Astérisque

WELLINGTON DE MELO
MARCELO VIANA
JEAN-CHRISTOPHE YOCCOZ (éd.)

Geometric methods in dynamics (I) : Volume in honor of Jacob Palis

Astérisque, tome 286 (2003)

<http://www.numdam.org/item?id=AST_2003__286_>

ASTÉRISQUE 286

GEOMETRIC METHODS IN DYNAMICS (I)

VOLUME IN HONOR OF JACOB PALIS

edited by

Welington de Melo
Marcelo Viana
Jean-Christophe Yoccoz

Société Mathématique de France 2003
Publié avec le concours du Centre National de la Recherche Scientifique

W. de Melo

IMPA, Estrada Dona Castorina, 110, Jardim Botânico, Rio de Janeiro 22460-320, Brazil.

E-mail : demelo@impa.br

Url : www.impa.br/~demelo

M. Viana

IMPA, Estrada Dona Castorina, 110, Jardim Botânico, Rio de Janeiro 22460-320, Brazil.

E-mail : viana@impa.br

Url : www.impa.br/~viana

J.-C. Yoccoz

Collège de France, 11, Place Marcelin Berthelot, 75005 Paris, France.

E-mail : jean-c.yoccoz@college-de-france.fr

Url : www.college-de-france.fr/site/equ_dif/p999000715275.htm

2000 *Mathematics Subject Classification.* — 37-XX, 34-XX, 60-XX, 35-XX.

Key words and phrases. — Dynamical systems, ergodic theory, bifurcation theory, differential equations.

GEOMETRIC METHODS IN DYNAMICS (I)

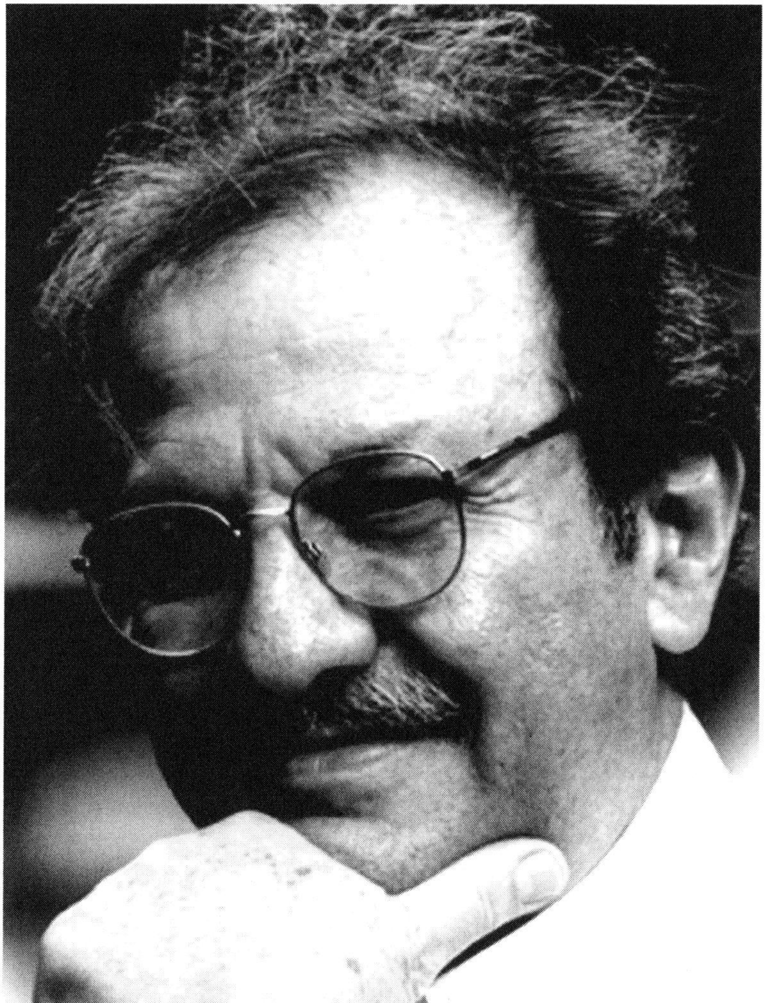
VOLUME IN HONOR OF JACOB PALIS

edited by Welington de Melo, Marcelo Viana, Jean-Christophe Yoccoz

***Abstract.* —** This is the first of two volumes collecting original research articles, on several aspects of dynamics, mostly by participants in the International Conference on Dynamical Systems held at IMPA (Rio de Janeiro), in July 2000, to celebrate Jacob Palis' 60th birthday.

***Résumé* (Méthodes géométriques en dynamique (I). Volume en l'honneur de Jacob Palis)**

Ceci est le premier de deux volumes regroupant des articles originaux de recherche concernant des aspects variés de la théorie des systèmes dynamiques, écrits par certains des participants à la Conférence Internationale sur les Systèmes Dynamiques qui s'est tenue à l'IMPA (Rio de Janeiro), en juillet 2000 pour commémorer le 60^{e} anniversaire de Jacob Palis.



CONTENTS

CONTENTS OF VOLUME II

ABSTRACTS

On the Mathematical Contributions of Jacob Palis
SHELDON NEWHOUSE 1

A Conference on Dynamical Systems celebrating the 60th birthday of Jacob Palis was held at IMPA (Instituto de Matemática Pura e Aplicada) in Rio de Janeiro from July 19-28, 2000. This article is a revised and expanded version of a lecture I gave at the Conference. Many additions, including the list of references and the entire sections below on *Homoclinic Bifurcations*, *Cantor Sets and Fractal Invariants*, *Non-Hyperbolic Systems*, and *A Unifying View of Dynamics*, were made later by Marcelo Viana. It was decided to preserve the flavor of the lecture by keeping the narrative in the first person. I am grateful to Marcelo for his contributions to this paper. In my opinion, they greatly improved the presentation of the mathematical scope and influence of Jacob Palis.

Random perturbations of nonuniformly expanding maps
JOSÉ FERREIRA ALVES & VÍTOR ARAÚJO 25

We give both sufficient conditions and necessary conditions for the stochastic stability of nonuniformly expanding maps either with or without critical sets. We also show that the number of probability measures describing the statistical asymptotic behaviour of random orbits is bounded by the number of SRB measures if the noise level is small enough. As an application of these results we prove the stochastic stability of certain classes of nonuniformly expanding maps introduced in [**Vi1**] and [**ABV**].

The minimal entropy problem for 3-manifolds with zero simplicial volume
JAMES W. ANDERSON & GABRIEL P. PATERNAIN 63

In this note, we consider the *minimal entropy problem*, namely the question of whether there exists a smooth metric of minimal (topological) entropy, for certain classes of closed 3-manifolds. Specifically, we prove the following two results.

Theorem A. *Let M be a closed orientable irreducible* 3*-manifold whose fundamental group contains a* $\mathbb{Z} \oplus \mathbb{Z}$ *subgroup. The following are equivalent:*
(1) *the simplicial volume* $\|M\|$ *of* M *is zero and the minimal entropy problem for* M *can be solved;*
(2) M *admits a geometric structure modelled on* $\mathbb{E}^3$ *or* Nil*;*
(3) M *admits a smooth metric* g *with* $\mathrm{h}_{\mathrm{top}}(g) = 0$.
Theorem B. *Let M be a closed orientable geometrizable* 3*-manifold. The following are equivalent:*
(1) *the simplicial volume* $\|M\|$ *of* M *is zero and the minimal entropy problem for* M *can be solved;*
(2) M *admits a geometric structure modelled on* $\mathbb{S}^3$, $\mathbb{S}^2 \times \mathbb{R}$, $\mathbb{E}^3$, *or* Nil*;*
(3) M *admits a smooth metric* g *with* $\mathrm{h}_{\mathrm{top}}(g) = 0$.

We prove that there is a residual set of families of smooth or analytic unimodal maps with quadratic critical point and negative Schwarzian derivative such that almost every non-regular parameter is Collet-Eckmann with subexponential recurrence of the critical orbit. Those conditions lead to a detailed and robust statistical description of the dynamics. This proves the Palis conjecture in this setting.

Geometric properties of multi-dimensional dispersing billiards are studied in this paper. On the one hand, non-smooth behaviour in the singularity submanifolds of the system is discovered (this discovery applies to the more general class of semi-dispersing billiards as well). On the other hand, a self-contained geometric description for unstable manifolds is given, together with the proof of important regularity properties. All these issues are highly relevant to studying the ergodic and statistical behaviour of the dynamics.

We study a class of Hamiltonian systems on a 4 dimensional symplectic manifold which have a saddle-center fixed point and satisfy the following property: All the periodic orbits in the center manifold of the fixed point have an orbit homoclinic to them, although the fixed point itself does not. In addition, we prove that these systems have a chaotic behavior in the neighborhood of the energy shell of the fixed point.

We describe an order on the set of scaling ratios of the generic universal smooth period doubling Cantor set and prove that this set of ratios forms itself a Cantor set, a Conjecture formulated by Coullet and Tresser in 1977. This result establishes explicitly the geometrical complexity of the universal period doubling Cantor set. We also show a convergence result for the two period doubling renormalization operators, acting on the codimension one space of period doubling maps. In particular they form an iterated function system whose limit set contains a Cantor set.

It is known that all non-hyperbolic robustly transitive sets Λ_φ have a dominated splitting and, generically, contain periodic points of different indices. We show that, for a $\mathcal{C}^1$-dense open subset of diffeomorphisms φ, the indices of periodic points in a robust transitive set Λ_φ form an interval in $\mathbb{N}$. We also prove that the homoclinic classes of two periodic points in Λ_φ are robustly equal. Finally, we describe what sort of homoclinic tangencies may appear in Λ_φ by studying its dominated splittings.

In this note examples are presented of vector fields depending on parameters and determined by the 3-jet, which display persistent occurrence of n-quasiperiodicity. In the parameter space this occurrence has relatively large measure. A leading example consists of weakly coupled Hopf bifurcations. This example, however, is extended to full generality in the space of all 3-jets.

We study dynamical systems generated by a motion of a particle in an array of scatterers distributed in a lattice. Such deterministic cellular automata are called Lorentz-type lattice gases or walks in rigid environments. It is shown that these models can be completely solved in the one-dimensional case. The corresponding regimes of motion can serve as the simple dynamical examples of diffusion, sub- and super-diffusion.

The perverse solutions of the n-body problem are the solutions which satisfy the equations of motion for at least two distinct systems of masses. I contribute

RÉSUMÉS DES ARTICLES

Une conférence sur les systèmes dynamiques s'est tenue à l'IMPA (Instituto de Matemática Pura e Aplicada) à Rio de Janeiro, à l'occasion du 60$^{\mathrm{e}}$ anniversaire de Jacob Palis, du 19 au 28 juillet 2000. Cet article est une version révisée et élargie d'un exposé que j'ai donné lors de la conférence. Plusieurs ajouts, incluant une liste de références et les paragraphes intitulés *Homoclinic Bifurcations*, *Cantor Sets and Fractal Invariants*, *Non-Hyperbolic Systems* et *A Unifying View of Dynamics*, ont été introduits plus tard par Marcelo Viana. Il a été décidé de préserver l'ambiance de l'exposé en conservant une narration à la première personne. Je remercie Marcelo pour ses contributions à cet article. À mon avis, celles-ci ont beaucoup amélioré la présentation de l'envergure mathématique et de l'influence de Jacob Palis.

Nous donnons des conditions suffisantes et des conditions nécessaires pour la stabilité stochastique de transformations non uniformément dilatantes, avec ou sans ensembles critiques. Nous prouvons aussi que le nombre de mesures de probabilité qui décrit le comportement statistique asymptotique des orbites aléatoires est borné par le nombre de mesures de SRB si le niveau de bruit est assez petit. Comme application de ces résultats nous prouvons la stabilité stochastique de certaines classes de transformations non uniformément dilatantes présentées dans [**Vi1**] et [**ABV**].

Dans cet article, nous considérons le *problème de l'entropie minimale*, c'est-à-dire la question de l'existence d'une métrique lisse d'entropie (topologique) minimale, pour certaines classes de variétés fermées de dimension 3. Précisément, nous montrons les deux résultats suivants.

Théorème A. *Soit M une variété fermée de dimension 3, orientable et irréductible, dont le groupe fondamental contient un sous-groupe $\mathbb{Z} \oplus \mathbb{Z}$. Les propriétés suivantes sont équivalentes:*
(1) *le volume simplicial $\|M\|$ de M est nul et le problème de l'entropie minimale pour M peut être résolu;*
(2) *M admet une structure géométrique modelée sur $\mathbb{E}^3$ ou* Nil*;*
(3) *M admet une métrique lisse g avec $\mathrm{h}_{\mathrm{top}}(g) = 0$.*

Théorème B. *Soit M une variété fermée de dimension 3, orientable et géométrisable. Les propriétés suivantes sont équivalentes:*
(1) *le volume simplicial $\|M\|$ de M est nul et le problème de l'entropie minimale pour M peut être résolu;*
(2) *M admet une structure géométrique modelée sur $\mathbb{S}^3$, $\mathbb{S}^2 \times \mathbb{R}$, $\mathbb{E}^3$, ou* Nil*;*
(3) *M admet une métrique lisse g avec $\mathrm{h}_{\mathrm{top}}(g) = 0$.*

Nous montrons que l'ensemble des familles d'applications unimodales telles que presque tout paramètre non régulier est Collet-Eckmann avec récurrence sous-exponentielle de l'orbite critique est résiduel. Ceci nous amène à donner une description statistique détaillée et robuste de la dynamique. Nos résultats démontrent la conjecture de Palis dans ce contexte.

Dans cet article, on étudie les propriétés géométriques des billards dispersifs multi-dimensionnels. D'une part, on découvre un comportement non régulier dans les variétés singulières du système (cette découverte concerne aussi la catégorie plus générale des billards semi-dispersifs). D'autre part, on donne une description géométrique cohérente pour les variétés instables, puis on démontre d'importantes propriétés de régularité. Toutes ces questions sont particulièrement en rapport avec l'étude du comportement ergodique et statistique de la dynamique.

On étudie une classe de systèmes hamiltoniens sur une variété symplectique de dimension 4 qui admettent un point fixe de type selle-centre et vérifient la propriété suivante: chaque orbite périodique de la variété centrale du point fixe a une orbite homocline, mais le point fixe lui-même n'a pas d'orbite homocline. On montre de plus que ces systèmes ont un comportement chaotique au voisinage de la surface d'énergie du point fixe.

Nous décrivons un ordre sur l'ensemble des rapports d'échelle de l'ensemble de Cantor du doublement de période générique universel lisse, et montrons que cet ensemble de rapports forme lui-même un ensemble de Cantor, ce qui est une conjecture formulée par Coullet et Tresser en 1977. Ce résultat établit explicitement la complexité géométrique de l'ensemble de Cantor du doublement de période universel. Nous montrons aussi un résultat de convergence pour les deux opérateurs de renormalisation du doublement de période, agissant sur l'espace de codimension 1 des applications de doublement de période.

On sait que les ensembles robustement transitifs non hyperboliques possèdent une décomposition dominée et contiennent génériquement des points périodiques de différents indices. Nous montrons que, sur une partie C^1-ouverte et dense de difféomorphismes φ, les indices des points périodiques d'un ensemble Λ_φ robustement transitif forment un intervalle dans $\mathbb{N}$. Nous montrons aussi que les classes homoclines de deux points périodiques de Λ_φ sont robustement égales. Finalement, nous décrivons les types de tangences homoclines qui peuvent apparaître dans Λ_φ, en analysant les différentes décompositions dominées de Λ_φ.

Dans cet article, on présente des exemples de champs de vecteurs dépendant de paramètres et déterminés par leur 3-jet, qui présentent une n-quasi-périodicité persistante. Dans l'espace des paramètres, ce phénomène apparaît sur un ensemble de mesure relativement grande. Les bifurcations de Hopf couplées en sont l'exemple principal. On étend cet exemple en toute généralité à l'espace de tous les 3-jets.

Nous étudions des systèmes dynamiques engendrés par le mouvement d'une particule sur un ensemble de dispersions distribuées dans un réseau. Ces automates cellulaires déterministes sont appelés gaz de réseau de type Lorentz ou marches en environnements rigides. Nous démontrons que ces modèles peuvent être complètement résolus en dimension 1. Les régimes de mouvement peuvent servir d'exemples dynamiques simples de diffusion, sous-diffusion et supra-diffusion.

Les solutions perverses du problème des n corps sont celles qui satisfont aux équations du mouvement pour au moins deux systèmes distincts de masses. Je fais quelques remarques simples sur la question de leur existence, question qui curieusement semble nouvelle.

Nous étudions des familles C^3 d'applications unimodales de l'intervalle avec une dérivée de Schwarz négative et un point critique quadratique, qui déploient transversalement une bifurcation de Misiurewicz, et nous démontrons, pour ces familles, que l'existence d'une mesure de probabilité invariante absolument continue ("chaos") et l'existence d'une renormalisation sont prévalentes en mesure dans l'espace des paramètres. D'autre part, la méthode montre aussi que l'existence d'une renormalisation est dense et le chaos a lieu avec une mesure positive.

PREFACE

These two volumes collect original research articles submitted by participants of the International Conference on Dynamical Systems held at IMPA, Rio de Janeiro, in July 19-28, 2000 to commemorate the 60th birthday of Jacob Palis.

These articles cover a wide range of subjects in Dynamics, reflecting the Conference's broad scope, itself a tribute to the diversity and influence of Jacob's contributions to the mathematical community worldwide, and most notably in Latin America, through his scientific work, his role as an educator of young researchers, his responsibilities in international scientific bodies, and the efforts he has always devoted to fostering the development of Mathematics in all regions of the globe.

His own mathematical work, which extends for more than 80 publications, is described in Sheldon Newhouse's opening article. It is, perhaps, best summarized by the following quotation from Jacob's recent nomination for the French Academy of Sciences: "sa vision, en constante évolution, a considérablement élargi le sujet".

As Jacob does not seem willing to slow down, we should expect much more from him in the years to come...

Rio de Janeiro and Paris,
May 20, 2003
Welington de Melo, Marcelo Viana, Jean-Christophe Yoccoz

Jacob Palis Mathematical Tree

Location of present institution indicated.

SCIENTIFIC WORKS OF JACOB PALIS

1. J. Palis. On Morse-Smale diffeomorphisms. *Bull. Amer. Math. Soc.*, 74:985–987, 1968.
2. J. Palis. On Morse-Smale dynamical systems. *Topology*, 8:385–404, 1968.
3. J. Palis. On the local structure of hyperbolic points in Banach spaces. *An. Acad. Brasil. Ci.*, 40:263–266, 1968.
4. M. Hirsch, J. Palis, C. Pugh, and M. Shub. Neighborhoods of hyperbolic sets. *Invent. Math.*, 9:121–134, 1969/1970.
5. J. Palis and S. Smale. Structural stability theorems. In *Global Analysis (Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968)*, pages 223–231. Amer. Math. Soc., Providence, R.I., 1970.
6. J. Palis. A note on Ω-stability. In *Global Analysis (Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968)*, pages 221–222. Amer. Math. Soc., Providence, R.I., 1970.
7. J. Palis. Ω-explosions. *Bulletin of the Brazilian Mathematical Society*, 1:55–57, 1970.
8. J. Palis. Ω-stability and explosions. In *Lecture Notes in Mathematics, 1206*, pages 40–42. Springer–Verlag, 1971.
9. J. Palis. Ω-explosions. *Proc. Amer. Math. Soc.*, 27:85–90, 1971.
10. J. Palis, editor. *Sistemas Dinâmicos (Portuguese)*. IMPA, 1971.
11. S. Newhouse and J. Palis. Hyperbolic nonwandering sets on two-dimensional manifolds. In *Dynamical systems (Proc. Sympos., Univ. Bahia, Salvador, 1971)*, pages 293–301, New York, 1973. Academic Press.
12. S. Newhouse and J. Palis. Bifurcations of Morse-Smale dynamical systems. In *Dynamical systems (Proc. Sympos., Univ. Bahia, Salvador, 1971)*, pages 303–366, New York, 1973. Academic Press.
13. J. Palis. Vector fields generate few diffeomorphisms. *Bull. Amer. Math. Soc.*, 80:503–505, 1974.

14. J. Palis, C. Pugh, and R. C. Robinson. Nondifferentiability of invariant foliations. In *Dynamical systems—Warwick 1974 (Proc. Sympos. Appl. Topology and Dynamical Systems, Univ. Warwick)*, pages 234–240. Lecture Notes in Math., Vol. 468, Berlin, 1975. Springer-Verlag.
15. J. Palis, C. Pugh, M. Shub, and D. Sullivan. Genericity theorems in topological dynamics. In *Dynamical systems—Warwick 1974 (Proc. Sympos. Appl. Topology and Dynamical Systems, Univ. Warwick)*, pages 241–250, Lecture Notes in Math., Vol. 468, Berlin, 1975. Springer-Verlag.
16. J. Palis and C. C. Pugh, editors. *Fifty problems in dynamical systems.* Proc. Sympos. Appl. Topology and Dynamical Systems, Univ. Warwick Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1975.
17. J. Palis. Arcs of dynamical systems: bifurcations and stability. Proc. Sympos. Appl. Topology and Dynamical Systems, Univ. Warwick In *Lecture Notes in Mathematics, 468*, pages 48–53. Springer–Verlag, 1975.
18. J. Palis. The Moinho Santista Prize in Mathematics. *Bol. Soc. Brasil. Mat.*, 7(2):85–87, 1976.
19. S. Newhouse and J. Palis. Cycles and bifurcation theory. *Astérisque*, 31:43–140, 1976.
20. S. Newhouse, J. Palis, and F. Takens. Stable arcs of diffeomorphisms. *Bull. Amer. Math. Soc.*, 82(3):499–502, 1976.
21. C. Camacho, N. H. Kuiper, and J. Palis. La topologie du feuilletage d'un champ de vecteurs holomorphe près d'une singularité. *C. R. Acad. Sci. Paris Sér. A-B*, 282(17):Ai, A959–A961, 1976.
22. J. Palis and F. Takens. Topological equivalence of normally hyperbolic dynamical systems. *Topology*, 16(4):335–345, 1977.
23. J. Palis. Some developments on stability and bifurcations. In *Geometry and topology (Proc. III Latin Amer. School of Math., Inst. Mat. Pura Aplicada CNPq, Rio de Janeiro, 1976)*, pages 495–509. Lecture Notes in Math., Vol. 597, Berlin, 1977. Springer-Verlag.
24. J. Palis and M. do Carmo, editors. *Geometry and topology.* Springer-Verlag, Berlin, 1977. Lecture Notes in Mathematics, Vol. 597.
25. J. Palis and W. de Melo. *Introdução aos sistemas dinâmicos.* Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1978.
26. C. Camacho, N. H. Kuiper, and J. Palis. The topology of holomorphic flows with singularity. *Inst. Hautes Études Sci. Publ. Math.*, 48:5–38, 1978.
27. J. Palis. Rigidity of the centralizers of diffeomorphisms and structural stability of suspended foliations. In *Differential topology, foliations and Gelfand-Fuks cohomology (Proc. Sympos., Pontifícia Univ. Católica, Rio de Janeiro, 1976)*, pages 114–121. Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1978.

28. J. Palis. Dynamical systems. In *Proceedings of the Tenth Brazilian Mathematical Colloquium (Poços de Caldas, 1975)*, Vol. I , pages 69–75. Inst. Mat. Pura Apl., Rio de Janeiro, 1978.
29. J. Palis. Invariantes de conjugação e módulos de estabilidade dos sistemas dinâmicos. In *Proceedings of the XI Brazilian Mathematical Colloquium*, Inst. Mat. Pura Apl., Rio de Janeiro, 1978.
30. J. Palis. A differentiable invariant of topological conjugacies and moduli of stability. *Astérisque*, 51:335–346, 1978.
31. J. Palis. Moduli of stability and bifurcation theory. In *Proceedings of the International Congress of Mathematicians (Helsinki, 1978)*, pages 835–839, Helsinki, 1980. Acad. Sci. Fennica.
32. W. de Melo and J. Palis. Moduli of stability for diffeomorphisms. In *Global theory of dynamical systems (Proc. Internat. Conf., Northwestern Univ., Evanston, Ill., 1979)*, pages 318–339. Springer-Verlag, Lecture Notes, Berlin, 1980.
33. W. de Melo, J. Palis, and S. J. van Strien. Characterising diffeomorphisms with modulus of stability one. In *Dynamical systems and turbulence, Warwick 1980 (Coventry, 1979/1980)*, pages 266–285. Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1981.
34. I. P. Malta and J. Palis. Families of vector fields with finite modulus of stability. In *Dynamical systems and turbulence, Warwick 1980 (Coventry, 1979/1980)*, pages 212–229. Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1981.
35. J. Palis and W. de Melo. *Geometric theory of dynamical systems.* Lecture Notes in Mathematics, Springer-Verlag, New York, 1982. An introduction, Translated from the Portuguese by A. K. Manning.
36. S. Newhouse, J. Palis, and F. Takens. Bifurcations and stability of families of diffeomorphisms. *Inst. Hautes Études Sci. Publ. Math.*, 57:5–71, 1983.
37. J. Palis, editor. *Geometric dynamics*, Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1983.
38. J. Palis. A note on the inclination lemma (λ-lemma) and Feigenbaum's rate of approach. In *Geometric dynamics*, pages 630–635. Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1983.
39. J. Palis and F. Takens. Stability of parametrized families of gradient vector fields. *Ann. of Math. (2)*, 118(3):383–421, 1983.
40. J. Palis. The dynamics of a diffeomorphism and the rigidity of its centralizer. In *Singularities and dynamical systems (Iráklion, 1983)*, pages 15–21. North-Holland, Amsterdam, 1985.
41. J. Palis and R. Roussarie. Topological invariants as numbers of translation. In *Dynamical systems and bifurcations (Groningen, 1984)*, pages 64–86. Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1985.

42. J. Palis and F. Takens. Cycles and measure of bifurcation sets for two-dimensional diffeomorphisms. *Invent. Math.*, 82(3):397–422, 1985.
43. Zh. Palis and V. di Melu. *Geometricheskaya teoriya dinamicheskikh sistem.* "Mir", Moscow, 1986. Vvedenie. [An introduction], Translated from the English by V. N. Kolokoltsov, Translation edited and with notes and an afterword by D. V. Anosov.
44. J. Palis. Homoclinic orbits, hyperbolic dynamics and dimension of Cantor sets. In *The Lefschetz centennial conference, Part III (Mexico City, 1984)*, pages 203–216. Amer. Math. Soc., Providence, R.I., 1987.
45. J. Palis. Dimensões fracionárias de conjuntos de Cantor e dinâmica hiperbólica. In *Proceedings of the XV Brazilian Mathematical Colloquium*, pages 341–353, 1987.
46. J. Palis and F. Takens. Hyperbolicity and the creation of homoclinic orbits. *Ann. of Math. (2)*, 125(2):337–374, 1987.
47. J. Palis and F. Takens. *Homoclinic bifurcations and hyperbolic dynamics. Proceedings of the XV Brazilian Mathematical Colloquium*, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1987.
48. J. Palis. On the solution of the C^1 stability conjecture (Mañé) and Ω-stability conjecture. In *Proceedings of the XVI Brazilian Mathematical Colloquium*, pages 599–606, 1988.
49. J. Palis and M. Viana. On the continuity of Hausdorff dimension and limit capacity for horseshoes. In *Dynamical systems, Valparaiso 1986*, pages 150–160. Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1988.
50. R. Bamón, R. Labarca, and J. Palis, editors. *Dynamical systems*, volume 1331 of *Lecture Notes in Mathematics*, Berlin, 1988. Springer-Verlag.
51. J. Palis. On the C^1 Ω-stability conjecture. *Inst. Hautes Études Sci. Publ. Math.*, 66:211–215, 1988.
52. J. Palis. On the solution of the stability conjecture and the Ω-stability conjecture. In *IXth International Congress on Mathematical Physics (Swansea, 1988)*, pages 469–471. Hilger, Bristol, 1989.
53. J. Palis and J.-C. Yoccoz. Rigidity of centralizers of diffeomorphisms. *Ann. Sci. École Norm. Sup. (4)*, 22(1):81–98, 1989.
54. J. Palis and J.-C. Yoccoz. Centralizers of Anosov diffeomorphisms on tori. *Ann. Sci. École Norm. Sup. (4)*, 22(1):99–108, 1989.
55. J. Palis. Homoclinic bifurcation and fractional dimensions. *Publicaciones Matematicas del Uruguay*, 1:55–66, 1989.
56. J. Palis. Gradient flows. In *Stability theory and related topics in dynamical systems (Nagoya, 1988)*, pages 142–144. World Sci. Publishing, Teaneck, NJ, 1989.

57. J. Palis. Sistemas caóticos ou turbulentos, atratores e bifurcações homoclínicas (Portuguese). *Revista Matemática Universitária, Brazilian Mathematical Society*, 10:167–192, 1989.
58. M. J. Dias Carneiro and J. Palis. Bifurcations and global stability of families of gradients. *Inst. Hautes Études Sci. Publ. Math.*, 70:103–168 (1990), 1989.
59. J. Palis. Centralizers of diffeomorphisms. In *Workshop on Dynamical Systems (Trieste, 1988)*, pages 19–22. Longman Sci. Tech., Harlow, 1990.
60. J. Palis and J.-C. Yoccoz. Differentiable conjugacies of Morse-Smale diffeomorphisms. *Bol. Soc. Brasil. Mat. (N.S.)*, 20(2):25–48, 1990.
61. J. Palis. Homoclinic bifurcations, sensitive-chaotic dynamics and strange attractors. In *Dynamical systems and related topics (Nagoya, 1990)*, pages 466–472. World Sci. Publishing, River Edge, NJ, 1991.
62. J. Palis. A glimpse at dynamical systems: the long trajectory from the sixties to present developments. In *Proceedings of the TWAS*, 1992. Prize talk at the Third World Academy of Sciences.
63. J. Palis. New developments in dynamics: hyperbolicity and chaotic dynamics. In *Chaos, resonance and collective dynamical phenomena in the solar systems*, pages 363–369. International Astron. Union, Kluwer Acad. Publ., 1992.
64. R. Bamón, R. Labarca, J. Lewowicz, and J. Palis, editors. *Dynamical systems*, volume 285 of *Pitman Research Notes in Mathematics*, Harlow, 1993. Longman Scientific & Technical.
65. J. Palis and F. Takens. *Hyperbolicity and sensitive chaotic dynamics at homoclinic bifurcations*. Cambridge University Press, Cambridge, 1993. Fractal dimensions and infinitely many attractors.
66. J. Palis. On the contribution of Smale to dynamical systems. In *From Topology to Computation: Proceedings of the Smalefest (Berkeley, CA, 1990)*, pages 165–178, New York, 1993. Lecture Notes in Mathematics, Springer-Verlag, .
67. J. Palis and J.-C. Yoccoz. Homoclinic tangencies for hyperbolic sets of large Hausdorff dimension. *Acta Math.*, 172(1):91–136, 1994.
68. J. Palis and M. Viana. High dimension diffeomorphisms displaying infinitely many periodic attractors. *Ann. of Math. (2)*, 140(1):207–250, 1994.
69. J. Palis. A view on chaotic dynamical systems. *Brazilian Journal of Physics*, 24:926–930, 1994.
70. J. Palis. Chaotic and complex systems. *Science International*, 58, November:27–31, 1995.
71. J. Palis. A view on chaotic dynamical systems. In *Dynamical systems and chaos, Vol. 1 (Hachioji, 1994)*, pages 217–225. World Sci. Publishing, River Edge, NJ, 1995.

72. J. Palis. From dynamical stability and hypebolicity to finitude of ergodic attractors. In *Proceedings of the Third World Academy of Sciences,* 11^{th} *General Conference, Italy*, 1996.
73. J. Palis and J.-C. Yoccoz. On the arithmetic sum of regular Cantor sets. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 14(4):439–456, 1997.
74. J. Palis. Sistemas caóticos e sistemas complexos (Portuguese). In *Caos e complexidade*, pages 27–38. Editora UFRJ, 1999.
75. J. Palis. Uncertainty-chaos in dynamics, a global view. In *Proceedings of* 10^{th} *General Meeting*, pages 33–38, 1999. Medal Lecture Third World Academy of Sciences.
76. J. Palis. A global view of dynamics and a conjecture on the denseness of finitude of attractors. *Astérisque*, 261:xiii–xiv, 335–347, 2000. Géométrie complexe et systèmes dynamiques (Orsay, 1995).
77. J. Palis. Homoclinic bifurcations: from Poincaré to present time. *The mathematical sciences after the year 2000*, World Scientific, (Beirut 1999), 123-134, 2000.
78. J. Palis and J. C. Yoccoz. Nonuniformily hyperbolic horseshoes unleashed by homoclinic bifurcations and zero density of attractors. *C. R. Ac. Sc. Paris*, 333 (2001), 867–871.
79. C. G. Moreira, J Palis, and M. Viana. Fractal invariants and homoclinic bifurcations in arbitrary dimension. *C. R. Ac. Sc. Paris*, 333 (2001), 475–480.
80. J. Palis and J. C. Yoccoz. Implicit formalism for affine-like maps and parabolic compositions. *Global analysis of dynamical systems*, Inst. Phys. Bristol, 67–87, 2001.
81. J. Palis. Chaotic and complex systems. *Current Science*, vol.82, 4 (2002), 403–406.

Astérisque
286, 2003, p. 1–24

ON THE MATHEMATICAL CONTRIBUTIONS OF JACOB PALIS

by

Sheldon Newhouse

Abstract. — A Conference on Dynamical Systems celebrating the 60th birthday of Jacob Palis was held at IMPA (Instituto de Matemática Pura e Aplicada) in Rio de Janeiro from July 19-28, 2000. This article is a revised and expanded version of a lecture I gave at the Conference. Many additions, including the list of references and the entire sections below on *Homoclinic Bifurcations*, *Cantor Sets and Fractal Invariants*, *Non-Hyperbolic Systems*, and *A Unifying View of Dynamics*, were made later by Marcelo Viana. It was decided to preserve the flavor of the lecture by keeping the narrative in the first person. I am grateful to Marcelo for his contributions to this paper. In my opinion, they greatly improved the presentation of the mathematical scope and influence of Jacob Palis.

Introduction

Let me begin just by saying that Jacob has made many, many contributions to Mathematics. I will not talk about all of them because, in fact, in one hour it's impossible to discuss in any detail all of them. I pick some of what I consider to be the main contributions, and there will be relatively little that is new for experts, but I hope you will be reminded of many experiences during the last thirty or some years of the development of Dynamical Systems.

First, to my mind his primary mathematical contributions fit into three categories:

– global stability related to the concepts of structural stability and Ω-stability;

– bifurcation theory, which is how systems depending on parameters change, how their structure changes.

– formulation of some general ideas and conjectures, that motivated several very interesting recent results in this field.

I will talk about these aspects of his work a little bit later. Besides these types of subjects there are many other ancillary results, many interesting kinds of things.

2000 ***Mathematics Subject Classification.*** — 37Dxx.
Key words and phrases. — Hyperbolic, Morse-Smale, structural stability, Axiom A, bifurcation theory, homoclinic tangency, tubular families, Ω-stability, stability conjecture.

But, together with the mathematical contributions that he has been making, one has to appreciate and understand the overview and direction of research that Jacob is responsible for. At the present time he is at

– 35 graduate students, and some 30 grand-students, originating from 10 different countries mainly in Latin America, as you can see in his academic tree (attached to this paper).

Some of these students have become main figures in the whole theory of Dynamical Systems, in fact in the world of Mathematics. You know who they are as well as I do, I don't need to mention names. It's a testimony to his vision, his generosity, and the freedom of ideas that he's encouraged, that he is such an inspiration to so many people.

In addition, I think it's really fair to say that in our time Jacob Palis has been one of the main figures responsible for the development of Mathematics and Science, primarily in Latin America[(1)] and, in fact, in many other places, through his

– organization of meetings, symposia, workshops, and the support of sciences and Mathematics in developing countries, most notably, that I'm familiar with, in Trieste.

He has facilitated the contacts between scientists who have had great difficulty in traveling to the west for political or other reasons. They were able to establish contacts with western mathematicians in the settings of meetings, workshops, and schools where one can get to meet many people. I myself met a number of people from mainland China in Trieste, at a time when it was extremely difficult for them to travel to Western Europe. Jacob has been one of the primary organizers and supporters of such occasions.

Moreover, he has been responsible, in great measure, for

– the tremendous growth of IMPA, this wonderful institute, as a researcher and, more recently, also as the Director.

I think it's fair to say that IMPA has become the principal center for Mathematics in Latin America and, certainly, one of the world centers for Dynamical Systems. In no small measure is this due to his efforts and, again, his vision.

I want to go now toward some of the mathematical developments Jacob has accompanied in his many years of activity.

Structural Stability

Let me go back to 1960. Let M be a compact connected smooth manifold without boundary, and let us consider the space $\mathcal{D}^r(M)$ of C^r diffeomorphisms on M, and the

[(1)] The impact of Jacob Palis's work throughout Latin America was the subject of another lecture at the Conference, by Alberto Verjovsky.

space $\mathcal{X}^r(M)$ of C^r vector fields on M, as well as certain distinguished well-known subsets of these

$\mathcal{D}^r_{ss}(M)$ = set of C^1 structurally stable diffeomorphisms on M,
$\mathcal{X}^r_{ss}(M)$ = set of C^1 structurally stable vector fields on M.

This notion of *structural stability* means that under any small C^1 perturbation, the entire orbit structure persists after a global continuous coordinate change. As far as I know, it was first presented by Andronov and Pontrjagin in 1937. They introduced these systems, that they called rough systems, or coarse systems, and the primary part of the paper [**2**] was to characterize them among vector fields in the two dimensional disk which were nowhere tangent to the boundary. And what they described in that paper was that a vector field X is structurally stable if and only if

(a) X has only finitely many critical elements (singular points and periodic orbits), all hyperbolic,

(b) and there are no saddle connections.

The next principal result connected to structural stability we will mention was due to Maurício Peixoto in a paper [**53**] that was published in 1959. There, he studied various general properties of structurally stable systems and proved that the Andronov-Pontrjagin systems formed an open and dense subset of the set of all vector fields on the two dimensional disk which were nowhere tangent to the boundary. Later, in [**54**], in a somewhat surprising way, he proved the following theorem: on a compact oriented surface M^2,

– the structurally stable vector fields $\mathcal{X}^r_{ss}(M^2)$ form a dense open set in the space $\mathcal{X}^r(M^2)$ and

– they are completely characterized by the Andronov-Pontrjagin conditions (a) and (b), and the additional condition that the α- and ω-limit sets of every point x are critical elements.

As far as I know, originally this paper was thought to prove that the result is true for all surfaces (not necessarily orientable), but that's still not known, except in the case of genus two, where Carlos Gutierrez [**18**] proved the general result, and in the C^1 topology, where it is a consequence of Pugh's closing-lemma [**56**].

This led to two main questions at the time:

– Is $\mathcal{X}^r_{ss}(M)$ non-empty, that is, do structurally stable systems exist on any manifold?

– Is $\mathcal{X}^r_{ss}(M)$ always dense in the space $\mathcal{X}^r(M)$ of all vector fields?

Also the analogous questions for C^r diffeomorphisms on compact manifolds.

Well, to some people's disappointment, the second question, the density, has a negative answer. That was discovered by Smale around 1964 or 65. He found out that on any manifold in dimension bigger than or equal to 4 there were open sets of vector fields which were not structurally stable. That dimension was then made optimal by Bob Williams in the end of the 60's [**68**]: he found more detailed versions of Smale's theorem, and a counter-example in dimension 3.

Around the same time, in the 60's, in the Soviet Union, Anosov studied other kinds of structurally stable systems. The systems that he called C-diffeomorphisms [**3**], where the entire space had a splitting into two continuous distributions invariant by the derivative, one of which was exponentially expanded and the other exponentially contracted under iterates. These systems, now well known, were coined the name Anosov diffeomorphisms by Smale in his 1967 paper [**65**] in the Bulletin of the AMS. What Anosov was able to to prove for these systems was that

- they formed an open subset of the set of all C^1 diffeomorphisms on a manifold
- and they were structurally stable systems.

The methods were related (I don't know, in fact, in which order) to his celebrated result that geodesic flows on manifolds with negative curvature were structurally stable and had the flow version of these Anosov conditions.

At this time, in the mid 60's, what was then the status of this kind of mathematics? We had high dimensional examples of structurally stable systems. They exhibited very complicated recurrence, and they were only known in special manifolds. In fact, for the Anosov systems the existence of the invariant bundles of course brings with it topological obstructions. So, for example on surfaces, Anosov diffeomorphisms only exist on the torus. And in higher dimensions, also only on very special manifolds. In fact, for a while it was felt that the only manifolds that admitted Anosov diffeomorphisms were the tori, of any dimension. Smale found examples using other kinds of Lie groups, non-Abelian Lie groups, but still they were very special in the kinds of manifolds that can exhibit them.

What about simple recurrence, that is, systems that don't have complicated recurrent orbits? Motivated by gradient systems, which Smale sort of used for going back and forward between dynamical systems and topology, a special class of dynamical systems, which we now call Morse-Smale systems, was defined. In the diffeomorphism case, these are systems where the non-wandering set consists of a finite number of hyperbolic periodic orbits, and if you have two such orbits their stable and unstable manifolds are transverse. Analogous definitions were given for vector fields, where the non-wandering set consists of finitely many critical points and periodic orbits all hyperbolic, and with the transversality conditions.

Smale was able to prove that there was a residual set of gradient systems (a residual set of functions) on any compact manifold that were Morse-Smale, and their time-one maps were Morse-Smale diffeomorphisms. The easy part of this is to realize that a Morse function has only hyperbolic critical points as its non-wandering set. But it's not so obvious to get the transversality condition: that is a consequence of a general approximation theorem, the Kupka-Smale theorem, which was done in those days. And Smale conjectured that,

– Morse-Smale systems form an open set in the space of all dynamical systems, both $\mathcal{D}^r(M)$ and $\mathcal{X}^r(M)$

– and every Morse-Smale system is structurally stable.

And then, in a remarkable result in 1967, in his thesis [**38**] Jacob Palis proved that the first statement, the openness statement, held in general. And he proved the second statement, that Morse-Smale systems were structurally stable, for any systems, diffeomorphisms and vector fields, in dimension less or equal to 3.

A Geometric Approach

To indicate some of the difficulties which Jacob had to overcome in proving this theorem, let's take a simple example of a Morse-Smale diffeomorphism on the 2-sphere as indicated in Figure 1, where we have six fixed points as the non-wondering set. The

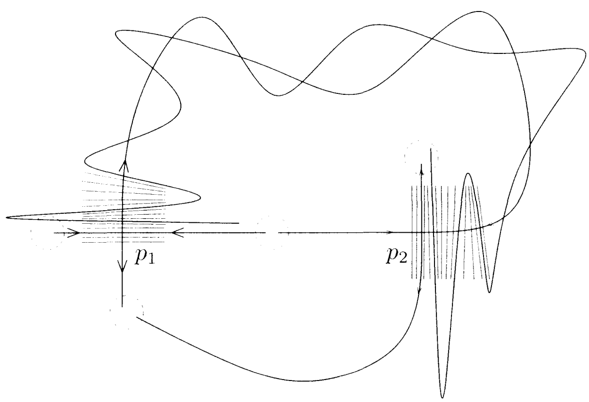


FIGURE 1. Tubular families

circles represent sources and sinks, and we have two saddle points, I denote p_1 and p_2, such that the unstable manifold of p_1 has some transverse intersection, a heteroclinic saddle connection, with the stable manifold of p_2.

Well, it was known earlier that there was a local stability phenomenon for hyperbolic fixed or periodic points, the Grobman-Hartman theorem. Locally, the system can be topologically linearized, that is, on a neighborhood of each periodic point the map is topologically conjugate to its derivative at the periodic point. But you need to do much more to get a global conjugacy, of course, you have to preserve stable and unstable manifolds globally. And orbits near the saddle points in the past go near the sources, and in the future go near the sinks. So, to have some conjugacy between a system like this and its perturbation it's not enough to look at local pictures, you have to glue them together in a special way. And the gluing is not obvious at all, because the local linearizations are very special, so how you glue this in some compatible way was a major problem.

And here there was the first major development that Palis came up with, which were the so-called tubular families, or invariant foliations, that I'll describe in some detail. They turned out to be very important for many later developments, as we'll see. These were invariant foliations defined in a neighborhood of each periodic point, one family for the stable direction and another for the unstable direction, and they were compatible: if two leaves from different periodic points intersect, then one contains the other. The construction of this is not at all obvious, it's still technically quite difficult — a very intricate geometric construction. The tubular families have different dimensions, in general. And the intricacies of this construction is what forced the restriction to dimension 3 in Jacob's thesis, the higher dimension analogue only came later.

In particular, initially it was thought that topological questions would arise in this connection, since one has to extend maps defined on certain subsets to bigger sets. It was thought that the annulus conjecture, a major unsolved problem at the time, was related to the higher dimension analogue of this tubular families method. Well, I'm not sure about the exact details of how these problems were overcome, but together with Smale in 1968 or 69, the general construction of tubular families was given, and the general structural stability of Morse-Smale systems in any dimension was proved [**42**].

It's important to notice that there is a lot of freedom in the construction of these tubular families. The conjugacies are not unique. The existence of invariant manifolds covering the whole manifold was crucial to Anosov in his treatment of structural stability. Those invariant manifolds are unique, and so the conjugacies, if they are near the identity, are unique for Anosov systems. Here they are highly non-unique, and in fact the flexibility of the choice is very much related to the freedom one has in the construction of tubular families. So this was a major breakthrough at the time and still is, in my opinion, a major contribution, that came quite early in his career.

This had two main corollaries. The first one was that

– an open dense subset of the set of gradient systems on any manifold consists of structurally stable vector fields;

Even more, the time-one maps of such vector fields are structurally stable diffeomorphisms. That's much stronger. Indeed, as we know, the usual equivalence relation for vector fields is homeomorphisms taking orbits to orbits. A stronger equivalence relation is conjugacy, actual one parameter group conjugacy. And structural stability for the time-one maps gives stability under this stronger equivalence relation, for gradient flows. So, as an extension of this, the problem of the existence of structural stability was solved in a very positive way:

– every manifold has structurally stable vector fields and diffeomorphisms.

The Stability Conjectures

Around this time, in the late 60's, having proved that structurally stable systems are not dense, Smale was looking for a more general kind of system, that would still have some good structure and have the chance to form a dense subset in the space of all dynamical systems. And so he formulated what was called the Ω-stability theorem.

Our system is Ω-stable if when you take a C^1 perturbation of it you have a conjugacy from the non-wandering set of the first system to the non-wandering set of the second one (not a global conjugacy on the whole manifold, as in the definition of structural stability). He studied special systems, the so-called Axiom A diffeomorphisms, where the non-wandering sets are hyperbolic sets, and the periodic points are dense in the non-wandering set. He also assumed an additional property, the no-cycle property, that gives the ability to construct so-called filtrations for the system, that is, to isolate the recurrent orbits in individual indecomposable pieces. And he proved the theorem that Axiom A and the no-cycle property implied that the diffeomorphism was Ω-stable.

Around the same time, Jacob proved that if you have an Axiom A system and it has a cycle, then it is not Ω-stable. And that led to the Stability Conjectures, which were also present in the Palis and Smale paper of 1969 [**42**]:

(1) a diffeomorphism $f \in \mathcal{D}^r(M)$ is structurally stable if and only if it satisfies the Axiom A and the so-called strong transversality condition: stable and unstable manifolds are in general position at each point wherever they meet;

(2) and $f \in \mathcal{D}^r(M)$ is Ω-stable if and only if it satisfies the Axiom A and the no-cycle property.

And they made analogous conjectures for flows.

Let me mention a little personal anecdote in connection with this theorem and the formulation of these conjectures. For those who were around that time, you remember that the first formulation of the Ω-stability theorem had another stronger condition, called Axiom B. Axiom B said that if you have two basic sets and the unstable manifold of one accumulates on the other, then there is a periodic point in the first whose unstable manifold has a transversal intersection with the stable manifold of the other. And the first formulation of the Ω-stability theorem, in fact the formulation that is in the Bulletin paper [**65**], says: Axiom A plus Axiom B implies Ω-stability, or something to that effect.

I remember Smale giving a lecture in the seminar in Berkeley in 1966 or maybe 1967. I was a new graduate student just sort of going to this seminar from time to time, but it was a very active and energetic seminar, many questions, comments, discussions. I remember Charles Pugh was there, and Mike Shub, Morris Hirsch, Jacob Palis. As a young graduate student we look around at all those famous people in the room, and just watch what they were doing. Well, Terry Wall had just come in from England and was interested, so he went to the seminar. In fact, he was under jet-lag so he was asleep in a large part of the talk. So, Smale was doing the construction of the local conjugacy of the Ω-stability for the basic sets. Then, with Axiom B, he constructed this partial order on the basic sets, and hence a filtration to isolate each piece, so that one can get the global conjugacy. And, suddenly, Terry woke up and looked and said, quietly: "Is all you need, the partial order relation, in order to get the stability?" This was an agitated seminar with many people. Steve turned and said: "Well, maybe, I'm not sure about that, I'm not sure."

At that instant, I didn't know who Jacob Palis was, but he became very animated and said: "That's right, that's it, that is all you need!" And the next day, as I recall, he proved that if you had a cycle then you had Ω-explosions, and so, in fact, this no-cycle condition was necessary for stability. Later on, in the paper that actually appears in the proceedings of the symposium [**42**], you see Axiom A and no-cycle condition, not Axiom A and Axiom B, Axiom B disappeared. So, as part of this discussion, Jacob had a significant part in the formulation of the Ω-stability theorem as it now sits.

From Hyperbolicity to Stability

How does one go beyond toward more general stability theorems and proving these conjectures? What did people know at that time? They knew that the Morse-Smale systems were structurally stable. They knew that Axiom A and no-cycle property implies Ω-stability. How does one to get more general structurally stable systems? One idea at the time was to take Jacob's tubular family construction and extend it to Axiom A systems. That is, to get an invariant foliation on neighborhoods of complicated hyperbolic sets. It turned out to be quite a complicated thing to do and, in fact, this is still not known in general, it's not known how to do that for high dimensional systems. But that program did succeed for two-dimensional diffeomorphisms, with the thesis of Welington de Melo in 1971.

The next progress came in what might seem a curious way. Jürgen Moser gave a second proof of the stability of Anosov systems, using the so-called infinitesimal methods. His idea was the following: you want to solve the equation $h \circ f = g \circ h$ for a homeomorphism h. You rewrite this as

$$f^{-1} \circ h \circ f = f^{-1} \circ g \circ h.$$

Then you take a Riemannian metric on your manifold, and try to find h as the exponential of some continuous vector field v, which should be C^0-small so that the homeomorphism is close to the identity. So, writing $h = \exp(v)$, and also $f^{-1} \circ g = \exp(w)$ for a C^1-small vector field w, you get

$$f^{-1} \circ \exp(v) \circ f = \exp(w) \circ \exp(v).$$

Linearizing this equation (or using infinitesimal methods, which is the term I use), you get

$$\exp(Df^{-1} \circ v \circ f) = \exp(w + v),$$

up to a small error. So, taking $\exp^{-1}$ in the previous relation, it becomes

$$Df^{-1} \circ v \circ f + s(v, w) = w + v,$$

where $s(v, w)$ is small. Denoting $F\,v = Df^{-1} \circ v \circ f$, this may be rewritten as

$$(\mathrm{I} - F)v = v - Df^{-1} \circ v \circ f = s(v, w) - w.$$

So, we know w, which is a C^1-small vector field, and we are looking for v, a small continuous vector field. Moser realized that if you could invert this operator $(\mathrm{I} - F)$ on the space of continuous vector fields, then you could solve this functional relation for v, using the contraction mapping theorem. And, in fact, the Anosov condition was precisely the condition you need to make $(\mathrm{I} - F)$ invertible.

So, he was able to give a new proof of the stability of Anosov systems using vector field methods, infinitesimal methods, whereas Anosov's proof made strong use of the existence of integral manifolds for the expanding and contracting distributions, the stable and unstable manifolds. Well, at the time this was interesting because it made Anosov's proof understandable to people in the West, there was no published English version of it. And also I think it was thought of as a useful addition, a curious new proof of a known result. One thing that came out of it is that you get unique solutions near the the identity, which you can also prove by other methods.

There is an other development that I should mention. In the group of people who were in Berkeley and in the West at the time, the way that Moser's methods became known was through an implicit function theorem argument that John Mather produced. It turned out that, in detail, Mather's argument was actually incorrect, because differentiability assumptions were not satisfied. What the method gave you was a continuous solution to the functional equation, it didn't prove that the solution was a homeomorphism. But the arguments could be fixed up. I think it was Mike Shub who observed, and was well-known in the Soviet Union as well, that Anosov systems were expansive, and you can use that to show that solutions which are C^0-close to the identity actually have to be one-to-one. So you got the proof anyway, even if the implicit function theorem didn't work.

Far away, in the middle of the United States, Joel Robbin was learning about those things, and I think he shocked everybody by announcing that he could prove that, in the C^2 case, Axiom A diffeomorphisms satisfying the strong transversality condition are structurally stable. Well, how did he do it? He used infinitesimal adaptations of the tubular families constructions. Basically, the conjugacies were not unique, they involved choices, and he used the fact that Moser's transformation $(\mathrm{I} - F)$ had a continuous right inverse. You can see Jacob's influence again, even at that level: at the end of the paper [**60**] there's a ratio that says

$$(\text{Moser}) : (\text{Anosov}) = (\text{Robbin}) : (\text{Palis} - \text{Smale}).$$

The idea being that Moser produced an infinitesimal proof of the structural stability, removing the necessity of integrating the invariant subbundles for the construction, and Robbin produced an infinitesimal proof for Axiom A systems, removing the necessity of tubular families.

For technical reasons Robbin needed the C^2 assumption, not for the perturbations, but for the original diffeomorphisms. That was ultimately improved by Clark Robinson, who proved the general structural stability theorem, that Axiom A C^1 diffeomorphisms satisfying the strong transversality condition are structurally stable [**62**], and he also proved it for the vector field case [**61**]. Concerning Ω-stability, in Smale's paper [**65**] where he proves his Ω-stability theorem, he makes the statement that, presumably, similar methods can be used for flows. It was a highly non-trivial extension required to do it for flows, and it was carried out by Charles Pugh and Mike Shub [**57**]. So, at this stage, which I suppose is the mid-70's, we had general sufficient conditions for structural stability and Ω-stability, both for diffeomorphisms and for flows.

From Stability Back to Hyperbolicity

Remember the stability conjecture had a converse as well. So there was a lot of activity focussed on the converse. The initial efforts involved changing the definition of stability, to include conditions about dependence of the solution on the perturbation (whether it is continuous, whether is Lipschitz), and a number of people contributed with interesting works in that direction. John Franks [**14**] had a notion of time-dependent stability, with which he was able to characterize Axiom A and strong transversality systems. John Guckenheimer [**16**] had a notion of absolute stability, and so on. And then the full problem itself was treated in some special cases in low dimensions, by Liao [**21**], Mañé [**23, 24**], Pliss [**55**], and Sannami [**64**].

But the major breakthrough came in 1986, when Ricardo Mañé, one of Jacob's early graduate students, completely solved the problem! He proved what was the main remaining part, that is, that structurally stable systems had to satisfy the Axiom A [**25**].

Curiously enough, although this is a substantial result which uses much information about the non-wandering set, Ricardo was not able to prove the Ω-stability converse, he only proved the structural stability statement. It took some other intricate knowledge, and a fair amount of effort, for Jacob to prove that converse, and so complete the Ω-stability conjecture for diffeomorphisms, again around 1986. For the flow case, neither of the statements was known at the time, they were resolved only recently, by Shuhei Hayashi [**19**] in 1994.

So, in the development of this very important concept and theory, a period of almost 25 years was needed to accomplish what is now one of the crown jewels in the field of Dynamical Systems, the complete characterization of structurally stable systems. And as you saw, Jacob Palis played a very central role in that.

That's what I wanted to say about stability, the global stability issue. Now I want to go toward bifurcation theory.

Bifurcation Theory

In 1970 or so, I had the privilege to come to IMPA for two years, and to begin our program in bifurcation theory with Jacob. We started to work on the problem of understanding the structure of how hyperbolicity breaks down when you start with a Morse-Smale system. Basically, what we wanted to study was the so-called accessible part of the boundary of the Morse-Smale systems. The idea is the following. Let $\{\xi_\mu\}_\mu$ be an arc (a curve) of diffeomorphisms starting at a Morse-Smale system ξ_0. See Figure 2. You look at the first value $\mu = b$ of the parameter where the system

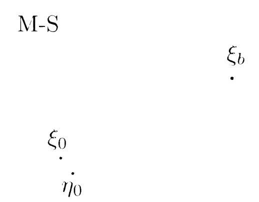


FIGURE 2. Bifurcations along parametrized families

fails to be structurally stable, the so-called first bifurcation point, and you want to describe the structure of such systems ξ_b.

Some ideas and problems were motivated by work done by Jorge Sotomayor [**66**] for one-parameter families of vector fields on surfaces, and also by a general periodic point description for one-parameter families of diffeomorphisms, which was obtained by Pavel Brunovsky [**6**]. In addition, there were mathematicians in the Soviet Union studying similar problems, Gavrilov and Shilnikov [**15**], although we didn't know that at the time, we only became aware of their work somewhat later.

During that period I wrote two papers with Jacob, [**35**] and [**36**], in which we basically proved the following. Assuming that at the first bifurcation point the limit set (the closure of the α- and ω-limit sets of the system) consists of a finite number of orbits, we completely described the structure at the bifurcation for generic arcs of diffeomorphisms. We also studied other issues related to stability as you move along the parameter, that I'll talk a bit more about later. But the main contents of the first paper [**35**] was this description at the bifurcation in the case when the limit set has finitely many orbits.

In the second paper [**36**] we considered systems where at the bifurcation point the limit set was actually hyperbolic, it stayed hyperbolic, but structural stability or Ω-stability failed all the same, because of the creation of a cycle. We studied the situation where the cycle was equidimensional, that is, the stable manifolds of all the periodic points in the cycle have the same dimension. We were able to prove that in that situation the bifurcation map ξ_b was accumulated by Axiom A, non Morse-Smale diffeomorphisms. That is,

– there existed parameter values $\mu_1 > \mu_2 > \cdots > \mu_i > \cdots$ converging to the first bifurcation point b, such that the diffeomorphisms ξ_{μ_i} satisfied the Axiom A and the strong transversality condition, and the non-wandering sets $\Omega(\xi_{\mu_i})$ were infinite.

Moreover, the non-wandering sets were all topologically distinct, so that ξ_{μ_i} could not be Ω-conjugate to each other. In fact, we proved that ξ_μ satisfies the Axiom A and the strong transversality condition for most parameters $\mu > b$ near b, in the sense that such parameters are a fraction close to 1, in measure, of small intervals $(\mu, \mu + \varepsilon)$.

Later, in a paper with Floris Takens and Jacob [**37**], we completely characterized the so-called stable arcs of diffeomorphisms, under the assumption that the limit set have finitely many orbits for each parameter value. An arc $\{\xi_\mu\}_\mu$ of diffeomorphisms is called stable if, given any perturbation $\{\eta_\mu\}_\mu$, as represented in Figure 2, then

(1) every diffeomorphism ξ_μ in the arc is conjugate to a diffeomorphism η_ν in the perturbed one, with a nearby parameter ν,

(2) and the conjugacy varies continuously with the parameter.

That's the condition of stability for arcs of diffeomorphisms. In [**37**] we characterized this condition and, as part of that work, a number of new concepts and ideas were introduced. In particular, a notion of rotation interval for circle endomorphisms was introduced. Strong rigidity for saddle-node bifurcations also came up in this work. One consequence of this strong rigidity phenomenon for saddle-node bifurcations is that the strong-stable and strong-unstable manifolds have to be preserved under conjugacy that varies continuously with the parameter (in general, topological conjugacies don't preserve strong-stable and strong-unstable manifolds).

Then these works were extended in a very significant way by Palis and Takens [**43**], who proved in 1983 that

– an open dense set of one-parameter families of gradients systems on any manifold were stable in the sense I've just described (continuous variation of the conjugacy with the parameter).

And somewhat later, in 1990, Mário Jorge Dias Carneiro and Jacob [**8**] proved that one can extend that to two-parameter families: an open and dense subset of families of gradient systems depending on two parameters are stable.

One might have hoped, in fact the hope around that time and earlier was that k-parameter families of gradient systems in a dense open set would be stable. That was shown to be false by Takens, who proved that for 8 or more parameters the stable families of gradient systems are not dense. I don't know how far down one has got yet, I think the conjecture still is that for k less than or equal to 4 the stable families should form an open and dense subset in the space of gradient systems.

In these constructions, the geometric freedom of tubular families and how you bring them up is, again, of fundamental importance. It's interesting to point out that at the time people discussed whether infinitesimal maps could be used for this theorems, but, as far as I know, they never managed to work. So far, infinitesimal methods have only been useful for the general structural stability theorem.

Homoclinic Bifurcations

Bifurcation theory continued to be one of Jacob's major projects during the 80's and afterwards. Initially, the goal was to extend some of these results, especially from [**36**], to the case where the limit set may have infinitely many orbits. In particular, now you want to consider more general arcs of systems starting inside the Axiom A, not just the Morse-Smale systems. But this also led to some very interesting new problems and ideas related, for instance, to fractal dimensions.

To explain this, let me consider a situation as described in Figure 3, a surface diffeomorphism with a non-transverse intersection between the stable and the unstable

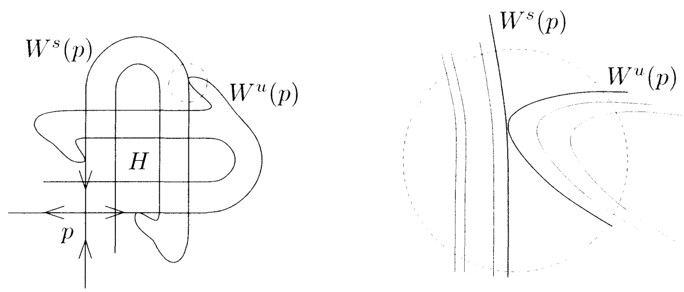


FIGURE 3. Homoclinic tangency associated to a hyperbolic set

manifold of a periodic saddle point p. We call that intersection a homoclinic tangency. And the periodic point p is contained in an infinite hyperbolic set H of the diffeomorphism, a horseshoe. This means that the homoclinic tangency is accumulated by a pair of laminations, or partial foliations, formed by the stable and unstable manifolds of all the points in H.

A diffeomorphism like this may be obtained as a first bifurcation ξ_b of an arc $\{\xi_\mu\}$ starting at an Axiom A system. The map ξ_b itself is not Axiom A, the homoclinic tangency implies that the non-wandering set is not hyperbolic. Then, as you increase the parameter, the stable and the unstable laminations move with respect to each other and, whenever there is a tangency between a leaf of one and a leaf of the other, the diffeomorphism can not be Axiom A.

Since these are just laminations, not full foliations of open sets, you might expect that such tangencies should be easy to avoid, taking advantage of the gaps between the leaves. However, I showed in my thesis [**32**] that it is not true in general. In fact,

– if the laminations are transversely thick, that is, if the gaps are relatively small, it is impossible to avoid tangencies between leaves of the two laminations, they exist for a whole open set of diffeomorphisms.

I'll call this phenomenon *persistent homoclinic tangencies.* Later, in [**34**], I proved that this phenomenon occurs near any surface diffeomorphism with a homoclinic tangency:

– there always exist open sets in the parameter space arbitrarily close to the bifurcation, that correspond to persistent tangencies.

And then Clark Robinson [**63**] deduced a version of this result for arcs of diffeomorphisms.

Palis and Takens wanted to understand this issue in more detail, and they came to establish a deep connection between homoclinic bifurcations and fractal dimensions of hyperbolic sets. Let me explain this.

In the paper [**36**], that I mentioned before, Jacob and I had shown that tangencies between the stable and the unstable laminations were, essentially, the only thing one has to worry about. We showed that if there were no tangencies and, in fact, the map was not too close to having a tangency, then the non-wandering set was hyperbolic. So this was a kind of converse to the fact that tangencies are an obstruction to hyperbolicity.

In the setting we were dealing with the limit set was finite, and we were able to show that parameters for which the map *is* too close to a tangency have small relative measure near the bifurcation. That's how we proved that hyperbolicity (Axiom A and strong transversality) prevails near these homoclinic tangencies, in terms of measure in parameter space. And the arguments suggested that it might be possible to avoid tangencies for most parameter values in more general situations, provided the laminations were not too thick.

Now, Palis and Takens realized that this should be formulated in terms of the transverse fractal dimensions of the laminations. The condition they required was that the sum of the transverse Hausdorff dimensions of the stable and unstable laminations should be less than 1. By definition, the transverse Hausdorff dimension is the Hausdorff dimension of the intersection of the lamination with some cross-section. It can be shown, in this context, that the definition doesn't depend on the choice of the cross-section.

It turns out that the sum of these transverse Hausdorff dimensions is equal to the Hausdorff dimension of the hyperbolic set H. So, their theorem, proved around 1984, has a very elegant statement [**45**]:

– if the Hausdorff dimension $HD(H)$ of the hyperbolic set involved in the tangency is less than 1, then ξ_μ is hyperbolic (Axiom A and strong transversality) for most nearby parameters $\mu > b$:

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon} m\left(\{\mu \in (b, b+\varepsilon) : \xi_\mu \text{ is hyperbolic }\}\right) = 1, \tag{1}$$

where $m(\cdot)$ is Lebesgue measure.

At about the same time they proved a similar result for the heteroclinic case [**44**], where the tangency is between stable and unstable manifolds of different periodic points. Actually, in those papers they used another notion of dimension, called limit capacity, or box dimension, instead of Hausdorff dimension. But then it became clear that the two notions of fractal dimension coincide for hyperbolic sets of surface diffeomorphisms. This is discussed in their book [**46**, Chapters 4-5], where they also explain why (1) can always be stated with the full limit, initially in the heteroclinic case they only had a lim sup.

Then, in a paper [**51**] that was published in 1994, Jacob and Jean-Christophe Yoccoz proved that the condition in the previous theorem is, in fact, optimal:

– if the Hausdorff dimension of H is larger than 1, then the conclusion (1) above no longer holds.

This statement and, to some extent, the proof itself were inspired on a result of John Marstrand [**26**] about arithmetic differences

$$K_1 - \lambda K_2 = \{a_1 - \lambda\, a_2 : a_1 \in K_1 \text{ and } a_2 \in K_2\}$$

of Cantor sets in the real line: if the sum $HD(K_1) + HD(K_2)$ is larger than 1 then the difference has positive Lebesgue measure, for almost every λ. So, at this point it was already clear that there was an important relation between this part of Dynamics and other topics, like Geometric Theory of Dimension and Harmonic Analysis.

Cantor sets and Fractal Invariants

Motivated by this, Jacob started asking several questions about arithmetic differences of Cantor sets, with an eye on their applications to Dynamical Systems and other areas. In particular, he conjectured that for generic regular Cantor sets K_1 and K_2, the arithmetic difference either has zero Lebesgue measure or contains some interval. A Cantor set is called regular if it is generated by a smooth expanding map. The set of such Cantor sets comes with a natural topology, inherited from the corresponding maps.

Well, this conjecture was proved by Carlos Gustavo Moreira and Yoccoz [**30**], around the beginning of 1995. Actually, they proved a rather strong version of the conjecture. Their result applied to an open and dense set of regular Cantor sets that has "full probability", in some natural sense. Moreover, they get stable intersections, which is much stronger than just having an interval contained in the arithmetic difference. Then, they proved the following substantial extension of the previous results about homoclinic tangencies [**31**]: for generic arcs of diffeomorphisms $\{\xi_\mu\}_\mu$ with a homoclinic tangency at $\mu = b$,

– for most parameters $\mu > b$ close to b, in the sense of (1), either ξ_μ is hyperbolic or μ is in some interval with persistent homoclinic tangencies.

In other words, if $PT + AT$ is the union of all the intervals of persistent tangencies with those parameters for which the map satisfies the Axiom A and the strong transversality condition, then

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}m\big(PT + AT \cap (b, b+\varepsilon)\big) = 1.$$

The theorem of Palis and Takens says that if the Hausdorff dimension of the horseshoe H is less than 1 then we have the same result already for the set of parameters corresponding to hyperbolic maps. So, the main novelty of this result is when the Hausdorff dimension is larger than 1.

There is a very natural question that arises, which is, what can we say about the dynamics when it's not hyperbolic. Well, Jacob has some recent joint work with Yoccoz [**52**] about this, that Yoccoz will talk about later in this Conference, so I won't discuss in any detail.(2) But the point is that they define so-called non-uniformly hyperbolic sets, or non-uniformly hyperbolic horseshoes, that are an extension of the hyperbolic sets that still have several nice properties. And they were able to show that if the Hausdorff dimension of the original hyperbolic set H is not much larger than 1 (they have a precise technical condition), then the diffeomorphisms ξ_μ are non-uniformly hyperbolic for most parameters $\mu > b$ near b. That is, if NUH is the set of parameters such that the non-wandering set is a non-uniformly hyperbolic set, then

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}m\big(NUH \cap (b, b+\varepsilon)\big) = 1,$$

as long as the Hausdorff dimension is not much larger than 1.

Now let me say a few words about the higher dimensional case. Most of this has been proved for surface diffeomorphisms, and there are several serious difficulties that appear in higher dimensions. The main reason is that the stable and unstable laminations need not be transversely smooth. So, in general, it's not even known whether the transverse Hausdorff dimension is well defined. In fact, the geometry of hyperbolic sets in high dimensions is much less understood than in the surface case.

(2) Abstracts of talks given at the Conference are available at www.impa.br/~dsconf/.

In general, the Hausdorff dimension and the limit capacity are not equal, and they do not vary continuously with the dynamical system.

However, and this is a development near my heart, Jacob and Marcelo Viana were able to overcome some of these difficulties and, around 1989, prove the higher dimensional extension of the result about persistent homoclinic tangencies. The result was published in [**47**].

And they have very recent results together with Moreira, as we heard in Moreira's talk in this Conference, which show that the relation between fractal dimensions and abundance of hyperbolicity in parameter space stays valid for families of diffeomorphisms in arbitrary dimension.

Non-Hyperbolic Systems

The study of bifurcations, and these results that I mentioned, are part of an effort to go beyond the hyperbolic systems and understand very general dynamical systems. I think that, from the beginning, Jacob was convinced that bifurcation theory was the right way to do that or, at least, an essential part of trying to understand systems that are not hyperbolic, that are not structurally stable. And as the theory of homoclinic bifurcations developed, he became more and more convinced that they should play a key role in this.

By 1989 there was a paper of Benedicks and Carleson [**4**] where they proved that non-uniformly hyperbolic dynamics is frequent in the so-called Hénon family of plane maps

$$h(x, y) = (1 - ax^2 + y, bx).$$

That is, for a set of values of the parameters a and b with positive Lebesgue measure, the maps have a non-uniformly hyperbolic attractor. This was a striking extension of a very important pioneering work of Jakobson [**20**], back in the late seventies, where he had obtained a similar result for the family of quadratic real maps $q(x) = 1 - ax^2$.

Even before their paper appeared, Palis suggested that this result should be true, more generally, for generic arcs $\{\xi_\mu\}$ of surface diffeomorphisms with a homoclinic tangency. You see, it was known that returns maps of ξ_μ to certain regions near the tangency look like the Hénon model, so that was the idea. So, he proposed this problem to two of his students at the time, Leonardo Mora and Marcelo Viana. And Mora and Viana [**27**] were able to show that the approach of Benedicks and Carleson extended to more general dissipative systems, that are called Hénon-like maps, and from this they could prove Jacob's conjecture, in 1990.

These kinds of results, there are many others, relating homoclinic tangencies to other types of complicated dynamics, convinced Jacob that homoclinic tangencies might be some sort of unifying notion for understanding non-hyperbolic systems, at least in low dimensions. So he made the following conjecture:

– the union of Axiom A diffeomorphisms with those that have a homoclinic tangency is dense in $\mathcal{D}^r(M)$, if M is a surface.

In other words, every C^r surface diffeomorphism that is not in the closure of the Axiom A systems is approximated by other diffeomorphisms that have homoclinic tangencies.

As you probably know, this conjecture was proved a couple of years ago by two other former students of Jacob, Enrique Pujals and Martin Sambarino, in the case $r = 1$. Their paper has just appeared [**58**]. In fact, the result had been announced by Araújo and Mañé in the early 90's, but they never provided a proof. As a consequence of their methods, Pujals and Sambarino also got another most interesting result [**59**]:

– any arc of surface diffeomorphisms such that the topological entropy is not constant on it must contain a homoclinic tangency.

There is a version of the previous conjecture for high dimensions, that says that the union of Axiom A diffeomorphisms with those that have a homoclinic tangency or a heterodimensional cycle should be dense in $\mathcal{D}^r(M)$. A cycle is called heterodimensional if the stable manifolds of the periodic points involved in the cycle are not all of the same dimension. It seems that several groups of people have made progress in the direction of this high dimensional conjecture, indeed there will be a couple of talks on this subject in this Conference, but a complete proof is not yet available.

Back in the late eighties, Jacob suggested the study of heterodimensional cycles to Lorenzo Díaz, as his thesis problem. The idea was to complement our own results in [**36**], as I said before, we studied the equidimensional case. Now, Díaz found out that the conclusions are quite different for heterodimensional cycles: most of the times the bifurcating diffeomorphism ξ_b is not accumulated by hyperbolic ones, in fact, there is a whole interval $(b, b + \varepsilon)$ such that ξ_μ is not hyperbolic, not structurally stable, for any parameter μ in this interval. These results appeared in his thesis [**11**] and were much developed in a series of joint papers with Jorge Rocha, another former student of Jacob. See for instance [**13**].

And, sometime later, it became clear that heterodimensional cycles also have an important connection with the phenomenon of robust non-hyperbolic attractors, which I'll mention again in a little while.

A Unifying View of Dynamics

By 1995, Jacob had put several ideas and conjectures together to form a coherent picture of what might be the typical kinds of behavior of non-hyperbolic systems. This appeared in a preprint that was published in Douady's volume of Astérisque [**41**]. The main point is a conjecture that every system can be approximated by another having only finitely many attractors, whose basins of attraction contain almost all points. In fact these systems should have large probability in parameter space, in some natural sense. And the attractors should have nice properties, such as the existence of so-called Sinai-Ruelle-Bowen measures.

It is interesting to observe that the idea that most dynamical systems should have a finite number of attractors goes back to René Thom, in the sixties, although he didn't make precise what "most" was supposed to mean. Certainly, he was motivated by Smale's ideas in hyperbolic theory at the time(3), where the point of view was, primarily, topological. Maybe because of this, it was widely understood that Thom had in mind a residual (second category of Baire) subset of all dynamical systems and, in this form, the finiteness statement turned out to be false [**33**]. So, Jacob's conjecture is a very interesting revival of this classical idea, in a new and more probabilistic framework. A key novelty in Palis' approach is to allow the existence of cycles occupying a small volume in the dynamical space. Indeed, cycles have been a main obstruction to the realization of previous global scenarios for Dynamics.

So far, it is known that this conjecture holds for quadratic maps of interval, as a consequence of work done by Lyubich, Martens, and Nowicki. See [**22**]. And both Misha Lyubich and Artur de Melo will speak in this conference about their recent work with Welington de Melo, where they extended this to general analytic families of unimodal maps.

In higher dimensions, there have been some very interesting results that, I believe, were at least partially motivated by Jacob's questions and conjectures.

There is the work of Díaz, Pujals, Ures, and Bonatti [**12, 5**] where they characterized the robust sets of diffeomorphisms in any dimension. An invariant set is robust if it is transitive and remains transitive under any C^1 small perturbation of the system. They proved that robust sets must have a so-called dominated splitting, which is a decomposition of the tangent space into two continuous distributions such that one is more expanding than the other at every point, by a definite factor. In dimension 3 at least one of the distribution is hyperbolic, either expanding or contracting. This is called partial hyperbolicity.

Moreover, Alves, Bonatti, and Viana proved existence and finiteness of ergodic attractors, or Sinai-Ruelle-Bowen measures, for certain types of partially hyperbolic systems, in a paper [**1**] that has just appeared.

And there is also very important work of Carlos Morales, Maria José Pacifico, and Enrique Pujals [**28, 29**], characterizing the robust sets of arbitrary flows in 3 dimensions. Robust sets containing only regular orbits must be hyperbolic, so the more interesting case is when the set contains some singularity. They proved that any robust set that contains a singularity is a Lorenz-like attractor, or repeller, meaning that it has all the main features of the geometric Lorenz models of Guckenheimer-Williams [**17**].

(3)"Toutefois, selon certaines idées récentes de S. Smale, si la varieté est compacte, presque tout champ X présenterait un nombre fini d'attracteurs isolément structurellement stables" [**67**, p. 56]

Many Other Results

There are many other important contributions that Palis has done. For instance, there is his work on moduli invariants, that is, characterizing systems with the property that the number of topological types of perturbations depends on a finite number of real parameters. In [**40**], he discovered a smooth invariant for topological conjugacy

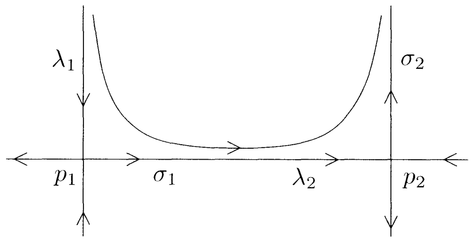


FIGURE 4. Moduli of conjugacy in saddle-connections

between flows with a saddle connection as in Figure 4. In fact, two such flows are conjugated if and only if they have the same ratio of eigenvalues

$$\frac{\lambda_1}{\sigma_2}$$

And, together with Welington de Melo and Sebastian van Strien [**9, 10**], he obtained a characterization of such systems with mild recurrence, in a wide variety of situations.

As a part of the development of moduli theory there was a description of typical holomorphic vectors fields, the topological types of linear holomorphic vector fields in CP^n, which was done by César Camacho, Nicolaas Kuiper, and Jacob in [**7**].

I should also mention his series of papers with Yoccoz, where they study rigidity of centralizers of diffeomorphisms, that are the sets of diffeomorphisms which commute with a given diffeomorphism. In a series of papers [**48, 49, 50**], they prove that, generically, the centralizer is trivial for a hyperbolic diffeomorphism, it just contains the iterates of the map.

Actually, even back in his thesis, Jacob had been interested in a related problem: how frequently diffeomorphisms embed in flows. He observed that there were open sets of diffeomorphisms where the natural topological conditions that you would need to embed in a flow were not sufficient: there were open sets of such diffeomorphisms that did not embed in flows. And, somewhat later, in [**39**], he was able to prove that, C^1 generically, diffeomorphisms do not embed in flows.

If you look at Jacob's list of scientific works attached to this paper, you'll see that I could still go on for a long time. So, let me just conclude with some personal remarks.

Conclusion

It's interesting to note that up to 1993 Jacob had 16 graduate students, whose theses appeared up to that year. He's been Director of IMPA since around 1993, and as of 2000 he has 35 graduate students. So one might conclude that administration is not so bad for someone with the talents of Jacob Palis...

In any event, he has exhibited leadership, as I indicated, direction and scope in formulating conjectures and stimulating many people throughout the world. The scope has increased dramatically as we get evidence of collaboration with Yoccoz, with Viana, with many other people, and of much activity, many interesting results, going deeply into the study of dynamical systems.

So, on the occasion of his 60th birthday, we all look forward to continued development for many, many years.

References

[1] J.F. Alves, C. Bonatti, and M. Viana. SRB measures for partially hyperbolic systems whose central direction is mostly expanding. *Invent. Math.*, 140(2):351–398, 2000.

[2] A. Andronov and L. Pontryagin. Systèmes grossiers. *Dokl. Akad. Nauk. USSR*, 14:247–251, 1937.

[3] D.V. Anosov. Geodesic flows on closed Riemannian manifolds of negative curvature. *Trudy Mat. Inst. Steklov.*, 90:209, 1967.

[4] M. Benedicks and L. Carleson. The dynamics of the Hénon map. *Ann. of Math. (2)*, 133(1):73–169, 1991.

[5] C. Bonatti, L.J. Díaz, and E.R. Pujals. A C^1-generic dichotomy for diffeomorphisms: weak forms of hyperbolicity or infinitely many sinks or sources. Preprint, 1999.

[6] P. Brunovský. On one-parameter families of diffeomorphisms. II. Generic branching in higher dimensions. *Comment. Math. Univ. Carolinae*, 12:765–784, 1971.

[7] C. Camacho, N.H. Kuiper, and J. Palis. The topology of holomorphic flows with singularity. *Inst. Hautes Études Sci. Publ. Math.*, 48:5–38, 1978.

[8] M.J. Dias Carneiro and J. Palis. Bifurcations and global stability of families of gradients. *Inst. Hautes Études Sci. Publ. Math.*, 70:103–168, 1990.

[9] W. de Melo and J. Palis. Moduli of stability for diffeomorphisms. In *Global theory of dynamical systems (Proc. Internat. Conf., Northwestern Univ., Evanston, Ill., 1979)*, pages 318–339. Springer, Berlin, 1980.

[10] W. de Melo, J. Palis, and S.J. van Strien. Characterising diffeomorphisms with modulus of stability one. In *Dynamical systems and turbulence, Warwick 1980 (Coventry, 1979/1980)*, pages 266–285. Springer, Berlin, 1981.

[11] L.J. Díaz. Robust nonhyperbolic dynamics and heterodimensional cycles. *Ergodic Theory Dynam. Systems*, 15(2):291–315, 1995.

[12] L.J. Díaz, E.R. Pujals, and R. Ures. Partial hyperbolicity and robust transitivity. *Acta Math.*, 183(1):1–43, 1999.

[13] L.J. Díaz and J. Rocha. Large measure of hyperbolic dynamics when unfolding heteroclinic cycles. *Nonlinearity*, 10(4):857–884, 1997.

[14] J.M. Franks. Time dependent stable diffeomorphisms. *Invent. Math.*, 24:163–172, 1974.

[15] N.K. Gavrilov and L.P. Šil'nikov. Three-dimensional dynamical systems that are close to systems with a structurally unstable homoclinic curve. I. *Mat. Sb. (N.S.)*, 88(130):475–492, 1972.

[16] J. Guckenheimer. Absolutely Ω-stable diffeomorphisms. *Topology*, 11:195–197, 1972.

[17] J. Guckenheimer and R.F. Williams. Structural stability of Lorenz attractors. *Inst. Hautes Études Sci. Publ. Math.*, 50:59–72, 1979.

[18] C. Gutiérrez. Structural stability for flows on the torus with a cross-cap. *Trans. Amer. Math. Soc.*, 241:311–320, 1978.

[19] S. Hayashi. Connecting invariant manifolds and the solution of the C^1 stability and Ω-stability conjectures for flows. *Ann. of Math. (2)*, 145(1):81–137, 1997.

[20] M.V. Jakobson. Absolutely continuous invariant measures for one-parameter families of one-dimensional maps. *Comm. Math. Phys.*, 81(1):39–88, 1981.

[21] S.-T. Liao. On the stability conjecture. *Chinese Ann. Math.*, 1(1):9–30, 1980.

[22] M. Lyubich. Almost every real quadratic map is either regular or stochastic. *Annals of Math.* To appear.

[23] R. Mañé. Contributions to the stability conjecture. *Topology*, 17(4):383–396, 1978.

[24] R. Mañé. An ergodic closing lemma. *Ann. of Math. (2)*, 116(3):503–540, 1982.

[25] R. Mañé. A proof of the C^1 stability conjecture. *Inst. Hautes Études Sci. Publ. Math.*, 66:161–210, 1988.

[26] J.M. Marstrand. Some fundamental geometrical properties of plane sets of fractional dimensions. *Proc. London Math. Soc. (3)*, 4:257–302, 1954.

[27] L. Mora and M. Viana. Abundance of strange attractors. *Acta Math.*, 171(1):1–71, 1993.

[28] C. A. Morales, M.J. Pacifico, and E.R. Pujals. Robust transitive singular sets for 3-flows are partially hyperbolic attractors and repellers. Preprint, 1999.

[29] C.A. Morales, M.J. Pacífico, and E.R. Pujals. On C^1 robust singular transitive sets for three-dimensional flows. *C. R. Acad. Sci. Paris Sér. I Math.*, 326(1):81–86, 1998.

[30] C.G. Moreira and J.-C. Yoccoz. Stable intersections of regular Cantor sets with large Hausdorff dimension. Preprint 1998.

[31] C.G. Moreira and J.-C. Yoccoz. Tangences homocliniques stables pour les ensembles hyperboliques de grande dimension fractale. Preprint 2000.

[32] S.E. Newhouse. Nondensity of Axiom A(a) on S^2. In *Global Analysis (Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968)*, pages 191–202. Amer. Math. Soc., Providence, R.I., 1970.

[33] S.E. Newhouse. Diffeomorphisms with infinitely many sinks. *Topology*, 13:9–18, 1974.

[34] S.E. Newhouse. The abundance of wild hyperbolic sets and nonsmooth stable sets for diffeomorphisms. *Inst. Hautes Études Sci. Publ. Math.*, 50:101–151, 1979.

[35] S.E. Newhouse and J. Palis. Bifurcations of Morse-Smale dynamical systems. In *Dynamical systems (Proc. Sympos., Univ. Bahia, Salvador, 1971)*, pages 303–366, New York, 1973. Academic Press.

[36] S.E. Newhouse and J. Palis. Cycles and bifurcation theory. *Astérisque*, 31:43–140, 1976.

[37] S.E. Newhouse, J. Palis, and F. Takens. Bifurcations and stability of families of diffeomorphisms. *Inst. Hautes Études Sci. Publ. Math.*, 57:5–71, 1983.

[38] J. Palis. On Morse-Smale dynamical systems. *Topology*, 8:385–404, 1968.

[39] J. Palis. Vector fields generate few diffeomorphisms. *Bull. Amer. Math. Soc.*, 80:503–505, 1974.

[40] J. Palis. A differentiable invariant of topological conjugacies and moduli of stability. *Astérisque*, 51:335–346, 1978.

[41] J. Palis. A global view of dynamics and a conjecture on the denseness of finitude of attractors. *Astérisque*, 261:xiii–xiv, 335–347, 2000. Géométrie complexe et systèmes dynamiques (Orsay, 1995).

[42] J. Palis and S. Smale. Structural stability theorems. In *Global Analysis (Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968)*, pages 223–231. Amer. Math. Soc., Providence, R.I., 1970.

[43] J. Palis and F. Takens. Stability of parametrized families of gradient vector fields. *Ann. of Math. (2)*, 118(3):383–421, 1983.

[44] J. Palis and F. Takens. Cycles and measure of bifurcation sets for two-dimensional diffeomorphisms. *Invent. Math.*, 82(3):397–422, 1985.

[45] J. Palis and F. Takens. Hyperbolicity and the creation of homoclinic orbits. *Ann. of Math. (2)*, 125(2):337–374, 1987.

[46] J. Palis and F. Takens. *Hyperbolicity and sensitive-chaotic dynamics at homoclinic bifurcations*. Cambridge University Press, 1993.

[47] J. Palis and M. Viana. High dimension diffeomorphisms displaying infinitely many periodic attractors. *Ann. of Math. (2)*, 140(1):207–250, 1994.

[48] J. Palis and J.-C. Yoccoz. Centralizers of Anosov diffeomorphisms on tori. *Ann. Sci. École Norm. Sup. (4)*, 22(1):99–108, 1989.

[49] J. Palis and J.-C. Yoccoz. Rigidity of centralizers of diffeomorphisms. *Ann. Sci. École Norm. Sup. (4)*, 22(1):81–98, 1989.

[50] J. Palis and J.-C. Yoccoz. Differentiable conjugacies of Morse-Smale diffeomorphisms. *Bol. Soc. Brasil. Mat. (N.S.)*, 20(2):25–48, 1990.

[51] J. Palis and J.-C. Yoccoz. Homoclinic tangencies for hyperbolic sets of large Hausdorff dimension. *Acta Math.*, 172(1):91–136, 1994.

[52] J. Palis and J. C. Yoccoz. Nonuniformily hyperbolic horseshoes unleashed by homoclinic bifurcations and zero density of attractors. *C. R. Ac. Sc. Paris*, 2000. To appear.

[53] M.M. Peixoto. On structural stability. *Ann. of Math. (2)*, 69:199–222, 1959.

[54] M.M. Peixoto. Structural stability on two dimensional manifolds. *Topology*, 1:101-120, 1962.

[55] V.A. Pliss. Analysis of the necessity of the conditions of Smale and Robbin for structural stability for periodic systems of differential equations. *Differencial'nye Uravnenija*, 8:972–983, 1972.

[56] C. Pugh. The closing lemma. *Amer. J. Math.*, 89:956–1009, 1967.

[57] C. Pugh and M. Shub. The Ω-stability theorem for flows. *Invent. Math.*, 11:150–158, 1970.

[58] E.R. Pujals and M. Sambarino. Homoclinic tangencies and hyperbolicity for surface diffeomorphisms: a conjecture of Palis. *Annals of Math.*, 2000. To appear.

[59] E.R. Pujals and M. Sambarino. On homoclinic tangencies, hyperbolicity, creation of homoclinic orbits and varation of entropy. *Nonlinearity*, 13(3):921–926, 2000.

[60] J.W. Robbin. A structural stability theorem. *Ann. of Math. (2)*, 94:447–493, 1971.

[61] C. Robinson. Structural stability of vector fields. *Ann. of Math. (2)*, 99:154–175, 1974.

[62] C. Robinson. Structural stability of C^1 diffeomorphisms. *J. Differential Equations*, 22(1):28–73, 1976.

[63] C. Robinson. Bifurcation to infinitely many sinks. *Comm. Math. Phys.*, 90(3):433–459, 1983.

[64] A. Sannami. The stability theorems for discrete dynamical systems on two-dimensional manifolds. *Nagoya Math. J.*, 90:1–55, 1983.

[65] S. Smale. Differentiable dynamical systems. *Bull. Amer. Math. Soc.*, 73:747–817, 1967.

[66] J. Sotomayor. Generic one-parameter families of vector fields on two-dimensional manifolds. *Inst. Hautes Études Sci. Publ. Math.*, 43:5–46, 1974.

[67] R. Thom. *Stabilité structurelle et morphogénèse.* W. A. Benjamin, Inc., Reading, Mass., 1972. Essai d'une théorie générale des modèles, Mathematical Physics Monograph Series.

[68] R.F. Williams. The "DA" maps of Smale and structural stability. In *Global Analysis (Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968)*, pages 329–334. Amer. Math. Soc., Providence, R.I., 1970.

S. NEWHOUSE, Department of mathematics, Michigan State University, USA
E-mail : sen@math.msu.edu

Astérisque
286, 2003, p. 25–62

RANDOM PERTURBATIONS OF NONUNIFORMLY EXPANDING MAPS

by

José Ferreira Alves & Vítor Araújo

Abstract. — We give both sufficient conditions and necessary conditions for the stochastic stability of nonuniformly expanding maps either with or without critical sets. We also show that the number of probability measures describing the statistical asymptotic behaviour of random orbits is bounded by the number of SRB measures if the noise level is small enough. As an application of these results we prove the stochastic stability of certain classes of nonuniformly expanding maps introduced in [**Vi1**] and [**ABV**].

1. Introduction

Dynamical systems theory has, among its main goals, the description of the typical behaviour of orbits as time goes to infinity, and understanding how this behaviour is modified under small perturbations of the system. This work refers to the study of the latter problem from a probabilistic point of view.

Given a map f from a manifold M into itself, let $(x_n)_{n\geqslant 1}$ be the orbit of a given point $x_0 \in M$, that is $x_{n+1} = f(x_n)$ for every $n \geqslant 1$. Consider the sequence of time averages of Dirac measures δ_{x_j} along the orbit of x_0 from time 0 to n. A special interest lies on the study of the convergence of such time averages for a "large" set of points $x_0 \in M$ and the properties of their limit measures. In this direction, we refer the work of Sinai [**Si**] for Anosov diffeomorphisms, later extended by Ruelle and Bowen [**BR, Ru**] for Axiom A diffeomorphisms and flows. In the context of systems with no uniform hyperbolic structure Jakobson [**Ja**] proved the existence of such measures for certain quadratic transformations of the interval exhibiting chaotic behaviour.

2000 ***Mathematics Subject Classification.*** — 37A05, 37A50, 37C40, 37D25.
Key words and phrases. — Random perturbations, physical measures, stochastic stability, non-uniform expansion.

Work partially supported by FCT through CMUP.

Another important contribution on this subject was given by Benedicks and Young [**BY1**], based on the previous work of Benedicks and Carleson [**BC1, BC2**], where this kind of measures were constructed for Hénon two dimensional maps exhibiting strange attractors. The recent work of Alves, Bonatti and Viana [**ABV**] shows that such measures exist in great generality for systems exhibiting some nonuniformly expanding behaviour.

The notion of stability that most concerns us can be formulated in the following way. Assume that, instead of time averages of Dirac measures supported on the iterates of $x_0 \in M$, we consider time averages of Dirac measures δ_{x_j}, where at each iteration we take x_{j+1} close to $f(x_j)$ with a controlled error. One is interested in studying the existence of limit measures for these time averages and their relation to the analogous ones for unperturbed orbits, that is, the stochastic stability of the initial system.

Systems with some uniformly hyperbolic structure are quite well understood and stability results have been established in general by Kifer and Young; see [**Ki1, Ki2**] and [**Yo**]. The knowledge of the stochastic behaviour of systems that do not exhibit such uniform expansion/contraction is still very incomplete. Important results on this subject were obtained by Katok, Kifer [**KK**], Benedicks, Young [**BY1**], Baladi and Viana [**BV**] for certain quadratic maps of the interval. Another important contribution is the announced work of Benedicks and Viana for Hénon-like strange attractors. As far as we know these are the only results of this type for systems with no uniform expanding behaviour.

In this work we present both sufficient conditions and necessary conditions for the stochastic stability of nonuniformly expanding dynamical systems. As an application of these results we prove that the classes of nonuniformly expanding maps introduced in [**Vi1**] and [**ABV**] are stochastically stable.

1.1. Statement of results. — Let $f : M \to M$ be a smooth map defined on a compact riemannian manifold M. We fix some normalized riemannian volume form m on M that we call *Lebesgue measure*.

Given μ an f-invariant Borel probability measure on M, we say that μ is an *SRB measure* if, for a positive Lebesgue measure set of points $x \in M$, the averaged sequence of Dirac measures along the orbit $(f^n(x))_{n\geqslant 0}$ converges in the weak* topology to μ, that is,

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{j=0}^{n-1} \varphi\big(f^n(x)\big) = \int \varphi\, d\mu \tag{1}$$

for every continuous map $\varphi : M \to \mathbb{R}$. We define the *basin* of μ as the set of those points x in M for which (1) holds for all continuous φ. The maps to be considered in this work will only have a finite number of SRB measures whose basins cover the whole manifold M, up to a set of zero Lebesgue measure.

We are interested in studying random perturbations of the map f. For that, we take a continuous map

$$\begin{aligned} \Phi : T &\longrightarrow C^2(M, M) \\ t &\longmapsto f_t \end{aligned}$$

from a metric space T into the space of C^2 maps from M to M, with $f = f_{t^*}$ for some fixed $t^* \in T$. Given $x \in M$ we call the sequence $\big(f_{\underline{t}}^n(x)\big)_{n\geqslant 1}$ a *random orbit* of x, where $\underline{t}$ denotes an element $(t_1, t_2, t_3, \dots)$ in the product space $T^{\mathbb{N}}$ and

$$f_{\underline{t}}^n = f_{t_n} \circ \cdots \circ f_{t_1} \quad \text{for } n \geqslant 1.$$

We also take a family $(\theta_\varepsilon)_{\varepsilon>0}$ of probability measures on T such that $(\operatorname{supp}\theta_\varepsilon)_{\varepsilon>0}$ is a nested family of connected compact sets and $\operatorname{supp}\theta_\varepsilon \to \{t^*\}$ when $\varepsilon \to 0$. We will also assume some quite general nondegeneracy conditions on Φ and $(\theta_\varepsilon)_{\varepsilon>0}$ (see the beginning of Section 3) and refer to $\{\Phi, (\theta_\varepsilon)_{\varepsilon>0}\}$ as a *random perturbation* of f.

In the context of random perturbations of a map we say that a Borel probability measure μ^ε on M is *physical* if for a positive Lebesgue measure set of points $x \in M$, the averaged sequence of Dirac probability measures $\delta_{f_{\underline{t}}^n(x)}$ along random orbits $\big(f_{\underline{t}}^n(x)\big)_{n\geqslant 0}$ converges in the weak* topology to μ^ε for $\theta_\varepsilon^{\mathbb{N}}$ almost every $\underline{t} \in T^{\mathbb{N}}$. That is,

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi\big(f_{\underline{t}}^n(x)\big) = \int \varphi \, d\mu^\varepsilon \quad \text{for all continuous } \varphi\colon M \to \mathbb{R} \tag{2}$$

and $\theta_\varepsilon^{\mathbb{N}}$ almost every $\underline{t} \in T^{\mathbb{N}}$. We denote the set of points $x \in M$ for which (2) holds by $B(\mu^\varepsilon)$ and call it the *basin of* μ^ε. The map $f\colon M \to M$ is said to be *stochastically stable* if the weak* accumulation points (when $\varepsilon > 0$ goes to zero) of the physical probability measures of f are convex linear combinations of the (finitely many) SRB measures of f.

