

# The differentiability of the conjugation of certain diffeomorphisms of the circle

Y. KATZNELSON† AND D. ORNSTEIN‡

*Mathematics Department, Stanford University, Stanford CA 94305, USA*

(Received 22 June 1987 and revised 27 August 1988)

*Abstract.* Our purpose in this paper is to present a more or less complete solution to the problem of the smoothness of the conjugation of aperiodic diffeomorphisms of the circle. We show that the rotation number and the smoothness of the diffeomorphism guarantee a certain smoothness for the homeomorphism which conjugates it with a rigid rotation, and obtain the best smoothness that can be guaranteed.

## 0. Introduction

We study orientation preserving aperiodic diffeomorphisms  $f$  of the circle. Poincaré (1885) noticed that the orbit structure of such  $f$  is determined by some irrational (mod 1), called ‘the rotation number’ of  $f$  and denoted  $\alpha = \rho(f)$ , in the following sense: for any point  $t \in \mathbf{T} = \mathbf{R}/\mathbf{Z}$ , the mapping  $f^j(t) \rightarrow j\alpha \pmod{1}$ ,  $j \in \mathbf{Z}$ , is order preserving (properly speaking, orientation preserving). Some fifty years later, Denjoy proved that if  $\log Df$  is of bounded variation, and in particular, if  $f \in \mathcal{H}^2$  (i.e. is  $C^2$ -diffeomorphism, see next section for notation) then the orbits  $\{f^j(t)\}_{j \in \mathbf{Z}}$  are dense and the mapping  $f^j(t) \rightarrow j\alpha \pmod{1}$  can therefore be extended by continuity to a homeomorphism  $h$  of  $\mathbf{T}$ , which conjugates  $f$  to the rigid rotation  $R_\alpha : t \rightarrow t + \alpha$ , that is, such that  $f = h^{-1}R_\alpha h$ .

The natural problem now was to obtain smoothness information for  $h$ ; more precisely to relate the smoothness of  $h$  to that of  $f$  and to the arithmetical properties of the rotation number  $\alpha$ . Knowing  $\alpha$  and assuming that  $f = h^{-1}R_\alpha h$  is  $C^k$ ,  $C^\infty$ , analytic etc, what can be said about the smoothness of  $h$ ?

To get an idea of what one may expect, one can consider the ‘corresponding’ linear difference equation

$$\varphi(t) = \eta(t + \alpha) - \eta(t)$$

where one gets a fairly precise idea almost trivially by writing the corresponding Fourier coefficient conditions

$$\hat{\varphi}(0) = 0, \quad \hat{\varphi}(n) = \hat{\eta}(n)(e^{2\pi i n \alpha} - 1)$$

† Research partially supported by NSF Grant No. DMS86-05098.

‡  $Df$  denotes the derivative of  $f$ .

§ We shall use the convention that for mappings iteration is written as product; we expand and write  $f \circ g$  only when we deem it essential for clarity.

or  $\hat{\eta}(n) = \hat{\varphi}(n)(e^{\pi i n \alpha} - 1)^{-1}$ . Smoothness of  $\varphi$  corresponds to a rate at which  $\hat{\varphi}(n)$  goes to zero as  $|n| \rightarrow \infty$ ; the ‘small divisors’  $(e^{2\pi i n \alpha} - 1)$  determine how much bigger than  $\hat{\varphi}(n)$  can  $\hat{\eta}(n)$  be, and that implies a degree of smoothness for  $\eta(t)$ . Thus if  $\alpha$  is such that, for some constant  $\beta$ ,  $|e^{2\pi i n \alpha} - 1| > c|n|^{-\beta-1}$ , ( $\alpha$  is then said to ‘satisfy a diophantine condition’, or to be ‘Diophantine’ or ‘Non-Liouville’) and does not satisfy the analogous condition for any  $\beta_0 < \beta$ , then  $\eta$  may have  $\beta + 1 + \varepsilon$  fewer derivatives than  $\varphi$ . (The only slightly delicate point in the harmonic analysis here comes from the fact that multiplying the Fourier coefficients by  $c_n \leq |n|^\gamma$  may result in a loss of  $\gamma + \frac{1}{2}$  derivatives, because of the difference between continuity and absolute convergence of the Fourier series; the special form of  $(e^{2\pi i n \alpha} - 1)^{-1}$  saves the extra ‘half a derivative’. Note also that if  $\hat{\varphi}(n)$  are small on the set of values of  $n$  for which  $n\alpha$  is close to one the loss of differentiability may be smaller.) On the other hand, if  $\alpha$  is a Liouville number, one may have an infinite loss of smoothness and even loss of analyticity (i.e.  $\varphi$  may be analytic and  $\eta$  not even absolutely continuous) depending on the rate at which  $n\alpha \pmod{1}$  may become small.

Arnold [A] (for  $f$  analytic) and Moser [M] (for  $f \in \mathcal{H}^k$ ) developed perturbation techniques that yielded analyticity resp. smoothness of  $h$  in the *local situation*, i.e. when one assumes  $f$  to be a perturbation of  $R_\alpha$  in an appropriate function space. In that case Moser showed that the loss of differentiability is no more than the loss for the linear difference equation above. In fact the difference equation above is the ‘linearized’ part of the perturbation equation  $(Id + \eta) \circ R_\alpha \circ (Id + \eta)^{-1} = (R_\alpha + d)$ , and has to be solved repeatedly in the so-called implicit function theorem technique. Herman [H] and Lazutkin [LA] showed that the loss of differentiability can be as bad as that of the difference equation.

The non-local case remained open until Herman [H] proved that for  $\alpha$  in some set  $A$  of full measure and  $f \in \mathcal{H}^k$  ( $k \geq 3$ ),  $h$  is in fact  $k - 1 - \varepsilon$  times differentiable, and is analytic if  $f$  is analytic. The elements of the set  $A$  all satisfy the diophantine conditions  $|e^{2\pi i n \alpha} - 1| > C_\varepsilon |n|^{-1-\varepsilon}$  for any  $\varepsilon > 0$ . Yoccoz [Y] extended Herman’s methods to obtain results for all Non-Liouville numbers with the loss of differentiability as one would expect from the linear difference equation, but only if one assumes substantially more differentiability than is needed for  $C^1$ -conjugation. Herman’s breakthrough depended crucially on an inequality obtained cleverly from properties of the Schwarzian derivative and the same basic estimates are also essential in Yoccoz’ work. The unfortunate aspect of this is that it only applies if  $f$  is sufficiently smooth; thus Herman needs to assume  $f \in \mathcal{H}^3$  even if  $\alpha$  entails the loss of a single (or  $1 + \varepsilon$ ) derivative, and cannot conclude  $h \in \mathcal{H}^1$  from  $f \in \mathcal{H}^{2+\varepsilon}$ . The same problem (due to the same technical reason) prevents Yoccoz from obtaining smoothness of  $h$  unless  $f \in \mathcal{H}^k$ ,  $k \geq k_0$ ,  $k_0$  being an integer satisfying  $k_0 > 2\beta + 1$ ,  $k_0 \geq 3$  (for that degree of smoothness he obtains the right result, namely  $h \in \mathcal{H}^{k-1-\beta-\varepsilon}$ ,  $\forall \varepsilon > 0$ ; no information, however, for  $\beta + 2 < k \leq \max(2\beta + 1, 3)$  is obtained).

In the opposite direction, easy constructions along the lines of [K] show that the expected differentiability loss does in fact happen. (Cf. Appendix 3 or [H], see also [LA] and [HS]. Lazutkin’s paper appeared in 1977, [LA] is the English translation. We thank the referee for pointing out this paper to us.)

What we propose to do in this paper, is to complete the picture obtained by Herman and Yoccoz when the rotation number is Diophantine (= Non Liouville), and to show that indeed one has the same loss of differentiability as in the linear case. Our main result, Theorem 4.\* (or 4.12), is the following (the diophantine condition is stated in § 4 in terms of the convergents of the continued fraction expansion of  $\alpha$ , equivalent to our statement here):

**THEOREM.** *Assume that  $f \in \mathcal{H}^k$  and the rotation number  $\alpha = \rho(f)$  satisfies the diophantine condition  $|e^{2\pi i n \alpha} - 1| > c|n|^{-\beta-1}$ , with  $\beta + 2 < k$ , then the homeomorphism  $h$  which conjugates  $f$  with  $R_\alpha$  is in  $\mathcal{H}^{k-1-\beta-\epsilon}$  for all  $\epsilon > 0$ .*

We do not use the Schwarzian derivative and obtain the necessary estimates by repeatedly using one basic idea which has its roots in Denjoy’s work with variants developed and used by Herman and by Carleson [C]. (In fact the same basic idea appears in many estimates of the growth, or decay, of cocycles in other dynamical systems, e.g. geodesic flow cf, Livshitz [LI]).

Much of the conceptual contents of the paper, including the complete proof of the  $\mathcal{H}^1$ -conjugation for almost all  $\alpha$  and  $f \in \mathcal{H}^{2+\epsilon}$ , is contained in the first three sections and the less than totally committed reader need not lose heart by the heavier technical aspect of § 4.

Preliminary versions of the paper, some containing the results of the first three sections, some containing all of it (as well as what became [KO]) have been in circulation for the past three years and several friends, including the referee, sent us comments which helped us with the preparation of the present version. We wish to thank them for their help. In the meantime a different approach to the problem [KS] also appeared.

**1. Background and notation**

We denote by  $\mathcal{H}$  the groups of orientation preserving homeomorphisms of the circle  $\mathbf{T} = \mathbf{R}/\mathbf{Z}$ , and by  $\mathcal{H}^k$ ,  $k \geq 1$ , the subgroup of  $C^k$ -diffeomorphisms. A conjugation  $f = h^{-1}gh$ ,  $f, g, h \in \mathcal{H}$  means, in terms of dynamics, that  $h$  maps  $f$ -orbits onto  $g$ -orbits, and since  $h$  is orientation preserving, the ‘order’ on the circle of an  $f$ -orbit has to agree with the ‘order’ of the corresponding  $g$ -orbit. The preceding sentence can be taken as a definition of ‘order’ on  $\mathbf{T}$ , namely: two doubly infinite sequences have the same ‘order’ if one can be mapped onto the other by an orientation preserving map. The key fact for orientation preserving homeomorphisms was observed and proved by Poincaré in 1885:

**THEOREM 1.1.** *There is a mapping  $\rho: \mathcal{H} \rightarrow \mathbf{T}$ , called the rotation number having the following properties:*

- (a)  $\rho(f) = 0$  if and only if  $f$  has a fixed point.
- (b) for any integer  $k$ ,  $\rho(f^k) = k\rho(f)$ . In particular,  $\rho(f)$  is of finite order (rational (mod 1)) if and only if  $f$  admits periodic orbits.
- (c) If  $f$  is aperiodic and hence by (b),  $\alpha = \rho(f)$  is of infinite order (irrational (mod 1)), then every  $f$  orbit has the same ‘order’ on  $\mathbf{T}$  as  $\{n\alpha \pmod{1}\}$ .

We omit the explicit reference to (mod 1) and often refer to  $\rho(f)$  as a number in  $[0, 1)$ ; in particular we discuss the continued fraction expansion of  $\rho(f)$  (see Appendix 1).

We denote by  $R_\alpha$  the rigid rotation on  $\mathbf{T}$  by angle  $\alpha$ :  $R_\alpha(t) = t + \alpha$ . Poincaré’s theorem mentioned above says that if  $f$  is conjugate to  $R_\alpha$ , then  $\alpha = \rho(f)$ . If  $\rho(f)$  is rational,  $f$  is typically *not* conjugate to  $R_{\rho(f)}$ . If  $\alpha = \rho(f)$  is irrational, then the orbit ‘order’ is consistent with conjugation but another condition has to be met as well: all the orbits of  $R_\alpha$  are dense and that property is again maintained by conjugation. If  $f$  has one dense orbit  $\{f^j(t_0)\}$ , we can define

$$h(f^j(t_0)) = j\alpha \pmod{1} \tag{1.1}$$

and extend  $h$  by continuity (it is orientation preserving from a sense subset onto a dense subset) to obtain  $h \in \mathcal{H}$  and  $f = h^{-1}R_\alpha h$ .

We obtained

**THEOREM 1.2.** *If  $\alpha = \rho(f)$  is irrational, then  $f$  is conjugate to  $R_\alpha$  if and only if some (all)  $f$ -orbits are dense.*

Unless specified otherwise, we shall assume henceforth  $\alpha = \rho(f)$  is irrational, if the orbits of  $f$  are not dense, we obtain, by taking the complement of the closure of an  $f$ -orbit, and  $f$ -invariant open set with infinitely many component intervals which are permuted by  $f$ . If we denote one of these intervals by  $I_0$  and  $f^j(I_0)$  by  $I_j$ , then  $\{I_j\}$ ,  $j \in \mathbf{Z}$ , are disjoint and the sum of their lengths  $|I_j|$  is therefore bounded by 1. If we assume that  $f \in \mathcal{H}^1$ , and write  $\|Df^j\|_\infty = \sup |Df^j|$  then  $|I_0| \leq \|Df^{-j}\|_\infty |I_j|$  and  $|I_0| \sum_j \|Df^j\|_\infty^{-1} \leq 1$ . We state this observation as

**THEOREM 1.3.** *If  $f \in \mathcal{H}^1$  and  $\sum_j \|Df^j\|_\infty^{-1} = \infty$ , then  $f$  is conjugate to  $R_\alpha$ .*

The divergence condition in Theorem 1.3 is satisfied in particular if the general term does not go to zero, that is, if  $f$  has an infinite collection of iterates which are uniformly bounded in  $\mathcal{H}^1$  (more generally, equicontinuous). Denjoy (1932) proved that this is always the case if  $f \in \mathcal{H}^1$  and  $Df$  is of bounded variation (see [H] for historical discussion). Denjoy’s inequality, which we state below, is based on some simple observations:

(a) The chain rule  $Df^q(t) = \prod_{j=0}^{q-1} Df(f^j(t))$  which we write ‘additively’ as

$$\log Df^q(t) = \sum_{j=0}^{q-1} \log Df(f^j(t)) \tag{1.2}$$

(another terminology:  $\log Df^q(t)$  is an additive 1-cocycle).

(b) Since  $f^q$  maps the circle onto itself, its derivative cannot be always bigger than 1, nor always smaller than 1. Thus there exists points  $t_q$  such that  $Df^q(t_q) = 1$  or

$$\log Df^q(t_q) = 0 \tag{1.3}$$

(c) If  $t, \tau \in \mathbf{T}$  and the sets  $\{f^j(t)\}_{j=0}^{q-1}$  and  $\{f^k(\tau)\}_{k=0}^{q-1}$  can be matched in such a way that the arcs  $(f^j(t), f^{k(j)}(\tau))$  are disjoint (where  $f^{k(j)}(\tau)$  is the point matched with  $f^j(t)$ ), then

$$|\log Df^q(t) - \log Df^q(\tau)| \leq \text{Var}(\log Df). \tag{1.4}$$

with  $\text{Var}(\log Df)$  denoting the variation of  $\log Df$  on  $T$ . This follows immediately from (1.2).

- (d) If (our standing assumption)  $\alpha = \rho(f)$  is irrational and  $q = q_n$  is a denominator of a convergent in the continued fraction expansion<sup>†</sup> of  $\alpha$ , then the matching condition of (c) is satisfied for any pair  $t, \tau$ . Thus taking  $\tau = t_{q_n}$  of (1.3) and  $t$  arbitrary, (1.4) reads

$$\|\log Df^{q_n}\|_\infty \leq \text{Var}(\log Df). \tag{1.5}$$

Inequality (1.5) is Denjoy’s inequality and combined with Theorem 1.3, it clearly implies

**THEOREM 1.4.** (Denjoy): *If  $f \in \mathcal{R}^1$  and  $Df$  has bounded variation, and if  $\alpha = \rho(f)$  is irrational, then  $f$  is conjugate to  $R_\alpha$ .*

The key idea which we broke above the observations (a)–(d) is to estimate the sum (1.2), using the fact that the terms have both signs and there is a fair amount of cancellation. We use the fact that for some point  $t = t_q$  the cancellation is perfect, (1.3), and compare other points to  $t_q$ .

We shall use the notation:

$$\begin{cases} K_n^0(t) = |\log Df^{q_n}(t)| \\ \mathbf{K}_n^0 = \|K_n^0(t)\|_\infty, \end{cases} \tag{1.6}$$

$$\eta_n(t) = |f^{q_n-1}(t) - t|, \quad \boldsymbol{\eta}_n = \|\eta_n(t)\|_\infty, \tag{1.7}$$

$$\eta_{m,n}(t) = \frac{\eta_n(t)}{\eta_m(t)}, \quad \boldsymbol{\eta}_{m,n} = \|\eta_{m,n}(t)\|_\infty. \tag{1.8}$$

We rephrase Lemma A.1.1 (cf, Appendix 1) as

**LEMMA 1.5.**  $\boldsymbol{\eta}_{m,n} = \mathcal{O}(\exp\{-c(n-m)\})$  with  $c > 0$ ,  $c$  depends only on  $\text{Var}(\log Df)$ .

*Definition.* An interval<sup>‡</sup>  $J = (t, \tau)$  is  $n$ -small and the pair  $t, \tau$   $q_n$ -close if  $\{f^j(J)\}_{j=0}^{q_n-1}$  are disjoint. We remark that the fact that two points on the circle determine two arcs, and not just one, leads to no confusion in the context of (c) above or the present definition; for later use we shall always take the shorter arc (in terms of the variation of  $h$ ).

The following is very much in the spirit of Denjoy’s inequality:

**LEMMA 1.6.** *If  $t$  and  $\tau$  are  $q_n$ -close and  $0 < m < q_n$ , then*

$$|\log Df^m(t) - \log Df^m(\tau)| \leq \text{Var}(\log Df).$$

*Proof.*

$$\log Df^m(t) - \log Df^m(\tau) = \sum_{j=0}^{m-1} [\log Df(f^j(t)) - \log Df(f^j(\tau))]$$

and the intervals  $(f^j(t), f^j(\tau))$  are disjoint. □

<sup>†</sup> See Appendix 1 for the notations and basic properties of the continued fraction expansion of  $\alpha$  and its relation to the dynamics of  $R_\alpha$ . We shall use a variation of the matching condition given by Lemma A.1.2.

<sup>‡</sup> We shall abuse the notation and write intervals as  $(a, b)$  without concern to determine whether  $a < b$  or  $b < a$ ; this because  $f^{q_n}(t)$  lies on both (alternate) sides of  $t$  and we choose not to modify the notation according to the parity of  $n$ .

LEMMA 1.7. *If  $t$  and  $\tau$  are  $q_n$ -close then*

$$|\log \eta_n(t) - \log \eta_n(\tau)| \leq \mathbf{K}_n^0. \tag{1.9}$$

*Proof.* Check (cf, Appendix 1) that  $t$  and  $\tau$  are  $q_n$ -close if either  $\tau \in (t, f^{q_n-1}(t))$  or  $t \in (\tau, f^{q_n-1}(\tau))$ . If, say,  $t < \tau < f^{q_n-1}(t) < f^{q_n-1}(\tau)$  we have

$$\begin{aligned} \eta_n(\tau) &= f^{q_n-1}(t) - \tau + f^{q_n-1}(\tau) - f^{q_n-1}(t), \\ \eta_n(t) &= f^{q_n-1}(t) - \tau + \tau - t, \end{aligned}$$

and the ratio  $\eta_n(\tau)/\eta_n(t)$  is closer to 1 than  $|f^{q_n-1}(\tau) - f^{q_n-1}(t)|/|\tau - t|$  which lies in the interval  $(\exp(-\mathbf{K}_n^0), \exp(\mathbf{K}_n^0))$ . □

An immediate corollary of (1.9), taken for  $n + 1$  rather than  $n$ , with  $\tau = f^{q_n}(t)$  and repeated  $j$  times is

$$|\log \eta_{n+1}(t) - \log \eta_{n+1}(f^{jq_n}(t))| \leq |j| \mathbf{K}_n^0.$$

In particular if  $a_{n,n+1} \mathbf{K}_n^0 \leq 1$  then the ratio of any two  $q_{n+1}$ -intervals within the same  $q_n$ -interval is bounded by  $e < 3$  and we obtain the estimate  $\eta_{n,n+1} < (3a_{n,n+1})^{-1}$ . If  $a_{n,n+1} \mathbf{K}_n^0 > 1$ , any  $q_{n+1}$ -interval has at least  $(\mathbf{K}_n^0)^{-1}$  comparable (within a factor of 3) disjoint images and we obtain

LEMMA 1.8.  $\eta_{n,n+1} \leq 3 \max(a_{n,n+1}^{-1}, \mathbf{K}_n^0)$ .

