Annals of Mathematics, **165** (2007), 89–170

Stretched exponential estimates on growth of the number of periodic points for prevalent diffeomorphisms I

By Vadim Yu. Kaloshin and Brian R. Hunt

Abstract

For diffeomorphisms of smooth compact finite-dimensional manifolds, we consider the problem of how fast the number of periodic points with period n grows as a function of n. In many familiar cases (e.g., Anosov systems) the growth is exponential, but arbitrarily fast growth is possible; in fact, the first author has shown that arbitrarily fast growth is topologically (Baire) generic for C^2 or smoother diffeomorphisms. In the present work we show that, by contrast, for a measure-theoretic notion of genericity we call "prevalence", the growth is not much faster than exponential. Specifically, we show that for each $\rho, \delta > 0$, there is a prevalent set of $C^{1+\rho}$ (or smoother) diffeomorphisms for which the number of periodic n points is bounded above by $\exp(Cn^{1+\delta})$ for some C independent of n. We also obtain a related bound on the decay of hyperbolicity of the periodic points as a function of n, and obtain the same results for 1-dimensional endomorphisms. The contrast between topologically generic and measure-theoretically generic behavior for the growth of the number of periodic points and the decay of their hyperbolicity show this to be a subtle and complex phenomenon, reminiscent of KAM theory. Here in Part I we state our results and describe the methods we use. We complete most of the proof in the 1-dimensional C^2-smooth case and outline the remaining steps, deferred to Part II, that are needed to establish the general case.

The novel feature of the approach we develop in this paper is the introduction of Newton Interpolation Polynomials as a tool for perturbing trajectories of iterated maps.

Table of contents

1. A problem of the growth of the number of periodic points and decay of hyperbolicity for generic diffeomorphisms

1.1. *Introduction.* Let $\mathrm{Diff}^r(M)$ be the space of C^r diffeomorphisms of a finite-dimensional smooth compact manifold M with the uniform C^r-topology, where $\dim M \geq 2$, and let $f \in \mathrm{Diff}^r(M)$. Consider the number of periodic points of period n

$$P_n(f) = \#\{x \in M : \quad x = f^n(x)\}. \tag{1.1}$$

The main question of this paper is:

Question 1.1.1. How quickly can $P_n(f)$ grow with n for a "generic" C^r diffeomorphism f?

We put the word "generic" in quotation marks because as the reader will see the answer depends on the notion of genericity.

For technical reasons one sometimes counts only *isolated* points of period n; let

$$P_n^i(f) = \#\{x \in M : \quad x = f^n(x) \text{ and } y \neq f^n(y) \text{ for } y \neq x \text{ in some neighborhood of } x\}. \tag{1.2}$$

We call a diffeomorphism $f \in \mathrm{Diff}^r(M)$ an *Artin-Mazur diffeomorphism* (or simply an A-M *diffeomorphism*) if the number of isolated periodic orbits of f grows at most exponentially fast, i.e. for some number $C > 0$,

$$P_n^i(f) \leq \exp(Cn) \quad \text{for all } n \in \mathbb{Z}_+. \tag{1.3}$$

Artin and Mazur [AM] proved the following result.

THEOREM 1.1.2. *For* $0 \leq r \leq \infty$, A-M *diffeomorphisms are dense in* $\mathrm{Diff}^r(M)$ *with the uniform* C^r*-topology.*

We say that a point $x \in M$ of period n for f is hyperbolic if $df^n(x)$, the linearization of f^n at x, has no eigenvalues with modulus 1. (Notice that a hyperbolic solution to $f^n(x) = x$ must also be isolated.) We call $f \in \mathrm{Diff}^r(M)$ a strongly Artin-Mazur diffeomorphism if for some number $C > 0$,

$$P_n(f) \leq \exp(Cn) \quad \text{for all } n \in \mathbb{Z}_+, \tag{1.4}$$

and all periodic points of f are hyperbolic (whence $P_n(f) = P_n^i(f)$). In [K1] an elementary proof of the following extension of the Artin-Mazur result is given.

THEOREM 1.1.3. *For* $0 \leq r < \infty$, *strongly* A-M *diffeomorphisms are dense in* $\mathrm{Diff}^r(M)$ *with the uniform* C^r*-topology.*

According to the standard terminology, a set in $\mathrm{Diff}^r(M)$ is called residual if it contains a countable intersection of open dense sets and a property is called (*Baire*) *generic* if diffeomorphisms with that property form a residual set. It turns out the A-M property is not generic, as is shown in [K2]. Moreover:

THEOREM 1.1.4 ([K2]). *For any* $2 \leq r < \infty$ *there is an open set* $\mathcal{N} \subset \mathrm{Diff}^r(M)$ *such that for any given sequence* $a = \{a_n\}_{n \in \mathbb{Z}_+}$ *there is a Baire generic set* $\mathcal{R}_a$ *in* $\mathcal{N}$ *depending on the sequence* a_n *with the property if* $f \in \mathcal{R}_a$, *then* $P_{n_k}^i(f) > a_{n_k}$ *for infinitely many* $n_k \in \mathbb{Z}_+$.

Of course since $P_n(f) \geq P_n^i(f)$, the same statement can be made about $P_n(f)$. But in fact it is shown in [K2] that $P_n(f)$ is infinite for n sufficiently large, due to a continuum of periodic points, for at least a dense set of $f \in \mathcal{N}$.

The proof of this theorem is based on a result of Gonchenko-Shilnikov-Turaev [GST1]. Two slightly different detailed proofs of their result are given in [K2] and [GST2]. The proof in [K2] relies on a strategy outlined in [GST1]. An example of a C^r smooth unimodal map of an interval $[0, 1]$ for which $P_n(f)$

grows faster than an arbitrary given sequence $\{a_n\}$ along a subsequence for any $2 \leq r < \infty$ appears in [KK]. In [KS], Theorem 1.1.4 is extended to the space of 3-dimensional volume-preserving diffeomorphisms also using ideas from [GST1].

However, it seems unnatural that if a diffeomorphism is picked at random then it may have arbitrarily fast growth of the number of periodic points. Moreover, Baire generic sets in Euclidean spaces can have zero Lebesgue measure. Phenomena that are Baire generic, but have a small probability are well-known in dynamical systems, KAM theory, number theory, etc. (see [O], [HSY], [K3] for various examples). This partially motivates the problem posed by Arnold [A]:

Problem 1.1.5. Prove that "with probability one" $f \in \text{Diff}^r(M)$ is an A-M diffeomorphism.

Arnold suggested the following interpretation of "with probability one": *for a* (*Baire*) *generic finite parameter family of diffeomorphisms* $\{f_\varepsilon\}$, *for Lebesgue almost every* ε *we have that* f_ε *is* A-M (compare with [K3]). As Theorem 1.3 shows, a result on the genericity of the set of A-M diffeomorphisms based on (Baire) topology is likely to be extremely subtle, if possible at all.[1] We use instead a notion of "probability one" based on prevalence [HSY], [K3], which is independent of Baire genericity. We also are able to state the result in the form Arnold suggested for generic families using this measure-theoretic notion of genericity.

For a rough understanding of prevalence, consider a Borel measure μ on a Banach space V. We say that a property holds "μ-almost surely for perturbations" if it holds on a Borel set $P \subset V$ such that *for all* $v \in V$ *we have* $v + w \in P$ *for almost every* w *with respect to* μ. Notice that if $V = \mathbb{R}^k$ and μ is Lebesgue measure, then "almost surely with respect to perturbations by μ" is equivalent to "Lebesgue almost everywhere". Moreover, the Fubini/Tonelli theorem implies that if μ is any Borel probability measure on $\mathbb{R}^k$, then a property that holds almost surely with respect to perturbations by μ must also hold Lebesgue almost everywhere. Based on this observation, we call a property on a Banach space "prevalent" if it holds almost surely with respect to perturbations by μ for some Borel probability measure μ on V, which for technical reasons we require to have compact support. In order to apply this notion to the Banach manifold $\text{Diff}^r(M)$, we must describe how we make perturbations in this space, which we will do in the next section.

[1] For example, using technique from [GST2] and [K2] one can prove that for a (Baire) generic finite-parameter family $\{f_\varepsilon\}$ and a (Baire) generic parameter value ε the corresponding diffeomorphism f_ε is not A-M. Unfortunately, how to estimate the measure of non-A-M diffeomorphisms from below is a, so far, unanswerable question.

Our first main result is a partial solution to Arnold's problem. It says that *for a prevalent diffeomorphism* $f \in \mathrm{Diff}^r(M)$, *with* $1 < r \leq \infty$, *and all* $\delta > 0$ *there exists* $C = C(\delta) > 0$ *such that for all* $n \in \mathbb{Z}_+$,

$$P_n(f) \leq \exp(Cn^{1+\delta}). \tag{1.5}$$

The results of this paper have been announced in [KH].

The Kupka-Smale theorem (see e.g. [PM]) states that for a generic diffeomorphism all periodic points are hyperbolic and all associated stable and unstable manifolds intersect one another transversally. [K3] shows that the Kupka-Smale theorem also holds on a prevalent set. So, the Kupka-Smale theorem, in particular, says that a Baire generic (resp. prevalent) diffeomorphism has only hyperbolic periodic points, but *how hyperbolic are the periodic points, as function of their period, for a Baire generic (resp. prevalent) diffeomorphism f*? This is the second main problem we deal with in this paper.

Recall that a linear operator $L : \mathbb{R}^N \to \mathbb{R}^N$ is *hyperbolic* if it has no eigenvalues on the unit circle $\{|z| = 1\} \subset \mathbb{C}$. Denote by $|\cdot|$ the Euclidean norm in $\mathbb{C}^N$. Then we define the *hyperbolicity* of a linear operator L by

$$\gamma(L) = \inf_{\phi\in[0,1)} \inf_{|v|=1} |Lv - \exp(2\pi i\phi)v|. \tag{1.6}$$

We also say that L is γ-hyperbolic if $\gamma(L) \geq \gamma$. In particular, if L is γ-hyperbolic, then its eigenvalues $\{\lambda_j\}_{j=1}^N \subset \mathbb{C}$ are at least γ-distant from the unit circle, i.e. $\min_j ||\lambda_j| - 1| \geq \gamma$. The *hyperbolicity* of a periodic point $x = f^n(x)$ of period n, denoted by $\gamma_n(x, f)$, equals the hyperbolicity of the linearization $df^n(x)$ of f^n at points x, i.e. $\gamma_n(x, f) = \gamma(df^n(x))$. Similarly the number of periodic points $P_n(f)$ of period n is defined, and

$$\gamma_n(f) = \min_{\{x:\ x=f^n(x)\}} \gamma_n(x, f). \tag{1.7}$$

The idea of Gromov [G] and Yomdin [Y] of measuring hyperbolicity is that a γ-hyperbolic point of period n of a C^2 diffeomorphism f has an $M_2^{-2n}\gamma$-neighborhood (where $M_2 = \|f\|_{C^2}$) free from periodic points of the same period.[2] In Appendix A we prove the following result.

Proposition 1.1.6. *Let M be a compact manifold of dimension N, let $f : M \to M$ be a $C^{1+\rho}$ diffeomorphism, where $0 < \rho \leq 1$, that has only hyperbolic periodic points, and let $M_{1+\rho} = \max\{\|f\|_{C^{1+\rho}}, 2^{1/\rho}\}$. Then there is a constant $C = C(M) > 0$ such that for each $n \in \mathbb{Z}_+$,*

$$P_n(f) \leq CM_{1+\rho}^{nN(1+\rho)/\rho}\gamma_n(f)^{-N/\rho}. \tag{1.8}$$

[2] In [Y] hyperbolicity is introduced as the minimal distance of eigenvalues to the unit circle. This way of defining hyperbolicity does not guarantee the existence of an $M_2^{-2n}\gamma$-neighborhood free from periodic points of the same period; see Appendix A.

Proposition 1.1.6 implies that a lower estimate on a decay of hyperbolicity $\gamma_n(f)$ gives an upper estimate on growth of the number of periodic points $P_n(f)$. Therefore, a natural question is:

Question 1.1.7. How quickly can $\gamma_n(f)$ decay with n for a "generic" C^r diffeomorphism f?

For a Baire generic $f \in \mathrm{Diff}^r(M)$, the existence of a lower bound on a rate of decay of $\gamma_n(f)$ would imply the existence of an upper bound on a rate of growth of the number of periodic points $P_n(f)$, whereas no such bound exists by Theorem 1.1.4. Thus again we consider genericity in the measure-theoretic sense of prevalence. Our second main result, which in view of Proposition 1.1.6 implies the first main result, is that *for a prevalent diffeomorphism* $f \in \mathrm{Diff}^r(M)$, *with* $1 < r \leq \infty$, *and for any* $\delta > 0$ *there exists* $C = C(\delta) > 0$ *such that*

$$\gamma_n(f) \geq \exp(-Cn^{1+\delta}). \tag{1.9}$$

Now we shall discuss in more detail our definition of prevalence ("probability one") in the space of diffeomorphisms $\mathrm{Diff}^r(M)$.

1.2. *Prevalence in the space of diffeomorphisms* $\mathrm{Diff}^r(M)$. The space of C^r diffeomorphisms $\mathrm{Diff}^r(M)$ of a compact manifold M is a Banach manifold. Locally we can identify it with a Banach space, which gives it a local linear structure in the sense that we can perturb a diffeomorphism by "adding" small elements of the Banach space. As we described in the previous section, the notion of prevalence requires us to make additive perturbations with respect to a probability measure that is independent of the place where we make the perturbation. Thus although there is not a unique way to put a linear structure on $\mathrm{Diff}^r(M)$, it is important to make a choice that is consistent throughout the Banach manifold.

The way we make perturbations on $\mathrm{Diff}^r(M)$ by small elements of a Banach space is as follows. First we embed M into the interior of the closed unit ball $B^N \subset \mathbb{R}^N$, which we can do for N sufficiently large by the Whitney Embedding Theorem [W]. We emphasize that our results hold for *every* possible choice of an embedding of M into $\mathbb{R}^N$. We then consider a closed neighborhood $U \subset B^N$ of M and Banach space $C^r(U, \mathbb{R}^N)$ of C^r functions from U to $\mathbb{R}^N$. Next we *extend* every element $f \in \mathrm{Diff}^r(M)$ to an element $F \in C^r(U, \mathbb{R}^N)$ that is strongly contracting in *all* the directions transverse to M.[3] Again the particular choice of how we make this extension is not important to our results; in Appendix C we describe how to extend a diffeomorphism and what conditions we need to ensure that the results of Sacker [Sac] and Fenichel [F] apply as follows. Since F has M as an invariant manifold, if we add to F a

[3]The existence of such an extension is not obvious, as pointed out by C. Carminati.

small perturbation in $g \in C^r(U, \mathbb{R}^N)$, the perturbed map $F+g$ has an invariant manifold in U that is C^r-close to M. Then $F+g$ restricted to its invariant manifold corresponds in a natural way to an element of $\mathrm{Diff}^r(M)$, which we consider to be the perturbation of $f \in \mathrm{Diff}^r(M)$ by $g \in C^r(U, \mathbb{R}^N)$. The details of this construction are described in Appendix C.

In this way we reduce the problem to the study of maps in $\mathrm{Diff}^r(U)$, the open subset of $C^r(U, \mathbb{R}^N)$ consisting of those elements that are diffeomorphisms from U to some subset of its interior. The construction we described in the previous paragraph ensures that the number of periodic points $P_n(f)$ and their hyperbolicity $\gamma_n(f)$ for elements of $\mathrm{Diff}^r(M)$ are the same for the corresponding elements of $\mathrm{Diff}^r(U)$. So the bounds that we prove on these quantities for almost every perturbation of any element of $\mathrm{Diff}^r(U)$ hold as well for almost every perturbation of any element of $\mathrm{Diff}^r(M)$. Another justification for considering diffeomorphisms in Euclidean space is that the problem of exponential/superexponential growth of the number of periodic points $P_n(f)$ for a prevalent $f \in \mathrm{Diff}^r(M)$ is a *local problem* on M and is not affected by a global shape of M.

The results stated in the next section apply to any compact domain $U \subset \mathbb{R}^N$, but for simplicity we state them for the closed unit ball B^N. In the previous section, we said that a property is *prevalent* on a Banach space such as $C^r(B^N)$ if it holds on a Borel subset S for which there exists a Borel probability measure μ on $C^r(B^N)$ with compact support such that for all $f \in C^r(B^N)$ we have $f+g \in S$ for almost every g with respect to μ. The complement of a prevalent set is said to be *shy*. We then say that a property is prevalent on an open subset of $C^r(B^N)$ such as $\mathrm{Diff}^r(B^N)$ if the exceptions to the property in $\mathrm{Diff}^r(B^N)$ form a shy subset of $C^r(B^N)$.

In this paper the perturbation measure μ that we use is supported within the analytic functions in $C^r(B^N)$. In this sense we foliate $\mathrm{Diff}^r(B^N)$ by analytic leaves that are compact and overlapping. The main result then says that *for every analytic leaf $L \subset \mathrm{Diff}^r(B^N)$ and every $\delta > 0$, for almost every diffeomorphism $f \in L$ in the leaf L both* (1.5) *and* (1.9) *are satisfied.* Now we define an analytic leaf as a "Hilbert Brick" in the space of analytic functions and a natural Lebesgue product probability measure μ on it.

1.3. *Formulation of the main result in the multidimensional case.* Fix a coordinate system $x = (x_1, \ldots, x_N) \in \mathbb{R}^N \supset B^N$ and the scalar product $\langle x, y \rangle = \sum_i x_i y_i$. Let $\alpha = (\alpha_1, \ldots, \alpha_N)$ be a multi-index from $\mathbb{Z}_+^N$, and let $|\alpha| = \sum_i \alpha_i$. For a point $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$ we write $x^\alpha = \prod_{i=1}^N x_i^{\alpha_i}$. Associate to a real analytic function $\phi : B^N \to \mathbb{R}^N$ the set of coefficients of its expansion:

$$\phi_{\vec{\varepsilon}}(x) = \sum_{\alpha \in \mathbb{Z}_+^N} \vec{\varepsilon}_\alpha x^\alpha. \tag{1.10}$$

Denote by $W_{k,N}$ the space of N-component homogeneous vector-polynomials of degree k in N variables and by $\nu(k,N) = \dim W_{k,N}$ the dimension of $W_{k,N}$. According to the notation of the expansion (1.10), denote coordinates in $W_{k,N}$ by

$$\vec{\varepsilon}_k = \big(\{\vec{\varepsilon}_\alpha\}_{|\alpha|=k}\big) \in W_{k,N}. \tag{1.11}$$

In $W_{k,N}$ we use a scalar product that is invariant with respect to orthogonal transformation of $\mathbb{R}^N \supset B^N$ (see Appendix B), defined as follows:

$$\langle \vec{\varepsilon}_k, \vec{\nu}_k \rangle_k = \sum_{|\alpha|=k} \binom{k}{\alpha}^{-1} \langle \vec{\varepsilon}_\alpha, \vec{\nu}_\alpha \rangle, \quad \|\vec{\varepsilon}_k\|_k = \big(\langle \vec{\varepsilon}_k, \vec{\varepsilon}_k \rangle_k\big)^{1/2}. \tag{1.12}$$

Denote by

$$B_k^N(r) = \{\vec{\varepsilon}_k \in W_{k,N} : \ \|\vec{\varepsilon}_k\|_k \le r\} \tag{1.13}$$

the closed r-ball in $W_{k,N}$ centered at the origin. Let $\mathrm{Leb}_{k,N}$ be Lebesgue measure on $W_{k,N}$ induced by the scalar product (1.12) and normalized by a constant so that the volume of the unit ball is one: $\mathrm{Leb}_{k,N}(B_k^N(1)) = 1$.

Fix a nonincreasing sequence of positive numbers $\vec{\mathbf{r}} = (\{r_k\}_{k=0}^\infty)$ such that $r_k \to 0$ as $k \to \infty$ and define a Hilbert Brick of size $\vec{\mathbf{r}}$

$$\begin{aligned} \mathrm{HB}^N(\vec{\mathbf{r}}) &= \{\vec{\varepsilon} = \{\vec{\varepsilon}_\alpha\}_{\alpha \in \mathbb{Z}_+^N} : \ \text{for all } k \in \mathbb{Z}_+, \|\vec{\varepsilon}_k\|_k \le r_k\} \\ &= B_0^N(r_0) \times B_1^N(r_1) \times \cdots \times B_k^N(r_k) \times \ldots \\ &\subset W_{0,N} \quad \times W_{1,N} \quad \times \cdots \times W_{k,N} \quad \times \ldots . \end{aligned} \tag{1.14}$$

Define *a Lebesgue product probability measure* $\mu_{\vec{\mathbf{r}}}^N$ *associated to the Hilbert Brick* $\mathrm{HB}^N(\vec{\mathbf{r}})$ *of size* $\vec{\mathbf{r}}$ by normalizing for each $k \in \mathbb{Z}_+$ the corresponding Lebesgue measure $\mathrm{Leb}_{k,N}$ on $W_{k,N}$ to the Lebesgue probability measure on the r_k-ball $B_k^N(r_k)$:

$$\mu_{k,r}^N = r^{-\nu(k,N)}\, \mathrm{Leb}_{k,N} \quad \text{and} \quad \mu_{\vec{\mathbf{r}}}^N = \times_{k=0}^\infty \mu_{k,r_k}^N. \tag{1.15}$$

Definition 1.3.1. Let $f \in \mathrm{Diff}^r(B^N)$ be a C^r diffeomorphism of B^N into its interior. We call $\mathrm{HB}^N(\vec{\mathbf{r}})$ a Hilbert Brick of an admissible size $\vec{\mathbf{r}} = (\{r_k\}_{k=0}^\infty)$ with respect to f if

A) for each $\vec{\varepsilon} \in \mathrm{HB}^N(\vec{\mathbf{r}})$, the corresponding function $\phi_{\vec{\varepsilon}}(x) = \sum_{\alpha \in \mathbb{Z}_+^N} \vec{\varepsilon}_\alpha x^\alpha$ is analytic on B^N;

B) for each $\vec{\varepsilon} \in \mathrm{HB}^N(\vec{\mathbf{r}})$, the corresponding map $f_{\vec{\varepsilon}}(x) = f(x) + \phi_{\vec{\varepsilon}}(x)$ is a diffeomorphism from B^N into its interior, i.e. $\{f_{\vec{\varepsilon}}\}_{\vec{\varepsilon} \in \mathrm{HB}^N(\vec{\mathbf{r}})} \subset \mathrm{Diff}^r(B^N)$;

C) for all $\delta > 0$ and all $C > 0$, the sequence $r_k \exp(Ck^{1+\delta}) \to \infty$ as $k \to \infty$.

Remark 1.3.2. The first and second conditions ensure that the family $\{f_{\vec{\varepsilon}}\}_{\vec{\varepsilon}\in \mathrm{HB}^N(\vec{\mathbf{r}})}$ lies inside an analytic leaf within the class of diffeomorphisms $\mathrm{Diff}^r(B^N)$. The third condition provides us enough freedom to perturb. It is important for our method to have infinitely many parameters to perturb. If r_k's were decaying too fast to zero it would make our family of perturbations essentially finite-dimensional.

An example of an admissible sequence $\vec{\mathbf{r}} = (\{r_k\}_{k=0}^{\infty})$ is $r_k = \tau/k!$, where τ depends on f and is chosen sufficiently small to ensure that condition (B) holds. Notice that the diameter of $\mathrm{HB}^N(\vec{\mathbf{r}})$ is then proportional to τ, so that τ can be chosen as some multiple of the distance from f to the boundary of $\mathrm{Diff}^r(B^N)$.

MAIN THEOREM. *For any* $0 < \rho \le \infty$ *(or even* $1+\rho = \omega$*) and any* $C^{1+\rho}$ *diffeomorphism* $f \in \mathrm{Diff}^{1+\rho}(B^N)$*, consider a Hilbert Brick* $\mathrm{HB}^N(\vec{\mathbf{r}})$ *of an admissible size* $\vec{\mathbf{r}}$ *with respect to* f *and the family of analytic perturbations of* f

$$\{f_{\vec{\varepsilon}}(x) = f(x) + \phi_{\vec{\varepsilon}}(x)\}_{\vec{\varepsilon}\in \mathrm{HB}^N(\vec{\mathbf{r}})} \tag{1.16}$$

with the Lebesgue product probability measure $\mu^N_{\vec{\mathbf{r}}}$ *associated to* $\mathrm{HB}^N(\vec{\mathbf{r}})$*. Then for every* $\delta > 0$ *and* $\mu^N_{\vec{\mathbf{r}}}$*-a.e.* $\vec{\varepsilon}$ *there is* $C = C(\vec{\varepsilon}, \delta) > 0$ *such that for all* $n \in \mathbb{Z}_+$

$$\gamma_n(f_{\vec{\varepsilon}}) > \exp(-Cn^{1+\delta}), \quad P_n(f_{\vec{\varepsilon}}) < \exp(Cn^{1+\delta}). \tag{1.17}$$

Remark 1.3.3. A relatively short (16 pages) exposition of ideas involved into the proof of this Theorem appears in Sections 2–6 of [GHK].

Remark 1.3.4. The fact that the measure $\mu^N_{\vec{\mathbf{r}}}$ depends on f does not conform to our definition of prevalence. However, we can decompose $\mathrm{Diff}^r(B^N)$ into a nested countable union of sets $\mathcal{S}_j$ that are each a positive distance from the boundary of $\mathrm{Diff}^r(B^N)$ and for each $j \in \mathbb{Z}^+$ choose an admissible sequence $\vec{\mathbf{r}}_j$ that is valid for all $f \in \mathcal{S}_j$. Since a countable intersection of prevalent subsets of a Banach space is prevalent [HSY], the Main Theorem implies the results stated in terms of prevalence in the introduction.

Remark 1.3.5. The Main Theorem holds also for diffeomorphisms defined on a closed subset of B^N, with essentially the same proof. This fact is used to prove Theorem 1.3.7 below.

Remark 1.3.6. Recently the first author along with A. Gorodetski [GK] applied the technique developed here and obtained partial solution of Palis' conjecture about finiteness of the number of coexisting sinks for surface diffeomorphisms. See also Sections 7 and 8 in [GHK].

In Appendix C we deduce from the Main Theorem the following result.

THEOREM 1.3.7. *Let* $\{f_\sigma\}_{\sigma\in B^m} \subset \mathrm{Diff}^{1+\rho}(M)$ *be a generic* m*-parameter family of* $C^{1+\rho}$ *diffeomorphisms of a compact manifold* M *for some* $\rho > 0$. *Then for every* $\delta > 0$ *and a.e.* $\sigma \in B^m$ *there is a constant* $C = C(\sigma, \delta)$ *such that* (1.17) *is satisfied for every* $n \in \mathbb{Z}_+$.

In Appendix C we also give a precise meaning to the term *generic*. See also Section 9 in [GHK] for a discussion of the notion of prevalence for diffeomorphisms that we use in this paper, and [HK] for a more general discussion of prevalence in nonlinear spaces.

Now we formulate the most general result we shall prove.

Definition 1.3.8. Let $\gamma \geq 0$ and $f \in \mathrm{Diff}^{1+\rho}(B^N)$ be a $C^{1+\rho}$ diffeomorphism for some $\rho > 0$. A point $x \in B^N$ is called (n, γ)*-periodic* if $|f^n(x) - x| \leq \gamma$ and (n, γ)*-hyperbolic* if $\gamma_n(x, f) = \gamma(df^n(x)) \geq \gamma$.

(Notice that a point can be (n, γ)-hyperbolic regardless of its periodicity, but this property is of interest primarily for (n, γ)-periodic points.) For positive C and δ let $\gamma_n(C, \delta) = \exp(-Cn^{1+\delta})$.

THEOREM 1.3.9. *Given the hypotheses of the Main Theorem, for every* $\delta > 0$ *and for* $\mu^N_{\mathbf{r}}$*-a.e.* $\vec{\varepsilon}$ *there is* $C = C(\vec{\varepsilon}, \delta) > 0$ *such that for all* $n \in \mathbb{Z}_+$, *every* $(n, \gamma_n^{1/\rho}(C, \delta))$*-periodic point* $x \in B^N$ *is* $(n, \gamma_n(C, \delta))$*-hyperbolic.* (*Here we assume* $0 < \rho \leq 1$; *in a space* $\mathrm{Diff}^{1+\rho}(B^N)$ *with* $\rho > 1$, *the statement holds with* ρ *replaced by* 1.)

This result together with Proposition 1.1.6 implies the Main Theorem, because any periodic point of period n is (n, γ)-periodic for any $\gamma > 0$.

Remark 1.3.10. In the statement of the Main Theorem and Theorem 1.3.9 the unit ball B^N can be replaced by an bounded open set $U \subset \mathbb{R}^N$. After scaling U can be considered as a subset of the unit ball B^N.

One can define a distance on a compact manifold M and almost periodic points of diffeomorphisms of M. Then one can cover $M = \cup_i U_i$ by coordinate charts and define hyperbolicity for almost periodic points using these charts $\{U_i\}_i$ (see [Y] for details). This gives a precise meaning to the following result.

THEOREM 1.3.11. *Let* $\{f_\sigma\}_{\sigma\in B^m} \subset \mathrm{Diff}^{1+\rho}(M)$ *be a generic* m*-parameter family of diffeomorphisms of a compact manifold* M *for some* $\rho > 0$. *Then for every* $\delta > 0$ *and almost every* $\sigma \in B^m$ *there is a constant* $C = C(\sigma, \delta)$ *such that every* $(n, \gamma_n^{1/\rho}(C, \delta))$*-periodic point* x *in* B^N *is* $(n, \gamma_n(C, \delta))$*-hyperbolic.* (*Here again we assume* $0 < \rho \leq 1$, *replacing* ρ *with* 1 *in the conclusion if* $\rho > 1$.)

The meaning of the term generic is the same as in Theorem 1.3.7 and is discussed in Appendix C.

1.4. *Formulation of the main result in the* 1*-dimensional case.* The proof of the main multidimensional result (Theorem 1.3.9) is quite long and complicated. In order to describe the general approach we develop in this paper we apply our method to the 1-dimensional maps which represent a nontrivial simplified model for the multidimensional problem. The statement of the main result for the 1-dimensional maps has another important feature: it clarifies the statement of the main multidimensional result.

Fix the interval $I = [-1, 1]$. Associate to a real analytic function $\phi : I \to \mathbb{R}$ the set of coefficients of its expansion

$$\phi_\varepsilon(x) = \sum_{k=0}^{\infty} \varepsilon_k x^k. \tag{1.18}$$

For a nonincreasing sequence of positive numbers $\vec{\mathbf{r}} = (\{r_k\}_{k=0}^{\infty})$ such that $r_k \to 0$ as $k \to \infty$ following the multidimensional notation we define a Hilbert Brick of size $\vec{\mathbf{r}}$

$$\mathrm{HB}^1(\vec{\mathbf{r}}) = \{\varepsilon = \{\varepsilon_k\}_{k=0}^{\infty} : \quad \text{for all} \quad k \in \mathbb{Z}_+, \quad |\varepsilon_k| \le r_k\} \tag{1.19}$$

and the product probability measure $\mu_{\vec{\mathbf{r}}}^1$ associated to the Hilbert Brick $\mathrm{HB}^1(\vec{\mathbf{r}})$ of size $\vec{\mathbf{r}}$ which considers each ε_k as a random variable uniformly distributed on $[-r_k, r_k]$ and independent from the other ε_k's.

Main 1-dimensional Theorem. *For any* $0 < \rho \le \infty$ (*or even* $1+\rho = \omega$) *and any* $C^{1+\rho}$ *map* $f : I \to I$ *of the interval* $I = [-1, 1]$ *consider a Hilbert Brick* $\mathrm{HB}^1(\vec{\mathbf{r}})$ *of an admissible size* $\vec{\mathbf{r}}$ *with respect to* f *and the family of analytic perturbations of* f

$$\{f_\varepsilon(x) = f(x) + \phi_\varepsilon(x)\}_{\varepsilon \in \mathrm{HB}^1(\vec{\mathbf{r}})} \tag{1.20}$$

with the Lebesgue product probability measure $\mu_{\vec{\mathbf{r}}}^1$ *associated to* $\mathrm{HB}^1(\vec{\mathbf{r}})$. *Then for every* $\delta > 0$ *and* $\mu_{\vec{\mathbf{r}}}^1$*-a.e.* ε *there is* $C = C(\varepsilon, \delta) > 0$ *such that for all* $n \in \mathbb{Z}_+$

$$\gamma_n(f_\varepsilon) > \exp(-Cn^{1+\delta}), \quad P_n(f_\varepsilon) < \exp(Cn^{1+\delta}). \tag{1.21}$$

Moreover, for $\mu_{\vec{\mathbf{r}}}^1$*-a.e.* ε, *we have that every* $(n, \exp(-Cn^{1+\delta}))$*-periodic point is* $(n, \exp(-Cn^{1+\delta}))$*-hyperbolic.*

In [MMS] Martens-de Melo-van Strien prove a stronger statement for C^2 maps. They show that for any C^2 map f of an interval without flat critical points there are $\gamma > 0$ and $n_0 \in \mathbb{Z}_+$ such that for any $n > n_0$ we have $|\gamma_n(f)| > \gamma$. This also implies that the number of periodic points is bounded by an exponential function of the period. The notion of a flat critical point used in [MMS] is a nonstandard one from the point of view of singularity theory; in particular, if 0 is a critical point, then the distance of $f(x)$ to $f(0)$ does not have to decay to 0 as $x \to 0$ faster than any degree of x.

In [KK] an example of a C^r-unimodal map with a critical point having tangency of order $2r+2$ and an arbitrary fast rate of growth of the number of periodic points is presented.

Let us point out again that the main purpose of discussing the 1-dimensional case in detail is to highlight ideas and explain the general method without overloading the presentation with technical details. The general N-dimensional case is highly involved and excessive amount of technical details make understanding of general ideas of the method not easily accessible.

Acknowledgments. It is great pleasure to thank the thesis advisor of the first author, John Mather, who regularly spent hours listening to oral expositions of various parts of the proof for nearly two years.[4] Without his patience and support this project would never have been completed. The authors are truly grateful to Anton Gorodetski, Giovanni Forni, and an anonymous referee who read the manuscript carefully and have made many useful remarks. The first author is grateful to Anatole Katok for providing an opportunity to give a minicourse on the subject of this paper at Penn State University during the fall of 2000. The authors have profited from conversations with Carlo Carminati, Bill Cowieson, Dima Dolgopyat, Anatole Katok, Michael Lyubich, Michael Shub, Yakov Sinai, Marcelo Viana, Jean-Christophe Yoccoz, Lai-Sang Young, and many others. The first author thanks the Institute for Physical Science and Technology, University of Maryland and, in particular, James Yorke for their hospitality. The second author is grateful in turn to the Institute for Advanced Study at Princeton for its hospitality. The first author acknowledges the support of a Sloan dissertation fellowship during his final year at Princeton, when significant parts of the work were done. The first author is supported by NSF-grant DMS-0300229 and the second author by NSF grant DMS-0104087.

2. Strategy of the proof

Here we describe the strategy of the proof of the Main Result (Theorem 1.3.9). See also Section 3 in [GHK] for a shorter description. The general idea is to fix $C > 0$ and prove an upper bound on the measure of the set of "bad" parameter values $\vec{\varepsilon} \in \mathrm{HB}^N(\vec{\mathbf{r}})$ for which the conclusion of the theorem does not hold. The upper bound we obtain will approach zero as $C \to \infty$, from which it follows immediately that the set of $\vec{\varepsilon} \in \mathrm{HB}^N(\vec{\mathbf{r}})$ that are "bad" for all $C > 0$ has measure zero. For a given $C > 0$, we bound the measure of "bad" parameter values inductively as follows.

Stage 1. We delete all parameter values $\vec{\varepsilon} \in \mathrm{HB}^N(\vec{\mathbf{r}})$ for which the corresponding diffeomorphism $f_{\vec{\varepsilon}}$ has an almost fixed point which is not sufficiently hyperbolic and bound the measure of the deleted set.

[4]This paper is based on the first author's Ph.D. thesis.

Stage 2. We consider only parameter values for which each almost fixed point is sufficiently hyperbolic. Then we delete all parameter values $\vec{\varepsilon}$ for which $f_{\vec{\varepsilon}}$ has an almost periodic point of period 2 which is not sufficiently hyperbolic and bound the measure of that set.

Stage n. We consider only parameter values for which each almost periodic point of period at most $n-1$ is sufficiently hyperbolic (we shall call this *the Inductive Hypothesis*). Then we delete all parameter values $\vec{\varepsilon}$ for which $f_{\vec{\varepsilon}}$ has an almost periodic point of period n which is not sufficiently hyperbolic and bound the measure of that set.

The main difficulty in the proof is then to prove a bound on the measure of "bad" parameter values at stage n such that the bounds are summable over n and that the sum approaches zero as $C \to \infty$. Let us formalize the problem. Fix positive ρ, δ, and C, and recall that $\gamma_n(C, \delta) = \exp(-Cn^{1+\delta})$ for $n \in \mathbb{Z}_+$. Assume $\rho \le 1$; if not, change its value to 1.

Definition 2.0.1. A diffeomorphism $f \in \mathrm{Diff}^{1+\rho}(B^N)$ satisfies the *Inductive Hypothesis of order* n with constants (C, δ, ρ), denoted $f \in \mathrm{IH}(n, C, \delta, \rho)$, if for all $k \le n$, every $(k, \gamma_k^{1/\rho}(C, \delta))$-periodic point is $(k, \gamma_k(C, \delta))$-hyperbolic.

For $f \in \mathrm{Diff}^{1+\rho}(M)$, consider the sequence of sets in the parameter space $\mathrm{HB}^N(\vec{\mathbf{r}})$

$$B_n(C, \delta, \rho, \vec{\mathbf{r}}, f) = \{\vec{\varepsilon} \in \mathrm{HB}^N(\vec{\mathbf{r}}) : f_{\vec{\varepsilon}} \in \mathrm{IH}(n-1, C, \delta, \rho) \text{ but } f_{\vec{\varepsilon}} \notin \mathrm{IH}(n, C, \delta, \rho)\}. \tag{2.1}$$

In other words, $B_n(C, \delta, \rho, \vec{\mathbf{r}}, f)$ is the set of "bad" parameter values $\vec{\varepsilon} \in \mathrm{HB}^N(\vec{\mathbf{r}})$ for which all almost periodic points of $f_{\vec{\varepsilon}}$ with period strictly less than n are sufficiently hyperbolic, but there is an almost periodic point of period n that is not sufficiently hyperbolic. Let

$$M_1 = \sup_{\vec{\varepsilon} \in \mathrm{HB}^N(\vec{\mathbf{r}})} \max\{\|f_{\vec{\varepsilon}}\|_{C^1}, \|f_{\vec{\varepsilon}}^{-1}\|_{C^1}\}; \tag{2.2}$$
$$M_{1+\rho} = \sup_{\vec{\varepsilon} \in \mathrm{HB}^N(\vec{\mathbf{r}})} \max\{\|f_{\vec{\varepsilon}}\|_{C^{1+\rho}}, M_1, 2^{1/\rho}\}.$$

Our goal is to find an upper bound

$$\mu_{\vec{\mathbf{r}}}^N \{B_n(C, \delta, \rho, \vec{\mathbf{r}}, f)\} \le \mu_n(C, \delta, \rho, \vec{\mathbf{r}}, M_{1+\rho}) \tag{2.3}$$

for the measure of the set of "bad" parameter values. Then the sum over n of (2.3) gives an upper bound

$$\mu_{\vec{\mathbf{r}}}^N \{\cup_{n=1}^{\infty} B_n(C, \delta, \rho, \vec{\mathbf{r}}, f)\} \le \sum_{n=1}^{\infty} \mu_n(C, \delta, \rho, \vec{\mathbf{r}}, M_{1+\rho}) \tag{2.4}$$

on the measure of the set of all parameters $\vec{\varepsilon}$ for which $f_{\vec{\varepsilon}}$ has for at least one n an $(n, \gamma_n^{1/\rho}(C,\delta))$-periodic point that is not $(n, \gamma_n(C,\delta))$-hyperbolic. If this sum converges and

$$\sum_{n=1}^{\infty} \mu_n(C,\delta,\rho,\vec{\mathbf{r}},M_{1+\rho}) = \mu(C,\delta,\rho,\vec{\mathbf{r}},M_{1+\rho}) \to 0 \text{ as } C \to \infty \tag{2.5}$$

for every positive ρ, δ, and $M_{1+\rho}$, then Theorem 1.3.9 follows. In the remainder of this chapter we describe the key construction we use to obtain a bound $\mu_n(C,\delta,\rho,\vec{\mathbf{r}},M_{1+\rho})$ that meets condition (2.5).