1.1.1. Local diffeomorphisms. — Let $f : M \to M$ be a C^2 local diffeomorphism of the manifold M. We say that f is *nonuniformly expanding* if there is some constant $c > 0$ for which

$$\limsup_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \log \|Df(f^j(x))^{-1}\| \leqslant -c < 0 \tag{3}$$

for Lebesgue almost every $x \in M$. It was proved in [**ABV**] that for a nonuniformly expanding local diffeomorphism f the following holds:

(P) *There is a finite number of ergodic absolutely continuous* (SRB) *f-invariant probability measures $\mu_1, \dots, \mu_p$ whose basins cover a full Lebesgue measure subset of M. Moreover, every absolutely continuous f-invariant probability measure μ may be written as a convex linear combination of $\mu_1, \dots, \mu_p$: there are real numbers $w_1, \dots, w_p \geqslant 0$ with $w_1 + \cdots + w_p = 1$ for which $\mu = w_1\mu_1 + \cdots + w_p\mu_p$.*

The proof of the previous result was based on the existence of α-hyperbolic times for the points in M: given $0 < \alpha < 1$, we say that $n \in \mathbb{Z}^+$ is a *α-hyperbolic time* for the point $x \in M$ if

$$\prod_{j=n-k}^{n-1} \|Df(f^j(x))^{-1}\| \leqslant \alpha^k \quad \text{for every} \quad 1 \leqslant k \leqslant n. \tag{4}$$

The existence of (a positive frequency of) α-hyperbolic times for points $x \in M$ is a consequence of the hypothesis of nonuniform expansion of the map f and permits us to define a map $h : M \to \mathbb{Z}^+$ giving the first hyperbolic time for m almost every $x \in M$.

In the context of random perturbations of a nonuniformly expanding map we are also able to prove a result on the finitness of physical measures.

Theorem A. — *Let $f\colon M \to M$ be a C^2 nonuniformly expanding local diffeomorphism. If $\varepsilon > 0$ is sufficiently small, then there are physical measures $\mu_1^\varepsilon, \dots, \mu_\ell^\varepsilon$ (with ℓ not depending on ε) such that:*

(1) *for each $x \in M$ and $\theta_\varepsilon^{\mathbb{N}}$ almost every $\underline{t} \in T^{\mathbb{N}}$, the average of Dirac measures $\delta_{f_{\underline{t}}^n(x)}$ converges in the weak* topology to some μ_i^ε with $1 \leqslant i \leqslant \ell$;*

(2) *for each $1 \leqslant i \leqslant \ell$ we have*

$$\mu_i^\varepsilon = w^*\text{-}\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \int \big(f_{\underline{t}}^j\big)_* \big(m \| B(\mu_i^\varepsilon)\big)\, d\theta_\varepsilon^{\mathbb{N}}(\underline{t}),$$

where $m\|B(\mu_i^\varepsilon)$ is the normalization of the Lebesgue measure restricted to $B(\mu_i^\varepsilon)$;

(3) *if f is topologically transitive, then $\ell = 1$.*

We say that the map f is *nonuniformly expanding for random orbits* if there is some constant $c > 0$ such that for $\varepsilon > 0$ small enough

$$\limsup_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \log \|Df(f_{\underline{t}}^j(x))^{-1}\| \leqslant -c < 0, \tag{5}$$

for $\theta_\varepsilon^{\mathbb{N}} \times m$ almost every $(\underline{t}, x) \in T^{\mathbb{N}} \times M$. Similarly to the deterministic situation, condition (5) permits us to introduce a notion of α-hyperbolic times for points in $T^{\mathbb{N}} \times M$ and define a map

$$h_\varepsilon \colon T^{\mathbb{N}} \times M \longrightarrow \mathbb{Z}^+$$

by taking $h_\varepsilon(\underline{t}, x)$ the first α-hyperbolic time for the point $(\underline{t}, x) \in T^{\mathbb{N}} \times M$ (see Section 2). Assuming that h_ε is integrable with respect to $\theta_\varepsilon^{\mathbb{N}} \times m$, then

$$\|h_\varepsilon\|_1 = \sum_{k=0}^{\infty} k\, (\theta_\varepsilon^{\mathbb{N}} \times m)\big(\{(\underline{t}, x)\colon h_\varepsilon(\underline{t}, x) = k\}\big) < \infty. \tag{6}$$

We say that the family $(h_\varepsilon)_{\varepsilon>0}$ has *uniform L^1-tail*, if the series in (6) converges uniformly to $\|h_\varepsilon\|_1$ (as a series of functions of the variable ε).

Theorem B. — *Let $f\colon M \to M$ be a nonuniformly expanding C^2 local diffeomorphism.*

(1) *If f is stochastically stable, then f is nonuniformly expanding for random orbits.*

(2) *If f is nonuniformly expanding for random orbits and $(h_\varepsilon)_\varepsilon$ has uniform L^1-tail, then f is stochastically stable.*

We should emphasize that we do not know if condition (2) in Theorem B is really necessary. No example of a stochastically stable map which does not satisfy the uniform L^1-tail property is known.

1.1.2. Maps with critical sets. — Similar results to those presented for random perturbations of local diffeomorphisms will also be obtained for maps with critical sets in the sense of [**ABV**]. We start by describing the class of maps that we are going to consider. Let $f\colon M \to M$ be a continuous map of the compact manifold M that fails to be a C^2 local diffeomorphism on a critical set $\mathcal{C} \subset M$ with zero Lebesgue measure. We assume that f *behaves like a power of the distance* close to the critical set $\mathcal{C}$: there are constants $B > 1$ and $\beta > 0$ for which

(S1) $\displaystyle \frac{1}{B}\operatorname{dist}(x,\mathcal{C})^\beta \leqslant \frac{\|Df(x)v\|}{\|v\|} \leqslant B\operatorname{dist}(x,\mathcal{C})^{-\beta};$

(S2) $\displaystyle \big|\log\|Df(x)^{-1}\| - \log\|Df(y)^{-1}\|\,\big| \leqslant B\frac{\operatorname{dist}(x,y)}{\operatorname{dist}(x,\mathcal{C})^\beta};$

(S3) $\displaystyle \big|\log|\det Df(x)^{-1}| - \log|\det Df(y)^{-1}|\,\big| \leqslant B\frac{\operatorname{dist}(x,y)}{\operatorname{dist}(x,\mathcal{C})^\beta};$

for every $x, y \in M \setminus \mathcal{C}$ with $\operatorname{dist}(x,y) < \operatorname{dist}(x,\mathcal{C})/2$ and $v \in T_xM$. Given $\delta > 0$ we define the *δ-truncated distance* from $x \in M$ to $\mathcal{C}$

$$\operatorname{dist}_\delta(x,\mathcal{C}) = \begin{cases} 1 & \text{if } \operatorname{dist}(x,\mathcal{C}) \geqslant \delta, \\ \operatorname{dist}(x,\mathcal{C}) & \text{otherwise.}\end{cases}$$

Assume that f is a nonuniformly expanding map, in the sense that there is $c > 0$ such that the limit in (3) holds for Lebesgue almost every $x \in M$ (recall that we are taking $\mathcal{C}$ with zero Lebesgue measure) and, moreover, suppose that the orbits of f have *slow approximation to the critical set*: given small $\gamma > 0$ there is $\delta > 0$ such that

$$\limsup_{n\to+\infty} \frac{1}{n}\sum_{j=0}^{n-1} -\log\operatorname{dist}_\delta(f^j(x),\mathcal{C}) \leqslant \gamma \tag{7}$$

for Lebesgue almost every $x \in M$. The results in [**ABV**] show that in this situation we obtain the same conclusion on the finiteness of SRB measures for such an f, also holding property (P).

In order to prove the stochastic stability of maps with critical sets we need to restrict the class of perturbations we are going to consider: we take maps f_t with the same critical set $\mathcal{C}$ and impose that

$$Df_t(x) = Df(x) \quad \text{for every } x \in M \setminus \mathcal{C} \text{ and } t \in T. \tag{8}$$

This may be implemented, for instance, in parallelizable manifolds (with an additive group structure, e.g. tori $\mathbb{T}^d$ or cylinders $\mathbb{T}^{d-k} \times \mathbb{R}^k$) by considering

$$T = \{t \in \mathbb{R}^d \colon \|t\| \leqslant \varepsilon_0\}$$

for some $\varepsilon_0 > 0$, θ_ε the normalized Lebesgue measure on the ball of radius $\varepsilon \leqslant \varepsilon_0$, and taking $f_t = f + t$; that is, adding at each step a random noise to the unperturbed dynamics.

For the case of maps with critical sets we also need to impose an analog of condition (7) for random orbits; we assume *slow approximation of random orbits to the critical set*: given any small $\gamma > 0$ there is $\delta > 0$ such that

$$\limsup_{n \to +\infty} \frac{1}{n} \sum_{j=0}^{n-1} -\log \operatorname{dist}_\delta(f_{\underline{t}}^j(x), \mathcal{C}) \leqslant \gamma \tag{9}$$

for $\theta_\varepsilon^{\mathbb{N}} \times m$ almost every $(\underline{t}, x) \in T^{\mathbb{N}} \times M$ and small $\varepsilon > 0$. Results similar to those presented for local diffeomorphisms on the finiteness of physical measures can also be obtained in this case.

Theorem C. — *Let $f\colon M \to M$ be a C^2 nonuniformly expanding map behaving like a power of the distance close to the critical set $\mathcal{C}$, and whose orbits have slow approximation to $\mathcal{C}$. If f is nonuniformly expanding for random orbits and random orbits have slow approximation to $\mathcal{C}$, then we arrive at the same conclusions as in Theorem A.*

The property of nonuniform expansion for random orbits, together with the slow approximation of random orbits to the critical set permit us to introduce a notion of (α, δ)-hyperbolic times for points in $(\underline{t}, x) \in T^{\mathbb{N}} \times M$ and define a map

$$h_\varepsilon \colon T^{\mathbb{N}} \times M \longrightarrow \mathbb{Z}^+,$$

by taking $h_\varepsilon(\underline{t}, x)$ the first (α, δ)-hyperbolic time for the point $(\underline{t}, x) \in T^{\mathbb{N}} \times M$, see Section 2. Assuming that h_ε is integrable with respect to $\theta_\varepsilon \times m$, then we obtain an analog to (6), which enables us to define a notion of *uniform L^1-tail* exactly in the same way as before.

Due to the fact that $\log \|Df^{-1}\|$ is not a continuous map (it is not even everywhere defined) we are not able to present in this setting a result similar to Theorem B in all its strength. However, we obtain the same kind of conclusion of the second item of Theorem B.

Theorem D. — *Let $f\colon M \to M$ be nonuniformly expanding C^2 map behaving like a power of the distance close to its critical set $\mathcal{C}$ and whose orbits have slow approximation to $\mathcal{C}$. Assume that f is nonuniformly expanding for random orbits and random orbits have slow approximation to $\mathcal{C}$. If $(h_\varepsilon)_\varepsilon$ has uniform L^1-tail, then f is stochastically stable.*

As a major application of the previous theorem we are thinking of a class of maps on the cylinder $S^1 \times \mathbb{R}$ introduced in [**Vi1**]. Subsequent works [**Al**] and [**AV**] showed that such systems are topologically mixing (thus transitive) and have a unique SRB measure. The work [**AV**] also shows that these SRB measures vary continuously with the map, which means that time averages of continuous functions are only slightly affected when the system is perturbed. Although this points in a direction of statistical stability, this does not imply the stochastic stability of such systems as we defined above.

The class of nonuniformly expanding maps (with critical sets) introduced by M. Viana can be described as follows. Let $a_0 \in (1,2)$ be such that the critical point $x=0$ is pre-periodic for the quadratic map $Q(x) = a_0 - x^2$. Let $S^1 = \mathbb{R}/\mathbb{Z}$ and $b : S^1 \to \mathbb{R}$ be a Morse function, for instance, $b(s) = \sin(2\pi s)$. For fixed small $\alpha > 0$, consider the map

$$\begin{aligned} \widehat{f} : S^1 \times \mathbb{R} &\longrightarrow \quad S^1 \times \mathbb{R} \\ (s,x) &\longmapsto \big(\widehat{g}(s), \widehat{q}(s,x)\big) \end{aligned}$$

where $\widehat{g}$ is the uniformly expanding map of the circle defined by $\widehat{g}(s) = ds \pmod{\mathbb{Z}}$ for some $d \geqslant 16$, and $\widehat{q}(s,x) = a(s) - x^2$ with $a(s) = a_0 + \alpha b(s)$. It is easy to check that for $\alpha > 0$ small enough there is an interval $I \subset (-2,2)$ for which $\widehat{f}(S^1 \times I)$ is contained in the interior of $S^1 \times I$. Thus, any map f sufficiently close to $\widehat{f}$ in the C^0 topology has $S^1 \times I$ as a forward invariant region. We consider from here on these maps f close to $\widehat{f}$ restricted to $S^1 \times I$. Taking into account the expression of $\widehat{f}$ it is not difficult to check that $\widehat{f}$ (and any map f close to $\widehat{f}$ in the C^2 topology) behaves like a power of the distance close to the critical set.

Theorem E. — *If f is sufficiently close to $\widehat{f}$ in the C^3 topology then f is nonuniformly expanding and its orbits have slow approximation to the critical set. Moreover, if the noise level of a random perturbation of f is sufficiently small, then*

(1) *f is nonuniformly expanding for random orbits;*
(2) *random orbits have slow approximation to the critical set;*
(3) *the family of hyperbolic time maps $(h_\varepsilon)_\varepsilon$ has uniform L^1-tail.*

As an immediate consequence of Theorems C, D and E we have that Viana maps are stochastically stable. An application of Theorems A and B will also be given in Section 6 for an open class of local diffeomorphisms introduced in [**ABV**, Appendix A].

2. Distortion bounds

In this section we generalize some of the results in [**Al**] and [**ABV**] for the setting of stochastic perturbations of a nonuniformly expanding map. These results will be proved in the setting of maps with critical sets. Then everything follows in the same way for local diffeomorphisms if we think of $\mathcal{C}$ as being equal to the empty set, with the only exception of a particular point that we clarify in Remark 2.4 below (due to the fact that we are not assuming condition (8) for maps with no critical sets). For the next definition we take $0 < b < \min\{1/2, 1/(2\beta)\}$.

Definition 2.1. — Given $0 < \alpha < 1$ and $\delta > 0$, we say that $n \in \mathbb{Z}^+$ is a (α, δ)-hyperbolic time for $(\underline{t}, x) \in T^{\mathbb{N}} \times M$ if

$$\prod_{j=n-k}^{n-1} \|Df_{t_{j+1}}(f_{\underline{t}}^j(x))^{-1}\| \leqslant \alpha^k \quad \text{and} \quad \operatorname{dist}_\delta(f_{\underline{t}}^{n-k}(x), \mathcal{C}) \geqslant \alpha^{bk}$$

for every $1 \leqslant k \leqslant n$.

The following lemma, due to Pliss [**Pl**], provides the main tool in the proof of the existence of hyperbolic times for points with nonuniform expansion on random orbits.

Lemma 2.2. — *Let $H \geqslant c_2 > c_1 > 0$ and $\zeta = (c_2 - c_1)/(H - c_1)$. Given real numbers $a_1, \ldots, a_N$ satisfying*

$$\sum_{j=1}^{N} a_j \geqslant c_2 N \quad \textit{and} \quad a_j \leqslant H \ \textit{ for all } \ 1 \leqslant j \leqslant N,$$

there are $\ell > \zeta N$ and $1 < n_1 < \cdots < n_\ell \leqslant N$ such that

$$\sum_{j=n+1}^{n_i} a_j \geqslant c_1 \cdot (n_i - n) \ \textit{ for each } \ 0 \leqslant n < n_i, \ i = 1, \ldots, \ell.$$

Proof. — See [**ABV**, Lemma 3.1]. □

Proposition 2.3. — *There are $\alpha > 0$ and $\delta > 0$ for which $\theta_\varepsilon^{\mathbb{N}} \times m$ almost every $(\underline{t}, x) \in T^{\mathbb{N}} \times M$ has some (α, δ)-hyperbolic time.*

Proof. — Let $(\underline{t}, x) \in T^{\mathbb{N}} \times M$ be a point satisfying (5). For large N we have

$$-\sum_{j=0}^{N-1} \log \left\| Df(f_{\underline{t}}^j(x))^{-1} \right\| \geqslant \frac{c}{2} N > 0,$$

by definition of nonuniform expansion on random orbits. Fixing $\rho > \beta$ we see that condition (S1) implies

$$\left|\log \left\|Df(x)^{-1}\right\|\right| \leqslant \rho \left|\log \operatorname{dist}(x, \mathcal{C})\right| \tag{10}$$

for every x in a neighborhood V of $\mathcal{C}$. Now we take $\gamma_1 > 0$ so that $\rho\gamma_1 \leqslant c/10$ and let $\delta_1 > 0$ be small enough to get

$$-\sum_{j=0}^{N-1} \log \operatorname{dist}_{\delta_1}(f_{\underline{t}}^j(x), S) \leqslant \gamma_1 N \quad \text{for large } N, \tag{11}$$

which is possible after property (7) of slow approximation to $\mathcal{C}$. Moreover, fixing $H \geqslant \rho|\log\delta|$ sufficiently large in order that it be also an upper bound for for the set $\{-\log\|Df_{\underline{t}}^{-1}\| : t\in T,\ x\in M\setminus V\}$, then the set

$$E=\{1\leqslant j\leqslant N : -\log\|Df(f_{\underline{t}}^{j-1}(x))^{-1}\|>H\}$$

is such that $f_{\underline{t}}^{j-1}(x)\in V$ for all $j\in E$ and

$$\rho\left|\log\operatorname{dist}(f_{\underline{t}}^{j-1}(x),\mathcal{C})\right| > -\log\left\|Df(f_{\underline{t}}^{j-1}(x))^{-1}\right\| > H \geqslant \rho|\log\delta|$$

i.e., $\operatorname{dist}(f_{\underline{t}}^{j-1}(x),\mathcal{C})<\delta_1$, in particular $\operatorname{dist}_{\delta_1}(f_{\underline{t}}^{j-1}(x),\mathcal{C})=\operatorname{dist}(f_{\underline{t}}^{j-1}(x),\mathcal{C})<\delta_1$ for all $j\in E$. Hence, defining

$$a_j=\begin{cases}-\log\left\|Df(f_{\underline{t}}^{j-1}(x))^{-1}\right\| & \text{if } j\notin E\\ 0 & \text{if } j\in E\end{cases}$$

it holds $a_j\leqslant H$ for $1\leqslant j\leqslant N$, and (10) and (11) imply

$$-\sum_{j\in E}\log\left\|Df(f_{\underline{t}}^{j-1}(x))^{-1}\right\|\leqslant\rho\sum_{j\in E}\left|\log\operatorname{dist}(f_{\underline{t}}^{j-1}(x),\mathcal{C})\right|\leqslant\rho\gamma_1 N.$$

Since $\rho\gamma_1\leqslant c/10$ we deduce

$$\sum_{j=1}^{N}a_j=\sum_{j=1}^{N}\Big(-\log\left\|Df(f_{\underline{t}}^{j-1}(x))^{-1}\right\|\Big)-\sum_{j\in E}\Big(-\log\left\|Df(f_{\underline{t}}^{j-1}(x))^{-1}\right\|\Big)\geqslant\frac{2}{5}cN.$$

By the previous arguments we may apply Lemma 2.2 to the sequence a_j with $c_1=c/5$ and $c_2=2c/5$ (we may suppose $H>c_1$ too by increasing H if needed). Thus there are $\zeta_1>0$ and $\ell_1>\zeta_1 N$ times $1\leqslant q_1<\cdots<q_{\ell_1}\leqslant N$ such that

$$\sum_{j=n+1}^{q_i}-\log\left\|Df(f_{\underline{t}}^{j-1}(x))^{-1}\right\|\geqslant\sum_{j=n+1}^{q_i}a_j\geqslant\frac{c}{2}(q_i-n)\tag{12}$$

for every $0\leqslant n<q_i$, $i=1,\dots,\ell_1$. We observe that (12) is just the first part of the requirements on (α,δ)-hyperbolic times for $(\underline{t},x)$ if $\alpha=\exp(c/5)$.

Now we apply again Lemma 2.2, this time to the sequence $a_j=\log\operatorname{dist}_{\delta_2}(f_{\underline{t}}^{j-1}(x),\mathcal{C})$, where $\delta_2>0$ is small enough so that for $\gamma_2>0$ with $2\gamma_2(bc)^{-1}<\zeta_1$ we have by assumption (7)

$$\sum_{j=0}^{N-1}\log\operatorname{dist}_{\delta_2}(f_{\underline{t}}^{j}(x),\mathcal{C})\geqslant-\gamma_2 N\quad\text{for large } N.$$

Defining $c_1=bc/2$, $c_2=-\gamma_2$, $H=0$ and

$$\zeta_2=\frac{c_2-c_1}{H-c_1}=1-\frac{2\gamma_2}{bc},$$

Lemma 2.2 ensures that there are $\ell_2 \geqslant \zeta_2 N$ times $1 \leqslant r_1 < \cdots < r_{\ell_2} \leqslant N$ satisfying

$$\sum_{j=n+1}^{r_i} \log \operatorname{dist}_{\delta_2}(f_{\underline{t}}^{j+1}(x), \mathcal{C}) \geqslant \frac{bc}{2}(r_i - n) \tag{13}$$

for every $0 \leqslant n < r_i$, $i = 1, \ldots, \ell_2$. Let us note that the condition on γ_2 assures $\zeta_1 + \zeta_2 > 1$. So if $\zeta = \zeta_1 + \zeta_2 - 1$, then there must be $\ell = (\ell_1 + \ell_2 - N) \geqslant \zeta N$ and $1 \leqslant n_1 < \cdots < n_\ell \leqslant N$ for which (12) and (13) both hold. This means that for $1 \leqslant i \leqslant \ell$ and $1 \leqslant k \leqslant n_i$ we have

$$\prod_{j=n_i-k}^{n_i} \left\| Df(f_{\underline{t}}^j(x))^{-1} \right\| \leqslant \alpha^k \quad \text{and} \quad \operatorname{dist}_{\delta_2}(f_{\underline{t}}^{n_i-k}(x), \mathcal{C}) \geqslant \alpha^{bk},$$

and hence these n_i are (α, δ)-hyperbolic times for $(\underline{t}, x)$, with $\delta = \delta_2$ and $\alpha = \exp(c/5)$. It follows that for $\theta_\varepsilon^{\mathbb{N}} \times m$ almost every $(\underline{t}, x) \in T^{\mathbb{N}} \times M$ there are (positive frequency of) times $n \in \mathbb{Z}^+$ for which

$$\prod_{j=n-k}^{n-1} \| Df(f_{\underline{t}}^j(x))^{-1} \| \leqslant \alpha^k \quad \text{and} \quad \operatorname{dist}_{\delta}(f_{\underline{t}}^{n-k}(x), \mathcal{C}) \geqslant \alpha^{bk} \tag{14}$$

for every $1 \leqslant k \leqslant n$. Now the conclusion of the lemma is a direct consequence of assumption (8). $\square$

Remark 2.4. — In the setting of random perturbations of a local diffeomorphism f we may also derive from the first part of (14) the existence of hyperbolic times for $\theta_\varepsilon^{\mathbb{N}} \times m$ almost every $(\underline{t}, x) \in T^{\mathbb{N}} \times M$ without assuming condition (8). Actually, let $(\underline{t}, x)$ be a point in $T^{\mathbb{N}} \times M$ for which the first part of (14) holds. Taking the perturbations f_t in a sufficiently small C^1-neighborhood of f, then

$$\|Df_t(y)^{-1}\| \leqslant \frac{1}{\sqrt{\alpha}} \|Df(y)^{-1}\|$$

for every $y \in M$, which together with (14) gives

$$\prod_{j=n-k}^{n-1} \|Df_t(f_{\underline{t}}^j(x))^{-1}\| \leqslant \prod_{j=n-k}^{n-1} \frac{1}{\sqrt{\alpha}} \|Df(f_{\underline{t}}^j(x))^{-1}\| \leqslant \alpha^{k/2}.$$

In the context of maps with no critical sets this n may be defined as a $\sqrt{\alpha}$-hyperbolic time for $(\underline{t}, x)$ and all the results that we present below hold with $\sqrt{\alpha}$-hyperbolic times replacing (α, δ)-hyperbolic times for maps with critical sets.

Proposition 2.3 allows us to introduce a map

$$h_\varepsilon \colon T^{\mathbb{N}} \times M \longrightarrow \mathbb{Z}^+,$$

by taking $h_\varepsilon(\underline{t}, x)$ as the first (α, δ)-hyperbolic time for $(\underline{t}, x) \in T^{\mathbb{N}} \times M$. We assume henceforth that the family $(h_\varepsilon)_{\varepsilon > 0}$ has uniform L^1-tail. For the next lemma we fix $\delta_1 > 0$ in such a way that $4\delta_1 < \min\{\delta, \delta^\beta |\log \alpha|\}$.

Lemma 2.5. — *Given any $1 \leqslant j \leqslant n$, we have*

$$\|Df(y)^{-1}\| \leqslant \alpha^{-1/2}\|Df(f_{\underline{t}}^{n-j}(x))^{-1}\|$$

for every y in the ball of radius $2\delta_1\alpha^{j/2}$ around $f_{\underline{t}}^{n-j}(x)$.

Proof. — We are assuming $\mathrm{dist}_\delta(f_{\underline{t}}^{n-j}(x), \mathcal{C}) \geqslant \alpha^j$ since n is a (α,δ)-hyperbolic time for $(\underline{t}, x)$. This means that

$$\mathrm{dist}(f_{\underline{t}}^{n-j}(x), \mathcal{C}) = \mathrm{dist}_\delta(f_{\underline{t}}^{n-j}(x), \mathcal{C}) \geqslant \alpha^{bj} \quad \text{or else} \quad \mathrm{dist}(f_{\underline{t}}^{n-j}(x), \mathcal{C}) \geqslant \delta.$$

Either way it holds $\mathrm{dist}(y, f_{\underline{t}}^{n-j}(x)) \geqslant \mathrm{dist}(f_{\underline{t}}^{n-j}(x), \mathcal{C})/2$ because $b < 1/2$ and $\delta_1 < \delta/4 < 1/4$ for all y in the ball of radius $2\delta_1\alpha^{j/2}$ around $f_{\underline{t}}^{n-j}(x)$. Therefore condition (S2) implies

$$\log \frac{\|Df(y)^{-1}\|}{\|Df(f_{\underline{t}}^{n-j}(x))^{-1}\|} \leqslant B\frac{\mathrm{dist}(f_{\underline{t}}^{n-j}(x), y)}{\mathrm{dist}(f_{\underline{t}}^{n-j}(x), \mathcal{C})^\beta} \leqslant B\frac{2\delta_1\alpha^{j/2}}{\min\{\alpha^{b\beta j}, \delta^\beta\}}.$$

But $\alpha, \delta < 1$ and $b\beta < 1/2$ so $\alpha^{j/2} < \alpha^{b\beta j}$ and thus the right hand side of the last expression is bounded from above by $2B\delta_1\delta^{-\beta}$. The assumptions on δ_1 assure this last bound to be smaller than $\log \alpha^{-1/2}$, which implies the statement. □

Proposition 2.6. — *There is $\delta_1 > 0$ such that if n is (α,δ)-hyperbolic time for $(\underline{t}, x) \in T^{\mathbb{N}} \times M$, then there is a neighborhood $V_n(\underline{t}, x)$ of x in M such that*

(1) *$f_{\underline{t}}^n$ maps $V_n(\underline{t}, x)$ diffeomorphically onto the ball of radius δ_1 around $f_{\underline{t}}^n(x)$;*
(2) *for every $1 \leqslant k \leqslant n$ and $y, z \in V_k(\underline{t}, x)$*

$$\mathrm{dist}(f_{\underline{t}}^{n-k}(y), f_{\underline{t}}^{n-k}(z)) \leqslant \alpha^{k/2}\mathrm{dist}(f_{\underline{t}}^n(y), f_{\underline{t}}^n(z)).$$

Proof. — The proof will be by induction on $j \geqslant 1$. First we show that there is a well defined branch of f^{-j} on a ball of small enough radius around $f_{\underline{t}}^j(x)$. Now we observe that Lemma 2.5 gives for $j = 1$

$$\|Df(y)^{-1}\| \leqslant \alpha^{-1/2}\|Df(f_{\underline{t}}^{n-1}(x))^{-1}\| \leqslant \alpha^{1/2},$$

because n is a (α,δ)-hyperbolic time for $(\underline{t}, x)$. This means that f is a $\alpha^{-1/2}$-dilation in the ball of radius $2\delta_1\alpha^{1/2}$ around $f_{\underline{t}}^{n-1}(x)$. Consequently there is some neighborhood $V_1(\underline{t}, x)$ of $f_{\underline{t}}^{n-1}(x)$ inside the ball of radius $2\delta_1\alpha^{1/2}$ that is diffeomorphic to the ball of radius δ_1 around $f_{\underline{t}}^n(x)$ through f_{t_n}, when f is a map with critical set satisfying (8).

For $j \geqslant 1$ let us suppose that we have obtained a neighborhood $V_j(\underline{t}, x)$ of $f_{\underline{t}}^{n-j}(x)$ such that $f_{t_n} \circ \cdots \circ f_{t_{n-j+1}} \mid V_j(\underline{t}, x)$ is a diffeomorphism onto the ball of radius δ_1 around $f_{\underline{t}}^n(x)$ with

$$\|Df(f_{t_{n-j+i+1}} \circ \cdots \circ f_{t_{n-j+1}}(z))^{-1}\| \leqslant \alpha^{-1/2}\|Df(f_{\underline{t}}^{n-j+i+1}(x))^{-1}\| \tag{15}$$

for all $z \in V_j(\underline{t}, x)$ and $0 \leqslant i < j$. Then, by Lemma 2.5 and under the assumption that n is a (α, δ)-hyperbolic time for x,

$$\begin{aligned}\left\|D\big(f_{t_n} \circ \cdots \circ f_{t_{n-j}}(y)\big)^{-1}\right\| &\leqslant \prod_{i=0}^{j} \left\|Df_{t_{n-j+i}}\big(f_{t_{n-j+i-1}} \circ \cdots \circ f_{t_{n-j}}(y)\big)^{-1}\right\| \\ &\leqslant \prod_{i=0}^{j} \alpha^{-1/2}\left\|Df_{t_{n-j+i}}\big(f_{\underline{t}}^{n-j+i-1}(x)\big)^{-1}\right\| \\ &\leqslant (\alpha^{-1/2})^{j+1} \cdot \alpha^{j+1} = \alpha^{(j+1)/2}\end{aligned}$$

for every y on the ball of radius $2\delta_1\alpha^{(j+1)/2}$ around $f_{\underline{t}}^{n-j-1}(x)$ whose image $f_{t_{n-j}}(y)$ is in $V_j(\underline{t}, x)$ (above we convention $f_{t_{n-j+i-1}} \circ \cdots \circ f_{t_{n-j}}(y) = y$ for $i = 0$).

This shows that the derivative of $f_{t_n} \circ \cdots \circ f_{t_{n-j}}$ is a $\alpha^{-(j+1)/2}$-dilation on the intersection of $f_{t_{n-j}}^{-1}\big(V_j(\underline{t}, x)\big)$ with the ball of radius $2\delta_1\alpha^{(j+1)/2}$ around $f_{\underline{t}}^{n-j-1}(x)$, and hence there is an inverse branch of $f_{t_n} \circ \cdots \circ f_{t_{n-j}}$ defined on the ball of radius δ_1 around $f_{\underline{t}}^{n}(x)$. Thus we may define $V_{j+1}(\underline{t}, x)$ as the image of the ball of radius δ_1 around $f_{\underline{t}}^{n}(x)$ under this inverse branch, and recover the induction hypothesis for $j + 1$. In this manner we get neighborhoods $V_j(\underline{t}, x)$ of $f_{\underline{t}}^{n-j}(x)$ as above for all $1 \leqslant j \leqslant n$. □

Corollary 2.7. — *There is a constant $C_1 > 0$ such that if $\underline{t} \in T^{\mathbb{N}}$, n is a (α, δ)-hyperbolic time for $x \in M$ and $y, z \in V_n(\underline{t}, x)$, then*

$$\frac{1}{C_1} \leqslant \frac{|\det Df_{\underline{t}}^{n}(y)|}{|\det Df_{\underline{t}}^{n}(z)|} \leqslant C_1.$$

Proof. — For $1 \leqslant k \leqslant n$ the distance between $f_{\underline{t}}^{k}(x)$ and either $f_{\underline{t}}^{k}(y)$ or $f_{\underline{t}}^{k}(z)$ is smaller than $\alpha^{(n-k)/2}$ which is smaller than $\alpha^{b(n-k)} \leqslant \operatorname{dist}(f_{\underline{t}}^{k}(x), \mathcal{C})$. So, by (S3) we have

$$\begin{aligned}\log\frac{|\det Df_{\underline{t}}^{n}(y)|}{|\det Df_{\underline{t}}^{n}(z)|} &= \sum_{k=0}^{n-1} \log\frac{|\det Df_{t_{k+1}}(f_{\underline{t}}^{k}(y))|}{|\det Df_{t_{k+1}}(f_{\underline{t}}^{k}(z))|} \\ &\leqslant \sum_{k=1}^{n-1} \log\frac{|\det Df(f_{\underline{t}}^{k}(y))|}{|\det Df(f_{\underline{t}}^{k}(z))|} \\ &\leqslant \sum_{k=0}^{n-1} 2B\frac{\alpha^{(n-k)/2}}{\alpha^{b\beta(n-k)}},\end{aligned}$$

and it is enough to take $C_1 \leqslant \exp\big(\sum_{i=1}^{\infty} 2B\alpha^{(1/2-b\beta)i}\big)$, recalling that $b\beta < 1/2$ and also (8). □

3. Stationary measures

As mentioned before, we will assume the random perturbations of the nonuniformly expanding map f satisfy some *nondegeneracy conditions*: there exists $0 < \varepsilon_0 < 1$ such that for every $0 < \varepsilon < \varepsilon_0$ we may take $n_0 = n_0(\varepsilon) \in \mathbb{N}$ for which the following holds:

(1) there is $\xi = \xi(\varepsilon) > 0$ such that $\left\{f_{\underline{t}}^n(x) \colon \underline{t} \in (\operatorname{supp}\theta_\varepsilon)^{\mathbb{N}}\right\}$ contains the ball of radius ξ around $f^n(x)$ for all $x \in M$ and $n \geqslant n_0$;

(2) $(f_x^n)_*\theta_\varepsilon^{\mathbb{N}} \ll m$ for all $x \in M$ and $n \geqslant n_0$.

Here $(f_x^n)_*\theta_\varepsilon^{\mathbb{N}}$ is the push-forward of $\theta_\varepsilon^{\mathbb{N}}$ to M via $f_x^n : T^{\mathbb{N}} \to M$, defined as $f_x^n(\underline{t}) = f_{\underline{t}}^n(x)$. Condition (1) means that perturbed iterates cover a full neighborhood of the unperturbed ones after a threshold for all sufficiently small noise levels. Condition (2) means that sets of perturbation vectors of positive $\theta_\varepsilon^{\mathbb{N}}$ measure must send any point $x \in M$ onto subsets of M with positive Lebesgue measure after a finite number of iterates.

In [**Ar**, Examples 1 & 2] it was shown that given any smooth map $f : M \to M$ of a compact manifold we can always construct a random perturbation satisfying the nondegeneracy conditions (1) and (2), if we take $T = \mathbb{R}^p$, $t^* = 0$ and θ_ε is equal to the normalized restriction of the Lebesgue measure to the ball of radius ε around 0, for a sufficiently big number $p \in \mathbb{N}$ of parameters. For parallelizable manifolds the random perturbations which consist in adding at each step a random noise to the unperturbed dynamics, as described in the Introduction, clearly satisfy nondegeneracy conditions (1) and (2) for $n_0 = 1$.

In the context of random perturbations of a map, we say that a set $A \subset M$ is *invariant* if $f_t(A) \subset A$, at least for $t \in \operatorname{supp}(\theta_\varepsilon)$ with $\varepsilon > 0$ small. The usual invariance of a measure with respect to a transformation is replaced by the following one: a probability measure μ is said to be *stationary*, if for every continuous $\varphi : M \to \mathbb{R}$ it holds

$$\int \varphi\, d\mu = \iint \varphi\big(f_t(x)\big)\, d\mu(x)\, d\theta_\varepsilon(t). \tag{16}$$

Remark 3.1. — If $(\mu^\varepsilon)_{\varepsilon>0}$ is a family of stationary measures having μ_0 as a weak* accumulation point when ε goes to 0, then it follows from (16) and the convergence of $\operatorname{supp}(\theta_\varepsilon)$ to $\{t^*\}$ that μ_0 must be invariant by $f = f_{t^*}$.

It is not difficult to see (cf. [**Ar**]) that a stationary measure μ satisfies

$$x \in \operatorname{supp}(\mu) \implies f_t(x) \in \operatorname{supp}(\mu) \text{ for all } t \in \operatorname{supp}(\theta_\varepsilon)$$

just by continuity of Φ. This means that if μ is a stationary measure, then $\operatorname{supp}(\mu)$ is an invariant set. Nondegeneracy condition (1) ensures that the interior of $\operatorname{supp}(\mu)$ is nonempty.

Let us write $\operatorname{supp}(\mu)$ as a disjoint union $\bigcup_i C_i$ of connected components and consider only those C_i for which $m(C_i) > 0$ — this collection is nonempty since

$\mathrm{supp}(\mu)$ contains open sets. Moreover each f_t must permute these components for $t \in \mathrm{supp}(\theta_\varepsilon)$, because $f_t(C_i)$ is connected by continuity, $f_t(C_i) \subset \mathrm{supp}(\mu)$ by invariance, and $m(f_t(C_i)) > 0$ since we have $(f_t)_* m \ll m$.

The connectedness of C_i and continuity of Φ guarantee that the above-mentioned perturbation of the components C_i induced by f_t does not depend on $t \in \mathrm{supp}(\theta_\varepsilon)$. Indeed, supposing that $t, t' \in \mathrm{supp}(\theta_\varepsilon)$ are such that

$$f_t(C_i) \subset C_j \quad \text{and} \quad f_{t'}(C_i) \subset C_{j'},$$

then fixing some $z \in C_i$ we have that $\{f_t(z) \colon t \in \mathrm{supp}(\theta_\varepsilon)\}$ is a connected set intersecting both C_j and $C_{j'}$ inside $\mathrm{supp}(\mu)$, and so $C_j = C_{j'}$.

We will show that these connected components are periodic under the action induced by f_t with $t \in \mathrm{supp}(\theta_\varepsilon)$. After this, we may use nondegeneracy condition (1) to conclude that each component contains a ball of uniform radius and thus that each component satisfies $m(C_i) > \mathrm{const} > 0$. Hence there existing only a finite number of such components.

At this point it is useful to introduce the skew-product map

$$\begin{aligned} F : T^{\mathbb{N}} \times M &\longrightarrow \quad T^{\mathbb{N}} \times M \\ (\underline{t}, z) \quad &\longmapsto \big(\sigma(\underline{t}), f_{t_1}(z)\big) \end{aligned}$$

where σ is the left shift on sequences $\underline{t} = (t_1, t_2, \dots) \in T^{\mathbb{N}}$. It is easy to check that the product measure $\theta_\varepsilon^{\mathbb{N}} \times \mu$ is F-invariant, as so is the set $\mathrm{supp}(\theta_\varepsilon^{\mathbb{N}} \times \mu) = \mathrm{supp}(\theta_\varepsilon)^{\mathbb{N}} \times \mathrm{supp}(\mu)$.

Lemma 3.2. — *The support of a stationary measure μ contains a finite number of connected components arranged in cycles permuted by the action of f_t for $t \in \mathrm{supp}(\theta_\varepsilon)$.*

Proof. — Is is enough to obtain that each connected component C_i is periodic under the action of f_t for $t \in \mathrm{supp}(\theta_\varepsilon)$, in the sense that $f_{\underline{t}}^p(C_i) \subset C_i$ for some $p \in \mathbb{N}$ and all $\underline{t} \in \mathrm{supp}(\theta_\varepsilon^{\mathbb{N}})$. There are components C_i with nonempty interior, since the interior of $\mathrm{supp}(\mu)$ is nonempty. So we may take a component C_i that contains some ball B. Then we have $m(B) > 0$ and so $(\theta_\varepsilon^{\mathbb{N}} \times \mu)(\mathrm{supp}(\theta_\varepsilon^{\mathbb{N}}) \times B) > 0$. Poincaré Recurrence Theorem now guarantees there is $(\underline{t}, x) \in \mathrm{supp}(\theta_\varepsilon^{\mathbb{N}}) \times B$ such that the F-orbit of $(\underline{t}, x)$ has the same $(\underline{t}, x)$ as an accumulation point. We see that there must exist some $p \in \mathbb{N}$ such that $f_{\underline{t}}^p(x) \in B \subset C_i$. In view of the independence of the permutation on the choice of $\underline{t}$, we conclude that C_i is sent inside itself by $f_{\underline{t}}^p$ for all $\underline{t} \in \mathrm{supp}(\theta_\varepsilon^{\mathbb{N}})$. $\square$

It is clear that the cycles obtained above are invariant sets. We are now ready to decompose μ into some simpler measures. For that we need the following result.

Lemma 3.3. — *The normalized restriction of a stationary measure to an invariant set is a stationary measure.*

Proof. — See [**Ar**, Lemma 8.2]. $\square$

We define an *invariant domain* in M as a finite collection $(U_0,\dots,U_{p-1})$ of pairwise separated open sets, that is, $\overline{U}_i\cap\overline{U}_j=\varnothing$ if $i\neq j$, such that $f_{\underline{t}}^k(U_i)\subset U_{(k+i)\bmod p}$ for all $k\geqslant 1$, $i=0,\dots,p-1$ and $\underline{t}\in\operatorname{supp}(\theta_\varepsilon^{\mathbb{N}})$.

In order to get the separation of the connected components in a cycle, we may unite those components C_i and C_j such that $\overline{C}_i\cap\overline{C}_j\neq\varnothing$ and observe that the permutation now induced in the new sets by f_t also does not depend on the choice of $t\in\operatorname{supp}(\theta_\varepsilon)$. In this manner we construct invariant domains inside the support of any stationary probability measure.

The next step is to look for *minimal invariant domains* with respect to the natural order relation of inclusion of sets. Let $D=(U_0,\dots,U_{p-1})$ and $D'=(W_0,\dots,W_{q-1})$ be invariant domains. On the one hand, $D=D'$ if there are $i,j\in\mathbb{N}$ such that $U_{(i+k)\bmod p}=W_{(j+k)\bmod q}$ for all $k\geqslant 1$, which implies $p=q$ because the open sets that form each invariant domain are pairwise disjoint. On the other hand, we say $D\prec D'$ if there are $i,j\in\mathbb{N}$ such that $U_{i\bmod p}\subsetneq W_{j\bmod q}$ and $U_{(i+k)\bmod p}\subset W_{(j+k)\bmod q}$ for all $k\geqslant 1$.

Lemma 3.4. — *In the partially ordered family of all invariant domains in M, with respect to the relation $\prec$, the number of $\prec$-minimal domains is finite. Moreover, every invariant domain contains at least one minimal domain.*

Proof. — The proof relies in showing that Zorn's Lemma can be applied to this partially ordered set and that minimal domains are pairwise separated. See [**Ar**, Section 3]. □

Let us now fix $x\in M$ and consider

$$\mu_n(x)=\frac{1}{n}\sum_{j=0}^{n-1}(f_x^j)_*\theta_\varepsilon^{\mathbb{N}}. \tag{17}$$

Since this is a sequence of probability measures on the compact manifold M, then it has weak* accumulation points.

Lemma 3.5. — *Every weak* accumulation point of $\big(\mu_n(x)\big)_n$ is stationary and absolutely continuous with respect to the Lebesgue measure.*

Proof. — Let μ be a weak* accumulation point of $\big(\mu_n(x)\big)_n$. We may write

$$\iint\varphi\big(f_t(x)\big)\,d\mu(x)\,d\theta_\varepsilon(t)=\int\lim_{k\to+\infty}\frac{1}{n_k}\sum_{j=0}^{n_k-1}\int\varphi\left(f_t\big(f_{\underline{t}}^j(x)\big)\right)d\theta_\varepsilon^{\mathbb{N}}(\underline{t})\,d\theta_\varepsilon(t)$$

for each continuous $\varphi:M\to\mathbb{R}$. Moreover dominated convergence ensures that we may exchange the limit and the outer integral sign and, by definition of $f_{\underline{t}}^j(x)$, we get

$$\lim_{k\to\infty}\frac{1}{n_k}\sum_{j=0}^{n_k-1}\int\varphi\big(f_{\underline{t}}^{j+1}(x)\big)\,d\theta_\varepsilon^{\mathbb{N}}(\underline{t})=\int\varphi\,d\mu,$$

according to the definition of μ. Thus (16) must hold and μ is stationary.

Noting that $C^0(M,\mathbb{R})$ is dense in $L^1(M,\mu)$ with the L^1 norm, we see that (16) holds for all μ-integrable functions $\varphi : M \to \mathbb{R}$. In particular, if $E \subset M$ is such that $m(E) = 0$, then

$$\begin{aligned}\int 1_E\, d\mu &= \iint 1_E\big(f_t(x)\big)\, d\mu(x)\, d\theta_\varepsilon(t)\\ &= \iint 1_E\big(f_t(x)\big)\, d\theta_\varepsilon(t)\, d\mu(x)\\ &= \iiint 1_E\big(f_t(f_s(x))\big)\, d\theta_\varepsilon(t)\, d\mu(x)\, d\theta_\varepsilon(s)\\ &= \iint 1_E\big(f^2_{\underline{t}}(x)\big)\, d\theta^{\mathbb{N}}_\varepsilon(\underline{t})\, d\mu(x)\\ &= \int (f^2_x)_*\theta^{\mathbb{N}}_\varepsilon(E)\, d\mu(x).\end{aligned}$$

This process may be iterated to yield

$$\mu(E) = \int (f^{n_0}_x)_*\theta_\varepsilon(E)\, d\mu(x)$$

and, since $(f^{n_0}_x)_*\theta_\varepsilon \ll m$ by nondegeneracy condition 2, we must have $\mu(E) = 0$. $\square$

Clearly if $x \in M$ belongs to some set of an invariant domain $(U_0,\dots,U_{p-1})$, then $\mu_n(x)$ have supports contained in $\overline{U}_0 \cup \cdots \cup \overline{U}_{p-1}$ for all $n \geqslant 1$ and any weak* accumlation point μ of $(\mu_n(x))_n$ is a stationary measure with $\operatorname{supp}(\mu) \subset \overline{U}_0 \cup \cdots \cup \overline{U}_{p-1}$. We will now see these measures are physical.

Lemma 3.6. — *If $(U_0,\dots,U_{p-1})$ is a minimal invariant domain, then there is a unique absolutely continuous stationary measure ν such that* $\operatorname{supp}(\nu) \subset \overline{U}_0 \cup \cdots \cup \overline{U}_{p-1}$. *Moreover, this ν is a physical measure and* $\operatorname{supp}(\nu) = \overline{U}_0 \cup \cdots \cup \overline{U}_{p-1}$.

Proof. — Let us assume $n_0 = 1$ for simplicity (see [**Ar**, Section 7] for the general case) and let us consider a stationary absolutely continuous probability measure ν with $\operatorname{supp}(\nu) \subset \overline{U}_0 \cup \cdots \cup \overline{U}_{p-1}$. We first show the ergodicity of ν, in the sense that $\theta^{\mathbb{N}}_\varepsilon \times \nu$ is F-ergodic. It turns out that to be F-ergodic it suffices that either $\nu(G) = 0$ or $\nu(G) = 1$ for every Borel set $G \subset M$ satisfying

$$1_G(x) = \int 1_G\left(f_t(x)\right)\, d\theta_\varepsilon(t) \tag{18}$$

for ν almost every x (cf. [**Ar**] and [**Vi2**]). So let us take G such that $\nu(G) > 0$ and G satisfies the left hand side of (18). Then it must be $m(G) > 0$ because $\nu \ll m$ and there is a closed set $J \subset G$ such that $m(G \setminus J) = 0$ and also $\nu(G \setminus J) = 0$. Hence J also satisfies the left hand side of (18) because of nondegeneracy condition (2) (with $n_0 = 1$), since

$$\int 1_E(f_t(x))\, d\theta_\varepsilon(t) = (f_x)_*\theta^{\mathbb{N}}_\varepsilon(E).$$

This means that when $x \in J$ we have $f_t(x) \in J$ for θ_ε almost all $t \in \operatorname{supp}(\theta_\varepsilon)$. Since a set of θ_ε measure 1 is dense in $\operatorname{supp}(\theta_\varepsilon)$ (we are supposing θ_ε to be positive on open sets) and $f_t(x)$ varies continuously with t, we see that $f_t(x) \in J$ for all $t \in \operatorname{supp}(\theta_\varepsilon)$ because J is closed. We then have that the interior of J is nonempty by condition (1) on random perturbations and we may apply the methods of decomposition into connected components as before (Lemma 3.2). In this manner we construct an invariant domain inside J which, in turn, is inside a minimal invariant domain. This contradicts minimality and so we conclude that J must contain $\overline{U}_0 \cup \cdots \cup \overline{U}_{p-1}$. Thus we have $\nu(G) = \nu(J) = 1$ proving $\theta_\varepsilon^{\mathbb{N}} \times \nu$ to be F-ergodic.

Now, given $\varphi : M \to \mathbb{R}$ continuous we consider the map $\psi = \varphi \circ \pi$ from $T^{\mathbb{N}} \times M$ to $\mathbb{R}$, where $\pi : T^{\mathbb{N}} \times M \to M$ is the natural projection. The Ergodic Theorem then ensures

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \psi(F^j(\underline{t}, x)) = \int \psi \, d(\theta_\varepsilon^{\mathbb{N}} \times \nu)$$

for $\theta_\varepsilon^{\mathbb{N}} \times \nu$ almost all $(\underline{t}, x)$, which is just the same as

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f_{\underline{t}}^j(x)) = \int \varphi \, d\nu \tag{19}$$

for $\theta_\varepsilon^{\mathbb{N}} \times \nu$ almost all $(\underline{t}, x)$. Finally considering the ergodic basin $B(\nu)$, defined as the set of points $x \in M$ for which

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f_{\underline{t}}^j(x)) = \int \varphi \, d\nu$$

for all $\varphi \in C^0(M, \mathbb{R})$ and $\theta_\varepsilon^{\mathbb{N}}$ almost every $\underline{t} \in T^{\mathbb{N}}$, it is easy to see that $B(\nu)$ satisfies (18) in the place of G and we must have as before $B(\nu) \supset \overline{U}_0 \cup \cdots \cup \overline{U}_{p-1}$.

This shows that if another stationary absolutely continuous probability measure $\widetilde{\nu}$ is such that $\operatorname{supp}(\widetilde{\nu}) \subset \overline{U}_0 \cup \cdots \cup \overline{U}_{p-1}$, then the basins of ν and $\widetilde{\nu}$ must have nonempty intersection. Thus these measures must be equal. Moreover $\nu\big(B(\nu)\big) = 1$ and so, by absolute continuity, $m\big(B(\nu)\big) > 0$ and thus ν is a physical probability. $\square$

4. The number of physical measures

In this section we will prove that the number ℓ of physical measures is bounded by the number p of SRB measures. Moreover we will present examples of dynamical systems for which $\ell = p$ and $\ell < p$.

Let $\mu_1, \ldots, \mu_\ell$ be the physical measures supported on the minimal invariant domains in M, which exist by Lemmas 3.2 and 3.4 through 3.6. If μ is an absolutely continuous stationary measure, its restrictions to the minimal invariant domains of M, normalized when not equal to the constant zero measure, are absolutely continuous stationary measures by Lemma 3.3. After Lemma 3.6 these restrictions must be the physical measures $\mu_1, \ldots, \mu_\ell$ of the minimal domains. Hence μ must decompose into a linear combination of physical measures. Moreover, the union of $\operatorname{supp}(\mu_1), \ldots, \operatorname{supp}(\mu_\ell)$ must contain $\operatorname{supp}(\mu)$, except possibly for a μ null set. In fact, if the following set function

$$\mu - \mu\big(\operatorname{supp}(\mu_1)\big)\mu_1 - \cdots - \mu\big(\operatorname{supp}(\mu_\ell)\big)\mu_\ell$$

were nonzero, then its normalization μ' would be an absolutely continuous stationary measure, and the above decomposition could be applied to μ', thus giving another minimal domain inside $\operatorname{supp}(\mu)$. Clearly this cannot happen. We then have a convex linear decomposition

$$\mu = \alpha_1\mu_1 + \cdots + \alpha_\ell\mu_\ell \tag{20}$$

where $\alpha_i = \mu(\operatorname{supp}(\mu_i)) \geqslant 0$ and $\alpha_1 + \cdots + \alpha_\ell = 1$. We will see that this decomposition is uniquely defined.

We remark that so far we did not use more than the continuity of the map f. For the next result we assume that $f : M \to M$ is a C^2 nonuniformly expanding map whose orbits have slow approximation to the critical $\mathcal{C}$ (possibly the emptyset) with $m(\mathcal{C}) = 0$. This result contains the assertions of the first two items of Theorem A (if we think of $\mathcal{C} = \varnothing$) and Theorem C.

Proposition 4.1. — *If $\varepsilon > 0$ is small enough, then there exist physical measures $\mu_1^\varepsilon, \ldots, \mu_\ell^\varepsilon$ (with ℓ not depending on ε) such that*

(1) *for $x \in M$ there is a $\theta_\varepsilon^{\mathbb{N}}$ mod 0 partition $T_1(x), \ldots, T_\ell(x)$ of $T^{\mathbb{N}}$ such that*

$$\mu_i^\varepsilon = w^*\text{-}\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n-1} \delta_{f_{\underline{t}}^j(x)} \quad \textit{if and only if} \quad \underline{t} \in T_i(x);$$

(2) *for each $i = 1, \ldots, \ell$ we have*

$$\mu_i^\varepsilon = w^*\text{-}\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \int (f_{\underline{t}}^j)_*\big(m \mid B(\mu_i^\varepsilon)\big)\, d\theta_\varepsilon^{\mathbb{N}}(\underline{t}),$$

where $m \mid B(\mu_i^\varepsilon)$ is the normalized restriction of Lebesgue measure to $B(\mu_i^\varepsilon)$.

Proof. — Take $x \in M$ and let μ be a weak* accumulation point of the sequence $(\mu_n(x))_n$ defined in (17). We will prove that this is the only accumulation point of (17) by showing that the values of the $\alpha_1, \ldots, \alpha_\ell$ in decomposition (20) depend only on x and not on the subsequence that converges to μ. The definition of the average in (17) implies that there is a subset of parameter vectors $\underline{t} \in \operatorname{supp}(\theta_\varepsilon^{\mathbb{N}})$ with positive $\theta_\varepsilon^{\mathbb{N}}$ measure for which there is $j \geqslant 1$ such that $f_{\underline{t}}^j(x) \in \operatorname{supp}(\mu_i)$. We define for $i = 1, \ldots, \ell$

$$T_i(x) = \left\{ \underline{t} \in \operatorname{supp}(\theta_\varepsilon^{\mathbb{N}}) : f_{\underline{t}}^j(x) \in \operatorname{supp}(\mu_i) \quad \text{for some} \quad j \geqslant 1 \right\}.$$

We clearly have

$$T_i(x) = \bigcup_{j\geqslant 1} T_i^j(x) \quad \text{where} \quad T_i^j(x) = \{\underline{t} \in \operatorname{supp}(\theta_\varepsilon^{\mathbb{N}}) : f_{\underline{t}}^j(x) \in \operatorname{supp}(\mu_i)\}$$

and $T_i^j(x) \subset T_i^{j+1}(x)$ for all $i, j \geqslant 1$, since the supports of stationary measures are themselves invariant. In addition, since μ is a regular (Borel) probability measure, we may find for each $\eta > 0$ an open set U and a closed set K such that $K \subset \operatorname{supp}(\mu_i) \subset U$ with $\mu(U \setminus K) < \eta$ and $\mu(\partial U) = \mu(\partial K) = 0$. In fact, there is an at most countable number of δ-neighborhoods of $\operatorname{supp}(\mu_i)$ whose boundaries have positive μ measure, and likewise for the compacts coinciding with the complement of the δ-neighborhood of $M \setminus \operatorname{supp}(\mu_i)$. Then, taking $\alpha_i = \mu(\operatorname{supp}(\mu_i))$ we have

$$\begin{aligned}\alpha_i + \eta \geqslant \mu(U) &= \lim_{k\to+\infty} \frac{1}{n_k} \sum_{j=0}^{n_k-1} \theta_\varepsilon^{\mathbb{N}}\{\underline{t} \in T^{\mathbb{N}} : f_{\underline{t}}^j(x) \in U\} \\ &\geqslant \limsup_{k\to+\infty} \frac{1}{n_k} \sum_{j=0}^{n_k-1} \theta_\varepsilon^{\mathbb{N}}\big(T_i^j(x)\big)\end{aligned}$$

for some sequence of integers $n_1 < n_2 < n_3 < \cdots$, and likewise for

$$\begin{aligned}\alpha_i - \eta \leqslant \mu(K) &= \lim_{k\to+\infty} \frac{1}{n_k} \sum_{j=0}^{n_k-1} \theta_\varepsilon^{\mathbb{N}}\{\underline{t} \in T^{\mathbb{N}} : f_{\underline{t}}^j(x) \in K\} \\ &\leqslant \liminf_{k\to+\infty} \frac{1}{n_k} \sum_{j=0}^{n_k-1} \theta_\varepsilon^{\mathbb{N}}\big(T_i^j(x)\big),\end{aligned}$$

where $\eta > 0$ is arbitrary. This shows

$$\alpha_i = \mu(\operatorname{supp}(\mu_i)) = \lim_{k\to\infty} \frac{1}{n_k} \sum_{j=0}^{n_k-1} \theta_\varepsilon^{\mathbb{N}}\big(T_i^j(x)\big).$$

We also have

$$\theta_\varepsilon^{\mathbb{N}}\big(T_i(x)\big) = \lim_{j\to\infty} \theta_\varepsilon^{\mathbb{N}}\big(T_i^j(x)\big) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \theta_\varepsilon^{\mathbb{N}}\big(T_i^j(x)\big) = \alpha_i$$

which shows that the α_i depend only on the random orbits of x and not on the particular sequence $(n_k)_k$. Thus we see that the sequence of measures in (17) converges in the weak* topology. Moreover the sets $T_1(x), \ldots, T_\ell(x)$ are pairwise disjoint by definition and their total $\theta_\varepsilon^{\mathbb{N}}$ measure equals $\alpha_1 + \cdots + \alpha_\ell = 1$, thus forming a $\theta_\varepsilon^{\mathbb{N}}$ modulo zero partition of $T^{\mathbb{N}}$. We observe that if $\underline{t} \in T_i(x)$, then $f_{\underline{t}}^n(x) \in \operatorname{supp}(\mu_i) \subset B(\mu_i)$ for some $n \geqslant 1$ and $i = 1, \ldots, \ell$. This means this $\theta_\varepsilon^{\mathbb{N}}$ modulo zero partition of $T^{\mathbb{N}}$ satisfies the first item of the proposition.

Now fixing $i = 1, \ldots, \ell$, for all $x \in B(\mu_i)$ (the ergodic basin of μ_i) it holds that

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(f_{\underline{t}}^j(x)) = \int \varphi \, d\mu_i$$

for $\theta_\varepsilon^{\mathbb{N}}$ almost every $\underline{t} \in T^{\mathbb{N}}$. Recall that $m(B(\mu_i)) > 0$ by the definition of physical measure. Using dominated convergence and integrating both sides of the above equality twice, first with respect to the Lebesgue measure m, and then with respect to $\theta_\varepsilon^{\mathbb{N}}$, we arrive at the statement of item 2.

Recall that up until now the noise level $\varepsilon > 0$ was kept fixed. For small enough $\varepsilon > 0$ the measures $\mu_i = \mu_i^\varepsilon$ depend on the noise level, but we will see that the number of physical measures is constant.

Fixing $i \in \{1, \ldots, \ell\}$ we let x in the interior of $\operatorname{supp}(\mu_i^\varepsilon)$ be such that the orbit $(f^j(x))_j$ has infinitely many hyperbolic times. Recall that $f \equiv f_{t^*}$ is nonuniformly expanding (possibly with criticalities). Then there is a big enough hyperbolic time n so that $V_n(\underline{t}^*, x) \subset \operatorname{supp}(\mu_i^\varepsilon)$, by Proposition 2.6, where we take $\underline{t}^* = (t^*, t^*, t^*, \ldots)$. Since $t^* \in \operatorname{supp}(\theta_\varepsilon)$ and $\operatorname{supp}(\mu_i^\varepsilon)$ is invariant under f_t for all $t \in \operatorname{supp}(\theta_\varepsilon)$, we must have

$$f_{\underline{t}^*}^n\big(V_n(\underline{t}^*, x)\big) = B\big(f_{t^*}^n(x), \delta_1\big) \subset \operatorname{supp}(\mu_i^\varepsilon), \tag{21}$$

where $\delta_1 > 0$ is the constant given by Proposition 2.6 and $B\big(f_{t^*}^n(x), \delta_1\big)$ is the ball of radius δ_1 around $f_{t^*}^n(x)$.

On the one hand, we deduce that the number $\ell = \ell(\varepsilon)$ is bounded from above by some uniform constant N since M is compact. On the other hand, since each invariant set must contain some physical measure (by Lemma 3.4), we see that for $0 < \varepsilon' < \varepsilon$ there must be some physical measure $\mu^{\varepsilon'}$ with $\operatorname{supp}(\mu^{\varepsilon'}) \subset \operatorname{supp}(\mu^\varepsilon)$. In fact $\operatorname{supp}(\mu^\varepsilon)$ is invariant under f_t for every $t \in \operatorname{supp}(\theta_{\varepsilon'}) \subset \operatorname{supp}(\theta_\varepsilon)$. This means the number $\ell(\varepsilon)$ of physical measures is a nonincreasing function of $\varepsilon > 0$. Thus we conclude that there must be $\varepsilon_0 > 0$ such that $\ell = \ell(\varepsilon)$ is constant for $0 < \varepsilon < \varepsilon_0$, ending the proof of the proposition. $\square$

Remark 4.2. — Let us point out that from (21) one easily deduces that the Lesbesgue measure of the basin of each physical measure is uniformly bounded from below, since the support of such a measure is always contained in its basin.

Remark 4.3. — Observe that if the map $f : M \to M$ is topologically transitive, then every stationary measure must be supported on the whole of M, since the support is invariant and has nonempty interior. According to the discussion above, there must be only one such stationary measure, which must be physical.

We note that the number ℓ of physical measures for small $\varepsilon > 0$ and the number p of SRB measures for f are obtained by different existential arguments. It is natural to ask if there is any relation between ℓ and p.

Proposition 4.4. — *If $p \geqslant 1$ is the number of SRB measures of f and $\ell \geqslant 1$ is the number of physical measures of the random perturbation of f, then for $\varepsilon > 0$ small enough we have $\ell \leqslant p$.*

Proof. — We observe that $\mathrm{supp}(\mu^\varepsilon)$ is forward invariant under $f = f_{t^*}$ and, moreover, condition (3) holds for Lebesgue almost every x in $\mathrm{supp}(\mu^\varepsilon)$ because holds almost everywhere in M (by assumption) and $\mathrm{supp}(\mu^\varepsilon)$ has nonempty interior. Thus from [**ABV**, Theorem C] we assure the existence of at least one SRB measure μ with $\mathrm{supp}(\mu) \subset \mathrm{supp}(\mu^\varepsilon)$.

We have seen that each support of a physical measure μ^ε must contain at least the support of one SRB measure for the unperturbed map f. Since the number of SRB measures is finite we have $\ell \leqslant p$, where p is the number of those measures. □

The reverse inequality does not hold in general, as the following examples show: it is possible for two distinct SRB measures to have intersecting supports and, in this circumstance, the random perturbations will mix their basins and there will be some physical measure whose support overlaps the supports of both SRB measures.

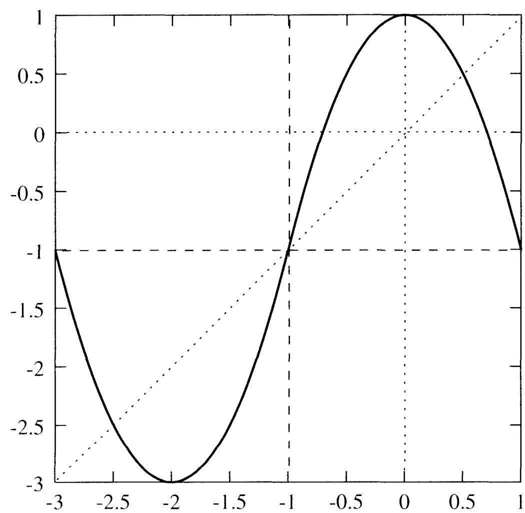


FIGURE 1. Map for which $1 = \ell < p = 2$

The first example is the map $f : [-3, 1] \to [-3, 1]$ whose graph is figure 1:

$$f(x) = \begin{cases} 1 - 2x^2 & \text{if } -1 \leqslant x \leqslant 1 \\ 2(x+2)^2 - 3 & \text{if } -3 \leqslant x \leqslant -1 \end{cases}.$$

The dynamics of f on $[-1, 1]$ and $[-3, -1]$ is conjugated to the tent map $T(x) = 1 - 2|x|$ on $[-1, 1]$. Thus understanding f as a circle map through the identification $S^1 = [-3, 1]/\{-3, 1\}$, this is a nonuniformly expanding map with a critical set satisfying conditions (S1)-(S3) and there are two ergodic absolutely continuous (thus SRB) invariant measures μ_1, μ_2 whose supports are $[-3, -1]$ and $[-1, 1]$ respectively. Moreover defining $\Phi(t) = R_t \circ f$, where $R_t : S^1 \to S^1$ is the rotation of angle t and $\theta_\varepsilon = (2\varepsilon)^{-1}(m \mid [-\varepsilon,\varepsilon])$ for small $\varepsilon > 0$, we have that $\{\Phi,(\theta_\varepsilon)_{\varepsilon>0}\}$ is a random perturbation satisfying nondegeneracy conditions (1) and (2). Since $\operatorname{supp}(\mu_1)\cap\operatorname{supp}(\mu_2) = \{-1\}$ we have that for $\varepsilon > 0$ small enough there must be a single physical measure μ^ε. Indeed, by property (P) any weak* accumulation point of a family of physical measures must have -1 in its support.

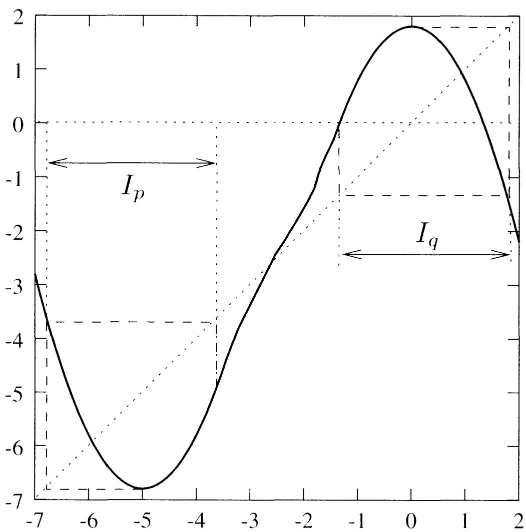


FIGURE 2. Map for which $\ell = p = 2$

The second example is defined on the interval $I = [-7,2]$. We take the map $q_a(x) = a - x^2$ on $[-2,2]$ for some parameter $a\in(1,2)$ satisfying Benedicks-Carleson conditions (see [**BC1**] and [**BC2**]), and the "same" map on $[-7,-3]$ conveniently conjugated: $p_a(x) = (x+5)^2 - 5 - a$. Then the two pieces of graph are glued together in such a way that we obtain a smooth map $f : I \to I$ sending I into its interior, as figure 2 shows. The intervals $I_q = [q_a^2(0), q_a(0)]$ and $I_p = [p_a(-5), p_a^2(-5)]$ are forward invariant for f, and then we can find slightly larger intervals $I_1 \supset I_p$ and $I_2 \supset I_q$ that become trapping regions for f. So, taking $\Phi(t) = f + t$, and θ_ε as in the previous example with $0 < \varepsilon < \varepsilon_0$ for some $\varepsilon_0 > 0$ small enough, then $\{\Phi,(\theta_\varepsilon)_\varepsilon\}$ is a random perturbation of f leaving the intervals I_1 and I_2 invariant by each $\Phi(t)$. Moreover, Lebesgue almost every $x \in I$ eventually arrives at one of these intervals. Then by [**BC1**] and [**BY1**] the map f is nonuniformly expanding and has two SRB measures with supports contained in each trapping region. Finally f admits two distinct physical measures whose supports are contained in I_1 and I_2 respectively, for $\varepsilon_0 > 0$ small enough. Moreover, these SRB measures are stochastically stable; see [**BV**].

5. Stochastic stability

In this section we will prove the first item of Theorem B and Theorem D. The second item of Theorem B may be obtained in the same way as Theorem D, if we think of $\mathcal{C}$ as being equal to the empty set and take into account Remark 2.4.

We start by proving the first item of Theorem B. Assume that f is a stochastically stable nonuniformly expanding local diffeomorphism. We know from Proposition 4.1 that there is a finite number of physical measures $\mu_1^\varepsilon, \dots \mu_\ell^\varepsilon$ and for each $x \in M$ there is a $\theta_\varepsilon^{\mathbb{N}}$ mod 0 partition $T_1(x), \dots, T_\ell(x)$ of $T^{\mathbb{N}}$ for which

$$\mu_i^\varepsilon = w^*\text{-}\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n-1} \delta_{f_{\underline{t}}^j(x)} \quad \text{for each} \quad \underline{t} \in T_i(x).$$

Furthermore, since we are taking f a local diffeomorphism, then $\log \|(Df)^{-1}\|$ is a continuous map. Thus, we have for each $x \in M$ and $\theta_\varepsilon^{\mathbb{N}}$ almost every $\underline{t} \in T^{\mathbb{N}}$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \log \|Df\big(f_{\underline{t}}^j(x)\big)^{-1}\| = \int \log \|(Df)^{-1}\| d\mu_i^\varepsilon$$

for some physical measure μ_i^ε with $1 \leqslant i \leqslant \ell$. Hence, for proving the nonuniform expansion of f on random orbits it suffices to show that there is $c_0 > 0$ such that if $\mu^\varepsilon = \mu_i^\varepsilon$ for some $1 \leqslant i \leqslant \ell$ then

$$\int \log \|(Df)^{-1}\| d\mu^\varepsilon < c_0 \quad \text{for small } \varepsilon > 0.$$

Lemma 5.1. — *Let $\varphi\colon M \to \mathbb{R}$ be a continuous map. Given $\delta > 0$ there is $\varepsilon_0 > 0$ such that if $\varepsilon \leqslant \varepsilon_0$, then*

$$\left| \int \varphi d\mu^\varepsilon - \int \varphi d\mu_\varepsilon \right| < \delta,$$

for some absolutely continuous f-invariant probability measure μ_ε.

Proof. — We will use the following auxiliary result: *Let X be a compact metric space, $K \subset X$ a closed (compact) subset and $(x_t)_{t>0}$ a curve in X (not necessarily continuous) such that all its accumulation points (as $t \to 0^+$) lie in K. Then for every open neighborhood U of K there is $t_0 > 0$ such that $x_t \in U$ for every $0 < t < t_0$.* Indeed, supposing not, there is a sequence $(t_n)_n$ with $t_n \to 0^+$ when $n \to \infty$ such that $x_{t_n} \notin U$. Since X is compact this means that $(x_t)_{t>0}$ has some accumulation point in $X \setminus U$, thus outside K, contrary to the assumption.

Now, the space $X = \mathbb{P}(M)$ of all probability measures in M is a compact metric space with the weak* topology, and the convex hull K of the (finitely many) SRB measures of f is closed. Hence, considering the curve $(\mu^\varepsilon)_\varepsilon$ in $\mathbb{P}(M)$, we are in the context of the above result, since we are supposing f to be stochastically stable.

A metric on X topologically equivalent to the weak* topology may be given by

$$\mathrm{d}_{\mathbb{P}}(\mu,\nu)=\sum_{k=1}^{\infty}\frac{1}{2^n}\left|\int\varphi_n\,d\mu-\int\varphi_n\,d\nu\right|$$

where $\mu,\nu\in\mathbb{P}(M)$ and $(\varphi_n)_{n\geqslant1}$ is a dense sequence of functions in $C^0(M,\mathbb{R})$, see [**Ma**].

Let $\varphi:M\to\mathbb{R}$ continuous be given and let us fix some $\delta>0$. There must be $n\in\mathbb{N}$ such that $\|\varphi-\varphi_n\|_0<\delta/3$ and, by the auxiliary result in the beginning of the proof, there exists, for some $\varepsilon_0>0$ and every $0<\varepsilon<\varepsilon_0$, a probability measure $\mu_\varepsilon\in\mathbb{P}(M)$ for which $\mathrm{d}_{\mathbb{P}}(\mu^\varepsilon,\mu_\varepsilon)<\delta(3\cdot2^n)^{-1}$. This in particular means that

$$\frac{1}{2^n}\left|\int\varphi_n\,d\mu^\varepsilon-\int\varphi_n\,d\mu_\varepsilon\right|<\frac{\delta}{3\cdot2^n},$$

by the definition of the distance $\mathrm{d}_{\mathbb{P}}$, which implies

$$\left|\int\varphi_n\,d\mu^\varepsilon-\int\varphi_n\,d\mu_\varepsilon\right|<\frac{\delta}{3}.$$

Hence we get

$$\begin{aligned}&\left|\int\varphi\,d\mu^\varepsilon-\int\varphi\,d\mu_\varepsilon\right|\leqslant\\&\quad\leqslant\left|\int\varphi\,d\mu^\varepsilon-\int\varphi_n\,d\mu^\varepsilon\right|+\left|\int\varphi_n\,d\mu^\varepsilon-\int\varphi_n\,d\mu_\varepsilon\right|+\left|\int\varphi_n\,d\mu_\varepsilon-\int\varphi\,d\mu_\varepsilon\right|\\&\quad<\frac{\delta}{3}+\frac{\delta}{3}+\frac{\delta}{3}=\delta,\end{aligned}$$

which completes the proof of the lemma. $\square$

Now we take $\varphi=\log\|(Df)^{-1}\|$ and $\delta=c/2$ in the previous lemma, where $c>0$ is the constant given by the nonuniform expansion of f (recall (3)). For each $\varepsilon\leqslant\varepsilon_0$ let μ_ε be the measure given by Lemma 5.1. Since property (P) holds, there are real numbers $w_1(\varepsilon),\dots,w_p(\varepsilon)\geqslant0$ with $w_1(\varepsilon)+\cdots+w_p(\varepsilon)=1$ for which $\mu_\varepsilon=w_1(\varepsilon)\mu_1+\cdots+w_p(\varepsilon)\mu_p$. Since each μ_i is an SRB measure for $1\leqslant i\leqslant p$, we have for Lebesgue almost every $x\in B(\mu_i)$

$$\int\log\|(Df)^{-1}\|d\mu_i=\lim_{n\to+\infty}\frac{1}{n}\sum_{j=0}^{n-1}\log\|Df(f^j(x))^{-1}\|\leqslant-c<0.$$

This implies

$$\int\log\|(Df)^{-1}\|d\mu_\varepsilon\leqslant-c,$$

and so, by Lemma 5.1 and the choice of δ,

$$\int\log\|(Df)^{-1}\|d\mu^\varepsilon\leqslant-c/2.$$

This completes the proof of the first item of Theorem B.

Now we go into the proof of Theorem D. In order to prove that f is stochastically stable, and taking into account property (P), it suffices to prove that the weak* accumulation points of any family $(\mu^\varepsilon)_{\varepsilon>0}$, where each μ^ε is a physical measure of level ε, are absolutely continuous with respect to the Lebesgue measure. Let μ^ε be a physical measure of level ε for some small $\varepsilon > 0$ and define for each $n \geqslant 1$

$$\mu_n^\varepsilon = \frac{1}{n}\sum_{j=0}^{n-1}\frac{1}{m(B(\mu^\varepsilon))}\int (f_{\underline{t}}^j)_*\big(m \mid B(\mu^\varepsilon)\big)\,d\theta_\varepsilon^{\mathbb{N}}(\underline{t}).$$

We know from Proposition 4.1 that each μ^ε is the weak* limit of the sequence $(\mu_n^\varepsilon)_n$. We will prove Theorem D by providing some useful estimates on the densities of the measures μ_n^ε. Define for each $\underline{t} \in T^{\mathbb{N}}$ and $n \geqslant 1$

$$H_n(\underline{t}) = \{x \in B(\mu^\varepsilon)\colon\ n \text{ is a } (\alpha,\delta)\text{-hyperbolic time for } (\underline{t}, x)\},$$

and

$$H_n^*(\underline{t}) = \{x \in B(\mu^\varepsilon)\colon\ n \text{ is the first } (\alpha,\delta)\text{-hyperbolic time for } (\underline{t}, x)\}.$$

$H_n^*(\underline{t})$ is precisely the set of those points $x \in B(\mu^\varepsilon)$ for which $h_\varepsilon(\underline{t}, x) = n$ (recall the definition of the map h_ε). For $n, k \geqslant 1$ we also define $R_{n,k}(\underline{t})$ as the set of those points $x \in M$ for which n is a (α,δ)-hyperbolic time and $n+k$ is the first (α,δ)-hyperbolic time after n, i.e.

$$R_{n,k}(\underline{t}) = \big\{x \in H_n(\underline{t})\colon\ f_{\underline{t}}^n(x) \in H_k^*(\sigma^n \underline{t})\big\},$$

where $\sigma\colon T^{\mathbb{N}} \to T^{\mathbb{N}}$ is the shift map $\sigma(t_1, t_2, \dots) = (t_2, t_3, \dots)$. Considering the measures

$$\nu_n^\varepsilon = \int (f_{\underline{t}}^n)_*\big(m \mid H_n(\underline{t})\big)d\theta_\varepsilon^{\mathbb{N}}(\underline{t})$$

and

$$\eta_n^\varepsilon = \sum_{k=2}^{\infty}\sum_{j=1}^{k-1}\int (f_{\underline{t}}^{n+j})_*\big(m \mid R_{n,k}(\underline{t})\big)d\theta_\varepsilon^{\mathbb{N}}(\underline{t}),$$

we may write

$$\mu_n^\varepsilon \leqslant \frac{1}{n}\sum_{j=0}^{n-1}\frac{1}{m(B(\mu^\varepsilon))}(\nu_j^\varepsilon + \eta_j^\varepsilon).$$

Proposition 5.2. — *There is a constant $C_2 > 0$ such that for every $n \geqslant 0$ and $\underline{t} \in T^{\mathbb{N}}$*

$$\frac{d}{dm}(f_{\underline{t}}^n)_*\big(m \mid H_n(\underline{t})\big) \leqslant C_2.$$

Proof. — Take $\delta_1 > 0$ given by Proposition 2.6. It is sufficient to prove that there is some uniform constant $C > 0$ such that if A is a Borel set in M with diameter smaller than $\delta_1/2$ then

$$m\big(f_{\underline{t}}^{-n}(A) \cap H_n(\underline{t})\big) \leqslant Cm(A).$$

Let A be a Borel set in M with diameter smaller than $\delta_1/2$ and B an open ball of radius $\delta_1/2$ containing A. We may write

$$f_{\underline{t}}^{-n}(B) = \bigcup_{k \geqslant 1} B_k,$$

where $(B_k)_{k\geqslant 1}$ is a (possibly finite) family of two-by-two disjoint open sets in M. Discarding those B_k that do not intersect $H_n(\underline{t})$, we choose for each $k \geqslant 1$ a point $x_k \in H_n(\underline{t}) \cap B_k$. For $k \geqslant 1$ let $V_n(\underline{t}, x_k)$ be the neighborhood of x_k in M given by Proposition 2.6. Since B is contained in $B\big(f_{\underline{t}}^n(x_k), \delta_1\big)$, the ball of radius δ_1 around $f_{\underline{t}}^n(x_k)$, and $f_{\underline{t}}^n$ is a diffeomorphism from $V_n(\underline{t}, x_k)$ onto $B\big(f_{\underline{t}}^n(x_k), \delta_1\big)$, we must have $B_k \subset V_n(\underline{t}, x_k)$ (recall that by our choice of B_k we have $f_{\underline{t}}^n(B_k) \subset B$). As a consequence of this and Corollary 2.7, we have for every k that the map $f_{\underline{t}}^n \mid B_k \colon B_k \to B$ is a diffeomorphism with bounded distortion:

$$\frac{1}{C_1} \leqslant \frac{|\det Df_{\underline{t}}^n(y)|}{|\det Df_{\underline{t}}^n(z)|} \leqslant C_1$$

for all $y, z \in B_k$. This finally gives

$$\begin{aligned} m\big(f_{\underline{t}}^{-n}(A) \cap H_n(\underline{t})\big) &\leqslant \sum_k m\big(f_{\underline{t}}^{-n}(A \cap B) \cap B_k\big) \\ &\leqslant \sum_k C_1 \frac{m(A \cap B)}{m(B)} m(B_k) \\ &\leqslant C_2 m(A), \end{aligned}$$

where $C_2 > 0$ is a constant only depending on C_1, on the volume of the ball B of radius $\delta_1/2$, and on the volume of M. □

It follows from Proposition 5.2 that

$$\frac{d\nu_n^\varepsilon}{dm} \leqslant C_2 \tag{22}$$

for every $n \geqslant 0$ and small $\varepsilon > 0$. Our goal now is to control the density of the measures η_n^ε in such a way that we may assure the absolute continuity of the weak* accumulation points of the measures μ^ε when ε goes to zero.

Proposition 5.3. — *Given $\zeta > 0$, there is $C_3(\zeta) > 0$ such that for every $n \geqslant 0$ and $\varepsilon > 0$ we may bound η_n^ε by the sum of two non-negative measures, $\eta_n^\varepsilon \leqslant \omega^\varepsilon + \rho^\varepsilon$, with*

$$\frac{d\omega^\varepsilon}{dm} \leqslant C_3(\zeta) \quad \textit{and} \quad \rho^\varepsilon(M) < \zeta.$$

Proof. — Let A be some Borel set in M. We have for each $n \geqslant 0$

$$
\begin{aligned}
\eta_n^\varepsilon(A) &= \sum_{k=2}^{\infty}\sum_{j=1}^{k-1}\int m\big(f_{\underline{t}}^{-n-j}(A)\cap R_{n,k}(\underline{t})\big)d\theta_\varepsilon^{\mathbb{N}}(\underline{t})\\
&\leqslant \sum_{k=2}^{\infty}\sum_{j=1}^{k-1}\int m\big(f_{\underline{t}}^{-n}\big(f_{\sigma^n\underline{t}}^{-j}(A)\cap H_k^*(\sigma^n\underline{t})\big)\cap H_n(\underline{t})\big)d\theta_\varepsilon^{\mathbb{N}}(\underline{t})\\
&\leqslant \sum_{k=2}^{\infty}\sum_{j=1}^{k-1}C_2\int m\big(f_{\underline{t}}^{-j}(A)\cap H_k^*(\underline{t})\big)d\theta_\varepsilon^{\mathbb{N}}(\underline{t}).
\end{aligned}
$$

(in this last inequality we used Proposition 5.2 and the fact that $\theta_\varepsilon^{\mathbb{N}}$ is σ-invariant). Let now $\zeta > 0$ be some fixed small number. Since we are assuming $(h_\varepsilon)_\varepsilon$ with uniform L^1-tail, then there is some integer $N = N(\zeta)$ for which

$$
\sum_{j=N}^{\infty} k\int m\big(H_k^*(\underline{t})\big)d\theta_\varepsilon^{\mathbb{N}}(\underline{t}) < \frac{\zeta}{C_2}.
$$

We take

$$
\omega^\varepsilon = C_2\sum_{k=2}^{N-1}\sum_{j=1}^{k-1}\int (f_{\underline{t}}^j)_*\big(m \mid H_k^*(\underline{t})\big)d\theta_\varepsilon^{\mathbb{N}}(\underline{t})
$$

and

$$
\rho^\varepsilon = C_2\sum_{k=N}^{\infty}\sum_{j=1}^{k-1}\int (f_{\underline{t}}^j)_*\big(m \mid H_k^*(\underline{t})\big)d\theta_\varepsilon^{\mathbb{N}}(\underline{t}).
$$

For this last measure we have

$$
\rho^\varepsilon(M) = C_2\sum_{k=N}^{\infty}\sum_{j=1}^{k-1}\int m\big(H_k^*(\underline{t})\big)d\theta_\varepsilon^{\mathbb{N}}(\underline{t}) \leqslant C_2\sum_{k=N}^{\infty} k\int m\big(H_k^*(\underline{t})\big)d\theta_\varepsilon^{\mathbb{N}}(\underline{t}) < \zeta.
$$

On the other hand, it follows from the definition of (α,δ)-hyperbolic times that there is some constant $a = a(N) > 0$ such that $\operatorname{dist}\big(H_k(\underline{t}),\mathcal{C}\big) \geqslant a$ for $1 \leqslant k \leqslant N$. Defining $\Delta \subset M$ as the set of those points in M whose distance to $\mathcal{C}$ is greater than a, we have

$$
\omega^\varepsilon \leqslant C_2\sum_{k=2}^{N-1}\sum_{j=1}^{k-1}\int (f_{\underline{t}}^j)_*(m \mid \Delta)\, d\theta_\varepsilon^{\mathbb{N}}(\underline{t}),
$$

and this last measure has density bounded by some uniform constant, as long as we take the maps f_t in a sufficiently small neighborhood of f in the C^1 topology. $\square$

It follows from Remark 4.2, Proposition 5.3 and (22) that the weak* accumulation points of μ^ε when $\varepsilon \to 0$ cannot have singular part, thus being absolutely continuous with respect to the Lebesgue measure. Moreover, the weak* accumulation points of a family of stationary measures are always f-invariant measures, cf. Remark 3.1. This together with (P) gives the stochastic stability of f.

6. Applications

In this section we will apply Theorems B and D to certain classes of nonuniformly expanding maps. Before we describe the examples we have in mind let us give a practical criterion for proving that the family of hyperbolic time maps $(h_\varepsilon)_\varepsilon$ has uniform L^1-tail.

If we look at the proof of Proposition 2.3 we see that what we did was fixing some positive number c_0 smaller than c, and then, for $\theta_\varepsilon^{\mathbb{N}} \times m$ almost every $(\underline{t}, x) \in T^{\mathbb{N}} \times M$, we took a positive integer $N_\varepsilon = N_\varepsilon(\underline{t}, x)$ for which

$$\sum_{j=0}^{N_\varepsilon - 1} \log \|Df(f_{\underline{t}}^j(x))^{-1}\| \leqslant -c_0 N_\varepsilon \quad \text{and} \quad \sum_{j=0}^{N_\varepsilon - 1} -\log \operatorname{dist}_\delta(f_{\underline{t}}^j(x), \mathcal{C}) \leqslant \gamma N_\varepsilon,$$

for suitable choices of $\delta > 0$ and $\gamma > 0$. This permits us to introduce a map

$$N_\varepsilon \colon T^{\mathbb{N}} \times M \longrightarrow \mathbb{Z}^+$$

whose existence provides a first hyperbolic time map

$$h_\varepsilon \colon T^{\mathbb{N}} \times M \longrightarrow \mathbb{Z}^+ \quad \text{with} \quad h_\varepsilon \leqslant N_\varepsilon$$

(recall the proof of Proposition 2.3). Thus, the integrability of the map h_ε is implied by the integrability of the map N_ε, which is in practice easier to handle.

Remark 6.1. — In the examples we are going to study below we will show that there is a sequence of positive real numbers $(a_k^\varepsilon)_k$ for which

$$(\theta_\varepsilon^{\mathbb{N}} \times m)\left(\left\{(\underline{t}, x) \in T^{\mathbb{N}} \times M \colon N_\varepsilon(\underline{t}, x) > k\right\}\right) \leqslant a_k^\varepsilon \quad \text{and} \quad \sum_{k=1}^{\infty} k a_k^\varepsilon < \infty,$$

This gives the integrability of h_ε with respect to the measure $\theta_\varepsilon^{\mathbb{N}} \times m$. The fact the family $(h_\varepsilon)_\varepsilon$ has uniform L^1-tail can be proved by showing that the sequence $(a_k^\varepsilon)_k$ may be chosen not depending on $\varepsilon > 0$.

Now we are ready for the applications of Theorems B and D. We will describe first a class of local diffeomorphisms introduced in [**ABV**, Appendix A] that satisfies the hypotheses of Theorem B, and then a class of maps (with critical sets) introduced in [**Vi1**] satisfying the hypotheses of Theorem D.

6.1. Local diffeomorphisms. — Now we follow [**ABV**, Appendix A] and describe robust classes of maps (open in the C^2 topology) that are nonuniformly expanding local diffeomorphisms and stochastically stable. Let M be a compact Riemannian manifold and consider

$$\begin{aligned} \Phi : T &\longrightarrow C^2(M, M) \\ t &\longmapsto f_t \end{aligned}$$

a continuous family of C^2 maps, where T is a metric space. We begin with an essentially combinatorial lemma.

Lemma 6.2. — *Let $p, q \geqslant 1$ be integers and $\sigma > q$ a real number. Assume M admits a measurable cover $\{B_1, \dots, B_p, B_{p+1}, \dots, B_{p+q}\}$ such that for all $t \in T$ it holds*

(1) $|\det Df_t(x)| \geqslant \sigma$ *for all* $x \in B_{p+1} \cup \dots \cup B_{p+q}$*;*

(2) $(f_t \mid B_i)$ *is injective for all* $i = 1, \dots, p$.

Then there is $\zeta > 0$ such that for every Borel probability θ on T we have

$$\#\{0 \leqslant j < n : f_{\underline{t}}^j(x) \in B_1 \cup \dots \cup B_p\} \geqslant \zeta n \tag{23}$$

for $\theta^{\mathbb{N}} \times m$ almost all $(\underline{t}, x) \in T^{\mathbb{N}} \times M$ and large enough $n \geqslant 1$. Moreover the set I_n of points $(\underline{t}, x) \in T^{\mathbb{N}} \times M$ whose orbits do not spend a fraction ζ of the time in $B_1 \cup \dots \cup B_p$ up to iterate n is such that $(\theta^{\mathbb{N}} \times m)(I_n) \leqslant \tau^n$ for some $0 < \tau < 1$ and for large $n \geqslant 1$.

Proof. — Let us fix $n \geqslant 1$ and $\underline{t} \in T^{\mathbb{N}}$. For a sequence

$$\underline{i} = (i_0, \dots, i_{n-1}) \in \{1, \dots, p+q\}^n$$

we write

$$[\underline{i}] = B_{i_0} \cap (f_{\underline{t}}^1)^{-1}(B_{i_1}) \cap \dots \cap (f_{\underline{t}}^{n-1})^{-1}(B_{i_{n-1}})$$

and define $g(\underline{i}) = \#\{0 \leqslant j < n : i_j \leqslant p\}$.

We start by observing that for $\zeta > 0$ the number of sequences $\underline{i}$ such that $g(\underline{i}) < \zeta n$ is bounded by

$$\sum_{k < \zeta n} \binom{n}{k} p^k q^{n-k} \leqslant \sum_{k \leqslant \zeta n} \binom{n}{k} p^{\zeta n} q^n.$$

Using Stirling's formula (cf. [**BV**, Section 6.3]) the expression on the right hand side is bounded by $(e^{\gamma} p^{\zeta} q)^n$, where $\gamma > 0$ depends only on ζ and $\gamma(\zeta) \to 0$ when $\zeta \to 0$.

Assumptions (1) and (2) ensure $m([\underline{i}]) \leqslant \sigma^{-(1-\zeta)n}$ (recall that $m(M) = 1$). Hence the measure of the union $I_n(\underline{t})$ of all the sets $[\underline{i}]$ with $g(\underline{i}) < \zeta n$ is bounded by

$$\sigma^{-(1-\zeta)n} (e^{\gamma} p^{\zeta} q)^n.$$

Since $\sigma > q$ we may choose ζ so small that $e^{\gamma} p^{\zeta} q < \sigma^{(1-\zeta)}$. Then $m(I_n(\underline{t})) \leqslant \tau^n$ with $\tau = e^{\gamma + \zeta - 1} \cdot p^{\zeta} \cdot q < 1$ for big enough $n \geqslant N$. Note that τ and N do not depend on $\underline{t}$. Setting

$$I_n = \bigcup\nolimits_{\underline{t} \in T^{\mathbb{N}}} (\{\underline{t}\} \times I_n(\underline{t}))$$

we also have $(\theta^{\mathbb{N}} \times m)(I_n) \leqslant \tau^n$ for all big $n \geqslant N$ and for every Borel probability θ on T, by Fubini's Theorem. Since $\sum_n (\theta^{\mathbb{N}} \times m)(I_n) < \infty$ then Borel-Cantelli's Lemma implies

$$(\theta^{\mathbb{N}} \times m) \left(\bigcap\nolimits_{n \geqslant 1} \bigcup\nolimits_{k \geqslant n} I_k \right) = 0$$

and this means that $\theta^{\mathbb{N}} \times m$ almost every $(\underline{t}, x) \in T^{\mathbb{N}} \times M$ satisfies (23). □

Lemma 6.3. — *Let $\{B_1, \dots, B_p, B_{p+1}, \dots, B_{p+q}\}$ be a measurable cover of M satisfying conditions* (1) *and* (2) *of Lemma 6.2. For $0 < \lambda < 1$ there are $\eta > 0$ and $c_0 > 0$ such that, if f_t also satisfies for all $t \in T$*

(3) $\|Df_t(x)^{-1}\| \leqslant \lambda < 1$ *for* $x \in B_1, \ldots, B_p$;
(4) $\|Df_t(x)^{-1}\| \leqslant 1+\eta$ *for* $x \in B_{p+1}, \ldots, B_{p+q}$;

then we have for $f \equiv f_{t^*}$, *where* t^* *is some given point in* T,

$$\limsup_{n\to+\infty} \frac{1}{n}\sum_{j=0}^{n-1} \log \|Df(f_{\underline{t}}^j(x))^{-1}\| \leqslant -c_0 \tag{24}$$

for $\theta^{\mathbb{N}} \times m$ *almost all* $(\underline{t}, x) \in T^{\mathbb{N}} \times M$, *where* θ *is any Borel probability measure on* T. *Moreover the first hyperbolic time map* $h : T^{\mathbb{N}} \times M \to \mathbb{Z}^+$ *satisfies*

$$(\theta^{\mathbb{N}} \times m)\{(\underline{t}, x) \in T^{\mathbb{N}} \times M : h(\underline{t}, x) > k\} \leqslant a_k \quad \text{and} \quad \sum_{k=1}^{\infty} k a_k < \infty$$

with $(a_k)_k$ *independent of the choice of* θ.

Proof. — Let $\zeta > 0$ be the constant provided by Lemma 6.2. We fix $\eta > 0$ sufficiently small so that $\lambda^\zeta(1+\eta) \leqslant e^{-c_0}$ holds for some $c_0 > 0$ and take $(\underline{t}, x)$ satisfying (23). Conditions (3) and (4) now imply

$$\prod_{j=0}^{n-1} \|Df(f_{\underline{t}}^j(x))^{-1}\| \leqslant \lambda^{\zeta n}(1+\eta)^{(1-\zeta)n} \leqslant e^{-c_0 n}. \tag{25}$$

for large enough n. This means (25) holds for $\theta^{\mathbb{N}} \times m$ almost every $(\underline{t}, x) \in T^{\mathbb{N}} \times M$.

We observe that if $h(\underline{t}, x) = k$, then $1 \leqslant n < k$ cannot be hyperbolic times for $(\underline{t}, x)$. Hence $(\underline{t}, x) \in I_n$ for all $n = 1, \ldots, k-1$. In particular

$$(\theta^{\mathbb{N}} \times m)\{(\underline{t}, x) \in T^{\mathbb{N}} \times M : h(\underline{t}, x) = k\} \leqslant (\theta^{\mathbb{N}} \times m)(I_{k-1}) \equiv a_k$$

and $\sum_k k a_k \leqslant \sum_k k\tau^{k-1} < \infty$. □

Now we will show that families of C^2 maps satisfying conditions (1) through (4) of Lemmas 6.2 and 6.3 contain open sets of families in the C^2 topology. Let M be a n-dimensional torus $\mathbb{T}^n$ and $f_0 : M \to M$ a uniformly expanding map: there exists $0 < \lambda < 1$ such that $\|Df_0(x)v\| \geqslant \lambda^{-1}\|v\|$ for all $x \in M$ and $v \in T_xM$. Let also W be some small compact domain in M where $f_0 \mid W$ is injective. Observe that f_0 is a volume expanding local diffeomorphism due to the uniform expansion.

Modifying f_0 by an isotopy inside W we may obtain a map f_1 which coincides with f_0 outside W, is volume expanding in M, i.e., $|\det Df_1(x)| > 1$ for all $x \in M$, and has bounded contraction on W near 1: $\|Df_1(x)^{-1}\| \leqslant 1+\eta$ for every $x \in W$ and some $\eta > 0$ small. This new map f_1 may be taken C^1 close to f_0 and we may consider a C^2 map f_2 arbitrarily C^1 close to f_1.

Now any map f in a small enough C^2 neighborhood of f_2 admits $\sigma > 1$ such that $|\det Df(x)| \geqslant \sigma$ for all $x \in M$ and, for x outside W, we have $\|Df(x)^{-1}\| \leqslant \lambda$. If the C^2 neighborhood is taken sufficiently small then we maintain $\|Df(x)^{-1}\| \leqslant 1+\eta$ for $x \in W$ and for some small $\eta > 0$. Let us take $B_1, \ldots, B_p, B_{p+1} = W$ a partition of M into measurable sets where the restriction $f \mid B_i$ is injective for $i = 1, \ldots, p+1$. Then any continuous family of C^2 maps $\Phi : T \to C^2(M, M)$ together with a family $(\theta_\varepsilon)_{\varepsilon>0}$ of Borel probability measures in the metric space T, satisfying $\operatorname{supp}(\theta_\varepsilon) \to \{t^*\}$ when $\varepsilon \to 0$ and $f_{t^*} \equiv f$, for some $t^* \in T$, is such that f is nonuniformly expanding for random orbits and $(h_\varepsilon)_{\varepsilon>0}$ has uniform L^1-tail — by Lemma 6.3 with $q = 1$ and $T = \operatorname{supp}(\theta_\varepsilon)$ for small enough $\varepsilon > 0$. Theorem B then shows

Corollary 6.4. — *There are open sets $\mathcal{U} \subset C^2(M, M)$ such that every $f \in \mathcal{U}$ is a stochastically stable nonuniformly expanding local diffeomorphism.*

6.2. Viana maps. — In what follows we study the class of nonuniformly expanding maps with critical sets introduced by M. Viana and prove Theorem E.