*Remark.* If  $a_{n,n+1} \mathbf{K}_n^0 < \varepsilon < 1$ , then the bound for  $\eta_{n,n+1}$  is  $e^\varepsilon a_{n,n+1}^{-1}$  and if we have  $a_{n,n+1} \mathbf{K}_n^0 \rightarrow 0$ , we may ignore the factor 3 (or  $e^\varepsilon$ )

Some more notation: for integers  $s \geq 0, n > 0, 0 < l \leq q_n$  we write

$$K_n^s(l) = K_n^s(l, t) = \eta_n^s(t) |D^s \log Df^l(t)|, \tag{1.10}$$

$$K_n^s = K_n^s(q_n); \quad \mathbf{K}_n^s = \|K_n^s(t)\|_\infty, \tag{1.11}$$

$$\bar{K}_n^s \text{ denotes a term of the form } K_n^s(jq_{n-1}) \text{ with } j \leq a_{n-1,n}, \tag{1.12}$$

$$\tilde{K}_n^s \text{ denotes a term of the form } K_n^s(l), l \leq q_n; \tag{1.13}$$

$$\tilde{\mathbf{K}}_n^s = \sup \|K_n^s(l)\|_\infty, \quad 0 < l \leq q_n.$$

A more useful version of Lemma 1.6 is

LEMMA 1.9. *If  $t$  and  $\tau$  are  $q_n$ -close and  $0 < l \leq q_n$ , then*

$$|\log Df^l(t) - \log Df^l(\tau)| \leq \tilde{\mathbf{K}}_n^1 \exp \mathbf{K}_n^0.$$

*Proof.* For some  $\bar{t} \in (t, \tau)$

$$\begin{aligned} |\log Df^l(t) - \log Df^l(\tau)| &= |t - \tau| |D \log Df^l(\bar{t})| \\ &\leq \eta_n(t) |D \log Df^l(\bar{t})| \leq \tilde{\mathbf{K}}_n^1 \exp \mathbf{K}_n^0 \end{aligned}$$

by Lemma 1.7. □

*Remark.* Denjoy's inequality (1.5) gives the bound  $b_0 = \exp(\text{Var } \log Df)$  for  $\exp \mathbf{K}_n^0$  and, for the sake of shorter formulae we sometimes use the constant  $b_0$  instead of the better estimate  $\exp \mathbf{K}_n^0$ . Thus we shall sometimes replace Lemma 1.8 by

$$|\log Df^l(t) - \log Df^l(\tau)| \leq b_0 \tilde{\mathbf{K}}_n^1. \tag{1.14}$$

Similarly, if  $t$  and  $\tau$  are  $q_n$ -close and  $l < q_m$  with  $m < n$  we have

$$|\log Df^l(t) - \log Df^l(\tau)| \leq b_0 \tilde{\mathbf{K}}_m^1 \eta_{m,n}(t). \tag{1.15}$$

Denjoy’s inequality (1.5) can be improved as follows: We want to estimate  $\log Df^{q_n}(t)$ ; let  $\bar{t}$  be such that  $\log Df^{q_n}(\bar{t}) = 0$ . There exists a  $\tau$  which is  $q_n$ -close to  $t$  and such that  $\bar{t} = f^l(\tau)$  with  $0 \leq l < q_n$ . Now

$$\begin{aligned} \log Df^{q_n}(t) &= \log Df^{q_n}(t) - \log Df^{q_n}(\bar{t}) \\ &= \log Df^l(t) - \log Df^l(f^{q_n}(\tau)) + \log Df^{q_n-l}(f^l(t)) - \log Df^{q_n-l}(\bar{t}) \end{aligned}$$

which gives, by (1.14)

$$\mathbf{K}_n^0 \leq 3b_0 \tilde{\mathbf{K}}_n^1. \tag{1.16}$$

For a variant see (4.35).

### 2. The smoothness of the conjugation

The groups  $\mathcal{H}^k$  have a natural translation invariant metric which is defined as follows: the  $\mathcal{H}^k$ -distance of  $\phi \in \mathcal{H}^k$  to the identity is  $\|\phi\|_k = \sup |\phi(t) - t| + \|D\phi\|_{C^{k-1}}$  where  $\|g\|_{C^{k-1}} = \sum_{j=0}^{k-1} \|D^j g\|_\infty$  (and for non integer values of  $k$  the appropriate Hölder constant of the last derivative is added), and the  $\mathcal{H}^k$ -distance between  $\psi, \phi \in \mathcal{H}^k$  is  $\|\phi_0 \psi^{-1}\|_k$ .

The differentiability class of  $h$  can be determined by

**THEOREM 2.1.** *Assume  $f = h^{-1}R_\alpha h$ ,  $\alpha$  irrational. Consider the following conditions:*

- (i)  $h \in \mathcal{H}^k$  ( $C^k$ -diffeomorphisms)
- (ii)  $\{f^j\}$  bounded in  $\mathcal{H}^k$
- (iii)  $\{f^j\}$  precompact in  $\mathcal{H}^k$ .

*Then, if  $k$  is a positive integer, all three conditions are equivalent; if  $k > 1$  is not an integer, conditions (i) and (ii) are equivalent.*

Even though the theorem offers necessary and sufficient conditions and is not hard to prove, it is not, methodologically, optimal. The only part that requires an argument (cf, [H]) in addition to Lemma 2.2 below is the implication (ii)  $\Rightarrow$  (i) which is useful though seldom essential since, often, proofs of (ii) give (iii) as a bonus. Worse than that: if we want to use it to prove  $h \in \mathcal{H}^k$  when  $k$  is not an integer we need uniform estimates of the modulus of continuity of  $D^{[k]}f^j$ , which may be tedious or difficult or both. The variation that we propose in Lemma (2.2), and Theorem 2.3 are much better adapted to our needs.

In the following lemma  $\tilde{h}$  and  $\tilde{f}$  are the liftings to  $\mathbf{R}$  of  $h$  and  $f$ , respectively.

**LEMMA 2.2.**  $\tilde{h}(x) = \text{const} + \lim_{N \rightarrow \infty} N^{-1} \sum_{j=1}^N (\tilde{f}^j(x) - j\alpha)$ .

*Proof.* Recall that  $f^j = h^{-1}R_{j\alpha}h$ , so that

$$\begin{aligned} \tilde{f}^j(x) &= \tilde{h}^{-1}(\tilde{h}(x) + j\alpha), \\ \tilde{f}^j(x) - j\alpha - \tilde{h}(x) &= (\tilde{h}^{-1} - Id) \circ (\tilde{h}(x) + j\alpha). \end{aligned}$$

Averaging for  $1 \leq j \leq N$  and letting  $N \rightarrow \infty$ , we have

$$\tilde{h}(x) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N (\tilde{f}^j(x) - j\alpha) - \int_0^1 (\tilde{h}^{-1}(u) - u) du. \quad \square$$

**COROLLARY.** *The implication (iii)  $\Rightarrow$  (i) of Theorem 2.1.*

**THEOREM 2.3.** *Assume that  $f^j = h^{-1}R_{j\alpha}h$  are uniformly bounded in  $\mathcal{X}^k$ ,  $k$  a positive integer, and that the mapping  $j\alpha \rightarrow f^j$  from  $\{j\alpha\} \subset \mathbf{T}$  (with the metric induced by  $\mathbf{T}$ ) into  $\mathcal{X}^k$  (with its natural metric) satisfies the modulus of continuity  $\phi$ . Then  $h \in \mathcal{X}^k$  and  $D^k h$  satisfies the same modulus of continuity.*

*Proof.* By Lemma 2.2 (or by its consequence, Theorem 2.1 implication (iii)  $\Rightarrow$  (i)) we obtain  $h \in \mathcal{X}^k$ . We can extend the mapping by continuity to all of  $\mathbf{T}$ , the extended mapping being  $\tau \rightarrow h^{-1}R_\tau h$ , and it satisfies the same modulus of continuity as its restriction to  $\{j\alpha\}$  namely  $\phi$ . All that we need to do is show that  $D^k h$  satisfies  $\phi$  and we start with the case  $k = 1$  which is completely transparent: For an arbitrary  $t \in \mathbf{T}$  write  $t_0 = h(t)$  and consider

$$D(h^{-1}R_\tau h)(t) = Dh^{-1}(t_0 + \tau)Dh(t) = \frac{Dh^{-1}(t_0 + \tau)}{Dh^{-1}(t_0)}.$$

Since the distance in  $\mathcal{X}^1$  of  $h^{-1}R_\tau h$  from the identity is bounded by  $\phi(\tau)$ , we obtain

$$|Dh^{-1}(t_0 + \tau) - Dh^{-1}(t_0)| \leq C\phi(\tau).$$

Thus  $Dh^{-1}$  satisfies  $\phi$  and hence so does  $Dh$ .

For  $k > 1$  we point out that the modulus of continuity of  $D^k h$  is the same as (bounded by a constant multiple of) the modulus of continuity of  $D^{k-1} \log Dh$  which we evaluate using formula (4.11) writing:

$$\begin{aligned} D^{k-1} \log D(h^{-1}R_\tau h) &= D^{k-1} \log D(h^{-1}(\tau + h)) \\ &= (D^{k-1} \log Dh^{-1})(\tau + h)(Dh)^{k-1} + D^{k-1} \log D(\tau + h) \\ &\quad + \text{terms in } C^1 \tag{2.1} \\ &= (D^{k-1} \log Dh^{-1})(\tau + h)(Dh)^{k-1} + D^{k-1} \log Dh + \text{terms in } C^1. \end{aligned}$$

Our assumption implies that  $D^{k-1} \log Dh^{-1}R_\tau h(0)$  satisfies, as a function of  $\tau$ , the modulus of continuity  $\phi$ ; and this implies the same for  $D^{k-1} \log Dh^{-1}(\tau + h(0))$  since, by (2.1), the two differ by a  $C^1$ -function.  $\square$

In the following theorem, as in Appendix A-1,  $d_n$  denotes the distance of  $q_n\alpha$  to 0.

**THEOREM 2.4.** *Assume  $\sum a_{m,m+1} \mathbf{K}_n^0 < \infty$ . Set*

$$\varphi(d_n) = \sum_{m=n}^{\infty} a_{n,n+1} \mathbf{K}_m^0 \tag{2.2}$$

*and extend  $\varphi$  by linearity in every  $(d_{n+1}, d_n)$ . Then the mapping  $\tau \rightarrow h^{-1}R_\tau h$  from  $\mathbf{T}$  to  $\mathcal{X}^1$  satisfies the modulus of continuity  $\varphi$ ,  $h \in \mathcal{X}^1$  and  $Dh$  satisfies the modulus of continuity  $\varphi$ .*



*Proof.* Since the mapping in question is a representation of  $T$  (resp.  $\{j\alpha\}$ ) in  $\mathcal{H}^1$ , it is enough to check the continuity at the identity. Let  $\tau$  be close to zero on the right, say,

$$d_{n+1} < \tau \leq d_n.$$

We can write

$$\tau = \sum_{n+1}^{\infty} c_m d_m,$$

with  $c_m$  integers and  $c_m < a_{m,m+1}$  (recall that  $a_{m,m+1} = [d_{m-1}/d_m]$ ), and hence

$$h^{-1}R_\tau h = \lim_{N \rightarrow \infty} f \sum_{n+1}^N \pm c_m q_m \quad \text{in } \mathcal{H}^1, \tag{2.3}$$

$$\log D(h^{-1}R_\tau h) = \lim_{N \rightarrow +\infty} \sum_{n+1}^N c_m \log Df^{\pm q_m} \circ f^{j_m}, \tag{2.4}$$

computed at appropriate points (the chain rule), and finally

$$\sup |\log D(h^{-1}R_\tau h)| \leq \sum_{n+1}^{\infty} c_m \mathbf{K}_m^0 \leq \sum_{n+1}^{\infty} a_{m,m+1} \mathbf{K}_m^0. \tag{2.5}$$

Remembering that  $\log x \sim 1-x$  for  $x \sim 1$  (here  $x = \log Dh^{-1}R_\tau h$ ), (2.5) clearly gives us the right estimate for the ‘derivative part’ of the  $\mathcal{H}^1$  distance of  $h^{-1}R_\tau h$  to the identity, but we also need to control the ‘uniform part’. The fact that  $h$  is a homeomorphism gives some estimate which, with (2.5), implies that  $\tau \rightarrow h^{-1}R_\tau h$  is a continuous representation in  $\mathcal{H}^1$ , and by Theorem 2.3 we may conclude  $h \in \mathcal{H}^1$ . This now gives the uniform estimate  $|h^{-1}R_\tau h - Id| = \mathcal{O}(\tau)$ ,  $\tau \rightarrow 0$ , so that in the  $\mathcal{H}^1$ -distance the uniform part is small compared to the distance in terms of the derivative which is controlled by (2.5). □

### 3. Assuming $f \in \mathcal{H}^{2+\psi}$

The notation  $f \in \mathcal{H}^{2+\psi}$  means that  $f \in \mathcal{H}^2$  and that the modulus of continuity of  $D^2f$  (or, equivalently, of  $D \log Df = D^2f/Df$ ) is bounded by  $\psi$ . If  $\psi(x) = x^\delta$  we write, as usual,  $\mathcal{H}^{2+\delta}$ . Our goal in this section is to prove that if  $f \in \mathcal{H}^{2+\delta}$ ,  $\delta > 0$ , then  $\mathbf{K}_n^0 \rightarrow 0$  exponentially fast, independently of  $\alpha$ . By the results of the previous section this implies  $C^1$ -conjugation provided the coefficients  $a_{n,n+1}$  of  $\alpha$  do not grow too fast, a proviso which is satisfied for almost all  $\alpha$  (see Appendix 2). The exponential decay of  $\mathbf{K}_n^0$  will also be a good starting point in § 4.

Even the weaker continuity condition  $\psi(x) = \mathcal{O}((\log x^{-1})^{-1-\epsilon}) \in > 0$ , provides an estimate,  $\mathbf{K}_n^0 = \mathcal{O}(n^{-1-\epsilon})$ , which is still sufficient for  $C^1$  conjugation for almost all  $\alpha$ .

In order to estimate  $\mathbf{K}_n^0$  efficiently we need to estimate also  $\mathbf{K}_n^1$  and  $\tilde{\mathbf{K}}_n^1$ , estimates that will also serve later, when we investigate higher differentiability of  $h$  under appropriate conditions. The method consists in applying the basic procedure for  $\mathbf{K}_n^1$  (the original Denjoy was for  $\mathbf{K}_n^0$ ) and, in the following section, to  $\mathbf{K}_n^s$  for  $s = 1, 2, 3, \dots$ . Our presentation here (§ 3) differs slightly from the general scheme of § 4, due to the need to establish that  $\mathbf{K}_n^1 \rightarrow 0$  as  $n \rightarrow \infty$ . As we shall presently see

this can be obtained easily from the ratio ergodic theorem (which is invoked in [H]) and had we chosen to use that, we could have continued with the format of § 4, as we could have also under the assumption  $\text{Var}(\log Df) < \frac{1}{4}$ . Instead we chose to adapt the presentation so as to avoid either the ergodic theorem or unnecessary assumptions (see also [C]).

We denote

$$W_n(t) = \sum_{j=0}^{q_n-1} Df^j(t) \tag{3.1}$$

and notice that if  $t$  and  $\tau$  are  $q_n$ -close

$$e^{-b_0 \tilde{K}_n^1} \leq \frac{W_n(t)}{W_n(\tau)} \leq e^{b_0 \tilde{K}_n^1} \tag{3.2}$$

as (1.14) implies the same for each of the summands  $Df^j$  which make  $W_n$ .

LEMMA 3.1.  $W_n(t) \sim (\eta_n(t))^{-1}$ .

(The notation  $\sim$  means that the ratios of the sides are bounded from zero and above, uniformly in the parameters, here  $n$  and  $t$ .)

*Proof.* Set  $I = (t, f^{q_n-1}(t))$ . By (3.2)

$$\int_I W_n(\tau) d\tau \sim W_n(t) \eta_n(t). \tag{3.3}$$

The integral (3.3) expresses the measure of the set  $J = \bigcup_{j=0}^{q_n-1} f^j(I)$  and hence is  $< 1$ . On the other hand  $J \cup f^{q_n}(J) = T$  and by (1.5)  $J$  and  $f^{q_n}(J)$  have comparable measures. □

Lemma 3.1 sheds some light on the definition of  $K_n^1$ . In fact, replacing  $\eta_n(t)$  by  $(W_n(t))^{-1}$  we get the expression

$$\underline{K}_n^1 = \frac{\sum_{j=0}^{q_n-1} D \log Df(f^j(t)) Df^j(t)}{\sum_{j=0}^{q_n-1} Df^j(t)}, \tag{3.4}$$

which is a weighted average of  $D \log Df$  along a  $q_n$ -piece of the orbit of  $t$ , the relative weights being  $Df^j(t)$ . (At this point we can invoke the ergodic theorem and show that  $\underline{K}_n^1 \rightarrow 0$  a.e.)

By Lemma 3.1,  $\underline{K}_n^1 \sim K_n^1$  and similarly for  $\tilde{K}_n^1 = (W_n \eta_n)^{-1} \tilde{K}_n^1$ . It will be somewhat simpler to write the main estimates of this section in terms of  $\underline{K}_n^1$  and this is what we do.

LEMMA 3.2. For  $m < n$

$$W_m(t) = \mathcal{O}(W_n(t) \eta_{m,n}), \tag{3.5}$$

$$\tilde{K}_n^1 \leq \underline{K}_m^1 + \mathcal{O}(\eta_{m,n}). \tag{3.6}$$

Both constants in the  $\mathcal{O}$ -estimates depend only on  $\|D \log Df\|_\infty$ .

*Proof.* (3.5) follows immediately from Lemma 3.1. For (3.6) we take  $l < q_n$  and write

it as  $l = bq_m + k$ ,  $k < q_m$ . For  $t \in \mathbf{T}$ , write  $t_i = f^{iq_m}(t)$ ;

$$\begin{aligned}
 D \log Df^l(t) &= \sum_{j=0}^{l-1} D \log Df(f^j(t)) Df^j(t) \\
 &= \sum_{i=0}^{b-1} Df^{iq_m}(t) \sum_{j=0}^{q_m-1} D \log Df(f^j(t_i)) Df^j(t_i) \\
 &\quad + Df^{bq_m}(t) \sum_{j=0}^{k-1} D \log Df(f^j(t_b)) Df^j(t_b).
 \end{aligned}
 \tag{3.7}$$

Dividing (3.7) on both sides by  $W_n(t)$ , we estimate the right hand side:

$$W_n^{-1}(t) \sum_{i=0}^{b-1} Df^{iq_m}(t) |D \log Df^{q_m}(t_i)| = \sum_{i=0}^{b-1} W_n^{-1}(t) Df^{iq_m}(t) W_n(t_i) \mathbf{K}_m^1(t_i) \leq \mathbf{K}_m^1$$

(3.8)

while, remembering that by Denjoy’s inequality (1.5), we have

$$W_n(t) \sim Df^{bq_m}(t) W_n(t_b),$$

and the last term coming from (3.7) is bounded by

$$\|D \log Df\|_\infty W_m(t_b) W_n^{-1}(t_b) = \mathcal{O}(\boldsymbol{\eta}(m, n)). \quad \square$$

If we denote  $b_1 = b_0 \sup(\tilde{\mathbf{K}}_n^1 / \tilde{\mathbf{K}}_n^1)$  we can rewrite (1.15) in the form

$$|\log Df^l(t) - \log Df^l(\tau)| \leq b_1 \tilde{\mathbf{K}}_m^1 \eta_{m,n}$$

(3.9)

(for  $0 < l \leq q_m$ ,  $m \leq n$  and  $t, \tau q_n$ -close) or, for  $m = n$ ,

$$|\log Df^l(t) - \log Df^l(\tau)| \leq b_1 \tilde{\mathbf{K}}_n^1$$

(3.9\*)

**PROPOSITION 3.3.**

(a) Assume  $f \in \mathcal{H}^2$ ; then  $\mathbf{K}_n^1 \rightarrow 0$ .