2.1. *Various perturbations of recurrent trajectories by Newton interpolation polynomials.* The approach we take to estimate the measure of "bad" parameter values in the space of perturbations $\mathrm{HB}^N(\vec{\mathbf{r}})$ is to choose a coordinate system for this space and for a finite subset of the coordinates to estimate the amount that we must change a particular coordinate to make a "bad" parameter value "good". Actually we will choose a coordinate system that depends on a particular point $x_0 \in B^N$, the idea being to use this coordinate system to estimate the measure of "bad" parameter values corresponding to initial conditions in some neighborhood of x_0, then cover B^N with a finite number of such neighborhoods and sum the corresponding estimates. For a particular set of initial conditions, a diffeomorphism will be "good" if every point in the set is either sufficiently nonperiodic or sufficiently hyperbolic.

In order to keep the notation and formulas simple as we formalize this approach, we consider the case of 1-dimensional maps, but the reader should always have in mind that our approach is designed for multidimensional diffeomorphisms. Let $f: I \to I$ be a C^1 map on the interval $I = [-1,1]$. Recall that a trajectory $\{x_k\}_{k\in\mathbb{Z}}$ of f is called *recurrent* if it returns arbitrarily close to its initial position — that is, for all $\gamma > 0$ we have $|x_0 - x_n| < \gamma$ for some $n > 0$. A very basic question is how much one should perturb f to make x_0 periodic. Here is an elementary Closing Lemma that gives a simple partial answer to this question.

Closing Lemma. *Let $\{x_k = f^k(x_0)\}_{k=0}^n$ be a trajectory of length $n+1$ of a map $f: I \to I$. Let $u = (x_0 - x_n)/\prod_{k=0}^{n-2}(x_{n-1} - x_k)$. Then x_0 is a periodic point of period n of the map*

$$f_u(x) = f(x) + u\prod_{k=0}^{n-2}(x - x_k). \tag{2.6}$$

Of course f_u is close to f if and only if u is sufficiently small, meaning that $|x_0 - x_n|$ should be small compared to $\prod_{k=0}^{n-2}|x_{n-1} - x_k|$. However, this product is likely to contain small factors for recurrent trajectories. In general, it is difficult to control the effect of perturbations for recurrent trajectories. The simple reason why this is so is because *one cannot perturb f at two nearby points independently*.

The Closing Lemma above also gives an idea of how much we must change the parameter u to make a point x_0 that is (n,γ)-periodic not be (n,γ)-periodic for a given $\gamma > 0$, which as we described above is one way to make a map that is "bad" for the initial condition x_0 become "good". To make use of the other part of our alternative we must determine how much we need to perturb a map f to make a given x_0 be (n,γ)-hyperbolic for some $\gamma > 0$.

Perturbation of hyperbolicity. *Let* $\{x_k = f^k(x_0)\}_{k=0}^{n-1}$ *be a trajectory of length* n *of a* C^1 *map* $f : I \to I$. *Then for the map*

$$f_v(x) = f(x) + v(x - x_{n-1}) \prod_{k=0}^{n-2} (x - x_k)^2 \tag{2.7}$$

such that $v \in \mathbb{R}$ *and*

$$\big||(f_v^n)'(x_0)| - 1\big| = \left|\left|\prod_{k=0}^{n-1} f'(x_k) + v \prod_{k=0}^{n-2} (x_{n-1} - x_k)^2 \prod_{k=0}^{n-2} f'(x_k)\right| - 1\right| > \gamma, \tag{2.8}$$

x_0 *is an* (n,γ)*-hyperbolic point of* f_v.

One more time we can see the product of distances $\prod_{k=0}^{n-2} |x_{n-1} - x_k|$ along the trajectory is an important quantitative characteristic of how much freedom we have to perturb.

The perturbations (2.6) and (2.7) are reminiscent of Newton interpolation polynomials. Let us put these formulas into a general setting using singularity theory.

Given $n > 0$ and a C^1 function $f : I \to \mathbb{R}$ we define an associated function $j^{1,n} f : I^n \to I^n \times \mathbb{R}^{2n}$ by

$$j^{1,n} f(x_0, \dots, x_{n-1}) = \Big(x_0, \dots, x_{n-1}, f(x_0), \dots, f(x_{n-1}), f'(x_0), \dots, f'(x_{n-1})\Big). \tag{2.9}$$

In singularity theory this function is called the *n-tuple 1-jet* of f. The ordinary 1-jet of f, usually denoted by $j^1 f(x) = (x, f(x), f'(x))$, maps I to the *1-jet space* $\mathcal{J}^1(I,\mathbb{R}) \simeq I \times \mathbb{R}^2$. The product of n copies of $\mathcal{J}^1(I,\mathbb{R})$, called the *multijet space*, is denoted by

$$\mathcal{J}^{1,n}(I,\mathbb{R}) = \underbrace{\mathcal{J}^1(I,\mathbb{R}) \times \cdots \times \mathcal{J}^1(I,\mathbb{R})}_{n \text{ times}}, \tag{2.10}$$

and is equivalent to $I^n \times \mathbb{R}^{2n}$ after coordinates are rearranged . The n-tuple 1-jet of f associates with each n-tuple of points in I^n all the information necessary to determine how close the n-tuple is to being a periodic orbit, and if so, how close it is to being nonhyperbolic.

The set

(2.11)
$$\Delta_n(I) = \Big\{(x_0, \dots, x_{n-1}) \times \mathbb{R}^{2n} \subset \mathcal{J}^{1,n}(I,\mathbb{R}) : \exists\ i \neq j \text{ such that } x_i = x_j\Big\}$$

is called the *diagonal* (or sometimes the *generalized diagonal*) in the space of multijets. In singularity theory the space of multijets is defined outside of the diagonal $\Delta_n(I)$ and is usually denoted by $\mathcal{J}^1_n(I,\mathbb{R}) = \mathcal{J}^{1,n}(I,\mathbb{R}) \setminus \Delta_n(I)$ (see [GG]). It is easy to see that *a recurrent trajectory* $\{x_k\}_{k\in\mathbb{Z}_+}$ *is located in a neighborhood of the diagonal* $\Delta_n(I) \subset \mathcal{J}^{1,n}(I,\mathbb{R})$ *in the space of multijets for a sufficiently large* n. If $\{x_k\}_{k=0}^{n-1}$ is a part of a recurrent trajectory of length n, then the product of distances along the trajectory

$$\prod_{k=0}^{n-2} |x_{n-1} - x_k| \tag{2.12}$$

measures how close $\{x_k\}_{k=0}^{n-1}$ is to the diagonal $\Delta_n(I)$, or how independently one can perturb points of a trajectory. One can also say that (2.12) is a quantitative characteristic of how recurrent a trajectory of length n is. Introduction of this *product of distances along a trajectory into analysis of recurrent trajectories* is a new point of our paper.

2.2. *Newton interpolation and blow-up along the diagonal in multijet space.* Now we present a construction due to Grigoriev and Yakovenko [GY] which puts the “Closing Lemma” and “Perturbation of Hyperbolicity” statements above into a general framework. It is an interpretation of Newton interpolation polynomials as an algebraic blow-up along the diagonal in the multijet space. In order to keep the notation and formulas simple we continue in this section to consider only the 1-dimensional case.

Consider the $2n$-parameter family of perturbations of a C^1 map $f : I \to I$ by polynomials of degree $2n-1$:

$$f_\varepsilon(x) = f(x) + \phi_\varepsilon(x), \qquad \phi_\varepsilon(x) = \sum_{k=0}^{2n-1} \varepsilon_k x^k, \tag{2.13}$$

where $\varepsilon = (\varepsilon_0, \dots, \varepsilon_{2n-1}) \in \mathbb{R}^{2n}$. The perturbation vector ε consists of coordinates from the Hilbert Brick $\mathrm{HB}^1(\vec{\mathbf{r}})$ of analytic perturbations defined in Section 1.3. Our goal now is to describe how such perturbations affect the n-tuple 1-jet of f. Since the operator $j^{1,n}$ is linear in f, for the time being we consider only the perturbations ϕ_ε and their n-tuple 1-jets. For each n-tuple $\{x_k\}_{k=0}^{n-1}$ there is a natural transformation $\mathcal{J}^{1,n} : I^n \times \mathbb{R}^{2n} \to \mathcal{J}^{1,n}(I,\mathbb{R})$ from ε-coordinates to jet-coordinates, given by

$$\mathcal{J}^{1,n}(x_0, \dots, x_{n-1}, \varepsilon) = j^{1,n}\phi_\varepsilon(x_0, \dots, x_{n-1}). \tag{2.14}$$

Instead of working directly with the transformation $\mathcal{J}^{1,n}$, we introduce intermediate u-coordinates based on Newton interpolation polynomials. The relation between ε-coordinates and u-coordinates is given implicitly by

$$\phi_\varepsilon(x) = \sum_{k=0}^{2n-1} \varepsilon_k x^k = \sum_{k=0}^{2n-1} u_k \prod_{j=0}^{k-1} (x - x_{j (\mathrm{mod}\ n)}). \tag{2.15}$$

Based on this identity, we will define functions $\mathcal{D}^{1,n} : I^n \times \mathbb{R}^{2n} \to I^n \times \mathbb{R}^{2n}$ and $\pi^{1,n} : I^n \times \mathbb{R}^{2n} \to \mathcal{J}^{1,n}(I,\mathbb{R})$ so that $\mathcal{J}^{1,n} = \pi^{1,n} \circ \mathcal{D}^{1,n}$, or in other words the diagram in Figure 1 commutes. We will show later that $\mathcal{D}^{1,n}$ is invertible, while $\pi^{1,n}$ is invertible away from the diagonal $\Delta_n(I)$ and defines a blow-up along it in the space of multijets $\mathcal{J}^{1,n}(I,\mathbb{R})$.

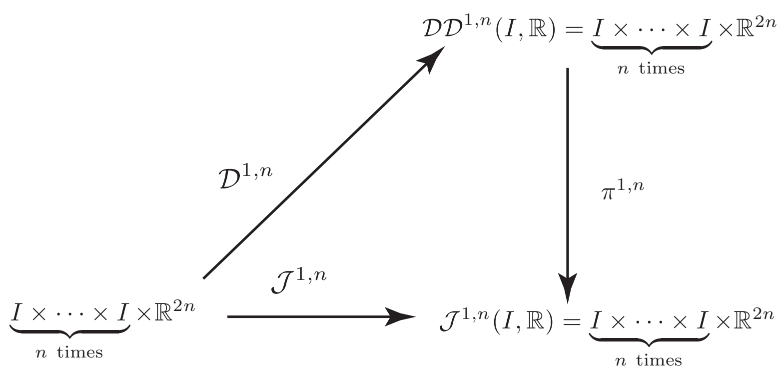

Figure 2.1: Algebraic blow-up along the diagonal $\Delta_n(I)$

The intermediate space, which we denote by $\mathcal{DD}^{1,n}(I,\mathbb{R})$, is called *the space of divided differences* and consists of n-tuples of points $\{x_k\}_{k=0}^{n-1}$ and $2n$ real coefficients $\{u_k\}_{k=0}^{2n-1}$. Here are explicit coordinate-by-coordinate formulas defining $\pi^{1,n} : \mathcal{DD}^{1,n}(I,\mathbb{R}) \to \mathcal{J}^{1,n}(I,\mathbb{R})$. This mapping is given by

$$\begin{aligned} &\pi^{1,n}(x_0, \dots, x_{n-1}, u_0, \dots, u_{2n-1}) \\ &\quad = \Big(x_0, \dots, x_{n-1}, \phi_\varepsilon(x_0), \dots, \phi_\varepsilon(x_{n-1}), \phi'_\varepsilon(x_0), \dots, \phi'_\varepsilon(x_{n-1})\Big), \end{aligned} \tag{2.16}$$

where

$$\begin{aligned} \phi_\varepsilon(x_0) &= u_0, \\ \phi_\varepsilon(x_1) &= u_0 + u_1(x_1 - x_0), \\ \phi_\varepsilon(x_2) &= u_0 + u_1(x_2 - x_0) + u_2(x_2 - x_0)(x_2 - x_1), \\ &\vdots \\ \phi_\varepsilon(x_{n-1}) &= u_0 + u_1(x_{n-1} - x_0) + \dots \\ &\quad + u_{n-1}(x_{n-1} - x_0) \dots (x_{n-1} - x_{n-2}), \end{aligned} \tag{2.17}$$

$$\phi_\varepsilon'(x_0) = \frac{\partial}{\partial x}\left(\sum_{k=0}^{2n-1} u_k \prod_{j=0}^{k}(x - x_{j(\mathrm{mod}\ n)})\right)\Big|_{x=x_0},$$

$$\vdots$$

$$\phi_\varepsilon'(x_{n-1}) = \frac{\partial}{\partial x}\left(\sum_{k=0}^{2n-1} u_k \prod_{j=0}^{k}(x - x_{j(\mathrm{mod}\ n)})\right)\Big|_{x=x_{n-1}}.$$

These formulas are very useful for dynamics. For a given base map f and initial point x_0, the image $f_\varepsilon(x_0) = f(x_0) + \phi_\varepsilon(x_0)$ of x_0 depends only on u_0. Furthermore the image can be set to any desired point by choosing u_0 appropriately — we say then that it depends only and nontrivially on u_0. If x_0, x_1, and u_0 are fixed, the image $f_\varepsilon(x_1)$ of x_1 depends only on u_1, and as long as $x_0 \neq x_1$ it depends nontrivially on u_1. More generally for $0 \leq k \leq n-1$, if distinct points $\{x_j\}_{j=0}^{k}$ and coefficients $\{u_j\}_{j=0}^{k-1}$ are fixed, then the image $f_\varepsilon(x_k)$ of x_k depends only and nontrivially on u_k.

Suppose now that an n-tuple of points $\{x_j\}_{j=0}^{n}$ not on the diagonal $\Delta_n(I)$ and Newton coefficients $\{u_j\}_{j=0}^{n-1}$ are fixed. Then derivative $f_\varepsilon'(x_0)$ at x_0 depends only and nontrivially on u_n. Likewise for $0 \leq k \leq n-1$, if distinct points $\{x_j\}_{j=0}^{n-1}$ and Newton coefficients $\{u_j\}_{j=0}^{n+k-1}$ are fixed, then the derivative $f_\varepsilon'(x_k)$ at x_k depends only and nontrivially on u_{n+k}.

As Figure 2 illustrates, these considerations show that for any map f and any desired trajectory of distinct points with any given derivatives along it, one can choose Newton coefficients $\{u_k\}_{k=0}^{2n-1}$ and explicitly construct a map $f_\varepsilon = f + \phi_\varepsilon$ with such a trajectory. Thus we have shown that $\pi^{1,n}$ is invertible away from the diagonal $\Delta_n(I)$ and defines a blow-up along it in the space of multijets $\mathcal{J}^{1,n}(I,\mathbb{R})$.

Next we define the function $\mathcal{D}^{1,n} : I^n \times \mathbb{R}^{2n} \to \mathcal{DD}^{1,n}(I,\mathbb{R})$ explicitly using so-called divided differences. Let $g : \mathbb{R} \to \mathbb{R}$ be a C^r function of one real variable.

Definition 2.2.1. The *first order divided difference of* g is defined as

$$\Delta g(x_0, x_1) = \frac{g(x_1) - g(x_0)}{x_1 - x_0} \tag{2.18}$$

for $x_1 \neq x_0$ and extended by its limit value as $g'(x_0)$ for $x_1 = x_0$. Iterating this construction we define divided differences of the m-th order for $2 \leq m \leq r$,

$$\Delta^m g(x_0, \dots, x_m) = \frac{\Delta^{m-1} g(x_0, \dots, x_{m-2}, x_m) - \Delta^{m-1} g(x_0, \dots, x_{m-2}, x_{m-1})}{x_m - x_{m-1}} \tag{2.19}$$

for $x_{m-1} \neq x_m$ and extended by its limit value for $x_{m-1} = x_m$.

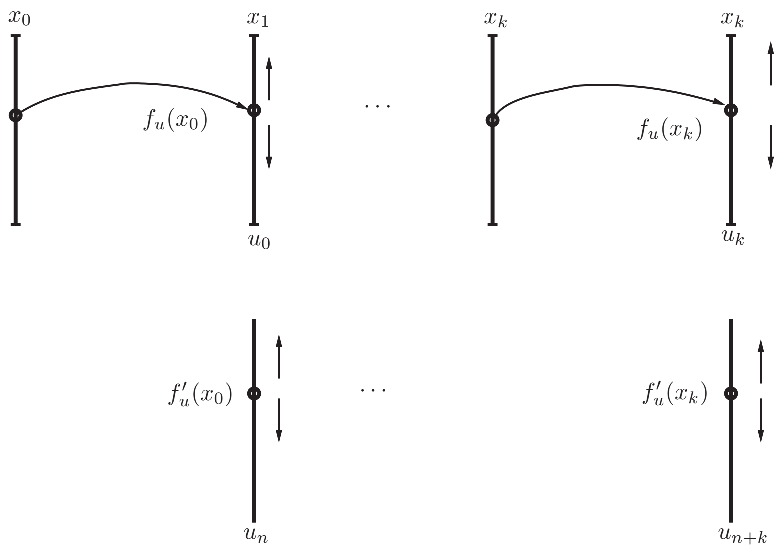

Figure 2.2: Newton coefficients and their action

A function loses at most one derivative of smoothness with each application of Δ, and so $\Delta^m g$ is at least C^{r-m} if g is C^r. Notice that Δ^m is linear as a function of g, and one can show that it is a symmetric function of $x_0, \dots, x_m$; in fact, by induction it follows that

$$\Delta^m g(x_0, \dots, x_m) = \sum_{i=0}^{m} \frac{g(x_i)}{\prod_{j \neq i}(x_i - x_j)}. \tag{2.20}$$

Another identity that is proved by induction will be more important for us, namely

$$\Delta^m \; x^k(x_0, \dots, x_m) = p_{k,m}(x_0, \dots, x_m), \tag{2.21}$$

where $p_{k,m}(x_0, \dots, x_m)$ is 0 for $m > k$ and for $m \leq k$ is the sum of all degree $k - m$ monomials in $x_0, \dots, x_m$ with unit coefficients,

$$p_{k,m}(x_0, \dots, x_m) = \sum_{r_0 + \dots + r_m = k-m} \prod_{j=0}^{m} x_j^{r_j}. \tag{2.22}$$

The divided differences are the right coefficients for the Newton interpolation formula. For all C^∞ functions $g : \mathbb{R} \to \mathbb{R}$ we have

$$\begin{aligned} g(x) = {} & \Delta^0 g(x_0) + \Delta^1 g(x_0, x_1)(x - x_0) + \dots \\ & + \Delta^{n-1} g(x_0, \dots, x_{n-1})(x - x_0) \dots (x - x_{n-2}) \\ & + \Delta^n g(x_0, \dots, x_{n-1}, x)(x - x_0) \dots (x - x_{n-1}) \end{aligned} \tag{2.23}$$

identically for all values of $x, x_0, \dots, x_{n-1}$. All terms of this representation are polynomial in x except for the last one which we view as a remainder term.

The sum of the polynomial terms is the degree $(n-1)$ *Newton interpolation polynomial* for g at $\{x_k\}_{k=0}^{n-1}$. To obtain a degree $2n-1$ interpolation polynomial for g and its derivative at $\{x_k\}_{k=0}^{n-1}$, we simply use (2.23) with n replaced by $2n$ and the $2n$-tuple of points $\{x_{k(\mathrm{mod}\ n)}\}_{k=0}^{2n-1}$.

Recall that $\mathcal{D}^{1,n}$ was defined implicitly by (2.15). We have described how to use divided differences to construct a degree $2n-1$ interpolating polynomial of the form on the right-hand side of (2.15) for an arbitrary C^∞ function g. Our interest then is in the case $g = \phi_\varepsilon$, which as a degree $2n-1$ polynomial itself will have no remainder term and coincide exactly with the interpolating polynomial. Thus $\mathcal{D}^{1,n}$ is given coordinate-by-coordinate by

$$u_m = \Delta^m \left(\sum_{k=0}^{2n-1} \varepsilon_k x^k \right) (x_0, \dots, x_{m\ (\mathrm{mod}\ n)}) = \varepsilon_m + \sum_{k=m+1}^{2n-1} \varepsilon_k p_{k,m}(x_0, \dots, x_{m\ (\mathrm{mod}\ n)}) \tag{2.24}$$

for $m = 0, \dots, 2n-1$.

Equation (2.24) defines a transformation $(u_0, \dots, u_{2n-1}) = \mathcal{L}^1_{\mathbf{X}_n}(\varepsilon)$ on $\mathbb{R}^{2n}$, where $\mathbf{X}_n = (x_0, \dots, x_{n-1}) \in I^n$. We call $\mathcal{L}^1_{\mathbf{X}_n}$ the *Newton map*. This map is simply a restriction of $\mathcal{D}^{1,n}$ to its final $2n$ coordinates:

$$\mathcal{D}^{1,n}(\mathbf{X}_n, \varepsilon) = (\mathbf{X}_n, \mathcal{L}^1_{\mathbf{X}_n}(\varepsilon)). \tag{2.25}$$

Notice that for fixed $\mathbf{X}_n$, the Newton map is linear and given by an upper triangular matrix with units on the diagonal. Hence it is Lebesgue measure-preserving and invertible, whether or not $\mathbf{X}_n$ lies on the diagonal $\Delta_n(I)$.

Furthermore, the Newton map $\mathcal{L}^1_{\mathbf{X}_n}$ preserves the class of *scaled Lebesgue product measures* introduced in (1.15). In general, a measure μ on $\mathbb{R}^{2n}$ is a scaled Lebesgue product measure if it is the product $\mu = \mu_0 \times \cdots \times \mu_{2n-1}$, where each μ_j is Lebesgue measure on $\mathbb{R}$ scaled by a constant factor (which may depend on the coordinate j). Since the $\mathcal{L}^1_{\mathbf{X}_n}$ only shears in coordinate directions, we have the following lemma.

LEMMA 2.2.2. *The Newton map $\mathcal{L}^1_{\mathbf{X}_n}$ given by* (2.24) *preserves all scaled Lebesgue product measures.*

This lemma will be used in Chapter 3. In the next section, we will introduce the particular scaled Lebesgue product measure to which the lemma will be applied.

We call the basis of monomials

$$\prod_{j=0}^{k} (x - x_{j(\mathrm{mod}\ n)}) \quad \text{for} \quad k = 0, \dots, 2n-1, \tag{2.26}$$

in the space of polynomials of degree $2n-1$ the *Newton basis* defined by the n-tuple $\{x_k\}_{k=0}^{n-1}$. The Newton map and the Newton basis, and their analogues in dimension N, are useful tools for perturbing trajectories and estimating the measure $\mu_n(C,\delta,\rho,\vec{\mathbf{r}},M_{1+\rho})$ of "bad" parameter values $\vec{\varepsilon}\in \mathrm{HB}^N(\mathbf{r})$.

2.3. *Estimates of the measure of "bad" parameters and Fubini reduction to finite-dimensional families.* We return now to the the general case of $C^{1+\rho}$ diffeomorphisms on $\mathbb{R}^N$. In order to bound $\mu^N_{\vec{\mathbf{r}}}\{B_n(C,\delta,\rho,\vec{\mathbf{r}},f)\}$ we decompose the infinite-dimensional Hilbert Brick $\mathrm{HB}^N(\vec{\mathbf{r}})$ into the direct sum of a finite-dimensional brick of polynomials of degree $2n-1$ in N variables and its orthogonal complement.

Recall that $\vec{\mathbf{r}}=(\{r_m\}_{m=0}^{\infty})$ denotes the nonincreasing sequence $\{r_m\}_{m\in\mathbb{Z}_+}$ of sizes of the Hilbert Brick. With the notation (1.11) and (1.12), define

(2.27)
$$\begin{aligned}
\mathrm{HB}^N_{<k}(\vec{\mathbf{r}}) &= \big\{\{\vec{\varepsilon}_m\}_{m<k}:\ \text{for every } 0\le m<k,\ \|\vec{\varepsilon}_m\|_m\le r_m\big\}\\
&= B^N_0(r_0)\times\cdots\times B^N_{k-1}(r_{k-1})\subset W_{0,N}\times W_{1,N}\times\cdots\times W_{k-1,N};\\
\mathrm{HB}^N_{\ge k}(\vec{\mathbf{r}}) &= \big\{\{\vec{\varepsilon}_m\}_{m\ge k}:\ \text{for every } m\ge k,\ \|\vec{\varepsilon}_m\|_m\le r_m\big\}\\
&= B^N_k(r_k)\times B^N_{k+1}(r_{k+1})\times\cdots\subset W_{k,N}\times W_{k+1,N}\times\ldots;\\
\mathrm{HB}^N(\vec{\mathbf{r}}) &= \mathrm{HB}^N_{<k}(\vec{\mathbf{r}})\oplus \mathrm{HB}^N_{\ge k}(\vec{\mathbf{r}}).
\end{aligned}$$

Each parameter $\vec{\varepsilon}\in \mathrm{HB}^N(\vec{\mathbf{r}})$ has a unique decomposition into

$$\begin{aligned}
\vec{\varepsilon} &= (\vec{\varepsilon}_{<k},\vec{\varepsilon}_{\ge k})\in \mathrm{HB}^N_{<k}(\vec{\mathbf{r}})\oplus \mathrm{HB}^N_{\ge k}(\vec{\mathbf{r}}),\\
\phi_{\vec{\varepsilon}}(x) &= \phi_{\vec{\varepsilon}_{<k}}(x)+\phi_{\vec{\varepsilon}_{\ge k}}(x)=\sum_{|\alpha|<k}\vec{\varepsilon}_\alpha x^\alpha+\sum_{|\alpha|\ge k}\vec{\varepsilon}_\alpha x^\alpha,
\end{aligned}\tag{2.28}$$

where $\phi_{\vec{\varepsilon}_{<k}}(x)$ is a vector-polynomial of degree $k-1$ and $\phi_{\vec{\varepsilon}_{\ge k}}(x)$ is an analytic function with all Taylor coefficients of order less than k being equal to zero. Recall the notation (1.15), and decompose the measure $\mu^N_{\vec{\mathbf{r}}}$ on the brick $\mathrm{HB}^N(\vec{\mathbf{r}})$ into the product

$$\mu^N_{<k,\vec{\mathbf{r}}}=\times_{m=0}^{k-1}\,\mu^N_{m,r_m},\ \ \mu^N_{\ge k,\vec{\mathbf{r}}}=\times_{m=k}^{\infty}\,\mu^N_{m,r_m},\ \ \mu^N_{\vec{\mathbf{r}}}=\mu^N_{<k,\vec{\mathbf{r}}}\times\mu^N_{\ge k,\vec{\mathbf{r}}}.\tag{2.29}$$

Thus, each component of the decomposition of the brick $\mathrm{HB}^N_{<k}(\vec{\mathbf{r}})$ (resp. $\mathrm{HB}^N_{\ge k}(\vec{\mathbf{r}})$) is supplied with the Lebesgue product probability measure $\mu^N_{<k,\vec{\mathbf{r}}}$ (resp. $\mu^N_{\ge k,\vec{\mathbf{r}}}$). Denote by

$$W_{<k,N}=\times_{m=0}^{k-1}\,W_{m,N},\quad W_{\ge k,N}=\times_{m=k}^{\infty}\,W_{m,N}\tag{2.30}$$

the spaces to which the brick $\mathrm{HB}^N_{<k}(\vec{\mathbf{r}})$ and the Hilbert Brick $\mathrm{HB}^N_{\ge k}(\vec{\mathbf{r}})$ belong.

Consider the decomposition with $k=2n$. Suppose we can get an estimate

$$\mu^N_{<2n,\vec{\mathbf{r}}}\{B_n(C,\delta,\rho,\vec{\mathbf{r}},f,\vec{\varepsilon}_{\ge 2n})\}\le\mu_n(C,\delta,\rho,\vec{\mathbf{r}},M_{1+\rho})\tag{2.31}$$

of the measure of the "bad" set

$$\begin{aligned}(2.32)\quad & B_n(C,\delta,\rho,\vec{\mathbf{r}},f,\vec{\varepsilon}_{\geq 2n}) \\ & = \{\vec{\varepsilon}_{<2n} \in \mathrm{HB}^N_{<2n}(\vec{\mathbf{r}}) : f_{\vec{\varepsilon}} \in \mathrm{IH}(n-1,C,\delta,\rho) \text{ but } f_{\vec{\varepsilon}} \notin \mathrm{IH}(n,C,\delta,\rho)\}.\end{aligned}$$

in each slice $\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}}) \times \{\vec{\varepsilon}_{\geq 2n}\} \subset \mathrm{HB}^N(\vec{\mathbf{r}})$, uniformly over $\vec{\varepsilon}_{\geq 2n} \in \mathrm{HB}^N_{\geq 2n}(\vec{\mathbf{r}})$. Then by the Fubini/Tonelli theorem and by the choice of the probability measure (2.29), estimate (2.31) implies (2.3). Thus we reduce the problem of estimating the measure of the "bad" set (2.1) in the infinite-dimensional Hilbert Brick $\mathrm{HB}^N(\vec{\mathbf{r}})$ to estimating the measure of the "bad" set (2.32) in the finite-dimensional brick $\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}})$ of vector-polynomials of degree $2n-1$. Now our main goal is to get an estimate for the right-hand side of (2.31).

Fix a parameter value $\vec{\varepsilon}_{\geq 2n} \in \mathrm{HB}^N_{\geq 2n}(\vec{\mathbf{r}})$ and the corresponding parameter slice $\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}}) \times \{\vec{\varepsilon}_{\geq 2n}\}$ in the Hilbert Brick $\mathrm{HB}^N(\vec{\mathbf{r}})$. Let $\tilde{f} = f_{(0,\vec{\varepsilon}_{\geq 2n})}$ be the center of this slice. In this slice we have the finite-parameter family

$$\{\tilde{f}_{\vec{\varepsilon}_{<2n}}\}_{\vec{\varepsilon}_{<2n}\in \mathrm{HB}^N_{<2n}(\vec{\mathbf{r}})} = \{f_{(\vec{\varepsilon}_{<2n},\vec{\varepsilon}_{\geq 2n})}\}_{\vec{\varepsilon}_{<2n}\in \mathrm{HB}^N_{<2n}(\vec{\mathbf{r}})} \tag{2.33}$$

of perturbations by polynomials of degree $2n-1$. This is the family we consider at the n-th stage of the induction. We redenote the "bad" set of parameter values $B_n(C,\delta,\rho,\vec{\mathbf{r}},f,\vec{\varepsilon}_{\geq 2n})$ by $B_n(C,\delta,\rho,\vec{\mathbf{r}},\tilde{f})$.

2.4. *Simple trajectories and the Inductive Hypothesis.* Based on the discussion in Section 2.1, we make the following definition.

Definition 2.4.1. A *trajectory* $x_0,\dots,x_{n-1}$ of length n of a diffeomorphism $f \in \mathrm{Diff}^r(B^N)$, where $x_k = f^k(x_0)$, is called (n,γ)-*simple* if

$$\prod_{k=0}^{n-2} |x_{n-1}-x_k| \geq \gamma^{1/(4N)}. \tag{2.34}$$

A point x_0 is called (n,γ)-*simple* if its *trajectory* $\{x_k = f^k(x_0)\}_{k=0}^{n-1}$ of length n is (n,γ)-*simple*. Otherwise a *point* (resp. a *trajectory*) is called non-(n,γ)-*simple*.

If a trajectory is simple, then perturbation of this trajectory by Newton Interpolation Polynomials is effective as the Closing Lemma and perturbation of hyperbolicity examples of Section 2.1 show. To evaluate the product of distances it is important to choose *a "good" starting point* x_0 of an almost periodic trajectory $\{x_k\}_k$ in order to have the largest possible value of the product in (2.34); for some starting points the product of distances may be artificially small.

Consider the following example of a homoclinic intersection: Let $f : B^2 \hookrightarrow B^2$ be a diffeomorphism with a hyperbolic saddle point at the origin $f(0)=0$. Suppose that the stable manifold $W^s(0)$ and the unstable manifold $W^u(0)$ intersect at some point $q \in W^s(0)\cap W^u(0)$. Then for a sufficiently large n there is a periodic point x of period n in a neighborhood of q going once nearby 0. It is clear that the trajectory $\{f^k(x)\}_{k=1}^n$ spends a lot of time in a neighborhood of the origin. Choose two starting points $x_0' = f^{k'}(x)$ and $x_0'' = f^{k''}(x)$ for the product (2.34). If x_0' is not in an $\exp(-\varepsilon n)$-neighborhood of the origin for some $\varepsilon > 0$, but x_0'' is, then it might happen that $\prod_{k=0}^{n-2} |f^{n-1}(x_0') - f^k(x_0')| \sim \exp(-\delta n)$ and $\prod_{k=0}^{n-2} |f^{n-1}(x_0'') - f^k(x_0'')| \sim \exp(-\delta' n^2)$ for some $\delta, \delta' > 0$. Indeed, if we pick out of $\{f^k(x)\}_{k=1}^n$ only the $n/2$ closest to the origin, then a simple calculation shows that all of them are in an $\exp(-\varepsilon n)$-neighborhood of the origin, where ε is some positive number depending on the eigenvalues of $df(0)$. So the first product might be significantly larger than the second one. This is because the trajectory $\{f^k(x_0'')\}_{k=0}^{n-1}$ has many points in a neighborhood of the origin and all of the corresponding terms in the product are small. This shows that sometimes the product of distances along a trajectory (2.34) might be small not because the trajectory is too recurrent, but because we chose a "bad" starting point. This motivates the following definition.

Definition 2.4.2. A point x is called *essentially* (n, γ)-*simple* if for some nonnegative $j < n$, the point $f^j(x)$ is (n, γ)-simple. Otherwise a point is called *essentially non-*(n, γ)-*simple*.

Let us return to the strategy of the proof of Theorem 1.3.9. At the n-th stage of the induction over the period we consider the family of polynomial perturbations $\{\tilde{f}_{\vec{\varepsilon}_{<2n}}\}_{\vec{\varepsilon}_{<2n} \in \mathrm{HB}^N_{<2n}(\vec{\mathbf{r}})}$ of the form (2.33) of the diffeomorphism $\tilde{f} \in \mathrm{Diff}^{1+\rho}(B^N)$ by polynomials of degree $2n-1$. Consider among them only diffeomorphisms $\tilde{f}_{\vec{\varepsilon}_{<2n}}$ that satisfy the Inductive Hypothesis of order $n-1$ with constants (C, δ, ρ); i.e., $\tilde{f}_{\vec{\varepsilon}_{<2n}} \in \mathrm{IH}(n-1, C, \delta, \rho)$ as we proposed earlier. To simplify notation we redenote the set $B_n(C, \delta, \rho, \vec{\mathbf{r}}, f, \vec{\varepsilon}_{\geq 2n})$ by $B_n(C, \delta, \rho, \vec{\mathbf{r}}, \tilde{f})$ with $\tilde{f} = f_{\vec{\varepsilon}_{\geq 2n}}$. Our main goal is to estimate the measure of "bad" parameter values $B_n(C, \delta, \rho, \vec{\mathbf{r}}, \tilde{f})$, given by (2.32), for which the corresponding diffeomorphism has an $(n, \gamma_n^{1/\rho}(C, \delta))$-periodic, but not $(n, \gamma_n(C, \delta))$-hyperbolic, point $x \in B^N$.

We split the set of all possible almost periodic points of period n into two classes: *essentially* $(n, \gamma_n(C, \delta))$-*simple* and *essentially non-*$(n, \gamma_n(C, \delta))$-*simple*. Decompose the set of "bad" parameters $B_n(C, \delta, \rho, \vec{\mathbf{r}}, \tilde{f})$ into two sets of "bad" parameters with simple and nonsimple almost periodic points that are not sufficiently hyperbolic:

$$(2.35)\quad \begin{aligned} B_n^{\mathrm{sim}}(C, \delta, \rho, \vec{\mathbf{r}}, \tilde{f}) = \{\vec{\varepsilon} \in \mathrm{HB}^N(\vec{\mathbf{r}}) : \tilde{f}_{\vec{\varepsilon}_{<2n}} \in \mathrm{IH}(n-1, C, \delta, \rho),\\ \tilde{f}_{\vec{\varepsilon}_{<2n}} \text{ has an } (n, \gamma_n^{1/\rho}(C, \delta))\text{-periodic, essentially}\\ (n, \gamma_n(C, \delta))\text{-simple, but not } (n, \gamma_n(C, \delta))\text{-hyperbolic point } x\} \end{aligned}$$

and

(2.36)
$$B_n^{\rm non}(C,\delta,\rho,\vec{\mathbf{r}},\tilde f) = \{\vec\varepsilon \in {\rm HB}^N(\vec{\mathbf{r}}) : \tilde f_{\vec\varepsilon_{<2n}} \in {\rm IH}(n-1,C,\delta,\rho),$$
$$\tilde f_{\vec\varepsilon_{<2n}} \text{ has an } (n,\gamma_n^{1/\rho}(C,\delta))\text{-periodic, essentially}$$
$$\text{non-}(n,\gamma_n(C,\delta))\text{-simple, but not } (n,\gamma_n(C,\delta))\text{-hyperbolic point } x\}.$$

It is clear that we have

$$B_n(C,\delta,\rho,\vec{\mathbf{r}},\tilde f) = B_n^{\rm sim}(C,\delta,\rho,\vec{\mathbf{r}},\tilde f) \cup B_n^{\rm non}(C,\delta,\rho,\vec{\mathbf{r}},\tilde f). \tag{2.37}$$

We shall estimate the measures of the sets of simple orbits $B_n^{\rm sim}(C,\delta,\rho,\vec{\mathbf{r}},\tilde f)$ and nonsimple orbits $B_n^{\rm non}(C,\delta,\rho,\vec{\mathbf{r}},\tilde f)$ separately, but using very similar methods.