6.2.1. Nonuniform expansion. — Let $\widehat{f}$ be defined as in Subsection 1.1.2. The results in [**Vi1**] show that if the map f is sufficiently close to $\widehat{f}$ in the C^3 topology then f has two positive Lyapunov exponents almost everywhere: there is a constant $\lambda > 0$ for which

$$\liminf_{n\to+\infty} \frac{1}{n} \log \|Df^n(s, x)v\| \geqslant \lambda$$

for Lebesgue almost every $(s, x) \in S^1 \times I$ and every non-zero $v \in T_{(s,x)}(S^1 \times I)$. As mentioned in [**ABV**], this does not necessarily imply that f is nonuniformly expanding. However a slight modification in Viana's arguments enables us to prove the nonuniform expansion of f.

For the sake of clearness, we start by assuming that f has the special form

$$f(s, x) = (g(s), q(s, x)), \quad \text{with} \quad \partial_x q(s, x) = 0 \quad \text{if and only if} \quad x = 0, \tag{26}$$

and describe how the conclusions in [**Vi1**] are obtained for each C^2 map f satisfying

$$\|f - \widehat{f}\|_{C^2} \leqslant \alpha \quad \text{on} \quad S^1 \times I. \tag{27}$$

Then we explain how these conclusions extend to the general case, using the existence of a central invariant foliation, and we show how the results in [**Vi1**] give the nonuniform expansion and slow approximation of orbits to the critical set for each map f as in (27).

The estimates on the derivative rely on a statistical analysis of the returns of orbits to the neighborhood $S^1 \times (-\sqrt{\alpha}, \sqrt{\alpha}\,)$ of the critical set $\mathcal{C} = \{(s, x) : x = 0\}$. We set

$$J(0) = I \setminus (-\sqrt{\alpha}, \sqrt{\alpha}) \quad \text{and} \quad J(r) = \{x \in I : |x| < e^{-r}\} \quad \text{for } r \geqslant 0.$$

From here on we only consider points $(s, x) \in S^1 \times I$ whose orbit does not hit the critical set $\mathcal{C}$. This constitues no restriction in our results, since the set of those points has full Lebesgue measure.

For each integer $j \geqslant 0$ we define $(s_j, x_j) = f^j(s, x)$ and

$$r_j(s, x) = \min\{r \geqslant 0 : x_j \in J(r)\}.$$

Consider, for some small constant $0 < \eta < 1/4$,

$$G = \left\{ 0 \leqslant j < n : r_j(s,x) \geqslant \left(\frac{1}{2} - 2\eta\right) \log \frac{1}{\alpha} \right\}.$$

Fix some integer $n \geqslant 1$ sufficiently large (only depending on $\alpha > 0$). The results in [**Vi1**] show that if we take

$$B_2(n) = \big\{ (s,x) : \text{ there is } 1 \leqslant j < n \text{ with } x_j \in J([\sqrt{n}])\, \big\},$$

where $[\sqrt{n}]$ is the integer part of $\sqrt{n}$, then we have

$$m(B_2(n)) \leqslant \text{const}\, e^{-\sqrt{n}/4} \tag{28}$$

and, for every small $c > 0$ (only depending on the quadratic map Q),

$$\log \prod_{j=0}^{n-1} |\partial_x q(s_j, x_j)| \geqslant 2cn - \sum_{j \in G} r_j(s,x) \quad \text{for } (s,x) \notin B_2(n), \tag{29}$$

see [**Vi1**, pp. 75 & 76]. Moreover, if we define for $\gamma > 0$

$$B_1(n) = \left\{ (s,x) \notin B_2(n) : \sum_{j \in G} r_j(s,x) \geqslant \gamma n \right\},$$

then, for small $\gamma > 0$, there is a constant $\xi > 0$ for which

$$m\big(B_1(n)\big) \leqslant e^{-\xi n}, \tag{30}$$

see [**Vi1**, p. 77]. Taking into account the definitions of $J(r)$ and r_j, this shows that if we take $\delta = (1/2 - 2\eta)\log(1/\alpha)$, then

$$\sum_{j=0}^{n-1} -\log \operatorname{dist}_\delta(f^j(x), \mathcal{C}) \leqslant \gamma n \quad \text{for } (s,x) \notin B_1(n) \cup B_2(n).$$

This in particular gives that almost all orbits have slow approximation to $\mathcal{C}$.

On the other hand, we have for $(s,x) \in S^1 \times I$

$$\big(Df(s,x)\big)^{-1} = \frac{1}{\partial_x q(s,x) \partial_s g(s)} \begin{pmatrix} \partial_x q(s,x) & 0 \\ -\partial_s q(s,x) & \partial_s g(s) \end{pmatrix}. \tag{31}$$

Since all the norms are equivalent in finite dimensional Banach spaces, it is no restriction for our purposes to take the norm of $\big(Df(s,x)\big)^{-1}$ as the maximum of the absolute values of its entries. From (26) and (27) we deduce that for small α

$$|\partial_s g| \geqslant d - \alpha, \quad |\partial_s q| \leqslant \alpha |b'| + \alpha \leqslant 8\alpha \quad \text{and} \quad |\partial_x q| \leqslant |2x| + \alpha \leqslant 4,$$

which together with (31) gives

$$\big\| \big(Df(s,x)\big)^{-1} \big\| = |\partial_x q(s,x)|^{-1},$$

as long as $\alpha > 0$ is taken sufficiently small. This implies

$$\sum_{j=0}^{n-1} \log \|Df(s_j,x_j))^{-1}\| = -\sum_{j=0}^{n-1} \log |\partial_x q(s_j, x_j)| \tag{32}$$

for every $(s,x)\in S^1\times I$. If we choose $\gamma<c$, then we have

$$\sum_{j=0}^{n-1}\log|\partial_x q(s_j,x_j)| = \log\prod_{j=0}^{n-1}|\partial_x q(s_j,x_j)| \geqslant cn \tag{33}$$

for every $(s,x)\notin B_1(n)\cup B_2(n)$ (recall (29) and the definition of $B_1(n)$). We conclude from (32) and (33) that

$$\sum_{j=0}^{n-1}\log\|Df(s_j,x_j))^{-1}\| \leqslant -cn \quad \text{for } (s,x)\notin B_1(n)\cup B_2(n),$$

which, in view of the estimates on the Lebesgue measure of $B_1(n)$ and $B_2(n)$, proves that f is a nonuniformly expanding map.

Now we describe how in [**Vi1**] the same conclusions are obtained without assuming (26). Since $\widehat{f}$ is strongly expanding in the horizontal direction, it follows from the methods of [**HPS**] that any map f sufficiently close to $\widehat{f}$ admits a unique invariant central foliation $\mathcal{F}^c$ of $S^1\times I$ by smooth curves uniformly close to vertical segments, see [**Vi1**, Section 2.5]. Actually, $\mathcal{F}^c$ is obtained as the set of integral curves of a vector field $(\xi^c,1)$ in $S^1\times I$ with ξ^c uniformly close to zero. The previous analysis can then be carried out in terms of the expansion of f along this central foliation $\mathcal{F}^c$. More precisely, $|\partial_x q(s,x)|$ is replaced by

$$|\partial_c q(s,x)| \equiv |Df(s,x)v_c(s,x)|,$$

where $v_c(s,x)$ is a unit vector tangent to the foliation at (s,x). The previous observations imply that v_c is uniformly close to $(0,1)$ if f is close to $\widehat{f}$. Moreover, cf. [**Vi1**, Section 2.5], it is no restriction to suppose $|\partial_c q(s,0)|\equiv 0$, so that $\partial_c q(s,x)\approx|x|$, as in the unperturbed case. Indeed, if we define the *critical set* of f by

$$\mathcal{C}=\{(s,x)\in S^1\times I : \partial_c q(s,x)=0\}.$$

by an easy implicit function argument it is shown in [**Vi1**, Section 2.5] that $\mathcal{C}$ is the graph of some C^2 map $\eta: S^1\to I$ arbitrarily C^2-close to zero if α is small. This means that up to a change of coordinates C^2-close to the identity we may suppose that $\eta\equiv 0$ and, hence, write for $\alpha>0$ small

$$\partial_c q(s,x)=x\psi(s,x) \quad \text{with } |\psi+2| \text{ close to zero.}$$

This provides an analog to the second part of assumption (26). At this point, the arguments apply with $\partial_x q(s,x)$ replaced by $\partial_c q(s,x)$, to show that orbits have slow approximation to the critical set $\mathcal{C}$ and $\prod_{i=0}^{n-1}|\partial_c q(s_i,x_i)|$ grows exponentially fast for Lebesgue almost every $(s,x)\in S^1\times I$. A matrix formula for $\big(Df^n(s,x)\big)^{-1}$ similar to that in (31) can be obtained if we replace the vector $(0,1)$ in the canonical basis of the space tangent to $S^1\times I$ at (s,x) by $v_c(s,x)$, and consider the matrix of $\big(Df^n(s,x)\big)^{-1}$ with respect to the new basis.

For future reference, let us make some considerations on the way the sets $B_1(n)$ and $B_2(n)$ are obtained. Let $X : S^1 \to I$ be a smooth map whose graph in $S^1 \times I$ is nearly horizontal (see the notion of admissible curve in [**Vi1**, Section 2] for a precise definition). Denote $\widehat{X}_n(s) = f^n\big(s, X(s)\big)$ for $n \geqslant 0$ and $s \in S^1$. Take some leaf L_0 of the foliation $\mathcal{F}^c$. Letting $L_n = f^n(L_0)$ for $n \geqslant 1$, we define a sequence of Markov partitions $(\mathcal{P}_n)_n$ of S^1 in the following way:

$$\mathcal{P}_n = \Big\{[s', s'')\colon (s', s'') \text{ is a connected component of } \widehat{X}_n^{-1}\big((S^1 \times I) \setminus L_n\big)\Big\}.$$

It is easy to check that $\mathcal{P}_{n+1}$ refines $\mathcal{P}_n$ for each $n \geqslant 1$ and

$$(d + \operatorname{const}\alpha)^{-n} \leqslant |\omega| \leqslant (d - \operatorname{const}\alpha)^{-n}$$

for each $\omega \in \mathcal{P}_n$. Due to the large expansion of f in the horizontal direction, we have that if $J \subset I$ is an interval with $|J| \leqslant \alpha$, then for each $\omega \in \mathcal{P}_n$

$$m\big(\{s \in \omega\colon \widehat{X}_j(s) \in S^1 \times J\}\big) \leqslant \operatorname{const}\sqrt{|J|}\, m(\omega) \tag{34}$$

see [**Vi1**, Corollary 2.3]. The estimate (28) on the Lebesgue measure of $B_2(n)$ is now an easy consequence of (34). For that we only have to compute the Lebesgue measure of $B_2(n)$ on each horizontal line of $S^1 \times I$ and integrate. The estimate (28) on the Lebesgue measure of $B_1(n)$ is obtained by means of a large deviations argument applied to the horizontal curves in $S^1 \times I$; see [**Vi1**, pp. 76 & 77].

Remark 6.5. — The choice of the constants c, ξ, γ and δ only depends on the quadratic map Q and $\alpha > 0$. In particular the decay estimates on the Lebesgue measure of $B_1(n)$ and $B_2(n)$ only depend on the quadratic map Q and $\alpha > 0$.

6.2.2. Random perturbations. — Let f be close to $\widehat{f}$ in the C^3 topology. As we have seen before, it is no restriction to assume that $\mathcal{C} = \{(s, x) \in S^1 \times I\colon x = 0\}$ is the critical set of f. Fix $\{\Phi, (\theta_\varepsilon)_\varepsilon\}$ a random perturbation of f for which (8) holds. Our goal now is to prove that any such f satisfies the hypotheses of Theorems C and D for $\varepsilon > 0$ sufficiently small, and thus conclude that f is stochastically stable. So, we want to show that if $\varepsilon > 0$ is small enough then

- f is nonuniformly expanding for random orbits;
- random orbits have slow approximation to the critical set $\mathcal{C}$;
- the family of hyperbolic time maps $(h_\varepsilon)_\varepsilon$ has uniform L^1-tail.

We remark that in the estimates we have obtained for $\log \|(Df(s_j, x_j))^{-1}\|$ and $\log \operatorname{dist}_\delta(x_j, \mathcal{C})$ over the orbit of a given point $(s, x) \in S^1 \times I$, we can easily replace the iterates (s_j, x_j) by random iterates $(s^j_{\underline{t}}, x^j_{\underline{t}}) = f^j_{\underline{t}}(s, x)$. Actually, the methods used for obtaining estimate (29) rely on a delicate decomposition of the orbit of a given point (s, x) from time 0 until time n into finite pieces according to its returns to the neighborhood $S^1 \times (-\sqrt{\alpha}, \sqrt{\alpha})$ of the critical set. The main tools are [**Vi1**, Lemma 2.4] and [**Vi1**, Lemma 2.5] whose proofs may easily be mimicked for random orbits. Indeed, the important fact in the proof of the referred lemmas is that orbits of points in the central direction stay close to orbits of the quadratic map Q for long periods, as long as $\alpha > 0$ is taken sufficiently small. Hence, such results can easily be obtained for random orbits as long as we take $\varepsilon > 0$ with $\varepsilon \ll \alpha$ and perturbation vectors $\underline{t} \in \operatorname{supp}(\theta_\varepsilon)$.

Thus, the procedure of [**Vi1**] described in Subsection 6.2.1 applies to this situation, and we are able to prove that there is $c > 0$, and for $\gamma > 0$ there is $\delta > 0$, such that

$$\sum_{j=0}^{n-1} \log \|Df(s_{\underline{t}}^j, x_{\underline{t}}^j))^{-1}\| \leqslant -cn \quad \text{and} \quad \sum_{j=0}^{n-1} -\log \operatorname{dist}_\delta(x_{\underline{t}}^j, \mathcal{C}) \leqslant \gamma n$$

for $(s, x) \notin B_1(n) \cup B_2(n)$, where $B_1(n)$ and $B_2(n)$ are subsets $S^1 \times I$ with

$$m\big(B_1(n)\big) \leqslant e^{-\xi n} \quad \text{and} \quad m(B_2(n)) \leqslant \operatorname{const} e^{-\sqrt{n}/4}$$

for some constant $\xi > 0$ only depending on γ. This gives the nonuniform expansion and slow approximation to the critical set for random orbits. Moreover, the arguments show that we may take the map N_ε with

$$(\theta_\varepsilon^{\mathbb{N}} \times m)\left(\left\{(\underline{t}, x) \in T^{\mathbb{N}} \times M \colon N_\varepsilon(\underline{t}, x) > n\right\}\right) \leqslant \operatorname{const} e^{-\sqrt{n}/4},$$

thus giving that the family of first hyperbolic time maps has uniform L^1-tail; cf. Remark 6.1.

For the sake of completeness, an explanation is required on the way the Markov partitions $\mathcal{P}_n$ of S^1 can be defined in this case, in order to obtain the estimates on the Lebesgue measure of $B_1(n)$ and $B_2(n)$. We consider $M = S^1 \times I$ and define the skew-product map

$$\begin{aligned} F : T^{\mathbb{N}} \times M &\longrightarrow \quad T^{\mathbb{N}} \times M, \\ (\underline{t}, z) \quad &\longmapsto \big(\sigma(\underline{t}), f_{t_1}(z)\big) \end{aligned}$$

where σ is the left shift map. Writing $f_t(z) = \big(g_t(z), q_t(z)\big)$ for $z = (s, x) \in S^1 \times I$, we have that $q_t(s, \cdot)$ is a unimodal map close to $\widehat{q}$ for all $s \in S^1$ and $t \in \operatorname{supp}(\theta_\varepsilon)$ with $\varepsilon > 0$ small.

Proposition 6.6. — *Given $\underline{t} \in T^{\mathbb{N}}$ there is a C^1 foliation $\mathcal{F}_{\underline{t}}^c$ of M such that if $L_{\underline{t}}(z)$ is the leaf of $\mathcal{F}_{\underline{t}}^c$ through a point $z \in M$, then*

(1) *$L_{\underline{t}}(z)$ is a C^1 submanifold of M close to a vertical line in the C^1 topology;*
(2) *$f_{t_1}\big(L_{\underline{t}}(z)\big)$ is contained in $L_{\sigma \underline{t}}\big(f_{t_1}(z)\big)$.*

Proof. — This will be obtained as a consequence of the fact that the set of vertical lines constitutes a normally expanding invariant foliation for $\widehat{f}$. Let $\mathcal{H}$ be the space of continuous maps $\xi : T^{\mathbb{N}} \times M \to [-1, 1]$ endowed with the sup norm, and define the map $A : \mathcal{H} \to \mathcal{H}$ by

$$A\xi(\underline{t}, z) = \frac{\partial_x q_{t_1}(z)\xi(F(\underline{t}, z)) - \partial_x g_{t_1}(z)}{-\partial_s q_{t_1}(z)\xi(F(\underline{t}, z)) + \partial_s g_{t_1}(z)}, \quad \underline{t} = (t_1, t_2, \dots) \in T^{\mathbb{N}} \quad \text{and} \quad z \in M.$$

Note that A is well-defined, since

$$|A\xi(\underline{t},z)| \leqslant \frac{(4+\alpha+\varepsilon)+\alpha+\varepsilon}{-(\text{const }\alpha+\varepsilon)+(d-\alpha-\varepsilon)} < 1$$

for small $\alpha > 0$ and $\varepsilon > 0$. Moreover, A is a contraction on $\mathcal{H}$: given $\xi, \zeta \in \mathcal{H}$ and $(\underline{t}, z) \in T^{\mathbb{N}} \times M$ then

$$\begin{aligned}
&|A\xi(\underline{t},z) - A\zeta(\underline{t},z)| \\
&\quad \leqslant \frac{|\det Df_{t_1}(z)| \cdot |\xi(\underline{t},z) - \zeta(\underline{t},z)|}{\left|\left(-\partial_s q_{t_1}(z)\xi(F(\underline{t},z)) + \partial_s g_{t_1}(z)\right) \cdot \left(-\partial_s q_{t_1}(z)\zeta(F(\underline{t},z)) + \partial_s g_{t_1}(z)\right)\right|} \\
&\quad \leqslant \frac{\left((d+\alpha+\varepsilon)(4+\alpha+\varepsilon)+\alpha+\varepsilon\right) \cdot \left|\xi(\underline{t},z) - \zeta(\underline{t},z)\right|}{(d - \text{const}\alpha - \varepsilon)^2}.
\end{aligned}$$

This last quantity can be made smaller than $|\xi(\underline{t},z) - \eta(\underline{t},z)|/2$, as long as α and ε are chosen sufficiently small. This shows that A is a contraction on the Banach space $\mathcal{H}$, and so it has a unique fixed point $\xi^c \in \mathcal{H}$.

It is no restriction for our purposes if we think of T as being equal to $\operatorname{supp}(\theta_\varepsilon)$ for some small ε. Note that the map A depends continuously on F and for $\varepsilon > 0$ small enough the fixed point of A is close to the zero constant map. This holds because we are choosing $\operatorname{supp}(\theta_\varepsilon)$ close to $\{t^*\}$, $f_{t^*} = f$ and f close to $\widehat{f}$. Then, for $\varepsilon > 0$ small enough, we have $\xi^c(\underline{t}, \cdot)$ uniformly close to $\xi^c(\underline{t}^*, \cdot)$ and it is not hard to check that $\xi^c_0 = \xi^c(\underline{t}^*, \cdot)$ is precisely the map whose integral leaves of the vector field $(\xi^c_0, 1)$ give the invariant foliation $\mathcal{F}^c$ associated to $f_{t^*} = f$. Since this foliation depends continuously on the dynamics and for $f = \widehat{f}$ we have $\xi^c_0 \equiv 0$ (see [**Vi1**, Section 2.5]), we finally deduce that $\xi^c(\underline{t}, \cdot)$ is uniformly close to zero for small $\varepsilon > 0$.

We have defined A in such a way that if we take $E^c(\underline{t}, z) = \operatorname{span}\{(\xi^c(\underline{t}, z), 1)\}$, then for every $\underline{t} \in T^{\mathbb{N}}$ and $z \in S^1 \times I$

$$Df_{t_1}(z)E^c(\underline{t},z) \subset E^c(F(\underline{t},z)). \tag{35}$$

Now, for fixed $\underline{t} \in T^{\mathbb{N}}$, we take $\mathcal{F}^c_{\underline{t}}$ to be the set of integral curves of the vector field $z \mapsto (\xi^c(\underline{t}, z), 1)$ defined on $S^1 \times I$. Since the vector field is taken of class C^0, it does not follow immediately that through each point in $S^1 \times I$ passes only one integral curve. We will prove uniqueness of solutions by using the fact that the map f has a big expansion in the horizontal direction.

Assume, by contradiction, that there are two distinct integral curves $Y, Z \in \mathcal{F}^c_{\underline{t}}$ with a common point. So we may take three distinct nearby points $z_0, z_1, z_2 \in S^1 \times I$ such that $z_0 \in Y \cap Z$, $z_1 \in Y$, $z_2 \in Z$ and z_1, z_2 have the same x-coordinate. Let X be the horizontal curve joining z_1 to z_2. If we consider $X_n = \pi_2 \circ F^n(\underline{t}, X)$ for $n \geqslant 1$, where π_2 is the projection from $T^{\mathbb{N}} \times S^1 \times I$ onto $S^1 \times I$, we have that the curves X_n are nearly horizontal and grow in the horizontal direction (when n increases) by a factor close to d for small α and ε, see [**Vi1**, Section 2.1]. Hence, for large n, X_n wraps many times around the cylinder $S^1 \times I$. On the other hand, since $Y_n = \pi_2 \circ F^n(\underline{t}, Y)$ and $Z_n = \pi_2 \circ F^n(\underline{t}, Z)$ are always tangent to the vector field $z \mapsto \big(\xi^c(\sigma^n \underline{t}, z), 1\big)$

on $S^1 \times I$, it follows that all the iterates of Y_n and Z_n have small amplitude in the s-direction. This gives a contradiction, since the closed curve made by Y, Z and X is homotopic to zero in $S^1 \times I$ and the closed curve made by Y_n, Z_n and X_n cannot be homotopic to zero for large n. Thus, for fixed $\underline{t} \in T^{\mathbb{N}}$ we have uniqueness of solutions of the vector field $z \to (\xi^c(\underline{t}, z), 1)$, and from (35) it follows that $\mathcal{F}^c_{\underline{t}}$ is an F-invariant foliation of M by nearly vertical leaves. □

Now, using the foliations given by the previous proposition we are also able to define the Markov partitions of S^1 in this setting. Given any smooth map $X : S^1 \to I$ whose graph is nearly horizontal, denote $\widehat{X}^n_{\underline{t}}(s) = f^n_{\underline{t}}\big(s, X(s)\big)$ for $n \geqslant 0$ and $s \in S^1$. Take some leaf $L^0_{\underline{t}}$ of the foliation $\mathcal{F}^c_{\underline{t}}$. Letting $L^n_{\underline{t}} = f^n_{\underline{t}}(L_{\underline{t}})$ for $n \geqslant 1$, we define the sequence of Markov partitions $(\mathcal{P}^n_{\underline{t}})_n$ of S^1 as

$$\mathcal{P}^n_{\underline{t}} = \Big\{[s', s'') \colon (s', s'') \text{ is a connected component of } (\widehat{X}^n_{\underline{t}})^{-1}\big((S^1 \times I) \smallsetminus L^n_{\underline{t}}\big)\Big\}.$$

It is easy to check that $\mathcal{P}^{n+1}_{\underline{t}}$ refines $\mathcal{P}^n_{\underline{t}}$ for each $n \geqslant 1$ and, taking $\varepsilon \ll \alpha$,

$$(d + \operatorname{const} \alpha)^{-n} \leqslant |\omega| \leqslant (d - \operatorname{const} \alpha)^{-n}$$

for each $\omega \in \mathcal{P}^n_{\underline{t}}$. This permits to obtain estimates (28) and (30) for the Lebesgue measure of the sets $B_1(n)$ and $B_2(n)$ exactly in the same way as in Subsection 6.2.1, also with the constants only depending on the quadratic map Q (cf. Remark 6.5).

References

[Al] J. F. Alves, *SRB measures for non-hyperbolic systems with multidimensional expansion*, Ann. Scient. Éc. Norm. Sup., 4^e série, 33 (2000), 1-32.

[ABV] J. F. Alves, C. Bonatti, M. Viana, *SRB measures for partially hyperbolic systems with mostly expanding central direction*, Invent. Math. 140 (2000), 351-398.

[AV] J. F. Alves, M. Viana, *Statistical stability for robust classes of maps with nonuniform expansion*, Erg. Th. & Dyn. Sys 22 (2002), 1-32.

[Ar] V. Araújo, *Attractors and time averages for random maps*, Ann. de l'Inst. H. Poincaré - Anal. non-Lin. 17 (2000), 307-369.

[BV] V. Baladi, M. Viana, *Strong stochastic stability and rate of mixing for unimodal maps*, Ann. Scient. Éc. Norm. Sup., 4^e série, 29 (1996), 483-517.

[BC1] M. Benedicks and L. Carleson, *On iterations of* $1 - ax^2$ *on* $(-1, 1)$, Ann. Math. 122 (1985), 1-25.

[BC2] M. Benedicks and L. Carleson, *The dynamics of the Hénon map*, Ann. Math. 133 (1991), 73-169.

[BY1] M. Benedicks, L.-S. Young, *Absolutely continuous invariant measures and random perturbations for certain one-dimensional maps*, Erg. Th. & Dyn. Sys. 12 (1992), 13-37.

[BY2] M. Benedicks and L.-S. Young, *SRB-measures for certain Hénon maps*, Invent. Math. 112 (1993), 541-576.

[BV] C. Bonatti and M. Viana, *SRB measures for partially hyperbolic systems with mostly contracting central direction*, Isr. J. Math. 115 (2000), 157-193.

[BR] R. Bowen and D. Ruelle, *The ergodic theory of Axiom A flows*, Invent. Math. 29 (1975), 181-202.

[HPS] M. Hirsch, C. Pugh and M. Shub, *Invariant manifolds*, Lect. Notes in Math. 583, Springer Verlag, 1977.

[Ja] M. Jakobson, *Absolutely continuous invariant measures for one-parameter families of one-dimensional maps*, Comm. Math. Phys. 81 (1981), 39-88.

[KK] A. Katok and Y. Kifer, *Random perturbations of transformations of an interval*, J. Anal. Math. 47 (1986), 193-237.

[Ki1] Y. Kifer, *Ergodic theory of random transformations*, Birkhäuser, Boston Basel, 1986.

[Ki2] Y. Kifer, *Random perturbations of dynamical systems*, Birkhäuser, Boston Basel, 1988.

[Ma] R. Mañé, *Ergodic theory and differentiable dynamics*, Springer-Verlag, 1987.

[Pl] V. Pliss, *On a conjecture due to Smale*, Diff. Uravnenija, 8 (1972), 262-268.

[Ru] D. Ruelle, *A measure associated with Axiom A attractors*, Amer. Jour. Math. 98 (1976), 619-654.

[Si] Ya. Sinai, *Gibbs measures in ergodic theory*, Russ. Math. Surv. 27, n. 4, (1972), 21-69.

[Vi1] M. Viana, *Multidimensional non-hyperbolic attractors*, Publ. Math. IHES 85 (1997), 63-96.

[Vi2] M. Viana, *Stochastic dynamics of deterministic systems*, Lect. Notes XXI Braz. Math Colloq., IMPA, 1997.

[Yo] L.-S. Young, *Stochastic stability of hyperbolic attractors*, Erg. Th. & Dyn. Sys. 6 (1986), 311-319.

J.F. Alves, Centro de Matemática da Universidade do Porto, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal • *E-mail :* jfalves@fc.up.pt
Url : http://www.fc.up.pt/cmup/jfalves

V. Araújo, Centro de Matemática da Universidade do Porto, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal
E-mail : vdaraujo@fc.up.pt • *Url :* http://www.fc.up.pt/cmup/vdaraujo

Astérisque
286, 2003, p. 63–79

THE MINIMAL ENTROPY PROBLEM FOR 3-MANIFOLDS WITH ZERO SIMPLICIAL VOLUME

by

James W. Anderson & Gabriel P. Paternain

Dedicated to Jacob Palis on his sixtieth birthday

Abstract. — In this note, we consider the *minimal entropy problem*, namely the question of whether there exists a smooth metric of minimal (topological) entropy, for certain classes of closed 3-manifolds. Specifically, we prove the following two results.

Theorem A. *Let M be a closed orientable irreducible 3-manifold whose fundamental group contains a $\mathbb{Z} \oplus \mathbb{Z}$ subgroup. The following are equivalent:*
(1) *the simplicial volume $\|M\|$ of M is zero and the minimal entropy problem for M can be solved;*
(2) *M admits a geometric structure modelled on $\mathbb{E}^3$ or* Nil*;*
(3) *M admits a smooth metric g with $\mathrm{h_{top}}(g) = 0$.*

Theorem B. *Let M be a closed orientable geometrizable 3-manifold. The following are equivalent:*
(1) *the simplicial volume $\|M\|$ of M is zero and the minimal entropy problem for M can be solved;*
(2) *M admits a geometric structure modelled on $\mathbb{S}^3$, $\mathbb{S}^2 \times \mathbb{R}$, $\mathbb{E}^3$, or* Nil*;*
(3) *M admits a smooth metric g with $\mathrm{h_{top}}(g) = 0$.*

1. Introduction and statement of results

Let M^n be a closed orientable n-dimensional manifold. For a smooth Riemannian metric g on M, let $\mathrm{Vol}(M, g)$ denote the volume of M calculated with respect to g.

Let $\mathrm{h_{top}}(g)$ be the *topological entropy* of the geodesic flow of g, as defined in Section 2.6. Set the *minimal entropy* of M to be

$$\mathrm{h}(M) := \inf\{\mathrm{h_{top}}(g) \mid g \text{ is a smooth metric on } M \text{ with } \mathrm{Vol}(M, g) = 1\}.$$

A smooth metric g_0 with $\mathrm{Vol}(M, g_0) = 1$ is *entropy minimizing* if

$$\mathrm{h_{top}}(g_0) = \mathrm{h}(M).$$

2000 ***Mathematics Subject Classification.*** — 53D25, 37D40.
Key words and phrases. — Minimal entropy, simplicial volume, geometrizable 3-manifold.

The second author was partially supported by CIMAT, Guanajuato, Mexico.

The *minimal entropy problem* for M is whether or not there exists an entropy minimizing metric on M. Say that the *minimal entropy problem can be solved* for M if there exists an entropy minimizing metric on M. Smooth manifolds are hence naturally divided into two classes: those for which the minimal entropy problem can be solved and those for which it cannot.

There are a number of classes of manifolds for which the minimal entropy problem can be solved. For instance, the minimal entropy problem can always be solved for a closed orientable surface M. For the 2-sphere and the 2-torus, this follows from the fact that both a metric with constant positive curvature and a flat metric have zero topological entropy. For surfaces of higher genus, A. Katok [**11**] proved that each metric of constant negative curvature minimizes topological entropy, and conversely that any metric that minimizes topological entropy has constant negative curvature.

This result of Katok has been generalized to higher dimensions by Besson, Courtois, and Gallot [**1**], as follows. Suppose that M^n ($n \geqslant 3$) admits a locally symmetric metric g_0 of negative curvature, normalized so that $\mathrm{Vol}(M, g_0) = 1$. Then g_0 is the unique entropy minimizing metric up to isometry. Unlike the case of a surface, a locally symmetric metric of negative curvature on a closed orientable n-manifold ($n \geqslant 3$) is unique up to isometry, by the rigidity theorem of Mostow [**18**].

The result of Besson, Courtois, and Gallot [**1**] has itself been generalized by Connell and Farb [**4**] to n-manifolds that admit a complete, finite-volume metric which is locally isometric to a product of negatively curved (rank 1) symmetric spaces of dimension at least 3.

A positive solution to the minimal entropy problem appears to single out manifolds that have either a high degree of symmetry or a low topological complexity. What this means in the context of 3-manifolds will become apparent below. A similar phenomena is observed for closed simply connected manifolds of dimensions 4 and 5: there are essentially only nine manifolds for which the minimal entropy problem can be solved and they can be explicitly listed. These nine manifolds share the property that their loop space homology grows polynomially for any coefficient field, see Paternain and Petean [**21**].

The goal of this note is to classify those closed orientable geometrizable 3-manifolds with zero simplicial volume for which the minimal entropy problem can be solved. Specifically, in Section 4, we prove:

Theorem A. — *Let M be a closed orientable irreducible 3-manifold whose fundamental group contains a $\mathbb{Z} \oplus \mathbb{Z}$ subgroup. The following are equivalent:*

(1) *the simplicial volume $\|M\|$ of M is zero and the minimal entropy problem for M can be solved;*

(2) *M admits a geometric structure modelled on $\mathbb{E}^3$ or* Nil*;*

(3) *M admits a smooth metric g with $\mathrm{h_{top}}(g) = 0$.*

In Section 5 we prove:

Theorem B. — *Let M be a closed orientable geometrizable* 3*-manifold. The following are equivalent:*

(1) *the simplicial volume $\|M\|$ of M is zero and the minimal entropy problem for M can be solved;*

(2) *M admits a geometric structure modelled on $\mathbb{S}^3$, $\mathbb{S}^2 \times \mathbb{R}$, $\mathbb{E}^3$, or* Nil*;*

(3) *M admits a smooth metric g with $\mathrm{h}_{\mathrm{top}}(g) = 0$.*

Recall that the *simplicial volume* of a closed orientable manifold M is defined as the infimum of $\sum_i |r_i|$ where the r_i are the coefficients of a real cycle that represents the fundamental class of M. For 3-manifolds, the positivity of the simplicial volume (which is a homotopy invariant) is closely related to the existence of compact hyperbolizable submanifolds in M. This is described in more detail in Section 2.5.

We close the introduction by describing some of the elements of the proofs of Theorems A and B, and by describing a conjectural picture. We will see in Section 2 that a closed orientable geometrizable 3-manifold M has zero simplicial volume if and only if M has zero minimal entropy. Therefore, the minimal entropy problem can be solved if and only if M admits a smooth metric with zero topological entropy. This in turn forces the fundamental group of M to have subexponential growth. We end up showing that this can occur only if M admits one of the four geometric structures listed in the statement of Theorem B. On the other hand, it is a calculation that the manifolds in the statement of Theorem B carry a metric of zero entropy. The proof of Theorem A follows a similar line, and makes use of the remarkable theorem, due essentially to Thurston, that a manifold satisfying the hypothesis of the theorem is geometrizable. The precise definition of geometrizable manifold is given in Subsection 2.4. Thurston's geometrization conjecture states that every closed orientable 3-manifold is geometrizable.

From this discussion and the above mentioned result of Besson, Courtois and Gallot it seems quite reasonable to speculate that the following statement holds:

Let M be a closed orientable geometrizable 3*-manifold. Then, the minimal entropy problem for M can be solved if and only if M admits a geometric structure modelled on $\mathbb{S}^3$, $\mathbb{S}^2 \times \mathbb{R}$, $\mathbb{E}^3$,* Nil*, or $\mathbb{H}^3$.*(1)

Indeed, suppose that the simplicial volume of M were *not* zero. This would imply that M contains a disjoint collection $H_1, \ldots, H_p$ of compact submanifolds whose interiors each admit a complete hyperbolic structure of finite volume. In particular, it should be that the minimal entropy of M is the maximum of the minimal entropies of the H_k. It

(1)Note added in proof: J. Souto (Geometric structures on 3-manifolds and their deformations. Dissertation, Rheinische Friedrich-Wilhelms-Universität Bonn 2001) has proven this conjecture for all geometrizable prime 3-manifolds would then seem reasonable that an entropy minimizing metric on M would try to be as hyperbolic as possible on the interiors of the H_k and would try as much as possible to be one of the other seven standard 3-dimensional geometries on the components of $M - (H_1 \cup \cdots \cup H_p)$. However, it would seem that the minimizer would have to be singular along the ∂H_k, and so there should be no metric of minimal entropy. Unfortunately, we do not yet know how to make this argument rigorous.

We would like to thank the referees for their careful reading of this note.

2. Preliminaries

The purpose of this Section is to present some of the basic material from 3-manifold theory that we will need. We refer the interested reader to Hempel [**8**] for a more detailed introduction to 3-manifold topology. For a more detailed description of Seifert fibered spaces, and of the torus decomposition and the geometrization of 3-manifolds, we also refer the interested reader to the survey articles of Scott [**26**] and Bonahon [**2**], and the references contained therein.

2.1. 3-manifold basics. — We begin with some basic definitions. A 3-manifold is *closed* if it is compact with empty boundary.

An embedded 2-sphere $\mathbb{S}^2$ in a 3-manifold M is *essential* if M does not bound a closed 3-ball in M. A 3-manifold is *irreducible* if it contains no essential 2-sphere.

A 3-manifold is *prime* if it cannot be decomposed as a non-trivial connected sum. That is, M is prime if for every decomposition $M = M_1 \# M_2$ of M as a connected sum, one of M_1 or M_2 is homeomorphic to the standard 3-sphere $\mathbb{S}^3$. Every irreducible 3-manifold is prime, but the converse does not hold. However, the only closed orientable 3-manifold that is prime but not irreducible is $\mathbb{S}^2 \times \mathbb{S}^1$.

We note here that if the closed orientable 3-manifold M contains a non-separating essential 2-sphere, then M can be expressed as the connected sum $M = P \# (\mathbb{S}^2 \times \mathbb{S}^1)$ for some 3-manifold P. Hence, in what follows, we need only consider separating essential 2-spheres in 3-manifolds.

There is an inverse to the operation of connected sum for 3-manifolds, called the *prime decomposition*. The following statement is adapted from Bonahon [**2**], and follows from work of Kneser [**12**] and Milnor [**16**].

Let M be a closed orientable 3-manifold. Then, there exists a compact 2-submanifold Σ of M, unique up to isotopy, so that two conditions hold. First, each component of Σ is an embedded essential separating 2-sphere, and the 2-spheres in Σ are pairwise non-parallel, in that no two 2-spheres in Σ bound an embedded $\mathbb{S}^2 \times [0, 1]$ in M. Second, if $M_0, M_1, \ldots, M_p$ are the closures of the components of $M - \Sigma$, then M_0 is homeomorphic to the 3-sphere $\mathbb{S}^3$ minus finitely many disjoint open 3-balls; while for $k \geqslant 1$, each M_k contains a unique component of Σ, and every separating essential 2-sphere in M_k is parallel to ∂M_k.

The *prime decomposition* of M is the collection of 3-manifolds that results by taking the complements of the 2-submanifold Σ in M as just described, and filling in each 2-sphere boundary component of $M_0, M_1, \ldots, M_p$ with a 3-ball. Each of the resulting 3-manifolds is then prime. (Note that both $\mathbb{S}^3$ and $\mathbb{S}^2 \times \mathbb{S}^1$ have trivial prime decompositions, as they do not contain a separating essential 2-sphere.) The prime decomposition is one of two standard decompositions of a closed orientable 3-manifold, the other being the torus decomposition, which is discussed in detail in Section 2.3.

In general, a closed orientable embedded surface S in a 3-manifold M is *2-sided* if there exists an embedding f of $S \times [-1, 1]$ into M so that $f(S \times \{0\}) = S$. A closed orientable embedded surface S in a 3-manifold M is *incompressible* if the fundamental group of S is infinite and if the inclusion $S \hookrightarrow M$ induces an injection on fundamental groups. An incompressible surface S is *essential* if S is not homotopic into ∂M.

A compact orientable irreducible 3-manifold M is *sufficiently large* if it contains a 2-sided incompressible surface. Sufficiently large 3-manifolds are also known as *Haken* 3-manifolds.

2.2. Seifert fibered spaces. — A *Seifert fibration* of a 3-manifold M is a decomposition of M into disjoint simple closed curves, called the *fibers* of the fibration, so that each fiber c has a neighborhood U in M of the following form: U is diffeomorphic to the quotient of $\mathbb{S}^1 \times \mathbb{B}^2$ by the free action of a finite group respecting the product structure, where the fibers of the fibration correspond to the curves $\{x\} \times \mathbb{B}^2$ for $x \in \mathbb{S}^1$. (In this note, we only consider Seifert fibrations of closed 3-manifolds and of 3-manifolds without boundary that are homeomorphic to the interior of a compact 3-manifold with 2-torus boundary components.)

Since we are considering only orientable 3-manifolds in this note, the group of covering transformations of $\mathbb{S}^1 \times \mathbb{B}^2$ in the above definition is generated by $\tau_{p,q}$ for some pair (p, q) of relatively prime integers, where

$$\tau_{p,q}\left(e^{i\varphi}, re^{i\theta}\right) = \left(e^{i(\varphi+2\pi/p)}, re^{i(\theta+2\pi q/p)}\right).$$

A fiber is a *regular fiber* if it has a neighborhood diffeomorphic to $\mathbb{S}^1 \times \mathbb{B}^2$, and is a *singular fiber* otherwise. Note that the singular fibers of a Seifert fibration are necessarily isolated.

Let S be the space of fibers of a Seifert fibration of a 3-manifold M, equipped with the quotient topology coming from the projection map $p : M \to S$. We often refer to S as the *base orbifold* of the Seifert fibered space M. Using the neighborhoods of the fibers in M, we see that S is an orientable surface with one cone point for each singular fiber.

Let $p_1, \ldots, p_s$ be the cone points on S, and let n_j be the *order* at the cone point p_j, so that a neighbhorhood of p_j is diffeomorphic to the quotient $\mathbb{B}^2/\mathbb{Z}_{n_j}$, where $\mathbb{Z}_{n_j}$

acts by rotation. The *orbifold Euler characteristic* $\chi(S)$ of S is the quantity

$$\chi(S) = 2 - 2\,\mathrm{genus}(S) - \sum_{k=1}^{s} \left(1 - \frac{1}{n_j}\right).$$

(This discussion is also valid in the case that M is a 3-manifold without boundary that is homeomorphic to the interior of a compact 3-manifold with 2-torus boundary components. In this case, the base orbifold has punctures as well as cone points, and we view each puncture as a cone point of infinite order.)

There are two cases of particular interest. In the case that $\chi(S) < 0$, S has a hyperbolic structure, so that we can express S as the quotient $S = \mathbb{H}^2/\Gamma$, where $\mathbb{H}^2$ is the hyperbolic plane and Γ is a discrete subgroup of $\mathrm{Isom}(\mathbb{H}^2)$, where the fixed points of the action of non-trivial elements of Γ descend to the cone points on S. We refer to Γ as the *orbifold fundamental group* of S. In this case, we have that Γ contains a free subgroup of rank 2, and in particular Γ contains an element of infinite order.

In the case that $\chi(S) = 0$, S has a Euclidean structure, so that we can express S as the quotient $S = \mathbb{E}^2/\Gamma$, where $\mathbb{E}^2$ is the Euclidean plane and Γ is a discrete subgroup of $\mathrm{Isom}(\mathbb{E}^2)$, where the fixed points of the action of non-trivial elements of Γ descend to the cone points on S. As above, we refer to Γ as the *orbifold fundamental group* of S. In this case, we have that Γ contains an element of infinite order, but not a free subgroup of rank two.

In both of these cases, the orbifold fundamental group of the base orbifold S of the Seifert fibered space M is a subgroup of $\pi_1(M)$. In fact, there is a short exact sequence

$$1 \longrightarrow \mathbb{Z} \longrightarrow \pi_1(M) \longrightarrow \pi_1(S) \longrightarrow 1,$$

where $\pi_1(S)$ is the orbifold fundamental group of S and where $\mathbb{Z}$ is generated by any regular fiber of the Seifert fibration.

The following follows immediately from this discussion.

Lemma 2.1. — *Let M be a Seifert fibered space as above with base orbifold S. If $\chi(S) \leqslant 0$, then $\pi_1(M)$ contains a $\mathbb{Z} \oplus \mathbb{Z}$ subgroup.*

Proof. — The proof of Lemma 2.1 is standard, but we sketch it here for the sake of completeness. Let $p : M \to S$ be the quotient map. Since $\chi(S) \leqslant 0$, there is a closed curve c, not necessarily simple, on S that represents an infinite order element of the orbifold fundamental group of S. Let $T = p^{-1}(c)$ in M be the subset of M that consists of all the fibers in M corresponding to points of c. Then, T is an incompressible 2-torus in M, though not necessarily embedded. However, this is sufficient to guarantee that there exists a $\mathbb{Z} \oplus \mathbb{Z}$ subgroup of $\pi_1(M)$, namely the fundamental group of T. □

2.3. The torus decomposition. — Let M be a closed orientable irreducible 3-manifold with infinite fundamental group. There is then a canonical decomposition of M along embedded essential 2-tori, due to Jaco and Shalen [**9**] and Johannson [**10**].

(Note that the restriction to irreducible 3-manifolds causes no loss of generality, as we may first apply the prime decomposition to M, as described in Section 2.1. Also, we tend to not take the torus decomposition of $\mathbb{S}^2 \times \mathbb{S}^1$.) The statement given below is adapted from Theorem 3.4 of Bonahon [**2**].

Theorem 2.2 ([**2**]). — *Let M be a closed orientable irreducible 3-manifold. Then, up to isotopy, there is a unique compact 2-submanifold T of M such that:*

(1) *every component of T is a 2-sided essential 2-torus;*

(2) *every component of $M - T$ either contains no essential embedded 2-torus or Klein bottle, or else admits a Seifert fibration (or possibly both);*

(3) *property (2) fails when any component of T is removed.*

We refer to this 2-submanifold T as the *torus decomposition* of M. Note that condition (3) implies that no two of the 2-tori in the torus decomposition are isotopic.

Let M be a compact orientable 3-manifold, and let $M_0, M_1, \dots, M_p$ be the components of its prime decomposition. Let T_k be the torus decomposition of M_k. Say that M is a *graph manifold* if, for each $1 \leqslant k \leqslant p$, every component of $M_k - T_k$ admits a Seifert fibration. Clearly, every Seifert fibered space is trivially a graph manifold. Also, every 2-torus bundle over $\mathbb{S}^1$ is a graph manifold.

Theorem 2.2 is a small part of the machinary of the *characteristic submanifold* of a 3-manifold developed by Jaco and Shalen and by Johannson. Note that this discussion includes the possibility that the torus decomposition T is empty, even though $\pi_1(M)$ may contain a $\mathbb{Z} \oplus \mathbb{Z}$ subgroup.

A closely related result is the following torus theorem. For a discussion and proof of this result, see Scott [**27**].

Theorem 2.3 ([**27**]). — *Let M be a closed orientable irreducible 3-manifold whose fundamental group contains a $\mathbb{Z}\oplus\mathbb{Z}$ subgroup. Then, either M contains an incompressible embedded 2-torus or M is a Seifert fibered space.*

2.4. Geometric structures and geometrization. — A *3-dimensional geometry* is a pair (X, G), where X is a simply connected Riemannian 3-manifold with a complete homogeneous metric and G is a maximal transitive group of orientation-preserving isometries of X, with the proviso that there exists a subgroup H of G with compact quotient X/H. Note that since G is a maximal group of isometries, it suffices to specify X and set $G = \mathrm{Isom}(X)$.

It is a result of Thurston that there exist exactly eight 3-dimensional geometries, namely $\mathbb{E}^3$, $\mathbb{S}^3$, $\mathbb{H}^3$, $\mathbb{S}^2 \times \mathbb{R}$, $\mathbb{H}^2 \times \mathbb{R}$, $\widetilde{\mathrm{SL}_2}$, Nil, and Sol, with their respective groups of (orientation preserving) isometries. (A proof of this result, and a detailed description of the eight geometries, is given in Scott [**26**].)

Let M be an orientable 3-manifold that is homeomorphic to the interior of a compact 3-manifold with 2-torus boundary components. (This includes the possibility that M is closed.) Say that M *admits a geometric structure modelled on* X if M is diffeomorphic to the quotient X/Γ, where X is one of the eight 3-dimensional geometries and Γ is a fixed point free subgroup of $\mathrm{Isom}(X)$. It is known that if a 3-manifold admits a geometric structure, then it admits a unique geometric structure.

More generally, let M be a closed orientable irreducible 3-manifold with torus decomposition T. Say that M is *geometrizable* if each component of $M - T$ admits a geometric structure. (Note that we do not require that different components of $M - T$ admit the same geometric structure.)

Finally, say that a closed orientable 3-manifold is *geometrizable* if every component of its prime decomposition is geometrizable. (This causes no difficulties, as $\mathbb{S}^2 \times \mathbb{S}^1$, which may arise as a component of the prime decomposition but is not irreducible, admits a geometric structure modelled on $\mathbb{S}^2 \times \mathbb{R}$.)

Thurston's *geometrization conjecture* states that every closed orientable 3-manifold is geometrizable. For a more complete discussion of the geometrization conjecture, see Scott [**26**], Bonahon [**2**], or Thurston [**30**].

There are a number of manifolds for which the geometrization conjecture is known to be true. If M is a closed orientable irreducible sufficiently large 3-manifold, then M is geometrizable; this is Thurston's geometrization theorem; see Morgan [**17**] or Otal [**19**] for a discussion of this theorem.

In particular, if M has a non-empty torus decomposition, then it is geometrizable. In this case, each component of the complement of the torus decomposition of M either is a Seifert fibered space or admits a hyperbolic structure, that is the geometric structure modelled on $\mathbb{H}^3$. We encode in the following theorem the parts of this discussion we make the most use of.

Theorem 2.4. — *Let M be a closed orientable irreducible sufficiently large 3-manifold. Then, M admits a torus decomposition T. Moreover, each component of $M - T$ either is a Seifert fibered space or admits a hyperbolic structure.*

Additionally, the geometrization of Seifert fibered spaces, and in fact of irreducible graph manifolds, is completely understood.

Theorem 2.5 ([**26**, Theorem 5.3]). — *Let M be a closed orientable 3-manifold. Then,*

(1) *M possesses a geometric structure modelled on* Sol *if and only if M is finitely covered by a 2-torus bundle over $\mathbb{S}^1$ with hyperbolic glueing map;*

(2) *M possesses a geometric structure modelled on one of $\mathbb{S}^3$, $\mathbb{E}^3$, $\mathbb{S}^2 \times \mathbb{R}$, $\mathbb{H}^2 \times \mathbb{R}$, $\widetilde{\mathrm{SL}}_2$, or* Nil *if and only if M is a Seifert fibered space.*

We note here that the two unresolved cases of the geometrization conjecture are that the fundamental group of M is finite, in which case M should admit a geometric structure modelled on $\mathbb{S}^3$ [the Poincaré conjecture and the spherical space form problem], and that the fundamental group of M is infinite, does not contain $\mathbb{Z} \oplus \mathbb{Z}$, and does not contain a normal cyclic subgroup, in which case M should admit a geometric structure modelled on $\mathbb{H}^3$ [the hyperbolization conjecture].

2.5. Simplicial volume. — Let M be a closed manifold. Denote by C_* the real chain complex of M: a chain $c \in C_*$ is a finite linear combination $\sum_i r_i \sigma_i$ of singular simplices σ_i in M with real coefficients r_i. Define the *simplicial* ℓ^1*-norm* in C_* by setting $|c| = \sum_i |r_i|$. This norm gives rise to a pseudo-norm on the homology $H_*(M, \mathbb{R})$ by setting

$$|[\alpha]| = \inf\{|z| \,:\, z \in C_* \text{ and } [z] = [\alpha]\}.$$

When M is orientable, define the *simplicial volume* of M, denoted $\|M\|$, to be the simplicial norm of the fundamental class. The simplicial volume is also called *Gromov's invariant*, since it was first introduced by Gromov [**7**].

The following lower bound on $\|M\|$ is due to Thurston [**29**].

Theorem 2.6 ([**29**, Theorem 6.5.5]). — *Suppose that M is a closed orientable 3-manifold and that $H \subset M$ is a 3-dimensional submanifold whose interior admits a complete hyperbolic structure of finite volume. Suppose further that $\overline{H}$ is embedded in M and that $\partial\overline{H}$ is incompressible in M. Then,*

$$\|M\| \geqslant \frac{\mathrm{Vol}(H)}{v_3} > 0,$$

where v_3 is the volume of the regular ideal tetrahedron in $\mathbb{H}^3$.

The next theorem follows immediately from Theorems 2.6, 2.4, and 2.5.

Theorem 2.7. — *Let M be a closed orientable geometrizable 3-manifold. Suppose that $\|M\| = 0$. Then M is a graph manifold.*

Proof. — The proof of Theorem 2.7 is essentially contained in Soma [**28**]; we include it here solely for the sake of completeness.

We begin by considering the prime decomposition of M. That is, write M as the connected sum $M = M_0 \# \cdots \# M_p$, where each M_i is a prime 3-manifold. (Note that we are including in this discussion the case that M is itself prime, and so has trivial prime decomposition.)

Since simplicial volume behaves additively with respect to connected sums (cf. Gromov [**7**]), the hypothesis that M has zero simplicial volume implies that each M_i has zero simplicial volume as well. Since the connected sum of graph manifolds is again a graph manifold (cf. Soma [**28**]), it suffices to show that each M_i is a graph manifold. Since each M_i is prime, it is either irreducible or diffeomorphic to $\mathbb{S}^2 \times \mathbb{S}^1$, which is a Seifert fibered space. So, we may assume without loss of generality that M is irreducible.

Let T be the torus decomposition of M. Recall that M is assumed to be geometrizable. If T is empty, then M admits a geometric structure other than the one modelled on $\mathbb{H}^3$ (which is excluded by the assumption on the simplicial volume of M), and so M is a graph manifold, by Theorem 2.5.

If T is non-empty, then M is sufficiently large, and so Thurston's geometrization conjecture holds for M. Since $\|M\| = 0$, each component of $M - T$ is a Seifert fibered space, as no piece can be hyperbolic, by Theorem 2.6. It follows that M must be a graph manifold. □

2.6. Topological entropy. — We recall in this subsection the definition of the topological entropy of the geodesic flow of a smooth Riemannian metric g on a closed manifold M. For a more detailed discussion, we refer the interested reader to Paternain [**20**].

The geodesic flow of g is a flow ϕ_t that acts on SM, the unit sphere bundle of M, which is a closed hypersurface of the tangent bundle of M. Let d be any distance function compatible with the topology of SM. For each $T > 0$ we define a new distance function

$$d_T(x, y) := \max_{0 \leqslant t \leqslant T} d(\phi_t(x), \phi_t(y)).$$

Since SM is compact, we can consider the minimal number of balls of radius $\varepsilon > 0$ in the metric d_T that are necessary to cover SM. Let us denote this number by $N(\varepsilon, T)$. We define

$$\mathrm{h}(\phi, \varepsilon) := \limsup_{T \to \infty} \frac{1}{T} \log N(\varepsilon, T).$$

Observe now that the function $\varepsilon \mapsto \mathrm{h}(\phi, \varepsilon)$ is monotone decreasing and therefore the following limit exists:

$$\mathrm{h}_{\mathrm{top}}(g) := \lim_{\varepsilon \to 0} \mathrm{h}(\phi, \varepsilon).$$

The number $\mathrm{h}_{\mathrm{top}}(g)$ thus defined is the *topological entropy* of the geodesic flow of g. Intuitively, this number is a measure of the orbit complexity of the flow. The positivity of $\mathrm{h}_{\mathrm{top}}(\phi)$ indicates complexity or 'chaos' of some kind in the dynamics of ϕ_t.

There is a formula, known as Mañé's formula, that gives a nice alternative description of $\mathrm{h}_{\mathrm{top}}(g)$. Given points p and q in M and $T > 0$, define $n_T(p, q)$ to be the number of geodesic arcs joining p and q with length $\leqslant T$. Mañé [**14**] showed that

$$\mathrm{h}_{\mathrm{top}}(g) = \lim_{T \to \infty} \frac{1}{T} \log \int_{M \times M} n_T(p, q)\, dp\, dq.$$

Finally we note that entropy behaves well under scaling of the metric. Namely, if c is any positive constant, then $\mathrm{h}_{\mathrm{top}}(cg) = \mathrm{h}_{\mathrm{top}}(g)/\sqrt{c}$.

2.7. Minimal volume and collapsing. — The *minimal volume* $\mathrm{MinVol}(M)$ of a Riemannian manifold M is defined to be the infimum of $\mathrm{Vol}(M, g)$ over all smooth metrics g such that the sectional curvature K_g of g satisfies $|K_g| \leqslant 1$. This differential invariant was introduced by M. Gromov in [**7**].

We shall need the following result, see Cheeger and Gromov [**3**, Example 0.2 and Theorem 3.1] and Rong [**23**].

Proposition 2.8. — *Let M be a closed orientable 3-manifold. If M is a graph manifold, then M admits a polarized F-structure, and hence* $\mathrm{MinVol}(M) = 0$.

We will not give here the precise definition of a polarized F-structure, because it is too technical. Instead we give an informal description, and we refer the interested reader to Cheeger and Gromov [**3**] for a more detailed discussion.

An *F-structure* on a manifold M is a natural generalization of a torus action on M. Different tori, possibly of different dimensions, act on subsets of M in such a way that M is partioned into disjoint orbits. The F-structure is said to be *polarized* if the local actions are locally free.

Consider the following example of a polarized F-structure on a graph manifold. Take a compact surface S with non-empty connected boundary, and consider two copies of $S \times \mathbb{S}^1$, each of which has a 2-torus boundary. Fixing an identification of ∂S with $\mathbb{S}^1$, glue the boundaries of two copies of $S \times \mathbb{S}^1$ by a map that interchanges the $\mathbb{S}^1$ factors, so that $(x, z) \in \partial S \times \mathbb{S}^1$ on one copy is glued to $(z, x) \in \partial S \times \mathbb{S}^1$ on the other copy.

The resulting manifold admits a free circle action on each copy of $\mathrm{int}(S) \times \mathbb{S}^1$, but at their common boundary the actions do not agree. However, they do generate a 2-torus action which acts locally near their common boundary, thus defining a polarized F-structure on the whole manifold.

2.8. An important chain of inequalities. — Let M be a closed Riemannian manifold with smooth metric g, and let $\widetilde{M}$ be its universal covering endowed with the induced metric. For each $x \in \widetilde{M}$, let $V(x, r)$ be the volume of the ball with center x and radius r. Set

$$\lambda(g) := \lim_{r \to +\infty} \frac{1}{r} \log V(x, r).$$

Manning [**13**] showed that this limit exists and is independent of x.

Set

$$\lambda(M) := \inf\{\lambda(g) \mid g \text{ is a smooth metric on } M \text{ with } \mathrm{Vol}(M, g) = 1\}.$$

It is well known, see Milnor [**15**], that $\lambda(g)$ is positive if and only if $\pi_1(M)$ has exponential growth. Manning's inequality [**13**] asserts that for any metric g,

$$\lambda(g) \leqslant \mathrm{h_{top}}(g). \tag{1}$$

In particular, it follows that if $\pi_1(M)$ has exponential growth, then $\mathrm{h_{top}}(g)$ is positive for any metric g. (This fact was first observed by Dinaburg [**5**]). Gromov [**7**] showed that if $\mathrm{Vol}(M, g) = 1$, then

$$\frac{1}{C_n\, n!} \|M\| \leqslant [\lambda(g)]^n, \tag{2}$$

where

$$C_n = \Gamma\left(\frac{n}{2}\right) / \sqrt{\pi}\, \Gamma\left(\frac{n+1}{2}\right).$$

Finally it was observed by Paternain [**20**] that

$$[\mathrm{h}(M)]^n \leqslant (n-1)^n \mathrm{MinVol}(M). \tag{3}$$

Combining equations (1), (2), and (3), we obtain the following chain of inequalities:

$$\frac{1}{C_n n!}\|M\| \leqslant [\lambda(M)]^n \leqslant [\mathrm{h}(M)]^n \leqslant (n-1)^n \mathrm{MinVol}(M). \tag{4}$$

We note here that the only known 3-manifolds with $\mathrm{h}(M) > 0$ are those with $\|M\| \neq 0$. In fact it follows from Theorem 2.7, Proposition 2.8, and the chain of inequalities (4) that if M is a closed orientable geometrizable 3-manifold, then the vanishing of the simplicial volume implies that $\mathrm{h}(M) = 0$.

We encode this information in the following theorem.

Theorem 2.9. — *Let M a closed orientable geometrizable 3-manifold. Then the following are equivalent:*

(1) *the minimal volume* $\mathrm{MinVol}(M)$ *of M vanishes;*
(2) *the minimal entropy* $\mathrm{h}(M)$ *of M vanishes;*
(3) *the simplicial volume* $\|M\|$ *of M vanishes;*
(4) *M is a graph manifold.*

3. Geometric structures and the minimal entropy problem

In this section, we consider the minimal entropy problem for those 3-manifolds that admit a single geometric structure. Namely, we prove the following.

Proposition 3.1. — *Let M be a closed orientable 3-manifold. Suppose that M admits a geometric structure. Then, the minimal entropy problem for M can be solved if and only if M admits a geometric structure modelled on $\mathbb{S}^3$, $\mathbb{E}^3$, $\mathbb{S}^2 \times \mathbb{R}$,* Nil, *or $\mathbb{H}^3$. Moreover, if M admits a geometric structure modelled on $\mathbb{S}^3$, $\mathbb{E}^3$, $\mathbb{S}^2 \times \mathbb{R}$, or* Nil, *then M admits a smooth metric g with $\mathrm{h}_{\mathrm{top}}(g) = 0$.*

Proof. — We start by showing that if M admits a geometric structure modelled on one of these 5 geometries, then the minimal entropy problem for M can be solved. Observe first that if M admits a geometric structure modelled on $\mathbb{H}^3$, then the minimal entropy problem can be solved by the results of Besson, Courtois and Gallot [**1**].

It follows immediately from Theorem 2.5 that if M admits a geometric stucture modelled on one of the seven geometries $\mathbb{S}^3$, $\mathbb{E}^3$, $\mathbb{S}^2 \times \mathbb{R}$, $\mathbb{H}^2 \times \mathbb{R}$, $\widetilde{\mathrm{SL}}_2$, Nil, or Sol, then M is a graph manifold. Hence by Proposition 2.8 and the chain of inequalities (4), we have that for such an M, the minimal entropy satisfies $\mathrm{h}(M) = 0$.

We now show that if M admits a geometric structure modelled on one of $\mathbb{S}^3$, $\mathbb{E}^3$, $\mathbb{S}^2 \times \mathbb{R}$, or Nil, then the minimal entropy problem for M can be solved. To do this, we need to show that M admits a smooth metric g with $\mathrm{h}_{\mathrm{top}}(g) = 0$.

(1) $\mathbb{S}^3$, $\mathbb{E}^3$, $\mathbb{S}^2 \times \mathbb{R}$: All the Jacobi fields in these geometries grow at most linearly (in the case of $\mathbb{S}^3$ they are actually bounded), and hence all the Liapunov exponents of *every* geodesic in M are zero. It follows from Ruelle's inequality [**24**] that all the measure entropies are zero. Hence, by the variational principle, the topological entropy of the geodesic flow of M must be zero.

(2) Nil: This geometry can be described as $\mathbb{R}^3$ with the metric

$$ds^2 = dx^2 + dy^2 + (dz - xdy)^2.$$

Here, not all the Jacobi fields grow linearly, but they certainly grow polynomially. Again this implies that all the Liapunov exponents of *every* geodesic in M are zero and hence the topological entropy of the geodesic flow of M must be zero.

Since we have assumed that M admits a geometric structure, we complete the proof by showing that if M admits a geometric structure modelled on one of the remaining geometries, namely $\mathbb{H}^2 \times \mathbb{R}$, $\widetilde{\mathrm{SL}}_2$, and Sol, then M cannot admit a metric of zero topological entropy. To do this, we use the next lemma, together with the fact described in Subsection 2.8, that if $\pi_1(M)$ grows exponentially, then $\mathrm{h_{top}}(g) > 0$ for any smooth metric g on M.

Lemma 3.2. — *Let M be a closed orientable 3-manifold, and suppose that M admits a geometric structure modelled on one of $\mathbb{H}^2 \times \mathbb{R}$, $\widetilde{\mathrm{SL}}_2$, or* Sol. *Then $\pi_1(M)$ grows exponentially.*

Proof. — In the case that M admits a geometric structure modelled on $\mathbb{H}^2 \times \mathbb{R}$ or $\widetilde{\mathrm{SL}}_2$, we start by recalling from Theorem 2.5 that M is then a Seifert fibered space. The base orbifold of the Seifert fiber space admits a hyperbolic structure, and so the orbifold fundamental group of the base orbifold contains a free subgroup of rank 2, and hence so does $\pi_1(M)$. Hence, $\pi_1(M)$ grows exponentially.

In the case that M admits a geometric structure modelled on Sol, we have that M is finitely covered by the mapping torus N of a hyperbolic automorphism of a 2-torus. Note that a hyperbolic automorphism of a 2-torus is an Anosov diffeomorphism, and so the suspension flow on N is an Anosov flow. It is known that the fundamental group of a 3-manifold with an Anosov flow has exponential growth (see for example Plante and Thurston [**22**]). □

This completes the proof of Proposition 3.1. □

4. Proof of Theorem A

Up to this point, we have been considering the minimal entropy problem for closed 3-manifolds that admit a single geometric structure. In this section, we consider a more general geometrizable 3-manifold.

Theorem A. — *Let M be a closed orientable irreducible 3-manifold whose fundamental group contains a $\mathbb{Z} \oplus \mathbb{Z}$ subgroup. The following are equivalent:*

(1) *the simplicial volume $\|M\|$ of M is zero and the minimal entropy problem for M can be solved;*

(2) *M admits a geometric structure modelled on $\mathbb{E}^3$ or* Nil*;*

(3) *M admits a smooth metric g with $\mathrm{h}_{\mathrm{top}}(g) = 0$.*

Proof. — Let us show that item 1 implies item 2. Suppose then that M has zero simplicial volume and that the minimal entropy problem for M can be solved. We show that M must then admit a geometric structure modelled on either $\mathbb{E}^3$ or Nil. Since the fundamental group of M contains a $\mathbb{Z} \oplus \mathbb{Z}$ subgroup, Theorem 2.3 ensures that either M contains an incompressible embedded 2-torus or M is a Seifert fibered space. We now split the proof into two cases:

– Suppose first that M contains an incompressible embedded 2-torus, and so is sufficiently large. Since we have assumed that $\|M\| = 0$, Theorem 2.7 yields that M is a graph manifold. Hence, by Theorem 2.9, we have that $\mathrm{h}(M) = 0$.

However, using work of Evans and Moser [**6**], specifically Theorem 4.2 and Corollary 4.10 in [**6**], we see that either $\pi_1(M)$ contains a free subgroup of rank 2 or M is finitely covered by a 2-torus bundle over $\mathbb{S}^1$. In the former case, $\pi_1(M)$ grows exponentially and therefore the minimal entropy problem cannot be solved for M.

In the latter case, M admits a geometric structure modelled on one of $\mathbb{E}^3$, Nil, or Sol (cf. Theorem 5.5 of Scott [**26**]). However, in the case that M admits a geometric structure modelled on Sol, we know from Proposition 3.1 that the minimal entropy problem cannot be solved for M.

Hence, if the minimal entropy problem can be solved for M and if M contains an incompressible embedded 2-torus, then M admits a geometric structure modelled on either $\mathbb{E}^3$ or Nil.

– The other case is that M is a Seifert fibered space. Here, Theorem 2.5 ensures that M possesses a geometric structure modelled on one of $\mathbb{S}^3$, $\mathbb{E}^3$, $\mathbb{S}^2 \times \mathbb{R}$, $\mathbb{H}^2 \times \mathbb{R}$, $\widetilde{\mathrm{SL}}_2$ or Nil.

Since the fundamental group of M admits a $\mathbb{Z} \oplus \mathbb{Z}$ subgroup, the geometric structure on M cannot be modelled on $\mathbb{S}^3$ or $\mathbb{S}^2 \times \mathbb{R}$. Since we have assumed that the minimal entropy problem can be solved for M, Proposition 3.1 yields that M must admit a geometric structure modelled on either $\mathbb{E}^3$ or Nil, as desired.

To see that item 2 implies item 3, recall from Proposition 3.1 that if M admits a geometric structure modelled on $\mathbb{E}^3$ or Nil, then M admits a smooth metric g with $\mathrm{h}_{\mathrm{top}}(g) = 0$.

Finally to prove that item 3 implies item 1, observe that if M admits a smooth metric g with $\mathrm{h}_{\mathrm{top}}(g) = 0$ it then follows from inequalities (1) and (2) that M has zero simplicial volume.

This completes the proof of Theorem A. □

5. Proof of Theorem B

We are now ready to consider the minimal entropy problem for a general geometrizable 3-manifold with zero simplicial volume.

Theorem B. — *Let M be a closed orientable geometrizable 3-manifold. The following are equivalent:*

(1) *the simplicial volume $\|M\|$ of M is zero and the minimal entropy problem for M can be solved;*

(2) *M admits a geometric structure modelled on $\mathbb{S}^3$, $\mathbb{S}^2 \times \mathbb{R}$, $\mathbb{E}^3$, or* Nil*;*

(3) *M admits a smooth metric g with $\mathrm{h_{top}}(g) = 0$.*

Proof. — Let us prove that item 1 implies item 2. Suppose that M has zero simplicial volume and that the minimal entropy problem for M can be solved. Since M is geometrizable and its simplicial volume vanishes, Theorem 2.7 tells us that M is a graph manifold. Hence, by Theorem 2.9, M has zero minimal entropy.

Since we are assuming that the minimal entropy problem can be solved for M, the fact that M has zero minimal entropy in turn implies there exists a smooth metric on M with zero topological entropy. This in turn implies, by the discussion in Section 2.8, that $\pi_1(M)$ does not have exponential growth.

However, it is a fact from combinatorial group theory (which follows immediately from the existence of normal forms for free products, for instance) that if A and B are two finitely generated groups, then the free product $A * B$ contains a free subgroup of rank two unless A is trivial or B is trivial, or A and B are both of order two. Since the fundamental group of a connected sum is the free product of the fundamental groups of the summands, we conclude that either the prime decomposition is trivial or there are only two summands both of which have fundamental group $\mathbb{Z}_2$.

In the former case, it follows that M must be either irreducible or $\mathbb{S}^2 \times \mathbb{S}^1$, while in the latter case M must be $\mathbb{P}^3 \# \mathbb{P}^3$, where $\mathbb{P}^3$ is the 3-dimensional real projective space. Since $\mathbb{S}^2 \times \mathbb{S}^1$ and $\mathbb{P}^3 \# \mathbb{P}^3$ both admit a geometric structure modelled on $\mathbb{S}^2 \times \mathbb{R}$, we may assume from now on that M is irreducible.

There are now several cases, depending on $\pi_1(M)$. Suppose first that $\pi_1(M)$ is finite. Since M is geometrizable, we have that M admits a geometric structure modelled on $\mathbb{S}^3$.

In the case that $\pi_1(M)$ is infinite and contains a $\mathbb{Z} \oplus \mathbb{Z}$ subgroup, the assumption that the simplicial volume of M is zero, together with the fact that the minimal entropy problem can be solved for M, allows us to apply Theorem A to see that M admits a geometric structure modelled on $\mathbb{E}^3$ or Nil.

The remaining case is that $\pi_1(M)$ is infinite and does not contain a $\mathbb{Z}\oplus\mathbb{Z}$ subgroup. Since M is geometrizable, either M admits a hyperbolic structure or M is Seifert fibered. (Since $\pi_1(M)$ does not contain a $\mathbb{Z}\oplus\mathbb{Z}$ subgroup, M cannot admit a geometric structure modelled on Sol, as Sol manifolds are finitely covered by 2-torus bundles over the circle.) However, since $\|M\| = 0$, M cannot admit a hyperbolic structure.

Note though that M cannot admit a geometric structure modelled on $\mathbb{H}^2 \times \mathbb{R}$, $\mathbb{E}^3$, $\widetilde{\mathrm{SL}_2}$, or Nil, as such manifolds always have a $\mathbb{Z} \oplus \mathbb{Z}$ in their fundamental groups, by Lemma 2.1. Hence, the only possibilities remaining are that M admits a geometric structure modelled on either $\mathbb{S}^2 \times \mathbb{R}$ or $\mathbb{S}^3$, as desired.

To see that item 2 implies item 3, recall from Proposition 3.1 that if M admits a geometric structure modelled on $\mathbb{S}^3$, $\mathbb{S}^2 \times \mathbb{R}$, $\mathbb{E}^3$, or Nil, then M admits a smooth metric g with $\mathrm{h}_{\mathrm{top}}(g) = 0$.

Finally to prove that item 3 implies item 1, observe that if M admits a smooth metric g with $\mathrm{h}_{\mathrm{top}}(g) = 0$, it then follows from inequalities (1) and (2) that M has zero simplicial volume.

This completes the proof of Theorem B. □

References

[1] G. Besson, G. Courtois, S. Gallot, 'Entropies et rigidités des espaces localement symétriques de courbure strictement négative', *Geom. Funct. Anal.* **5** (1995) 731–799.

[2] F. Bonahon, 'Geometric structures on 3-manifolds', *Handbook of Geometric Topology*, edited by R. Daverman and R. Sher, North-Holland, Amsterdam, 2002, 93–164.

[3] J. Cheeger, M. Gromov, 'Collapsing Riemannian manifolds while keeping the curvature bounded. I', *J. Diff. Geometry* **23** (1986) 309–364.

[4] C. Connell and B. Farb, 'Minimal entropy rigidity for lattices in products of rank one symmetric spaces', to appear in Comm. Anal. Geom.

[5] E.I. Dinaburg, 'On the relations among various entropy characteristics of dynamical systems', *Math. USSR Izv.* **5** (1971) 337–378.

[6] B. Evans, L. Moser, 'Solvable fundamental groups of compact 3-manifolds', *Trans. Amer. Math. Soc.* **168** (1972), 189–210.

[7] M. Gromov, 'Volume and bounded cohomology', *Publ. Math. I.H.E.S.* **56** (1982) 213–307.

[8] J. Hempel, *3-manifolds*, Annals of Mathematics Studies **86**, Princeton University Press, 1976.

[9] W. Jaco and P.B. Shalen, *Seifert fibered spaces in 3-manifolds*, Memoirs of the AMS **220**, American Mathematical Society, Providence, 1979.

[10] K. Johannson, 'Homotopy equivalences of 3-manifolds with boundary', Lecture Notes in Mathematics **761**, Springer-Verlag, 1979.

[11] A. Katok, 'Entropy and closed geodesics', *Ergod. Th. and Dynam. Syst.* **2** (1982) 339–369.

[12] H. Kneser, 'Geschlossene Flächen in dreidimensionale Mannigflatigkeiten', Jahresber. Deutsch. Mat. Verein **38** (1929) 248–260.

[13] A. Manning, 'Topological entropy for geodesic flows', *Ann. Math.* **110** (1979), 567–573.

[14] R. Mañé, 'On the topological entropy of geodesic flows', *J. Differential Geom.* **45** (1997) 74–93.

[15] J. Milnor, 'A note on curvature and fundamental group', *J. Diff. Geometry* **2** (1968) 1–7.

[16] J. Milnor, 'A unique factorisation theorem for 3-manifolds', *Amer. J. Math.* **84** (1962) 1–7.

[17] J. Morgan, 'On Thurston's uniformization theorem for 3-dimensional manifolds', in *The Smith conjecture*, ed. by H. Bass and J. Morgan, Academic Press, 1984, 37–126.

[18] G. Mostow, 'Strong rigidity of locally symmetric spaces', Annals of Mathematics Studies **78**, Princeton University Press, 1973.

[19] J.-P. Otal, 'Thurston's hyperbolization of Haken manifolds', in *Surveys in differential geometry III*, ed. by C.-C. Hsiung and S.-T. Yau, International Press, Boston, 1998, 77–194.

[20] G.P. Paternain, *Geodesic Flows*, Progress in Mathematics **180**, Birkhäuser, 1999.

[21] G.P. Paternain, J. Petean, Minimal entropy and collapsing with curvature bounded from below, *Invent. Math.* **151** (2003) 415–450.

[22] J.F. Plante, W.P. Thurston, 'Anosov flows and the fundamental group', *Topology* **11** (1972) 147–150.

[23] X. Rong, 'The limiting eta invariants of collapsed three-manifolds', *J. Diff. Geometry* **37** (1993) 535–568.

[24] D. Ruelle, 'An inequality for the entropy of differentiable maps', *Bol. Soc. Brasil. Mat.* **9** (1978) 83–87.

[25] G.P. Scott, 'A new proof of the annulus and torus theorems', *Amer. J. Math.* **102** (1980) 241–277.

[26] G.P. Scott, 'The geometries of 3-manifolds', *Bull. Lond. Math. Soc.* **15** (1983) 401–487.

[27] G.P. Scott, Math. Review 96f:57011, 1996.

[28] T. Soma, 'The Gromov invariant for links', *Invent. Math.* **64** (1981) 445–454.

[29] W.P. Thurston, 'The geometry and topology of three-manifolds', electronic version 1.0-October 1997.

[30] W.P. Thurston, 'Three dimensional manifolds, Kleinian groups, and hyperbolic geometry', *Bull. Amer. Math. Soc.* **6** (1982), 357–381.

J.W. ANDERSON, Faculty of Mathematical Studies, University of Southampton, Southampton SO17 1BJ, England • *E-mail :* `j.w.anderson@maths.soton.ac.uk`

G.P. PATERNAIN, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge CB3 0WB, England • *E-mail :* `G.P.Paternain@dpmms.cam.ac.uk`

Astérisque
286, 2003, p. 81–118

STATISTICAL PROPERTIES OF UNIMODAL MAPS: SMOOTH FAMILIES WITH NEGATIVE SCHWARZIAN DERIVATIVE

by

Artur Avila & Carlos Gustavo Moreira

Abstract. — We prove that there is a residual set of families of smooth or analytic unimodal maps with quadratic critical point and negative Schwarzian derivative such that almost every non-regular parameter is Collet-Eckmann with subexponential recurrence of the critical orbit. Those conditions lead to a detailed and robust statistical description of the dynamics. This proves the Palis conjecture in this setting.

1. Introduction

'The main strategy of the study of all mathematical models is, according to Poincaré, the consideration of each model as a point of a space of different but similar admissible systems' (V. Arnold in [**Ar**]). One of the main concerns of dynamical systems is to establish properties valid for typical systems. Since the space of such systems is usually infinite dimensional, there are of course many concepts of 'typical'. According to [**Ar**] again, 'The most physical genericity notion is defined by Kolmogorov (1954), who suggested to call a property of dynamical systems exceptional, if it holds only on Lebesgue measure zero set of values of the parameters in every (topologically) generic family of systems, depending on sufficiently many parameters'.

In the last decade Palis [**Pa**] described a general program for (dissipative) dynamical systems in any dimension. He conjectured that a typical dynamical system has a finite number of attractors described by physical measures, the union of their basins has full Lebesgue measure, and those physical measures are stochastically stable. Typical was to be interpreted in the Kolmogorov sense: full measure in generic families. Our aim here is to give a proof of this conjecture for an important class of one-dimensional dynamical systems.

Here we consider unimodal maps, that is, continuous maps from an interval to itself which have a unique turning point. More specifically, we consider S-unimodal maps, that is, we assume that the map is C^3 with negative Schwarzian derivative and that the critical point is non-degenerate.

2000 ***Mathematics Subject Classification.*** — 37E05, 37C40, 37A25.
Key words and phrases. — Unimodal maps, decay of correlations, stochastic stability.

Partially supported by Faperj and CNPq, Brazil.

1.1. The quadratic family. — The basic model for unimodal maps is the quadratic family, $q_a(x) = a - x^2$, where $-1/4 \leqslant a \leqslant 2$ is a parameter. Despite its simple appearance, the dynamics of those maps presents many remarkable phenomena. Restricting to the probabilistic point of view, its richness first became apparent with the work of Jakobson [**J**], where it was shown that a positive measure set of parameters corresponds to quadratic maps with stochastic behavior. More precisely, those parameters possess an absolutely continuous invariant measure (the physical measure) with positive Lyapunov exponent. On the other hand, it was later shown by Lyubich [**L2**] and Graczyk-Swiatek [**GS**] that regular parameters (with a periodic hyperbolic attractor) are (open and) dense. So at least two kinds of very distinct observable behavior are present on the quadratic family, and they alternate in a complicate way.

Besides regular and stochastic behavior, different behavior was shown to exist, including examples with bad statistics, like absence of a physical measure or a physical measure concentrated on a hyperbolic repeller. Those pathologies were shown to be non-observable in [**L3**] and [**MN**]. Finally in [**L4**] it was proved that almost every real quadratic map is either regular or stochastic.

Among stochastic maps, a specific class grabbed lots of attention in the 90's: Collet-Eckmann maps. They are characterized by a positive Lyapunov exponent for the critical value, and gradually they were shown to have 'best possible' near hyperbolic properties: exponential decay of correlations, validity of central limit and large deviations theorems, good spectral properties and zeta functions ([**KN**], [**Y**]). Let us call attention to the robustness of the statistical description, with a good understanding of stochastic perturbations: strong stochastic stability ([**BV**]), rates of convergence to equilibrium ([**BBM**]).

In [**AM1**] the regular or stochastic dichotomy was extended by showing that almost every stochastic map is actually Collet-Eckmann and has polynomial recurrence of its critical point, in particular implying the validity of the above mentioned results.

The position of the quadratic family in the borderline of real and complex dynamics made it a meeting point of many different techniques: most of the deeper results depend on this interaction. It gradually became clear however that studying the quadratic family allows one to obtain results on more general unimodal maps.

1.2. Universality. — Starting with the works of Milnor-Thurston, and also through the discoveries of Feigenbaum and Coullet-Tresser, the quadratic family was shown to be a prototype for other families of unimodal maps which presents universal combinatorial and geometric features. More recently, the result of density of hyperbolicity among unimodal maps was obtained in [**K**] exploiting the validity of this result for quadratic maps.

In [**ALM**], a general method was developed to transfer information from the quadratic family to real analytic families of unimodal maps. It was shown that the decomposition of spaces of analytic unimodal maps according to combinatorial behavior is essentially a codimension-one lamination.

Thinking of two analytic families as transversals to this lamination, one may try to compare the parameter space of both families via the holonomy map. A straightforward application of this method allows one to conclude that the bifurcation pattern of a general analytic family is locally the same as in the quadratic family from the topological point of view (outside of countably many 'bad parameters').

The 'holonomy' method was then successfully applied to extend the regular or stochastic dichotomy from the quadratic family to a general analytic family. The probabilistic point of view presents new difficulties however. First, the statistical properties of two topologically conjugate maps need not correspond by the (generally not absolutely continuous) conjugacy. Fortunately many properties are preserved, in particular the criteria used by Lyubich in his result.

The second difficulty is that the holonomy map is usually not absolutely continuous, so typical combinatorics for the quadratic family may not be typical for other families: it has to be shown that the class of regular or stochastic maps is still typical after application of the holonomy map.

1.3. Results and outline of the proof. — Let us call a k-parameter family good if almost every non-regular parameter is Collet-Eckmann (and satisfies some additional technical conditions). Our goal will be to prove that good families are generic. This question naturally makes sense in different spaces of unimodal maps (corresponding to different degrees of smoothness). We only deal with the last steps of this problem (going from the quadratic family to analytic and then smooth categories), basing ourselves on the building blocks [**L3**], [**L4**], [**ALM**], and [**AM1**].

We start by describing how the holonomy method of [**ALM**] can be applied to generalize the results of [**AM1**] to general analytic families (to put together those two papers we need to do a non-trivial strengthening of [**AM1**]). As a consequence we conclude that essentially all analytic families are good.

To get to the smooth setting (at least C^3, since we are assuming negative Schwarzian derivative), our strategy is different: we show a certain robustness of good families, which together with their denseness (due to the analytic case) will yield genericity. Our main tool is one of the nice properties of Collet-Eckmann maps: persistence of the Collet-Eckmann condition under generic unfolding (a result of [**T1**]). By means of some general argument, we reduce the global result to this local one.

Let us mention that the results of this paper are still valid without the negative Schwarzian derivative assumption (also allowing one to get to C^2 smoothness), see [**A**], [**AM4**]. The techniques are very different however, since we replace the global holonomy method we use here by a local holonomy analysis based on a "macroscopic" version of the infinitesimal perturbation method of [**ALM**]. For analytic maps this also allowed us to obtain better asymptotic estimates which have interesting consequences, for instance pathological measure-theoretical behavior of the lamination by combinatorial classes (see [**AM2**]).

Acknowledgements. — We thank Viviane Baladi, Mikhail Lyubich, and Marcelo Viana for helpful discussions and suggestions.

2. General definitions

2.1. Notation. — Let $I = [-1, 1]$ and let B^k be the closed unit ball in $\mathbb{R}^k$ (we will use the notation I for the *dynamical* interval, while B^1 will be reserved for the one-dimensional *parameter* space). We will consider B^k endowed with the Lebesgue measure normalized so that $|B^k| = 1$. Let $C^r(I)$ denote the space of C^r maps $f : I \to \mathbb{R}$.

By a *unimodal map* we will mean a smooth (at least C^2) symmetric (even) map $f : I \to I$ with a unique critical point at 0 such that $f(-1) = -1$, $Df(-1) \geqslant 1$, and if $Df(-1) = 1$ then $D^2f(-1) < 0$. If f is C^3, we define the Schwarzian derivative on $I \setminus \{0\}$ as

$$Sf = \frac{D^3f}{Df} - \frac{3}{2}\left(\frac{D^2f}{Df}\right)^2.$$

For $a > 0$, let $\Omega_a \subset \mathbb{C}$ denote an a neighborhood I.

Let $\mathcal{A}_a$ denote the space of holomorphic maps on Ω_a which have a continuous extension to $\partial\Omega_a$, satisfying $\phi(z) = \phi(-z)$, $\phi(-1) = \phi(1) = -1$ and $\phi'(0) = 0$.

Notice that $\mathcal{A}_a$ is a closed affine subspace of the Banach space of bounded holomorphic maps of Ω_a. We endow it with the induced metric and affine structure.

We define $\mathcal{A}_a^{\mathbb{R}} \subset \mathcal{A}_a$ the space of maps which are real symmetric.

2.2. More on unimodal maps. — A C^3 unimodal map such that $Sf < 0$ on $I \setminus \{0\}$ and such that its critical point is non-degenerate (that is, $D^2f \neq 0$) will be called a *S-unimodal map*.

We say that x is a periodic orbit (of period n) for f if $f^n(x) = x$ and $n \geqslant 1$ is minimal with this property. In this case we define $Df^n(x)$ as the multiplier of x. Notice that this definition depends only on the orbit of x. We say that x is hyperbolic if $|Df^n(x)| \neq 1$.

A unimodal map is called *regular* (or hyperbolic) if all periodic orbits are hyperbolic and the iterates of the critical point converge to an attracting periodic orbit. This condition is C^2-open, moreover a S-unimodal map is regular if and only if it has a hyperbolic periodic attractor (see [**MvS**]).

A *k-parameter family* of unimodal maps is a map $F : B^k \times I \to I$ such that for $p \in B^k$, $f_p(x) = F(p, x)$ is a unimodal map. Such a family is said to be C^n or analytic, according to F being C^n or analytic. We introduce the natural topology in spaces of smooth families (C^n with $n = 2, \ldots, \infty$), but do not introduce any topology in the space of analytic families (however, we will refer from time to time to induced C^n topologies).

An analytic family of S-unimodal maps F will be called *non-trivial* if there exists a regular parameter. Notice that this condition is C^3-open.

A unimodal map f is called *Collet-Eckmann* (CE) if there exists constants $C > 0$, $\lambda > 1$ such that for every $n > 0$,

$$|Df^n(f(0))| > C\lambda^n.$$

This means that the map is strongly hyperbolic along the critical orbit. It is also useful to study the hyperbolicity of backward iterates of the critical point, so we say that f is *Backwards Collet-Eckmann* (BCE) if there exists $C > 0$, $\lambda > 1$ such that for any $n > 0$ and any x with $f^n(x) = 0$, we have

$$|Df^n(x)| > C\lambda^n.$$

By a result of Nowicki (see [**MvS**]), for S-unimodal maps CE implies BCE, so we will mostly discuss the Collet-Eckmann condition (except for the last section where we consider C^2 unimodal maps as well).

Very often it is useful to estimate how fast is the recurrence of the critical orbit. We will be mainly interested in two kinds of control: *Polynomial Recurrence* (P) if there exists $\alpha > 0$ such that

$$|f^n(0)| > n^{-\alpha}$$

for big enough n and *Subexponential Recurrence* (SE) if for all $\alpha > 0$,

$$|f^n(0)| > e^{-\alpha n}$$

for n big enough.

We will say that f is *Weakly Regular* (WR) if

$$\lim_{\delta \to 0} \liminf_{n \to \infty} \frac{1}{n} \sum_{\substack{1 \leqslant k \leqslant n \\ f^k(0) \in (-\delta, \delta)}} \ln |Df(f^k(0))| = 0.$$

This condition is used in proofs of stochastic stability for C^2 maps, see [**T2**].

We will consider spaces of S-unimodal maps: we define $\mathcal{U}^r \subset C^r(I)$ the set of S-unimodal maps. Spaces of analytic unimodal maps are now easily defined: $\mathcal{U}_a = \mathcal{U}^3 \cap \mathcal{A}_a^{\mathbb{R}}$.

2.3. The quadratic family. — The quadratic family is the most studied family of unimodal maps. It is usually parametrized by

$$q_t(x) = t - x^2,$$

so that for $-1/4 \leqslant t \leqslant 2$, there exists a unique symmetric interval $I_t = [-\beta_t, \beta_t]$ such that $q_t(I_t) \subset I_t$ and $q_t(-\beta_t) = -\beta_t$, so q_t can be seen as a unimodal map of I_t (which depends on t). Moreover $Sq_t(x) < 0$ if $x \neq 0$.

By an affine reparametrization of the parameter t and of each interval I_t, we obtain a canonical one-parameter family of S-unimodal maps in the interval I, which we denote p_t, $t \in B^1$, which will be called the quadratic family as well.

2.4. Quasisymmetric maps. — Let $\gamma \geqslant 1$ be given. We say that a homeomorphism $f : \mathbb{R} \to \mathbb{R}$ is *quasisymmetric* (qs) if there exists a constant $k > 1$ such that for all $x \in \mathbb{R}$ and any $h > 0$

$$\frac{1}{k} \leqslant \frac{f(x+h) - f(x)}{f(x) - f(x-h)} \leqslant k.$$

A homeomorphism h is quasisymmetric if and only if it admits a real-symmetric extension to a quasiconformal map $\widetilde{h} : \mathbb{C} \to \mathbb{C}$ (Ahlfors-Beurling). We will say that h is γ-qs (or that γ is a qs constant for h) if the dilatation of $\widetilde{h}$ is bounded by γ. This definition of the quasisymmetric constant is convenient since the composition of quasisymmetric maps g and f is readily seen to be quasisymmetric and the qs constant of $g \circ f$ is bounded by the product of the qs constants of g and f.

If $X \subset \mathbb{R}$ and $h : X \to \mathbb{R}$ has a γ-quasisymmetric extension to $\mathbb{R}$ we will also say that h is γ-qs.

3. Statement of the results

3.1. A dichotomy for generic families of S-unimodal maps. — We would like to classify the typical behavior in generic families of unimodal maps. This classification should reveal refined information on the stochastic description of the dynamics of those typical parameters.

We will therefore consider a smooth enough family of unimodal maps F. The techniques of the present paper will need the fact that F is a family of S-unimodal maps. This includes two main restrictions: the negative Schwarzian derivative and the quadratic critical point. The first one is serious, since this condition is not dense, but can be removed with more refined techniques (see [**A**]). The second one (which is not present in the usual definition of S-unimodal map, but is rather a convention in this paper) is no serious loss of generality, since quadratic critical point is certainly typical among unimodal maps.

Remark 3.1. — Families of unimodal maps with a fixed critical exponent different from 2 have also been subject of much study. This theory has many similarities, but also some important differences and new features, and is not nearly as complete as the case of criticality 2. It is however widely expected that the Palis conjecture (and indeed our Theorems A, B and C) still holds in this setting.

We first consider the analytic case.

Theorem A. — *Let F be a non-trivial k-parameter analytic family of S-unimodal maps. Then for almost every non-regular parameter $p \in B^k$, f_p satisfies the Collet-Eckmann and Polynomial Recurrence conditions.*

Notice that the set of non-trivial analytic families is indeed generic in any meaningful sense: its complement has "infinite codimension", see Proposition 4.3. Moreover, if an analytic family is non-trivial, it is possible to verify the non-triviality in finite time (with an infinite precision computer [1]).

Our second result about non-trivial analytic families is the robustness of a slightly weaker dichotomy under C^2 perturbations of the family.

Theorem B. — *Let F be a non-trivial k-parameter analytic family of S-unimodal maps. Let $F^{(n)}$ be a sequence of C^2 families such that $F^{(n)} \to F$ in the C^2 topology. For each n, let X_n be the set of parameters $p \in B^k$ where $F^{(n)}$ is either regular or has only repelling periodic orbits and satisfies simultaneously the Backwards Collet-Eckmann, Collet-Eckmann, Subexponential Recurrence and Weak Regularity conditions. Then $|X_n| \to 1$. In particular, almost every parameter of F is Weakly Regular.*

As a consequence, we can use a Baire argument to conclude that the dichotomy is still valid among topologically generic smooth families (that is, belonging to some residual set), obtaining the following corollary of Theorems A and B.

***Theorem C* (Smooth Dichotomy)**. — *In topologically generic k-parameter C^r, $r = 3, 4, \dots, \infty$ families of S-unimodal maps, almost every non-regular parameter satisfies the Backwards Collet-Eckmann, Collet-Eckmann, Subexponential Recurrence and Weak Regularity conditions.*

It is good to recall that both types of behavior described by the dichotomy are indeed observable for open sets of families of unimodal maps ([**J**], [**BC**]).

Remark 3.2. — The space of S-unimodal maps is easy to describe and easier to work with but has some disadvantages. One of them is that it is not an intrinsic condition, in particular it is not invariant by analytic change of coordinates. A more natural class to work with is the space of quasiquadratic unimodal maps as defined by [**ALM**]. A unimodal map f is called quasiquadratic if there exists a C^3-neighborhood of f where all maps are topologically conjugate to some quadratic map. The results of this paper are still valid in spaces of quasiquadratic unimodal maps (which includes S-unimodal maps). The proofs are unchanged, since the results we need from [**ALM**] are stated and proved for quasiquadratic maps. We remark further that the description of quasiquadratic unimodal maps can be used to describe all unimodal maps: it is proved

[1] Since regular parameters form an open set (non-empty if the family is non-trivial), and any regular parameter one can be also checked in finite time (by locating the attracting hyperbolic periodic orbit).

in [**A**], [**AM4**] that (Kolmogorov) typical (analytic or smooth) unimodal maps have either a quasiquadratic renormalization or a quasiquadratic unimodal restriction.

3.2. Ergodic consequences. — The importance of the above dichotomy is the fact that each of the two possibilities has very well defined stochastic properties. We quickly recall those (we assume that maps are S-unimodal).

Regular maps have a periodic attractor whose basin is big both topologically (open and dense set) as in the measure-theoretical sense (full measure). Moreover the attractor and its basin are stable under C^1 perturbations. The dynamics of such maps can be described in deterministic terms.

Maps satisfying CE and SE have non-deterministic dynamics. They can be however described through their stochastic properties, and it turns out that such maps have the main good properties usually found in hyperbolic maps. First, there is a physical measure, that is an invariant probability which describes asymptotic behavior of orbits: for almost every x and for every continuous $\phi : I \to \mathbb{R}$,

$$\lim \frac{1}{n} \sum_{k=0}^{n-1} \phi(f^k(x)) = \int \phi d\mu.$$

This physical measure has a positive Lyapunov exponent and is indeed absolutely continuous and supported on a cycle of intervals, so the asymptotic behavior is non-deterministic. The convergence to the asymptotic stochastic model is exponential, see the results on decay of correlations and convergence to equilibrium ([**KN**], [**Y**]). Those properties are beautifully related to a spectral gap of a transfer operator and to zeta functions, see [**KN**]. Notice finally that exponential decay of correlations is actually equivalent to the Collet-Eckmann condition (see [**NS**]).

While the dynamics is highly unstable under deterministic perturbations (nearby maps can be regular for instance), the stochastic description given by the physical measure μ is robust under stochastic perturbations: the perturbed system has a stationary measure which is close to μ in the sense of the L^1 distance between their densities ([**BV**]). For studies of decay of correlations for the perturbed systems, see [**BBM**].

4. Analytic families

4.1. Hybrid classes and holonomy maps. — Two S-unimodal maps $f, \widetilde{f}$ are said to be *hybrid equivalent* if they are topologically conjugate and, in case they are regular, their attracting periodic orbits have the same multiplier.

The set of all maps which are hybrid equivalent to some f is called the *hybrid class* of f. The partition of S-unimodal maps into hybrid classes is thus a refinement of the partition in topological conjugacy classes.

It follows from a result of Guckenheimer (see [**MvS**]) that any S-unimodal map f is topologically conjugate to some quadratic map. It turns out that if f has a hyperbolic attractor, we can select the quadratic map with a hyperbolic attractor with the same multiplier[(2)]. In particular, each hybrid class intersects the quadratic family in at least one point.