(b) Assume  $f \in \mathcal{H}^{2+\psi}$ ; then, for  $m < n$  and a constant  $b_1$  which depends only on  $\|D \log Df\|_\infty$

$$\mathbf{K}_n^1 \leq 2b_1(\mathbf{K}_m^1)^2 + \mathcal{O}(\boldsymbol{\eta}_{m,n}) + \mathcal{O}(\psi(\boldsymbol{\eta}_n)).$$

(3.10)

*Proof.* We shall first obtain the estimate

$$\mathbf{K}_n^1 \leq (1 - \exp(-2b_1 \mathbf{K}_m^1 - 2b_2 \boldsymbol{\eta}_{mn,n})) \mathbf{K}_m^1 + \mathcal{O}(\boldsymbol{\eta}_{m,n}) + \mathcal{O}(\psi(\boldsymbol{\eta}_n)).$$

(3.10\*)

In order to do that we need to estimate  $D \log Df^{q_n}(t)$  for arbitrary  $t$ . Being the derivative of a periodic function,  $D \log Df^{q_n}$  has mean value zero and hence there are points  $\bar{t}$  for which  $D \log Df^{q_n}(\bar{t}) = 0$ ; we can write, for any choice of  $c$ ,

$$D \log Df^{q_n}(t) = \sum_{j=0}^{q_n-1} D \log Df(f^j(t)) Df^j(t) - c \sum_{j=0}^{q_n-1} D \log Df(f^j(\bar{t})) Df^j(\bar{t}).$$

(3.11)

There exists  $\tau$  which is  $q_n$ -close to  $t$  and such that  $\bar{t} = f^l(\tau)$  with  $0 \leq l \leq q_n$ . Write

$$j^* = \begin{cases} j, & j \geq l, \\ j + q_n, & j < l, \end{cases}$$

then  $f^j(t)$  and  $f^{j^*}(\tau)$  are  $q_n$ -close for  $j \geq l$  and, at worst, ‘twice the distance of  $q_n$ -close’ for  $j < l$  (which does not affect the following estimates).

We now write

$$D \log Df^{q_n}(t) = \sum_{j=0}^{q_n-1} [D \log Df(f^j(t))Df^j(t) - c_1 D \log Df(f^{j^*}(\tau))Df^{j^*}(\tau)], \tag{3.12}$$

with  $c_1 = cDf^{-1}(\bar{t})$  still completely arbitrary, and rewrite it as

$$\left\{ \begin{aligned} D \log Df^{q_n}(t) &= \sum_{j=0}^{q_n-1} (D \log Df)(f^j(t))[Df^j(t) - c_1 Df^{j^*}(\tau)] \\ &\quad + c_1 \sum_{j=0}^{q_n-1} [(D \log Df)(f^j(t)) - (D \log Df)(f^{j^*}(\tau))]Df^{j^*}(\tau) \\ &= S_1 + S_2. \end{aligned} \right. \tag{3.13}$$

The constant  $c_1$  will be bounded by 1 and  $|f^j(t) - f^{j^*}(\tau)| < \eta_n$  so that the term  $S_2$  in (3.13) is  $\mathcal{O}(\psi(\eta_n)W_n(\tau))$  and contributed the last summand in (3.10).

To estimate  $S_1$  in (3.13) we break the block  $0 \leq j < l$  into blocks of length  $q_m$  and possibly a shorter remainder block; similarly for the sum from  $l$  to  $q_n - 1$ .

The  $q_m$ -blocks,  $\sum_s^{s+q_m-1}$ , can be written (with  $t_s = f^s(t)$  and  $\tau_s = f^{s^*}(\tau)$ ) as

$$\sum_{j=0}^{q_m-1} (D \log Df)(f^j(t_s))[Df^s(t)Df^j(t_s) - c_1 Df^{s^*}(\tau)Df^j(\tau_s)] \tag{3.14},$$

and we compare separately  $Df^s(t)$  with  $Df^{s^*}(\tau)$  and  $Df^j(t_s)$  with  $Df^j(\tau_s)$ .

For the first pair we evaluate (by (3.9\*) and Lemma 3.2)

$$|\log Df^s(t) - \log Df^{s^*}(\tau)| \leq b_1 \mathbf{K}_m^1 + b_2 \eta_{m,n} \tag{3.15}$$

and if  $s^* = s + q_n$  we add the estimate

$$|\log Df^s(t) - \log Df^{s^*+q_n}(\tau)| = |\log Df^{q_n}(f^s(\tau))| \leq \mathbf{K}_n^0 \leq b_1 \mathbf{K}_m^1 + b_2 \eta_{m,n}. \tag{3.16}$$

For the second pair we have, by (3.9),

$$|\log Df^j(t_s) - \log Df^j(\tau_s)| \leq \tilde{\mathbf{K}}_m^1 \eta_{m,n} \leq b_1 \|D \log Df\|_\infty \eta_{m,n}$$

and since  $b_1 \|D \log Df\|_\infty$  is fixed and  $\eta_{m,n}$  is small (exponentially in  $n - m$ ) we have

$$|Df^j(t_s) - Df^j(\tau_s)| \leq \mathcal{O}(Df^j(t_s)\eta_{m,n})$$

and we can write (3.14), as

$$\left\{ \begin{aligned} &\sum_{j=0}^{q_m-1} (D \log Df)(f^j(t_s))Df^j(t_s)[Df^s(t) - c_1 Df^{s^*}(\tau)] + \mathcal{O}(Df^s(t)W_m(t_s)\eta_{m,n}) \\ &= D \log Df^{q_m}(t_s)[Df^s(t) - c_1 Df^{s^*}(\tau)] + \mathcal{O}(Df^s(t)W_m(t_s)\eta_{m,n}). \end{aligned} \right. \tag{3.17}$$

We now take†  $c_1 = \inf(Df^s(t)/Df^{s^*}(\tau))$  the infimum being taken for all the configurations of  $s, t, \tau$ . By (3.15), (3.16) etc . . . we have

$$1 > c_1 \geq e^{-b_1 \mathbf{K}_m^1 - b_2 \eta_{m,n}}, \tag{3.18}$$

$$0 \leq Df^s(t) - c_1 Df^{s^*}(\tau) \leq (1 - c_1^2) Df^s(t). \tag{3.19}$$

† Had we used the ratio ergodic theorem or assumed that  $\mathbf{K}_m^1$  is small, we could simply take  $c_1 = 1$  and replace (3.19) by  $|Df^s(t) - Df^{s^*}(\tau)| \leq (b_1 \mathbf{K}_m^1 + b_2 \eta_{m,n}) Df^s(t)$ .

Adding the absolute values of (3.17)<sub>s</sub> for all *s* we obtain

$$S_1 \leq \sum \underline{K}_m^1(t_s) W_m(t_s) Df^s(t) (1 - c_1^2) + \mathcal{O}(\eta_{m,n} \sum Df^s(t) W_m(t_s))$$

+ the absolute value of the 2 'remainder' blocks.

Each one of the 2 remainder blocks is bounded by  $\|D \log Df\| \sum_{j=u}^v Df^j(t)$  with  $0 \leq u \leq v < q_n$  and  $v - u \leq q_m$ , which is bounded by  $\mathcal{O}(\eta_{m,n} W_n(t))$  (notice that we can assume that either  $v = q_{n-1}$ , or  $u = 0$ ).

Since  $\sum_s Df^s(t) W_m(t_s) \leq W_n(t)$  we obtain

$$S_1 \leq [(1 - c_1^2) \underline{K}_m^1 + \mathcal{O}(\eta_{m,n})] W_n(t),$$

which, together with (3.18) and the estimate we had for  $S_2$ , completes the proof of (3.10\*).

If  $f \in \mathcal{H}^2$  we have (3.10\*) with  $\psi$  denoting whatever modulus of continuity  $D \log Df$  has. Now the first term on the right hand side of (3.10\*) is at least a fixed proportion less than  $\underline{K}_m^1$  while the rest tends to zero as  $n \rightarrow \infty$ . Thus if  $n$  is sufficiently bigger than  $m$ ,  $\underline{K}_n^1 < \rho \underline{K}_m^1$  with  $\rho < 1$ . This proves (a).

Once we know that  $\underline{K}_m^1$  is small, (3.10) follows directly from (3.10\*). □

**COROLLARY 3.4.** *If  $f \in \mathcal{H}^2$  then  $\underline{K}_n^1 \rightarrow 0$ .*

Remembering that  $\eta_n < \eta^n$  and  $\eta_{m,n} < \eta^{n-m}$  for some  $\eta < 1$ , we obtain

**COROLLARY 3.5.**

- (a) *If  $\psi(x) = \mathcal{O}(x^\delta)$ ,  $0 < \delta \leq 1$ , then  $\underline{K}_n^1 \rightarrow 0$  exponentially fast.*
- (b) *If  $\psi(x) = \mathcal{O}((\log x^{-1})^{-\delta})$ ,  $0 < \delta$ , then  $\underline{K}_n^1 = \mathcal{O}(n^{-\delta})$ .*

We may replace the assumption  $f \in \mathcal{H}^{2+\psi}$  by  $f \in \mathcal{H}^{2+bv}$ , that is instead of modulus of continuity  $\psi$  for  $D \log Df$  we assume simply that it is of bounded variation. That only affects the estimates of  $S_2$  in (3.13) and instead of  $\mathcal{O}(\psi(\eta_n) W_n(\tau))$  obtained there we now get  $\text{Var}(D \log Df) \sup_j Df^j(\tau)$ . After dividing by  $W_n(t)$ , i.e. multiplying by  $\eta_n(t) \sim \eta_n(\tau)$  we obtain the estimate  $S_2 = \mathcal{O}(\eta_n)$ , which is the same we would obtain under the assumption  $D \log Df \in \text{Lip}(1)$ . We record it as

**COROLLARY 3.5.**

- (c) *If  $f \in \mathcal{H}^{2+bv}$ , then  $\underline{K}_n^1 \rightarrow 0$  exponentially fast.*

**PROPOSITION 3.6.**

- (a) *If  $\psi(x) = \mathcal{O}(x^\delta)$ ,  $0 < \delta \leq 1$ , then  $\tilde{\mathbf{K}}_n^1 \rightarrow 0$  exponentially fast.*
- (b) *If  $\psi(x) = \mathcal{O}((\log x^{-1})^{-\delta})$ ,  $0 < \delta$ , then  $\tilde{\mathbf{K}}_n^1 = \mathcal{O}(n^{-\delta})$ .*
- (c) *If  $f \in \mathcal{H}^{2+bv}$  then  $\tilde{\mathbf{K}}_n^1 \rightarrow 0$  exponentially fast.*

*Proof.* Corollary 3.5 and estimate (3.6) with  $m = n/2$ . □

By (1.16) we have

$$\mathbf{K}_n^0 = \mathcal{O}(\tilde{\mathbf{K}}_n^1) \tag{3.20}$$

and this, combined with Proposition 3.6, gives

**THEOREM 3.7.** *Assume  $f \in \mathcal{H}^{2+\psi}$*

- (a) *If  $\psi(x) = \mathcal{O}(x^\delta)$ ,  $0 < \delta \leq 1$ , then  $\mathbf{K}_n^0 \rightarrow 0$  exponentially fast.*
- (b) *If  $\psi(x) = \mathcal{O}((\log 1/x)^{-\delta})$ ,  $0 < \delta$ , then  $\mathbf{K}_n^0 = \mathcal{O}(n^{-\delta})$ .*
- (c) *If  $D^2f$  is of bounded variation then  $\mathbf{K}_n^0 \rightarrow 0$  exponentially fast.*

COROLLARY 3.8. For  $\alpha \in F$  (cf, Appendix 2), if  $\psi = \mathcal{O}((\log 1/x)^{-\delta})$ ,  $\delta > 1$ , or else† if  $D^2f$  is of bounded variation, then  $f$  is  $C^1$  conjugate to  $R_\alpha$ .

We end this section by remarking that for the estimates involving  $\tilde{K}_n^1$  we don't really need uniform estimates for  $D \log Df^l$ ,  $l < q_n$ . If we go back to the proof of Lemma 1.9 and evaluate  $|\log Df^l(t) - \log Df^l(\tau)|$  not by the mean value theorem, but by integrating  $D \log Df^l$  on  $(\tau, t)$ , we would obtain

$$|\log Df^l(t) - \log Df^l(\tau)| \leq \tilde{K}_n^1,$$

where  $\tilde{K}_n^1 = \sup \int_I |D \log Df^l| dt$ , the supremum being taken for all integers  $l$ ,  $0 \leq l \leq q_n$ , and  $I$  which are  $q_n$ -small.

Assuming that the  $\log Df$  is absolutely continuous we can adjust slightly the proof of Lemma 3.2 and obtain

$$\tilde{K}_n^1 \leq \sup \int_{J_m} |D \log f^{q_m}| dt + o(1) \quad \text{as } n - m \rightarrow \infty$$

the supremum being taken on all  $J_m$  which are  $q_m$  small.

Either by invoking the ratio ergodic theorem or by reviewing our proof of Proposition 3.3, writing  $D \log Df = \Phi + \theta$  with  $\Phi$  continuous of mean value zero and  $\|\theta\|_{L^1}$  small, one obtains

THEOREM 3.9. If  $f \in \mathcal{H}^1$  and  $D \log Df$  is absolutely continuous then, (the supremum relative to all  $q_m$ -small intervals)

$$\sup \int_{J_m} |D \log Df^{q_m}| dt \rightarrow 0$$

and consequently  $\tilde{K}_n^1 \rightarrow 0$  and  $K_n^0 \rightarrow 0$ .

4. The case  $f \in \mathcal{H}^k$ ,  $k > 2$

Our standing assumption in this section is that  $f \in \mathcal{H}^k$  with  $k > 2$ . Since we would have to word things slightly differently depending on whether  $k$  is an integer or not, and when  $k$  is an integer we use the  $k$ th derivative only to give a modulus of continuity (namely  $\text{Lip}(1)$ ) for the previous derivative, and since for the purpose of our theorem below we lose nothing if we replace the  $\mathcal{H}^k$ -condition by  $f \in \mathcal{H}^{k-\epsilon}$  for all  $\epsilon > 0$ , we stipulate that  $k$  is not an integer and rewrite our standing assumption as  $f \in \mathcal{H}^{r+1+\delta}$ ,  $r \geq 1$  is an integer and  $0 < \delta < 1$ . The condition is also written as  $\log Df \in C^{r+\delta}$ .

Our main result is:

THEOREM 4.\* If  $f \in \mathcal{H}^{r+1+\delta}$  and the rotation number  $\alpha = \rho(f)$  satisfies the diophantine condition  $q_{n+1} = \mathcal{O}(q_n^{\beta+1})$  with  $\beta + 1 < r + \delta$ , then the homeomorphism  $h$  which conjugates  $f$  with  $R_\alpha$  is in  $\mathcal{H}^{r+\delta-\beta-\epsilon}$  for all  $\epsilon > 0$ .

The result is sharp (see Appendix 3).

The case  $r = 1$  was treated in section 3 but not completely, our goal there being the exponential decay of  $K_n^1$  and  $K_n^0$  with the implied  $C^1$  conjugation for almost all

† For  $f \in \mathcal{H}^{2+h\nu}$  and the rotation number with bounded coefficients this was done in [H<sub>2</sub>].

rotation numbers. For the complete result, namely that  $0 \leq \beta < \delta < 1$  implies  $C^{1+\delta-\beta-\varepsilon}$  conjugation, we need the methods of the present section. In fact we strongly recommend that the entire section be read first assuming that  $r = 1$ , assumption that cuts out much of the formalism and reveals the strategy of the proof, (especially if one compares it to § 3). The general case will be less forbidding in a second reading.

We shall outline the main ideas of the proof after introducing some additional notation and review the notation introduced in § 1 namely:

$$\eta_n(t) = |f^{q_n-1}(t) - t| \quad \eta_n = \|\eta_n\|_\infty, \tag{4.1}$$

$$\eta_{n,n+1}(t) = \eta_{n+1}(t) / \eta_n(t) \quad \eta_{n,n+1} = \|\eta_{n,n+1}\|_\infty, \tag{4.2}$$

$$K_n^s(m) = K_n^s(m, t) = \eta_n^s(t) |D^s \log Df^m(t)| \quad (\eta_n^s \text{ means } (\eta_n)^s), \tag{4.3}$$

$$K_n^s = K_n^s(q_n) \quad K_n^s = \|K_n^s\|_\infty, \tag{4.4}$$

$\bar{K}_n^s$  denotes a term of the form  $K_n^s(jq_{n-1})$

$$\text{with } j \leq a_{n-1,n} \bar{K}_n^s = \sup \| \bar{K}_n^s \| \text{ for all such terms.} \tag{4.5}$$

$$\tilde{K}_n^s \text{ denotes a term of the form } K_n^s(m), m < q_n. \quad \tilde{K}_n^s = \sup \| K_n^s(m) \|_\infty, \quad m < q_n \tag{4.6}$$

$$\Delta_i K_n^r(t, \tau) = \eta_n^r(t) |D^r \log Df^{q_n}(t) - D^r \log Df^{q_n}(\tau)|$$

(the subscript  $i$  on  $\Delta$  limits its domain to  
' $t$  and  $\tau$  are  $q_i$ -neighbors'; it serves as a reminder,  
especially if  $t$  and  $\tau$  are not explicitly specified)

$$\tag{4.7}$$

Similarly

$$\Delta_i \tilde{K}_n^r(m, t, \tau) = \eta_n^r(t) |D^r \log Df^m(t) - D^r \log Df^m(\tau)|$$

where  $m \leq q_n$  (the  $\tilde{K}_n^r$  convention)  
and  $t, \tau$  are  $q_i$ -neighbors (the  $\Delta_i$  convention).

$$\tag{4.8}$$

Notice that the difference operator is applied before the multiplication by  $\eta_n^r$ .

$$\bar{\eta}_{n,n+1} = \max(\alpha_n, M, K_n^r) \text{ with } \alpha_n = [a_{n,n+1}, a_{n+1,n+2}, \dots] \quad (\text{cf, Appendix 1}),$$

$M$ , a constant to be specified later.

$$\tag{4.9}$$

$$\bar{\eta}_n = \prod_{j=1}^{n-1} \bar{\eta}_{j,j+1} \tag{4.10}$$

We also introduce the following notation convention:

$$A \leq_n B$$

will mean that  $A \leq (1 + \theta_n)B$  with  $\theta_n$  exponentially small (in  $n$ ). The exponent is not specified and we obtain  $A \leq_n C$  from  $\{A \leq_n B \text{ and } B \leq_n C\}$ . We have longer chains of transitivity of the relation  $\leq_n$  but they will be of uniformly bounded length, hence legitimate. We also multiply inequalities like  $A_j \leq_n B_j$ ; and as long as the number of values of  $j$  for which  $n_j = m$  is bounded, uniformly in  $m$ , we can conclude that  $\prod A_j \leq_{m_0} \prod B_j$  where  $m_0 = \min \{n_j\}$ . Similarly,

$$A \sim_n B \text{ means that } A \leq_n B \text{ and } B \leq_n A.$$

*Remark.* The reason for the notation is that in many of our estimates we shall need a multiplicative correction factor which is exponentially close to one. Since the estimates that we are seeking now are typically much smaller than ‘exponentially small’ we certainly cannot afford exponentially small *additive errors*; however multiplicative errors, the product of which converges, can be ignored since their total effect on the final estimates is a constant factor. There are three sources, in all that follows, for the introduction of these multiplicative errors:

(a) replacing  $\eta_n^s \circ g$  by  $\eta_n^s(Dg)^s$  or vice versa, where  $g = f^j$  with  $j < q_n$  requires a multiplicative correction factor  $(Dg(\tau)/Dg(t))^s$  which, by (1.14) and Corollary 3.6, is of the form  $1 + \vartheta_n$  with  $\vartheta_n = O(e^{-cn})$  for some  $c > 0$ . (b) replacing  $\eta_n(t)$  or  $\eta_{n,n+1}(t)$  by the same function at  $\tau$ , where  $\tau$  is a  $q_n$ -neighbor of  $t$ , see Lemma 1.7; and finally (c) in the course of the proof of Proposition 4.5 we evaluate  $K_n^s(1 - \delta_n)$ ,  $\delta_n$  exponentially small, and simply ignore the factor  $(1 - \delta_n)$  in the sequel (cf, (4.38)).

