \sigma_n$, thus $\hat{h}_f \in \mathbb{C}\{z\}$. □

Remark 1.11 Since the bound $|\lambda^n - \lambda|^{-1} \leq c_2$ is uniform w.r.t $\lambda \in D(\lambda_0, \delta)$, where $\lambda_0 \in \mathbb{C}^* \setminus \mathbb{S}^1$ and $\delta < \mathrm{dist}\,(\lambda_0, \mathbb{S}^1)$, the above given proof of the Poincaré–Koenigs Theorem shows that the map

$$\begin{aligned} \mathbb{C}^* \setminus \mathbb{S}^1 &\to G_1 \\ \lambda &\mapsto h_{\tilde{f}}(\lambda) \end{aligned}$$

is analytic[1] for all $\tilde{f} \in z^2\mathbb{C}\{z\}$, where $h_{\tilde{f}}(\lambda)$ is the linearization of $\lambda z + \tilde{f}(z)$.

The Poincaré–Koenigs Theorem has the following straightforward generalization :

[1] This notion needs a little comment since $\mathbb{C}\{z\}$ is a rather wild space : it is

Theorem 1.12 (Koenigs–Poincaré with parameters) *Let $r > 0$, let $f : \mathbb{D}_r^n \times \mathbb{D}_r \subset \mathbb{C}^n \times \mathbb{C} \to \mathbb{C}$, $(t,z) \mapsto f(t,z) = f_t(z)$ be an holomorphic map such that $f_0(z) = \lambda(0)z + O(z^2)$, with $|\lambda(0)| \notin \{0,1\}$. Then there exists $r_0 \in (0,r)$, a unique holomorphic function $z_0 : \mathbb{D}_{r_0}^n \to \mathbb{C}$ and a unique $h : \mathbb{D}_{r_0}^n \times \mathbb{D}_{r_0} \to \mathbb{C}$, $(t,z) \mapsto h_t(z) = h(t,z)$ holomorphic such that for $t \in \mathbb{D}_{r_0}^n$ one has the following properties :*

(i) $f_t(z_0(t)) = z_0(t)$, $f_t(z) = \lambda(t)(z - z_0(t)) + O((z - z_0(t))^2)$, $|\lambda(t)| \notin \{0,1\}$;

(ii) $h_t(0) = z_0(t)$, $h_t'(0) = \frac{\partial}{\partial z} h_t|_{z=0} = 1$;

(iii) $h_t^{-1} \circ f \circ h_t = R_{\lambda(t)}$.

Proof. (sketch) The existence of z_0 and (i) follows easily from the implicit function theorem applied to $F(t,z) = f(t,z) - z$ at the point $(t,z) = (0,0)$ (note that $F(0,0) = 0$ and $\frac{\partial}{\partial z} f_t(z)|_{(t,z)=(0,0)} = \lambda(0) - 1 \neq 0$). Therefore there exists a unique fixed point for f_t close to $z = 0$ when t is close to 0 depending analytically on t as t varies in a neighborhood of $(t,z) = (0,0)$. Then one can consider $g_t(z) = f_t(z + z_0(t)) - z_0(t)$ and apply the proof given above of the Koenigs–Poincaré Theorem to $g_t(z)$. It is easy to convince oneself that the linearizing map depends analytically on t. □

1.5 Centralizers and Linearizations

The study of centralizers generalizes the study of linearizability as the following exercises show :

Exercise 1.13 Prove that if $f \in G_\lambda$ is linearizable and λ is not a root of unity then Cent $(f) \simeq \mathbb{C}^*$. [Hint : use the fact that the centralizer of f is conjugate to the centralizer of R_λ which is completely known.]

Exercise 1.14 Prove that if $g \in$ Cent (f), $g \in G_\mu$ is linearizable and μ is not a root

---

an inductive limit of Banach spaces, thus it is a locally convex topological vector space and it is complete but it is not metrisable, thus it is not a Fréchet space (see Section 9.1). Here we simply mean that if λ varies in some relatively compact open connected subset of $\mathbb{C}^* \setminus \mathbb{S}^1$ then $h_{\tilde{f}}(\lambda)$ belongs to some fixed Banach space of holomorphic functions (e.g. the Hardy space $H^\infty(\mathbb{D}_r)$ of bounded analytic functions on the disk $\mathbb{D}_r = \{z \in \mathbb{C}\,, |z| < r\}$, where $r > 0$ is fixed and small enough) and depends analytically on λ in the usual sense.

of unity then f is linearizable. [Hint : use that $f \in \mathrm{Cent}\,(g) = \{h_g R_\nu h_g^{-1}\,,\ \nu \in \mathbb{C}^*\}$ and that ν is invariant under conjugacy.]

Exercise 1.15 Prove that if $f \in G_\lambda$ and λ is not a root of unity then f is linearizable if and only if $\mathrm{Cent}\,(f) \simeq \mathbb{C}^*$. [Hint : apply exercises 1.13, 1.4 and the Koenigs–Poincaré Theorem]

1.6 Cremer's Non–Linearizable Germs

When $|\lambda| = 1$ and λ is not a root of unity we can write

$$\lambda = e^{2\pi i\alpha} \ \text{ with } \alpha \in \mathbb{R} \setminus \mathbb{Q} \cap (-1/2, 1/2)\ , \tag{1.7}$$

and whether $f \in G_\lambda$ is linearizable or not depends crucially on the arithmetical properties of α. Let $\{x\}$ denote the fractional part of a real number x : $\{x\} = x - [x]$, where $[x]$ is the integer part of x.

Theorem 1.16 (Cremer) *If* $\limsup_{n\to+\infty} |\{n\alpha\}|^{-1/n} = +\infty$ *then there exists* $f \in G_{e^{2\pi i\alpha}}$ *which is not linearizable.*

Proof. First of all note that $\limsup_{n\to+\infty} |\{n\alpha\}|^{-1/n} = +\infty$ if and only if

$$\limsup_{n\to+\infty} |\lambda^n - 1|^{-1/n} = +\infty$$

since

$$|\lambda^n - 1| = 2|\sin(\pi n\alpha)| \in (2|\{n\alpha\}|, \pi|\{n\alpha\}|)\ .$$

Then we construct f in the following manner : for $n \geq 2$ we take $|f_n| = 1$ and we choose inductively $\arg f_n$ such that

$$\arg f_n = \arg \sum_{j=2}^{n-1} f_j \sum_{n_1+\ldots+n_j=n} \hat{h}_{n_1} \cdots \hat{h}_{n_j}\ , \tag{1.8}$$

(recall the induction formula (1.4) for the coefficients of the formal linearization of f and note that the r.h.s. of (1.8) is a polynomial in $n-2$ variables $f_2, \ldots, f_{n-1}$ with coefficients in the field $\mathbb{C}(\lambda)$). Thus

$$|\hat{h}_n| \geq \frac{|f_n|}{|\lambda^n - 1|} = \frac{1}{|\lambda^n - 1|}$$

and $\limsup_{n\to+\infty} |\hat{h}_n|^{1/n} = +\infty$: the formal linearization $\hat{h}$ is a divergent series. □

Exercise 1.17 Write the decimal expansion of an irrational number α satisfying the assumption of Cremer's Theorem.

Exercise 1.18 Show that the set of irrational numbers satisfying the assumption of Cremer's Theorem is a dense G_δ with zero Lebesgue measure (following Baire, a set is a dense G_δ if it is a countable intersection of dense open sets. These sets are "big" from the point of view of topology).

In the next Chapter we will continue our study of the problem of the existence of a linearization of germs of holomorphic diffeomorphisms. To this purpose the following "normalization" will be useful.

1.7 Normalized Germs

Let us note that there is an obvious action of $\mathbb{C}^*$ on G by homotheties :

$$(\mu, f) \in \mathbb{C}^* \times G \mapsto \mathrm{Ad}_{R_\mu} f = R_\mu^{-1} f R_\mu \ . \tag{1.9}$$

Note that this action leaves the fibers G_λ invariant by Exercise 1.3. Also, $f \in G_\lambda$ is linearizable if and only if $\mathrm{Ad}_{R_\mu} f$ is also linearizable for all $\mu \in \mathbb{C}^*$ (indeed if h_f linearizes f then $\mathrm{Ad}_{R_\mu} h_f$ linearizes $\mathrm{Ad}_{R_\mu} f$). Therefore, in order to study the problem of the existence of a linearization, it is enough to consider $G/\mathbb{C}^*$, i.e. we identify two germs of holomorphic diffeomorphisms which are conjugate by a homothety.

Consider the space S of univalent maps $F : \mathbb{D} \to \mathbb{C}$ such that $F(0) = 0$ and the projection

$$G \to S$$
$$f \mapsto F = \begin{cases} f & \text{if } f \text{ is univalent in } \mathbb{D} \\ \mathrm{Ad}_{R_r} f & \text{if } f \text{ is univalent in } \mathbb{D}_r \end{cases}$$

This map is clearly onto and two germs have the same image only if they coincide or if they are conjugate by some homothety. Thus this projection induces a bijection from $G/\mathbb{C}^*$ onto S.

In what follows we will always consider the topological space S of germs of holomorphic diffeomorphisms $f : \mathbb{D} \to \mathbb{C}$ such that $f(0) = 0$ and f is univalent in $\mathbb{D}$. We will denote

- S_λ the subspace of f such that $f'(0) = \lambda\,$;
- $S_{\mathbb{T}}$ the subspace of f such that $|f'(0)| = 1$.

Clearly the projection above induces a bijection between $G_\lambda/\mathbb{C}^*$ and S_λ.

2. Topological Stability vs. Analytic Linearizability

The purpose of this Chapter is to connect the study of the conjugacy classes of germs of holomorphic diffeomorphisms to the theory of one–dimensional conformal dynamical systems and in particular to the notion of stability of a fixed point. The extremely remarkable fact is that stability, which is a topological property, will turn out to be equivalent to linearizability, which is an analytic property.

2.1 Dynamics of Rational Maps

Let us first of all recall the notion of normal family on an open subset U of the Riemann sphere $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. To this purpose we recall the usual system of coordinates on $\overline{\mathbb{C}}$ determined by the stereographic projection : $z : \overline{\mathbb{C}} \setminus \{\infty\} \to \mathbb{C}$, $z(0) = 0$ and $w : \overline{\mathbb{C}} \setminus \{0\} \to \mathbb{C}$, $w(\infty) = 0$, related by $zw = 1$. The *spherical metric* on $\overline{\mathbb{C}}$ is defined as follows :

$$ds_{\overline{\mathbb{C}}} = \begin{cases} \frac{2|dz|}{1+|z|^2} & \text{in the } z\text{–chart ;} \\ \frac{2|dw|}{1+|w|^2} & \text{in the } w\text{–chart ;} \end{cases} \tag{2.1}$$

Let $U \subset \overline{\mathbb{C}}$ be open and $\mathcal{F}_U = \{f : U \to \overline{\mathbb{C}},\ f \text{ meromorphic}\}$. We endow $\overline{\mathbb{C}}$ with the spherical metric and $\mathcal{F}_U$ with the topology of uniform convergence on compact subsets of U. It is a classical result of Weierstrass that the limit of a convergent sequence in $\mathcal{F}_U$ still belongs to $\mathcal{F}_U$ (note that the constant function $f \equiv \infty$ is considered meromorphic).

Definition 2.1 *A family $\mathcal{F} \subset \mathcal{F}_U$ is* normal *if it is relatively compact in $\mathcal{F}_U$, i.e. any sequence $\{f_n\} \subset \mathcal{F}$ contains a subsequence which converges uniformly in the spherical metric on compact subsets of U.*

Warning ! If $\{f_n\}$ is a normal family then $\{f'_n\}$ needs not be normal : e.g. $f_n(z) = n(z^2 - n)$ on $\mathbb{C}$.

By means of the Ascoli–Arzelà theorem one gets :

Proposition 2.2

(I) A family of meromorphic functions on U is normal on U if and only if it is equicontinuous on every compact subset of U ;

(II) A family of analytic functions on U is normal on U if and only if it is locally uniformly bounded (i.e. uniformly bounded on every compact subset of U).

Proof. The first statement is obvious since the compactness of $\overline{\mathbb{C}}$ guarantees that the family is uniformly bounded. The second statement follows from Cauchy's integral theorem. $\square$

The notion of normal family allows us to introduce the basic notions of one–dimensional holomorphic dynamics. Here we are interested in studying the dynamics of a discrete dynamical system (i.e. an action of $\mathbb{N}$) on the Riemann sphere $\overline{\mathbb{C}}$ generated by a holomorphic transformation $R : \overline{\mathbb{C}} \to \overline{\mathbb{C}}$, i.e. and element of $\mathrm{End}\,(\overline{\mathbb{C}})$.

Let d denote the topological degree of R. We will assume $d \geq 2$ thus R is a d–fold branched covering of the Riemann sphere and can be written in a unique way in the form $R(z) = \frac{P(z)}{Q(z)}$, where $P(z) \in \mathbb{C}[z]$, $Q(z) \in \mathbb{C}[z]$ have no common factors and $d = \max(\deg P, \deg Q)$. In fact every $d : 1$ conformal branched covering of $\overline{\mathbb{C}}$ comes from some such rational function and

$$\mathrm{End}\,(\overline{\mathbb{C}}) = \{R : \overline{\mathbb{C}} \to \overline{\mathbb{C}}\,\text{holomorphic}\} = \{R : \overline{\mathbb{C}} \to \overline{\mathbb{C}}\,,\ R(z) = P(z)/Q(z)\}\,.$$

Note that

$$R(z) = \begin{cases} \frac{P(z)}{Q(z)} & \text{if } Q(z) \neq 0, \\ \infty & \text{if } Q(z) = 0, \\ \lim_{z\to\infty} \frac{P(z)}{Q(z)} & \text{if } z = \infty. \end{cases}$$

We define the iterates R^n of R as usual : $R^n = R \circ R^{n-1}$. Note that R^n has degree d^n.

Given a point $z_0 \in \overline{\mathbb{C}}$ the sequence of points $\{z_n\}_{n\geq 0}$ defined by $z_{n+1} = R(z_n)$ is called the orbit of z_0. A point z_0 is a *fixed point* of R if $R(z_0) = z_0$, *periodic* if $z_n = R^n(z_0) = z_0$ for some n (the minimal n is the period). The orbit $\{z_1, \ldots, z_n = z_0\}$ is called a *cycle*. The point z_0 is called *preperiodic* if z_k is periodic for some $k > 0$.

The fundamental dichotomy of $\overline{\mathbb{C}}$ associated to the dynamics of R is the following :

Definition 2.3 *The* Fatou set *$F(R)$ of R is the set of points $z_0 \in \overline{\mathbb{C}}$ such that $\{R^n\}_{n\geq 0}$ is a normal family in some disk $D(z_0, r)$ (w.r.t. the spherical metric). The complement of the Fatou set is the* Julia set *$J(R)$.*

Exercise 2.4 Show that $J(R) = J(R^n)$ for all $n \geq 1$ and that $J(R)$ is nonempty and closed. [Hint : if $J(R) = \emptyset$ then $\{R^n\}_{n\geq 0}$ is a normal family on all $\overline{\mathbb{C}}$ thus $R^{n_j} \to S \in \mathrm{End}\,(\overline{\mathbb{C}})$. Compare degrees.]

2.2 Stability

Definition 2.5 *A point $z_0 \in \overline{\mathbb{C}}$ is* stable *if for all $\delta > 0$ there exists a neighborhood W of z_0 such that for all $z \in W$ and for all $n \geq 0$ one has $d(R^n(z), R^n(z_0)) \leq \delta$ (here, as usual, d denotes the spherical metric).*

Exercise 2.6 Show that a point is stable if and only if it belongs to the Fatou set.

Exercise 2.7 Let $R \in \mathrm{End}\,(\overline{\mathbb{C}})$ and assume that $R(0) = 0$. If R is linearizable and $R'(0) = \lambda$, $|\lambda| \leq 1$ then 0 is a stable fixed point.

If we consider the more general situation of a germ $f \in S$, i.e. $f : \mathbb{D} \to \mathbb{C}$ injectively and holomorphically, $f(0) = 0$, the definition of stability must be slightly generalized so as to take into account the fact that the iterates of f are not necesaarily defined for all n.

Definition 2.8 *0 is* stable *if and only if there exists a neighborhood U of 0 such that f^n is defined on U for all $n \geq 0$ and for all $z \in U$ and $n \geq 0$ one has $|f^n(z)| < 1$.*

Exercise 2.9 Show that if f is a rational map with a fixed point 0 then definitions 2.8 and 2.5 are equivalent.

Exercise 2.10 If $f'(0) = \lambda$ and $|\lambda| > 1$ then 0 is not stable.

To each germ $f \in S$, $|f'(0)| \leq 1$, one can associate a natural f–invariant compact set

$$0 \in K_f := \bigcap_{n\geq 0} f^{-n}(\mathbb{D}) \ . \tag{2.2}$$

Let U_f denote the connected component of the interior of K_f which contains 0. Then 0 is stable if and only if $U_f \neq \emptyset$, i.e. if and only if 0 belongs to the interior of K_f.

Exercise 2.11 Show that if $f \in S$ and $|f'(0)| < 1$ then 0 is stable. [Hint : consider a small disk around 0 on which the inequality $|f(z)| \leq \rho|z|$ with $\rho < 1$ holds.]

2.3 Stability vs. Linearizability

The main result of this section is the equivalence (for $|f'(0)| \leq 1$) of stability (a topological notion) and linearizability (an analytical notion) :

Theorem 2.12 *Let $f \in S$, $|f'(0)| \leq 1$. 0 is stable if and only if f is linearizable.*

Proof. The statement is non–trivial only if $\lambda = f'(0)$ has unit modulus. If f is linearizable then the linearization h_f maps a small disk $\mathbb{D}_r$ around zero conformally into $\mathbb{D}$. Since $h_f(0) = 0$ and $|f^n(z)| < 1$ for all $z \in h_f(\mathbb{D}_r)$ one sees that 0 is stable.

Conversely assume now that 0 is stable. Then $U_f \neq \emptyset$ and one can easily see that it must also be simply connected (otherwise, if it had a hole V, surrounding it with some closed curve γ contained in U_f since $|f^n(z)| < 1$ for all $z \in \gamma$ and $n \geq 0$ the maximum principle leads to the same conclusion for all the points in V thus $V \subset U_f$). Applying the Riemann mapping theorem to U_f one sees that by conjugation with the Riemann map f induces a univalent map g of the disk into itself with the same linear part λ. By Schwarz' Lemma one must have $g(z) = \lambda z$ thus f is analytically linearizable. □

When $\lambda = f'(0)$ has modulus one, is not a root of unity and 0 is stable then U_f is conformally equivalent to a disk and is called the *Siegel disk* of f (at 0). Thus the Siegel disk of f is the maximal connected open set containing 0 on which f is conjugated to R_λ. The conformal representation $\tilde{h}_f : \mathbb{D}_{c(f)} \to U_f$ of U_f which satisfies $\tilde{h}_f(0) = 0$, $\tilde{h}'_f(0) = 1$ linearizes f thus the power series of $\tilde{h}_f$ and h_f coincide. If $r(f)$ denotes the radius of convergence of the linearization h_f (whose power series coefficients are recursively determined as in (1.4)), recalling the definition of conformal capacity (Exercise A1.4, Appendix 1) we see that :

(i) if $r(f) > 0$ then $0 < c(U_f, 0) = c(f) \leq r(f)$;

(ii) if $r(f) = 0$ then $c(f) = r(f) = 0$.

Exercise 2.13 Show that the map $c : S_{\mathbb{T}} \to [0, 1]$ which associates to each germ f the conformal capacity of U_f w.r.t. 0 is upper semicontinuous.

When 0 is not stable and λ is not a root of unity K_f is called a *hedgehog.*

We conclude this introduction to Siegel disks with two results on their conformal capacity.

Proposition 2.14 *One has $c(f) = r(f)$ when at least one of the two following conditions is satisfied :*
(i) U_f is relatively compact in $\mathbb{D}$;
(ii) each point of $\mathbb{S}^1$ is a singularity of f.

Proposition 2.15 *Let $\lambda \in \mathbb{T}$ and assume that λ is not a root of unity. Then*

$$\inf_{f \in S_\lambda} c(f) = \inf_{f \in S_\lambda} r(f) \ .$$

For the proofs see [Yo2, p.19].

3. The Quadratic Polynomial. Yoccoz's Proof of Siegel Theorem

In this Chapter we will show the special role played by the quadratic polynomial

$$P_\lambda(z) = \lambda\left(z - \frac{z^2}{2}\right) . \tag{3.1}$$

Indeed P_λ is the "worst possible perturbation of the linear part R_λ" as the following theorem shows

Theorem 3.1 (Yoccoz) *Let $\lambda = e^{2\pi i\alpha}$, $\alpha \in \mathbb{R}\setminus\mathbb{Q}$. If P_λ is linearizable then every germ $f \in G_\lambda$ is also linearizable.*

Proof. Let $f \in G_\lambda$, $f(z) = \lambda z + \sum_{n=2}^\infty f_n z^n$. By conjugating with some homothety one has $|f_n| \le 10^{-3}4^{-n}$. We now consider the one–parameter family $f_b(z) = \lambda z + bz^2 + \sum_{n=2}^\infty f_n z^n$. Note that $f_0 = f$. Since λ is not a root of unity there exists a unique formal germ $\hat{h}_b \in \hat{G}_1$ such that $\hat{h}_b^{-1} f_b \hat{h}_b = R_\lambda$. Its power series expansion is $\hat{h}_b(z) = z + \sum_{n=2}^\infty h_n(b) z^n$ with $h_n(b) \in \mathbb{C}[b]$. Thus by the maximum principle one has $|h_n(0)| \le \max_{|b|=1/2} |h_n(b)|$. If $|b| = 1/2$ then (possibly after conjugation with a rotation) $f_b(z) = P_\lambda(z) + \sum_{n=2}^\infty f_n z^n = P_\lambda(z) + \psi(z)$ and it is immediate to check that $\sup_{z\in\mathbb{D}_3} |\psi(z)| < 10^{-2}$. From Douady–Hubbard's theorem on the stability of the quadratic polynomial (Appendix 1) it follows that f_b is quasiconformally conjugated to P_λ. If P_λ is linearizable then 0 is stable for P_λ thus also for f_b since a quasiconformal conjugacy is in particular a topological conjugacy. But we know that this implies that f_b is linearizable. Therefore there exists two positive constants C and r such that $|h_n(b)| \le Cr^{-n}$ for all b of modulus 1/2, thus $|h_n(0)| \le Cr^{-n}$. Then $\hat{h}_0$ converges and $f_0 = f$ is linearizable. □

3.1 Yoccoz's Linearization Theorem for the Quadratic Polynomial

Once one has established that the linearizability of the quadratic polynomial for a certain λ implies that G_λ is a conjugacy class the following remarkable theorem of Yoccoz shows that G_λ is a conjugacy class for almost all $\lambda \in \mathbb{T}$.

Theorem 3.2 *Let $\lambda = e^{2\pi i\alpha}$, $\alpha \in \mathbb{R}\setminus\mathbb{Q}$. For almost all $\lambda \in \mathbb{T}$ the quadratic polynomial P_λ is linearizable.*

This statement deserves a comment. As we will see in Chapter 5 this theorem of Yoccoz is indeed weaker than the Siegel [S] and Brjuno [Br] theorems which date respectively to 1942 and 1970. What is *very remarkable* is the proof of Theorem 3.2 which does not need any subtle estimate on the growth of the coefficients of the formal linearization as provided by (1.4) : compare with the proof of the Siegel–Brjuno Theorem given in Section 5.1.

Let us note that P_λ has a unique critical point $c = 1$ (apart from $z = \infty$) and that the corresponding critical value is $v_\lambda = P_\lambda(c) = \lambda/2$. If $|\lambda| < 1$ by Koenigs–Poincaré theorem we know that there exists a unique analytic linearization H_λ of P_λ and that it depends analytically on λ as λ varies in $\mathbb{D}$. Let $r_2(\lambda)$ denote the radius of convergence of H_λ. One has the following

Proposition 3.3 *Let* $\lambda \in \mathbb{D}$. *Then :*

(1) $r_2(\lambda) > 0$ *;*

(2) $r_2(\lambda) < +\infty$ *and* H_λ *has a continuous extension to* $\overline{\mathbb{D}_{r_2(\lambda)}}$. *Moreover the map* $H_\lambda : \overline{\mathbb{D}_{r_2(\lambda)}} \to \mathbb{C}$ *is conformal and verifies* $P_\lambda \circ H_\lambda = H_\lambda \circ R_\lambda$.

(3) On its circle of convergence $\{z\,,\, |z| = r_2(\lambda)\}$, H_λ *has a unique singular point which will be denoted* $u(\lambda)$.

(4) $H_\lambda(u(\lambda)) = 1$ *and* $(H_\lambda(z) - 1)^2$ *is holomorphic in* $z = u(\lambda)$.

Proof. The first assertion is just a consequence of Koenigs–Poincaré theorem.