The problem of uniqueness is much harder. The following result is due to Lyubich [**L2**] and Graczyk-Swiatek [**GS**], and is a consequence of (the proof of) the equivalent rigidity result for quadratic maps:

Theorem 4.1. — *Let h be a topological conjugacy between two analytic S-unimodal maps f and $\widetilde{f}$ which have all periodic orbits repelling. Then h is quasisymmetric.*

Remark 4.1. — Although we won't use it here, a similar theorem still holds for maps with non-repelling periodic orbits: if f and $\widetilde{f}$ are two topologically conjugate S-unimodal maps and have non-repelling periodic orbits then we can select a topological conjugacy which is quasisymmetric (the choice of the topological conjugacy is not unique). This result is considerably easier than the case where all periodic orbits are repelling, and does not use analyticity.

This rigidity result has a remarkable consequence for quadratic maps: each hybrid class intersects the quadratic family at a unique parameter. Thus, any S-unimodal map f is hybrid equivalent to a unique quadratic map $\chi(f)$. The map χ is called the *straightening* [(3)].

Lemma 4.2. — *Let f be an analytic S-unimodal map. Then $\chi(f)$ is regular/CE/P if and only if f also satisfies the corresponding property.*

Proof. — The property of being regular is clearly invariant under hybrid equivalence, so we only have to analyze invariance of the conditions CE and P.

By [**NP2**], the Collet-Eckmann condition is topologically invariant, so it is preserved under hybrid equivalence.

To check invariance of polynomial recurrence of the critical orbit, first assume that f has some non-repelling periodic point p. In this case, the the orbit of p must attract the critical point. In particular, the critical point is either non-recurrent (in

(2) This follows for instance from Milnor-Thurston kneading theory and the fact that the quadratic family is a full family. Another way to see this is to notice that in each "hyperbolic window" of quadratic maps (a maximal parameter interval (a, b) such that p_t is hyperbolic for $t \in (a, b)$), the multiplier of the hyperbolic attractor induces a homeomorphism from (a, b) to $(-1, 1)$ (this is a consequence for instance of the work of Douady-Hubbard on the complex quadratic family).

(3) We should point out that there is also a notion of hybrid class in complex dynamics. In that context, the fact that each hybrid class (of quadratic-like maps with connected Julia set) contains exactly one quadratic polynomial is a consequence of the Straightening Theorem of Douady-Hubbard. Our definition of hybrid class is motivated precisely by the possibility of defining an analogous straightening map (whose existence is proved by quite different methods).

which case both f and $\chi(f)$ satisfy P in a trivial way) or periodic (in which case f and $\chi(f)$ do not satisfy P also in a trivial way).

If f has all periodic orbits repelling, by Theorem 4.1, the conjugacy between f and $\chi(f)$ is quasisymmetric, and in particular Hölder. It is easy to see that P is invariant by Hölder conjugacy. □

Remark 4.2. — By [**NP1**], two S-unimodal Collet-Eckmann maps which are topologically conjugate are Hölder conjugate, so using [**NP2**] we see that the joint conditions CE and P are topologically invariant. This joint invariance of CE and P is all that will be used in the further arguments. Notice that [**NP1**] and [**NP2**] do not assume analyticity, and are more elementary than Theorem 4.1.

4.2. Hybrid laminations. — It is natural to study the hybrid class of some map f. This is what is done in Theorem A of [**ALM**] in the analytic setting, where it is shown that in $\mathcal{U}_a$, every hybrid class is a codimension-one analytic submanifold. Moreover, different hybrid class fit together in some nice structure, called *hybrid lamination.*

Remark 4.3. — It is not known if the hybrid lamination is really a lamination everywhere. In [**ALM**], it is shown that the hybrid lamination is a lamination (in the usual sense) "almost everywhere" (more precisely, if restricted to an open set containing the complement of countably many classes corresponding to existence of neutral periodic orbits), which is enough for our purposes.

A k-parameter analytic family of S-unimodal maps can be thought as an analytic map from B^k to some $\mathcal{U}_a$. As a consequence, the structure of the hybrid lamination implies that non-trivial analytic families are indeed quite frequent.

***Lemma 4.3* (Most analytic families are non-trivial)**. — *If a k-parameter analytic family of S-unimodal maps is not contained in some non-regular hybrid class then it is non-trivial. In particular, non-trivial analytic families are dense in the space of C^n families of S-unimodal maps, $n = 3, \dots, \infty$.*

Proof. — Let us consider an analytic family of S-unimodal maps F. By the theory of Milnor-Thurston, see [**MvS**], either all parameters have the same non-periodic *kneading sequence*, or there exists a parameter with periodic critical point. In the latter case, the family is of course non-trivial, so let us consider the former case. Two S-unimodal maps with the same kneading sequence are either topologically conjugate, or one of them possess a neutral periodic orbit (see Corollary, Chapter 2, page 157 of [**MvS**]), and it follows that the other is necessarily regular. Thus, if the family F does not have regular parameters, all maps are non-regular and topologically conjugate, that is, F is contained in a non-regular hybrid class.

For the denseness result, given a C^r family F, approximate it by an analytic family $\widetilde{F}$. If such an analytic family is contained in a hybrid class, we can perturb it further in order to intersect two hybrid classes, since each hybrid class is a codimension-one submanifold. □

Let us consider the case where F is a one-parameter analytic family of S-unimodal maps, that is, an analytic curve in some $\mathcal{U}_a$. A consequence of the nice structure of the hybrid lamination is the following result:

***Lemma 4.4* (see the proof of Theorem C of [ALM]).** — *If F is a one-parameter analytic family of S-unimodal maps which is not contained in some hybrid class then there is an open set of parameters, with countable complement, where F is transverse to the hybrid lamination.*

Define the map χ_F on B^1 by $\chi_F(t) = \chi(f_t)$. In [**ALM**] the map χ_F is considered as the holonomy map from F to the quadratic family along the hybrid lamination in some $\mathcal{U}_a$. Using this interpretation, they obtain the following result:

***Theorem 4.5* (Theorem C of [ALM]).** — *Let F be a one-parameter family of unimodal maps which is not contained in some hybrid class. Then there is an open set $U \subset B^1$ with countable complement such that the straightening χ_F is quasisymmetric in any compact interval $J \subset U$.*

4.3. Dichotomy in the quadratic family. — The main result of [**AM1**] is that almost every parameter in the quadratic family is either regular or Collet-Eckmann with a polynomial recurrence of the critical orbit. To obtain the same result for a non-trivial analytic family using Theorem 4.5, we will need a stronger estimate, since quasisymmetric maps are not in general absolutely continuous.

Let us say that a set $X \subset B^1$ has *total qs-probability* if the image of $B^1 \setminus X$ by any quasisymmetric map $h : B^1 \to B^1$ has zero Lebesgue measure.

By an improvement of the proofs in [**AM1**] (see appendix), it is possible to obtain the following result:

***Theorem 4.6*.** — *The set of quadratic maps which are either regular or simultaneously CE and P has total qs-probability.*

***Remark 4.4*.** — In [**AM1**] a better result than polynomial recurrence is obtained in the quadratic family. Namely it is shown that the asymptotic exponent of the recurrence

$$\limsup_{n\to\infty} \frac{-\ln|f^n(0)|}{\ln n}$$

is exactly 1 for almost every non-regular map. However, for a set of total qs-probability, we are only able to show that the asymptotic exponent is bounded.

4.4. Proof of Theorem A. — Let F be a non-trivial analytic family. If all parameters are regular, there is nothing to prove, so assume that there is a non-regular parameter.

First assume F is one-parameter. By Theorems 4.6 and 4.5, for almost every $t \in B^1$, $\chi_F(t)$ is either regular or satisfies CE and P. By Lemma 4.2, this implies that f_t is either regular or CE and P.

Assume now that F is a k-parameter family. Let $p \in B^k$ be a regular parameter. Let $L : B^1 \to B^k$ be an affine map such $p \in L(B^1)$. Let F^L be the one-parameter family defined by $f_t^L = f_{L(t)}$. Then F^L is a non-trivial one-parameter analytic family and hence for almost every t, f_t^L is either regular or CE and P. The result follows by application of Fubini's Theorem.

5. Robustness of the dichotomy

To obtain the robustness claimed on Theorem B our approach will be to exploit an important result of Tsujii, whose core is a strong generalization of Benedicks-Carleson result and techniques. This result establishes that the CE and SE conditions are infinitesimally persistent in one-parameter families unfolding generically: they are density points of CE and SE parameters. The connection with our robustness result, which has a global nature, is done using some general argument.

5.1. Tsujii's theorem. — Let F be a C^2 k-parameter family of unimodal maps. Assume that p_0 is a parameter such that f_{p_0} satisfies CE, BCE, SE, has a quadratic critical point and all periodic orbits repelling. Tsujii's Theorem considers the case where F is a generic unfolding at p_0. For one-parameter families, generic unfolding means precisely

$$\sum_{j=0}^{\infty} \frac{v(f_{p_0}^j(0))}{Df_{p_0}^j(f_{p_0}(0))} \neq 0, \quad \text{where} \quad v = \frac{d}{dp} f_p \Big|_{p=p_0}. \tag{5.1}$$

This transversality condition will be called *Tsujii transversality*.

If F is a one-parameter family, we will say that (F, p_0) satisfies the *Tsujii conditions* if all above requirements are satisfied.

The following is an immediate consequence of the main theorem of Tsujii in [**T1**].

Theorem 5.1. — *Let F be a C^2 one-parameter family of unimodal maps. Assume (F, t_0) satisfies the Tsujii conditions. Then t_0 is a density point of parameters t for which (F, t) satisfies the Tsujii conditions and for which f_t is WR.*

5.2. A higher dimensional version. — In order to pass from one-parameter to k-parameters, we will need the following easy proposition. Let us say that $p \in B^k$ is a density point of a set X along a line l through p if p is a density point of $l \cap X$ in l (endowed with the linear Lebesgue measure).

Proposition 5.2. — *If $p \in B^k$ is a density point of X along almost every line, then p is a density point of X in B^k.*

Proof. — Let E be the characteristic function of X. For each line l through p, let $A_l : \mathbb{R} \to l$ be an isometric parametrization of l taking 0 into p. Let P^{k-1} be the space of such lines with the natural probability measure (obtained by identification with the $k-1$ dimensional projective space). Let

$$\rho_\varepsilon(l) = \int_{-1}^{1} |r| E(A_l(\varepsilon r)) dr.$$

Assuming that p is a density point of X along almost every l we have, for almost every l

$$\lim_{\varepsilon \to 0} \rho_\varepsilon(l) = 1.$$

Using polar coordinates, the relative measure of X in an ε ball around p is given by

$$\int_{P^{k-1}} \rho_\varepsilon(l) dl.$$

By the Lebesgue Convergence Theorem,

$$\lim_{\varepsilon \to 0} \int \rho_\varepsilon(l) dl = \int \lim_{\varepsilon \to 0} \rho_\varepsilon(l) dl = 1.$$

This shows that p is a density point of X. □

We say that a k-parameter F satisfies the Tsujii transversality at p_0 if there exists a line through p_0 along which the one-parameter Tsujii transversality condition is satisfied. In other words, there exists an affine map $L : B^1 \to B^k$ such that $L(t_0) = p_0$ for some $t_0 \in \operatorname{int} B^1$ and such that the induced one-parameter family F^L defined by $f_t^L = f_{L(t)}$ is Tsujii transverse at the parameter t_0.

By linearity of (5.1) with respect to v, if (F, p_0) is Tsujii transverse then all lines passing through p_0 are Tsujii transverse except the lines parallel to a certain codimension-one space of $\mathbb{R}^k$.

Lemma 5.3. — *Let F be a C^2 k-parameter family of unimodal maps. Assume (F, p_0) satisfies the Tsujii conditions. Then p_0 is a density point of parameters p for which (F, p) satisfies the Tsujii conditions and for which f_p is WR.*

Proof. — If F is Tsujii transverse at p_0 then it is Tsujii transverse along almost every line through p_0. Along such a line it is a density point of parameters satisfying the Tsujii conditions and WR. The result follows from Proposition 5.2. □

5.2.1. Tsujii transversality and hybrid lamination. — Let us take a closer look at the Tsujii transversality for an analytic F. Let $f_p = f$.

Assuming the summability condition,

$$\sum_{k=0}^{\infty} \frac{1}{|Df^k(f(0))|} < \infty \tag{5.2}$$

(in particular if f is CE), let

$$\nu_f(v) = \sum_{k=0}^{\infty} \frac{v(f^k(0))}{Df^k(f(0))} = v(0) + \sum_{k=1}^{\infty} \frac{v(f^k(0))}{Df^k(f(0))}$$

be a functional defined on continuous vector fields v on the interval.

Lemma 5.4. — *If f satisfies the summability condition then there exists an even polynomial vector field v, with $v(-1) = v(1) = 0$ and such that $\nu_f(v) \neq 0$.*

Proof. — Let $S = \sum |Df^k(f(0))|^{-1}$. Let ε be so small that

$$\sum_{\substack{k>0 \\ f^k(0)\in(-\varepsilon,\varepsilon)}} \frac{1}{|Df^k(f(0))|} < 1/3.$$

Let v be an even polynomial vector field satisfying $v(-1) = v(1) = 0$,

$$\begin{aligned} |v(x)| &< 2, &&\text{for} \quad x \in I, \\ v(x) &> 1, &&\text{for} \quad x \in (-\varepsilon/2, \varepsilon/2), \\ |v(x)| &< \frac{1}{10S}, &&\text{for} \quad x \in I \setminus (-\varepsilon, \varepsilon). \end{aligned}$$

Then $\nu_f(v) > 1 - 2/3 - 1/10 > 0$. □

Lemma 5.5. — *The kernel of ν_f intersected with $T\mathcal{A}_a^{\mathbb{R}}$ is the tangent space to the hybrid class of f.*

Proof. — By the previous lemma, ν_f is non-trivial over $T\mathcal{A}_a^{\mathbb{R}}$, so the above intersection is a closed codimension-one subspace of $T\mathcal{A}_a^{\mathbb{R}}$. So it is enough to show that if v is tangent then $\nu_f(v) = 0$. Assuming that v is tangent, consider an analytic family f_t contained in the hybrid class of f, such that $f_0 = f$ and

$$\left.\frac{d}{dt} f_t\right|_{t=0} = v.$$

It is remarked in [**ALM**] that

$$\alpha_{n+1} = Df^n(f(0)) \sum_{k=0}^{n} \frac{v(f^k(0))}{Df^k(f(0))} = Df^n(f(0))\nu_f(v)$$

is precisely

$$\left.\frac{d}{dt} f_t^{n+1}(0)\right|_{t=0}.$$

Moreover, $t \mapsto f_t^{n+1}(0)$ are holomorphic functions of the complex parameter t, taking values in Ω_a, and whose domain is some definite neighborhood of 0. It follows by Cauchy estimates on the derivative that this sequence is bounded independently of n. By the summability condition (5.2), $|Df^n(f(0))| \to \infty$, so we have necessarily $\nu_f(v) = 0$. $\square$

Remark 5.1. — It is shown in [**ALM**] that the sequence α_n is not only bounded (for tangent vector fields v), but that the vector field defined on the orbit of the critical value by $w(f^k(0)) = \alpha_k$, $k > 0$, extends to a quasiconformal vector field on $\mathbb{C}$.

So Tsujii transversality can be interpreted for such a map (satisfying the summability condition (5.2)) as transversality of the family to the hybrid class of f_p.

Since for maps with negative Schwarzian derivative CE implies the BCE and that all periodic orbits are repelling, we can conclude from Theorem A, Lemma 4.4 and this discussion the following result:

Lemma 5.6. — *If F is a non-trivial k-parameter analytic family of S-unimodal maps then almost every parameter is regular or satisfies the Tsujii conditions.*

5.3. Estimates of density in perturbed families. — Let K be the space of C^2 k-parameter families of unimodal maps (without, naturally, the hypothesis of negative Schwarzian derivative).

Let $X \subset K \times B^k$ be the set of (F, p) such that either f_p is regular or satisfies the Tsujii conditions and WR. For $F \in K$, let $X_F = \{p \in B^k | (F, p) \in X\}$.

Let $Y \subset B^k$ be measurable with $|Y| > 0$. We define the density of X along F on Y as

$$d(F, Y) = \frac{|Y \cap X_F|}{|Y|}.$$

Instead of defining the classical infinitesimal density:

$$\liminf_{\varepsilon \to 0} d(F, B_\varepsilon(p))$$

we will need to consider the stability of the density with respect to perturbations of F. With this in mind we introduce two parameters. Let

$$D^-(F, p) = \liminf_{\substack{\widehat{F} \to F \\ \widehat{f}_p = f_p}} \liminf_{\varepsilon \to 0} d(\widehat{F}, B_\varepsilon(p)),$$

$$D^+(F, p) = \liminf_{\varepsilon \to 0} \liminf_{\widehat{F} \to F} d(\widehat{F}, B_\varepsilon(p)).$$

Remark 5.2. — Notice that in the definition of $D^-(F, p)$ we only consider families through a fixed map, while in the definition of $D^+(F, p)$ we do not make this restriction.

Theorem A and Tsujii's result give a direct way to estimate D^-:

Lemma 5.7. — *Let F be a non-trivial analytic family of S-unimodal maps. Then for almost every $p \in B^k$, $D^-(F,p) = 1$.*

Proof. — Indeed, by Lemma 5.6, almost every parameter is either regular or satisfies the Tsujii conditions. Since the set of regular maps is C^2 open, $D^-(F,p) = 1$ at any regular parameter p.

Let us show that this still holds for parameters p satisfying the Tsujii conditions. Since Tsujii transversality *through a fixed CE map* is clearly an open condition, if $\widehat{F}$ is any C^2 family near F with $\widehat{f}_p = f_p$ then $(\widehat{F}, p)$ also satisfies the Tsujii conditions. By Lemma 5.3,

$$\lim_{\varepsilon \to 0} d(\widehat{F}, B_\varepsilon(p)) = 1.$$

Thus $D^-(F,p) = 1$. $\square$

However, for measure estimates in perturbed families, $D^+(F,p)$ is more relevant. We proceed to discuss the effect of the interchange of limits in the definitions of $D^-(F,p)$ and $D^+(F,p)$.

Lemma 5.8. — *In this setting,*

$$D^+(F,p) \geqslant D^-(F,p).$$

Proof. — The idea is to construct, arbitrarily near F, a family $\widetilde{F}$ with $\widetilde{f}_p = f_p$ and

$$\lim_{j \to \infty} d(\widetilde{F}, B_{\varepsilon_j}(p)) = D^+(F,p),$$

for some sequence $\varepsilon_j \to 0$, which implies $D^+(F,p) \geqslant D^-(F,p)$. To construct $\widetilde{F}$, we will interpolate F with a certain sequence $F^{(n)}$ which realizes the limit in the definition of $D^+(F,p)$.

Let $\varepsilon_j \to 0$ be a sequence such that

$$\lim_{j \to \infty} \liminf_{\widehat{F} \to F} d(\widehat{F}, B_{\varepsilon_j}(p)) = D^+(F,p).$$

Passing to a subsequence, we may assume that

$$\lim_{j \to \infty} \frac{\varepsilon_{j+1}}{\varepsilon_j} = 0.$$

Let $K_j \subset B_{\varepsilon_j}(p) \setminus \overline{B_{\varepsilon_{j+1}}(p)}$ be compact sets such that

$$\lim_{j \to \infty} \frac{|\operatorname{int} K_j|}{|B_{\varepsilon_j}(p)|} = 1. \tag{5.3}$$

Let $\phi_j : \mathbb{R}^k \to \mathbb{R}$ be a C^∞ function supported in $B_{\varepsilon_j}(p) \setminus \overline{B_{\varepsilon_{j+1}}(p)}$ such that $\phi_j | K_j = 1$.

For a sequence $F^{(n)} \to F$, let us define $\widetilde{F} : B^k \times I \to I$ by

$$\widetilde{f}_q = f_q + \sum_{j=1}^{\infty} \phi_j(q)(f_q^{(j)} - f_q).$$

It is easy to see that for every $\delta > 0$ there exists a sequence $\delta_n > 0$, $n \geqslant 1$, such that, if $\|F^{(n)} - F\|_{C^2} < \delta_n$ then $\|\widetilde{F} - F\|_{C^2} < \delta$ (and in particular $\widetilde{F}$ is C^2). In other words, if $F^{(n)} \to F$ sufficiently fast then $\widetilde{F}$ is C^2 and close to F in the C^2 topology.

Notice that $\widetilde{F}$ interpolates F and the sequence $F^{(n)}$ in such a way that inside each $B_{\varepsilon_n}(p)$, $\widetilde{f}_p = f_p^{(n)}$ for p in $\operatorname{int} K_n$. Thus,

$$X_{\widetilde{F}} \cap K_n = X_{F^{(n)}} \cap K_n. \tag{5.4}$$

Fix $\delta > 0$ and select $F^{(n)}$ such that

$$\lim_{n\to\infty} d(F^{(n)}, B_{\varepsilon_n}(p)) = D^+(F^{(0)}, p) \tag{5.5}$$

and moreover $\|F^{(n)} - F\|_{C^2} < \delta_n$, so that $\|\widetilde{F} - F\|_{C^2} < \delta$. By (5.3), (5.4), and (5.5),

$$\liminf_{\varepsilon\to 0} d(\widetilde{F}, B_\varepsilon(p)) \leqslant \lim_{n\to\infty} d(\widetilde{F}, B_{\varepsilon_n}) = \lim_{n\to\infty} d(F^{(n)}, B_{\varepsilon_n}) = D^+(F, p).$$

Making $\delta \to 0$, $\widetilde{F}$ converges to F and we obtain $D^+(F, p) \geqslant D^-(F, p)$. □

5.4. Proof of Theorem B. — Let F be a non-trivial analytic family of S-unimodal maps. Then almost every parameter satisfies $D^-(F, p) = 1$. Hence, for almost every p we have $D^+(F, p) = 1$.

Fix $\varepsilon > 0$. Let $p \in B^k$ be such that $D^+(F, p) = 1$. By definition of D^+ there exists a sequence of balls $U^n(p)$ centered at p and converging to p, and neighborhoods $\mathcal{V}^n(p) \subset K$ of F such that if $\widetilde{F} \in \mathcal{V}^n(p)$ then

$$d(\widetilde{F}, U^n(p)) > 1 - \varepsilon/2.$$

By Vitali's Lemma, there exist sequences p_j, n_j such that $U^{n_j}(p_j)$ are disjoint and $|\cup U^{n_j}(p_j)| = 1$. Let m be such that $\cup_{j=1}^m U^{n_j}(p_j) > 1 - \varepsilon/2$. Let $\mathcal{V} = \cap_{j=1}^m \mathcal{V}^{n_j}(p_j)$. Then if $\widetilde{F} \in \mathcal{V}$, $d(\widetilde{F}, B^k) \geqslant 1 - \varepsilon$. If $F^{(n)} \to F$ in the C^2 topology then $F^{(n)} \in \mathcal{V}$ for n large enough and the set of parameters for $F^{(n)}$ which are either regular or satisfy the Tsujii conditions and Weak Regularity have measure at least $1 - \varepsilon$, as required.

Moreover, considering the sequence $F^{(n)} \equiv F$, we conclude that almost every parameter for F is Weakly Regular, hence the last claim of Theorem B.

5.5. Proof of Theorem C (Smooth Dichotomy). — By Proposition 4.3 non-trivial analytic families are dense among C^n families of S-unimodal maps, $n = 3, \dots, \infty$. Theorem B implies that for all ε the set D_ε of C^n families of S-unimodal maps for which the set of bad parameters (not regular or BCE, CE, SE and WR) has measure less then ε, contains a neighborhood of all non-trivial analytic families, that is, an open and dense set. Therefore $\cap D_{1/2^n}$ is a residual set. Clearly any family in $\cap D_{1/2^n}$ satisfies the stated dichotomy.

Appendix
Quasisymmetric robustness of Collet-Eckmann and polynomial recurrence

The aim of this Appendix is to sketch a proof of Theorem 4.6. This proof is similar in strategy to the one of the main results of [**AM1**], however non-trivial modifications are needed. To avoid too much intersection, this will be a concise exposition concentrated mainly on the new steps needed for this improvement: the reader can find a full proof of this result in [**AM3**].

A.1. Quasisymmetric maps

A.1.1. Quasisymmetric reparametrization. — Let now H be an arbitrary but fixed $\widehat{\gamma}$-quasisymmetric map from B^1 to the parameter space of the quadratic family. To prove Theorem 4.6, it will be enough to show that almost every $t \in B^1$ correspond under H to a parameter of the quadratic family which is either regular or satisfies the Collet-Eckmann and Polynomial Recurrence conditions.

From now on, all mentions to parameter space will (unless explicitly stated otherwise) refer to the above reparametrization.

A.1.2. Quasisymmetric capacities. — The γ-capacity of a set $X \subset \mathbb{R}$ in an interval I is defined as follows:

$$p_\gamma(X|I) = \sup \frac{|h(X \cap I)|}{|h(I)|}$$

where the supremum is taken over all γ-qs maps $h : \mathbb{R} \to \mathbb{R}$.

Notice that if I^j are disjoint subintervals of I and $X \subset \cup I^j$ then

$$p_\gamma(X|I) \leqslant p_\gamma(\cup_j I^j|I) \sup_j p_\gamma(X|I^j).$$

A.2. Sequence of first return maps. — The statistical analysis of [**AM1**] concerns mainly the following objects: we are given a unimodal map (which we will assume finitely renormalizable and with a recurrent critical point) $f : I \to I$ and a sequence of nested intervals $I_n \subset I$. The inductive relation between the I_n is as follows: the domain of the first return map R_n to I_n consists of countably many intervals $\{I_n^j\}_{j \in \mathbb{Z}}$, with the convention that $0 \in I_n^0$ (the central component), and we let $I_n^0 = I_{n+1}$.

The special sequence of intervals I_n that we consider is called the principal nest, see [**L2**]. Since we assume f to be finitely renormalizable, there exists a smallest symmetric interval $T \subset I$ which is periodic (say, of period m). For the principal nest, $I_1 = [-p, p]$, where p is the orientation reversing fixed point of $f^m : T \to T$. A level n of the principal nest is called central if $R_n(0) \in I_{n+1}$. Let us say that f is a *simple map* if its principal nest has at most finitely many central levels.

Each non-central branch of R_n is a diffeomorphism onto I_n. Let us introduce some convenient notation related to the iteration of the non-central branches of R_n. Let Ω

be the set of finite sequences of non-zero integers (the empty sequence is included), an element of Ω is denoted $\underline{d} = (j_1, \dots, j_m)$. If $\underline{d} \in \Omega$ has length $|\underline{d}| = m$, we denote $R_n^{\underline{d}}$ the branch of $R_n^{|\underline{d}|}$ with combinatorics $\underline{d}$, that is, the domain of $R_n^{\underline{d}}$ is the set

$$I_n^{\underline{d}} = \{x \in I | R_n^{k-1}(x) \in I_n^{j_k}, 1 \leqslant k \leqslant m\}.$$

We let $C_n^{\underline{d}} = (R_n^{\underline{d}})^{-1}(I_{n+1})$.

Let us denote by L_n the first landing map from I_n to I_{n+1}. This map relates easily to R_n using the above description: the domain of L_n is $\cup C_n^{\underline{d}}$, and $L_n | C_n^{\underline{d}} = R_n^{\underline{d}}$. The reader should think of L_n as a high iterate of R_n. This leads to the following inductive relation between return maps: $R_{n+1} = L_n \circ R_n | I_{n+1}$.

The return time of a point x belonging to an interval I_n^j is denoted by $r_n(x)$ (or $r_n(j)$, since it does not depend on $x \in I_n^j$), that is, $R_n | I_n^j = f^{r_n(j)}$. The landing time is denoted by $l_n(x) \equiv l_n(j)$. The combinatorics at level n of a point x is denoted $\underline{d}^{(n)}(x)$, so that $x \in C_n^{\underline{d}^{(n)}(x)}$. Let $j^{(n)}(x)$ be such that $x \in I_n^{j^{(n)}(x)}$. We let $\tau_n = j^{(n)}(R_n(0))$, so that $R_n(0) \in I_n^{\tau_n}$. The return time of the critical point is denoted $v_n = r_n(0)$. Let $s_n = |\underline{d}^{(n)}(R_n(0))|$.

Notice that $I_{n+1} = R_{n-1}^{-1}(C_{n-1}^{\underline{d}})$ for some $\underline{d}$. The interval $\widetilde{I}_{n+1} = R_{n-1}^{-1}(I_{n-1}^{\underline{d}}) \subset I_n$ is a big neighborhood of I_{n+1} which will be useful later. This choice of neighborhood is particularly good for simple maps, and it turns out that in this case $\widetilde{I}_{n+1}$ is still much smaller than I_n for big n.

A.2.1. Phase-parameter relation. — The starting point of [**AM1**] are two theorems of Lyubich describing the (unreparametrized) parameter space of the quadratic family: infinitely renormalizable maps have zero Lebesgue measure [**L4**] and almost every finitely renormalizable non-regular map is simple [**L3**]. We will need the following remark of [**ALM**]: Lyubich's proof actually allows one to conclude that the set of regular or simple maps has full measure after any quasisymmetric reparametrization.

In view of those results, Theorem 4.6 is reduced to proving that the set of parameters which are Collet-Eckmann and polynomially recurrent have full measure (*after reparametrization by* H) among simple maps. From now on we exclude non-simple maps from measure-theoretic considerations, and we will use "with total probability" to refer to a set of parameters with full measure (*after reparametrization by* H) among simple maps.

To estimate the probability in the parameter corresponding to a certain behavior of the n-th stage of the principal nest, we make use of the Phase-Parameter Lemmas of [**AM1**]. They describe how the partition of the phase space induced by return and landing maps R_n and L_n induce parameter partitions of certain parameter windows J_n.

The topological part of the phase-parameter relation is described in the following:

Theorem A.1. — *For each non-renormalizable quadratic map f with a recurrent critical point, there exists a sequence of parameter intervals $\{J_n\}$ such that:*

(1) *J_n is the maximal interval containing f such that for all $g \in J_n$, there exists a continuation $I_{n+1}[g]$ of I_{n+1} with the "same combinatorics" in the following sense. There exists a continuous family of homeomorphisms $h_n[g] : I \to I$, $g \in J_n$ which is equivariant with respect to the actions of $g|(I \setminus I_{n+1}[g])$ and $f|(I \setminus I_{n+1})$, so that if $x \in I \setminus I_{n+1}[f]$ then $g \circ h_n[g](x) = h_n[g] \circ f(x)$.*

(2) *There exists a homeomorphism $\Xi_n : I_n \to J_n$ such that $\Xi_n(C_n^{\underline{d}})$ is the set of all $g \in J_n$ such that $R_n[g](0) \in h_n[g](C_n^{\underline{d}})$.*

This result follows immediately from the Topological Phase-Parameter relation for the unreparametrized quadratic family (Theorem 2.2 of [**AM1**]), since the reparametrization is a homeomorphism.

In words, the sequence J_n in Theorem A.1 denotes the maximal interval containing f where we can consider a continuation of I_n (recall that the boundary of I_n is preperiodic), and such that the first return map to this continuation does not change combinatorics, so that its domain changes continuously. When we change the map g inside the interval J_n, the critical value of $R_n[g]$ varies inside the interval $I_n[g]$ "properly", that is, moves from one boundary point to the other. In doing so, it goes through the partition induced by the $C_n^{\underline{d}}$ in a well behaved ("monotonic") way: it goes through each member of the partition exactly once, and thus defines a partition in the parameter interval J_n, corresponding topologically to the partition in the phase interval I_n. Theorem A.1 thus establishes that the "diagonal" motion of the critical value and the "horizontal" motion of the partition of the phase space are "transversal". This is indeed how the proof of Lyubich goes (using complex analysis). This result can also be established using the Milnor-Thurston's combinatorial theory of unimodal maps together with the monotonicity property of the quadratic family.

The next component of the phase-parameter relation is a quantitative estimate on the regularity of the phase-parameter homeomorphisms Ξ_n. While the topological part is based on a very general transversality argument, the quantitative part depends on the delicate geometric estimates of Lyubich.

We let $J_n^{\tau_n} = \Xi_n(I_n^{\tau_n})$. The correspondence Ξ_n is uniquely defined if restricted to $K_n = I_n \setminus \cup C_n^{\underline{d}}$. More importantly, it is quasisymmetric if restricted to certain subsets of K_n. To make this precise, let $K_n^\tau = K_n \cap I_n^{\tau_n}$ (forgetting information outside $I_n^{\tau_n}$) and $\widetilde{K}_n = I_n \setminus (\cup I_n^j \cup \widetilde{I}_{n+1})$ (forgetting information inside each I_n^j and also inside $\widetilde{I}_{n+1}$).

Theorem A.2. — *Let f be a simple map. Then, for all $\gamma = (1+\delta)\widehat{\gamma} > \widehat{\gamma}$, there exist $n_0 > 0$ such that for all $n > n_0$,*

PhPa1: *$\Xi_n|K_n^\tau$ is γ-qs;*
PhPa2: *$\Xi_n|\widetilde{K}_n$ is γ-qs;*
PhPh1: *$h_n[g]|K_n$ is $1+\delta$-qs for all $g \in J_n^{\tau_n}$;*
PhPh2: *$h_n[g]|\widetilde{K}_n$ is $1+\delta$-qs for all $g \in J_n$.*

This theorem is a straightforward consequence of the Phase-Parameter relation for the unreparametrized quadratic family (Theorem 2.3 of [**AM1**]). While in [**AM1**] the quasisymmetric constants in PhPa1 and PhPa2 could be taken arbitrarily close to 1 (the unreparametrized case corresponds to taking $H = \mathrm{id}$, that is, $\widehat{\gamma} = 1$) for deeper levels of the principal nest, this does not hold here due to introduction of reparametrization, which multiply all phase-parameter constants by $\widehat{\gamma}$ (notice that PhPh1 and PhPh2 are estimates which do not depend on reparametrization, so we can still choose constants close to 1). This will be the source of many difficulties addressed in this Appendix.

A.3. The statistical argument. — For the remaining of this Appendix we fix some constant $\gamma > \widehat{\gamma}$, and we will start our consideration with levels of the principal nest where the reparametrized phase-parameter relation is already γ-qs. We will also need some very large constants $\widetilde{b} < b$ which depend only on γ (the relation can be computed explicitly following the proof, in particular, $\widetilde{b}$ should be at least so big that $\widetilde{b}^{-1}$ is a lower bound on the Hölder constant of γ-qs maps). We let $a = b^{-1}$ and $\widetilde{a} = \widetilde{b}^{-1}$.

From now on we will always estimate the γ-capacity of bad sets in the phase space. To conclude results for the parameter we will use the following variation of the Borel-Cantelli Lemma (this is Lemma 3.1 of [**AM1**]).

Lemma A.3. — *Let $X \subset \mathbb{R}$ be a measurable set such that for each $x \in X$ there is a sequence $D_n(x)$ of nested intervals converging to x such that for all $x_1, x_2 \in X$ and any n, $D_n(x_1)$ is either equal or disjoint to $D_n(x_2)$. Let Q_n be measurable subsets of $\mathbb{R}$ and $q_n(x) = |Q_n \cap D_n(x)|/|D_n(x)|$. Let Y be the set of x in X which belong to finitely many Q_n. If $\sum q_n(x)$ is finite for almost any $x \in X$ then $|Y| = |X|$.*

In practice, the D_n will be the parameter windows defined before (either J_n or $J_n^{\tau_n}$), and Q_n will be certain subsets of J_n or $J_n^{\tau_n}$ corresponding (under the phase-parameter map) to branches of the return map (in the case of J_n) or landings (in the case of $J_n^{\tau_n}$), whose behavior we want to avoid. We will then show that such bad events have summable γ-capacity in the phase space, which will yield the conclusion for Lebesgue measure of the parameter using PhPa1 (for landings) or PhPa2 (for returns).

A.3.1. A simple application: torrential decay of geometry. — We will now illustrate the use of Lemma A.3 and the phase-parameter relation with an estimate on the decay of geometry. More precisely, we will consider the *scaling factor*

$$c_n = \frac{|I_{n+1}|}{|I_n|}.$$

The scaling factor is a particularly important parameter in the subsequent analysis: all statistical estimates that follow will be related to c_n.

One initial information on the scaling factors is provided by the following result of Lyubich:

***Theorem A.4* (see [L1])**. — *If f is simple than there exists $C > 0$, $\lambda < 1$ such that $c_n < C\lambda^n$.*

We will now show that, with total probability, the decay of c_n is much faster than exponential. To express this decay, let us consider the tower function defined by recursion $T(1) = 2$, $T(n+1) = 2^{T(n)}$. We will show that, with total probability, the c_n decrease torrentially to 0, that is, there exists $k > 0$ such that $c_n^{-1} > T(n-k)$ for n big enough. More precisely, we will show that c_{n+1}^{-1} behaves as an exponential of (a bounded power of) c_n^{-1}.

This very fast decay implies that the landing map to I_{n+1} is essentially a very high iterate of the return map to I_n (since it takes a long time to hit a very small interval). This very high iteration time will allow us to conclude that the characteristics (say, return time) of each level tend to be better behaved than in the previous one due to fast convergence to some average (some kind of Law of Large Numbers). The fact that we must deal with qs-capacity instead of Lebesgue measure will essentially reflect in the presence of errors terms (whose size depend on $\widehat{\gamma}$) in certain exponents in the above description.

In order to estimate c_n, we first consider the related quantity $s_n = |\underline{d}^{(n)}(R_n(0))|$, which denotes the number of times the critical orbit visits I_n before hitting I_{n+1}.

If the critical orbit behaved as a sequence of random points (uniformly distributed with respect to Lebesgue), the expectation of this first hitting time should be c_n^{-1}. More relevant for us, the distribution of the first hitting time (for the random model) should be concentrated about c_n^{-1}: with large probability (say, less than 2^{-n}), the first hitting time is in some "neighborhood" of c_n^{-1} (say, $[4^{-n}c_n^{-1}, 4^n c_n^{-1}]$). The corresponding statement for our actual dynamical system is that the distribution of $|\underline{d}^{(n)}(x)|$, with respect to Lebesgue measure on $x \in I_n$ is concentrated around c_n^{-1}, which can be easily checked by the reader: the estimates are not significantly affected in the non-random case.

However, due to the nature of the phase-parameter relation, we must estimate the distribution of $|\underline{d}^{(n)}(x)|$ in terms of capacities. This will affect drastically the estimates. To understand why, keep in mind that γ-qs maps are only Hölder (with some constant bounded from below by $\widetilde{b}^{-1}$), so they can potentially distort the logarithm of the ratio between I_{n+1} and I_n by such a constant. Aside from this problem, the information we need can be computed quite easily and is summarized below.

Lemma A.5. — *With total probability, for all n sufficiently big we have*

$$p_{2\gamma}(|\underline{d}^{(n)}(x)| \leqslant k|I_n) < kc_n^{\widetilde{a}}, \tag{1}$$

$$p_{2\gamma}(|\underline{d}^{(n)}(x)| \geqslant k|I_n) < e^{-kc_n^{\widetilde{b}}}. \tag{2}$$

We also have

$$p_{2\gamma}(|\underline{d}^{(n)}(x)| \leqslant k|I_n^{\tau_n}) < kc_n^{\tilde{a}}, \tag{3}$$

$$p_{2\gamma}(|\underline{d}^{(n)}(x)| \geqslant k|I_n^{\tau_n}) < e^{-kc_n^{\tilde{b}}}. \tag{4}$$

This lemma corresponds to Lemma 4.2 of [**AM1**].

The phase-parameter lemmas (specially PhPa1) allow us to transfer the last pair of estimates to the parameter space: for n sufficiently big, (Lebesgue) most parameters in $J_n^{\tau_n}$ satisfy

$$c_n^{-\tilde{a}/2} < s_n < c_n^{-2\tilde{b}}.$$

Here 'most' means that the complement has probability bounded by $c_n^{\tilde{a}/3}$. But c_n (and thus $c_n^{\tilde{a}/3}$) decays exponentially for every simple map (by Theorem A.4). So $\sum c_n^{\tilde{a}/3} < \infty$ and we are able to apply Lemma A.3 to obtain the following:

Lemma A.6. — *With total probability, for n sufficiently big we have*

$$c_n^{-\tilde{a}/2} < s_n < c_n^{-2\tilde{b}}.$$

This lemma corresponds to Lemma 4.3 of [**AM1**].

Remark A.1. — This result implies easily torrential decay of c_n: $\ln c_{n+1}^{-1}$ can be easily bounded from below by Ks_n for some universal $K > 0$, and thus for big n,

$$c_{n+1}^{-1} \geqslant e^{c_n^{-\tilde{a}/3}}.$$

A.4. Derivatives. — We proceed to estimate derivatives of branches of the return map. All lemmas in this section can be proved using the same argument as in [**AM1**].

The first step is to exclude the possibility of a 'too recurrent' or 'too low' return. It is analogous to Lemma 4.8 of [**AM1**], being a simple application of PhPa2.

Lemma A.7. — *With total probability, the distance between $R_n(0)$ and $\partial I_n \cup \{0\}$ is at least $|I_n|n^{-\tilde{b}}$. In particular $R_n(0) \notin \tilde{I}_{n+1}$ for all n large enough.*

Recall that the *distortion* of a diffeomorphism ϕ on an interval T is defined by

$$\operatorname{Dist}(\phi|T) = \frac{\sup_T |D\phi|}{\inf_T |D\phi|}.$$

Lemma A.7 allows us to start estimating the distortion of iterates of f. The following estimate corresponds to Lemma 4.9 of [**AM1**]. It is based on the fact that the distortion of branches of return maps is due to the position of the branch with respect to the critical point. Using PhPa1, we are able to give polynomial lower bounds on the distance between the critical point with respect to non-central branches, which are valid with total probability.

Lemma A.8. — *With total probability, for n big enough and $j \neq 0$*

$$\operatorname{Dist}(f|I_n^j) \leqslant n^{\tilde{b}}.$$

The following estimate is analogous to Lemma 4.10 of [**AM1**]. It is based on the previous one and the observation that return branches are torrentially expansive in average (from the decay of geometry).

Lemma A.9. — *With total probability, for n big enough and for all $\underline{d} \in \Omega$*

$$\mathrm{Dist}(R_n^{\underline{d}}) \leqslant n^{\tilde{b}}.$$

In particular, for n big enough, $|DR_n(x)| > 2$ if $x \in \cup_{j \neq 0} I_n^j$.

Lemma A.9 gives estimates of derivatives under iterates of R_n. To obtain estimates of derivatives under iterates of f, we will need the following very general result of Guckenheimer which shows that quadratic maps are hyperbolic away from critical points and parabolic points (this actually generalizes to very general one-dimensional systems by a result of Mañé), see [**MvS**]. We state just a consequence adapted to our particular setting.

Theorem A.10. — *Let f be a quadratic map without non-repelling periodic orbits (in particular if f is a simple map). For every $\varepsilon > 0$, there exists $C > 0$, $\lambda > 1$ such that if $|f^k(x)| > \varepsilon$ for $0 \leqslant k \leqslant m$ then $Df^{m+1}(x) > C\lambda^m$.*

With this information we are now able to give a lower bound on the derivative of iterates of f. The next lemma is identical to Lemma 4.11 of [**AM1**], and is based on the idea that full returns to sufficiently deep levels cause expansion (from the previous lemma), while the dynamics outside a definite neighborhood of the critical point is hyperbolic (by Theorem A.10).

Lemma A.11. — *With total probability, if n is sufficiently big and if $x \in I_n^j$, $j \neq 0$, and $R_n|I_n^j = f^r$, then for $1 \leqslant k \leqslant r$, $|Df^k(x)| > |x| c_{n-1}^3$.*

A.5. How to deal with hyperbolicity. — Keeping in mind that our analysis of the statistical properties of the dynamics of f is made in terms of the induced return maps R_n, we see that in order to estimate the hyperbolicity along the critical orbit (to obtain the Collet-Eckmann condition) we must have a convenient way to quantify the hyperbolicity of (for instance) non-central return branches. To do so, for $j \neq 0$, we define the quantity

$$\lambda_n(j) = \inf_{x \in I_n^j} \frac{\ln |DR_n(x)|}{r_n(j)}.$$

We let $\lambda_n = \inf_{j \neq 0} \lambda_n(j)$.

To analyze the behavior of λ_n, we start with the general information provided by Theorem A.10. Coupled with exponential upper bounds on distortion for returns (which competes with torrential expansion of each non-central branch from the decay of c_n), the hyperbolicity of f in the complement of I_{n+1} immediately implies the following estimate (identical to Lemma 7.9 of [**AM1**]).

Lemma A.12. — *With total probability, for all n sufficiently big, $\lambda_n > 0$.*

The "minimum hyperbolicity" $\liminf \lambda_n$ of the parameters we will obtain will in fact be positive, as it follows from one of the properties of Collet-Eckmann parameters (uniform hyperbolicity on periodic orbits), together with our estimates on distortion.

Our strategy however is not to show that the minimum hyperbolicity is positive, but that the typical value of $\lambda_n(j)$ stays big as n grows (and is in fact bigger than $\lambda_{n_0}/2$ for $n > n_0$ big). In this sense, it is convenient to think of $\lambda_n(j)$ as a random variable whose distribution we are interested in.

There is an inductive relation between the random variables $\lambda_n(j)$ for different values of n: this is related to the fact that if $R_n(I^j_{n+1}) \subset C^{\underline{d}}_n$, $\underline{d} = (j_1, \ldots, j_m)$, we have $R_{n+1}|I^j_{n+1} = L_n|C^{\underline{d}}_n \circ R_n|I^j_{n+1}$. The hyperbolicity of the "landing part" $L_n|C^{\underline{d}}_n$ is essentially a weighted sum

$$\frac{\sum_{i=1}^m \lambda_n(j_i) r_n(j_i)}{\sum_{i=1}^m r_n(j_i)}. \tag{A.1}$$

So if the "return part" $R_n|I^j_{n+1}$ does not carry a big weight on the computation of $\lambda_{n+1}(j)$ (outside a set of branches with small γ-qs capacity), we can think of $\lambda_{n+1}(j)$ as distributed according to the weighted sum (A.1). This turns out to be the case as the return part does not affect much the denominator (time) and does not have a bad effect on the numerator (derivative). Indeed, in the next section we will see that the return time of $R_n|I^j_{n+1}$ (given by v_n) is much smaller (of order c_{n-1}^{-1}) than the total return time $R_{n+1}|I^j_{n+1}$ (of order c_n^{-1}). Moreover, if I^j_{n+1} is outside a small neighborhood of 0, $|DR_n|I^j_{n+1}|$ is bigger than 1.

Since we also have to estimate the hyperbolicity of truncated branches (as the Collet-Eckmann condition is a condition along the full critical orbit, and not only at full returns), it will not be enough to just obtain that the distribution of $\lambda_n(j)$ is concentrated around some value bigger than $\lambda_{n_0}/2$. In order to state exactly what kind of hyperbolicity estimate we need, it is convenient to introduce a certain class of branches: good returns.

We define the set of good returns $G(n_0, n) \subset \mathbb{Z} \setminus \{0\}$, $n_0, n \in \mathbb{N}$, $n \geqslant n_0$ as the set of all j such that

G1: (hyperbolic return)

$$\lambda_n(j) \geqslant \lambda_{n_0} \frac{1 + 2^{n_0 - n}}{2},$$

G2: (hyperbolicity in truncated return) for $c_{n-1}^{-3/(n-1)} \leqslant k \leqslant r_n(j)$ we have

$$\inf_{I^j_n} \frac{\ln |Df^k|}{k} \geqslant \lambda_{n_0} \frac{1 + 2^{n_0 - n + 1/2}}{2} - c_{n-1}^{2/(n-1)}.$$

Of course we still have to show that the set of returns which fail to be good has small γ-qs capacity. In order to do so, we will construct explicitly a class of branches whose complement has small γ-qs capacity and then show that this class of branches is contained in good branches (see Lemma A.20). Before doing so, we must first estimate the distribution of return times, since they have an important role in the computation of $\lambda_n(j)$.

A.6. Distribution of return and landing times. — To estimate the distribution of return and landing times, it is convenient to also think of $r_n(j)$ and $l_n(j)$ as "random variables" which are related by some simple rules: if $\underline{d} = (j_1, \ldots, j_m)$ then $l_n(\underline{d}) = \sum_{i=1}^m r_n(j_i)$ and $r_{n+1}(j) = v_n + l_n(\underline{d})$ where $R_n(I^j_{n+1}) \subset C^{\underline{d}}_n$. In particular, since the distribution of $|\underline{d}^{(n)}|$ is concentrated around c_n^{-1} which is torrentially big, the random variable l_n behaves like a very large sum of random variables distributed as r_n. On the other hand, r_{n+1} should have distribution approximately like l_n itself, once we show that v_n does not make an important contribution.

The main tool to do the actual analysis is to prove first a Large Deviation Estimate for r_n using only the torrential decay of c_n, and then show that such estimate leads to much more precise control of the subsequent levels.

Since the transition between different levels introduces some distortion (although torrentially small), we are forced to deal with a sequence of quasisymmetric constants in our estimates: instead of just estimating γ-qs capacities for some fixed γ, we must consider a sequence $\gamma_n = \gamma(n+1)/n$ and $\widetilde{\gamma}_n = \gamma(2n+3)/(2n+1)$. The basic idea is that control of the distribution of r_n with respect to γ_n-capacities will provide control of the distribution of l_n with respect to $\widetilde{\gamma}_n$ capacities which in turn will allow to estimate the distribution of r_{n+1} with respect to γ_{n+1} capacities. Notice that $\inf \gamma_n = \inf \widetilde{\gamma}_n = \gamma$. (This ideas are introduced in § 5 of [**AM1**].)

Although very technical, this part is very similar to the analysis made on (the several lemmas of) § 6 of [**AM1**] (differing only by change of constants), so we will only state the final estimate which summarizes the results of that section and provide a short outline of the argument.

Lemma A.13. — *With total probability, for all n sufficiently large we have*

(1) $p_{\widetilde{\gamma}_n}(l_n(x) < c_n^{-s} | I_n) < c_n^{\tilde{a}^2 - s} < c_n^{a-s}$, *with* $s > 0$,
(2) $p_{\widetilde{\gamma}_n}(l_n(x) < c_n^{-s} | I_n^{\tau_n}) < c_n^{a-s}$, *with* $s > 0$,
(3) $p_{\widetilde{\gamma}_n}(l_n(x) > c_n^{-s} | I_n) < e^{-c_n^{b-s}}$, *with* $s > b$,
(4) $p_{\widetilde{\gamma}_n}(l_n(x) > c_n^{-s} | I_n^{\tau_n}) < e^{-c_n^{b-s}}$, *with* $s > b$,
(5) $p_{\gamma_n}(r_n(x) < c_{n-1}^{-s} | I_n) < c_{n-1}^{\tilde{a}^2 - s} < c_{n-1}^{a-s}$, *with* $s > 0$,
(6) $p_{\gamma_n}(r_n(x) > c_{n-1}^{-s} | I_n) < e^{-c_{n-1}^{\sqrt{b}-s}} < e^{-c_{n-1}^{b-s}}$ *with* $s > b$.
(7) $c_{n-1}^{-a} < r_n(\tau_n) < c_{n-1}^{-b}$.
(8) $c_{n-1}^{-a} < v_n < c_{n-1}^{-b}$.
(9) $c_{n-1}^{-a} < \ln(c_n^{-1}) < c_{n-1}^{-b}$.

A.6.1. Outline of the proof of Lemma A.13. — The estimates from below are relatively easy. Estimates (1) and (2) follow directly from $l_n(\underline{d}) \geqslant |\underline{d}|$ and Lemma A.5. Estimate (5) follows from (1) using the relation between r_{n+1} and l_n. The estimate from below in (8) follows from (2) and PhPa1, and the estimate from below in (7) follows from (5) and PhPa2. The estimate from below on (9) was computed on Remark A.1.

The estimates from above are much more delicate. In what follows we will ignore the difference between I_n and $I_n^{\tau_n}$, since it is not substantial for the argument. The key estimate is (6), which says that the tail $p_{\gamma_n}(r_n(x) > k)$ decays exponentially fast (in k) with some specific rate (polynomial in c_{n-1}). On the other hand, decay with *some* rate is easy: f is hyperbolic outside I_{n+1} (see Theorem A.10), so there exists some (small) $\alpha_n > 0$ with $p_{\gamma_n}(r_n(x) > k\alpha_n^{-1}) < e^{-k}$ for $k \geqslant 1$. This exponential decay implies that it is very unlikely that a large sequence $\underline{d} = (j_1, \dots, j_m)$ will have a landing time $l_n(\underline{d}) = \sum_{i=1}^m r_n(j_i)$ much bigger than $m\alpha_n^{-1}$.

From this relation between r_n and l_n, we see that there exists some β_n with $p_{\widetilde{\gamma}_n}(l_n(x) > k\beta_n^{-1}) < e^{-k}$, and moreover we can estimate β_n in terms of α_n and the size of a typical $\underline{d}^{(n)}$ (which is given by a polynomial on c_n^{-1}): β_n^{-1} is bounded by a polynomial (this polynomial error is related to γ) on $\alpha_n^{-1}c_n^{-1}$. From the relation between l_n and r_{n+1} we obtain an estimate on α_{n+1} in terms of v_n and β_n, which we can rewrite in terms of v_n, c_n and α_n: $\alpha_{n+1}^{-1} - v_n$ is bounded by some polynomial on $\alpha_n^{-1}c_n^{-1}$.

Since $p_{\widetilde{\gamma}_n}(l_n(x) > \beta_n^{-1}c_n^{-1})$ is summable (by definition of β_n), it follows that $v_{n+1} - v_n$ is bounded by a polynomial on $\alpha_n^{-1}c_n^{-1}$ with total probability (use PhPa1), in particular, for n big we can bound v_{n+1} with a polynomial on $\alpha_n^{-1}c_n^{-1}$.

In particular, if $\alpha_n^{-1} > c_n^{-1}$, α_{n+1}^{-1} is bounded by a *polynomial* in α_n^{-1}. Although initially we did not have any control on the value of α_n, we know that c_{n+1}^{-1} behaves as an *exponential* on c_n^{-1} (torrential growth), so eventually it catches up with α_n^{-1}: for n big, $c_n^{-1} > \alpha_n^{-1}$.

So for n big α_n^{-1} can be bounded exclusively by a polynomial on c_{n-1}^{-1} as stated in (6). This automatically implies the estimate from above in (7) using PhPa2. Since β_n^{-1} and v_{n+1} are bounded by a polynomial on $\alpha_n^{-1}c_n^{-1}$ we obtain (3) and (4) and the estimate from above in (8).

Since f^{v_n} expands I_{n+1} to an interval of size at least $2^{-n}|I_n|$, and the derivative of f is bounded by 4, we have $2^n c_n^{-1} < 4^{v_n}$, so the estimate from above on (9) follows from the estimate from above in (8).

A.7. Constructing hyperbolic branches. — In this section we show by an inductive process that the great majority of branches are reasonably hyperbolic (good branches). In order to do that, in the following subsection, we define some classes of branches with 'very good' distribution of times and which are not too close to the critical point. The definition of 'very good' distribution of times has an inductive component: they are composition of many 'very good' branches of the previous level. The fact that most branches are 'very good' is related to the validity of some kind of Law of Large Numbers estimate. The inductive definition will guarantee that the 'very good' distribution of times holds in all scales and allows us to preserve hyperbolicity from one step to the other: very good branches are good.

Remark A.2. — The several classes of branches that we will define do not correspond exactly to the same classes in [**AM1**], although classes with the same name have essentially the same function in the proof. There are some non-trivial steps to make this adaptation work, since the previous proof uses strongly small quasisymmetric constants. This will lead to consideration of extra classes below (bad returns and fast landings).

Remark A.3. — This section contains the main modifications with respect to [**AM1**] (precisely the introduction of bad returns and fast landings). The role of those modifications is explained in Remark A.4.

A.7.1. Standard landings. — Let us define the set of standard landings at time n, $LS(n) \subset \Omega$ as the set of all $\underline{d} = (j_1, \dots, j_m)$ satisfying the following:

LS1: (m is not too small or large) $c_n^{-a/2} < m < c_n^{-2b}$,
LS2: (No very large times) $r_n(j_i) < c_{n-1}^{-3b}$ for all i.
LS3: (Short times are sparse in large enough initial segments) For $c_{n-1}^{-2b} \leqslant k \leqslant m$

$$\#\{1 \leqslant i \leqslant k,\, r_n(j_i) < c_{n-1}^{-a/2}\} < (6 \cdot 2^n) c_{n-1}^{a/2} k.$$

We also define the set of fast landings at time n, $LF(n) \subset \Omega$ by the following conditions

LF1: (m is small) $m < c_n^{-a/2}$.
LS2: (No very large times) $r_n(j_i) < c_{n-1}^{-3b}$ for all i.

It is easy to convince oneself that most landings are standard. Indeed, the distribution of $|\underline{d}^{(n)}(x)|$ is concentrated around c_n^{-1} as requested by LS1. Moreover, branches with very large times (larger than c_{n-1}^{-3b}) are so few that even a long sequence $(j_1, \dots, j_m)$ with $m < c_{n-1}^{-2b}$ is not likely to contain such an event, as required by LS2. Finally, the Law of Large Numbers indicates that a long sequence $(j_1, \dots, j_m)$ will seldom contain a proportion of short times much bigger than their frequency as given by Lemma A.13, as required by LS3.

Since fast landings are not standard, they must be few. However, they correspond to most of the branches which are not standard. The reason for this comes from the requirements of LS1, which imposes two conditions (an upper and a lower bound on m). The upper bound condition is much more rarely violated (by one exponential order of magnitude) than the lower bound (just check Lemma A.5). Fast landings essentially capture the violations of the lower bound (LF1).

The actual estimates for the frequency of standard and fast landings are provided below. They can be obtained from the estimates of distribution of return times (contained in Lemma A.13) following the general lines of Lemma 7.1 of [**AM1**]. This step is purely dynamical (no further parameter exclusion is made).

Lemma A.14. — *With total probability, for all n sufficiently big,*

(1) $p_{\tilde{\gamma}_n}(\underline{d}^{(n)}(x) \notin LS(n)|I_n) < c_n^{a/3}/2$,

(2) $p_{\tilde{\gamma}_n}(\underline{d}^{(n)}(x) \notin LS(n) \cup LF(n)|I_n) < c_n^{n^2}/2$,

(3) $p_{\tilde{\gamma}_n}(\underline{d}^{(n)}(x) \notin LS(n)|I_n^{\tau_n}) < c_n^{a/3}/2$,

(4) $p_{\tilde{\gamma}_n}(\underline{d}^{(n)}(x) \notin LS(n) \cup LF(n)|I_n^{\tau_n}) < c_n^{n^2}/2$.

A.7.2. Very good returns, bad returns and excellent landings. — Define the set of very good returns, $VG(n_0, n) \subset \mathbb{Z} \setminus \{0\}$, $n_0 \leqslant n \in \mathbb{N}$ and the set of bad returns, $B(n_0, n) \subset \mathbb{Z} \setminus \{0\}$, $n_0 \leqslant n \in \mathbb{N}$, by induction as follows. We let $VG(n_0, n_0) = \mathbb{Z} \setminus \{0\}$, $B(n_0, n_0) = \varnothing$ and supposing $VG(n_0, n)$ and $B(n_0, n)$ defined, define the set of excellent landings $LE(n_0, n) \subset LS(n)$ satisfying the following extra assumptions.

LE1: (Not very good moments are sparse in large enough initial segments) For all $c_{n-1}^{-2b} < k \leqslant m$

$$\#\{1 \leqslant i \leqslant k,\ j_i \notin VG(n_0, n)\} < (6 \cdot 2^n) c_{n-1}^{a^2} k,$$

LE2: (Bad moments are sparse in large enough initial segments) For all $c_n^{-1/n} < k \leqslant m$

$$\#\{1 \leqslant i \leqslant k,\ j_i \notin B(n_0, n)\} < (6 \cdot 2^n) c_{n-1}^{n} k,$$

We define $VG(n_0, n+1)$ as the set of j such that $R_n(I_{n+1}^j) = C_n^{\underline{d}}$ with $\underline{d} \in LE(n_0, n)$ and the extra condition:

VG: (distant from 0) The distance of I_{n+1}^j to 0 is bigger than $c_n^{n^2}|I_{n+1}|$.

And we define $B(n_0, n+1)$ as the set of $j \notin VG(n_0, n+1)$ such that $R_n(I_{n+1}^j) = C_n^{\underline{d}}$ with $\underline{d} \notin LF(n)$.

Very good returns are designed to carry hyperbolicity from level to level: since they are composed of many very good returns of the previous level (LE1), and are not too close to 0 (VG), they should keep most of the hyperbolicity of level n_0 (given by $\lambda_{n_0} > 0$). For this to work, we must control the distribution of return times of the previous level inside a very good branch. The risky situation is the presence of not very good branches which have a large return time: those are contained in the bad branches defined above. It turns out that they can not spoil the hyperbolicity because they are too few (LE2). This basic idea will be carried out in detail through a series of lemmas.

Very good and bad returns can be estimated in an inductive fashion analogously to the estimate of Lemmas 7.2 and 7.3 of [**AM1**]: initially all branches are very good and no branches are bad, and as n grows the Law of Large Numbers indicates that conditions LE1 and LE2 should be rarely violated so that very good branches should continue to be frequent and bad branches rare. This estimate is again purely dynamical.

Lemma A.15. — *With total probability, for all n_0 sufficiently big,*

(1) $p_{\gamma_n}(j^{(n)}(x) \notin VG(n_0, n)|I_n) < c_{n-1}^{a^2}$,

(2) $p_{\gamma_n}(j^{(n)}(x) \in B(n_0, n)|I_n) < 2c_{n-1}^{2n}$,
(3) $p_{\tilde{\gamma}_n}(\underline{d}^{(n)}(x) \notin LE(n_0, n)|I_n) < c_n^{2a/5}$,
(4) $p_{\tilde{\gamma}_n}(\underline{d}^{(n)}(x) \notin LE(n_0, n) \cup LF(n)|I_n) < c_n^{bn}$,
(5) $p_{\tilde{\gamma}_n}(\underline{d}^{(n)}(x) \notin LE(n_0, n)|I_n^{\tau_n}) < c_n^{2a/5}$.

This translates immediately using PhPa2 to a parameter estimate analogous to Lemma 7.4 of [**AM1**]:

Lemma A.16. — *With total probability, for all n_0 big enough, for all n big enough (depending on n_0), $\tau_n \in VG(n_0, n)$.*

Before going on we will need two simple estimates: one is for the return time of very good branches and another is for the return time of branches which are neither very good or bad. The first of those estimates is analogous to Lemma 7.5 of [**AM1**], and follows directly from the definitions of very good and bad branches.

Lemma A.17. — *With total probability, for all n_0 big enough and for all $n \geqslant n_0$, if $j \in VG(n_0, n+1)$ then*

$$m < r_{n+1}(j) < mc_{n-1}^{-4b},$$

where, as usual, m is such that $R_n(I_{n+1}^j) = C_n^{\underline{d}}$ and $\underline{d} = (j_1, \ldots, j_m)$.

Lemma A.18. — *With total probability for all n_0 sufficiently big, if $n > n_0$, if $j \notin VG(n_0, n) \cup B(n_0, n)$ then $r_n(j) < c_{n-1}^{-a/2} c_{n-2}^{-4b}$.*

Proof. — Indeed, if $j \notin VG(n_0, n) \cup B(n_0, n)$ then $R_{n-1}(I_n^j) \subset C_{n-1}^{\underline{d}}$ with $\underline{d} \in LF(n-1)$. By definition of fast landing, $l_{n-1}(\underline{d}) < c_{n-1}^{-a/2} c_{n-2}^{-3b}$, so

$$r_n(j) = v_{n-1} + l_{n-1}(\underline{d}) < c_{n-1}^{-a/2} c_{n-2}^{-3b} + c_{n-2}^{-b}. \qquad \square$$

At this stage we have most of the tools to show that almost every parameter is "Collet-Eckmann at first returns", that is, $|Df^{k_n}(f(0))|$ is exponentially big for the sequence k_n of first landings of $f(0)$ in I_n. To obtain the full Collet-Eckmann condition (exponential growth for all k), we will need to analyze truncations of branches or landings, that is, we will consider iterates of the type $f^k|I_n^j$ (or $f^k|C_n^{\underline{d}}$) for k less then the return time $r_n(j)$ (or $l_n(\underline{d})$).

We now show that very good branches are well behaved when truncated at a reasonably big time. Here "well behaved" means "spending most of the time in very good branches of the previous level". So if we are able to control the hyperbolicity of very good branches in some level we will have a good possibility of controlling truncated very good branches in the next level. This lemma corresponds to Lemma 7.6 of [**AM1**], but the proof must be modified, with the use of bad returns and fast landings.

Lemma A.19. — *With total probability, for all n_0 big enough and for all $n \geqslant n_0$, the following holds.*

Let $j \in VG(n_0, n+1)$, as usual let $R_n(I^j_{n+1}) \subset C^{\underline{d}}_n$ and $\underline{d} = (j_1, \dots, j_m)$. Let m_k be biggest possible with

$$v_n + \sum_{j=1}^{m_k} r_n(j_i) \leqslant k$$

(the amount of full returns to level n before time k) and let

$$\beta_k = \sum_{\substack{1 \leqslant i \leqslant m_k, \\ j_i \in VG(n_0, n)}} r_n(j_i).$$

(the total time spent in full returns to level n which are very good before time k) Then $1 - \beta_k/k < c_{n-1}^{a^2/3}$ if $k > c_n^{-2/n}$.

Proof. — Let us estimate first the time i_k which is not spent on non-critical full returns:

$$i_k = k - \sum_{j=1}^{m_k} r_n(j_i).$$

This corresponds exactly to v_n plus some incomplete part of the return $j_{m_{k+1}}$. This part can be bounded by $c_{n-1}^{-b} + c_{n-1}^{-3b}$ (use the estimate of v_n and LS2 to estimate the incomplete part).

Using LS2 we conclude now that

$$m_k > (k - c_{n-1}^{-b}) c_{n-1}^{3b} > c_n^{-1/n}$$

so m_k is not too small.

Let us now estimate the contribution h_k from bad full returns j_i. The number of such returns must be less than $c_{n-1}^{n/2} m_k$ by LE2 and the estimate on m_k. By LS2 their total time is at most $c_{n-1}^{(n/2)-3b} m_k < m_k$.

The non very good full returns on the other hand can be estimated by LE1 (given the estimate on m_k), they are at most $c_{n-1}^{a^2} m_k$. So we can estimate the total time l_k of non very good or bad full returns (with time less then $c_{n-1}^{-a/2} c_{n-2}^{-4b}$ by Lemma A.18) by

$$c_{n-1}^{a^2} c_{n-1}^{-a/2} c_{n-2}^{-4b} m_k,$$

while β_k can be estimated from below by

$$(1 - c_{n-1}^{a/4}) c_{n-1}^{-a/2} m_k.$$

It is easy to see then that $i_k/\beta_k \ll c_{n-1}^{a/5}$, $h_k/\beta_k \ll c_{n-1}^{a/5}$. We also have

$$\frac{l_k}{\beta_k} < 2 c_{n-1}^{a^2/2}.$$

So $(i_k + h_k + l_k)/\beta_k$ is less then $c_{n-1}^{a^2/3}$. Since $i_k + h_k + l_k + \beta_k = k$ we have $1 - \beta_k/k < (i_k + h_k + l_k)/\beta_k$. □

Remark A.4. — This lemma illustrates the main reason why the original argument of [**AM1**] must be changed in order to deal with big quasisymmetric constants. Indeed, in [**AM1**], we do not need to split the branches which are not very good in bad branches and otherwise (fast). The reason is that in [**AM1**] the distribution of $r_n(j)$ is concentrated in a much narrower window around c_{n-1}^{-1} (say, $(c_{n-1}^{-1+2\varepsilon}, c_{n-1}^{-1-2\varepsilon})$). In particular, in a large sequence $(j_1, \ldots, j_k)$ (which should be thought as an initial segment of an excellent landing), we can estimate the proportion of the total return time due to very good branches essentially by considering the proportion of very good branches in the sequence.

In this Appendix, the distribution of $r_n(j)$ is located in a much larger window $(c_{n-1}^{-a}, c_{n-1}^{-b})$. The risky situation is to have a large sequence $(j_1, \ldots, j_k)$ with a large proportion of very good branches, but whose return time is near the bottom of the window (c_{n-1}^{-a}), while the not very good branches in the sequence have all return time near the top (c_{n-1}^{-b}). In this case, the proportion of the total time due to very good branches could be very small.

The solution given in this Appendix is based on the idea that the not very good branches *with large time* (bad branches) are really very few: most of the not very good branches are indeed fast. Paying attention to this asymmetry, we can indeed prove that in such a sequence $(j_1, \ldots, j_k)$, most of the total time is due to very good branches.

This argument (most branches with atypical time are fast) is based implicitly in the following asymmetry which appeared already in our first statistical estimate, Lemma A.5, when we showed that the distribution of $|\underline{d}^{(n)}(x)|$ is concentrated around c_n^{-1}: there is a big difference (one extra exponential) in the estimates on the upper tail (γ-qs capacity of $\{|\underline{d}^{(n)}(x)| > c_n^{-k\tilde{b}}\}$) and the lower tail ($\gamma$-qs capacity of $\{|\underline{d}^{(n)}(x) < c_n^{-k\tilde{a}}\}$).

(Essentially the same problem, with the same solution, appears in Lemma A.22.)

Now we conclude that very good (that is, most) branches are good, justifying our previous hints.

Lemma A.20. — *With total probability, for n_0 big enough and for all $n > n_0$, $VG(n_0, n) \subset G(n_0, n)$.*

The proof is the same as for Lemma 7.10 of [**AM1**], the two main features of very good branches exploited here are their good distribution of return times and the condition VG which allows us to avoid drastic losses of derivative due to starting very close to the critical point. The argument is by induction: first, all very good branches of level n_0 satisfy condition G1 of a good branch, that is, a full return is very hyperbolic (this follows from the definition of λ_{n_0}). Then, supposing that all very good branches of level n satisfy G1, we conclude that very good branches of level $n+1$ have enough hyperbolic branches in its composition (even if truncated) to satisfy both conditions G1 and G2.

A.7.3. Cool landings. — As we hinted in the last section, very good branches play the role of building blocks of hyperbolicity. We must now show that the critical point spends most of its time in very good branches. To do so, we will define a class of landings which are composed by many very good branches, but which are controlled to an ever greater detail than excellent landings. Their design will allow to estimate their hyperbolicity if truncated outside a relatively small initial segment.

We define the set of cool landings $LC(n_0,n) \subset \Omega$, $n_0, n \in \mathbb{N}$, $n \geqslant n_0$ as the set of all $\underline{d} = (j_1, \ldots, j_m)$ in $LE(n_0,n)$ satisfying

LC1: (Starts very good) $j_i \in VG(n_0,n)$, $1 \leqslant i \leqslant c_{n-1}^{-a^2/2}$.

LC2: (Not very good moments are sparse in large enough initial segments) For all $c_{n-1}^{-a^2/4} < k \leqslant m$

$$\#\{1 \leqslant i \leqslant k,\, r_n(j_i) < c_{n-1}^{-a/2}\} < (6 \cdot 2^n) c_{n-1}^{a/3} k,$$

LC3: (Bad moments are sparse in large enough initial segments) For $c_{n-1}^{-n/3} \leqslant k \leqslant m$

$$\#\{1 \leqslant i \leqslant k,\, j_i \in B(n_0,n)\} < (6 \cdot 2^n) c_{n-1}^{n/6} k,$$

LC4: (Starts with no bad moments) $j_i \notin B(n_0,n)$, $1 \leqslant i \leqslant c_{n-1}^{-n/2}$.

As in Lemma 7.7 of [**AM1**], cool landings are frequent and we get the following parameter estimate analogous to Lemma 7.8 of [**AM1**]. The ideas of this estimate are quite similar to the case of standard landings.

Lemma A.21. — *With total probability, for all n_0 big enough, for all n big enough we have $R_n(0) \in LC(n_0,n)$.*

Let us now show that cool landings inherit hyperbolicity from very good returns. This result corresponds to Lemma 7.11 of [**AM1**], but the proof of this fact needs adjustments for big quasisymmetric constants, so we provide it here.

Lemma A.22. — *With total probability, if n_0 is sufficiently big, for all n sufficiently big, if $\underline{d} \in LC(n_0,n)$ then for all $c_{n-1}^{-4/(n-1)} < k \leqslant l_n(\underline{d})$,*

$$\inf_{C_n^{\underline{d}}} \frac{\ln |Df^k|}{k} \geqslant \frac{\lambda_{n_0}}{2}.$$

Proof. — Fix such $\underline{d} \in LC(n_0,n)$, and let $\underline{d} = (j_1, \ldots, j_m)$.

Let

$$a_k = \inf_{C_n^{\underline{d}}} \frac{\ln |Df^k|}{k}.$$

Analogously to Lemma A.19, we define m_k as the number of full returns before k, that is, the biggest integer such that

$$\sum_{i=1}^{m_k} r_n(j_i) \leqslant k.$$

We define

$$\beta_k = \sum_{\substack{1\leqslant i\leqslant m_k,\\ j_i\in VG(n_0,n+1)}} r_n(j_i),$$

(counting the time up to k spent in complete very good returns) and

$$i_k = k - \sum_{i=1}^{m_k} r_n(j_i).$$

(counting the time in the incomplete return at k).

Let us then consider two cases: small m_k ($m_k < c_{n-1}^{-a^2/2}$) and otherwise.

Case 1 ($m_k < c_{n-1}^{-a^2/2}$). The idea of the first case is that all full returns are very good by LC1, and the incomplete time is also part of a very good return.

Since full very good returns are very hyperbolic by G1 and very good returns are good, we just have to worry about possibly losing hyperbolicity in the incomplete time. To control this, we introduce the queue (or tail) $q_k = \inf_{C_n^d} \ln |Df^{i_k} \circ f^{k-i_k}|$. We have $-q_k < -\ln(c_{n-1}^{1/3} c_{n-1}^3)$ by VG and Lemma A.11. Let us split again in two cases: i_k big or otherwise.

Subcase 1a ($i_k > c_{n-1}^{-4/(n-1)}$). If the incomplete time is big, we can use G2 to estimate the hyperbolicity of the incomplete time (which is part of a very good return). The reader can easily check the estimate in this case.

Subcase 1b ($i_k < c_{n-1}^{-4/(n-1)}$). If the incomplete time is not big, we can not use G2 to estimate q_k, but in this case i_k is much less than k: since $k > c_{n-1}^{-4/(n-1)}$, at least one return was completed ($m_k \geqslant 1$), and since it must be very good we conclude that $k > c_{n-1}^{-a/2}$ by LS1, so

$$a_k > \lambda_{n_0} \frac{(1+2^{n_0-n})}{2} \cdot \frac{k-i_k}{k} - \frac{-q_k}{k} > \frac{\lambda_{n_0}}{2}.$$

Case 2 ($m_k > c_{n-1}^{-a^2/2}$). For an incomplete time we still have $-q_k < -\ln(c_n c_{n-1}^3)$, so $-q_k/k < c_{n-1}^{a^2/3}$.

Arguing as in Lemma A.19, we split $k - \beta_k - i_k$ (time of full returns which are not very good) in part relative to bad returns h_k and in part relative to returns that are not very good or bad (which must be fast) l_k. Using LC3 and LC4 to bound the number of bad returns and LS2 to bound their time, we get

$$h_k < c_{n-1}^{-3b} c_{n-1}^{n/7} m_k,$$

and using LC1 and LC2 we have

$$l_k < c_{n-1}^{-a/2} c_{n-2}^{-4b} (6 \cdot 2^n) c_{n-1}^{a^2} m_k,$$

By LC1 and LC2 again, using LS1 to estimate the time of a very good return by $c_{n-1}^{-a/2}$, we have that $\beta_k > c_{n-1}^{-a/2} m_k/2$, thus we get

$$\frac{h_k + l_k}{\beta_k} < c_{n-1}^{a^2/2}, \tag{A.2}$$

which is very small.

On the other hand, $\beta_k > c_{n-1}^{-a/2} c_{n-1}^{-a^2/2}/2$ by hypothesis on m_k. Let us split in three cases according to the behavior of i_k.

Subcase 2a (i_k not very good or bad). In this case, $i_k < c_{n-1}^{-a/2} c_{n-2}^{-4b}$, so i_k/β_k is very small, and we actually have $1 - \beta_k/k < c_{n-1}^{a^2/10}$. Since very good returns are good and even not very good returns have derivative at least 1,

$$a_k > \lambda_{n_0} \frac{1 + 2^{n_0 - n}}{2} \cdot \frac{\beta_k}{k} - \frac{-q_k}{k} > \frac{\lambda_{n_0}}{2}. \tag{A.3}$$

Subcase 2b (i_k very good). If i_k is very good and $i_k > c_{n-1}^{-4/(n-1)}$, we can reason as in Subcase 1a that $G2$ can be used for the estimate of q_k so that we have

$$a_k > \lambda_{n_0} \frac{1 + 2^{n_0 - n}}{2} \cdot \frac{\beta_k}{k} + \frac{i_k}{k} \cdot \frac{\lambda_{n_0}}{2} < \frac{\lambda_{n_0}}{2}$$

by (A.2).

If $i_k \leqslant c_{n-1}^{-4/(n-1)}$, then i_k/β_k is very small and so $1 - \beta_k/k < c_{n-1}^{a^2/10}$, and we obtain (as in Subcase 2a) estimate (A.3).

Subcase 2c (i_k bad). If i_k is bad, by LC4 we have that $m_k > c_{n-1}^{-n/2}$, but $i_k < c_{n-1}^{-3b}$ by LS2, so i_k/β_k is very small again and we have $1 - \beta_k/k < c_{n-1}^{a^2/10}$, so estimate (A.3) applies and we are done. $\square$

A.8. Collet-Eckmann. — Since the critical point always falls in cool landings (see Lemma A.21), the Collet-Eckmann condition follows easily from Lemma A.22 (which guarantees gain of derivative after large truncations), together with Lemma A.11, which controls loss of derivative at small truncations. This argument is identical to the one in § 8.1 of [**AM1**], but we reproduce it here for the convenience of the reader.

Let

$$a_k = \frac{\ln |Df^k(f(0)))}{k}$$

and $e_n = a_{v_n - 1}$.

It is easy to see that if n_0 is big enough such that both Lemmas A.21 and A.22 we obtain for n big enough that

$$e_{n+1} \geqslant e_n \frac{v_n - 1}{v_{n+1} - 1} + \frac{\lambda_{n_0}}{2} \cdot \frac{v_{n+1} - v_n}{v_{n+1} - 1}$$

and so

$$\liminf_{n \to \infty} e_n \geqslant \frac{\lambda_{n_0}}{2}. \tag{A.4}$$

Let now $v_n - 1 < k < v_{n+1} - 1$. Define $q_k = \ln |Df^{k - v_n}(f^{v_n}(0))|$.

Assume first that $k < v_n + c_{n-1}^{-4/(n-1)}$. From LC1 we know that τ_n is very good, so by LS1 we have $r_n(\tau_n) > c_{n-1}^{-a/2}$, so k is in the middle of this branch (that is, $v_n \leqslant k \leqslant v_n + r_n(\tau_n) - 1$). Using that $|R_n(0)| > |I_n|/2^n$ (by Lemma A.7), we get by Lemma A.11 that $-q_k < -\ln(2^{-n}c_{n-1}c_{n-1}^5)$. We then get from $v_n > c_{n-1}^{-a}$ that

$$a_k \geqslant e_n \frac{v_n - 1}{k} - \frac{-q_k}{k} > \left(1 - \frac{1}{2^n}\right) e_n - \frac{1}{2^n}. \tag{A.5}$$

If $k > v_n + c_{n-1}^{-4/(n-1)}$ using Lemma A.22 we get

$$a_k \geqslant e_n \frac{v_n - 1}{k} + \frac{\lambda_{n_0}}{2} \cdot \frac{k - v_n + 1}{k}. \tag{A.6}$$

Estimates (A.4), (A.5), and (A.6) imply that $\liminf a_k \geqslant \lambda_{n_0}/2$ and so f is Collet-Eckmann.

A.9. Recurrence. — To show that the critical point is polynomially recurrent, we can follow the same lines from [**AM1**]. First we look at the essentially Markov process $R_n|(I_n \setminus I_{n+1})$, which shows that with total probability, most (in the γ-qs sense) points in I_n approach 0 with a polynomial rate (the exponent must be chosen according to γ) until the first time they fall in I_{n+1}. More precisely, we show (after transferring to the parameter) the following estimate (analogous to Corollary 8.3 of [**AM1**]).

Lemma A.23. — *With total probability, for n big enough and for $1 \leqslant i \leqslant s_n$,*

$$\frac{\ln |R_n^i(0)|}{\ln(c_{n-1})} < b^2 \left(1 + \frac{\ln(i)}{\ln(c_{n-1}^{-1})}\right).$$

To obtain the polynomial recurrence for f we relate the return times in terms of R_n to return times in terms of f. In other words, letting k_i be such that $R_n^i(0) = f^{k_i}(0)$, we must relate k_i and i. It is enough to do the estimate for a cool landing and we obtain the following estimate (as in Corollary 8.5 of [**AM1**]).

Lemma A.24. — *With total probability, for n big enough and for $1 \leqslant i \leqslant s_n$,*

$$\frac{\ln(k_i)}{\ln(c_{n-1}^{-1})} > a/3 \left(1 + \frac{\ln(i)}{\ln(c_{n-1}^{-1})}\right).$$

Let now $v_n \leqslant k < v_{n+1}$. If $|f^k(0)| < k^{-3b^3}$ we have $f^k(0) \in I_n$ and so $k = k_i$ for some i. It follows from Lemmas A.23 and A.24 that

$$|f^{k_i}(0)| > k_i^{-3b^3}.$$

This concludes the proof of polynomial recurrence. We notice that polynomial lower bounds are easily obtained: considering $|R_n(0)| = |f^{v_n}(0)| < c_{n-1}$ and using $v_n < c_{n-1}^{-b}$ we get

$$\limsup_{n \to \infty} \frac{-\ln |f^n(0)|}{\ln n} \geqslant a.$$

References

[Ar] V. Arnold. Dynamical systems. In "Development of mathematics 1950–2000", 33–61, Birkhäuser, Basel, 2000.

[A] A. Avila. Bifurcations of unimodal maps: the topologic and metric picture. IMPA Thesis (2001) www.math.sunysb.edu/~artur/.

[ALM] A. Avila, M. Lyubich and W. de Melo. Regular or stochastic dynamics in real analytic families of unimodal maps. Preprint IMS at Stony Brook, #2001/15. To appear in Invent. Math.

[AM1] A. Avila, C.G. Moreira. Statistical properties of unimodal maps: the quadratic family. Preprint www.arXiv.org. To appear in Annals of Math.

[AM2] A. Avila, C.G. Moreira. Statistical properties of unimodal maps: physical measures, periodic orbits and pathological laminations. Preprint www.arXiv.org.

[AM3] A. Avila, C.G. Moreira. Quasisymmetric robustness of the Collet-Eckmann condition in the quadratic family. Preprint www.arXiv.org.

[AM4] A. Avila, C.G. Moreira. Phase-Parameter relation and sharp statistical properties in general families of unimodal maps. Preprint www.arXiv.org.

[BBM] V. Baladi, M. Benedicks and V. Maume. Almost sure rates of mixing for i.i.d. unimodal maps. Ann. Sci. Ecole Norm. Sup. (4), v. 35 (2002), no. 1, 77-126.

[BV] V. Baladi and M. Viana. Strong stochastic stability and rate of mixing for unimodal maps. Ann. Sci. Ecole Norm. Sup. (4), v. 29 (1996), no. 4, 483-517.

[BC] M. Benedicks and L. Carleson. On iterations of $1 - ax^2$ on (-1,1). Ann. Math., v. 122 (1985), 1-25.

[GS] J. Graczyk and G. Swiatek. Generic hyperbolicity in the logistic family. Ann. of Math., v. 146 (1997), 1-52.

[J] M. Jakobson. Absolutely continuous invariant measures for one-parameter families of one-dimensional maps. Comm. Math. Phys., v. 81 (1981), 39-88.

[KN] G. Keller and T. Nowicki. Spectral theory, zeta functions and the distribution of periodic points for Collet-Eckmann maps. Comm. Math. Phys., 149 (1992), 31-69.

[K] O.S. Kozlovski. Structural stability in one-dimensional dynamics. Thesis (1998).

[L1] M. Lyubich. Combinatorics, geometry and attractors of quasi-quadratic maps. Ann. Math, 140 (1994), 347-404.

[L2] M. Lyubich. Dynamics of quadratic polynomials, I-II. Acta Math., 178 (1997), 185-297.

[L3] M. Lyubich. Dynamics of quadratic polynomials, III. Parapuzzle and SBR measure. Preprint IMS at Stony Brook, # 1995/5. Astérisque, v. 261 (2000), 173 - 200.

[L4] M. Lyubich. Almost every real quadratic map is either regular or stochastic. Ann. of Math. (2) 156 (2002), no. 1, 1-78.

[MN] M. Martens and T. Nowicki. Invariant measures for Lebesgue typical quadratic maps. Preprint IMS at Stony Brook, # 1996/6. Astérisque, v. 261 (2000), 239–252.

[MvS] W. de Melo and S. van Strien. One-dimensional dynamics. Springer, 1993.

[NP1] T. Nowicki and F. Przytycki. The conjugacy of Collet-Eckmann's map of the interval with the tent map is Hölder continuous. Ergodic Theory Dynam. Systems 9 (1989), no. 2, 379–388.

[NP2] T. Nowicki and F. Przytycki. Topological invariance of the Collet-Eckmann property for S-unimodal maps. Fund. Math. 155 (1998), no. 1, 33–43.

[NS] T. Nowicki and D. Sands. Non-uniform hyperbolicity and universal bounds for S-unimodal maps. Invent. Math. 132 (1998), no. 3, 633–680.

[Pa] J. Palis. A global view of dynamics and a conjecture of the denseness of finitude of attractors. Astérisque, v. 261 (2000), 335–347.

[T1] M. Tsujii. Positive Lyapunov exponents in families of one dimensional dynamical systems. Invent. Math. 111 (1993), 113–137.

[T2] M. Tsujii. Small random perturbations of one dimensional dynamical systems and Margulis-Pesin entropy formula. Random & Comput. Dynamics. Vol.1 No.1 59–89, (1992).

[Y] L.-S. Young. Decay of correlations for certain quadratic maps. Comm. Math. Phys., 146 (1992), 123–138.

A. AVILA, Collège de France, 3, Rue d'Ulm, 75005 Paris, France • *E-mail :* avila@impa.br
Url : www.math.sunysb.edu/~artur

C.G. MOREIRA, IMPA, Estr. D. Castorina 110, 22460-320 Rio de Janeiro, Brazil
E-mail : gugu@impa.br • *Url :* www.impa.br/∼gugu

Astérisque
286, 2003, p. 119–150

GEOMETRY OF MULTI-DIMENSIONAL DISPERSING BILLIARDS

by

Péter Bálint, Nikolai Chernov, Domokos Szász & Imre Péter Tóth

Abstract. — Geometric properties of multi-dimensional dispersing billiards are studied in this paper. On the one hand, non-smooth behaviour in the singularity submanifolds of the system is discovered (this discovery applies to the more general class of semi-dispersing billiards as well). On the other hand, a self-contained geometric description for unstable manifolds is given, together with the proof of important regularity properties. All these issues are highly relevant to studying the ergodic and statistical behaviour of the dynamics.

1. Introduction

Let $\mathbb{Q}$ be an open connected domain in $\mathbb{R}^d$ or on the d-dimensional torus $\mathbb{T}^d$. Assume that the boundary $\partial\mathbb{Q}$ consists of a finite number of C^k smooth ($k \geqslant 3$) compact hypersurfaces (possibly, with boundary). Now let a pointwise particle move freely (along a geodesic line with constant velocity) in $\mathbb{Q}$ and reflect elastically at the boundary $\partial\mathbb{Q}$ (by the classical rule "the angle of incidence is equal to the angle of reflection"). This is what is commonly refered to as a billiard dynamical system.

Billiards make an important class in the modern theory of dynamical systems. Many classical and quantum models in physics belong to this class, most notably, the Lorentz gas [**Si**] and hard ball gases studied as early as the XIX century by L. Boltzmann [**Bo**].

The periodic Lorentz process is obtained by fixing a finite number of disjoint convex bodies $B_1, \ldots, B_s \subset \mathbb{T}^d$ with smooth boundary and putting the moving particle in

2000 ***Mathematics Subject Classification.*** — 37D50.
Key words and phrases. — Dispersing billiards, hyperbolicity, singularities, curvature bounds.

N. Chernov acknowledges partial support from NSF grant DMS-9732728. The financial support of the Hungarian National Foundation for Scientific Research (OTKA), grants T26176 and T32022; and of the Research Group Stochastics@TUB of the Hungarian Academy of Sciences, affiliated to the Technical University of Budapest is also acknowledged.

the exterior domain $\mathbb{Q} = \mathbb{T}^d \setminus (\cup B_i)$. This system models the motion of an electron among a periodic array of molecules in a metal, as it was introduced by H. Lorentz in 1905.

Mathematical studies of billiards have begun long ago. Ya. Sinai in his seminal paper of 1970 [**Si**] described the first large class of billiards with truly chaotic behavior — with nonzero Lyapunov exponents, positive entropy, enjoying ergodicity, mixing, and (as was later discovered by G. Gallavotti and D. Ornstein [**GO**]) the Bernoulli property. Sinai billiards are defined in two dimensions ($d = 2$), i.e. for $\mathbb{Q} \subset \mathbb{R}^2$ or $\mathbb{Q} \subset \mathbb{T}^2$, and the boundary of $\mathbb{Q}$ must be concave (i.e., convex inward $\mathbb{Q}$), similarly to the Lorentz process (where the bodies B_i are convex). Due to the geometric concavity, the boundary $\partial\mathbb{Q}$ scatters or disperses bundles of geodesic lines falling upon it, see Fig. 1. For this reason, Sinai billiards are said to be *dispersing*.

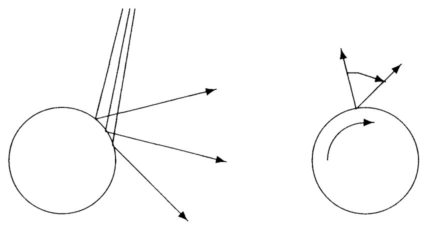


FIGURE 1. Scattering effect

Lorentz processes in two dimension have been studied very thoroughly since 1970. Many fine ergodic and statistical properties have been established by various researchers, including P. Bleher, L. Bunimovich, N. Chernov, J. Conze, C. Dettmann, G. Gallavotti, A. Krámli, J. Lebowitz, D. Ornstein, K. Schmidt, N. Simányi, Ya. Sinai, D. Szász, and others (see the references). The latest major result for this model (the exponential decay of correlations) was obtained by L.-S. Young [**Y1**]. The success in these studies had significant impact on modern statistical mechanics. The methods and ideas originally developed for the planar Lorentz process were applied to many other classes of physical models — see recent reviews by Cohen, Gallavotti, Ruelle and Young [**GC, Ru, Y2**].

On the other hand, the progress in the study of the multidimensional Lorentz process (where $d > 2$) has been much slower and somewhat controversial. Relatively few papers were published covering specifically the case $d > 2$, especially in contrast to the big number of works on the 2-D case. Furthermore, the arguments in the published articles were usually rather sketchy, as in Chernov's paper [**Ch1**]. It was commonly assumed that the geometric properties of the multidimensional Lorentz process were essentially similar to those of the 2-D system, and so the basic methods of study should be extended from 2-D to any dimension at little cost. Thus, the authors rarely elaborated on details.

Recent discoveries proved that spatial dispersing billiards are *very much* different from planar ones. Bunimovich and Reháček studies of *astigmatism* [**BR**], in the somewhat different context of focusing billiards, emphasized the known fact that the billiard trajectories may focus very rapidly in one plane and very slowly in the orthogonal planes. Astigmatism is unique to 3-D (and higher dimensional) billiards, it cannot occur on a plane. It plays an improtant role in higher dimensional focusing billiards as investigated in [**BR**].

In this paper we consider multidimensional *dispersing* billiards. We show that multi-dimensionality has great effect on the dynamics in the dispersing case as well — the system requires much more elaborated study than the 2D process. What is worse (cf. section 3), the singularity manifolds in the phase space of a spatial Lorentz process have pathologies — points exist where the sectional curvature is unbounded (blows up). Actually, singularity manifolds are in these pathologies — which form two-codimensional submanifolds of them — not even differentiable. Indeed, as it will be shown in section 3, the unit normal vector to the singularity manifold has different directional limits at the pathological points — the geometry is pretty much like the classical Whitney umbrella $x^2z = y^2$ in $\mathbb{R}^3$. This phenomenon is again unique to billiards in dimension $d \geqslant 3$. All these facts call for a revision of some earlier arguments and results on the multidimensional Lorentz process. This is much the more important since the studies of physically relevant multiparticle systems will require the same methods as those used for the high-dimensional Lorentz process.

Throughout the paper we conduct a systematic study of the geometry of the Lorentz process in any dimension $d > 2$, aiming at the future investigation of its ergodic and statistical properties (in particular, the decay of correlations). First we describe our recent discovery — pathological behavior of singularity manifolds — and show exactly where it occurs (in order to "localize the pathology"). Then we develop tools for the study of basic geometric properties of the dynamics — operator techniques in the Poincaré section of the phase space. By applying these geometric tools we provide rigorous proofs of important properties for unstable manifolds: we show absolute continuity, distorsion bounds, curvature bounds and alignment. All these facts are absolutely important for the studies of ergodic and statistical properties of the Lorentz gas, but strangely enough, their proofs (in the case of dimension $d > 2$) have never been published before. Lastly, we show how our results can be used in the study of the decay of correlations, which will be done in a separate paper.

2. Preliminaries

There are two ways of considering billiard dynamics, the motion of a point particle in a connected, compact domain $\mathbb{Q} \subset \mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$, $d \geqslant 2$ with a piecewise C^3-smooth boundary. The phase space of the *flow* can be identified with the unit tangent bundle over $\mathbb{Q}$ — the configuration space is $\mathbb{Q}$ while the phase space is $\mathcal{M} := \mathbb{Q} \times \mathbb{S}^{d-1}$

($\mathbb{S}^{d-1}$ is the surface of the unit d-ball). In other words, every phase point x is of the form (q, v) where $q \in \mathbb{Q}$ and $v \in \mathbb{S}^{d-1}$. We denote the flow by $S^t : -\infty < t < \infty$.

On the other hand there is a naturally defined cross-section for this flow. The phase space of the Poincaré section *map* (or simply, of the billiard map) is $M := \partial\mathbb{Q} \times \mathbb{S}_+^{d-1}$, where $+$ means that we only take into account the hemisphere of the outgoing velocities (for a more precise definition of the phase space, see subsection 4.1). For any $x \in \mathcal{M}$ we set $t^+(x) := \inf\{t > 0 \mid S^t x \in M\}$, and $T^+ x := S^{t^+(x)} x$ (of course, $T^+ : \mathcal{M} \to M$). Then the Poincaré section map $T : M \to M$ is defined as follows: $Tx := T^+ x$ for $x \in M$.

We require the following properties from the system to be studied:

– Our billiard is *dispersing* (a Sinai-billiard): each $\partial\mathbb{Q}_i$ is strictly convex (had we required convexity only, our billiard would be *semi-dispersing*).

– The scatterers B_i are disjoint. This ensures the C^3-smoothness of the boundary $\partial\mathbb{Q}$, i.e. that there are *no corner points*.

– The condition that *the horizon is finite* says exactly that $t^+(x) < \infty$ for any $x \in M$.

Finally, some more notation. Let $n(q)$ be the unit normal vector of the boundary component $\partial\mathbb{Q}_i$ at $q \in \partial\mathbb{Q}_i$ directed inwards $\mathbb{Q}$. Then the invariant Liouville-measure of the discretized map is

$$d\mu(q, v) := \text{const.}\, \langle n(q), v\rangle\, dq\, dv \tag{2.1}$$

where dq is the induced Riemannian measure on $\partial\mathbb{Q}$ whereas dv is the Lebesgue-measure on $\mathbb{S}_+^{d-1}$.

Throughout the paper, unless otherwise emphasized, we are considering this discretized dynamics.

2.1. Fronts. — In billiard theory, several basic constructions and concepts are based on the notion of a local orthogonal manifold, which - for simplicity - we will call front. A front $\mathcal{W}$ is defined in the whole phase space rather than in the Poincaré section. Take a smooth 1-codim submanifold E of the whole configuration space, and add the unit normal vector $v(r)$ of this submanifold at every point r as a velocity, continuously. Consequently, at every point the velocity points to the same side of the submanifold E. Then

$$\mathcal{W} = \{(r, v(r)) \mid r \in E\} \subset \mathcal{M},$$

where $v : E \to \mathbb{S}^{d-1}$ is continuous (smooth) and $v \perp E$ at every point of E. The derivative of this function v, called B plays a crucial role: $dv = Bdr$ for tangent vectors (dr, dv) *of the front*. B acts on the tangent plane $\mathcal{T}_r E$ of E, and takes its values from the tangent plane $\mathcal{J} = \mathcal{T}_{v(r)}\mathbb{S}^{d-1}$ of the velocity sphere. These are both naturally embedded in the configuration space $\mathbb{Q}$, and can be identified through this embedding. So we just write $B : \mathcal{J} \to \mathcal{J}$. B is nothing else than the curvature operator of the submanifold E. Yet we will prefer to call it second fundamental form (s.f.f.), in order to avoid confusion with other curvatures that are coming up. Obviously, B is symmetric.

Notice that fronts remain fronts during time evolution - at least locally, and apart from some singularity lines.

When we talk about a front, we sometimes think of it as the part of the (whole) phase space just described (for example, when we talk about time evolution under the flow), but sometimes just as the submanifold E (for example, when we talk about the tangent space or the curvature of the front). This should cause no confusion.