The reader is left to bear some of the responsibility to check that this notation is in fact not abused.

The general strategy of the proof is similar to that of the previous two sections. We obtain the smoothness of  $h$  by invoking Theorem 2.3. Thus we need to show that if  $m + 1 + \gamma < r + \delta - \beta$  ( $m \geq 0$  an integer and  $0 < \gamma < 1$ ; the notation matches that of the proof of Theorem 4.11 in which this step of the proof is carried out), then the mapping  $j\alpha \mapsto f^j$ ,  $j \in \mathbf{Z}$  from  $\{j\alpha\} \subset \mathbf{T}$  (with the induced metric) into  $\mathcal{X}^{m+1}$ , satisfies a  $\text{Lip}(\gamma)$  modulus of continuity.

This is done at the very end of the proof, Theorem 4.12, and that is the only place in which the diophantine condition enters. The first part of the proof, all the way to Theorem 4.8, is preparing the needed estimates of the various derivatives of all the iterates of  $f$ . As begun in § 3, we work with

$$K_n^s(m) = K_n^s(m, t) = \eta_n^s(t) |D^s \log Df^m(t)|$$

rather than with  $D^{s+1}f^m$ . This is done because of the better formal behaviour of  $D^s \log Df^m$  and the ‘equalizing effect’ of the factor  $\eta_n$  (recall Lemma 3.1 and (3.4)).

There are two key elements that are used together and repeatedly in the proof of Theorem 4.8. The first ‘leg’ is  $a$  an almost purely formal extension of the chain rule and we sometime refer to it as ‘breaking into shorter runs’. The only part that is not purely formal in this step, which goes all the way to Theorem 4.2’, is the use of the continuity properties of  $\eta_n$  and  $\eta_{n-1,n}$  which permits the use of ‘ $\leq_n$ ’.

The second leg, Lemma 4.4 until Lemma 4.7, is what we refer to as ‘the basic procedure’ and is the extension of the proof of Denjoy’s inequality and Proposition 3.3 to our setup. The key idea is that it may be easier to evaluate a cocycle-like expression by comparing it to the same expression at a different point and using the fact, if one has it, that there exist points at which the expression must vanish.

Theorem 4.8 is the heart of the proof, where the two legs work together (with hands, nails and huffing and puffing) to give the estimates needed for 4.11.

We denote by  $P_l$  polynomials  $P_l(x_1, \dots, x_l)$  which are homogeneous of weight  $l$  if  $x_j$  is of weight  $j$ . Notice that  $P_l$  may denote different polynomials in different places. By definition  $P_0 \equiv 1$ .



LEMMA 4.1. For  $g, h \in \mathcal{H}^{s+1}$

$$D^s \log D(g \circ h) = (D^s \log Dg \circ h)(Dh)^s + D^s \log Dh + \sum_{l=1}^{s-1} (D^{s-l} \log Dg \circ h)(Dh)^{s-l} P_l(\dots, D^l \log Dh, \dots). \tag{4.11}$$

$$D^s \log Dg^m = \sum_{l=0}^{s-1} \sum_{j=0}^{m-1} (D^{s-l} \log Dg \circ g^j)(Dg^j)^{s-l} P_l(\dots, D^l \log Dg^j, \dots). \tag{4.12}$$

*Proof.* This is formal, checked by induction (recall that  $D \log DF = D^2 F / DF$ ). □

In our context both  $g$  and  $h$  will be powers (iterates) of  $f$  and our goal is inequality (4.19) below. If  $g = f^{q_{n-1}}$ , (4.12) becomes

$$D^s \log Df^{mq_{n-1}} = \sum_{j=0}^{m-1} (D^s \log Df^{q_{n-1}} \circ f^{jq_{n-1}})(Df^{jq_{n-1}})^s + \sum_{l=1}^{s-1} \sum_{j=1}^{m-1} (D^{s-l} \log Df^{q_{n-1}} \circ f^{jq_{n-1}}) \times (Df^{jq_{n-1}})^{s-l} P_l(\dots, D^l \log Df^{jq_{n-1}}, \dots). \tag{4.13}$$

If we multiply both sides of (4.13) by  $\eta_n^s$ , and keeping in mind that  $\eta_n^s \circ g \underset{n}{\sim} \eta_n^s (Dg)^s$ , we obtain:

$$\eta_n^s |D^s \log Df^{mq_{n-1}}| \leq \sum_{j=0}^{m-1} (\eta_{n-1,n}^s K_{n-1}^s) \circ f^{jq_{n-1}} + \sum_{l=1}^{s-1} \sum_{j=1}^{m-1} (\eta_{n-1,n}^{s-l} K_{n-1}^{s-l}) \circ f^{jq_{n-1}} P_l(\dots, \eta_n^l |D^l \log Df^{jq_{n-1}}|, \dots). \tag{4.14}$$

The factors  $\eta_{n-1,n}$  appear because of the different normalization of  $K_{n-1}^s$  and  $K_n^s$  (namely  $\eta_{n-1}^s$  for the first,  $\eta_n^s$  for the latter).

The value of  $m$  in (4.14) will always be between 2 and  $a_{n-1,n}$ ; the range needed for evaluating  $\bar{K}_n, \tilde{K}_n$ , and  $K_n$  in terms of  $K_{n-1}$ , etc. When we evaluate  $K_n$  or  $\tilde{K}_n$  we use (4.11) as well.

We can get rid of the sum over  $j$  in (4.14) by observing that:

$$\sum_{j=0}^{a_{n-1,n}-1} \eta_n \circ f^{jq_{n-1}} < \eta_{n-1} \circ f^{a_{n-1,n}q_{n-1}} \underset{n}{\sim} \eta_{n-1} \tag{4.15}$$

(the intervals  $f^{jq_{n-1}}(t), f^{(j+1)q_{n-1}}(t)$  are disjoint, their union is

$$(t, f^{a_{n-1,n}q_{n-1}}(t)) \subset (f^{q_{n-1}}(t), f^{a_{n-1,n}q_{n-1}}(t)) \tag{4.16}$$

whose length defines  $\eta_{n-1} \circ f^{a_{n-1,n}q_{n-1}}$ , the ' $\underset{n}{\sim} \eta_{n-1}$ ' is justified by the fact that  $t$  and  $f^{a_{n-1,n}q_{n-1}}(t)$  are  $q_{n-1}$ -close, see (b) in the Remark above). This implies

$$\sum_{j=0}^{a_{n-1,n}} \eta_{n-1,n}^{s-l} \circ f^{jq_{n-1}} \leq \eta_{n-1,n}^{s-l-1}, \quad l = 0, 1, \dots, s-1 \tag{4.17}$$

and transforms (4.14) into

$$\bar{\mathbf{K}}_n^s \leq \sum_{l=0}^{s-1} \mathbf{K}_{n-1}^{s-l} \boldsymbol{\eta}_{n-1,n}^{s-l-1} P_l(\bar{\mathbf{K}}_n^1, \dots, \bar{\mathbf{K}}_n^l). \tag{4.18}$$

For  $s = 1$  (4.18) takes the pleasant form

$$\bar{\mathbf{K}}_n^1 \leq \mathbf{K}_{n-1}^1 \tag{4.18}_1$$

and for  $s = 2$  the estimate is

$$\bar{\mathbf{K}}_n^2 \leq \mathbf{K}_{n-1}^2 \boldsymbol{\eta}_{n-1,n} + \mathbf{K}_{n-1}^1 P_1(\bar{\mathbf{K}}_n^1), \tag{4.18}_2$$

where  $P_1$  is now just a constant multiple of the variable. We now replace the term  $\bar{\mathbf{K}}_n^1$  by (4.18) and (4.18)<sub>2</sub> becomes

$$\bar{\mathbf{K}}_n^2 \leq \mathbf{K}_{n-1}^2 \boldsymbol{\eta}_{n-1,n} + C(\mathbf{K}_{n-1}^1)^2. \tag{4.19}_2$$

In the general case we obtain

$$\bar{\mathbf{K}}_n^s \leq \sum \left( \prod \mathbf{K}_{d-1}^{c_j} \boldsymbol{\eta}_{d-1,n}^{c_j-1} \right), \quad \sum c_j = s, \tag{4.19}$$

where the number of summands in (4.19) depends on  $s$  but not on  $n$  nor on  $\alpha = \rho(f)$ . Notice that we did not write the constants – they are replaced by repetitions, so that when we say that the number of terms is bounded we mean the sum of the coefficients in a more conventional writing.

This is a notation convention that we use throughout the rest of the paper, namely writing ‘a sum of terms of certain forms’ with  $\sum$  with unspecified limits; the convention being that the number of summands (or the sum of the coefficients) depends only on specified parameters and nothing else (here on  $s$ ).

The cases  $s = 1, 2$  of (4.19) were done above, the general case is obtained by induction on  $s$  (replacing in (4.18) the terms appearing in  $P_l$  by (4.19) for  $j \leq l < s$ ). For most of our needs (4.19) is sufficient, but for some we shall have to go back to (4.14).

Our next goal is the analogous inequality for  $\tilde{\mathbf{K}}_n$ , (4.22) below. We obtain expressions for  $D^s \log Df^k$ ,  $k \leq q_n$ , by writing  $k = mq_{n-1} + w$ ,  $0 \leq w < q_{n-1}$ , setting  $h = f^{mq_{n-1}}$ ,  $g = f^w$  and applying (4.11) and (4.13). Thus

$$\begin{aligned} D^s \log Df^k &= (D^s \log Df^w \circ f^{mq_{n-1}})(Df^{mq_{n-1}})^s + D^s \log Df^{mq_{n-1}} \\ &+ \sum_{l=1}^{s-1} (D^{s-l} \log Df^w \circ f^{mq_{n-1}})(Df^{mq_{n-1}})^{s-1} P_l(\dots, D^l \log Df^{mq_{n-1}}, \dots) \\ &= (D^s \log Df^w \circ f^{mq_{n-1}})(Df^{mq_{n-1}})^s \\ &+ \sum_{j=0}^{m-1} (D^s \log Df^{q_{n-1}} \circ f^{jq_{n-1}})(Df^{jq_{n-1}})^s \tag{4.20} \\ &+ \sum_{l=1}^{s-1} \sum_{j=1}^{m-1} (D^{s-l} \log Df^{q_{n-1}} \circ f^{jq_{n-1}})(Df^{jq_{n-1}})^{s-1} \\ &\times P_l(\dots, D^l \log Df^{jq_{n-1}}, \dots) \\ &+ \sum_{l=1}^{s-1} (D^{s-l} \log Df^w \circ f^{mq_{n-1}})(Df^{mq_{n-1}})^{s-1} P_l(\dots, D^l \log Df^{mq_{n-1}}, \dots) \end{aligned}$$

and using (4.14)

$$\begin{aligned} \tilde{K}_n^s &\leq (\tilde{K}_{n-1}^s \eta_{n-1,n}^s) \circ f^{mq_{n-1}} + \sum_{j=0}^{m-1} (\eta_{n-1,n}^s K_{n-1}^s) \circ f^{jq_{n-1}} \\ &+ \sum_{l=1}^{s-1} \sum_{j=0}^{m-1} (\eta_{n-1,n}^{s-l} K_{n-1}^{s-l}) \circ f^{jq_{n-1}} P_l(\dots, \eta_n^i | D^i \log Df^{jq_{n-1}} |, \dots) \\ &+ \sum_{l=1}^{s-1} (\eta_{n-1,n}^{s-l} \tilde{K}_{n-1}^{s-l}) \circ f^{mq_{n-1}} P_l(\dots, \eta_n^i | D^i \log Df^{mq_{n-1}} |, \dots). \end{aligned} \tag{4.21}$$

Notice that  $\tilde{K}_n^s$  and  $\tilde{K}_{n-1}^{s-l}$  in (4.21) are related functions:

$$\begin{aligned} \tilde{K}_n^s &= \eta_n^s D^s \log Df^k, \\ \tilde{K}_{n-1}^{s-l} &= \eta_{n-1}^{s-l} D^{s-l} \log Df^w \end{aligned}$$

and  $k = mq_{n-1} + w$ . We can replace the various functions by their norms, use (4.17), and obtain an inequality analogous to (4.21):

PROPOSITION 4.2.

$$\tilde{K}_n^s \leq \sum_{\substack{n \ d+\sum c_j=s \\ 0 \leq d \leq s}} (\prod \mathbf{K}_{n-1}^{c_j} \eta_{n-1,n}^{c_j}) \tilde{K}_{n-1}^d \eta_{n-1,n}^d. \tag{4.22}$$

Recall that the right hand side of (4.21) means ‘a sum of terms’ of the form

$$(\prod \mathbf{K}_{n-1}^{c_j} \eta_{n-1,n}^{c_j}) \tilde{K}_{n-1}^d \eta_{n-1,n}^d$$

the number of terms depends only on  $s$  and the integer parameters  $c_j$  and  $d$  are non-negative integers satisfying the conditions appearing under the summation sign.

We can also keep the terms involving  $\tilde{K}_{n-1}^s$  and

$K_{n-1}^s$  in (4.21) as they are and apply (4.17) to the rest to obtain:

$$\begin{aligned} \tilde{K}_n^s &\leq (\tilde{K}_{n-1}^s \eta_{n-1,n}^s) \circ f^{mq_{n-1}} + \sum_{j=0}^{m-1} (\eta_{n-1,n}^s K_{n-1}^s) \circ f^{jq_{n-1}} \\ &+ \sum_{\substack{d+\sum c_j=s \\ 0 \leq d, c_j < s}} (\prod \mathbf{K}_{n-1}^{c_j} \eta_{n-1,n}^{c_j-1}) \tilde{K}_{n-1}^d \eta_{n-1,n}^d. \end{aligned} \tag{4.23}$$

PROPOSITION 4.2'. Assume that  $k = mq_{n-1} + v \leq q_n$  and that  $t, \tau$  are  $q_n$ -neighbors, then,

$$\left\{ \begin{aligned} &|\Delta_n \tilde{K}_n^r = \eta_n^r(t) | D^r \log Df^k(t) - D^r \log Df^k(\tau) | \\ &\leq (\eta_n^r \Delta_n D^r \log Df^v) \circ f^{mq_{n-1}} + \sum (\eta_n^r \Delta_n D^r \log Df^{q_{n-1}}) \circ f^{jq_{n-1}} \\ &+ \sum_{\substack{d+\sum c_j=r+1 \\ 0 \leq d, c_j \leq r}} (\prod \mathbf{K}_{n-1}^{c_j} \eta_{n-1,n}^{c_j-1}) \tilde{K}_{n-1}^d \eta_{n-1,n}^d. \end{aligned} \right. \tag{4.24}$$

*Proof.* We estimate  $\Delta_n \tilde{K}_n^r$  by expressing  $D^r \log Df^w$ , both at  $t$  and at  $\tau$ , by (4.20), and subtracting the corresponding terms. We keep the leading terms as differences

and bound the other differences by the mean value theorem, remembering that

$$|f^j(t) - f^j(\tau)| \leq \eta_n \circ f^j(t).$$

After multiplying by  $\eta_n^r$  we obtain an expression that differs from the one expressing  $\tilde{K}_n^{r+1}$ , (if the additional derivative exists), only in that the terms  $\eta_n K_{n-1}^{r+1}$  and  $\eta_n \tilde{K}_{n-1}^{r+1}$  are replaced by the corresponding  $\Delta K^r$ . □

We shall refer to the sum of the terms of order  $s$  as it appears in (4.23) or of order  $r$  in (4.24) above, as *the leading term* (the linear term) in (4.23) or (4.24), respectively. Notice that (4.23) and (4.24) are point estimates; the left-hand side and the leading term are functions, the other terms are constants. Notice also that  $\sum c_j > 0$  in each term of the last row of either (4.23) or (4.24).

Next comes a ‘temporary result’ which uses the formal results obtained in conjunction with Corollary 3.6 to give (at least) exponential decay for  $\tilde{K}_n^s$ . It is convenient to have, and will be superseded eventually.

**PROPOSITION 4.3.** *For  $1 \leq s \leq r$ , there exists  $\xi_s, 0 < \xi_s < 1$ , such that*

$$\tilde{K}_n^s = O(\xi_s^n) \quad \text{as } n \rightarrow \infty. \tag{4.25}$$

*Proof.* For  $s = 1$  we have the result in Corollary 3.6. We now use induction on  $s$ . For appropriate  $t$  and  $m < q_n$  we have, by (4.12)

$$\begin{aligned} \tilde{K}_n^s &= |\eta_n^s(t) D^s \log Df^m(t)| \\ &\leq \sum_n^{s-1} \sum_{j=0}^{m-1} (\eta_n^{s-1} D^{s-1} \log Df) \circ f^j(t) P_l(\dots \eta_n^i(t) D^i \log Df^j(t), \dots) \\ &\leq C \sum_n^{s-1} \sum_{j=0}^{m-1} |\eta_n^{s-1} \circ f^j(t) P_l(\dots \eta_n^i(t) D^i \log Df^j(t), \dots)|. \end{aligned} \tag{4.26}$$

We use two estimates

- (a) As  $\sum_{j=0}^m \eta_n \circ f^j \leq 1$ , we have  $\sum_{j=0}^m \eta_n^{s-1} \circ f^j \leq \eta_n^{s-1}$ .
- (b) by the induction hypothesis, for  $l \geq 1$ ,  $P_l(\dots, \eta_n^i D^i \log Df^j, \dots) = O(\xi^n)$  for some  $\xi < 1$ .

The sum (over  $j$ ) for  $l = 0$  in (4.26) is bounded by  $\eta_n^{s-1}$ , exponentially small since  $s > 1$ . The other sums are exponentially small because of (b). □

The next step of the proof (through Lemma 4.7) provides us with the second leg discussed in the outline earlier. We estimate  $K_n^r$  at any point by comparing it, formally, to its value at a point where it vanishes (such points always exist for derivatives of periodic functions). The notation is:

- $t$  - an arbitrary point
- $t_s$  - a point for which  $D^s \log Df^{q_n}(t_s) = 0$
- $\tau$  - a  $q_n$ -neighbor of  $t$  such that for some  $m < q_n$ ,  $t_s = f^m(\tau)$ .