Let $\tilde f_{\vec\varepsilon_{<2n}} \in {\rm IH}(n-1,C,\delta,\rho)$ be a diffeomorphism that satisfies the Inductive Hypothesis of order $n-1$ with constants (C,δ,ρ). It turns out that if $\tilde f_{\vec\varepsilon_{<2n}}$ has an $(n,\gamma_n^{1/\rho}(C,\delta))$-periodic and essentially non-$(n,\gamma_n(C,\delta))$-simple point x_0, then the trajectory of x_0 has a close return $\tilde f^k_{\vec\varepsilon_{<2n}}(x_0) = x_k$ for $k<n$ such that distance $|x_0-x_k|$ is much smaller of all the previous $|x_0-x_j|,\ 1\le j<k$. Let us formulate more precisely what we mean here by "much smaller".

Definition 2.4.3. Let $g\in {\rm Diff}^{1+\rho}(B^N)$ be a diffeomorphism and let $D>1$ be some number. A point $x_0\in B^N$ (resp. a trajectory $x_0,\dots,x_{n-1}=g^{n-1}(x_0)$ $\subset B^N$ of length n) has a *weak (D,n)-gap* at a point $x_k=g^k(x_0)$ if

$$|x_k-x_0| \le D^{-n} \min_{0<j\le k-1} |x_0-x_j| \tag{2.38}$$

and there is no $m<k$ such that x_0 has a weak (D,n)-gap at $x_m=g^m(x_0)$.

Remark 2.4.4. The term "gap" arises by consideration of the sequence $-\log|x_0-x_1|,\ -\log|x_0-x_2|,\dots,-\log|x_0-x_k|$. Definition 2.4.3 implies that the last term is significantly larger then all the previous terms.

Let us show that n should be divisible by k for an almost periodic point of period n with a weak gap at x_k. This feature of a weak gap allows us to treat almost periodic trajectories of length n with a weak gap at x_k as n/k almost identical parts of length k each.

LEMMA 2.4.5. *Let $g\in {\rm Diff}^{1+\rho}(B^N)$ be a diffeomorphism, M_1 be an upper bound on the C^1-norm of g and g^{-1}, $D>M_1^2$, and let x_0 have a weak (D,n)-gap at x_k and $|x_0-x_n|\le|x_0-x_k|$. Then n is divisible by k.*

Sketch of Proof. Denote by $\gcd(k,n)$ the greatest common divisor of k and n. Then using the bound on the C^1-norm of g and g^{-1} for any $x,y\in B^N$ we have

$$M_1^{-1}\ |g^{-1}(x)-g^{-1}(y)| \le |x-y| \le M_1\ |g(x)-g(y)|. \tag{2.39}$$

Using the Euclidean division algorithm developed in Part II of this paper, one can show that

$$|x_0 - x_{\gcd(k,n)}| \le M_1^{2n} D^{-n} \min_{0<j\le k-1} |x_0 - x_j|. \tag{2.40}$$

Since $D > M_1^2$, we cannot have $\gcd(k,n) < k$, so n must be divisible by k. Q.E.D.

In Part II of this paper we prove the following result.

THEOREM 2.4.6. *Let $g \in \mathrm{Diff}^{1+\rho}(B^N)$ be a diffeomorphism for some $\rho > 0$ and satisfy the Inductive Hypothesis of order $n-1$ with constants (C,δ,ρ), i.e. $g \in \mathrm{IH}(n-1, C, \delta, \rho)$ and let $M_{1+\rho} = \max\{\|g^{-1}\|_{C^1},\ \|g\|_{C^{1+\rho}},\ 2^{1/\rho}\}$, $C > 100\rho^{-1}\delta^{-1}\log M_{1+\rho}$, and $D = \max\{M_{1+\rho}^{30/\rho}, \exp(C/100)\}$. Suppose the diffeomorphism g has an $(n, \gamma_n^{1/\rho}(C,\delta))$-periodic and essentially non-$(n, \gamma_n(C,\delta))$-simple point $x_0 \in B^N$. Then either x_0 is $(n,\gamma_n(C,\delta))$-hyperbolic or x_0 has a weak (D,n)-gap at $x_k = g^k(x_0)$ for some k dividing n and x_j is $(k,\gamma_n(C,\delta))$-simple for some $j < n$.*

Remark 2.4.7. As a matter of fact we need a sharper result, but Theorem 2.4.6 is a nice starting point.

Theorem 2.4.6 implies that the set of "bad" parameters with an essentially nonsimple trajectory can be decomposed into the following finite union: Define the set of parameters with an almost periodic point of period n with a weak gap at the k-th point of its trajectory.

(2.41)

$$\begin{aligned} B_n^{\mathrm{wgap}(k)}(C,\delta,\rho,\vec{\mathbf{r}},\tilde f; D) = \{\vec\varepsilon \in \mathrm{HB}^N(\vec{\mathbf{r}}) : \tilde f_{\vec\varepsilon} \in \mathrm{IH}(n-1,C,\delta,\rho),\\ \tilde f_{\vec\varepsilon} \text{ has an } (n,\gamma_n^{1/\rho}(C,\delta))\text{-periodic, but not}\\ (n,\gamma_n(C,\delta))\text{-hyperbolic point } x_0 \text{ with a weak } (D,n)\text{-gap at } x_k = \tilde f_{\vec\varepsilon}^k(x)\}. \end{aligned}$$

Then for $D = \max\{M_{1+\rho}^{30/\rho}, \exp(C/100)\}$, Theorem 2.4.6 gives

$$B_n^{\mathrm{non}}(C,\delta,\rho,\vec{\mathbf{r}},\tilde f) \subseteq \left(\cup_{k|n} B_n^{\mathrm{wgap}(k)}(C,\delta,\rho,\vec{\mathbf{r}},\tilde f; D)\right). \tag{2.42}$$

Combining inclusions (2.35) and (2.42), we have

$$B_n(C,\delta,\rho,\vec{\mathbf{r}},\tilde f) \subseteq B_n^{\mathrm{sim}}(C,\delta,\rho,\vec{\mathbf{r}},\tilde f) \cup \left(\cup_{k|n} B_n^{\mathrm{wgap}(k)}(C,\delta,\rho,\vec{\mathbf{r}},\tilde f; D)\right). \tag{2.43}$$

Thus we need to get estimates on the measures of bad parameters associated with essentially simple trajectories $B_n^{\mathrm{sim}}(C,\delta,\rho,\vec{\mathbf{r}},\tilde f)$ and trajectories with a weak gap $B_n^{\mathrm{wgap}(k)}(C,\delta,\rho,\vec{\mathbf{r}},\tilde f; D)$, where k divides n. In Chapter 3, we describe the Discretization method for the 1-dimensional model problem. This method will allow us to estimate the measure of parameters $B_n^{\mathrm{sim}}(C,\delta,\rho,\vec{\mathbf{r}},\tilde f)$ associated with simple almost periodic points. At the end of Chapter 3, we show how using the Discretization method one can estimate the measure of parameters $B_n^{\mathrm{wgap}(k)}(C, \delta, \rho, \vec{\mathbf{r}}, \tilde{f}; D)$ associated with almost periodic trajectories with a weak gap. Loosely speaking, it is because those trajectories have the simple parts of length k (see the end of Theorem 2.4.6), and hyperbolicity of the simple part of length k enforces hyperbolicity of the trajectories of length n. See also diagrams (12) and (13) in [GHK].

3. A model problem: C^2-smooth maps of the interval $I = [-1, 1]$

In Section 2.4 we concluded that the key to the proof of Theorem 1.3.9 (which implies the Main Theorem) is to get an estimate of the measure of "bad" parameters. Recall that the set of "bad" parameters (2.32) consists of those parameters $\vec{\varepsilon} \in \mathrm{HB}^N(\vec{\mathbf{r}})$ for which the corresponding diffeomorphism $f_{\vec{\varepsilon}} : B^N \hookrightarrow B^N$ has an almost periodic point x of period n that is not sufficiently hyperbolic. In this chapter we present a detailed discussion of C^2-smooth 1-dimensional noninvertible maps ($N = 1$ and $\rho = 1$) with a Hilbert Brick of a "nice" size. This 1-dimensional model gives a useful insight into the general approach of estimating the measure of "bad" parameters for the N-dimensional $C^{1+\rho}$-smooth diffeomorphisms and allows us to avoid several technical complications that will arise in Part II of this paper [K5]. These complications are outlined in the next chapter.

3.1. *Setting up of the model.* Let $C^2(I, I)$ be the space of C^2-smooth maps of the interval $I = [-1, 1]$ into its interior. Consider a C^2-smooth map of the interval $f \in C^2(I, I)$ and the family of perturbations of f by analytic functions represented as their power series

$$f_\varepsilon(x) = f(x) + \sum_{k=0}^{\infty} \varepsilon_k x^k. \tag{3.1}$$

Fix a positive $\tau > 0$. Define a range of parameters of this family in the form of a Hilbert Brick

$$\mathrm{HB}^{\mathrm{st}}(\tau) = \left\{ \{\varepsilon_m\}_{m=0}^{\infty} : \ \forall\, m \geq 0,\ |\varepsilon_m| < \frac{\tau}{m!} \right\}. \tag{3.2}$$

We call $\mathrm{HB}^{\mathrm{st}}(\tau)$ *a Hilbert Brick of standard thickness* with width τ. If we choose τ small enough, then the whole family $\{f_\varepsilon\}_{\varepsilon \in \mathrm{HB}^{\mathrm{st}}(\tau)} \subset C^2(I, I)$ consists of C^2-smooth maps of the interval I.

Define the Lebesgue product probability measure, denoted by μ_τ^{st}, on the Hilbert Brick of parameters $\mathrm{HB}^{\mathrm{st}}(\tau)$ by normalizing the 1-dimensional Lebesgue measure along each component to the 1-dimensional Lebesgue probability measure

$$\mu_{m,\tau}^{\mathrm{st}} = \left(\frac{m!}{2\tau}\right) Leb_1, \quad \mu_{<k,\tau}^{\mathrm{st}} = \times_{m=0}^{k-1} \mu_{m,\tau}^{\mathrm{st}} \quad \mu_\tau^{\mathrm{st}} = \times_{m=0}^{\infty} \mu_{m,\tau}^{\mathrm{st}}. \tag{3.3}$$

The main result of this chapter is the following 1-dimensional analogue of Theorem 1.3.9.

THEOREM 3.1.1. *Let $f \in C^2(I, I)$ be a C^2-smooth map of the interval I into its interior and let $\tau > 0$ be so small that the family of analytic perturbations $\{f_\varepsilon\}_{\varepsilon \in \mathrm{HB}^{\mathrm{st}}(\tau)} \subset C^2(I, I)$ consists of C^2-smooth maps of the interval I. Then for any $\delta > 0$ and μ_τ^{st}-a.e. $\varepsilon \in \mathrm{HB}^{\mathrm{st}}(\tau)$ there exists $C = C(\varepsilon, \delta) > 0$ such that the number of periodic points $P_n(f_\varepsilon)$ of f_ε of period n and their minimal hyperbolicity $\gamma_n(f_\varepsilon)$, defined in* (1.7), *for all $n \in \mathbb{Z}_+$ satisfy*

$$\gamma_n(f_\varepsilon) > \exp(-Cn^{1+\delta}), \quad P_n(f_\varepsilon) < \exp(Cn^{1+\delta}). \tag{3.4}$$

The strategy for the proof of this theorem is the same as the strategy of the proof of Theorem 1.3.9 described in Chapter 2. Denote the supremum C^2 and C^1-norms of the family (3.1)

$$M_1 = \sup_{\varepsilon \in \mathrm{HB}^{\mathrm{st}}(\tau)} \{\|f_\varepsilon\|_{C^1}\}, \quad M_2 = \sup_{\varepsilon \in \mathrm{HB}^{\mathrm{st}}(\tau)} \{\|f_\varepsilon\|_{C^2}, M_1, 2\}. \tag{3.5}$$

By analogy with the direct decomposition of the Hilbert Brick in the N-dimensional case (2.27), for each positive integer $k \in \mathbb{Z}_+$ define the direct decomposition of the Hilbert Brick of standard thickness $\mathrm{HB}^{\mathrm{st}}(\tau)$

$$\begin{aligned} \mathrm{HB}^{\mathrm{st}}_{<k}(\tau) &= \Big\{\{\varepsilon_m\}_{m=0}^{k-1} :\ \forall\ 0 \le m < k,\ |\varepsilon_m| < \frac{\tau}{m!}\Big\}, \\ \mathrm{HB}^{\mathrm{st}}_{\ge k}(\tau) &= \Big\{\{\varepsilon_m\}_{m=k}^{\infty} :\ \forall\ m \ge k,\ |\varepsilon_m| < \frac{\tau}{m!}\Big\}. \end{aligned} \tag{3.6}$$

We call $\mathrm{HB}^{\mathrm{st}}_{<k}(\tau)$ *a* (k*-dimensional*) *Brick of standard thickness* with width τ. The product measure μ_τ^{st} on the whole Hilbert Brick $\mathrm{HB}^{\mathrm{st}}(\tau)$ induces the measure of the product of Lebesgue probability $\mu^{\mathrm{st}}_{<k,\tau}$ on the k-dimensional Brick $\mathrm{HB}^{\mathrm{st}}_{<k}(\tau)$.

Fix $n \in \mathbb{Z}_+$ and consider the n-th stage of the induction over the period (see the beginning of Chapter 2). Let

$$\tilde{f}(x) = f(x) + \sum_{k=2n}^{\infty} \varepsilon_k x^k \tag{3.7}$$

for some $\{\varepsilon_k\}_{k=2n}^{\infty} \in \mathrm{HB}^{\mathrm{st}}_{\ge 2n}(\tau)$, and consider the $2n$-parameter family of perturbations by polynomials of degree $2n-1$ with coefficients in the brick of standard thickness $\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$,

$$\tilde{f}_\varepsilon(x) = \tilde{f}(x) + \sum_{k=0}^{2n-1} \varepsilon_k x^k, \quad \varepsilon = (\varepsilon_0, \dots, \varepsilon_{2n-1}) \in \mathrm{HB}^{\mathrm{st}}_{<2n}(\tau). \tag{3.8}$$

The bounds M_1 and M_2 from (3.5) apply to this subfamily of (3.1).

Using the Fubini reduction to finite-dimensional families from Section 2.3 right after (2.31), for the proof of Theorem 3.1.1 it is sufficient to estimate the measure of "bad" parameters in each such family.

To fit the notation of our model we choose a sufficiently small positive γ_n and we introduce sets of all "bad" parameters (compare with (2.32)):

$$B_{n,\tau}(C,\delta,\tilde{f},\gamma_n) = \{\varepsilon \in \mathrm{HB}^{\mathrm{st}}_{<2n}(\tau) : \tilde{f}_\varepsilon \in \mathrm{IH}(n-1,C,\delta,1),\ \tilde{f}_\varepsilon \text{ has an } (n,\gamma_n)\text{-periodic, but not } (n,\gamma_n)\text{-hyperbolic point } x_0\}, \tag{3.9}$$

and define the sets $B^{\mathrm{sim}}_{n,\tau}(C,\delta,\tilde{f},\gamma_n)$ and $B^{\mathrm{non}}_{n,\tau}(C,\delta,\tilde{f},\gamma_n)$ of "bad" parameters with essentially simple (respectively nonsimple) trajectories as in (2.35) and (2.36):

$$B^{\mathrm{sim}}_{n}(C,\delta,\tilde{f},\gamma_n) = \{\vec{\varepsilon} \in \mathrm{HB}^{\mathrm{st}}_{<2n}(\tau) : \tilde{f}_\varepsilon \in \mathrm{IH}(n-1,C,\delta,1),\ \tilde{f}_\varepsilon \text{ has an } (n,\gamma_n)\text{-periodic, essentially } (n,\gamma_n)\text{-simple, but not } (n,\gamma_n)\text{-hyperbolic point } x_0\}, \tag{3.10}$$

and

$$B^{\mathrm{non}}_{n}(C,\delta,\tilde{f},\gamma_n) = \{\vec{\varepsilon} \in \mathrm{HB}^{\mathrm{st}}_{<2n}(\tau) : \tilde{f}_\varepsilon \in \mathrm{IH}(n-1,C,\delta,1),\ \tilde{f}_\varepsilon \text{ has an } (n,\gamma_n)\text{-periodic, essentially non-}(n,\gamma_n)\text{-simple, but not } (n,\gamma_n)\text{-hyperbolic point } x_0\}. \tag{3.11}$$

For sufficiently small γ_n, e.g., $\gamma_n \le \gamma_n(C,\delta)$, similarly to (2.37) we have the following decomposition,

$$B_{n,\tau}(C,\delta,\tilde{f},\gamma_n) \ = \ B^{\mathrm{sim}}_{n,\tau}(C,\delta,\tilde{f},\gamma_n) \cup B^{\mathrm{non}}_{n,\tau}(C,\delta,\tilde{f},\gamma_n). \tag{3.12}$$

The main result of the next three sections is the following estimate.

PROPOSITION 3.1.2. *Let $\{\tilde{f}_\varepsilon\}_{\varepsilon\in \mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)}$ be the family of polynomial perturbations* (3.8) *with bound M_2 on the C^2-norm. Then with the notation above, for any $C>2$, $\delta>0$, and $\tau>0$ and a sufficiently small positive γ_n, e.g., $\gamma_n \le \gamma_n(C,\delta)$, the following estimate on the measure of parameters associated with maps $\tilde{f}_\varepsilon$ with an (n,γ_n)-periodic, essentially (n,γ_n)-simple, but not (n,γ_n)-hyperbolic, point holds*:

$$\mu^{\mathrm{st}}_{<2n,\tau}\big\{B^{\mathrm{sim}}_{n,\tau}(C,\delta,\tilde{f},\gamma_n)\big\} \le 6^{2n} M_2^{6n+1}\ \frac{(n-1)!}{\tau}\ \frac{(2n-1)!}{\tau}\ \gamma_n^{1/4}. \tag{3.13}$$

It is clear that for any $C>0$ and $\delta>0$, if $\gamma_n = \exp(-Cn^{1+\delta})$, then the right-hand side of (3.13) tends to 0 as $n\to\infty$ superexponentially fast in n. An estimate on the measure of essentially nonsimple trajectories $\mu^{\mathrm{st}}_{<2n,\tau}\big\{B^{\mathrm{non}}_{n}(C,\delta,\tilde{f},\gamma_n)\big\}$ is obtained in Section 3.5, Proposition 3.5.1. Application of these two propositions and arguments (2.3, 2.4) will prove Theorem 3.1.1. The method of obtaining an estimate for $\mu^{\mathrm{st}}_{<2n,\tau}\big\{B^{\mathrm{non}}_{n,\tau}(C,\delta,\tilde{f},\gamma_n)\big\}$ is similar to the one we shall develop now to prove (3.13). See also Sections 3–5 in [GHK] or Section 11 in [GK].

3.2. *Decomposition into pseudotrajectories.* In this section, we decompose the set of "bad" parameters $B^{\rm sim}_{n,\tau}(C,\delta,\tilde f,\gamma_n)$ for which there exists a simple, almost periodic, but not sufficiently hyperbolic trajectory into a finite union of sets of "bad" parameters. Each set will be associated with a particular simple, almost periodic, but not sufficiently hyperbolic pseudotrajectory. In the next section we will estimate the measure of "bad" parameters associated with a *particular* trajectory, and in the subsequent section we will extend this estimate to the set of "bad" parameters associated with *all possible* simple trajectories, obtaining estimate (3.13).

Fix a sufficiently small $\gamma_n > 0$ and $\tilde\gamma_n = \gamma_n M_2^{-2n}$. Consider the $2\tilde\gamma_n$-grid in the interval I

$$I_{\tilde\gamma_n} = \{x \in I :\ \exists k \in \mathbb{Z} \text{ such that } x = (2k+1)\tilde\gamma_n\} \subset I. \tag{3.14}$$

Definition 3.2.1. We call a k-tuple $\{x_j\}_{j=0}^{k-1} \in I^k_{\tilde\gamma_n}$ *a $\tilde\gamma_n$-pseudotrajectory associated to ε (or to the map $\tilde f_\varepsilon$)* if for each $j = 1,\dots,k-1$ we have $|\tilde f_\varepsilon(x_{j-1}) - x_j| \le \tilde\gamma_n$, and we call it *a $\tilde\gamma_n$-pseudotrajectory associated to* $\mathrm{HB}^{\rm st}_{<2n}(\tau)$ (or to the family $\{\tilde f_\varepsilon\}_{\varepsilon\in \mathrm{HB}^{\rm st}_{<2n}(\tau)}$) if it is associated to some $\varepsilon \in \mathrm{HB}^{\rm st}_{<2n}(\tau)$.

A $\tilde\gamma_n$-pseudotrajectory $x_0,\dots,x_{n-1}$ of length n associated to some parameter $\varepsilon \in \mathrm{HB}^{\rm st}_{<2n}(\tau)$ for some $\gamma > 0$ is

- (n,γ)-*periodic* if $|\tilde f_\varepsilon(x_{n-1}) - x_0| \le \gamma$,
- (n,γ)-*simple* if $\prod_{j=0}^{n-2} |x_{n-1} - x_j| \ge \gamma^{1/4}$,
- (n,γ)-*hyperbolic* if $\left|\prod_{j=0}^{n-1} \left|(\tilde f_\varepsilon)'(x_j)\right| - 1\right| \ge \gamma$.

Remark 3.2.2. For fixed $\varepsilon \in \mathrm{HB}^{\rm st}_{<2n}(\tau)$, each initial point $x_0 \in I_{\tilde\gamma_n}$ generates a $\tilde\gamma_n$-pseudotrajectory $\tilde x_0, \tilde x_1, \dots, \tilde x_{n-1}$ of length n as follows. For each successive $k = 1,\dots,n-1$, choose $\tilde x_k \in I_{\tilde\gamma_n}$ such that $|\tilde x_k - f_\varepsilon(\tilde x_{k-1})| \le \tilde\gamma_n$. Notice that this choice is unique unless $\tilde f_\varepsilon(\tilde x_{k-1})$ happens to lie halfway between two points of $I_{\tilde\gamma_n}$. It may be helpful in understanding the upcoming arguments to think of each initial point $x_0 \in I_{\tilde\gamma_n}$ as generating a unique $\tilde\gamma_n$-pseudotrajectory for a given f_ε, though for a measure zero set of ε there are exceptions to this rule. In fact, for our estimates it is important only that the number of $\tilde\gamma_n$-pseudotrajectories per initial point be bounded by an exponential function of n, which is true in this case even if there is a choice of two grid points at each iteration.

We would like to contain the set of "bad" parameters $B^{\rm sim}_{n,\tau}(C,\delta,\tilde f,\gamma_n)$ in a finite collection of subsets each of "bad" parameters corresponding to a single

$\tilde{\gamma}_n$-pseudotrajectory

$$(3.15)\quad B_{n,\tau}^{\mathrm{sim},\tilde{\gamma}_n}(\tilde{f},\gamma_n,M_2;x_0,\dots,x_{n-1})$$
$$= \{\varepsilon \in \mathrm{HB}^{\mathrm{st}}_{<2n}(\tau) : \{x_k\}_{k=0}^{n-1} \text{ is a } \tilde{\gamma}_n\text{-pseudotrajectory associated to } \varepsilon \text{ and is } \left(n,\frac{\gamma_n}{2}\right)\text{-simple and } (n, M_2^n\gamma_n)\text{-periodic, but not } (n, M_2^{3n}\gamma_n)\text{-hyperbolic}\}.$$

We introduce the union of all "bad" sets associated with $\tilde{\gamma}_n$-pseudotrajectories

$$(3.16)\quad B_{n,\tau}^{\mathrm{sim},\tilde{\gamma}_n}(\tilde{f},\gamma_n,M_2) = \cup_{\{x_0,\dots,x_{n-1}\}\subset I^n_{\tilde{\gamma}_n}} B_{n,\tau}^{\mathrm{sim},\tilde{\gamma}_n}(\tilde{f},\gamma_n,M_2;x_0,\dots,x_{n-1}).$$

Most of the sets in the right-hand side are empty, and one of our goals is to estimate the number of nonempty ones.

In comparison to the definition (3.10) of $B_{n,\tau}^{\mathrm{sim}}(C,\delta,\tilde{f},\gamma_n)$, we increase periodicity and hyperbolicity for pseudotrajectories and decrease simplicity. This will allow us to prove that

$$(3.17)\qquad B_{n,\tau}^{\mathrm{sim}}(C,\delta,\tilde{f},\gamma_n) \subset B_{n,\tau}^{\mathrm{sim},\tilde{\gamma}_n}(\tilde{f},\gamma_n,M_2).$$

Intuitively this is true because each trajectory of length n can be approximated by a pseudotrajectory of length n which has almost the same periodicity, simplicity, and hyperbolicity as the original one. We will make this argument precise at the end of Section 3.4.

Remark 3.2.3. Unlike $B_{n,\tau}^{\mathrm{sim}}(C,\delta,\tilde{f},\gamma_n)$, we do not assume in the definition (3.15) of $B_{n,\tau}^{\mathrm{sim},\tilde{\gamma}_n}(\tilde{f},\gamma_n,M_2)$ that $\tilde{f}_\varepsilon \in \mathrm{IH}(n-1,C,\delta,1)$. This is because we only need the Inductive Hypothesis to estimate the measure of "bad" parameters in the case of nonsimple trajectories.

Our goal is then to estimate the measure $\mu^{\mathrm{st}}_{<2n,\tau}\{B_{n,\tau}^{\mathrm{sim},\tilde{\gamma}_n}(\tilde{f},\gamma_n,M_2)\}$ in order to prove Proposition 3.1.2. Loosely speaking, this measure will be estimated in two steps:

Step 1. Estimate the number of different $\tilde{\gamma}_n$-pseudotrajectories $\#_n(\tilde{\gamma}_n,\tau)$ associated to $\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$;

Step 2. Estimate the measure

$$(3.18)\qquad \mu^{\mathrm{st}}_{<2n,\tau}\{B_{n,\tau}^{\mathrm{sim},\tilde{\gamma}_n}(\tilde{f},\gamma_n,M_2;x_0,\dots,x_{n-1})\} \le \mu_n(M_2,\gamma_n,\tilde{\gamma}_n,\tau)$$

uniformly for an (n,γ_n)-simple $\tilde{\gamma}_n$-pseudotrajectory $\{x_0,\dots,x_{n-1}\} \in I^n_{\tilde{\gamma}_n}$.

Then the product of two numbers $\#_n(\tilde{\gamma}_n,\tau)$ and $\mu_n(M_2,\gamma_n,\tilde{\gamma}_n,\tau)$ that are obtained in Steps 1 and 2 gives the required estimate (3.13).

Actually the procedure of estimating $\mu^{\mathrm{st}}_{<2n,\tau}\{B_{n,\tau}^{\mathrm{sim},\tilde{\gamma}_n}(\tilde{f},\gamma_n,M_2)\}$ is a little more complicated. Based on the definition (3.15) of the set

$$B_{n,\tau}^{\mathrm{sim},\tilde{\gamma}_n}(\tilde{f},\gamma_n,M_2;x_0,\dots,x_{n-1})$$

of parameters ε for which the diffeomorphism f_ε has a prescribed $\tilde{\gamma}_n$-pseudotrajectory $\{x_0, \ldots, x_{n-1}\} \in I^n_{\tilde{\gamma}_n}$ that is almost periodic and not sufficiently hyperbolic, define a set of parameters ε for which only a part of the $\tilde{\gamma}_n$-pseudotrajectory $\{x_0, \ldots, x_{m-1}\} \in I^m_{\tilde{\gamma}_n}$ is prescribed for f_ε:

$$B^{\mathrm{sim},\tilde{\gamma}_n}_{n,\tau}(\tilde{f}, \gamma_n, M_2; x_0, \ldots, x_{m-1}) = \{\varepsilon \in \mathrm{HB}^{\mathrm{st}}_{<2n}(\tau) : \text{ there exist } x_m, \ldots, x_{n-1} \in I_{\tilde{\gamma}_n} \text{ such that } \{x_j\}_{j=0}^{n-1} \text{ is a } \tilde{\gamma}_n\text{-pseudotrajectory associated to } \varepsilon, \text{ and } \varepsilon \in B^{\mathrm{sim},\tilde{\gamma}_n}_{n,\tau}(\tilde{f}, \gamma_n, M_2; x_0, \ldots, x_{n-1})\}. \tag{3.19}$$

For each $m = 1, 2, \ldots, n-1$ it is clear that

$$B^{\mathrm{sim},\tilde{\gamma}_n}_{n,\tau}(\tilde{f}, \gamma_n, M_2; x_0, \ldots, x_{m-1}) = \cup_{x_m \in I_{\tilde{\gamma}_n}} B^{\mathrm{sim},\tilde{\gamma}_n}_{n,\tau}(\tilde{f}, \gamma_n, M_2; x_0, \ldots, x_m). \tag{3.20}$$

Inductive application of this formula to the definition (3.16) yields

$$B^{\mathrm{sim},\tilde{\gamma}_n}_{n,\tau}(\tilde{f}, \gamma_n, M_2) = \cup_{\tilde{x}_0 \in I_{\tilde{\gamma}_n}} B^{\mathrm{sim},\tilde{\gamma}_n}_{n,\tau}(\tilde{f}, \gamma_n, M_2; \tilde{x}_0). \tag{3.21}$$

The estimate of Step 1 then breaks down as follows:

$$\#_n(\tilde{\gamma}_n, \tau) \approx \boxed{\begin{array}{c}\#\text{ of initial}\\ \text{points of } I_{\tilde{\gamma}_n}\end{array}} \times \boxed{\begin{array}{c}\#\text{ of } \tilde{\gamma}_n\text{-pseudotrajectories}\\ \text{per initial point}\end{array}}. \tag{3.22}$$

And up to an exponential function of n, the estimate of Step 2 breaks down as:

$$\mu_n(M_2, \gamma_n, \tilde{\gamma}_n, \tau) \approx \frac{\boxed{\begin{array}{c}\text{Measure of}\\ \text{periodicity}\end{array}} \times \boxed{\begin{array}{c}\text{Measure of}\\ \text{hyperbolicity}\end{array}}}{\boxed{\begin{array}{c}\#\text{ of } \tilde{\gamma}_n\text{-pseudotrajectories}\\ \text{per initial point}\end{array}}}. \tag{3.23}$$

(Roughly speaking, the terms in the numerator represent respectively the measure of parameters for which a given initial point will be (n, γ_n)-periodic and the measure of parameters for which a given n-tuple is (n, γ_n)-hyperbolic; they correspond to estimates (3.30) and (3.33) in the next section.) Thus after cancellation, the estimate of the measure of "bad" set $B^{\mathrm{sim}}_{n,\tau}(C, \delta, \tilde{f}, \gamma_n)$ associated to simple, almost periodic, not sufficiently hyperbolic trajectories becomes:

$$\boxed{\begin{array}{c}\text{Measure of bad}\\ \text{parameters}\end{array}} \le \boxed{\begin{array}{c}\#\text{ of initial}\\ \text{points of } I_{\tilde{\gamma}_n}\end{array}} \times \boxed{\begin{array}{c}\text{Measure of}\\ \text{periodicity}\end{array}} \times \boxed{\begin{array}{c}\text{Measure of}\\ \text{hyperbolicity}\end{array}}. \tag{3.24}$$

The first term on the right-hand side of (3.24) is of order γ_n^{-1} (up to an exponential function in n). In Section 3.3, we will show that the second term is at most of order $n!\gamma_n^{3/4}$, and the third term is at most of order $(2n)!\gamma_n^{1/2}$, so that the product on the right-hand side of (3.24) is of order at most $n!(2n)!\gamma_n^{1/4}$

(up to an exponential function in n) and is superexponentially small in n. These bounds use the change of parameter coordinates by Newton interpolation polynomials that was introduced in Section 2.2, and they *do not depend on whether the parameters are associated with the brick* $\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$, except in that we use the bound M_2 on the C^2 norm of the maps involved.

In Section 3.4, we complete the proof of Proposition 3.1.2 by bounding the total measure of "bad" parameters for all pseudotrajectories associated to $\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$. Since we use the Fubini/Tonelli theorem in the Newton coordinates $u_0, \dots, u_{2n-1}$, we need to know the maximum range of each of these parameters in the image of $\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$ under this coordinate change. In the "Distortion Lemma", we show that the image of $\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$ is contained in a brick 3 times as large in each direction. Then, in the "Collection Lemma", we show in effect that the cancellation in going from (3.22) and (3.23) to (3.24) is valid. In fact, the number of $\tilde{\gamma}_n$-pseudotrajectories for a given initial point may depend significantly on the initial point, and we do not bound it explicitly. Rather, we show that in the decomposition (3.21), the measure of each term $B^{\mathrm{sim},\tilde{\gamma}_n}_{n,\tau}(\tilde{f}, \gamma_n, M_2; \tilde{x}_0)$ is bounded (up to a factor exponential in n) by the product of the "measure of periodicity" and "measure of hyperbolicity" derived in Section 3.3, thus yielding a final estimate of the form (3.24).

3.3. *Application of Newton interpolation polynomials to estimate the measure of "bad" parameters for a single trajectory.* In this section we fix an n-tuple of points $\{x_j\}_{j=0}^{n-1} \in I^n$, denoted by $\mathbf{X}_n$, and estimate the measure of "bad" parameters $B^{\mathrm{sim},\tilde{\gamma}_n}_{n,\tau}(\tilde{f}, \gamma_n, M_2; x_0, \dots, x_{n-1})\big\}$ associated with *this particular* trajectory. See also Section 4 in [GHK]. Recall that $\tilde{\gamma}_n = M_2^{-2n}\gamma_n$.

Problem 3.3.1. Estimate the measure of $\varepsilon \in \mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$ for which the n-tuple $\{x_j\}_{j=0}^{n-1}$ is

(3.25)

A) a $\tilde{\gamma}_n$-pseudotrajectory, i.e., $|\tilde{f}_\varepsilon(x_j) - x_{j+1}| \le \tilde{\gamma}_n$ for $j = 0, \dots, n-2$;

B) (n, γ_n)-periodic, i.e., $|\tilde{f}_\varepsilon(x_{n-1}) - x_0| \le \gamma_n$; and

C) not (n, γ_n)-hyperbolic, i.e., $\left| \prod_{j=0}^{n-1} |(f_\varepsilon)'(x_j)| - 1 \right| \le \gamma_n$.

Recall the definitions and notation of Sections 2.2 and 2.3. In particular, $W_{<2n,1}$ is the space of polynomials of degree $2n-1$ with the standard basis $\{x^m\}_{m=0}^{2n-1}$. The measure $\mu^{\mathrm{st}}_{<2n,\tau}$ defined on the brick $\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau) \in W_{<2n,1}$ by (3.3) extends naturally to $W_{<2n,1}$ using the same formulas. Denote by $W^{u,\mathbf{X}_n}_{<2n,1}$ the same space of polynomials of degree $2n-1$, but with the Newton basis (2.26). Lemma 2.2.2 implies that the Newton Map $\mathcal{L}^1_{\mathbf{X}_n} : W_{<2n,1} \to W^{u,\mathbf{X}_n}_{<2n,1}$

defined by (2.24) preserves the measure $\mu^{\mathrm{st}}_{<2n,\tau}$. In other words, the definition (3.3) produces the same measure whether the standard basis or Newton basis is used.

Now we will estimate the measure of "bad" parameters for a particular trajectory using the Newton basis, without regard (except in the final hyperbolicity estimate) to whether the parameters $u = (u_0, u_1, \ldots, u_{2n-1})$ lie in the image $\mathcal{L}^1_{\mathbf{X}_n}(\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau))$ of the brick we are concerned with. For a fixed n-tuple of points $\mathbf{X}_n = \{x_j\}_{j=0}^{n-1}$, consider the Newton family of polynomial perturbations

$$\tilde{f}_{u,\mathbf{X}_n}(x) = \tilde{f}(x) + \sum_{m=0}^{2n-1} u_m \prod_{j=0}^{m-1} (x - x_{j \pmod n}). \tag{3.26}$$

Notice that in (2.17) and Figure 2.2, the image $\tilde{f}_{u,\mathbf{X}_n}(x_0)$ of x_0 is independent of u_k for all $k > 0$. Therefore, the position of $\tilde{f}_{u,\mathbf{X}_n}(x_0)$ *depends only on* u_0. Recall that $\mu^{\mathrm{st}}_{m,\tau}$ is 1-dimensional Lebesgue measure scaled by $m!/(2\tau)$. This gives

(3.27)
$$\mu^{\mathrm{st}}_{0,\tau}\left\{u_0 : \left|\tilde{f}_{u,\mathbf{X}_n}(x_0) - x_1\right| = \left|\tilde{f}(x_0) + u_0 - x_1\right| \le \tilde{\gamma}_n\right\} \le \frac{0!}{2\tau} 2\tilde{\gamma}_n = \frac{0!\ \tilde{\gamma}_n}{\tau}.$$

Fix u_0. Similarly, the position of $\tilde{f}_u(x_1)$ *depends only on* u_1 (see (2.17) and Figure 2.2). Thus, we have

$$\begin{aligned} \mu^{\mathrm{st}}_{1,\tau}\{u_1 : \left|\tilde{f}_{u,\mathbf{X}_n}(x_1) - x_2\right| = \left|\tilde{f}(x_1) + u_0 + u_1(x_1 - x_0) - x_2\right| \le \tilde{\gamma}_n\} \\ \le \frac{1!\ \tilde{\gamma}_n}{\tau\ |x_1 - x_0|}. \end{aligned} \tag{3.28}$$

Inductively for $k = 2, \ldots, n-1$, fix $u_0, \ldots, u_{k-1}$. Then the position of $\tilde{f}_{u,\mathbf{X}_n}(x_k)$ *depends only on* u_k. Moreover, for $k = 2, \ldots, n-2$,

(3.29)
$$\begin{aligned} \mu^{\mathrm{st}}_{k,\tau}\Big\{u_k : \left|\tilde{f}_{u,\mathbf{X}_n}(x_k) - x_{k+1}\right| = \left|\tilde{f}(x_k) + \sum_{m=0}^{k} u_m \prod_{j=0}^{m-1}(x_k - x_j) - x_{k+1}\right| \le \tilde{\gamma}_n\Big\} \\ \le \frac{k!}{\tau}\ \frac{\tilde{\gamma}_n}{\prod_{j=0}^{k-1}|x_k - x_j|}, \end{aligned}$$

and for $k = n-1$,

(3.30)
$$\mu^{\mathrm{st}}_{n-1,\tau}\{u_{n-1} : \left|\tilde{f}_{u,\mathbf{X}_n}(x_{n-1}) - x_0\right| \le \gamma_n\} \le \frac{(n-1)!}{\tau} \frac{\gamma_n}{\prod_{j=0}^{n-2}|x_{n-1} - x_j|}.$$

In particular, the parameter u_{n-1} *is responsible* for (n, γ_n)-periodicity of the n-tuple $\mathbf{X}_n$. Formula (3.30) estimates the "measure of periodicity" discussed in the previous section.