2.2. Evolution of fronts. — The evolution of a front during free propagation (that is, from one collision to the other) is described by the formula

$$B_1^- = ((B^+)^{-1} + \tau Id)^{-1} \tag{2.2}$$

where τ is the length of the free run between the two collisions, B^+ is the s.f.f. of the front just after the first collision, and B_1^- is the s.f.f. just before the next one.

For this formula — and the next one — to make sense, we need to identify the tangent planes of the front at different moments of time. Let $\mathcal{T} = \mathcal{T}_r\partial\mathbb{Q}$ be the tangent plane of the scatterer at a collision point r. Just like $\mathcal{J}$, $\mathcal{T}$ is viewed together with its natural embedding into $\mathbb{Q}$. The identification of different $\mathcal{J}$'s is done in the usual way (cf. [**SCh**], [**KSSz**]):

– by translation parallel to v from one collision to the other,

– by reflection with respect to $\mathcal{T}$ (or, equivalently, by projection parallel to n) from pre-collision to post-collision moments.

Notation for the unitary operator that executes this identification is U, however, for brevity, we will often omit U if it causes no confusion.

At a moment of collision the curvature of the front changes non-continuously (the front is "scattered"):

$$B^+ = B^- + 2\Theta = B^- + 2\langle n, v\rangle V^*KV \tag{2.3}$$

where[(1)]

– $B^- : \mathcal{J} \to \mathcal{J}$ is the s.f.f. just before collision,

– $B^+ : \mathcal{J} \to \mathcal{J}$ is the s.f.f. just after collision,

– $V : \mathcal{J} \to \mathcal{T}$ is the projection parallel to v: $Vdv = dv - \frac{\langle dv, n\rangle}{\langle v, n\rangle}v \in \mathcal{T}$ for $dv \in \mathcal{J}$,

– $V^* : \mathcal{T} \to \mathcal{J}$ (the adjoint of V) is the projection parallel to n: $V^*dq = dq - \frac{\langle dr, v\rangle}{\langle n, v\rangle}n \in \mathcal{J}$ for $dq \in \mathcal{T}$,

– $K : \mathcal{T} \to \mathcal{T}$ is the s.f.f. of the scatterer at the collision point,

– $\langle n, v\rangle = \cos\phi$, where $\phi \in [0, \frac{\pi}{2}]$ is the so-called collision angle,

– and the operator $\Theta : \mathcal{J} \to \mathcal{J}$: $\Theta = \langle n, v\rangle V^*KV$ is the so-called collision term.

(1) This convention on the collision term ($\Theta = \langle n, v\rangle V^*K$) will be useful in the geometric description of the phase space, see section 4.

2.3. Singularities . — As it can be easily seen the billiard map T is discontinuous at pre-images of tangential reflections. Indeed, consider the set of tangential reflections:

$$\mathcal{S}_0 := \partial M = \{(q, v) \mid \langle v, n(q) \rangle = 0\}$$

(which is nothing else than the boundary of the phase space). Its pre-images are:

$$\mathcal{S}_k = T^{-k}\mathcal{S}_0 \quad (k > 0).$$

(From section 4 on it will be useful to introduce the notation $\mathcal{S}^{(k)}$ for the set of all singularities up to k, i.e. $\mathcal{S}^{(k)} = \cup_{i=1}^{k}\mathcal{S}_i$.) The map T is discontinuous precisely at the points of $\mathcal{S}_1 (= \mathcal{S}^{(1)})$. Furthermore — related to the smallness of the term $\langle n, v \rangle$ — the derivative DT is unbounded near $\mathcal{S}_1$. As a consequence, to get a well-behaved dynamics, the phase space is partitioned into homogeneity layers by introducing secondary singularities (for a detailed discussion see [**BSC2**] or subsection 4.1).

To consider higher iterates of the dynamics — the maps T^k $(k \geqslant 1)$ — the sets $\mathcal{S}_k$ are to be investigated. We view all these sets as (finite unioins of) topologically embedded one codimensional compact submanifolds with boundary. They have smooth manifold structure in the interior, however, in the multi-dimensional case (as it is demonstrated in subsection 3.1) the behaviour at the boundary is irregular (the curvature diverges). This behaviour is related to the fact that in the multi-dimensional case, in addition to unbounded derivatives, the dynamics is highly non-isotropic near singularities.

3. Geometry of singularities

In several papers that appeared, singularities were assumed — either explicitely or implicitely — to consist of smooth 1-codim submanifolds of the phase space. Often, even a uniform bound on the curvature was assumed, independent of the order of the singularity. This is true in 2-dimensional billiards. However, it is not true in higher dimensions. In this section we present a counter-example in a 3-dimensional dispersing billiard. In correspondence with the notations introduced in subsection 2.3, we will use the notation $\mathcal{S}_1$ and $\mathcal{S}_2$ for the set of those phase points the trajectories of which have tangential first and second collisions, respectively. We will demonstrate that already the curvature of $\mathcal{S}_2$ has no upper bound, i.e. the curvature blows up near a point where the singularity manifold is not even differentiable.

To avoid confusion let us make one further remark. As already mentioned, billiard dynamics has singularities: points where the billiard map is not continuous. These singularities occur on one codimensional submanifolds of the phase space. The development of the theory is based on considering connected and essentially smooth components of the singularity manifolds. The recently discovered phenomenon described below shows that these components are, indeed, only essentially smooth. On certain two-codimensional submanifolds of them pathologies occur: singularities in the sense of algebraic singularity theory. To avoid confusion we will refer to these singular two-codimensional submanifolds as *pathologies* (in contrast to the *singularities*, the singularity manifolds of the dynamics themselves).

3.1. Counter-example for bounded curvature . — In this section we prove that even in a $3D$ dispersing billiard, already the two-step singularities have no bounded curvature. The proof is rather implicit. We start with the indirect assumption that the curvature is bounded, and find that the two-step singularity intersects the one-step singularity tangentially at every point of their intersection, except for a one-codimensional degeneracy, where the intersection is not tangent. However — as a consequence of bounded curvature — our indirect asumption implies that the unit normal vector of $\mathcal{S}_2$ is a continuously differentiable function of its base point. Thus the set of those points where the two singularity manifolds intersect non-tangentially is open in $\mathcal{S}_1 \cap \mathcal{S}_2$. This way we get a contradiction.

Consider the situation demonstrated on Figure 2. To perform as transparent an argument as possible

– the parameters on the figure and in the calculations below are different,

– the first scatterer, the surface where the trajectories start out is a plane — thus it is not strictly convex.

Nevertheless the reader can easily see that these modifications have no real significance. We are in 3 dimensions, so take a standard 3D Cartesian coordinate system. Let the first 'scatterer' be the $\{z = 0\}$ plane. Let the second scatterer be the sphere with centre $O_1 = (0, -1, 1)$ and radius $R = 1$. Let the third scatterer be the sphere with centre $O_2 = (1, 0, 2)$ and radius $R = 1$. We look at the component of the phase space corresponding to the first scatterer, near the phase point $(x_0 = 0, y_0 = 0, v_{x0} = 0, v_{y0} = 0)$. Of course, $v_{z0} = 1$, and the trajectory is the z axis. We are interested in the singularity manifold belonging to a tangent *second* collision. To describe this, let $D \in \mathbb{R}^4$ be the set of those points (x, y, v_x, v_y) the trajectories of which hit the first sphere. Let $r : D \to \mathbb{R}$ be the distance of the trajectory and O_2. That is, the singularity manifold we are looking at is the set $\mathcal{S}_2 = \{(x, y, v_x, v_y) \in D \mid r(x, y, v_x, v_y) = 1\}$. So, if we want to construct the normal vector of the singularity manifold, we just need to calculate the gradient of r. We will directly calculate the partial derivatives. Since $(x_0, y_0, v_{x0}, v_{y0}) = (0, 0, 0, 0)$ is on the boundary of D, we can only hope to find one-side partial derivatives. What is even worse: $(x, y, v_x, v_y) = (x, 0, 0, 0) \in D$ only if $x = 0$, so we cannot differentiate with respect to x. The same is true for v_x. What we can do is take these partial derivatives at the points $(0, y, 0, v_y)$ and than the limits

$$\lim_{y\to 0}\lim_{v_y\to 0}\frac{\partial}{\partial x} r(x, y, v_x, v_y)|_{x=v_x=0}.$$

(we will see that it is important to fix $x = v_x = 0$).

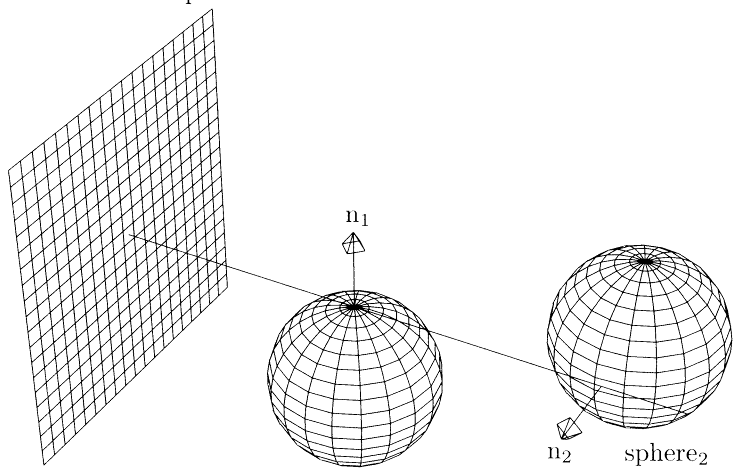


FIGURE 2. The studied billiard configuration

We start with the indirect assumption that $\mathcal{S}_2$ has bounded curvature. This implies that the unit normal vector of $\mathcal{S}_2$ is a continuously differentiable function of its base point with bounded derivative. In this way it makes sense to define the normal vector of $\mathcal{S}_2$ on the boundary points of $\mathcal{S}_2$ as the limit of (unit) normal vectors on the interior. For us the indirect assumption will mean that the limit

$$\operatorname{grad} r(0,0,0,0) := \lim_{(x,y,v_x,v_y)\to(0,0,0,0)} \operatorname{grad} r(x,y,v_x,v_y)$$

exists.

The closer a reflection is to tangential, the less effect it has on the "neutral" direction. In our case, the reflection on the first sphere causes "no scattering" in the x direction. That is, let (v'_x, v'_y, v'_z) be the velocity after the first collision. The "x" direction being the "neutral" direction means that

$$\lim_{y\to 0} \frac{\partial}{\partial v_x} v'_x(0,y,0,0) = 1$$

which implies that

$$\lim_{y\to 0} \frac{\partial}{\partial v_x} r(0,y,0,0) = -2$$

Similarly,

$$\lim_{y\to 0} \frac{\partial}{\partial x} v'_x(0,y,0,0) = 0$$

which implies that

$$\lim_{y\to 0} \frac{\partial}{\partial v_x} r(0,y,0,0) = -1.$$

According to our indirect assumption, this means that

$$\frac{\partial}{\partial x} r(0,0,0,0) = -1 \quad \text{and} \quad \frac{\partial}{\partial v_x} r(0,0,0,0) = -2.$$

For the other two components, fix $x = v_x = 0$. So the trajectory is in the $\{x = 0\}$ plane, the scattering is just a 2D problem. We will calculate the one-side partial derivatives $\frac{\partial}{\partial y} r(0,0,0,0)$ and $\frac{\partial}{\partial v_y} r(0,0,0,0)$.

To find out about v'_y, let ϕ be the angle of the first sphere's radius at the first collision point and the $(0,1,0)$ vector. If $v_y = 0$, then $1 - \cos\phi = -y$ ($y < 0$, of course), which, in leading order, gives $\phi = \sqrt{-2y}$. It can be seen that after the reflection $v'_y = \sin 2\phi$. That is, the trajectory is far from being a line. However, it is diverted in the very direction which - in the first order - does not affect its distance from O_2. Instead, in leading terms, $r^2 = 1 + (v'_y)^2$.

Putting these together, we get $r = \sqrt{1 - 8y}$, that is,

$$\frac{\partial}{\partial y} r(0,0,0,0) = -4.$$

If we fix $y = 0$, the exact same consideration gives $r = \sqrt{1 - 8v_y}$, that is,

$$\frac{\partial}{\partial v_y} r(0,0,0,0) = -4$$

as well. All together, we get

$$\operatorname{grad} r(0,0,0,0) = (-1,-4,-2,-4).$$

This is (the limit of) the normal vector of the singularity at the point ($x = 0$, $y = 0$, $v_x = 0$, $v_y = 0$).

It is easy to see that the singularity corresponding to a tangent reflection on the first sphere has the normal vector

$$\operatorname{grad} r_0(x,y,v_x,v_y) = (0,-1,0,-1).$$

That is, the two singularities are not tangent at this point.

The previous consideration for $\operatorname{grad} r$ also shows that this behaviour is exceptional. It is the result of the fact that in the first order r was unaffected by v'_y. If the radii at the reflection points $(x,y,z) = (0,0,1)$ and $(x,y,z) = (0,0,2)$ had not been orthogonal, the result would have been

$$\frac{\partial r}{\partial y} = \infty, \quad \frac{\partial r}{\partial v_y} = \infty,$$

corresponding to a normal vector $(0,1,0,1)$, meaning that the two singularities are tangent. Non-tangentiallity of the two singularities is a one-codimensional degeneracy.

As we have pointed out at the beginning of the subsection, this contradicts our indirect assumption on the boundedness of the curvature. In this way we have only proven that the assumption was false. However, we believe that the picture of the singularity suggested above is correct, the singularities *are* tangent almost everywhere, and their curvature only blows up near the pathological points described.

3.2. Discussion. — For a rigorous proof of some finer properties (such as correlation decay) of multi-dimensional dispersing billiards it seems essential to characterize singularities in a systematic way. Such a characterization should be subject to future research (some possible ideas related to this question are discussed in [**BChSzT**]). In this subsection we do not plan to give rigorous proofs; we would like to point out some analogies to and emphasize some interesting features of the irregularities demonstrated above.

The Whitney-umbrella. — Consider the one-codimensional set in $\mathbb{R}^3$ defined by the polynomial equation:

$$\{(x, y, z) \in \mathbb{R}^3 \mid x^2 z = y^2\},$$

the Whitney-umbrella (for more details see [**AGV**]). 'One half' of this set (its intersection with the quadrants $xy \geqslant 0$) is shown on Figure 3. For simplicity we use the notations: W_2 for this 'half-umbrella' and W_1 for the $\{z = 0\}$ plane. Clearly

– W_2 terminates on W_1 (in the points of the x-axis), thus $W_1 \cap W_2 = \partial W_2$.

– at every point of the x-axis where $x \neq 0$ the intersection of W_2 and W_1 is tangential.

– W_2 has smooth manifold structure in its interior; nevertheless, near the origin its curvature is unbounded as the normal vector changes rapidly (actually, the normal vector does not even have a well-defined limit at the origin).

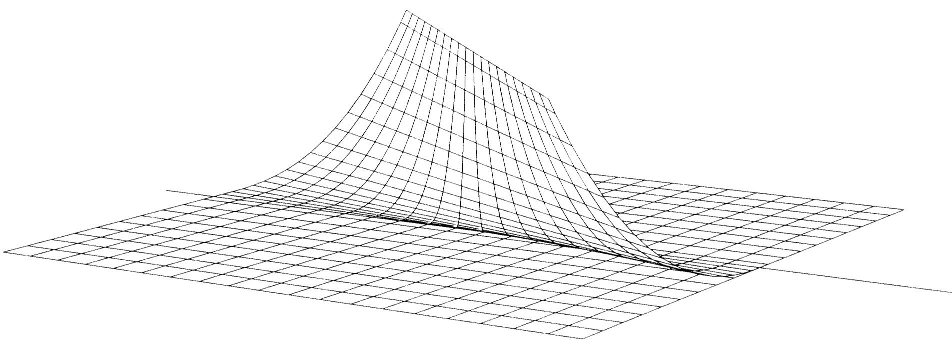


FIGURE 3. The Whitney Umbrella

By these properties the geometry of singularities described in subsection 3.1 is analogous to Figure 3.(2) W_1 corresponds to $\mathcal{S}_1$, W_2 corresponds to $\mathcal{S}_2$ while the

(2)To be precise, the situation on Figure 3. has one dimension less — in contrast to W_2 the singularities are 3-dimensional manifolds — but this has little significance to the analogy.

origin corresponds to the set of those doubly tangential reflections where the two radii are orthogonal (this set is one-codimensional in $\mathcal{S}_1 \cap \mathcal{S}_2$).

Generalization I. — First let us consider the first-step singularity $\mathcal{S}_1$. By the notations of the previous subsection we may characterize the points of (x, y, v_x, v_y) belonging to $\mathcal{S}_1$ easily. These are precisely those for which $d(x, y, v_x, v_y) = 1$, where $d(.,.,.,.)$ is the distance of the point $O_1 = (0, -1, 1)$ from the line that passes through the point $(x, y, 0)$ and has direction specified by the velocity components v_x, v_y. As d is a smooth function of its variables there is no curvature blow-up for $\mathcal{S}_1$ — and, for first-step singularities in general. Thus $\mathcal{S}_2$ is a *pre-image of a smooth one-codimensional compact submanifold*, however, *the map* under which the pre-image is taken *has unbounded derivatives and is highly an-isotropic.* Curvature blow-up occurs only at those points of $\mathcal{S}_2$ (near its intersection with $\mathcal{S}_1$) where the map behaves irregularly.

In correspondence with the above observation we conjecture that curvature blow-up is not a peculiar feature of $\mathcal{S}_2$, it is present in the pre-images of one-codimensional smooth submanifolds in general. Consider for example two-step *secondary* singularities Γ_2 — those phase points for which at the second iterate instead of tangentiality the collision angle $(\langle n, v\rangle)$ is a given constant (see section 4 for more detail). In the specific example of subsection 3.1 such secondary singular trajectories are precisely those that touch tangentially a sphere of radius R' $(R' < 1)$ at the second iterate. It is clear that the geometry of Γ_2 is completely analogous to $\mathcal{S}_2$.

Generalization II. — Our calculations in subsection 3.1 do not use any speciality of the explicitly given billiard configuration. Doubly tangential reflections for which the normal vectors of the scatterers at the consecutive collisions are orthogonal can be found in *any multi-dimensional semi-dispersing billiard.* Near such trajectories a similar calculation can be performed.

Generalization III. — All in all, the discovered pathology is general. In addition, *the higher step singularities* $\mathcal{S}_k$; $(k \geqslant 3)$ may show even wilder behaviour near their intersections. Nevertheless, we strongly conjecture that a nice geometric characterization — suggested by the analogy with the Whitney-umbrella in the case of $\mathcal{S}^{(2)}$ — can be performed. This question is subject to future research.

4. Geometric properties of u-manifolds

Throughout sections 4 and 5 we investigate u-manifolds (their counterparts, s-manifolds can be treated similarly). u-manifolds are $d-1$-dimensional submanifolds of the phase space with tangent planes in the (appropriately defined) unstable cone. Possibly the most important tools in studying ergodic and statistical properties, local unstable manifolds (or LUMs for short) are suitable limits of u-manifolds (for details see [**Y1, Ch2, Ch3**]). In contrast to the $2d-3$-dimensional (one-codimensional) singularity manifolds, u-manifolds behave in a uniformly regular way. In section 4 we introduce a natural geometrical description that turns out to be very useful for studying multi-dimensional dispersing billiards. Proofs for some basic properties of u-manifolds are also included. More involved technicalities — that play a crucial role in investigating the statistical behaviour of a billiard system (cf. [**Y1, Ch2, Ch3**]) — are discussed in section 5.

4.1. The phase space. — We shall work with the discrete time (collision to collision) dynamical system, thus our phase space — which we denote by M — is the Poincaré phase space, the collection of possible collision points supplied with outgoing velocities. Mathematically this space is a bundle over the scatterers $\partial\mathbb{Q}$, the fibers of which consist of the possible outgoing velocities. At every base point q the fiber is the $(d-1)$-dimensional hemisphere with boundary which we shall denote by $\mathbf{S}_+^{d-1}$. Note that this bundle can be viewed as a subbundle (of vectors of unit length) in the direct sum of the tangent and normal bundles over the scatterers. Thus, by the Riemannian structure of $\partial\mathbb{Q}$, there is a naturally defined parallel translation on our bundle (see the description of the tangent plane below). Local coordinates on our phase space will be denoted $x = (q, v)$. Additionally we shall use all the notations for local quantities introduced in the previous section(s) (eg. $n(q)$, ϕ).

Some conventions. — Throughout the paper the superscripts '+' and '−' denote post- and precollisional values, respectively, for certain functions, operators, hyperplanes etc. (e.g. v^+ and v^-). The dynamics and its derivative are denoted by T and DT, respectively. In correspondence with $x_1 = Tx$ ($\delta x_1 = DT\delta x$), the subscript '1' means the value of a certain quantity at the first iterate. We shall usually prime the points, trajectories, operators etc. infinitesimally close to a reference point or trajectory.

The tangent plane. — At any point $x = (q, v)$ the tangent plane has a natural splitting $\mathcal{T}_x M = \mathcal{T}_q\partial\mathbb{Q} + \mathcal{T}_v\mathbf{S}_+^{d-1} = \mathcal{T} + \mathcal{J}$. The two planes $\mathcal{J}$ and $\mathcal{T}$ are related by the projection operator $V : \mathcal{J} \to \mathcal{T}$ and its adjoint V^* (for their description see the section 2).

For two points $x = (q, v)$ and $x' = (q', v')$ infinitesimally close, the tangent vector pointing from x to x' is

$$\delta x = (\delta q, \delta v) \qquad \delta q = q - q'; \ \delta v = Q_0^{-1}v' - v$$

where Q_0 is the rotator that takes $\mathcal{T}$ to $\mathcal{T}'$. Up to first order:

$$Q_0 u = u - \langle u, dn\rangle n + \langle u, n\rangle dn \quad \text{for } u \in \mathbb{R}^d; \tag{4.1}$$

$$Q_0^{-1} u = u + \langle u, dn\rangle n - \langle u, n\rangle dn \quad \text{for } u \in \mathbb{R}^d \tag{4.2}$$

and thus:

$$\delta v = dv - \langle v, n\rangle V^* dn$$

Here $dv = v' - v$ and $dn = n' - n$. These formulas execute (up to first order) the parallel translation of the bundle M.

4.2. Important submanifolds

Singularity manifolds. — The dynamics T is discontinuous, the singularity manifold is $\mathcal{S}^{(1)} = \mathcal{S}_1 = T^{-1}\mathcal{S}_0$ where $\mathcal{S}_0 = \partial M = \{(q,v) \mid \langle v, n\rangle = 0\}$ is just the boundary of the phase space. However, as already mentioned, to get a well-behaved dynamics we should partition the original phase space into homogeneity layers:

$$I_k = \{(q,v) \in M \mid (k+1)^{-2} < \langle v, n(q)\rangle < k^{-2}\} \quad \text{and}$$
$$I_0 = \{(q,v) \in M \mid \langle v, n(q)\rangle > k_0^{-2}\} \tag{4.3}$$

Here the integer constant k_0 is arbitrary. The boundary of this partitioned phase space, $\overline{M}$ is

$$\Gamma_0 = \partial\overline{M} = \cup_{k=k_0}^{\infty}\{(q,v) \mid \langle v, n\rangle = k^{-2}\}$$

Correspondingly, the countably many manifolds in the set $\Gamma^{(1)} = T^{-1}\Gamma_0$ are the so called *secondary singularities.* For a higher iterate of the dynamics, T^n, the primary and secondary singularities are, respectively:

$$\mathcal{S}^{(n)} = \mathcal{S}^{(1)} \cup T^{-1}\mathcal{S}^{(1)} \cup \cdots T^{-n+1}\mathcal{S}^{(1)}; \quad \Gamma^{(n)} = \Gamma^{(1)} \cup T^{-1}\Gamma^{(1)} \cup \cdots T^{-n+1}\Gamma^{(1)}.$$

Fronts. — As introduced in section 2, $(d-1)$-dimensional submanifolds in $\mathbb{Q}$, the configurational space of the flow, everywhere orthogonal to the flow direction will be referred to as fronts. When supplied with their normal vectors v (the velocities), fronts can be viewed as submanifolds of the flow phase space $\mathcal{M}$. Vectors (in the tangent bundle over $\mathcal{M}$) tangent to fronts are denoted by $(dr, dv) = (dr, Bdr)$ where B is the second fundamental form (s.f.f.) of our submanifold in $\mathbb{Q}$ (here, of course, $dr \perp v$).

Let us consider a front directly after (before) collision. It leaves a trace of velocities on the scatterer which can be viewed either as a (unit) vector field over $\partial\mathbb{Q}$ or as a $(d-1)$- dimensional submanifold in the Poincaré phase space. Direct calculations show that for a vector $(dr, dv) = (dr, B^+dr)$, tangent to the post-collisional front, the corresponding vector in the Poincaré phase space is $\delta x = (\delta q, \delta v)$ where:

$$\begin{aligned}\delta q &= Vdr;\\ \delta v &= dv - \langle v, n\rangle V^*dn = dv - \langle v, n\rangle V^*K\delta q\\ &= (B^+V^{-1} - \langle v, n\rangle V^*K)\delta q = F\delta q.\end{aligned} \tag{4.4}$$

The operator $F : \mathcal{T} \to \mathcal{J}$ plays an important role, it describes the tangent plane of our $(d-1)$-dimensional manifold in the Poincaré phase space.

A front will be called *convex/diverging* whenever B^+ is positive semi-definite ($B^+ \geqslant 0$). Convex fronts remain convex under time evolution. The *convex cone* consists of those tangent vectors δx that are tangent to some convex front.

Lemma 4.1. — *There are constants $m_0 \in \mathbb{N}$ and $\phi_0 < \pi/2$ that depend only on the billiard domain itself such that out of m_0 consecutive reflections at least for one of them for the collison angle ϕ we have: $\phi < \phi_0$.*

Proof. — Let us assume the contrary: there is a sequence x_n of phase points which have trajectories with n consecutive collisions, all with collision angle $\phi > \pi/2 - 1/n$. By compactness there is a limit phase point with infinitely many consecutive tangential reflections. This, however, contradicts the finite horizon assumption. □

u-manifolds and homogeneous u-manifolds. — We shall consider C^u_x, the m_0-image of the convex cone as our *unstable cone.* A manifold is a *u-manifold* if it has all tangent vectors in C^u_x. u-manifolds remain u-manifolds as C^u_x is invariant under the positive powers of T.

A u-manifold is said to be *homogeneous* if it is contained in one homogeneity layer.

There will be two metrics used on u-manifolds. Before their introduction we mention that for any vector dz in $\mathcal{T}$ or in $\mathcal{J}$ $\|dz\|$ is the notation for the Euclidean length and for oprators O acting on these spaces $\|O\|$ denotes the naturally induced norm.

The *p-metric*

$$\|\delta x\|_p = \|dr\|$$

measures distances on the corresponding front while the *Euclidean metric*

$$\|\delta x\|_e = \sqrt{\delta q^2 + \delta v^2}$$

in the Poincaré phase space. A priori the p-metric seems to be degenerate but as we shall see it is a good metric on the cone C^u_x. Time evolution in the p-metric is given by:

$$\|\delta x_1\|_p = \|dr_1\| = \|dr + \tau dv\| = \|(I + \tau B^+)dr\| \tag{4.5}$$

Some further notation. — For any u-manifold W; the quantities $J^p_W(x)$ and $J^e_W(x)$ are the Jacobians of the dynamics in the p- and e-metrics, respectively.

Remark. — All the above introduced concepts have their natural counterparts (with the corresponding nice properties) for the reversed dynamics: concave/convergent fronts, s-manifolds etc.

4.3. Properties of F and equivalence of metrics

Some conventions. — Constants that depend only on the billiard table itself (like $\tau_{\min}$, ϕ_0...) will be called *global constants.*

For an invertible operator O the meaning of the relations $c \prec O \prec C$ is that there are two positive global constants C_1 and C_2 that bound the norms of the operator and its inverse:

$$\|O\| < C_1; \qquad \|O^{-1}\| < C_2.$$

Note that the operator O is not necessarily symmetric, even more, it need not be an automorphism. The values of the constants C_1 and C_2 are usually irrelevant.

Two quantities f and g defined on the unstable cones will be called *equivalent* ($f \sim g$) if there are some global constants C_1 and C_2 such that $C_1 f \leqslant g \leqslant C_2 f$.

Throughout this subsection we restrict our considerations on the vectors of the unstable cone.

Sublemma 4.2. — *Let us consider any u-front with incoming and outgoing s.f.f.-s B^- and B^+, respectively. Then $c \prec B^+$ and $c \prec B^- \prec C$.*

Proof. — By the collision equations the operator $B^+ - B^-$ is always positive semi-definite, thus it is enough to prove $c \prec B^- \prec C$ as it implies $c \prec B^+$. The upper bound is trivial by (2.2) and the lack of corner points (there is a lower bound on the free path: $\tau \geqslant \tau_{\min}$). Thus it remains to prove $c \prec B^-$, what is an easy consequence of Lemma 4.1. Indeed, our submanifold is an m_0-iterate of a convex front . By the lemma out of these m_0 reflections there is definitely at least one with collision angle smaller than ϕ_0. We shall denote the collision term that corresponds to this particular reflection by Θ_0. Of course, $c \prec \Theta_0$ as the spectrum of Θ_0 is bounded below by $k_{\min} \cos\phi_0$ (here $k_{\min}$ is the lower bound on the spectrum of K — the curvature operator of the scatterers $\partial\mathbb{Q}$). Now let us consider any $dr \in \mathcal{J}$. By the evolution equations (2.2) and (2.3):

$$\langle dr, B^- dr\rangle \geqslant \langle dr, ((\Theta_0)^{-1} + m_0\tau_{\max} I)^{-1} dr\rangle \geqslant ((k_{\min}\cos\phi_0)^{-1} + m_0\tau_{\max})^{-1}\langle dr, dr\rangle.$$

Thus we have the desired lower bound. □

Now we can formulate our most important technical lemma.

Lemma 4.3. — *Assume $K' : \mathcal{T} \to \mathcal{T}$ and $B' : \mathcal{J} \to \mathcal{J}$ are both symmetric, positive definite and $c \prec B', K' \prec C$. Then:*

$$c \prec B'V^{-1} + \langle v, n\rangle V^* K' \prec C.$$

Proof. — The upper bound is obvious since $\|V^{-1}\| = 1$ and $\langle v, n\rangle \|V^*\| = 1$.

By the definition of V, we have

$$Vu = u - \frac{\langle u, n\rangle}{\langle v, n\rangle} v \qquad \text{for } u \in \mathcal{J}$$

and

$$V^{-1}u = u - \langle u, v\rangle v \qquad \text{for } u \in \mathcal{T}$$

Similarly,

$$V^* u = u - \frac{\langle u, v\rangle}{\langle v, n\rangle} n \qquad \text{for } u \in \mathcal{T} \tag{4.6}$$

and

$$(V^*)^{-1} u = u - \langle u, n\rangle n \qquad \text{for } u \in \mathcal{J}$$

It is then easy to arrive at

$$(V^*)^{-1} V^{-1} u = u - \langle u, v\rangle v + \langle u, v\rangle\langle v, n\rangle n$$

and

$$\langle v, n\rangle^2 V V^* u = \langle v, n\rangle^2 u + \langle u, v\rangle v - \langle u, v\rangle\langle v, n\rangle n$$

Adding the two equations above yields

$$(V^*)^{-1}V^{-1} + \langle v,n\rangle^2 VV^* = (1+\langle v,n\rangle^2)I \tag{4.7}$$

where I is the identity operator in $\mathcal{T}$.

Another useful observation: Since $\|(B')^{-1}\| \leqslant C$ and $\|(K')^{-1}\| \leqslant C$ for a global constant $C>0$, all the eigenvalues of B' and K' are bounded below by $c'=1/C$. Hence

$$\langle B'u,u\rangle > c'\|u\|^2 \qquad \text{for} \quad u\in\mathcal{J} \tag{4.8}$$

and

$$\langle K'u,u\rangle > c'\|u\|^2 \qquad \text{for} \quad u\in\mathcal{T} \tag{4.9}$$

Now, let $u\in\mathcal{T}$, $\|u\|=1$. Then $\|V^{-1}u\|\leqslant 1$, and

$$\langle B'V^{-1}u + \langle v,n\rangle V^*K'u, V^{-1}u\rangle = \langle B'V^{-1}u, V^{-1}u\rangle + \langle v,n\rangle\langle K'u,u\rangle$$

Here all three scalar products are positive, hence

$$\|B'V^{-1}u + \langle v,n\rangle V^*K'u\| \geqslant c'\|V^{-1}u\| \tag{4.10}$$

due to (4.8). Next, we have $\langle v,n\rangle\|V^*u\|\leqslant 1$, and

$$\langle B'V^{-1}u + \langle v,n\rangle V^*K'u, \langle v,n\rangle V^*u\rangle = \langle B'V^{-1}u, \langle v,n\rangle V^*u\rangle + \langle K'u, \langle v,n\rangle^2 VV^*u\rangle$$

Substitution of (4.7) and using (4.9) gives

$$\|B'V^{-1}u + \langle v,n\rangle V^*K'u\| \geqslant c'\|u\|^2 - c''\|V^{-1}u\| = c' - c''\|V^{-1}u\|$$

for some global constant $c''>0$. Combining this with (4.10) yields

$$\|B'V^{-1}u + \langle v,n\rangle V^*K'u\| \geqslant c$$

with $c = c'/(1+c''/c')$. The lower bound is proved. □

Corollary 4.4. — *There are global constants c and C such that for any u-front $c \prec F \prec C$. As a consequence, for all vectors of the unstable cone, $\delta x\in C^u_x$ the norm $\|\delta x\|_e$ is uniformly equivalent to both $\|\delta q\|$ and $\|\delta v\|$. Furthermore, the p-metric is non-degenerate on the cone C^u_x (nonzero vectors in C^u_x have nonzero p-length).*

Proof. — This is an easy application of Lemma 4.3 with $B'=B^-$ and $K'=K$ (see also formula (4.4)). □

Corollary 4.5. — *The p-metric and the e-metric are equivalent in a 'dynamical' sense: for any $\delta x\in C^u_x$: $\|DT\delta x\|_p \sim \|\delta x\|_e$.*

Proof. — Indeed, by the evolution equation (4.5):

$$\|DT\delta x\|_p = \|(I+\tau B^+)dr\| = \|(I+\tau B^+)V^{-1}\delta q\|.$$

Now we may apply Lemma 4.3 with $K' = 2K$ and $B' = I + \tau B^-$ (remember that the free path τ is uniformly bounded from below and above). Together with Corollary 4.4 we get:

$$\|(I + \tau B^+)V^{-1}\delta q\| \sim \|\delta q\| \sim \|\delta x\|_e.$$

The two equations together give Corollary 4.5. □

Before going into further details we would like to make an important remark.

Remark 4.6. — From the next section on we turn to a closer investigation of u-manifolds. We will see that — as long as the properties discussed in the rest of the paper are concerned — u-manifolds are no less regular in multi-dimensional billiards than in the planar ones. This can be easily checked if our results are compared to those proved in the literature for the two-dimensional case, see especially [**Ch2**], Section 6 and the references cited there.

Nevertheless, there are important differences from planar billiards in the way how u-manifolds are actually described. Anisotropy of the geometry is reflected in the use of linear operators. It is of course much more difficult to handle operators than numbers, thus the proof of the very same regularity properties becomes more technical as one switches from dimension two to three.

4.4. Geometry and hyperbolicity of u-manifolds. — Now we would like to turn to the hyperbolic and geometric properties of the unstable cone. Unless otherwise stated, any vector δx mentioned is an element of the u-cone C^u_x.

Uniform *hyperbolicity in the p-metric* is guaranteed by the uniform bound $\tau > \tau_{\min}$ and Sublemma 4.2. Indeed:

$$\|DT\delta x\|_p = \|(Id + \tau B^+)dr\| > \Lambda\|\delta x\|_p.$$

Here $\Lambda > 1$ is a global constant. On the other hand, by Sublemma 4.2 again (together with the evolution equations) for the $(d-1)$ eigenvalues of the symmetric operator B^+:

$$\lambda_1 \sim (\cos\phi)^{-1}; \qquad \lambda_i \sim 1, \quad i = 2, \dots, d-1.$$

As a consequence, for an arbitrary u-manifold W the Jacobian in the p-metric behaves as

$$J^p_W(x) \sim (\cos(\phi))^{-1}.$$

In the *e-metric* we have by Corollary 4.5:

$$\|DT^n\delta x\|_e \geqslant \|DT^n\delta x\|_p > \Lambda^{n-1}\|DT\delta x\|_p > C\Lambda^n\|\delta x\|_e. \tag{4.11}$$

This implies that for a sufficiently high fixed power of the dynamics, $T_1 = T^{m_1}$:

$$\|DT_1\delta x\|_e > \Lambda_1\|\delta x\|_e \qquad \text{with } \Lambda_1 > 1 \text{ global.} \tag{4.12}$$

To calculate $J^e_W(x)$ for any u-manifold W consider the operator $G : \mathcal{T} \to \mathcal{T}_x W$ that acts by the rule $\delta q \mapsto (\delta q, F(\delta q)) = \delta x$. Then one can easily check that in our notation

$$DT\mid_W(x) = G_1 \circ V_1 \circ U_1 \circ (I + \tau B^+) \circ V^{-1} \circ G^{-1}$$

in correspondence with equation (4.5) that describes evolution in the p-metric. Now we may get a formula for the Jacobian in the e-metric:

$$J^e_W(x) = \det G_1 \, \det V_1 \, J^p_W(x) \, (\det V)^{-1} \, (\det G)^{-1}. \tag{4.13}$$

We observe that

$$(\det G)^2 = \det(I + F^* F) \tag{4.14}$$

Indeed, there is an orthonormal basis in $\mathcal{T}$ and an orthonormal basis in $\mathcal{J}$ such that $F : \mathcal{T} \to \mathcal{J}$ is represented, in those bases, by a diagonal matrix (this follows from the singular value decomposition theorem in linear algebra). For a diagonal matrix F, the relation (4.14) is easily verified by direct inspection.

Now it is easy to see that there are global constants c and C such that: $c < \det G < C$ for the operator G at any u-manifold. Direct calculation gives:

$$J^e_W(x) \sim \det(V_1) \sim (\cos(\phi_1))^{-1}. \tag{4.15}$$

Let us consider a further restriction of DT onto a subspace $R \subset \mathcal{T}_x W$ of the tangent plane. Applying the above argument for the restriction $DT\mid_R$ we get:

$$\det(DT\mid_R) \sim \det(V_1\mid_{R'}) \tag{4.16}$$

where $R' = (V_1^{-1} \circ G_1^{-1} \circ DT)(R)$.

Now we turn to some geometric properties of our submanifolds. *Transversality* — the property that the stable and unstable cones are uniformly transversal — is justified by the following theorem:

Theorem 4.7. — *The u-manifolds and s-manifolds in M are uniformly transversal. Precisely, there is a global constant $c_0 > 0$ such that for any u-manifold W_u and any s-manifold W_s at any point of intersection $x \in W_u \cap W_s$ the angle between W_u and W_s is greater than c_0.*

Proof. — We use the subscripts u and s to denote various quantities and operators related to the submanifolds W_u and W_s, respectively. According to (4.4),

$$F_u = UB_u^- U^{-1} V^{-1} + \langle v, n \rangle V^* K$$

and

$$F_s = B_s^+ V^{-1} - \langle v, n \rangle V^* K$$

Note that the operator $-B_s^+$ is symmetric, positive definite and satisfies $c \prec -B_s^+ \prec C$ (this is the counterpart of the previously established property $c \prec B_u^- \prec C$). Hence, the operator $B' := UB_u^- U^{-1} - B_s^+$ is symmetric, positive definite and satisfies $c \prec B' \prec C$. Now Lemma 4.3 implies

$$c \prec F_u - F_s \prec C \tag{4.17}$$

Next assume that Theorem 4.7 is false. Then, by using Corollary 4.4, one can easily conclude that for any $\varepsilon > 0$ there are a u-manifold W_u, an s-manifold W_s intersecting W_u at some point $x = (q, v)$, and a nonzero vector $\delta q \in \mathcal{T}$ such that

$$\|F_u(\delta q) - F_s(\delta q)\| \leqslant \varepsilon \, \|\delta q\|$$

This clearly contradicts (4.17). Theorem 4.7 is proved. □

Remark. — Observe that the above proof goes through even if instead of the s-manifold W_s we have just an arbitrary convergent front W_0. Indeed, for the crucial equation (4.17) it is enough to have the upper bound $-B_0^+ \prec C$ (which trivially holds for any convergent front W_0), the lower bound $c \prec -B_s^+$ — which is only true for s-manifolds — is, however, not essential.

As a consequence we are able to prove the so-called *alignment* property.

Corollary 4.8. — *The u-manifolds are uniformly transversal to all the singularity manifolds $S \subset \mathcal{S}^{(n)}$ and $S \subset \Gamma^{(n)}$, $n \geqslant 1$. Precisely, there is a global constant $c_0 > 0$ such that for any u-manifold W_u intersecting any manifold $S \subset \mathcal{S}^{(n)}$ or $S \subset \Gamma^{(n)}$ at a point x there is a $(d-1)$-dimensional submanifold $S' \subset S$ through x such that the angle between W_u and S' is greater than c_0.*

Proof. — We have $S = T^{-k}S_0$ for some $1 \leqslant k \leqslant n$ and a domain $S_0 \subset \mathcal{S}_0$ (or $S_0 \subset \Gamma_0$). Let $x_0 = (q_0, v_0) = T^k x \in S_0$. Define a small $(d-1)$-dimensional submanifold $S_0' \subset S_0$ through x_0 by $S_0' = \{y = (r, v) \in M \mid v = Q_0 v_0\}$, where Q_0 is the rotator of $\mathbb{R}^d$ taking $n(q_0)$ to $n(q)$, as defined by (4.1).

First let us discuss the primary singularities (i.e. the case $S_0 \subset \mathcal{S}_0$). We claim that $S' = T^{-k}S_0'$ is a limit, in C^0 metric, of a sequence of convergent fronts. Indeed, we first approximate S_0' by a sequence of $(d-1)$-dimensional manifolds $S_0^{(i)}$ defined as follows. Pick a sequence of vectors $v_0^{(i)} \in S^{d-1}$ such that $v_0^{(i)} \to v_0$ as $i \to \infty$ and $\langle v_0^{(i)}, n(q_0)\rangle > 0$ for all i. Then we put $S_0^{(i)} = \{y = (q, v) \in M \mid v = Q_0 v_0^{(i)}\}$. For each submanifold $S_0^{(i)}$, the tangent plane at every point $(q, v) \in S_0^{(i)}$ is characterized by $\delta v = 0$, hence $F = 0$ in our notation. According to (4.4), we now have $UB^- U^{-1} = -\langle v, n\rangle V^* K V^{-1}$, which is a negative definite operator. So, the trajectories of $S_0^{(i)}$, as they flow backward in time, make a convergent front. Therefore, $T^{-k}S_0^{(i)}$ is a convergent front for every i. As $i \to \infty$, these fronts converge to $S' = T^{-k}S_0'$, as we claimed. Now, Theorem 4.7 (in view of the remark above) completes the proof for the case of primary singularities.

In the secondary case (i.e. $S \subset \Gamma^{(n)}$) the $(d-1)$-dimensional manifold $S' = T^{-k}S'_0$ is a convergent front itself. Thus we may refer to the theorem and the remark directly. $\square$

Remark. — Recall that singularity manifolds are $2d-3$-dimensional. The above Corollary roughly states that there is a $d-1$-dimensional subbundle in their tangent bundle that lies in the stable cone field. However, the tangent space may behave wildly in the further $d-2$ directions, in correspondence with the curvature blow-up discussed in section 3.

5. Technical bounds on u-manifolds

After introducing the basic structures and tools now we would like to turn to the discussion of some more complicated technical properties. Unless otherwise stated, all calculations refer to the unstable cone (field) C^u_x and we use all other conventions from the previous section as well (e.g. quantities corresponding to a trajectory infinitesimally close to a reference one are primed).

Our main reference will be Lemma 4.3. Before discussing the important specific properties in the subsections, we record a few immediate consequences of this Lemma. For every u-manifold W, at every reflection we have

$$c \prec B^+V^{-1} \prec C. \tag{5.1}$$

This bound has its adjoint version

$$c \prec (V^*)^{-1}B^+ \prec C. \tag{5.2}$$

Let τ be the time between the current and the next reflections (or, more generally, any number satisfying $\tau_{\min}/10 < \tau \leqslant \tau_{\max}$). Then

$$c \prec (I+\tau B^+)V^{-1} \prec C \tag{5.3}$$

and we also have an adjoint version of (5.3)

$$c \prec (V^*)^{-1}(I+\tau B^+) \prec C. \tag{5.4}$$

Note that if $c \prec A \prec C$ for any operator A, then also $c \prec A^{-1} \prec C$. Hence, all the above inequalities remain true for the inverse operators as well. For example, we have

$$(I+\tau B^+)^{-1}V^* \prec C \quad \text{and} \quad V(I+\tau B^+)^{-1} \prec C. \tag{5.5}$$

5.1. Curvature bounds on u-manifolds. — In this subsection we would like to prove that there is a uniform bound on the curvature of u-manifolds. More precisely we prove that the tangent plane of a u-manifold is a Lipschitz function of the base point, with a uniform (global) Lipschitz constant. The tangent plane is described by the operator F, thus we should prove that F depends smoothly enough on the base point.

First we will get the relevant *curvature bounds in the phase space of the flow*; in other words, we investigate the smoothness of the dependence for s.f.f.-s B that describe any front corresponding to some u-manifold (which we refer to as u-fronts for short). Let $\mathcal{W}$ be any such u-front and $x = (r, v) \in \mathcal{W}$. Let $x' = (r', v') \in \mathcal{W}$ be infinitesimally close to x, and $dr = r' - r$, $dv = v' - v$ the infinitesimal displacement vectors in $\mathbb{Q}$ and $\mathbf{S}^{d-1}$, respectively. Clearly, $dr, dv \in \mathcal{J}$ and $dv = [B_{\mathcal{W}}(x)](dr)$. Consider the evolution of the displacement vector $(dr_t, dv_t) = S^t(dr, dv)$. If no collisions occur on an interval $(t, t + \Delta t)$, then $dv_{t+\Delta t} = dv_t$ and

$$dr_{t+\Delta t} = dr_t + \Delta t\, dv_t = [I + \Delta t\, B_t](dr_t) \tag{5.6}$$

where $B_t = B_{\mathcal{W}_t}(x_t)$. By Sublemma 4.2 we know that $\langle B_t u, u\rangle \geqslant b_{\min}\|u\|^2$ for all $u \in \mathcal{J}$. Therefore

$$\|dr_{t+\Delta t}\| \geqslant (1 + \Delta t\, b_{\min})\|dr_t\| \tag{5.7}$$

hence

$$\|(I + \Delta t\, B_t)^{-1}\| \leqslant (1 + \Delta t\, b_{\min})^{-1}. \tag{5.8}$$

Now consider a moment of reflection. The tangent vector $dx_t = (dr_t, dv_t)$ changes discontinuously, in correspondence with (2.3): $dr = dr^+ = U dr^-$ and $dv = dv^+ = U(dv^-) + \Theta(dr^+)$. The two trajectories reflect at the points $q, q' \in \partial Q$ in the time moments t, t' , respectively. For the infinitesimal differences we use the notations $dt \in \mathbb{R}$, $\delta q \in \mathcal{T}$ and $dn = n(q') - n(q) = K\delta q \in \mathcal{T}$. As to their relations:

$$\|dr^+\| \leqslant \|\delta q\|; \qquad |dt| \leqslant 2\|\delta q\|; \qquad \|dn\| \leqslant C\|\delta q\| \quad \text{and} \quad \|dv\| \leqslant C\|\delta q\|. \tag{5.9}$$

Indeed, these bounds are straight consequences of the formulas (2.3) and (4.4), the boundedness of K, the triangle inequality $|dt| \leqslant \|dq\| + \|dr^+\|$ and our crucial Lemma 4.3.

We need to *compare the operators* Θ *and* Θ' *taken at the points* (q, v) *and* (q', v'), respectively. They act in the hyperplanes $\mathcal{J}$ and $\mathcal{J}'$ orthogonal to v and v', respectively. Consider the operators V^*, K, V entering (2.3) at the reference point (q, v) and their counterparts $(V')^*, K', V'$ at the nearby point (q', v'). Let $Q = Q_{v,v'}$ be the rotation in $\mathbb{R}^d$ taking v to v' and leaving invariant all the vectors perpendicular to v and v'. Then Q takes $\mathcal{J}$ to $\mathcal{J}'$. More specifically, Q acts by the rule

$$Qu = u - \langle u, dv\rangle v \qquad \text{for } \ u \in \mathcal{J} \tag{5.10}$$

and its inverse acts by

$$Q^{-1}u = u + \langle u, dv\rangle v \qquad \text{for } \ u \in \mathcal{J}' \tag{5.11}$$

where the terms of the second order in dv are dropped. Furthermore we shall use another rotator, Q_0, that takes $\mathcal{T}$ to $\mathcal{T}'$: this later one we have already introduced at the description of the parallel translation of the tangent bundle (see (4.1), (4.2)).

Instead of V and V^*, it is now more convenient to work with more "tame" operators $\widetilde{V} = \langle v, n\rangle V$ and $\widetilde{V}^* = \langle v, n\rangle V^*$. They act by the rules

$$\widetilde{V}u = \langle v, n\rangle u - \langle u, n\rangle v \qquad \text{for} \quad u \in \mathcal{J} \tag{5.12}$$

and

$$\widetilde{V}^*u = \langle v, n\rangle u - \langle u, v\rangle n \qquad \text{for} \quad u \in \mathcal{T} \tag{5.13}$$

Similar formulas hold for $\widetilde{V}'$ and $(\widetilde{V}')^*$, where v', n' are substituted for v, n.

Put $\Delta\widetilde{V} = Q_0^{-1}\widetilde{V}'Q - \widetilde{V}$, $\Delta\widetilde{V}^* = Q^{-1}(\widetilde{V}')^*Q_0 - \widetilde{V}^*$ and $\Delta K = Q_0^{-1}K'Q_0 - K$. Direct calculations based on (5.12), (5.10) and (4.2) yield

$$[\Delta\widetilde{V}](u) = (\langle dv, n\rangle + \langle v, dn\rangle)u + (\langle v, n\rangle\langle u, dn\rangle - \langle u, n\rangle\langle v, dn\rangle)n - \langle u, dn\rangle v - \langle u, n\rangle dv$$

hence

$$\|\Delta\widetilde{V}\| \leqslant 2\,\|dv\| + 4\,\|dn\| \tag{5.14}$$

Note that $\Delta\widetilde{V}^*$ is the adjoint of $\Delta\widetilde{V}$, hence

$$\|\Delta\widetilde{V}^*\| = \|\Delta\widetilde{V}\| \leqslant 2\,\|dv\| + 4\,\|dn\| \tag{5.15}$$

Now, because $\partial\mathbb{Q}$ is C^3 smooth we have

$$\|\Delta K\| \leqslant C\|\delta q\| \tag{5.16}$$

for some global constant $C > 0$.

Sublemma 5.1. — *There is a global constant $C > 0$ such that for any $\tau \in (\tau_{\min}/10, \tau_{\max})$*

$$\|(I + \tau B^+)^{-1}(Q^{-1}\Theta'Q - \Theta)(I + \tau B^+)^{-1}\| \leqslant C\,\|\delta q\|$$

Proof. — Recall that

$$\Theta = 2\langle v, n\rangle V^*KV = 2\langle v, n\rangle^{-1}\widetilde{V}^*K\widetilde{V}$$

and similar formulas hold for Θ'. We have, to the first order of $\|\delta q\|$,

$$\begin{aligned} Q^{-1}\Theta'Q - \Theta &= 2(\langle v', n'\rangle - \langle v, n\rangle)V^*KV \\ &\quad + 2\langle v, n\rangle^{-1}(\Delta\widetilde{V}^*K\widetilde{V} + \widetilde{V}^*\Delta K\widetilde{V} + \widetilde{V}^*K\Delta\widetilde{V}) \end{aligned} \tag{5.17}$$

Note that $\langle v', n'\rangle - \langle v, n\rangle = (\langle dv, n\rangle + \langle v, dn\rangle)$, to the first order in $\|\delta q\|$. Thus we can rewrite (5.17) in this way:

$$Q^{-1}\Theta'Q - \Theta = 2(\langle dv, n\rangle + \langle v, dn\rangle)V^*KV + 2(\Delta\widetilde{V}^*KV + V^*\Delta K\widetilde{V} + V^*K\Delta\widetilde{V})$$

Now we apply (5.5) and then (5.14)-(5.16) with (5.9). This completes the proof of the sublemma. □

After so much preparation we are ready to *discuss curvature bounds for the flow*, i.e. for u-fronts $\mathcal{W}$.

We need to estimate the 'derivative' of the second fundamental form $B_{\mathcal{W}}(x)$ with respect to $x \in \mathcal{W}$. The operator $B_{\mathcal{W}}(x)$ acts in the hyperplane $\mathcal{J}$ that also depends on x. For points $x' = (r', v') \in \mathcal{W}$ infinitesimally close to x, let $Q = Q_{v,v'}$ be the rotator in $\mathbb{R}^d$ that takes $\mathcal{J}$ to $\mathcal{J}'$ as defined by (5.10). Then the 'increment' of B is defined by $Q^{-1}B'Q - B$, where $B = B_{\mathcal{W}}(x)$ and $B' = B_{\mathcal{W}}(x')$. Now consider

$$D_{\mathcal{W}}(x) := \max_{dr \neq 0} \|Q^{-1}B'Q - B\| / \|dr\|$$

where the maximum is taken over all nonzero infinitesimal displacement vectors $dr = r' - r$.

***Lemma 5.2* (Curvature bounds - I).** — *There is a constant $D_{\max}$ such that for any divergent wave front $\mathcal{W}$ and $x \in \mathcal{W}$ there is a $t_0 = t_0(\mathcal{W}, x)$ such that for all $t > t_0$ we have the following: if no collisions occur in the interval $(t - \tau_{\min}/2, t)$, then $D_{\mathcal{W}_t}(x_t) \leqslant D_{\max}$.*

Proof. — For short, we put $D_t = D_{\mathcal{W}_t}(x_t)$. First we show that D_t decreases during free runs between collisions.

Sublemma 5.3. — *If there are no collisions in a time interval $(t, t + \Delta t)$, then*

$$D_{t+\Delta t} \leqslant (1 + \Delta t\, b_{\min})^{-3} D_t$$

Proof. — For short, we put $B = B_{\mathcal{W}_t}(x_t)$ and $B_1 = B_{\mathcal{W}_{t+\Delta t}}(x_{t+\Delta t})$. Similarly, we define B' and B_1' at the points x_t' and $x_{t+\Delta t}'$. Now, if A_1 and A_2 are two invertible linear operators acting in the same space, then obviously

$$A_1 - A_2 = -A_1(A_1^{-1} - A_2^{-1})A_2 \tag{5.18}$$

Applying this trick twice and using (2.2) yields

$$Q^{-1}B_1'Q - B_1 = Q^{-1}(I + \Delta t\, B')^{-1} Q\, [Q^{-1}B'Q - B]\, (I + \Delta t\, B)^{-1}$$

Now the sublemma easily follows, with the help of (5.7) and (5.8). □

Sublemma 5.4. — *If there is a collision in a time interval $(t, t + \tau_{\min}/4)$, then*

$$D_{t+\tau_{\min}/2} \leqslant D_t + \overline{D}$$

where $\overline{D} > 0$ is a global constant.

Proof. — Let $s = t + \tau_{\min}/2$. Note that there are no collisions in the interval $(t + \tau_{\min}/4, s)$. For short, we put $B = B_{\mathcal{W}_s}(x_s)$ and $B' = B_{\mathcal{W}_s}(x_s')$. Denote by t_1 and t_1' the moments of reflection of the trajectories of the points x_t and x_t', respectively, that occur in the interval $(t, t + \tau_{\min}/4)$. Put $dt = t_1' - t_1$, $\tau = s - t_1$ and $\tau' = s - t_1'$. Note that $\tau > \tau_{\min}/4$ and $\tau' > \tau_{\min}/4$. Put $B^+ = B_{\mathcal{W}_{t_1+0}}(x_{t_1+0})$ and

$B'^+ = B_{\mathcal{W}_{t'_1+0}}(x_{t'_1+0})$. Let Q be the rotation of $\mathbb{R}^d$ that takes $v = v_s$ to $v' = v'_s$. It acts on $\mathcal{J} = \mathcal{J}_{x_s}$ by the rule (5.10). Applying the trick (5.18) twice yields

$$\begin{aligned} Q^{-1}B'Q - B &= -Q^{-1}B'Q(dt\, I)B \\ &\quad + Q^{-1}(I+\tau' B'^+)^{-1}Q\,[Q^{-1}B'^+Q - B^+]\,(I+\tau B^+)^{-1} \end{aligned} \tag{5.19}$$

Note that $\|B\| \leqslant 1/\tau \leqslant 4/\tau_{\min}$, and likewise $\|B'\| \leqslant 4/\tau_{\min}$. Hence,

$$\| - Q^{-1}B'Q(dt\, I)B\| \leqslant C|dt|$$

for a global constant $C > 0$. Next, we have $B^+ = UB^-U^{-1} + \Theta$ by (2.3), and, similarly $B'^+ = U'B'^-U'^{-1} + \Theta'$. Then we can further decompose the last term in (5.19):

$$\begin{aligned} \|Q^{-1}B'Q - B\| &\leqslant C|dt| + \|Q^{-1}U'B'^-U'^{-1}Q - UB^-U^{-1}\| \\ &\quad + \|Q^{-1}(I+\tau' B'^+)^{-1}Q\,[Q^{-1}\Theta' Q - \Theta]\,(I+\tau B^+)^{-1}\| \end{aligned}$$

Using Sublemma 5.1 (and its notation) gives, up to the first order in $\|\delta q\|$,

$$\begin{aligned} &\|Q^{-1}(I+\tau' B'^+)^{-1}Q\,[Q^{-1}\Theta' Q - \Theta]\,(I+\tau B^+)^{-1}\| \\ &\qquad = \|(I+\tau B^+)^{-1}[Q^{-1}\Theta' Q - \Theta]\,(I+\tau B^+)^{-1}\| \leqslant C\|\delta q\| \end{aligned}$$

Note that

$$\|Q^{-1}U'B'^-U'^{-1}Q - UB^-U^{-1}\| = \|Q_1^{-1}B'^-Q_1 - B^-\| \tag{5.20}$$

where $Q_1 = U'^{-1}QU$ is the rotator that takes the hyperplane $\mathcal{J}^- = \mathcal{J}_{x_{t_1-0}}$ to $\mathcal{J}'^- = \mathcal{J}_{x'_{t'_1-0}}$. We apply the trick (5.18) twice and act as in (5.19) and easily obtain

$$\|Q_1^{-1}B'^-Q_1 - B^-\| \leqslant \|B'^-\|\,|dt|\,\|B^-\| + \|Q_1^{-1}B'_1Q_1 - B_1\| \tag{5.21}$$

where $B_1 = B_{\mathcal{W}_t}(x_t)$ and $B'_1 = B_{\mathcal{W}_t}(x'_t)$.

Combining the above estimates gives

$$\|Q^{-1}B'Q - B\| \leqslant C|dt| + C\|\delta q\| + \|Q_1^{-1}B'_1Q_1 - B_1\|$$

for some global constant $C > 0$. Note that $dr_s = (I + \tau B^+)dr^+ = (I + \tau B^+)V^{-1}\delta q$, and due to (5.3) we have $\|\delta q\| \leqslant C\|dr_s\|$. Lastly, $|dt| \leqslant 2\|\delta q\|$ by (5.9) and $\|dr_t\| < \|dr_s\|$, which easily follows from (5.7). Therefore,

$$\|Q^{-1}B'Q - B\|/\|dr_s\| \leqslant \overline{D} + \|Q_1^{-1}B'_1Q_1 - B_1\|/\|dr_t\|$$

where $\overline{D}$ is a global constant, which proves the sublemma. □

We now complete the proof of Lemma 5.2. Let $t > 0$ satisfy the condition of the Lemma, and n be the number of collisions on the interval $(0, t)$. Then combining Sublemmas 5.3 and 5.4 gives

$$D_t \leqslant \lambda^n D_0 + (1 + \lambda + \cdots + \lambda^n)\overline{D}$$

where $\lambda = (1 + \tau_{\min} b_{\min}/4)^{-3} < 1$. Since $\overline{D}$ is a global constant, the Lemma follows. □

In all that follows we will only consider u-fronts $\mathcal{W}$ for which $D_{\mathcal{W}}(x) \leqslant D_{\max}$ for all $x \in \mathcal{W}$ provided the trajectory $S^t x$, $-\tau_{\min}/2 < t < 0$, does not collide with $\partial\mathbb{Q}$. As we are mainly interested in those u-manifolds that approximate LUM-s, this convention is justified by Lemma 5.2. Indeed, if the front $\mathcal{W}$ corresponds to a LUM, than $S^{-t}\mathcal{W}$ is a divergent front for any $t > 0$.

Remark. — A useful estimate (5.21) obtained in the proof of Sublemma 5.4 can now be restated. Recall that $|dt| \leqslant 2\|\delta q\|$, $\|B'^{-}\| \cdot \|B^{-}\| \leqslant 1/\tau_{\min}^2$ (a global bound) and

$$\|Q_1^{-1} B_1' Q_1 - B_1\| \leqslant D_{\max} \|dr_t\|$$

by the above convention. Also note that $\|dr_t\| \leqslant \|dr^-\| = \|dr^+\| = \|V^{-1}\delta q\| \leqslant \|\delta q\|$. Hence,

$$\|Q_1^{-1} B'^{-} Q_1 - B^{-}\| \leqslant C\|dr\| \tag{5.22}$$

with a global constant $C > 0$.

Finally we should prove the *curvature bounds on u-manifolds W in the Poincaré phase space*, in other words, that the 'derivative' of F along u-manifolds is uniformly bounded.

We will denote by $\operatorname{dist}_W(x, y)$ the distance between $x, y \in W$ in the Euclidean metric on W. Let $x = (q, v)$ and $x' = (q', v')$ be two infinitesimally close points of a u-manifold W, and F and F' the corresponding operators at x and x'. Using our previous notation, we consider the increment of F defined by $Q^{-1}F'Q_0 - F$. Here again Q_0 is the rotator taking $n = n(q)$ to $n' = n(q')$ and Q is the rotator taking v to v'.

***Theorem 5.5* (Curvature bounds - II)**. — *There is a global constant $C > 0$ such that*

$$\|Q^{-1}F'Q_0 - F\| \leqslant C\,\|\delta q\|$$

Proof. — Using the second formula in (4.4) and our earlier notation $\widetilde{V}^* = \langle v, n\rangle V^*$ gives

$$\begin{aligned}\|Q^{-1}F'Q_0 - F\| \leqslant\ & \|Q^{-1}\widetilde{V}'^{*}Q_0Q_0^{-1}K'Q_0 - \widetilde{V}^{*}K\| \\ & + \|Q^{-1}U'B'^{-}U'^{-1}QQ^{-1}V'^{-1}Q_0 - UB^{-}U^{-1}V^{-1}\|\end{aligned}$$

The first term is bounded by $C\,\|\delta q\|$ for some global constant $C > 0$, according to our earlier estimates (5.15) and (5.16). To bound the second term we need two more estimates. One is

$$\|Q^{-1}V'^{-1}Q_0 - V^{-1}\| \leqslant 4\,\|dv\| + 2\,\|dn\| \leqslant C\|\delta q\| \tag{5.23}$$

which is proved just like (5.14) and (5.15), we omit the details. The other is

$$\|Q^{-1}U'B'^{-}U'^{-1}Q - UB^{-}U^{-1}\| \leqslant C\|\delta q\| \tag{5.24}$$

for a global constant $C > 0$. In the proof of Sublemma 5.4 we introduced the rotator $Q_1 = U'^{-1}QU$ that takes the hyperplane $\mathcal{J}^-$ to $\mathcal{J}'^-$. With this, (5.24) is simply equivalent to our early estimate (5.22). Theorem 5.5 is now proved. □

5.2. Distorsion bounds. — This subsection is devoted to the question, how smoothly the volume expansion rates vary at nearby points on the same u-manifold (distorsion bounds) and at different u-manifolds joint by holonomy maps along s-manifolds (absolute continuity). Actually, the reason for introducing homogeneity strips and secondary singularities (see (4.3)) is that we would like to control these distorsions. Let us consider the evolution under T^n of a u-manifold W. Due to (4.11) distances grow exponentially in n, and the same is true for the $(d-1)$-dimensional volume of T^nW. However, at almost grazing reflections, when $\langle v, n\rangle \approx 0$, the expansion of u-manifolds is highly nonuniform, and so distortions are unbounded. Nevertheless, as we shall prove in Theorem 5.7, the situation is much better with homogeneous u-manifolds.

Throughout the subsection all metric quantities (norms, distances, volume elements, Jacobians) are understood in the e-metric, thus we often drop the sub- or superscripts e.

Sublemma 5.6. — *If W is a homogeneous u-manifold, then for any two points $x = (q, v)$ and $\overline{x} = (\overline{q}, \overline{v})$ of W we have*

$$|\langle \overline{v}, \overline{n}\rangle - \langle v, n\rangle| \leqslant C \langle v, n\rangle \left[\mathrm{dist}_W(x, \overline{x})\right]^{1/3}$$

where $\overline{n} = n(\overline{q})$ and $C > 0$ is a global constant.

Proof. — Let $W \cap I_k \neq \varnothing$ for some k. Then

$$|\langle \overline{v}, \overline{n}\rangle - \langle v, n\rangle| \leqslant C_1 (k+1)^{-3} \tag{5.25}$$

with a global constant C_1, according to our construction of I_k. Next, for any point $x' = (q', v')$ infinitesimally close to x, we have, up to the first order in $\|\delta x\| (= \|\delta x\|_e)$,

$$|\langle v', n'\rangle - \langle v, n\rangle| = |\langle dv, n\rangle + \langle v, dn\rangle| \leqslant C_2\|\delta q\| \leqslant C_3\|\delta x\| \tag{5.26}$$

with some global constants C_2, C_3, see (5.9) and Corollary 4.4. Integrating (5.26) from x to $\overline{x}$ yields

$$|\langle \overline{v}, \overline{n}\rangle - \langle v, n\rangle| \leqslant C_3 \,\mathrm{dist}(x, \overline{x}) \tag{5.27}$$

Now (5.25) and (5.27) give

$$|\langle \overline{v}, \overline{n}\rangle - \langle v, n\rangle|^3 \leqslant C_1^2 C_3 (k+1)^{-6} \,\mathrm{dist}(x, \overline{x})$$

Lastly, recall that $\langle v, n\rangle \geqslant (k+1)^{-2}$ if $k > 0$ and $\langle v, n\rangle \geqslant k_0^{-2}$ if $k = 0$, hence $\langle v, n\rangle \geqslant k_0^{-2}(k+1)^{-2}$ for any k. Therefore,

$$|\langle \overline{v}, \overline{n}\rangle - \langle v, n\rangle|^3 \leqslant C_1^2 C_3 k_0^6 \langle v, n\rangle^3 \,\mathrm{dist}(x, \overline{x})$$

This proves the sublemma. □

Let W be a u-manifold, $x \in W$ and T^n continuous at x. Denote by $J_{W,n}(x)$ the expansion factor of the $(d-1)$-dimensional volume of the manifold W under T^n at the point x, i.e. $J_{W,n}(x) := |\det DT^n|_W(x)|$.

***Theorem 5.7* (Distorsion bounds).** — *Let W be a small u-manifold on which T^n is continuous. Assume that $W_i := T^i W$ is a homogeneous u-manifold for each $0 \leqslant i \leqslant n$. Then for all $x, \overline{x} \in W$*

$$|\ln J_{W,n}(\overline{x}) - \ln J_{W,n}(x)| \leqslant C \cdot \Big[\operatorname{dist}_{W_n}(T^n x, T^n \overline{x})\Big]^{1/3}$$

for a global constant $C > 0$.

Proof. — Note that $J_{W,n}(x) = \prod_{i=0}^{n-1} J_{W_i,1}(T^i x)$. Hence, it is enough to prove the lemma for $n = 1$, because $\operatorname{dist}(T^i x, T^i \overline{x})$ grows uniformly exponentially in i due to (4.11). So we put $n = 1$.

Denote $x_1 = Tx$ and $\overline{x}_1 = T\overline{x}$. We will also use a variable point $x' \in W$ infinitesimally close to x, and put $x'_1 = Tx'$. For convenience, we will use the subscript 1 to denote quantities (including operators, hyperplanes, etc.) related to the points $x_1, \overline{x}_1$ and x'_1. In a similar way, bars are used to denote quantities related to the points $\overline{x}$ and $\overline{x}_1$, and primes are used for quantities related to x' and x'_1. For example, we denote by B^+, $\overline{B}^+$ and B'^+ the second fundamental forms of the wave front (corresponding to the u-manifold W) at the points x, $\overline{x}$, and x', respectively. Similarly, F, $\overline{F}$, and F' denote the F operator (4.4) taken at x, $\overline{x}$ and x', respectively. In a similar way, F_1, $\overline{F}_1$, and F'_1 are the F operators taken at x_1, $\overline{x}_1$ and x'_1, respectively, etc.

Note that the basic quantity, $J_{W,1}(x)$ was already calculated as $J^e_W(x)$ in the previous section (formula (4.13)) where we also introduced the operator G. In view of this formula, to prove Theorem 5.7 with $n = 1$, it is now enough to prove three claims:

Claim 1. — $|\ln\det\overline{V} - \ln\det V| \leqslant C \cdot \Big[\operatorname{dist}_W(x, \overline{x})\Big]^{1/3}$.

Claim 2. — $|\ln\det\overline{G} - \ln\det G| \leqslant C \cdot \Big[\operatorname{dist}_W(x, \overline{x})\Big]$.

Claim 3. — $|\ln\det(I + \overline{\tau}\overline{B}^+) - \ln\det(I + \tau B^+)| \leqslant C \cdot \Big[\operatorname{dist}_{TW}(x_1, \overline{x}_1)\Big]^{1/3}$.

By C we denote some global constants. Indeed, the bounds in Claims 1 and 2 will also hold at the points x_1 and $\overline{x}_1$, because TW is a homogeneous u-manifold, and Theorem 5.7 will then easily follow.

Proof of Claim 1. — Since $\det V = \langle v, n\rangle^{-1}$, the claim is a direct consequence of Sublemma 5.6.

Our proofs of Claims 2 and 3 use the following

Sublemma 5.8. — *Let A be an invertible linear operator in an m-dimensional space, and ΔA an infinitesimal operator. Then, up to the first order of $\|\Delta A\|$,*

$$|\ln\det(A+\Delta A)-\ln\det A| = |\operatorname{tr}(A^{-1}\cdot\Delta A)| \leqslant m\,\|A^{-1}\cdot\Delta A\|$$

Proof. — We have $\ln\det(A+\Delta A)=\ln\det A+\ln\det(I+A^{-1}\cdot\Delta A)$, and the rest is straightforward. □

Proof of Claim 2. — It is enough to prove

$$|\ln\det G'-\ln\det G| \leqslant C\,\|\delta x\| \tag{5.28}$$

for infinitesimally close points $x, x' \in W$, then the integration from x to $\overline{x}$ will give the bound in Claim 2.

As to the value of $\det G$, we refer to formula (4.14). Now, by Sublemma 5.8, we have

$$\begin{aligned}|\ln\det G'-\ln\det G| &\leqslant |\ln\det(I+F'^{*}F')-\det(I+F^{*}F)|\\ &= |\ln\det(I+Q_0^{-1}F'^{*}F'Q_0)-\det(I+F^{*}F)|\\ &\leqslant (d-1)\,\|(I+F^{*}F)^{-1}(Q_0^{-1}F'^{*}F'Q_0-F^{*}F)\|\end{aligned}$$

(the introduction of Q_0 defined by (4.1) was necessary to ensure that both operators act in the same space). It is obvious that $\|(I+F^{*}F)^{-1}\| \leqslant 1$, and by Corollary 4.4 and Theorem 5.5 we have

$$\|Q_0^{-1}F'^{*}F'Q_0-F^{*}F\| \leqslant C\,\|dr\|$$

This proves (5.28), and so Claim 2 is proved.

Proof of Claim 3. — To shorten some formulas, we put $R=I+\tau B^{+}$ (and, respectively, define $\overline{R}$ and R' at the points $\overline{x}$ and x'). It will be enough to prove that

$$|\ln\det R'-\ln\det R| \leqslant C|\langle v',n'\rangle-\langle v,n\rangle|\,\langle v,n\rangle^{-1}+C\,\|\delta x\|+C\,\|\delta x_1\| \tag{5.29}$$

for infinitesimally close points $x, x' \in W$. Note that $\|\delta x\| \leqslant C\|\delta x_1\|$ by (4.11). Then the integration of (5.29) from x to $\overline{x}$ (and, respectively, from x_1 to $\overline{x}_1$) will give

$$|\ln\det\overline{R}-\ln\det R| \leqslant C|\langle\overline{v},\overline{n}\rangle-\langle v,n\rangle|\,\langle v,n\rangle^{-1}+C\left[\operatorname{dist}_{TW}(x_1,\overline{x}_1)\right]$$

After that Claim 3 will follow by Sublemma 5.6.

We now prove (5.29). By Sublemma 5.8 we have, to the first order in $\|\delta x\|$,

$$\begin{aligned}\ln\det R'-\ln\det R &= \ln\det Q^{-1}R'Q-\ln\det R\\ &= \operatorname{tr}\left[R^{-1}(\tau'Q^{-1}B'^{+}Q-\tau B^{+})\right]\end{aligned} \tag{5.30}$$

(the introduction of Q defined by (5.10) was necessary to ensure that both operators act in the same space). Note that $\|R^{-1}\| \leqslant C$ by (5.8). Next, we have, again to the first order in $\|\delta x\|$,

$$\begin{aligned}\tau'Q^{-1}B'^{+}Q-\tau B^{+} &= d\tau\,B^{+}+\tau(Q^{-1}U'B'^{-}U'^{-1}Q-UB^{-}U^{-1})\\ &\quad+\tau(Q^{-1}\Theta'Q-\Theta)\end{aligned} \tag{5.31}$$

Observe that

$$\|VR^{-1}\| \leqslant C \qquad \text{and} \qquad \|R^{-1}V^*\| \leqslant C \tag{5.32}$$

according to (5.3) and (5.4). Using (2.3) now yields

$$\|R^{-1}B^+\| \leqslant \|R^{-1}\|\,\|B^-\| + 2\,\|R^{-1}V^*K\widetilde{V}\| \leqslant C \tag{5.33}$$

Now recall that $|d\tau| \leqslant 2\|\delta q\| + 2\|\delta q_1\|$ by (5.9). Hence we have, by (5.33),

$$|\operatorname{tr}(d\tau\, R^{-1}B^+)\,| \leqslant (d-1)\,|d\tau|\,\|R^{-1}B^+\| \leqslant C(\|\delta q\| + \|\delta q_1\|)$$

so the first term in the right hand side of (5.31) is properly taken care of.

Denote $\Delta B^- = Q^{-1}U'B'^-U'^{-1}Q - UB^-U^{-1}$. We then have, using (5.20) and (5.22),

$$\begin{aligned}|\operatorname{tr}(\tau\, R^{-1}\,\Delta B^-)\,| &\leqslant (d-1)|\tau|\,\|R^{-1}\,\Delta B^-\| \\ &\leqslant \tau_{\max}\|R^{-1}\|\,\|Q_1^{-1}B'^-Q_1 - B^-\| \\ &\leqslant C\,\|\delta q\|\end{aligned}$$

which takes care of the second term in (5.31).

Lastly, we use (5.17) to handle the third term in (5.31):

$$\begin{aligned}|\operatorname{tr}(R^{-1}(Q^{-1}\Theta'Q - \Theta))\,| \leqslant{} & 2\,|\langle v',n'\rangle - \langle v,n\rangle|\,|\operatorname{tr}(R^{-1}V^*KV)| \\ & +2\,|\operatorname{tr}(R^{-1}\Delta\widetilde{V}^*KV)| + 2\,|\operatorname{tr}(R^{-1}V^*\Delta K\widetilde{V})| \\ & +2\,|\operatorname{tr}(R^{-1}V^*K\Delta\widetilde{V})|\end{aligned} \tag{5.34}$$

We note that

$$\operatorname{tr}(R^{-1}\Delta\widetilde{V}^*KV) = \operatorname{tr}(\Delta\widetilde{V}^*KVR^{-1}) = \operatorname{tr}(R^{-1}V^*K\Delta\widetilde{V})$$

where the first equation follows from a general formula $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ in linear algebra, and the second is due to the fact that the operators $\Delta\widetilde{V}^*KVR^{-1}$ and $R^{-1}V^*K\Delta\widetilde{V}$ are adjoint to each other. Using this observation, we can rewrite (5.34) as

$$\begin{aligned}|\operatorname{tr}(R^{-1}(Q^{-1}\Theta'Q - \Theta))\,| \leqslant{} & C\,|\langle v',n'\rangle - \langle v,n\rangle|\,\langle v,n\rangle^{-1}\,\|R^{-1}V^*K\widetilde{V}\| \\ & +C\,\|\Delta\widetilde{V}^*KVR^{-1}\| + C\,\|R^{-1}V^*\Delta K\widetilde{V}\|\end{aligned}$$

We now apply (5.32) and (5.15)-(5.16) with (5.9) and obtain

$$|\operatorname{tr}(R^{-1}(Q^{-1}\Theta'Q - \Theta))\,| \leqslant C|\langle v',n'\rangle - \langle v,n\rangle|\,\langle v,n\rangle^{-1} + C\,\|\delta x\|$$

This completes the proof of (5.29) and hence Claim 3. Theorem 5.7 is now proved. □

After proving that the expansion factors vary nicely between nearby points on the same u-manifold, we now investigate their behaviour at points of different u-manifolds that lie on the same s-manifold. This is the absolute continuity property. Just like it was with the distorsion bounds, it is important to consider homogeneous manifolds.

***Theorem 5.9* (Absolute continuity).** — *Let W_s be a small s-manifold, $x, \overline{x} \in W_s$, and $W_u, \overline{W}_u$ two u-manifolds crossing W_s at x and $\overline{x}$, respectively. Assume that T^k is continuous on W_s and $T^i W_s$ is a homogeneous s-manifold for each $0 \leqslant i \leqslant k$. Then*

$$|\ln J_{W_u,k}(x) - \ln J_{\overline{W}_u,k}(\overline{x})| \leqslant C$$

where C is a global constant.

Proof. — For any $z \in W_s$, let $J_{W_s,k}(z)$ be the volume expansion factor of W_s under T^k at the point z, i.e. $J_{W_s,k}(z) = |\det DT^k|_{W_s}(z)|$. By the analogue of Theorem 5.7 for homogeneous s-manifolds, we have

$$|\ln J_{W_s,k}(x) - \ln J_{W_s,k}(\overline{x})| \leqslant C' \tag{5.35}$$

for a global constant C'.

Let $|DT^k(x)|$ denote the Jacobian of T^k at a point $x = (q, v) \in M$ with respect to the Lebesgue measure $\delta q\, \delta v$ on M in our local coordinates (q, v). Note that the T-invariant measure is $d\nu = \langle v, n\rangle\, \delta q\, \delta v$. Hence, $|DT^k(x)| = \langle v, n\rangle / \langle v_k, n_k\rangle$ where $x_k = (q_k, v_k) = T^k x$ and $n_k = n(q_k)$. Similarly, $|DT^k(\overline{x})| = \langle \overline{v}, \overline{n}\rangle / \langle \overline{v}_k, \overline{n}_k\rangle$, where the notation is quite clear. Since both W_s and $T^k W_s$ are small homogeneous s-manifolds, Sublemma 5.6 implies that the quantity $\langle v, n\rangle$ does not vary much over either W_s or $T^k W_s$. In fact, $c < \langle v, n\rangle / \langle \overline{v}, \overline{n}\rangle < C$ and $c < \langle v_k, n_k\rangle / \langle \overline{v}_k, \overline{n}_k\rangle < C$ for global constants $C > c > 0$. Hence,

$$0 < c < |DT^k(x)| / |DT^k(\overline{x})| < C < \infty \tag{5.36}$$

for some global constants c and C. Now Theorem 5.9 follows easily from (5.35), (5.36), and Theorem 4.7. □

6. Outlook

The results of this paper can be summarized as follows. We have some bad news (non-smooth behaviour) related to the singularity submanifolds in multi-dimensional hyperbolic billiards. On the other hand, there are important good news related to the u-manifolds in the multi-dimensional dispersing case. It is proved that practically all important regularity properties (uniform hyperbolicity, alignment, curvature and distorsion bounds) are just as valid as they are in the multi-dimensional case (cf. Remark 4.6).

In billiard theory one is mainly interested in the ergodic and statistical properties of the dynamical system. We emphasize that the above results are highly relevant to these issues. As to the ergodic properties, a major breakthrough was achieved with the proof of the Fundamental (or Local Ergodicity) Theorem ([**SCh, KSSz**]). However, for some measure theoretic estimates, the original arguments in these papers implicitly assumed uniform curvature bounds on the singularities. Thus these proofs have to be checked. In a separate paper ([**BChSzT**]) we will show that — at least, for billiards with algebraic scatterers — the original proofs of local ergodicity remain valid if some suitable modifications are performed.

Much less is known about statistical properties. As to the multi-dimensional dispersing case, no optimal result (exponential decay of correlations) has been achieved so far. Nevertheless, we conjecture that the rate of mixing is, indeed, exponential. The recently developed method of Markov-returns ([**Y1**]) turned out to be especially powerful in the study of decay rates for planar billiards ([**Ch2, Ch3**]). It is the growth of unstable manifolds that is to be investigated for Young's method to work. Essentially all important features of unstable manifolds have been checked in sections 4 and 5 to control growth of LUMs, the only thing we do not know yet how to handle is the irregular behaviour of singularities. We conjecture that, given a systematic geometric characterization of singularities, exponential decay of correlations for multi-dimensional dispersing billiards could be proved.

Acknowledgements. — The authors express their sincere gratitude to Nándor Simányi and András Szűcs for illuminating discussions. Thanks are due to the referees for their careful reading of the manuscript and for valuable remarks.

References

[AGV] V.I. Arnold, S.M. Gusein-Zade, A.N. Varchenko; *Singularities of differentiable maps I.*, Monographs in Mathematics 82, Birkhauser Boston, 1985

[BChSzT] P. Bálint, N.I. Chernov, D. Szász and I.P. Tóth, *Multi-dimensional Semi-dispersing Billiards: Singularities and the Fundamental Theorem*, Ann. Henri Poincaré **3** (2002), 451–482.

[Bl] P. Bleher, *Statistical properties of two-dimensional periodic Lorentz gas with infinite horizon*, J. Statist. Phys. **66** (1992), 315–373.

[Bo] L. Boltzmann, *Über die Eigenschaften monocyklischer und anderer damit verwandter Systeme*, Kronecker J. = J. Reine Angew. Math. **98** (1884) 68–94, Jahrbuch Fortschritte Math. **16** 863

[BSC1] L.A. Bunimovich, Ya.G. Sinai and N.I. Chernov, *Markov partitions for two-dimensional billiards*, Russ. Math. Surv. **45** (1990), 105–152.

[BSC2] L.A. Bunimovich, Ya.G. Sinai and N.I. Chernov, *Statistical properties of two-dimensional hyperbolic billiards*, Russ. Math. Surv. **46** (1991), 47–106.

[BR] L.A. Bunimovich and J. Reháček, *How high dimensional stadia look like*, Commun. Math. Phys. **197** (1998) 277–301.

[CELS1] N.I. Chernov, G.L. Eyink, J.L. Lebowitz and Ya.G. Sinai, *Steady-state electrical conduction in the periodic Lorentz gas*, Comm. Math. Phys. **154** (1993), 569–601.

[CELS2] N.I. Chernov, G.L. Eyink, J.L. Lebowitz and Ya.G. Sinai, *Derivation of Ohm's law in a deterministic mechanical model*, Phys. Rev. Let. **70** (1993), 2209–2212.

[Ch1] N. Chernov, *Statistical properties of the periodic Lorentz gas. Multidimensional case.* J. Statist. Phys. **74** (1994), 11–53

[Ch2] N. Chernov, *Decay of correlations and dispersing billiards*, J. Statist. Phys. **94** (1999), 513–556.

[Ch3] N. Chernov, *Sinai billiards under small external forces*, Ann. Henri Poincaré **2** (2001), 197–236.

[Co] J.P. Conze, Sur un critère de récurrence en dimension 2 pour les marches stationnaires, applications, Ergod. Th. Dynam. Sys. **19** (1999), 1233–1245.

[DM] C.P. Dettmann and G.P. Morriss, *Crisis in the periodic Lorentz gas*, Phys. Rev. E. **54** (1996), 4782–4790.

[GO] G. Gallavotti and D. Ornstein, *Billiards and Bernoulli schemes*, Comm. Math. Phys. **38** (1974), 83–101.

[GC] G. Gallavotti and E.D.G. Cohen, *Dynamical ensembles in stationary states*, J. Stat. Phys. **80** (1995), 931–970.

[KSSz] A. Krámli, N. Simányi and D. Szász, *A "Transversal" Fundamental Theorem for Semi-Dispersing Billiards*, Comm. Math. Phys. **129** (1990), 535–560.

[KSz] A. Krámli and D. Szász, *The problem of recurrence for the Lorentz process*, Commun. Math. Phys. **98** (1985), 539–552.

[Ru] D. Ruelle, *Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics*, J. Statist. Phys. **94** (1999)

[Schm] K. Schmidt, *On joint recurrence*, C.R. Acad. Sci. Paris Séc. I. Math. **327** (1998) 837–842

[Si] Ya.G. Sinai, *Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards*, Russ. Math. Surv. **25** (1970), 137–189.

[SCh] Ya.G. Sinai and N. Chernov, *Ergodic Properties of Certain Systems of 2–D Discs and 3–D Balls.*, Russain Mathematical Surveys, **(3) 42**, 181–201, (1987).

[Y1] L.-S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, Annals of Math. **147** (1998), 585–650.

[Y2] L.-S. Young, *Ergodic theory of chaotic dynamical systems*, XIIth International Congress of Mathematical Physics (ICMP'97) (Brisbane), 131–143, Internat. Press, Cambridge, MA, 1999.

P. BÁLINT, Alfréd Rényi Institute of the H.A.S., H-1053 Reáltanoda u. 13-15, Budapest, Hungary
E-mail : `bp@renyi.hu` • *Url :* `http://www.renyi.hu/~bp`

N. CHERNOV, Department of Mathematics, University of Alabama at Birmingham, Birmingham, AL 35294, USA • *E-mail :* `chernov@math.uab.edu`
Url : `http://www.math.uab.edu/chernov`

D. SZÁSZ, Mathematical Institute, Technical University of Budapest, H-1111 Egry József u. 1, Budapest, Hungary • *E-mail :* `szasz@math.bme.hu`
Url : `http://www.math.bme.hu/~szasz`

I.P. TÓTH, Mathematical Institute, Technical University of Budapest, H-1111 Egry József u. 1, Budapest, Hungary • *E-mail :* `mogy@math.bme.hu` • *Url :* `http://www.math.bme.hu/~mogy`

Astérisque
286, 2003, p. 151–165

HOMOCLINIC ORBITS NEAR SADDLE-CENTER FIXED POINTS OF HAMILTONIAN SYSTEMS WITH TWO DEGREES OF FREEDOM

by

Patrick Bernard, Clodoaldo Grotta Ragazzo & Pedro A. Santoro Salomão

To Jacob Palis for his 60^{th} Birthday

Abstract. — We study a class of Hamiltonian systems on a 4 dimensional symplectic manifold which have a saddle-center fixed point and satisfy the following property: All the periodic orbits in the center manifold of the fixed point have an orbit homoclinic to them, although the fixed point itself does not. In addition, we prove that these systems have a chaotic behavior in the neighborhood of the energy shell of the fixed point.

Introduction

A fixed point of a Hamiltonian system with two degrees of freedom is called a Saddle-Center if the linearized vector field has one pair of purely imaginary eigenvalues and one pair of non zero real eigenvalues. A saddle-center fixed point is surrounded by a two-dimensional invariant manifold, the center manifold, filled by closed orbits. A saddle-center fixed point has also a one-dimensional stable manifold and a one-dimensional unstable manifold; the periodic orbits in the center manifold have two-dimensional stable and unstable manifolds. If a point belongs to the intersection of the stable and unstable manifold of the fixed point (resp. of one periodic orbit) then its orbit is biasymptotic to the fixed point (resp. the periodic orbit). We call such an orbit homoclinic.

Some consequences of the existence of an orbit homoclinic to the fixed point have been investigated in [**5**], [**9**], [**7**], [**8**], [**11**], [**18**] (specially section 7.2) and other papers. It should be noted, however, that the existence of such a homoclinic is exceptional, in contrast to the case of hyperbolic fixed points. Dimensional considerations show that orbits homoclinic to the periodic motions of the center manifold are more likely

2000 ***Mathematics Subject Classification.*** — 37J45, 34C37.
Key words and phrases. — Hamiltonian systems, saddle-center fixed point, homoclinic orbits, integrability.

to exist. The existence of such homoclinics has been studied in [**4**], [**14**] (see also [**11**], [**9**], [**10**], [**7**], [**12**]) by perturbation methods, and in [**2**] by global methods. In these papers, orbits homoclinic to periodic orbits sufficiently far away from the fixed point are found.

In the present work, we study analytic perturbations of an integrable system with a homoclinic loop. We prove the following interesting behavior : Given any periodic orbit sufficiently close to the equilibrium in the center manifold, there exists an orbit homoclinic to it, although in general there does not exist any orbit homoclinic to the fixed point. This illustrates a question asked in [**2**].

In addition, topological entropy near the energy shell of the fixed point is obtained as a consequence of the presence of these homoclinics. More precisely, we prove that every neighborhood of the energy shell of the fixed point contains an energy shell with chaotic behavior on it. A similar result for reversible Hamiltonian systems is claimed, with no proof, in [**14**] pg 116. Other results in this direction under the hypothesis of the system being far from integrable can be found in [**9**], [**7**], [**13**].

Our method is semi-global and heavily relies on the low dimension: We first use the perturbative setting to prove the existence of quasiperiodic invariant tori confining the system in a neighborhood of the unperturbed homoclinic loop. We then reduce the problem to an area preservation argument on appropriate Poincaré return maps. It would of course be very interesting to obtain similar results by global methods and in higher dimension, in the spirit of [**2**], and to understand to what extent the phenomenon described here is general.

This paper emanated from a discussion between the authors after a talk of one of them at the international conference on dynamical systems dedicated to Jacob Palis. The authors would like to thank the organizers of that conference, who made that encounter possible. The first author learned a lot during his numerous conversations with Michel Herman, and was moved a lot by his sudden death.

1. Notations and results

1.1. Let M be a four-dimensional analytic manifold, endowed with a symplectic form Ω, and let

$$\begin{aligned} H : M \times I &\longrightarrow \mathbb{R}, \\ (x, \mu) &\longmapsto H(x, \mu) = H_\mu(x) \end{aligned}$$

be an analytic one-parameter family of Hamiltonians, where I is some interval containing 0 in its interior. In all this paper, we shall assume that the Hamiltonian system H_μ has a saddle-center fixed point r_μ for all $\mu \in I$, and that $H_\mu(r_\mu) = 0$. It is by now classical (see [**15**], [**17**], [**5**], [**14**], [**7**]), that the system H_μ is integrable in the neighborhood of the saddle-center r_μ. More precisely, there exist a neighborhood U of 0 in $\mathbb{R}^4$ and an analytic mapping $\phi : U \times I \to M$ such that ϕ_μ is a symplectic embedding for each μ, $\phi_\mu(0) = r_\mu$, and

$$H_\mu \circ \phi_\mu(q_1, p_1, q_2, p_2) = h(I_1, I_2, \mu),$$

where

$$I_1 = p_1 q_1, \quad I_2 = (p_2^2 + q_2^2)/2,$$

and the function h is analytic (one may have to reduce I). Furthermore, one can be reduced via a change in time-scale and a canonical transformation to the case where

$$\partial_{I_1} h(0,0,\mu) = -1 \quad \text{and} \quad \partial_{I_2} h(0,0,\mu) = \omega(\mu) > 0.$$

The functions I_1 and I_2 are preserved by the flow restricted to the local chart, this flow is determined by the equations

$$\begin{aligned} \dot{p}_1 &= -\partial_{I_1} h(I_1, I_2, \mu) p_1 & \dot{p}_2 &= -\partial_{I_2} h(I_1, I_2, \mu) q_2 \\ \dot{q}_1 &= \ \ \partial_{I_1} h(I_1, I_2, \mu) q_1 & \dot{q}_2 &= \ \ \partial_{I_2} h(I_1, I_2, \mu) p_2. \end{aligned}$$

It follows that the center manifold of r_μ has equation $I_1 = 0$, its stable manifold has equation $I_2 = 0$, $p_1 = 0$ and its unstable manifold $I_2 = 0$, $q_1 = 0$. In the following, we will call $P_{E,\mu}$ the periodic orbit of H_μ at energy E, which in local coordinates is the circle $p_1 = q_1 = 0$, $I_2 = E$.

1.2. We shall also suppose that H_0 is integrable (namely, its associated Hamiltonian vector field has an additional real analytic first integral J such that $dH_0(x)$ and $dJ(x)$ are independent for almost every x) and that the vector field associated to H_0 has an orbit homoclinic to r_0 which connects the branch $p_1 > 0$ of the unstable manifold to the branch $q_1 > 0$ of the stable manifold. Integrable systems with a saddle-center and an orbit doubly asymptotic to it have been studied in [**9**], where it is explained that there exist two different kinds of homoclinics. For comparison, let us mention that we are here in case (A) of [**9**].

1.3. Theorem. — *Let us consider an analytic one-parameter family H_μ of Hamiltonian systems satisfying the above hypotheses. There exists a positive number ε such that for all $E \in]0, \varepsilon[$ and all $\mu \in]-\varepsilon, \varepsilon[\subset I$, there exists an orbit of H_μ homoclinic to the periodic orbit $P_{E,\mu}$. In fact, there even exist infinitely many geometrically distinct orbits homoclinic to $P_{E,\mu}$.*

1.4. Theorem. — *Let us fix $\mu \in]-\varepsilon, \varepsilon[$. For each $E \in]0, \varepsilon[$, either the stable and unstable manifolds of $P_{E,\mu}$ coincide, or the flow of H_μ on the energy shell $H_\mu = E$ has positive topological entropy.*

1.5. Theorem. — *Let us fix a value of μ satisfying the hypothesis of theorem 1.3. Assume in addition that the stable and unstable manifolds of the fixed point r_μ do not coincide. Then there exists a sequence $E_n > 0$ converging to 0 and such that the stable and unstable manifolds of $P_{E_n,\mu}$ do not coincide. It follows that, for each n, the flow of H_μ restricted to the energy surface $H_\mu = E_n$ has positive topological entropy.*

1.6. The main result of the present paper is Theorem 1.3. It is proved in section 3. Theorem 1.4 may be considered classical. However we include a proof in section 4 because we could not find any reference matching precisely our needs. Theorem 1.5 is a simple but, we believe, interesting consequence. It is proved in section 5. The main notations and tools that will be used throughout the paper are introduced in section 2

1.7. Remark. — In order to apply Theorem 1.5, one has to be able to decide whether there exists an orbit homoclinic to the fixed point. Let us mention a result in that direction. Under an additional hypothesis of reversibility of the family of Hamiltonian systems H_μ (see [**7**]) it is possible to prove that the set of values of μ for which a homoclinic orbit to the equilibrium point r_μ occurs is either a whole interval or it is countable ([**7**], section 6). The same result may hold for the non reversible case considered here but this is an open question.

2. Local sections and invariant curves

We analyze the orbit structure near the homoclinic loop in a rather usual way (see [**5**], [**9**], [**14**],...), via Poincaré sections. More details in these papers. The existence of invariant curves was already obtained in [**8**].

2.1. Let us define the two Poincaré sections given in local coordinates by

$$\Sigma_1 = \{q_1 = \delta\}, \quad \Sigma_2 = \{p_1 = \delta\},$$

where δ is a small positive number. Since $\partial_{I_1} h = -1$, the equation $h(I_1, I_2, \mu) = E$ can be solved in I_1 for sufficiently small I_2, E and μ *i.e.* there exists an analytic function v defined in a neighborhood of 0 in $\mathbb{R}^3$ and such that

$$h(I_1, I_2, \mu) = E \iff I_1 = v(I_2, E, \mu).$$

As a consequence, for sufficiently small E and μ, the intersection $\Sigma_i(E,\mu)$ of Σ_i with the energy shell $H_\mu = E$ is a graph over the (p_2, q_2)-plane. More precisely, the analytic mappings $\sigma_i^{E,\mu} : \mathbb{R}^2 \to \mathbb{R}^4$ given by

$$\sigma_1^{E,\mu}(p_2, q_2) = \sigma_1(p_2, q_2, E, \mu) = \big(v\big(I_2(p_2, q_2), E, \mu\big)/\delta,\ \delta,\ p_2,\ q_2\big),$$
$$\sigma_2^{E,\mu}(p_2, q_2) = \sigma_2(p_2, q_2, E, \mu) = \big(\delta,\ v\big(I_2(p_2, q_2), E, \mu\big)/\delta,\ p_2,\ q_2\big)$$

are symplectic charts of $\Sigma_i(E,\mu)$. In the following, we note $y = (p_2, q_2)$ and take it as coordinates of $\Sigma_i(E,\mu)$.