We apply (4.11) and write  $D^s \log Df^{q_n+m}(\tau)$  in two different ways: setting  $h = f^m$ ,  $g = f^{q_n}$  (4.11) becomes

$$\begin{aligned} D^s \log Df^{q_n+m}(\tau) &= D^s \log Df^m(\tau) \\ &+ \sum_{l=1}^{s-1} D^{s-l} \log Df^{q_n}(t_s) (Df^m(\tau))^{s-l} P_l(\dots, D^l \log Df^m(\tau), \dots) \end{aligned} \tag{4.27}$$

while writing  $h = f^{q_n}$ ,  $g = f^m$  in (4.27) gives:

$$\begin{aligned}
 D^s \log Df^{q_n+m}(\tau) &= (D^s \log Df^m \circ f^{q_n}(\tau))(Df^{q_n}(\tau))^s + D^s \log Df^{q_n}(\tau) \\
 &\quad + \sum_{l=1}^{s-1} (D^{s-l} \log Df^m \circ f^{q_n}(\tau)) \\
 &\quad \times (Df^{q_n}(\tau))^{s-l} P_l(\dots, D^l \log Df^{q_n}(\tau), \dots).
 \end{aligned}
 \tag{4.28}$$

Comparing (4.27) and (4.28) we obtain

$$\begin{aligned}
 D^s \log Df^{q_n}(\tau) &= D^s \log Df^m(\tau) - (D^s \log Df^m \circ f^{q_n}(\tau))(Df^{q_n}(\tau))^s \\
 &\quad + \sum_{l=1}^{s-1} D^{s-l} \log Df^{q_n}(t_s)(Df^m(\tau))^{s-l} P_l(\dots, D^l \log Df^m, \dots) \\
 &\quad - \sum_{l=1}^{s-1} (D^{s-l} \log Df^m \circ f^m \circ f^{q_n}(\tau)) \\
 &\quad \times (Df^{q_n}(\tau))^{s-l} P_l(\dots, D^l \log Df^{q_n}, \dots)
 \end{aligned}
 \tag{4.29}$$

and estimate  $D^s \log Df^{q_n}(t)$  by

$$D^s \log Df^{q_n}(t) - D^s \log Df^{q_n}(\tau) + \text{the right hand side of (4.29)}. \tag{4.29}'$$

The following inequalities are obvious.

$$|(Df^{q_n}(\tau))^s - 1| \leq s \mathbf{K}_n^0, \tag{4.30}$$

$$|f^{q_n}(\tau) - \tau| = \eta_{n+1}(\tau) \leq \eta_n \boldsymbol{\eta}_{n,n+1}, \tag{4.31}$$

$$|D^s \log Df^{q_n}(t) - D^s \log Df^{q_n}(\tau)| \leq \eta_n(\tau) \|D^{s+1} \log Df^{q_n}\|_\infty, \tag{4.32}$$

$$|D^s \log Df^m(\tau) - D^s \log Df^m \circ f^{q_n}(\tau)| \leq \eta_{n+1}(\tau) \|D^{s+1} \log Df^m\|_\infty. \tag{4.33}$$

Multiplying (4.29)' by  $\eta_n^s (= \eta_n^s(t) \sim_n \eta_n^s(\tau))$  and using (4.30), (4.31), and (4.32) we have proved

LEMMA 4.4. For  $1 \leq s < r$

$$\mathbf{K}_n^s \leq_n \begin{cases} \mathbf{K}_n^{s+1} + \boldsymbol{\eta}_{n,n+1} \tilde{\mathbf{K}}_n^{s+1} + s \mathbf{K}_n^0 \tilde{\mathbf{K}}_n^s \\ + \sum_{l=1}^{s-1} \mathbf{K}_n^{s-l} P_l(\dots, \tilde{\mathbf{K}}_n^j, \dots) \\ + \sum_{l=1}^{s-1} \tilde{\mathbf{K}}_n^{s-l} P_l(\dots, \mathbf{K}_n^j, \dots) \end{cases} \tag{4.34}_s$$

and

$$\mathbf{K}_n^r \leq_n \begin{cases} -\Delta_n \mathbf{K}_n^r(t, \tau) + \Delta_{n+1} \tilde{\mathbf{K}}_n^r(\cdot, \tau, f^{q_n}(\tau)) + r \mathbf{K}_n^0 \tilde{\mathbf{K}}_n^r \\ + \sum_{l=1}^{s-1} \mathbf{K}_n^{r-l} P_l(\dots, \tilde{\mathbf{K}}_n^j \dots) \\ + \sum_{l=1}^{s-1} \tilde{\mathbf{K}}_n^{r-l} P_l(\dots, \mathbf{K}_n^j \dots). \end{cases} \tag{4.34}_r$$

Notice that the inequalities (4.34)<sub>s</sub> deal with constants while (4.34)<sub>r</sub> is an inequality between functions. There is no intrinsic difference between the case  $s < r$  and  $r$ ; we write the inequalities simply in the form that we are going to use them later.

Denjoy's inequality is obtained from the case  $s = 0$  which we write separately:

$$\mathbf{K}_n^0 \leq \mathbf{K}_n^1 + \eta_{n,n+1} \tilde{\mathbf{K}}_n^1. \tag{4.35}$$

For  $s = 1$  (4.34) has only the first row on the right side and using (4.35) we obtain

$$\mathbf{K}_n^1 \leq \mathbf{K}_n^1 \tilde{\mathbf{K}}_n^1 + \eta_{n,n+1} (\tilde{\mathbf{K}}_n^2 + (\tilde{\mathbf{K}}_n^1)^2) + \mathbf{K}_n^2. \tag{4.36}$$

Since  $\tilde{\mathbf{K}}_n^1$  is exponentially small (Proposition 4.3) we can ignore the term  $\mathbf{K}_n^1 \tilde{\mathbf{K}}_n^1$  and (4.36) can be written

$$\mathbf{K}_n^1 \leq \eta_{n,n+1} P_2(\dots, \tilde{\mathbf{K}}_n^j, \dots) + \mathbf{K}_n^2, \tag{4.36'}$$

which gives also

$$\mathbf{K}_n^0 \leq \eta_{n,n+1} [\tilde{\mathbf{K}}_n^1 + P_2(\dots, \tilde{\mathbf{K}}_n^j, \dots)] + \mathbf{K}_n^2. \tag{4.35'}$$

These are special cases of

PROPOSITION 4.5. For  $0 \leq s < s_1 \leq r$

$$\mathbf{K}_n^s \leq \eta_{n,n+1} \sum_{l \geq s+1} P_l(\dots, \tilde{\mathbf{K}}_n^j, \dots) + \mathbf{K}_n^{s_1}. \tag{4.37}$$

*Proof.* Assume (4.37) for some  $s_1 < r$  and all  $s < s_1$ . By (4.34)<sub>s\_1</sub>

$$\begin{aligned} \mathbf{K}_n^{s_1} &\leq \mathbf{K}_n^{s_1+1} + \eta_{n,n+1} \tilde{\mathbf{K}}_n^{s_1+1} + s_1 \mathbf{K}_n^0 \tilde{\mathbf{K}}_n^{s_1} \\ &\quad + \sum_{l=1}^{s_1-1} \mathbf{K}_n^{s_1-l} P_l(\dots, \tilde{\mathbf{K}}_n^j \text{ dots}) \\ &\quad + \sum_{l=1}^{s_1-1} \tilde{\mathbf{K}}_n^{s_1-l} P_l(\dots, \mathbf{K}_n^j, \dots) \end{aligned}$$

by (4.37) replace  $\mathbf{K}_n^j, 0 \leq j < s_1$  by  $\mathbf{K}_n^{s_1} + \eta_{n,n+1}(\dots)$  and open brackets. The terms  $\mathbf{K}_n^{s_1}$  are all multiplied by some  $\tilde{\mathbf{K}}_n^j$  and can be collected and moved to the left which becomes  $\mathbf{K}_n^{s_1} \cdot (1 - \delta_n)$ , with  $\delta_n$  exponentially small.

So that

$$\mathbf{K}_n^{s_1} \leq \mathbf{K}_n^{s_1+1} + \eta_{n,n+1} (\tilde{\mathbf{K}}_n^{s_1+1} + P(\tilde{\mathbf{K}}_n^j)), \tag{4.38}$$

where  $P$  is a polynomial in  $\tilde{\mathbf{K}}_n^j, j \leq s_1$ , all whose terms have weight  $> s_1$ , in particular all the terms are products. Combine (4.38) with (4.37) to obtain

$$\mathbf{K}_n^s \leq \eta_{n,n+1} \left[ \sum_{s+1}^{s_1} P_l(\tilde{\mathbf{K}}_n^j) + P(\tilde{\mathbf{K}}_n^j) + \tilde{\mathbf{K}}_n^{s_1+1} \right] + \mathbf{K}_n^{s_1+1}$$

which is (4.37) for  $s_1 + 1$ . □

LEMMA 4.6. For appropriate  $t, \tau$  which is a  $q_n$ -neighbor of  $t$ , and  $0 \leq m < q_n$  we have

$$\mathbf{K}_n^r \leq \Delta_n \mathbf{K}_n^r(t, \tau) + \Delta_{n+1} \tilde{\mathbf{K}}_n^r(m, \tau, f^{q_n}(t)) + \eta_{n,n+1} P(\tilde{\mathbf{K}}_n^j, 1 \leq j \leq r, \text{wt} \geq r+1),$$

where 'wt  $\geq r + 1$ ' means that the weight of every term in  $P$  is  $\geq r + 1$ .

*Proof.* Write (4.34)<sub>r</sub> for  $t$  which maximizes  $K_n^r$ , apply Proposition 4.5 with  $s_1 = r$  to all the terms  $\mathbf{K}_n^s$  appearing there and move the (exponentially small) multiple of  $\mathbf{K}_n^r$  obtained on the right-hand side to the left. □

The functions  $\eta_n$  and of  $\eta_{n,n+1}$  and their norms  $\mathfrak{n}_n$  and  $\mathfrak{n}_{n,n+1}$  are the only geometric element in our discussion, and the motivation for the introduction of the formal  $\bar{\eta}_{n,n+1}$  in (4.9) and  $\bar{\eta}_n$  in (4.10) should become clearer in the next few pages. The formal relation comes from Lemma 1.8 and we make it more precise now. The two elements in which the following differs from Lemma 1.8 (considering the definition (4.9)) are that  $\mathbf{K}_n^0$  is replaced by a (large multiple of)  $\mathbf{K}_n^r$ , and that  $a_{n,n+1}^{-1}$  is replaced by  $\alpha_{n,n+1}$ . The reason for the pettiness is (Definition (4.10)) the fact that a constant factor for  $\bar{\eta}_{n,n+1}$  becomes an exponential factor for  $\bar{\eta}_n$  and we need to be able to bring it to arbitrarily slow exponential growth or lose† some of the smoothness of  $h$

LEMMA 4.7. *For  $n$  sufficiently large, and writing  $\varepsilon = 2M_r^{-1}$  (see (4.9)), we have*

$$\mathfrak{n}_{n,n+1} \leq (1 + \varepsilon) \bar{\eta}_{n,n+1}.$$

*Proof.* We repeat the proof of Lemma 1.8. A  $q_n$ -interval splits into  $a_{n,n+1}q_{n+1}$ -intervals and a single  $q_{n+2}$ -interval. The ratio between two adjacent  $q_{n+1}$ -intervals is bounded by  $\exp(\mathbf{K}_n^0)$  and if  $a_{n,n+1}\mathbf{K}_n^0 < M^{-1}$  with  $M > 1$  the intervals are all equal within a factor  $\exp M^{-1}$ . This gives a little better than

$$\mathfrak{n}_{n,n+1} \leq (1 + M^{-1}) \max(a_{n,n+1}^{-1}, M\mathbf{K}_n^0)$$

and by (4.37) (with  $s = 0$  and  $s_1 = r$ ) we may replace  $\mathbf{K}_n^0$  by  $\mathbf{K}_n^r$ , the additional summand being an exponentially small multiple of  $\mathfrak{n}_{n,n+1}$ .

To understand why we can replace  $a_{n,n+1}^{-1}$  by  $\alpha_n$  (with a correction factor  $1 + \varepsilon$ ) we notice that the ratio of the two is bounded by  $1 + (a_{n,n+1}a_{n+1,n+2})^{-1}$  and if  $a_{n,n+1}a_{n+1,n+2} \gg \varepsilon^{-1}$ , the replacement is clearly valid. If the coefficients  $a_{n,n+1}, a_{n+1,n+2}$  are small, we recall that for the rigid rotation  $R_\alpha$  we have  $\eta_{n,n+1} = \alpha_n$  and one can get an estimate (to within a factor  $(1 + \varepsilon/2)$ ) of the ratio of the lengths of a  $q_{n-1}$  and a  $q_n$  intervals for  $R_\alpha$  by counting how many disjoint  $q_{n+k}$ -intervals each of them contains, where  $k$  is the largest integer for which  $q_{n+k} < \varepsilon^{-5}q_n$ . Since  $\mathbf{K}_n^0$  is exponentially small all these intervals (for our picture, with powers of  $f$ ) are equal enough to give our estimate. □

*Remark.* If either  $\alpha_n$  or  $M_r K_n^r$  is substantially bigger than the other, the factor  $1 + \varepsilon$  can be brought substantially closer to 1.

We now come to the ‘*Pièce de résistance*’ of the entire paper, namely the estimates of  $\mathbf{K}_n^r$  and related quantities.

*Notation.* We denote  $\mathbf{T}_n^L = \sup \prod \tilde{\mathbf{K}}_n^l$  the supremum conditioned by  $l_j \leq r, \sum l_j = L$ .

† We are dealing now with rotation numbers whose coefficients may grow at superexponential rate but don't have to. If the growth is consistently superexponential the exponential factor will have no effect, but if we have long stretches of small coefficients the estimates may really suffer if the exponent is not arbitrarily small (multiple of  $n$ ).

**THEOREM 4.8.** Assume  $f \in \mathcal{H}^{r+1+\delta}$ ,  $0 < \delta < 1$ . There exist positive constants  $M$  and  $\xi$ ,  $\xi < 1$ , such that if  $M_r > M$  and we set  $\varepsilon = 2rM_r^{-1}$ , we have for appropriate constants  $\lambda_1, \lambda_2, \lambda_3$ , for  $r + 1 \leq L \leq 2r$  and all  $n$ :

$$\mathbf{K}_n^r \leq \lambda_1(1 + \varepsilon)^n \bar{\eta}_n^{r-1+\delta}, \tag{A}_n$$

$$\Delta_n \tilde{\mathbf{K}}_n^r \leq \lambda_2(1 + \varepsilon)^n \bar{\eta}_n^{r-1+\delta}, \tag{B}_n$$

$$\mathbf{T}_n^L \leq \lambda_3 \xi^n \bar{\eta}_n^{r-1+\delta}. \tag{C}_n$$

*Remark.* Estimate  $(C)_n$  for  $L \leq 2r$  implies exponentially better estimates for bigger values of  $L$  since we can factor out some of the factors  $\tilde{\mathbf{K}}_n^L$ , which are exponentially small, and still have the weight of the rest bigger than  $r$ .

*Remark.* We get Theorem 4.8 by repeatedly ‘breaking into shorter runs’ Propositions 4.2 and 4.2’ and applying our ‘basic procedure’ Propositions 4.5 and 4.6. The main difficulties are the loss of an  $\eta_{n-1,n}$  associated with the  $K_{n-1}$  (e.g., (4.40)) and the growing number of terms. The loss of the  $\eta_{n-1,n}$  is handled by using the ‘basic procedure’ either regaining the lost  $\eta_{n-1,n}$  or getting an  $K_{n-1}^{(r)}$  that can be replaced by an  $\bar{\eta}_{n-1,n}$  (e.g., (4.43)). The growth in the number of terms is compensated for by certain exponentially small factors. In some cases we get a higher power of  $\eta$  than we need. In other cases, the terms at the next (lower) level have more weight and this can happen at most  $r$  times before we can split off an exponentially small factor.

*Proof.* The constants  $M_r$  and  $\xi$  are determined as follows: there are various values  $\xi_s$  given by Proposition 4.3 and also a bound  $\eta < 1$  for  $\eta_{n-1,n}$  and  $\xi$  can be taken as the biggest of these, raised to the power  $\frac{1}{2}(1 - \delta)$  and then multiplied by  $1 + \varepsilon$ . It is less than 1 provided  $\varepsilon$  is small enough. This, with Lemma 4.7, gives (4.39)<sub>a</sub> below. In order to simplify the notation later we also impose  $(1 + \varepsilon)^r < \eta$ ; this defines the lower bound  $M$  for  $M_r$ ; (we may choose it much bigger if we want to lower the bound on  $\varepsilon$ , but we should notice that this affects the definition† of  $\bar{\eta}_{n-1,n}$ ).

The estimates are needed for large  $n$ , and we assume that  $n > n_0$  where  $n_0$  is large enough so that the various estimates which we import, and which we know to be true for large  $n$ , are in fact valid for  $n \geq n_0$ . Also, we obtain in the course of the proof various polynomials (in  $n$ ) with large coefficients and we want  $\xi^n$ ,  $n \geq n_0$ , to be small enough to neutralize them.

With these constraints on the constants we have:

$$\eta_{n-1,n}^{r+1} \leq \eta_{n-1,n}^r \leq \xi^2 \bar{\eta}_{n-1,n}^{r-1+\delta}, \tag{4.39}_a$$

$$\tilde{\mathbf{K}}_n^s \leq \xi^{2n} (1 + \varepsilon)^{-2n}. \tag{4.39}_b$$

Define:

$$\lambda_1^{(n)} = \mathbf{K}_n^r (1 + \varepsilon)^{-n} \bar{\eta}_n^{1-r-\delta}, \tag{A}_n^*$$

$$\lambda_2^{(n)} = \sup \Delta_n \tilde{\mathbf{K}}_n^r (1 + \varepsilon)^{-n} \bar{\eta}_n^{1-r-\delta}, \tag{B}_n^*$$

† For the  $\alpha$ 's that will interest us later, the value of  $\bar{\eta}_{n-1,n}$  will be determined, for all large  $n$ , by  $\alpha_n$  rather than  $M_r \mathbf{K}_n^r$ . Here we make no assumption on  $\alpha$  (except for being irrational).



the supremum for all legitimate values of  $\Delta_n \tilde{\mathbf{K}}_n^r$

$$\lambda_3^{(n)} = \max_{1 \leq L \leq 2r} \mathbf{T}_n^L \xi^{-n} \bar{\eta}_n^{1-r-\delta}. \tag{C)*}$$

We shall prove that  $\lambda_j^{(n)}$  are bounded and that for large values of  $n$  we have  $\lambda_j^{(n)} \leq \lambda_j$  with  $\lambda_2 = \|D^r \log Df\|_{\text{Lip}} \delta + 1$ ,  $\lambda_1 = 3\lambda_2$  and  $\lambda_3$  - some absolute multiple of  $\lambda_1$ . We use a pseudo-induction procedure, i.e. evaluate  $\lambda_j^{(n)}$  in terms of  $\lambda_j^{(m)}$ ,  $m < n$ ,

We have two tools which we combine, the ‘Basic procedure’ embodied in Lemmas 4.5 and 4.6, and the ‘break into shorter runs’ embodied in formulas (4.21)–(4.24). One important thing to remember is that the expressions in (4.22)–(4.24) are universal polynomials, independent of  $n$  and of the rotation number (which maintain a ‘low profile presence’, however, through the loss of the factor  $\eta_{n-1,n}$ , loss which replaced, (4.17), the summation over the iterates  $f^{jq_{n-1}}$ ).

A: 
$$\lambda_1^{(n)} \leq 2\lambda_2^{(n)} + \mathcal{O}(\xi^n) \lambda_3^{(n)}.$$

*Proof.* This follows from Lemma 4.6. □

We now estimate  $\lambda_3^{(n)}$ , leaving  $\lambda_2^{(n)}$  for last. In the formulas below  $C$  denotes some absolute constant, and we write  $\lambda_{1,3}^{(n-1)} = \max(\lambda_1^{(n-1)}, \lambda_3^{(n-1)})$ .