The functional equation $P_\lambda(H_\lambda(z)) = H_\lambda(\lambda z)$ is satisfied for all $z \in \mathbb{D}_{r_2(\lambda)}$. Moreover $H_\lambda : \mathbb{D}_{r_2(\lambda)} \to \mathbb{C}$ is univalent (if one had $H_\lambda(z_1) = H_\lambda(z_2)$ with $z_1 \neq z_2$ and $z_1, z_2 \in \mathbb{D}_{r_2(\lambda)}$ one would have $H_\lambda(\lambda^n z_1) = H_\lambda(\lambda^n z_2)$ for all $n \geq 0$ which is impossible since $|\lambda| < 1$ and $H'_\lambda(0) = 1$). Thus $r_2(\lambda) < +\infty$. On the other hand if H_λ is holomorphic in $\mathbb{D}_r$ for some $r > 0$ and the critical value $v_\lambda \notin H_\lambda(\mathbb{D}_r)$ the functional equation allows to continue analytically H_λ to the disk $\mathbb{D}_{|\lambda|^{-1}r}$. Therefore there exists $u(\lambda) \in \mathbb{C}$ such that $|u(\lambda)| = r_2(\lambda)$ and $H_\lambda(\lambda u(\lambda)) = v_\lambda$. Such a $u(\lambda)$ is unique since H_λ is injective on $\mathbb{D}_{r_2(\lambda)}$. If $|w| = |\lambda| r_2(\lambda)$ and $w \neq \lambda u(\lambda)$ one has $H_\lambda(w) = P_\lambda(H_\lambda(\lambda^{-1} w))$ and

$$H_\lambda(\lambda^{-1} w) = 1 - \sqrt{1 - 2\lambda^{-1} H_\lambda(w)} \tag{3.2}$$

which shows how to extend continuously and injectively H_λ to $\overline{\mathbb{D}_{r_2(\lambda)}}$. By construction the functional equation is trivially verified. This completes the proof of (2).

To prove (3) and (4) note that from $H_\lambda(\lambda u(\lambda)) = P_\lambda(H_\lambda(u(\lambda)))$ it follows that $H_\lambda(u(\lambda)) = 1$. Formula (3.2) shows that all points $z \in \mathbb{C}$, $|z| = r_2(\lambda)$ are regular except for $z = u(\lambda)$. Finally one has $(H_\lambda(z)-1)^2 = 1-2\lambda^{-1}H_\lambda(\lambda z)$ which is holomorphic also at $z = u(\lambda)$. □

The fact that H_λ is injective on $\overline{\mathbb{D}}_{r_2(\lambda)}$ implies that $r_2(\lambda) < +\infty$ (otherwise it would be a biholomorphism of $\mathbb{C}$ thus an affine map). A more precise upper bound is provided by the following

Lemma 3.4 (a priori estimate of $r_2(\lambda)$). $r_2(\lambda) \le 2$.

Proof. It is an easy consequence of Koebe 1/4–Theorem. Indeed if $\tilde{f} \in S_1$ and $t > 0$ then $f = R_t^{-1}\tilde{f}R_t : \mathbb{D}_{t^{-1}} \to \mathbb{C}$ is univalent and $f(\mathbb{D}_{t^{-1}}) = R_{t^{-1}}\tilde{f}(\mathbb{D})$. By Koebe 1/4–Theorem one has $\mathbb{D}_{1/4} \subset \tilde{f}(\mathbb{D})$ thus $\mathbb{D}_{t^{-1}/4} \subset f(\mathbb{D}_{t^{-1}})$. But we know that $v_\lambda \notin H_\lambda(\mathbb{D}_{|\lambda| r_2(\lambda)})$ thus $|v_\lambda| = \frac{|\lambda|}{2} \ge \frac{|\lambda| r_2(\lambda)}{4}$. □

Exercise 3.5 Show that $r_2(\lambda) \le 8/7$. [Hint : apply (A1.1) to the function

$$\tilde{H}_\lambda(z)\left(1 + \frac{2r_2(\lambda)}{\lambda}\tilde{H}_\lambda(z)\right)^{-1} \in S_1 ,$$

where $\tilde{H}_\lambda(z) = \frac{H_\lambda(r_2(\lambda)z)}{r_2(\lambda)}$.]

Exercise 3.6 Show that the image by H_λ of its circle of convergence is a Jordan curve, analytic except at $H_\lambda(u(\lambda)) = 1$ where it has a right angle.

Proposition 3.7 *$u : \mathbb{D}^* \to \mathbb{C}$ has a bounded analytic extension to $\mathbb{D}$. Moreover it is the limit of the sequence of polynomials $u_n(\lambda) = \lambda^{-n}P_\lambda^n(1)$ uniformly on compact subsets of $\mathbb{D}$. One has $u(0) = 1/2$.*

Proof. By Proposition 3.3 one has $P_\lambda^n(1) = H_\lambda(\lambda^n u(\lambda))$. From Lemma 3.4 and Koebe's distorsion estimates (specifically (A1.4)) applied to $\tilde{H}_\lambda(z) = \frac{H_\lambda(u(\lambda)z)}{u(\lambda)}$ one has

$$|P_\lambda^n(1)| = |u(\lambda)\tilde{H}_\lambda(\lambda^n)| \le r_2(\lambda)\frac{|\lambda|^n}{(1-|\lambda|^n)^2} \le 2\frac{|\lambda|^n}{(1-|\lambda|^n)^2} ,$$

thus $|u_n(\lambda)| \le 2(1-|\lambda|)^{-2}$ for all $\lambda \in \mathbb{D}$ and the polynomials u_n verify the recurrence relation

$$u_0(\lambda) = 1 , \quad u_{n+1}(\lambda) = u_n(\lambda) - \frac{\lambda^n}{2}(u_n(\lambda))^2 . \tag{3.3}$$

This shows that u_n converges uniformly on compact subsets of $\mathbb{D}$. The limit is u since

$$\lim_{n\to+\infty} u_n(\lambda) = \lim_{n\to+\infty} \lambda^{-n} H_\lambda(\lambda^n u(\lambda)) = u(\lambda)\ .$$

Finally from $|u(\lambda)| = r_2(\lambda)$ and Lemma 3.4 one has $|u(\lambda)| \le 2$ on $\mathbb{D}$. □

The function $u : \mathbb{D} \to \mathbb{C}$ will be called *Yoccoz's function*. It has many remarkable properties and it is the object of various conjectures (see Section 5.3).

Exercise 3.8 Check that :

1) $u(\lambda) - u_n(\lambda) = \mathrm{O}(\lambda^n)$;
2) $u(\lambda) = \frac{1}{2} - \frac{\lambda}{8} - \frac{\lambda^2}{8} - \frac{\lambda^3}{16} - \frac{9\lambda^4}{128} - \frac{\lambda^5}{128} - \frac{7\lambda^6}{128} + \frac{3\lambda^7}{256} - \frac{29\lambda^8}{1024} - \frac{\lambda^9}{256} + \frac{25\lambda^{10}}{2048} + \frac{559\lambda^{11}}{32768} + \ldots$;
3) $u(\lambda) \in \mathbb{Q}\{\lambda\}$ and all the denominators are a power of 2.

Write a computer program to calculate the power series expansion of u and use it to design the level sets of $\log|u|$ and $\arg u$. Try to compute the graph of $\theta \mapsto \arg u(re^{2\pi i\theta})$ as $r \to 1-$. You may use some formulas given in [Yo2, pp. 70–71] and to compare with [MMY2]. If you get nice pictures I would like to get a copy of them.

3.2 Radial Limits of Yoccoz's Function. Conclusion of the Proof

Proposition 3.9 *Let $\lambda_0 \in \mathbb{T}$ and assume that λ_0 is not a root of unity. Then $r_2(\lambda_0) \ge \limsup_{\mathbb{D}\ni\lambda\to\lambda_0} |u(\lambda)|$.*

Proof. Let $r = \limsup_{\mathbb{D}\ni\lambda\to\lambda_0} |u(\lambda)|$. It is not restrictive to assume $r > 0$. Let $(\lambda_n)_{n\ge1} \subset \mathbb{D}$ such that $\lambda_n \to \lambda_0$ and $|u(\lambda_n)| \to r$. Since the linearizations H_{λ_n} are univalent on their disks of convergence $\mathbb{D}_{r_2(\lambda_n)}$ one can extract a subsequence uniformly convergent on the compact susbets of $\mathbb{D}_r$. The limit function H verifies $H(0) = 0$, $H'(0) = 1$ and $H(\lambda_0 z) = P_{\lambda_0}(H(z))$ (this is immediate by taking the limit of the corresponding equations for λ_n). Thus $H_{\lambda_0} = H$ and $r_2(\lambda_0) \ge r$. □

Yoccoz has indeed proved the following stronger result [Yo2, pp. 65-69]

Theorem 3.10 *For all $\lambda_0 \in \mathbb{T}$, $|u(\lambda)|$ has a non–tangential limit in λ_0 which is equal to the radius of convergence $r_2(\lambda_0)$ of H_{λ_0}.*

Of course, if λ_0 is a root of unity then P_{λ_0} is not even formally linearizable and one poses $r_2(\lambda_0) = 0$.

Collecting Propositions 3.3, 3.7 and 3.9 together one can finally prove Theorem 3.2.

Proof. of Theorem 3.2. Applying Fatou's Theorem to $u : \mathbb{D} \to \mathbb{C}$ one finds that there exists $u^* \in L^\infty(\mathbb{T}, \mathbb{C})$ such that for almost all $\lambda_0 \in \mathbb{T}$ one has $|u^*(\lambda_0)| > 0$ and $u(\lambda) \to u(\lambda_0)$ as $\lambda \to \lambda_0$ non tangentially. From Proposition 3.7 one concludes that for almost all $\lambda_0 \in \mathbb{T}$ one has $r_2(\lambda_0) > 0$. □

Remark 3.11 Continuing the above argument of Yoccoz, L. Carleson and P. Jones prove that for almost all $\lambda \in \mathbb{T}$ the critical point $z = 1$ of P_λ belongs to the boundary of the Siegel disk (see, for example, [CG]). This has also been proved directly by M. Herman under the assumption that α is diophantine [He3]. M. Herman has also shown that there are λ's for which the critical point is not on the boundary of the Siegel disk (even though the boundary is a quasicircle) [Do].

4. Douady–Ghys' Theorem. Continued Fractions and the Brjuno Function

From Yoccoz's theorem it follows that G_λ, $\lambda = e^{2\pi i\alpha}$, $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, is a conjugacy class for almost all values of α. Let $\mathcal{Y}$ denote the set of $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ such that $G_{e^{2\pi i\alpha}}$ is a conjugacy class. Then we already know that $\mathcal{Y}$ has full measure but that its complement in $\mathbb{R} \setminus \mathbb{Q}$ is a G_δ–dense (Exercise 1.18). The goal of this Section is to prove a result due to Douady and Ghys on the structure of $\mathcal{Y}$ (Section 4.1) and to introduce various sets of irrational numbers (Sections 4.3 and 4.4) which have the same properties of $\mathcal{Y}$. Our main tool will be the use of continued fractions (see Appendix 2 for a short introduction).

4.1 Douady–Ghys' Theorem

We recall that $\mathrm{SL}(2,\mathbb{Z})$ is the group of matrices $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with integer coefficients a, b, c, d such that $ad - bc = 1$. It acts on $\mathbb{R} \cup \{\infty\}$ (thus on $\mathcal{Y}$ too) as usual : $g \cdot \alpha = \frac{a\alpha+b}{c\alpha+d}$. $\mathrm{SL}(2,\mathbb{Z})$ is generated by $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $T \cdot \alpha = \alpha + 1$, and $U = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, $U \cdot \alpha = -1/\alpha$.

Further information on the structure of $\mathcal{Y}$ is provided by the following

Theorem 4.1 (Douady–Ghys) *$\mathcal{Y}$ is $SL(2,\mathbb{Z})$–invariant.*

Proof. (sketch). $\mathcal{Y}$ is clearly invariant under T, thus we only need to show that if $\alpha \in \mathcal{Y}$ then also $U \cdot \alpha = -1/\alpha \in \mathcal{Y}$.

Let $f \in S_{e^{2\pi i\alpha}}$ and consider a domain V' bounded by

1) a segment l joining 0 to $z_0 \in \mathbb{D}^*$, $l \subset \mathbb{D}$;
2) its image $f(l)$;
3) a curve l' joining z_0 to $f(z_0)$.

We choose l' and z_0 (sufficiently close to 0) so that l, l' and $f(l)$ do not intersect except at their extremities. Note that l and $f(l)$ form an angle of $2\pi\alpha$ at 0. Then glueing l to $f(l)$ one obtains a topological manifold $\overline{V}$ with boundary which is homeomorphic to $\overline{\mathbb{D}}$. With the induced complex structure its interior is biholomorphic to $\mathbb{D}$. Let us now consider the *first return map* $g_{V'}$ to the domain V' (this is well defined if z is choosen with $|z|$ small enough) : if $z \in V'$ (and $|z|$ is small enough) we define $g_{V'}(z) = f^n(z)$ where n (depends on z) is defined asking that

$f(z), \ldots, f^{n-1}(z) \notin V'$ and $f^n(z) \in V'$, i.e. $n = \inf\{k \in \mathbb{N}, k \geq 1, f^k(z) \in V'\}$. Then it is easy to check that $n = \left[\frac{1}{\alpha}\right]$ or $n = \left[\frac{1}{\alpha}\right] + 1$. The first return map $g_{V'}$ induces a map $g_{\overline{V}}$ on a neighborhood of $0 \in \overline{V}$ and finally a germ g of holomorphic diffeomorphism at $0 \in \mathbb{D}$ ($gp = pg_{\overline{V}}$, where p is the projection from $\overline{V}$ to the disk $\mathbb{D}$). It is easy to check that $g(z) = e^{-2\pi i/\alpha} z + \mathrm{O}(z^2)$ (note that in the passage from V' to $\mathbb{D}$ through $\overline{V}$ the angle $2\pi\alpha$ at the origin is mapped in 2π).

To each orbit of f near 0 corresponds an orbit of g near 0. In particular

- f is linearizable if and only if g is linearizable ;
- if f has a periodic orbit near 0 then also g has a periodic orbit ;
- if f has a point of instability (i.e. a point which does not belong to K_f) then also g has a point of instability (which, after having normalized g so as to be univalent on $\mathbb{D}$, will leave the unit disk even more rapidly).

In particular these statements show that $\alpha \in \mathcal{Y}$ if and only if $-1/\alpha \in \mathcal{Y}$. □

4.2 SL(2,Z) and Continued Fractions

To better understand the action of $\mathrm{SL}\,(2, \mathbb{Z})$ on $\mathbb{R} \setminus \mathbb{Q}$ we can introduce a fundamental domain $[0, 1)$ for one of the two generators (the translation T) and restrict our attention to the inversion $\alpha \mapsto 1/\alpha$ restricted to $[0, 1)$. This gives us a "microscope" since $\alpha \mapsto 1/\alpha$ is expanding on $[0, 1)$, i.e. its derivative is always greater than 1. Our microscope magnifies more and more as $\alpha \to 0+$ and leads to the introduction of continued fractions discussed in Appendix A2.

Exercise 4.2 Show that given any pair $x, y \in \mathbb{R} \setminus \mathbb{Q}$ there exists $g \in \mathrm{SL}\,(2, \mathbb{Z})$ such that $x = g \cdot y$ *if and only if* $x = [a_0, a_1, \ldots, a_m, c_0, c_1, \ldots]$ and $y = [b_0, b_1, \ldots, b_n, c_0, c_1, \ldots]$. [Hint : it is easy to check that the condition is sufficient for having $x = g \cdot y$; necessity is more tricky, see [HW] pp. 141–143.]

Exercise 4.3 Show that if x is a quadratic irrational, i.e. $x \in \mathbb{R} \setminus \mathbb{Q}$ is a zero of a monic quadratic polynomial equation with coefficients in $\mathbb{Q}$, then there exists $N \in \mathbb{N}$ such that the partial fractions a_n of x are bounded $a_n \leq N$ for all $n \geq 0$.

The two main results which make continued fractions so useful in the study of one–dimensional small divisors problems are the following

Theorem 4.4 (Best approximation) *Let $x \in \mathbb{R} \setminus \mathbb{Q}$ and let p_n/q_n denote its n–th convergent. If $0 < q < q_{n+1}$ then $|qx - p| \geq |q_n x - p_n|$ for all $p \in \mathbb{Z}$ and* *equality can occur only if $q = q_n$, $p = p_n$.*

Theorem 4.5 *If $\left| x - \frac{p}{q} \right| < \frac{1}{2q^2}$ then $\frac{p}{q}$ is a convergent of x.*

For the proofs see [HW], respectively Theorems 182, p. 151 and 184, p. 153.

4.3 Classical Diophantine Conditions

Let $\gamma > 0$ and $\tau \geq 0$ be two real numbers.

Definition 4.6 *$x \in \mathbb{R} \setminus \mathbb{Q}$ is* diophantine *of exponent τ and constant γ if and only if for all $p, q \in \mathbb{Z}$, $q > 0$, one has $\left| x - \frac{p}{q} \right| \geq \gamma q^{-2-\tau}$.*

Remark 4.7 Note that Theorem 4.5 implies that given any irrational number there are infinitely many solutions to $\left| x - \frac{p}{q} \right| < \frac{1}{q^2}$ with p and q coprime. This explains why the previous definition would never be satisfied if $\tau < 0$.

We denote $\mathrm{CD}\,(\gamma, \tau)$ the set of all irrationals x such that $\left| x - \frac{p}{q} \right| \geq \gamma q^{-2-\tau}$ for all $p, q \in \mathbb{Z}$, $q > 0$. $\mathrm{CD}\,(\tau)$ will denote the union $\cup_{\gamma > 0} \mathrm{CD}\,(\gamma, \tau)$ and $\mathrm{CD} = \cup_{\tau \geq 0} \mathrm{CD}\,(\tau)$.

Exercise 4.8 Show that

$$\begin{aligned} \mathrm{CD}\,(\tau) &= \{x \in \mathbb{R} \setminus \mathbb{Q} \mid \; q_{n+1} = \mathrm{O}(q_n^{1+\tau})\} = \{x \in \mathbb{R} \setminus \mathbb{Q} \mid \; a_{n+1} = \mathrm{O}(q_n^{\tau})\} \\ &= \{x \in \mathbb{R} \setminus \mathbb{Q} \mid \; x_n^{-1} = \mathrm{O}(\beta_{n-1}^{-\tau})\} = \{x \in \mathbb{R} \setminus \mathbb{Q} \mid \; \beta_n^{-1} = \mathrm{O}(\beta_{n-1}^{-1-\tau})\} \end{aligned}$$

Exercise 4.9 Show that if x is an algebraic number of degree $n \geq 2$, i.e. $x \in \mathbb{R} \setminus \mathbb{Q}$ is a zero of a monic polynomial with coefficients in $\mathbb{Q}$ and degree n, then $x \in \mathrm{CD}\,(n-2)$ (Liouville's theorem). Thue improved this result in 1909 showing that $x \in \mathrm{CD}\,(\tau - 1 + n/2)$ for all $\tau > 0$ (see [ST], Chapter V, for a very nice discussion of the proof in the cubic case). Actually one can prove that if x is algebraic then $x \in \mathrm{CD}\,(\tau)$ for all $\tau > 0$ regardless of the degree, but this is difficult (Roth's theorem).

Exercise 4.10 Using the fact that the continued fraction of $e = \sum_{n=0}^{\infty} \frac{1}{n!}$ is

$$[2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, \ldots]$$

show that $e \in \cap_{\tau>0} \mathrm{CD}\,(\tau)$. A proof of the continued fraction expansion of e, which is due to L. Euler, can be found in [L], Chapter V. Perhaps you may like to try to obtain it yourself starting from the knowledge of the continued fraction of $\frac{e+1}{e-1} = [2, 6, 10, 14, \ldots]$.

Exercise 4.11 Use the result of Exercise 4.9 to exhibit explicit examples of trascendental numbers, e.g. $x = \sum_{n=0}^{\infty} 10^{-n!}$.

The complement in $\mathbb{R} \setminus \mathbb{Q}$ of CD is called the set of Liouville numbers.

Exercise 4.12 Show that $\mathrm{CD}\,(\tau)$ and CD are both $\mathrm{SL}\,(2, \mathbb{Z})$–invariant.

Proposition 4.13 *For all $\gamma > 0$ and $\tau > 0$ the Lebesgue measure of $CD\,(\gamma, \tau)(mod\, 1)$ is at least $1 - 2\gamma\zeta(1 + \tau)$, where ζ denotes the Riemann zeta function.*

Proof. The complement of $\mathrm{CD}\,(\gamma, \tau)(\mathrm{mod}\, 1)$ is contained in

$$\cup_{p/q \in \mathbb{Q} \cap [0,1]} \left(\frac{p}{q} - \gamma q^{-2-\tau}, \frac{p}{q} + \gamma q^{-2-\tau} \right)$$

whose Lebesgue measure is bounded by

$$\sum_{q=1}^{\infty} \sum_{p=1}^{q} 2\gamma q^{-2-\tau} \leq 2\gamma \sum_{q=1}^{\infty} q^{-1-\tau} \ .$$

□

From the point of view of dimension one has (see [Fa], p. 142 for a proof)

Theorem 4.14 (Jarnik) *Let $\tau > 0$ and let F_τ be the set of real numbers $x \in [0, 1]$ such that $\{qx\} \leq q^{-1-\tau}$ for infinitely many positive integers q. The Hausdorff dimension of F_τ is $2/(2 + \tau)$.*

Exercise 4.15 The set $\mathrm{CD}\,(0)$ is also called the set of numbers of *constant type* since $x \in \mathrm{CD}\,(0)$ if and only if the sequence of its partial fractions is bounded. Show that $\mathrm{CD}\,(0)$ has Hausdorff dimension 1 and zero Lebesgue measure.

Exercise 4.16 Show that the set of Liouville numbers has zero Lebesgue measure, zero Hausdorff dimension but it is a dense G_δ–set

4.4 Brjuno Numbers and the Brjuno Function

Let $x \in \mathbb{R}\backslash\mathbb{Q}$, let $\left(\frac{p_n}{q_n}\right)_{n\geq 0}$ denote the sequence of its convergents and let $(\beta_n)_{n\geq -1}$ be defined as in (A2.14).

Definition 4.17 *x is a* Brjuno number *if* $B(x) := \sum_{n=0}^{\infty} \beta_{n-1} \log x_n^{-1} < +\infty$. *The function* $B : \mathbb{R}\setminus\mathbb{Q} \to (0, +\infty]$ *is called the* Brjuno function.

Exercise 4.18 Show that all diophantine numbers are Brjuno numbers.

Exercise 4.19 Show that there exists $C > 0$ such taht for all Brjuno numbers x one has

$$\left| B(x) - \sum_{n=0}^{\infty} \frac{\log q_{n+1}}{q_n} \right| \leq C \ .$$

Exercise 4.20 (see [MMY]) Show that the Brjuno function satisfies

$$\begin{aligned} &B(x) = B(x+1) \ , \quad \forall x \in \mathbb{R}\setminus\mathbb{Q} \\ &B(x) = -\log x + xB\left(\frac{1}{x}\right) \ , \quad x \in \mathbb{R}\setminus\mathbb{Q}\cap(0,1) \end{aligned} \tag{4.1}$$

Deduce from this that the set of Brjuno numbers is $\mathrm{SL}\,(2,\mathbb{Z})$–invariant. Use the above given functional equation to compute $B(x_p)$, where $x_p = \frac{\sqrt{p^2+4}-p}{2}$, $p \in \mathbb{N}$.

Exercise 4.21 (see [MMY]) Show that the linear operator $(Tf)(x) = xf\left(\frac{1}{x}\right)$, $x \in (0,1)$, acting on periodic functions which belong to $L^p\left(\mathbb{T}, \frac{dx}{(1+x)\log 2}\right)$ has spectral radius bounded by $g = \frac{\sqrt{5}-1}{2}$. Conclude that the Brjuno function $B \in \cap_{p\geq 1} L^p(\mathbb{T})$. Note that $B \notin L^\infty(\mathbb{T})$.

Exercise 4.22 Write the continued fraction expansion of a Brjuno number which is not a diophantine number. The same for the decimal expansion. Is $\sum_{n=0}^{\infty} 10^{-n!}$ a Brjuno number ? What about $\sum_{n=0}^{\infty} 10^{-10^{n!}}$?

Exercise 4.23 Let $\sigma > 0$. Use the results of Exercise 4.21 to study the functions

$$\begin{aligned} &B^{(\sigma)}(x) = B^{(\sigma)}(x+1) \ , \quad \forall x \in \mathbb{R}\setminus\mathbb{Q} \\ &B^{(\sigma)}(x) = x^{-1/\sigma} + xB^{(\sigma)}\left(\frac{1}{x}\right) \ , \quad x \in \mathbb{R}\setminus\mathbb{Q}\cap(0,1) \end{aligned}$$

Show that if $B^{(\sigma)}(x) < +\infty$ then $x \in \mathrm{CD}\,(\sigma)$. Viceversa, if $x \in \mathrm{CD}\,(\tau)$ then $B^{(\sigma)}(x) < +\infty$ for all $\sigma > \tau$.

5. Siegel–Brjuno Theorem, Yoccoz's Theorem and Some Open Problems

Recall that $\mathcal{Y}$ denotes the set of $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ such that $G_{e^{2\pi i\alpha}}$ is a conjugacy class. Here we list what we already know about it

- $\mathcal{Y}$ has full measure (Chapter 3) ;
- the complement of $\mathcal{Y}$ in $\mathbb{R} \setminus \mathbb{Q}$ is a G_δ–dense (Exercise 1.18) ;
- $\mathcal{Y}$ is invariant under the action of $\mathrm{SL}\,(2, \mathbb{Z})$ (Douady–Ghys' Theorem, Chapter 4).

The purpose of this Chapter is to prove the classical results of Siegel [S] and Brjuno [Br] which show that :

all Brjuno numbers belong to $\mathcal{Y}$

Moreover we will state the Theorems of Yoccoz [Yo2] and in particular his celebrated result :

$\mathcal{Y}$ *is equal to the set of Brjuno numbers.*

We will also mention several open problems.

For the sake of brevity, starting with this section all the proofs will be only sketched : the reader can try autonomously to fill in the details but we will always refer to the original literature where complete proofs are given.

5.1 Siegel–Brjuno Theorem

The theorem of Siegel and Brjuno says that the set of Brjuno numbers is a subset of $\mathcal{Y}$. Indeed in 1942 C.L. Siegel was the first to show that $\mathcal{Y}$ is not empty showing that $\mathrm{CD} \subset \mathcal{Y}$.