Choose $u_0, \dots, u_{n-1}$ so that the n-tuple $\mathbf{X}_n$ is a $\tilde{\gamma}_n$-pseudotrajectory and is (n, γ_n)-periodic. Notice that parameters $u_n, u_{n+1}, \dots, u_{2n-1}$ do not change the $\tilde{\gamma}_n$-pseudotrajectory $\{x_k\}_{k=0}^{n-1}$. Fix now parameters $u_0, \dots, u_{2n-2}$ and vary only u_{2n-1}. Then for any C^1-smooth map $g : I \to I$, consider the 1-parameter family

$$g_{u_{2n-1}}(x) = g(x) + (x - x_{n-1}) \prod_{j=0}^{n-2} (x - x_j)^2. \tag{3.31}$$

Since the corresponding monomial $(x - x_{n-1}) \prod_{j=0}^{n-2} (x - x_j)^2$ has zeroes of the second order at all points x_k, except the last one x_{n-1},

$$\prod_{j=0}^{n-1} (g_{u_{2n-1}})'(x_j) = \left(g'(x_{n-1}) + u_{2n-1} \prod_{j=0}^{n-2} |x_{n-1} - x_j|^2 \right) \prod_{j=0}^{n-2} g'(x_j). \tag{3.32}$$

To get the final estimate, we use the fact that we are interested only in maps from the family $\{\tilde{f}_\varepsilon\}_{\varepsilon \in \mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)}$. Therefore, $|\tilde{f}'_{u,\mathbf{X}_n}(x_{n-1})| \le M_1 \le M_2$. For condition (C) of (3.25) to hold, $\left| \prod_{j=0}^{n-1} \tilde{f}'_{u,\mathbf{X}_n}(x_j) \right|$ must lie in $[1-\gamma_n, 1+\gamma_n]$. If this occurs for any $u_{2n-1} = \varepsilon_{2n-1} \in \mathrm{HB}^{\mathrm{st}}_{2n}(\tau)$, then $\left| \prod_{j=0}^{n-2} \tilde{f}'_{u,\mathbf{X}_n}(x_j) \right| \ge (1 - \gamma_n)/M_2$ for all u_{2n-1}, because this product does not depend on u_{2n-1}. Using (3.32) and the fact that $1 - \gamma_n \ge 1/2$, we get

$$\begin{aligned} &\mu^{\mathrm{st}}_{2n-1,\tau} \left\{ u_{2n-1} : \left| \prod_{j=0}^{n-1} |(\tilde{f}_{u,\mathbf{X}_n})'(x_j)| - 1 \right| \le \gamma_n \right\} \\ &\quad \le 2M_2 \frac{(2n-1)!}{2\tau} \, \frac{4\gamma_n}{\prod_{j=0}^{n-2} |x_{n-1} - x_j|^2} = 4M_2 \frac{(2n-1)!}{\tau} \, \frac{\gamma_n}{\prod_{j=0}^{n-2} |x_{n-1} - x_j|^2}. \end{aligned} \tag{3.33}$$

This formula estimates the "measure of hyperbolicity" discussed in the previous section.

By Lemma 2.2.2, we can combine all these estimates and get

$$\begin{aligned} &\mu^{\mathrm{st}}_{<n,\tau} \times \mu^{\mathrm{st}}_{2n-1,\tau} \{ (u_0, \dots, u_{n-1}, u_{2n-1}) \in W^{u,\mathbf{X}_n}_{<n,1} \times W^{u,\mathbf{X}_n}_{2n-1,1} : \\ &\quad \tilde{f}_{u,\mathbf{X}_n} \text{ satisfies conditions (3.25) and } \|\tilde{f}_{u,\mathbf{X}_n}\|_{C^2} \le M_2 \} \\ &\quad \le 4M_2 \, \frac{(n-1)! \; \gamma_n}{\tau \; \prod_{j=0}^{n-2} |x_{n-1} - x_j|} \, \frac{(2n-1)! \; \gamma_n}{\tau \; \prod_{j=0}^{n-2} |x_{n-1} - x_j|^2} \prod_{m=0}^{n-2} \frac{m! \; \tilde{\gamma}_n}{\tau \; \prod_{j=0}^{m-1} |x_m - x_j|}, \end{aligned} \tag{3.34}$$

where the spaces $W^{u,\mathbf{X}_n}_{<n,1}$ and $W^{u,\mathbf{X}_n}_{2n-1,1}$ are as discussed in the beginning of this section. This estimate corresponds loosely to (3.23) in the previous section. The final term is an upper bound on measure of parameters for which $\mathbf{X}_n$ is

a $\tilde{\gamma}_n$-pseudotrajectory for $f_{u,\mathbf{X}_n}$. Roughly speaking, since almost every initial point x_0 has exactly one $\tilde{\gamma}_n$-pseudotrajectory $\mathbf{X}_n \in I^n_{\tilde{\gamma}_n}$ for each set of parameters, and the total measure of parameters in $\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$ is 1, the sum over all $\tilde{\gamma}_n$-pseudotrajectories $\mathbf{X}_n$ associated to x_0 and $\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$ of the parameter measure associated with $\mathbf{X}_n$ should be 1. Thus the final term on the right-hand side of (3.34) also represents an upper bound on the inverse of the number of $\tilde{\gamma}_n$-pseudotrajectories per initial point, which appears in (3.23). However, we need the upper bound to be sharp in order to cancel this term with that in (3.22), and the heuristic explanation of this paragraph is complicated by the fact that the parametrization we are using *depends on the $\tilde{\gamma}_n$-pseudotrajectory* $\mathbf{X}_n$. These challenges will be resolved in the Collection Lemma of the next section.

3.4. *The Distortion and Collection Lemmas.* (See also Section 5 in [GHK].) In this section we formulate the Distortion Lemma for the Newton map $\mathcal{L}^1_{\mathbf{X}_n}$, and complete the estimate of the measure of all "bad" parameters with a simple, almost periodic, but not sufficiently hyperbolic trajectory (3.13), by collecting all possible "bad" pseudotrajectories (see the Collection Lemma below).

Consider an ordered n-tuple of points $\mathbf{X}_n = \{x_j\}_{j=0}^{n-1} \in I^n$ and the Newton map $\mathcal{L}^1_{\mathbf{X}_n} : W_{<2n,1} \to W^{u,\mathbf{X}_n}_{<2n,1}$, defined by (2.24). We now estimate the distortion of the Newton map $\mathcal{L}^1_{\mathbf{X}_n}$ as the map from the standard basis $\{\varepsilon_k\}_{k=0}^{2n-1}$ in $W_{<2n,1}$ to the Newton basis $\{u_k\}_{k=0}^{2n-1}$ in $W^{u,\mathbf{X}_n}_{<2n,1}$. It helps to have in mind the following picture characterizing the distortion of the Newton map.

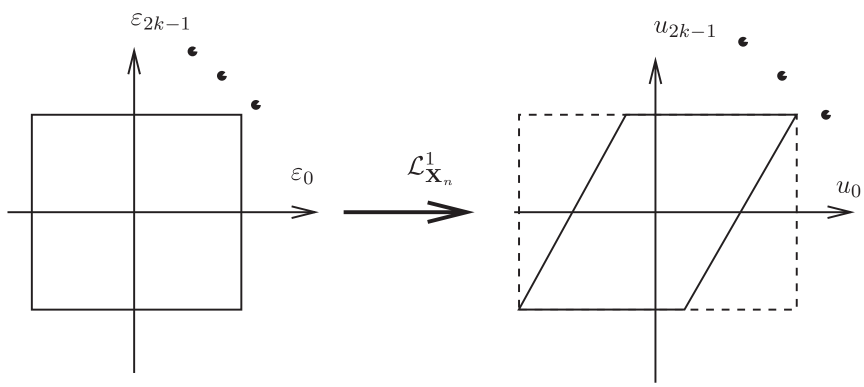

Figure 3.1: Distortion by the Newton map

THE DISTORTION LEMMA. *Let $\mathbf{X}_n = \{x_j\}_{j=0}^{n-1} \in I^n$ be an ordered n-tuple of points in the interval $I = [-1,1]$ and $\mathcal{L}^1_{\mathbf{X}_n} : W_{<2n,1} \to W^{u,\mathbf{X}_n}_{<2n,1}$ be the Newton map, defined by (2.24). Then the image of the Brick of standard thickness $\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$ with width $\tau > 0$ is contained in the Brick of standard thickness*

$\mathrm{HB}^{\mathrm{st}}_{<2n}(3\tau)$ *with width* 3τ:

$$\mathcal{L}^1_{\mathbf{X}_n}(\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)) \subset \mathrm{HB}^{\mathrm{st}}_{<2n}(3\tau) \subset W^{u,\mathbf{X}_n}_{<2n,1}. \tag{3.35}$$

In other words, independently of the choice of an n*-tuple* $\{x_j\}_{j=0}^{n-1} \in I^n$ *for any* $0 \le m < 2n$*, the coefficient* u_m *has at most the range of values* $|u_m| \le \frac{3\tau}{m!}$ *in the image* $\mathcal{L}^1_{\mathbf{X}_n}(\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau))$.

Remark 3.4.1. For this lemma, the sides of the Brick $\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$ have to decay at least as fast as a factorial in the order of the side, i.e., $r_n \le \frac{\tau}{n!}$ for some $\tau > 0$. If the sides of a Brick under investigation decay, say, as an exponential function, i.e., $r_n = \exp(-Kn)$ for some $K > 0$, then the Distortion Lemma fails and there is no uniform estimate on distortion. In terms of formula (2.24), if the range of values of ε_k does not decay fast enough with k, then u_m depends significantly on ε_k with k much larger than m.

Proof. Recall that for $\{\varepsilon_m\}_{m=0}^{2n-1} \in \mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$, for each m, that $|\varepsilon_m| \le \tau/m!$. By definition (2.24) of the Newton map $\mathcal{L}^1_{\mathbf{X}_n}$,

$$u_m = \varepsilon_m + \sum_{k=m+1}^{2n-1} \varepsilon_k p_{k,m}(x_0, \dots, x_{m \pmod n}), \tag{3.36}$$

where $p_{k,m}$ is the homogeneous polynomial of degree $k-m$ defined by (2.22). Notice that every monomial of $p_{k,m}$ is uniformly bounded by 1, provided all points $\{x_j\}_{j=0}^{n-1}$ are bounded in absolute value by 1. Therefore, $|p_{k,m}|$ is uniformly bounded by the number of its monomials $\binom{k}{m}$. This implies that

$$|u_m| \le |\varepsilon_m| + \sum_{k=m+1}^{2n-1} |\varepsilon_k| \binom{k}{m} \le \frac{\tau}{m!} \left(1 + \sum_{k=m+1}^{2n-1} \frac{1}{(k-m)!}\right) \le \frac{3\tau}{m!}. \tag{3.37}$$

This completes the proof of the lemma. Q.E.D.

For a given n-tuple $\mathbf{X}_n = \{x_j\}_{j=0}^{n-1} \in I^n$, the parallelepiped

$$\mathcal{P}^{\mathrm{st}}_{<2n,\mathbf{X}_n}(\tau) = \mathcal{L}^1_{\mathbf{X}_n}(\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)) \subset W^{u,\mathbf{X}_n}_{<2n,1} \tag{3.38}$$

is the set of parameters $(u_0, \dots, u_{2n-1})$ that correspond to parameters $(\varepsilon_0, \dots, \varepsilon_{2n-1}) \in \mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$. In other words, these are the Newton parameters *allowed by the family* (3.8) for the n-tuple $\mathbf{X}_n$. We already knew by Lemma 2.2.2 that $\mathcal{P}^{\mathrm{st}}_{<2n,\mathbf{X}_n}(\tau)$ has the same volume as $\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$, but the Distortion Lemma tells us in addition that the projection of $\mathcal{P}^{\mathrm{st}}_{<2n,\mathbf{X}_n}(\tau)$ onto any coordinate axis is at most a factor of 3 longer than the projection of $\mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)$.

Let $\mathbf{X}_m = \{x_j\}_{j=0}^{m-1}$ be the m-tuple of first m points of the n-tuple $\mathbf{X}_n$. We now consider which Newton parameters are allowed by the family (3.8) when $\mathbf{X}_m$ is fixed but $x_m, \dots, x_{n-1}$ are arbitrary. Since we will only be using the definitions below for discretized n-tuples $\mathbf{X}_n \in I^n_{\tilde{\gamma}_n}$, we consider only the (finite number of) possibilities $x_m, \ldots, x_{n-1} \in I_{\tilde{\gamma}_n}$. Let

$$\pi^{u,\mathbf{X}_n}_{<2n,\leq m} : W^{u,\mathbf{X}_n}_{<2n,1} \to W^{u,\mathbf{X}_m}_{\leq m,1} \text{ and } \pi^{u,\mathbf{X}_n}_{<2n,m} : W^{u,\mathbf{X}_n}_{<2n,1} \to W^{u,\mathbf{X}_m}_{m,1}$$

be the natural projections onto the space $W^{u,\mathbf{X}_m}_{\leq m,1} \simeq \mathbb{R}^m$ of polynomials of degree m and the space $W^{u,\mathbf{X}_m}_{m,1} \simeq \mathbb{R}$ of homogeneous polynomials of degree m respectively. Denote the unions over all $x_m, \ldots, x_{n-1} \in I_{\tilde{\gamma}_n}$ of the images of $\mathcal{P}^{\mathrm{st}}_{<2n,\mathbf{X}_n}(\tau)$ under the respective projections $\pi^{u,\mathbf{X}_m}_{<2n,\leq m}$ and $\pi^{u,\mathbf{X}_n}_{<2n,m}$ by

$$\begin{aligned} \mathcal{P}^{\mathrm{st}}_{<2n,\leq m,\mathbf{X}_m}(\tau) &= \cup_{x_m,\ldots,x_{n-1}\in I_{\tilde{\gamma}_n}} \ \pi^{u,\mathbf{X}_n}_{<2n,\leq m}(\mathcal{P}^{\mathrm{st}}_{<2n,\mathbf{X}_n}(\tau)) \subset W^{u,\mathbf{X}_m}_{\leq m,1}, \\ \mathcal{P}^{\mathrm{st}}_{<2n,m,\mathbf{X}_m}(\tau) &= \cup_{x_m,\ldots,x_{n-1}\in I_{\tilde{\gamma}_n}} \ \pi^{u,\mathbf{X}_n}_{<2n,m}(\mathcal{P}^{\mathrm{st}}_{<2n,\mathbf{X}_n}(\tau)) \subset W^{u,\mathbf{X}_m}_{m,1}. \end{aligned} \tag{3.39}$$

For each $m < n$, the set $\mathcal{P}^{\mathrm{st}}_{<2n,\leq m,\mathbf{X}_m}(\tau)$ is a polyhedron and $\mathcal{P}^{\mathrm{st}}_{<2n,m,\mathbf{X}_m}(\tau)$ is a segment of length at most $6\tau/m!$ by the Distortion Lemma. Both depend only on the m-tuple $\mathbf{X}_m$ and width τ. The set $\mathcal{P}^{\mathrm{st}}_{<2n,\leq m,\mathbf{X}_m}(\tau)$ consists of all Newton parameters $\{u_j\}_{j=0}^m \in W^{u,\mathbf{X}_m}_{m,1}$ that are allowed by the family (3.8) for the m-tuple $\mathbf{X}_m$.

For each $m < n$, we introduce the family of diffeomorphisms

$$\tilde{f}_{u(m),\mathbf{X}_m}(x) = \tilde{f}(x) + \sum_{s=0}^{m} u_s \prod_{j=0}^{s-1}(x - x_j), \tag{3.40}$$

where $u(m) = (u_0, \ldots, u_m) \in \mathcal{P}^{\mathrm{st}}_{<2n,\leq m,\mathbf{X}_m}(\tau)$. For each possible continuation $\mathbf{X}_n$ of $\mathbf{X}_m$, the family $\tilde{f}_{u(m),\mathbf{X}_m}$ includes the subfamily of $\tilde{f}_{u,\mathbf{X}_n}$ (with $u \in \mathcal{P}^{\mathrm{st}}_{<2n,\mathbf{X}_n}(\tau)$) corresponding to $u_{m+1} = u_{m+2} = \cdots = u_{2n-1} = 0$. However, the action of $\tilde{f}_{u,\mathbf{X}_n}$ on $x_0, \ldots, x_m$ does not depend on $u_{m+1}, \ldots, u_{2n-1}$, and so for these points the family $\tilde{f}_{u(m),\mathbf{X}_m}$ is representative of the entire family $\tilde{f}_{u,\mathbf{X}_n}$. This motivates the definition

$$\begin{aligned} &T^{1,\tilde{\gamma}_n}_{<2n,\leq m,\tau}(\tilde{f}; x_0, \ldots, x_{m-1}, x_m, x_{m+1}) \\ &\quad = \Big\{ u(m) \in \mathcal{P}^{\mathrm{st}}_{<2n,\leq m,\mathbf{X}_m}(\tau) \subset W^{u,\mathbf{X}_m}_{\leq m,1} : \\ &\qquad\qquad |\tilde{f}_{u(m),\mathbf{X}_m}(x_{j-1}) - x_j| \leq \tilde{\gamma}_n \text{ for } j = 1, \ldots, m+1 \Big\}. \end{aligned} \tag{3.41}$$

The set $T^{1,\tilde{\gamma}_n}_{<2n,\leq m,\tau}(\tilde{f}; x_0, \ldots, x_{m-1}, x_m, x_{m+1})$ represents the set of Newton parameters $u(m) = (u_0, \ldots, u_m)$ allowed by the family (3.8) for which $x_0, \ldots, x_{m+1}$ is a $\tilde{\gamma}_n$-pseudotrajectory of $\tilde{f}_{u(m),\mathbf{X}_m}$ (and hence of $\tilde{f}_{u,\mathbf{X}_n}$ for all valid extensions u and $\mathbf{X}_n$ of $u(m)$ and $\mathbf{X}_m$).

In the following lemma, we collect all possible $\tilde{\gamma}_n$-pseudotrajectories and estimates of "bad" measure corresponding to those $\tilde{\gamma}_n$-pseudotrajectories.

The Collection Lemma. *With notation above, for all $x_0 \in I_{\tilde\gamma_n}$, the measure of the "bad" parameters satisfies*

$$\mu^{\mathrm{st}}_{<2n,\tau}\big\{B^{\mathrm{sim},\tilde\gamma_n}_{n,\tau}(\tilde f,\gamma_n,M_2;x_0)\big\} \le 6^{2n} M_2^{4n+1}\ \frac{(n-1)!}{\tau}\ \frac{(2n-1)!}{\tau}\ \gamma_n^{5/4}. \tag{3.42}$$

Remark 3.4.2. Figure 3 in [GHK] is a good illustration of the proof below.

Proof of the Collection Lemma. We prove by backward induction on m that for $x_0,\dots,x_m \in I_{\tilde\gamma_n}$,

(3.43)
$$\mu^{\mathrm{st}}_{<2n,\tau}\big\{B^{\mathrm{sim},\tilde\gamma_n}_{n,\tau}(\tilde f,\gamma_n,M_2;x_0,\dots,x_m)\big\}$$
$$\le 6^{2n-m} M_2^{4n+1}\ \frac{(n-1)!}{\tau}\ \frac{(2n-1)!}{\tau}\mu^{\mathrm{st}}_{<m,\tau}\big\{T^{1,\tilde\gamma_n}_{<2n,\le m-1,\tau}(\tilde f;x_0,\dots,x_m)\big\}\ \gamma_n^{5/4},$$

resulting when $m=0$ in (3.42).

Consider the case $m = n-1$. Fix an $(n,\gamma_n/2)$-simple n-tuple $\mathbf{X}_n = \{x_j\}_{j=0}^{n-1} \in I^n_{\tilde\gamma_n}$. Using formulas (3.30) and (3.33), we get

$$\mu^{\mathrm{st}}_{n-1,\tau}\{u_{n-1} : \left|\tilde f_{u,\mathbf{X}_n}(x_{n-1}) - x_0\right| \le M_2^n\gamma_n\} \tag{3.44}$$
$$\le \frac{(n-1)!}{\tau}\frac{M_2^n\gamma_n}{\prod_{m=0}^{n-2}|x_{n-1}-x_m|} \le \frac{2^{1/4}M_2^n(n-1)!}{\tau}\ \gamma_n^{3/4}$$

and

$$\mu^{\mathrm{st}}_{2n-1,\tau}\{u_{2n-1} : \left|\prod_{j=0}^{n-1}|(\tilde f_{u,\mathbf{X}_n})'(x_j)| - 1\right| \le M_2^{3n}\gamma_n\} \tag{3.45}$$
$$\le 4M_2\frac{(2n-1)!}{\tau}\ \frac{M_2^{3n}\gamma_n}{\prod_{m=0}^{n-2}|x_{n-1}-x_m|^2} \le \frac{2^{5/2}M_2^{3n+1}(2n-1)!}{\tau}\gamma_n^{1/2}.$$

The Fubini theorem, Lemma 2.2.2, and the definition of the product measure $\mu^{\mathrm{st}}_{<2n,\tau}$ imply that

(3.46)
$$\mu^{\mathrm{st}}_{<2n,\tau}\left\{B^{\mathrm{sim},\tilde\gamma_n}_{n,\tau}(\tilde f,\gamma_n,M_2;x_0,\dots,x_{n-1})\right\}$$
$$\le \mu^{\mathrm{st}}_{<n-1,\tau}\left\{T^{1,\tilde\gamma_n}_{<2n,\le n-2,\tau}(\tilde f;x_0,\dots,x_{n-1})\right\}$$
$$\times\mu^{\mathrm{st}}_{n-1,\tau}\left\{u_{n-1} : \left|\tilde f_{u,\mathbf{X}_n}(x_{n-1}) - x_0\right| \le M_2^n\gamma_n\right\} \times \prod_{s=n}^{2n-2}\mu^{\mathrm{st}}_{s,\tau}\{\mathcal{P}^{\mathrm{st}}_{<2n,s,\mathbf{X}_n}(\tau)\}$$
$$\times\mu^{\mathrm{st}}_{2n-1,\tau}\left\{u_{2n-1} : \left|\prod_{j=0}^{n-1}|(f_{u,\mathbf{X}_n})'(x_j)| - 1\right| \le M_2^{3n}\gamma_n\right\}$$

$$\leq 2^{11/4} 3^{n-1} M_2^{4n+1} \ \frac{(n-1)!}{\tau} \frac{(2n-1)!}{\tau}$$
$$\times \mu^{\mathrm{st}}_{<n-1,\tau} \big\{ T^{1,\tilde{\gamma}_n}_{<2n,\leq n-2,\tau}(\tilde{f}; x_0, \dots, x_{n-1}) \big\} \ \gamma_n^{5/4}.$$

The last inequality follows from the Distortion Lemma, which says that for each $s = n, n+1, \dots, 2n-2$

$$\mu^{\mathrm{st}}_{s,\tau} \{ \mathcal{P}^{\mathrm{st}}_{<2n,s,\mathbf{X}_n}(\tau) \} \leq 3. \tag{3.47}$$

Since $2^{11/4} 3^{n-1} < 6^{n+1}$, this yields the required estimate (3.43) for $m = n-1$.

Suppose now that for $m+1$, (3.43) is true and we would like to prove it for m. Denote by $G^{1,\tilde{\gamma}_n}_{<2n,m,\tau}(\tilde{f}, u(m-1); x_0, \dots, x_m) \subset I_{\tilde{\gamma}_n}$ the set of points x_{m+1} of the $2\tilde{\gamma}_n$-grid $I_{\tilde{\gamma}_n}$ such that the $(m+2)$-tuple $x_0, \dots, x_{m+1}$ is a $\tilde{\gamma}_n$-pseudotrajectory associated to some extension $u(m) \in \mathcal{P}^{\mathrm{st}}_{<2n,\leq m,\mathbf{X}_m}(\tau)$ of $u(m-1)$. In other words, $G^{1,\tilde{\gamma}_n}_{<2n,m,\tau}(\tilde{f}, u(m-1); x_0, \dots, x_m)$ is the set of all possible continuations of the $\tilde{\gamma}_n$-pseudotrajectory $x_0, \dots, x_m$ using all possible Newton parameters u_m allowed by the family (3.8).

Now if $x_0, \dots, x_m$ is a $\tilde{\gamma}_n$-pseudotrajectory associated to $u(m) = (u_0, \dots, u_m)$, then at most two values of $x_{m+1} \in I_{\tilde{\gamma}_n}$ are within $\tilde{\gamma}_n$ of $\tilde{f}_{u(m),\mathbf{X}_m}(x_m)$. Thus for fixed $u(m-1) = (u_0, \dots, u_{m-1}) \in \mathcal{P}^{\mathrm{st}}_{<2n,\leq m-1,\mathbf{X}_n}(\tau)$, each value of $u_m \in \mathcal{P}^{\mathrm{st}}_{<2n,m,\mathbf{X}_n}(\tau)$ corresponds to at most two points in

$$G^{1,\tilde{\gamma}_n}_{<2n,m,\tau}(\tilde{f}, u(m-1); x_0, \dots, x_m).$$

It follows that

$$\sum_{x_{m+1} \in G^{1,\tilde{\gamma}_n}_{<2n,m,\tau}(\tilde{f}, u(m-1); x_0, \dots, x_m)} \mu^{\mathrm{st}}_{\leq m,\tau} \big\{ T^{1,\tilde{\gamma}_n}_{<2n,\leq m,\tau}(\tilde{f}; x_0, \dots, x_{m+1}) \big\} \tag{3.48}$$
$$\leq 2 \ \mu^{\mathrm{st}}_{m,\tau} \{ \mathcal{P}^{\mathrm{st}}_{<2n,m,\mathbf{X}_n}(\tau) \} \ \mu^{\mathrm{st}}_{\leq m-1,\tau} \big\{ T^{1,\tilde{\gamma}_n}_{<2n,\leq m-1,\tau}(\tilde{f}; x_0, \dots, x_m) \big\}.$$

The Distortion Lemma then implies that

$$\sum_{x_{m+1} \in G^{1,\tilde{\gamma}_n}_{<2n,m,\tau}(\tilde{f}, u(m-1); x_0, \dots, x_m)} \mu^{\mathrm{st}}_{\leq m,\tau} \big\{ T^{1,\tilde{\gamma}_n}_{<2n,\leq m,\tau}(\tilde{f}; x_0, \dots, x_{m+1}) \big\} \tag{3.49}$$
$$\leq 6 \ \mu^{\mathrm{st}}_{\leq m-1,\tau} \big\{ T^{1,\tilde{\gamma}_n}_{<2n,\leq m-1,\tau}(\tilde{f}; x_0, \dots, x_m) \big\}.$$

Inductive application of this formula completes the proof of the Collection Lemma. Q.E.D.

Proof of Proposition 3.1.2. The number of starting points $\tilde{x}_0 \subset I_{\tilde{\gamma}_n}$ for a $\tilde{\gamma}_n$-pseudotrajectory equals $1/\tilde{\gamma}_n$. Therefore, multiplying the estimate (3.42) by $1/\tilde{\gamma}_n = M_2^{2n}/\gamma_n$ and using (3.21) we get

$$\mu^{\mathrm{st}}_{<2n,\tau} \big\{ B^{\mathrm{sim},\tilde{\gamma}_n}_{n,\tau}(\tilde{f}, \gamma_n, M_2) \big\} \leq 6^{2n} M_2^{6n+1} \ \frac{(n-1)!}{\tau} \ \frac{(2n-1)!}{\tau} \ \gamma_n^{1/4}. \tag{3.50}$$

To prove the required inequality (3.13), it remains only to prove (3.17).

If a parameter ε belongs to $B^{\mathrm{sim}}_{n,\tau}(C,\delta,\tilde{f},\gamma_n)$, then ε is associated with an essentially (n,γ_n)-simple, (n,γ_n)-periodic, and not (n,γ_n)-hyperbolic trajectory $\{x_k = f^k_\varepsilon(x_0)\}_{k=0}^{n-1}$. Our goal is to show that $\varepsilon \in B^{\mathrm{sim}}_{n,\tau}(\tilde{f},\gamma_n,M_2)$; recall the definitions (3.15) and (3.16).

The bound M_2 on the C^2-norm of $\tilde{f}_\varepsilon$ implies that for all $x, y \in I$ we have $|\tilde{f}_\varepsilon(x) - \tilde{f}_\varepsilon(y)| < M_2|x-y|$. Essential simplicity of the trajectory $\{x_m\}_{m=0}^{n-1}$ implies that for some $j < n$, the shifted trajectory $\{x_{j+m} = \tilde{f}^{j+m}_\varepsilon(x_0)\}_{m=0}^{n-1}$ is (n,γ_n)-simple and $(n, M_2^j\gamma_n)$-periodic. We approximate the shifted trajectory $\{x_{j+m}\}_{m=0}^{n-1}$ by a $\tilde{\gamma}_n$-pseudotrajectory $\{\tilde{x}_{j+m}\}_{m=0}^{n-1} \in I^n_{\tilde{\gamma}_n}$ associated to the (fixed above) parameter ε. Consider the $\tilde{\gamma}_n$-pseudotrajectory $\{\tilde{x}_{j+m}\}_{m=0}^{n-1} \in I^n_{\tilde{\gamma}_n}$ starting at $\tilde{x}_j$. Choose $\tilde{x}_j$ such that $|x_j - \tilde{x}_j| \le \tilde{\gamma}_n$ and choose $\tilde{x}_{j+1}$ such that $|\tilde{f}_\varepsilon(\tilde{x}_j) - \tilde{x}_{j+1}| \le \tilde{\gamma}_n$. Then

$$
\begin{aligned}
(3.51)\qquad |x_{j+1} - \tilde{x}_{j+1}| &\le |\tilde{f}_\varepsilon(x_j) - \tilde{f}_\varepsilon(\tilde{x}_j)| + |\tilde{f}_\varepsilon(\tilde{x}_j) - \tilde{x}_{j+1}| \\
&\le (M_2+1)\tilde{\gamma}_n = \frac{M_2^2 - 1}{M_2 - 1}\tilde{\gamma}_n.
\end{aligned}
$$

By induction on m, choosing $\tilde{x}_{j+m}$ such that $|\tilde{f}_\varepsilon(\tilde{x}_{j+m-1}) - \tilde{x}_{j+m}| \le \tilde{\gamma}_n$, we have

$$
\begin{aligned}
(3.52)\quad |x_{j+m} - \tilde{x}_{j+m}| &\le |\tilde{f}_\varepsilon(x_{j+m-1}) - \tilde{f}_\varepsilon(\tilde{x}_{j+m-1})| + |\tilde{f}_\varepsilon(\tilde{x}_{j+m-1}) - \tilde{x}_{j+m}| \\
&\le \frac{M_2^{m+1} - 1}{M_2 - 1}\tilde{\gamma}_n.
\end{aligned}
$$

Using this estimate with $m = n-1$, we have

$$
\begin{aligned}
(3.53)\qquad |\tilde{f}_\varepsilon(\tilde{x}_{j+n-1}) - \tilde{x}_j| &\le |\tilde{f}_\varepsilon(\tilde{x}_{j+n-1}) - \tilde{f}_\varepsilon(x_{j+n-1})| \\
+ |\tilde{f}_\varepsilon(x_{j+n-1}) - x_j| + |x_j - \tilde{x}_j| &\le M_2^{n+1}\tilde{\gamma}_n + M_2^{n-1}\gamma_n \le M_2^n\gamma_n.
\end{aligned}
$$

So, the $\tilde{\gamma}_n$-pseudotrajectory $\{\tilde{x}_{j+m}\}_{m=0}^{n-1}$ is $(n, M_2^n\gamma_n)$-periodic.

Next, since $\{x_{j+m}\}_{m=0}^{n-1}$ is (n,γ_n)-simple, this means

$$
\prod_{m=0}^{n-2} |x_{j+n-1} - x_{j+m}| \ge \gamma_n^{1/4}.
$$

Each term in the product must then be at least $2^{-(n-2)}\gamma_n^{1/4}$. For $m \le n-1$, we have already shown that $|x_{j+m} - \tilde{x}_{j+m}| \le M_2^n\tilde{\gamma}_n$. Then since $\tilde{\gamma}_n = M_2^{-2n}\gamma_n$,

$$
\begin{aligned}
(3.54)\qquad |\tilde{x}_{j+n-1} - \tilde{x}_{j+m}| &\ge |x_{j+n-1} - x_{j+m}| - 2M_2^n\tilde{\gamma}_n \\
&\ge \frac{2^{-(n-2)}\gamma_n^{1/4} - 2M_2^n\tilde{\gamma}_n}{2^{-(n-2)}\gamma_n^{1/4}}|x_{j+n-1} - x_{j+m}| \\
&= (1 - 2^{n-1}M_2^{-n}\gamma_n^{3/4})|x_{j+n-1} - x_{j+m}| \\
&\ge (1 - \gamma_n^{3/4})|x_{j+n-1} - x_{j+m}|.
\end{aligned}
$$

Because $\gamma_n \le \gamma_n(C,\delta) = \exp(-Cn^{1+\delta})$ with $C > 2$, a simple calculation shows that $(1-\gamma_n^{3/4})^{n-1} \ge 2^{-1/4}$. Then taking the product of (3.54) over

$m = 0, 1, \dots, n-2$ proves that the approximating pseudotrajectory $\{\tilde{x}_{j+m}\}_{m=0}^{n-1}$ is $(n, \gamma_n/2)$-simple.

Now consider the difference of derivatives

$$\left| \prod_{m=0}^{n-j-1} \tilde{f}'_\varepsilon(\tilde{x}_{j+m}) \prod_{m=0}^{j-1} \tilde{f}'_\varepsilon(\tilde{x}_{n+m}) - \prod_{m=0}^{n-1} \tilde{f}'_\varepsilon(x_m) \right|. \tag{3.55}$$

Since $\|\tilde{f}\|_{C^2} \le M_2$, for $0 \le m \le n-j-1$,

$$|\tilde{f}'_\varepsilon(\tilde{x}_{j+m}) - \tilde{f}'_\varepsilon(x_{j+m})| \le M_2|\tilde{x}_{j+m} - x_{j+m}| \le M_2^{n+1}\tilde{\gamma}_n \le \gamma_n \tag{3.56}$$

and for $0 \le m \le j-1$,

(3.57)
$$\begin{aligned} |\tilde{f}'_\varepsilon(\tilde{x}_{n+m}) - \tilde{f}'_\varepsilon(x_m)| &\le M_2|\tilde{x}_{n+m} - x_m| \le M_2(|\tilde{x}_{n+m} - x_{n+m}| + |x_{n+m} - x_m|) \\ &\le M_2^{2n}\tilde{\gamma}_n + M_2^n\gamma_n \le 2M_2^n\gamma_n. \end{aligned}$$

Then when $|\tilde{f}'_\varepsilon(x)| \le M_2$ for all $x \in I$, we get that the difference of derivatives (3.55) is bounded by $2nM_2^{2n-1}\gamma_n \le M_2^{3n-1}\gamma_n$. Therefore, if the initial exact trajectory $\{x_k\}_{j=0}^{n-1}$ is not (n, γ_n)-hyperbolic, then the pseudotrajectory is not $(n, M_2^{3n}\gamma_n)$-hyperbolic, and the parameter ε belongs to the set $B_{n,\tau}^{\mathrm{sim},\tilde{\gamma}_n}(\tilde{f}, \gamma_n, M_2)$. Q.E.D.

3.5. *Discretization method for trajectories with a gap.* Recall that we consider the C^2-smooth 1-dimensional model, defined in Section 3.1. Our goal is to estimate the measure of the set (3.9) of all "bad" parameter values $B_{n,\tau}(C, \delta, \tilde{f}, \gamma_n)$. This set belongs to the union (3.12) of "bad" parameter values $B_{n,\tau}^{\mathrm{sim}}(C, \delta, \tilde{f}, \gamma_n)$ and $B_{n,\tau}^{\mathrm{non}}(C, \delta, \tilde{f}, \gamma_n)$ associated with essentially simple and essentially nonsimple almost periodic points. In the last three sections, we developed the Discretization Method and estimated the measure of the set $B_{n,\tau}^{\mathrm{sim}}(C, \delta, \tilde{f}, \gamma_n)$ associated with essentially simple, almost periodic trajectories (3.13). In this and the next section, we consider "bad" parameters $B_{n,\tau}^{\mathrm{non}}(C, \delta, \tilde{f}, \gamma_n)$ associated with essentially nonsimple, almost periodic trajectories (3.11) and get an estimate on the measure of this set.

It is helpful to read Section 6 in [GHK] and have in mind Figure 4 there while reading this section.

PROPOSITION 3.5.1. *Let $\{\tilde{f}_\varepsilon\}_{\varepsilon \in \mathrm{HB}^{\mathrm{st}}_{<2n}(\tau)}$ be the family of polynomial perturbations* (3.8) *and M_2 be an upper bound on the C^2-norm of the family. Then with the notation above, for any $\delta > 0$, $C > 100\delta^{-1}\log M_2$, $\tau > 0$ and a sufficiently small positive γ_n, e.g., $\gamma_n \le \gamma_n(C, \delta)$, there exists the following estimate on the measure of parameters associated with maps $\tilde{f}_\varepsilon$ with an (n, γ_n)-periodic, essentially non-(n, γ_n)-simple, but not (n, γ_n)-hyperbolic point*:

(3.58)
$$\mu^{\mathrm{st}}_{<2n,\tau}\{B_{n,\tau}^{\mathrm{non}}(C, \delta, \tilde{f}, \gamma_n)\} \le 128n^{1+\delta}\, 6^{2n}\, M_2^{10n+1}\, \exp(-Cn\log_2 n/200).$$

Remark 3.5.2. Though we have stated Proposition 3.5.1 only for the C^2 case in $\mathbb{R}$, we will make our upcoming definitions for the general $C^{1+\rho}$ case in $\mathbb{R}^N$ so that they may be used later. For the time being, the reader may keep in mind the case $\rho = N = 1$. The scheme of the proof of this proposition is in Section 3.5.2.

According to the decomposition (2.42), the set $B^{\mathrm{non}}_{n,\tau}(C, \delta, \tilde{f}, \gamma_n)$ of parameters associated with essentially nonsimple trajectories can be decomposed into a finite union of sets of parameters with a trajectory that has a weak gap. This decomposition follows from Theorem 2.4.6 which needs improvement. Let us sharpen Definition 2.4.3 of a weak gap for almost periodic trajectories.

Definition 3.5.3. Let $g \in \mathrm{Diff}^{1+\rho}(B^N)$ be a $C^{1+\rho}$-smooth diffeomorphism of the ball B^N (respectively $g \in C^{1+\rho}(I, I)$ be a $C^{1+\rho}$-smooth map of the interval I) and let $\vec{\mathbf{r}} = \{r_k\}_{k=0}^{\infty}$ be a nonincreasing sequence of sizes of the Brick $\mathrm{HB}^N(\vec{\mathbf{r}})$ that tend to zero, and $D > 2$ be some number. A point $x_0 \in B^N$ (respectively $x_0 \in I$) or a trajectory $x_0, \dots, x_{n-1} = g^{n-1}(x_0)$ in B^N (respectively I) of length n has a *(D, n, r_{2k})-gap* at a point $x_k = g^k(x_0)$ if

$$(3.59)\quad |x_k - x_0| \le \min\left\{ D^{-n\log_2 n} \min_{0<j\le k-1} |x_0 - x_j|,\ r_{2k}^{4(N+N^2)},\ \prod_{j=0}^{k-2} |x_{k-1} - x_j|^{4(N+2)} \right\}.$$

In the case of the model when $r_k = \frac{\tau}{k!}$, we denote a *(D, n, r_{2k})-gap* by a *(D, n, τ)-gap.*

This definition is designed to fit the induction over the period n outlined in Chapter 2. Recall that the first term in the minimum (3.59) corresponds to Definition 2.4.3 of a weak (D, n)-gap. For $D = M_{1+\rho}^{30/\rho}$, if a trajectory $\{x_j\}_{j=0}^{n-1}$ is $(n, |x_k - x_0|)$-periodic and has a weak (D, n)-gap at x_k, then by Lemma 2.4.5 the fraction $p = n/k$ is an integer, and one can split the trajectory of length n into p almost identical parts of length k each. Then by perturbing the linearization $dg_u^k(x_0)$ at x_{k-1} using the family $g_u(x) = g(x) + u(x - x_{k-1}) \prod_{j=0}^{k-2} (x - x_j)^2$, one can reach sufficient hyperbolicity for $dg_u^k(x_0)$. Moreover, the first and the third terms in the minimum (3.59) allow us to extract sufficient hyperbolicity of $dg_u^n(x_0)$ from sufficient hyperbolicity of $dg_u^k(x_0)$. Roughly, this is because we have $dg_u^n(x_0) \approx \big(dg_u^k(x_0)\big)^p$.