C: 
$$\lambda_3^{(n)} \leq Cn^r \xi^{n/2} \left[ \max_{n/2 \leq m < n} \lambda_{1,3}^{(m)} \right].$$

*Proof.* We estimate  $\mathbf{T}_n^L = \prod \tilde{\mathbf{K}}_n^{l_j}$  ( $l_j \leq r, \sum l_j = L \geq r + 1$ ) by applying (4.22) to each of the factors and multiplying the corresponding expressions

$$\mathbf{T}_n^L \leq \sum_{\sum c_j + \sum d_k = L} \left( \prod \mathbf{K}_{n-1}^{c_j} \eta_{n-1,n}^{c_j-1} \right) \left( \prod \tilde{\mathbf{K}}_{n-1}^{d_k} \eta_{n-1,n}^{d_k} \right). \tag{4.40}$$

In particular, there is a single term in (4.40) for which  $\sum c_j = 0$ , namely the term  $\prod \tilde{\mathbf{K}}_{n-1}^{l_j} \eta_{n-1,n}^{l_j}$ , which is a replica of the original expression; and the number of the other summands is bounded by some constant which depends only on  $L$  (recall that in (4.22) we have universal polynomials so that the exact form of (4.40) is completely determined by  $\{l_j\}$  of which there is a finite number with  $\sum l_j = L$ , and  $L$  limited to  $[r + 1, 2r]$ ).

In each of these (other) summands we have at least one factor  $\mathbf{K}_{n-1}^{c_j}$ ,  $c_j \geq 1$ , and we apply Lemma 4.4, with  $s_1 = r$ , to every such factor with  $c_j < r$ ,

$$\mathbf{K}_{n-1}^{c_j} \leq \mathbf{K}_{n-1}^r + \eta_{n-1,n} P(\tilde{\mathbf{K}}_{n-1}^i; 1 \leq i \leq r, \text{wt} \geq c_j + 1)$$

and obtain a bound for  $\mathbf{T}_n^L$  by a sum of terms of the form

$$(\mathbf{K}_{n-1}^r)^b \prod \tilde{\mathbf{K}}_{n-1}^{v_i} \eta_{n-1,n}^{\sum l_j - b} \tag{4.41}$$

whose number is bounded by some universal constant which we denote by  $N_r$ ; with  $0 \leq b$  bounded by the number of factors, (which is certainly bounded by  $L$ ).

We now evaluate separately the terms (4.41) for which  $b = 0$  and those for which  $b \geq 1$ .

The terms for which  $b = 0$  all have weight  $\geq L + 1$  and the exponent on  $\eta_{n-1,n}$  is at least  $L$ . The sum of the one ‘replica’ pulled out of (4.40) and the terms with  $b = 0$

and weight  $\leq 2r$  in (4.41), (with the factors  $\eta_{n-1,n}$  in excess of  $L$  removed), namely

$$\eta_{n-1,n}^L \left[ \prod \tilde{\mathbf{K}}_{n-1}^{l_i} + \sum_{\substack{d_j \leq r \\ 2r \geq \sum d_j \geq L+1}} (\prod \tilde{\mathbf{K}}_{n-1}^{d_j}) \right] \tag{4.42}$$

will be referred to as the *first generation*, and is the part that we'll have to develop further.

The rest, that is, the terms in (4.41) for which  $b \geq 0$  as well as the terms whose weight exceeds  $2r$ , we estimate right here and now.

If  $b = 1$  we have the term  $\mathbf{K}_{n-1}^r \prod \tilde{\mathbf{K}}_{n-1}^{v_j} \eta_{n-1,n}^r$  with  $\sum v_j \geq 1$  which we can estimate (using  $(A)_{n-1}$  and  $(4.39)_b$ ) by

$$\xi^{2n} \lambda_1^{(n-1)} \bar{\eta}_n^{r-1+\delta}. \tag{4.43}$$

If  $b \geq 2$  we replace  $b - 2 + \delta$  of the factors  $\mathbf{K}_{n-1}^r$  by  $(M_r^{-1} \bar{\eta}_{n-1,n})^{b-2+\delta}$ , (see (4.9)), estimate one factor  $\mathbf{K}_{n-1}^r$  using  $(A)_{n-1}$ , and the remaining  $\mathbf{K}_{n-1}^{1-\delta}$  by (4.40) to obtain again the bound (4.43). We have several such terms (the number bounded by  $N_r$ ). Thus the sum of all the terms (4.41) with  $b \geq 1$  is bounded by  $N_r$  times the estimate (4.43).

Finally, terms whose weight exceeds  $2r$  can be written as  $\eta_{n-1,n}^L \tilde{\mathbf{K}}_{n-1}^d \prod_{c_j \leq r, \sum c_j \geq r+1} \tilde{\mathbf{K}}_{n-1}^{c_j}$  and can be estimated as follows: use (4.40) to replace  $\eta_{n-1,n}^{r+1}$  by  $\xi^{2-r-1+\delta}$  and  $\tilde{\mathbf{K}}_{n-1}^d$  by  $\xi^{n-1}$ . Replace the remaining product by  $\lambda_3^{(n-1)} \xi^{n-1} \bar{\eta}_{n-1}^{r-1+\delta}$ . This gives the estimate  $\lambda_3^{(n-1)} \xi^{2n} \bar{\eta}_{n-1}^{r-1+\delta}$  and their number is bounded by  $N_r$  (which bounds the total number of terms we obtain at this stage).

Putting all of this together, we obtain (writing  $\lambda_{1,3}^{(n-1)} = \max(\lambda_1^{(n-1)}, \lambda_3^{(n-1)})$ )

$$\mathbf{T}_n^L \leq \eta_{n-1,n}^L \left[ \prod \tilde{\mathbf{K}}_{n-1}^{l_i} + \sum_{\substack{d_j \leq r \\ \sum d_j \geq L+1}} (\prod \tilde{\mathbf{K}}_{n-1}^{d_j}) \right] + N_r \lambda_{1,3}^{(n-1)} \xi^{2n} \bar{\eta}_n^{r-1+\delta}. \tag{4.44}$$

Now we have to repeat the same procedure on every term in the first generation, i.e., the terms in the square brackets in (4.44). The first is just the replica of our original, (for  $n - 1$  instead of  $n$ ) and it's contribution to the next generation is a replica of (4.42). The other terms are all of weight  $\geq L + 1$  and the contribution of each to the next generation consists in a replica of itself, and at most  $N^r$  additional terms, all of weight  $\geq L + 2$  and  $\leq 2r$ . Each of these also spins off other terms, the ones with  $b > 0$  above, and those with weight  $\geq 2r + 1$ , which are estimated directly as in (4.43), in other words:

The term  $\eta_{n-1,n}^L \prod \tilde{\mathbf{K}}_{n-1}^{l_i}$  is estimated by:

$$\eta_{n-1,n}^L \eta_{n-2,n-1}^L \left[ \prod \tilde{\mathbf{K}}_{n-2}^{l_i} + \sum_{\substack{d_j \leq r \\ 2r \geq \sum d_j \geq L+1}} (\prod \tilde{\mathbf{K}}_{n-2}^{d_j}) \right] + \eta_{n-1,n}^L N_r \lambda_{1,3}^{(n-2)} \xi^{2n-2} \bar{\eta}_{n-1}^{r-1+\delta} \tag{4.45}$$

and similarly each  $\eta_{n-1,n}^L \prod \tilde{\mathbf{K}}_{n-1}^{d_j}$  to:

$$\eta_{n-1,n}^L \eta_{n-2,n-1}^{L+1} \left[ \prod \tilde{\mathbf{K}}_{n-2}^{d_j} + \sum_{\substack{f_j \leq r \\ 2r \geq \sum f_j \geq L+2}} (\prod \tilde{\mathbf{K}}_{n-2}^{f_j}) \right] + \eta_{n-1,n}^L N_r \lambda_{1,3}^{(n-2)} \xi^{2n-2} \bar{\eta}_{n-1}^{r-1+\delta} \tag{4.45}^*$$

and putting them all together we have

$$\mathbf{T}_n^L \leq \boldsymbol{\eta}_{n-1,n}^L \boldsymbol{\eta}_{n-2,n-1}^L \sum_{s=0}^2 Q_{n-2}^{(s)} + N_r \lambda_{1,3}^{(n-1)} \xi^{2n} \bar{\eta}_{n-1}^{r-1+\delta} + N_r^2 \lambda_{1,3}^{(n-2)} \xi^{2n-2} \boldsymbol{\eta}_{n-1,n}^{r+1} \bar{\eta}_{n-s}^{r-1+\delta},$$

where  $Q_{n-2}^{(0)} = \prod \tilde{\mathbf{K}}_{n-j}^l$  is a single term,

$$Q_{n-2}^{(1)} = 2 \sum_{\substack{d_j \leq r \\ 2r \geq \sum d_j \geq L+1}} (\prod \tilde{\mathbf{K}}_{n-j}^{d_j})$$

is the ‘new terms in the first generation’, each comes now twice, being reproduced both by itself and by the original term of order  $L$ . We also have the ‘new’ terms of weight  $\geq L+2$ , whose number is bounded by  $N_r^2$ . This is the second generation.

Repeating  $J$  times and being somewhat wasteful in our estimates, we obtain

$$\mathbf{T}_n^L \leq \boldsymbol{\eta}_{n-1,n}^L \cdots \boldsymbol{\eta}_{n-J,n-J+1}^L \sum_{s=0}^J Q_{n-J}^{(s)} + \sum_{s=1}^J N_r^s \lambda_{1,3}^{(n-s)} \xi^{2n-2s+2} \boldsymbol{\eta}_{n-1,n}^L \cdots \boldsymbol{\eta}_{n-s,n-s+1}^L \bar{\eta}_{n-s}^{r-1+\delta}, \tag{4.46}$$

where  $Q_{n-J}^{(0)} = \prod \tilde{\mathbf{K}}_{n-j}^l$  is a single term, and  $Q_{n-J}^{(1)} = J \sum_{d_j \leq r, 2r \geq \sum d_j \geq L+1} (\prod \tilde{\mathbf{K}}_{n-j}^{d_j})$  are as described above, the factor  $J$  in  $Q_{n-J}^{(1)}$  is there since with each iteration of the procedure not only do these terms reproduce themselves (with the index  $n$  reduced by one) but  $Q^{(0)}$  produces another copy. Similarly  $Q_{n-J}^{(2)}$  are the terms of weight  $\geq L+2$  (and  $\leq 2r$ ), which make their first appearance in the second generation, (4.45)\*; their number increases quadratically with  $J$ .

In general  $Q_{n-J}^{(s)}$  are the terms which appear first at the  $s$  iteration of our procedure as the terms (4.41) (for which  $b=0$ , and whose weight is between  $L+s$  and  $2r$ ) as well as their replications in the steps from  $s+1$  to  $J$ . Their number is bounded by  $J^s$  times their number at their first appearance, in  $Q_{n-s}^{(s)}$ . That number is no more than  $N_r$  times the number of terms in  $Q_{n-s+1}^{(s-1)}$  and induction on  $s$  shows it to be  $\leq N_r^s$ .

Since  $Q^{(s)}$  are identically zero for  $s \geq r$  (in fact for  $s > 2r - L$ , terms with weight exceeding  $2r$  are estimated directly), and after we exchange some  $\boldsymbol{\eta}$ 's for  $\xi$ 's, we have

$$\mathbf{T}_n^L \leq \boldsymbol{\eta}_{n-1,n}^L \cdots \boldsymbol{\eta}_{n-J,n-J+1}^L \sum_{s=0}^{r-1} Q_{n-J}^{(s)} + \left( \sum_{s=1}^J R_r^{(s)} \lambda_{1,3}^{(n-s)} \xi^n \right) \xi^n \bar{\eta}_n^{r-1+\delta}. \tag{4.47}$$

$R_r^{(s)}$  is the number of terms obtained at the  $s$ -iteration which are estimated directly rather than developed further, and is bounded by a constant multiple of the number of terms in  $\sum_{t=1}^r Q_{n-s+1}^{(t)}$  which is  $O(s^{r-1})$ . It replaces the factor  $N_r^s$  in (4.46) which is far too wasteful for large values of  $s$ .

We take  $J = [n/2]$ , evaluate the terms in  $\sum Q_{n-J}^{(s)}$  by  $(C)_{n-J}$ ; replace  $L$  by  $r+1$  and convert the factors  $\boldsymbol{\eta}_{m,m+1}^{r+1}$  into  $\xi^2 \bar{\eta}_{m,m+1}^{r-1+\delta}$ ; and obtain (with some absolute constant  $C$ ):

$$\mathbf{T}_n^L \leq C n^{r-1} \xi^{n/2} \lambda_3^{(J)} \cdot \xi^n \bar{\eta}_n^{r-1+\delta} + C n^r \xi^n \max_{n/2 \leq m < n} \lambda_{1,3}^{(m)} \xi^n \cdot \xi^n \bar{\eta}_n^{r-1+\delta}, \tag{4.48}$$

which we write, after combining the first two rows, as

$$\lambda_3^{(n)} \leq C n^r \xi^{n/2} \left[ \max_{n/2 \leq m < n} \lambda_{1,3}^{(m)} \right]. \tag{4.48*}$$

□

Finally we turn to estimate  $\lambda_2^{(n)}$ :

**B:** 
$$\lambda_2^{(n)} \leq \|D^r \log Df\|_{\text{Lip } \delta} + \xi^n \sum_{j=1}^n \lambda_3^{(j)}.$$

*Proof.* We estimate  $\Delta_n \tilde{K}_n^r$  by (4.24)

$$\begin{aligned} \Delta_n \tilde{K}_n^r &\leq (\eta_n^r \Delta_n D^r \log Df^v) \circ f^{mq_{n-1}} + \sum (\eta_n^r \Delta_n D^r \log Df^{q_{n-1}}) \\ &\quad \circ f^{jq_{n-1}} \sum (\prod \mathbf{K}_{n-1}^{c_i} \boldsymbol{\eta}_{n-1,n}^{c_i-1}) \tilde{\mathbf{K}}_{n-1}^d \boldsymbol{\eta}_{n-1,n}^d \end{aligned} \tag{4.24}$$

and the second row, the non-leading terms, have the form† of the right-hand side of (4.40) which we have just ((4.48)) shown to be bounded by  $\lambda_3^{(n)} \xi^n \bar{\eta}_n^{r-1+\delta}$ .

All that we need now is a good estimate for the leading term in (4.24). We evaluate each of the terms  $\Delta_n D^r \log Df^v$  and  $\Delta_n D^r \log Df^{q_{n-1}}$ , computed at their appropriate points, by the analogue to (4.24) (the difference is that we break the runs into  $q_{n-2}$ -runs) and obtain a new leading term

$$\sum (\eta_n^r \Delta_n D^r \log Df^{\mu_i}) \circ f^{\sigma_i}, \tag{4.49}$$

where  $\mu_j$  are mostly equal to  $q_{n-2}$  with a single  $q_{n-3}$  per each  $f^{q_{n-1}}$  in (4.24) and the last  $\mu$  being the remainder after dividing  $v$  by  $q_{n-2}$ . The  $\sigma_j$  are the starting points, in the orbit  $\{f^l(t)\}$ , of the runs  $\mu_j$ . Notice that every  $\sigma_j$  has the form  $aq_{n-1} + bq_{n-2}$  with  $a \leq a_{n-1,n}$ ,  $b \leq a_{n-2,n-1}$ , so that the points  $f^{\sigma_i}(t)$  all lie within one  $(n-2)$  interval.

We also get the additional terms

$$\eta_{n-1,n}^{r+1} \sum_{\substack{\sum c_i = r+1 \\ d \leq r, c_i \leq r}} (\prod \mathbf{K}_{n-2}^{c_i} \boldsymbol{\eta}_{n-2,n-1}^{c_i-1}) \tilde{\mathbf{K}}_{n-2}^d \boldsymbol{\eta}_{n-2,n-1}^d \tag{4.50}$$

for the term  $\eta_n^r \Delta_n D^r \log Df^v$  and

$$\eta_{n-1,n}^{r+1} \sum (\prod \mathbf{K}_{n-2}^{c_i} \boldsymbol{\eta}_{n-2,n-1}^{c_i-1}) \mathbf{K}_{n-3}^d \boldsymbol{\eta}_{n-3,n-2}^d \boldsymbol{\eta}_{n-2,n-1}^d \tag{4.51}$$

for each of the summands  $(\eta_n^r \Delta_n D^r \log Df^{q_{n-1}}) \circ f^{jq_{n-1}}$ . Using (4.17) we replace the sum of all the expressions (4.51) by  $\eta_{n-1,n}^{-1}$  times one of them so that the leading term of (4.24) is bounded by (4.49) plus

$$\eta_{n-1,n}^r \sum_{d+\sum c_i=r+1} (\prod \mathbf{K}_{n-2}^{c_i} \boldsymbol{\eta}_{n-2,n-1}^{c_i-1}) \tilde{\mathbf{K}}_{n-2}^d \boldsymbol{\eta}_{n-2,n-1}^d$$

which we can estimate as before (using (4.39)<sub>a</sub>) by

$$\lambda_3^{(n-1)} \xi^{n-1} \boldsymbol{\eta}_{n-1,n}^r \bar{\eta}_{n-1}^{r-1+\delta} \leq \lambda_3^{(n-1)} \xi^n \bar{\eta}_n^{r-1+\delta}. \tag{4.52}$$

Repeatedly reducing the length of the runs to  $q_{n-3}$ ,  $q_{n-4}$  etc. . . . we finally reach ‘runs’ of length 1, namely the leading term

$$\sum (\eta_n^r \Delta_n D^r \log D) \circ f^j$$

which is bounded by

$$\begin{aligned} \|D^r \log Df\|_{\text{Lip } \delta} \sum \eta_n^{r+\delta} \circ f^j &\leq \|D^r \log Df\|_{\text{Lip } \delta} \boldsymbol{\eta}_n^{r-1+\delta} \\ &\leq \|D^r \log Df\|_{\text{Lip } \delta} (1 + \varepsilon)^n \bar{\eta}_n^{r-1+\delta}. \end{aligned}$$

† Notice that  $\Delta_n \tilde{K}_n^r$  behaves formally like  $\tilde{K}_n^{r+1}$ , the only difference being in the leading terms. All the terms spun off in the course of the proof are well within the class estimated in C above.

The additional terms, evaluated along the way (e.g. (4.52)), add up to  $\sum_{j=1}^n \lambda_3^{(j)} \xi^n \bar{\eta}_n^{r-1+\delta}$  which gives

$$\lambda_2^{(n)} \leq \|D^r \log Df\|_{\text{Lip}\delta} + \xi^n \sum_{j=1}^n \lambda_3^{(j)}. \quad \square$$

In order to conclude the proof of theorem 4.8 we write  $\lambda^{(n)} = \max (\lambda_j^{(m)})_{j=1, 2, 3, m < n}$ ; and write the estimates **A**, **B** and **C** as:

$$\begin{aligned} \lambda_3^{(n)} &\leq \mathcal{O}(\xi^{n/3})\lambda^{(n)}, \\ \lambda_2^{(n)} &\leq \|D^r \log Df\|_{\text{Lip}\delta} + \mathcal{O}(\xi^{n/3})\lambda^{(n)}, \\ \lambda_1^{(n)} &\leq 2\|D^r \log Df\|_{\text{Lip}\delta} + \mathcal{O}(\xi^{n/3})\lambda^{(n)}, \end{aligned}$$

hence  $\lambda^{(n+1)} \leq \max (\lambda^{(n)}, 3\|D^r \log Df\|_{\text{Lip}\delta})$  provided  $n$  is large enough. This shows that  $\lambda^{(n)}$  is eventually constant and the proof is complete.  $\square$

*Remark.* We shall sometimes replace  $(A)_n$  by:

$$\mathbf{K}_n^r \leq \bar{\eta}_n^{r-1+\delta''} \tag{A}_n^{**}$$

for any  $\delta'' < \delta$  and  $n > n(\delta'')$ .