We will sketch the proof of a more precise result which follows from the Theorem of Yoccoz which we will discuss in the next section but which can also be proved following the classical majorant series method (see [CM]). Let us recall that S_λ denotes the topological space of all germs of holomorphic diffeomorphisms $f : \mathbb{D} \to \mathbb{C}$ such that $f(0) = 0$, f is univalent on $\mathbb{D}$ and $f'(0) = \lambda$. By Theorem A1.19 it is a compact space. Given a germ $f \in S_\lambda$ we let $r(f)$ indicate the radius of convergence of the linearization h_f of f and we set

$$r(\alpha) = \inf_{f \in S_{e^{2\pi i\alpha}}} r(f) \, . \tag{5.1}$$

Theorem 5.1 (Yoccoz's lower bound)

$$\log r(\alpha) \geq -B(\alpha) - C \tag{5.2}$$

where $C > 0$ is a universal constant (independent of α) and B is the Brjuno function.

Before of sketching the proof let us briefly mention what is the main difficulty which was first overcome by Siegel and which was clearly well–known among mathematicians at the end of the 19th and at the beginning of the 20th century (in 1919 Gaston Julia even claimed, in an incorrect paper, to disprove Siegel's theorem).

Assume that $\alpha \in \mathrm{CD}\,(\tau)$ for some $\tau \geq 0$. Recalling the recurrence (1.2) for the power series coefficients of the linearization $h_f(z) = \sum_{n=1}^{\infty} h_n z^n$

$$h_1 = 1\ , \quad h_n = \frac{1}{\lambda^n - \lambda} \sum_{j=2}^{n} f_j \sum_{n_1+\ldots+n_j=n} h_{n_1} \cdots h_{n_j} \quad n \geq 2 \tag{5.3}$$

one sees that h_n is a polynomial in $f_2, \ldots, f_n$ with coefficients which are rational functions of λ : $h_n \in \mathbb{C}(\lambda)[f_2, \ldots, f_n]$ for all $n \geq 2$.

Let us compute explicitely the first few terms of the recurrence

$$\begin{aligned} h_2 &= (\lambda^2 - \lambda)^{-1} f_2\ , \\ h_3 &= (\lambda^3 - \lambda)^{-1}[f_3 + 2f_2^2(\lambda^2 - \lambda)^{-1}]\ , \\ h_4 &= (\lambda^4 - \lambda)^{-1}[f_4 + 3f_3 f_2(\lambda^2 - \lambda)^{-1} + 2f_2 f_3(\lambda^3 - \lambda)^{-1} \\ &\qquad 4f_2^3(\lambda^3 - \lambda)^{-1}(\lambda^2 - \lambda)^{-1} + f_2^3(\lambda^2 - \lambda)^{-2}]\ , \end{aligned} \tag{5.4}$$

and so on. It is not difficult to see that among all contributes to h_n there is always a term of the form $2^{n-2} f_2^{n-1}[(\lambda^n - \lambda)\ldots(\lambda^3 - \lambda)(\lambda^2 - \lambda)]^{-1}$. If one then tries to estimate $|h_n|$ by simply summing up the absolute values of each contribution then one term will be

$$2^{n-2}|f_2|^{n-1}[|\lambda^n - \lambda|\ldots|\lambda^3 - \lambda||\lambda^2 - \lambda|]^{-1} \leq 2^{n-2}|f_2|^{n-1}(2\gamma)^{(n-1)\tau}[(n-1)!]^{\tau} \tag{5.5}$$

if $\alpha \in \mathrm{CD}\,(\gamma, \tau)$ and one obtains a divergent bound. Note the difference with the case $|\lambda| \neq 1$: in this case the bound would be $|\lambda|^{-(n-1)} 2^{n-2} |f_2|^{n-1} c^{n-1}$ for some positive constant c independent of f. Thus one must use a more subtle majorant series method.

The key point is that the estimate (5.5) is far too pessimistic :

Exercise 5.2 Show that the series $\sum_{n=1}^{\infty}\frac{z^n}{(\lambda^n-1)\ldots(\lambda-1)}$, with $\lambda = e^{2\pi i\alpha}$, has positive radius of convergence whenever $\limsup_{n\to\infty}\frac{\log q_{k+1}}{q_k} < +\infty$. [Hint : see [HL] for a proof.]

Indeed when a small divisor is really small then, for a certain time, all other small divisors cannot be too small. This vague idea is made clear by the two following lemmas of A.M. Davie [Da] which extend and improve previous results of A.D. Brjuno.

Let $x \in \mathbb{R}$, $x \neq 1/2$, we denote $\|x\|_{\mathbb{Z}} = \min_{p\in\mathbb{Z}}|x+p|$.

Lemma 5.3 *Let $\alpha \in \mathbb{R}\setminus\mathbb{Q}$, $(p_j/q_j)_{j\geq 0}$ denote the sequence of its convergents, $k \in \mathbb{N}$, $n \in \mathbb{N}$, $n \neq 0$, and assume that $\|n\alpha\|_{\mathbb{Z}} \leq 1/(4q_k)$. Then $n \geq q_k$ and either q_k divides n or $n \geq q_{k+1}/4$.*

Proof. From Theorem 4.5 it follows that if r is an integer and $0 < r < q_k$ then $\|r\alpha\|_{\mathbb{Z}} \geq (2q_k)^{-1}$. Thus $n \geq q_k$. Assume that q_k does not divide n and that $n < q_{k+1}/4$. Then $n = mq_k + r$ where $0 < r < q_k$ and $m < q_{k+1}/(4q_k)$. Since $\|q_k\alpha\|_{\mathbb{Z}} \leq q_{k+1}^{-1}$ one gets $\|mq_k\alpha\|_{\mathbb{Z}} \leq mq_{k+1}^{-1} < (4q_k)^{-1}$. But $\|r\alpha\|_{\mathbb{Z}} \geq (2q_k)^{-1}$ thus $\|n\alpha\|_{\mathbb{Z}} > (4q_k)^{-1}$. $\square$

Using this information on the sequence $(\|n\alpha\|_{\mathbb{Z}})_{n\geq 0}$ Davie shows the following : Let $A_k = \left\{n \geq 0 \mid \|n\alpha\| \leq \frac{1}{8q_k}\right\}$, $E_k = \max(q_k, q_{k+1}/4)$ and $\eta_k = q_k/E_k$. Let A_k^* be the set of non negative integers j such that either $j \in A_k$ or for some j_1 and j_2 in A_k, with $j_2 - j_1 < E_k$, one has $j_1 < j < j_2$ and q_k divides $j - j_1$. For any non negative integer n define :

$$l(n) = \max\left\{(1+\eta_k)\frac{n}{q_k} - 2, (m_n\eta_k + n)\frac{1}{q_k} - 1\right\} \tag{5.6}$$

where $m_n = \max\{j \mid 0 \leq j \leq n, j \in A_k^*\}$. We then define a function $h_k : \mathbb{N} \to \mathbb{R}_+$ as follows

$$h_k(n) = \begin{cases} \frac{m_n+\eta_k n}{q_k} - 1 & \text{if } m_n + q_k \in A_k^* \\ l(n) & \text{if } m_n + q_k \notin A_k^* \end{cases} \tag{5.7}$$

The function $h_k(n)$ has some properties collected in the following proposition

Proposition 5.4 *The function $h_k(n)$ verifies*
(1) $\frac{(1+\eta_k)n}{q_k} - 2 \leq h_k(n) \leq \frac{(1+\eta_k)n}{q_k} - 1$ *for all* n.
(2) *If* $n > 0$ *and* $n \in A_k^*$ *then* $h_k(n) \geq h_k(n-1) + 1$.
(3) $h_k(n) \geq h_k(n-1)$ *for all* $n > 0$.
(4) $h_k(n+q_k) \geq h_k(n) + 1$ *for all* n.

Now we set $g_k(n) = \max\left(h_k(n), \left[\frac{n}{q_k}\right]\right)$ and we state the following proposition

Proposition 5.5 *The function* g_k *is non negative and verifies :*
(1) $g_k(0) = 0$ *;*
(2) $g_k(n) \leq \frac{(1+\eta_k)n}{q_k}$ *for all* n *;*
(3) $g_k(n_1) + g_k(n_2) \leq g_k(n_1+n_2)$ *for all* n_1 *and* n_2 *;*
(4) *if* $n \in A_k$ *and* $n > 0$ *then* $g_k(n) \geq g_k(n-1) + 1$.

The proof of these propositions can be found in [Da].

Let $k(n)$ be defined by the condition $q_{k(n)} \leq n < q_{k(n)+1}$. Note that k is non–decreasing.

Lemma 5.6 (Davie's lemma) *Let*

$$K(n) = n\log 2 + \sum_{k=0}^{k(n)} g_k(n)\log(2q_{k+1})\ . \tag{5.8}$$

The function $K(n)$ verifies :
(a) *There exists a universal constant* $c_0 > 0$ *such that*

$$K(n) \leq n\left(\sum_{k=0}^{k(n)} \frac{\log q_{k+1}}{q_k} + c_0\right)\ ; \tag{5.9}$$

(b) $K(n_1) + K(n_2) \leq K(n_1+n_2)$ *for all* n_1 *and* n_2 *;*
(c) $-\log|\lambda^n - 1| \leq K(n) - K(n-1)$.

Proof. We will apply Proposition 5.4. By (2) we have

$$\begin{aligned} K(n) &\leq n\left[\log 2 + \sum_{k=0}^{k(n)} \frac{(1+\eta_k)}{q_k}\log(2q_{k+1})\right] \\ &\leq n\left[\sum_{k=0}^{k(n)} \frac{\log q_{k+1}}{q_k} + \log 2 + \log 2\sum_{k=0}^{\infty}\left(\frac{1}{q_k} + \frac{4}{q_{k+1}}\right) + 4\sum_{k=0}^{\infty}\frac{\log q_{k+1}}{q_{k+1}}\right] \end{aligned}$$

since $\eta_k \le 4q_k q_{k+1}^{-1}$.

By Remark A2.4 the series $\sum \frac{\log q_{k+1}}{q_{k+1}}$ and $\sum q_k^{-1}$ are uniformly bounded by some constant independent of α, thus (a) follows.

(b) is an immediate consequence of Proposition 5.5, (3), and the fact that $k(n)$ is not decreasing.

Finally recall that

$$-\log|\lambda^n - 1| = -\log 2|\sin \pi n\alpha| \in (-\log \pi \|n\alpha\|_{\mathbb{Z}}, -\log 2\|n\alpha\|_{\mathbb{Z}}) \, .$$

For all n we have either $\|n\alpha\|_{\mathbb{Z}} > 1/4$ or there exists some non–negative integer k such that $(8q_k)^{-1} > \|n\alpha\|_{\mathbb{Z}} \ge (8q_{k+1})^{-1}$, thus $n \in A_k$, which implies by Proposition 5.5, (4), that $g_k(n) \ge g_k(n-1) + 1$, and $-\log|\lambda^n - 1| \le 2q_{k+1}$. Combining these facts together one gets (c). □

Following [CM] we can now prove Theorem 5.1.

Proof. of Theorem 5.1. Let $s(z) = \sum_{n\ge 1} s_n z^n$ be the unique solution analytic at $z = 0$ of the equation $s(z) = z + \sigma(s(z))$, where $\sigma(z) = \frac{z^2(2-z)}{(1-z)^2} = \sum_{n\ge 2} n z^n$. The coefficients satisfy

$$s_1 = 1 \, , \; s_n = \sum_{m=2}^{n} m \sum_{n_1+\ldots+n_m = n \, , \, n_i \ge 1} s_{n_1} \ldots s_{n_m} \, . \tag{5.10}$$

Clearly there exist two positive constants c_1, c_2 such that

$$|s_n| \le c_1 c_2^n \, .$$

From the recurrence relation and Bieberbach–De Branges's bound $|f_n| \le n$ for all $n \ge 2$ we obtain

$$|h_n| \le \frac{1}{|\lambda^n - \lambda|} \sum_{m=2}^{n} m \sum_{n_1+\ldots+n_m = n \, , \, n_i \ge 1} |h_{n_1}| \ldots |h_{n_m}| \, .$$

We now deduce by induction on n that $|h_n| \le s_n e^{K(n-1)}$ for $n \ge 1$, where $K : \mathbb{N} \to \mathbb{R}$ is defined in (5.8). If we assume this holds for all $n' < n$ then the above inequality gives

$$|h_n| \le \frac{1}{|\lambda^n - \lambda|} \sum_{m=2}^{n} m \sum_{n_1+\ldots+n_m = n \, , \, n_i \ge 1} s_{n_1} \ldots s_{n_m} e^{K(n_1-1)+\ldots K(n_m-1)} \, .$$

But $K(n_1-1)+\dots K(n_m-1) \le K(n-2) \le K(n-1)+\log|\lambda^n-\lambda|$ and we deduce that

$$|h_n| \le e^{K(n-1)} \sum_{m=2}^{n} m \sum_{n_1+\ldots+n_m=n\,,\,n_i\ge 1} s_{n_1}\ldots s_{n_m} = s_n e^{K(n-1)}\ ,$$

as required. Theorem 5.1 then follows from the fact that $n^{-1}K(n) \le B(\omega)+c_0$ for some universal constant $c_0>0$ (Davie's lemma). □

Exercise 5.7 Consider the quadratic polynomial $P_\lambda(z)=\lambda(z-z^2)$ (we have conjugated (3.1) by an homothehty so as to eliminate a factor 1/2 in what follows). Its formal linearization $H_\lambda(z)=\sum_{n=1}^{\infty} H_n(\lambda)z^n$ is given by the recurrence

$$H_1(\lambda)=1\ ,\ \ H_n(\lambda)=(1-\lambda^{n-1})^{-1}\sum_{i+j=n} H_i(\lambda)H_j(\lambda)\ .$$

Define the sequence of positive real numbers $(h_n(\lambda))_{n\ge 1}$ by the recurrence

$$h_1(\lambda)=1\ ,\ \ h_n(\lambda)=|1-\lambda^{n-1}|^{-1}\sum_{i+j=n} h_i(\lambda)h_j(\lambda)\ .$$

Clearly $|H_n(\lambda)|\le h_n(\lambda)$. Show that if $\lambda=e^{2\pi i\alpha}$, $\alpha\in\mathbb{R}\setminus\mathbb{Q}$ is *not* a Brjuno number then $\limsup_{n\to\infty} n^{-1}\log h_n(\lambda)=+\infty$, i.e. the majorant series $\sum_{n\ge 1} h_n(\lambda)z^n$ is divergent. Under what assumptions on α one can show that there exist two positive constants c_0, c_1 and $s>0$ such that $h_n(\lambda)\le c_0c_1^n(n!)^s$ for all $n\ge 1$? [Hint : First show that the set $\{n\ge 0 \mid q_{n+1}\ge (q_n+1)^2\}$ is infinite and denote $(q_i')_{i\ge 0}$ its elements. Then show that the sequence $h_n(\lambda)$ is increasing and $h_{q'_{r+1}+1}(\lambda)\ge |1-\lambda^{q'_{r+1}}|^{-1}(h_{q'_r+1}(\lambda))^{\left[\frac{q'_{r+1}}{q'_r+1}\right]}$. One can also consult [Yo2, Appendice 2, pp. 83–85] and [CM].]

Exercise 5.8 (Linearization and Gevrey classes, see [CM] for solutions and more information.) Between $\mathbb{C}[[z]]$ and $\mathbb{C}\{z\}$ one has many important algebras of "ultradifferentiable" power series (i.e. asymptotic expansions at $z=0$ of functions which are "between" $\mathcal{C}^\infty$ and $\mathbb{C}\{z\}$). Consider two subalgebras $A_1\subset A_2$ of $z\mathbb{C}\,[[z]]$ closed with respect to the composition of formal series. For example Gevrey–s classes, $s>0$ (i.e. series $F(z)=\sum_{n\ge 0} f_nz^n$ such that there exist $c_1,c_2>0$ such that $|f_n|\le c_1c_2^n(n!)^s$ for all $n\ge 0$). Let $f\in A_1$ being such that $f'(0)=\lambda\in\mathbb{C}^*$. We say that f *is linearizable in* A_2 if there exists $h_f\in A_2$ tangent to the identity and such that $f \circ h_f = h_f \circ R_\lambda$. Show that if one requires $A_2 = A_1$, i.e. the linearization h_f to be as regular as the given germ f, once again the Brjuno condition is sufficient. It is quite interesting to notice that given any algebra of formal power series which is closed under composition (as it should if one whishes to study conjugacy problems) a germ in the algebra is linearizable *in the same algebra* if the Brjuno condition is satisfied. If the linearization is allowed to be less regular than the given germ (i.e. A_1 is a proper subset of A_2) one finds new arithmetical conditions, weaker than the Brjuno condition. Let $(M_n)_{n\geq 1}$ be a sequence of positive real numbers such that :

0. $\inf_{n\geq 1} M_n^{1/n} > 0$;
1. There exists $C_1 > 0$ such that $M_{n+1} \leq C_1^{n+1} M_n$ for all $n \geq 1$;
2. The sequence $(M_n)_{n\geq 1}$ is logarithmically convex ;
3. $M_n M_m \leq M_{m+n-1}$ for all $m, n \geq 1$.

Let $f = \sum_{n\geq 1} f_n z^n \in z\mathbb{C}[[z]]$; f belongs to the algebra $z\mathbb{C}[[z]]_{(M_n)}$ if there exist two positive constants c_1, c_2 such that

$$|f_n| \leq c_1 c_2^n M_n \text{ for all } n \geq 1 .$$

Show that the condition 3 above implies that $z\mathbb{C}[[z]]_{(M_n)}$ is closed for composition. Show that if $f \in z\mathbb{C}[[z]]_{(N_n)}$, $f_1 = e^{2\pi i\alpha}$ and α verifies

$$\limsup_{n\to+\infty} \left(\sum_{k=0}^{k(n)} \frac{\log q_{k+1}}{q_k} - \frac{1}{n} \log \frac{M_n}{N_n} \right) < +\infty$$

where $k(n)$ is defined by the condition $q_{k(n)} \leq n < q_{k(n)+1}$, then the linearization $h_f \in z\mathbb{C}[[z]]_{(M_n)}$. (We of course assume that the sequence $(N_n)_{n\geq 0}$ is asymptotically bounded by the sequence (M_n), i.e. $M_n \geq N_n$ for all sufficiently large n).

5.2 Yoccoz's Theorem

The main result of Yoccoz can be very simply stated as

$$\mathcal{Y} = \{\alpha \in \mathbb{R} \setminus \mathbb{Q} \mid B(\alpha) < +\infty\} = \text{Brjuno numbers} ,$$

but he proves much more than the above :

Theorem 5.9

(a) If $B(\alpha) = +\infty$ there exists a non–linearizable germ $f \in S_{e^{2\pi i\alpha}}$;

(b) If $B(\alpha) < +\infty$ then $r(\alpha) > 0$ and

$$|\log r(\alpha) + B(\alpha)| \le C \ , \tag{5.11}$$

where C is a universal constant (i.e. independent of α) ;

(c) For all $\varepsilon > 0$ there exists $C_\varepsilon > 0$ such that for all Brjuno numbers α one has

$$-B(\alpha) - C \le \log r(P_{e^{2\pi i\alpha}}) \le -(1-\varepsilon)B(\alpha) + C_\varepsilon \tag{5.12}$$

where C is a universal constant (i.e. independent of α and ε.

The remarkable consequence of (5.11) and (5.12) is that the Brjuno function not only identifies the set $\mathcal{Y}$ but also gives a rather precise estimate of the size of the Siegel disks. When α is not a Brjuno number the problem of a complete classification of the conjugacy classes of germs in $G_{e^{2\pi i\alpha}}$ is open and quite difficult (perhaps unreasonable) as the following result of Yoccoz shows :

Theorem 5.10 *Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, $B(\alpha) = +\infty$. There exists a set with the power of the continuum of conjugacy classes of germs of $G_{e^{2\pi i\alpha}}$, each of which does not contain an entire function.*

The proof of the Theorem of Yoccoz (5.9 above) uses a method, invented by Yoccoz himself, known as *geometric renormalization*. Roughly speaking it is a quantitative version of the topological construction of Douady–Ghys described in Chapter 4 and which shows that the set $\mathcal{Y}$ is $\mathrm{SL}\,(2,\mathbb{Z})$–invariant. Whereas the construction of non–linearizable germs $f \in G_{e^{2\pi i\alpha}}$ when $B(\alpha) = +\infty$ and Yoccoz's upper bound

$$\log r(\alpha) \le C - B(\alpha)$$

go far beyond the scope of these lectures, it is not too difficult to give an idea of how Yoccoz proves Theorem 5.1, i.e. the lower bound

$$\log r(\alpha) \ge -C - B(\alpha) \ .$$

Let $f \in S_{e^{2\pi i\alpha}}$ and let $E : \mathbb{H} \to \mathbb{D}^*$ be the exponential map $E(z) = e^{2\pi i z}$. Then f lifts to a map $F : \mathbb{H} \to \mathbb{C}$ such that

$$E \circ F = f \circ E \ , \tag{5.13}$$

$$F : \mathbb{H} \to \mathbb{C} \text{ is univalent} \ , \tag{5.14}$$

$$F(z) = z + \alpha + \varphi(z) \text{ where } \varphi \text{ is } \mathbb{Z}--\text{periodic and } \lim_{\Im m\, z \to +\infty} \varphi(z) = 0 \tag{5.15}$$

We will denote $S(\alpha)$ the space of univalent functions F verifying (5.13), (5.14) and (5.15).

Exercise 5.11 Show that $S(\alpha)$ is compact and that it is the universal cover of $S_{e^{2\pi i\alpha}}$.

Exercise 5.12 Show that if $f \in S_1$ then one has the following distorsion estimate :

$$\left| z\frac{f'(z)}{f(z)} - 1 \right| \leq \frac{2|z|}{1-|z|} \text{ for all } z \in \mathbb{D} \ . \tag{5.16}$$

Exercise 5.13 Use the result of the previous exercise to show that if φ is as in (5.15) and $z \in \mathbb{H}$ then

$$|\varphi'(z)| \leq \frac{2\exp(-2\pi\, \Im m\, z)}{1-\exp(-2\pi\, \Im m\, z)} \ , \tag{5.17}$$

$$|\varphi(z)| \leq -\frac{1}{\pi}\log(1-\exp(-2\pi\, \Im m\, z)) \ . \tag{5.18}$$

Let $r > 0$, $\mathbb{H}_r = \mathbb{H} + ir$. It is clear that if $F \in S(\alpha)$ and r is sufficiently large then F is very close to the translation $z \mapsto z + \alpha$ for $z \in \mathbb{H}_r$. Indeed using the compactness of $S(\alpha)$ and Exercise 5.13 one can prove the following :

Exercise 5.14 Let $\alpha \neq 0$. Show that there exists a universal constant $c_0 > 0$ (i.e. independent of α) such that for all $F \in S(\alpha)$ and for all $z \in \mathbb{H}_{t(\alpha)}$ where

$$t(\alpha) = \frac{1}{2\pi}\log \alpha^{-1} + c_0 \ , \tag{5.19}$$

one has

$$|F(z) - z - \alpha| \leq \frac{\alpha}{4} \ . \tag{5.20}$$

[Hint : Let $\varphi(z) = F(z) - z - \alpha = \sum_{n=1}^{\infty} \varphi_n e^{2\pi i n z}$. If $\Im m\, z > t(\alpha)$ then $|F(z) - \alpha - z| \leq \sum_{n=1}^{\infty} |\varphi_n| \alpha^n e^{-2\pi n c_0}$ thus ....]

Given F, the lowest admissible value $t(F,\alpha)$ of $t(\alpha)$ such that (5.20) holds for all $z \in \mathbb{H}_{t(F,\alpha)}$ represents the height in the upper half plane $\mathbb{H}$ at which the strong nonlinearities of F manifest themselves. When $\Im m\, z > t(F,\alpha)$, F is very close to the translation $z \mapsto z + \alpha$. An example of strong nonlinearity is of course a fixed point : if $F(z) = z + \alpha + \frac{1}{2\pi i}e^{2\pi i z}$, $\alpha > 0$, then $z = -\frac{1}{4} + \frac{i}{2\pi}\log(2\pi\alpha)^{-1}$ is fixed and $t(F,\alpha) \geq \frac{1}{2\pi}\log(2\pi\alpha)^{-1}$.

The estimates (5.19) and (5.20) are the fundamental ingredient of Yoccoz's proof of the lower bound (5.2) together with Proposition 2.15 and the following elementary properties of the conformal capacity.