2.2. The intersection between the stable manifold of $P_{E,\mu}$ and Σ_1, as well as the intersection between the unstable manifold and Σ_2, are the circles $I_2(y) = I^c(E,\mu)$ in coordinates, where $I^c(E,\mu)$ is the solution of the equation

$$h(0, I^c(E,\mu), \mu) = E \iff v(I^c(E,\mu), E, \mu) = 0.$$

The orbits starting in $\Sigma_1(E,\mu)$ outside of this circle hit $\Sigma_2(E,\mu)$ after a time

$$T(y,E,\mu)=t(I_2(y),E,\mu)=\frac{1}{\partial_{I_1}h\big(v(I_2(y),E,\mu),I_2(y),\mu\big)}\log\frac{v(I_2(y),E,\mu)}{\delta^2}.$$

Notice in the previous expression that $v(I_2(y),E,\mu)$ is positive if and only if y lies outside of the stable circle. The local transition map $l_{E,\mu}:\Sigma_1(E,\mu)\cap\{p_1>0\}\to\Sigma_2(E,\mu)$ is defined outside of the stable circle and can be computed in local coordinates

$$l_{E,\mu}(y)=l(y,E,\mu)=R\big(\theta(I_2(y),E,\mu)\big)\,y$$

where $R(\theta)$ is the matrix of the rotation of angle θ, and

$$\theta(I_2,E,\mu)=t(I_2)\,\partial_{I_2}h\big(v(I_2,E,\mu),I_2,\mu\big).$$

The outer transition map $g_{E,\mu}:\Sigma_2(E,\mu)\to\Sigma_1(E,\mu)$ is defined by following the flow along the homoclinic loop.

2.3. The following estimate will be useful (see [**7**]):

$$\theta(I_2,E,\mu)=-\omega(\mu)\log|I_2-I^c(E,\mu)|+\Lambda_{E,\mu}(I_2),\quad I_2>I^c(E,\mu)$$

where

$$I^c(E,\mu)=\frac{E}{\omega(\mu)}+O(E^2)$$

and where the function $I_2\longmapsto\Lambda_{E,\mu}(I_2)$ is analytic around $I^c(E,\mu)$ for each E and μ. To see this, just write $v(I_2,E,\mu)=(I_2-I_c(E,\mu))w(I_2,E,\mu)$, where w is analytic and $w(I_c,E,\mu)\neq 0$.

2.4. The local transition maps $l_{E,\mu}$ seen in coordinates as mappings of $\mathbb{R}^2$ preserve the circles centered at the origin. Since the unperturbed Hamiltonian H_0 is assumed to be integrable, the outer transition map $g_{0,0}$ also preserves these circles, hence this symplectic map can be written

$$g_{0,0}(y)=R(\psi(I_2(y)))\,y,$$

where ψ is a real map analytic in a neighborhood of 0. Let us now define the mapping

$$F_{E,\mu}=g_{E,\mu}\circ l_{E,\mu},$$

we have

$$F_{0,0}=R\big(\psi\circ I_2+\theta\circ I_2\big).$$

In view of the estimates of 2.3, it is possible to choose positive numbers $I^-<I^+$ such that $F_{0,0}$ is an integrable analytic twist area preserving diffeomorphism of the annulus $\mathcal{A}=\big\{y \text{ s.t. } I^-\leqslant I_2(y)\leqslant I^+\big\}$. For sufficiently small E and μ, $F_{E,\mu}$ is a two-parameter analytic family of exact area preserving diffeomorphisms between $\mathcal{A}$ and its image in $\mathbb{R}^2$. Here exact means that there exists a rotational Jordan curve C in the annulus $\mathcal{A}$ with the following property : The image $F_{E,\mu}(C)$ is also a rotational Jordan curve in $\mathcal{A}$ and the area of the domain between $\{I_2=I^-\}$ and $F_{E,\mu}(C)$ is

equal to the area of the domain between $\{I_2 = I^-\}$ and C. A direct application of KAM theorem now proves the following proposition.

2.5. Proposition. — *There exist positive numbers ε and I such that, for all $E \in]0, \varepsilon[$ and $\mu \in] - \varepsilon, \varepsilon[$, there exists an analytic rotational Jordan curve $C(E, \mu)$ contained in $\{y \text{ s.t. } I/2 \leqslant I_2(y) \leqslant I\}$ and invariant under $F_{E,\mu}$. Let us denote $C'(E, \mu) = l_{E,\mu}(C(E, \mu)) = g^{-1}_{E,\mu}(C(E, \mu))$.*

3. Homoclinic orbits and multiplicity

We now prove Theorem 1.3. We have to study the dynamics of the flow of H_μ on the energy surface $\{H_\mu = E\}$, where E and μ satisfy the hypotheses of Proposition 2.5. We will not mention any more the parameters E and μ.

3.1. The map $F = g \circ l$ has an invariant circle C. Let S be the intersection between the stable manifold and Σ_1, and S' be the intersection between the unstable manifold and Σ_2. Both S and S' are the circle $\{I_2(y) = I^c\}$ in coordinates. The local transition map l is defined in the open annulus A in Σ_1 enclosed between S and C, and takes values in the annulus A' of Σ_2 enclosed between S' and C'. The outer transition map g is defined and analytic in D', the open disk enclosed in C', and takes values in D, the open disk enclosed in C. We call B the closed disk bounded by S and B' the closed disk bounded by S'. Both l and g preserve area (see figure 1).

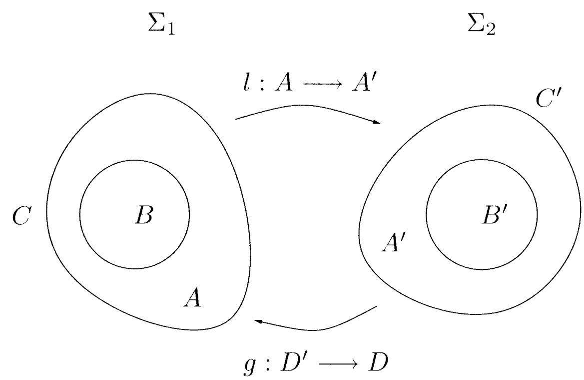


FIGURE 1. The mappings

3.2. The existence of a homoclinic is a consequence of the facts recalled above, as we shall see now. If $g(S')$ intersects S, then these intersection points are homoclinic points, since S is contained in the stable manifold on one hand, and S', hence $g(S')$, are contained in the unstable manifold on the other hand. We call such intersections 1-bump homoclinic points. If $g(S')$ does not intersect S, then $g(B')$ $(g^{-1}(B))$ is contained in A (A') since it can't be contained in B (B'), by area preservation. It follows that there exists a neighborhood U of B' such that $F \circ g$ is well defined in U.

3.3. Lemma. — *Suppose that for each $n \leqslant N-2$ the map $F^n \circ g$ is well defined in a neighborhood of B' and satisfies $F^n \circ g(S') \cap S = \varnothing$. Then $F^i \circ g(B') \cap F^j \circ g(B') = \varnothing$, for all $0 \leqslant i < j \leqslant N-2$.*

Proof. — The hypothesis $F^n \circ g(S') \cap S = \varnothing$ and the area preservation property of F and g imply that $F^n \circ g(B') \cap B = \varnothing$ for all $n \leqslant N-2$. We also have that $F^n \circ g(B') \cap g(B') = \varnothing$ for all $1 \leqslant n \leqslant N-2$. To prove this, we observe that the image $\mathrm{Im}(F)$ of the map F is $g(\mathrm{Im}(l)) = g(A')$. Since A' is disjoint from B', the image of F is disjoint from $g(B')$. Let us now take $0 \leqslant i < j \leqslant N-2$, we have $F^i \circ g(B') \cap F^j \circ g(B') = F^i\big(g(B') \cap F^{j-i} \circ g(B')\big) = F^i(\varnothing) = \varnothing$. □

3.4. Proposition. — *There exists an integer $N \geqslant 1$ satisfying the hypotheses of Lemma 3.3 and such that*

$$F^{N-1} \circ g(S') \cap S \neq \varnothing.$$

The intersection points seen as points of Σ_1, are homoclinic points, we call them N-bump homoclinic points. We have the following alternative: Either $F^{N-1} \circ g(S') = S$ and there are infinitely many N-bumps homoclinics, or $F^{N-1} \circ g(S') \not\subset S$ and there are infinitely many $2N$-bumps homoclinics.

Proof. — Since the annulus A has bounded area, and since all the domains $F^n \circ g(B')$ have the same positive area, only finitely many of them can be disjoint, hence the existence of N. It is quite clear that there exist infinitely many N-bumps homoclinic orbits in the case where $F^{N-1} \circ g(S') = S$. We shall now see that there exist infinitely many $2N$-bumps homoclinic points in the second case, *i.e.* if

$$F^{N-1} \circ g(S') \neq S.$$

***3.5. Definition* (see [1]).** — Let Δ be a compact topological disk in $\mathbb{R}^2$. We say that a continuous curve $\delta \subset \mathbb{R}^2 - \Delta$ has the obstruction property with respect to Δ if any continuous curve γ containing a point in Δ and a point outside Δ intersects the curve δ. It follows that any such curve γ must intersect δ infinitely many times.

Let us note $G = F^{N-1} \circ g$. In view of the estimates 2.3, and since $G(S')$ is not contained in B, the curve

$$\delta = l\big(G(S') \cap A\big)$$

has the obstruction property with respect to B'. It follows that the curve $G(\delta \cap \mathrm{dom}(G))$ has the obstruction property with respect to $G(B')$, where $\mathrm{dom}(G)$ is the domain of definition of G. We have supposed that $G(S')$ intersects S (and thus B), and that $G(S')$ is not S, hence is not contained in B, by area preservation. It follows from the obstruction property that $G(\delta)$ has to intersect S infinitely many times. We have proved that the set $G \circ l \circ G(S') = F^{2N-1} \circ g(S')$ has infinitely many points of intersection with S. These points clearly represent geometrically distinct $2N$-bumps homoclinics. □

4. Bernoulli shift

In order to prove Theorem 1.4, we are now going to build a Bernoulli shift. Our construction is quite similar to the one described in [**16**], chapter III, for the Sitnikov map. However, we only look for a semiconjugacy, instead of a conjugacy in [**16**]. This avoids many calculations and allows weaker hypotheses.

4.1. We use the notations of 3.1. The mapping $G = F^{N-1} \circ g$ is defined in a neighborhood of B'. We suppose that $G(S')$ and S are neither disjoint nor equal *i.e.* that there exists an N-bump homoclinic to the periodic orbit under interest, but that its stable and unstable manifolds do not coincide. The local transition map l is defined outside of B and satisfies the estimate of 2.3.

4.2. Under the hypotheses recalled in 4.1, the mapping $F^N = G \circ l$ has the Bernoulli shift as a topological factor. As a consequence, the mapping F has positive topological entropy, and there exist infinitely many kN-bump homoclinic orbits for all $k \geqslant 2$.

In order to be more explicit, let us consider the set $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$, endowed with the following topology : A subset $U \subset \overline{\mathbb{N}}$ is open if and only if either it does not contain ∞, or it contains the subset $\{n \in \mathbb{N}, \text{ s.t. } n \geqslant N\}$ for some $N \geqslant 1$. This is the classical compactification of $\mathbb{N}$. Let us consider the set $\overline{\Lambda}$ of the sequences $s \in \overline{\mathbb{N}}^{\mathbb{Z}}$ of the form

$$\ldots, \infty, \infty, s_{-m}, \ldots, s_0, \ldots, s_n, \infty, \infty, \ldots$$

with $\infty \geqslant m \geqslant -1$, $\infty \geqslant n \geqslant 0$, and $s_i < \infty$ for all $-m \leqslant i \leqslant n$. It has to be understood that $m = -1$ and $n = 0$ in the above expression stand for the sequence $\ldots, \infty, \infty, \ldots$. The set $\overline{\Lambda}$ is a compact subset of $\overline{\mathbb{N}}^{\mathbb{Z}}$ containing

$$\Lambda = \mathbb{N}^{\mathbb{Z}}.$$

In addition, Λ is dense in $\overline{\Lambda}$, which justifies the notations. The map $\lambda : \Lambda \to \Lambda$ is defined by $\lambda(s)_i = s_{i-1}$. Note that the continuous extension $\overline{\lambda}$ of λ to $\overline{\mathbb{N}}^{\mathbb{Z}}$ does not preserve $\overline{\Lambda}$.

We shall prove that there exist a compact set $\overline{X} \subset \overline{A}$, an invariant set X contained in $\overline{X}$ and dense in it, and a surjective continuous mapping $\overline{\tau} : \overline{X} \to \overline{\Lambda}$ satisfying $\overline{\tau}(X) = \Lambda$ and such that the diagram

$$\begin{tikzcd} X \arrow[r, "F^N"] \arrow[d, "\tau"'] & X \arrow[d, "\tau"] \\ \Lambda \arrow[r, "\lambda"] & \Lambda \end{tikzcd}$$

commutes, where τ is the restriction of $\overline{\tau}$ to X. In addition, the points of $\overline{\tau}^{-1}(s)$ are kN-bump homoclinic points when s is a sequence

$$\ldots, \infty, \infty, s_{-m}, \ldots, s_0, \ldots s_n, \infty, \infty, \ldots$$

with $\infty > m \geqslant 0$, $\infty > n \geqslant 0$, $k = m+n+2$ and $s_i < \infty$ for all $-m \leqslant i \leqslant n$. To finish this description, the preimage $\overline{\tau}^{-1}(\ldots,\infty,\infty,\ldots)$ consists of N-bump homoclinic points.

4.3. In order to prove the statements of 4.2, we shall introduce the notion of vertical and horizontal strips, following [**16**] for the main lines. However, as we already mentioned, we work under weaker hypotheses, and we will need more topological notions, in the spirit of works of Conley, Easton and McGehee, see for example [**6**]. See also [**3**] for related work. Let us consider the square Q as drawn in figure 2, where V_0 is the right edge, U_0 is the lower edge, V_∞ and U_∞ are the left and upper edges, and P is the vertex $V_\infty \cap U_\infty$. We shall also note Q any domain of the plane homeomorphic

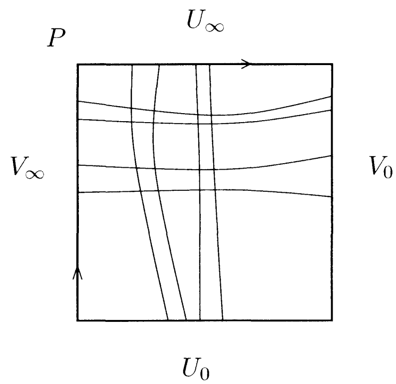


FIGURE 2. The square

to this square, and define the following distinguished subsets:

Definition. — A vertical strip is a compact subset V of Q such that $V \cup U_0 \cup U_\infty$ is connected. A horizontal strip is a compact subset U of Q such that $U \cup V_0 \cup V_\infty$ is connected.

Lemma. — *If V_i is a decreasing sequence of vertical strips, the intersection $\cap_i V_i$ is a vertical strip. The same holds for horizontal strips. A vertical strip and a horizontal strip have non empty intersection.*

Proposition. — *There exists a square Q in Σ_1 such that $U_\infty \subset G(S')$, $V_\infty \subset S$ (hence $P \subset G(S') \cap S$) and int$(Q) \cap (G(B') \cup B) = \varnothing$, where int$(Q)$ is the interior of Q. In this square Q, there exists a sequence U_i, $i \in \mathbb{N}$ of disjoint horizontal strips, and a sequence V_i, $i \in \mathbb{N}$ of disjoint vertical strips such that*

$$F^N(V_i) = U_i.$$

The strip U_{i+1} is above U_i in Q and U_i is converging to U_∞ for the Hausdorff metric. Seemingly, the strips V_i are ordered from the right to the left and converge to V_∞. In addition, we have the following property:
If V is a vertical strip, then each of the sets $F^{-N}(V) \cap V_i$ contains a vertical strip. If U is a horizontal strip, then each $F^N(U) \cap U_i$ contains a horizontal strip.

4.4. The structure described in 4.3 implies the existence of a Bernoulli shift as defined in 4.2. We shall prove this fact now, and delay the proof of Proposition 4.3 up to 4.5. We closely follow the presentation of [**16**], which may be consulted for more details. Let us consider a sequence $s_i \in \overline{\Lambda}$, and define the sets

$$V_{s_0 s_{-1} \cdots s_{-n}} = \bigcap_{i=0}^{j} F^{-iN}(V_{s_{-i}}),$$

where $j = n$ if $s_{-n} < \infty$, and $j = \min\{k \leqslant n, \text{ s.t. } s_{-k} = \infty\}$ otherwise. These sets are vertical strips, as can be proved by induction using Proposition 4.3 and noticing that

$$V_{s_0 s_{-1} \cdots s_{-n}} = V_{s_0 s_{-1} \cdots s_{-j}} = V_{s_0} \cap F^{-N}(V_{s_{-1} \cdots s_{-j}}).$$

In the same way, we define the horizontal strips

$$U_{s_1 \cdots s_n} = \bigcap_{i=1}^{j} F^{iN}(U_{s_i}),$$

where $j = n$ if $s_n < \infty$, and $j = \min\{k \leqslant n, \text{ s.t. } s_k = \infty\}$ otherwise. It follows from Lemma 4.3 that

$$V(s) = \bigcap_{n=0}^{\infty} V_{s_0 s_{-1} \cdots s_{-n}}$$

is a vertical strip, and that

$$U(s) = \bigcap_{n=1}^{\infty} U_{s_1 \cdots s_n}$$

is a horizontal strip. The set $V(s) \cap U(s)$ is thus a non empty compact set. If $s \in \Lambda$, we have

$$V(s) \cap U(s) = \{p \in Q \text{ s.t. } F^{-iN}(p) \in V_{s_i}\}.$$

We can now define the invariant set

$$X = \bigcup_{s \in \Lambda} V(s) \cap U(s),$$

the compact set

$$\overline{X} = \bigcup_{s \in \overline{\Lambda}} V(s) \cap U(s)$$

and the mapping $\overline{\tau}$ which, to each point of $V(s) \cap U(s)$, associates the sequence $s \in \overline{\Lambda}$. This mapping is well defined since the sets $V(s) \cap U(s)$ and $V(s') \cap U(s')$ are obviously disjoint for different sequences s and s'. It is straightforward with these definitions to check the statements of 4.2.

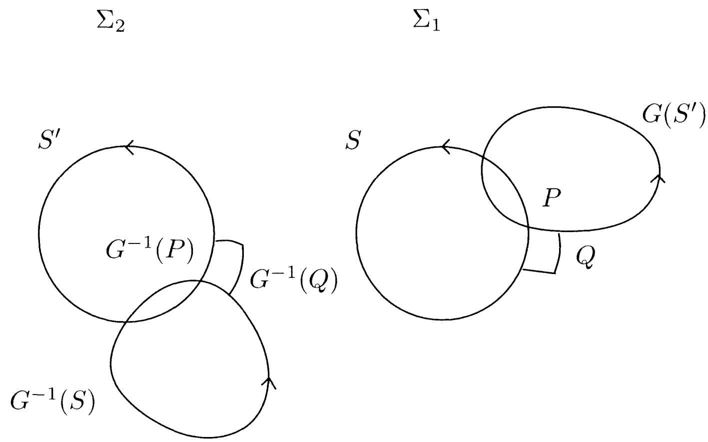


FIGURE 3. Construction of Q

4.5. In order to prove Proposition 4.3, we shall first build the square Q. Let us choose a point P of $G(S') \cap S$. There are two cases.

i. The curves $G(S')$ and S are outer tangent, *i.e.* $G(B') \cap B \subset S$ and we can take any point $P \in G(S') \cap S$.

ii. The curves $G(S')$ and S are crossing each other. In this case, we choose P such that the curves $G(S')$ and S locally cross each other at P.

In both cases, P is isolated in $G(S') \cap S$ since both curves are analytic. Let us consider the action-angle coordinates (I_2, θ) on Σ_2, defined by the relations

$$p_2 = \sqrt{2I_2}\cos\theta, \quad q_2 = \sqrt{2I_2}\sin\theta.$$

There exists a positive integer a, a positive real number δ and an analytic function $h : [I_c, I_c+\delta] \to \mathbb{R}$ such that the curve $(I_2, h((I_2-I_c)^{1/a}))$, $I_2 \in [I_c, I_c+\delta]$ is contained in $G^{-1}(S) \cap A$. Recall that the circle S' has the equation $I_2 = I_c$. In the case where P is a point of transversal intersection, we can take $a = 1$. It is possible to choose P, δ and h in such a way that the open set

$$\big\{I_c < I_2 < I_c + \delta, h((I_2 - I_c)^{1/a}) < \theta < h((I_2 - I_c)^{1/a}) + \delta\big\}$$

is disjoint from $G^{-1}(S)$. We then set

$$Q = G\big(\big\{I_c \leqslant I_2 \leqslant I_c + \delta, h((I_2 - I_c)^{1/a}) \leqslant \theta \leqslant h((I_2 - I_c)^{1/a}) + \delta\big\}\big).$$

We orient the curves S, S', $G(S')$ and $G^{-1}(S)$ positively, and give U_0 and U_∞ the induced orientation. In order to prove that Proposition 4.3 holds with this square Q, it is enough to prove the following proposition.

Proposition. — *Any sufficiently small neighborhood of U_∞ in Q contains a horizontal strip U which is the image by F^N of a vertical strip V of Q, and satisfies the following property : If $\widetilde{V}$ is a vertical strip of Q, then $F^{-N}(\widetilde{V}\cap U)\subset V$ contains a vertical strip of Q, and if $\widetilde{U}$ is a horizontal strip of Q, then $F^N(\widetilde{U}\cap V)\subset U$ contains a horizontal strip.*

Proof. — We need a Lemma.

Lemma. — *Let $c:[0,1]\to\overline{A}$, be an analytic curve such that $c(]0,1])\subset A$ and $c(0)\in S$. Then for ε small enough, the curve $F^N\circ c:\,]0,\varepsilon]\to A$ is an analytic spiral that accumulates on $G(S')$ and that crosses Q infinitely many times. Moreover, every connected component of $F^N\circ c(]0,\varepsilon])\cap Q$ crosses Q from V_0 to V_∞ (the orientation of $F^N\circ c$ is that defined by the parameterization).*

To prove this lemma we first write $F^N\circ c$ as $G\circ l\circ c$. Then using estimate 2.3 and recalling that l is explicitly given by (see 2.2)

$$l_{E,\mu}(y)=l(y,E,\mu)=R\big(\theta(I_2(y),E,\mu)\big)\,y$$

we conclude that $l\circ c$ is an infinite spiral turning monotonically around S and accumulating on S. In addition, easy explicit estimates show that, when ε is small enough, each connected component of $l(c(]0,\varepsilon]))\cap G^{-1}(Q)$ is crossing $G^{-1}(Q)$ from $G^{-1}(V_0)$ to $G^{-1}(V_\infty)$. The lemma follows from the fact that G is a local diffeomorphism in a neighborhood of P (see figure 4).

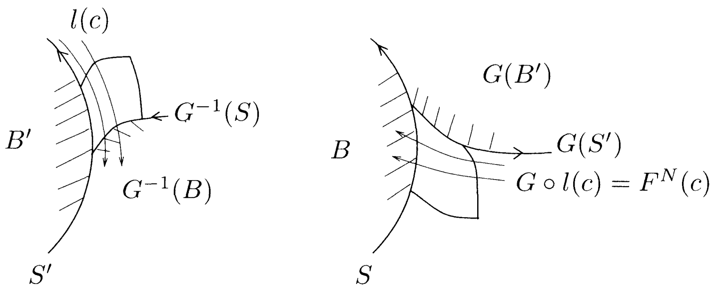


FIGURE 4. Spirals

This lemma implies that the set $F^N(Q)\cap Q$ has infinitely many connected components which are horizontal strips accumulating on U_∞. Each of these strips is bounded by two horizontal arcs, a lower and an upper one, which are contained in $F^N(U_\infty)$ and $F^N(U_0)$, respectively, and two small sub-arcs of V_0 and V_∞. Let U be one of these horizontal strips, sufficiently close to U_∞. Using that: $F^{-N}(U)$ is connected, $F^{-N}(U)\cap U_\infty\neq\varnothing$ and $F^{-N}(U)\cap U_0\neq\varnothing$, we conclude that $F^{-N}(U)=V$ is a vertical strip. In addition, we see that the vertical strip V is a topological square bounded on one side by a connected component of $F^{-N}(V_0)\cap Q$ crossing Q, and on the other side by a connected component of $F^{-N}(V_\infty)\cap Q$. To finish the proof of the proposition we need another lemma

Lemma. — *Let us consider a compact curve γ in Q connecting U_0 and U_∞. There exist connected components of $\gamma\cap U$ intersecting both $F^N(U_0)$ and $F^N(U_\infty)$.*

Proof. — To prove this fact, let us orient γ from U_∞ to U_0, and consider the last point of intersection of γ with $F^N(U_0)\cap U$. Just after this last intersection, γ lies inside U, hence has to leave U through $F^N(U_\infty)$. This proves the lemma. □

Let $\widetilde{V}$ be a vertical strip. It intersects U, by lemma 4.3. We are going to prove that $F^{-N}(\widetilde{V}\cap U)$ is a vertical strip. Assume that this is not true. In this case, the compact set $\widetilde{V}\cap U$ is disconnected, and is the union of two disjoint compact sets K_1 and K_2, where K_1 is the union of the connected components of $\widetilde{V}\cap U$ which intersect $F^N(U_0)$, and K_2 the union of those which intersect $F^N(U_\infty)$. We can find two disjoint open sets of Q, Ω_1 and Ω_2, containing respectively K_1 and K_2. In addition, since $F^N(U_0)\cap U$ and $F^N(U_\infty)\cap U$ are compact, we can choose Ω_1 and Ω_2 such that Ω_1 does not intersect $F^N(U_\infty)\cap U$ and Ω_2 does not intersect $F^N(U_0)\cap U$. The sets $U-(\Omega_1\cup\Omega_2)$ and $\widetilde{V}$ are compact and disjoint. It follows that one can find a connected open neighborhood Ω of $\widetilde{V}$ such that $\Omega\cap U\subset\Omega_1\cup\Omega_2$. The open set Ω contains a curve γ connecting U_0 and U_∞. Each connected component of $\gamma\cap U$ is contained either in Ω_1 or in Ω_2, which is in contradiction with the conclusion of the lemma. The intersection between V and horizontal strips can be studied exactly in the same way. □

5. Chaos near the energy shell of the fixed point

5.1. In this section, we fix a value of the parameter μ and work with a fixed Hamiltonian H. We suppose that the conditions of existence of invariant curves (see Proposition 2.5) is satisfied, hence there exists a critical energy $\eta>0$ such that, for all $E\in\,]0,\eta[$, there exists a homoclinic orbit to the periodic orbit P_E of energy E contained in the center manifold. We also suppose that the stable and unstable manifolds of the fixed point do not coincide.

5.2. Theorem. — *Under the hypotheses recalled above, there exists a sequence $E_n\to 0$ of positive numbers such that, for each n, the stable manifold of P_{E_n} and its unstable manifold do not coincide.*

5.3. In order to prove this theorem, let us define the function $N(E)$ which, to each value of energy $E\in\,]0,\eta[$, associates the minimal number of bumps of an orbit homoclinic to P_E

$$N(E)=\min\{n\in\mathbb{N}\text{ s.t. } F_E^{n-1}\circ g_E(S'_E)\cap S_E\neq\varnothing\},$$

which is finite in view of Theorem 1.3. See 3.1 for the definition of S_E.

Lemma. — *The function* $]0, \eta[\ \ni E \longmapsto N(E)$ *is lower semi-continuous and continuous at each point* E_0 *such that* $F_{E_0}^{N(E_0)-1} \circ g_{E_0}(S'_{E_0}) = S_{E_0}$. *In addition,* $\lim_{E\to 0} N(E) = \infty$.

This lemma implies the desired result. Assume by contradiction that the stable and unstable manifolds of P_E coincide for all energies E in an interval $]0, \varepsilon[$. By the lemma the function N would be continuous, hence constant on this interval, and N would have a finite limit in 0, which is in contradiction with the last part of the lemma. There remains to prove the lemma:

Proof of the lemma. — Let us fix a value E_0 of the energy, and consider a sequence $E_n \to E_0$ such that $N(E_n) = N$ is constant. We have

$$F_{E_n}^{N-1} \circ g_{E_n}(S'_{E_n}) \cap S_{E_n} \neq \varnothing.$$

for each n. This clearly implies that

$$F_{E_0}^{N-1} \circ g_{E_0}(S'_{E_0}) \cap S_{E_0} \neq \varnothing.$$

hence $N(E_0) \leqslant N$. This proves lower semi-continuity of N. If the stable and unstable manifolds of P_{E_0} coincide, there holds

$$F_{E_0}^{N(E_0)-1} \circ g_{E_0}(S'_{E_0}) = S_{E_0}.$$

It is then clear, by area preservation, that

$$F_{E}^{N(E_0)-1} \circ g_{E}(S'_{E}) \cap S_{E} \neq \varnothing$$

for E sufficiently close to E_0, hence $N(E) \leqslant N(E_0)$. As a consequence, E_0 is a point of upper semi-continuity of N, hence a point of continuity. To end the proof, we note that if there existed a sequence $E_n \to 0$ with $N(E_n)$ bounded, there would exist a homoclinic orbit to the fixed point. This can be checked by a compactness argument similar to the proof of lower semi-continuity above. □

References

[1] Arnold, V.I., (1964) *Instability of dynamical systems with several degrees of freedom*, Dokl. Akad. Nauk. 156, pp 9-12.

[2] Bernard, P., *Homoclinic orbit to a center manifold*, to appear in Calc. Var. & P. D. E.

[3] Buffoni, B., (1999) *Shooting methods and topological transversality* Proc. R. Soc. Edimb., sec A, Math. 125 No6, pp 1137-1155

[4] Conley, C., (1968) *Twist mappings, linking, analyticity, and periodic solutions which pass close to an unstable periodic solution*, Topological dynamics (Sympos., Colorado State Univ., Ft. Collins, Colo., 1967), pp. 129–153. Benjamin, New York.

[5] Conley, C., (1969), *On the ultimate behavior of orbits with respect to an unstable critical point I. Oscillating, asymptotic and capture orbits*, Journ. Diff. Eqns 5, pp 136-158.

[6] Easton, R.W., McGehee, R., (1979) *Homoclinic phenomena for orbits doubly asymptotic to an invariant three-sphere*, Indiana Univ. Math. J. 28, pp 211-240.

[7] Grotta-Ragazzo, C., (1997) *Irregular Dynamics and Homoclinic Orbits to Hamiltonian Saddle Centers*, Comm. Pure App. Math. L, pp 105-147.

[8] Grotta-Ragazzo, C., (1997) *On the Stability of Double Homoclinic Loops*, Comm. Math. Phys. 184, pp 251-272.

[9] Lerman, L.M., (1991) *Hamiltonian systems with loops of a separatrix of saddle center*, Selecta Math. Sov. 10, 1991, pp 297-306.

[10] Lerman, L.M., Koltsova, O.Yu., (1995) *Periodic and homoclinic orbits in two-parameter unfolding of a Hamiltonian system with a homoclinic orbit to a saddle-center*, Intern. Jour. Biff. Chaos 5, pp. 397-408.

[11] LLibre, J., Martinez, R., Simó, C., (1985) *Transversality of the invariant manifolds associated to the Lyapunov family of periodic orbits near L_2 in the restricted three body problem*, Jour. Diff. Eq. 58, pp. 104-156.

[12] Lombardi, E., (2000) *Oscillatory integrals and phenomena beyond all algebraic orders*, Lec. Notes in Maths 1741, Springer.

[13] Mielke, A., (1990) *Topological methods for discrete dynamical systems*, GAMM-Mitteilungen, Heft-2, pp. 19-37.

[14] Mielke, A., Holmes, P., and O´Reilly, O., (1992) *Cascades of homoclinic orbits to, and chaos near, a Hamiltonian saddle center*, J. Dyn. Diff. Eqns. 4, pp. 95-126.

[15] Moser, J., (1958) *On the generalization of a theorem of A. Liapunoff*, Comm. Pure Appl. Math. v. 11, pp. 257-271.

[16] Moser, J., (1973) *Stable and random motions in dynamical systems*, Ann. Math. Stud. 77, Princeton University Press.

[17] Rüssmann, H., (1964) *Über das Verhalten analytischer Hamiltonscher Differentialgleichungen in der Nähe einer Gleichgewichtslösung*, Math. Ann. v. 154, pp. 285-300. Princeton University Press.

[18] Simó, C., Stuchi, T., (2000) *Central stable/unstable manifolds and the destruction of KAM tori in the restricted planar Hill problem*, Physica D 140, pp 1-32.

P. Bernard, Institut Fourier, BP74, 38402 Saint Martin d'Hères cedex, France
E-mail : `Patrick.Bernard@ujf-grenoble.fr` • *Url :* `http://www-fourier.ujf-grenoble.fr/~pbernard/`

C. Grotta Ragazzo, Universidade de Sao Paulo, Instituto de Matematica e Estatistica, Rua do Matao 1010, Butanta-Cidade Universitaria, 05508-090, Sao Paulo SP, Brasil
E-mail : `ragazzo@ime.usp.br`

P.A. Santoro Salomão, Universidade de Sao Paulo, Instituto de Matematica e Estatistica, Rua do Matao 1010, Butanta-Cidade Universitaria, 05508-090, Sao Paulo SP, Brasil
E-mail : `psalomao@ime.usp.br)`

Astérisque
286, 2003, p. 167–186

ON THE SCALING STRUCTURE FOR PERIOD DOUBLING

by

Garrett Birkhoff, Marco Martens & Charles Tresser

Abstract. — We describe an order on the set of scaling ratios of the generic universal smooth period doubling Cantor set and prove that this set of ratios forms itself a Cantor set, a Conjecture formulated by Coullet and Tresser in 1977. This result establishes explicitly the geometrical complexity of the universal period doubling Cantor set. We also show a convergence result for the two period doubling renormalization operators, acting on the codimension one space of period doubling maps. In particular they form an iterated function system whose limit set contains a Cantor set.

1. Definitions and Statement of the Results

A *unimodal map with critical exponent* $\alpha > 1$ is an interval map that can be written in the form $f = \psi \circ q_t \circ \phi$, where ψ and ϕ are orientation preserving C^3 diffeomorphisms of $[0, 1]$, and $q_t : [0, 1] \to [0, 1]$ with $t \in (0, \frac{1}{2}]$ is the *standard folding map (with critical exponent* $\alpha > 1$*)* defined by

$$q_t(x) = 1 - \frac{|x - t|^\alpha}{|1 - t|^\alpha},$$

that "folds" the interval at its unique critical point t, $q_t(t) = 1$ and $q_t^2(t) = 0$.

The space of orientation preserving diffeomorphisms of the interval $[0, 1]$ with fixed smoothness is denoted by $\mathrm{Diff}^k([0, 1])$. The space of unimodal maps with fixed critical

2000 ***Mathematics Subject Classification.*** — 58F03, 58F13.
Key words and phrases. — Universality, renormalization, period doubling, scaling function, Cantor sets.

A part of the work has been done when M.M. and C.T. were participating in the conference dedicated to Jacob Palis 60th birthday in IMPA. Both authors acknowledge the hospitality of IMPA. M.M. and C.T. are partially supported by NSF Grant DMS-0073069, and G.B. acknowledges the hospitality of the Mathematical Sciences department of the T.J. Watson Research Center.

exponent $\alpha > 1$ and fixed smoothness can be represented by

$$\mathcal{U} = \mathrm{Diff}^k([0,1]) \times (0, \tfrac{1}{2}] \times \mathrm{Diff}^k([0,1]).$$

It carries what we call *C^k-distances* d_k, $k \geqslant 3$, which combines the two C^k distances on each of the two diffeomorphisms ψ and ϕ with the distance between the parameters t of the folding parts. Notice that in general, the critical point of f is $c_f = \phi^{-1}(t) \neq t$. Let p_f be the unique fixed point of $f \in \mathcal{U}$. A map on the interval is *renormalizable* if it exchanges some number N_1 of subintervals. The return map on one of these subintervals can again be renormalizable, exchanging this time N_2 intervals. If the process continues forever, one says the map is *infinitely renormalizable.* For precise definitions and an account of the theory, see for instance [**dMvS**]. Except otherwise specified when we say renormalizable, we mean renormalizable in the sense of period doubling, *i.e.*, the map exchanges two intervals. We will only consider infinitely renormalizable maps with $N_1 = N_2 = \cdots = 2$.

Fix a critical exponent $\alpha > 1$. We consider the set W of maps $f : [0,1] \to [0,1]$ with $f(c_f) = 1$ and $f(1) = 0$ which are infinitely renormalizable. The critical point defines two invariant intervals

$$U_f = [f^2(c_f), f^4(c_f)] \quad \text{and} \quad V_f = [f^3(c_f), f(c_f)].$$

To these two intervals correspond two *renormalization operators* $R_0 : W \to W$ and $R_1 : W \to W$ defined by:

$$R_0 f = [f^2 | V_f], \quad \text{and} \quad R_1 f = [f^2 | U_f],$$

where $[\cdot]$ means *affine rescaling to obtain a unimodal map on* $[0,1]$ *that sends its critical point to 1 and 1 to 0.*

Observe, both operators preserve W and R_1 is the critical point period doubling renormalization operator which has been most studied in the literature (see in particular [**La**], [**Ly**], [**Mc**], [**dMvS**], [**S2**], and references therein for the case when α is an even integer, and [**E1**], [**E2**] and [**Ma2**] for arbitrary $\alpha > 1$).

Let T_n be the set of all words of length n over the alphabet $\{0,1\}$. We denote by T the set of all infinite words of the form $w1^\infty$ over the alphabet $\{0,1\}$, and by $\overline{T}$ the set of all infinite words over the alphabet $\{0,1\}$, equipped with the usual metric. Notice that each T_n naturally embeds into T. For any word $\tau \in \overline{T}$, we will write $\tau_{\{n\}} \in T_n$ for the initial segment of length n of τ. We are going to consider the iterated function system generated by R_0 and R_1. To this end, we define:

$$R_{\tau_{\{n\}}} = R_{\tau(1)} \circ \cdots \circ R_{\tau(n)} : W \longrightarrow W,$$

and we will prove the following convergence result for this iterated function system.

Theorem 1.1. — *For any fixed point f_0 of R_0, there is a Hölder-continuous map $h : \overline{T} \to W$ such that for any $\tau \in \overline{T}$*

$$\lim_{n\to\infty} R_{\tau_{\{n\}}} f_0 = h(\tau).$$

Moreover, the convergence of the sequence $\{R_{\tau_{\{n\}}} f_0\}$ *is exponential in the* C^2*-metric.*

A similar statement holds for any fixed point f_1 *of* R_1.

Remark 1.2. — For any $\alpha > 1$, the existence of a fixed point f_1 of R_1 is proven in [**E1, E2**] and [**Ma2**]. We will show (see Lemma 2.4) that the existence of a fixed point f_1 for R_1 is equivalent to the existence of a fixed point f_0 for R_0. The uniqueness of f_1 in the case when α is an even integer was proven in [**S2**]. In the sequel we will fix f_0 and f_1 to be fixed points of respectively R_0 and R_1.

Remark 1.3. — The set $h(\overline{T})$ of limits $\lim_{n\to\infty} R_{\tau_{\{n\}}} f_0$ is denoted by $A \subset W$. Here the notation A represents the fact that we believe, but do not prove, that the set A is indeed the attractor of the iterated function system generated by R_0 and R_1, and in particular does not depend on the initial point, chosen here to be f_0.

The second Main result, Theorem 1.10, describes the structure of the set A in the case when $\alpha = 2$. It relies on convexity properties of f_0 and $R_1(f_0)$.

Convexity Conditions 1.4. — *We assume that:*

C1 $\quad f_0|[(f_0)^3(c_{f_0}), 1)$ *is strictly convex,*

C2 $\quad R_1(f_0)|[(R_1(f_0))^3(c_{R_1(f_0)}), 1]$ *is strictly convex.*

Remark 1.5. — In section 4 we will show that **C1** actually holds true in the case when successive R_1 renormalizations of a convex function converge to f_1: this is known to be the case when α is an even integer. Furthermore, as we will explain, one can check that both **C1** and **C2** hold true in the most important case of generic (quadratic) critical points, $\alpha = 2$.

Recall that a Cantor set is a perfect and totally disconnected compact metric space.

Proposition 1.6. — *If the Convexity Conditions C1 and C2 hold true, then the limit set A of orbits of f_0 under the interated function system defined by R_0 and R_1 is a Cantor set.*

For completeness and to fix notations and definitions, we include some basic discussion of the scaling function, whose origin is rather diffuse: first conjectures about a form of it appeared in [**CT**], the name and a form of it come from [**F**], while what was arguably the first theorem about it was in a never circulated work by Feigenbaum and Sullivan cited in [**S1**]. The literature on scaling functions is extensive and discusses scaling functions beyond the context of dynamics. In particular, in [**KSV**] a relation with the thermodynamic formalism appeared.

Let Λ be the invariant Cantor set of f_0. In the sequel we will remind the dynamical construction of covers of Λ by finitely many intervals. These covers, called cycles, form a refining nest of covers of this Cantor set. The scaling function contains the infinitesimal geometrical information on how these covers refine. It will be shown that the Cantor set Λ is, from a geometrical point of view, very different from the well known middle third Cantor set, in which each refinement is done everywhere in the same manner.

Although, the Cantor set Λ is the invariant set of a non expanding map, it is also the invariant Cantor set of an expanding interval map, the so-called *presentation function* [**R**], [**S1**], a great remark that Rand attributes to Misiurewicz. As we next recall, this directly follows from f_0 being a renormalization fixed point that is expanding to the right of p_{f_0}.

Let $U = U_{f_0}$ and $V = V_{f_0} = [1-v, 1]$. The affine (scaling) map $s : [0,1] \to [0,1]$ defined by $s : x \mapsto v \cdot (x-1) + 1$ is a homeomorphism from Λ to $\Lambda \cap V$. This is a direct consequence of the fact that s conjugates $f_0 = R_0(f_0) = s^{-1} \circ f_0^2 \circ s$ to f_0^2. Also the restriction,

$$f_0|V : \Lambda \cap V \longrightarrow \Lambda \cap U,$$

is a homeomorphism so that the map $g : [0,1] \to U$ defined by $g = (f_0|V) \circ s$ is a homeomorphism from Λ to $\Lambda \cap U$. Let $F : [0,1] \to [0,1]$ be the multivalued function defined by the two branches

$$F_0 = s : [0,1] \longrightarrow [0,1] \quad \text{and} \quad F_1 = g : [0,1] \longrightarrow [0,1].$$

The branch $F_0 = s$ is affine, contracting, and orientation preserving while the branch $F_1 = g$ is orientation reversing. Furthermore, the absolute value of the derivative of F_1 strictly increases as a consequence of the Convexity Condition C1, so that F_1 is also contracting (as p_{f_0} is an expanding fixed point). It follows that the invariant set of the iterated function system $F = \{F_0, F_1\}$ is Λ, the invariant Cantor set of f_0.

The cover $\{U, V\}$ of Λ is called *the cycle of the first generation.* The two intervals of this cycle are permuted by the map f_0. The Cantor set Λ is the intersection of a decreasing sequence of covers we call respectively the *cycles of generation* n: the cycle of generation n is the cover of Λ consisting of 2^n intervals which are permuted by f_0. The intervals that form the n^{th} cycle can be described as follows.

The construction of the cycles is made by using the iterated function system generated by F_0 and F_1. We will use a notation for the words describing sequences of compositions of these maps that will be different from the one we used in the definition of the iterated function system generated by R_0 and R_1. Namely, we write Σ_n for the set of words $w = w(1)w(2)\dots w(n)$ of length $|w| = n$ over the alphabet $\{0,1\}$, and Σ for the set of infinite sequences over the alphabet $\{0,1\}$ with the usual metric. Let

$$I_w = F_{w(n)} \circ \cdots \circ F_{w(1)}([0,1]).$$

The n^{th} cycle consists of the intervals I_w with w a word of length n.

Lemma 1.7. — *The way f_0 permutes these intervals is described by addition* mod 2^n *on the words indexing the intervals. In particular, if c is the critical point of f_0 then*

$c \in I_{1^n}$ and $f_0(c) \in I_{0^n}$. Moreover, $f(I_{1^n}) = I_{0^n}$ and

$$f_0 : I_w \longrightarrow I_{w+1},$$

is a diffeomorphism for each word not equal to 1^n, $n \geqslant 1$.

Proof. — Let w be a word of length $n-1$. Then

$$I_{w1} = F_1(I_w) = f_0 \circ s(I_w) = f_0(I_{w0}),$$

which proves that f_0 permutes the intervals as stated. □

The *orientation* of an interval I_w is defined to be the number

$$o(w) = (-1)^{\#(w)},$$

where $\#(w)$ is the number of 1's in w. The shift of a word $w = w(1)w(2)\dots w(n)$ is defined as

$$\sigma(w) = w(2)w(3)\dots w(n).$$

Observe, that

$$I_w \subset I_{\sigma(w)}.$$

In particular, the n^{th} cycle has two intervals in each interval of the $(n-1)^{th}$ cycle:

$$I_{0w}, I_{1w} \subset I_w.$$

The scaling function $q_n : w \mapsto (0,1)$ assigns to each word w of length n the ratio

$$q_n(w) = \frac{|I_w|}{|I_{\sigma(w)}|}.$$

The a priori bounds on the possible values of q_n, as presented in [**Ma1**] for example imply

$$|I_w| \leqslant \rho^{|w|}$$

for some fixed $\rho < 1$. From this and the smoothness of f_0 it follows that the sequence q_n converges to a Hölder function $q : \Sigma = \{0,1\}^{\mathbf{N}} \to (0,1)$. This function q is what we call the *scaling function*, in minor departure from some previous authors.

The next proposition describes properties of the scaling function. To formulate this proposition we need an order on Σ: with w standing for the maximal word such that $w_1 = ww^1$ and $w_2 = ww^2$, we say that w_1 is strictly smaller than w_2 (or $w_1 \prec w_2$)if and only if

$$(-1)^{\#(w)} \cdot w^1(1) < (-1)^{\#(w)} \cdot w^2(1).$$

Proposition 1.8. — *If the Convexity Conditions hold true then q is strictly monotone.*

Furthermore, under the same hypothesis, there exists constants $C > 0$ and $r < 1$ such that if $w_1 \prec w_2$ and $w_1(k) = w_2(k)$ whenever $k \leqslant n$ then

$$q(w_2) \geqslant q(w_1) + Cr^n$$

Remark 1.9. — If the Convexity Conditions C1 and C2 hold true, Proposition 1.8 confirms the 1977 Conjecture in [**CT**] that the limit set of the ratios $q_n(w)$ defining the period doubling Cantor set is itself a Cantor set.

In particular, we thus have the following

Theorem 1.10. — *In the case of quadratic critical point, $\alpha = 2$, we have the following.*

– *The Convexity Conditions holds true.*

– *The universal period doubling scaling function q is strictly monotone and the range forms a Cantor set.*

– *The limit set A of orbits of f_0 under the interated function system defined by R_0 and R_1 is a Cantor set.*

This Theorem establishes explicitly the geometrical complexity of the universal period doubling Cantor set: for related matters, see [**GT**] and [**T**].

Acknowledgements. — H. Epstein and O.E. Lanford discovered a relation between the fixed points of R_0 and R_1. Roughly speaking this relation states that if $f(x) = h(x^2)$ represents the fixed point of R_1 then $g(x) = (h(x))^2$ represents the fixed point of R_0. This result was not published. However, it was the main inspiration for Section 2. In particular, Lemma 2.4 contains this result.

2. Decompositions and Convergence

The notion of decomposition, introduced in [**Ma2**], is a tool to describe the combinatorial aspects of universality. In this section, after some background on decompositions, we prove the convergence properties stated in Theorem 1.1.

The set T_n is ordered by the embedding into the natural numbers defined by

$$\tau(1)\tau(2)\dots\tau(n) \longmapsto \sum_{i=1}^{n} \tau(i)\cdot 2^{n-i}.$$

Consider also the embedding $j_n : T_n \to T_{n+1}$ defined by

$$j_n : \tau \longmapsto \tau 1.$$

This embedding preserves the order. Observe that T inherites an order from the orders on the sets T_n, which extends to the order on $\overline{T}$ such that $\tau^1 \leqslant \tau^2$ iff $\tau^1_{\{n\}} \leqslant \tau^2_{\{n\}}$ for all $n \geqslant 1$. The elements of $\overline{T}$ are called *decomposition times.*

For the order $<$, the successor in T_n of $1^n \in T_n$ is $0^n \in T_n$ and the predecessor in T_n of $0^n \in T_n$ is 1^n. The successor of $\tau \in T$ in T_n is denoted by τ^{n+} and the predecessor is denoted by τ^{n-}.

The nonlinearity of an orientation preserving diffeomorphism $\phi \in \mathrm{Diff}^2([0,1])$ is

$$\eta_\phi = D \ln D\phi \in C^0([0,1]).$$

A *decomposed unimodal map* is a map

$$\widetilde{f} : T \longrightarrow \mathrm{Diff}^3([0,1]) \cup (0, \tfrac{1}{2}]$$

with the following properties

– $\widetilde{f}(1^\infty)$, the *folding* part of $\widetilde{f}$ represents an element q_t of the standard folding family, so we have $\widetilde{f}(1^\infty) = t \in (0, \frac{1}{2}]$,

– $\widetilde{f}(\tau) \in \mathrm{Diff}^3([0,1])$ for $\tau \neq 1^\infty$, (the *diffeomorphic parts* of $\widetilde{f}$).

– $\sum_{\tau \in T \smallsetminus \{1^\infty\}} |\eta_{\widetilde{f}(\tau)}|_0 < \infty$.

– $\sum_{\tau \in T \smallsetminus \{1^\infty\}} |D\eta_{\widetilde{f}(\tau)}|_0 < \infty$.

The set U of decomposed unimodal maps carries the metric d defined by

$$d(\widetilde{f}, \widetilde{g}) = \sum_{\tau \in T \smallsetminus \{1^\infty\}} |\eta_{\widetilde{f}(\tau)} - \eta_{\widetilde{g}(\tau)}|_1 \ + \ |\widetilde{f}(1^\infty) - \widetilde{g}(1^\infty)|.$$

The two summability conditions for decomposed unimodal maps allow to define what we call *compositions* associated to decomposed unimodal maps. Namely, if one considers a finite set T_n of decomposition times, the composition associated to $\widetilde{f}$ and T_n is defined as

$$O(\widetilde{f}, n) = \widetilde{f}(1^{n-1}0) \circ \cdots \circ \widetilde{f}(0^{n-1}1) \circ \widetilde{f}(0^n) \circ q_{\widetilde{f}(1^n)},$$

otherwise speaking, the folding part followed by the diffeomorphic parts in the order of the decomposition times (so that the end result of the composition is a unimodal map). In [**Ma2**] it is shown that this composition, when defined for decomposed unimodal maps over the sets T_n, extends to a composition operator still denoted O:

$$O : U \longrightarrow \mathcal{U},$$

where $\mathcal{U}$ is equipped with the C^2 metric, which is a Lipschitz map. This composition operator is based on a choice. Namely, the composition starts with the folding part $q_{\widetilde{f}(1^n)}$. We could as well start at any decomposition time $\tau \in T_N$, $N \geqslant 1$ and consider for each $n \geqslant N$ the compositions defined by

$$O(\tau, \widetilde{f}, n) = \widetilde{f}(\tau^{n-}) \circ \cdots \circ \widetilde{f}(0^{n-1}1) \circ \widetilde{f}(0^n) \circ q_{\widetilde{f}(1^n)} \circ \widetilde{f}(1^{n-1}0) \circ \cdots \circ \widetilde{f}(\tau^{n+}) \circ \widetilde{f}(\tau).$$

The same proof which was used in [**Ma2**] to construct $O(\widetilde{f})$ shows the pointwise convergence of the sequence $O(\tau, \widetilde{f}, n)$ as $n \to \infty$, thus defining a map denoted O again:

$$O : T \times U \longrightarrow \mathcal{U}.$$

Observe that $O(1^\infty, \widetilde{f})$ is the operator studied in [**Ma2**].

This construction can be generalized even more. Fix $\widetilde{f} \in U$ and choose $\tau_2 > \tau_1$ in T_N. For each $n \geqslant N$ define the diffeomorphism

$$O_{\tau_1}^{\tau_2}(\widetilde{f}, n) = \widetilde{f}(\tau_2^{n-}) \circ \cdots \circ \widetilde{f}(\tau^{n+}) \circ \widetilde{f}(\tau) \circ \cdots \circ \widetilde{f}(\tau_1^{n+}) \circ \widetilde{f}(\tau_1).$$

It follows from [**Ma2**] that these maps converge, and we set

$$O_{\tau_1}^{\tau_2}(\widetilde{f}) = \lim_{n \to \infty} O_{\tau_1}^{\tau_2}(\widetilde{f}, n).$$

Moreover, there is a constant $K_{\widetilde{f}}$ such that

$$|O_{\tau_1}^{\tau_2}(\widetilde{f}) - \mathrm{id}|_2 \leqslant K_{\widetilde{f}} \cdot \sum_{\{\tau \in T | \tau_2 > \tau \geqslant \tau_1\}} |\eta_{\widetilde{f}(\tau)}|_0.$$

Lemma 2.1. — *The operator O extends continuously to an operator*

$$O : \overline{T} \times U \longrightarrow \mathcal{U}.$$

In particular, for each $\widetilde{f} \in U$ there exists a constant $K_{\widetilde{f}} > 0$ such that for any pair $\tau_2, \tau_1 \in \overline{T}$ with $\tau_2 \geqslant \tau_1$,

$$d_2(O(\tau_2, \widetilde{f}), O(\tau_1, \widetilde{f})) \leqslant K_{\widetilde{f}} \cdot \sum_{\{\tau \in T | \tau_2 > \tau \geqslant \tau_1\}} |\eta_{\widetilde{f}(\tau)}|_0.$$

Moreover for each $\tau_3 > \tau_2 > \tau_1 \in \overline{T}$ and $\widetilde{f} \in U$

$$O_{\tau_1}^{\tau_3}(\widetilde{f}) = O_{\tau_2}^{\tau_3}(\widetilde{f}) \circ O_{\tau_1}^{\tau_2}(\widetilde{f}).$$

Proof. — Fix $\widetilde{f} \in U$ and choose $\tau_2 > \tau_1$ in T_N. Let $h = O_{\tau_1}^{\tau_2}(\widetilde{f})$. The construction of h implies directly

$$h \circ O(\tau_1, \widetilde{f}) = O(\tau_2, \widetilde{f}) \circ h.$$

This construction can be done for every pair of $\tau_1', \tau_2' \in [\tau_2, \tau_1] \cap T$. Hence, there is a constant which only depends on $\widetilde{f}$ such that

$$d_2(O(\tau_2', \widetilde{f}), O(\tau_1', \widetilde{f})) \leqslant \text{ Const } \cdot \sum_{\{\tau \in T | \tau_2 > \tau \geqslant \tau_1\}} |\eta_{\widetilde{f}(\tau)}|_0.$$

From this we get the continuous extension of O to $\overline{T} \times U$, together with the estimate stated in the Lemma. The composition rule clearly holds for the operators $O_{\tau_1}^{\tau_2}(\widetilde{f}, n)$ and hence for the continuous extension of O. $\square$

We will also write $O_\tau(\cdot)$ for $O(\tau, \cdot)$. Let $\mathcal{U}_0$ be the set of renormalizable unimodal maps and $U_0 = (O_{1^\infty})^{-1}(\mathcal{U}_0)$. A renormalization operator $R : U_0 \to U$ is constructed in [**Ma2**] such that

$$O_{1^\infty} \circ R = R_1 \circ O_{1^\infty}.$$

A decomposed unimodal map $\widetilde{f} \in U_0$ is said to be n *times renormalizable* iff $f = O(\widetilde{f}) \in \mathcal{U}$ is n times renormalizable: we then set $f = \phi \circ q_t$ with $t \in (0, \frac{1}{2}]$. This means there are pairwise disjoint intervals $I_\tau^{f,n}$, $\tau \in T_n$, forming the n^{th} cycle of f, such that

- $t \in I_{1^n}^{f,n}$,
- $f : I_\tau^{f,n} \to I_{\tau^{n+}}^{f,n}$ is a diffeomorpishm, whenever $\tau \neq 1^n$,
- $f : I_{1^n}^{f,n} \to I_{0^n}^{f,n}$ is onto.

Let $g : I \to J$ be an endormorphism which has either one or zero critical point. Then $[g] : [0,1] \to [0,1]$ is a either a unimodal map or an orientation preserving diffeomorphism obtained by affine scaling of the domain and image of g.

Lemma 2.2. — *Let $\widetilde{f} \in U$ be n times renormalizable and $O(\widetilde{f}) = f = \phi \circ q_t \in \mathcal{U}_0$ with $t \in (0, \frac{1}{2}]$. For $n \geqslant 1$ and $\tau \in T_n \subset T$*

- $O_\tau^{\tau^{n+}}(R^n \widetilde{f}) = [f|I_\tau^{f,n}]$,
- $O_\tau^{\tau^{n+}0^\infty}(R^n \widetilde{f}) = [q_t|I_\tau^{f,n}]$,
- $O_{\tau^{n+}0^\infty}^{\tau^{n+}}(R^n \widetilde{f}) = [\phi|q_t(I_\tau^{f,n})]$.

The reader is refered to [**Ma2**] for the precise definition of the renormalization operator $R : U_0 \to U$, from which the Lemma immediately follows. This lemma indeed captures all the properties of the renormalization operator R that we will need.

Proposition 2.3. — *For every $\tau \in T_n \subset T$*

$$O_\tau \circ R^n = R_\tau \circ O_{1^\infty}, \quad \textit{and} \quad O_{\tau 0^\infty} \circ R^n = R_\tau \circ O_{0^\infty}.$$

Proof. — Let $\widetilde{f} \in U$ be $n \geqslant 1$ times renormalizable and

$$O(\widetilde{f}) = O_{1^\infty}(\widetilde{f}) = f = \phi \circ q_t \in \mathcal{U}_0$$

with $t \in (0, \frac{1}{2}]$. As in the proof of Lemma 3.1 shows that for every $n \geqslant 1$ and $\tau \in T_n \subset T$

$$R_\tau(f) = [f^{2^n}|I_\tau^{f,n}].$$

Let $\tau_1 = \tau$, $\tau_k = \tau_{k-1}^{n+}$, for $k = 2, 3, \ldots, 2^n$. The composition rule for the operators $O_{\tau_1}^{\tau_2}$ and Lemma 2.2 imply

$$\begin{aligned} O_\tau \circ R^n(\widetilde{f}) &= O_{\tau_{2^n}}^{\tau_1}(R^n \widetilde{f}) \circ \cdots \circ O_{\tau_2}^{\tau_3}(R^n \widetilde{f}) \circ O_{\tau_1}^{\tau_2}(R^n \widetilde{f}) \\ &= [f|I_{\tau_{2^n}}^{f,n}] \circ \cdots \circ [f|I_{\tau_2}^{f,n}] \circ [f|I_{\tau_1}^{f,n}] \\ &= [f^{2^n}|I_{\tau_1}^{f,n}] \\ &= R_{\tau_1}(f) \\ &= R_\tau \circ O_{1^\infty}(\widetilde{f}). \end{aligned}$$

The second equation is proved similarly. □

Lemma 2.4. — *The operators R_0 and R_1 have fixed points. Furthermore, for any even integer α, both operators R_0 and R_1 have a unique fixed point.*

Proof. — It was shown in [**Ma2**] that the operator R has a fixed point. The previous proposition implies that a fixed point $\widetilde{f} \in U_0$ of R produces fixed points of R_0 and R_1. Namely,

$$R_1(O_{1^\infty}(\widetilde{f})) = O_{1^\infty}(\widetilde{f}) \quad \text{and} \quad R_0(O_{0^\infty}(\widetilde{f})) = O_{0^\infty}(\widetilde{f}).$$

Claim 2.5. — *For each fixed point $f \in \mathcal{U}$ of R_1 (or R_0) there exists a unique fixed point of R, say $\widetilde{f} \in U$ such that $O_{1^\infty}(\widetilde{f}) = f$ (or $O_{0^\infty}(\widetilde{f}) = f$).*

Proof. — Let $f = \phi \circ q_t \in \mathcal{U}$ be a fixed point of R_1 (the case of a fixed point for R_0 can be treated the similarly). Choose $\widetilde{f} \in U$ such that

$$O_{1^\infty}(\widetilde{f}) = f.$$

For example, consider $\widetilde{f} \in U$ defined by

– $\widetilde{f}(1^\infty) = q_t$,
– $\widetilde{f}(01^\infty) = \phi$,
– $\widetilde{f}(\tau) = \mathrm{id}$ for $\tau \neq 1^\infty, 01^\infty$.

The definition of $\widetilde{f}$ and the fact that $O_{1^\infty} \circ R = R_1 \circ O_{1^\infty}$, implies

$$O_{1^\infty}(R^n \widetilde{f}) = f, n \geqslant 1.$$

We will show

$$\lim_{n\to\infty} R^n \widetilde{f} = \widehat{f} \in U,$$

with

$$R\widehat{f} = \widehat{f} \text{ and } O_{1^\infty}(\widehat{f}) = f.$$

Let $n \geqslant 1$ and $\tau_3 > \tau_2 > \tau_1 \in T_{n+1}$ three consecutive decomposition times in T_{n+1} with $\tau_3, \tau_1 \in T_n$. Observe, that τ_3 and τ_1 are consecutive points in T_n. From Lemma 2.2 we get

$$\begin{aligned} O_{\tau_1}^{\tau_3}(R^{n+1}\widetilde{f}) &= O_{\tau_2}^{\tau_3}(R^{n+1}\widetilde{f}) \circ O_{\tau_1}^{\tau_2}(R^{n+1}\widetilde{f}) \\ &= [f|I_{\tau_2}^{f,n+1}] \circ [f|I_{\tau_1}^{f,n+1}] \\ &= [f^2|I_{\tau_1}^{f,n+1}] \\ &= [f|I_{\tau_1}^{f,n}], \end{aligned}$$

where we used that f is a fixed point of R_1. Again from Lemma 2.2 we get $[f|I_{\tau_1}^{f,n}] = O_{\tau_1}^{\tau_3}(R^n\widetilde{f})$. Hence,

$$O_{\tau_1}^{\tau_3}(R^{n+1}\widetilde{f}) = O_{\tau_1}^{\tau_3}(R^n\widetilde{f}).$$

This should be interpreted as $R^{n+1}\widetilde{f}$ being a refinement of $R^n\widetilde{f}$. In [**AMM**] it has been shown that there is a constant $K > 0$ and $\rho < 1$ such that

$$\sum_{\tau_1 \in T_n} |(O_{\tau_2}^{\tau_3}(R^{n+1}\widetilde{f}) - \mathrm{id})|_2 \leqslant K \cdot \rho^n.$$

This implies that $\lim_{n\to\infty} R^n\widetilde{f} = \widehat{f} \in U$. In particular, this implies that $\widehat{f}$ is a fixed point of R which projects by O_{1^∞} to f. This concludes the existence part of the Claim.

We can use Lemma 2.2 to identify $\widetilde{f}(\tau)$, $\tau \in T_N$. Namely,

$$\begin{aligned}\widetilde{f}(\tau) &= \lim_{n\to\infty} R^n \widetilde{f}(\tau)\\ &= \lim_{n\to\infty} O_\tau^{\tau^{n+0^\infty}}(R^n \widetilde{f})\\ &= \lim_{n\to\infty} [q_t | I_\tau^{f,n}]\\ &= [q_t | I_\tau^{f,N}],\end{aligned}$$

where we used that f is a fixed point of R_1 to obtain the last equality. This implies the uniqueness part of the Claim. □

It has been shown in [**S2**] that the operator R_1 has a unique fixed point when α is an even integer. Now the uniqueness part of Lemma 2.4 follows by using the Claim. □

Proof of Theorem 1.1. — Let f_0 be a fixed point of R_0 and $\widetilde{f}_0 \in U$ the unique fixed of R with $O_{0^\infty}(\widetilde{f}_0) = f_0$. Let $h : \overline{T} \to W$ be defined by

$$h(\tau) = O_\tau(\widetilde{f}_0).$$

For any $\tau_1, \tau_2 \in \overline{T}$ let $|\tau_2 - \tau_1|$ be the maximal length for which initial segments of the word τ_1 and τ_2 of that length agree. In [**AMM**] it has been shown that there is a constant $K > 0$ and $\rho < 1$ such that

$$\sum_{\tau_2 > \tau > \tau_1} |\eta_{\widetilde{f}_0(\tau)}|_0 \leqslant K \cdot \rho^{|\tau_2 - \tau_1|}.$$

Recall that $\tau_{\{n\}}$ is the word consisting of the first n symbols of a word $\tau \in \overline{T}$. From Lemma 2.1 we get

$$d_2(h(\tau_{\{n\}}0^\infty), h(\tau)) \leqslant K \cdot \rho^n.$$

Theorem 1.1 follows from Proposition 2.3. Namely,

$$\begin{aligned}R_{\tau_{\{n\}}} f_0 &= R_{\tau_{\{n\}}} \circ O_{0^\infty} \widetilde{f}_0\\ &= O_{\tau_{\{n\}}0^\infty} \circ R^n \widetilde{f}_0\\ &= O_{\tau_{\{n\}}0^\infty} \widetilde{f}_0\\ &= h(\tau_{\{n\}}0^\infty) \longrightarrow h(\tau),\end{aligned}$$

where the convergence is exponential.

3. The monotonicity of the scaling function

The monotonicity of the scaling function q, as formulated in Proposition 1.8 is based on the following combinatorial Lemmas. First we will concentrate on these Lemmas and prove Proposition 1.8. Secondly, Proposition 1.8 is used to prove Proposition 1.6.

Although decomposition times and the words used to define the intervals I_w are conceptually different, the following Lemma shows that they are strongly related.

Lemma 3.1. — *For every word w of length n*

$$R_w(f_0) = [f_0^{2^n}|I_w].$$

Proof. — The proof is by induction in n. For $n = 1$ the Lemma restates the definition of R_0 and R_1. Assume the Lemma holds for some $n \geqslant 1$. Choose a word w of length n and consider the two intervals I_{0w} and I_{1w}. These intervals are contained in I_w and each contains a boundary point of I_w. Using the induction hypothesis $R_w(f_0) = [f_0^{2^n}|I_w]$ and the fact that $f_0^{2^n}|I_w$ permutes I_{0w} and I_{1w} we get that $R_{0w}(f_0) = R_0(R_w(f_0))$ and $R_{1w}(f_0) = R_1(R_w(f_0))$ correspond to either of $f^{2^{n+1}}|I_{0w}$ or $f^{2^{n+1}}|I_{1w}$.

It is left to identify which of the two intervals corresponds to $U_{R_w(f_0)}$ (resp. to $V_{R_w(f_0)}$). The map f_0 permutes the intervals $I_{w'}$ with $|w'| = n+1$ according to addition mod.2^n on the words indexing the intervals, as described in Lemma 1.7. Observe that

$$1w = 0w + 2^n \cdot 1.$$

This means that $f_0^{2^n}|I_{0w}$ is monotone because $0w + k \cdot 1$, $k < 2^n$ never equals the word 1^{n+1} and $f_0|I_{1^{n+1}}$ is the only place where monotonicity of f_0 fails. Hence,

$$R_{0w}(f_0) = R_0([f_0^{2^n}|I_w]) = [(f_0^{2^n})^2|I_{0w}]$$

and

$$R_{1w}(f_0) = R_1([f_0^{2^n}|I_w]) = [(f_0^{2^n})^2|I_{1w}]. \qquad \square$$

In the sequel we will identify $R_w(f_0)$ with $f_0^{2^n}|I_w$.

Lemma 3.2. — *For every pair of words w and w^0, the map*

$$R_{w^0}(f_0) : I_{w0w^0} \longrightarrow I_{w1w^0},$$

is monotone and onto.

Proof. — Let $|w^0| = n$. The action of f_0 on the intervals of length $|w| + 1 + |w^0|$ is described by addition mod.2^n on the words indexing the intervals (see Lemma 1.7). In particular,

$$w1w^0 = w0w^0 + 2^n \cdot 1.$$

Hence

$$f_0^{2^n}(I_{w0w^0}) = I_{w1w^0}.$$

By construction we have

$$I_{w1w^0}, I_{w0w^0} \subset I_{w^0}.$$

Now the Lemma follows from $R_{w^0}(f_0) = [f_0^{2^n}|I_{w^0}]$, which we know from Lemma 3.1. $\square$

Lemma 3.3. — *If the Convexity Condition holds then there exist constants $C > 0$ and $r \in (0,1)$ with the following property. Let w be a word of with $|w| = n$.*

If $o(w) = +1$ then

– $w0 < w1$ *and* $w00 < w01 < w11 < w10$
– $q_{n+1}(w0) < q_{n+1}(w1)$
– $q_{n+2}(w00) < q_{n+2}(w01) < q_{n+2}(w11) < q_{n+2}(w10)$
– $q_{n+2}(w11) > q_{n+2}(w01) + Cr^n$

If $o(w) = -1$ *then*

– $w1 < w0$ *and* $w10 < w11 < w01 < w00$
– $q_{n+1}(w1) < q_{n+1}(w0)$
– $q_{n+2}(w10) < q_{n+2}(w11) < q_{n+2}(w01) < q_{n+2}(w00)$
– $q_{n+2}(w01) > q_{n+2}(w11) + Cr^n$

Proof. — The construction of the intervals I_w imply immediately the following. If $o(w) = +1$ then the interval I_w contains the right boundary point of $I_{\sigma(w)}$. And if $o(w) = -1$ then I_w contains the left boundary point of $I_{\sigma(w)}$. Using this, the convexity of F_1 and the fact that F_0 is affine we get

Claim 3.4. — $o(w) \cdot q_{n+1}(w0) < o(w) \cdot q_{n+1}(w1)$, *for every word* w *with* $|w| = n$.

The case when $o(w) = -1$ of the Lemma can be proved similarly as the first case. We will only present the proof in the case $o(w) = +1$. The first statement is merely the definition of the order on the symbol space. The second follows directly from Claim 3.4. This Claim also implies

$$q_{n+2}(w00) < q_{n+2}(w01), \quad \text{and} \quad q_{n+2}(w11) < q_{n+2}(w10).$$

To study the middle inequality, observe that

$$I_{\sigma(w)01} \cup I_{\sigma(w)11} \subset I_1.$$

First observe that $o(w01) = -1$ (and $o(w11) = 1$). In particular the negatively oriented interval I_{w01} contains the left boundary point of the interval $I_{\sigma(w)01}$. Moreover,

$$I_{\sigma(w)01} \subset I_{01} \subset [0, f_0^4(c_{f_0})],$$

where $0 \in I_1$ is the left boundary point of I_1.

By Lemma 3.2 we have

$$R_1(f_0) : I_{\sigma(w)01} \longrightarrow I_{\sigma(w)11}.$$

The Convexity Condition states that the absolute value of the derivative of this map decreases strictly on the interval $[0, f_0^6(c_{f_0})]$. Now using

$$I_{w01} \subset I_{\sigma(w)01} \subset [0, f_0^6(c_{f_0})]$$

and that the interval $I_{w01} \subset I_{\sigma(w)01}$ contains the left boundary point of $I_{\sigma(w)01}$, we get

$$q_{n+2}(w01) < q_{n+2}(w11).$$

From the a priori bounds described for example in [**Ma1**], we know that there are constants $C > 0$ and $r \in (0,1)$ such that

$$|I_w| \geqslant C \cdot r^{|w|}$$

for all words w. This implies the final estimate of Lemma 3.3. □

Let w be a word with $|w| = k$. Then define the interval

$$J_w = [q_{k+1}(w0), q_{k+1}(w1)].$$

Proof of Proposition 1.8. — The proposition 1.8 is reformulated in

Claim 3.5. — *Let w be a word with $|w| = k$ and $|wh| = n$. Then*

$$q_n(wh) \in J_w.$$

In particular,

$$J_{wh} \subset J_w.$$

Moreover, if w^1 and w^2 are distinct words of length k then $J(w^1)$ and $J(w^2)$ are disjoint and the distance between them is larger than Cr^k.

Proof. — The proof of the first part of the Claim is by induction in n. For $n = 2$ the statement follows from the Lemma 3.3. Assume the Claim holds for all words wh with $|wh| \leqslant n$.

Consider a word $wh = w\widehat{h}h^1h^2$ with $|wh| = n+1$ and $|h^1| = |h^2| = 1$. Then Lemma 3.3 implies that for every pair of symbol x, y

$$q_{n+1}(w\widehat{h}xy) \in [q_{n+1}(w\widehat{h}10), q_{n+1}(w\widehat{h}00)].$$

In particular,

$$\begin{aligned} q_{n+1}(wh) &\in [q_{n+1}(w\widehat{h}10), q_{n+1}(w\widehat{h}00)] \\ &= [q_n(w\widehat{h}1), q_n(w\widehat{h}0)] \\ &\subset J_w, \end{aligned}$$

The above equality follows from the fact that $q_{n+1}(w\widehat{h}10) = q_n(w\widehat{h}1)$ because the interval $I_{w\widehat{h}10}$ is obtained from $I_{w\widehat{h}1}$ by applying the affine branch F_0. The other boundary is treated similarly. The last inclusion follows from the induction hypothesis.

The proof of the second part of the Claim is by induction in $k = |w|$. For $k = 1$ the Claim considering the distance between J_0 and J_1 is a reformulation of the previous Lemma. Assume, the Claim is proved up to some $k \geqslant 1$. Let w^1 and w^2 be two words of length $k+1$, say $w^1 = \widetilde{w}^1 x$ and $w^2 = \widetilde{w}^2 y$ with $|\widetilde{w}^1| = |\widetilde{w}^2| = k$.

If $\widetilde{w}^1$ differs from $\widetilde{w}^2$ then the Claim follows because

$$J_{w^1} \subset J_{\widetilde{w}^1}, J_{w^2} \subset J_{\widetilde{w}^2}$$

and the induction hypothesis. So we may assume that

$$w^1 = w0, w^2 = w1.$$

Apply Lemma 3.3 again to conclude that J_{w^1} and J_{w^2} are disjoint with the appropriate distance between them. □

Proof of Proposition 1.6. — The proof of Proposition 1.6 relies on the relation between the two iterated function systems generated by respectively $\{R_0, R_1\}$ and $\{F_0, F_1\}$ as formulated in Lemma 3.1. Notice, the only difference between Σ and $\overline{T}$ is that they carry different orders. The order does not play any role in the proof of Proposition 1.6. We will use the symbol w for words which are in $\Sigma = \overline{T}$. In Section 2 we constructed the continuous map $h : \Sigma \to A$ (see Remark 1.3). Namely, for $w \in \Sigma$, let

$$h(w) = \lim_{n\to\infty} R_{w_{\{n\}}}(f_0).$$

In particular, this map is onto. It is left to show that h is injective.

Observe that every word w with $|w| = n$

$$R_{0w}(f_0) = R_0(R_w(f_0)).$$

In particular,

$$q_{n+1}(0w) = |V_{R_w(f_0)}|.$$

Recall that for $w \in \Sigma$ we denote the word consisting of the first n symbols of $w \in \Sigma$ by $w_{\{n\}}$. Let $w^1, w^2 \in \Sigma$ be such that $h(w^1) = h(w^2)$. Then

$$\begin{aligned}
|q(0w^1) - q(0w^2)| &= \lim_{n\to\infty} |q_{n+1}(0w^1_{\{n\}}) - q_{n+1}(0w^2_{\{n\}})| \\
&= \lim_{n\to\infty} ||V_{R_{w^1_{\{n\}}}(f_0)}| - |V_{R_{w^2_{\{n\}}}(f_0)}|| \\
&\leqslant \text{Const} \lim_{n\to\infty} \text{dist}(R_{w^1_{\{n\}}}(f_0), R_{w^2_{\{n\}}}(f_0)) \\
&= \text{Const} \cdot \text{dist}(h(w^1), h(w^2)) = 0.
\end{aligned}$$

The strict monotonicity of the scaling function, Proposition 1.8, implies $w^1 = w^2$. This proves that $h : \Sigma \to A$ is a homeomorphism.

4. The Convexity Condition

In this section the Convexity Condition will be studied.

Lemma 4.1. — *Let $f : (-1, 1) \to (-1, 1)$ be C^2. If*

- $f(0) = 0$,
- $Df(0) < -1$,
- $D^2f(0) < 0$

then

$$D^2(f^2)(0) < 0.$$

Proof. — The chain rule applied to f^2 gives

$$D^2(f^2)(x) = D^2f(f(x)) \cdot (Df(x))^2 + Df(f(x)) \cdot D^2f(x).$$

Using the properties of f in $x = 0$ we get

$$D^2(f^2)(0) = D^2f(0) \cdot Df(0) \cdot [Df(0) + 1] < 0. \qquad \square$$

Lemma 4.2. — *Let $C \subset W$ consisting of unimodal maps $f \in W$, with negative Schwarzian derivative (see* [**dMvS**] *for the definition), and the following property: $f|[0,c]$ is convex, where c is the critical point of f, and $f|[c,1]$ is strictly convex (The derivative of f is decreasing over $[0,1]$ but strictly decreasing on $[c,1]$). Then*

$$R_0(C) \subset C.$$

Proof. — Let $f \in C$ with critical point $c \in [0,1]$ and let p_f be its fixed point. Let $V_f = P \cup Q$, where P, Q are the two intervals on which $R_0 f$ is monotone. Choose $Q \subset V_f$ such that $f(Q) \subset [0,c]$. The convexity property of f implies directly the strict convexity of $R_0(f)|Q$.

The Schwarzian derivative of f is negative. This implies that p_f is an expanding fixed point, otherwise it would attract the critical point (see [**dMvS**]). Hence, $Df(p_f) < -1$. The convexity condition of f allows us to apply the previous Lemma:

$$D^2R_0(f)(p_f) < 0,$$

i.e. the derivative of f^2 is decreasing in p_f. Now, the Minimum Principle for maps with negative Schwarzian derivative (again see [**dMvS**]), implies that Df^2 is decreasing monotonically to zero on the interval $[p_f, P]$, hence $R_0 f \in C$. $\square$

Lemma 4.3. — *The convexity condition* **C1** *holds true for any even critical exponent α, the map $f_0|[p_{f_0}, 1]$ is strictly convex.*

Proof. — Let $q_t \in W$ be a standard folding map. Clearly, $q_t \in C$. From [**S2**] we have

$$\lim_{n\to\infty} R_1^n q_t = f_1.$$

Let $\widehat{f}$ be the unique fixed point of R (with $O_{1^\infty}(\widehat{f}) = f_1$). As in the proof of Claim 2.5 we get for every $\widetilde{f} \in U$ with $O_{1^\infty}(\widetilde{f}) = q_t$ that

$$\lim_{n\to\infty} R^n \widetilde{f} = \widehat{f}.$$

Hence,

$$\begin{aligned} \lim_{n\to\infty} R_0^n q_t &= \lim_{n\to\infty} O_{0^\infty} R^n \widetilde{f} \\ &= O_{0^\infty}(\widehat{f}) \\ &= f_0, \end{aligned}$$

where f_0 is the fixed point of R_0. This implies that the derivative of f_0 is decreasing because of the previous Lemma . The renormalization fixed point f_0 is real analytic.

Hence, the set $E \subset [0,1]$ consisting of the flat points of f_0, points where $D^2 f_0$ vanishes, is finite.

The map f_0 is the fixed point of R_0. Hence, $s(E)$, the map s is the affine scaling of the interval $[0,1]$ to V_{f_0}, is the set of flat points of $R_0 f_0 (= f_0)$. Let $Q \subset V_{f_0}$ be the maximal interval such that $f_0(Q) \subset [0,c]$. Any non-flat point $x \in Q$ will be a non-flat point of $R_0 f_0$, this follows from the convexity, maybe not strict, of f_0. Hence

$$s(E) \cap Q \subset E.$$

Assume, $E \cap [c,1) \neq \varnothing$ and let $x \in E \cap [c,1)$ be the rightmost point. The fact that f_0 is a renormalization fixed point implies that $s(c)$ is the left boundary point of Q. In particular we get

$$x < s(x) \in s(E) \cap s([c,1]) \subset E \cap Q,$$

contradicting the fact that x was chosen to be the right most point in $E \cap [c,1)$. This proves that f_0 does not have flat points in $[p_{f_0}, 1) \subset [c,1)$. □

Lemma 4.4. — *The convexity condition* **C2** *holds true for* $\alpha = 2$.

Proof. — In the case $\alpha = 2$, an approximation of f_1 can be found in [**La**]. We will use the notation of [**La**]. The fixed point f_1 is represented as $g(z) = h(z^2)$ where $|z|^2 \leqslant 1.5$. Actually, the map h defined on the disk $D_0 = \{z \| z| \leqslant 1.5\}$ where it is analytic. The map h is approximated by a polynomial of degree 40.

$$h_0(z) = 1 + \sum_{n=1}^{40} g_n^{(0)} z^n$$

where

$$|g_n^{(0)}| \leqslant 10^{-(n-2)}$$

It is also shown in [**La**] that

$$|h(z) - h_0(z)| \leqslant 1.5 \cdot 10^{-23}, \quad z \in D_0.$$

From Lemma 2.4 we get that the map $f(z) = (h(z))^2$, $z \in D_0$ represents the fixed point f_0 of R_0, the maps are equal up to an affine scaling. The map $P(z) = (h_0(z))^2$, $z \in D_0$, approximates this fixed point. For both maps the dynamically relevant interval is $[0,1]$: $f([0,1]) = [0,1]$ and $P([0,1]) = [0,1]$.

To prove the convexity condition **C2** it suffices to show the strict convexity of f^2 restricted to the interval $[f^6(c_f), 1]$.

Claim 4.5. — *The derivative of of* h_0 *restricted to* D_0 *satisfies*

$$|Dh_0(z)| \leqslant 14.0$$

This estimate follows from the bounds on the coefficients of the polynomial h_0.

Claim 4.6. — *For $z \in D_0$*

$$|P(z) - f(z)| \leqslant 1.0 \cdot 10^{-20}$$

and the derivative of P restricted to the disk D_0 satisfies

$$|DP(z)| \leqslant 700.$$

The bound on the coefficients of the polynomial h_0 imply that $h_0(\{z||z| \leqslant 1.5\})$ is contained in a disk of radius 23 around 0. The bounds on the distance between h and h_0 and the fact that the derivative of the map $z \mapsto z^2$ is bounded by 50 on the disk of radius 23 around 0, finishes the proof of this Claim.

Let D be the $\frac{1}{1500}$−neighborhood of the interval $[0, 1]$.

Claim 4.7. — *The map f^2 is defined on D (and is analytic). Moreover*

$$|f^2(z) - P^2(z)| \leqslant 10^{-17}, \quad z \in D.$$

The fact that f^2 is well defined on D follows from the fact that P maps D well inside the disk of radius D_0 and that P and f are close on D_0. The estimate on the distance between f^2 and P^2 on D follows from the bound on the derivative of P and the very small distance between f and P.

Claim 4.8. — *For every $z \in [0, 1]$*

$$|D^2P^2(z) - D^2f^2(z)| \leqslant 1.0 \cdot 10^{-7}$$

and

$$|D^3P^2(z) - D^3f^2(z)| \leqslant 1.0 \cdot 10^{-4}$$

These bounds follow by applying the Cauchy integral formula for derivatives. Let $z_0 \in [0, 1]$. Then

$$\begin{aligned}|D^2P^2(z) - D^2f^2(z)| &\leqslant \frac{1}{2\pi}\int_{\partial D}\frac{|P^2(z) - f^2(z)|}{|z - z_0|^3}dz \\ &\leqslant \frac{1}{2\pi}\cdot 10^{-20}\cdot\frac{1}{0.001^3}\cdot 2\pi(1.5) \leqslant 1.0 \cdot 10^{-10}.\end{aligned}$$

The third derivative is treated similarly.

It is left to find a lower bound for $|D^2P^2(z)|$ larger than 10^{-9}. We will use traditional cross ratio technology [**dMvS**] to reduce this question to a calculation in finitely many points. Let $h : T \to h(T)$ be a diffeomorphism of the interval T to its image and suppose it has negative Schwarzian derivative. Let $M \subset T$ be a subinterval and let $L, R \subset T \setminus M$ be the two connected components of $T \setminus M$. Let

$$\tau = \min\left\{\frac{|h(L)|}{|h(M)|}, \frac{|h(R)|}{|h(M)|}\right\}.$$

Then

$$|Dh(x)| \leqslant \frac{1+\tau}{\tau}\cdot\frac{|h(M)|}{|M|}, \quad x \in M.$$

The nonlinearity of h is $\eta = D\ln Dh = D^2h/Dh$. We have the following estimate

$$|\eta(x)| \leqslant 2\frac{(1+\tau)|h(M)|}{\tau^2|M|^2}, \quad x \in M.$$

The third inequality we will use is

$$D^3h \geqslant \frac{3}{2}(\frac{D^2h}{Dh})^2 \cdot Dh$$

in the case when Dh is negative.

We will apply these three estimates to the map f^2 restricted to the interval $[c_f, 1]$ with $M = [f^6(c_f), 1]$. The period two point of f and the position of $f^4(c_f)$ can be precisely estimated with the help of P. Using estimates for these two points gives the following estimates

$$\tau \geqslant 0.2 \quad \text{and} \quad |M| \geqslant 0.1, \ |f^2(M)| \leqslant 0.6.$$

This implies

$$|D(D^2P^2)(x)| \geqslant -2.2 \cdot 10^8, \quad x \in M.$$

Claim 4.9. — *For every $z \in [f^6(c_f), 1]$*

$$|D^2P^2(z)| \geqslant 0.5.$$

This is shown by numerical analysis. The second derivative of P^2 is calculated in a sequence of points with increment 10^{-9} over the interval $[f^6(c_f), 1]$. In these point the second derivative of P^2 is smaller than -1. The derivative estimate of D^2P^2 leads to the lower bound as stated in the Claim. □

The quadratic case, $\alpha = 2$, as described in Theorem 1.10 follows from propositions 1.6 and 1.8, and lemmas 4.3 and 4.4.

References

[AMM] A. Avila, M. Martens, W. de Melo, *On the dynamics of the renormalization operator*, *Global Analysis of Dynamical Systems*, (Festschrift dedicated to Floris Takens for his 60th birthday), W. Broer, B. Krauskopf, and G. Vegter Eds. IOP Publishing Ltd 2001, 449-461.

[CT] P. Coullet, C. Tresser, *Itérations d'endomorphismes et groupe de renormalisation,* J. Phy. **C5** (1978), 25-28.

[E1] H. Epstein, *New proofs of the existence of the Feigenbaum fuctions,* Commun. Math. Phys. **106** 395-426 (1986).

[E2] H. Epstein, *Fixed points of composition operators II,* Nonlinearity **2** 305-310 (1989).

[EL] H. Epstein, O.E. Lanford, personal communications.

[F] M.J. Feigenbaum, *The transition to aperiodic behavior in turbulent systems,* Commun. Math. Phys. **77** (1980), 65-86.

[GT] J.M. Gambaudo, C. Tresser, *Self-similar constructions in smooth dynamics: Rigidity, Smoothness, and Dimension,* Commun. Math. Phys. **150** (1992), 45-58.

[KSV] K.M. Khanin, Ya. Sinai, E.B. Vul, *Feigenbaum universality and thermodynamic formalism,* Uspekhi Mat. Nauk **39** (1984) 3-37.

[La] O.E. Lanford III, *A computer assisted proof of the Feigenbaum conjecture,* Bull. Amer. Math. Soc. (N.S.) **6** (1984), 427-434.

[Ly] M. Lyubich, *Feigenbaum-Coullet-Tresser Universality and Milnor's Hairriness Conjecture,* Ann. of Math. **149** (1999), 319-420.

[Ma1] M. Martens, *Distortion Results and Invariant Cantor Sets of Unimodal Maps*, Erg. Th. & Dyn. Sys. **14** (1994), 331-349.

[Ma2] M. Martens, *The Periodic Points of Renormalization*, Ann. of Math. **147** (1998), 543-584.

[dMvS] W. de Melo, S. van Strien, *One-dimensional Dynamics*, (Springer-Verlag, Berlin, 1993).

[Mc] C. McMullen, *Complex Dynamics and Renormalization*, Ann. of Math., Studies **135**, (Princeton University Press, Princeton, 1994).

[R] D.A. Rand, *Global phase space universality, smooth conjugacy and renormalization: I. The $C^{1+\alpha}$ case,* Nonlinearity **1** (1988), 181-202.

[S1] D. Sullivan, *Differentiable structure on fractal like sets, determined by intrinsic scaling functions on dual Cantor sets,* The mathematical heritage of Hermann Weyl (Durham, NC, 1987), Proc. Sympos. in Pure Math. **48**, 15-23. (Amer. Math. Soc., Providence, RI, 1988).

[S2] D. Sullivan, *Bounds, Quadratic Differentials, and Renormalization Conjectures,* in A.M.S. Centennial Publication **Vol 2** *Mathematics into the Twenty-first Century* (Am.Math. Soc., Providence, RI, 1992).

[T] C. Tresser, *Fine structure of universal Cantor sets,* in *Instabilities and Nonequilibrium Structures III*, E. Tirapegui and W. Zeller Eds., (Kluwer, Dordrecht-Boston-London, 1991).

G. BIRKHOFF, Englewood, 311 Morrow Road, N.J. 07631, USA • *E-mail :* gbirkhoff@groton.org

M. MARTENS, Department of Mathematics, IBM T.J. Watson Research Center, Yorktown Heights, P.O. Box 21817, N.Y. 10598, USA • *E-mail :* mmartens@us.ibm.com

C. TRESSER, Department of Mathematics, IBM T.J. Watson Research Center, Hawthorne, 17 skyline drive, N.Y. 10532, USA • *E-mail :* tresser@us.ibm.com

Astérisque
286, 2003, p. 187–222

ROBUSTLY TRANSITIVE SETS AND HETERODIMENSIONAL CYCLES

by

Christian Bonatti, Lorenzo J. Díaz, Enrique R. Pujals & Jorge Rocha

Abstract. — It is known that all non-hyperbolic robustly transitive sets Λ_φ have a dominated splitting and, generically, contain periodic points of different indices. We show that, for a $\mathcal{C}^1$-dense open subset of diffeomorphisms φ, the indices of periodic points in a robust transitive set Λ_φ form an interval in $\mathbb{N}$. We also prove that the homoclinic classes of two periodic points in Λ_φ are robustly equal. Finally, we describe what sort of homoclinic tangencies may appear in Λ_φ by studying its dominated splittings.

1. Introduction

When a diffeomorphism ϕ is *hyperbolic*, i.e., it verifies the Axiom A, the Spectral Decomposition Theorem of Smale says that its limit set (set of non-wandering points) is the union of finitely many *basic pieces* satisfying *nice* properties, each piece is invariant, compact, *transitive* (i.e., it contains an orbit which is a dense subset), pairwise disjoint and isolated (each piece is the maximal invariant set in a neighborhood of itself). Moreover, by construction, a basic piece is the *homoclinic class* of a hyperbolic periodic point, i.e., the closure of the transverse intersections of its invariant manifolds.

Even if the dynamics is non-hyperbolic, the homoclinic classes of hyperbolic periodic points seem to be the natural elementary pieces of the dynamics, satisfying many of the properties of the basic sets of the Smale's theorem: invariance, compactness, transitivity and density of hyperbolic periodic points. Recent results in [**BD**$_2$], [**Ar**] and [**CMP**] show that, for $\mathcal{C}^1$-generic diffeomorphisms (i.e., those belonging to

2000 Mathematics Subject Classification. — 37 C20, 37 D30, 37 C29.
Key words and phrases. — Dominated splitting, heterodimensional cycles, periodic orbits, transitivity.

This work was partially supported by CNPQ, FAPERJ, FINEP, and PRONEX-Dynamical Systems (Brazil), Laboratoire de Topologie of Université de Bourgogne (Dijon, France), and Centro de Matemática of Universidade do Porto (Portugal).

a residual subset of $\mathrm{Diff}^1(M)$) two homoclinic classes are either disjoint or equal and they are maximal transitive sets (i.e., every transitive set intersecting a homoclinic class is contained in it). Notice that, in general, the homoclinic classes fail to be hyperbolic, isolated and pairwise disjoint.

In [**BDP**] it is shown that, for $\mathcal{C}^1$-generic diffeomorphisms, a homoclinic class is either contained in the closure of an infinite set of sinks or sources, or satisfies some weak form of hyperbolicity (partial hyperbolicity or, at least, existence of a *dominated splitting*). The first situation (called the *Newhouse phenomenon*) can be locally generic, in the residual sense: there exist open sets in $\mathrm{Diff}^r(M)$ where the diffeomorphisms with infinitely many sinks or sources are (locally) residual for the $\mathcal{C}^r$-topology. The case $r \geqslant 2$ for surface diffeomorphisms can be found in [**N**], see [**PV**] for the case $r \geqslant 2$ in higher dimensions, and [$\mathbf{BD}_1$] for $r = 1$ in dimensions greater than or equal to 3. Certainly, the Newhouse phenomenon exhibits very wild behavior and it is conjectured that (in some sense) diffeomorphisms satisfying this phenomenon are very rare (for instance, for generic parametrized families of diffeomorphisms, the Lebesgue measure of the parameters corresponding to diffeomorphisms satisfying the Newhouse phenomenon is zero), see [**Pa**].

We focus here on the opposite behavior. More precisely, we restrict our attentions to the so-called *robustly transitive sets* introduced in [**DPU**] as a non-hyperbolic generalization of the basic sets of the *Spectral Decomposition of Smale.* A *robustly transitive set* Λ of a diffeomorphism ϕ is a transitive set which is locally maximal in some neighbourhood U of it and such that, for every C^1-perturbation ψ of the diffeomorphism ϕ, the maximal invariant set of ψ in U is transitive. From the results in [$\mathbf{M}_2$], [**DPU**] and [**BDP**] every robustly transitive set Λ admits a *dominated splitting,* say $T_\Lambda M = E_1 \oplus \cdots \oplus E_k$, and by [$\mathbf{BD}_2$], $\mathcal{C}^1$-generically, it is a *homoclinic class.* An invariant set may admit more than one dominated splitting, since one can always sum up some bundles of the original dominated splitting, obtaining a new dominated splitting with less bundles, or, conversely, split some bundle of the splitting in a dominated way. So it is natural to consider the *finest dominated splitting* of the set Λ (i.e., the one that does not admit any dominated sub-splitting).

In this paper we study the interrelation between the dominated splittings (especially the finest one) of a robustly transitive set Λ and its dynamics, answering questions about the *indices* (dimension of the stable manifold) of the periodic points of Λ, the possible bifurcations (saddle-node and homoclinic tangencies) occurring in this set as well as its dynamical structure.

Let us recall some definitions, necessary for what follows.

In what follows, M denotes a compact, closed Riemannian manifold and $\mathrm{Diff}^1(M)$ the space of $\mathcal{C}^1$-diffeomorphisms of M endowed with the usual topology.

Let Λ be a compact invariant set of a diffeomorphism ϕ. A ϕ_*-invariant splitting $T_\Lambda M = E \oplus F$ over Λ is said to be *dominated* if the fibers of E and F have constant dimension and there exists $k \in \mathbb{N}$ such that, for every $x \in \Lambda$, one has

$$||\phi_*^k|_{E(x)}|| \cdot ||\phi_*^{-k}|_{F(\phi^k(x))}|| < \frac{1}{2},$$

that is, the vectors in F are uniformly more expanded than the vectors in E by the action of ϕ_*^k. If it occurs we say that F *dominates* E and write $E \prec F$.

An invariant bundle E over Λ is *uniformly contracting* if there exists k such that, for every $x \in \Lambda$, one has:

$$||\phi_*^k|_{E(x)}|| < \frac{1}{2}.$$

An invariant bundle E over Λ is *uniformly expanding* if it is uniformly contracting for ϕ_*^{-1}.

Let $T_\Lambda M = E_1 \oplus E_2 \oplus \cdots \oplus E_m$ be a ϕ_*-invariant splitting over Λ such that the fibers of the bundles E_i have constant dimension. Denote by $E_i^j = \bigoplus_i^j E_k$ the direct sum of $E_i, \ldots, E_j$. Note that $E_1^{k-1} \oplus E_k^m$ is a splitting of $T_\Lambda M$ for all $k \in \{2, \ldots, m\}$. We say that $E_1 \oplus E_2 \oplus \cdots \oplus E_m$ is the *finest dominated splitting* of Λ if $E_1^{k-1} \oplus E_k^m$ is a dominated splitting for each $k \in \{2, \ldots, m\}$ and every E_k is *indecomposable* (i.e., it does not admit any nontrivial dominated splitting). See [**BDP**] for the existence and uniqueness of the finest dominated splitting.

Consider a set $V \subset M$ and a diffeomorphism $\varphi\colon M \to M$. We denote by $\Lambda_\varphi(V)$ the *maximal invariant set* of φ in V, i.e., $\Lambda_\varphi(V) = \bigcap_{i \in \mathbb{Z}} \varphi^i(V)$. Given an open set $U \subset M$ the set $\Lambda_\varphi(U)$ is *robustly transitive* if $\Lambda_\psi(U)$ is equal to $\Lambda_\psi(\overline{U})$ and is *transitive* for all ψ in a $\mathcal{C}^1$-neighbourhood of φ. We say that a ψ-invariant closed set K is *transitive* if there exists some $x \in K$ having a positive orbit which is dense in K.