The second and the third terms guarantee that with respect to the normalized Lebesgue measure $\mu^N_{\vec{\mathbf{r}}}$, "bad" parameters occupy a small measure set; loosely speaking, they counteract terms like $(2n-1)!/\tau$ and $\prod_{j=0}^{n-2} |x_{n-1} - x_j|^{-1}$ in (3.30) and (3.33).

Recall that in our notation, $\gamma_n(C, \delta) = \exp(-Cn^{1+\delta})$ for each $n \in \mathbb{Z}_+$. Notice that we use both $\gamma_k(C, \delta)$ and $\gamma_n(C, \delta)$ in the definition below.

Definition 3.5.4. Let $g \in \mathrm{Diff}^{1+\rho}(B^N)$ be a $C^{1+\rho}$-smooth diffeomorphism of the ball B^N (respectively $g \in C^{1+\rho}(I, I)$ be a $C^{1+\rho}$-smooth map of the interval I) for some $\rho > 0$. Let also $C > 0$, $\delta > 0$ and $k < n$ be positive integers. We say that a point x_0 has a (k, n, C, δ, ρ)-*leading saddle* if $|x_0 - x_k| \le n^{-1/\rho}\gamma_k^{4N/\rho}(C, \delta)$. Also if x_0 is $(n, \gamma_n^{1/\rho}(C, \delta))$-periodic, we say that x_0 has *no* (n, C, δ, ρ)-*leading saddles* if for all $k < n$ we have that x_0 has *no* (k, n, C, δ, ρ)-*leading saddles*.

Lemma 3.5.5. *Let a $C^{1+\rho}$-smooth diffeomorphism* (*respectively $C^{1+\rho}$-smooth map*) *g satisfy the Inductive Hypothesis of some order $n-1$ with some constants $\rho > 0$, $C > 30\rho^{-1}\log M_{1+\rho}$, and $\delta > 0$, i.e., $g \in \mathrm{IH}(n-1, C, \delta, \rho)$. Assume that g has a point $x_0 \in B^N$* (*respectively $x \in I$*) *that has a (k, n, C, δ, ρ)-leading saddle. Then there is a periodic point $x^* = g^k(x^*)$ of period k such that $|x^* - x_0| \le 2n^{-1/\rho}\gamma_k^{3N/\rho}(C, \delta)$. Moreover, by the Inductive Hypothesis, x^* is $(k, \gamma_k(C, \delta))$-hyperbolic.*

This lemma follows from a lemma in Part II of this paper [K5] that is used to prove Theorem 2.4.6 and the Shift Theorem below. It turns out that Theorem 2.4.6 can be improved to

The Shift Theorem. *Let $g \in \mathrm{Diff}^{1+\rho}(B^N)$ be a $C^{1+\rho}$-smooth diffeomorphism* (*respectively $g \in C^{1+\rho}(I, I)$ be a $C^{1+\rho}$-smooth map of the interval I*) *with some $\rho > 0$, $M_{1+\rho} = \max\{\|g^{-1}\|_{C^1}, \|g\|_{C^{1+\rho}}, 2^{1/\rho}\}$, and let $\vec{\mathbf{r}} = \{r_k\}_{k=0}^{\infty}$ be an admissible sequence* (*see Definition* 1.3.1). *Assume that g satisfies the Inductive Hypothesis of some order $n-1$ with some constants $\delta > 0$, $\rho > 0$, and $C > 100\rho^{-1}\delta^{-1}\log M_{1+\rho}$, i.e., $g \in \mathrm{IH}(n-1, C, \delta, \rho)$. If a point $x_0 \in B^N$* (*respectively $x_0 \in I$*) *is $(n, \gamma_n^{1/\rho}(C, \delta))$-periodic, then either x_0 is $(n, \gamma_n(C, \delta))$-hyperbolic, or for some $j < n\log_2 n$ the point $x_j = g^j(x_0)$ has no (n, C, δ, ρ)-leading saddles.*

Moreover, if x_0 has no (n, C, δ, ρ)-leading saddles, then either it is $(n, \gamma_n(C, \delta))$-simple or it is $(k, \gamma_n(C, \delta))$-simple and has a (D, n, r_{2k})-gap at x_k for some k dividing n and $D = \max\{M_{1+\rho}^{30/\rho}, \exp(C/100)\}$.

Remark 3.5.6. The Shift Theorem is the key for splitting all almost periodic trajectories of period n into groups and is *the main reason why* the estimate on the number of periodic points is $\exp(Cn^{1+\delta})$, not, say, $\exp(Cn\log n)$. In view of the importance of the theorem, we present an outline of its proof.

Outline of the proof of the Shift Theorem. Let x_0 be $(n, \gamma_n^{1/\rho}(C, \delta))$-periodic.

- If x_0 has a (k, n, C, δ, ρ)-leading saddle, then by Lemma 3.5.5 there is a periodic point $x^* = g^k(x^*)$ such that $|x_0 - x^*| \le 2n^{-1/\rho}\gamma_k^{3N/\rho}(C, \delta)$, and by the Inductive Hypothesis x^* is $(k, \gamma_k(C, \delta))$-hyperbolic. If x_{pk} remains in the $2n^{-1/\rho}\gamma_k^{3N/\rho}(C,\delta)$-neighborhood of x^* for $1 \le p < n/k$, then the approximation $df^n(x_0) \approx df^n(x^*) = (df^k(x^*))^{n/k}$ is good enough to extract from $(k, \gamma_k(C,\delta))$-hyperbolicity of x^* sufficient hyperbolicity of $df^n(x_0)$. (If n/k is not an integer, we use a similar argument with k replaced by $\gcd(n,k)$; the Euclidean Algorithm in Part II of this paper [K5] shows that x_0 is almost $\gcd(n,k)$-periodic.)

• If x_0 has a (k,n,C,δ,ρ)-leading saddle, but x_{pk} leaves the $2n^{-1/\rho}\gamma_k^{3N/\rho}(C,\delta)$-neighborhood of x^* for some $p < n/k$, then we show (choosing the smallest k and then the smallest p with these properties) that x_{pk} has no (k',n,C,δ,ρ)-leading saddles for $k' < 2k$. If x_{pk} has a (k',n,C,δ,ρ)-leading saddle for some $k' \ge 2k$, then we proceed inductively. Either we can bound the hyperbolicity of $df^n(x_{pk})$ sufficiently well to conclude that x_0 is $(n,\gamma_n(C,\delta))$-hyperbolic (because x_{pk} is still quite close to x_0), or with a further shift we can eliminate all leading saddles of period less than $2k'$. Thus, either we prove the necessary hyperbolicity at some step of the induction, or after at most $\log_2 n$ shifts of at most $n > pk$ iterates per shift we eliminate all (n,C,δ,ρ)-leading saddles.

• If we cannot prove $(n,\gamma_n(C,\delta))$-hyperbolicity of x_0 using the arguments above, then for some $j < n\log_2 n$ we have that $x_j = \tilde{x}_0$ has no (n,C,δ,ρ)-leading saddles. Notice that $\tilde{x}_0$ is a $(n, M_{1+\rho}^{n\log_2 n}\gamma_n^{1/\rho}(C,\delta))$-periodic point. Put $D = \max\{M_{1+\rho}^{30/\rho}, \exp(C/100)\}$. It turns out that $\tilde{x}_0$ can have at most N weak (D,n)-gaps at $\tilde{x}_{k_1},\dots,\tilde{x}_{k_s}$ for $k_1 < \cdots < k_s < n$, $s \le N$. The reason is that each weak (D,n)-gap $\tilde{x}_{k_j}$ after the first one at k_1 implies that the linearization $df^{k_1}(\tilde{x}_0)$ has an almost eigenvalue that is a k_j/k_1-root of unity, and the same is true for $k_{s+1} = n$. (The Euclidean Algorithm from Part II of this paper [K5] implies that k_j is divisible by k_{j-1} for all j.) Heuristically, this follows from replacing f^{k_1} in a neighborhood of $\tilde{x}_0$ by its linearization, and observing that its (k_j/k_1)-th iteration nearly fixes $(\tilde{x}_{k_1} - \tilde{x}_0)$.

• One can prove that for all $k < n$, if there is a (D,n,r_{2k})-gap at $\tilde{x}_k$ but no $(D,n,r_{2k'})$-gap at $\tilde{x}_{k'}$ for $k' < k$, then $\tilde{x}_0$ is $(k,\gamma_n(C,\delta))$-simple; this completes the proof of the theorem in this case. Otherwise, $\tilde{x}_0$ has at most N weak (D,n)-gaps, none of which is a (D,n,r_{2k})-gap. In this case, we show that $\tilde{x}_0$ is $(n,\gamma_n(C,\delta))$-simple.

• First we show that $\tilde{x}_0$ is $(n,\gamma_n(C,\delta))$-simple assuming that $\tilde{x}_0$ has no weak (D,n)-gaps; later we will extend the argument to the general case. With no weak (D,n)-gaps, the Euclidean Algorithm gives that

$$\min_{0<j\le n-1} |\tilde{x}_0 - \tilde{x}_j| \ge D^{-n\log_2 n}. \tag{3.60}$$

Consider concentric balls B_k centered at $\tilde{x}_0$ of radii $R_k = D^{-k\log_2 k}$. A pigeonhole argument shows that the number of visits m_k of $\{\tilde{x}_j\}_{j=1}^{n-1}$ to B_k is not too large. To be precise, let

$$A_k = \{1 \le j < n : \ |\tilde{x}_0 - \tilde{x}_j| < R_k\}$$

be the collection of indices of points of the trajectory $\{\tilde{x}_j\}_{j=0}^n$ which visit the R_k-ball B_k around $\tilde{x}_0$. In part II of this paper [K5] we show that

$$m_k = \#\{A_k\} \le \frac{16NC}{\rho \log D} \quad \frac{n}{(k \log_2 k)^{\frac{1}{1+\delta}}}. \tag{3.61}$$

Otherwise, there are consecutive visits $\tilde{x}_j$ and $\tilde{x}_{j+\ell}$ to B_k with ℓ being small relative to n, more exactly, of order $(k \log_2 k)^{\frac{1}{1+\delta}}$. Thus, $\tilde{x}_j$ has an $(\ell, n, C, \delta, \rho)$-leading saddle and, therefore, a periodic point $\tilde{x}_j^* = g^\ell(\tilde{x}_j^*)$ close to $\tilde{x}_j$ and also to $\tilde{x}_0$. This implies that $\tilde{x}_0$ has an $(\ell, n, C, \delta, \rho)$-leading saddle too, which is a contradiction.

• Knowing that there are no visits to B_n, by (3.60), and given the bound (3.61) on the number of visits m_k to B_k for $1 \le k < n$, we can get a lower estimate on $\prod_{j=1}^{n-1} |\tilde{x}_0 - \tilde{x}_j|$ according the following scheme. By (3.60) we have $A_n = \emptyset$ and $m_n = 0$. Rewrite the product of distances as follows

$$\prod_{j=1}^{n-1} |\tilde{x}_j - \tilde{x}_0| = \prod_{j \notin A_1} |\tilde{x}_j - \tilde{x}_0| \prod_{j \in A_1 \setminus A_2} |\tilde{x}_j - \tilde{x}_0| \cdots \prod_{j \in A_{n-1}} |\tilde{x}_j - \tilde{x}_0|. \tag{3.62}$$

By definition of A_k, for each $j \in A_{k-1} \setminus A_k$ we have $|\tilde{x}_0 - \tilde{x}_j| \ge R_k$. Put $a_k = k \log_2 k$. Then the product (3.62) admits the following lower bound:

$$\begin{aligned} &R_1^{n-1-m_1} R_2^{m_1-m_2} \dots R_n^{m_{n-1}-m_n} \\ &\quad = \exp\Big(-\log D \,[a_1(n-1-m_1) + a_2(m_1 - m_2) + \dots + a_n(m_{n-1} - m_n)]\Big). \end{aligned} \tag{3.63}$$

Using Abel's resummation and $a_1 = m_n = 0$, we can rewrite the last expression in the form

$$\exp\left(-\log D \,[m_1(a_2 - a_1) + m_2(a_3 - a_2) + \dots + m_{n-1}(a_n - a_{n-1})]\right). \tag{3.64}$$

By definition of a_k, we have $a_{k+1} - a_k \le 3 \log_2 k$ for $k \ge 2$, while for $k = 1$ we simply estimate $m_1(a_2 - a_1) < 2n$ since $m_1 < n$. Using inequality (3.61) above, we get the following lower bound for the product (3.64)

$$\begin{aligned} \exp\left(-\frac{48NC}{\rho} n \log_2 n \sum_{k=1}^n k^{-\frac{1}{1+\delta}}\right) &\ge \exp\left(-\frac{48NC(1+\delta)}{\rho\delta} n^{1+\frac{\delta}{1+\delta}} \log_2 n\right) \\ &\ge \exp\left(-\frac{Cn^{1+\delta}}{4N}\right), \end{aligned} \tag{3.65}$$

where $192N^2(1+\delta)\; n^{-\frac{\delta^2}{1+\delta}} \log_2 n < \rho\delta$ for sufficiently large n. This shows that $\tilde{x}_0$ is $(n, \gamma_n(C, \delta))$-simple.

• We return now to the case where $\tilde{x}_0$ has at most N weak (D,n)-gaps. It turns out that if $k^{1+\delta} < \frac{\rho \log D}{2C} n$ and $\tilde{x}_0$ has a weak (D,n)-gap at $\tilde{x}_k$, then x_0 is $(n, \gamma_n(C,\delta))$-hyperbolic. Indeed, in this case $|\tilde{x}_0 - \tilde{x}_k| < D^{-n} < \gamma_k^{1/\rho}(C,dt)$ and, therefore, by inductive hypothesis $\tilde{x}_0$ is $(k, \gamma_k(C,\delta))$-hyperbolic. Points $\tilde{x}_k$ and $\tilde{x}_0$ are so close that n is divisible by k, so we can approximate $df^n(\tilde{x}_0)$ with $(df^k(\tilde{x}_0))^{n/k}$ as we do in the first item of the proof. Since $\tilde{x}_0$ is shifted from x_0 by at most $n \log_2 n$ iterates, we get sufficient hyperbolicity of $df^n(x_0)$ as well.

• Let $\tilde{x}_0$ have weak (D,n)-gaps at $\tilde{x}_{k_1},\ \tilde{x}_{k_2}, \dots, \tilde{x}_{k_s} = \tilde{x}_n$, none of which is a (D,n,r_{2k})-gap. Then $s < N$ and $k_1^{1+\delta} \geq \frac{\rho \log D}{2C} n$. An estimate similar to (3.61) leads to a lower bound of the form

$$\prod_{j=0}^{k_1-1} |\tilde{x}_0 - \tilde{x}_j| \geq \exp\left(-\frac{48NC(1+\delta)}{\rho\delta} k_1 n^{\frac{\delta}{1+\delta}} \log_2 n \right). \tag{3.66}$$

Notice that because of the lower bound on k_1, the exponent on the right-hand side of (3.66) is as least of order $n \log_2 n$; thus multiplication below by factors that are exponential in n or $n \log_2 n$ do not affect the general form of the estimate. Next, $\tilde{x}_{k_1}$ is so close to $\tilde{x}_0$ that the Euclidean Algorithm from Part II of this paper [K5] implies that k_2 is divisible by k_1; denote $p_1 = k_2/k_1$. Moreover, $\tilde{x}_{k_1}$ is so close to $\tilde{x}_0$ that the following approximation holds true:

$$\prod_{j=1}^{k_2-1} |\tilde{x}_0 - \tilde{x}_j| \geq \frac{1}{2} \prod_{\ell=1}^{p_1-1} |\tilde{x}_0 - \tilde{x}_{\ell k_1}| \left(\prod_{j=1}^{k_1} |\tilde{x}_0 - \tilde{x}_j| \right)^{p_1}. \tag{3.67}$$

Absence of a (D,n,r_{2k_1})-gap at x_{k_1} implies that

$$|\tilde{x}_0 - \tilde{x}_{k_1}| \geq \min \left\{ D^{-n \log_2 n} \min_{0<j\leq k_1-1} |\tilde{x}_0 - \tilde{x}_j|,\ r_{2k_1}^{4(N+N^2)},\ \prod_{j=0}^{k_1-2} |\tilde{x}_{k_1-1} - \tilde{x}_j|^{4(N+2)} \right\}.$$

Since the next weak (D,n)-gap is at $\tilde{x}_{k_2}$, we have

$$\min_{1\leq \ell \leq p_1-1} |\tilde{x}_0 - \tilde{x}_{\ell k_1}| \geq D^{-n \log_2 n} |\tilde{x}_0 - \tilde{x}_{k_1}|. \tag{3.68}$$

Combining (3.66), (3.67), and (3.68) we get a lower bound on $\prod_{j=1}^{k_2-1} |\tilde{x}_0 - \tilde{x}_j|$. To see that such a lower bound has the same form as (3.66), notice that since $k_1^{1+\delta} \geq \frac{\rho \log D}{2C} n$, all three terms on the right-hand side of (3.67) are bounded from below by $\exp(-CKk_1 n^{\frac{\delta}{1+\delta}} \log_2 n)$ for some constant $K = K(N,\rho,\delta)$. Therefore, $\prod_{j=1}^{k_2-1} |\tilde{x}_0 - \tilde{x}_j|$ has a lower bound of the same form $\exp(-CK'k_2 n^{\frac{\delta}{1+\delta}} \log_2 n)$ for some constant $K' = K'(N,\rho,\delta)$. Iterating this argument s times and letting $k_{s+1} = n$, we obtain the lower bound

$$\prod_{j=1}^{n-1} |\tilde{x}_0 - \tilde{x}_j| \geq \exp(-CK'' n^{1+\frac{\delta}{1+\delta}} \log_2 n) \tag{3.69}$$

for some constant $K'' = K''(N, \rho, \delta)$. This proves $(n, \gamma_n(C))$-simplicity of $\tilde{x}_0$ as before.

• Unfortunately, using this scheme *only for* $\gamma_n(C,\delta) = \exp(-Cn^{1+\delta})$ *we can prove that above* $\tilde{x}_0 = x_j$ *is* $(n, \gamma_n(C,\delta))$*-simple.* For example, if $\gamma_n(C,\delta) = \exp(-Cn \log n)$, then $\tilde{x}_0$ *as above might be non-*$(n, \gamma_n(C,\delta))$*-simple.* In Appendix D, we construct examples of trajectories of a full shift on two symbols that have no (n, C, δ, ρ)-leading saddles and are not $(n, \gamma_n(C,\delta))$-simple for $\gamma_n(C,\delta) = \exp(-Cn\log n)$. This shows that additional ideas are required to improve $\gamma_n(C,\delta) = \exp(-Cn^{1+\delta})$ to $\gamma_n(C,\delta) = \exp(-Cn\log n)$ or better.

3.5.1. *Decomposition of nonsimple parameters into groups.* With the notation of the general problem (2.42) and (2.43), for any $D > 2$ we introduce the set of parameters associated with an almost periodic point of period n having a gap at the k-th point of its trajectory:

$$\begin{aligned} B_n^{\mathrm{gap}(k)}(C,\delta,\rho,\vec{\mathbf{r}},\tilde{f};D) = \{&\vec{\varepsilon} \in \mathrm{HB}^N(\vec{\mathbf{r}}) : \tilde{f}_{\vec{\varepsilon}} \in \mathrm{IH}(n-1, C, \delta, \rho), \\ &\tilde{f}_{\vec{\varepsilon}} \text{ has an } (n, \gamma_n^{1/\rho}(C,\delta))\text{-periodic, but not } (n, \gamma_n(C,\delta))\text{-hyperbolic} \\ &\text{point } x_0 \text{ with a } (D, n, r_{2k})\text{-gap at } x_k = \tilde{f}^k_{\vec{\varepsilon}}(x)\}. \end{aligned} \tag{3.70}$$

Choose $D = \max\{M_{1+\rho}^{30/\rho}, \exp(C/100)\}$. Theorem 2.4.6 implies existence of inclusions (2.42) and (2.43). Similarly, the Shift Theorem implies that the following inclusions hold:

$$\begin{aligned} &B_n^{\mathrm{non}}(C,\delta,\rho,\vec{\mathbf{r}},\tilde{f}) \subseteq \cup_{k|n} B_n^{\mathrm{gap}(k)}(C,\delta,\rho,\vec{\mathbf{r}},\tilde{f};D); \\ &B_n(C,\delta,\rho,\vec{\mathbf{r}},\tilde{f}) \subseteq B_n^{\mathrm{sim}}(C,\delta,\rho,\vec{\mathbf{r}},\tilde{f}) \cup \left(\cup_{k|n} B_n^{\mathrm{gap}(k)}(C,\delta,\rho,\vec{\mathbf{r}},\tilde{f};D)\right). \end{aligned} \tag{3.71}$$

Return to our C^2-smooth 1-dimensional model, $N = \rho = 1$. To fit the notation of the model, for a sufficiently small γ_n, e.g. $\gamma_n \leq \gamma_n(C,\delta)$, we introduce the set

$$\begin{aligned} B_{n,\tau}^{\mathrm{gap}(k)}(C,\delta,\tilde{f},\gamma_n;D) = \{&\varepsilon \in \mathrm{HB}^{\mathrm{st}}_{<2n}(\tau) : \tilde{f}_\varepsilon \in \mathrm{IH}(n-1, C, \delta, 1), \\ &\tilde{f}_\varepsilon \text{ has an } (n,\gamma_n)\text{-periodic, but not } (n,\gamma_n)\text{-hyperbolic} \\ &\text{point } x_0 \text{ with a } (D, n, \tau)\text{-gap at } x_k = \tilde{f}^k_\varepsilon(x)\}, \end{aligned} \tag{3.72}$$

and rewrite inclusion (3.71) in the notation of the 1-dimensional model:

$$\begin{aligned} &B_{n,\tau}^{\mathrm{non}}(C,\delta,\tilde{f},\gamma_n) \subseteq \cup_{k|n} B_{n,\tau}^{\mathrm{gap}(k)}(C,\delta,\tilde{f},\gamma_n;D); \\ &B_{n,\tau}(C,\delta,\tilde{f},\gamma_n) \subseteq B_{n,\tau}^{\mathrm{sim}}(C,\delta,\tilde{f},\gamma_n) \cup \left(\cup_{k|n} B_{n,\tau}^{\mathrm{gap}(k)}(C,\delta,\tilde{f},\gamma_n;D)\right). \end{aligned} \tag{3.73}$$

This inclusion shows that to prove Proposition 3.5.1, it is sufficient to prove

PROPOSITION 3.5.7. *With the conditions of Proposition* 3.5.1, *let k be some integer that divides n. Then for $D = \max\{M_2^{30}, \exp(C/100)\}$ and a sufficiently small positive γ_n, e.g., $\gamma_n \le \gamma_n(C,\delta)$, we have the following estimate on the measure of maps $\tilde f_\varepsilon$'s associated with an (n,γ_n)-periodic, but not (n,γ_n)-hyperbolic point that has a (D,n,τ)-gap at the k-th point of its trajectory.*

$$\mu^{\rm st}_{<2n,\tau}\{B^{{\rm gap}(k)}_{n,\tau}(C,\delta,\tilde f,\gamma_n;D)\} \le 128 n^\delta\, 6^{2n}\, M_2^{10n+1}\, \exp(-Cn\log_2 n/200). \tag{3.74}$$

Let us give a name to the right-hand side of the inequality (3.59) for the Brick of parameters $\mathrm{HB}^{\rm st}_{<2k}(\tau)$ of the standard thickness. Let $\mathbf{X}_k = \{x_j\}_{j=0}^{k-1}$ be a k-tuple and $D > 2$ be some number. We call the number

$$\begin{aligned}(3.75)\quad &\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k, D\}\\ &= \min\left\{D^{-n\log_2 n}\min_{0<j\le k-1}|x_0-x_j|,\ \left(\frac{\tau}{(2k)!}\right)^{4(N+N^2)},\ \prod_{j=0}^{k-2}|x_{k-1}-x_j|^{4(N+2)}\right\}\end{aligned}$$

the (D,n,τ)-gap number associated to the k-tuple $\mathbf{X}_k$ and the Brick $\mathrm{HB}^{\rm st}_{<2n}(\tau)$ of standard thickness. In the case under current consideration, $N = 1$. Similarly, one can define the (D,n,r_{2k})-gap number for a nonincreasing sequence $\vec{\mathbf{r}} = \{r_m\}_{m=0}^\infty$ replacing $\tau/(2k)!$ by r_{2k}.

Let $\mathbf{X}_k(x_0,g) = \{g^j(x_0)\}_{j=0}^{k-1}$. By the definition of the (D,n,τ)-gap number, if x_0 has a (D,n,τ)-gap at x_k, then x_0 is $(k,\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k(x_0,g),D\})$-periodic. Introduce the set of "bad" parameters

$$\begin{aligned}(3.76)\quad &\\ &B^{{\rm gap}(k),\Delta^{\rm st}}_{n,\tau}(C,\delta,\tilde f,M_2;D) = \{\varepsilon\in \mathrm{HB}^{\rm st}_{<2n}(\tau):\ \tilde f_\varepsilon \in \mathrm{IH}(n-1,C,\delta,1), \tilde f_\varepsilon \text{ has}\\ &\text{a point } x_0 \text{ that is } (n,\gamma_n(C,\delta))\text{-periodic, has a } (D,n,\tau)\text{-gap at } x_k = \tilde f^k_\varepsilon(x),\\ &\text{is } (k,\gamma_n(C,\delta))\text{-simple, and is not } (k, M_2^{3n}\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k(x_0,\tilde f_\varepsilon),D\})\text{-hyperbolic}\}.\end{aligned}$$

In Section 3.6 using the Shift Theorem above we shall prove the following lemma.

LEMMA 3.5.8. *Let $C > 0$, $\delta > 0$, and $k,n\in\mathbb{Z}_+$ be some positive integer, and let k divide n. Then with the notation above for any $\gamma_n \le \gamma_n(C,\delta)$ and $D = \max\{M_2^{30}, \exp(C/100)\}$,*

$$B^{{\rm gap}(k)}_{n,\tau}(C,\delta,\tilde f,\gamma_n;D) \subset B^{{\rm gap}(k),\Delta^{\rm st}}_{n,\tau}(C,\delta,\tilde f,M_2;D). \tag{3.77}$$

The proof is postponed until Section 3.6.

Remark 3.5.9. $B^{{\rm gap}(k),\Delta^{\rm st}}_{n,\tau}(C,\delta,\tilde f,M_2;D)$ depends only on properties of trajectories of length $k < n$, whereas $B^{{\rm gap}(k)}_{n,\tau}(C,\delta,\tilde f,\gamma_n;D)$ depends on properties of those of length n. This reduces consideration of sufficiently recurrent, nonsimple, and not hyperbolic, trajectories of length n to studying only their initial *simple* parts of length k, where x_k is a point of the first sufficiently close return to x_0 with n divisible by k. In this case one can effectively perturb such a trajectory at the k-th return using the Newton family $\tilde{f}(x) + u_0 + u_1(x - x_0) + \cdots + u_{2k-1}(x - x_{k-1}) \prod_{j=0}^{k-2}(x - x_j)^2$.

3.5.2. *Decomposition into i-th recurrent pseudotrajectories.* We now split the set $B_{n,\tau}^{\mathrm{gap}(k),\Delta^{\mathrm{st}}}(C, \delta, \tilde{f}, M_2; D)$ into a finite union. Let

$$\tilde{\gamma}_{n,i}(C, \delta, \rho) = M_2^{4ni/\rho} \gamma_n^{1/4\rho}(C, \delta).$$

Definition 3.5.10. Let $g \in \mathrm{Diff}^{1+\rho}(B^N)$ be a $C^{1+\rho}$-smooth diffeomorphism (respectively $g \in C^{1+\rho}(I, I)$ be a $C^{1+\rho}$-smooth map), $C > 0$, $\delta > 0$, $\tau > 0$, and $D > 2$ be some constants. We call a point $x_0 \in B^N$ (respectively $x_0 \in I$) the i-th *recurrent with constants* $(C, \delta, \rho, \tau, D)$ if for $\mathbf{X}_k(x_0, g) = \{g^j(x_0)\}_{j=0}^{n-1}$,

$$M_2^{2n/\rho}\ \tilde{\gamma}_{n,i}(C, \delta, \rho)\ \le\ \Delta_{k,n,\tau}^{\mathrm{st}}\{\mathbf{X}_k(x_0, g), D\}\ < M_2^{2n/\rho}\ \tilde{\gamma}_{n,i+1}(C, \delta, \rho). \tag{3.78}$$

Remark 3.5.11. Recall that the (D, n, r_{2k})-gap number is defined by replacing, in the definition of the (D, n, τ)-gap number, $\tau/(2k)!$ by r_{2k}. Similarly, we define an i-th recurrent point with constants $(C, \delta, \rho, r_{2k}, D)$ by replacing, in Definition 3.5.10 above, the (D, n, τ)-gap number $\Delta_{k,n,\tau}^{\mathrm{st}}\{\mathbf{X}_k, D\}$ by the (D, n, r_{2k})-gap number.

For the purpose of this section, $\rho = 1$, and for brevity redenote $\tilde{\gamma}_{n,i} = \tilde{\gamma}_{n,i}(C, \delta, 1)$. Define the set of parameters from $B_{n,\tau}^{\mathrm{gap}(k),\Delta^{\mathrm{st}}}(C, \delta, \tilde{f}, M_2; D, i)$ associated to i-th order recurrent with respect to $(C, \delta, 1, D)$ trajectories that satisfy conditions (3.76):

$$B_{n,\tau}^{\mathrm{gap}(k),\Delta^{\mathrm{st}}}(C, \delta, \tilde{f}, M_2; D, i) = \{\varepsilon \in \mathrm{HB}_{<2n}^{\mathrm{st}}(\tau) :\ \tilde{f}_\varepsilon \text{ has a point } x_0 \text{ as in (3.76) and } i\text{-th recurrent with constants } (C, \delta, 1, \tau, D)\}. \tag{3.79}$$

Lemma 3.5.12. *With the notation of Lemma* 3.5.8, *for* $L = [Cn^\delta/(4 \log M_2)]$,

$$B_{n,\tau}^{\mathrm{gap}(k),\Delta^{\mathrm{st}}}(C, \delta, \tilde{f}, M_2; D) \subseteq \bigcup_{i=0}^{L-1} B_{n,\tau}^{\mathrm{gap}(k),\Delta^{\mathrm{st}}}(C, \delta, \tilde{f}; M_2, D, i). \tag{3.80}$$

Lemma 3.5.13. *Let $C > 0$, $\delta > 0$, and $n \in \mathbb{Z}_+$ be some numbers, and let $k \in \mathbb{Z}_+$ divide n. Then with the notation as above, for*

$$D = \max\{M_2^{30}, \exp(C/100)\},$$

any $\gamma_n \le \gamma_n(C, \delta)$, and any $i \in \mathbb{Z}_+$ such that $0 \le i < L$,

$$\mu_{<2n,\tau}^{\mathrm{st}}\{B_{n,\tau}^{\mathrm{gap}(k),\Delta^{\mathrm{st}}}(C, \delta, \tilde{f}, M_2; D, i)\} \le 128 \cdot 6^{2k}\ M_2^{10n+1}\ D^{-n \log_2 n/2}. \tag{3.81}$$

We postpone the proof of these lemmas until the next section.

Proof of Proposition 3.5.1. Since there are at most n numbers $k < n$ dividing n, Proposition 3.5.1 is a corollary of Proposition 3.5.7 and the Shift Theorem. Let γ_n be sufficiently small, e.g., $\gamma_n \le \gamma_n(C,\delta)$, and let $D = \max\{M_2^{30}, \exp(C/100)\}$. Then by the Shift Theorem, the set of all parameter values $B_{n,\tau}^{\text{non}}(C,\delta,\tilde f,\gamma_n)$ associated to maps $\tilde f_\varepsilon$ with an (n,γ_n)-periodic, essentially non-(n,γ_n)-simple, but not (n,γ_n)-hyperbolic point, is contained in the union (3.73) over all k dividing n of $B_{n,\tau}^{\text{gap}(k)}(C,\delta,\tilde f,\gamma_n;D)$ associated to maps $\tilde f_\varepsilon$ with an (n,γ_n)-periodic, but non-(n,γ_n)-hyperbolic point with a (D,n,τ)-gap at x_k. Q.E.D.

Proof of Proposition 3.5.7. Let γ_n and D be as in the proof above. We combine Lemmas 3.5.8, 3.5.12, and 3.5.13 as follows.

• By Lemma 3.5.8, we have that $B_{n,\tau}^{\text{gap}(k)}(C,\delta,\tilde f,\gamma_n;D)$ is contained in the set of parameters $B_{n,\tau}^{\text{gap}(k),\Delta^{\text{st}}}(C,\delta,\tilde f,M_2;D)$ associated to maps $\tilde f_\varepsilon$ which have a non-$(k,\Delta_{k,n,\tau}^{\text{st}}\{\mathbf{X}_k(x_0,\tilde f_\varepsilon),D\})$-hyperbolic point x_0 with a (D,n,τ)-gap at x_k.

• By Lemma 3.5.12, in turn, $B_{n,\tau}^{\text{gap}(k),\Delta^{\text{st}}}(C,\delta,\tilde f,M_2;D)$ is contained in the union of $\{B_{n,\tau}^{\text{gap}(k),\Delta^{\text{st}}}(C,\delta,\tilde f,M_2;D,i)\}_{i=0}^{L-1}$ such that the i-th set

$$B_{n,\tau}^{\text{gap}(k),\Delta^{\text{st}}}(C,\delta,\tilde f,M_2;D,i)$$

is associated to maps $\tilde f_\varepsilon$ that have a non-$(k,\Delta_{k,n,\tau}^{\text{st}}\{\mathbf{X}_k(x_0,\tilde f_\varepsilon),D\})$-hyperbolic point x_0 with a (D,n,τ)-gap at x_k and such that the k-tuple $\{x_j = \tilde f_\varepsilon^j(x_0)\}_{j=0}^{k-1}$ is i-th recurrent.

• Lemma 3.5.13 then estimates the measures of $B_{n,\tau}^{\text{gap}(k),\Delta^{\text{st}}}(C,\delta,\tilde f,M_2;D,i)$ sufficiently well to prove Proposition 3.5.7. Q.E.D.

In the next section, we shall prove Lemma 3.5.13 using the Discretization Method and then we shall prove Lemmas 3.5.8 and 3.5.12 using simple approximation arguments similar to the one given in the proof of Proposition 3.1.2.

3.6. *The measure of maps $\tilde f_\varepsilon$ having i-th recurrent, insufficiently hyperbolic trajectories with a gap and proofs of auxiliary lemmas.* We shall prove Lemma 3.5.13 in three steps.

Step 1. *Reduction to polynomial perturbations of degree* $2k-1$. The measure $\mu_{<2n,\tau}^{\text{st}}$ is Lebesgue product probability measure and each of its components $\mu_{m,\tau}^{\text{st}}$ is Lebesgue probability measure (see (3.3))

$$\mu_{<2n,\tau}^{\text{st}} = \mu_{<2k,\tau}^{\text{st}} \times \left(\times_{m=2k}^{2n-1}\mu_{m,\tau}^{\text{st}}\right). \tag{3.82}$$

Therefore, by the Fubini/Tonelli theorem it is sufficient to prove that

$$\mu^{\rm st}_{<2k,\tau}\left\{B^{{\rm gap}(k),\Delta^{\rm st}}_{n,\tau}(C,\delta,\tilde f,M_2;D,i)\cap\{(\varepsilon_{2k},\dots,\varepsilon_{2n-1})\}\right\} \le 128\cdot 6^{2k}\ M_2^{10n+1}\ D^{-n\log_2 n/2}$$

uniformly over $\{(\varepsilon_{2k},\dots,\varepsilon_{2n-1})\}\in {\rm HB}^{\rm st}_{2k}(\tau)\times{\rm HB}^{\rm st}_{2k+1}(\tau)\times\cdots\times{\rm HB}^{\rm st}_{2n-1}(\tau)$. To simplify notation we omit $(\varepsilon_{2k},\dots,\varepsilon_{2n-1})$ and write as if the set of parameters $B^{{\rm gap}(k),\Delta^{\rm st}}_{n,\tau}(C,\delta,\tilde f,M_2;D,i)$ is a subset of the Brick ${\rm HB}^{\rm st}_{<2k}(\tau)$.

From now on we consider the $2k$-parameter family of polynomial perturbations

$$\left\{\tilde f_{\varepsilon(2k-1)}=\tilde f(x)+\sum_{m=0}^{2k-1}\varepsilon_m x^m\right\}, \tag{3.83}$$

where $\varepsilon(2k-1)=(\varepsilon_0,\dots,\varepsilon_{2k-1})\in{\rm HB}^{\rm st}_{<2k}(\tau)$. Recall that ${\rm HB}^{\rm st}_{<2k}(\tau)$ is supplied with the Lebesgue product probability measure $\mu^{\rm st}_{<2k,\tau}$.

Step 2. *An estimate of the measure of parameters associated with a trajectory* $\{x_j\}_{j=0}^{n-1}$ *with a gap at* x_k *that is* i*-th recurrent and not sufficiently hyperbolic.* With the notation of Section 3.2 we give the following:

Definition 3.6.1. Let $D>2$. We say that a $\tilde\gamma_n$*-pseudotrajectory* $\mathbf{X}_k=\{x_j\}_{j=0}^{k-1}\subset I^k_{\tilde\gamma_n}$ associated to some map $g: I\hookrightarrow I$ has a (D,n,τ)-gap at a point $x_k=g(x_{k-1})$ if (3.59) is satisfied and there is no $m<k$ such that x_0 has a (D,n,τ)-gap at x_m.

By analogy with (3.15), for each i satisfying $0\le i<L$ of Lemma 3.5.12 we define

(3.84)

$B^{{\rm gap}(k),\Delta^{\rm st}}_{n,\tau}(\tilde f,\tilde\gamma_{n,i},M_2;D,i;x_0,\dots,x_{k-1})=\{\varepsilon\in{\rm HB}^{\rm st}_{<2k}(\tau):$
$\mathbf{X}_k=\{x_m\}_{m=0}^{k-1}\subset I^k_{\tilde\gamma_{n,i}}$ is a $\tilde\gamma_{n,i}$-pseudotrajectory associated to ε that is
i-th recurrent with constants $(C,\delta,1,\tau,D)$, $\left(k,2M_2^n\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k,D\}\right)$-periodic
and not $\left(k,2M_2^{3n}\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k,D\}\right)$-hyperbolic$\}$.