Exercise 5.15 Let $U \subset C$ be a simply connected open set, $U \neq \mathbb{C}$, and let $z_0 \in U$. Let d be the distance of z_0 from the complement of U in $\mathbb{C}$ and let $C(U, z_0)$ denote the conformal capacity of U w.r.t. z_0. Then

$$d \leq C(U, z_0) \leq 4d \,. \tag{5.21}$$

As in the proof of Douady–Ghys theorem we can now construct the first return map in the strip B delimited by $l = [it(\alpha), +i\infty[$, $F(l)$ and the segment $[it(\alpha), F(it(\alpha))]$. Given z in B we can iterate F until $\Re e\, F^n(z) > 1$. If $\Im m\, z \geq t(\alpha) + c$ for some $c > 0$ then $z' = F^n(z) - 1 \in B$ and $z \mapsto z'$ is the first return map in the strip B. Glueing l and $F(l)$ by F one obtains a Riemann surface S corresponding to $\operatorname{int} B$ and biholomorphic to $\mathbb{D}^*$. This induces a map $g \in S_{e^{2\pi i/\alpha}}$ which lifts to $G \in S(\alpha^{-1})$. One can then show the following (see [Yo2], pp. 32–33)

Proposition 5.16 *Let $\alpha \in (0,1)$, $F \in S(\alpha)$ and $t(\alpha) > 0$ such that if $\Im m\, z \geq t(\alpha)$ then $|F(z) - z - \alpha| \leq \alpha/4$. There exists $G \in S(\alpha^{-1})$ such that if $z \in \mathbb{H}$, $\Im m\, z \geq t(\alpha)$ and $F^i(z) \in \mathbb{H}$ for all $i = 0, 1, \ldots, n-1$ but $F^n(z) \notin \mathbb{H}$ then there exists $z' \in \mathbb{C}$ such that*

1. *$\Im m\, z' \geq \alpha^{-1}(\Im m\, z - t(\alpha) - c_1)$, where $c_1 > 0$ is a universal constant ;*
2. *There exists an integer m such that $0 \leq m < n$ and $G^m(z') \notin \mathbb{H}$.*

From this Proposition one can conclude the proof as follows. Let us recall that $K_f = \cap_{n \geq 0} f^{-n}(\mathbb{D})$ is the maximal compact f–invariant set containing 0. Let $F \in S(\alpha)$ be the lift of $f \in S_{e^{2\pi i \alpha}}$ and let $K_F \subset \mathbb{C}$ be defined as the cover of K_f : $K_F = E^{-1}(K_f)$. It is immediate to check that

$$d_F = \sup\{\Im m\, z \mid z \in \mathbb{C} \setminus K_F\} = -\frac{1}{2\pi} \log \operatorname{dist}(0, \mathbb{C} \setminus K_f) \,.$$

Thus by (5.21) one gets

$$\exp(-2\pi d_F) \leq C(K_f, 0) \leq 4 \exp(-2\pi d_F) \,.$$

Theorem 5.1 is therefore equivalent to the lower bound

$$\sup_{F \in S(\alpha)} d_F \leq \frac{1}{2\pi} B(\alpha) + C \tag{5.22}$$

for some universal constant $C > 0$.

Assume that (5.22) is not true and that there exist $\alpha \in \mathbb{R} \setminus \mathbb{Q} \cap (0, 1/2)$ with $B(\alpha) < +\infty$, $F \in S(\alpha)$, $z \in \mathbb{H}$ and $n > 0$ such that

$$\Im m\, F^n(z) \leq 0 \,,$$
$$\Im m\, z \geq \frac{1}{2\pi} B(\alpha) + C \,.$$

Let us choose α, F and z so that n is as small as possible. By Proposition 5.16, if $C > c_0$, one gets

$$\begin{aligned} \Im m\, z' &\geq \alpha^{-1}[\Im m\, z - t(\alpha) - c_1] \\ &\geq \alpha^{-1}\left[\frac{1}{2\pi}(B(\alpha) - \log \alpha^{-1}) + C - c_0 - c_1\right] \,. \end{aligned}$$

By the functional equation of B one gets

$$\Im m\, z' \geq \frac{1}{2\pi} B(\alpha^{-1}) + \alpha^{-1}[C - c_0 - c_1] \geq \frac{1}{2\pi} B(\alpha^{-1}) + C$$

provided that $C \geq 2(c_0 + c_1)$. But Proposition 5.16 shows that this contradicts the minimality of n and we must therefore conclude that (5.22) holds. □

A nice description of the proof of the upper bound $\log r(\alpha) \leq C - B(\alpha)$ can be found in the Bourbaki seminar of Ricardo Perez–Marco [PM1].

5.3 Some Open Problems

The first open problem we want to address is whether or not the infimum in (5.1) is attained by the quadratic polynomial $P_\lambda(z) = \lambda z \left(1 - \frac{z}{2}\right)$:

Question 5.17 Does $r(\alpha) = \inf_{f \in S_{e^{2\pi i\alpha}}} r(f) = r_2(e^{2\pi i\alpha})$, i.e. the radius of convergence of the quadratic polynomial ?

If this were true then the very precise bound (5.11) would hold also for P_λ : recalling that $r_2(\lambda) = |u(\lambda)|$, where $u : \mathbb{D} \to \mathbb{C}$ is the function defined in Section 3.1, one can ask

Question 5.18 Does the function $\alpha \mapsto \log |u(e^{2\pi i\alpha})| + B(\alpha) \in L^\infty(\mathbb{S}^1)$?

Indeed there is a good numerical evidence that much more could be true :

Conjecture 5.19 The function $\alpha \mapsto \log|u(e^{2\pi i\alpha})| + B(\alpha)$ extends to a Hölder 1/2 function.

We refer to [Ma1] and to [MMY2] and references therein for a discussion of Conjecture 5.19. The next Exercises give an idea of how to compute approximately but effectively the function $\alpha \mapsto \log|u(e^{2\pi i\alpha})|$ on a computer. More informations can be found in [He4] (where one can also find many problems, most of which are still open) and [Ma1].

Exercise 5.20 Let $f \in G_\lambda$ be linearizable, $\lambda = e^{2\pi i\alpha}$. Let U_f be the Siegel disk of f, h_f be the linearization of f, $z \in U_f$, $z = h_f(w)$, where $w \in \mathbb{D}_{r(f)}$, $|w| = r < r(f)$. Show that

$$\lim_{m\to+\infty} \frac{1}{m} \sum_{j=0}^{m-1} \log|f^j(z)| = \log r \tag{5.23}$$

[Solution : Since h_f conjugates f to R_λ one has $f^j(z) = f^j(h_f(w)) = h_f(\lambda^j w)$ for all $j \geq 0$ and $w \in \mathbb{D}_{r(f)}$, thus

$$\frac{1}{m} \sum_{j=0}^{m-1} \log|f^j(z)| = \frac{1}{m} \sum_{j=0}^{m-1} \log|h_f(\lambda^j w)| .$$

h_f has neither poles nor zeros but $w = 0$ thus by the mean property of harmonic functions one has $\int_0^1 \log|h_f(re^{2\pi ix})|dx = \log r$ for all $r \leq r(f)$. Finally note that $w \mapsto \lambda w$ is uniquely ergodic on $|w| = r$, and in this case Birkhoff's ergodic theorem holds for all initial points, thus

$$\begin{aligned} \lim_{m\to+\infty} \frac{1}{m} \sum_{j=0}^{m-1} \log|f^j(z)| &= \lim_{m\to+\infty} \frac{1}{m} \sum_{j=0}^{m-1} \log|h_f(\lambda^j w)| \\ &= \int_0^1 \log|h_f(re^{2\pi ix})|dx = \log r . \end{aligned}$$

Exercise 5.21 Deduce from the previous exercise that for almost every $z \in \partial U_f$ with respect to the harmonic measure one has

$$\lim_{m\to+\infty} \frac{1}{m} \sum_{j=0}^{m-1} \log|f^j(z)| = \log r(f) . \tag{5.24}$$

Let us now consider the quadratic polynomial P_λ once more. According to (5.24) to compute $|u(\lambda)|$ one needs to know that some point belongs to the boundary of the Siegel disk of P_λ (and hope ...). The critical point cannot be contained in U_{P_λ} because $f|_{U_{P_\lambda}}$ is injective, and from the classical theory of Fatou and Julia one knows that ∂U_{P_λ} is contained in the closure of the forward orbit $\{P_\lambda^k(1) \mid k \geq 0\}$ of the critical point $z = 1$. Finally Herman proved if α verifies an arithmetical condition $\mathcal{H}$, weaker than the Diophantine condition but stronger than the Brjuno condition (see, for example, [Yo1] for its precise formulation) the critical point belongs to ∂U_{P_λ}.

Exercise 5.22 If $\alpha \in \mathrm{CD}\,(0)$ Herman has also proved that ∂U_{P_λ} is a *quasicircle*, that is the image of $\mathbb{S}^1$ under a quasiconformal homeomorphism. In this case h_{P_λ} admits a quasiconformal extension to $|w| = r_2(\lambda)$ and therefore is Hölder continuous [Po] :

$$|h_{P_\lambda}(w_1) - h_{P_\lambda}(w_2)| \leq 4|w_1 - w_2|^{1-\chi} \tag{5.25}$$

for all $w_1\,, w_2 \in \partial\mathbb{D}_{r_2(\lambda)}$, where $\chi \in [0, 1[$ depends on λ is the so–called Grunsky norm [Po] associated with the univalent function $g(z) = r_2(\lambda)/h_{P_\lambda}(r_2(\lambda)/z)$ on $|z| > 1$. Using this information show that

$$|\frac{1}{q_k}\sum_{j=0}^{q_k-1} \log|P_\lambda^j(z)| - \log r_2(\lambda)| \leq \frac{8}{r_2(\lambda)}(\frac{2\pi}{q_k})^{1-\chi}\ , \tag{5.26}$$

where p_k/q_k is a convergent of the continued fraction expansion of α. Note that (5.26) implies convergence to $\log r_2(\lambda)$ *for all* $z \in \partial U_{P_\lambda}$, thus also for the critical point $z = 1$.

6. Small divisors and loss of differentiability

In this Chapter we will (very!) briefly illustrate other two completely different approaches to the problem of linearization of germs of holomorphic diffeomorphisms with an indifferent fixed point.

In the previous chapter we saw how the optimal sufficient condition can be obtained by the classical majorant series method as Siegel and Brjuno did and that Yoccoz was able to show that it is also necessary with his ingenious creation of "geometric renormalization". Here we will give an idea of two proofs of the Siegel theorem, one due to Herman [He1, He2] and the other due essentially to Kolmogorov [K] (see also Arnol'd [Ar3] and Zehnder [Ze2]).

Herman's method is far from giving the optimal number–theoretical condition, the idea is simply so original and beautiful that it deserves being known. It also illustrates how in one–dimensional small divisor problems the problem known as "loss of differentiability" does not prevent from the application of simple tools like the contraction principle. Herman's method can also be extended to (and it is actually described by him for) the problem of local conjugacy to rotation of smooth orientation–preserving diffeomorphisms of the circle.

The idea of Kolmogorov is do adapt Newton's method for finding the roots of algebraic equations so as to apply it for finding the solution of the conjugacy equation. This method has been shown by Rüssmann [Rü] to be adaptable so as to prove the sufficiency of a condition (slightly stronger than) Brjuno's. The main reason for sketching Kolmogorov's argument in this rather limited setting is that in the second part of this monograph we will illustrate Nash–Moser's implicit function theorem which is essentially the abstract and flexible formulation of Kolmogorov's idea.

6.1 Hardy–Sobolev spaces and loss of differentiability

Let $k \in \mathbb{N}$, $r > 0$. Following [He2], we introduce the Hardy–Sobolev spaces

$$\mathcal{O}_r^{k,2} = \{f(z) = \sum_{n=0}^{\infty} f_n z^n \ \mid \ \|f\|_{\mathcal{O}_r^{k,2}} = (|f_0|^2 + \sum_{n=1}^{\infty} n^{2k}|f_n|^2 r^{2n})^{1/2} < +\infty\} \ . \quad (6.1)$$

Exercise 6.1

(a) Show that $\mathcal{O}_r^{k,2}$ is a Hilbert space.

(b) Show if $f \in \mathcal{O}_r^{k,2}$, $f(0) = 0$, one has

$$\sup_{|z| \le r} |f(z)| \le \sqrt{\zeta(2k)} \|f\|_{\mathcal{O}_r^{k,2}} ,$$

where ζ denotes Riemann's zeta function.

(c) Show that if $k \ge 1$ then $\mathcal{O}_r^{k,2}$ is a Banach algebra.

(d) Show that if $f \in \mathcal{O}_r^{k,2}$ and ϕ is holomorphic in a neighborhood of $f(\mathbb{D}_r)$ then $\phi \circ f \in \mathcal{O}_r^{k,2}$ and on a sufficiently small neighborhood V of f in $\mathcal{O}_r^{k,2}$ the map $\psi \in V \mapsto \phi \circ \psi \in \mathcal{O}_r^{k,2}$ is holomorphic.

The following very elementary proposition well illustrates the phenomenon of "loss of differentiability" due to the small divisors which already arises at the level of the linearized conjugacy equation (6.2).

Proposition 6.2 *Let $0 \le \tau \le \tau_0$, $\tau_0 \in \mathbb{N}$, $k \ge 1 + \tau_0$, $f \in \mathcal{O}_r^{k,2}$, $f(0) = 0$. If $\alpha \in CD(\tau)$ then the unique solution $g := D_\lambda^{-1} f$ verifying $g(0) = 0$ of*

$$g \circ R_\lambda - g = f \ , \tag{6.2}$$

belongs to $\mathcal{O}_r^{k-1-\tau_0,2}$. Moreover there exists a universal constant $C > 0$ such that

$$\|\frac{d^{k-1-\tau_0} g}{dz}\|_{\mathcal{O}_r^{0,2}} \le \frac{C}{\gamma} \|\frac{d^k f}{dz}\|_{\mathcal{O}_r^{0,2}} \ ,$$

where $\gamma = \inf_{n \ge 1} n^{1+\tau} |\lambda^n - 1|$.

Proof. It is a straightforward computation starting from the identity $g(z) = \sum_{n=1}^{\infty} (\lambda^n - 1)^{-1} f_n$. □

This Proposition shows that solving the linear equation (6.2) the small divisors cause the loss of $1 + \tau_0$ derivatives. This loss of differentiability phenomenon is typical of small divisors problems and it will be crucial in the discussions in the second part of this monograph. The most annoying consequence of this phenomenon is the impossibility of using fixed points methods to solve conjugacy equations, simply because the operator D_λ is *unbounded* if regarded on a *fixed* Hardy–Sobolev space. However under some restriction on τ_0 one can actually use the contraction principle to solve the conjugacy problem, thanks to an ingenious idea of Herman we will shortly describe in the next section.

6.2 Herman's Schwarzian derivative trick

Let Ω be a region in the complex plane and $f : \Omega \to \mathbb{C}$ be holomorphic.

Definition 6.3 *The* Schwarzian derivative $S(f)$ *of* f *is*

$$\begin{aligned} S(f) &:= (\log f')'' - \frac{1}{2}((\log f')')^2 = \frac{f'''}{f'} - \frac{3}{2}\left(\frac{f''}{f'}\right)^2 \\ &= -2\sqrt{f'}\left(\frac{1}{\sqrt{f'}}\right)'' \end{aligned} \tag{6.3}$$

Exercise 6.4

(a) Prove that the following "chain rule" holds : $S(f \circ g) = (S(f) \circ g)(g')^2 + S(g)$.

(b) Show that $S(f) \equiv 0$ if and only if f is a Möbius map.

The idea of Herman is to apply the Schwarzian derivative to the conjugacy equation $f \circ h = h \circ R_\lambda$: one obtains

$$\lambda^2(S(h) \circ R_\lambda) - S(h) = (S(f) \circ h)(h')^2 \ . \tag{6.4}$$

At the r.h.s. appears h', thus one has already lost one derivative and this does not seem to lead to anything good. Assuming the r.h.s. as given one could solve for $S(h)$ but this would cost $1 + \tau_0$ derivatives according to Proposition 6.2. However if $\tau_0 = 1$ (which is true for almost all α as we saw in Section 4.2) the total loss of derivatives is three. The idea now is that applying S^{-1} one should recuperate three derivatives and this would imply that the map

$$\mathcal{O}_r^{k,2} \ni h \mapsto S^{-1}D_\lambda^{-1}[(S(f) \circ h)(h')^2] \tag{6.5}$$

takes its values in $\mathcal{O}_r^{k,2}$ too. Note that we have slightly modified the definition of D_λ with respect to the previous Section : here $D_\lambda f = \lambda^2 f \circ R_\lambda - f$. This does not change the conclusions of Section 6.1.

On a disk $\mathbb{D}_r$ of sufficiently small radius f is close to R_λ thus $S(f)$ must be small and one can hope to conclude using a fixed point theorem (the contraction principle, say). This strategy indeed works : see [He1, He2] for the details.

The inversion of S is achieved as follows. First of all note that it is not restrictive to assume $f''(0) = 0$, so that $h''(0) = 0$ (one can preliminarly conjugate

f by the polynomial $z - \frac{f''(0)}{2(\lambda-\lambda^2)} z^2)$: this implies $[(S(f) \circ h)(h')^2]_{z=0} = 0$. Let $\psi = D_\lambda^{-1}[(S(f) \circ h)(h')^2]$. If ψ is small enough (this is always the case if one considers a sufficiently small disk) then one can write ψ uniquely in the form

$$\psi = \psi_1' - \frac{1}{2}\psi_1^2$$

with $\psi_1(0) = 0$. Now one can easily solve the problem $S(h_1) = \psi = \psi_1' - \frac{1}{2}\psi_1^2$ just looking for h_1 such that $(\log h_1')' = \psi_1$. This is achieved in three steps :

$$\begin{aligned} \psi_2' &= \psi_1 \ , \\ \psi_3 &= e^{c+\psi_2} \text{ where } c \text{ is chosen s.t. } \psi_3(0) = 1 \ , \\ h_1 &= \int_0^z \psi_3(\zeta) d\zeta \ . \end{aligned}$$

Then it is immediate to check that $S(h_1) = \psi$.

Exercise 6.5 Let $r > 0$, ψ and h_1 as above. Show that if $\psi \in \mathcal{O}_r^{0,2}$ then $h_1 \in \mathcal{O}_r^{3,2}$.

6.3 Kolmogorov's modified Newton method

Here we follow quite closely [St], Volume II, Chapter III, Section 7. We suggest however the reader to look also at [Ze2] for a complete proof.

Let $f \in S_\lambda$, $\lambda = e^{2\pi i\alpha}$ and assume that α is a diophantine number. We want to construct h tangent to the identity such that $R_\lambda - h^{-1} \circ f \circ h = 0$. Let $\tilde{h} = h - \mathrm{id}$ and let us define the composition law $\odot$ as

$$(\mathrm{id} + \tilde{h}_1) \circ (\mathrm{id} + \tilde{h}_2) = \mathrm{id} + \tilde{h}_1 \odot \tilde{h}_2 \ . \qquad (6.6)$$

Clearly one expects that

$$\tilde{h}_1 \odot \tilde{h}_2 = \tilde{h}_1 + \tilde{h}_2 + \text{quadratic terms} \ .$$

Let $\tilde{f} = f - R_\lambda$ and let us define a second composition law $\otimes$ as

$$R_\lambda + \tilde{h} \otimes \tilde{f} = (\mathrm{id} + \tilde{h})^{-1} \circ (R_\lambda + \tilde{f}) \circ (\mathrm{id} + \tilde{h}) \ . \qquad (6.7)$$

Of course one needs $\tilde{h}$ be small so as to assure the existence of the inverse in (6.7) but this is not difficult to obtain considering a small enough disk since $\tilde{h} = h_2 z^2 + \ldots$.

Exercise 6.6 Recall Lagrange's Theorem on the inversion of analytic functions (see [Di], p. 250) : if $h : \mathbb{D}_r \to \mathbb{C}$ is holomorphic and tangent to the identity then choosing r small enough there exists a unique solution $z = \kappa(w)$ of the equation $w = h(z)$. Moreover κ is holomorphic in a neighborhood of 0 and is explicitly given by

$$\kappa(w) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{d^{n-1}}{dw^{n-1}} (h(w))^n .$$

Get some precise estimate on the size of the domain and of the norm of κ if h belongs to some Hardy–Sobolev space.

We then define

$$\mathcal{R}(\tilde{h}, \tilde{f}) = (\mathrm{id} + \tilde{h}) \circ R_\lambda - (R_\lambda + \tilde{f}) \circ (\mathrm{id} + \tilde{h}) . \tag{6.8}$$

Clearly $\mathcal{R}(0,0) = 0$, $\mathcal{R}(0, \tilde{f}) = -\tilde{f}$ and

$$\mathcal{R}(\tilde{h}_1 \odot \tilde{h}_2, \tilde{f}) = (\mathrm{id} + \tilde{h}_1) \circ (\mathrm{id} + \tilde{h}_2) \circ R_\lambda - (R_\lambda + \tilde{f}) \circ (\mathrm{id} + \tilde{h}_1) \circ (\mathrm{id} + \tilde{h}_2) , \tag{6.9}$$

$$\mathcal{R}(\tilde{h}_2, \tilde{h}_1 \otimes \tilde{f}) = (\mathrm{id} + \tilde{h}_2) \circ R_\lambda - (\mathrm{id} + \tilde{h}_1)^{-1} \circ (R_\lambda + \tilde{f}) \circ (\mathrm{id} + \tilde{h}_1) \circ (\mathrm{id} + \tilde{h}_2) . \tag{6.10}$$

Comparing (6.9) with (6.10) we have (for z small enough)

$$(\inf |1+\tilde{h}_1'|) |\mathcal{R}(\tilde{h}_2, \tilde{h}_1 \otimes f)(z)| \le |\mathcal{R}(\tilde{h}_1 \odot \tilde{h}_2, \tilde{f})(z)| \le (\sup |1+\tilde{h}_1'|) |\mathcal{R}(\tilde{h}_2, \tilde{h}_1 \otimes f)(z)| ,$$

thus one should get

$$C^{-1} \|\mathcal{R}(\tilde{h}_2, \tilde{h}_1 \otimes f)\| \le \|\mathcal{R}(\tilde{h}_1 \odot \tilde{h}_2, \tilde{f})\| \le C \|\mathcal{R}(\tilde{h}_2, \tilde{h}_1 \otimes f)\| . \tag{6.11}$$

for a suitably chosen norm $\| \ \|$ and some $C > 0$ (see Exercise 6.6).

Let us now try to solve the equation $\mathcal{R}(\tilde{h}, \tilde{f}) = 0$ by taking a sequence of approximations defined as follows :

(0) Let $\tilde{h}_0 = \tilde{g}_0 = 0$, $\tilde{f}_0 = \tilde{f}$: thus $\mathcal{R}(\tilde{h}_0, \tilde{f}_0) = -\tilde{f}_0$;

(1) Let $\tilde{f}_1 = \tilde{g}_0 \otimes \tilde{f}_0 = \tilde{f}_0$. Choose $\tilde{g}_1$ to be the solution of the linearized equation

$$\partial_1 \mathcal{R}(0,0) \tilde{g}_1 + \partial_2 \mathcal{R}(0,0) \tilde{f}_1 = 0 , \tag{6.12$_1$}$$

where ∂_j denotes the partial derivative w.r.t. the j–th argument. Finally we set $\tilde{h}_1 = \tilde{h}_0 \odot \tilde{g}_1$.

(i+1) We choose $\tilde{f}_{i+1} = \tilde{g}_i \otimes \tilde{f}_i$ and $\tilde{g}_{i+1}$ to be the solution of

$$\partial_1 \mathcal{R}(0,0)\tilde{g}_{i+1} + \partial_2 \mathcal{R}(0,0)\tilde{f}_{i+1} = 0 \ , \qquad (6.12)_{i+1}$$

and we set $\tilde{h}_{i+1} = \tilde{h}_i \odot \tilde{g}_{i+1}$.

It is immediate to check that the linearized equations $(6.12)_i$ have the form

$$\tilde{g}_i \circ R_\lambda - R_\lambda \circ \tilde{g}_i = \tilde{f}_i \ , \qquad (6.13)$$

which we studied in Section 6.2. Note that we do *not* linearize at the point $(0, f)$ since we would get a difference equation *without* constant coefficients :

$$\partial_1 \mathcal{R}(0, \tilde{f}_{i-1})\tilde{g}_i + \partial_2 \mathcal{R}(0, \tilde{f}_{i-1})\tilde{f}_i = \tilde{g}_i \circ R_\lambda - R_\lambda \circ \tilde{g}_i - \tilde{f}'_{i-1}\tilde{g}_i = 0 \ .$$

If one could solve (6.13) at each step with a bound

$$\|\tilde{g}_i\| \le C\|\tilde{f}_i\| \ , \qquad (6.14)$$

by (6.11) and (6.14) one would have

$$\|\mathcal{R}(\tilde{g}_i, \tilde{f}_i)\| \le \sup \|d^2\mathcal{R}\|(\|\tilde{g}_i\|^2 + \|\tilde{f}_i\|^2 \le C^3\|\mathcal{R}(\tilde{g}_{i-1}, \tilde{f}_{i-1})\|^2 \ . \qquad (6.15)$$

This would imply the convergence of the iterative scheme to a solution of $\mathcal{R}(\tilde{h}, \tilde{f}) = 0$ provided that one chooses $\|\tilde{f}\|$ small enough (i.e. one considers the restriction of f to a small enough disk $\mathbb{D}_r$) : indeed iterating (6.15) one gets

$$\|\mathcal{R}(\tilde{g}_i, \tilde{f}_i)\| \le (C^{3/2}\|\mathcal{R}(\tilde{g}_0, \tilde{f}_0)\|)^{2^i}$$

thus again by (6.11) one has

$$\begin{aligned}\|\mathcal{R}(\tilde{h}_i, \tilde{f})\| &= \|\mathcal{R}(\tilde{g}_0 \odot \tilde{g}_1 \odot \ldots \odot \tilde{g}_i, \tilde{f})\| \\ &\le C\|\mathcal{R}(\tilde{g}_1 \odot \ldots \odot \tilde{g}_i, \tilde{g}_0 \otimes \tilde{f})\| \\ &\le C^i\|\mathcal{R}(\tilde{g}_i, \tilde{f}_i)\| \\ &\le (C^2\|\mathcal{R}(\tilde{g}_0, \tilde{f}_0)\|)^{2^i}\end{aligned}$$

Exercise 6.7 Assuming the estimates above show that the sequence $\tilde{h}_n$ converges thus by continuity of $\mathcal{R}$ one gets the desired result.

Exercise 6.8 Use the above scheme to give an alternative proof of Koenigs–Poincaré theorem.