If a robustly transitive set $\Lambda_\phi(U)$ is not (uniformly) hyperbolic then, by a $\mathcal{C}^1$-small perturbation of ϕ, one can create non-hyperbolic periodic points, and thus hyperbolic periodic points with different indices in $\Lambda_\phi(U)$ (see [$\mathbf{M}_2$]). Our first two results describe the possible indices of the periodic points of $\Lambda_\phi(U)$, in terms of the finest dominated splitting of $\Lambda_\phi(U)$:

Theorem A. — *Let U be an open set of M and $\mathcal{M}(U)$ a $\mathcal{C}^1$-open subset of* $\mathrm{Diff}^1(M)$ *such that $\Lambda_\varphi(U)$ is robustly transitive for every $\varphi \in \mathcal{M}(U)$. Then there is a dense open subset $\mathcal{N}(U)$ of $\mathcal{M}(U)$ such that, for every $\varphi \in \mathcal{N}(U)$, the set of indices of the hyperbolic periodic points of $\Lambda_\varphi(U)$ is an interval of integers (i.e., if P and Q are hyperbolic periodic points of indices p and q, $p \geqslant q$, of $\Lambda_\varphi(U)$, $\varphi \in \mathcal{N}(U)$, and $j \in [q,p]$, then $\Lambda_\varphi(U)$ has a hyperbolic periodic point of index j).*

In the next result, we use the arguments in [$\mathbf{M}_2$] to relate the uniform contraction or expansion of the extremal bundles of the finest dominated splitting of a robustly transitive set with the indices of the periodic points of this set.

Theorem B. — *Consider an open subset U of a compact manifold M and an integer $q \in \mathbb{N}^*$. Let $\mathcal{U}$ be a $\mathcal{C}^1$-open subset of* $\mathrm{Diff}^1(M)$ *such that for every $\phi \in \mathcal{U}$ the maximal invariant set $\Lambda_\phi(\overline{U})$ satisfies the following properties:*

(1) *the set* $\Lambda_\phi(\overline{U})$ *is contained in* U *and admits a dominated splitting* $E_\phi \oplus F_\phi$, $E_\phi \prec F_\phi$, *with* $\dim E_\phi(x) = q$,

(2) *the set* $\Lambda_\phi(\overline{U})$ *has no periodic points of index* $k < q$.

Then the bundle E_ϕ *is uniformly contracting for every* $\phi \in \mathcal{U}$.

We can summarize the two results above, in order to get a characterization of the set of indices of the periodic points of the set $\Lambda_\phi(\overline{U})$, as follows.

Let $U \subset M$ be open and φ a diffeomorphism such that $\Lambda_\varphi(U)$ is robustly transitive with a finest dominated splitting of the form $T_{\Lambda_\varphi(U)}M = E_1 \oplus \cdots \oplus E_{k(\varphi)}$, $E_i \prec E_{i+1}$. Denote by E^s the sum of all uniformly contracting bundles of this splitting and let E_α be the first non-uniformly contracting bundle, i.e., $E^s = E_1 \oplus \cdots \oplus E_{\alpha-1}$. In the same way, denote by E^u the sum of all uniformly expanding bundles of the splitting and let E_β be the last non-uniformly expanding bundle, i.e., $E^u = E_{\beta+1} \oplus \cdots \oplus E_{k(\varphi)}$. Let $\mathcal{U}$ be a $\mathcal{C}^1$-neighborhood of φ such that, for every $\psi \in \mathcal{U}$, the set $\Lambda_\psi(U)$ has the same properties as $\Lambda_\varphi(U)$ (i.e., robustly transitive and the number $k(\psi)$ of bundles of the finest dominated splitting is equal to $k(\varphi)$) and the dimensions of bundles $E^s(\psi)$, $E_\alpha(\psi)$, $E_\beta(\psi)$ and $E^u(\psi)$, defined in the obvious way, are constant in $\mathcal{U}$ and equal to corresponding bundles for ϕ.

Corollary C. — *With the notation above, there exist a* $\mathcal{C}^1$*-open and dense subset* $\mathcal{V}$ *of* $\mathcal{U}$ *and locally constant functions* $i, j\colon \mathcal{V} \to \mathbb{N}^*$ *such that*

$$i(\psi) \in [\dim(E^s), \dim(E^s) + \dim(E_\alpha)] \cap \mathbb{N}^*,$$
$$j(\psi) \in [\dim(E^u), \dim(E^u) + \dim(E_\beta)] \cap \mathbb{N}^*,$$

and, for every $\psi \in \mathcal{V}$, *the set of indices of the hyperbolic periodic points of* $\Lambda_\psi(\overline{U})$ *is the interval* $[i(\psi), \dim(M) - j(\psi)] \cap \mathbb{N}^*$.

The first known examples of non-hyperbolic robustly transitive sets had a one-dimensional central direction, see [**M**$_1$] and [**Sh**]. As a consequence, these examples do not present *homoclinic tangencies* (non-transverse homoclinic intersections between the invariant manifolds of some periodic point). Observe that if a periodic point has a homoclinic tangency then, after a perturbation of the diffeomorphism, one create a Hopf bifurcation (a periodic point whose derivative has a pair of conjugate nonreal eigenvalues of modulus one), see [**YA**] and [**R**], hence points whose central direction has dimension at least two. Currently examples of robustly transitive sets having a central direction of dimension two or more are known, see [**BD**$_1$], [**B**] and [**BV**]. Moreover, in some cases these sets exhibit homoclinic tangencies, see [**B**] and [**BV**]. Our next result explains what sort of dominated splitting of a robustly transitive set prevents homoclinic bifurcations.

We say that a robustly transitive set $\Lambda_\varphi(U)$ is $\mathcal{C}^1$*-far from homoclinic tangencies* if there are no *homoclinic tangencies* in $\Lambda_\psi(U)$, for all ψ in a $\mathcal{C}^1$-neighbourhood of φ.

Theorem D. — *Given an open set U of M let $\mathcal{P}(U) \subset \mathrm{Diff}^1(M)$ be an open set of diffeomorphisms φ such that:*

(1) *The set $\Lambda_\varphi(U)$ is robustly transitive and the minimum and the maximum of the indices of the hyperbolic periodic points of $\Lambda_\varphi(U)$ are constant in $\mathcal{P}(U)$. Denote these numbers by i_s and i_c, respectively.*

(2) *The set $\Lambda_\varphi(U)$ is $\mathcal{C}^1$-far from homoclinic tangencies.*

Then there is an open and dense subset $\mathcal{O}(U)$ of $\mathcal{P}(U)$ such that, for every $\varphi \in \mathcal{O}(U)$, the set $\Lambda_\varphi(U)$ has a dominated splitting $T_{\Lambda_\varphi(U)} = E^s \oplus E_1 \oplus \cdots \oplus E_r \oplus E^u$, such that

– *E^s is uniformly contracting and has dimension $i_s \geqslant 1$,*

– *E^u is uniformly expanding and has dimension $\dim(M) - i_c \geqslant 1$,*

– *$r = i_c - i_s$ and the bundle E_i has dimension one and it is not uniformly hyperbolic for every $i = 1, \ldots, r$.*

In fact, from the proof of this theorem, we get more: given any robustly transitive set $\Lambda_\phi(U)$, for diffeomorphisms in a $\mathcal{C}^1$-neighbourhood of ϕ, the dimensions of the non-hyperbolic bundles of its finest dominated splitting determine the *ranks* of the homoclinic tangencies (that is, the indices of the periodic points exhibiting the tangency) that can occur in $\Lambda_\psi(U)$. The precise statement of this result is in Section 6, see Theorem F.

Finally, for robustly transitive sets which are far from homoclinic tangencies, we prove that the *(relative) homoclinic classes* of two periodic points of this set are equal in a $\mathcal{C}^1$-robust way. More precisely, let P_φ be a hyperbolic periodic point of a diffeomorphism φ. We denote by H_{P_φ} the set of transverse intersections of the invariant manifolds of P_φ. Observe that the homoclinic class of P_φ is the closure of H_{P_φ}. Given an open set U, the *relative homoclinic class of P_φ in U* is the closure of the set $H_{P_\varphi}(U)$ of transverse homoclinic points of P_φ whose orbits are contained in U.

Theorem E. — *Let U be an open subset of M and $\mathcal{S}(U) \subset \mathrm{Diff}^1(M)$ an open set of diffeomorphisms φ such that*

– *the set $\Lambda_\varphi(U)$ is robustly transitive, and*

– *there are no homoclinic tangencies (in the whole manifold) associated to periodic points of $\Lambda_\varphi(U)$.*

Consider any pair of hyperbolic periodic points P_φ and Q_φ of $\Lambda_\varphi(U)$ with indices p and q whose continuations are defined for every ψ in $\mathcal{S}(U)$. Then there is an open and dense subset $\mathcal{D}(U)$ of $\mathcal{S}(U)$ such that

$$\overline{H_{P_\psi}(U)} = \overline{H_{Q_\psi}(U)}$$

for every ψ in $\mathcal{D}(U)$.

Unfortunately, in the theorem above we cannot ensure that the relative homoclinic classes of P_ψ and Q_ψ are equal to $\Lambda_\psi(U)$, although by the results in [**BD**$_2$] this is true for a residual subset of $\mathcal{S}(U)$.

Let us now say a few words about the proofs of our results. One of the main tools is the notion of *heterodimensional cycle.* Given a diffeomorphism ϕ with two hyperbolic periodic points P_ϕ and Q_ϕ with different indices, say index$(P_\phi) >$ index(Q_ϕ), we say that ϕ has a *heterodimensional cycle associated to P_ϕ and Q_ϕ,* denoted by $\Gamma(\phi, P_\phi, Q_\phi)$, if $W^s(P_\phi)$ and $W^u(Q_\phi)$ have a (nontrivial) transverse intersection and $W^u(P_\phi)$ and $W^s(Q_\phi)$ have a quasi-transverse intersection along the orbit of some point x, i.e., $T_xW^u(P_\phi) + T_xW^s(Q_\phi)$ is a direct sum. Notice that, in this case, $\dim(M) - \dim(T_xW^u(P_\phi) + T_xW^s(Q_\phi))$ is equal to index$(P_\phi)-$ index(Q_ϕ), this number being the *codimension of the cycle.*

The proof of Theorem A has two main ingredients. The first is Theorem 3.1, which implies that, by unfolding a heterodimensional cycle associated to points of indices q and p as above, one gets hyperbolic periodic points of some index in between q and p (a priori, we do not know the index of such a point). The second ingredient of the proof is the Connecting Lemma of Hayashi (see Theorem 2.1 and [**H**]) which allows us to create (after a $\mathcal{C}^1$-perturbation) heterodimensional cycles associated to any pair of periodic points of a robustly transitive set.

Two other important tools are the constructions in [**M**$_2$] and in [**BDP**] (specially the periodic linear systems with transitions). In this paper we need to introduce *transitions between points of different indices* in the same homoclinic class, generalizing the construction in [**BDP**], in which only transitions between points with the same index were considered.

Finally, to prove Theorem E, the main ingredient, besides the Connecting Lemma, is the proposition below concerning the structure of the homoclinic classes of hyperbolic points having a heterodimensional cycle.

We say that a hyperbolic periodic point R_ϕ is *$\mathcal{C}^1$-far from tangencies* if there is a $\mathcal{C}^1$-neighbourhood $\mathcal{W}$ of ϕ in $\mathrm{Diff}^1(M)$ such that every $\psi \in \mathcal{W}$ has no homoclinic tangencies associated to R_ψ. A heterodimensional cycle $\Gamma(\phi, P_\phi, Q_\phi)$ is *$\mathcal{C}^1$-far from homoclinic tangencies* if the points P_ϕ and Q_ϕ in the cycle are $\mathcal{C}^1$-far from homoclinic tangencies.

Finally, we say that two points x and y are *transitively related by ϕ* if there exists a transitive set of ϕ containing x and y. The points x and y are *transitively related in an open set U* if there exists a transitive set of ϕ contained in U that contains x and y.

Proposition 1.1. — *Let U be an open set, φ a diffeomorphism, and P_φ and Q_φ a pair of hyperbolic periodic points of φ of indices p and $q = p - 1$, respectively. Consider a neighbourhood $\mathcal{W}$ of ϕ in $\mathrm{Diff}^1(M)$ such that, for all $\psi \in \mathcal{W}$,*

– *the continuations P_ψ and Q_ψ are defined and $\mathcal{C}^1$-far from tangencies,*

– *the points* P_ψ *and* Q_ψ *are transitively related in* U.

Then there is a $\mathcal{C}^1$*-open subset* $\mathcal{W}_\varphi$ *of* $\mathcal{W}$, *with* $\varphi \in \overline{\mathcal{W}}$, *such that the relative homoclinic classes of* P_ψ *and* Q_ψ *in* U *are equal for every* $\psi \in \mathcal{W}_\varphi$.

[**DR**, Theorem A] asserts that, given any heterodimensional cycle $\Gamma(\phi, P_\phi, Q_\phi)$ of codimension one, far from homoclinic tangencies, there exists a $\mathcal{C}^1$-open set, whose closure contains ϕ, of diffeomorphisms φ such that P_φ and Q_φ are transitively related. Thus, for any diffeomorphism ϕ with a heterodimensional cycle which is far from homoclinic tangencies, there are diffeomorphisms φ arbitrarily close to ϕ satisfying the hypotheses of the proposition. The proof of Proposition 1.1 follows from the results in [**DR**] and the Connecting Lemma of Hayashi.

This paper is organized as follows. In Section 2 we get some results concerning heterodimensional cycles, robustly transitive sets and homoclinic classes using the Hayashi's Connecting Lemma. In Section 3 we prove Theorem A. For that, we study the creation of periodic points in the unfolding of heterodimensional cycles (of any codimension). In Section 4 we prove Theorem B, for that we recall some folklore results concerning dominated splittings and reformulate some results in [$\mathbf{M}_2$]. In Sections 5 and 6, we study the relationship between the finest dominated splitting of a robustly transitive set and the creation of homoclinic tangencies inside this set. Finally, in Section 7 we prove the results concerning (relative) homoclinic classes.

Acknowledgements. — We acknowledge Bianca Santoro for her comments that helped us to make the english redaction compatible with the high standards of one of the referees.

2. Transitively related points

We begin the proofs of our results by recalling the Hayashi's Connecting Lemma and deducing some consequences from it.

2.1. Connecting lemma and transitively related points

***Theorem 2.1* (Hayashi's Connecting Lemma, [H])**. — *Let* P_φ *and* Q_φ *be a pair of hyperbolic periodic points of a* C^1*-diffeomorphism* φ *such that there are sequences of points* x_n *and of natural numbers* k_n *such that the sequences* x_n *and* $\varphi^{k_n}(x_n)$ *accumulate on* $W^u_{loc}(P_\varphi)$ *and on* $W^s_{loc}(Q_\varphi)$, *respectively.*

Then there is a diffeomorphism ψ *arbitrarily* $\mathcal{C}^1$*-close to* φ *such that* $W^u(P_\psi)$ *and* $W^s(Q_\psi)$ *have a nonempty intersection.*

Remark 2.2. — Every pair of hyperbolic periodic points P_φ and Q_φ which are transitively related satisfy the hypotheses of the Connecting Lemma (Theorem 2.1).

Proof of the remark. — Consider a transitive set Λ containing P_φ and Q_φ and a point x of Λ whose positive orbit is dense in Λ. Then there are sequences of natural numbers m_n and r_n, $m_n, r_n \to \infty$ as $n \to \infty$, such that $\varphi^{m_n}(x) \to P_\varphi$ and $\varphi^{r_n}(x) \to Q_\varphi$. Then it is immediate to get new sequences m'_n and r'_n, with $m'_n, r'_n \to \infty$, such that $\varphi^{m'_n}(x)$ and $\varphi^{r'_n}(x)$ converge to some point of $W^u_{loc}(P_\varphi)$ and of $W^s_{loc}(Q_\varphi)$, respectively. Taking subsequences, if necessary, we can assume that $r'_n = m'_n + k_n$ for some $k_n > 0$. Now it suffices to take $x_n = \varphi^{m'_n}(x)$ and consider the sequences x_n and k_n. $\square$

2.2. Homoclinic relative classes and robustly transitive sets. — By [**BD**$_2$, Theorem B], there is a residual subset of $\mathrm{Diff}^1(M)$ consisting of diffeomorphisms such that the homoclinic classes of any two transitively related hyperbolic periodic points are the same. The proof of this result is based on the Hayashi's Connecting Lemma. Using the *relative version* of the connecting lemma, we get a relative version of [**BD**$_2$, Theorem B] whose proof is here omitted.

***Theorem 2.3* (Relative version of [BD$_2$, Theorem B]).** — *Given an open subset U of M, there exists a residual set $\mathcal{G}(U) \subset \mathrm{Diff}^1(M)$ such that, for every $\varphi \in \mathcal{G}(U)$, two hyperbolic periodic points P_φ and Q_φ of φ are transitively related in U if and only if the relative homoclinic class in U of P_φ and Q_φ are equal, i.e., $\overline{H_{P_\varphi}(U)} = \overline{H_{Q_\varphi}(U)}$.*

Let $\mathcal{A}(U) \subset \mathrm{Diff}^1(M)$ be an open set such that $\Lambda_\varphi(U)$ is robustly transitive for all $\varphi \in \mathcal{A}(U)$. By Pugh closing lemma (see [**Pu**]) and a Kupka-Smale argument, there is a residual subset $\mathcal{R}(U)$ of $\mathcal{A}(U)$ of diffeomorphisms φ such that, for all $\varphi \in \mathcal{R}(U)$, the hyperbolic periodic points form a dense subset of $\Lambda_\varphi(U)$. Taking $\mathcal{T}(U) = \mathcal{G}(U) \cap \mathcal{R}(U)$, where $\mathcal{G}(U)$ and $\mathcal{R}(U)$ are as above, we get the following:

Proposition 2.4. — *Let $U \subset M$ and $\mathcal{A}(U) \subset \mathrm{Diff}^1(M)$ be open sets such that $\Lambda_\varphi(U)$ is robustly transitive for all $\varphi \in \mathcal{A}(U)$. Then there exists a residual subset $\mathcal{T}_{\mathcal{A}}(U)$ of $\mathcal{A}(U)$ such that*

$$\overline{H_{P_\varphi}(U)} = \Lambda_\varphi(U)$$

for every $\varphi \in \mathcal{T}_{\mathcal{A}}(U)$ and every hyperbolic periodic point $P_\varphi \in \Lambda_\varphi(U)$.

2.3. Heterodimensional cycles. — We will use the following lemma, which follows from the Connecting Lemma and an argument of transversality:

Lemma 2.5. — *Let P_φ and Q_φ be hyperbolic periodic points of a diffeomorphism φ of indices p and q, $p \geqslant q$. Suppose that P_ψ and Q_ψ are transitively related for every ψ in a $\mathcal{C}^1$-neighbourhood $\mathcal{V}$ of φ. Then there is a dense subset $\mathcal{W}$ of $\mathcal{V}$ such that every ϕ in $\mathcal{W}$ has a heterodimensional cycle $\Gamma(\phi, P_\phi, Q_\phi)$ of codimension $(p-q)$.*

Proof. — Let $\psi \in \mathcal{V}$. Since P_ψ and Q_ψ are transitively related, by Remark 2.2, we can apply Theorem 2.1 to get ξ arbitrarily close to ψ (hence ξ is in $\mathcal{V}$) such that

$W^s(P_\xi) \cap W^u(Q_\xi) \neq \varnothing$. Since

$$\dim(W^s(P_\xi)) + \dim(W^u(Q_\xi)) = p + (\dim(M) - q) \geqslant \dim(M),$$

we can assume that $W^s(P_\xi)$ and $W^u(Q_\xi)$ intersect transversely.

Since ξ belongs to $\mathcal{V}$, the points P_ξ and Q_ξ are transitively related. Thus, again by Remark 2.2, we can apply Theorem 2.1 to get ϕ arbitrarily close to ξ (ϕ in $\mathcal{V}$) such that $W^s(P_\phi)$ and $W^u(Q_\phi)$ have (non empty) transverse intersection and $W^u(P_\phi) \cap W^s(Q_\phi) \neq \varnothing$. After a new perturbation, if necessary, we can assume that the last intersection is quasi-transverse, obtaining a heterodimensional cycle $\Gamma(\phi, P_\phi, Q_\phi)$ of codimension $(p-q)$, finishing the proof of the lemma. □

Let us state two remarks about the proof above that will be used in Section 7.

Remark 2.6. — Let P_φ and Q_φ be hyperbolic periodic points of a diffeomorphism φ of indices p and q, $p \geqslant q$. Suppose that P_ψ and Q_ψ are transitively related for every ψ in a neighbourhood $\mathcal{V}$ of φ. Then there is an open and dense subset $\mathcal{D}$ of $\mathcal{V}$ such that $W^s(P_\psi)$ and $W^u(Q_\psi)$ have a nontrivial transverse intersection, for every ψ in $\mathcal{D}$.

If in Theorem 2.3 we assume that the points P_φ and Q_φ have the same index, we get the following stronger version of it:

Remark 2.7. — Let P_φ and Q_φ be hyperbolic periodic points of the same index of a diffeomorphism φ and U an open set containing the orbits of P_φ and Q_φ. Suppose that P_ψ and Q_ψ are transitively related for every ψ in a neighbourhood $\mathcal{V}$ of φ. Then, there exists an open dense subset $\mathcal{O}$ of $\mathcal{V}$ such that, for every ψ in $\mathcal{O}$, the relative homoclinic classes of P_ψ and Q_ψ in U are equal.

3. Proof of Theorem A: unfolding heterodimensional cycles

3.1. Transitions for heterodimensional cycles. — We begin this section by stating a technical result, which introduces the concept of *transition between periodic points of different indices.*

Theorem 3.1. — *Let P and Q be two hyperbolic periodic points of a diffeomorphism φ of indices p and q, $p > q$, and periods $n(P)$ and $n(Q)$, respectively. Denote by M_P and M_Q the linear maps*

$$\varphi_*^{n(P)}(P)\colon T_PM \longrightarrow T_PM \quad \textit{and} \quad \varphi_*^{n(Q)}(Q)\colon T_Q \longrightarrow T_QM.$$

Assume that there exist dominated splittings

$$T_PM = E_1(P) \oplus E_2(P) \oplus E_3(P) \quad \textit{and} \quad T_QM = E_1(Q) \oplus E_2(Q) \oplus E_3(Q),$$

with $\dim(E_1(P)) = \dim(E_1(Q)) = q$ *and* $\dim(E_3(P)) = \dim(E_3(Q)) = \dim(M) - p$, *which are invariant by M_P and M_Q, respectively. Assume, in addition, that there is a heterodimensional cycle $\Gamma(\varphi, U, P, Q)$ in some open subset U of M.*

Then, for any fixed $\varepsilon > 0$, there are matrices T_0 and T_1 and $\delta > 0$ such that, for every n and $m \geqslant 0$, and every family of matrices $(I_i), i = 0, \ldots, (n+m)+2$, δ-close to identity, there exists a diffeomorphism ψ ε-$\mathcal{C}^1$-close to φ having a periodic orbit R of period $n(R)$ such that the linear map $M_R = \psi_*^{n(R)}$ is conjugate to

$$I_{n+m+2} \circ T_1 \circ I_{n+m+1} \circ M_Q \circ I_{n+m} \circ \cdots \circ I_{n+2} \circ M_Q \circ I_{n+1} \circ T_0 \circ I_n \circ M_P \circ I_{n-1} \circ \cdots \circ I_1 \circ M_P \circ I_0.$$

Moreover, $n(R) = t_1 + t_2 + n \cdot n(P) + m \cdot n(Q)$, where t_1 and t_2 are constants depending only on T_0 and T_1.

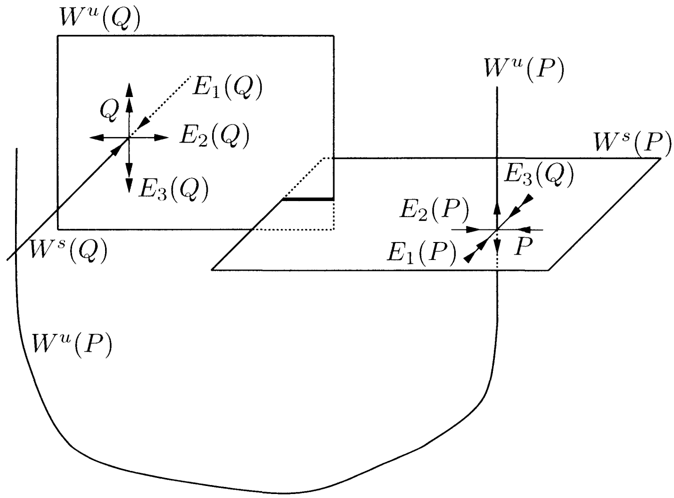


FIGURE 1. A heterodimensional cycle

The maps T_0 and T_1 are called *transitions* (from P to Q and from Q to P, respectively). These maps are a generalization of the transitions introduced in [**BDP**] for hyperbolic periodic points which are homoclinically related.

Theorem 3.1 is the main step for the proof of Theorem A. Taking appropriate n and m, and assuming that $\operatorname{index}(P) > \operatorname{index}(Q) + 1$, we get, using the theorem, that the index of R is between the indices of P and Q, see Corollary 3.6. This construction will also allow us to get points R corresponding to saddle-node bifurcations.

Proof. — For simplicity, assume that P and Q are fixed points. Notice that $E_1(Q)$ is the stable direction of Q, $E_1(P)$ is the strong stable direction of P, $E_3(Q)$ is the strong unstable direction of Q and $E_3(P)$ is the unstable direction of P.

We now make a $\mathcal{C}^1$-perturbation of the diffeomorphism φ to get appropriate linearizing coordinates of the cycle. The properties of this linearization are summarized in the next lemma:

Lemma 3.2. — *Let φ be a diffeomorphism satisfying the hypotheses of Theorem 3.1. Then, there exists ϕ, arbitrarily $\mathcal{C}^1$-close to φ, with a heterodimensional cycle $\Gamma(\phi, U, P, Q)$ such that:*

(1) *There are smooth linearizing charts*

$$U_P,\, U_Q \simeq [-1,1]^q \times [-1,1]^{p-q} \times [-1,1]^{\dim(M)-p}$$

(defined on neighbourhoods of P and Q), where ϕ is a linear map such that, for every $x \in U_P \cap \phi^{-1}(U_P)$ or $x \in U_Q \cap \phi^{-1}(U_Q)$, we have:

(a) *In these charts, both P and Q correspond to $\{0\}^{\dim(M)}$ and $\phi_*(P) = \varphi_*(P)$ and $\phi_*(Q) = \varphi_*(Q)$,*

(b) *The foliation by q-planes parallel to $[-1,1]^q \times \{0\}^{p-q} \times \{0\}^{\dim(M)-p}$ (called the* strong stable foliation, *$\mathcal{F}^s$) is locally invariant and corresponds to the smallest (in modulus) eigenvalues of the linear maps induced by ϕ in U_P and U_Q.*

(c) *The foliation by $(p-q)$-planes parallel to $\{0\}^q \times [-1,1]^{p-q} \times \{0\}^{\dim(M)-p}$ (called the* central foliation, *$\mathcal{F}^c$) is locally invariant.*

(d) *The foliation by $(\dim(M) - p)$-planes parallel to $\{0\}^q \times \{0\}^{p-q} \times [-1,1]^{\dim(M)-p}$ (called the* strong unstable foliation, *$\mathcal{F}^u$) is locally invariant and corresponds to the biggest (in modulus) eigenvalues of the linear maps induced by ϕ in U_P and U_Q.*

(2) *There exist points $X_0 \in (W^u(Q) \pitchfork W^s(P)) \cap U_Q$ and $Y_0 = \phi^{k_0}(X_0) \in U_P$, $k_0 > 0$, such that, in these coordinates, $X_0 \in \{0\}^q \times [-1,1]^{p-q} \times \{0\}^{\dim(M)-p}$ (the* central leaf *through Q) and $Y_0 \in \{0\}^q \times [-1,1]^{p-q} \times \{0\}^{\dim(M)-p}$ (the* central leaf *through P).*

(3) *There exist points $X_1 \in (W^s(Q) \cap W^u(P)) \cap U_P$ and $Y_1 = \phi^{k_1}(X_1) \in U_Q$, $k_1 > 0$, such that, in these coordinates, $X_1 \in \{0\}^q \times \{0\}^{p-q} \times [-1,1]^{\dim(M)-p}$ (the* local unstable manifold *of P, $W^u_{loc}(P)$) and $Y_1 \in [-1,1]^q \times \{0\}^{p-q} \times \{0\}^{\dim(M)-p}$ (the* local stable manifold *$W^s_{loc}(Q)$ of Q).*

(4) *There exist small cubes $C_0 \subset U_Q$ and $C_1 \subset U_P$ centered at X_0 and X_1, respectively, such that*

(a) *$\phi^{k_0}(C_0) \subset U_P$ and $\phi^{k_1}(C_1) \subset U_Q$,*

(b) *the restrictions $T_0 = \phi^{k_0}|_{C_0}$ and $T_1 = \phi^{k_1}|_{C_1}$ are affine maps which preserve the strong stable, central and strong unstable foliations above.*

Proof. — We first consider a heteroclinic point $X \in W^u(Q) \pitchfork W^s(P)$. After an arbitrarily small perturbation of φ, we can assume that X does not belong to the strong unstable manifold of Q nor in the strong stable manifold of P. After a new perturbation, we can assume that φ is linear in small neighbourhoods U_P of P and U_Q of Q. So, we now consider the foliations $\mathcal{F}^s_P$, $\mathcal{F}^u_P$ and $\mathcal{F}^c_P$ (resp. $\mathcal{F}^s_Q$, $\mathcal{F}^u_Q$ and $\mathcal{F}^c_Q$) defined in these linearizing chart U_P (resp. U_Q) as in the item (1) in the lemma.

Considering a heteroclinic point X as above and, using the domination, we have that the backward orbit of X approaches the central leaf through Q. Similarly, the forward iterates of X approach the central leaf through P. We will make a $\mathcal{C}^1$-perturbation of φ in such a way that, after a sufficiently large number of backward (resp. forward) iterations, the orbit of X is in the central leaf of Q (resp. P), that is, in coordinates, these points are in $\{0\}^q \times [-1,1]^{p-q} \times \{0\}^{\dim(M)-p}$. To get this perturbation for the backward orbit, first observe that, due to the domination, the distance between two consecutive iterates of $\varphi^{-n}(X)$ and $\varphi^{-n-1}(X)$, big n, is larger than their distances to the central leaf through Q. More precisely, the ratio between these two distances goes to infinite. So, taking a large $i > 0$, there is a diffeomorphism θ, $\mathcal{C}^1$-close to the identity, coinciding with the identity outside a small neighbourhood U of $\varphi^{-i}(X)$ intersecting the orbit of X only at $\varphi^{-i}(X)$, and such that $\theta(\varphi^{-i}(X))$ belongs to the central leaf through Q. Then $\psi = (\varphi \circ \theta^{-1})$ is a $\mathcal{C}^1$-perturbation of φ such that $\psi^{-n}(X)$ belongs to the central leaf through Q, for every n big enough. Moreover, the forward orbit of X is not modified. We now repeat the previous construction for the forward orbit of X obtaining the announced perturbation (already denoted by φ). Observe that we can perform all the previous perturbations without breaking the cycle (i.e., preserving the non-transverse intersection between $W^s(Q)$ and $W^u(P)$).

Now, there exist some backward iterate X_0 of X in the central leaf through Q and $k_0 > 0$ such that $Y_0 = \varphi^{k_0}(X_0)$ belongs to the central leaf through P. Observe now that the points X_1 and $Y_1 = \varphi^{k_1}(X_1)$ in the lemma are directly given by the intersection $W^s(Q) \cap W^u(P)$.

Recall that φ was constructed to be linear in small neighbourhoods of P and of Q. By a new small $\mathcal{C}^1$-perturbation, we can assume that φ^{k_0} and φ^{k_1} are both affine in small neighbourhoods of X_0 and X_1.

The only thing that remains to do in order to prove the lemma is to notice that (after new perturbations, if necessary) these affine maps can be chosen preserving the foliations (strong stable, central and strong unstable). The proof of this fact follows from a similar argument, actually, it follows as in the proof of [**BDP**, Lemma 4.13] using the domination. More precisely, in our linearizing charts U_P and U_Q, we consider the center-stable foliations $\mathcal{F}_P^{cs}$ and $\mathcal{F}_Q^{cs}$ (resp., center-unstable foliations $\mathcal{F}_P^{cu}$ and $\mathcal{F}_Q^{cu}$) tangent to the sum $E_1 \oplus E_2$ of the stable and central directions (resp., the sum $E_2 \oplus E_3$ of the central and unstable directions). By genericity, we can assume that the images by φ^{k_0} of the foliations $\mathcal{F}_Q^s$, $\mathcal{F}_Q^u$, $\mathcal{F}_Q^c$, $\mathcal{F}_Q^{cs}$ and $\mathcal{F}_Q^{cu}$ are in general position with respect to the foliations $\mathcal{F}_P^s$, $\mathcal{F}_P^u$, $\mathcal{F}_P^c$, $\mathcal{F}_P^{cs}$ and $\mathcal{F}_P^{cu}$ in a neighbourhood of Y_0. Now, the forward iterates of the images by φ^{k_0} of the leaves of $\mathcal{F}_Q^{cu}$ become closer to the center-unstable leaves in U_P. Replacing the initial k_0 by $k_0+\ell$, for some ℓ large enough, and making a small perturbation, one gets an invariant center-unstable foliation. More precisely, as above, we compose φ with a small $\mathcal{C}^1$-perturbation of the identity supported on a small neighbourhood of $\varphi^\ell(Y_0)$ mapping the foliation $\varphi^{k_0+\ell}(\mathcal{F}_Q^{cu})$ into $\mathcal{F}_P^{cu}$. Moreover, we choose this perturbation of the identity in order to fix the point $\varphi^\ell(Y_0)$. We now replace φ by the resulting composition, k_0 by $k_0+\ell$ and Y_0 by $\varphi^\ell(Y_0)$.

To get the invariance of the strong stable foliation we consider negative iterates of the foliations in the neighbourhood of Y_0. By the previous construction, the center-unstable foliation is preserved by negative iterations. So the negative iterates of the strong stable foliation are transverse to the center-unstable one. As above, the backward iterates of the strong stable leaves approach the leaves of the strong stable foliation in U_Q. So we can replace X_0 by some (large) negative iterate of it, say $-\ell'$, and perform a small perturbation (preserving the center-unstable foliation) such that the transition map $\varphi^{k_0+\ell+\ell'}$ from a neighbourhood of X_0 to a neighbourhood of Y_0 preserves the strong stable and center-unstable foliations.

To get the invariance of the strong unstable and center foliations (keeping the invariance of the strong stable), we repeat all the arguments above inside the center-unstable foliation. We omit the details of this construction. This gives the transition T_0.

The transition T_1 is obtained using the same arguments. The proof of the lemma is now complete. □

Definition 3.1. — Consider a $\dim(M)$-cube $C = I^s \times I^c \times I^u$, where I^s is a q-cube, I^c a $(p-q)$-cube and I^u a $(\dim(M)-p)$-cube. In these cubes we define coordinates (x^s, x^c, x^u) as above.

A subset Δ of C is s-complete if, for every $Z = (z^s, z^c, z^u) \in \Delta$, the horizontal q-cube $I^s \times \{(z^c, z^u)\}$ is contained in Δ. Similarly, a subset Δ of C is u-complete if, for every point $Z \in \Delta$, the vertical $(\dim(M)-p)$-cube $\{z^s, z^c\} \times I^u$ is contained in Δ.

By shrinking, if necessary, the size of the neighbourhood U_Q in the strong unstable direction and taking an appropriate cube C_1 around X_1, we can assume that the image by T_1 of any u-complete disk Δ of C_1 (contained in a leaf of the strong unstable foliation) is a u-complete disk of U_Q.

For simplicity, let us denote A and B the restrictions of ϕ to U_Q and U_P, respectively.

Lemma 3.3. — *There exists a natural number $\ell_0 \geqslant 0$ satisfying the following conditions:*

(1) *Consider any $Z \in W^u_{loc}(Q)$ and any s-complete disk Δ^s of C_0 (contained in a leaf of the strong stable foliation) containing Z. Then the connected component of $A^{-n}(\Delta^s) \cap U_Q$ containing $A^{-n}(Z)$ is a s-complete disk in U_Q for all $n \geqslant \ell_0$.*

(2) *Consider any u-complete disk Δ^u of C_0 (in a leaf of the strong unstable foliation). Then the intersection between Δ^u and $T_0^{-1}(W^s_{loc}(P))$ is a unique point W. Let Δ^u_m be the connected component of $(B^m \circ T_0(\Delta^u)) \cap U_P$ containing $B^m \circ T_0(W)$. Then $\Delta^u_m \cap C_1$ is a complete u-disk (in C_1) for every $m \geqslant \ell_0$.*

Proof. — Recall that both foliations are invariant by the action of A and B. So, the proof follows, since A^{-1} expands the s-direction and B expands the u-direction. □

We are now ready to finish the proof of Theorem 3.1. Given $\varepsilon > 0$, there is an $\varepsilon/2$-perturbation ϕ of φ satisfying Lemmas 3.2 and 3.3. We will now obtain the final diffeomorphism considering a perturbation of ϕ obtained by composing the transition T_1 with a small translation T_v generated by a vector v, parallel to the central direction (in U_Q). Let us now go through the details of this construction.

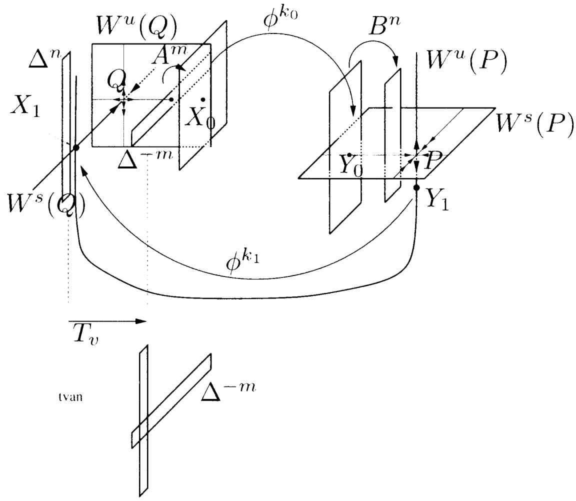


FIGURE 2. A periodic orbit

In our coordinates, $X_0 = (0^s, x_0^c, 0^u)$. Consider now the su-disk

$$\Delta = \left([-1,1]^q \times \{x_0^c\} \times [-1,1]^{\dim(M)-p}\right) \cap C_0.$$

With the terminology above, the disk Δ is u and s-complete in C_0.

Given n and m bigger than ℓ_0 (ℓ_0 as in Lemma 3.3), let Δ^{-m} and Δ_0^n be the connected components of $A^{-m}(\Delta) \cap U_Q$ containing $A^{-m}(X_0)$ and of $(B^n \circ T_0(\Delta)) \cap U_P$ containing $B^n(T_0(X_0))$, respectively. Let $\Delta_1^n = \Delta_0^n \cap C_1$. Write $\Delta^n = T_1(\Delta_1^n)$. By Lemma 3.3 and the observation before, Δ^{-m} and Δ^n are s-complete and u-complete disks in U_Q and C_1, respectively.

Observe that there is a unique vector v, parallel to the central direction, such that the intersection between $T_v(\Delta^n)$ and Δ^{-m} is non-empty. Moreover, since these sets are both su-disks of U_Q, such an intersection is a sub-rectangle R intersecting *completely* Δ^{-m} in the u-direction and Δ^n in the s-direction. Here by a *complete intersection in the u-direction* we mean that, for every $Z \in R$, the leaf $F^u(Z)$ of the strong unstable foliation containing Z is such that the connected components of $F^u(Z) \cap R$ and of $\Delta^{-m} \cap F^u(Z)$ containing Z are equal. The definition of *complete intersection in the s-direction* is totally analogous (considering strong stable leaves).

Now a classical argument of hyperbolicity implies that the map $T = T_v \circ T_1 \circ B^n \circ T_0 \circ A^m$ has a fixed point W in Δ^{-m}. Observe that the derivative of T at W is $\widetilde{T}_1 \circ B^n \circ \widetilde{T}_0 \circ A^m$ (where $\widetilde{T}_i$ is the linear part of the affine map T_i).

So it remains to check that the size of the translation T_v can be chosen to be smaller than $\varepsilon/2$. For that, first observe that the disks Δ^{-m} and Δ^n can be taken arbitrarily close to the heteroclinic intersection Y_1 (it is enough to take n and m large enough). Thus, there exists $n_0 \in \mathbb{Z}$ such that the distance between Δ^{-m} and Δ^n is less than $\varepsilon/2$, for every n and m greater than n_0. Fixing this n_0 and replacing T_0 by $T_0 \circ A^{n_0}$ and T_1 by $T_1 \circ B^{n_0}$, we get that, for every positive n and m, there exists a translation T_v, $v = v(n, m)$, such that the modulus of v is less than $\varepsilon/2$.

The diffeomorphism ψ in the statement of the theorem is obtained from ϕ by composing T_1 with T_v. By construction, ψ has a periodic point R of period $n_R = t_0 + t_1 + n + m$, where $t_0 = k_0 + n_0$ and $t_1 = k_1 + n_1$, such that

$$\psi_*^{n_R}(R) = \widetilde{T}_1 \circ B^n \circ \widetilde{T}_0 \circ A^m.$$

Notice that t_0 and t_1 depend exclusively on the transitions T_0 and T_1. The theorem now follows from the definition of A and B and the lemma below, that allows us to perform any small perturbation of the derivative of a diffeomorphism along the orbit of a periodic point in a dynamical way.

Lemma 3.4 **([F], [M$_2$])**. — *Consider a $\mathcal{C}^1$-diffeomorphism φ and a φ-invariant finite set Σ. Let A be an ε-perturbation of φ_* along Σ (i.e., the linear maps $A(x)$ and $\varphi_*(x)$ are ε-close for all $x \in \Sigma$). Then, for every neighbourhood U of Σ, there is a diffeomorphism ϕ, $\mathcal{C}^1$-ε-close to φ, such that*

- *$\varphi(x) = \phi(x)$ if $x \in \Sigma$ or if $x \notin U$, and*
- *$\phi_*(x) = A(x)$ for all $x \in \Sigma$.*

The proof of Theorem 3.1 is now complete. □

We end this subsection by stating a lemma that follows from the proof of [**BDP**, Lemma 4.13]:

Lemma 3.5. — *Let M_P and M_Q be linear maps as in the statement of Theorem 3.1. Suppose that M_P and M_Q preserve the dominated splittings $T_P M = E_P^1 \oplus \cdots \oplus E_P^k$ and $T_Q M = E_Q^1 \oplus \cdots \oplus E_Q^k$, where $\dim(E_P^i) = \dim(E_Q^i)$ for every i. Then the matrices T_0 and T_1 in Theorem 3.1 can be chosen in such a way that:*

$$T_0(E_P^i) = E_Q^i \text{ and } T_1(E_Q^i) = E_P^i, \quad \text{for every } i \in \{1, \ldots, k\}.$$

3.2. Periodic points in the unfolding of heterodimensional cycles. — Using Lemma 3.4 we get the following two corollaries of Theorem 3.1. First we use the notation $\Gamma(\varphi, U, P, Q)$ to localize a cycle, that is, if we are only concerned with the intersection between the invariant manifolds of P and Q whose orbit is contained in U.

Corollary 3.6. — *Consider a heterodimensional cycle $\Gamma(\varphi, U, P, Q)$ associated to the hyperbolic periodic points P and Q of indices p and q, where $p > q$, having positive real eigenvalues of multiplicity one. Then, for every integer $\ell \in [q, p]$, there is a diffeomorphism ϕ arbitrarily close to φ with a hyperbolic periodic point of index ℓ in $\Lambda_\phi(U)$.*

Proof. — This corollary is trivial when $\ell = p$ or q. So let us fix some $\ell \in]q, p[$. Define the matrices M_P and M_Q as in the statement of Theorem 3.1 and denote by $\lambda_P^1, \cdots, \lambda_P^{\dim(M)}$ the eigenvalues of M_P, where $0 < \lambda_P^1 < \cdots < \lambda_P^{\dim(M)}$, and by $\lambda_Q^1, \cdots, \lambda_Q^{\dim(M)}$ the eigenvalues of M_Q, where $0 < \lambda_Q^1 < \cdots < \lambda_Q^{\dim(M)}$.

For each $i \in \{1, \dots, \dim(M)\}$, let $E^i(P)$ and $E^i(Q)$ be the eigenspaces corresponding to λ_P^i and λ_Q^i, respectively. We now consider the invariant splittings (of M_P and M_Q) given by

$$\begin{aligned} E_1(P) &= E^1(P) \oplus \cdots \oplus E^{\ell-1}(P), \quad E_2(P) = E^\ell(P), \\ E_1(Q) &= E^1(Q) \oplus \cdots \oplus E^{\ell-1}(Q), \quad E_2(Q) = E^\ell(Q), \end{aligned}$$

$$\begin{aligned} E_3(P) &= E^{\ell+1}(P) \oplus \cdots \oplus E^{\dim(M)}(P), \\ E_3(Q) &= E^{\ell+1}(Q) \oplus \cdots \oplus E^{\dim(M)}(Q). \end{aligned}$$

Observe that, by the hypotheses on the eigenvalues of P and Q, the splittings $E_1(R), E_2(R)$ and $E_3(R)$, $R = P, Q$, are dominated (for M_P and M_Q), therefore they satisfy the hypotheses of Theorem 3.1.

Since $q < \ell < p$, we have that $\lambda_P^\ell < 1 < \lambda_Q^\ell$. Thus, there are constants C and C', $0 < C < 1 < C'$, and arbitrarily large natural numbers n_0 and m_0 such that

$$(\lambda_P^{\ell-1})^{n_0}\,(\lambda_Q^{\ell-1})^{m_0} < C < (\lambda_P^{\ell})^{n_0}\,(\lambda_Q^{\ell})^{m_0} < C' < (\lambda_P^{\ell+1})^{n_0}\,(\lambda_Q^{\ell+1})^{m_0}.$$

Applying Theorem 3.1 to the matrices M_P and M_Q, $n = n_0$, $m = m_0$, and the matrices $I_0, \dots, I_{n+m+2}$ equal to the identity, we get transitions T_0 and T_1 and a diffeomorphism ϕ close to φ having a periodic point $R \in \Lambda_\phi(U)$ of period $n(R) \simeq n_0 + m_0$ such that $\phi_*^{n(R)}$ is conjugate to

$$M_R = T_1 \circ M_Q^{m_0} \circ T_0 \circ M_P^{n_0}.$$

By Lemma 3.5, we can suppose that T_0 and T_1 preserve the splittings $E_1 \oplus E_2 \oplus E_3$. Hence, the ℓ^{th}-eigenvalue λ_R^ℓ of M_R is such that

$$\frac{C}{k_1} < |\lambda_R^\ell| < k_2\, C',$$

where k_1 is the product of the norms of T_0^{-1} and T_1^{-1}, and k_2 is the product of the norms of T_0 and T_1. Observe that, a priori, we cannot guarantee that this eigenvalue is positive (we do not know if the transitions preserve the orientation). Thus, taking n_0 and m_0 large enough, we can assume that $|\log(\lambda_R^\ell)|/(n_0 + m_0)$ is arbitrarily close to zero.

Applying now Lemma 3.4 to the derivative of ϕ along the orbit of R, we can assume that the eigenvalues $\lambda_R^1, \dots, \lambda_R^{\dim(M)}$ of $\phi_*^{n(R)}(R)$ satisfy

$$(1) \qquad 0 < |\lambda_R^1| < \cdots < |\lambda_R^{\ell-1}| < 1 = |\lambda_R^{\ell}| < |\lambda_R^{\ell+1}| < \cdots < |\lambda_R^{\dim(M)}|.$$

After a final perturbation, we have that R has index ℓ, finishing the proof of the corollary. □

Finally, a minor modification of the proof of Corollary 3.6 gives the following:

Corollary 3.7. — *Consider a heterodimensional cycle $\Gamma(\varphi, U, P, Q)$ satisfying the hypothesis of Theorem 3.1. Moreover, suppose that there is a dominated splitting $F_1 \oplus \cdots \oplus F_i \oplus \cdots \oplus F_k$ over $\Lambda_\varphi(U)$ such that the moduli of the Jacobians of φ restricted to F_i along the orbits of Q and P are strictly bigger and less than one, respectively.*

Then, there exists a diffeomorphism ϕ, arbitrarily $\mathcal{C}^1$-close to φ, with a hyperbolic periodic point $R \in \Lambda_\phi(U)$ such that the modulus of the Jacobian of $\phi^{n(R)}$ over F_i at R is equal to one.

Proof. — Consider the dominated splittings

$$E_1 = F_1 \oplus \cdots \oplus F_{i-1}, \quad E_2 = F_i \quad E_3 = F_{i+1} \oplus \cdots \oplus F_k.$$

Just observe that by Lemma 3.5 we can choose the transitions T_i preserving the dominated splitting $E_1 \oplus E_2 \oplus E_3$. The result follows from a similar argument we gave in Corollary 3.6. □

3.3. End of the proof of Theorem A. — We need the following lemma:

Lemma 3.8 (**[BDP**, Lemma 5.4]). — *Let V be an open subset of M and R_φ a hyperbolic periodic point of a diffeomorphism φ, such that its relative homoclinic class in V, $\overline{H_{R_\varphi}(V)}$, is non trivial. Then there is a diffeomorphism ϕ arbitrarily $\mathcal{C}^1$-close to φ such that $\overline{H_{R_\phi}(V)}$ contains a hyperbolic periodic point of the same index of R_ϕ, whose eigenvalues are all real, positive and of multiplicity one.*

Under the hypothesis of Theorem A, this lemma allows us to assume that, after perturbing the original diffeomorphism and replacing the initial points P_φ and Q_φ by other points of $\Lambda_\varphi(U)$ of the same index, we can assume that the points P_φ and Q_φ of $\Lambda_\varphi(U)$ have real positive eigenvalues of multiplicity one. To check this just notice that, by Theorem 2.3, after a $\mathcal{C}^1$-perturbation of φ, we can assume that $\overline{H_{P_\varphi}(U)} = \overline{H_{Q_\varphi}(U)} \subset \Lambda_\varphi(U)$. Therefore, these two relative homoclinic classes are non-trivial. Hence, we can now apply Lemma 3.8 to such homoclinic classes to get the periodic points (of indices p and q) in $\Lambda_\varphi(U)$ with real positive eigenvalues of multiplicity one. So there is no loss of generality if we assume that the points P_φ and Q_φ in Theorem A have real positive eigenvalues of multiplicity one. Using Lemma 2.5 and Corollary 3.6 one gets:

Lemma 3.9. — *Given $p > q$ and $\ell \in]q, p]$ let $\varphi \in \mathcal{M}(U)$ be a diffeomorphism with two hyperbolic periodic points P_φ and Q_φ in $\Lambda_\varphi(U)$ (of indices p and q), having positive real eigenvalues of multiplicity one. Then there is $\phi \in \mathcal{M}(U)$ arbitrarily $\mathcal{C}^1$-close to φ having a hyperbolic periodic point of index ℓ in $\Lambda_\phi(U)$.*

Proof. — By hypothesis, the continuations P_ϕ and Q_ϕ of P_φ and Q_φ are transitively related for every ϕ in a neighbourhood of φ in $\mathcal{M}(U)$ (just observe that set $\Lambda_\phi(U)$ is robustly transitive and P_ϕ and Q_ϕ belong to $\Lambda_\phi(U)$). Hence we can apply Lemma 2.5 to P_φ and Q_φ to create a heterodimensional cycle $\Gamma(\psi, U, P_\psi, Q_\psi)$ for some ψ arbitrarily close to φ. Corollary 3.6 now gives ϕ close to ψ (thus close to φ) with a periodic point of index ℓ in $\Lambda_\phi(U)$, finishing the proof of the lemma. $\square$

Given $\varphi \in \mathcal{M}(U)$, consider a neighbourhood $\mathcal{U}_\varphi$ of φ in $\mathcal{M}(U)$ such that every $\psi \in \mathcal{U}_\varphi$ has hyperbolic periodic points of indices q and p. Let $\mathcal{H}_j$ be the set of diffeomorphisms $\psi \in \mathcal{U}_\varphi$ having some hyperbolic periodic point of index j in $\Lambda_\varphi(U)$. Applying Lemma 3.9 finitely many times, one gets that the sets $\mathcal{H}_j$, $j \in [q, p]$, are dense in $\mathcal{U}_\varphi$.

Theorem A now follows by observing that, for every j, the set $\mathcal{H}_j$ is open. Now it is enough to consider the set $\cap_q^p \mathcal{H}_j$, which is a dense open subset of $\mathcal{U}_\varphi$. So, we have just finished the proof of Theorem A.

4. Hyperbolicity of the extremal bundles

In this section, we will prove Theorem B. For that, as in the hypotheses of this theorem, consider an open subset U of a compact manifold M and $q \in \mathbb{N}^*$. Let $\mathcal{U}$ be a $\mathcal{C}^1$-open set of $\mathrm{Diff}^1(M)$ such that, for every diffeomorphism $\phi \in \mathcal{U}$, the set $\Lambda_\phi(\overline{U})$ has a dominated splitting $E_\phi \oplus F_\phi$ with $\dim(E_\phi(x)) = q$ for all $x \in \Lambda_\phi(\overline{U})$. Suppose that every $\phi \in \mathcal{U}$ has no periodic points of index $r < q$. Then we prove that the bundle E_ϕ is uniformly contracting for every $\phi \in \mathcal{U}$.

The proof of this result follows using the arguments in [$\mathbf{M}_2$] after some small technical modifications. Therefore, we will just sketch this proof, emphasizing the main modifications that we need to introduce.

The results in [$\mathbf{M}_2$] are formulated in terms of *families of periodic sequences of linear maps.* It is considered the family obtained by taking all the diffeomorphism ϕ in an open set of $\mathrm{Diff}^1(M)$ and the restrictions of the derivatives of these diffeomorphisms to their periodic orbits. It is considered perturbations of this system of linear maps without paying attention if such perturbations come from perturbations of the initial diffeomorphism. However, a Lemma of Franks' (see Lemma 3.4 above) allows one to perform dynamically the perturbation of the derivative: given a diffeomorphism φ and a periodic point x of φ, to each perturbation A of the derivative φ_* throughout the orbit of x corresponds a diffeomorphism ψ $\mathcal{C}^1$-close to φ which preserves the φ-orbit of x and such that $A(z) = \psi_*(z)$ for all z in the φ-orbit of x.

We begin by recalling some results about dominated splittings, see next section. In Section 4.2 we recall the terminology about families of periodic linear systems and some results in [$\mathbf{M}_2$]. Finally, in Section 4.3 we prove Theorem B.

4.1. Remarks on dominated splittings. — In this subsection, we state precisely some folklore results on dominated splittings. Before that, let us observe that, if $\Lambda_\varphi(U)$ is robustly transitive, then, by definition, it is a φ-invariant compact subset of U which is the maximal φ-invariant set of $\overline{U}$. This implies that, for any neighbourhood V of $\Lambda_\varphi(U)$ and every diffeomorphism ϕ close to φ, the set $\Lambda_\phi(U)$ coincides with $\Lambda_\phi(\overline{U})$ and is contained in V. Thus $\Lambda_\phi(U)$ depends lower-semi-continuously on ϕ. We say that $\Lambda_\phi(U)$ is the *continuation* of $\Lambda_\varphi(U)$ for ϕ.

Lemma 4.1. — *Let φ be a diffeomorphism and U an open subset of M such that $\Lambda_\varphi(U)$ coincides with $\Lambda_\varphi(\overline{U})$ and admits a dominated splitting $T_{\Lambda_\varphi(U)}M = E \oplus F$, $E \prec F$. Then, for every diffeomorphism ψ close enough to φ, there is a unique dominated splitting $E_\psi \oplus F_\psi$, $E_\psi \prec F_\psi$, defined on $\Lambda_\psi(U)$, such that* $\dim(E_\psi) = \dim(E)$.

The splitting $E_\psi \oplus F_\psi$ above is the *continuation* of $E \oplus F$. Moreover, the continuations E_ψ and F_ψ depend continuously on ψ. This lemma also holds for dominated splittings with an arbitrary number of bundles.

Proof. — Let us just sketch the proof of the lemma. By the definition of domination, there is a strictly φ_*-invariant continuous cone field $\mathcal{C}^+$ defined over $\Lambda_\varphi(U)$ such that the bundle F is obtained as the intersection of the forward φ_*-iterates of the cones of $\mathcal{C}^+$. Similarly, there is a strictly (φ_*^{-1})-invariant continuous cone field $\mathcal{C}^-$ defined over $\Lambda_\varphi(U)$ such that the intersections of the backward iterates of $\mathcal{C}^-$ define E. These cone fields can be extended continuously to invariant cone fields $\mathcal{C}_0^+$ and $\mathcal{C}_0^-$ defined on a compact neighbourhood V of $\Lambda_\varphi(U)$.

Observe that every ψ close to φ leaves invariant the cone fields $\mathcal{C}_0^+$ and $\mathcal{C}_0^-$ and recall that $\Lambda_\psi(U) \subset V$. We now define the bundles E_ψ and F_ψ as the intersection of the (backward and forward, respectively) iterates by ψ_* of the cones of $\mathcal{C}_0^-$ and $\mathcal{C}_0^+$, respectively. By construction, the splitting $E_\psi \oplus F_\psi$ is dominated and satisfies $\dim E_\psi = \dim E$.

For the continuous dependence of the bundles E_ψ and F_ψ on the diffeomorphism ψ we refer the reader to [**BDP**, Lemma 1.4], for instance. This ends the sketch of the proof. □

Lemma 4.2 ([**BDP**, Lemma 1.4]). — *Let ϕ be a diffeomorphism and Σ a ϕ-invariant set having a dominated splitting $E \oplus F$. Then this splitting can be extended (in a dominated way) to the closure of Σ.*

Remark 4.3. — Let φ be a diffeomorphism, K a transitive φ-invariant compact set, $T_K M = E_1 \oplus E_2 \oplus \cdots \oplus E_m$ the finest dominated splitting of φ over K, and $\Sigma \subset K$

a φ-invariant dense subset of K. Then the finest dominated splitting of φ over Σ is given by the restriction to Σ of the bundles E_i.

Proof of the remark. — We argue by contradiction. Suppose that there is a dominated splitting over Σ which refines the splitting given by the restrictions to Σ of the bundles E_i. Then, by Lemma 4.2, such a splitting can be extended to the whole K, contradicting that the splitting $E_1 \oplus \cdots \oplus E_m$ is the finest one. □

Let us state a final result, whose proof is here omitted.

Remark 4.4. — Let φ be a diffeomorphism and E a φ_*-invariant bundle defined on a φ-invariant compact set K_1. Consider any φ-invariant dense subset K_2 of K_1. Then,

– the bundle E is uniformly hyperbolic over K_1 if and only if its restriction to K_2 is uniformly hyperbolic,

– the diffeomorphism φ contracts (resp. expands) uniformly the volume in E over K_1 if and only if it contracts uniformly (resp. expands) the volume in E over K_2.

4.2. Families of periodic sequences of linear maps and dominated splittings. — We begin this section by recalling some definitions in [$\mathbf{M}_2$].

Definition 4.1

(1) A *periodic sequence of linear maps* is a periodic map $\xi\colon \mathbb{Z} \to GL(N,\mathbb{R})$, $n \mapsto \xi_n$. We denote this family by $\{\xi_n\}$.

(2) A periodic sequence of linear maps $\{\xi_n\}$ of period n is called *contracting* if the product $\xi_{n-1} \circ \cdots \circ \xi_0$ is an uniform contraction, i.e., all its eigenvalues have modulus strictly less than 1.

(3) Consider a family $\Xi = \{\xi^\alpha = (\xi^\alpha_n)_{n\in\mathbb{Z}}\}_{\alpha\in\mathcal{A}}$ of periodic sequences of linear maps, such that the norms $\|\xi^\alpha_n\|$ and $\|(\xi^\alpha_n)^{-1}\|$ are uniformly bounded (independently of n and α). The family Ξ is *robustly contracting*[(1)] if there is $\varepsilon > 0$ such that any family $\Theta = \{\theta^\alpha\}_{\alpha\in\mathcal{A}}$ having the same period function $n(\alpha)$ and ε-close to Ξ (i.e., $\|\theta^\alpha_n - \xi^\alpha_n\| < \varepsilon$ for all $\alpha \in \mathcal{A}$ and $n \in \mathbb{Z}$) is contracting.

The example of family of periodic sequence of linear maps that will be play a key role in the proof of Theorem B is obtained as follows. Let $\phi \in \mathcal{U}$, $\mathcal{U}$ as in Theorem B, and $\delta > 0$ such that every diffeomorphism ψ which is 2δ-$\mathcal{C}^1$-close to ϕ belongs to $\mathcal{U}$. Now let $\mathcal{A}_\phi$ be the set of pairs $\alpha = (x,\psi)$ such that ψ is δ-close to ϕ and the ψ-orbit of x is contained in U and periodic. Consider now some trivialization of the bundles E_ψ (as in Theorem B) over the set of periodic points (by choosing an orthonormal basis of $E_\psi(x)$) and, for each $\alpha = (x,\psi) \in \mathcal{A}_\phi$, define ξ^α as being the restrictions of the differential ψ_* to $\{E_\psi(\psi^i(x))\}_{i\in\mathbb{Z}}$. We now have that $\Xi_\phi = \{\xi^\alpha\}_{\alpha\in\mathcal{A}_\phi}$ is a family of periodic sequences of linear maps.

[(1)]This notion is called *uniformly contracting* in [$\mathbf{M}_2$], but we rename it to avoid ambiguity with the now usually accepted notion of uniform hyperbolicity or uniform contraction.

Lemma 4.5. — *The family Ξ_ϕ defined above is robustly contracting.*

Proof. — The proof is by contradiction. Otherwise, there exist $(x,\psi) \in \mathcal{A}_\phi$ and a linear map ν corresponding to a perturbation of the restriction of the differential of ψ to E_ψ along the periodic ψ-orbit of x, having an eigenvalue of modulus bigger or equal than one, i.e.,

$$\nu(\psi^{n(x)-1}(x)) \circ \cdots \circ \nu(x) \colon E_\psi(x) \longrightarrow E_\psi(x)$$

has an eigenvalue λ such that $|\lambda| \geqslant 1$, where $n(x)$ is the ψ-period of x.

Using Lemma 3.4, we get a diffeomorphism ζ close to ψ, thus in $\mathcal{U}$, such that x is a periodic point of $\Lambda_\zeta(\overline{U})$ and

$$\zeta_*^{n(x)}(x) = \zeta_*(\zeta^{n(x)-1}) \circ \cdots \zeta_*(x) = \nu(\psi^{n(x)-1}(x)) \circ \cdots \circ \nu(x).$$

Therefore, the restriction of $\zeta_*^{n(x)}(x)$ to $E_\zeta(x)$ has at most $(q-1)$ eigenvalues of modulus (strictly) less than one. On the other hand, by the domination $E_\psi \prec F_\psi$, the eigenvalues of the restriction of $\zeta_*^{n(x)}(x)$ to $F_\zeta(x)$ are all strictly bigger than one in modulus. This implies that there is a periodic point x in $\Lambda_\zeta(\overline{U})$ of index (strictly) less than q, contradicting the definition of $\mathcal{U}$. This contradiction finishes the proof of the lemma. □

We now borrow the following lemma from [**M**$_2$].

Lemma 4.6 ([**M**$_2$, Lemma II.7]). — *Let $\{\xi^{(\alpha)}, \alpha \in \mathcal{A}\}$ be a robustly contracting family of periodic sequences of isomorphisms of $\mathbb{R}^N$. Then, there exist $K > 0$, $0 < \lambda < 1$ and $m \in \mathbb{N}^*$ such that:*

a) *if $\alpha \in \mathcal{A}$ and ξ^α has minimum period $n \geqslant m$, then*

$$\prod_{j=0}^{k-1} \left\| \prod_{i=0}^{m-1} \xi^{(\alpha)}_{i+mj} \right\| \leqslant K\lambda^k,$$

where k is the integer part of n/m;

b) *for all $\alpha \in \mathcal{A}$*

$$\limsup_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \log \left(\left\| \prod_{i=0}^{m-1} \xi^{(\alpha)}_{i+mj} \right\| \right) < 0.$$

Applying Lemma 4.6 to the family Ξ_ϕ defined above, we get the next proposition, which is a reformulation of [**M**$_2$, Proposition II.1]:

Proposition 4.7. — *Let $\phi \in \mathcal{U}$ ($\mathcal{U}$ as in Theorem B). Then, there exist a neighborhood $\mathcal{V}$ of ϕ and constants $K > 0$, $m \in \mathbb{N}^*$ and $0 < \lambda < 1$ such that, for every $g \in \mathcal{V}$ and every periodic point x of ψ whose orbit is contained in U,*

a) *If x has minimum period $n \geqslant m$ then*

$$\prod_{i=0}^{k-1} \left\|(\psi^m)_*(\psi^{mi}(x))|_{E_\psi(\psi^{mi}(x))}\right\| \leqslant K\lambda^k,$$

where k is the integer part of n/m.

b) *Moreover,*

$$\limsup_{r\to+\infty} \frac{1}{r} \sum_{i=0}^{r-1} \log\left(\left\|(\psi^m)_*(\psi^{mi}(x))|_{E_\psi(\psi^{mi}(x))}\right\|\right) < 0.$$

Theorem B will be a consequence of Proposition 4.7 and the Mañé's Ergodic Closing Lemma, that we now recall, for completeness:

***Theorem 4.8* (Ergodic Closing Lemma, [$\mathbf{M}_2$, Theorem A]).** — *Consider a diffeomorphism ϕ defined on a compact manifold. Then there is a ϕ-invariant set $\Sigma(\phi)$ (named set of* well closable points *of ϕ) such that:*

(1) *The set $\Sigma(\phi)$ has total measure (i.e. $\mu(\Sigma(\phi)) = 1$ for every ϕ-invariant probability measure μ).*

(2) *For every $x \in \Sigma(\phi)$ and $\varepsilon > 0$ there is a diffeomorphism ψ, which is ε-close to ϕ in the C^1-topology, such that x is periodic for ψ and the distance* $\operatorname{dist}(\phi^i(x), \psi^i(x)) < \varepsilon$ *for all $i \in [0, n(x,\psi)]$, where $n(x,\psi)$ is the period of x for ψ.*

4.3. End of the proof of Theorem B. — The proof of the theorem now follows through the same lines as the proof of [$\mathbf{M}_2$, Theorem B], see pages 520-524. We will recall the main steps of this proof and point out the changes we need to introduce.

Proof. — Let $\phi \in \mathcal{U}$. By compactness of the set $\Lambda_\phi(\overline{U})$, as in [$\mathbf{M}_2$] to get the uniform contraction of the bundle E_ϕ, it is enough to check that

$$\liminf_{n\to+\infty} \|\phi_*^n|_{E_\phi(x)}\| = 0.$$

We argue by contradiction. If ϕ_* is not uniformly contracting on E_ϕ over $\Lambda_\phi(\overline{U})$ then there exist a constant $\kappa > 0$, a point $x \in \Lambda_\phi(\overline{U})$ and $n_0 \in \mathbb{N}$ such that

$$\|\phi_*^n|_{E_\phi(x)}\| > \kappa > 0$$

for every $n > n_0$. We now choose a sequence j_n, $j_n \to +\infty$, such that the sequence of probabilities μ_n defined by

$$\mu_n = \frac{1}{j_n} \sum_{i=0}^{j_n-1} \delta(\phi^{mi}(x))$$

converges (in the weak topology) to a probability μ, where $\delta(z)$ is the Dirac measure at the point z and m is as in Proposition 4.7.

Let $\varphi^\phi = \log\|\phi_*^m|_{E_\phi}\|$. By Lemma 4.1 the bundle E_ϕ is continuous on $\Lambda_\phi(\overline{U})$, so φ^ϕ is continuous on $\Lambda_\phi(\overline{U})$. By the choice of x, one has $\int \varphi^\phi\, d\mu_n \geqslant 0$ for every n

sufficiently large. So $\int \varphi^\phi \, d\mu \geqslant 0$. Using Birkhoff's Theorem and the Ergodic Closing Lemma, we get a point $p \in \Lambda_\phi(\overline{U}) \cap \Sigma(\phi)$ such that

$$\lim_{n\to+\infty} \frac{1}{j_n} \sum_{i=0}^{j_n-1} \log \|\phi_*^m|_{E_\phi(\phi^{mi}(p))}\| \geqslant 0.$$

By item (b) of Proposition 4.7, the point p is not periodic. Now, by Theorem 4.8, there is ψ arbitrarily $\mathcal{C}^1$-close to ϕ (so $\psi \in \mathcal{V} \subset \mathcal{U}$, $\mathcal{V}$ as in Proposition 4.7) such that p is a periodic point of ψ of period $n(p)$ and the distance $\operatorname{dist}(\phi^i(p), \psi^i(p))$ is less than an arbitrarily small $\varepsilon > 0$, for every $i \in [0, n(p)]$. Observe that since p is not periodic for ϕ, the period $n(p)$ goes to infinity as ε goes to zero, i.e., ψ tends to ϕ.

Since the fibers $E_\psi(y)$ vary continuously with (y, ψ) (recall Lemma 4.1), the function

$$\Upsilon^\psi(y) = \log ||\psi_*^m|_{E_\psi(y)}||$$

is continuous. Now for λ as in Proposition 4.7 take λ_0 and $n_0 \in \mathbb{N}^*$ such that $\lambda < \lambda_0 < 1$ and for every $n \geqslant n_0$ one has

$$\frac{1}{n} \sum_{i=0}^{n-1} \Upsilon^\phi(\phi^{mi}(p)) \geqslant \frac{1}{2} \log(\lambda_0).$$

We can also assume that $K\lambda^n < \lambda_0^n$, for every $n \geqslant n_0$. So, if ψ is close enough to ϕ, then

$$|\Upsilon^\psi(\psi^i(p)) - \Upsilon^\phi(\phi^i(p))| < |\frac{1}{2} \log(\lambda_0)|$$

for every $i \in [0, n(p)]$. Moreover, the integer part k of $n(p)/m$ is greater than n_0. Therefore,

$$\frac{1}{k} \sum_{i=0}^{k-1} \Upsilon^\psi(\psi^{mi}(p)) \geqslant \log(\lambda_0) > \frac{1}{k} \log(K\lambda^k),$$

contradicting item (a) of Proposition 4.7. This contradiction finishes the proof of Theorem B. □

5. Proof of Theorem D

5.1. Perturbation of the derivative at periodic points. — In this section, we recall some results from [**BDP**]. These results are formulated in terms of *families of periodic linear systems,* that is, considering the differential of the diffeomorphism as an abstract linear cocycle over the set $\Lambda_\varphi(U)$ and perturbations of this cocycle, without taking in consideration if such perturbations come from perturbations of the diffeomorphism. However, as in Section 4, Lemma 3.4 allows us to perform dynamically the final abstract cocycle. Let us explain these results in detail.

Given a diffeomorphism φ and a hyperbolic periodic point P_φ of φ of index p, denote by Σ_{P_φ} the subset of $\overline{H_{P_\varphi}(U)}$ of hyperbolic periodic points R of index p homoclinically related to P_φ, i.e., $W^s(R) \pitchfork W^u(P_\varphi) \neq \emptyset$ and $W^u(R) \pitchfork W^s(P_\varphi) \neq \emptyset$. Observe that, in our setting, we can assume that Σ_{P_φ} is not trivial (different to the orbit of P_φ).

As above, given $x \in \Sigma_{P_\varphi}$, denote by M_x the matrix $M_x = \varphi_*^{n(x)}(x)\colon T_xM \to T_xM$, where $n(x)$ is the period of x. The first important property formalized in [**BDP**] is that the matrices M_x corresponding to different points of Σ_{P_φ} (the derivatives of $\varphi^{n(x)}$ at these points x) can be multiplied *essentially* how many times as one wants, and the resulting product corresponds to a matrix of the system at some different point. More precisely,

Lemma 5.1. — *Let φ be a diffeomorphism and P_φ a hyperbolic periodic point of φ. Consider any pair of periodic points of x and y of φ in Σ_{P_φ} and $\varepsilon > 0$. Suppose that M_x and M_y preserve invariant dominated splittings*

$T_xM = E_x^1 \oplus \cdots \oplus E_x^k,\ E_i(x) \prec E_{i+1}(x),$ *and* $T_yM = E_y^1 \oplus \cdots \oplus E_y^k,\ E_i(y) \prec E_{i+1}(y),$

such that $\dim(E_x^i) = \dim(E_y^i)$ *for every i. Then there is $\delta \in]0, \varepsilon[$ satisfying the following property:*

Given any pair of δ-perturbations $\widetilde{M}_x$ and $\widetilde{M}_y$ of M_x and M_y, respectively, $\widetilde{M}_x\colon T_xM \to T_xM$ and $\widetilde{M}_y\colon T_yM \to T_yM$, there exist linear maps

$$T_1 : T_xM \longrightarrow T_yM \quad \text{and} \quad T_2 \colon T_yM \longrightarrow T_xM$$

preserving the dominated splittings above (i.e., $T_1(E_x^i) = E_y^i$ and $T_2(E_y^i) = E_x^i$ for every i) and such that, for any $n \geqslant 0$ and $m \geqslant 0$, there exist a periodic point $z \in \Sigma_{P_\varphi}$ and an ε-perturbation of φ_* along the orbit of z,

$$A^i \colon T_{\varphi^i(z)}M \longrightarrow T_{\varphi^{i+1}(z)}M, \quad i = 0, \dots, n(z) - 1,$$

such that

$$\widetilde{M}_z = A^{n(z)-1} \circ \cdots \circ A^0 \colon T_zM \longrightarrow T_zM$$

is conjugate to the product $T_2 \circ M_y^m \circ T_1 \circ M_x^n$.

Remark 5.2. — In fact, in [**BDP**], it is shown that Lemma 5.1 holds for any finite number of orbits $x_1, \dots, x_k$ of Σ_{P_φ}. This allows us to get linear maps $T_i\colon T_{x_i}M \to T_{x_{i+1}}M$ preserving a dominated splitting such that, for every $n_1, \dots, n_k$, there exist a point $z \in \Sigma_{P_\varphi}$ and perturbations A^i of the derivative of φ_* at $\varphi^i(z)$ such that $\widetilde{M}_z = A^{n(z)-1} \circ \cdots \circ A^0$ is conjugate to $T_k \circ M_{x_k}^{n_k} \circ \cdots \circ T_2 \circ M_{x_2}^{n_2} \circ T_1 \circ M_{x_1}^{n_1}$.

The maps T_i correspond to the *transitions*, recall also Theorem 3.1. The fact that the transitions can be chosen preserving a dominated splitting has been proved in [**BDP**, Lemma 4.13]. This property is the basis of the proof of the following result:

Lemma 5.3. — *Let $E_1 \oplus \cdots \oplus E_m$, $E_i \prec E_{i+1}$, be the finest dominated splitting of TM over Σ_{P_φ} of φ_*. Then, for every $\varepsilon > 0$, there exist a dense subset Σ_ε of Σ_{P_φ} and an*

ε-perturbation A_ε of φ_* preserving the splitting $E_1 \oplus \cdots \oplus E_m$ such that, for every $R \in \Sigma_\varepsilon$, the restriction of the linear maps

$$M_{A_\varepsilon}(R) = A_\varepsilon(\varphi^{n(R)-1}(x)) \circ \cdots \circ A_\varepsilon(\varphi(x)) \circ A_\varepsilon(x)$$

to $E_i(R)$ is a homothety.

Moreover, if there exist $i \in \{1, \ldots, m\}$ and $Q \in \Sigma_{P_\varphi}$ such that the modulus of the Jacobian of the restriction of $\varphi_*^{n(Q)}$ to $E_i(Q)$ is one, then R can be chosen in such a way that the restriction of $M_{A_\varepsilon}(R)$ to $E_i(R)$ is the identity map.

This lemma is a consequence of [**BDP**, Propositions 2.4 and 2.5]. To see that these propositions can be applied in our context, we just need to observe that the restriction of φ_* to each bundle E_i (over Σ_{P_φ}) defines a periodic linear system with transitions. For that, it is enough to recall that the transitions of φ_* can be chosen preserving the bundles E_j of the dominated splitting (see [**BDP**, Section 4]).

Given a hyperbolic linear map A of an Euclidean space (i.e., without eigenvalues of modulus equal to 1) the *index of A* is the number of eigenvalues of A of modulus less than 1, counted with multiplicity.

Lemma 5.4 **([BDP**, Lemma 4.16]). — *Given $\varepsilon > 0$ there exist $x \in \Sigma_{P_\varphi}$ and an ε-perturbation of φ_* along the orbit of x such that the corresponding matrix M_x has index p, $p = index(P_\varphi)$, and all the eigenvalues of M_x are real, positive and with multiplicity* 1.

5.2. Tangencies and codimension one heterodimensional cycles. — The existence of non-real eigenvalues in the *central direction* of the saddles in a (codimension one) heterodimensional cycle produces homoclinic tangencies. That is formalized in the following result we export from [**DR**].

Let A be a linear map of an n-dimensional Euclidean space E, we say that a non-real eigenvalue $\lambda \in (\mathbb{C} \setminus \mathbb{R})$ of A has *rank ℓ* if there are $(\ell - 1)$ eigenvalues (counted with multiplicity) of A of modulus strictly less than $|\lambda|$ and $(n - \ell - 1)$ eigenvalues of modulus strictly bigger than $|\lambda|$. A periodic point P of a diffeomorphism φ has a *non-real eigenvalue of rank ℓ* if its derivative $\varphi_*^{n(P)}(P)$ has a non-real eigenvalue of rank ℓ.

Lemma 5.5. — *Let $\Gamma(\phi, U, R^1_\phi, R^2_\phi)$ be a codimension one heterodimensional cycle associated to hyperbolic periodic points of indices $(r+1)$ and r. Suppose that R^1_ϕ (resp. R^2_ϕ) has a non-real eigenvalue of rank r (resp. $r+1$). Then, there is ψ arbitrarily close to ϕ, with a homoclinic tangency associated to R^2_ψ (resp. R^1_ψ) in $\Lambda_\psi(U)$.*

Proof. — Just observe that, if R^1_ϕ has a non-real eigenvalue of rank r, then the unstable manifold of R^2_ϕ spirals around $W^u(R^1_\phi)$. Now, unfolding the cycle $\Gamma(\phi, U, R^1_\phi, R^2_\phi)$, we get a homoclinic tangency associated to the continuation of R^2_ϕ. See [**DR**, Section 8.1] for details. □

5.3. Proof of Theorem D. — Consider $\varphi \in \mathcal{P}(U)$ and its finest dominated splitting $E_1(\varphi) \oplus \cdots \oplus E_{m(\varphi)}(\varphi)$ over $\Lambda_\varphi(U)$. By Lemma 4.1, the continuation of this splitting over $\Lambda_\phi(U)$ is uniquely defined for every ϕ close to φ. Denote such a continuation by $E_1(\phi) \oplus \cdots \oplus E_{m(\varphi)}(\phi)$. By Lemma 4.1, the number $m(\varphi)$ of bundles of the finest dominated splitting of $\Lambda_\varphi(U)$ is lower semi-continuous, thus locally constant in an open and dense subset $\mathcal{P}_1(U)$ of $\mathcal{P}(U)$. Moreover, the dimensions of the bundles of the finest dominated splitting are also locally constant in $\mathcal{P}_1(U)$. So there is an open and dense subset $\mathcal{O}(U)$ of $\mathcal{P}(U)$ where $m(\varphi)$ and the dimensions of the bundles of the finest dominated splitting are continuous functions. This set $\mathcal{O}(U)$ is the open and dense subset of $\mathcal{P}(U)$ announced in Theorem D.

Observe that it is enough to prove the theorem for a connected component of $\mathcal{O}(U)$. So, from now on, we restrict our attention to a fixed connected component $\mathcal{O}_0$ of $\mathcal{O}(U)$.

Given $\varphi \in \mathcal{O}_0$, consider the finest dominated splitting of $\Lambda_\varphi(U)$, say $T_{\Lambda_\varphi(U)}M = E_1(\varphi) \oplus E_2(\varphi) \oplus \cdots \oplus E_{m(\varphi)}(\varphi)$. Since the dimensions and the number of bundles of the splitting do not depend on $\varphi \in \mathcal{O}_0$, from now on we will omit such dependence on φ.

Let us now introduce some notations. For simplicity, write $p = i_c$ and $q = i_s$ (the maximum and minimum indices of the hyperbolic periodic points of $\Lambda_\varphi(U)$). Given i and j in $\{1, \ldots, m\}$, with $i < j$, let

$$E_i^j = E_i \oplus E_{i+1} \oplus \cdots \oplus E_j.$$

Denote by d_i and d_i^j the dimensions of E_i and E_i^j, respectively (thus, $d_i^j = \sum_{k=i}^j d_k$). We define i_q and i_p by the relations

$$d_1^{i_q - 1} < q \leqslant d_1^{i_q} \quad \text{and} \quad d_1^{i_p - 1} < p \leqslant d_1^{i_p}.$$

To prove Theorem D it is enough to check the following:

(A) $d_1^{i_q} = q$ and $d_{i_p+1}^m = \dim(M) - p$,

(B) $d_j = 1$ and the bundle E_j is not uniformly hyperbolic for all $j \in \{i_q+1, \ldots, i_p\}$,

(C) $E_1^{i_q}$ and $E_{i_p+1}^m$ are uniformly contracting and expanding, respectively.

The proof of the items will be given in Lemmas 5.6, 5.7 and 5.8.

***Lemma 5.6* (Proof of (A)).** — *$d_1^{i_q} = q$ and $d_{i_p+1}^m = \dim(M) - p$.*

Proof. — Let us prove the first part of the lemma. The proof is by contradiction. Assume that $d_1^{i_q} > q$. Then, by definition of $d_1^{i_q}$, one has

$$d_1^{i_q - 1} < q < q + 1 \leqslant d_1^{i_q} = d_1^{i_q - 1} + d_{i_q},$$

hence,

$$d_{i_q} \geqslant 2. \tag{2}$$

By Proposition 2.4 and the definition of $\mathcal{O}_0$, there is a diffeomorphism $\varphi \in \mathcal{O}_0$ with a hyperbolic periodic point Q_φ of index q such that Σ_{Q_φ} is dense in $\Lambda_\varphi(U)$. By

Remark 4.3, the finest dominated splitting of φ over Σ_{Q_φ} is the restriction to Σ_{Q_φ} of the bundles E_i.

By equation (2), E_{i_q} is indecomposable and has dimension d_{i_q} greater than or equal to 2. Applying Lemma 5.3 to the set Σ_{Q_φ} and the bundle E_{i_q}, we get $R_\varphi \in \Sigma_{Q_\varphi}$ of period $n(R_\varphi)$ and a perturbation A of φ_* throughout the φ-orbit of R_φ such that

$$M_A(R_\varphi)) = A(\varphi^{n(R_\varphi)-1}(R_\varphi)) \circ \cdots \circ A(\varphi(R_\varphi)) \circ A(R_\varphi)$$

is a homothety in $E_{i_q}(R_\varphi)$. We observe that the perturbation A of φ_* can be obtained (and that is what is done here) such that its restrictions to the bundles $E_k(R_\varphi)$, $k \neq i_q$, coincide with φ_*. Thus, since all points of Σ_{Q_φ} have index q, we have that, for every $T_\varphi \in \Sigma_{Q_\varphi}$, the bundles $E_j(T_\varphi)$, $j > i_q$, correspond to expanding eigenvalues of $\varphi_*^{n(T_\varphi)}$. Hence, the number of contracting eigenvalues of $M_A(R_\varphi)$ is at most $d_1^{i_q}$.

First, if the ratio of this homothety (the restriction of $M_A(R_\varphi)$ to $E_{i_q}(R_\varphi)$) is bigger or equal than one, using Lemma 3.4, one gets ϕ close to φ ($\phi \in \mathcal{O}_0$) with a hyperbolic periodic point $R_\phi \in \Lambda_\phi(U)$ having at most $d_1^{i_q-1}$ contracting eigenvalues. By hypothesis, $d_1^{i_q-1} < q$, thus the index of R_ϕ is strictly less than q, contradicting the definition of q (minimality of the index of the points of $\Lambda_\phi(U)$, $\phi \in \mathcal{P}(U)$).

So, we can assume that the ratio of the homothety $M_A(R_\varphi)|_{E_{i_q}(R_\varphi)}$ is less than one. As the restriction of $\varphi_*^{n(R_\varphi)}$ to each $E_i(R_\varphi)$, $i > i_q$, has expanding eigenvalues, the index of R_ϕ is exactly $d_1^{i_q}$. Now, the definition of p implies that $d_1^{i_q} \leqslant p$.

Write $\ell = d_1^{i_q} \leqslant p$. Since all the eigenvalues of the restriction of $\phi_*^{n(R_\varphi)} = M_A(R_\varphi)$ to $E_{i_q}(R_\phi)$ are equal and $\dim(E_{i_q}(R_\phi)) \geqslant 2$, using again Lemma 3.4, one gets a diffeomorphism υ (close to ϕ) such that R_υ has index ℓ and $\upsilon_*^{n(R_\upsilon)}(R_\upsilon)$ has a contracting non-real eigenvalue of rank $(\ell - 1)$.

By Theorem A, since $q \leqslant \ell - 1$, there is a diffeomorphism ζ (close to υ) with a periodic point $S_\zeta \in \Lambda_\zeta(U)$ of index $(\ell - 1)$. Using Lemma 2.5, we obtain η close to ζ with a codimension one heterodimensional cycle in U associated to R_η and S_η, say $\Gamma(\eta, U, R_\eta, S_\eta)$. Since η can be taken arbitrarily close to υ, we can assume that R_η has index ℓ and a non-real eigenvalue of rank $\ell - 1$ and that S_η has index $(\ell - 1)$. Finally, by Lemma 5.5, there is a diffeomorphism $\xi \in \mathcal{O}_0$ arbitrarily close to η with a homoclinic tangency in $\Lambda_\xi(U)$ associated to the point S_ξ of index $(\ell-1)$, contradicting the definition of $\mathcal{P}(U)$. This finishes the proof of the first assertion in the lemma.

Using the same arguments, we get that $d^m_{i_p+1} = (\dim(M) - p)$, so we omit this proof. $\square$

***Lemma 5.7* (Proof of (B))**. — *The bundle E_i is one dimensional and non-uniformly hyperbolic for all $i \in \{i_q + 1, \ldots, i_p\}$.*

Proof. — Given $k \in \{i_q + 1, \cdots, i_p\}$, let $\ell = d_1^k = \dim(E_1^k)$. Observe that by, Lemma 5.6, $q < \ell \leqslant p$.

The bundle E_k is not uniformly hyperbolic. — We argue by contradiction. Otherwise, since E_k is indecomposable, it would be either uniformly contracting or expanding. In the first case, using the domination of the splitting, one has that every periodic point of $\Lambda_\varphi(U)$ has index bigger or equal than $\ell > q$, contradicting the definition of q. In the second case, again by the domination of the splitting, every periodic point of $\Lambda_\varphi(U)$ has index strictly less than $\ell \leqslant p$, contradicting the definition of p.

The bundle E_k is one-dimensional. — The proof is by contradiction, assuming that $\dim(E_k) = d_k \geqslant 2$. By Theorem A and Proposition 2.4, there exists $\varphi \in \mathcal{O}_0$ with a hyperbolic periodic point $R_\varphi \in \Lambda_\varphi(U)$ of index ℓ such that Σ_{R_φ} is dense in $\Lambda_\varphi(U)$. By Lemma 5.3, there exist a perturbation A of φ_* and a point $S_\varphi \in \Sigma_{R_\varphi}$ such that the restriction of $M_A(S_\varphi)$ to $E_k(S_\varphi)$ is a homothety. Moreover, as before, we can take A such that its restrictions to the bundles $E_i(S_\varphi)$, $i \neq k$, coincide with the one of φ_*.

Suppose, for instance, that the ratio of such a homothety is bigger than one. From $S_\varphi \in \Sigma_{R_\varphi}$ and the definition of Σ_{R_φ}, the restrictions of $\varphi_*^{n(S_\varphi)}$ to the bundles $E_i(S_\varphi)$, $i > k$, have only expanding eigenvalues. Thus, the matrix $M_A(S_\varphi)$ has exactly $r = d_1^{k-1}$ contracting eigenvalues, where

$$q \leqslant d_1^{i_q} \leqslant d_1^{k-1} = r \leqslant d_1^{i_p - 1} < d_1^{i_p} = p \quad \text{and} \quad r < r + d_k \leqslant r + 2 \leqslant p.$$

Using Lemma 3.4, we get $\phi \in \mathcal{O}_0$ with a hyperbolic periodic point $S_\phi \in \Lambda_\phi(U)$ of index r such that the restriction of ϕ_* to $E_k(S_\phi)$ is equal to A. After a new perturbation, if necessary, we can assume that $\phi_*^{n(S_\phi)}(S_\phi)$ has a expanding non-real eigenvalue of rank $(r+1)$.

As in the proof of Lemma 5.6, by Theorem A and Lemma 2.5, there is $\psi \in \mathcal{O}_0$ close to ϕ with a periodic point $T_\psi \in \Lambda_\psi(U)$ of index $(r+1) < p$ and a heterodimensional cycle $\Gamma(\psi, U, T_\psi, S_\psi)$, where S_ψ has index r and a (expanding) non-real eigenvalue of rank $(r+1)$. Finally, by Lemma 5.5, there is $\xi \in \mathcal{O}_0$ close to ψ with a homoclinic tangency associated to T_ξ, contradicting the definition of $\mathcal{O}_0$. This finishes the proof of the lemma in this case. If the homothety given by the restriction of $M_A(S_\varphi)$ to E_k has ratio less than one the proof follows similarly. □

***Lemma 5.8* (Proof of (C)).** — *The bundles $E_1^{i_q}$ and $E_{i_p+1}^m$ are uniformly volume contracting and volume expanding, respectively.*

Proof. — This lemma follows from Theorem B. To check that $E = E_1^{i_q}$ is uniformly contracting, observe that the set $\mathcal{O}_0$ and the dominated splitting $E_1^{i_q} \oplus E_{i_q+1}^m$ satisfy the hypotheses of Theorem B (recall that, by Lemma 5.6, $q = d_1^{i_q} = \dim(E_1^{i_q})$).

The uniform expansion of $E_{i_p+1}^m$ follows analogously. This completes the proof of the lemma and of the theorem. □

6. Homoclinic tangencies

We now analyze the dimensions of the bundles of finest dominated splitting of a robust transitive set to deduce the different types of homoclinic bifurcations that this set may exhibit.

We consider an open subset U of M and $\mathcal{N}(U) \subset \mathrm{Diff}^1(M)$ an open set such that, for every $\varphi \in \mathcal{N}(U)$, the set $\Lambda_\varphi(U)$ is robustly transitive and

– the maximum and the minimum of the indices of the periodic points of $\Lambda_\varphi(U)$ are constant, equal to p and q, respectively,

– the dimensions of the bundles of the finest dominated splitting of $\Lambda_\varphi(U)$ do not depend on $\varphi \in \mathcal{N}(U)$.

Notice that, in this section, it is not assumed that there are no homoclinic tangencies in $\Lambda_\varphi(U)$, as in the previous section.

We use the notation introduced in Section 5.3 for the dimensions of the bundles of the finest dominated splitting. Recall that, with this notation and by definition, $q \leqslant d_1^{i_q}$ and $p \leqslant d_1^{i_p}$.

We say that a robustly transitive set $\Lambda_\varphi(U)$ has a *homoclinic tangency of rank* r if there is a periodic point $R_\varphi \in \Lambda_\varphi(U)$ of index r having a homoclinic tangency and such a point of tangency belongs to $\Lambda_\varphi(U)$.

Theorem F. — *Let U, $\mathcal{N}(U)$, p and q as before. Consider any $\varphi \in \mathcal{N}(U)$.*

– If $d_1^{i_q} > q$, then there is ϕ arbitrarily close to φ such that $\Lambda_\phi(U)$ has a homoclinic tangency of rank $(d_1^{i_q} - 1)$.

– If $d_1^{i_p} > p$, then there is ϕ arbitrarily close to φ such that $\Lambda_\phi(U)$ has a homoclinic tangency of rank $(d_1^{i_p-1} + 1)$.

– If $d_j \geqslant 2$ for some $j \in \{i_q+1, \dots, i_p\}$, then, for every $k \in [d_1^{j-1}+1, d_1^j)$, there is ϕ arbitrarily close to φ such that $\Lambda_\phi(U)$ has a homoclinic tangency of rank k.

This theorem is a generalization of the result [**DPU**, Corollary G] for three dimensional robustly transitive sets, which says that the existence of an indecomposable bundle of dimension strictly greater than one leads to the creation of homoclinic tangencies in a (non-hyperbolic) robustly transitive set.

The proof of Theorem F follows from a small modification of the the proofs of Lemmas 5.6 and 5.7 and involves heterodimensional cycles.

Denote by $\mathcal{T}_k(U)$, $k = 1, \dots, \dim(M)-1$, the subset of $\mathcal{N}(U)$ of diffeomorphisms ϕ such that $\Lambda_\phi(U)$ has a homoclinic tangency of rank k. Theorem F now follows from the next two lemmas.

Lemma 6.1. — *Under the hypothesis of Theorem F, we have the following*

– If $d_1^{i_q} > q$, then $\mathcal{T}_{d_1^{i_q}-1}(U)$ is dense in $\mathcal{N}(U)$.

– If $d_1^{i_p} > p$, then $\mathcal{T}_{d_1^{i_p-1}+1}(U)$ is dense in $\mathcal{N}(U)$.

Proof. — First, observe that, by definition, if $d_1^{i_q} > q$ (resp. $d_1^{i_p} > p$) then $d_{i_q} > 1$ (resp. $d_{i_p} > 1$).

To prove the first part of the lemma, it is enough to check that if $\varphi \in \mathcal{N}(U)$ and $d_1^{i_q} > q$ then there is v arbitrarily close to φ such that $\Lambda_v(U)$ has a homoclinic tangency of rank $(d_1^{i_q} - 1)$. Recall that, in the proof of Lemma 5.6, under the assumption that $\ell = d_1^{i_q} > q$, we got v close to φ having a hyperbolic periodic point $R_v \in \Lambda_\nu(U)$ of index ℓ with a non-real eigenvalue of rank $(\ell - 1)$.

Since $q \leqslant \ell - 1 < p$, by Theorem A and Lemma 2.5, after a $\mathcal{C}^1$-perturbation of v, we can assume that v has a periodic point S_v of index $(\ell - 1)$ and a (codimension one) heterodimensional cycle $\Gamma(v, U, R_v, S_v)$ (R_v of index ℓ with a non-real eigenvalue of rank $(\ell - 1)$). By Lemma 5.5, there is ξ close to v with a homoclinic tangency associated to S_ξ. This finishes the first part of the lemma.

The second part of the lemma follows similarly. $\square$

Lemma 6.2. — *Under the hypotheses of Theorem F, suppose that $d_j \geqslant 2$, $j \in \{i_q + 1, \ldots, i_p - 1\}$. Then, for every $k \in [d_1^{j-1} + 1, d_1^j)$, the set $\mathcal{T}_k(U)$ is dense in $\mathcal{N}(U)$.*

Proof. — As in the previous lemma, given any $\varphi \in \mathcal{N}(U)$ with $d_j \geqslant 2$ and $k \in [d_1^{j-1} + 1, d_1^j)$ we will obtain ϕ arbitrarily close to φ such that $\Lambda_\varphi(U)$ has a homoclinic tangency or rank k. By Theorem A, and since

$$q \leqslant d_1^{j-1} < d_1^j \leqslant d_1^{i_p - 1} < p,$$

after perturbing φ, we can assume that φ has a pair of hyperbolic periodic points $S_\varphi, T_\varphi \in \Lambda_\varphi(U)$ of indices d_1^j and d_1^{j-1}, respectively.

By Lemma 2.5, there is ψ close to φ with a heterodimensional cycle $\Gamma(\psi, U, S_\psi, T_\psi)$. Observe that the modulus of the restriction of the Jacobian of $\psi_*^{n(T_\psi)}$ to $E_j(T_\psi)$ is greater than one and the modulus of the restriction of the Jacobian of $\psi_*^{n(S_\psi)}$ to $E_j(S_\psi)$ is less than one. By Corollary 3.7, unfolding this cycle, we get ϕ close to φ with a hyperbolic periodic point $R_\phi \in \Lambda_\phi(U)$ with index r, $r \in [d_1^{j-1}, d_1^j]$, such that the modulus of the Jacobian of $\phi_*^{n(R_\phi)}$ to $E_j(R_\phi)$ is exactly one.

By Proposition 2.4, after a perturbation of ϕ, we can assume that $\Sigma_{R_\phi}(\phi)$ is dense in $\Lambda_\phi(U)$. Since $E_j(R_\phi)$ is indecomposable and has dimension equal to or greater than 2, arguing exactly as in the proof of Lemma 5.7, but now applying the final part of Lemma 5.3, we get ξ (arbitrarily close to ϕ) with a periodic point $A_\xi \in \Lambda_\xi(U)$ such that the restriction of $\xi_*^{n(A_\xi)}$ to $E_j(A_\xi)$ is the identity.

Take now any $k \in [d_1^{j-1} + 1, d_1^j)$. After a perturbation of ξ we can assume that the index of A_ξ is $k - 1$, and that $\xi_*^{n(A_\xi)}(A_\xi)$ has an expanding non-real eigenvalue of rank k. Again, by Theorem A, we can assume there is a periodic point $B_\xi \in \Lambda_\xi(U)$ of index k, where $k > q$. Finally, by Lemma 2.5, there is η close to ξ with a codimension one cycle $\Gamma(\eta, U, B_\eta, A_\eta)$, A_η of index $(k-1)$ and with an expanding non-real eigenvalue of rank k and B_η of index k. Now the lemma follows from Lemma 5.5. $\square$

7. Proof of Theorem E

As we have mentioned in the introduction, Theorem E follows from Proposition 1.1. So, before proving the proposition let us deduce the theorem from it.