To show an analogy with the simple trajectory case consider the following

Problem 3.6.2. Estimate the measure of $\varepsilon\in{\rm HB}^{\rm st}_{<2k}(\tau)$ for which the k-tuple $\mathbf{X}_k=\{x_j\}_{j=0}^{k-1}$ is

(3.85)

$A)$ a $\tilde\gamma_{n,i}$-pseudotrajectory, i.e., $|\tilde f_\varepsilon(x_j)-x_{j+1}|\le\tilde\gamma_{n,i}$ for $j=0,\dots,k-2$;

$B)$ i-th recurrent with constants $(C,\delta,1,\tau,D)$, i.e.,
$M_2^{2n/\rho}\ \tilde\gamma_{n,i}\le\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k,D\}<M_2^{2n/\rho}\ \tilde\gamma_{n,i+1}$;

$C)$ $(k, 2M^n\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k, D\})$-periodic, i.e.,

$|\tilde f_\varepsilon(x_{k-1}) - x_0| < 2M^n\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k, D\}$;

$D)$ not $(n, 2M_2^{3n}\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k, D\})$-hyperbolic, i.e.,

$$\left|\left|\prod_{j=0}^{k-1}(\tilde f_\varepsilon)'(x_j)\right| - 1\right| \le 2M_2^{3n}\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k, D\}.$$

For a fixed k-tuple of points $\mathbf{X}_k = \{x_j\}_{j=0}^{k-1}$, consider the Newton family of polynomial perturbations

$$\tilde f_{u,\mathbf{X}_k}(x) = \tilde f(x) + \sum_{m=0}^{2k-1} u_m \prod_{j=0}^{m-1}(x - x_{j \pmod k}). \tag{3.86}$$

Recalling (2.17) and Figure 2.2, we notice that for any $m < k$ and $s > 0$ the image $\tilde f_{u,\mathbf{X}_k}(x_m)$ (respectively derivative $\tilde f'_{u,\mathbf{X}_k}(x_m)$) of (respectively at) the point x_m is independent of the Newton coefficients u_{m+s} (respectively u_{m+k+s}) with $s > 0$. This implies that the Newton coefficients $u_0, \dots, u_{k-2}$ determine if property (A) holds. Fix $u_0, \dots, u_{k-2}$. Now the Newton coefficient u_{k-1} determines if property (C) of almost periodicity holds. Once u_{k-1} is fixed, the Newton coefficient u_{2k-1} determines if property (D) of almost nonhyperbolicity holds.

Following formulas (3.27), (3.28), and formulas (3.30), (3.33) with n replaced by k and γ_n replaced by $2M^n\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k, D\}$ for periodicity and $2M^{3n}\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k, D\}$ for hyperbolicity, we have that for fixed $\mathbf{X}_k$, the measure of $(u_0, \dots, u_{k-1}, u_{2k-1})$ with conditions (3.85) is bounded as in (3.34). Then we apply the Distortion Lemma with n replaced by k. This gives an additional factor 6^k. Thus we get a bound

$$\begin{aligned} &\mu^{\rm st}_{<2k,\tau}\Big\{u(2k-1) \in W^{u,\mathbf{X}_n}_{<2k,1} : \ \tilde f_{u,\mathbf{X}_k} \text{ satisfies conditions (3.85)}\Big\} \\ &\le 4\ M_2\ 6^k \prod_{m=0}^{k-2} \frac{m!\ \tilde\gamma_{n,i}}{\tau\ \prod_{j=0}^{m-1}|x_m - x_j|} \\ &\frac{(k-1)!}{\tau}\ \frac{2M_2^n\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k, D\}}{\prod_{j=0}^{k-2}|x_{k-1} - x_j|}\ \frac{(2k-1)!}{\tau}\ \frac{2M_2^{3n}\ \Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k, D\}}{\prod_{j=0}^{k-2}|x_{k-1} - x_j|^2}. \end{aligned} \tag{3.87}$$

Consider the last two terms in the right-hand side product. Definitions of the (D, n, τ)-gap number and the inequality $\min(a, b, c) \le a^{1/2}\ b^{1/4}\ c^{1/4}$ show that these two terms are bounded by

$$M_2^{4n}\frac{k!}{\tau}\frac{(2k)!}{\tau}\frac{\left(\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k, D\}\right)^2}{(\prod_{j=0}^{k-2}|x_{k-1} - x_j|)^3} \le 2M_2^{4n}D^{-n\log_2 n/2}\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k, D\}. \tag{3.88}$$

Therefore,

(3.89)

$$\mu^{\rm st}_{<2k,\tau}\Big\{u(2k-1)\in W^{u,\mathbf{X}_n}_{<2k,1}:\ \tilde f_{u,\mathbf{X}_k} \text{ satisfies conditions (3.85)}\Big\}$$
$$\le 32\ M_2^{4n+1}\ 6^k\ \prod_{m=0}^{k-2}\frac{m!\ \tilde\gamma_n}{\tau\ \prod_{j=0}^{m-1}|x_m-x_j|}\ M_2^{4n}\ D^{-n\log_2 n/2}\ \Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k,D\}.$$

The next step of the proof of Lemma 3.5.13, as in the case of simple trajectories (§3.2), is to collect all possible "bad" pseudotrajectories and make sure that those pseudotrajectories indeed approximate sufficiently well all "bad" true trajectories.

Step 3. *Collection of "bad" i-th recurrent, not sufficiently hyperbolic trajectories $\{x_j\}_{j=0}^{k-1}$ with a gap at x_k using grids $\{I_{\tilde\gamma_{n,i}}\}_i$ of variable size in i.* In the case of simple trajectories in Section 3.2, we considered only one $\tilde\gamma_n$-grid of a fixed size and collected all simple "bad" trajectories in the Collection Lemma (§3.4). In the case of i-th recurrent trajectories with a gap at x_k we define grids $\{I_{\tilde\gamma_{n,i}}\}_i$ of i dependent size $\tilde\gamma_{n,i}$. Then we prove that $\tilde\gamma_{n,i}$-pseudotrajectories approximate real i-th recurrent trajectories with a gap at x_k sufficiently well. Finally, we collect all possible i-th recurrent $\tilde\gamma_{n,i}$-pseudotrajectories with a gap at x_k and sum the estimates of the measures of "bad" sets. Let's realize this program. Recall that $\tilde\gamma_{n,i}=M_2^{4ni}\gamma_n(C,\delta)$ and call the i-th grid

$$I_{\tilde\gamma_{n,i}}=\{x\in I:\exists k\in\mathbb{Z}\text{ such that }x=(2k+1)\tilde\gamma_{n,i}\}\subset I. \tag{3.90}$$

Definition 3.6.3. Let $\{x_j\}_{j=0}^{k-1}\in I^k_{\tilde\gamma_{n,i}}$ be a k-tuple for some $i\in\mathbb{Z}_+$. Then the k-tuple $\mathbf{X}_k=\{x_j\}_{j=0}^{k-1}$ is called i-th recurrent with constants $D>2$ and $M_2>1$ if

$$\frac{1}{2}\ M_2^{2n}\ \tilde\gamma_{n,i}\le\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k,D\}\le 2\ M_2^{2n}\ \tilde\gamma_{n,i+1}. \tag{3.91}$$

For simplicity if the k-tuple $\mathbf{X}_k=\{x_j\}_{j=0}^{k-1}$ is a $\tilde\gamma_{n,i}$-pseudotrajectory associated to some parameter $\varepsilon\in{\rm HB}^{\rm st}_{<2k}(\tau)$ which is i-th recurrent with some constants $D>2$ and the C^2-norm M_2 of the family (3.7), then we say that $\mathbf{X}_k$ is an i-th recurrent $\tilde\gamma_{n,i}$-pseudotrajectory.

Remark 3.6.4. Definition of an i-th recurrent pseudotrajectory is reasonable only for a grid of a sufficiently small size. Indeed, for any i-th recurrent trajectory we need to find a $\tilde\gamma_{n,i}$-pseudotrajectory whose periodicity, hyperbolicity, and product of distances along itself approximate well enough those of the trajectory.

Define a discretized version of the set $B^{{\rm gap}(k),\Delta^{\rm st}}_{n,\tau}(C,\delta,\tilde f,M_2;D,i)$, given by (3.79) associated with all i-th recurrent trajectories of length k which has a weak (D, n, τ)-gap at x_k and is not $M_2^{3n} \Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k(x_0, \tilde f_\varepsilon), D\}$-hyperbolic.

$$\begin{aligned} (3.92)\quad & B^{{\rm gap}(k),\Delta^{\rm st}}_{n,\tau}(C, \delta, \tilde f, M_2, \tilde\gamma_{n,i}; D, i; x_0) \\ & = \cup_{\mathbf{X}_k = \{x_j\}_{j=1}^{k-1} \in I^{k-1}_{\tilde\gamma_{n,i}}} B^{{\rm gap}(k),\Delta^{\rm st}}_{n,\tau}(C, \delta, \tilde f, M_2, \tilde\gamma_{n,i}; D, i; x_0, \dots, x_{k-1}). \end{aligned}$$

Similarly to the case of simple trajectories, we prove that after discretization all real "bad" trajectories can be sufficiently well approximated by pseudo-trajectories of a certain grid so that quantities of periodicity, hyperbolicity, existence of a gap, and product of distances along the trajectory are almost the same. Namely,

LEMMA 3.6.5. *In the notation above,*

$$(3.93)\quad B^{{\rm gap}(k),\Delta^{\rm st}}_{n,\tau}(C, \delta, \tilde f; D, i) \subset \cup_{x_0 \in I_{\tilde\gamma_{n,i}}} B^{{\rm gap}(k),\Delta^{\rm st}}_{n,\tau}(C, \delta, \tilde f, M_2, \tilde\gamma_{n,i}; D, i; x_0).$$

The proof of this lemma is very similar to the proof of Lemma 3.5.8, given below, and is omitted.

To complete the proof of Lemma 3.5.13 we apply the Collection Lemma to (3.89) to estimate the measure of $B^{{\rm gap}(k),\Delta^{\rm st}}_{n,\tau}(C, \delta, \tilde f, M_2, \tilde\gamma_{n,i}; D, i; x_0)$. We get

$$\begin{aligned} & \mu^{\rm st}_{<2k,\tau}\{B^{{\rm gap}(k),\Delta^{\rm st}}_{n,\tau}(C, \delta, \tilde f, M_2, \tilde\gamma_{n,i}; D, i; x_0)\} \\ & \qquad \le 32 \cdot 6^{2k} M_2^{4n+1} D^{-n \log_2 n/2} \Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k, D\}. \end{aligned}$$

The number of grid points is $2\gamma_{n,i}^{-1}$. By definition of i-th recurrent pseudotrajectory we have $2\gamma_{n,i}^{-1} \Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k, D\} \le 4M_2^{6n}$. This implies (3.81), which in turn implies Lemma 3.5.13.

Proof of Lemma 3.5.8. Fix a parameter $\varepsilon \in B^{{\rm gap}(k)}_{n,\tau}(C, \delta, \tilde f, \gamma_n; D)$; we wish to show that $\varepsilon \in B^{{\rm gap}(k),\Delta^{\rm st}}_{n,\tau}(C, \delta, \tilde f, M_2; D)$. By definition (3.72) there is an (n, γ_n)-periodic, non-(n, γ_n)-hyperbolic point x_0 with a (D, n, τ)-gap at $x_k = \tilde f^k_\varepsilon(x_0)$. By definition (3.76), we want to show that x_0 is not $(k, M_2^{3n} \Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k(x_0, \tilde f_\varepsilon), D\})$-hyperbolic.

Since n is divisible by k, we can split the trajectory $\{x_j = \tilde f^j_\varepsilon(x_0)\}_{j=0}^{n-1}$ of length n into $p = n/k$ parts of length k each. Consider the linearization

$$(3.94)\quad (\tilde f^n_\varepsilon)'(x_0) = (\tilde f^k_\varepsilon)'(x_{(p-1)k}) \cdot \dots \cdot (\tilde f^k_\varepsilon)'(x_k) \cdot (\tilde f^k_\varepsilon)'(x_0).$$

Definition 3.5.3 of a (D, n, τ)-gap at x_k says that $|x_0 - x_k| \le \Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k(x_0, \tilde f_\varepsilon), D\}$, and hence for every $1 \le j \le p-1$ and $0 \le s < k$,

$$(3.95)\quad |x_s - x_{jk+s}| \le M_1^n \ \Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k(x_0, \tilde f_\varepsilon), D\}.$$

This implies that the trajectory $\{x_j\}_{j=0}^{n-1}$ of length n consists of p almost identical parts of length k each. Thus, for each $0 \le m < n$,

$$\left|(\tilde{f}_\varepsilon)'(x_m) - (\tilde{f}_\varepsilon)'(x_{m \pmod k})\right| \le M_2^{n+1}\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k(x_0,\tilde{f}_\varepsilon),D\}. \tag{3.96}$$

Considering the product over $m = 0, 1, \dots, n-1$ of each term on the left-hand side above, we get that

$$\left|(\tilde{f}_\varepsilon^n)'(x_0) - \Big((\tilde{f}_\varepsilon^k)'(x_0)\Big)^p\right| \le nM_2^{2n}\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k(x_0,\tilde{f}_\varepsilon),D\}. \tag{3.97}$$

Therefore, since $\left|\left|(\tilde{f}_\varepsilon^n)'(x_0)\right| - 1\right| \le \gamma_n \le \gamma_n(C,\delta)$, and by the proof of Lemma 3.5.12 below, $\gamma_n(C,\delta) < \Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k(x_0,\tilde{f}_\varepsilon),D\}$, it follows that

$$\left||(\tilde{f}_\varepsilon^k)'(x_0)| - 1\right| \le \left|\,|\Big((\tilde{f}_\varepsilon^k)'(x_0)\Big)^p| - 1\right| \le 2nM_2^{2n}\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k(x_0,\tilde{f}_\varepsilon),D\}, \tag{3.98}$$

and ε belongs to $B_{n,\tau}^{{\rm gap}(k),\Delta^{\rm st}}(C,\delta,\tilde{f},M_2;D)$. Q.E.D.

Proof of Lemma 3.5.12. It follows from the Shift Theorem (last sentence) with $N = 1$ that if a point x_0 has no $(n,C,\delta,1)$-leading saddles and a (D,n,τ)-gap at x_k, then x_0 is $(k,\gamma_n(C,\delta))$-simple. Therefore, for the (D,n,τ)-gap number (3.75) we have the following bounds

$$\gamma_n(C,\delta) = \exp\Big(-Cn^{1+\delta}\Big) \le \Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k(x_0,\tilde{f}_\varepsilon),D\} \le 2D^{-n}. \tag{3.99}$$

Thus since $D = M_2^{30} \ge 2$, each such x_0 with a (D,n,τ)-gap at x_k is i-th recurrent with constants (C,δ,ρ,τ,D) for some $i \ge 0$. If $i \ge L = Cn^\delta/(4\log M_2)$, then $\Delta^{\rm st}_{k,n,\tau}\{\mathbf{X}_k(x_0,\tilde{f}_\varepsilon),D\} \ge M_2^{4ni}\gamma_n(C,\delta) \ge 1 > 2D^{-n}$. This contradicts the above inequality. Q.E.D.

4. Comparison of the discretization method in 1-dimensional and N-dimensional cases

In this section we discuss changes that one has to make and difficulties that arise while generalizing the Discretization Method for C^2-smooth 1-dimensional maps to a Discretization Method for the $C^{1+\rho}$-smooth N-dimensional diffeomorphisms. Recall that the proof of the Main Theorem (see Section 1.3) reduces to estimating the measure of the set of "bad" parameters (2.1). In Section 2.4, we show how to decompose this set into "simple" pieces. Namely, the set $B_n(C,\delta,\rho,\vec{\mathbf{r}},\tilde{f})$ belongs to a finite union of sets. Each of these sets consists of "bad" parameters, which are bad due to existence of a simple nonhyperbolic trajectory of length either n or some k dividing n. To estimate the measure of such sets, we need to carry out the Discretization Method. The decomposition into pseudotrajectories of Section 3.2, remains essentially the same. Therefore, we focus primarily on the obstacles to generalizing the estimates of Sections 3.3 and 3.4 for simple trajectories. In Sections 10 and 11 of [GK], we develop the Discretization Method for $N = 2$ in a situation very similar to ours.

4.1. *Dependence of the main estimates on N and ρ.* Recall that the Inductive Hypothesis (Definition 2.0.1) asserts that points that are $(n, \gamma_n^{1/\rho}(C, \delta))$-periodic are also $(n, \gamma_n(C, \delta))$-hyperbolic. The reason for the exponent $1/\rho$ is as follows. In the Discretization Method, we approximate trajectories of length n with $\tilde{\gamma}_n$-pseudotrajectories on a grid. In order that γ_n-hyperbolicity of the pseudotrajectory imply hyperbolicity of the true trajectory for a $C^{1+\rho}$ map, we need that $\tilde{\gamma}_n \le \gamma_n^{1/\rho}$ (up to a factor exponential in n). In our heuristic estimate (3.24) on the measure of "bad" parameters, the number of initial points in the $\tilde{\gamma}_n$-grid is (again up to an exponential factor) $\tilde{\gamma}_n^{-N} \ge \gamma_n^{-N/\rho}$. The best possible bounds one can get (for nonrecurrent trajectories) on the "measure of periodicity" and the "measure of hyperbolicity" in (3.24) are respectively the N-th power of the periodicity and the hyperbolicity γ_n. Thus, in order for the right side of (3.24) to be small, we need that the periodicity be approximately bounded by $\gamma_n^{1/\rho}$.

Next, we explain the exponent $1/(4N)$ in Definition 2.4.1 of simple trajectories. The actual estimate we obtain on the "measure of periodicity" discussed in the previous paragraph is not the N-th power of the periodicity $\gamma_n^{N/\rho}$, but instead $\gamma_n^{N/\rho}\ r_{n-1}^{-N}\ (\prod_{j=0}^{n-2} |x_{n-1} - x_j|)^{-N}$, where $\{x_j\}_{j=0}^{n-1}$ is a trajectory and r_{n-1} is the width of an admissible Hilbert Brick in the direction of degree $n-1$ polynomials (see Definition 1.3.1). This estimate reduces to (3.30) in the case $N = \rho = 1$ and $r_{n-1} = \tau/(n-1)!$, and is obtained in a similar fashion, treating each of the N coordinates independently. In Appendix A, we will show (Proposition A.5) that the analogue of the bound (3.33) on the "measure of hyperbolicity" is, up to a factor exponential in n, $\gamma_n\ r_{2n-1}^{-N^2}\ (\prod_{j=0}^{n-2} |x_{n-1} - x_j|)^{-2}$. Again ignoring exponential factors, with $\tilde{\gamma}_n = \gamma_n^{1/\rho}$ the number of initial points in (3.24) is $\gamma_n^{-N/\rho}$, making the right side of (3.24)

$$\gamma_n\ r_{n-1}^{-N}\ r_{2n-1}^{-N^2} \left(\prod_{j=0}^{n-2} |x_{n-1} - x_j| \right)^{-N-2}. \tag{4.1}$$

By Definition 1.3.1, for admissible Hilbert Bricks, γ_n decays faster than any power of r_{2n}. Thus to make the bound (3.24) on the measure of "bad" parameters small, we basically need γ_n to dominate $(\prod_{j=0}^{n-2} |x_{n-1} - x_j|)^{-N-2}$. This is certainly true if $\prod_{j=0}^{n-2} |x_{n-1} - x_j| \ge \gamma_n^{1/(4N)}$. Though our choice of the particular exponent $1/(4N)$ is somewhat arbitrary, the factor of N in the exponent is necessary.

4.2. *The multidimensional space of divided differences and dynamically essential parameters.* In the 1-dimensional case, with a fixed n-tuple of points $\{x_j\}_{j=0}^{n-1} \subset I$, the space of Newton Interpolation Polynomials (the Divided Differences Space) is $2n$-dimensional. In Section 2.2, in formulas (2.17), we noticed that there is a simple relation between parameters of Newton Interpolation Polynomials and dynamical properties of the trajectory. In terms of the family (3.86), we have that u_0 determines the position of $f_{u(2n-1);\mathbf{X}_n}(x_0)$, that u_1 determines the position of $f_{u(2n-1);\mathbf{X}_n}(x_1)$, provided that u_0 (and hence $f_{u(2n-1);\mathbf{X}_n}(x_0)$) is fixed, and so on — for $k = 2, \dots, n-1$, we have that u_k determines the position of $f_{u(2n-1);\mathbf{X}_n}(x_k)$, provided that $\{u_j\}_{j=0}^{k-1}$ (hence $\{f_{u(2n-1);\mathbf{X}_n}(x_j)\}_{j=0}^{k-1}$) are fixed. Similarly, for $k = 0, \dots, n-1$, we have that u_{n+k} determines the derivative $(f_{u(2n-1);\mathbf{X}_n})'(x_k)$, provided that $\{u_j\}_{j=0}^{n+k-1}$ (hence the positions $\{f_{u(2n-1);\mathbf{X}_n}(x_j)\}_{j=0}^{n-1}$ and derivatives $\{(f_{u(2n-1);\mathbf{X}_n})'(x_j)\}_{j=0}^{k-1}$) are fixed.

This correspondence makes transparent estimates of the measure of parameters associated with a particular (pseudo-)trajectory having a given property of periodicity and hyperbolicity (see (3.27)–(3.33)). In the multidimensional case ($N > 1$), such a correspondence between dynamical properties of trajectories and coefficients of Newton Interpolation Polynomials becomes much less transparent.

In Section 2.2, we define the space of Divided Differences $\mathcal{DD}^{1,n}(I, \mathbb{R}) = I^n \times \mathbb{R}^{2n}$ in the 1-dimensional case, where $I = [-1, 1]$. In this case, we estimated the measure of "bad" parameters in Sections 3.3–3.4.

Recall now the notation of Section 1.3. In N dimensions, we define the space of Divided Differences

$$
\begin{aligned}
(4.2)\qquad \mathcal{DD}^{N,n}(B^N, \mathbb{R}^N) &= \Big\{ \big(x_0, \dots, x_{n-1}; \{\vec{u}_\alpha\}_{|\alpha|=0}, \dots, \{\vec{u}_\alpha\}_{|\alpha|=2n-1}\big) \\
&\qquad \in \underbrace{B^N \times \cdots \times B^N}_{n \text{ times}} \times \mathbb{R}^{\nu(0,N)} \times \cdots \times \mathbb{R}^{\nu(2n-1,N)} \Big\} \\
&= \underbrace{B^N \times \cdots \times B^N}_{n \text{ times}} \times W_{0,N}^{u,\mathbf{X}_0} \times W_{1,N}^{u,\mathbf{X}_1} \cdots \times W_{n-1,N}^{u,\mathbf{X}_n} \times W_{n,N}^{u,\mathbf{X}_n} \times \cdots \times W_{2n-1,N}^{u,\mathbf{X}_n},
\end{aligned}
$$

where B^N is the N-dimensional unit ball, $\nu(k, N)$ is N times the number of N-dimensional multiindices α with $|\alpha| = k$, $\mathbf{X}_k = \{x_j\}_{j=0}^{k-1}$, and $W_{k,N}^{u,\mathbf{X}_{k (\mathrm{mod}\ n)}}$ is the space of homogeneous polynomials of degree k from $\mathbb{R}^N$ to $\mathbb{R}^N$ with the Newton basis defined below. There are two issues we face that were not a concern for the 1-dimensional Newton basis (2.26).

Nonuniqueness. It turns out that the choice of a basis in the space of Divided Differences $\mathcal{DD}^{N,n}(B^N, \mathbb{R}^N)$ and the definition of the Newton map

$$
(4.3)\qquad \mathcal{L}_{\mathbf{X}_n}^N : \big(\{\vec{\varepsilon}_\alpha\}_{|\alpha| \le 2n-1}\big) \to \big(\{\vec{u}_\alpha\}_{|\alpha| \le 2n-1}\big)
$$

(defined by (2.24) for $N = 1$) for a multi-index $\alpha \in \mathbb{Z}_+^N$ is far from unique. In the 1-dimensional case, the standard basis is $\{x^k\}_{k=0}^{2n-1}$ and the Newton basis is $\left\{\prod_{j=0}^{k-1} |x - x_j|\right\}_{k=0}^{2n-1}$. In the N-dimensional case, $(x - x_j) \in \mathbb{R}^N$ is an N-dimensional vector. For a fixed basis in $\mathbb{R}^N$, let $(x - x_j)_s$ denote the s-th coordinate of the vector $(x - x_j)$. The number of different monomials of the form

$$\left\{\prod_{j=0}^{k-1} (x_k - x_j)_{i(j)}\right\}_{\{i(0),\dots,i(k-1)\}\in\{1,\dots,N\}^k} \tag{4.4}$$

is N^k, while the number of homogeneous monomials in N variables of degree k, i.e. $\{x^\alpha\}_{|\alpha|=k}$, is bounded above by k^N, which is much smaller than N^k for $k \gg N$.

Therefore, among the monomials (4.4) we need to choose an appropriate basis and define an appropriate Newton map $\mathcal{L}_{\mathbf{X}_n}^N$. The *standard* way to choose a Newton basis [GY] is as follows. For $\alpha \in \mathbb{Z}_+^N$, let the Newton basis monomial for the multi-index α be

$$\begin{aligned}(x; x_0, \dots, x_{(|\alpha|-1) \pmod n})^\alpha &= \prod_{i_1=0}^{\alpha_1-1} (x - x_{i_1})_1 \\ &\times \prod_{i_2=0}^{\alpha_2-1} (x - x_{\alpha^1+i_2})_2 \times \cdots \times \prod_{i_N=0}^{\alpha_N-1} (x - x_{\alpha^{N-1}+i_N})_N,\end{aligned} \tag{4.5}$$

where $\alpha^j = \sum_{i=1}^j \alpha_i$ for $j = 1, \dots, N-1$. The Newton basis for $W_{k,N}^{u,\mathbf{X}_n}$, the space of homogeneous vector-polynomials of degree k, consists of N such monomials (one for each basis vector of $\mathbb{R}^N$) for each α with $|\alpha| = k$.

Dynamically essential coordinates. After a Newton basis is chosen, one needs to make sure that it is effective for dynamical purposes. Earlier, we noticed in (3.30) and (3.33) that in order to perturb by Newton Interpolation Polynomials in an effective way, we need to make sure that the product of distances $\prod_{j=0}^{n-2} |x_{n-1} - x_j|$ is not too small. Similarly, in the multidimensional case we need at least one Newton monomial $(x; x_0, \dots, x_{n-2})^\alpha$ with $|\alpha| = n-1$ not to be too small. The most natural way to choose a "good" monomial is by taking the maximal coordinates of corresponding vectors. Let $v \in \mathbb{R}^N$ be a nonzero vector and $\|v\| = (\sum_{i=1}^N v_i^2)^{1/2}$. Denote by

$$m(v) = \min\{i : \ 1 \le i \le N, \ |v_i| = \max_{j=1}^N |v_j|\}$$

the minimal index of one of the largest components v_i of v. Then

$$\left|\prod_{j=0}^{n-2} (x_{n-1} - x_j)_{m(x_{n-1}-x_j)}\right| \ge N^{\frac{1-n}{2}} \prod_{j=0}^{n-2} |x_{n-1} - x_j|. \tag{4.6}$$

This is a satisfactory estimate, because $\gamma_n(C,\delta)$ is a stretched exponential in n, so we can neglect factors that are exponential in n. For given $\mathbf{X}_n = \{x_j\}_{j=0}^{n-1}$, we call the monomials

$$\prod_{j=0}^{k-1}(x - x_j)_{m(x_k - x_j)}, \qquad k = 0, \dots, n-1 \tag{4.7}$$

dynamically essential. Using these monomials, we shall imitate estimates (3.27)–(3.34) in the multidimensional Discretization Method.

Consider the $(nN + N^2)$-parameter family of dynamically essential perturbations

$$\begin{aligned} f_U(x) = {} & f(x) + \vec{u}_0 + \vec{u}_1 (x - x_0)_{m(x_1 - x_0)} \\ & + \vec{u}_2 \prod_{j=0}^{1}(x - x_j)_{m(x_2 - x_j)} + \cdots + \vec{u}_{n-1} \prod_{j=0}^{n-2}(x - x_j)_{m(x_{n-1} - x_j)} \\ & + U_{2n-1}(x - x_{n-1}) \prod_{j=1}^{n-2} \left((x - x_j)_{m(x_{n-1} - x_j)}\right)^2, \end{aligned} \tag{4.8}$$

where $U = (\vec{u}_0, \dots, \vec{u}_{n-1}, U_{2n-1})$ consists of n vectors in $\mathbb{R}^N$ and an $N \times N$ matrix. This family is a natural candidate for the multidimensional Newton family in the sense that $\vec{u}_k$ moves the image of x_k (respectively U_{2n-1} moves the derivative at x_{n-1}) at the maximal speed.

However, the dynamically essential monomials do not necessarily belong to the standard Newton basis (4.5), so we will need to define the basis differently depending on which monomials are dynamically essential for the given pseudo-trajectory $\mathbf{X}_n$. This is not a major obstacle, since we already use a different basis for each $\mathbf{X}_n$ in the Discretization Method, but it does further complicate the argument.

The necessity of altering the basis is illustrated by the following example for $N = 2$: $x_0 = (1, 0),\ x_1 = (0, 1),\ x_2 = (1, 1)$. Then for all α with $|\alpha| = 2$, we have $(x_2; x_0, x_1)^\alpha = 0$. Thus, the monomial $(x; x_0, x_1)^\alpha$ is useless to perturb the image of x_2 (see also Section 10 in [GK]).

Dynamically essential Newton basis. In view of the example above, monomials from (4.7) do not always belong to the standard Newton basis (4.5). Here we show how to modify the construction of the Newton basis so that it contains dynamically essential monomials.

Recall that $\mathbf{X}_k = (x_0, \dots, x_{k-1})$ for $k = 1, \dots, n$. For $k \le n - 1$ and $j = 1, \dots, N$, using the functional $m(v)$ given above, we let

$$s_{j,\mathbf{X}_{k+1}} = \{0 \le i \le k - 1 :\ m(x_k - x_i) = j\} \tag{4.9}$$

be ordered by increasing values of i. Define a permutation $\sigma(\mathbf{X}_{k+1})$ on k elements

$$\sigma(\mathbf{X}_{k+1}) : \big(0, 1, \dots, k-1\big) \to \big(s_{1,\mathbf{X}_{k+1}}, s_{2,\mathbf{X}_{k+1}}, \dots, s_{N,\mathbf{X}_{k+1}}\big). \tag{4.10}$$

Now we define a *dynamical Newton basis* in the space $W_{<2n,N}$ as follows. For $\alpha \in \mathbb{Z}_+$ with $|\alpha| < 2n$, let $k = |\alpha| (\text{mod } n)$ and $\alpha^j = \sum_{i=1}^{j} \alpha_i$. Then

$$\begin{aligned} (x; x_0, \dots, x_{|\alpha|-1(\text{mod } n)})^{\alpha}_{\sigma(\mathbf{X}_{k+1})} = \prod_{i_1=0}^{\alpha_1-1} \Big(x - x_{\sigma(\mathbf{X}_{k+1})\left(i_1\right)}\Big)_1 \\ \prod_{i_2=0}^{\alpha_2-1} \Big(x - x_{\sigma(\mathbf{X}_{k+1})\left(\alpha^1+i_2\right)}\Big)_2 \cdots \prod_{i_N=0}^{\alpha_N-1} \Big(x - x_{\sigma(\mathbf{X}_{k+1})\left(\alpha^{N-1}+i_N\right)}\Big)_N . \end{aligned} \tag{4.11}$$

The space $W_{<2n,N}$ with this basis is denoted by $W^{u,\text{dyn}}_{<2n,N}(\mathbf{X}_n)$ and is called the *dynamical Newton space* with the *trajectory respecting coordinate system*. We call the Newton map

$$\mathcal{L}^{N,\text{dyn}}_{\mathbf{X}_n} : W_{<2n,N} \to W^{u,\text{dyn}}_{<2n,N}(\mathbf{X}_n) \tag{4.12}$$

implicitly defined by

$$\begin{aligned} \sum_{|\alpha|<2n} \vec{\varepsilon}_\alpha x^\alpha = \sum_{|\alpha|<2n} \vec{u}^{\,\text{dyn}}_\alpha (x; x_0, \dots, x_{|\alpha|-1(\text{mod } n)})^{\alpha}_{\sigma(\mathbf{X}_{k+1})}, \\ \mathcal{L}^{N,\text{dyn}}_{\mathbf{X}_n}(\{\vec{\varepsilon}_\alpha\}_{|\alpha|<2n}) = \{\vec{u}^{\,\text{dyn}}_\alpha\}_{|\alpha|<2n} \end{aligned} \tag{4.13}$$

the *dynamical Newton map*.

Notice that if α_j is the number of elements of $s_{j,\mathbf{X}_{k+1}}$ for each $j = 1, \dots, N$, then the monomial (4.11) is dynamically essential. This multiindex, which we denote by $\alpha(\mathbf{X}_{k+1})$, is the unique multiindex such that

$$\prod_{j=0}^{k-1} (x - x_j)_{m(x_k - x_j)} = (x; x_0, \dots, x_{k-1})^{\alpha(\mathbf{X}_k)}_{\sigma(\mathbf{X}_k)}. \tag{4.14}$$

The last vector-monomial $(x - x_{n-1}) \prod_{j=0}^{n-1} (x - x_j)^2_{m(x_{n-1}-x_j)}$ in (4.8) has N multiindices corresponding to it. Denote by $\alpha_1(\mathbf{X}_n), \alpha_2(\mathbf{X}_n), \dots, \alpha_N(\mathbf{X}_n)$ the multiindices corresponding to $(x - x_{n-1})_1 \prod_{j=0}^{n-1} (x - x_j)^2_{m(x_{n-1}-x_j)}, \dots, (x - x_{n-1})_N \prod_{j=0}^{n-1} (x - x_j)^2_{m(x_{n-1}-x_j)}$, respectively.

Definition 4.2.1. The set of multiindices $\alpha(\mathbf{X}_1), \dots, \alpha(\mathbf{X}_n), \{\alpha_i(\mathbf{X}_n)\}_{i=1}^N$ defined above is called the dynamically essential set. The $(nN + N^2)$-dimensional plane generated by the corresponding monomials in $W^{\text{dyn},u}_{<2n}(\mathbf{X}_n)$ is called the dynamically essential plane and its orthogonal complement is called dynamically nonessential. These planes are denoted by $D^{\text{dyn,ess}}_{\mathbf{X}_n}$ and by $D^{\text{dyn,non}}_{\mathbf{X}_n}$, respectively.

We have

$$W^{u,\text{dyn}}_{<2n,N}(\mathbf{X}_n) = D^{\text{dyn,ess}}_{\mathbf{X}_n} \bigoplus D^{\text{dyn,non}}_{\mathbf{X}_n}. \tag{4.15}$$

Denote by

$$\begin{aligned} \pi^{\text{dyn,ess}}_{\mathbf{X}_n} &: W^{u,\text{dyn}}_{<2n,N}(\mathbf{X}_n) \to D^{\text{dyn,ess}}_{\mathbf{X}_n} \\ \pi^{\text{dyn,non}}_{\mathbf{X}_n} &: W^{u,\text{dyn}}_{<2n,N}(\mathbf{X}_n) \to D^{\text{dyn,non}}_{\mathbf{X}_n} \end{aligned} \tag{4.16}$$

the natural projections along the complement.

4.3. *The multidimensional Distortion Lemma.* One of the key ingredients of the Discretization Method is the Distortion Lemma from Section 3.4. Fix an n-tuple of points $\mathbf{X}_n = \{x_j\}_{j=0}^{n-1} \subset B^N$. The Distortion Lemma in dimension 1 shows that the Newton map $\mathcal{L}^1_{\mathbf{X}_n} : W_{<2n,1} \to W^{u,\mathbf{X}_n}_{<2n,1}$ expands a brick $\text{HB}^{\text{st}}_{<2n}(\tau)$ of standard thickness along each direction by at most a factor of 3 in each direction. Since, the space of 1-dimensional divided differences $W_{<2n,1}$ is $2n$-dimensional, this gives a total volume distortion of at most 3^{2n}. This factor is part of the estimates (3.13) and (3.42), but is ultimately unimportant since it is exponential in n.

In the multidimensional case, this naive approach does not work. One can, indeed, show that the Newton map expands a brick of standard thickness along each direction by at most a factor 3^N. However, the space of divided differences $W_{<2n,N}$ has dimension of order $(2n)^N$. So, the naive estimate of distortion is $3^{N(2n)^N}$, which is highly unaffordable. In the multidimensional case, we need a more precise estimate on distortion.

First, we define a Cubic Brick of at most standard thickness, which differs from the brick $\text{HB}^{\text{st}}_{<2n}(\vec{\mathbf{r}})$ defined by (2.27) in two ways. First, it is a parallelepiped for all N, whereas $\text{HB}^{\text{st}}_{<2n}(\vec{\mathbf{r}})$ is a product of balls whose dimension is greater than 1 when $N > 1$. Second, we require that the side lengths of a Cubic Brick decay very rapidly. We will estimate distortion of such Cubic Bricks by the Newton map, then in the next section we will cover a brick with Cubic Bricks.

Definition 4.3.1. Let $\vec{\lambda}_{<k} = (\lambda_0, \dots, \lambda_{k-1}) \in \mathbb{R}^k_+$ be a vector with strictly positive components. If for every $0 \le m < k$ we have $\lambda_m \ge \lambda_{m+1}(m+1)^{4N}$, we call the rectangular parallelepiped

$$\begin{aligned} \text{CHB}^{N,\text{st}}_{<k}(\vec{\lambda}_{<k}, \vec{\delta}\,) = \big\{ \{\vec{\varepsilon}_\alpha\}_{|\alpha|<k} :\ & \forall\, 0 \le m < k, \\ & \forall\, \alpha \in \mathbb{Z}^N_+,\ |\alpha| = m,\ \forall\, j = 1, \dots, N,\ |\varepsilon^j_\alpha - \delta^j_\alpha| \le \lambda_m \big\} \end{aligned} \tag{4.17}$$

a *Cubic Brick* of at most standard thickness centered at $\vec{\delta}$ (here the superscript j denotes the j-th coordinate). Similarly, for $\lambda_k \in \mathbb{R}_+$, let

$$\begin{aligned} \text{CHB}^{N,\text{st}}_{k}(\vec{\lambda}_{k}, \vec{\delta}\,) = \big\{ \{\vec{\varepsilon}_\alpha\}_{|\alpha|=k} :\ & \forall\, \alpha \in \mathbb{Z}^N_+, \\ & \forall\, |\alpha| = k,\ \forall\, j = 1, \dots, N,\ |\varepsilon^j_\alpha - \delta^j_\alpha| \le \lambda_k \big\}. \end{aligned} \tag{4.18}$$

An example of a suitable thickness vector $\vec\lambda_{<k}$ is $\{\lambda_m = \frac{\tau}{(m!)^{4N}}\}_{m=0}^{k-1}$.