The above discussion shows how to prove the existence of the linearization disregarding the problem of loss of differentiability due to small divisors. This makes impossible to get an estimate like (6.14) *unless one regularizes the r.h.s.*. The simplest method of regolarization, which is adapted to the analytic case, is to *consider restrictions* of the domains :

Exercise 6.9 Show that if $f \in \mathcal{O}_r^{0,2}$, $f(0) = 0$, $k \in \mathbb{N}$, for all $\delta > 0$ one has

$$\|f\|_{\mathcal{O}^{k,2}_{re^{-\delta}}} \leq \left(\frac{k}{\delta}\right)^k e^{-k} \|f\|_{\mathcal{O}^{0,2}_r} \, . \tag{6.16}$$

Combining the above given discussion with a suitable choice of restrictions (i.e. a sequence $(\delta_n)_{n\geq 0}$ such that $\sum_{n=0}^{\infty} \delta_n < +\infty$) one can indeed prove Siegel's Theorem following the iteration method.

Part II. Implicit Function Theorems and KAM Theory

7. Hamiltonian Systems and Integrable Systems

In this Chapter we will very briefly recall some well–known facts on symplectic manifolds and Hamiltonian systems. Very good references are [AKN] and [AM].

7.1 Symplectic Manifolds and Hamiltonian Systems

Definition 7.1 *A $\mathcal{C}^\infty$* symplectic manifold *is a $2l$–dimensional $\mathcal{C}^\infty$ manifold M equipped with a non–degenerate $\mathcal{C}^\infty$ two–form (the symplectic form) ω. A $\mathcal{C}^\infty$ map $f : U \to M'$ where $U \subset M$ is open and M' is also symplectic (with symplectic form ω') is* symplectic *(or* canonical*) if $f^*\omega' = \omega$.*

The simplest (but important) examples of symplectic manifolds are :

- $M = \mathbb{R}^{2l} \ni (p_1, \ldots, p_l, q_1, \ldots, q_l)$, $\omega = \sum_{i=1}^{l} dp_i \wedge dq_i$ (*standard symplectic structure*). If U and V are two open sets in $\mathbb{R}^{2l}$ and $f : U \to V$ then f is symplectic if and only if its Jacobian matrix $J_f \in \mathrm{Sp}\,(l, \mathbb{R})$, the Lie group of $2l \times 2l$ real matrices A such that $A^T \mathcal{I} A = \mathcal{I}$, where $\mathcal{I} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.
- $M = T^*N$ where N is a $\mathcal{C}^\infty$ Riemannian manifold. This is the typical situation in classical mechanics. If $(q_1, \ldots, q_l)$ are local coordinates in N and $(p_1, \ldots, p_l)$ are the corresponding local coordinates in the cotangent space at a point, then $\omega = \sum_{i=1}^{l} dp_i \wedge dq_i$.
- $M = \mathbb{T}^{2l}$, $\omega = \sum_{i=1}^{l} d\theta_i \wedge d\theta_{i+l}$.

Theorem 7.2 (Darboux) *Each symplectic manifold M has an atlas $(U_\alpha, \varphi_\alpha)_{\alpha \in \mathcal{A}}$ such that on $\varphi_\alpha(U_\alpha) \subset \mathbb{R}^{2l}$ one has $\omega = \varphi_\alpha^* \sum_{i=1}^{l} dp_i \wedge dq_i$ (the standard symplectic structure on $\mathbb{R}^{2l}$). The transition maps $\varphi_\alpha \circ \varphi_\beta^{-1}$ are symplectic diffeomorphisms, i.e. their Jacobians $J_{\alpha\beta}(x) \in Sp\,(l, \mathbb{R})$ for all $x \in \varphi_\beta(U_\alpha \cap U_\beta)$.*

The atlas given by Darboux's Theorem and the corresponding local coordinates are called symplectic.

Definition 7.3 *A* Hamiltonian function *on a symplectic manifold* (M,ω) *is a function* $H \in \mathcal{C}^\infty(M,\mathbb{R})$*. The* Hamiltonian vector field *associated to* H *is the unique* $X_H \in \mathcal{C}^\infty(M,TM)$ *such that* $i_{X_H}\omega = dH$*.*

Note that in symplectic local coordinates a Hamiltonian vector field takes the form

$$X_H = \sum_{i=1}^{l} -\frac{\partial H}{\partial q_i}\frac{\partial}{\partial p_i} + \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q_i} \ , \tag{7.1}$$

and the associated ordinary differential equations are the classical Hamilton's equations of the motion of a conservative mechanical system with l degrees of freedom :

$$\dot{p}_i = -\frac{\partial H}{\partial q_i} \ , \ \ \dot{q}_i = \frac{\partial H}{\partial p_i} \ , \ \ 1 \le i \le l \ . \tag{7.2}$$

Clearly the Hamiltonian function is a first integral of (7.2). The coordinates q_i are also called "generalized coordinates" and the p_i their "conjugate momenta". In many problems arising from celestial mechanics the flow is not complete due to the unavoidable occurance of collisions, but we will always assume completeness of the Hamiltonian flow.

Definition 7.4 *The* Poisson bracket *of two functions* $f,g \in \mathcal{C}^\infty(M,\mathbb{R})$ *defined on an open subset of* (M,ω) *is* $\{F,G\} := X_G F = \omega(X_F,X_G) = -X_F G$ *, thus* $X_{\{F,G\}} = -[X_F,X_G]$ *. Two functions* F,G *are* in involution *if* $\{F,G\} = 0$*, i. e. when their hamiltonian flows commute.*

Exercise 7.5 Show that the *Hamiltonian flow* $\Phi \ : \mathbb{R} \times M \to M$ is symplectic : for all $t \in \mathbb{R}$ one has $\Phi(t,\cdot)^*\omega = \omega$. [Hint : use Cartan's formula $\frac{d}{dt}|_{t=0}\Phi(t,\cdot)^*\omega = d(i_{X_H}\omega) + i_{X_H}d\omega$, where $X_H = \frac{d}{dt}|_{t=0}\Phi(t,\cdot)$ is the Hamiltonian vector field associated to Φ.]

The importance of Exercise 7.5 is that to make symplectic coordinate changes of a Hamiltonian vector field it is sufficient to change the varables in the corresponding Hamiltonian function. This is a simpler operation, both conceptually and computationally.

As we will see in the next Section, among the possible orbits of Hamiltonian systems, *quasiperiodic* orbits are of special interest.

Definition 7.6 *A continuous function $F : \mathbb{R} \to \mathbb{R}$ is* quasiperiodic *if there exist $n \geq 2$, $f : \mathbb{T}^n \to \mathbb{R}$ continuous and $\nu \in \mathbb{R}^n \setminus \{0\}$ such that $F(t) = f(\nu_1 t, \dots \nu_n t)$.*

Let $\mathcal{M} = \{k \in \mathbb{Z}^n \mid k \cdot \nu = 0\}$. Note that $\mathcal{M}$ is a $\mathbb{Z}$–module. If $\dim \mathcal{M} = n$ then $\nu = 0$, if $\dim \mathcal{M} = n - 1$ then there exists $\alpha \in \mathbb{R}$ and $k \in \mathbb{Z}^n$ such that $\nu = \alpha k$. If $\dim \mathcal{M} = 0$ then ν is called *non–resonant*.

Exercise 7.7 Show that the closure of any orbit of the linear flow $\dot{\theta} = \nu$ on $\mathbb{T}^n$ is diffeomorphic to the torus $\mathbb{T}^{n-\dim \mathcal{M}}$.

Exercise 7.8 Show that if $\dim \mathcal{M} \in \{1, \dots n - 1\}$ there exists $A \in \mathrm{SL}\,(n, \mathbb{Z})$ such that posing $\varphi = A\theta$ the linear flow $\dot{\theta} = \nu$ on $\mathbb{T}^n$ becomes $\dot{\varphi}_i = 0$ for $i = 1, \dots, m$ and $\dot{\varphi}_i = \nu'_i$ for $i = m + 1, \dots, n$ with $(\nu'_{m+1}, \dots, \nu'_n) \in \mathbb{R}^{n-m}$ non–resonant.

Exercise 7.9 Show that if ν is non–resonant then the Haar measure on $\mathbb{T}^n$ is uniquely ergodic (see [Mn] for its definition) for the linear flow $\dot{\theta} = \nu$ on $\mathbb{T}^n$.

7.2 Integrable Systems

An especially interesting example of symplectic manifold is $M = \mathbb{R}^l \times \mathbb{T}^l$ which can be identified with the cotangent bundle of the l–dimensional torus $\mathbb{T}^l = \mathbb{R}^l/(2\pi\mathbb{Z})^l$. This manifold has a natural symplectic structure defined by the closed 2–form $\omega = \sum_{i=1}^{l} dJ_i \wedge d\vartheta_i$ where $(J_1, \dots J_l, \vartheta_1, \dots \vartheta_l)$ are coordinates on $\mathbb{R}^l \times \mathbb{T}^l$.

Definition 7.10 *Let U denote an open connected subset of $\mathbb{R}^l$. Whenever an Hamiltonian system can be reduced by a symplectic change of coordinates to a function $H : U \times \mathbb{T}^l \to \mathbb{R}$ which* does not depend on the angular variables ϑ *one says that the system is* completely canonically integrable *and the variables J are called* action variables.

Note that in this case Hamilton's equations (7.2) take the particularly simple form

$$\dot{J}_i = -\frac{\partial H}{\partial \vartheta_i} = 0 \ , \quad \dot{\vartheta}_i = \frac{\partial H}{\partial J_i} \ , \quad i = 1, \dots, l$$

and the flow leaves invariant the l–dimensional torus $J =$ constant. The motion is therefore bounded and quasiperiodic (or periodic).

Being completely canonically integrable is a stronger requirement than integrability by quadratures or complete integrability (see [AKN] for their discussion).

In the latter case one requires the existence of l independent first integrals in involution but their joint level–set may well be non compact (this is already the case in the two body problem for non negative energy values) and the flow does not need to be quasiperiodic (scattering states).

The main risult in the theory of completely canonically integrable systems is the celebrated

Theorem 7.11 (Arnol'd–Liouville) *Let $H \in \mathcal{C}^\infty(M, \mathbb{R})$ and assume that $F_1, \ldots, F_l \in \mathcal{C}^\infty(M, \mathbb{R})$ are l first integrals in involution for the Hamiltonian flow associated to H. Let $a \in \mathbb{R}^l$ be such that $M_a = \{m \in M \mid F_i(m) = a_i \, \forall i = 1, \ldots, l\}$ is not empty and assume that the l functions $F_1, \ldots, F_l$ are independent*[1] *in a neighborhood of M_a. Then if M_a is compact and connected it is diffeomorphic to the l–torus. Moreover there exists an invariant open subset V of M which contains M_a and is symplectically diffeomorphic to $U \times \mathbb{T}^l$, where U is an open subset of $\mathbb{R}^l$.*

Arnol'd–Liouville's Theorem thus assures that the existence of sufficiently many first integrals together with the compactness and connectedness of their level set guarantees complete canonical integrability.

7.3 Examples of completely canonically integrable systems

In this section we will briefly describe some examples of completely canonical integrable systems.

Example 7.12 : Harmonic oscillators. Let $M = \mathbb{R}^{2l}$ with the standard symplectic structure, $S \in \mathrm{GL}\,(2l, \mathbb{R})$ be symmetric and positive definite. Consider the Hamiltonian system $H(x) = \frac{1}{2} x^T S x$. This is completely integrable. Indeed if J is a symplectic matrix which diagonalizes S, in the variables $y = J^{-1}x$ the Hamiltonian will be

$$H(y) = \sum_{i=1}^{2l} \frac{\lambda_i y_i^2 + \lambda_{i+l} y_{i+l}^2}{2} \ ,$$

where λ_i , $i = 1, \ldots, 2l$ are the eigenvalues of S. Then the functions $F_i = \frac{\lambda_i y_i^2 + \lambda_{i+l} y_{i+l}^2}{2}$, $i = 1, \ldots, l$, are independent first integrals in involution and their common level set is compact and connected (since $\lambda_i > 0$ for all i). The symplectic

[1] As usual $F_1, \ldots, F_l$ are independent if $dF_1 \wedge \ldots \wedge dF_l \neq 0$.

transformation to action–angle variables is

$$y_i = \sqrt{2J_i\sqrt{\lambda_{i+l}/\lambda_i}}\cos\chi_i \ , \quad y_{i+l} = \sqrt{2J_i\sqrt{\lambda_i/\lambda_{i+l}}}\sin\chi_i \ , \quad i = 1, \dots, l \ .$$

Example 7.13 : The two body problem. The Hamiltonian $\mathcal{H} : T^*(\mathbb{R}^3\backslash\{0\}) \mapsto \mathbb{R}$ of the two-body problem in the center of mass frame is (we have assumed $G = 1$, where G is the universal gravitational constant)

$$\mathcal{H}(p,q) = \frac{1}{2\mu}\|p\|^2 - \frac{m_0 m}{\|q\|}$$

where $\mu = m_0 m/(m_0 + m)$ is the reduced mass of the system.

It is well-known that for negative energy the solutions are ellipses with one focus at the origin (i. e. the center of mass). These are called *keplerian orbits*. The shape and the position of the ellipse in space are determined from the knowledge of the major semiaxis a, the eccentricity e, the angle of inclination i of its plane w.r.t. the horizontal plane $q_3 = 0$, the argument of perihelion ω and the longitude of the ascending node Ω. The position of the planet along the ellipse is determined by the mean anomaly l, which is proportional to the area swept by the position vector q of the planet starting from the perihelion.

The systems admits 5 *independent* first integrals : the total energy $\mathcal{H}$, the three components of the angular momentum $q \wedge p$ and one of the components of the Laplace–Runge–Lenz vector $A = p \wedge q \wedge p - \frac{m_0 m q}{\|q\|}$. Among these integrals one can choose three integrals in involution and construct the completely canonical transformation to action–angle variables. The other two integrals are responsible for the proper complete degeneration of the Kepler problem : one can choose action–angle variables so that the Hamiltonian depends only on one of the actions. Indeed the Delaunay action–angle variables $(L, G, \Theta, l, g, \theta)$ are related to the orbital elements as follows :

$$L = \mu\sqrt{(m_0 + m)a} \ , \ \ G = L\sqrt{1 - e^2} \ , \ \ \Theta = G\cos i \ , \ \ l \ , \ \ g = \omega \ , \ \ \theta = \Omega \ .$$

Note that G is the modulus of angular momentum $q \wedge p$, thus Θ is its projection along the q_3–axis. One has the obvious limitation $|\Theta| \le G$. The new Hamiltonian reads $\mathcal{H} = -\frac{\mu^3(m_0+m)^2}{2L^2}$.

The relation among Delaunay variables and the original momentum–position (p, q) variables is much more subtle and will not be discussed here.

The two–body problem is the modelization of the motion of a planet around the Sun. But the Delaunay variables are not suitable for the description of the orbits of the planets of the solar system since they are singular for circular orbits ($e = 0$, thus $L = G$ anf the argument of the perihelion g is not defined) and for horizontal orbits ($i = 0$ or $i = \pi$, thus $G = \Theta$ and the longitude of the ascending node θ is not defined). But all the planets of the solar system have almost circular orbits (with the exception of Mercury and Mars) and small inclinations.

Poincaré solved the problem first introducing a new set of action–angle variables $(\Lambda, H, Z, \lambda, h, \zeta)$: $\Lambda = L$, $H = L - G$, $Z = G - \Theta$, $\lambda = l + g + \theta$, $h = -g - \theta$, $\zeta = -\theta$ (λ is called the mean longitude, $-h$ is the longitude of the perihelion) then considering the couples (H, h) and (Z, ζ) as polar symplectic coordinates :

$$\xi_1 = \sqrt{2H} \cos h \ , \quad \eta_1 = \sqrt{2H} \sin h \ , \qquad \xi_2 = \sqrt{2Z} \cos \zeta \ , \quad \eta_2 = \sqrt{2Z} \sin \zeta \ .$$

The variables $(\Lambda, \xi, \lambda, \eta)$ are called *Poincaré variables*. They are well defined also in the case of circular $(H = 0)$ or horizontal $(Z = 0)$ orbits.

Example 7.14 : Motion of a "heavy" particle on a surface of revolution. Let $S \subset \mathbb{R}^3$ be a surface of revolution with the Riemannian metric induced by its embedding into $\mathbb{R}^3$ and assume that x_3 is its symmetry axis. Let $f \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R})$. If the surface never meets the x_3 axis then it is diffeomorphic to the cylinder $S \approx \mathbb{R} \times \mathbb{T}^1$ and its cotangent bundle will be $\mathbb{R}^3 \times \mathbb{T}^1$. If (p_1, p_2, q_1, q_2) are symplectic coordinates the Hamiltonian of a (heavy) point mass constrained to move on S is $H(p, q) = \frac{1}{2}\|p\|^2 + f(q_1)$. Here f is the "weight" and p_2 (which corresponds to the projection of the angular momentum of the particle along the x_3–axis) is an independent integral of the motion. Complete integrability is assured if the curve $\{(p_1, q_1) \in \mathbb{R}^2 \ | \ H(p_1, a, q_1, q_2) = E\}$ is closed for some value of a and E.

8. Quasi–integrable Hamiltonian Systems

The importance of completely canonically integrable Hamiltonian systems is due both to the fact that their flows can be studied in great detail and that many problems in mathematical physics can be considered as *perturbations* of integrable systems. The most famous example is given by the motion of the planets in the Solar System (see [Ma2] and references therein for an introduction). If the (weak) mutual attraction between the planets is neglected the system decouples into several independent Kepler problems and it is completely integrable. Exactly this problem gave origin in the 18th century to "perturbation theory" whose modern formulation is mainly due to the monumental work of Henri Poincaré [P]. The goal of pertubation theory is to understand the dynamics of a "perturbed" system which is close to a well–understood one (usually an integrable system).

8.1 Quasi–integrable Systems

Following Poincaré [P], the *fundamental problem of dynamics* is the study of quasi–integrable Hamiltonian systems : let $\varepsilon_0 > 0$,

Definition 8.1 *A* quasi–integrable *Hamiltonian system is a function $\mathcal{H} \in \mathcal{C}^\infty((-\varepsilon_0, \varepsilon_0) \times M, \mathbb{R})$ such that the Hamiltonian function $H = \mathcal{H}(0, \cdot) : M \to \mathbb{R}$ is completely canonically integrable.*

Using the canonical transformation to action–angle variables associated to $\mathcal{H}(0, \cdot)$, Hamiltonians $\mathcal{H} : (-\varepsilon_0, \varepsilon_0) \times U \times \mathbb{T}^l \mapsto \mathbb{R}$ (smooth or analytic) of the form

$$\mathcal{H}(\varepsilon, J, \chi) = h_0(J) + \varepsilon f(J, \chi) \ , \tag{8.1}$$

where $f \in \mathcal{C}^\infty(U \times \mathbb{T}^l, \mathbb{R})$, are typical examples of quasi–integrable Hamiltonian systems.

The most ambitious program would be to prove that quasi–integrable Hamiltonian systems are indeed integrable : i.e. to show that there exists a one–parameter family V_ε of open connected invariant subsets of M which are symplectically diffeomorphic to $U_\varepsilon \times \mathbb{T}^l$ where $U_\varepsilon \subset \mathbb{R}^l$ is open, connected and such that if $(\tilde{J}, \tilde{\chi})$ are the coordinates in $U_\varepsilon \times \mathbb{T}^l$ one has $\mathcal{H}(\varepsilon, \cdot, \cdot) \mid_{V_\varepsilon} = h_\varepsilon(\tilde{J})$ for some smooth one–parameter family of smooth function $h_\varepsilon : U_\varepsilon \times \mathbb{T}^l \to \mathbb{R}$.

In general this is asking too much : a result of Poincaré shows that in general quasi–integrable Hamiltonian systems are not completely integrable (in addition to [P], Tome I, Chapitre V, see [BFGG] for a nice discussion of the consequences of this problem and a related result of Fermi).

Theorem 8.2 (Poincaré) *Consider a quasi–integrable Hamiltonian of the form (8.1), $l \geq 2$. Assume that the two following genericity assumptions are satisfied : (1)* non–degeneracy *:* $\det\left(\frac{\partial^2 h_0}{\partial J_i \partial J_k}\right) \neq 0$ *on U ; (2)* generic perturbations *: for all $J \in U$ and for all $k \in \mathbb{Z}^l \setminus \{0\}$ either the k–th Fourier coefficient $\hat{f}_k(J)$ of f does not vanish or there exists $k' \in \mathbb{Z}^l \setminus \{0\}$ parallel to k such that $\hat{f}_{k'}(J) \neq 0$. Then the system is not a smooth one–parameter family of completely canonically integrable Hamiltonians.*

One can also recall the following theorem of Markus and Meyer [MM]

Theorem 8.3 *Generically hamiltonian systems are neither completely canonically integrable nor ergodic*

Exercise 8.4 Prove Poincaré's Theorem following these lines. Using the notations introduced above, if the system were completely canonically integrable then the new actions $\tilde{J}$ would be a system of l independent first integrals of the Hamiltonian flow of $\mathcal{H}$ in involution. Writing them explicitly in terms of the old local coordinates (J, χ) one has

$$\tilde{J} = J + \varepsilon \tilde{J}_1(J, \chi) + \mathcal{O}(\varepsilon^2) \;, \tag{8.2}$$

for some smooth function $\tilde{J}_1 \,:\, V_\varepsilon \to U_\varepsilon$. Imposing that $\{\tilde{J}, \mathcal{H}\} = \mathcal{O}(\varepsilon^2)$ leads to the system of linear partial differential equations

$$\sum_{i=1}^{l} \frac{\partial h_0}{\partial J_i} \frac{\partial \tilde{J}_{1j}}{\partial \chi_i} = \frac{\partial f}{\partial \chi_j} \;, \quad j = 1, \ldots, l. \tag{8.3}$$

Using Fourier series try to find a smooth solution to these equations ….

8.2 Constant Coefficients Linear PDE on $\mathbb{T}^n$ and Loss of Differentiability.

The (very) short sketch of the proof of Poincaré's Theorem led us to consider the general constant coefficients linear partial differential equation on $\mathbb{T}^n$

$$D_\mu u := \mu \cdot \partial u = v \;, \tag{8.4}$$

where $\mu \in \mathbb{R}^n$, $\partial u = (\partial_1 u, \ldots, \partial_n u)$, is the gradient of u, $v \in \mathcal{C}^{0,\infty}(\mathbb{T}^n, \mathbb{R}^m)$ (i.e. $v \in \mathcal{C}^\infty(\mathbb{T}^n, \mathbb{R}^m)$ *and* $\int_{\mathbb{T}^n} v(\theta) d\theta = 0$). Indeed for all fixed value of J the equation (8.3) is a special case of (8.4) with $n = m = l$.

It is easy to check (see Appendix A3 for a detailed discussion of the case $n = 2$) that D_μ is hypoelliptic [1] if and only if μ is a diophantine vector, i.e. there exist two constants $\gamma > 0$ and $\tau \geq n-1$ such that

$$|\mu \cdot k| \geq \gamma |k|^{-\tau} \ \forall k \in \mathbb{Z}^n \setminus \{0\} \ , \tag{8.5}$$

where $k = (k_1, \ldots k_n)$, $|k| = |k_1| + \ldots + |k_n|$.

Exercise 8.5 Prove that almost all $\mu \in \mathbb{R}^n$ is diophantine of exponent $\tau > n-1$.

Exercise 8.6 Have a look to the book of Y. Meyer [Me]. Among many interesting things one finds the following theorem (Proposition 2, p. 16) : Let $\mathcal{R}$ be a real algebraic number field and let n be its degree over $\mathbb{Q}$. Let σ be the $\mathbb{Q}$–isomorphism of $\mathcal{R}$ such that $\sigma(\mathcal{R}) \subset \mathbb{R}$ and let $\mu_1, \ldots \mu_n$ be any basis of $\mathcal{R}$ over $\mathbb{Q}$. Then $(\sigma(\mu_1), \ldots, \sigma(\mu_n)) \in \mathbb{R}^n$ is diophantine of exponent $\tau = n-1$. Try to prove it if you remember a tiny bit of Galois theory. Apply it to $(1, \sqrt{2}, \sqrt{3}, \sqrt{6})$ and $(1, 2^{1/3}, 2^{2/3})$. There exist also higher dimensional generalizations of Roth's Theorem quoted in Exercise 4.9 : see, for example, the Subspace Theorem [Sch1, Sch2].

In addition to knowing that $u \in \mathcal{C}^\infty(\mathbb{T}^n, \mathbb{R}^m)$ one has the following more precise estimate :

Proposition 8.7 *Let* $\| \ \|_k$ *denote the* $\mathcal{C}^k$ *norm. If* μ *is diophantine with exponent* τ *then for all* $r > \tau + n - 1$ *and for all* $i \in \mathbb{N}$ *there exists a positive constant* A_i *such that*

$$\|u\|_i \leq A_i \|v\|_{i+r} \ . \tag{8.6}$$

Proof. Let $u(\theta) = \sum_{k \in \mathbb{Z}^n} \hat{u}_k e^{2\pi i k \cdot \theta}$, where obvoiusly one has

$$\hat{u}_k = \int_{\mathbb{T}^n} u(\theta) e^{-2\pi i k \cdot \theta} d\theta \ .$$

[1] A constant coefficients linear partial differential operator P is hypoelliptic if all u such that $Pu = v$ are $\mathcal{C}^\infty$ on all open sets where v is $\mathcal{C}^\infty$ (see [H1], p.109).