This can be interpreted and checked as follows: the factor  $\bar{\eta}_n^{\delta-\delta''}$  compensates for  $\lambda_1(1+\varepsilon)^n$ . As  $\bar{\eta}_n$  decreases at least exponentially fast, knowing  $\delta - \delta''$  enables us to choose  $M_r$  large enough and hence  $\varepsilon$  small enough to guarantee  $\lambda_1(1+\varepsilon)^n \bar{\eta}_n^{\delta-\delta''} \rightarrow 0$ . A little closer look enables us to obtain  $(A)_n^*$  without adjusting the  $\varepsilon$  (which also affects  $\bar{\eta}_n$ ) by noticing that the source of the factor  $(1+\varepsilon)^n$  is the conversion, in the estimate of the leading term in the bound for  $\lambda_2^{(n)}$ , of  $\eta_n^{r-1+\delta}$  into  $\bar{\eta}_n^{r-1+\delta}$ . This is done by writing  $\eta_n \leq \prod \eta_{i,j+1}$  and applying Lemma 4.7 to each factor and we have remarked after the proof of that lemma that the factor  $(1+\varepsilon)$  appears only if  $\alpha_j$  and  $\mathbf{K}_j^r$  are comparable. Thus the conversion factor (which we estimated by  $(1+\varepsilon)^n$ ) is in fact exponentially big in  $n$  only if  $\bar{\eta}_n \rightarrow 0$  superexponentially and it follows that for any  $\delta'' < \delta$ ,  $\bar{\eta}_n^{\delta-\delta''}$  is more than enough to compensate for the conversion.

The basic method used in the proof above yields other estimates which we shall find useful. The next lemma and Proposition 4.10 are examples.

LEMMA 4.9. Assume  $f \in \mathcal{H}^{r+1+\delta}$ ,  $0 < \delta < 1$ . For  $s \leq r$  and with  $\xi$  as above we have

$$\begin{aligned} \tilde{\mathbf{K}}_n^s &= \mathcal{O}(\xi^n) \bar{\eta}_n^{s-1}, \\ \prod_{\sum c_j \geq s} \tilde{\mathbf{K}}_n^{c_j} &= \mathcal{O}(\xi^n) \bar{\eta}_n^{s-1}. \end{aligned}$$

*Proof.* We can estimate  $\tilde{\mathbf{K}}_n^r$  by (4.23) and (4.17)

$$\begin{aligned} \tilde{\mathbf{K}}_n^r &\leq \tilde{\mathbf{K}}_{n-1}^r \eta_{n-1,n}^r + \mathbf{K}_{n-1}^r \eta_{n-1,n}^{r-1} \\ &\quad + \sum_{\substack{1 < d \leq r-1 \\ d + \sum c_j = r}} \eta_{n-1,n}^d \tilde{\mathbf{K}}_{n-1}^d \prod \mathbf{K}_{n-1}^{c_j} \eta_{n-1,n}^{c_j-1}, \end{aligned} \tag{4.53}$$

apply Proposition 4.5 to the terms  $\mathbf{K}_{n-1}^{c_j}$  and write the terms obtained as in (4.41):

$$(\mathbf{K}_{n-1}^r)^b \prod \tilde{\mathbf{K}}_{n-1}^{v_i} \eta_{n-1,n}^{r-b}$$

The terms with  $b=0$  have weight  $\geq r+1$  and can be estimated by  $(C)_{n-1}$  of Theorem 4.8. In the terms with  $b \geq 1$  replace  $b-1$  of the factors  $\mathbf{K}'_{n-1}$  by  $(A)_{n-1}$  of Theorem 4.8. Do the same to the term  $\mathbf{K}'_{n-1}$  in the first row of (4.53) and obtain, after trading  $(1+\varepsilon)^n \bar{\eta}_n^\delta$  for  $\xi^n$ ,

$$\tilde{\mathbf{K}}_n^r \leq \tilde{\mathbf{K}}'_{n-1} \boldsymbol{\eta}'_{n-1,n} + C \xi^n \bar{\eta}_n^{r-1}. \tag{4.54}$$

Running (4.54) for  $n-1, n-2, \dots, 1$  we finally obtain

$$\tilde{\mathbf{K}}_n^r \leq C \sum \xi^{n-j} \boldsymbol{\eta}'_{n-1,n} \cdots \boldsymbol{\eta}'_{n-j,n-j+1} \bar{\eta}_n^{r-1} \leq C_1 \xi^n \bar{\eta}_n^{r-1}. \tag{4.55}$$

The same argument applies also to expressions of the form  $\prod \tilde{\mathbf{K}}_n^{c_j}$  whose weight  $\sum c_j \geq r$  and gives for them the estimate  $C \xi^n \bar{\eta}_n^{r-1}$ .

Finally, we can replace  $r$  by  $r-1, r-2$  etc, and obtain the inequalities for all  $s \leq r$ . □

These estimates can be improved, especially for small values of  $s$ .

**PROPOSITION 4.10.** *For  $1 \leq s \leq r$  we have the following,*

$$\tilde{\mathbf{K}}_n^s = \mathcal{O}\left(\bar{\eta}_n^s \sum_1^{n-1} \bar{\eta}_j^{r-s+\delta-1} \bar{\eta}_{j,j+1}^{-1}\right).$$

*Remark.* The ghost of the diophantine condition is showing itself here in the form of comparing  $\boldsymbol{\eta}_{j,j+1}$  with a power of  $\boldsymbol{\eta}_j$ , and any estimate of the form  $\boldsymbol{\eta}_{j,j+1} \geq \boldsymbol{\eta}_j^\omega$  makes the present proposition better than Lemma 4.9.

*Proof.* By Proposition 4.2,

$$\tilde{\mathbf{K}}_n^s \leq \tilde{\mathbf{K}}'_{n-1} \boldsymbol{\eta}'_{n-1,n} + \mathbf{K}'_{n-1} \boldsymbol{\eta}'_{n-1,n} \sum_{\substack{1 \leq d \leq s-1 \\ d + \sum c_i = s}} \boldsymbol{\eta}'_{n-1,n} \tilde{\mathbf{K}}'_{n-1} \prod \mathbf{K}'_{n-1} \boldsymbol{\eta}'_{n-1,n}^{c_i-1} \tag{4.55}$$

and by Proposition 4.5,

$$\mathbf{K}'_{n-1} \boldsymbol{\eta}'_{n-1,n}^{s-1} \leq \mathbf{K}'_{n-1} \boldsymbol{\eta}'_{n-1,n}^{s-1} + \boldsymbol{\eta}'_{n-1,n} P(\tilde{\mathbf{K}}'_{n-1}; \text{wt} \geq s+1). \tag{4.56}$$

By Lemma 4.9

$$\boldsymbol{\eta}'_{n-1,n} P(\tilde{\mathbf{K}}'_{n-1}; \text{wt} \geq s+1) = \mathcal{O}(\xi^n \bar{\eta}_n^s)$$

similarly, applying Proposition 4.5 to each of the factors  $\mathbf{K}'_{n-1}$  in (4.55) and continuing as before, we estimate the last sum in (4.55) by a sum of a (universally) bounded number of terms whose form is either

$$\bar{\eta}_{n-1,n}^{s-1} \tilde{\mathbf{K}}'_{n-1} \mathbf{K}'_{n-1}, \quad d \geq 1 \tag{4.57}$$

or

$$\boldsymbol{\eta}'_{n-1,n} \prod \tilde{\mathbf{K}}'_{n-1}^{l_j}, \quad \sum l_j \geq s+1. \tag{4.58}$$

The terms (4.57) are exponentially smaller than the first term in (4.56) while the terms (4.58) are  $\mathcal{O}(\xi^n \bar{\eta}_n^s)$ ; thus,

$$\tilde{\mathbf{K}}_n^s \leq \tilde{\mathbf{K}}'_{n-1} \boldsymbol{\eta}'_{n-1,n} + \mathbf{K}'_{n-1} \bar{\eta}_{n-1,n}^{s-1} + \xi^n \bar{\eta}_n^s \tag{4.59}$$

and by  $(A)_{n-1}^{**}$

$$\mathbf{K}'_{n-1} \bar{\eta}_{n-1,n}^{s-1} \leq \bar{\eta}_{n-1}^{r-1+\delta} \bar{\eta}_{n-1,n}^{s-1} = \bar{\eta}_n^s (\bar{\eta}_{n-1}^{r-s+\delta-1} \bar{\eta}_{n-1,n}^{-1})$$

and the proof goes by induction. □

We are finally ready to prove the main result stated in the beginning of this section. Assume that  $\alpha$  satisfies a ‘Diophantine condition of order  $\beta$ ’ that is:

$$q_{n+1} = \mathcal{O}(q_n^{\beta+1}) \tag{4.60}$$

or, what amounts to the same

$$a_{n,n+1} = \mathcal{O}(q_n^\beta),$$

or again

$$\alpha_n \geq cq_n^{-\beta}.$$

LEMMA 4.11. *If  $f$  is in  $\mathcal{H}^{r+1+\delta}$ , and  $\alpha$  satisfied condition (4.60) with  $\beta + 1 < r + \delta$ , then for all large  $n$ ,  $\bar{\eta}_{n,n+1} = \alpha_n$  and consequently*

$$\bar{\eta}_n \sim q_n^{-1}. \tag{4.61}$$

*Proof.* We introduce the numbers  $\gamma_n$  by the condition  $\bar{\eta}_n = q_n^{-\gamma_n}$ . Thus  $0 < \gamma_n \leq 1$ . We also introduce the numbers  $c_n$  by

$$q_{n+1} = q_n^{c_n}.$$

Condition (4.60) says basically that  $c_n \leq \beta + 1$ .

By (4.9), either  $\bar{\eta}_{n,n+1} = \alpha_{n,n+1}$  or,

$$\bar{\eta}_{n,n+1} = M_r \mathbf{K}_n^r \leq \bar{\eta}_n^{r-1+\delta''}$$

with  $\delta'' < \delta$  close enough so that  $\beta + 1 < r + \delta''$ .

In the first case  $\gamma_{n+1}$  is a convex combination of  $\gamma_n$  and 1, so  $\gamma_{n+1} > \gamma_n$ . In the second case

$$\bar{\eta}_{n+1} = \bar{\eta}_n M_r \mathbf{K}_n^r \leq \bar{\eta}_n^{r+\delta''} = q_n^{-\gamma_n(r+\delta'')} = q_{n+1}^{-\gamma_n(r+\delta'')/c_n}$$

so that  $\gamma_{n+1} = \gamma_n(r + \delta'') / (\beta + 1)$  and we went up by a factor greater than 1. As  $\gamma_n \leq 1$ , this can only happen a finite number of times; thus for all large  $n$ :  $\bar{\eta}_{n,n+1} = \alpha_n$  and  $\bar{\eta}_n \sim q_n^{-1}$ . □

THEOREM 4.12.† *If  $f$  is in  $\mathcal{H}^{r+1+\delta}$ , and  $\alpha$  satisfied condition (4.60) with  $\beta + 1 < r + \delta$ , then the homeomorphism  $h$  which conjugates  $f$  to  $R_\alpha$  is in  $\mathcal{H}^{\delta+r-\beta-\epsilon}$  for all  $\epsilon > 0$ .*

*Proof.* Let  $m$  be the largest integer less than  $r + \delta - \beta - 1$ . By (4.60)

$$\bar{\eta}_j^{r+\delta-m-1} \bar{\eta}_{j,j+1}^{-1} \leq \bar{\eta}_j^{r+\delta-m-\beta-1} \sim q_j^{-(r+\delta-m-\beta-1)},$$

by Proposition 4.10,

$$\tilde{\mathbf{K}}_n^l = \mathcal{O}(q_n^{-l}) \tag{4.62}$$

and by Lemma 4.11 this means

$$\|D^l \log Df^j\|_\infty = \mathcal{O}(1)$$

for all  $l \leq m$ , which implies that  $h \in \mathcal{H}^{m+1}$ .

† The overall structure of our results is especially clear if we focus on a special case of Theorem 4.12: the boundedness of  $\tilde{\mathbf{K}}_n^0, \tilde{\mathbf{K}}_n^0 < \sum_{j=0}^{n-1} a_{j,j+1} \mathbf{K}_j^0, \mathbf{K}_j^0 \leq_j \mathbf{\eta}_{j,j+1} \sum P_j(\dots \tilde{\mathbf{K}}_j^l \dots) + \mathbf{K}_j^r$  (by the basic procedure, Proposition 4.5). Proposition 4.8 and the diophantine condition show that  $a_{j,j+1} \mathbf{K}_j^r$  is exponentially small.  $a_{j,j+1}(\mathbf{\eta}_{j,j+1} \sum P_j(\mathbf{K}_j^l))$  is exponentially small by Lemma 4.11, Proposition 4.3 and Lemma 4.7.

Since  $h \in \mathcal{H}^{m+1}$ , it follows that the conjugation by  $h$  is a differentiable representation of  $\mathbf{T}$  in  $\mathcal{H}^m$ . We have  $f^{q_n} = h^{-1} \circ R_{q_n \alpha} \circ h$  and  $R_{q_n \alpha}$  is within  $q_{n+1}^{-1}$  from the identity, which implies, for  $l < m$ ,

$$\|D^l \log Df^{q_n}\|_\infty = \mathcal{O}(q_{n+1}^{-1}), \tag{4.63}$$

that is

$$\mathbf{K}_n^l = \mathcal{O}(q_{n+1}^{-1} q_n^{-1}). \tag{4.64}$$

Similarly

$$\|D^l \log Df^{c_n q_n}\|_\infty = \mathcal{O}(c_n q_{n+1}^{-1}). \tag{4.65}$$

In order to complete the proof we need to estimate the modulus of continuity of the representation  $\tau \rightarrow h^{-1} \circ R_\tau \circ h$  in  $\mathcal{H}^{m+1}$ , and it is enough to check it for a dense subset of  $\mathbf{T}$ , (and in fact only in a neighborhood of the identity).

We take  $\tau$  of the form  $k\alpha$ , but instead of taking all  $k \in \mathbf{Z}$  or all positive  $k$ , we take those of the form  $k = c_b q_b + \sum_{j \geq b+1} c_j q_j$  with  $1 \leq |c_b| \leq a_{b,b+1}$ ,  $0 \leq |c_j| \leq a_{j,j+1}$  for  $j > b$  so that

$$\|k\alpha\| \sim |c_b| q_{b+1}^{-1} \quad (\|k\alpha\| \text{ denotes the distance of } k\alpha \text{ to the nearest integer}). \tag{4.66}$$

Allowing the coefficients  $c_j$  to be negative comes to guarantee efficient† approximation, and in particular (4.66). As the estimates for  $f^{-1}$  and  $f$  are equivalent, we write everything that follows as if  $c_j > 0, j \geq b$ .

We need to show that for any  $\gamma > 0$  such that

$$m + \gamma < r + \delta - \beta - 1$$

we have

$$\|D^m \log Df^k\|_\infty = \mathcal{O}((c_b q_{b+1}^{-1})^\gamma). \tag{4.67}$$

By (4.13) we have

$$\begin{aligned} D^m \log Df^{c_n q_n} &= \sum_{j=0}^{c_n-1} (D^m \log Df^{q_n} \circ f^{jq_n})^m \\ &+ \sum_{l=1}^{m-1} \sum_{j=0}^{c_n-1} (D^{m-l} \log Df^{q_n} \circ f^{jq_n})(Df^{jq_n})^{m-l} P_l(\dots, D^l \log Df^{jq_n}, \dots) \end{aligned}$$

and we observed already ((4.64)) that for  $0 < l < m$  we have

$$\begin{aligned} \|D^{m-l} \log Df^{q_n}\|_\infty &= \mathcal{O}(q_{n+1}^{-1}), \\ \|D^{m-l} \log Df^{jq_n}\|_\infty &= \mathcal{O}(jq_{n+1}^{-1}), \end{aligned}$$

so that the double sum above is bounded by  $\mathcal{O}(c_n^2 q_{n+1}^{-2})$  and

$$\|D^m \log Df^{c_n q_n}\|_\infty = \mathcal{O}(c_n \|D^m \log Df^{q_n}\|_\infty + c_n^2 q_{n+1}^{-2}). \tag{4.68}$$

By Proposition 4.5

$$\mathbf{K}_n^m \leq \mathbf{K}_n^r + \eta_{n,n+1} P(\tilde{\mathbf{K}}_n^j, \text{wt} \geq m + 1).$$

† For example, if in the usual expansion,  $c_b = 1$  and  $c_{b+1} = a_{b+1,b+2} - 1$  then  $\|k\alpha\| \sim q_{b+2}^{-1}$  and (4.66) may fail. We propose to start the expansion in this case with  $-q_{b+1}\alpha$ .



We use the estimates given by Theorem 4.8 and Proposition 4.10 as well as (4.62), namely

$$\begin{aligned} \mathbf{K}_n^r &= \mathcal{O}(\bar{\eta}_n^{r+\delta-1}) = \mathcal{O}(q^{-(r+\delta-1)}), \\ \tilde{\mathbf{K}}_n^{m+1} &= \mathcal{O}\left(\bar{\eta}_n^{m+1} \sum_1^{n-1} \bar{\eta}_j^{r-m-2+\delta} \eta_{j,j+1}^{-1}\right) \\ &= \mathcal{O}\left(q_n^{-m-1} \sum_1^{n-1} q_j^{(r+\delta-m-2)} q_{j+1} q_j^{-1}\right) \\ &= \mathcal{O}\left(q_n^{-m-1} \sum_1^{n-1} q_j^{(r+\delta-m-2)} q_j^\beta\right) \\ &= \mathcal{O}(q_n^{-m-1} q_n^{1-\gamma}) = \mathcal{O}(q_n^{-m-\gamma}), \end{aligned}$$

terms of weight  $\geq m + 2$  are  $\mathcal{O}(q_n^{-m-1})$  and so are terms of weight  $m + 1$  which are products of terms of weight  $\leq m$ .

Combining all of this we obtain

$$\begin{aligned} \mathbf{K}_n^m &= \mathcal{O}(q_n^{-(r+\delta-1)} + q_n q_{n+1}^{-1} q_n^{-m-\gamma}), \\ \bar{\eta}_n^{-m} \mathbf{K}_n^m &= \mathcal{O}(q_n^{-(r+\delta-m-1)} + q_{n+1}^{-1} q_n^{1-\gamma}) \\ &= \mathcal{O}(q_n^{-(\beta+\gamma)} + q_{n+1}^{-1} q_n^{1-\gamma}) = \mathcal{O}(q_{n+1}^{-1} q_n^{1-\gamma}), \end{aligned}$$

and finally, by (4.68) and since  $c_n q_{n+1}^{-1} \leq q_n^{-1}$

$$\begin{aligned} \|D^m \log Df^{c_n q_n}\|_\infty &\leq \text{const.} [c_n q_{n+1}^{-1} q_n^{1-\gamma} + c_n^2 q_{n+1}^{-2}] \\ &\leq \text{const.} (c_n q_{n+1}^{-1} q_n^{1-\gamma}) \leq \text{const.} (c_n q_{n+1}^{-1})^\gamma. \end{aligned} \tag{4.69}$$

We use this estimate for  $n \geq b$ , keeping it as for  $n = b$ , (this is the expression we need, see (4.67)), while for  $n > b$  we simplify, replacing  $(c_n q_{n+1}^{-1})^\gamma$  by  $q_n^{-\gamma}$ .

By (4.11) and (4.69)

$$\begin{aligned} \|D^m \log Df^{c_b q_b + c_{b+1} q_{b+1}}\|_\infty &\leq \text{const.} [(c_b q_{b+1}^{-1})^\gamma + q_{b+1}^{-\gamma}] \\ &\quad + \sum_{l=1}^{m-1} \|D^{m-l} \log Df^{c_b q_b}\|_\infty \|P_l \\ &\quad \times (\dots, D^j \log Df^{c_{b+1} q_{b+1}}, \dots)\|_\infty \end{aligned}$$

and as before we invoke (4.65) and bound the sum, which involves only derivatives of order less than  $m$ , by a constant multiple of

$$c_b q_{b+1}^{-1} c_{b+1} q_{b+2}^{-1} \leq c_b q_{b+1}^{-2}$$

which can be ignored.