To formulate a multidimensional version of the Distortion Lemma, we extend the definition of the parameters (3.38) and (3.39) allowed by the family (3.8) in Section 3.4. Consider a $C^{1+\rho}$-smooth diffeomorphism $f \in \mathrm{Diff}^{1+\rho}(B^N)$ of the unit ball B^N into its interior and some positive integer n. Let $\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec\lambda_{<2n}, \vec\delta\,)$ be a Cubic Brick of at most standard thickness, which defines the family of diffeomorphisms of the unit ball B^N into its interior

$$\left\{ f_{\vec\varepsilon}(x) = f(x) + \sum_{|\alpha|\le 2n-1} \vec\varepsilon_\alpha x^\alpha \right\}_{\{\vec\varepsilon_\alpha\}_{|\alpha|\le 2n-1} \in \mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec\lambda_{<2n},\vec\delta\,)}. \tag{4.19}$$

Fix an n-tuple of points $\mathbf{X}_n = \{x_j\}_{j=0}^{n-1} \subset B^N$, and fix the standard Newton basis $\big\{(x; x_0, \dots, x_{(|\alpha|-1) \pmod n})^\alpha\big\}_{|\alpha|\le 2n-1}$ in the multidimensional space of divided differences $\mathcal{DD}^{N,n}(B^N, \mathbb{R}^N)$, defined by (4.5). Denote by

$$\mathcal{L}^N_{\mathbf{X}_n} : \big(\{\vec\varepsilon_\alpha\}_{|\alpha|\le 2n-1}\big) \to \big(\{\vec u_\alpha\}_{|\alpha|\le 2n-1}\big) \tag{4.20}$$

the Newton map that corresponds to rewriting an N-vector polynomial of degree $2n-1$ given in the basis $\{x^\alpha\}_{|\alpha|\le 2n-1}$ in the Newton basis (4.5).

Consider the reparametrization of family (4.19) by the Newton parameters

$$f_{\vec u(2n-1),\mathbf{X}_n}(x) = f(x) + \sum_{|\alpha|\le 2n-1} \vec u_\alpha (x; x_0, \dots, x_{(|\alpha|-1) \pmod n})^\alpha. \tag{4.21}$$

Denote by $\vec u(m) = (\{\vec u_\alpha\}_{|\alpha|\le m})$ (respectively $\vec u_m = (\{\vec u_\alpha\}_{|\alpha|=m})$) the set of all Newton parameters of degree at most m (respectively of degree exactly m).

In order to state the Discretization Lemma in the N-dimensional case, we have to define the set of values of Newton parameters $\vec u(2n-1)$ that are allowed by the family (4.19) and the Cubic Brick (4.17).

Let $\mathbf{X}_k = \{x_j\}_{j=0}^{k-1}$ for $0 \le k \le n$, and for convenience below let $\mathbf{X}_k = \mathbf{X}_n$ for $n < k \le 2n-1$. Let

$$\pi^{N,u,\mathbf{X}_k}_{<2n,\le k} : W^u_{<2n,N} \to W^{u,\mathbf{X}_k}_{\le k,N} \quad \text{and} \quad \pi^{N,u,\mathbf{X}_k}_{<2n,k} : W^u_{<2n,N} \to W^{u,\mathbf{X}_k}_{k,N}$$

be the natural projection onto the space $W^{u,\mathbf{X}_k}_{\le k,N}$ of N-vector polynomials in N variables of degree k and the space $W^{u,\mathbf{X}_k}_{k,N}$ of homogeneous N-vector polynomials in N variables of degree k. As for $N = 1$, this projection depends only on $\mathbf{X}_k$, not on $\mathbf{X}_n \setminus \mathbf{X}_k$.

Notice that the image of the Cubic Brick $\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec\lambda_{<2n}, \vec\delta\,)$ under the Newton map $\mathcal{L}^{N,\mathrm{dyn}}_{\mathbf{X}_n}$ is a parallelepiped

$$\mathcal{P}^{N,\mathrm{dyn}}_{<2n,\mathbf{X}_n}(\vec\lambda_{<2n}, \vec\delta\,) = \mathcal{L}^{N,\mathrm{dyn}}_{\mathbf{X}_n}\big(\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec\lambda_{<2n}, \vec\delta\,)\big), \tag{4.22}$$

because the map $\mathcal{L}^{N,\mathrm{dyn}}_{\mathbf{X}_n}$ is linear.

We call the parallelepiped $\mathcal{P}^{N,\mathrm{dyn}}_{<2n,\mathbf{X}_n}(\vec{\lambda}_{<2n},\vec{\delta}\,)$ the set of parameters *allowed by the family* (4.19). Notice that the values of $\vec{u}^{\,\mathrm{dyn}}(2n-1)=(\vec{u}_0^{\,\mathrm{dyn}},\dots,\vec{u}_{2n-1}^{\,\mathrm{dyn}})$ $\in W^{u,\mathrm{dyn}}_{<2n,N}(\mathbf{X}_n)$ that do not belong to $\mathcal{P}^{N,\mathrm{dyn}}_{<2n,\mathbf{X}_n}(\vec{\lambda}_{<2n},\vec{\delta}\,)$ are of *no interest* for us, because they lie outside of the range of the family (4.19) under consideration. Now we define the range of allowed parameters $\vec{u}^{\,\mathrm{dyn}}(2n-1)\in W^{u,\mathrm{dyn}}_{<2n,N}(\mathbf{X}_n)$.

Denote the images of the Cubic Brick $\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n},\vec{\delta}\,)$ under composition of the Newton map $\mathcal{L}^{N,\mathrm{dyn}}_{\mathbf{X}_n}$ and the projections $\pi^{\mathrm{dyn,ess}}_{\mathbf{X}_n}$ and $\pi^{\mathrm{dyn,non}}_{\mathbf{X}_n}$ by

(4.23)
$$\begin{aligned}\mathcal{P}^{N,\mathrm{dyn,ess}}_{<2n,\mathbf{X}_n}(\vec{\lambda}_{<2n},\vec{\delta}\,)&=\pi^{\mathrm{dyn,ess}}_{\mathbf{X}_n}\circ\mathcal{L}^{N,\mathrm{dyn}}_{\mathbf{X}_n}(\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n},\vec{\delta}\,))\subset D^{\mathrm{dyn,ess}}_{\mathbf{X}_n}\\ \mathcal{P}^{N,\mathrm{dyn,non}}_{<2n,\mathbf{X}_n}(\vec{\lambda}_{<2n},\vec{\delta}\,)&=\pi^{\mathrm{dyn,non}}_{\mathbf{X}_n}\circ\mathcal{L}^{N,\mathrm{dyn}}_{\mathbf{X}_n}(\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n},\vec{\delta}\,))\subset D^{\mathrm{dyn,non}}_{\mathbf{X}_n}.\end{aligned}$$

It follows from the definition of the Newton map $\mathcal{L}^{N,\mathrm{dyn}}_{\mathbf{X}_n}$ that the sets $\mathcal{P}^{N,\mathrm{dyn,ess}}_{<2n,\mathbf{X}_n}(\vec{\lambda}_{<2n},\vec{\delta}\,)$ and $\mathcal{P}^{N,\mathrm{dyn,non}}_{<2n,\mathbf{X}_n}(\vec{\lambda}_{<2n},\vec{\delta}\,)$ are polyhedra depending on $\mathbf{X}_n$. We call $\mathcal{P}^{N,\mathrm{dyn,ess}}_{<2n,\mathbf{X}_n}(\vec{\lambda}_{<2n},\vec{\delta}\,)$ (respectively $\mathcal{P}^{N,\mathrm{dyn,non}}_{<2n,\mathbf{X}_n}(\vec{\lambda}_{<2n},\vec{\delta}\,)$) the set of dynamically essential (respectively nonessential) parameters *allowed by the family* (4.19).

Similarly to the 1-dimensional Distortion Lemma from Section 3.4, its N-dimensional generalization gives an estimate on the ratio of volumes of the polyhedra $\mathcal{P}^{N,\mathrm{dyn,ess}}_{<2n,\mathbf{X}_n}(\vec{\lambda}_{<2n},\vec{\delta}\,)$ and $\mathcal{P}^{N,\mathrm{dyn,non}}_{<2n,\mathbf{X}_n}(\vec{\lambda}_{<2n},\vec{\delta}\,)$ of allowed Newton parameters $\vec{u}(k)$ and the corresponding Cubic Bricks $CHB^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n},\vec{\delta}\,)$. Let $\mathrm{Leb}_{<2n,N}$, $\mathrm{Leb}^{\mathrm{dyn,ess}}_{<2n,N}$, and $\mathrm{Leb}^{\mathrm{dyn,non}}_{<2n,N}$) be the Lebesgue measures on the spaces $W_{<2n,N}$, $D^{\mathrm{dyn,ess}}_{\mathbf{X}_n}$, and $D^{\mathrm{dyn,non}}_{\mathbf{X}_n}$, respectively.

THE N-DIMENSIONAL DISTORTION LEMMA. *Let $\mathbf{X}_n=\{x_j\}_{j=0}^{n-1}\subset B^N$ be an n-tuple of points in the unit ball B^N, and let $\mathcal{L}^{N,\mathrm{dyn}}_{\mathbf{X}_n}:W_{<2n,N}\to W^{u,\mathrm{dyn}}_{<2n,N}(\mathbf{X}_n)$ be the dynamical Newton map* (4.12). *Let $\vec{\lambda}_{<2n}=(\lambda_0,\dots,\lambda_{2n-1})\in\mathbb{R}^{2n}_+$ be a vector with strictly positive components that defines a Cubic Brick $\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n},\vec{\delta}\,)$ of at most standard thickness. Then we have the volume ratio estimate*

(4.24)
$$\frac{\mathrm{Leb}^{\mathrm{dyn,ess}}_{<2n,N}\left(\mathcal{P}^{N,\mathrm{dyn,ess}}_{<2n,\mathbf{X}_n}(\vec{\lambda}_{<2n},\vec{\delta}\,)\right)\ \mathrm{Leb}^{\mathrm{dyn,non}}_{<2n,N}\left(\mathcal{P}^{N,\mathrm{dyn,non}}_{<2n,\mathbf{X}_n}(\vec{\lambda}_{<2n},\vec{\delta}\,)\right)}{\mathrm{Leb}_{<2n,N}\left(\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n},\vec{\delta}\,)\right)}\le 3^{2N^2n}.$$

Remark 4.3.2. As in the 1-dimensional case, the Newton map $\mathcal{L}^N_{\mathbf{X}_n}$ is *volume-preserving*, and hence the parallelepipeds

$$\mathcal{P}^{N,\mathrm{st}}_{<2n,\mathbf{X}_n}(\vec{\lambda}_{<2n},\vec{\delta}\,)\ \text{and}\ \mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n},\vec{\delta}\,)$$

have the same volume. However, the estimate (4.24) concerns the volumes of projections of these parallelepipeds. Since $\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n},\vec{\delta}\,)$ is a rectangular parallelepiped (aligned with the coordinate axes) and $\mathcal{P}^{N,\mathrm{st}}_{<2n,\mathbf{X}_n}(\vec{\lambda}_{<2n},\vec{\delta})$ is not, the projections of the latter set have larger volume according to the amount of shear in the Newton map. The Distortion Lemma bounds the effect of the shear.

The lemma above will be proved in Part II of this paper in a slightly different form required for the proof of the Main Result (Theorem 1.3.9). Precise formulation requires the introduction of product measures of type (1.15) on $W_{<2n,N}$ and $W^{u,\mathrm{dyn}}_{<2n,N}(\mathbf{X}_n)$ and the induced measures on the subspaces of dynamically essential and nonessential monomials (see also Section 10.1.2 in [GK]). The other main ingredient of Section 3.4, the Collection Lemma, proceeds in much the same way for general N as it does for $N = 1$. (See also a similar 2-dimensional Collection Lemma in Section 11.6 of [GK].)

4.4. *From a brick of at most standard thickness to an admissible brick.* This section is very similar to Section 11.2 in [GK]. In order to apply the method developed in this paper to other problems about generic properties of dynamical systems, we need to have a sufficiently rich space of parameters (enough freedom to perturb). Various dynamical phenomena have a "size" that is exponential in the period or number of iterations. Since we perturb trajectories of length n with polynomials whose degree is proportional to n, it seems essential to have a Hilbert Brick of parameters with sides decaying at most exponentially in the period of the polynomials that the respective parameters multiply. That is, with the notation of Definition 1.3.1, we would like to have $r_n \geq \exp(-Cn)$ for some $C > 0$. However, even for the 1-dimensional model, if we consider a brick of parameters with exponentially decaying sides, then we cannot control the distortion properties of the Newton map (see Remark 3.4.1).

In order to circumvent this problem, we do the following. Consider a Hilbert Brick $\mathrm{HB}^N(\vec{\mathbf{r}})$ of an admissible size $\vec{\mathbf{r}} = \{r_k\}_{k=0}^{\infty}$ (see Definition 1.3.1). Using Fubini's reduction from Section 2.3 at the n-th stage of induction over the period, we reduce consideration of an infinite-dimensional Hilbert Brick $\mathrm{HB}^N(\vec{\mathbf{r}})$ to a finite-dimensional brick $\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}})$.

Recall that $\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}})$ belongs to the space of N-vector polynomials of degree $2n-1$, denoted by $W_{<2n,N}$.

Definition 4.4.1. We call the *Cubic Brick in* $W_{<2n,N}$ of at most standard thickness centered at a point $\vec{\delta}_{<2n} \in W_{<2n,N}$ and associated to the brick $\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}})$ the one denoted by $\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n}(\vec{\mathbf{r}}),\vec{\delta}_{<2n})$ and defined by (4.17) where $\lambda_m(\vec{\mathbf{r}}) = r_m/(m!)^{4N}$ and $\vec{\lambda}_{<2n}(\vec{\mathbf{r}}) = (\lambda_0(\vec{\mathbf{r}}),\dots,\lambda_{2n-1}(\vec{\mathbf{r}}))$.

Remark 4.4.2. It follows from the nonincreasing property of the sequence r_m (see Definition 1.3.1) that

$$(m+1)^{4N}\lambda_{m+1}(\vec{\mathbf{r}}) = (m+1)^{4N}\frac{r_{m+1}}{((m+1)!)^{4N}} \leq \frac{r_m}{(m!)^{4N}} = \lambda_m(\vec{\mathbf{r}}), \tag{4.25}$$

so that the Cubic Brick $\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n}(\vec{\mathbf{r}}), \vec{\delta}_{<2n})$ is of at most standard thickness.

We wish to cover the brick $\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}})$ in $W_{<2n,N}$ by a collection of Cubic Bricks $\{\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n}(\vec{\mathbf{r}}), \vec{\delta}_{<2n})\}_{\vec{\delta}_{<2n}\in G(\vec{\mathbf{r}})}$ of at most standard thickness associated with it. First we define a grid in $W_{<2n,N}$ on which the centers $\vec{\delta}_{<2n}$ should lie in order that the Cubic Bricks fit together without overlap.

Recall that $W_{<2n,N} = \times_{m=0}^{2n-1} W_{m,N}$, $\dim W_{m,N} = \nu(m,N)$, and denote $\eta(k,N) = \sum_{m=0}^{k} \nu(m,N)$. Let

$$\begin{aligned}\mathbb{Z}^{\eta(2n-1,N)}_{2\vec{\lambda}_{<2n}(\vec{\mathbf{r}})} &= \mathbb{Z}^{\nu(0,N)}_{2\lambda_0(\vec{\mathbf{r}})} \times \mathbb{Z}^{\nu(1,N)}_{2\lambda_1(\vec{\mathbf{r}})} \times \cdots \times \mathbb{Z}^{\nu(2n-1,N)}_{2\lambda_{2n-1}(\vec{\mathbf{r}})} \\ &\subset W_{0,N} \times W_{1,N} \times \cdots \times W_{2n-1,N} = W_{<2n,N},\end{aligned} \tag{4.26}$$

where $\mathbb{Z}^{\nu(m,N)}_{2\lambda_m(\vec{\mathbf{r}})}$ is the grid in $\mathbb{R}^{\nu(m,N)}$ with spacing $2\lambda_m(\vec{\mathbf{r}})$ in each coordinate. Let

$$G(\vec{\mathbf{r}}) = \left\{ \vec{\delta}_{<2n} \in \mathbb{Z}^{\eta(2n-1,N)}_{2\vec{\lambda}_{<2n}(\vec{\mathbf{r}})} : \mathrm{HB}^N_{<2n}(\vec{\mathbf{r}}) \cap \mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n}(\vec{\mathbf{r}}), \vec{\delta}_{<2n}) \neq \emptyset \right\}. \tag{4.27}$$

The Cubic Bricks with centers in $G(\vec{\mathbf{r}})$ are the ones needed for our covering:

$$\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}}) \subset \cup_{\vec{\delta}\in G(\vec{\mathbf{r}})} \mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n}(\vec{\mathbf{r}}), \vec{\delta}_{<2n}). \tag{4.28}$$

THE COVERING LEMMA. *The ratio of the volume of the covering* (4.28) *by Cubic Bricks and the volume of the brick* $\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}})$ *is bounded by* e^{6N}.

Proof. We claim that the covering (4.28) is contained in the slightly larger brick $\mathrm{HB}^N_{<2n}(\{(1+\frac{2\sqrt{N}}{(m!)^{3N}})r_m\}_{m=0}^{2n-1})$. The amount that the covering extends beyond $\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}})$ in the direction of a given multi-index α with $|\alpha| = m$ is at most the diameter in the α direction of one of the Cubic Bricks $\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n}(\vec{\mathbf{r}}), \vec{\delta}^{\,j}_{<2n})$, which by Definition 4.4.1 is

$$2\sqrt{N}\,\lambda_m(\vec{\mathbf{r}}) = 2\sqrt{N}\,r_m/(m!)^{4N}.$$

Recall the definitions (2.27) of $\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}})$ and (1.12) of the norm used therein. In this norm, the diameter of $\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n}(\vec{\mathbf{r}}), \vec{\delta}^{\,j}_{<2n})$ in the directions with $|\alpha| = m$ is $2\sqrt{N} r_m/(m!)^{4N}$ times the square root of $\sum_{|\alpha|=m} \binom{m}{\alpha}^{-1} < (m!)^N$, from which our claim follows. The ratio of volumes that we wish to bound is then at most $\prod_{m=0}^{2n-1} \exp(2\sqrt{N}/(m!)^{3N}) \leq e^{2e\sqrt{N}} < e^{6N}$. Q.E.D.

Recall now that our main goal in the proof of Theorem 1.3.9 is to get an estimate on the measure of the "bad" set of parameters (2.1) inside of the Hilbert Brick $\mathrm{HB}^N(\vec{\mathbf{r}})$. Using the Fubini reduction from Section 2.3, we know that it is sufficient to get an estimate on the measure of the "bad" set in a finite-dimensional slice of the form $\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}}) \times \{\vec{\varepsilon}_{\geq 2n}\}$ that is uniform over

$\vec{\varepsilon}_{\geq 2n} \in \mathrm{HB}^N_{\geq 2n}(\vec{\mathbf{r}})$ see (2.31)–(2.32). Notice now that if we can prove that for each Cubic Brick slice $\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n}(\vec{\mathbf{r}}), \vec{\delta}_{<2n}) \times \{\vec{\varepsilon}_{\geq 2n}\}$ from the covering collection $\vec{\delta}_{<2n} \in G(\vec{\mathbf{r}})$ that the fraction of parameters in the slice that are "bad" is uniformly small over $\vec{\delta}_{<2n} \in G(\vec{\mathbf{r}})$, then the same fraction is small in the whole slice $\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}}) \times \{\vec{\varepsilon}_{\geq 2n}\}$. (By the Covering Lemma, we must increase the bound on the fraction of "bad" parameters only by the factor e^{6N}, which is independent of n.) By the Fubini reduction, this shows that the measure of the "bad" set in $\mathrm{HB}^N(\vec{\mathbf{r}})$ is small too. Thus it is sufficient to prove the following estimate.

(4.29)
$$\frac{\mu^N_{<2n,\vec{\mathbf{r}}}\left(B_n(C,\delta,\rho,\vec{\mathbf{r}},f,\vec{\varepsilon}_{\geq 2n}) \cap \mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n}(\vec{\mathbf{r}}), \vec{\delta}_{<2n})\right)}{\mu^N_{<2n,\vec{\mathbf{r}}}\left(\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n}(\vec{\mathbf{r}}), \vec{\delta}_{<2n})\right)} \leq \mu'_n(C,\delta,\rho,M_{1+\rho})$$

uniformly over all $\delta_{<2n} \in J(\vec{\mathbf{r}})$ and so $\sum_{n=1}^{\infty} \mu'_n(C,\delta,\rho,M_{1+\rho})$ converges for all positive C, δ, ρ, and $M_{1+\rho}$, and tends to zero as C tends to infinity.

4.5. *The main estimate on the measure of "bad" parameters.* In this subsection we formulate the main theorem of the rest of the paper which implies (the Main) Theorem 1.3.9. It will be proved in Part II of this paper.

THEOREM 4.5.1. *For any $\rho > 0$ and any diffeomorphism $f \in \mathrm{Diff}^{1+\rho}(B^N)$, consider a Hilbert Brick $\mathrm{HB}^N(\vec{\mathbf{r}})$ of an admissible size $\vec{\mathbf{r}}$ with respect to f and the family of analytic perturbations of f*

$$\{f_{\vec{\varepsilon}}(x) = f(x) + \phi_{\vec{\varepsilon}}(x)\}_{\vec{\varepsilon} \in \mathrm{HB}^N(\vec{\mathbf{r}})} \tag{4.30}$$

(see (1.10)) with the Lebesgue product probability measure $\mu^N_{\vec{\mathbf{r}}}$ (see (1.15)) associated to $\mathrm{HB}^N(\vec{\mathbf{r}})$.

Then for any positive integer n and any $\vec{\varepsilon}_{\geq 2n} \in \mathrm{HB}^N_{\geq 2n}(\vec{\mathbf{r}})$, consider a slice (2.27) of the Hilbert Brick $\mathrm{HB}^N(\vec{\mathbf{r}})$ of the form

$$\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}}) \times \{\vec{\varepsilon}_{\geq 2n}\} \subset \mathrm{HB}^N_{<2n}(\vec{\mathbf{r}}) \times \mathrm{HB}^N_{\geq 2n}(\vec{\mathbf{r}}) = \mathrm{HB}^N(\vec{\mathbf{r}}). \tag{4.31}$$

Inside of this slice fix a grid point $\vec{\delta}_{<2n} \in G(\vec{\mathbf{r}})$, and consider the Cubic Brick

$$\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n}(\vec{\mathbf{r}}), \vec{\delta}_{<2n}) \times \{\vec{\varepsilon}_{\geq 2n}\} \ \subset \mathrm{HB}^N_{<2n}(\vec{\mathbf{r}}) \times \{\vec{\varepsilon}_{\geq 2n}\}. \tag{4.32}$$

from the covering (4.28) associated to the brick $\mathrm{HB}^N_{<2n}(\vec{\mathbf{r}})$.

Let $\tilde{f} = f_{(\vec{\delta}_{<2n},\vec{\varepsilon}_{\geq 2n})}$ be a diffeomorphism corresponding to the center of the Cubic Brick $\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n}(\vec{\mathbf{r}}), \vec{\delta}_{<2n}) \times \{\vec{\varepsilon}_{\geq 2n}\}$. Consider the family of diffeomorphisms

$$\begin{aligned}\{\tilde{f}_{\vec{\varepsilon}_{<2n}}\}&_{\vec{\varepsilon}_{<2n} \in \mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n}(\vec{\mathbf{r}}),0)} \\ &\supset \{f_{(\vec{\varepsilon}_{<2n}+\vec{\delta}_{<2n},\vec{\varepsilon}_{\geq 2n})}\}_{(\vec{\varepsilon}_{<2n}+\vec{\delta}_{<2n} \in \mathrm{HB}^N_{<2n}(\vec{\mathbf{r}}) \cap \mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{<2n}(\vec{\mathbf{r}}),\vec{\delta}_{<2n})}.\end{aligned} \tag{4.33}$$

Then for $C \geq 100 \log M_{1+\rho}/(\rho\delta)$, *the fraction of the measure of "bad" parameters* $B_n(C, \delta, \rho, \mathbf{\vec{r}}, f, \vec{\varepsilon}_{\geq 2n})$, *defined in* (2.32), *inside* $\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{2n-1}(\mathbf{\vec{r}}), \vec{\delta}_{<2n}) \times \{\vec{\varepsilon}_{\geq 2n}\}$ *satisfies the bound*

$$\frac{\mu^N_{<2n,\mathbf{\vec{r}}}\left(B_n(C, \delta, \rho, \mathbf{\vec{r}}, f, \vec{\varepsilon}_{\geq 2n}) \cap \mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{2n-1}(\mathbf{\vec{r}}), \vec{\delta}_{<2n})\right)}{\mu^N_{<2n,\mathbf{\vec{r}}}\left(\mathrm{CHB}^{N,\mathrm{st}}_{<2n}(\vec{\lambda}_{2n-1}(\mathbf{\vec{r}}), \vec{\delta}_{<2n})\right)} \leq 512 e^{6N} n^{1+\delta} 6^{2nN^2} M_{1+\rho}^{10Nn+1} \exp(-Cn \log_2 n/200). \tag{4.34}$$

The discussion from the end of the previous section shows that this theorem implies (the Main) Theorem 1.3.9.

Appendix A: Properties of hyperbolicity

In this paper we have quantified the hyperbolicity of periodic points in order to bound from below the distance between periodic points of the same period. Recall the definitions of $\gamma(L)$, $\gamma_n(x, f)$, and $\gamma_n(f)$ from (1.6) and (1.7) and the text between them. In this appendix we will prove Proposition 1.1.6, along with a result that says that the hyperbolicity of a power of a linear operator is no smaller than the hyperbolicity of the operator.

Proposition 1.1.6 follows immediately from the next lemma.

LEMMA A.1. *Given the hypotheses of Proposition* 1.1.6, *for every pair of distinct period n points of f, say $x = f^n(x) \neq y = f^n(y)$, the distance $|x - y|$ between them is at least* $(M_{1+\rho}^{-n(1+\rho)} \gamma_n(f))^{1/\rho}$.

Proof. Let $v = (y - x)/|x - y|$. Then

$$v = \frac{f^n(y) - f^n(x)}{|x-y|} = \frac{1}{|x-y|} \int_0^{|x-y|} \frac{d}{d\lambda} f^n(x + \lambda v) d\lambda = \frac{1}{|x-y|} \int_0^{|x-y|} df^n(x + \lambda v) v d\lambda. \tag{A.1}$$

It follows that

$$\frac{1}{|x-y|} \int_0^{|x-y|} (df^n(x + \lambda v)v - v) d\lambda = 0. \tag{A.2}$$

Let $w = df^n(x)v - v$; by hypothesis $|w| \geq \gamma_n(f)$. Also,

$$|df^n(x + \lambda v)v - v - w| = |(df^n(x + \lambda v) - df^n(x))v| \leq \|df^n(x + \lambda v) - df^n(x)\|. \tag{A.3}$$

Now

$$
\begin{aligned}
\text{(A.4)}\quad & df^n(x+\lambda v) - df^n(x) \\
&= [df(f^{n-1}(x+\lambda v)) - df(f^{n-1}(x))]df^{n-1}(x+\lambda v) \\
&\quad + df(f^{n-1}(x))[df(f^{n-2}(x+\lambda v)) - df(f^{n-2}(x))]df^{n-2}(x+\lambda v) \\
&\quad + \cdots + df^{n-1}(f(x))[df(x+\lambda v) - df(x)].
\end{aligned}
$$

Since $M_{1+\rho}$ is an upper bound on the $C^{1+\rho}$ norm of f, it is an upper bound on the norm of $df(z)$ for all $z \in M$. It follows that $|f^k(x+\lambda v) - f^k(x)| \le M_{1+\rho}^k \lambda$ for $k = 0, 1, \dots, n-1$, and hence

$$
\begin{aligned}
\text{(A.5)}\qquad \|df^n(x+\lambda v) - df^n(x)\| &\le \sum_{k=0}^{n-1} M_{1+\rho}(M_{1+\rho}^k \lambda)^\rho M_{1+\rho}^{n-1} \\
&= M_{1+\rho}^n \frac{M_{1+\rho}^{n\rho} - 1}{M_{1+\rho}^\rho - 1} \lambda^\rho < M_{1+\rho}^{n(1+\rho)} \lambda^\rho.
\end{aligned}
$$

(Recall that we assumed $M_{1+\rho} > 2^{1/\rho}$ in the definition of $M_{1+\rho}$.)

By the results above we then have

$$
\begin{aligned}
\text{(A.6)}\qquad 0 &= \frac{1}{|x-y|} \int_0^{|x-y|} (df^n(x+\lambda v)v - v) \cdot \frac{w}{|w|} d\lambda \\
&\ge \frac{1}{|x-y|} \int_0^{|x-y|} (|w| - \|df^n(x+\lambda v) - df^n(x)\|) d\lambda \\
&> \frac{1}{|x-y|} \int_0^{|x-y|} (\gamma_n(f) - M_{1+\rho}^{n(1+\rho)} \lambda^\rho) d\lambda \\
&> (\gamma_n(f) - M_{1+\rho}^{n(1+\rho)} |x-y|^\rho).
\end{aligned}
$$

From this we get the desired upper bound on $|x-y|$. This completes the proof of the lemma. Q.E.D.

Notice that the notion of hyperbolicity $\gamma(L)$ of a linear operator L as a lower bound on $|Lv - v|$ for unit vectors v occurs naturally in the proof above. It is not possible to make an analogous estimate with the same power on the period n hyperbolicity of f if the hyperbolicity is defined in the more usual manner, as in [Y], if we take the minimum distance of the eigenvalues of L from the unit circle in $\mathbb{C}$. To see this, consider the following C^2 map $f : \mathbb{R}^N \to \mathbb{R}^N$ for small $\gamma > 0$:

(A.7)

$$
\begin{aligned}
& f(x_1, x_2, \dots, x_N) \\
&\quad = ((1+\gamma)x_1 - x_2, (1+\gamma)x_2 - x_3, \dots, (1+\gamma)x_n - x_{n-1}, (1+\gamma)x_n - x_1^2).
\end{aligned}
$$

Notice that f has two nearby fixed points, 0 and $(\gamma^N, \gamma^{N+1}, \dots, \gamma^{2N-1})$, that are within roughly γ^N of each other. Notice also that $df(0)$ is upper triangular and hence all N of its eigenvalues are $1+\gamma$, so that by the eigenvalue notion of hyperbolicity, 0 is an (n,γ)-hyperbolic fixed point of f. (Though df has eigenvalues closer to the unit circle at the other fixed point, they are still much farther away than γ^N for large N.) On the other hand, for $v = (1, \gamma, \dots, \gamma^{N-1})$ we have $|v|$ slightly larger than 1 while

$$|Lv - v| = |(0, 0, \dots, \gamma^N)| = \gamma^N, \tag{A.8}$$

so that our notion of hyperbolicity is commensurate with the spacing between the fixed points.

To this point, when using the hyperbolicity of a linear operator L, it has only been important that $|Lv - v|$ not be small for unit vectors v. The reason for estimating from below $|Lv - \exp(2\pi i\phi)v|$ for $\phi \in [0, 1)$ in (1.6) is that the hyperbolicity of L will be a lower bound on the hyperbolicity of L^k for positive integers k. In terms of diffeomorphisms, this estimate gives a bound on how close points of period nk can lie to a hyperbolic point of period n. For the eigenvalue-based notion of hyperbolicity, the estimate is trivial, but for our notion it must be proved.

Lemma A.2. *For every linear operator $L : \mathbb{C}^N \to \mathbb{C}^N$ and $k \in \mathbb{Z}^+$, there exists $\gamma(L^k) \geq \gamma(L)$.*

Proof. Suppose $\gamma(L^k) < \gamma(L)$; then for some $\phi \in [0, 1)$ and unit vector $v \in \mathbb{C}^N$ we have $|L^k v - \exp(2\pi i\phi)v| < \gamma(L)$. Without loss of generality we may assume that $\phi = 0$; otherwise replace L with $\exp(-2\pi i\phi/k)L$, so that $\gamma(L)$ and $\gamma(L^k)$ are unaffected and $|L^k v - v| < \gamma(L)$. Let $\omega = \exp(2\pi i/k)$, and for $j = 0, 1, \dots, k-1$ let

$$u_j = v + \omega^j Lv + \omega^{2j} L^2 v + \cdots + \omega^{(k-1)j} L^{k-1} v. \tag{A.9}$$

Notice that $u_0 + u_1 + \cdots + u_{k-1} = kv$, and since v is a unit vector we must have $|u_j| \geq 1$ for some j. But

$$\begin{aligned} Lu_j - \omega^{-j} u_j &= Lv - \omega^{-j} v + \omega^j L^2 v - Lv + \cdots \omega^{(k-1)j} L^k v - \omega^{(k-2)j} L^{k-1} v \\ &= \omega^{(k-1)j} L^k v - \omega^{-j} v = \omega^{-j}(Lv - v), \end{aligned} \tag{A.10}$$

the last step because ω is a k-th root of unity. This yields $|Lu_j - \omega^{-j} u_j| = |Lv - v| < \gamma(L)$, contradicting the definition of $\gamma(L)$. This completes the proof of the lemma. Q.E.D.

The next lemma is a simple estimate on how much a small perturbation of a linear operator can change its hyperbolicity.

Lemma A.3. *For any pair of linear operators L and Δ of $\mathbb{R}^N$ into itself, hyperbolicity satisfies the estimate*

$$\gamma(L + \Delta) \geq \gamma(L) - \|\Delta\|. \tag{A.11}$$

Proof. By the definition of hyperbolicity,

$$\gamma(L+\Delta) = \inf_{\phi\in[0,1)} \inf_{\|v\|=1} |(L+\Delta)v - \exp(2\pi i\phi)v|. \tag{A.12}$$

By triangle inequality, for all $v \in \mathbb{R}^N$,

$$|(L+\Delta)v - \exp(2\pi i\phi)v| \geq |Lv - \exp(2\pi i\phi)v| - |\Delta v|. \tag{A.13}$$

This implies the statement of the lemma. Q.E.D.

The following lemma generalizes the previous two lemmas. The proof is very similar to that of Lemma A.2, but we need to be a bit more careful.

LEMMA A.4. *For all linear operators $L, L_1, L_2, \dots, L_k : \mathbb{C}^N \to \mathbb{C}^N$,*

$$\gamma(L_k L_{k-1} \cdots L_1) \geq \gamma(L) - \sum_{j=1}^{k} \|L - L_j\|. \tag{A.14}$$

Proof. Choose $v_0 \in \mathbb{C}^N$ and $\phi \in \mathbb{R}$ such that

$$|L_k L_{k-1} \cdots L_1 v_0 - e^{i\phi} v_0| = \gamma(L_k L_{k-1} \cdots L_1)|v_0|. \tag{A.15}$$

Let $v_1 = L_1 v_0$, $v_2 = L_2 L_1 v_0$, $\dots$, $v_k = L_k L_{k-1} \cdots L_1 v_0$. For $j = 0, 1, \dots, k-1$ let

$$\omega_j = e^{i(-\phi + 2\pi j)/k} \tag{A.16}$$

and

$$u_j = v_0 + \omega_j v_1 + \omega_j^2 v_2 + \cdots + \omega_j^{k-1} v_{k-1}. \tag{A.17}$$

Choose ℓ for which $|v_\ell| = \max(|v_0|, |v_1|, \dots, |v_{k-1}|)$, and notice that

$$\sum_{j=0}^{k-1} \omega_j^{-\ell} u_j = k v_\ell. \tag{A.18}$$

Thus there exists j such that

$$|u_j| \geq |v_\ell| = \max(|v_0|, |v_1|, \dots, |v_{k-1}|). \tag{A.19}$$

Then we have

$$\begin{aligned}
\gamma(L) &\leq |Lu_j - \omega_j^{-1} u_j| / |u_j| \\
&= \frac{|-\omega_j^{-1} v_0 + Lv_0 - v_1 + L\omega_j v_1 - \omega_j v_2 + \cdots + L\omega_j^{k-1} v_{k-1}|}{|u_j|} \\
&= \frac{|\omega_j^{k-1}(v_k - e^{i\phi} v_0) + (L - L_1)v_0 + (L - L_2)\omega_j v_1 + \cdots (L - L_k)\omega_j^{k-1} v_{k-1}|}{|u_j|} \\
&\leq \frac{\gamma(L_k L_{k-1} \cdots L_1)|v_0| + \|L - L_1\||v_0| + \|L - L_2\||v_1| + \cdots + \|L - L_k\||v_{k-1}|}{|u_j|} \\
&\leq \gamma(L_k L_{k-1} \cdots L_1) + \|L - L_1\| + \|L - L_2\| + \cdots + \|L - L_k\|,
\end{aligned}$$

which is equivalent to the desired inequality. Q.E.D.

Proposition A.5. *Let $r \le 1 \le K$ be positive numbers and A, B be linear operators of $\mathbb{R}^N$ into itself given by $N \times N$ matrices from $M_N(\mathbb{R})$ with real entries. Consider an N^2-parameter family $\{A_U = A + UB\}_{U \in C^{N^2}(r)}$, where $C^{N^2}(r)$ is the cube in $M_N(\mathbb{R})$ whose entries are bounded in absolute value by r. Suppose that $\|B\|, \|B^{-1}\| \le K$. Then for the Lebesgue product probability measure μ_{r,N^2} on the cube $C^{N^2}(r)$ and all $0 < \gamma \le \min(r, 1)$,*

$$\mu_{r,N^2}\left\{U \in C^{N^2}(r): \ \gamma(A_U) \le \gamma\right\} \le \frac{C(N)K^{2N^2}\gamma}{r^2}, \tag{A.20}$$

where the constant $C(N)$ depends only on N.

Proof. For $0 < \gamma \le 1$ and $\phi \in [0, 1)$, define the sets of non-γ-hyperbolic matrices by

$$\begin{aligned} NH_N^{\gamma}(\mathbb{R}) &= \{L \in M_N(\mathbb{R}): \ \gamma(L) \le \gamma\}, \\ NH_N^{\gamma,\phi}(\mathbb{R}) &= \{L \in M_N(\mathbb{R}): \ \inf_{|v|=1} |(L - \exp(2\pi i\phi)v| \le \gamma\}. \end{aligned} \tag{A.21}$$

Then

$$NH_N^{\gamma}(\mathbb{R}) = \cup_{\phi \in [0,1)} NH_N^{\gamma,\phi}(\mathbb{R}). \tag{A.22}$$

We claim that

$$NH_N^{\gamma}(\mathbb{R}) \subset \cup_{j=0,\dots,[5/\gamma]-1} NH_N^{2\gamma, j/[5/\gamma]}(\mathbb{R}). \tag{A.23}$$

Indeed, suppose that $L \in NH_N^{\gamma}(\mathbb{R})$. Then for some number $\phi \in [0, 1)$ and vector $v \in \mathbb{R}^N$ with $|v| = 1$, we have $|(L - \exp(2\pi i\phi))v| \le \gamma$. Let j be the nearest integer to $[5/\gamma]\phi$ and let $\phi_\gamma = j/[5/\gamma]$; then $\phi - \phi_\gamma \le 1/(2(5/\gamma - 1)) < \gamma/(2\pi)$. Thus

$$|(L - \exp(2\pi i\phi_\gamma))v| \le |(L - \exp(2\pi i\phi))v| + |\exp(2\pi i\phi) - \exp(2\pi i\phi_\gamma)| \le 2\gamma \tag{A.24}$$

and $L \in NH_N^{2\gamma, j/[\pi/\gamma+1]}(\mathbb{R})$ as claimed.