Then the $\mathcal{C}^k$–norm can be equivalently given in terms of Fourier coefficients : for all $i \in \mathbb{N}$ there exists a positive constant B_i such that

$$B_i^{-1} \sup_{k\in\mathbb{Z}^n} [(1+|k|)^i |\hat{u}_k|] \le \|u\|_i \le B_i \sup_{k\in\mathbb{Z}^n} [(1+|k|)^{i+n+1} |\hat{u}_k|] \ . \tag{8.7}$$

Comparing the Fourier coefficients of u with those of v one has

$$\hat{u}_k = \frac{\hat{v}_k}{2\pi i k\cdot\mu} \quad \forall k \in \mathbb{Z}^n \setminus \{0\} \ , \tag{8.8}$$

The desired estimates are an easy consequence of (8.7), the assumption that μ is diophantine and of the elementary fact $\sum_{k\in\mathbb{Z}^n\setminus\{0\}} |k|^{-\delta} < +\infty$ for all $\delta > n$. $\square$

The fact that one needs r more derivatives to bound the norms of u in terms of those of v is what is called the "loss of differentiability". As we have already seen in Chapter 6 this is a typical phenomenon associated to small divisors. The analogue in the analytic case would be the necessary restriction of the domain to control the maximum norm of u in terms of v by means of Cauchy's estimates as we did in Section 6.3.

In both cases (smooth and analytic) these are not artefacts of the methods used but a concrete manifestation of the unboundedness of the linear operator D_μ^{-1}. The main consequence of this fact is that one cannot use Banach spaces techniques to study semilinear equations like $D_\mu u = v + \varepsilon f(u)$, where ε is some small parameter. These semilinear equations are however typical of perturbation theory.

8.3 KAM Theory, Nekhoroshev Theorem, Arnol'd Diffusion

Despite Theorem 8.2, most results on quasi–integrable systems have been obtained under the assumption of non–degeneracy (i.e. the hessian matrix of h_0 is non degenerate thus the frequency map $J \mapsto \nu_0(J) = \frac{\partial h_0}{\partial J}(J) \in \mathbb{R}^l$ is a local diffeomorphism) but accepting the fact that one cannot hope for integrability on open sets.

The general picture is provided by KAM [Ar1,Ga, Bo, Yo1] and Nekhoroshev [Ne, Lo] theorems : if ε is sufficiently small, most initial conditions (w.r.t. Lebesgue measure) lie on invariant l–dimensional lagrangian tori carrying quasiperiodic motions with Diophantine frequencies. The action variables corresponding to these orbits will remain ϵ–close to their initial values for all times. The complement of this set is open and dense and it is connected if $l \geq 3$. It contains a connected ($l \geq 3$) web $\mathcal{R}$ of resonant zones corresponding to $\mathbb{Z}^l$–linearly dependent frequencies : $\cup_{k \in \mathbb{Z}^l} \{J \in U \,, \nu_0(J) \cdot k = 0\} \times \mathbb{T}^l$. Motion along these resonances cannot be excluded (see [Ar2] for an explicit example), resulting in a variation of $\mathcal{O}(1)$ of the actions in a finite time[2]. But if the hamiltonian is *analytic* and h_0 is *steep* (for example convex or quasi convex) then this variation is very slow : it takes a time at least $\mathcal{O}\left(\exp\left(\frac{1}{\varepsilon^a}\right)\right)$ to change the actions of $\mathcal{O}(\varepsilon^b)$, where a and b are two positive constants. Moreover each invariant torus has a neighborhood filled in with trajectiories which remain close to it for an even longer time. Indeed, if h_0 is quasi–convex one can prove [GM] that all trajectories starting at a distance of order $\rho < \rho^*$ from a Diophantine l–torus of exponent τ will remain close to it for a time $\mathcal{O}\left(\exp\left(\exp\left(\frac{\rho^*}{\rho}\right)^{1/\tau+1}\right)\right)$.

One of the consequences of KAM theorem [Pö] is the existence, for sufficiently small values of ε, of a *Cantor set* N_ε of values of the frequencies ν for which the Hamiltonian system (8.1) has smooth invariant tori with linear flow. Moreover there exists a homeomorphism $F_\varepsilon \, : N_\varepsilon \times \mathbb{T}^l \to U \times \mathbb{T}^l$ ε–close to the identity, Whitney smooth w.r.t. the first factor and analytic w.r.t. the second (if the Hamiltonian (8.1) is analytic) which transforms Hamilton's equations into $\dot{\nu} = 0$, $\dot{\varphi} = \nu$. This foliation into invariant tori is thus parametrized over a Cantor set and hence nowhere dense. It exhibits the phenomenon of "anisotropic differentiability" since it is much more regular tangentially to these tori than transversally to them (see also [BHS]).

[2] It is conjectured [AKN, p. 189] that generically quasi–integrable hamiltonians with more than two degrees of freedom are topologically unstable

9. The Inverse Function Theorem of Nash and Moser

The Inverse Function Theorem for Banach spaces is one of the extremely useful standard tools in the study of a variety of non–linear problems, ranging from the good position of the Cauchy problem for ordinary differential equations to non–linear elliptic equations. Unfortunately the "loss of differentiability" typical of small divisors problems prevents from its use (with some remarkable exceptions however, see Section 6.2 and [He2]). In the analytic case, Kolmogorov suggested the use of a modified Newton method to overcome this difficulty but in the differentiable case the need of an Inverse Function Theorem in Fréchet spaces has also other sources : its origin is the solution of the embedding problem for Riemannian manifolds by Nash [N]. Later Moser discovered how to adapt Kolmogorov's idea to the differentiable case creating a theory with a wide spectrum of applications [AG, Gr, H2, Ha, Ni, Ser, St, SZ, Ze1] : to geometry, to the study of foliations and deformations of complex and CR structures, to free boundary problems, etc.. In all these cases a non–linear partial differential equation is solved using a rapidly convergent iterative algorithm introducing at each step of the iteration a smoothing of the approximate solution.

In this Chapter we will follow the presentation of [Ha] very closely.

9.1 Calculus in Fréchet Spaces

Definition 9.1 *A* Fréchet space *is a locally convex topological vector space (lctvs) which is complete, Hausdorff and metrizable.*

Exercise 9.2 Show that a lctvs X is Hausdorff if and only if $x \in X$, $\|x\|_i = 0\ \forall i \in \mathcal{I}$ then $x = 0$ (where $(\|\cdot\|_i)_{i\in\mathcal{I}}$ is the collection of seminorms giving the topology of X). Show that X is metrizable if and only if $\mathcal{I}$ is countable.

Exercise 9.3 Show that $\mathbb{R}^\infty$ (space of all sequences of real numbers), $\mathcal{C}^\infty(M)$ (where M is a smooth compact manifold), $\mathcal{A}(\mathbb{C})$ (entire functions) are Fréchet spaces (thus the exercise asks you to define suitable seminorms). Show that $\mathcal{C}_0(\mathbb{R})$ (continuous functions with compact support) with the usual topology ($f_n \to f$ if and only if there exists a compact interval I such that $\operatorname{supp} f_n \subset I$ for all sufficiently large n, $\operatorname{supp} f \subset I$ and f_n converges uniformly to f on I) is a lctvs but it is not a Fréchet space since it is not metrizable.

Exercise 9.4 Prove that Hahn–Banach Theorem holds in Fréchet spaces : if X is a Fréchet space and x is a non–zero vector in X then there exists a continuous linear functional $l : X \to \mathbb{R}\,(\text{or}\,\mathbb{C})$ such that $l(x) = 1$. This allows to introduce quite straightforwardly X–valued analytic functions [Va]. A function $x : \Omega \to X$, where Ω is a region in $\mathbb{C}$, is analytic if and only if for all $l \in X^*$ the function $l \circ x$ is analytic. Show that this is equivalent to asking that, for all $z_0 \in \Omega$, x has a convergent power series expansion at z_0 : $x(z) = \sum_{n=0}^{\infty}(z - z_0)^n x_n$.

Exercise 9.5 Extend the theory of Riemann's integration, including the fundamental theorem of calculus, to continuous X–valued functions on $[a, b] \subset \mathbb{R}$.

Definition 9.6 *Let X, Y be two Fréchet spaces, $U \subset X$ be open, $f : U \to Y$ be continuous. The* derivative *of f at $x \in U$ in the direction of $h \in X$ is*

$$Df(x) \cdot h := \lim_{t \to 0} \frac{f(x + th) - f(x)}{t} \; . \tag{9.1}$$

f is $\mathcal{C}^1$ on U if and only if Df exists for all $x \in U$ and for all $h \in X$ and *$Df : U \times X \to Y$ is continuous.*

Remark 9.7 In the case of Banach spaces this definition of $\mathcal{C}^1$ is weaker than the usual one.

Exercise 9.8 Prove that the composition of $\mathcal{C}^1$ maps is $\mathcal{C}^1$ and that the chain rule holds : $D(g \circ f)(x) \cdot h = Dg(f(x)) \cdot (Df(x) \cdot h)$.

Exercise 9.9 Define higher order derivatives and $\mathcal{C}^k$ maps between Fréchet spaces.

Exercise 9.10 Let $f : \mathcal{C}^\infty([a, b]) \to \mathcal{C}^\infty([a, b])$, $f(x) = P(x, x', \ldots, x^{(n)})$, where $P \in \mathbb{R}[X_0, \ldots X_n]$, is $\mathcal{C}^\infty$. Is there a nice formula for $Df(x) \cdot h$? [Hint : start from monomials like $(x^{(i)})^k$.]

The following examples show why the extension of the inverse function theorem to Fréchet spaces is not a straightforward generalization of the inverse function theorem in Banach spaces but needs some extra assumption.

The map $x \mapsto f(x) = \sin x$, where $x \in X = L^2([0, 1])$, is of class $\mathcal{C}^1$ according to Definition 9.6. Its derivative $Df(0) =$identity but f is not invertible : $f(0) = 0$ and the functions $x_n(\xi) = \pi\chi_{[0,1/n]}(\xi)$, where $\chi_{[0,1/n]}$ denotes the characteristic function of the interval $[0, 1/n]$, converge to $x = 0$ but $f(x_n) = 0$ for all n.

Another example is obtained taking $X = \mathcal{C}^\infty([-1,1])$ and considering the map $f : X \to X$ defined as $f(x)(\xi) = x(\xi) - \xi x(\xi)x'(\xi)$ for all $\xi \in [-1,1]$. Then it is immediate to check that f is smooth and $Df(x)\cdot h = h - \xi x'h - \xi x h'$, thus $f(0) = 0$ and $Df(0) =$identity. But f is not invertible : the sequence $x_n(\xi) = \frac{1}{n} + \frac{\xi^n}{n!} \to 0$ in $\mathcal{C}^\infty([-1,1])$ as $n \to +\infty$ but one can check that it does not belong to $f(\mathcal{C}^\infty([-1,1]))$ for all $n \geq 1$. [Hint : use the fact that if $x \in \mathcal{C}^\infty([-1,1])$ one can take its Taylor series at 0 at any finite order and apply f.]

An even more interesting counterexample (see [Ha] for details) is the following : let M be a compact manifold, $X = \mathcal{C}^\infty(M, TM)$ be the Fréchet space of smooth vector fields on M, $\mathrm{Diff}^\infty(M)$ be the group of smooth diffeomorphisms of M (it is a Fréchet manifold, it's not very hard to figure out what this means, otherwise look in [Ha]). Then the usual exponential map

$$\begin{aligned} \exp : \mathcal{C}^\infty(M, TM) &\to \mathrm{Diff}^\infty(M) \\ v &\mapsto \exp(v) \end{aligned}$$

clearly verifies $\exp(0) = \mathrm{id}_M$ and $D\exp(0) =$identity, but the exponential map is not invertible in general. Note that this would have meant that any diffeomorphism extends to a one parameter flow.

For example a diffeomorphism of $\mathbb{S}^1$ without fixed points is the exponential of a vector field only if it is conjugate to a rotation. But there exist [Yo3] diffeomorphisms of $\mathbb{S}^1$ arbitrarily close to the identity which are not conjugate to a rotation.

What goes wrong in all these examples is that although the derivative of the map is the identity at the origin it fails to be invertible at nearby points. Indeed in the second example above one has $Df(1/n)\cdot\xi^k = \left(1 - \frac{k}{n}\right)\xi^k$, thus $Df(1/n)\xi^n = 0$.

Thus *one has to require the invertibility of Df on a neighborhood explicitly* and this is usually difficult to be checked. In a Banach space (with the usual definition of derivative of a map instead of Definition 9.6) this is not needed.

9.2 Tame Maps and Tame Spaces

Definition 9.11 *A* graded Fréchet space *X is a Fréchet space with a collection of seminorms $(\|\ \|)_{n\in\mathbb{N}}$ which define the topology and are increasing in strength*

$$\|x\|_0 \leq \|x\|_1 \leq \|x\|_2 \leq \ldots \quad \forall x \in X\ .$$

Definition 9.12 *Let X, Y be graded Fréchet spaces, $U \subset X$ open, $P : U \to Y$ be a continuous map. P is* tame *if for all $x_0 \in U$ there exists a neighborhood $V \subset U$ of x_0 and a non negative integer r such that for all $i \in \mathbb{N}$ there exists $C_i > 0$ such that*

$$\|P(x)\|_i \leq C_i(1 + \|x\|_{i+r}) \ \ \forall x \in V \ . \tag{9.2}$$

A $\mathcal{C}^k$ tame map is a $\mathcal{C}^k$ map P such that $D^j P$ is tame for all $0 \leq j \leq k$.

The most typical example of a tame operator between Fréchet spaces is given by nonlinear partial differential operators on compact manifolds. If $P : \mathcal{C}^\infty(M) \to \mathcal{C}^\infty(M)$ is a smooth function of $x \in \mathcal{C}^\infty(M)$ and its partial derivatives of degree at most r then we say that the degree of P is r and this will be the "loss of differentiability" in (9.2). The proof of this fact is given in [Ha] and uses Hadamard's inequalities for functions $x \in \mathcal{C}^\infty(M)$: for all $n \in N$ and for all integer k such that $0 \leq k \leq n$ there exists $C_{k,n} > 0$ such that

$$\|x\|_k \leq C_{k,n}\|x\|_n^{k/n}\|x\|_0^{1-k/n} \ \forall x \in \mathcal{C}^\infty(M) \text{ and } . \tag{9.3}$$

Exercise 9.13 Prove that the composition of two $\mathcal{C}^k$ tame maps is a $\mathcal{C}^k$ tame map.

Definition 9.14 *A graded Fréchet space X is* tame *if it admits* smoothing operators, *i.e. a one–parameter family $S(t) : X \to X$, $t \in [1, +\infty)$, of continuous linear operators such that there exists a non negative integer r and positive real constants $(C_{n,k})_{n,k\in\mathbb{N}}$ such that for all $x \in X$ and for all $t \in [1, +\infty)$ and for all $k \in \{0, 1, \ldots, n\}$ one has*

$$\begin{aligned} \|S(t)x\|_n &\leq C_{n,k}t^{n-k}\|x\|_k \\ \|x - S(t)x\|_k &\leq C_{k,n}t^{k-n}\|x\|_n \end{aligned} \tag{9.4}$$

Exercise 9.15 (convolution with regularizing kernels) Let $\psi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ and assume that $\psi \geq 0$, $\psi \equiv 1$ near 0. Let φ be the Fourier transform of ψ : $\varphi(\xi) = \int_{\mathbb{R}^n} \psi(\eta)e^{-2\pi i\xi\eta}d\eta$. Let $t \geq 1$, $\varphi_t(\xi) = t^n\varphi(t\xi)$. Define $S(t)f = \varphi_t \star f$, where $f \in \mathcal{C}^\infty(\mathbb{T}^n)$. Show that $(S(t))_{t\geq 1}$ is a family of smoothing operators on $\mathcal{C}^\infty(\mathbb{T}^n)$.

Exercise 9.16 Prove that in a tame Fréchet space Hadamard's inequalities hold :

$$\|x\|_l \leq C(k,n)\|x\|_k^{1-\alpha}\|x\|_n^\alpha \ \ \forall k \leq l \leq n \ \ , \ \ l = (1-\alpha)k + \alpha n \ .$$

[Hint : use (9.4) with $t = \|x\|_n^{1/(n-k)} \|x\|_k^{-1/(n-k)}$.]

9.3 The Nash–Moser Theorem

We can finally state Nash–Moser's [N,M] implicit and inverse function theorems.

Theorem 9.17 (implicit function) *Let X, Y, Z be three tame Fréchet spaces, $U \subset X \times Y$ open, $\Phi : U \to Z$ a tame $\mathcal{C}^r$ map, $2 \leq r \leq \infty$. Let $(x_0, y_0) \in U$. Assume that there exists a neighborhood V_0 of (x_0, y_0) and a continuous z–linear tame map $L : V_0 \times Z \to Y$, $((x,y),z) \mapsto L(x,y) \cdot z$, such that if $(x,y) \in V_0$ then $D_y\Phi(x,y)$ is invertible with inverse $L(x,y)$. Then x_0 has a neighborhood W on which $\Psi \in \mathcal{C}^r(W, Y)$ is defined and such that $\Psi(x_0) = y_0$ and for all $x \in W$ one has $(x, \Psi(x)) \in U$ and $\Phi(x, \Psi(x)) = \Phi(x_0, y_0)$.*

Theorem 9.18 (inverse function) *Let X, Y be two tame Fréchet spaces, $U \subset X$ open, $\Phi : U \to Y$ a tame $\mathcal{C}^r$ map, $2 \leq r \leq \infty$. Let $x_0 \in U$, $y_0 = \Phi(x_0)$. Assume that there exists a neighborhood V_0 of x_0 and a continuous y–linear tame map $L : V_0 \times Y \to X$, $(x, y) \mapsto L(x) \cdot y$, such that if $x \in V_0$ then $D\Phi(x)$ is invertible with inverse $L(x)$. Then x_0 has a neighborhood $V \subset V_0$ and y_0 has a neigborhood W such that $\Phi : V \to W$ is a tame $\mathcal{C}^r$ diffeomorphism.*

Exercise 9.19 Show that the two previous theorems are equivalent.

We refer the reader to [Ha] for the proofs of Theorems 9.17 and 9.18. The main idea of the proof is to use a modified Newton's method for finding the root of the equation $\Phi(x) = y$. It makes use of the smoothing operators $S(t)$ to guarantee convergence. Here we will content ourselves with a brief sketchy description of the argument.

Without loss of generality we can assume $x_0 = y_0 = 0$. An algorithm for constructing a sequence $x_j \in X$ which will converge to a solution x of $\Phi(x) = y$ (for small enough y) is the following : fix a sequence $t_j = e^{(3/2)^j}$, so that $t_{j+1} = t_j^{3/2}$, and let

$$\begin{aligned}
x_0 &= 0 \quad \text{(initial guess)}\ , \\
x_j &= \dots \quad \text{(j–th guess)}\ , \\
z_j &= y - \Phi(x_j) \quad \text{(j–th error)}\ , \\
\Delta x_j &= S(t_j)L(x_j)z_j \quad \text{(j–th correction)}\ , \\
x_{j+1} &= x_j + \Delta x_j \quad \text{($j+1$–th guess)}\ .
\end{aligned}$$

The idea to show convergence of this algorithm is the following : let

$$R(x,h) = \int_0^1 D^2\Phi(x+th)(h,h)dt$$

denote the quadratic integral remainder in Taylor's formula. Since

$$\Phi(x+h) = \Phi(x) + D\Phi(x)\cdot h + R(x,h)$$

one gets

$$\begin{aligned}
z_{j+1} &= y - \Phi(x_{j+1}) = y - \Phi(x_j + \Delta x_j) \\
&= z_j - D\Phi(x_j)S(t_j)L(x_j)z_j - R(x_j, \Delta x_j)\ .
\end{aligned}$$

Using the identity $z_j = D\Phi(x_j)L(x_j)z_j$ we find

$$z_{j+1} = D\Phi(x_j)[I - S(t_j)]L(x_j)z_j + R(x_j, \Delta x_j)\ .$$

The first term tends to zero very rapidly since $S(t_j) \to I$ as $j \to +\infty$ and the second term is quadratic.

This short description of the idea of the proof makes also clear why one needs the assumption Φ at least of class $\mathcal{C}^2$ (in Banach spaces $\mathcal{C}^1$ is enough).

10. From Nash–Moser's Theorem to KAM : Normal Form of Vector Fields on the Torus

Following Herman we will prove in this Chapter a normal form theorem for vector fields on the torus which can be considered as the basic KAM theorem in higher dimension (without taking the symplectic structure into account). The proof will be an application of Nash–Moser's Theorem. For a proof of KAM theorem see, for example, [Bo].

Let $\mathrm{Diff}^\infty(\mathbb{T}^l, 0)$ denote the group of $\mathcal{C}^\infty$ diffeomorphisms f of the torus $\mathbb{T}^l$ homotopic to the identity and such that $f(0) = 0$. This space can be identified to an open subset of the tame Fréchet space $\mathcal{C}^\infty(\mathbb{T}^l, \mathbb{R}^l, 0) = \{u \in \mathcal{C}^\infty(\mathbb{T}^l, \mathbb{R}^l)\,,\, u(0) = 0\}$: u corresponds to a diffeomorphism f if and only if for all $\chi \in \mathbb{T}^l$ one has $\mathrm{id} + \partial u(\chi) \in \mathrm{GL}\,(l, \mathbb{R})$. In this case one has $f = \mathrm{id}_{\mathbb{T}^l} + u$.

Since the tangent bundle of $\mathbb{T}^l$ is canonically isomorphic to $\mathbb{T}^l \times \mathbb{R}^l$ one can also identify the space of $\mathcal{C}^\infty$ vector fields on the torus with $\mathcal{C}^\infty(\mathbb{T}^l, \mathbb{R}^l)$.

Let $\mu \in \mathbb{R}^l$ be Diophantine with exponent τ and constant γ. We will denote R_μ the translation by μ on the torus $\mathbb{T}^l$: $R_\mu(\chi_1, \ldots, \chi_l) = (\chi_1 + \mu_1, \ldots, \chi_l + \mu_l)$.

Exercise 10.1 Show that the map

$$\begin{aligned} I\ :\ &\mathrm{Diff}^\infty(\mathbb{T}^l, 0) \to \mathrm{Diff}^\infty(\mathbb{T}^l, 0) \\ &f \mapsto I(f) = f^{-1} \end{aligned}$$

is a tame $\mathcal{C}^\infty$ map. Its derivative is

$$DI(f) \cdot h = -[(\partial f)^{-1} \cdot h] \circ f^{-1}\ . \tag{10.1}$$

The following statements (and proof) are taken from ([Bo], pp. 139–141) and [He5].

Theorem 10.2 *Let $\mu \in \mathbb{R}^l$. The map*

$$\begin{aligned} \Phi_\mu\ :\ &\mathit{Diff}^\infty(\mathbb{T}^l, 0) \times \mathbb{R}^l \to \mathit{Diff}^\infty(\mathbb{T}^l)\ , \\ &(f, \nu) \mapsto R_\nu \circ f \circ R_\mu \circ f^{-1}\ , \end{aligned} \tag{10.2}$$

is a tame $\mathcal{C}^\infty$ map. Moreover, if μ is a diophantine[1] *vector then Φ_μ is a tame $\mathcal{C}^\infty$ local diffeomorphism near $f = id_{\mathbb{T}^l}$, $\nu = 0$.*

[1] In this situation μ is diophantine if there exist two constants $\gamma > 0$ and $\tau \geq l$ such that $|\mu \cdot k + p| \geq \gamma |k|^{-\tau}$ for all $k \in \mathbb{Z}^l \setminus \{0\}$ and for all $p \in \mathbb{Z}$.

The meaning of the second part is that when μ is diophantine the diffeomorphisms of the torus $\mathbb{T}^l$ conjugate to the translation R_μ by a diffeomorphism close to the identity form a Fréchet submanifold of codimension l of $\mathrm{Diff}^\infty(\mathbb{T}^l)$ which is transverse in μ to the space $\mathbb{R}^l$ of the translations on the torus.

Exercise 10.3 Guess the statement of for vector fields equivalent to Theorem 10.2.

Here is the solution :

Theorem 10.3*Let $\mu \in \mathbb{R}^l$. The map*

$$\begin{aligned} \Psi_\mu \ : \mathit{Diff}^\infty(\mathbb{T}^l, 0) \times \mathbb{R}^l &\to \mathcal{C}^\infty(\mathbb{T}^l, \mathbb{R}^l) \ , \\ (f, \nu) &\mapsto \nu + f_* \mu = \nu + \partial f \circ f^{-1} \cdot \mu \ , \end{aligned} \tag{10.3}$$

is a tame $\mathcal{C}^\infty$ map. Moreover, if μ is a diophantine vector (see (8.5)) then Ψ_μ is a tame $\mathcal{C}^\infty$ local diffeomorphism near $f = id_{\mathbb{T}^l}$, $\nu = 0$.

Proof. First of all note that $\Psi_\mu(\mathrm{id}_{\mathbb{T}^l}, 0) = \mu$. The first assertion is an immediate consequence of Exercises 9.13 and 10.1. Moreover, using (10.1), one easily checks that

$$\begin{aligned} D\Psi_\mu(f, \nu) \cdot (\Delta f, \Delta \nu) &= \Delta \nu + (\partial \Delta f) \circ f^{-1} \cdot \mu \\ &\quad + \partial^2 f \circ f^{-1} \cdot (-(\partial f)^{-1} \circ f^{-1} \cdot \Delta f \circ f^{-1}, \mu) \\ &= \Delta \nu + [(\partial \Delta f) \cdot \mu \\ &\quad + \partial^2 f \cdot (-(\partial f)^{-1} \cdot \Delta f, \mu)] \circ f^{-1} \end{aligned} \tag{10.4}$$

(we recall that here one has $\Delta f \in \mathcal{C}^\infty(\mathbb{T}^l, \mathbb{R}^l, 0)$, $\Delta \nu \in \mathbb{R}^l$).