Adding  $c_{b+2} q_{b+2}$ ,  $c_{b+3} q_{b+2}$ , etc. . . . and using the same estimate, we obtain

$$\begin{aligned} \|D^m \log Df^k\|_\infty &\leq \text{const.} [(c_b q_{b+1}^{-1})^\gamma + q_{b+1}^{-\gamma} + q_{b+2}^{-\gamma} + \dots] \\ &\leq \text{const.} (c_b q_{b+1}^{-1})^\gamma = \text{const.} \|k\alpha\|^\gamma \end{aligned}$$

and the proof is complete. □

*Appendix 1. The orbits of  $R_\alpha$*

Let  $\alpha$  be irrational,  $0 < \alpha < \frac{1}{2}$ †. We want to relate the orbit structure of any  $t_0 \in \mathbf{T}$  under  $R_\alpha$  to the continued fraction expansion of  $\alpha$ . Since the orbit of  $t_0$  is obtained

† For  $\frac{1}{2} < \alpha < 1$  interchange ‘left’ and ‘right’ in the description.

from that of  $t = 0$ , for notation we might as well take  $t_0 = 0$ . We define a sequence of positive integers,  $\{q_n\}_{n=1}^\infty$  by taking in their natural order all integers  $m$  such that  $m\alpha$  is closer to zero on  $\mathbb{T}$  than any  $k\alpha$  with  $0 < k < m$ . Thus  $q_0 = 1$  and the first time we come closer to zero is just before crossing it, which means  $q_1 = [1/\alpha]$ . We denote by  $d_n$  the distance of  $q_n\alpha$  to zero ( $d_0 = \alpha$ ,  $d_1 = 1 - q_1\alpha$ , etc.). Since  $(q_1 + j)\alpha$  is at distance  $d_1$  to the 'left' of  $j\alpha$ , we have, for  $j < q_2$ ,  $(q_1 + j)\alpha$  close to zero than  $j\alpha$  is to the right, and farther from zero if  $j\alpha$  is to the 'left' of zero. Thus the first  $m\alpha$  closer to zero than  $q_1\alpha$  will fall in  $(0, \alpha)$  and we must have  $m = aq_1 + 1 = aq_1 + q_0$  where  $a$  is the first positive integer for which  $\alpha - ad_1 < d_1$ , that is,  $a = [\alpha/d_1] = [d_0/d_1]$ . We now write  $a_1 = [1/\alpha]$ ,  $\alpha_n = d_{n-1}/d_{n-2}$  and  $a_n = [1/\alpha_n]$  for  $n > 1$  and notice that  $q_1 = a_1$  and, by the remarks above,  $q_2 = a_2q_1 + q_0$ . We can now repeat the same argument for arbitrary  $n$ , and see that  $q_{n-1}\alpha$  and  $q_n\alpha$  lie on different sides of zero ( $q_n\alpha$  'right' of zero for even  $n$ , 'left' for odd) and one gets to  $q_{n+1}\alpha$  by moving from  $q_{n-1}\alpha$   $a_{n+1}$  steps of size  $d_n$ , that is, by  $a_{n+1}q_n\alpha$ . Thus

$$q_{n+1} = a_{n+1}q_n + q_{n-1}. \tag{A1.1}$$

Denoting by  $p_n$  the integer closest to  $q_n\alpha$  (on  $\mathbb{R}$ ), a moment's reflection shows that  $p_{n+1} = a_{n+1}p_n + p_{n-1}$ , and we recognize the recurrence formulas for the denominators  $q_n$  and the numerators  $p_n$  of the convergents to the infinite continued fraction

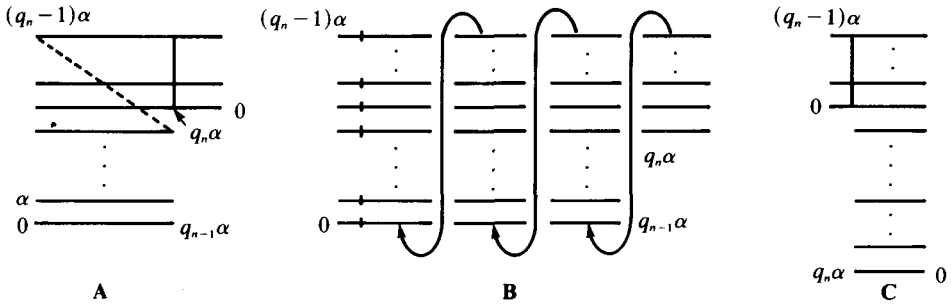
$$[a_1, a_2, \dots] = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 \dots}}}$$

We modify the notation slightly and write  $a_{n,n+1}$  instead of  $a_{n+1}$ , (this somewhat heavy notation is used in order to avoid confusion where shifting the index  $n$  by one may change the order of magnitude of the corresponding  $a$ ). Thus  $a_{n,n+1} = [q_{n+1}/q_n]$  and (A1.1) now reads

$$q_{n+1} = a_{n,n+1}q_n + q_{n-1} \tag{A1.1'}$$

and, similarly,  $\alpha = [a_{1,2}, a_{2,3}, \dots]$  and  $\alpha_n = [a_{n,n+1}, a_{n+1,n+2}, \dots]$ .

We denote by  $P_n = P_n(R_\alpha)$  the partition of  $\mathbb{T}$  by the points  $\{j\alpha\}, j = 0, \dots, q_n - 1$ . If we arrange the  $P_n$ -intervals by the orbits order of their left end points for  $n$  even, (right end points for  $n$  odd), we obtain 'tower with balconies' Diagram **A**, where the action of  $R_\alpha$  on all but the top level is to map any point to the point directly above it. Diagram **B** describes the action of  $R_\alpha$  on all but the leftmost part of the roof, and cutting and stacking according to **B** leads to **C** which is the analogue of **A** (but for the next value of  $n$ ). Notice that the balconies in **C** face the opposite direction from those in **A**. The size of the base of the tower is  $d_{n-1}$  and of the balconies is  $d_n$ . The balconies in **A** become the base for **C**, each level without balcony of **A** (i.e.  $P_n$ ) is divided in **C** (i.e.  $P_{n+1}$ ) into  $a_{n,n+1} - 3$  in our illustration, levels (intervals) and the levels with balcony into  $1 + a_{n,n+1}$ . This description of dynamics of  $R_\alpha$  by towers is a convenient rephrasing of the description in the beginning of this section.



For  $f \in H$  with  $\rho(f) = \alpha$  we denote by  $P_n(f)$  the partition of  $\mathbf{T}$  by  $\{f^j(0)\}$ ,  $j = 0, \dots, q_n - 1$ . The action of powers of  $f$  on  $P_n(f)$  is completely analogous (in fact conjugate) to that of the corresponding powers of  $R_\alpha$  on  $P_n(R_\alpha)$ . This holds whether  $f$  is conjugate to  $R_\alpha$  or not. It is conjugate precisely when the maximal length of the intervals in  $P_n(f)$  tends to zero as  $n \rightarrow \infty$ . We can use this observation to give a slightly different proof of Denjoy's Theorem 1.4 from his inequality (1.5).

Every  $P_n(f)$  interval  $I$  is divided in  $P_{n+1}(f)$  into  $a_{n,n+1}$  (or one more, if  $I$  has a balcony) intervals, one of which has a balcony. For simplicity we assume now that  $a_{n,n+1} \geq 2$ . Each of the  $P_{n+1}(f)$  intervals without balconies can be mapped into its adjacent by  $f^{q_n}$  (onto if the adjacent interval is without balcony) and their ratio is therefore bounded by  $e^v$  ( $v = \text{var log } Df$ ) an interval with balcony can be covered by two images  $f^{q_n}(J)$  and  $f^{2q_n}(J)$ , where  $J$  is an adjacent interval without balcony, which gives again bounds on the ratios. The outcome of all this is that no  $P_{n+1}(f)$  interval can fill more than a certain fraction of the  $P_n(f)$  interval which contains it. If  $a_{n,n+1} = 1$ , the  $P_n(f)$  intervals without balcony simply become  $P_{n+1}(f)$  intervals with balcony which do split in  $P_{n+2}$ . This plus induction yields

LEMMA A.1.1. *There exists a constant  $\eta = \eta(v) < 1$  such that for  $k \geq 2$  the relative length of a  $P_{n+k}(f)$  interval in the containing  $P_n(f)$  interval is bounded by  $\eta^k$ . In particular the length of any  $P_n(f)$  interval is bounded by  $\eta^k$ . In the notations (1.7), (1.8) we have*

$$\eta_{m,n} \leq \eta^{n-m} \quad (n - m \geq 2), \quad \eta_n \leq \eta^n.$$

COROLLARY. *Under the assumptions of Denjoy's Theorem 1.4, if  $\alpha \in G$  (see A.2) then the conjugating function  $h$  satisfies a  $\text{Lip}(\delta)$  condition, where  $\delta$  is such that  $\eta^n < q_n^{-\delta}$ .*

Definition.  $t, \tau$  are  $q_n$ -close if  $\{f^j(t, \tau)\}_{j=0}^{q_n}$  are disjoint.

It is clear that  $t, \tau$  are  $q_n$ -close if and only if  $\tau \in (f^{q_{n-1}}(t), f^{-q_{n-1}}(t), f^{-q_{n-1}}(t))$ . The tower diagrams make it transparent that

$$\left\{ \bigcup_{j=0}^{q_n-1} f^j(t, f^{q_{n-1}}(t)) \right\} \cup \left\{ \bigcup_{j=0}^{q_{n-1}-1} f^j(t, f^{q_n}(t)) \right\} = \mathbf{T};$$

this is just a decomposition of  $\mathbf{T}$  into 'rooms' and 'balconies'. An immediate consequence, which is used repeatedly in the Denjoy procedure, is the following:

LEMMA A.1.2. *Given  $t, t' \in \mathbf{T}$ , there exist  $\tau \in \mathbf{T}$  and an integer  $l, 0 \leq l < q_n$ , such that  $t, \tau$  are  $q_n$ -close, and  $t' = f^l(\tau)$ .*

*Appendix 2. Continued fraction statistics*

We recall some basic facts about the continued fraction algorithm. For  $0 < x < 1$  we write

$$a(x) = [1/x], \quad \phi(x) = \frac{1}{x} - a(x),$$

$$x = \frac{1}{a(x) + \phi(x)} = \frac{1}{a(x) + \frac{1}{a(\phi(x)) + \phi^2(x)}} = \frac{1}{a(x) + \frac{1}{a(\phi(x)) + \frac{1}{\ddots}}},$$

is the continued fraction expansion of  $x$ . We also write it as  $x = [a_1(x), a_2(x), \dots]$  where  $a_1(x) = a(x)$  and  $a_n(x) = a(\phi^{n-1}(x))$ .

The mapping  $\phi$  (of  $(0, 1)$  onto itself) was observed by Gauss to preserve the probability measure  $(\log 2)^{-1} dx/(1+x)$  (direct verification), and it is easy to check that its action is ergodic. As the coefficients  $a_n(x) = a(\phi^{n-1}(x))$  are read by sampling the function  $a$  along orbits, we get relatively precise information about the distribution of values of  $a_n(x)$  for almost all  $x$ . The function  $a(x)$  is not integrable but  $\log(1+a(x))$  is, and applying the ergodic theorem to it yields

$$\lim \frac{1}{N} \log \prod_1^N (1 + a_n(x)) = \text{const.} \left( = (\log 2)^{-1} \int_0^1 \frac{\log(1+a(x))}{1+x} dx \right). \tag{A2.1}$$

Since  $q_{n+1} = a_n q_n + q_{n-1} < (1 + a_n) q_n$ , (A2.1) gives a bound on the exponential growth of  $q_n$  for almost all  $x$ . We get finer estimates if we use the fact

$$q_n = a_{n-1} q_{n-1} + q_{n-2} < \left( a_{n-1} + \frac{1}{a_{n-2} - 1} \right) q_{n-1}$$

that is, apply the ergodic theorem to  $\log(a(\phi(x)) + (1/a(x) - 1))$  etc. and one shows this way that  $\lim q_n^{1/n} = Q$  for almost all  $x$  with  $Q$  an absolute constant.

We denote by  $G$  the set (of full Lebesgue measure) of all  $x$  in  $[0, 1]$  for which the ergodic averages  $N^{-1} \sum_1^N f \circ \phi^n(x)$  converge to the proper limit, namely

$$(\log 1)^{-1} \int_0^1 f(x) \frac{dx}{1+x},$$

for all  $f$  is some countable collection which include all the functions that may be of interest to us here (such as  $\log(1+a(x))$ , or the indicator functions of the sets  $\{x; a(x) = j\}$  etc.) points in  $G$  are referred to as ‘typical’.

Now turn to the question of the convergence a.e. of series of the forms

$$\sum_{n=1}^x a_n(x) K_n, \tag{A2.2}$$

where  $K_n \geq 0$ .

LEMMA A.2.1. *The series (A.2.2) converges a.e. if*

$$\sum K_n \log n < \infty. \tag{A2.3}$$

*Proof.* Write  $a_n^*(x) = \min(a_n(x), n^4)$ . Since  $a_n^*(x) = a_n(x)$  except on a set of measure  $\sim n^{-4}$  we have that for almost all  $x$   $a_n^*(x) = a_n(x)$  except possibly a finite number of times (values of  $n$ ), and the series (A2.2) converges a.e. if and only if

$$\sum a_n^*(x)K_n \tag{A2.4}$$

does also. Now,  $\int_0^1 a_n^*(x) dx \sim \log n$ , and the convergence of (A2.3) implies that of (A2.4) in  $L^1$ , and, the terms being positive, we obtain pointwise convergence a.e. □

*Remark.* The opposite implication is false unless we assume some regularity of the sequence  $\{K_n\}$ . It is not hard to check, e.g., that if  $\{K_n\}$  is monotone the a.e. convergence of (A2.2) is in fact equivalent to (A2.3).

**COROLLARY A2.2.** *For every  $\epsilon > 0$*

$$\begin{aligned} \sum a_n(x)n^{-1}(\log n)^{-2-\epsilon} &< \infty \quad \text{a.e.} \\ \sum a_n(x)n^{-1}(\log n)^{-2}(\log \log n)^{-1-\epsilon} &< \infty \quad \text{a.e.} \end{aligned}$$

*Appendix 3: Sharpness*

We describe briefly how one can construct examples that show that our main results are sharp. We assume that  $\alpha$  satisfies (4.60) and that the constants  $\beta$  cannot be reduced so that for  $\beta' < \beta$  there are infinitely many values of  $n$  such that

$$q_{n+1} > q_n^{\beta'+1}. \tag{A3.1}$$

On the interval  $[0, 1]$  consider a  $C^\infty$ -function  $\phi$  whose support is in  $(0.1, 0.9)$   $\phi = -1$  on  $(0.2, 0.4)$ ;  $\phi = 1$  on  $(0.6, 0.8)$ ;  $\int \phi(t) dt = 0$ ; and its norm in  $C^{r+\delta}$  is about as small as is consistent with all these conditions.  $\phi(\lambda t)$  will display the same basic behavior that  $\phi$  does on  $[0, 1]$ , on the interval  $[0, \lambda^{-1}]$  but its norm in  $C^{r+\delta}$  is  $\lambda^{r+\delta} \|\phi\|_{C^{r+\delta}}$ .

Let  $f$  be a smooth diffeomorphism with rotation number  $\alpha$ . Let  $n$  be such that (A3.1) is valid. Consider the partition  $P_n(f)$  (whose atoms are the levels of a tower such as described in diagram A) and define a perturbation  $f_1$  of  $f$  in the following manner: on each atom of  $P_n(f)$  (excluding the balconies) add to  $Df$  a constant multiple of  $\phi$  scaled so that the atom now plays the role of  $[0, 1]$  for  $\phi$ . Since the size of the atoms is  $\approx q_n^{-1}$ , the norm of the scaled  $\phi$  in  $C^{r+\delta}$  will be  $q_n^{r+\delta} \|\phi\|_{C^{r+\delta}}$  and if we multiply it by  $\theta = q_n^{-r-\delta}$ , the norm is brought back to  $\|\phi\|_{C^{r+\delta}}$ .  $f_1$  is defined as the diffeomorphism which agrees with  $f$  on the edges of the levels of the tower, and whose derivative is  $Df + \theta\phi$  (in the middle of the levels); we may have to adjust it slightly to keep the same rotation number. The diffeomorphism  $f_1$  is still  $C^\infty$  and its norm in  $C^{r+1+\delta}$  may have increased but only to the level of  $\|\phi\|_{C^{r+\delta}}$ . Furthermore, if we repeat this for a sequence of values of  $n$  which increases fast enough, the norms of the perturbations (which are really the Lip  $\delta$  norms of the  $r$ th derivatives) do not add up and we end up with a diffeomorphism in  $C^{r+1+\delta}$ .

The effect such perturbations have on the conjugation becomes clear when we notice that the ‘ $\theta\phi$ ’ added  $\theta$  to  $Df$  on a set that covers a run of order  $\frac{1}{10}q_{n+1}$  and

as we started with  $f$  that was  $C^\infty$ -conjugate to  $R_\alpha$  we can assume  $Df^{c q_n}$  very close to one for  $c < a_{n,n+1}$ . Taking  $c \sim \frac{1}{10} a_{n,n+1}$  we obtain

$$Df_1^{c q_n} > e^{\theta c q_n} \sim e^{\frac{1}{10} q_{n+1} q_n^{-r-\delta}} = e^{\frac{1}{10} q_n^{\beta+1-r-\delta}}$$

and we do not have  $C^1$ -conjugation if  $\beta + 1 - r - \delta > 0$ .

The same construction (with the appropriate scaling) gives that, generically, Theorem 4.11 is sharp, i.e. if  $\beta$  cannot be replaced by  $\beta' < \beta$  in (4.60) we'll have diffeomorphisms in  $\mathcal{H}^{r+1+\delta}$ , with rotation number  $\alpha$  such that the conjugating  $h$  is not in  $C^{r+\delta-\beta+\varepsilon}$  for  $\varepsilon > 0$ .

#### REFERENCES

- [A] V. I. Arnold. Small denominators I, on the mapping of a circle into itself. *Ivestija Akad. Nauk. serie Math.* **25**(1) (1961), 21–86. Translation *Amer. Math. Soc.* 2nd series, **46**, 213–284.
- [C] L. Carleson. A remark on Denjoy's inequality and Herman's theorem. *Publ. Math. I.H.E.S.* **49** (1979), 235–241.
- [H] M. R. Herman. Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations. *Publ. Math. I.H.E.S.* **49** (1979), 5–234.
- [H<sub>2</sub>] M. Herman. Sur les courbes invariantes par les difféomorphismes de l'anneau. *Astérisque* **104** (1983), chap. VIII.
- [HS] J. Hawkins & K. Schmidt. On  $C^2$ -diffeomorphisms which are of type III<sub>1</sub>. *Invent. Math.* **66** (1982), 511–518.
- [K] Y. Katznelson. The action of diffeomorphisms of the circle on the Lebesgue measure. *J. D'Anal. Math.* **36** (1979), 156–166.
- [KO] Y. Katznelson & D. Ornstein. The absolute continuity of the conjugation of certain diffeomorphisms of the circle. *Ergod. Th. & Dynam. Sys.*
- [KS] K. M. Khanin & Ya. G. Sinai. A new proof of M. Herman's theorem. *Comm. Math. Phys.* **112** (1987), 89–101.
- [LA] V. F. Lazutkin. Examples of diffeomorphisms of a circle. *Vestnik Leningrad Univ. Math.* **10** (1982), 55–62.
- [LI] A. N. Livčic. Some homology properties of  $U$ -systems. *Mat. Zametki* **10** (1971), 555–564; *Math. Notes* **10** (1971), 758–763.
- [M] J. Moser. A rapidly convergent iteration method, part II. *Ann. Scuola Norm. Sup. di Pisa* **20** (1966), 499–535; J. Moser. A new technique for the construction of solutions of non-linear differential equations. *Proc. Nat. Acad. Sci.* **41** (1963), 1824–1831; J. Moser. Perturbation theory for almost periodic solutions for undamped non linear differential equations. *Inter. Symp. Nonlinear Differential Equations and Nonlinear Mechanics*, pp. 71–79. New York, 1963.
- [Y] J.-C. Yoccoz. Conjugaison différentiable des difféomorphismes du cercle dont le nombre de rotation vérifie une condition diophantienne. *Ann. Sci. École Norm. Sup.* **4** **17** (1984), 333–359.