Next, we claim that every matrix in $NH_N^{2\gamma, j/[5/\gamma]}(\mathbb{R})$ lies within 2γ of a matrix in $NH_N^{0, j/[5/\gamma]}(\mathbb{R})$, where we use the Euclidean ($\mathbb{R}^{N^2}$) norm on $M_N(\mathbb{R})$ (not the matrix norm). Consider $L \in NH_N^{2\gamma, j/[5/\gamma]}(\mathbb{R})$, $\phi \in [0, 1)$, and $v \in \mathbb{R}^N$ with $|v| = 1$ and $|(L - \exp(2\pi ij/[5/\gamma]))v| \le 2\gamma$. Let $w = (L - \exp(2\pi ij/[5/\gamma]))v$ and let $M \in M_N(\mathbb{R})$ be the matrix whose k-th row is $w_k v$, where w_k is the k-th coordinate of w. Then the Euclidean norm of M is $|w| \le 2\gamma$ and $Mv = w$, so that $(L - M - \exp(2\pi ij/[5/\gamma]))v = 0$ and hence $L - M \in NH_N^{0, j/[5/\gamma]}(\mathbb{R})$.

We complete the estimate (A.20) by estimating for each j the number of γ-balls needed to cover $NH_N^{0, j/[5/\gamma]}(\mathbb{R})$ within an appropriate bounded domain. It then follows from the previous paragraph that if we inflate each of these balls to the concentric ball of radius 3γ, the collection of larger balls will cover

$NH_N^{2\gamma,j/[5/\gamma]}(\mathbb{R})$, and from the paragraph before that the union over j of these covers will then cover $NH_N^{\gamma}(\mathbb{R})$. To this end, we show that each $NH_N^{0,j/[5/\gamma]}(\mathbb{R})$ is a real algebraic set and compute its codimension.[5] Then we will apply an estimate of Yomdin [Y] on the number of γ-balls necessary to cover a given algebraic set by polynomials of known degree.

Notice that

$$NH_N^{0,\phi}(\mathbb{R}) = \{L \in M_N(\mathbb{R}) : \ \det(L - \exp(2\pi i\phi)\mathrm{Id}) = 0\}. \tag{A.25}$$

We split into the two cases $\exp(2\pi i\phi) \in \mathbb{R}$ (that is, $\phi = 0$ or $1/2$) and $\exp(2\pi i\phi) \notin \mathbb{R}$. In the first case, the equation $\det(L \pm \mathrm{Id}) = 0$ is a polynomial of degree N in the entries of L, so $NH_N^{0,0}(\mathbb{R})$ and $NH_N^{0,1/2}(\mathbb{R})$ are real algebraic sets defined by a single polynomial of degree N.

In the second case, decompose the equation $\det(L-\exp(2\pi i\phi)\mathrm{Id}) = 0$ into two parts: $\mathrm{Re}[\det(L - \exp(2\pi i\phi)\mathrm{Id})] = 0$ and $\mathrm{Im}[\det(L - \exp(2\pi i\phi)\mathrm{Id})] = 0$. Each part is given by a real polynomial of degree N. Furthermore, these two polynomials are algebraically independent, since otherwise $\mathrm{Re}[\det(L - \exp(2\pi i\phi)\mathrm{Id})]$ and $\mathrm{Im}[\det(L - \exp(2\pi i\phi)\mathrm{Id})]$ would satisfy some polynomial relationship and, thus, $\det(L- \exp(2\pi i\phi)\mathrm{Id})$ would take on values only in some real algebraic subset of the complex plane. However, for $N \geq 2$ (which is necessary for complex eigenvalues), by considering real diagonal matrices L we see that the values of $\det(L - \exp(2\pi i\phi)\mathrm{Id})$ contain an open set in $\mathbb{C}$. Therefore, $NH_N^{0,\phi}(\mathbb{R})$ is a real algebraic set given by two algebraically independent polynomials of degree N.

Covering Lemma for Algebraic Sets ([Y, Lemma 4.6]). *Let $V \subset \mathbb{R}^m$ be an algebraic set given by k algebraically independent polynomials $p_1, \ldots, p_k$ of degrees $d_1, \ldots, d_k$ respectively, i.e. $V = \{x \in \mathbb{R}^m : \ p_1(x) = 0, \ldots, \ p_k(x) = 0\}$. Let $C_A^m(s)$ be the cube in $\mathbb{R}^m$ with side $2s$ centered at some point A. Then for $\gamma \leq s$, the number of γ-balls necessary to cover $V \cap C_A^m(s)$ does not exceed $C(D,m)\,(2s/\gamma)^{m-k}$, where the constant $C(D,m)$ depends only on the dimension m and product of degrees $D = \prod_i d_i$.*

Remark A.6. Some additional arguments based on Bezout's Theorem give an upper estimate of $C(D,m)$ by $2^m D$ for γ sufficiently small.

To complete the proof of Proposition A.5, we apply the Covering Lemma for Algebraic Sets to each $NH^{0,j/[5/\gamma]}(\mathbb{R})$, where $j = 0, \ldots, [5/\gamma] - 1$, with $m = N^2$, $s = Kr$, and A as in the statement of the proposition. (Notice that if $U \in C^{N^2}(r)$ then $A + UB \in C_A^{N^2}(Kr)$, so we need only cover the part of $NH^{0,j/[5/\gamma]}(\mathbb{R})$ lying in the latter set.) In the case that $j/[5/\gamma] = 0$ or $1/2$, we have $k = 1$, $d_1 = N$, and $D = N$, so that the number of covering

[5] Unfortunately $NH_N^0(\mathbb{R})$, in contrast to $NH_N^{0,j/[5/\gamma]}(\mathbb{R})$, is not algebraic.

γ-balls is bounded by $C(N,N^2)(2Kr/\gamma)^{N^2-1}$. In the case of other j, we have $k=2$, $d_1=d_2=N$, and $D=N^2$, so that the number of covering γ-balls is bounded by $C(N^2,N^2)(2Kr/\gamma)^{N^2-2}$. The number of j's of the second type is less than $5/\gamma$. Combining all these estimates along with (A.23) we get that $NH^\gamma(\mathbb{R})\cap C_A^{N^2}(Kr)$ can be covered by $C(N^2,N^2)(2+5/(2Kr))(2Kr/\gamma)^{N^2-1}$ balls of radius 3γ.

Finally, notice that the preimage of a ball of radius 3γ under the map $U\mapsto A+UB$ is contained in a ball of radius $3K\gamma$, whose μ_{r,N^2}-measure is less than $(3K\gamma/r)^{N^2}$. Therefore the total measure of $3K\gamma$-balls needed to cover the set $\{U\in C^{N^2}(r):\gamma(A+UB)\le\gamma\}$ is at most $C(N)K^{2N^2}\gamma/r^2$, where the constant $C(N)$ depends only on N. Q.E.D.

Appendix B: Orthogonal transformations of R^N and the spaces of homogeneous polynomials

In this appendix, we prove that the scalar product (1.12) in the space $W_{k,N}$ of homogeneous N-vector polynomials of degree k in N variables is invariant with respect to orthogonal transformations of $\mathbb{R}^N$.

LEMMA B.1. *Let $x\in\mathbb{R}^N$ be given by N coordinates $x=(x_1,\dots,x_N)$. For $k\in\mathbb{Z}_+$, consider homogeneous polynomials $p_k(x)=\sum_{|\alpha|=k}\vec{\varepsilon}_\alpha x^\alpha\in W_{k,N}$ in N variables, where $x^\alpha=x_1^{\alpha_1}\dots x_N^{\alpha_N}$. Let $O\in\mathrm{SO}(N)$ be an orthogonal transformation of $\mathbb{R}^N$. Denote by $x'=(x'_1,\dots,x'_N)$ the new coordinate system induced by the relation $x=Ox'$. Write $p'_k(x')=p_k(Ox')=\sum_{|\alpha|=k}\vec{\varepsilon}'_\alpha(x')^\alpha$ in the new coordinate system. Then for all $\{\vec{\varepsilon}_\alpha\}_{|\alpha|=k}$ and $\{\vec{\nu}_\alpha\}_{|\alpha|=k}$,*

$$\sum_{|\alpha|=k}\binom{k}{\alpha}^{-1}\langle\vec{\varepsilon}_\alpha,\vec{\nu}_\alpha\rangle=\sum_{|\alpha|=k}\binom{k}{\alpha}^{-1}\langle\vec{\varepsilon}'_\alpha,\vec{\nu}'_\alpha\rangle. \tag{B.1}$$

Proof. For this lemma it will be helpful to use a different notation for monomials. Given a k-tuple $\beta=(\beta_1,\dots,\beta_k)\in\{1,2,\dots,N\}^k$, define $x^{(\beta)}=x_{\beta_1}x_{\beta_2}\cdots x_{\beta_k}$. Notice that $x^{(\beta)}=x^\alpha$ where α_i is the number of indices j for which $\beta_j=i$. Write $\alpha(\beta)$ for the multi-index corresponding in this manner to the k-tuple β, and notice that for each multiindex α there are $\binom{k}{\alpha}$ different k-tuples β for which $\alpha(\beta)=\alpha$. Let $|\beta|=k$.

Given a polynomial p_k as in the statement of the lemma, we can write $p_k(x)=\sum_{|\beta|=k}\vec{\varepsilon}_{(\beta)}x^{(\beta)}$, where $\vec{\varepsilon}_{(\beta)}=\binom{k}{\alpha}^{-1}\vec{\varepsilon}_{\alpha(\beta)}$. We can also rewrite the scalar product as follows:

$$\sum_{|\alpha|=k}\binom{k}{\alpha}^{-1}\langle\vec{\varepsilon}_\alpha,\vec{\nu}_\alpha\rangle=\sum_{|\beta|=k}\langle\vec{\varepsilon}_{(\beta)},\vec{\nu}_{(\beta)}\rangle. \tag{B.2}$$

(Remember that for each α, there are $\binom{k}{\alpha}$ corresponding terms on the right-hand side.)

Our goal is then to show that

$$\sum_{|\beta|=k} \langle \vec{\varepsilon}_{(\beta)}, \vec{\nu}_{(\beta)} \rangle = \sum_{|\beta|=k} \langle \vec{\varepsilon}'_{(\beta)}, \vec{\nu}'_{(\beta)} \rangle. \tag{B.3}$$

We prove this by induction on k. For $k = 0$, the identity is trivial. Assume that the identity now holds for some $k \geq 0$. Given β with $|\beta| = k$ and $i \in \{1, 2, \dots, N\}$, let (β, i) be the $(k+1)$-tuple $(\beta_1, \dots, \beta_k, i)$. Also, let $\vec{\varepsilon}^{\,i}_{(\beta)} = \vec{\varepsilon}_{(\beta,i)}$. The reason for this alternate notation is that we will mean something different below by $\vec{\varepsilon}^{\,i\prime}_{(\beta)}$ and $\vec{\varepsilon}^{\,\prime}_{(\beta,i)}$. In the former case, we fix i and apply the coordinate transformation O to the polynomial $\sum_{|\beta|=k} \vec{\varepsilon}^{\,i}_{(\beta)} x^{(\beta)}$ to get the coefficients $\vec{\varepsilon}^{\,i\prime}_{(\beta)}$. In the latter case, we transform the polynomial $\sum_{|(\beta,i)|=k+1} \vec{\varepsilon}_{(\beta,i)} x^{(\beta,i)}$ to get the coefficients $\vec{\varepsilon}^{\,\prime}_{(\beta,i)}$.

Next, notice that

$$\sum_{|(\beta,i)|=k+1} \vec{\varepsilon}_{(\beta,i)} x^{(\beta,i)} = \sum_{i=1}^{N} x_i \sum_{|\beta|=k} \vec{\varepsilon}^{\,i}_{(\beta)} x^{(\beta)}. \tag{B.4}$$

Applying the coordinate change $x = Ox'$ to both sides, we get

$$\sum_{|(\beta,i)|=k+1} \vec{\varepsilon}^{\,\prime}_{(\beta,i)} (x')^{(\beta,i)} = \sum_{i=1}^{N} \sum_{j=1}^{N} O_{ij} x'_j \sum_{|\beta|=k} \vec{\varepsilon}^{\,i\prime}_{(\beta)} (x')^{(\beta)}. \tag{B.5}$$

It follows that

$$\vec{\varepsilon}^{\,\prime}_{(\beta,j)} = \sum_{i=1}^{N} O_{ij} \vec{\varepsilon}^{\,i\prime}_{(\beta)}. \tag{B.6}$$

A similar identity holds with $\vec{\varepsilon}$ replaced by $\vec{\nu}$, whereupon

$$\sum_{|(\beta,j)|=k+1} \langle \vec{\varepsilon}^{\,\prime}_{(\beta,j)}, \vec{\nu}'_{(\beta,j)} \rangle = \sum_{j=1}^{N} \sum_{|\beta|=k} \sum_{i=1}^{N} \sum_{\ell=1}^{N} \langle O_{ij} \vec{\varepsilon}^{\,i\prime}_{(\beta)}, O_{\ell j} \vec{\nu}^{\,\ell\prime}_{(\beta)} \rangle. \tag{B.7}$$

Since O is an orthogonal matrix, $\sum_{j=1}^{N} O_{ij} O_{\ell j} = \delta_{i\ell}$. Exchanging the order of summation on the right-hand side above, we then have

$$\sum_{|(\beta,j)|=k+1} \langle \vec{\varepsilon}^{\,\prime}_{(\beta,j)}, \vec{\nu}'_{(\beta,j)} \rangle = \sum_{i=1}^{N} \sum_{|\beta|=k} \langle \vec{\varepsilon}^{\,i\prime}_{(\beta)}, \vec{\nu}^{\,i\prime}_{(\beta)} \rangle = \sum_{|(\beta,i)|=k+1} \langle \vec{\varepsilon}^{\,i}_{(\beta)}, \vec{\nu}^{\,i}_{(\beta)} \rangle \tag{B.8}$$

by the inductive hypothesis. Q.E.D.

Appendix C: Embedding of the space of diffeomorphisms of a compact manifold $\mathrm{Diff}^r(M)$ into that of an open set in a Euclidean space

In this appendix, we describe how to extend and perturb a diffeomorphism of a compact manifold embedded into a Euclidean space, and what conditions we need to ensure that the results of Sacker [Sac] and Fenichel [F] about persistence of invariant manifolds apply.

Recall that M is a smooth (C^∞) compact manifold, and let f be a diffeomorphism in $\mathrm{Diff}^r(M)$. First we consider a manifold $\tilde{M} = M \times [0,1]/\sim$, where the equivalence relation is defined by $(x,1) \sim (f(x),0)$ for all $x \in M$. Then $\tilde{M}$ is as smooth as f is and carries a naturally defined vector field X_f whose time one map, restricted to $M \times \{0\}$, coincides with f. Such a construction is usually called *suspension*. Now we embed $\tilde{M}$ into the interior of the closed unit ball $B^{N+1} \subset \mathbb{R}^{N+1}$ in such a way that $M \times \{0\}$ embeds into an N-dimensional subspace. Given a compact manifold $\tilde{M}$ of dimension D, for $N+1 > 2D$ the Whitney Embedding Theorem (see e.g. [GG]) says that a generic smooth map from $\tilde{M}$ to $\mathbb{R}^{N+1}$ is an embedding, i.e., a diffeomorphism between $\tilde{M}$ and its image. Fix coordinates in $\mathbb{R}^{N+1}$. Consider a smooth map $E' : \tilde{M} \to B^{N+1}$ such that $E'(M \times \{0\}) \subset \mathbb{R}^N \times \{0\}$, the hyperplane where the last coordinate is zero. By the Whitney Embedding Theorem, we can choose a small perturbation E of E' such that E is an embedding of $\tilde{M}$. Since $E(M \times \{0\})$ is close to $\mathbb{R}^N \times \{0\}$, we can change coordinates in $\mathbb{R}^{N+1}$ so that $E(M \times \{0\}) \subset \mathbb{R}^N \times \{0\}$ with a new coordinate system. The mapping $x \mapsto E(x,0)$ then provides an embedding of M into the hyperplane in $\mathbb{R}^{N+1}$ where the last coordinate equals zero, which we identify with $\mathbb{R}^N$. To simplify notation, we identify $\tilde{M}$ and M with their images, so that $\tilde{M}$ and M become submanifolds of $\mathbb{R}^{N+1}$ and $\mathbb{R}^N$ respectively, with $M \times \{0\}$ a submanifold of $\tilde{M}$.

Before we present a way to extend a vector field on $\tilde{M}$ to its tube neighborhood, we need to recall the notion of a k-normally hyperbolic manifold. For a linear transformation L, let

$$m(L) = \inf\{|Lx| : |x| = 1\}.$$

When L is invertible, $m(L) = \|L^{-1}\|^{-1}$. Fix $t > 0$. Let X be a C^r smooth vector field on $\mathbb{R}^{N+1}$ and X^t be the time t map along trajectories of X. Let $T_{\tilde{M}}\mathbb{R}^{N+1}$ be the tangent bundle of $\mathbb{R}^{N+1}$ over $\tilde{M}$. Suppose we have a dX^t-invariant splitting into three subspaces

$$T_{\tilde{M}}\mathbb{R}^{N+1} = W^u \bigoplus T\tilde{M} \bigoplus W^s,$$

i.e. for any $y \in \mathbb{R}^{N+1}$ we have $dX^t(y)W^s_y = W^s_{X^t(y)}$ and $dX^t(y)W^u_y = W^u_{X^t(y)}$. Moreover, for some $C > 0$ and $\lambda > 1$ we have $|dX^t(y)v| \ge C\lambda^t|v|$ for all $y \in \tilde{M}$, all $v \in W^u_y$ (respectively W^s_y), and all $t \ge 0$ (respectively $t \le 0$). Denote

$$d(X^t)^s(y) = dX^t(y)|_{W^s},\ d(X^t)^u(y) = dX^t(y)|_{W^u},\ d(X^t)^c(y) = dX^t(y)|_{T_y\tilde{M}}.$$

Let $0 \leq k \leq r$. We say that the vector field X is *k-normally hyperbolic* at $\tilde{M}$ if there is such a splitting that for all $y \in \tilde{M}$ we have:

$$m(d(X^t)^u(y)) > \|d(X^t)^c(y)\|^k \quad \text{and} \quad m(d(X^t)^c(y))^k > \|d(X^t)^s(y)\|.$$

Notice that if X^t is k-normally hyperbolic for small enough t, then it is k-normally hyperbolic for all positive t.

Let $\tilde{T} \subset \mathbb{R}^{N+1}$ be a closed neighborhood of $\tilde{M}$, chosen sufficiently small that there is a well-defined projection $\tilde{\pi} : \tilde{T} \to \tilde{M}$ for which $\tilde{\pi}(\tilde{x})$ is the closest point in $\tilde{M}$ to $\tilde{x}$. Then for each $\tilde{y} \in \tilde{M}$, by the Implicit Function Theorem, $\tilde{\pi}^{-1}(\tilde{y})$ is an $(N-D)$-dimensional disk. Then we can extend each vector field X on $\tilde{M}$ to a vector field $\mathcal{X}$ on $\tilde{T}$ so that the component tangent to $\tilde{\pi}^{-1}(\tilde{y})$ is directed toward $\tilde{y}$ and is r-dominated by the orthogonal one, meaning that the vector field $\mathcal{X}$ is r-normally hyperbolic at $\tilde{M}$. Such an extension is possible because $\tilde{M}$ is compact and one can keep increasing the "strength" of attraction toward $\tilde{M}$ by $\mathcal{X}$ until r-normal hyperbolicity is attained.

Consider the Poincaré return map of $\mathcal{X}_f$ from $T = \tilde{T} \cap \{B^N \times \{0\}\}$ into $B^N \times \{0\}$, which is well-defined by the construction. Denote this map by F. The vector field $\mathcal{X}_f$ is directed so that F maps T strictly inside itself. Now we shall use r-normally hyperbolicity of $\mathcal{X}_f$ to construct an Artin-Mazur approximation f_σ of f via approximating $\mathcal{X}_f$ and relying on persistence of M for the Poincaré return map of $\mathcal{X}_f$.

Now the closed neighborhood T of M can be considered as a subset of $\mathbb{R}^N$ and can be chosen sufficiently small that there is a well-defined projection $\pi : T \to M$ for which $\pi(x)$ is the closest point in M to x. Every small perturbation $F_\sigma \in C^r(T)$ of F can be suspended to a vector field $\mathcal{X}_{f_\sigma}$ close to $\mathcal{X}_f$. Then by Fenichel's Theorem [F], for σ sufficiently small F_σ has an invariant manifold $M_\sigma \subset T$ for which $\pi|_{M_\sigma}$ is a C^r diffeomorphism from M_σ to M. Then to such an F_σ we can associate a diffeomorphism $f_\sigma \in \mathrm{Diff}^r(M)$ by letting

$$f_\sigma(y) = \pi(F_\sigma(\pi|_{M_\sigma}^{-1}(y))).$$

Notice that the periodic points of F_σ all lie on M_σ and are in one-to-one correspondence with the periodic points of f_σ. Furthermore, because f_σ and $F_\sigma|_{M_\sigma}$ are conjugate, the hyperbolicity of each periodic orbit is the same for either map. Thus any estimate on $P_n(F_\sigma)$ or $\gamma_n(F_\sigma)$ applies also to f_σ.

The construction above defines a function Π from a neighborhood of $F \in C^r(T)$ to a neighborhood of $f \in \mathrm{Diff}^r(M)$ such that $f_\sigma = \Pi(F_\sigma)$. While Π is not one-to-one, as mentioned above each F_σ for which $\Pi(F_\sigma) = f_\sigma$ has the same periodic points as f_σ with the same hyperbolicity. Thus the properties of periodic orbits studied in this paper are the same for all elements of $\Pi^{-1}(f_\sigma)$. Furthermore, given f_σ sufficiently close to f, we can construct a canonical diffeomorphism $F_\sigma \in \Pi^{-1}(f_\sigma)$ as follows. First notice that while $\tilde{M}$ is defined in terms of f, if f_σ is sufficiently close to f then the manifold associated with the suspension flow X_{f_σ} of f_σ is diffeomorphic to $\tilde{M}$ (see Theorem 1 in [V]). Thus we can consider X_{f_σ} to be defined on $\tilde{M}$, extend it as above to a vector field $\mathcal{X}_{f_\sigma}$ on $\tilde{T}$, and define F_σ to be the Poincaré return map of $\mathcal{X}_{f_\sigma}$ on T.

The meaning of the phrase "generic m-parameter family" in Theorems 1.3.7 and 1.3.11 is based on the constructions above. Given an m-parameter family $\{f_\sigma\}_{\sigma\in B^m} \subset \mathrm{Diff}^r(M)$ for which the perturbations are sufficiently small, we construct the corresponding family $\{F_\sigma\}_{\sigma\in B^m} \subset C^r(T)$. Recalling the notation of Section 1.3, we say that a property holds for a generic m-parameter family in $\mathrm{Diff}^r(M)$ if there is a family of perturbations $\{\phi_{\vec{\varepsilon}}\}_{\vec{\varepsilon}\in HB^n(\vec{\mathbf{r}})} \subset C^r(T)$, independent of the family $\{f_\sigma\}_{\sigma\in B^m}$, such that for $\mu^n_{\vec{\mathbf{r}}}$-a.e. $\vec{\varepsilon}$, the m-parameter family $\{\Pi(F_\sigma + \phi_{\vec{\varepsilon}})\}_{\sigma\in B^m}$ has the desired property.

Theorem 1.3.7 then follows from the Main Theorem (and likewise, Theorem 1.3.11 from Theorem 1.3.9) by the Fubini/Tonelli theorem. Specifically, for each fixed σ, we know that for $\mu^n_{\vec{\mathbf{r}}}$-a.e. $\vec{\varepsilon}$, the conclusion of the Main Theorem holds for $\Pi(F_\sigma + \phi_{\vec{\varepsilon}})$. Therefore, for $\mu^n_{\vec{\mathbf{r}}}$-a.e. $\vec{\varepsilon}$, the same property holds for Lebesgue almost every σ, and this is what Theorem 1.3.7 says based on the discussion above.

Appendix D: Pathological examples of decay of product of distances of recurrent trajectories

In this appendix we present two types of orbits of a horseshoe diffeomorphism that show that with the methods in this paper, the estimate $\exp(Cn^{1+\delta})$ on the growth of the number of periodic points (the Main Theorem from §1.3) cannot be improved to $\exp(Cn(\log n)^\delta)$ for any real number δ. More exactly, the Shift Theorem, stated in Section 3.5, is crucial to split all almost periodic trajectories into classes as in (3.12). In Section 3.5, we outline the proof of this theorem and it might be helpful to review the strategy presented there, especially, the last remark right before Section 3.5.1. Suppose that we now set $\gamma_n(C,\delta)$, which is roughly the inverse of the bound we get on the number of periodic points, equal to $\exp(-Cn(\log n)^\delta)$. In Example 2,we construct a trajectory that for an infinite number of periods n is nonsimple, has no leading saddles, and no close returns (gaps), as defined in Section 3.5. Thus for such slowly decaying $\gamma_n(C,\delta)$, we cannot deal with these kinds of trajectories with our methods.

First we give an example that shows more simply that the product of distances along a period n orbit can be of order $\exp(-Cn\log n)$ even though the closest return along the orbit is of order $\exp(-Cn)$.

Example D.1. Consider the sequence of periodic orbits of a horseshoe map with symbolic dynamics

$$S_0 = 1$$
$$S_1 = 0$$

$$
\begin{aligned}
S_2 &= 01\\
S_3 &= 010\\
S_4 &= 01001\\
S_5 &= 01001010\\
S_6 &= 0100101001001\\
&\ \vdots\ .
\end{aligned}
$$

Each sequence is the concatenation of the previous two sequences; it can also be generated from the previous sequence by the substitution rules $0 \to 01$ and $1 \to 0$. The number of symbols in S_n is the n-th Fibonacci number F_n.

Notice also that

$$
\begin{aligned}
S_n &= S_{n-1}S_{n-2}\\
&= S_{n-2}S_{n-3}S_{n-2}\\
&= S_{n-3}S_{n-4}S_{n-3}S_{n-3}S_{n-4}\\
&= S_{n-4}S_{n-5}S_{n-4}S_{n-4}S_{n-5}S_{n-4}S_{n-5}S_{n-4}\\
&= \dots .
\end{aligned}
$$

More formally, the sequence S_n can be generated from S_k for any $0 \le k \le n$ by replacing each 0 in S_k by S_{n-k+1} and each 1 by S_{n-k}. We refer below to this decomposition of S_n into copies of S_{n-k} and S_{n-k+1} as "decomposition k".

Every three symbol subsequence of S_n is either 010, 100, 001, or 101. Furthermore, when each S_n is a cyclic sequence, each of the four triplets above occurs at least once in S_4, at least once in S_5, at least twice in S_6, and in general at least F_{n-4} times in S_n for $n \ge 4$. The importance of this observation below will be that in decomposition k for $4 \le k \le n$, each of the substrings $S_{n-k+1}S_{n-k}S_{n-k+1}$, $S_{n-k}S_{n-k+1}S_{n-k+1}$, $S_{n-k+1}S_{n-k+1}S_{n-k}$, and $S_{n-k}S_{n-k+1}S_{n-k}$ occurs at least F_{k-4} times.

Now let $x_0, x_1, \dots, x_{p-1}$ be points in the periodic orbit with symbolic dynamics S_n, where $p = F_n$ is the length of S_n. No matter where the symbolic sequence of x_0 starts within S_n, we claim that for n sufficiently large,

$$
\prod_{j=1}^{p-1} |x_0 - x_j| \le e^{(c_1 - c_2 n)p} \le e^{c_1 p - c_3 p \log p}
$$

for some positive constants c_1, c_2, c_3 independent of n. The latter inequality follows from the fact that $p \le 2^n$, so it remains to prove the first inequality.

Assume that the distance between any two points in the nonwandering set is at most 1. Say the symbolic sequence of x_0 starts at the m-th symbol of S_n. If for some positive integer q, the block of $2q - 1$ symbols centered at the m-th symbol is repeated centered at the ℓ-th symbol, then the distance between the points x_0 and $x_{\ell-m}$ is bounded above by e^{-cq} for an appropriate positive constant c. Here the index $\ell - m$ is taken modulo p.

Now for $4 \le k \le n$, in decomposition k the m-th symbol in S_n lies in a copy of either S_{n-k} or S_{n-k+1}, which in turn lies in the middle of one of the four substrings $S_{n-k+1}S_{n-k}S_{n-k+1}$, $S_{n-k}S_{n-k+1}S_{n-k+1}$, $S_{n-k+1}S_{n-k+1}S_{n-k}$, and $S_{n-k}S_{n-k+1}S_{n-k}$ described above. Each such substring occurs at least F_{k-4} times in S_n, and all but one of these occurrences contributes a factor of at most $e^{-cF_{n-k}}$ to the product of distances $|x_0 - x_j|$. Therefore for $n \ge 6$,

$$\begin{aligned}\prod_{j=1}^{p-1} |x_0 - x_j| &\le \prod_{k=6}^{n} (e^{-cF_{n-k}})^{F_{k-4}-F_{k-5}} \\ &= e^{-c\sum_{k=6}^{n} F_{n-k}F_{k-6}} \\ &\le e^{-c(n-5)F_n/F_8} \\ &= e^{c(5-n)p/34}.\end{aligned}$$

(The estimate $F_{n-k}F_{k-6} \ge F_n/F_8$ can be proved by induction, but heuristically this type of estimate follows from the fact that F_n is approximately an exponential function of n.)

Example D.2. Consider now the aperiodic nonwandering orbit of the horseshoe map whose symbolic dynamics are given as follows. Given a sequence of positive integers $k_1, k_2, \dots$, let $S_0 = 0$ and $S_n = 1S_{n-1}S_{n-1}\cdots S_{n-1}1$ where S_{n-1} occurs $2k_n + 1$ consecutive times. For example, if $k_n = n$ then

$$\begin{aligned}S_0 &= 0 \\ S_1 &= 10001 \\ S_2 &= 1100011000110001100011000 11 \\ &\ \ \vdots\ .\end{aligned}$$

Each sequence is symmetric, and for $n \ge 1$, each S_n contains a copy of S_{n-1} at its center. Let L_n be the length of S_n; then $L_0 = 1$ and $L_n = (2k_n+1)L_{n-1}+2$ for $n \ge 1$. One can easily check that $k_1k_2\cdots k_n \le L_n \le 5^n k_1k_2\cdots k_n$.

Let x_0 be the point on the nonwandering set whose symbolic sequence has middle L_n symbols S_n for each $n \ge 0$. By symmetry, to estimate the product of distances $|x_0 - x_j|$ as j goes from 1 to $L_n - 1$, we can estimate the product as j goes from $-(L_n - 1)$ to $L_n - 1$, excluding $j = 0$, and take the square root of the latter estimate.

As in the previous example, let c be a positive constant such that $|x_0 - x_j| \le e^{-cq}$, where q is the largest positive integer for which the middle $2q - 1$ symbols of the sequences for x_0 and x_j agree, or $q = 0$ if their middle symbols do not agree. Then for all $n \ge 1$ and $-k_n \le m \le k_n$ we have $|x_0 - x_{mL_{n-1}}| \le e^{-c(k_n-|m|+1/2)L_{n-1}}$. The square root of the product of these upper bounds, excluding $m = 0$, is

$$\prod_{m=1}^{k_n} e^{-c(k_n - m + 1/2)L_{n-1}} = e^{-ck_n^2 L_{n-1}/2} \le e^{-ck_n L_n/10}.$$

Here we used the inequality $5k_n L_{n-1} \ge (2k_n + 1)L_{n-1} + 2 = L_n$.

In addition, for n and m as above and all $-k_{n-1} \le p \le k_{n-1}$ we have $|x_0 - x_{mL_{n-1}+pL_{n-2}}| \le e^{-c(k_{n-1}-|p|+1/2)L_{n-2}}$. The square root of the product of these upper bounds, excluding $p = 0$, is

$$\prod_{m=-k_n}^{k_n} \prod_{p=1}^{k_{n-1}} e^{-c(k_{n-1}-p+1/2)L_{n-2}} = e^{-c(2k_n+1)k_{n-1}^2 L_{n-2}/2} \le e^{-c(2k_n+1)k_{n-1}(L_{n-1}+1)/12}.$$

Here we used the inequality $6k_{n-1}L_{n-2} \ge (2k_{n-2} + 1)L_{n-2} + 3 = L_{n-1} + 1$. Then in turn we can say $(2k_n + 1)(L_{n-1} + 1) \ge (2k_n + 1)L_{n-1} + 3 = L_n + 1$, so that the bound on the product above can be replaced by $e^{-ck_{n-1}L_n/12}$.

In a similar manner, we can bound above another set of terms contributing to the product of distances $|x_0 - x_j|$ by $e^{-ck_{n-\ell}L_n/12}$ for $\ell = 2, 3, \ldots, n-1$. Multiplying all these bounds together we conclude that

$$\prod_{j=1}^{L_n-1} |x_0 - x_j| \le e^{-c(k_1+k_2+\cdots+k_n)L_n/12}.$$

Notice that if $k_n = k$ independent of n, then $L_n \sim (2k+1)^n$ and $k_1 + k_2 + \cdots + k_n = nk \sim \log L_n$. Thus we get an estimate similar to Example 1.

If $k_n = n^\alpha$, then $\log L_n \sim n \log n$ and hence $k_1 + k_2 + \cdots + k_n \sim n^{\alpha+1} \sim (\log L_n)^{\alpha+1}$, loosely speaking. The closest returns to x_0 are of the form $-\log|x_0 - x_{L_n}| \sim k_{n+1}L_n \sim L_n(\log L_n)^\alpha$, loosely speaking again. Thus if we attempt to apply the Inductive Hypothesis with $\gamma_j(C, \delta) = \exp(-Cj(\log j)^\delta)$, this example with $\alpha = \delta - 1/2$ shows that the product of distances along a hyperbolic trajectory can be smaller than any fixed power of $\gamma_j(C, \delta)$ for arbitrarily large $j = L_n$, despite the fact that the closest return over j iterations is larger than any fixed power of $\gamma_j(C, \delta)$ for j sufficiently large.

California Institute of Technology, Pasadena, CA
E-mail address: kaloshin@its.caltech.edu

University of Maryland, College Park, MD
E-mail address: bhunt@math.umd.edu

References

[A] V. Arnold, Problems of Arnold's seminar, 1989.

[AM] M. Artin and B. Mazur, On periodic orbits, *Ann. of Math.* **81** (1965), 82–99.

[F] N. Fenichel, Persistence and smoothness of invariant manifolds for flows, *Indiana Univ. Math. J.* **21** (1971), 193–226.

[GG] M. Golubitsky and V. Guillemin, *Stable Mappings and their Singularities*, Springer-Verlag, New York (1973).

[GST1] S. Gonchenko, L. Shil'nikov, and D. Turaev, On models with non-rough Poincaré homoclinic curves, *Physica* D **62** (1993), 1–14.

[GST2] ———, Homoclinic tangencies of an arbitrary order in Newhouse regions (Russian), in *Dynamical Systems* (Russian) (Moscow, 1998), 69–128; *Itogi Nauki Tekh. Ser. Sovrem. Mat. Prilozh. Temat. Obz*. **67**, Vseross. Inst. Nauchn. i Tekhn. Inform. (VINITI), Moscow, 1999.

[GHK] A. Gorodetski, B. Hunt, and V. Kaloshin, Newton interpolation polynomials, discretization method, and certain prevalent properties in dynamical systems, *Proc. ICM*, Madrid, Spain, 2006, European Mathematical Society, vol. 2, 27–55, 2006.

[GK] A. Gorodetski and V. Kaloshin, How often surface diffeomorphisms have infinitely many sinks and hyperbolicity of periodic points near a homoclinic tangency, *Adv. in Math.* **208** (2007), 710–797.

[GY] A. Grigoriev and S. Yakovenko, Topology of generic multijet preimages and blow-up via Newton interpolation, *J. Diff. Eq.* **150** (1998), 349–362.

[G] M. Gromov, On entropy of holomorphic maps, preprint.

[HK] B. Hunt and V. Kaloshin, Prevalence, preprint.

[HSY] B. Hunt, T. Sauer, and J. Yorke, Prevalence: a translation-invariant almost every for infinite dimensional spaces, *Bull. Amer. Math. Soc*. **27** (1992), 217–238; Prevalence: an addendum, *Bull. Amer. Math. Soc*. **28** (1993), 306–307.

[K1] V. Kaloshin, An extension of the Artin-Mazur theorem, *Ann. of Math*. **150** (1999), 729–741.

[K2] ———, Generic diffeomorphisms with superexponential growth of number of periodic orbits, *Comm. Math. Phys*. **211** (2000), 253–271.

[K3] ———, Some prevalent properties of smooth dynamical systems, *Tr. Math. Inst. Steklova* **213** (1997), 123–151.

[K4] ———, Stretched exponential estimate on growth of the number of periodic points for prevalent diffeomorphisms, Ph.D. Thesis, Princeton University, 2001.

[K5] ———, Stretched exponential estimate on growth of the number of periodic points for prevalent diffeomorphisms II, preprint, 86pp, www.its.caltech.edu/∼kaloshin/.

[KH] V. Kaloshin and B. Hunt, A stretched exponential bound on the rate of growth of the number of periodic points for prevalent diffeomorphisms. I, II, *Electron. Res. Announc. Amer. Math. Soc*. **7** (2001), 17–27, 28–36.

[KK] V. Kaloshin and O. Kozlovski, A C^r unimodal map with an arbitrary fast growth of the number of periodic points, preprint, 5pp, www.its.caltech.edu/∼kaloshin/.

[KS] V. Kaloshin and M. Saprykina, Generic 3-dimensional volume-preserving diffeomorphisms with superexponential growth of number of periodic orbits, *Discrete Contin. Dyn. Syst.* **15** (2006), 611–640.

[MMS] M. Martens, W. de Melo, and S. van Strien, Julia-Fatou-Sullivan theory for real one-dimensional dynamics, *Acta Math.* **168** (1992), 273–318.

[O] J. C. Oxtoby, *Measure and Category*, Springer-Verlag, New York (1971).

[PM] J. Palis and W. de Melo, *Geometric Theory of Dynamical Systems*: *An Introduction*, Springer-Verlag, New York (1982).

[Sac] R. J. Sacker, A perturbation theorem for invariant manifolds and Hölder continuity, *J. Math. Mech.* **18** (1969), 705–762.

[San] L. Santal, *Integral Geometry and Geometric Probability*, *Encycl. Math and its Appl.* **1**, Addison-Wesley Publ. Co., Reading, Mass., 1976.

[Sar] A. Sard, The measure of the critical points of differentiable maps, *Bull. Amer. Math. Soc.* **48** (1942), 883–890.

[Sz] W. Szlenk, *An Introduction to the Theory of Smooth Dymanical Systems*, Wiley & Sons, New York, 1984.

[V] A. Verjovsky, Flows with cross sections, *Proc. Nat. Acad. Sci.* **66** (1970), 1154–1156.

[W] H. Whitney, Differentiable manifolds, *Ann. of Math.* **37** (1936), 645–680.

[Y] Y. Yomdin, A quantitative version of the Kupka-Smale Theorem, *Ergodic Theory Dynam. Systems* **5** (1985), 449–472.

(Received July 28, 2003)
(Revised January 4, 2005)