If one introduces u, writing $\Delta f = \partial f \cdot u$ then one gets

$$(\partial \Delta f) \circ f^{-1} \cdot \mu = [\partial^2 f \cdot u \cdot \mu + \partial f \cdot \partial u \cdot \mu] \circ f^{-1}$$

and

$$\partial^2 f \circ f^{-1} \cdot (-(\partial f)^{-1} \circ f^{-1} \cdot \Delta f \circ f^{-1}, \mu) = -[\partial^2 f \cdot u \cdot \mu] \circ f^{-1} \ .$$

Therefore (10.4) simplifies considerably and becomes

$$D\Psi_\mu(f, \nu) \cdot (\partial f \cdot u, \Delta \nu) = \Delta \nu + (\partial f \cdot \partial u \cdot \mu) \circ f^{-1} \ . \tag{10.5}$$

To prove the second assertion we will apply Theorem 9.18 to $\Phi = \Psi_\mu$ at the points $x_0 = (\mathrm{id}_{\mathbb{T}^l}, 0)$ and $y_0 = \mu$. We must just check that $D\Psi_\mu(f,\nu)$ is invertible *for all* (f,ν) in a neighborhood of $(\mathrm{id}_{\mathbb{T}^l}, 0)$. This leads us to the equation

$$\Delta\nu + (\partial f \cdot \partial u \cdot \mu) \circ f^{-1} = w \ . \tag{10.6}$$

Composing on the right with f and multiplying both sides by $(\partial f)^{-1}$ one gets

$$\mu \cdot \partial u = (\partial f)^{-1} \cdot [w \circ f - \Delta\nu] \ , \tag{10.7}$$

i.e. an equation of the form (8.4) with $v = (\partial f)^{-1} \cdot [w \circ f - \Delta\nu]$. This clarifies why one needs the term ν in the definition (10.3) of Ψ_μ : indeed one fixes it so as to assure that $v \in \mathcal{C}^{0,\infty}(\mathbb{T}^l, \mathbb{R}^l)$, i.e. it has zero average on the torus $\mathbb{T}^l$. One can also check that the map $(f,w) \mapsto \Delta\nu$ is tame.

Proposition 8.7 allows to conclude since it shows that the map $(f,\nu) \mapsto u = D_\mu^{-1} v$ is tame. □

Appendices

A1. Uniformization, Distorsion and Quasi–conformal maps

In this appendix we recall some elementary and less elementary facts from the theory of conformal and quasi–conformal maps of one complex variable.

A1.1 A nonempty connected open set is called a *region*.

Theorem A1.1 (The Maximum Principle) *If $f(z)$ is analytic and non–constant in a region Ω of the complex plane $\mathbb{C}$, then its absolute value $|f(z)|$ has no maximum in Ω.*

Proof. It is an easy consequence of the fact that non–constant analytic functions map open sets onto open sets. □

The maximum principle implies that if f is defined and continuous on a compact set K and analytic in the interior of K then the maximum of $|f(z)|$ on K is assumed on the boundary of K. Another easy consequence is the following

Exercise A1.2 (Schwarz's Lemma, automorphisms of the disk) *Schwarz's Lemma* : If $|f(z)|$ is analytic for $|z| < 1$ and satisfies the conditions $|f(z)| \leq 1$, $f(0) = 0$, then $|f(z)| \leq |z|$ and $|f'(0)| \leq 1$. If $|f(z)| = |z|$ for some $z \neq 0$ or if $|f'(0)| = 1$ then $f(z) = cz$ with $c \in \mathbb{C}$, $|c| = 1$. *Automorphisms of the disk* : Show that if $f : \mathbb{D} \to \mathbb{D}$ is an automorphism of the disk and $f(0) = 0$ then $|f'(0)| = 1$ and f is a rotation. Deduce from this that the group of automorphisms of the unit disk $\mathbb{D}$ is

$$\mathrm{Aut}\,(\mathbb{D}) = \{z \mapsto T(z) = \frac{az + b}{\bar{b}z + \bar{a}}\,,\ a, b \in \mathbb{C}\,,\ |a|^2 - |b|^2 = 1\}\ .$$

A1.2 A mapping f of a region Ω into $\mathbb{C}$ is called *conformal* if it is holomorphic and injective. Such maps are also called *univalent*. Since an analytic map is injective if and only if $f' \neq 0$ if f is univalent in Ω its derivative never vanishes, i.e. it has no critical points inside Ω. The most important result of the theory of conformal maps is certainly the

Theorem A1.3 (Riemann Mapping Theorem) *Given any simply connected region Ω which is not the whole plane, and a point $z_0 \in \Omega$ there exists a unique conformal map (the Riemann map) $f : \mathbb{D} \to \Omega$ such that f is onto, $f(0) = z_0$ and $f'(0) > 0$.*

Exercise A1.4 Drop the requirement $f'(0) > 0$. Then f is not unique but the number $|f'(0)|$ does not depend on f. It is called [Ah1] the *conformal capacity* of Ω with respect to z_0 and it will be denoted $C(\Omega, z_0)$. [Hint : use Exercise A1.3]

One should not think to an arbitrary simply connected region as the "potato" of PDEs but as a rather irregular object. For example the boundary needs not to be locally connected.

A1.3 Assume that the region Ω is bounded and $\partial\Omega$ is a closed Jordan curve (i.e. $\partial\Omega = \gamma([0,1])$, where $\gamma : [0,1] \to \mathbb{C}$ is continuous, $\gamma(0) = \gamma(1)$ and $\gamma(t_1) = \gamma(t_2)$ if and only if $t_1 = t_2$ or $t_1 = 0$, $t_2 = 1$). In this case the Riemann map $f : \mathbb{D} \to \Omega$ has a nice boundary behaviour :

Theorem A1.5 (Caratheodory) *A Riemann map $f : \mathbb{D} \to \Omega$ extends to a homeomorphism of $\overline{\mathbb{D}}$ onto $\overline{\Omega}$ if and only if $\partial\Omega$ is a closed Jordan curve.*

The topic of the boundary behaviour of conformal maps is very rich and it's an active research area : we refer to [Po] for more informations and references. We will only need two other results : the first extends Caratheodory's theorem dropping the assumption that the restriction of f to the boundary of the disk is injective.

Theorem A1.6 *Let $f : \mathbb{D} \to \Omega$ be a Riemann map. The following four conditions are equivalent :*

(i) *f has a continuous extension to $\overline{\mathbb{D}}$;*

(ii) *$\partial\Omega$ is a continuous curve, i.e. $\partial\Omega = \{\varphi(\zeta), \zeta \in \mathbb{T}\}$ with φ continuous ;*

(iii) *$\partial\Omega$ is locally connected ;*

(iv) *$\mathbb{C} \setminus \Omega$ is locally connected.*

Our second result, due to Fatou, applies to all $f : \mathbb{D} \to \mathbb{C}$ holomorphic *and bounded* (thus we are dropping the assumption of f being injective in $\mathbb{D}$).

Theorem A1.7 (Fatou) *Let $f : \mathbb{D} \to \mathbb{C}$ be holomorphic and bounded. Then f has a non–tangential limit at almost all points $\zeta \in \mathbb{T} = \partial\mathbb{D}$. Moreover if f is not identically zero then $\varphi(\zeta) = \lim_{r\to 1-} f(r\zeta)$ (which belongs to $L^\infty(\mathbb{T})$) is not zero almost everywhere.*

A1.4 Another fundamental result is the celebrated

Theorem A1.8 (Uniformization Theorem) *The only simply connected Riemann surfaces, up to biholomorphic equivalence, are the Riemann sphere $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, the complex plane $\mathbb{C}$ and the unit disk $\mathbb{D}$.*

Exercise A1.9 Prove that the group of automorphisms of the Riemann sphere is the group $\mathrm{PGL}\,(2,\mathbb{C})$ acting by homographies : if $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PGL}\,(2,\mathbb{C})$ then $z \mapsto g \cdot z = \frac{az+b}{cz+d}$. The group of automorphisms of the complex plane is simply the affine group.

A1.5 One can consider univalent functions f on regions of $\overline{\mathbb{C}}$ with values in $\overline{\mathbb{C}}$: in this case f must be meromorphic and injective. Here are some elementary properties :

Exercise A1.10

(a) If f is univalent on a region $\Omega \subset \overline{\mathbb{C}}$ then f is analytic except for at most a single simple pole and f' never vanishes.

(b) If $f : \Omega \to \Omega'$ is onto and univalent then $f^{-1} : \Omega' \to \Omega$ is also univalent.

(c) A univalent map is a homeomorphism.

(d) A univalent map preserves angles between curves and their orientation (that's why they're called conformal !).

(e) The composition of univalent maps is univalent ; f is univalent if and only if $1/f$ is univalent.

Exercise A1.11 Prove that if $f : \Omega \to \overline{\mathbb{C}}$ is univalent and $A \subset \Omega$ is measurable then

$$\text{Area}\,(f(A)) = \int_A |f'(x+iy)|^2 dxdy \ .$$

Exercise A1.12 (The Area Formula) Show that if $f : \mathbb{D} \to \mathbb{C}$ is univalent, letting $f(z) = \sum_{n=0}^{\infty} f_n z^n$, one has

$$\text{Area}\,(f(\mathbb{D})) = \pi \sum_{n=1}^{\infty} n|f_n|^2 \ .$$

[Hint : First consider the disk $\mathbb{D}_r$ of radius $r < 1$. If $f = u + iv$ then $\text{Area}\,(f(\mathbb{D})) = \int_{\partial \mathbb{D}_r} udv = \frac{i}{2} \int_{\partial \mathbb{D}_r} f d\overline{f}$. Then let $r \to 1-$.]

Let S_1 denote the collection of functions f univalent in $\mathbb{D}$ and such that $f(0) = 0$, $f'(0) = 1$, thus $f(z) = z + f_2 z^2 + \ldots$. With Σ_1 we will denote all functions $g(\zeta) = \zeta + g_0 + g_1 \zeta^{-1} + \ldots$ univalent in the outer disk $\mathbb{E} = \{\zeta \in \overline{\mathbb{C}}, |\zeta| > 1\}$. Clearly if $f \in S_1$ then $g(\zeta) = 1/f(\zeta^{-1})$ belongs to Σ_1 and omits 0. Conversely, if $g \in \Sigma_1$ and $g(\zeta) \neq 0$ for all $\zeta \in \mathbb{E}$ then $f(z) = 1/g(z^{-1})$ belongs to S_1. One of the most important results on univalent functions is the object of the following exercise :

Exercise A1.13 (Area Theorem) If $g \in \Sigma_1$ then $|g_1| \leq \sum_{n=1}^{\infty} n|g_n|^2 \leq 1$. Prove that equality holds if and only if $g(\zeta) = \zeta + g_0 + g_1 \zeta^{-1}$ with $|g_1| = 1$. [Hint : show that $\text{area}(\mathbb{C} \setminus g(\mathbb{E})) = \pi(1 - \sum_{n=1}^{\infty} n|g_n|^2)$.]

One of the main consequences is the following apparently innocent bound : if $f \in S_1$ then

$$|f_2| \leq 2 \ , \tag{A1.1}$$

as one can easily check applying the Area Theorem to $g(\zeta) = \sqrt{f(\zeta^{-2})}$. However this estimate will have many important consequences as we will see soon. A (much harder and for a long time conjectural) result is the celebrated [DeB]

Theorem A1.14 (Bieberbach–De Branges) *If $f \in S_1$ then $|f_n| \leq n$ for all n.*

Using (A1.1) one can easily show that the image of $\mathbb{D}$ through a univalent map cannot be too small :

Theorem A1.15 (Koebe $1/4$**–Theorem)** *If* $f \in S_1$ *then* $\mathbb{D}_{1/4} \subset f(\mathbb{D})$.

Proof. Let $w \in \mathbb{D}$ and assume $w \notin f(\mathbb{D})$. Then

$$\tilde{f}(z) = \frac{wf(z)}{w - f(z)} = z + (f_2 + w^{-1})z^2 + \dots$$

belongs to S_1. Applying (A1.1) to both f and $\tilde{f}$ one gets $|w|^{-1} \leq |f_2| + |f_2 + w^{-1}| \leq 4$. □

The *Koebe function* $f(z) = z(1-z)^{-2} = \sum_{n=1} nz^n$ maps the unit disk $\mathbb{D}$ conformally onto $\mathbb{C} \setminus (-\infty, -1/4)$. Therefore Bieberbach–De Branges' Theorem and Koebe 1/4–Theorem are optimal. If f is univalent and analytic in $\mathbb{D}$, given any $z_0 \in \mathbb{D}$ the *Koebe transform* of f at z_0

$$\begin{aligned} K_{z_0,f}(z) &= \frac{f\left(\frac{z+z_0}{1+\overline{z_0}z}\right) - f(z_0)}{(1-|z_0|)^2 f'(z_0)} \\ &= z + \left[\frac{(1-|z_0|)^2 f''(z_0)}{2f'(z_0)} - \overline{z_0}\right] z^2 + \dots \end{aligned} \qquad (A1.2)$$

belongs to S_1. This is a very useful tool in order to transfer the information at 0 to information at any point of the disk. Applying systematically this idea, from (A1.1) one deduces the following important distorsion estimates :

Exercise A1.16 (Koebe distortion theorems) If f maps $\mathbb{D}$ conformally into $\mathbb{C}$ then $\forall z \in \mathbb{D}$ one has :

$$\left|(1-|z|^2)\frac{f''(z)}{f'(z)} - 2\overline{z}\right| \leq 4\,, \qquad (A1.3)$$

$$|f'(0)|\frac{|z|}{(1+|z|)^2} \leq |f(z) - f(0)| \leq |f'(0)|\frac{|z|}{(1-|z|)^2}\,, \qquad (A1.4)$$

$$|f'(0)|\frac{1-|z|}{(1+|z|)^3} \leq |f'(z)| \leq |f'(0)|\frac{1+|z|}{(1-|z|)^3}\,, \qquad (A1.5)$$

$$\frac{1}{4}(1-|z|^2)|f'(z)| \leq \mathrm{dist}\,(f(z), \partial f(\mathbb{D})) \leq (1-|z|^2)|f'(z)|\,. \qquad (A1.6)$$

These estimates show that the growth of f as z approaches $\partial\mathbb{D}$ cannot be faster than $(1-r)^{-2}$, where $r = |z|$. The next theorem (see [Po]) for a proof) shows that the average growth is much lower than $(1-r)^{-2}$.

Theorem A1.17 *Let f map $\mathbb{D}$ conformally into $\mathbb{C}$. Then $f(\zeta) = \lim_{r \to 1} f(r\zeta) \neq \infty$ exists for almost all $\zeta \in \mathbb{T} = \partial\mathbb{D}$ and for $0 \leq r < 1$ one has*

$$\frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta} - f(0)|^{2/5} dt \leq 5|f'(0)|^{2/5} \ . \qquad (A1.7)$$

A1.6 We conclude our brief introduction to univalent functions with the proof of a fundamental property of S_1. In order to do this we recall ([Re], p. 163)

Lemma A1.18 (Hurwitz) *If a sequence $(f_n)_{n \in \mathbb{N}}$ of functions holomorphic in a region $\Omega \subset \mathbb{C}$ converges uniformly on compact subsets of Ω to a non–constant holomorphic function $f : \Omega \to \mathbb{C}$ then the following statements hold :*
(a) if all the images $f_n(\Omega)$ are contained in a fixed set A then $f(\Omega) \subset A$;
(b) if all the maps $f_n : \Omega \to \mathbb{C}$ are injective then so is $f : \Omega \to \mathbb{C}$;
(c) if all the maps $f_n : \Omega \to \mathbb{C}$ are locally biholomorphic, then so is $f : \Omega \to \mathbb{C}$.

This is the ingredient we missed for the proof of the following

Theorem A1.19 *S_1 endowed with the topology of uniform convergence on compact subsets of $\mathbb{D}$ is a compact topological space.*

Proof. Any sequence $(f_n)_{n \in \mathbb{N}} \subset S_1$ is equicontinuous and uniformly bounded on compact subsets of $\mathbb{D}$ by Koebe distortion theorems. Limit functions are in S_1 because they are univalent by Hurwitz's lemma and the normalisation $|f_n'(0)|$ for all $n \in \mathbb{N}$. □

Exercise A1.20 Prove that Theorem A1.19 is equivalent to the following (see [Mc]) : the space of *all* univalent maps $f : \mathbb{D} \to \overline{\mathbb{C}}$ is compact up to post–composition with automorphisms of $\overline{\mathbb{C}}$. This precisely means that any sequence of univalent maps contains a subsequence $f_n : \mathbb{D} \to \overline{\mathbb{C}}$ such that $M_n \circ f_n$ converges to a univalent map f, uniformly on compact subsets of $\mathbb{D}$, for some sequence of Möbius transforms $M_n \in \mathrm{PGL}\,(2, \overline{\mathbb{C}})$.

A1.7 Let $f : \mathbb{C} \to \mathbb{C}$ be a $\mathcal{C}^1$ orientation–preserving diffeomorphism. Then given any point $z_0 \in \mathbb{C}$ one has

$$f(z) = f(z_0) + f_z(z_0)(z - z_0) + f_{\bar{z}}(z_0)(\bar{z} - \bar{z}_0) + \mathrm{o}\,(|z - z_0|) \ , \qquad (A1.8)$$

where

$$f_z = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right) \ , \quad f_{\bar z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right) \ , \quad (z = x+iy) \ . \qquad (A1.9)$$

Note that if f is analytic in z_0 then $f_{\bar z}(z_0) = 0$ (Cauchy–Riemann). The Jacobian determinant of f is $J = |f_z|^2 - |f_{\bar z}|^2$. Since f is orientation–preserving one has $J > 0$, thus $|f_z| > |f_{\bar z}|$.

Definition A1.21 *The* dilatation *of f in z_0 is*

$$D_f(z_0) := \frac{|f_z(z_0)| + |f_{\bar z}(z_0)|}{|f_z(z_0)| - |f_{\bar z}(z_0)|} \geq 1 \ . \qquad (A1.10)$$

Note that if f is conformal then $D_f = 1$.

Here is a geometric interpretation of the meaning of the dilatation : the differential $df(z_0)$ maps a circle in the tangent space $T_{z_0}\mathbb{C}$ into an ellipse in $T_{f(z_0)}\mathbb{C}$. The dilatation measures the distorsion since it is the ratio of the major semiaxis amd the minor semiaxis. Indeed applying (A1.8) to an infinitesimal circle $\Delta z = \varepsilon e^{i\theta}$ centered at z_0 one finds an infinitesimal ellipse centered at $f(z_0)$ with major semiaxis $[|f_z| - |f_{\bar z}|]^{-1}\varepsilon$ and minor semiaxis $[|f_z| + |f_{\bar z}|]^{-1}\varepsilon$.

The *maximal dilatation* of f on $\mathbb{C}$ is $D_f = \sup_{z\in\mathbb{C}} D_f(z)$. If $D_f < +\infty$, let $\kappa_f = (D_f - 1)/(D_f + 1)$. Then one has $\frac{|f_{\bar z}|}{|f_z|} \leq \kappa_f < 1$.

Definition A1.22 *f is* quasiconformal *if $D_f < +\infty$, i.e. $\kappa_f < 1$.*

Clearly if f is conformal then $D_f = 1$, $\kappa_f = 0$.

We want now to extend the notion of quasiconformal map to homeomorphisms. We will follow the geometric approach outlined in [Ah2].

Exercise A1.23 Given two rectangles R_1 and R_2 respectively with sides $a_1 \leq b_1$ and $a_2 \leq b_2$, show that there exists a conformal map of R_1 onto R_2 which maps vertices on vertices if and only if $\frac{a_1}{b_1} = \frac{a_2}{b_2}$.

Definition A1.24 *A* quadrilateral *$Q(z_1, z_2, z_3, z_4)$ is a Jordan domain in $\mathbb{C}$ with four distinguished boundary points z_1, z_2, z_3, z_4. Its* modulus *$M(Q)$ is the ratio a/b of the lengths $a < b$ of the sides of any rectangle R which is the conformal image of Q and whose vertices are image of the distinguished points.*

Note that the modulus of a quadrilateral is a conformal invariant. Thus one can use its variation under a homeomorphism to measure the lack of conformality of a map.

Definition A1.25 *Let $f : \mathbb{C} \to \mathbb{C}$ be an orientation–preserving homeomorphism. Its* maximal dilatation *is $D_f := \sup_{Q \subset \mathbb{C}, Q \text{ quadrilateral}} \frac{M(f(Q))}{M(Q)}$. Let $\kappa_f = (D_f - 1)/(D_f + 1)$. f is a* quasiconformal homeomorphism *if $D_f, +\infty$, i.e. $\kappa_f < 1$.*

Exercise A1.26 Prove that f is conformal if and only if $D_f = 1$.

Theorem A1.27 *A quasiconformal homeomorphism f with maximal dilatation D_f is almost everywhere differentiable and at each point z_0 where f is differentiable one has*

$$\frac{|f_{\bar{z}}(z_0)|}{|f_z(z_0)|} \leq \kappa_f \ .$$

Perhaps the most useful result in the theory of quasiconformal maps is the following existence theorem also known as Measurable Riemann Mapping Theorem [Ah2, p.98]

Theorem A1.28 *Let μ be a complex–valued measurable function with $\|\mu\|_\infty < 1$. There exists a quasiconformal mapping f such that*

$$f_{\bar{z}} = \mu(z) f_z \text{ almost everywhere} \tag{A1.11}$$

and f leaves the points $0, 1, \infty$ fixed.

The equation (A1.11) is also known as *Beltrami equation.*

Quasiconformal maps have been introduced in the subject of holomorphic dynamics by Dennis Sullivan and Adrien Douady and have rapidly become a standard tool. What we will need in Chapter 3 is the following

Theorem A1.29 (Douady–Hubbard : stability of the quadratic polynomial) *Let $P_\lambda(z) = \lambda\left(z - \frac{z^2}{2}\right)$ and let $F(z) = P_\lambda(z) + \psi(z)$ where ψ is holomorphic and bounded in the disk $\mathbb{D}_3$, $\psi(z) = \sum_{n=2}^{\infty} \psi_n z^n$ (i.e. $\psi(0) = \psi'(0) = 0$). Assume that $\sup_{z \in \mathbb{D}_3} |\psi(z)| < 10^{-2}$. Then there exists a quasiconformal homeomorphism*

h such that on the disk $\mathbb{D}_2$ one has $h^{-1}Fh = P_\lambda$. If ψ is small enough then h is near the identity in the $\mathcal{C}^0$ topology.

A2. Continued Fractions

In this appendix we recall some elementary facts on standard real continued fractions (we refer to [MMY], and references therein, for more general continued fractions).

We will consider the iteration of the Gauss map

$$A : (0,1) \mapsto [0,1] , \tag{A2.1}$$

defined by

$$A(x) = \frac{1}{x} - \left[\frac{1}{x} \right] . \tag{A2.2}$$

A is piecewise analytic with branches

$$A(x) = x^{-1} - n \text{ if } \frac{1}{n+1} < x \leq \frac{1}{n} , n \geq 1 .$$

Exercise A2.1 Prove that $A^*(\rho(x)dx) = \rho(x)dx$ where $\rho(x) = [(1+x)\log 2]^{-1}$, i.e. ρ is an invariant probability density for the Gauss map.

Let

$$G = \frac{\sqrt{5}+1}{2} , \; g = G^{-1} = \frac{\sqrt{5}-1}{2} .$$

To each $x \in \mathbb{R} \setminus \mathbb{Q}$ we associate a continued fraction expansion by iterating A as follows. Let

$$\begin{aligned} x_0 &= x - [x] , \\ a_0 &= [x] , \end{aligned} \tag{A2.3}$$

then one obviously has $x = a_0 + x_0$. We now define inductively for all $n \geq 0$

$$\begin{aligned} x_{n+1} &= A(x_n) , \\ a_{n+1} &= \left[\frac{1}{x_n} \right] \geq 1 , \end{aligned} \tag{A2.4}$$

thus

$$x_n^{-1} = a_{n+1} + x_{n+1} \; . \tag{A2.5}$$

Therefore we have

$$x = a_0 + x_0 = a_0 + \cfrac{1}{a_1 + x_1} = \ldots = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n + x_n}}} \; , \tag{A2.6}$$

and we will write

$$x = [a_0, a_1, \ldots, a_n, \ldots] \; . \tag{A2.7}$$

The nth-convergent is defined by

$$\frac{p_n}{q_n} = [a_0, a_1, \ldots, a_n] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n}}} \; . \tag{A2.8}$$

Exercise A2.2 Show that the numerators p_n and denominators q_n are recursively determined by

$$p_{-1} = q_{-2} = 1 \;\; , \quad p_{-2} = q_{-1} = 0 \;\; , \tag{A2.9}$$

and for all $n \geq 0$ one has

$$\begin{aligned} p_n &= a_n p_{n-1} + p_{n-2} \; , \\ q_n &= a_n q_{n-1} + q_{n-2} \; . \end{aligned} \tag{A2.10}$$

Exercise A2.3 Show that for all $n \geq 0$ one has

$$x = \frac{p_n + p_{n-1} x_n}{q_n + q_{n-1} x_n} \; , \tag{A2.11}$$

$$x_n = -\frac{q_n x - p_n}{q_{n-1} x - p_{n-1}} \; , \tag{A2.12}$$

$$q_n p_{n-1} - p_n q_{n-1} = (-1)^n \;\; . \tag{A2.13}$$

Note that $q_{n+1} > q_n > 0$ and that the sequence of the numerators p_n has the same constant sign of x. Equation (A2.13) implies also that for all $k \geq 0$ and for all $x \in \mathbb{R} \setminus \mathbb{Q}$ one has $\frac{p_{2k}}{q_{2k}} < x < \frac{p_{2k+1}}{q_{2k+1}}$.

Let

$$\beta_n = \Pi_{i=0}^n x_i = (-1)^n (q_n x - p_n) \quad \text{for } n \geq 0, \quad \text{and } \beta_{-1} = 1 \;\; . \tag{A2.14}$$

Then $x_n = \beta_n \beta_{n-1}^{-1}$ and $\beta_{n-2} = a_n \beta_{n-1} + \beta_n$.

Proposition A2.4 *For all* $x \in \mathbb{R} \setminus \mathbb{Q}$ *and for all* $n \geq 1$ *one has*

(i) $\quad |q_n x - p_n| = \dfrac{1}{q_{n+1} + q_n x_{n+1}}$, *so that* $\frac{1}{2} < \beta_n q_{n+1} < 1$;

(ii) $\beta_n \leq g^n$ *and* $q_n \geq \frac{1}{2} G^{n-1}$.

Proof. Using (A2.11) one has

$$|q_n x - p_n| = \left| q_n \frac{p_{n+1} + p_n x_{n+1}}{q_{n+1} + q_n x_{n+1}} - p_n \right| = \frac{|q_n p_{n+1} - p_n q_{n+1}|}{q_{n+1} + q_n x_{n+1}} = \frac{1}{q_{n+1} + q_n x_{n+1}}$$

by (A2.13). This proves (i).

Let us now consider $\beta_n = x_0 x_1 \ldots x_n$. If $x_k \geq g$ for some $k \in \{0, 1, \ldots, n-1\}$, then, letting $m = x_k^{-1} - x_{k+1} \geq 1$,

$$x_k x_{k+1} = 1 - m x_k \leq 1 - x_k \leq 1 - g = g^2 .$$

This proves (ii). □

Remark A2.5 Note that from *(ii)* it follows that $\sum_{k=0}^{\infty} \frac{\log q_k}{q_k}$ and $\sum_{k=0}^{\infty} \frac{1}{q_k}$ are always convergent and their sum is uniformly bounded.

For all integers $k \geq 1$, the iteration of the Gauss map k times leads to the following partition of $(0,1)$; $\sqcup_{a_1, \ldots, a_k} I(a_1, \ldots, a_k)$, where $a_i \in \mathbb{N}$, $i = 1, \ldots, k$, and

$$I(a_1, \ldots, a_k) = \begin{cases} \left(\frac{p_k}{q_k}, \frac{p_k + p_{k-1}}{q_k + q_{k-1}} \right) & \text{if } k \text{ is even} \\ \left(\frac{p_k + p_{k-1}}{q_k + q_{k-1}}, \frac{p_k}{q_k} \right) & \text{if } k \text{ is odd} \end{cases}$$

is the branch of A^k determined by the fact that all points $x \in I(a_1, \ldots, a_k)$ have the first $k+1$ partial quotients exactly equal to $\{0, a_1, \ldots, a_k\}$. Thus

$$I(a_1, \ldots, a_k) = \left\{ x \in (0,1) \mid x = \frac{p_k + p_{k-1} y}{q_k + q_{k-1} y} , \; y \in (0,1) \right\} .$$

Note that $\frac{dx}{dy} = \frac{(-1)^k}{(q_k + q_{k-1} y)^2}$ is positive (negative) if k is even (odd). It is immediate to check that any rational number $p/q \in (0,1)$, $(p,q) = 1$, is the endpoint of exactly two branches of the iterated Gauss map. Indeed p/q can be written as $p/q = [\bar{a}_1, \ldots, \bar{a}_k]$ with $k \geq 1$ and $\bar{a}_k \geq 2$ in a unique way and it is the left (right) endpoint of $I(\bar{a}_1,\ldots,\bar{a}_k)$ and the right (left) endpoint of $I(\bar{a}_1,\ldots,\bar{a}_k-1,1)$ if k is even (odd).

A3. Distributions, Hyperfunctions, Formal Series. Hypoellipticity and Diophantine Conditions.

A3.1 We follow here [H1], Chapter 9 but we also recommend [Ph], especially the first few chapters, for a nice introduction to hyperfunctions and their applications.

Let K be a non empty compact subset of $\mathbb{R}$. A *hyperfunction with support in* K is a linear functional u on the space $\mathcal{O}(K)$ of functions analytic in a neighborhood of K such that for all neighborhood V of K there is a constant $C_V>0$ such that

$$|u(\varphi)|\le C_V\sup_V|\varphi|\ ,\quad \forall\varphi\in\mathcal{O}(V)\ .$$

We denote by $A'(K)$ the space of hyperfunctions with support in K. It is a Fréchet space : a seminorm is associated to each neighborhood V of K.

Let $\mathcal{O}^1(\overline{\mathbb{C}}\setminus K)$ denote the complex vector space of functions holomorphic on $(\overline{\mathbb{C}}\setminus K)$ and vanishing at infinity. One has the following

Proposition A3.1 *The spaces $A'(K)$ and $\mathcal{O}^1(\overline{\mathbb{C}}\setminus K)$ are canonically isomorphic. To each $u\in A'(K)$ corresponds $\varphi\in\mathcal{O}^1(\overline{\mathbb{C}}\setminus K)$ given by*

$$\varphi(z)=u(c_z)\ ,\ \forall z\in\mathbb{C}\setminus K\ ,$$

where $c_z(x)=\frac{1}{\pi}\frac{1}{x-z}$. Conversely to each $\varphi\in\mathcal{O}^1(\overline{\mathbb{C}}\setminus K)$ corresponds the hyperfunction

$$u(\psi)=\frac{i}{2\pi}\int_\gamma\varphi(z)\psi(z)dz\ ,\ \forall\psi\in A$$

where γ is any piecewise $\mathcal{C}^1$ path winding around K in the positive direction. We will also use the notation

$$u(x)=\frac{1}{2i}[\varphi(x+i0)-\varphi(x-i0)]$$

for short.

Proof. It is very easy : note that the function $x\mapsto c_z(x)$ is analytic in a neighborhood of K for all $z\notin K$. Then it is immediate to check applying Cauchy's formula that these two correspondences are surjective and are the inverse of one another. □

A3.2 Let $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z} \subset \mathbb{C}/\mathbb{Z}$. A *hyperfunction on* $\mathbb{T}$ is a linear funtional U on the space $\mathcal{O}(\mathbb{T}^1)$ of functions analytic in a complex neighborhood of $\mathbb{T}^1$ such that for all neighborhood V of $\mathbb{T}$ there exists $C_V > 0$ such that

$$|U(\Phi)| \leq C_V \sup_V |\varphi| \ , \quad \forall \Phi \in \mathcal{O}(V) \ .$$

We will denote $A'(\mathbb{T}^1)$. the Fréchet space of hyperfunctions with support in $\mathbb{T}$. For $U \in A'(\mathbb{T})$, let $\hat{U}(n) := U(e_{-n})$ with $e_n(z) = e^{2\pi i n z}$.

Exercise A3.2 Show that the doubly infinite sequence $(\hat{U}(n))_{n\in\mathbb{Z}}$ satisfies

$$|\hat{U}(n)| < C_\varepsilon e^{2\pi|n|\varepsilon} \ .$$

for all $\varepsilon > 0$ and for all $n \in \mathbb{Z}$ with a suitably chosen $C_\varepsilon > 0$. Conversely show that any such sequence is the Fourier expansion of a unique hyperfunction with support in $\mathbb{T}$.

Let $\mathcal{O}_\Sigma$ denote the complex vector space of holomorphic functions $\Phi : \mathbb{C} \setminus \mathbb{R} \to \mathbb{C}$, 1–periodic, bounded at $\pm i\infty$ and such that $\Phi(\pm i\infty) := \lim_{\Im m\, z \to \pm\infty} \Phi(z)$ exist and verify $\Phi(+i\infty) = -\Phi(-i\infty)$.

Exercise A3.3 Show that the spaces $A'(\mathbb{T}^1)$ and $\mathcal{O}_\Sigma$ are canonically isomorphic. Indeed to each $U \in A'(\mathbb{T}^1)$ corresponds $\Phi \in \mathcal{O}_\Sigma$ given by

$$\Phi(z) = U(C_z) \ , \ \forall z \in \mathbb{C} \setminus K \ ,$$

where $C_z(x) = \operatorname{cotg} \pi(x - z)$. Conversely to each $\Phi \in \mathcal{O}_\Sigma$ corresponds the hyperfunction

$$U(\Psi) = \frac{i}{2} \int_\Gamma \Phi(z)\Psi(z)dz \ , \ \forall \Psi \in A(\mathbb{T}^1)$$

where Γ is any piecewise $\mathcal{C}^1$ path winding around a closed interval $I \subset \mathbb{R}$ of length 1 in the positive direction. We will also use the notation

$$U(x) = \frac{1}{2i}[\Phi(x + i0) - \Phi(x - i0)]$$

for short.

The nice fact is that the following diagram commutes :

$$
\begin{array}{ccc}
A'([0,1]) & \longrightarrow & \mathcal{O}^1(\overline{\mathbb{C}} \setminus [0,1]) \\
{\scriptstyle \sum_z} \Big\downarrow & & \Big\downarrow {\scriptstyle \sum_z} \\
A'(\mathbb{T}^1) & \longrightarrow & \mathcal{O}_\Sigma
\end{array}
$$

the horizontal lines are the above mentioned isomorphisms, $\sum_Z$ is the sum over integer translates : $(\sum_{\mathbb{Z}} \varphi)(z) = \sum_{n\in\mathbb{Z}} \varphi(z-n)$.

A3.3 As we have seen in A3.2 periodic distributions and hyperfunctions are naturally identified with the two following subspaces of the complex vector space of *formal* Fourier series

$$
\begin{aligned}
\varphi \in \mathcal{D}'(\mathbb{T}) \Leftrightarrow \varphi(\theta) = \sum_{-\infty}^{+\infty} \hat{\varphi}(n) e^{2\pi i n\theta} & \text{ and there exists} M > 0\,,\ r > 0 \\
& \text{such that } |\hat{\varphi}(n)| \le M|n|^{+r}\ \forall n \in \mathbb{Z}^*\ , \\
\varphi \in \mathcal{A}'(\mathbb{T}) \Leftrightarrow \varphi(\theta) = \sum_{-\infty}^{+\infty} \hat{\varphi}(n) e^{2\pi i n\theta} & \text{ and for all}\, \varepsilon > 0 \text{ there exists} C_\varepsilon > 0 \\
& \text{such that } |\hat{\varphi}(n)| \le C_\varepsilon \exp(2\pi|n|\varepsilon)\ \forall n \in \mathbb{Z}\ .
\end{aligned}
$$

Let us now consider the following linear first–order difference equation on $\mathbb{T}^1$

$$
f_g(\theta+\alpha) - f_g(\theta) = g(\theta)
$$

where $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. A necessary condition for the existence of a solution is that $\int_0^{2\pi} g(\theta) d\theta = 0$. Thus we introduce the zero–mean Dirac delta function on $\mathbb{T}^1$

$$
\delta_{\mathbb{T},0}(\theta) = \sum_{n\in\mathbb{Z}\,, n\neq 0} e^{2\pi i n\theta}
$$

and we note that the corresponding f_δ plays the role of a fundamental solution since

$$f_g = f_\delta \odot g = \sum_{n=-\infty}^{+\infty} \hat{f}_\delta(n)\hat{g}(n)e^{2\pi i n\theta} = \frac{1}{2\pi}\int_0^{2\pi} f_\delta(\theta-\theta_1)g(\theta_1)d\theta_1$$

when the integral makes sense.

On the other hand one clearly has

$$f_\delta(\theta) = \sum_{n\neq 0} \frac{e^{2\pi i n\theta}}{e^{2\pi i n\alpha}-1}$$

as a formal power series. We have the following elementary Proposition which can also be taken as an equivalent definition of diophantine numbers

Proposition A3.4 *f_δ is a distribution if and only if $\alpha \in CD$. f_δ is a hyperfunction if and only if the denominators q_n of the convergents of α verify* $\lim_{n\to+\infty} \frac{log q_{n+1}}{q_n} = 0$.

The proof is immediate and it is left as an Exercise.

Note that the above discussion carries over easily to the linear PDE on the two–dimensional torus $\mathbb{T}^2$

$$(\partial_{\theta_1} + \alpha\partial_{\theta_2})f = g$$

(which is associated to the linear flow $\dot{\theta_1} = 1$, $\dot{\theta_2} = \alpha$). In this case one has $\delta_{\mathbb{T}^2,0} = \sum_{n\in\mathbb{Z}^2, n\neq 0} e^{2\pi i(n_1\theta_1+n_2\theta_2)}$ and the fundamental solution is $f_\delta = \sum_{n\in\mathbb{Z}^2, n\neq 0} \frac{e^{2\pi i(n_1\theta_1+n_2\theta_2)}}{2\pi i(n_1+n_2\alpha)}$. Then Proposition A3.4 holds also in this case, showing that the operator $\partial_{\theta_1} + \alpha\partial_{\theta_2}$ is hypoelliptic if and only if α is diophantine.

References

[AG] S. Alinhac, P. Gérard "Opérateurs pseudo–différentiels et thèoréme de Nash–Moser" Savoirs Actuels, CNRS Editions (1991)

[Ah1] L.V. Ahlfors "Conformal Invariants : Topics in Geometric Function Theory" McGraw–Hill (1973)

[Ah2] L.V. Ahlfors "Lectures on Quasiconformal Mappings" Van Nostrand (1966)

[AM] R. Abraham, J. Marsden "Foundations of Mechanics" Benjamin Cummings, New York (1978)

[Ar1] V. I. Arnol'd "Small denominators and problems of stability of motion in classical celestial mechanics" Russ. Math. Surv. **18** (1963), 85–193.

[Ar2] V. I. Arnol'd "Instability of dynamical systems with several degrees of freedom" Sov. Math. Dokl. **5** (1964), 581–585.

[Ar3] V. I. Arnol'd "Geometrical Methods in the Theory of Ordinary Differential Equations" Springer–Verlag (1983)

[AKN] V. I. Arnol'd, V. V. Kozlov and A. I. Neishtadt "Dynamical Systems III", Springer–Verlag (1988).

[Be] A. Beardon "Iteration of Rational Functions" Springer–Verlag (1991)

[BFGG] G. Benettin, G. Ferrari, L. Galgani and A. Giorgilli "An Extension of the Poincaré–Fermi Theorem on the Nonexistence of Invariant Manifolds in Nearly Integrable Hamiltonian Systems" Il Nuovo Cimento **72B** (1982) 137

[BHS] H. W. Broer, G. B. Huitema and M. B. Sevryuk "Quasi–Periodic Motions in Families of Dynamical Systems" Springer–Verlag (1996)

[Bo] J. B. Bost "Tores invariants des systèmes dynamiques hamiltoniens" Séminaire Bourbaki **639**, Astérisque **133–134** (1986), 113–157.

[Br] A. D. Brjuno "Analytical form of differential equations" *Trans. Moscow Math. Soc.* **25** (1971), 131-288 ; **26** (1972), 199-239.

[CG] L. Carleson and T. Gamelin "Complex Dynamics" Universitext, Springer–Verlag, Berlin Heidelber New York (1993)

[CM] T. Carletti and S. Marmi "Linearization of Analytic and Non–Analytic Germs of Diffeomorphisms of $(\mathbb{C}, 0)$" Bulletin de la Societé Mathématique de France (1999)

[Da] A.M. Davie "The critical function for the semistandard map" *Nonlinearity* **7** (1994), 219 - 229.

[DeB] L. de Branges "A proof of the Bieberbach conjecture" *Acta Math.* **154** (1985), 137–152.

[Di] J. Dieudonné "Calcul Infinitésimal" Hermann, Paris (1980)

[Do] A. Douady "Disques de Siegel et anneaux de Herman" Séminaire Bourbaki n. 677, *Astérisque* **152–153** (1987), 151–172

[Fa] K. Falconer "Fractal Geometry. Mathematical Foundations and Applications" John Wiley and Sons (1990)

[Ga] G. Gallavotti "Quasi–Integrable Mechanical Systems" in Phénomènes critiques, systèmes aléatoires, théories de jauge, Part I, II, Les Houches 1984, North–Holland, Amsterdam (1986) 539–624

[GM] A. Giorgilli and A. Morbidelli "Invariant KAM tori and global stability for Hamiltonian systems" ZAMP **48** (1997), 102–134.

[Gr] M.L. Gromov "Smoothing and Inversion of Differential Operators" *Math. USSR Sbornik* **17** (1972), 381–434

[Ha] R.S. Hamilton "The Inverse Function Theorem of Nash and Moser" *Bull. A.M.S.* **7** (1982), 65–222.

[H1] L. Hörmander "The Analysis of Linear Partial Differential Operators I" Grundlehren der mathematischen Wissenschaften **256**, Springer–Verlag, Berlin, Heidelberg, New York, Tokyo (1983)

[H2] L. Hörmander "The boundary problem of physical geodesy" *Arch. Rat. Mech. Anal.* **62** (1976), 1–52

[He1] M. R. Herman "Examples de fractions rationelles ayant une orbite dense sur la sphere de Riemann" Bulletin de la Societé Mathématique de France **112** (1984), 93–142

[He2] M. R. Herman "Simple proofs of local conjugacy theorems for diffeomorphisms of the circle with almost every rotation numbers" Bull. Soc. Bras. Mat. **16** (1985) 45–83

[He3] M. R. Herman "Are there critical points one the boundary of singular domains ?" Commun. Math. Phys. **99** (1985) 593–612.

[He4] M. R. Herman "Recent results and some open questions on Siegel's linearization theorem of germs of complex analytic diffeomorphisms of $\mathbf{C}^n$ near a fixed point" Proc. VIII Int. Conf. Math. Phys. Mebkhout and Seneor eds. (Singapore : World Scientific) (1986), 138–184.

[He5] M. R. Herman "Démonstration du théorème des courbes translatées par difféomorphismes de l'anneau ; démonstration du théorème des tores invariants" manuscripts (1980) and "Abstract methods in small divisors : implicit function theorems in Fréchet spaces", lectures given at the CIME conference on Dynamical Systems and Small Divisors, Cetraro 1998

[HL] G. H. Hardy and J. E. Littlewood "Notes on the theory of series (XXIV) : a curious power series" Proc. Cambridge Phil. Soc. **42** (1946), 85–90

[HW] G.H. Hardy and E.M. Wright "An introduction to the theory of numbers" Fifth Edition, Oxford Science Publications (1990)

[K] A. N. Kolmogorov "On the persistence of conditionally periodic motions under a small perturbation of the Hamilton function" Dokl. Akad. Nauk SSSR **98** (1954) 527–530 (in Russian : English translation in G. Casati and J. Ford, editors, Stochastic Behavior in Classical and Quantum Hamiltonian Systems, Lecture Notes in Physics **93** (1979) 51–56 Springer–Verlag).

[L] S. Lang "Introduction to Diophantine Approximation" Addison–Wesley (1966)

[Lo] P. Lochak "Canonical perturbation theory via simultaneous approximations", Russ. Math. Surv. **47** (1992), 57–133.

[Ma1] S. Marmi "Critical Functions for Complex Analytic Maps" *J. Phys. A : Math. Gen.* **23** (1990), 3447–3474

[Ma2] S. Marmi "Chaotic Behaviour in the Solar System (Following J. Laskar)" Séminaire Bourbaki n. 854, November 1998, to appear in Astérisque

[Me] Y. Meyer "Algebraic Numbers and Harmonic Analysis" North–Holland Mathematical Library **2** (1972)

[MM] L. Markus and K. R. Meyer "Generic Hamiltonian Systems are neither integrable nor ergodic" Memoirs of the A.M.S. **144** (1974)

[MMY] S. Marmi, P. Moussa and J.–C. Yoccoz "The Brjuno functions and their regularity properties" *Commun. Math. Phys.* **186** (1997), 265-293

[MMY2] S. Marmi, P. Moussa and J.–C. Yoccoz "Complex Brjuno Functions" preprint SPhT Saclay, France, 71 pages (1999)

[Mc] C. T. McMullen "Complex Dynamics and Renormalization" Ann. of Math. Studies, Princeton University Press (1994)

[Mn] R. Mañe "Ergodic Theory and Differentiable Dynamics" Springer–Verlag (1987)

[M] J. Moser "A rapidly convergent iteration method and nonlinear differential equations" *Ann. Scuola Norm. Sup. Pisa* **20** (1966) 499-535

[N] J. Nash "The embedding problem for Riemannian manifolds" *Ann. of Math.* **63** (1956) 20–63

[Ne] N.N. Nekhoroshev "An exponential estimate for the time of stability of nearly integrable Hamiltonian systems" *Russ. Math. Surveys* **32** (1977), 1–65.

[Ni] L. Niremberg “An abstract form of the non–linear Cauchy–Kowalewskaya theorem” *J. Diff. Geom.* **6** (1972) 561–576

[PM1] R. Pérez–Marco “Solution complète au problème de Siegel de linéarisation d’une application holomorphe au voisinage d’un point fixe (d’après J.–C. Yoccoz)” Séminaire Bourbaki n. 753, *Astérisque* **206** (1992), 273–310

[Ph] F. Pham (Editor) “Hyperfunctions and Theoretical Physics” Lecture Notes in Mathematics **449** Springer–Verlag (1975)

[P] H. Poincaré “Les Méthodes Nouvelles de la Mécanique Celeste”, tomes I–III Paris Gauthier–Villars (1892, 1893, 1899).

[Po] Ch. Pommerenke “Boundary Behaviour of Conformal Maps” Grundlehren der Mathematischent Wissenschaften **299**, Springer–Verlag (1992)

[Pö] J. Pöschel “Integrability of Hamiltonian systems on Cantor sets” Comm. Pure Appl. Math. **35** 653–696 (1982)

[Re] R. Remmert “Classical Topics in Complex Function Theory” Graduate Texts in Mathematics **172**, Springer–Verlag (1998)

[Rü] H. Rüssmann “Kleine Nenner II : Bemerkungen zur Newtonschen Methode” Nachr. Akad. Wiss. Göttingen Math. Phys. Kl (1972) 1–20

[S] C.L. Siegel “Iteration of analytic functions” Annals of Mathematics **43** (1942), 807-812.

[Sch1] W.M. Schmidt “Diophantine Approximation” Lecture Notes in Mathematics, **785**, Springer–Verlag (1980)

[Sch2] W.M. Schmidt “Diophantine Approximations and Diophantine Equations” Lecture Notes in Mathematics, **1467**, Springer–Verlag (1991)

[Ser] F. Sergeraert “Un théorème de fonctions implicites sur certains espaces de Fréchet et quelques applications” *Ann. Scient. Èc. Norm. Sup.* **5** (1972), 599–660.

[ST] J. Silverman and J. Tate “Rational Points on Elliptic Curves” Undergraduate Texts in Mathematics, Springer–Verlag (1992)

[SZ] D. Salomon and E. Zehnder “KAM theory in confuguration space” *Comm. Math. Helvetici* **64**, (1989), 84–132

[St] S. Sternberg “Celestial Mechanics” (two volumes) W.A. Benjamin, New York (1969).

[Va] F.H. Vasilescu “Analytic Functional Calculus” D. Reidel Publ. Co. (1982)

[Yo1] J.–C. Yoccoz “An introduction to small divisors problems” in “From number theory to physics”, M. Waldschmidt, P. Moussa, J.M. Luck and C. Itzykson (editors) Springer–Verlag (1992), 659–679

[Yo2] J.–C. Yoccoz "Théorème de Siegel, nombres de Bruno et polynômes quadratiques" *Astérisque* **231** (1995), 3-88.

[Yo3] J.–C. Yoccoz, lectures given at the CIME conference on Dynamical Systems and Small Divisors, Cetraro 1998, to appear in Lecture Notes in Mathematics

[Ze1] E. Zehnder "Generalized Implicit Function Theorems with Applications to some Small Divisor Problems (I and II)" *Commun. Pure Appl. Math.* **28** (1975) 91–140, **29** (1976) 49–113.

[Ze2] E. Zehnder "A simple proof of a generalization of a Theorem by C. L. Siegel" Lecture Notes in Mathematics **597** (1977) 855–866.

Analytical index

List of symbols

Ad : adjoint action
$c(\Omega, z_0)$: conformal capacity of Ω w.r.t. z_0
$\mathbb{C}$: complex plane
$\mathbb{C}^*$: $\mathbb{C} \setminus \{0\}$
$\overline{\mathbb{C}}$: Riemann sphere
$\mathbb{C}\{z\}$: ring of convergent power series in one complex variable
$\mathbb{C}[[z]]$: ring of formal power series in one complex variable
Cent : centralizer
$\widehat{\mathrm{Cent}}$: formal centralizer
$D_f(z_0)$, D_f : dilatiation of f (maximal)
$\mathbb{D}$: open unit disk $\{|z| < 1\}$
$\mathbb{D}_r$: open disk $\{|z| < r\}$ of radius $r > 0$.
$\mathbb{E}$: outer disk $\{|z| > 1\}$
$[f]$: orbit of f
$F(R)$: Fatou set of R
G : group of germs of holomorphic diffeomorphisms of $(\mathbb{C}, 0)$
$\hat{G}$: formal analogue of G
G_λ : elements of G with linear part λ
$\hat{G}_\lambda$: formal analogue of G_λ
$J(R)$: Julia set of R
$\mathbb{N}$: non–negative integers
Ω : a region of $\mathbb{C}$
$\mathbb{Q}$: rational integers
R_λ : the germ $R_\lambda(z) = \lambda z$
S : univalent maps on $\mathbb{D}$
S_λ : elements of S with linear part R_λ
$S_{\mathbb{S}^1}$: elements of S with linear part of unit modulus
$\mathbb{S}^1$: unit circle $\{|z| = 1\}$
u Yoccoz's function, see Section 3.1
$\mathcal{Y}$: see Chapter 4
$\mathbb{Z}$: integers

Stefano Marmi, Dipartimento di Matematica e Informatica, Università di Udine,
Via delle Scienze 206, Località Rizzi, 33100 Udine, Italy ; marmi@dimi.uniud.it