Recall that U and $\mathcal{S}(U)$ are open subsets of M and $\mathrm{Diff}^1(M)$ such that, for every diffeomorphism $\varphi \in \mathcal{S}(U)$, the set $\Lambda_\varphi(U)$ is robustly transitive and has no homoclinic tangencies (in the whole manifold). By Theorem D, there is an open and dense subset $\mathcal{I}(U)$ of $\mathcal{S}(U)$, such that if φ belongs to $\mathcal{I}(U)$ and $\Lambda_\varphi(U)$ contains periodic points of indices q and p, $q < p$, then $\Lambda_\varphi(U)$ contains points of every index between q and p. So it is enough to prove the theorem for the subset $\mathcal{I}(U)$ of $\mathcal{S}(U)$.

Consider the maps $i^+, i^- : \mathcal{I}(U) \to \mathbb{N}^*$ that associate to each $\varphi \in \mathcal{I}(U)$ the maximum and the minimum of the indices of the hyperbolic periodic points of $\Lambda_\varphi(U)$, respectively. These two functions are semi-continuous, so they are continuous in an open and dense subset $\mathcal{I}_0(U)$ of $\mathcal{I}(U)$. Now it is enough to fix a connected component $\mathcal{I}_0$ of $\mathcal{I}(U)$ where i^+ and i^- are both constant and to prove the theorem for this set. Suppose that $i^+(\varphi) = p$ and $i^-(\varphi) = q$ for all $\varphi \in \mathcal{I}_0$, $q \leqslant p$.

Assume that $q < p$ (the case $q = p$ follows from Remark 2.7, so we omit it). Let Q_φ and P_φ be points of indices q and p of $\Lambda_\varphi(U)$. For notational simplicity, let us assume that their continuations are defined in the whole $\mathcal{I}_0$. Since P_φ and Q_φ are transitively related in $\mathcal{I}_0$, by Remark 2.6, there is an open and dense subset $\mathcal{I}_1$ of $\mathcal{I}_0$ such that $W^s(P_\phi)$ and $W^u(Q_\phi)$ have nonempty transverse intersection for all $\phi \in \mathcal{I}_1$. So it is enough to prove the theorem for $\mathcal{I}_1$.

For each $j \geqslant 0$ with $q + j \leqslant p$, let $\mathcal{A}(j)$ be the subset of $\mathcal{I}_1$ of diffeomorphisms ψ such that $\Lambda_\psi(U)$ contains hyperbolic periodic points $R^0_\psi, R^1_\psi, \ldots, R^j_\psi$ such that

– $\mathrm{index}(R^i_\psi) = q + i$,

– $\overline{H_{R^0_\varphi}(U)} = \overline{H_{R^1_\varphi}(U)} = \cdots = \overline{H_{R^j_\varphi}(U)}$ for every φ in a neighbourhood of ψ

To finish the proof of Theorem E, it is enough to check the following.

Lemma 7.1. — *The set $\mathcal{A}(j)$ is open and dense in $\mathcal{I}_1$ for every $j \in (0, r]$, $r = p - q$.*

Before proving this lemma, let us assume it and prove the theorem.

Observe that, by Lemma 7.1, $\mathcal{A}(r)$ is open and dense in $\mathcal{I}_1$, and for every ψ in $\mathcal{A}(r)$, there exist hyperbolic periodic points R^0_ψ and R^r_ψ of $\Lambda_\psi(U)$ of indices q and $q + r = p$ such that

$$\overline{H_{R^r_\psi}(U)} = \overline{H_{R^0_\psi}(U)}.$$

As before, for notational simplicity, assume that the continuations of R^0_ψ and R^r_ψ are defined in the whole $\mathcal{A}(r)$. The points Q_ψ and R^0_ψ have index q and are transitively related in $\mathcal{A}(r)$. Thus, by Remark 2.7, there is an open and dense subset $\mathcal{D}_1$ of $\mathcal{A}(r)$

of diffeomorphisms ζ such that the relative homoclinic classes of Q_ψ and R^0_ψ in U are equal. Similarly, there is an open and dense subset $\mathcal{D}_2$ of $\mathcal{A}(r)$ of diffeomorphisms ζ such that the relative homoclinic classes of P_ψ and R^r_ψ in U are equal. Thus, for all $\zeta \in \mathcal{D}_1 \cap \mathcal{D}_2$, one has that

$$\overline{H_{P_\zeta}(U)} = \overline{H_{R^r_\zeta}(U)} = \overline{H_{R^0_\zeta}(U)} = \overline{H_{Q_\zeta}(U)}.$$

Since $\mathcal{D}_1 \cap \mathcal{D}_2$ is open and dense in $\mathcal{A}(r)$, thus in $\mathcal{I}_1$, and the result is proved.

Proof of the lemma. — We will argue by induction. To see that $\mathcal{A}(1)$ is open and dense in $\mathcal{I}_1$, it suffices to prove that, given any $\phi \in \mathcal{I}_1$, there is an open subset $\mathcal{A}_\phi$ of $\mathcal{I}_1$ such that

– ϕ belongs to the closure of $\mathcal{A}_\phi$,

– for every $\psi \in \mathcal{A}_\phi$, there exists a hyperbolic periodic point $R^1_\psi \in \Lambda_\psi(U)$ of index $(q+1)$ such that $\overline{H_{Q_\psi}(U)} = \overline{H_{R^1_\psi}(U)}$ (here we take $R^0_\psi = Q_\psi$).

Since ϕ is in $\mathcal{I}_1$ there is a periodic point $R^1_\phi \in \Lambda_\phi(U)$ of index $(q+1)$. Observe that Q_ϕ and R^1_ϕ are transitively related and $\operatorname{index}(Q_\psi)+1 = \operatorname{index}(R^1_\psi)$. Thus, by Lemma 2.5, after a perturbation of ϕ, we can assume that ϕ has a (codimension one) cycle $\Gamma(\phi, U, R^1_\phi, Q_\phi)$. By hypothesis, this cycle is far from homoclinic tangencies. Thus, by Proposition 1.1, there is an open set $\mathcal{B}_\phi$, whose closure contains ϕ, such that $\overline{H_{Q_\zeta}(U)} = \overline{H_{R^1_\zeta}(U)}$ for all $\zeta \in \mathcal{B}_\phi$. The first inductive step follows taking $\mathcal{A}_\phi = \mathcal{B}_\phi \cap \mathcal{I}_1$.

Suppose now defined inductively the open and dense subsets $\mathcal{A}(1), \mathcal{A}(2), \ldots, \mathcal{A}(j-1)$, $q+j \leqslant p$, of $\mathcal{I}_1$ satisfying the properties above. Then the set

$$\mathcal{A}'(j-1) = \mathcal{A}(1) \cap \cdots \cap \mathcal{A}(j-1)$$

is open and dense in $\mathcal{I}_1$. Now it is enough to get an open and dense subset $\mathcal{A}(j)$ of $\mathcal{A}'(j-1)$ with the announced properties. For that we argue exactly as in the step $j=1$.

Consider any $\phi \in \mathcal{A}'(j-1)$. Since $\phi \in \mathcal{I}_1$ the set $\Lambda_\phi(U)$ contains a hyperbolic periodic point R^j_ϕ of index $(q+j)$. As in the first step of the induction, using Lemma 2.5, we can assume (after a perturbation of ϕ) that ϕ has a (codimension one) cycle $\Gamma(\phi, U, R^j_\phi, R^{j-1}_\phi)$, where R^{j-1}_ϕ is the point of index $(q+j-1)$ in the inductive step $(j-1)$. By hypothesis, this cycle is far from homoclinic tangencies. Thus, by Proposition 1.1, there is an open set $\mathcal{B}_\phi \subset \mathcal{A}'(j-1)$ containing ϕ in its closure such that

$$\overline{H_{R^{j-1}_\zeta}(U)} = \overline{H_{R^j_\zeta}(U)}$$

for all $\zeta \in \mathcal{B}_\phi$. Since $\mathcal{B}_\phi \subset \mathcal{A}'(j-1)$, we have

$$\overline{H_{R^0_\zeta}(U)} = \overline{H_{R^1_\zeta}(U)} = \overline{H_{R^{j-1}_\zeta}(U)} = \overline{H_{R^j_\zeta}(U)}$$

for all $\zeta \in \mathcal{B}_\phi$, finishing the proof of the lemma. □

7.1. Proof of Proposition 1.1. — Suppose now that (as in the hypotheses of Proposition 1.1) the indices of P_φ and Q_φ are p and q with $p = (q+1)$. By [**BDP**, Lemma 5.4], we can assume that the robustly transitive set $\Lambda_\varphi(U)$ contains a pair of hyperbolic periodic points of indices q and $p+1$ having only real eigenvalues with multiplicity one and different moduli. For notational simplicity, assume that Q_φ and P_φ verify these hypotheses. In particular, these points verify the hypotheses of Corollary 3.6. By (1) in the proof of the corollary, after a small perturbation, we can assume that φ has a saddle-node periodic point (a point with an eigenvalue equal to one) with q contracting eigenvalues and $(\dim(M)-q-1)$ expanding eigenvalues. After a new perturbation, by unfolding the saddle-node, we can assume that φ has a pair of periodic points A_φ and B_φ of indices p and q, respectively, such that there is a curve γ whose extremes are A_φ and B_φ and whose interior is contained in $W^s(A_\varphi) \pitchfork W^u(B_\varphi)$. By Remark 2.7, we can assume that there is an open subset $\mathcal{V}$ of $\mathrm{Diff}^1(M)$ containing φ in its closure such that $\overline{H_{P_\psi}(U)} = \overline{H_{A_\psi}(U)}$ and $\overline{H_{Q_\psi}(U)} = \overline{H_{B_\psi}(U)}$ for all ψ in $\mathcal{V}$.

By Remark 2.6, there is a sequence of diffeomorphisms φ_k, $\varphi_k \to \varphi$ in the C^1-topology, such that φ_k has a codimension one heterodimensional cycle $\Gamma(\varphi_k, U, A_{\varphi_k}, B_{\varphi_k})$. By construction, these cycles are *connected*, i.e., $W^s(A_{\varphi_k}) \pitchfork W^u(B_{\varphi_k})$ has a periodic connected component whose extremes are contained in the orbits of A_{φ_k} and B_{φ_k} (here the connected component is the continuation of the curve γ above).

The proposition now follows directly from [**DR**]. For completeness let us state these results.

Lemma 7.2. — *Let ζ be a C^1-diffeomorphism with a codimension one connected heterodimensional cycle $\Gamma(\zeta, U, A_\zeta, B_\zeta)$ as above. Then given any C^1-neighbourhood $\mathcal{A}$ of ζ, there exists a C^1-open subset $\mathcal{U}(\zeta)$ of $\mathcal{A}$ such that $\overline{H_{A_\psi}(U)} = \overline{H_{B_\psi}(U)}$ for every $\psi \in \mathcal{U}(\zeta)$.*

By the lemma, for each φ_k as before, there is an open set $\mathcal{U}(\varphi_k) \subset \mathcal{V}$ containing φ_k in its closure, such that, for every $\psi \in \mathcal{U}(\varphi_k)$, we have $\overline{H_{A_\psi}(U)} = \overline{H_{B_\psi}(U)}$. Since $\psi \in \mathcal{V}$, we have that $\overline{H_{A_\psi}(U)} = \overline{H_{P_\psi}(U)}$ and $\overline{H_{Q_\psi}(U)} = \overline{H_{B_\psi}(U)}$. Proposition 1.1 now follows taking $\mathcal{W}_\varphi = \bigcup_k \mathcal{U}(\varphi_k)$.

Proof of the lemma. — Observe that the cycle $\Gamma(\zeta, U, A_\zeta, B_\zeta)$ is connected and far from homoclinic tangencies. In [**DR**], see the comments after Theorem A, it is proved that, given any neighbourhood $\mathcal{U}$ of ζ there is an open subset $\mathcal{U}_0$ of $\mathcal{U}$ such that every $\psi \in \mathcal{U}_0$ has a transitive set Λ_ψ containing A_ψ and B_ψ such that $\Lambda_\psi \subset \overline{H(B_\psi)}$. The main step to prove this result is the fact we state below.

First, observe that, by construction, there is a multiplicity one contracting eigenvalue $\lambda_c \in \mathbb{R}$ of the derivative of ζ at A_ζ such that $1 > |\lambda_c| > |\lambda|$ for every contracting eigenvalue λ of A_ζ different from λ_c (see condition (CE) in [**DR**, Section 3.1]). Thus, for every ψ close to ζ, the (codimension one) strong stable foliation $\mathcal{F}^s_\psi$ of $W^s(A_\psi)$ is

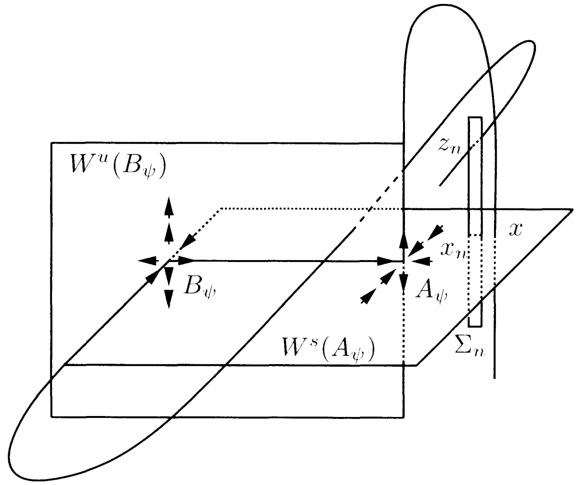


FIGURE 3. Homoclinic points

defined. Similarly, we have that the (codimension one) strong unstable foliation $\mathcal{F}^u_\psi$ of $W^u(A_\psi)$ is defined. Now the lemma will follow from the following fact.

Fact. — *Let $\mathcal{A}$ be as in Lemma 7.2.*

– Let $u = (\dim(M) - p)$ be the dimension of the unstable bundle of A_ζ. There is an open subset $\mathcal{A}_0$ of $\mathcal{A}$ of diffeomorphisms ψ such that $W^s(B_\psi)$ intersects transversely every $(u+1)$-disk Σ transverse to $\mathcal{F}^s_\psi$.

– Let s be the dimension of the stable bundle of A_ζ. There is an open subset $\mathcal{A}_0$ of $\mathcal{A}$ of diffeomorphisms ψ such that $W^u(A_\psi)$ intersects transversely every $(s+1)$-disk Σ transverse to $\mathcal{F}^u_\psi$.

This fact is a non-technical reformulation of [**DR**, Proposition 3.6 (b)]. Notice that (due to the context) in [**DR**] this proposition is stated for parametrized families of diffeomorphisms unfolding a connected cycle corresponding to a first bifurcation. But, as mentioned in [**DR**, Section 6], it holds in a much more general setting (including the case under consideration).

To see, for instance, that $\overline{H_{A_\psi}(U)}$ is contained in $\overline{H_{B_\psi}(U)}$, we use the first part of the fact. Take any x in $H_{A_\psi}(U)$. By the cycle configuration, $W^u(A_\psi)$ is contained in the closure of $W^u(B_\psi)$, thus there is a sequence $x_n \to x$ with $x_n \in W^s(A_\psi) \pitchfork W^u(B_\psi)$ for all n. Associated to each x_n, we have a $(u+1)$-disk Σ_n of diameter less than $1/n$ which is contained in $W^u(B_\psi)$ and transverse to $W^s(A_\psi)$ at x_n (see figure). The fact implies that, for each n, there is $z_n \in W^s(B_\psi) \pitchfork \Sigma_n$. By construction $z_n \in H_{A_\psi}$ (in fact one can take $z_n \in H_{B_\psi}(U)$) and $\lim z_n = \lim x_n = x$.

The inclusion $\overline{H_{B_\varphi}(U)} \subset \overline{H_{B_\varphi}(U)}$ follows similarly using the second part of the fact. This finishes the sketch of the proof of the lemma. □

References

[Ar] M.-C. Arnaud, *Creation de connexions en topologie C^1*, Ergod. Th. & Dynam. Syst., **21**, 339-381, (2001).

[AS] R. Abraham and S. Smale, *Nongenericity of Ω-stability,* Global Analysis I, Proc. Symp. Pure Math. A.M.S., **14**, 5-8, (1968).

[B] Ch. Bonatti, *A local mechanism for robust transitivity,* Lecture at the Seminar of Dynamical Systems, IMPA, August (1996).

[BD$_1$] Ch. Bonatti and L.J. Díaz, *Persistence of transitive diffeomorphims,* Ann. Math., **143**, 367-396, (1995).

[BD$_2$] Ch. Bonatti and L.J. Díaz, *Connexions hétéroclines et généricité d'une infinité de puits ou de sources,* Ann. Scient. Éc. Norm. Sup., 4^e série, t32, 135-150, (1999).

[BDP] Ch. Bonatti, L.J. Díaz, and E.R. Pujals, *A C^1-generic dichotomy for diffeomorphisms: Weak forms of hyperbolicicity or or infinitely many sinks or sources,* preprint PUC-Rio (1999) and to appear in Annals of Math..

[BV] Ch. Bonatti and M.Viana, *SRB for partially hyperbolic attractors: the contracting case,* Israel Journal of Math., **115**, 157-193, (2000).

[CMP] C. Carballo, C. Morales, and M.J. Pacífico, *Homoclinic classes for C^1-generic vector fields,* preprint PUC-Rio (2000) and to appear in Ergod. Th. & Dynam. Syst.

[D] L.J. Díaz, *Robust nonhyperbolic dynamics and heterodimensional cycles,* Ergod. Th. & Dynam. Syst., **15**, 291-315, (1995).

[DR] L.J. Díaz and J. Rocha, *Partially hyperbolic and transitive dynamics generated by heteroclinic cycles,* Ergod. Th. & Dynam. Syst., **25**, 25-76, (2001).

[DPU] L.J. Díaz, E.R. Pujals, and R. Ures, *Partial hyperbolicity and robust transitivity,* Acta Mathematica, **183**, 1-43, (1999).

[F] J. Franks, *Necessary conditions for stability of diffeomorphisms,* Trans. A.M.S., **158**, 301-308, (1971).

[H] S. Hayashi, *Connecting invariant manifolds and the solution of the C^1-stability and Ω-stability conjectures for flows,* Ann. of Math., **145**, 81-137, (1997).

[M$_1$] R. Mañé, *Contributions to the stability conjecture,* Topology, **17**, 386-396, (1978).

[M$_2$] R. Mañé, *An ergodic closing lemma,* Ann. of Math., **116**, 503-540, (1982).

[N] S. Newhouse, *Diffeomorphisms with infinitely many sinks,* Topology, **13**, 9-18, (1974).

[R] N. Romero, *Persistence of homoclinic tangencies in higher dimensions,* Ergod. Th. & Dynam. Syst., **15**, 735-759, (1995).

[Pa] J. Palis, *A global view of dynamics and a conjecture on the denseness and finitude of attractors,* Astérisque, **261**, 335-347, (2000).

[PV] J. Palis and M. Viana, *High dimension diffeomorphisms displaying infinitely many sinks,* Ann. of Math., **140**, 207-250, (1994).

[Pu] C. Pugh, *The closing lemma,* Amer. Jour. of Math., **89**, 956-1009, (1967).

[Sh] M. Shub, *Topological transitive diffeomorphism on T^4*, Lect. Notes in Math., **206**, 39 (1971).

[Si] R.C. Simon, *A 3-dimensional Abraham-Smale example,* Proc. A.M.S., **34**(2), 629-630, (1972).

[Sm] S. Smale, *Differentiable dynamical systems*, Bull. A.M.S., **73**, 747-817, (1967).

[YA] J. Yorke and K. Alligood, *Cascades of period doubling bifurcations: a prerequisite for horseshoes*, Bull. Amer. Math. Soc. (New Series), **9**, 319-332, (1983).

CH. BONATTI, Laboratoire de Topologie, UMR 5584 du CNRS, BP 47 870, 21078 Dijon Cedex, France • *E-mail :* bonatti @u-bourgogne.fr

L.J. DÍAZ, Dep. Matemática PUC-RJ, Marquês de S. Vicente 225, 22453-900 Rio de Janeiro RJ, Brazil • *E-mail :* lodiaz@mat.puc-rio.br

E.R. PUJALS, Instituto de Matemática UFRJ, Ilha do Fundão CP 68530, 21945-970 Rio de Janeiro RJ, Brazil • *E-mail :* enrique@tristan.lpim.ufrj.br

J. ROCHA, Departamento de Matemática Pura, Universidade do Porto, Rua do Campo Alegre núm. 687, 4169-007 Porto, Portugal • *E-mail :* jrocha@fc.up.pt

Astérisque
286, 2003, p. 223–229

COUPLED HOPF-BIFURCATIONS: PERSISTENT EXAMPLES OF n-QUASIPERIODICITY DETERMINED BY FAMILIES OF 3-JETS

by

Henk Broer

Abstract. — In this note examples are presented of vector fields depending on parameters and determined by the 3-jet, which display persistent occurrence of n-quasiperiodicity. In the parameter space this occurrence has relatively large measure. A leading example consists of weakly coupled Hopf bifurcations. This example, however, is extended to full generality in the space of all 3-jets.

1. Introduction

In the theory of coupled reaction diffusion equations the following is of interest, see Poláčik *et al.* [**12, 19**]. The problem is whether persistent examples exist of (parameter dependent) dynamical systems with the following properties:

(1) Occurrence of n-quasiperiodicity in a measure theoretically significant way.
(2) The system is local, and the property is determined by a low order jet.
(3) Preferably parameters are only needed in the linear part.

Below we present a solution to this problem by means of coupled Hopf families. To fix thoughts, we start with an example.

***Example 1* (Weakly coupled Hopf bifurcations).** — Consider a C^∞-system of n weakly coupled Hopf bifurcations, near the origin of $\mathbb{R}^{2n}$ given by

$$\begin{pmatrix} \dot{x}_j \\ \dot{y}_j \end{pmatrix} = \begin{pmatrix} \alpha_j & -\beta_j \\ \beta_j & \alpha_j \end{pmatrix} \begin{pmatrix} x_j \\ y_j \end{pmatrix} - (x_j^2 + y_j^2) \begin{pmatrix} x_j \\ y_j \end{pmatrix} + O(r^4), \tag{1}$$

$1 \leqslant j \leqslant n$, where $r^2 = \sum_{j=1}^n (x_j^2 + y_j^2)$. The lower order part (the 3-jet) consists of n completely decoupled Hopf bifurcations, as already considered in [**15**]. Presently, however, we include the coupling term $O(r^4)$. Moreover, we include $2n$ parameters

2000 ***Mathematics Subject Classification.*** — 37L10, 70K43.
Key words and phrases. — Coupled Hopf bifurcations, center manifold reduction, quasiperiodic tori, quasiperiodic bifurcation.

$(\alpha,\beta) = (\alpha_1,\ldots,\alpha_n,\beta_1,\ldots,\beta_n)$, where $\alpha \in \mathbb{R}^n_+$ is small and where β varies over any compact disc $L \subset \mathbb{R}^n_+$. In multi-polar coordinates $x_j = r_j \cos\varphi_j, y_j = r_j \sin\varphi_j$ the decoupled lower order part reads

$$
\begin{aligned}
\dot{\varphi}_j &= \beta_j \\
\dot{r}_j &= r_j(\alpha_j - r_j^2),
\end{aligned}
\tag{2}
$$

$1 \leqslant j \leqslant n$. Clearly for $\alpha_j > 0$, system (2) has an n-torus attractor $r_j = \sqrt{\alpha_j}$, where the dynamics is parallel given by $\dot{\varphi}_j = \beta_j$, $1 \leqslant j \leqslant n$. Our interest is with the fate of this dynamical phenomenon upon addition of the higher order perturbation $O(r^4)$.

Below we generalize the setting of Example 1, raising a similar problem. To solve this we apply both Center Manifold Theory [**14**] and KAM Theory in the dissipative setting [**17, 2, 6, 3, 7**]. We summarize the results of our investigation. First, for small $|\alpha|$, the family of n-tori is C^ℓ-persistent for such perturbations, where the bound on $|\alpha|$ depends on ℓ. Second, the continuum of parallel dynamics persists as a Whitney smooth family of quasiperiodic attractors, foliated over a Cantor set. Projected to the (α,β)-parameter space, this Cantor foliation has positive measure, expressed in terms of a Lebesgue density point of quasiperiodicity corresponding to $\alpha = 0$. Notably, the dynamics in between generically break up due to internal resonance: upon variation of parameters the dynamics can be asymptotically periodic (e.g., phase locked) or chaotic ($n \geqslant 3$), [**20, 18, 21, 23**].

Let us briefly outline the contents of this paper. We start by generalizing the setting of Example 1, and developing an appropriate perturbation model for the application of Center Manifold and KAM Theory. We end by a general discussion, pointing towards some interesting problems regarding quasiperiodic Hopf bifurcation that occur in a subordinate way.

2. Coupled Hopf-bifurcations

2.1. Setting of the problem. — Instead of weakly coupled case (1) we here consider the more general system

$$
\begin{pmatrix} \dot{x}_j \\ \dot{y}_j \end{pmatrix} = \begin{pmatrix} \alpha_j & -\beta_j \\ \beta_j & \alpha_j \end{pmatrix} \begin{pmatrix} x_j \\ y_j \end{pmatrix} + O(r^2),
\tag{3}
$$

which will be subject to suitable C^3-open conditions. As in Example 1 we include dependence on the parameter vector $(\alpha,\beta) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$. We apply a standard normal form procedure to system (3), e.g. compare [**22, 4, 24, 11**]. Note that the linear part of (3) has a $\mathbb{T}^n$-symmetry. Strong resonances are excluded by requiring that

$$
\sum_{j=1}^{n} \beta_j k_j \neq 0 \text{ whenever } 0 < \sum_{j=1}^{n} |k_j| \leqslant 4,
\tag{4}
$$

which amounts to the first open condition. Granted (4), by near identity, polynomial changes of variables this $\mathbb{T}^n$-symmetry can be pushed over the whole 3-jet, which then in appropriate multi-polar coordinates reads

$$\begin{aligned}\dot{\varphi}_j &= \beta_j - f_j(r_1^2, \ldots, r_n^2) + O(r^4) \\ \dot{r}_j &= r_j(\alpha_j - g_j(r_1^2, \ldots, r_n^2)) + O(r^4),\end{aligned} \tag{5}$$

$1 \leqslant j \leqslant n$. As in Example 1, we truncate the $O(r^4)$ term, so arriving at the present generalization of (2), which we explore for invariant n-tori. Therefore we expand

$$\begin{aligned}f_j &= f_{j1}r_1^2 + \cdots + f_{jn}r_n^2 \\ g_j &= g_{j1}r_1^2 + \cdots + g_{jn}r_n^2,\end{aligned}$$

with f_{ji}, g_{ji} constants, $i, j = 1, 2, \ldots, n$. Invariant n-tori then are determined by n equations

$$g_{j1}r_1^2 + \cdots + g_{jn}r_n^2 = \alpha_j, \tag{6}$$

$1 \leqslant j \leqslant n$. Consider the 'action-space' $\mathbb{R}_+^n = \{r_1^2, \ldots, r_n^2\}$, where the equations (6) determine n hyperplanes. Considering the $n \times n$-matrix

$$G = \left((g_{j,i})_{j,i=1}^{\;\;n} \right),$$

we impose further C^3-open conditions

$$\det G \neq 0 \text{ while } G^{-1}(\alpha) \in \mathbb{R}_+^n. \tag{7}$$

By $c^2 := G^{-1}(\alpha)$ denote the unique (transversal) intersection point of the hyperplanes. Then the equations (6) have the unique solution

$$r_1^2 = c_1^2, \ldots, r_n^2 = c_n^2, \text{ with } c_j^2 = c_j^2(\alpha), \tag{8}$$

$1 \leqslant j \leqslant n$, which determines our invariant n-torus, carrying parallel dynamics.

Remark. — The open conditions (7) are trivially satisfied in Example 1, where $G = \mathrm{Id}_n$. One easily detects other concrete examples that are C^3-nearby (1).

As announced in Example 1, the problem is to study the effect of the higher order perturbaton $O(r^4)$ on this family of tori. As said before, to answer this we shall apply both Center Manifold Theory [**14**] and KAM Theory in the dissipative setting [**6, 3, 7**].

2.2. An appropriate perturbative setting. — We formulate a perturbation problem suitable for our purposes. Introducing the small parameter ε and putting $I_j = r_j - c_j$, we scale

$$\begin{aligned}\alpha_j &= \varepsilon^2\overline{\alpha}_j & \beta_j &= \overline{\beta}_j \\ r_j &= \varepsilon\overline{r}_j & \varphi_j &= \overline{\varphi}_j,\end{aligned}$$

also writing $I_j = \varepsilon \overline{I}_j$, $1 \leqslant j \leqslant n$. This gives the estimates

$$\begin{aligned} \dot{\varphi}_j &= \overline{\beta}_j + \varepsilon O(\overline{I}) \\ \dot{\overline{I}}_j &= -2\varepsilon^2 \overline{\alpha}_j \overline{I}_j + \varepsilon^2 \overline{I}_j O(\sqrt{\overline{\alpha}} \overline{I}), \end{aligned}$$

$1 \leqslant j \leqslant n$, which are uniform for $(\overline{\alpha}, \overline{\beta}) \in K \times L$, for any given compact subsets $K, L \subseteq \mathbb{R}^n_+$. A further scaling

$$\overline{I} = \varepsilon^q \overline{\overline{I}},$$

for a fixed $q > 1$, leads to the following perturbation problem:

$$\begin{aligned} \dot{\varphi}_j &= \overline{\beta}_j + O(\varepsilon^{1+q}) \\ \dot{\overline{\overline{I}}}_j &= -2\varepsilon^2 \overline{\alpha}_j \overline{\overline{I}}_j + O(\varepsilon^{2+q}), \end{aligned} \tag{9}$$

$1 \leqslant j \leqslant n$, again with uniform estimates. Since $q > 1$, the form (9) is suitable for application of the Center Manifold Theorem [**14**], Thm. 4.1, implying the C^ℓ-persistence of the invariant n-torus for small values of ε.

To further investigate persistence of the quasiperiodic dynamics we apply dissipative KAM Theory as developed in [**6**] § § 4 and 8. In the unperturbed case

$$\begin{aligned} \dot{\varphi}_j &= \overline{\beta}_j \\ \dot{\overline{\overline{I}}}_j &= -2\varepsilon^2 \overline{\alpha}_j \overline{\overline{I}}_j, \end{aligned}$$

$1 \leqslant j \leqslant n$, we single out parameter vectors $\beta \in L$ such that for all $k \in \mathbb{Z}^n \setminus \{0\}$ Diophantine conditions

$$|\langle k, \beta \rangle| \geqslant \frac{\gamma}{|k|^\tau} \tag{10}$$

hold. Here $\tau > n - 1$ is a constant, while we choose $\gamma = c\varepsilon^q$, for an appropriate (sufficiently small) constant c, depending on K and L. These conditions define a Cantor foliation $\mathcal{C}_{\varepsilon,c} \subseteq K \times L$ for the unperturbed system. The complement $(K \times L) \setminus \mathcal{C}_{\varepsilon,c}$ has measure $O(\varepsilon^q)$ as $\varepsilon \downarrow 0$. The main result of the present paper is:

***Theorem 2* (Perturbation Theorem).** — *Consider system (9), with parameter vectors $(\overline{\alpha}, \overline{\beta}) \in K \times L$, for given compact subsets $K, L \subseteq \mathbb{R}^n_+$. Also let $\ell \in \mathbb{N}$ be sufficiently large. Then, for $\varepsilon > 0$ and sufficiently small the (9) has the following properties:*

(1) *The unperturbed n-torus family $I = 0$ persists as a unique C^ℓ-family of hyperbolic n-torus attractors $\mathcal{T}_\varepsilon$, also depending C^ℓ on ε.*

(2) *For parameter values $(\overline{\alpha}, \overline{\beta}) \in \mathcal{C}_{\varepsilon,c}$ as described above, with c sufficiently small, the unperturbed tori persist as quasiperiodic tori inside $\mathcal{T}_\varepsilon$.*

(3) *The union of tori inside $\mathcal{T}_\varepsilon$ with non-quasiperiodic dynamics has Lebesgue measure $O(\varepsilon^q)$, as $\varepsilon \downarrow 0$, $1 \leqslant j \leqslant n$, uniformly in $(\overline{\alpha}, \overline{\beta}) \in K \times L$.*

Proof (sketch). — For simplicity we perform the KAM Theory inside the center manifold, see [**6**], the Appendix. This means that for some $\ell' < \ell$, there exists a $C^{\ell'}$-reparametrization $\Phi_\varepsilon : K \times L \to \mathbb{R}^n_+ \times \mathbb{R}^n_+$, extending to a $C^{\ell'}$-diffeomorphism

$$\Psi_\varepsilon : \mathbb{T}^n \times \{0\} \times K \times L \longrightarrow \mathbb{R}^{2n} \times \mathbb{R}^n_+ \times \mathbb{R}^n_{\cdot}$$

such that for $(\overline{\alpha}, \overline{\beta}) \in \mathcal{C}_\varepsilon$, the map Ψ_ε is a conjugacy from the unperturbed system to the full (perturbed) system (9).[(1)] The dependence of the map Ψ_ε on ε is smooth.

This implies that (9) has a subsystem consisting of a Whitney smooth foliation of quasiperiodic n-tori. Morover, in the $C^{\ell'}$-topology $|\Psi_\varepsilon - \mathrm{Id}| = o(\varepsilon)$ as $\varepsilon \downarrow 0$, which implies that the the image $\Phi(\mathcal{C}_\varepsilon)$ has positive (almost full) measure as $\varepsilon \downarrow 0$. □

Returning to the context of system (3) we conclude

Corollary 3. — *Consider system (3), with parameter vectors $\overline{\alpha} \in K$, and $\overline{\beta} \in L$, for any given compact sets $K, L \subset \mathbb{R}^n_+$ with L not containing any strong resonances (4). Also let $\ell \in \mathbb{N}$ be sufficiently large. Then, up to condition (7), for $\varepsilon > 0$, sufficiently small, the system (3) has the properties 1., 2. and 3. of Theorem 2. Moreover, the behaviour described above is persistent under sufficiently C^3-small perturbations.*

The $\mathcal{T}_\varepsilon$-dynamics in between the Cantor foliation generically break up due to internal resonance. Finally notice that the weakly coupled case of Example 1 also is fully covered by the Corollary.

2.3. Conclusive remarks, towards quasiperiodic Hopf bifurcation. — Concerning the asymptotics of the measure estimate, we note that the KAM Theorem [**6, 7**] may well be applied in the C^∞-setting, keeping track of the normal linear part. This does not lead to contradictions, since the quasiperiodic tori by parallellity are of class C^∞, since they are ℓ normally hyperbolic for any ℓ. A further (quasiperiodic) normalizing and reparametrizing (see, e.g., [**3**]) leads to sharper estimates on the complement of quasiperiodicity, which become of arbitrary large order in ε, as $\varepsilon \downarrow 0$, compare [**7**], § 5.2. Assuming real analyticity, even exponentially small estimates can be obtained for the complement of quasiperiodicity. Compare [**7, 16**] and further references given there.

Next we point out a relationship with quasiperiodic bifurcation theory. Until now we restricted $\overline{\alpha}$ to compact subsets K of the open cone $\mathbb{R}^n_+$, thereby only covering interior points. Consider the $\overline{\alpha}$-regime near the boundary of this cone. In the unperturbed system (2) of Example 1, this gives normally elliptic subtori, and clearly the perturbative cases involve subordinate quasiperiodic Hopf bifurcations, compare [**3, 7, 8, 5, 25, 9, 10, 23**]. In the present, general case the theory will be even more interesting, compare [**13, 26**], for general reference also see [**1**]. Investigation of

(1) To be precise, $K \times L$ has to be shrunk at its boundary by a layer of thickness $O(\varepsilon^q)$.

this will require further research, as is also indicated by the following example, kindly provided by Florian Wagener.

***Example 4* (Two weakly coupled Hopf families).** — Consider the 'integrable' system

$$\begin{aligned}\dot{\varphi}_1 &= \beta_1\\ \dot{\varphi}_2 &= \beta_2\\ \dot{r}_1 &= r_1(\alpha_1 - r_1^2) - r_2^4\\ \dot{r}_2 &= r_2(\alpha_2 - r_2^2).\end{aligned}$$

A brief calculation reveals that the non-hyperbolic tori occur for

$$\begin{array}{lll} 27\alpha_2^4 = 4\alpha_1^3 & r_1 = \frac{1}{\sqrt{3}}\sqrt{\alpha_1} & r_2 = \sqrt{\alpha_2},\\ \alpha_2 = 0 & r_1 = \sqrt{\alpha_1} & r_2 = 0.\end{array}$$

The problem is to study the effects of integrable and non-integrable higher order terms. Notice that the hyperplanes in parameter space where subordinate quasiperiodic Hopf bifurcations take place, have shifted somewhat. However, by the scaling all interior points of the cone are drawn within the regime with invariant n-tori.

Observe that this program is reminiscent to the Hamiltonian case of elliptic subtori, [**6, 3, 7, 16**].

Acknowledgements. — Helpful discussions are acknowledged with Peter Poláčik, Florian Wagener, Carles Simó, Floris Takens and Bernold Fiedler. Also thanks is due to the Universitat de Barcelona for hospitality during part of the preparation of this paper.

References

[1] C. Baesens, J. Guckenheimer, S. Kim and R.S. MacKay, Three coupled oscillators: Mode-locking, global bifurcation and toroidal chaos, *Physica D*, (1991), 387–475.

[2] B.L.J. Braaksma & H.W. Broer, On a quasi-periodic Hopf bifurcation, *Ann. Institut Henri Poincaré, Analyse non linéaire*, **4**, no.2 (1987), 115–168.

[3] B.L.J. Braaksma, H.W. Broer and G.B. Huitema, Towards a Quasiperiodic Bifurcation Theory, *Mem. AMS*, **83**(421) (1990), 83–175.

[4] H.W. Broer, Formal normal forms for vector fields and some consequences for bifurcations in the volume preserving case. In: D. Rand and L.S. Young (eds.), *Dynamical Systems and Turbulence, Warwick 1980* LNM **898** (1981), 75–89, Springer-Verlag.

[5] H.W. Broer, KAM-Theory: Multi-Periodicity in conservative and dissipative systems, *Nieuw Arch. Wisk.* **14**(1), (1996), 1–15.

[6] H.W. Broer, G.B. Huitema and F. Takens, Unfoldings of Quasi-Periodic Tori, *Mem. AMS*, **83**(421) (1990), 1–82.

[7] H.W. Broer, G.B. Huitema and M.B. Sevryuk, Quasi-periodic tori in families of dynamical systems: order amidst chaos, *LNM* **1645**, Springer Verlag 1996.

[8] H.W. Broer & F. Takens, Mixed spectrum and rotational symmetry, *Arch. Rational Mech. An.* **124** (1993), 13–42.

[9] H.W. Broer, F. Takens and F.O.O. Wagener, Integrable and non-integrable deformations of the skew Hopf bifurcation, *Reg. and Chaot. Dyn.* **4** (1999), 17–43.

[10] H.W. Broer and F.O.O. Wagener, Quasi-periodic stability of subfamilies of an unfolded skew-Hopf bifurcation, *Arch. Rational Mech. An.* **152** (2000), 283–326.

[11] S-N. Chow, C. Li and D. Wang, *Normal Forms and Bifurcation of Planar Vector Fields.* Cambridge University Press 1994.

[12] B. Fiedler and P. Poláčik, Complicated dynamics of scalar reaction diffusion equations with a nonlocal term, *Proc. Roy. Soc. Edinburgh Sect. A, (Mathematical and Physical Sciences)* **115**(1-2) (1990), 167–192.

[13] J. Guckenheimer and P. Holmes, Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields, *Appl. Math. Sciences* **42**, Springer Verlag 1983.

[14] M.W. Hirsch, C.C. Pugh and M. Shub, Invariant Manifolds, *LNM* **583**, Springer Verlag 1977.

[15] E. Hopf, A mathematical example displaying features of turbulence. *Commun. Appl. Math.* **1** (1948), 303–322.

[16] A. Jorba and J. Villanueva, On the Normal Behaviour of Partially Elliptic Lower-Dimensional Tori of Hamiltonian Systems, *Nonlinearity* **10** (1997), 783–822.

[17] J.K. Moser, Convergent series expansions for quasi-periodic motions, *Math. Ann.* **169** (1967), 136–176.

[18] S.E. Newhouse, D. Ruelle and F. Takens, Occurrence of strange Axiom A attractors near quasiperiodic flows on $\mathbb{T}^m$, $m \leqslant 3$, *Commun. Math. Phys.* **64** (1978), 35–40.

[19] P. Poláčik and V. Šošovička, Stable periodic solutions of a spatially homogeneous nonlocal reaction-diffusion equation, *Proc. Roy. Soc. Edinburgh Sect. A, (Mathematical and Physical Sciences)* **126**(4) (1996), 867–884.

[20] D. Ruelle and F. Takens, On the nature of turbulence, *Commun. Math. Phys.* **20** (1971), 167–192; **23** (1971), 343–4.

[21] C. Simó, Perturbations of translations in the two-dimensional torus: the case near resonance, *Proceedings VI CEDYA*, Universidad de Zaragoza, 1984.

[22] F. Takens, Singularities of Vector Fields, *Commun. Math. Phys.* **20**, 1970, 167–192, **23**, (1971), 343–344. *Publ. I.H.E.S.* **43** (1974), 47–100.

[23] F. Takens and F.O.O. Wagener, Resonances in skew and reducible quasi-periodic Hopf bifurcations, *Nonlinearity* **13** (2000), 377–396.

[24] A. Vanderbauwhede, Centre Manifolds, Normal Forms and Elementary Bifurcations, *Dynamics Reported* 2 (1989), 89–170.

[25] F.O.O. Wagener, On the skew Hopf bifurcation. *PhD thesis,* University of Groningen, 1998.

[26] H. Zoladek, Bifurcation of a certain family of planar vector fields tangent to axes, *Journ. Diff. Eqns.* **67** (1987), 1–55.

H. Broer, Department of Mathematics and Computer Science, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands

Astérisque
286, 2003, p. 231–248

WALKS IN RIGID ENVIRONMENTS: SYMMETRY AND DYNAMICS

by

Leonid A. Bunimovich

Abstract. — We study dynamical systems generated by a motion of a particle in an array of scatterers distributed in a lattice. Such deterministic cellular automata are called Lorentz-type lattice gases or walks in rigid environments. It is shown that these models can be completely solved in the one-dimensional case. The corresponding regimes of motion can serve as the simple dynamical examples of diffusion, sub- and super-diffusion.

1. Introduction

Deterministic (dynamical systems) or stochastic (random processes) models are the ones which were used traditionally to model real phenomena and processes. The theory of these two types of models, purely deterministic and purely stochastic ones, is very rich and therefore the intuition on evolution of such systems is well developed. The intuition means a right expectation of what should happen in the course of evolution of some concrete system even though the rigorous mathematical analysis is usually lacking.

Such intuition is based on some explicitly solvable simple (but nontrivial) and visible examples, i.e., on the comprehensive mathematical analysis of the corresponding models. These fundamental models in the theory of stochastic processes include sequences of identically distributed independent random variables (Bernoulli shifts), a random walk, etc. In dynamical systems such fundamental models include a rotation of a circle, an algebraic toral automorphism, some billiard models, etc. Certainly, this class of completely solvable models is growing, and our intuition is essentially growing with it. I cannot resist to mention the quadratic family which now finally belongs to this class as well [**14**].

2000 ***Mathematics Subject Classification.*** — 37A50, 60K35, 82C20, 82C41.
Key words and phrases. — Deterministic walks in random environments, Lorentz lattice gases, scatterers, diffusion, sub-diffusion, super-diffusion.

This work was partially supported by NSF grant DMS-9970215.

However, dynamics of many (and actually of a majority) of real systems is neither purely stochastic nor purely deterministic but it rather has both these components. Certainly, it is the well known fact and traditional attempts to account for that is to study e.g., small random perturbations of dynamical systems or to add a small deterministic flow term (advection) to a diffusion process. Such small perturbations, while being very important to study, do not address the question on the behavior of hybrid systems with (nonsmall) deterministic and stochastic features in their evolution. In fact, in applications almost always models were chosen as stochastic ones (instead of hybrid ones) with the standard argument that each real phenomenon or process has infinitely many features neglected by any model and therefore it is, in fact, a random process.

There are large areas like e.g., operations research, logistics, etc., which still completely belong to the probability theory while already the first applications of the dynamical systems methods allowed to achieve very encouraging results by essentially increasing production rates of certain production lines [**2**].

Another class of hybrid systems goes back to the classical Lorentz gas. Recall that in the Lorentz gas (light) point, noninteracting between themselves, particles move by inertia in an array of immovable scatterers and collide with scatterers elastically. It is a dynamical system which can be reduced to Sinai billiard. This system has been comprehensively studied and until now it is the only one nontrivial system for which time irreversible macroscopic dynamics (governed by the diffusion equation) has been rigorously derived from the time reversible microscopic dynamics (governed by Newton equations). It is transparent that this result has been obtained only for periodic configurations of scatterers (under the condition that a free path of the point particle is bounded, see details in [**6**]).

The very interesting mathematically and important problem for various applications is to study this system in case when the scatterers are distributed randomly. It seems, at the first sight, that this problem should follow from the one with periodic distribution of scatterers because of some additional "self-averaging" generated by a random distribution of scatterers. Indeed, it seems that such "self-averaging" should just improve stochastic properties of the corresponding dynamical system with periodically placed scatterers. However, this idea is totally wrong. In fact, in the Lorentz gas with randomly distributed scatterers we encounter a hybrid system, which has both deterministic and stochastic features. (Certainly, the Lorentz gas with randomly distributed scatterers can be described as purely deterministic (dynamical) system. However, it does not make this system to be deterministic, as well as the representation of a stationary random process a shift in the space of its realizations does not transform this stochastic process into a deterministic one.)

If an interesting and important system does not allow a comprehensive analysis then it is natural to consider some simpler model which retains (some) principal features of this system. Such simplified Lorentz gas model has been introduced in [**18**]. In this model scatterers (usually of two different types, e.g., left and right mirrors aligned along the diagonals of the square lattice) are randomly distributed on vertices of the square lattice. The point particle moves with unit speed along the bonds of this lattice and get reflected by the scatterers. These systems were naturally called Lorentz Lattice Gases (LLG). It is worthwhile to mention that this model is the generalization of another classical model in nonequilibrium statistical mechanics, which is called the (Ehrenfests') Wind-Tree model. In the Wind-Tree model a (light) point particle moves in an array of randomly distributed scatterers, which are identical rhombuses with parallel diagonals. The particle moves parallel to one of the diagonals of rhombuses and therefore after (elastic) reflection from the boundary of some scatterer, its velocity becomes parallel to another diagonal of the rhombuses and so on.

The Lorentz Lattice Gases belong to the class of systems which can be naturally called Deterministic Walks in Random Environments (DWRE). Indeed the dynamics of these systems is generated by deterministic motion of the particle, where both the free motion and reflections from the boundary of scatterers are deterministic, while distribution of scatterers is random.

It occurred that the Lorentz Lattice Gases were studied (without using this name) in lots of applications, e.g., in material science, superconductivity, chemical kinetics, information transmission and especially in the theoretical computer science. All these studies were exclusively numerical and these systems were included in the class of systems which are conventionally called "complex systems" (and are often discussed in the journal with the same name).

In fact, in many applications there were considered so called flipping LLG, where the moving particle has impact on an environment as well. Formally dynamics of such models is defined by the rule that after reflection of the moving particle from a scatterer this scatterer instantly changes its type. Therefore in flipping LLG there is also a dynamics of an environment formed by the configuration of scatterers. Hence for such models it makes sense to consider dynamics of many particles moving along the bonds of a lattice rather than of a single one. Indeed, even though the moving particles do not interact directly they do, in fact, interact via changing the environment to each other. It allows to account for an "information exchange" between particles (signals, etc.) and environment (neurons, etc.), see. e.g., [**1, 7, 9, 10, 12**].

From the mathematical point of view all these models are dynamical systems. In fact, they belong to the class of deterministic cellular automata. However, this formal observation does not help much in studying these systems. In fact, it occurred that the much more productive approach is to consider all these models as Deterministic Walks in Random Environments. (To make clear distinction with purely stochastic models of this kind we mention that in the last ones a scatterer after colliding with particle "flips a coin" to decide whether it should change its type.)

In the studies of DWRE the important role is played both by the structure of a lattice where particles move (which could be e.g., the square, triangular, cubic, random, etc. lattice) and by the types of scatterers considered (e.g., there are 4^4 types of scatterers in a square lattice). It is not surprising, of course, because a lattice defines a configuration space and the types of scatterers (together with a lattice) define the dynamics (equations of motion).

The great majority of papers on DWRE are numerical. There are as well quite a few mathematical results on dynamics of DWRE. They usually use some specific features of the given model, which allow sometimes to come up with complete solution. For instance, it is possible to reduce a (purely deterministic) problem to a (purely probabilistic) percolation problem on some graph [**4**]. (It is worthwhile to mention that such graph is defined not only by the lattice but by the types of scatterers as well.) Sometimes it was possible to completely solve the problem by constructing some peculiar class of solutions and by proving that no other solutions exist (see e.g., [**5**]). However, in most cases the results were rather counterintuitive. Actually, in almost all cases when dealing with the hybrid (neither purely deterministic, nor purely stochastic) systems the authors confessed that they obtained results different from what they expected.

This situation clearly calls for some kind of a general view at these systems, especially the one which would allow to integrate the studies of DWRE in fixed and in evolving (e.g., flipping) environments. The corresponding approach has been developed in [**3**] where these two classes of DWRE, were integrated into one class of dynamical systems called Walks in Rigid Environments (WRE). (Observe that R in DWRE refers to "random," while in WRE it refers to "rigid").

WRE is also a dynamical system generated by motion of point particles in some graph (e.g., in a lattice). For the sake of simplicity we will consider here only one-particle systems. Some scatterers are randomly distributed along the vertices of this graph. (Again for the sake of simplicity we assume that the scatterers are distributed independently, even though one may assume that they interact via some potential.)

The crucial feature of WRE is the new parameter r which is called a rigidity of an environment. The rigidity determines how many times the particle must collide with the given scatterer in order to change its type. In the other words, the scatterer at a given site changes its type at the moment after the rth visit of the moving particle to this site. It is easy to see that the LLGs with fixed environment correspond to the case $r = \infty$, while the LLGs with flipping environment correspond to the case $r = 1$. Thus the two studied so far classes of LLG form, in fact, two extreme sub-classes of WRE.

Besides the introduction of Walks in Rigid Environments allowed to move rigorous studies of LLG to another level and to address the central problem of the theory of such systems which is the diffusion problem. Until [**3**] the mathematical papers on Deterministic Walks in Random Environments usually addressed the problem whether a typical path of a particle is bounded or unbounded. However, the most important question which one can ask about evolution of a system generated by a motion of some object (particle, signal, etc.) is where this object is going to be at a sufficiently large moment of time t. The quantity of interest is the mean square displacement $Ez^2(t)$ (or, in other words, the expectation, taken with respect to the distribution of environments, of a (squared) position $z(t)$ of the particle at time t). One distinguishes diffusive, subdiffusive and superdiffusive behavior which correspond to the linear, slower than linear and to faster than linear growth of $Ez^2(t)$ respectively.

It has been shown in [**3**] that the asymptotic behavior of the particles' position is determined by an interplay between the symmetries of the lattice and symmetries of scatterers.

The present paper deals with WRE where the problem of the particle's diffusion can be solved completely. We give the examples of all three situations, i.e., diffusion, sub- and super-diffusion. Moreover, in these examples it was possible to completely "separate" stochastic and deterministic elements of the evolution of these models.

Qualitatively the situation is the following one. Stochastic evolution of the system takes place when the particle visits some site of the lattice at the first time, while between two consecutive visits to the new (nonvisited before) sites the particle undergoes a deterministic evolution. This deterministic evolution is completely defined by the types of the scatterers allowed in the model under study and by their symmetries. It is exactly this deterministic evolution defines the speed of growth of visited (exited) domain.

Such separation of the evolution into random events and intermediate deterministic motion allowed to describe in one-dimensional case all three types of behavior in the same way.

It occurred that the evolution of the particle can be broken into the qualitatively similar stages. Each such stage is characterized by deterministic motion of the particle in some box of a random size. In cases of diffusion and of subdiffusion the sizes of these boxes are growing in time, while in case of super-diffusion the sizes of these boxes fluctuate and the boxes are moving along the lattice in one direction, which is defined by the initial distribution of scatterers near the origin.

Actually the analysis of all these three models is rather straight-forward and they could be used in the first courses of dynamical systems and/or random processes as completely solvable models which are neither purely deterministic nor purely stochastic to develop intuition on systems with such mixed type of behavior.

The structure of the paper is the following. In Sect. 2 we give the necessary definitions and formulate the results. The proofs are given in Sect. 3. The last Sect. 4 contains some concluding remarks.

2. Definitions and main results

Consider an one-dimensional regular lattice which, without any loss of generality, could be identified with the set of integers $\mathbb{Z}$. We assume that at each site $z \in \mathbb{Z}$ there is a scatterer of some type. A particle moves with the unit speed along the lattice $\mathbb{Z}$, i.e., $v(t) = 1$ or $v(t) = -1$ at each moment of time t. Denote by $z(t)$ position of the particle at time t. Then the position of the particle at the next moment of time is determined by $v(t)$ and by the type of scatterer located at the site $z(t)$. Certainly it is enough to consider a discrete time. To distinguish between two moments of time when the particle reached some site of the lattice but had not yet reflected by a scatterer at this site and the one when it was just reflected by a scatterer we will denote these moments by t and t_+ respectively. Hence $v(t)$ is the velocity with which the particle approaches a site $z(t)$ and $v(t_+)$ is the velocity with which the particle leaves this site.

It is clear that in dimension one there are 2^2 possible scatterers (or local scattering rules), which we will denote by BS, FS, LS and RS. Here BS is the backward scatterer, which changes the velocity of the particle to the opposite one. In other words, if BS is located at a site $z(t) \in \mathbb{Z}$ then $v(t_+) = -v(t)$. FS is the trivial, or forward scatterer which does not change the velocity of the particle, i.e., $v(t_+) = v(t)$ if at the site $z(t)$ was the forward scatterer. The last two types of scatterers, LS and RS, which we will refer to as the left and the right scatterer respectively, are the semi-transparent ones. Namely LS (RS) sends all scattered particles to the left (right), i.e., if a LS (RS) is located at a site $z(t) \in \mathbb{Z}$ then $z(t+1) = z(t) - 1(z(t+1) = z(t) + 1)$.

Now we will define the dynamics of our system. In order to do it we introduce an integer r, $1 \leqslant r \leqslant \infty$, which we will refer to as a *rigidity* of an environment. Let $\widehat{S}$ be a space of all possible scatterers on a lattice under consideration. (Recall that in this paper we discuss only WRE in one-dimensional lattice $\mathbb{Z}$, i.e., $\widehat{S} = \{BS, FS, LS, RS\}$.)

WRE is defined by three objects:

(1) A subspace $S \subset \widehat{S}$ of scatterers, which we will call a space of allowed scatterers.
(2) An integer $r > 0$ (rigidity).
(3) A function $e : S \to S$.

Let $S_r = S \times \{0, 1, \ldots, r-1\}$ and $\pi : S_r \to S$ is the natural projection. Denote a function $a : S_r \to S_r$ as

$$a(S, i) = \begin{cases} (S, i+1), & \text{if } 0 \leqslant i < r-1 \\ (e(s), 0), & \text{if } i = r-1, \end{cases} \tag{1}$$

where $s \in S$. We will call i an index of the corresponding scatterer.

We will denote by $s(z)$ a type of scatterer which is located at the site $z \in \mathbb{Z}$. The type of scatterer at z may change in the course of dynamics (if $r < \infty$). By $(s(z))_t$ we denote the type of a scatterer located at a site $z \in \mathbb{Z}$ at a moment of time t. The notation $s(z(t))$ will be referred to a type of scatterer located at a moment t at the site where the particle sits at this moment.

The configuration space of our system $W = S_r^{\mathbb{Z}} \times \mathbb{Z}$, where $S_r^{\mathbb{Z}}$ is a configuration of scatterers (together with a number of visits occurred to a site $z \in \mathbb{Z}$ while a scatterer of some fixed type was located there) and the second factor $\mathbb{Z}$ corresponds to the position of the particle. The phase space $\Omega = W \times \{-1, 1\}$.

Now we are able to write the equations governing the dynamics

$$
\begin{aligned}
v(t+1) &= g(v(t), s(z(t))), \\
z(t+1) &= z(t) + v(t+1), \\
((s(z))_{t+1}, i) &= ((s(z))_t, i) \qquad \text{if } z \neq z(t) \\
((s(z(t), i) &= a(s(z(t)), i) \qquad \text{if } z = z(t).
\end{aligned} \tag{2}
$$

The function $g(v(t), s(z(t)))$ in (2) is completely defined by the type of scatterer $s(z(t))$. (The formal expressions for an abstract scatterer are rather cumbersome. It would become simple though when we consider concrete models of WRE.)

We will introduce two such models. In the first model we will take semi-transparent scatterers LS and RS as the set S of admissible scatterers. The second model corresponds to $S = \{BS, FS\}$.

We describe now the dynamics of these two models informally (but precisely and in more visible way than it is formally defined by the relations (2)).

Each of the models under consideration deals with two types of scatterers. The particle moves with unit velocity along the lattice $\mathbb{Z}$. At each integer moment of time t it comes to some vertex $z(t) \in \mathbb{Z}$ and gets scattered by the scatterer located at this moment at $z(t)$. (A function $g(\cdot, \cdot)$ is immediately specified by the type of this scatterer.) If the particle was scattered r consecutive times by this scatterer located at $z(t)$ (i.e., if particle returned to this site with this very scatterer r times) then this scatterer gets changed to another type.

Now we need to specify initial conditions for our dynamical system. Without any loss of generality we can always assume that the particle starts at the origin with the initial velocity $v(0) = 1$. We take Bernoulli measure on space of scatterers' initial configurations, i.e., the types of scatterers at different sites are chosen independently and have the same distributions.

Two models under consideration have quite different symmetry properties. The only nontrivial symmetry of the lattice $\mathbb{Z}$ is the reflection with respect to the origin. (Indeed the probability distributions on initial configurations of scatterers are translationally invariant.) Observe now that LS and RS do respect this symmetry, while BS and FS do not. It is the key point why dynamical properties of these models are quite different as we will see later.

It is easy to see that an orbit of any WRE is completely defined by the initial configurations of scatterers. We will use sometimes the same notation ω to denote an orbit of a dynamical system and the corresponding configuration of scatterers. Another remark is that initially (at $t = 0$) all scatterers have indices zero.

For the sake of brevity we will refer to the model with $S = \{LS, RS\}$ as to the model with oriented scatterers (OS-model) and to the model with $S = \{BS, FS\}$ as to NOS-model (the model with non-oriented scatterers).

We start with the formulation of the results on the qualitative behavior of the OS- and NOS-models and then turn to their quantitative behavior.

The first simple remark is that the dynamics of both models is trivial (and similar) in case when the environment does not change in time ($r = \infty$). Indeed the particle will with probability one oscillate between two closest to the origin BS (for the NOS-model) with positive and non-positive coordinate respectively, or between the closest to the origin LS with positive coordinate and the closest to the origin RS with non-positive coordinate in the OS-model.

It is the characteristic feature of hybrid systems (intermediate ones between purely deterministic and purely stochastic) that an exceptional set of orbits of measure zero can often be completely characterized. For instance, if $r = \infty$ this set consists of initial configurations of scatterers where all scatterers with positive coordinates are RS (for the OS-model) or FS (for the NOS-model) or/and all scatterers with non-positive coordinates are LS (for the OS-model) or FS (for the NOS-model). The dynamics of the OS-model is characterized qualitatively by the following statement.

Theorem 1. — *In the OS-model for any value of rigidity $r < \infty$ the particle will almost surely visit each site of the lattice $\mathbb{Z}$ infinitely many times. Moreover, for almost every point $\omega \in \Omega$ of the phase space there exists a sequence of moments of time τ_i, $i = 0, 1, \ldots$, $\tau_0 = 0$, $\tau_i < \tau_{i+1}$, $\tau_i \to \infty$ as $i \to \infty$ and a corresponding sequence of closed intervals $B_i(\omega) = [a_i(\omega), b_i(\omega)] \subset \mathbb{Z}$, $i = 1, 2, \ldots$, $a_i(\omega) \leqslant 0$, $b_i(\omega) > 0$, $B_i(\omega) \subset B_{i+1}(\omega), B_i(\omega) \to (-\infty, \infty)$, as $i \to \infty$, such that within a time interval $\tau_{i-1} < t < \tau_i$, $i = 1, 2, \ldots$, the particle stays inside the interval $B_i(\omega)$ and visits the origin $z = 0$ $2r$ times.*

Thus Theorem 1 shows that in the OS-model for any finite value of rigidity the particle will oscillate about origin with an increasing amplitude. We will say that a point $\omega \in \Omega$ has a positive (negative) tail of scatterers of some type if there exists $z_+ > 0$ ($z_- \leqslant 0$) such that all scatterers at the sites $z \geqslant z_+$ ($z \leqslant z_-$) are of one and the same type.

Corollary 1. — *The exceptional set of measure zero in* Theorem 1 *consists of such points $\omega \in \Omega$, where the corresponding configurations of scatterers contains a positive tail of RS or/and a negative tail of LS.*

Denote by $z_{\max}(t)$ and $z_{\min}(t)$ the sites with the maximal and the minimal coordinates respectively visited by the particle to a moment t. The next theorem describes quantitative features of the dynamics of the OS-model. Namely, it says that the size of the region visited by the particle to a moment t grows diffusively.

Theorem 2. — *In OS-model $Ez^2_{\max}(t)$, $Ez^2_{\min}(t)$ and $Ez^2(t)$ grow linearly in t.*

In the NOS-model the scatterers are invariant with respect to reflections. This is the reason why the dynamics of this model is quite different from the one of the OS-model.

Theorem 3. — *In the NOS the particle visits almost surely all sites of the lattice $\mathbb{Z}$ infinitely many times if the rigidity r is an even number. Besides for almost every $\omega \in \Omega$ there exist sequences of moments of time τ_i, $i = 0, 1, 2, \dots$, and of closed intervals $B_i(\omega)$, $i = 1, 2, \dots$, with properties analogous to the ones in* Theorem 2. *If the rigidity r is an odd number then for all $\omega \in \Omega$, the particle visits all sites in $[0, \infty)$, $[-1, \infty)$, or $(-\infty, 1]$ and only these sites. Besides the particle visits each of these sites no more than $3r$ times. Moreover in this case there exist sequences of moments of time $\widehat{\tau}_i(\omega)$, $i = 0, 1, 2, \dots$, and of closed intervals $\widehat{B}_i(\omega) = [\widehat{a}_i(\omega), \widehat{b}_i(\omega)]$, $i = 1, 2, \dots$ such that $\widehat{\tau}_0(\omega) = 0$, $\widehat{\tau}_i(\omega) < \widehat{\tau}_{i+1}(\omega)$, $\widehat{a}_i(\omega) < \widehat{a}_{i+1}(\omega) < \widehat{b}_i(\omega)$, $\widehat{a}_i(\omega) \to \infty$ as $i \to \infty$ or $\widehat{b}_i(\omega) \to -\infty$ as $i \to \infty$ and the particle stays inside $\widehat{B}_i(\omega)$, within the time interval $[\widehat{\tau}_i(\omega), \widehat{\tau}_{i+1}(\omega)]$.*

We recall that the particle always starts at the origin with positive velocity. This explains why in Theorem 3 the semi-interval $(-\infty, 0)$ does not show up.

By comparison of Theorems 1 and 3 one can immediately see that a parity of rigidity does not play any role in the OS-model while in the NOS-model it completely defines its qualitative behavior.

Remark. — Observe that in case of odd rigidity Theorem 3 refers to the behavior of all (rather than of almost all) orbits.

Corollary 2. — *Let in the NOS-model the rigidity be even. Then the exceptional set of measure zero orbits in* Theorem 3 *corresponds to the configuration of scatterers with a positive tail of FS or/and with a negative tail of FS.*

The next statement immediately follows from Theorem 3.

Corollary 3. — *Let the rigidity r be an odd number. Then in the NOS-model the particle will for all $\omega \in \Omega$ propagate in one direction with a random velocity.*

Indeed the particle at any moment of time is confined to some segment (box) $B_i(\omega)$ where it goes back and forth. These boxes move in one direction and the particle eventually propagates with them. At each first visit to any site of $\mathbb{Z}$ the particle can be scattered backward or forward according to a random initial distribution of scatterers. Therefore the particle propagates with a random speed.

The next theorem gives the quantitative description of the dynamics of NOS-model.

Theorem 4. — *In NOS-model $Ez^2(t)$ grows as const t^2 if r is an odd number. Otherwise, if r is an even number, $Ez^2(t)$ grows as const* $\log t$.

Because of the deterministic evolution of WRE it is possible to give much more detailed description of the motion of the particle within random boxes in each of the models under study. On the other hand transition from one box where the particle gets confined for some time to the next such box is a random event.

We describe now a geometric nature of typical orbits in the OS- and NOS-models. At first we introduce some notions and notations.

We will denote by Ω_1 and Ω_2 the phase spaces of the OS-model and of the NOS-model respectively. It is convenient to introduce the reduced phase spaces $\widehat{\Omega}_1 = \{LS, RS\}^{\mathbb{Z}}$ and $\widehat{\Omega}_2 = \{BS, FS\}^{\mathbb{Z}}$. Thus $\widehat{\Omega}_1$ and $\widehat{\Omega}_2$ refer just to a type of scatterer at any site of $\mathbb{Z}$, without taking into account how many times the particle has already been reflected by this scatterer. Dynamics of the OS-model (NOS-model) we will define by $f_1 : \Omega_1 \to \Omega_1$ ($f_2 : \Omega_2 \to \Omega_2$). Let $\pi_1' : \Omega_1 \to \widehat{\Omega}_1$, $\pi_2' : \Omega_2 \to \widehat{\Omega}_2$, $\pi_1'' : \Omega_1 \to \{-1, 1\}$, $\pi_2'' : \Omega_2 \to \{-1, 1\}$ are the natural projections.

For $x, y \in \widehat{\Omega}_1$ ($x, y \in \widehat{\Omega}_2$) we define the distance $d(x, y)$ as $d(x, y) = 2^{-n}$ if $x_i = y_i$ for $|i| < n$ and $x_i \neq y_i$ for $i = n$ or $i = -n$, i.e., if configurations of scatterers restricted to $(-n, n)$ coincide for x and y.

Lemma 5. — *In the OS-model for almost every point $\omega \in \Omega_1$ there exists an infinite sequence of moments of time $\tau_k = \tau_k(\omega)$, $k = 1, 2, \ldots$, $\tau_k \to \infty$ as $k \to \infty$, such that*

(i) $\pi_1''(f_1^{\tau_k}(\omega)) = 1$

(ii) $(\pi_1'(f_1^{\tau_k}(\omega)))_i = (RS, 0)$ *if* $0 \leqslant i \leqslant k$

(iii) $(\pi_1'(f_1^{\tau_k}(\omega)))_i = (LS, 0)$ *if* $-k \leqslant i < 0$.

In other words, Lemma 5 states that a typical orbit of the OS-model returns into the smaller and smaller neighborhoods of the orbit, for which at $t = 0$ at all positive sites of $\mathbb{Z}$ were right scatterers with zero indices while at all nonpositive sites of the lattice were left scatterers with zero indices.

Lemma 6. — *Let the rigidity r be an even number. Then in the NOS-model for almost every point $\omega \in \Omega_2$ there exists an infinite sequence of moments of time $\tau_k = \tau_k(\omega)$, $k = 1, 2, \ldots, \tau_k < \tau_{k+1}$, $\tau_k \to \infty$ as $k \to \infty$, such that*

(i) $\pi_2''(f_2^{\tau_k}(\omega)) = 1$

(ii) $\pi_1''(f_2^{\tau_k}(\omega))_i = (BS, 0)$ *for* $|i| < k$.

Lemma 7. — *Consider the NOS-model. Let the rigidity r be an odd number. Then for any point $\omega \in \Omega_2$ and any $z \in \mathbb{Z}$ there exist a moment of time $\tau = \tau(\omega, z)$ such that:*

(i) *The type of scatterer at z never changes after the moment $\tau(\omega, z)$.*

(ii) *The type of scatterer located at z after the moment $\tau(\omega, z)$ is BS if at the next site of $\mathbb{Z}$ in the direction of propagation, it was initially FS, or FS if the next site of $\mathbb{Z}$ in the direction of propagation was initially occupied by BS.*

Lemma 7 states that the initial configuration of scatterers gets flipped (each FS becomes BS and vice versa) and shifted on one site in the direction opposite to the eventual direction of the particle's propagation.

3. Proofs

We have already mentioned that a WRE in a fixed environment ($r = \infty$) $\mathbb{Z}$ is always trivial, i.e., the particle will be moving forever back and forth in a segment between two closest BS (in NOS-model) or LS with a positive coordinate and RS with a negative coordinate (in OS-model).

Denote by $\eta(z,t)$ a number of visits of the particle to a site $z \in \mathbb{Z}$, which occurred between the last moment of time $\tau = \tau(z,t)$, $0 < \tau(z,t) < t$, when a scatterer at z flipped and t. It is easy to see that $\eta(z,t)$ equals the index of the scatterer at the site z at time t. Thus, to make the scatterer at the site z flip requires another $r - \eta(z,t)$ visit of the particle to this site.

Proof of Theorem 1. — Recall that we always assume that the particle starts at the origin $z = 0$ with velocity $v = 1$. Therefore the initial segment of any orbit is the motion of the particle with the unit speed until it will get to the closest to the origin site $z = b_1 > 0$ with LS. At such site z (at the moment $t = b_1$) the particle will turn, i.e., its velocity becomes $v = -1$.

Now (if $r > 1$) for some time the particle will be confined between the sites $z = b_1 - 1$ and $z = b_1$. Indeed, both indices $\eta(b_1 - 1, b_1)$ and $\eta(b_1, b_1)$ equal 1. Then (if $r > 1$) there is RS at the site $b_1 - 1$ and at the moment $t = b_1 + 1$ the particle gets reflected back and hits again LS at b_1 at the moment $b_1 + 2$ and so on.

It is easy to see that those oscillations between $z = b_1 - 1$ and $z = b_1$ will be over at the moment $t = b_1 + 2r - 1$. In fact at the moment $b_1 + 2r$ the particle will be in the site $z = b_1 - 2$ and $\eta(b_1, (b_1 + 2r - 1)) = \eta(b_1 - 1, b_1 + 2r - 1) = 0$, i.e., the scatterers at the sites $b_1 - 1$ and b_1 changed their type. (If $r = 1$ then the particle makes no oscillations in its way between $z = b_1 - 1$ and $z = b_1$. Instead the particle after the reflection at $z = b_1$ goes back and passes at the next step the site $z = b_1 - 1$.)

Therefore at the moment $t = b_1 + (2r - 1)(b_1 - 1) + 1$ the particle will return back to the origin $z = 0$. If there was at $t = 0$ RS at the origin then at the moment $t = 2rb_1$ it will start its travel from $z = 0$ into the positive semiaxis. In this case we set $a_1 = 0$. Otherwise, the particle starts to move from the origin into the negative semiaxis. The same consideration can obviously apply to this piece of trajectory, where one only needs to change LS into RS and vice versa and to change $v = 1$ into $v = -1$. Denote the site where the particle meets its first RS by $z = a_1 < 0$.

Observe now that at the moment of time when the particle goes through the origin with the velocity $v = -1$ (i.e., at this moment there is LS at the origin) all scatterers at the sites $z = 1, 2, \ldots, b_1$ are RS. Analogously, when the particle will cross the origin next time with the velocity $v = 1$ at all sites $z = 0, -1, -2, \dots, a_1 + 1, a_1$ there will be LS.

Therefore, at the moment of time $t = 2r(b_1 - a_1)$ the particle will be again (as at $t = 0$) at the origin with the velocity $v = 1$. Besides at this moment all sites $z = 1, 2, \dots, b_1$ will be occupied again by RS with the indices equal zero. Therefore, now the particle will travel into the positive semiaxis $\mathbb{Z}_+$ at least $b_1 + 1$ consecutive steps, i.e., it will penetrate into $\mathbb{Z}_+$ farther than at its first excursion to $\mathbb{Z}_+$ when it was backscattered by LS at the site $\mathbb{Z} = b_1$. Denote the closest (at this moment of time) to the origin (positive) site with LS by $z = b_2 > b_1$ and the closest to $z = 0$ negative site by $z = a_2 < a_1$. Then the same arguments as before are applied.

It is easy to see that in the same way we can construct segments $B_i = [a_i, b_i]$, $i = 1, 2, \dots$, with the properties satisfying to Theorem 1. Obviously these intervals as well as the corresponding intervals of time $[\tau_{i-1}, \tau_i]$, $i = 1, 2 \dots$, when the particle is confined within B_i are completely defined by the initial distribution of scatterers ω, i.e., $\tau_i = \tau_i(\omega)$ and $a_i = a_i(\omega)$, $b_i = b_i(\omega)$. □

Proof of Corollary 1. — It follows from the proof of Theorem 1 that the only case when there is no infinite sequence of closed intervals $B_{i+1}(\omega) \supset B_i(\omega)$ occurs when $b_k(\omega) = \infty$ or $a_k(\omega) = -\infty$ for some integer $k > 0$. But it means that the configuration of scatterers ω has a positive tail (where $z_+ = b_k(\omega)$) or it has a negative tail (where $z_- = a_k(\omega)$). □

Proof of Theorem 2. — It follows from the proof of Theorem 1 that for almost every initial configuration of scatterers there exists a sequence $\tau_i(\omega)$ such that within the interval $[\tau_i(\omega), \tau_{i+1}(\omega)]$, $i = 0, 1, 2, \dots$, the particle moves (starting at the origin) inside the interval $B_{i+1}(\omega) = [a_{i+1}(\omega), b_{i+1}(\omega)]$. Besides it follows from the proof of Theorem 1 that the length of the interval $[\tau_i(\omega), \tau_{i+1}(\omega)]$ equals

$$\Delta\tau_{i+1} = 2r(b_{i+1}(\omega) - a_{i+1}(\omega)).$$

Moreover we know exactly how the particle moves in this interval. Indeed, the particle visits within the interval $\Delta\tau_{i+1}$ each site in $B_{i+1}(\omega)$ exactly $2r$ times.

Hence to prove Theorem 2 we need to evaluate expected length of an interval $B_i(\omega) = [a_i(\omega), b_i(\omega)]$. Let us note first that now it would be more convenient to use the probabilistic approach and language. Indeed, $b_1(\omega), b_{i+1}(\omega) - b_i(\omega)$ and $-a_1(\omega), a_i(\omega) - a_{i+1}(\omega)$, $i = 1, 2, \dots$, are sequences of independent identically distributed random variables. These random variables are the ones we need to analyze because the proof of Theorem 1 provided us with the complete description of the deterministic motion of the particle inside the (random) intervals $B_i(\omega)$.

Let q $((1-q))$ be a probability that LS (RS) is located at any given site of the lattice $\mathbb{Z}$. Recall that according to our assumptions the scatterers were placed independently at the different sites. Therefore both $b_{i+1}(\omega) - b_i(\omega)$ and $a_i(\omega) - a_{i+1}(\omega)$ have the geometric probability distribution, i.e., for any $i = 0, 1, 2, \ldots,$

$$\mathrm{Prob}\,\{b_{i+1}(\omega) - b_i(\omega) = k\} = (1-q)^{k-1}q \tag{3}$$

and

$$\mathrm{Prob}\,\{a_i(\omega) - a_{i+1}(\omega) = k\} = q^{k-1}(1-q) \tag{4}$$

where $k \geqslant 1$ is an integer and $a_0(\omega) = b_0(\omega) = 0$.

Denote by $f(t, z)$ probability that the particle will visit a site $z > 0$ at the first time at some moment t. Then one can write the following recurrence equation

$$\begin{aligned} f(z,t) &= (1-q)f(z-1,t-1) \\ &\quad + (1-q)q\sum_{k=1}^{\infty} q^{k-1} f(z-1, t-1-[b(t-1)-a(t-1)+k]2r) \end{aligned} \tag{5}$$

where $b(t)$ and $a(t)$ are the maximal, and the minimal coordinates of sites visited by the particle to the moment t. It is easy to see that $b(t-1)$ in our case equals $z-1$ and therefore we can rewrite (5) as

$$\begin{aligned} f(z,t) &= (1-q)f(z-1,t-1) \\ &\quad + (1-q)q\sum_{k=1}^{\infty} q^{k-1} f(z-1, t-[z-a(t-1)+k]2r) \end{aligned} \tag{6}$$

Similar equations can be written for $z \leqslant 0$. Certainly $b(t) = b(t, \omega)$ and $a(t) = a(t, \omega)$, i.e., both these quantities depend upon the initial configuration of scatterers ω.

Let $m_+(t, \omega)$ be a number of LS located between the origin and $b(t-1)$ in a configuration ω. Then there are two possibilities. Either a number $m_-(t, \omega)$ of RS located between 0 (including the origin itself) and $a(t-1)$ equals $m_+(t, \omega)$ or $m_+(t, \omega) - m_-(t, \omega) = 1$. Because we assumed that $z > 0$ the second possibility holds.

It follows from Theorem 1 that at each moment of time τ the particle almost surely is confined in some segment $B(\tau, \omega) = [a(\tau, \omega), b(\tau, \omega)]$.

We will now compute the expected values of $a(\tau)$ and $b(\tau)$. It is enough to do it for $b(\tau)$ because the procedure of computing $a(\tau)$ is completely similar. One can write

$$b(\tau, \omega) = b_1(\omega) + (b_2(\omega) - b_1(\omega)) + \cdots + (b_{m_+(\tau,\omega)}(\omega) - b_{m_+(\tau,\omega)-1}(\omega)). \tag{7}$$

The probability distributions of the terms in this sum are given by (3). Therefore we just need to find the expected value of $m_+(\tau, \omega)$.

It follows from Theorem 1 and Corollary 1 that for almost every orbit ω of OS-model there exist infinite sequences of moments of time $\tau_i^+(\omega)$ $(\tau_i^-(\omega))$, $i = 1, 2, \ldots,$ such that at the moment τ_k^+ (τ_k^-) the orbit visits at the first time the right (left) end $b_k(\omega)$ $(a_k(\omega))$ of the interval $B_k(\omega)$.

We restrict the consideration to the set of orbits $\Omega_1' \subset \Omega_1$ of measure one described in Corollary 1. Then for any configuration $\omega \in \Omega_1'$ and for any moment of time $\tau > 0$ one can write the following identity

$$\tau = 2r \sum_{i=1}^{m_+(\tau,\omega)-1} (b_i(\omega) - a_i(\omega)) + \gamma(\tau,\omega), \tag{8}$$

where $\gamma(\tau,\omega)$ is the length of the interval of time between the moment when the particle returned to the origin with $v = 1$ after visiting $2r$ times all sites in the interval $B_{m_+(\tau,\omega)-1}(\omega)$ and the moment τ.

Indeed, it follows from the proof of Theorem 1 that any orbit $\omega \in \Omega_1'$ has the following structure. First, it visits $2r$ times all sites in the interval $B_1(\omega)$ and occurs at the origin after that with the positive velocity, then it visits $2r$ times all sites in the interval $B_2(\omega)$ and returns to $z = 0$ with $v = 1$ and so on.

Therefore we have

$$E\gamma(\tau,\omega) \leqslant 2rE(b_{m_+(\tau,\omega)} - a_{m_+(\tau,\omega)}) \tag{9}$$

Hence, we need to find $Eb_{m_+(\tau,\omega)}$ and $Ea_{m_+(\tau,\omega)}$. By making use of (3), (4) it is easy to compute

$$\begin{aligned} Eb_{m_+(\tau,\omega)} &= \frac{1}{q} Em_+(\tau,\omega) \\ Ea_{m_+(\tau,\omega)} &= \frac{1}{1-q} Em_+(\tau,\omega). \end{aligned} \tag{10}$$

Indeed, (3) and (4) imply that

$$\begin{aligned} Eb_1(\omega) = E(b_{i+1}(\omega) - b_i(\omega)) &= \frac{1}{q}, \\ E(-a_1(\omega)) = E(a_i(\omega) - a_{i+1}(\omega)) &= \frac{1}{1-q}, \end{aligned} \tag{11}$$

where $i = 1, 2, \ldots$.

Recall now, that $b_1(\omega)$, $(b_{i+1}(\omega) - b_i(\omega))$, and $-a_1(\omega)$, $(a_i(\omega) - a_{i+1}(\omega))$, $i = 1, 2, \ldots$, are two sequences of independent identically distributed random variables.

It follows from (8) that

$$\begin{aligned} 2r &\sum_{i=1}^{m_+(\tau,\omega)-1} (m_+(\tau,\omega) - i)\left[(b_i(\omega) - b_{i-1}(\omega)) + (a_{i-1}(\omega) - a_i(\omega))\right] \leqslant \tau \\ &\leqslant 2r \sum_{i=1}^{m_+(\tau,\omega)} (m_+(\tau,\omega) - i + 1)\left[(b_i(\omega) - b_{i-1}(\omega)) + (a_{i-1}(\omega) - a_i(\omega))\right] \end{aligned} \tag{12}$$

where again $a_0(\omega) = b_0(\omega) = 0$.

The relations (10)–(12) imply

$$\left(\frac{1}{q}+\frac{1}{1-q}\right)\frac{Em_+(\tau,\omega)(Em_+(\tau,\omega)-1)}{2}\leqslant\tau \leqslant \frac{Em_+(\tau,\omega)(Em_+(\tau,\omega)+1)}{2}\left(\frac{1}{q}+\frac{1}{1-q}\right). \tag{13}$$

Therefore there exist such positive constants C_1 and C_2 that

$$C_1 t^{1/2}\leqslant Em_+(t,\omega)\leqslant C_2 t^{1/2} \tag{14}$$

for sufficiently large t.

It follows from (14) that there exist positive constants C_1', C_2', C_1'', C_2'' such that for sufficiently large t one has

$$\begin{aligned} C_1' t &\leqslant Ez^2_{\max}(t)\leqslant C_2' t,\\ C_1'' t &\leqslant Ez^2_{\min}(t)\leqslant C_2'' t. \end{aligned} \tag{15}$$

It remains to prove that $Ez^2(t)$ has the same asymptotics. This fact immediately follows from (15) and Theorem 1. Indeed, it has been shown in the proof of Theorem 1 that within the interval of time $\tau_i(\omega)\leqslant t\leqslant \tau_{i+1}(\omega)$, $i=0,1,2,\dots$ the particle for any $\omega\in\Omega_1'$ spends the same amount of time (equal $2r$) at each site of the interval $B_i(\omega)$.

Therefore, position of the particle is uniformly distributed within $B_{i+1}(\omega)$ in the time interval $[\tau_i(\omega),\tau_{i+1}(\omega)]$, and the last statement of Theorem 2 follows. □

Proof of Theorem 3. — Theorem 3 follows from Theorem 1 and Theorem 2 in [**3**]. Therefore we just outline the proof.

Let us consider the NOS-model and assume first that the rigidity r is an odd number. Then the particle will travel from the origin till the closest to $z=0$ site $\widehat{b}_1=\widehat{b}_1(\omega)>0$ where in the configuration ω there is a back-scatterer BS. At $z=\widehat{b}_1$ the particle will turn back and travel now in the negative direction until it reaches the closest to $z=\widehat{b}_1$ site $z=\widehat{a}_1(\omega)$ with FS. Observe that if the rigidity $r=1$ then $\widehat{a}_1(\omega)=\widehat{b}_1(\omega)-1$, unless $\widehat{b}_1(\omega)=1$ and a FS is located at the origin in the configuration ω. In this case the scatterer at the origin becomes BS after the particle pass $z=0$ in the negative direction. Therefore, it is enough for $r=1$ to consider only such cases when there was a BS at the origin at $t=0$.

We return now to the general case of an odd rigidity. According to the dynamics the particle will move back and forth in the segment $\widehat{B}_1(\omega)=[\widehat{a}_1(\omega),\widehat{b}_1(\omega)]$ until it hits the BS located at the site $z=b_1$ at the $\left(\frac{r+1}{2}\right)$th time. Denote this moment by $\widehat{\tau}_1(\omega)$. Observe that to this moment of time the particle will visit all internal sites of $B_1(\omega)$ exactly r times. Recall that initially at all these sites were located forward scatterers. Therefore to $t=\widehat{\tau}_1$ all of those got substituted by BS.

It is easy to see that at the moment $t=\widehat{\tau}_1(\omega)$ the BS located at the site $z=\widehat{b}_1(\omega)$ has the index $(r+1)/2$ while the BS at $z=\widehat{b}_1(\omega)$ has the index zero. Therefore the particle will move now for $(r+1)$ moments of time between $z = \widehat{b}_1(\omega)$ and $z = \widehat{b}_1(\omega) - 1$. Finally, at the moment $\tau_1^* = \widehat{\tau}_1(\omega) + (r+1)$ the particle will pass the site $z = \widehat{b}_1(\omega)$ with positive velocity and travel until the closest site $z = \widehat{b}_2(\omega)$ with a backward scatterer. At the moment $t = \tau_1^*$ the BS located at $z = \widehat{b}_1(\omega) - 1$ will have the index equal $(r+1)/2$.

Therefore, after the moment $t = \tau_1^*$ the particle will move back and forth between the sites $z = \widehat{b}_1(\omega) - 1$ and $z = \widehat{b}_2(\omega)$ until the moment $t = \tau_2^*(\omega)$, when it will pass the site $z = \widehat{b}_2(\omega)$ with the positive velocity. We denote $\widehat{a}_2(\omega) = \widehat{b}_1(\omega) - 1$.

In the same manner one can construct intervals $\widehat{B}_i(\omega) = [\widehat{a}_i(\omega), \widehat{b}_i(\omega)]$ and the corresponding sequence of times $\widehat{\tau}_i(\omega)$, $i = 1, 2, \dots$.

Let now the rigidity r is an even number. Then again the first segment of any orbit will travel till the closest to $z = 0$ site $z = b_1(\omega)$ with a backward scatterer. Then the particle will travel from $z = b_1(\omega)$ in the negative direction till the closest site $z = a_1(\omega) \leqslant 0$ with BS. After it reaches $z = a_1(\omega)$ the particle continues to move back and forth within the segment $B_1(\omega) = [a_1(\omega), b_1(\omega)]$.

The crucial difference with the case of odd rigidity is that at all internal sites of $B_1(\omega)$ will appear BS (with index 0) at the moment $\widetilde{\tau}_1(\omega)$, when the particle will return to the origin at the rth time. At $t = \widetilde{\tau}_1(\omega)$ the indices of BSs at $z = a_1(\omega)$ and $z = b_1(\omega)$ equal $r/2$.

Therefore it will take now a very long time for the particle to get out of the segment $B_1(\omega)$. Indeed, all scatterers located in the internal sites of this segment must change their type before that back to FS.

At the moment $\tau_1'(\omega)$ when it happens the particle will start again to move back and forth in $B_1(\omega)$ from its left end $a_1(\omega)$ till its right end $b_1(\omega)$. It is easy to see that at the rth visit of the particle to the origin $z = 0$ in the process of these consecutive trespassing of $B_1(\omega)$ at all internal sites of $B_1(\omega)$ will be BS with the index 0 while at $z = a_1(\omega)$ and $z = b_1(\omega)$ will be FS with the index 0.

Therefore after the next repetition of the same process of turning all BS at the internal sites of $B_1(\omega)$ into FS the particle will get out of $B_1(\omega)$ and will become confined to some interval $B_2(\omega) = [a_2(\omega), b_2(\omega)]$, where the similar process will take place. Here $z = a_2(\omega)$ $(z = b_2(\omega))$ is the closest to $a_1(\omega)$ $(b_1(\omega))$ site with negative (positive) coordinate where there is a BS. In the same way one can construct a sequence of closed segments $B_i(\omega)$, $i = 1, 2, \dots$, with the required properties. The corresponding sequence $\tau_i(\omega)$, $i = 1, 2, \dots$, is naturally defined by the condition that the particle remains confined to $B_i(\omega)$ until the moment $t = \tau_i(\omega) + 1$, when it leaves this segment at the first time. □

Proof of Corollary 2. — Consider the NOS-model with an even rigidity. Then it follows from the proof of Theorem 3 that the particle will visit the origin infinitely many times unless $a_i(\omega) = -\infty$ or/and $b_j(\omega) = \infty$ for some positive integers i, j. □

Proof of Theorem 4. — The case of an odd rigidity r has been considered in [**3**].

Let the rigidity r is an even number. Consider any site $z \in \mathbb{Z}$, $z > 0$. Denote by τ_z the moment of time when the particle visits the site z at the first time in such state that there is a forward scatterer at z. In other words $\tau_z = \tau_z(\omega)$ is the moment of the first visit of the particle to the site z if there was a FS at $t = 0$ at this site, or it is the moment of the first visit of the particle to the site z after a BS at this site has been changed to a FS.

It has been shown in [**3**] that the expectation of the random variable $\tau_z(\omega)$ equals

$$E\tau_z = 2r + 1 + (z-1)[(1-q) + r(1 + 4q + qz)] + q(3^{z-1} - z) \tag{16}$$

The analogous formula holds for $z \geqslant 0$. The statement of Theorem 4 for even rigidity immediately follows from Theorem 3 and (16). $\square$

Lemma 5 is the immediate corollary of Theorem 1.

Lemma 6 is the immediate corollary of Theorem 3.

Lemma 7 follows from the proof of Theorem 3.

4. Concluding remarks

One may get the impression that the phenomena discussed and results obtained in this paper are essentially restricted to the one-dimensional case. It is, certainly, the simplest possible situation, when one studies walks in $\mathbb{Z}$ and, perhaps, it is not feasible to hope that the same type of comprehensive analysis would be possible for deterministic walks on some sufficiently general class of graphs.

However, various regimes of anomalous diffusion were observed in computer experiments with WRE (see e.g., [**8**]). For instance, the phenomenon of propagation in a random environment has been proven to exist [**13**] in the triangular lattice as well. It is worthwhile to mention that this propagation reminds very much the famous gliders in the Conway's Game of Life [**9**]. Observe though that the glider is just a particular solution to this dynamical system, while propagation in WRE takes place for any orbit of a certain deterministic walk in the triangular lattice. Moreover, this propagation occurs with random velocity, while in the Game of Life gliders always move with one and the same velocity. This and other features of WRE are currently explored in the theoretical computer sciences (see e.g., [**11**]).

We believe that the rigorous theory of Walks in Rigid Environments could be developed much farther. Although these dynamical systems demonstrate various features of stochastic (chaotic) behavior, their behavior is quite different from the one which we encounter in familiar classes of chaotic dynamical systems. For instance, these systems are nonexpansive [**5**].

On the other hand WRE provide clearer models than probabilistic models of various types of random walks and they do not require detailed assumptions about probability distributions involved on contrary to the purely probabilistic models (see e.g., [**15, 16, 17**]).

References

[1] R.K. Ahuja, T.L. Magnanti, J.B. Orlin. *Network Flows.* Prentice Hall, NY, 1993.

[2] L.A. Bunimovich. Controlling production lines. In *Handbook of Chaos Control,* H. Schuster, editor, Wiley-VCH, Berlin, 1999.

[3] L.A. Bunimovich. Walks in rigid environments. *Physica A* **279**: 169–179, 2000.

[4] L.A. Bunimovich and S.E. Troubetzkoy. Recurrence properties of Lorentz lattice gas cellular automata. *J. Statist. Phys.,* **67**: 289–302, 1992.

[5] L.A. Bunimovich and S.E. Troubetzkoy. Topological properties of flipping Lorentz lattice gases. *J. Statist. Phys.,* **72**: 297–307, 1993.

[6] L.A. Bunimovich. Billiards and other hyperbolic systems. In *Dynamical Systems, Ergodic Theory and Applications*, Ya.G. Sinai, editor, Springer, Berlin, 2000, pp. 192–233.

[7] B. Chopard. Complexity of cellular automata models. In *Physics of Complexity*, S. Ciliberto, T. Dauxois and M. Dros, eds., Editions Frontiers, Lyon, 1995, pp. 111–136.

[8] E.G.D. Cohen. New types of diffusion in lattice gas cellular automata. In *Microscopic Simulation of Complex Hydrodynamic Phenomena*, M. Marechal and B.L. Holian, eds., Plenum, NY, 1992.

[9] M. Delorme, J. Mazoyer. *Cellular Automata.* Kluwer, Dordrecht, 1999.

[10] *Dynamics of Complex Interacting Systems*, E. Goles and S. Martinez, eds., Kluwer, Dordrecht, 1996.

[11] A. Gajardo, A. Moreira and E. Goles. Complexity of Langton's ant. *Discrete Appl. Math.,* **117**: 41–50, 2002.

[12] General physical systems and the emergence of physical structures from information theory. *Int. J. Gen. Syst.,* **27**: 1–379, 1998.

[13] P. Grosfils, J.-P. Boon, E.G.D. Cohen and L.A. Bunimovich. Organization and propagation in deterministic Lorentz gases. *J. Statist. Phys.,* **97**: 375–401, 1999.

[14] M. Lyubich. The quadratic family as a qualitatively solvable model of chaos. *Notices of AMS,* **47**: 1042–1052, 2000.

[15] R.K. Mauldin, M. Monticino and H. von Weizsäcker. Directionally reinforced random walks. *Adv. in Math.,* **117**: 239–252, 1996.

[16] E.W. Montroll and M.F. Shlesinger. On the wonderful world of random walks. In *Studies in Statistical Mechanics XI*, E.W. Montroll and J.L. Lebowitz, eds. North-Holland Phys. Publ., Amsterdam, 1984.

[17] R. Pemantle. Phase transition in reinforced random walk and RWRE on trees. *Ann. Probab.,* **16**: 1229–1241, 1995.

[18] T.W. Ruijgrok and E.G.D. Cohen, Deterministic lattice gas models. *Phys. Lett. A.,* **133**: 415–418, 1988.

L.A. BUNIMOVICH, Georgia Institute of Technology, Atlanta GA 30332, USA
E-mail : `bunimovh@math.gatech.edu` • *Url :* `http://www.math.gatech.edu/~bunimovh`

Astérisque
286, 2003, p. 249–256

PERVERSE SOLUTIONS OF THE PLANAR n-BODY PROBLEM

by

Alain Chenciner

To Jacob, some questions for his 60th anniversary

Abstract. — The perverse solutions of the n-body problem are the solutions which satisfy the equations of motion for at least two distinct systems of masses. I contribute with some simple remarks concerning their existence, a question which curiously seems to be new.

Let $X(t) = \big(\vec{r}_1(t), \vec{r}_2(t), \ldots, \vec{r}_n(t)\big)$ be a solution of the n-body problem with newtonian potential and masses $m_1, m_2, \ldots, m_n$. We ask the following questions:

Question 1. — *Does there exist another system of masses, $(m'_1, m'_2, \ldots, m'_n)$, for which $X(t)$ is still a solution ?*

Question 2. — *Same as question 1 but insisting that the sum $M = \sum_{i=1}^n m_i$ of the masses and the center of mass $\vec{r}_G = (1/M)\sum_{i=1}^n m_i \vec{r}_i$ do not change.*

Definition. — *If the answer to the first (resp. second) question is yes, we shall say $X(t)$ is a* perverse *(resp.* truly perverse*) solution and the allowed systems of masses will be called* admissible.

Remark. — If the inverse problem raised by Question 1 may seem very natural, Question 2 needs some motivation. The possible existence of *choreographies* whose masses are not all equal is at the origin of the notion of perverse solution. Recall that a planar choreography is a periodic solution $\mathcal{C}(t) = \big(q(t+T/n), \ldots, q(t+(n-1)T/n), q(t+T) = q(t)\big)$ of the n-body problem such that all n bodies follow the same closed plane curve $q(t)$ with equal time spacing ([**S1, S2, CGMS**]). It is noticed in [**C**] that if a choreography exists whose masses are not all equal, it is a truly perverse choreography: by replacing each mass by the mean mass M/n we obtain new admissible masses, while keeping the center of mass and total mass unchanged.

2000 ***Mathematics Subject Classification.*** — 70F10.
Key words and phrases. — n-body problem, homographic solutions.

In the sequel, we shall consider only the planar problem. We shall identify the plane of motion with the complex plane $\mathbb{C}$, hence the positions $\vec{r}_G, \vec{r}_i$, $i=1,\dots,n$, with complex numbers z_G, z_i, $i=1,\dots,n$, and $X(t)$ with an element of $\mathbb{C}^n$. We shall use the following notations (we always assume that $z_i \neq z_j$):

$$\begin{cases} z_{ij} &= z_i - z_j, \quad a_{ij} = \dfrac{z_{ij}}{|z_{ij}|^3} \text{ if } i\neq j,\ a_{ii}=0, \quad m=(m_1,m_2,\dots,m_n),\\ \mathcal{A}_0 &= \big(z_{ij}\big)_{1\leqslant i,j\leqslant n}, \quad \mathcal{A}=\big(a_{ij}\big)_{1\leqslant i,j\leqslant n}\end{cases}$$

We shall identify a matrix as $\mathcal{A}_0$ or $\mathcal{A}$ with a linear map from $\mathbb{C}^n$ to $\mathbb{C}^n$. This will allow it to act on the vector m. The definition of the center of mass may be rewritten

$$\sum_{j=1}^{n} m_j z_{ij} = M(z_i - z_G), \quad M=\sum_{j=1}^{n} m_j,$$

that is

$$\mathcal{A}_0(t)m = M\big(X(t) - z_G(t)(1,\dots,1)\big),$$

and the equations of motion in a galilean frame are

$$\forall\, t, \forall\, i,\ \ddot{z}_i(t) = -\sum_{j\neq i} m_j \frac{z_i - z_j}{|z_i - z_j|^3}, \quad \text{that is} \quad \mathcal{A}(t)m = -\ddot{X}(t).$$

Hence, if another set $m'_1, m'_2, \dots, m'_n$ of masses admits the same solution $X(t)$, the difference

$$\mu = m - m' = (\mu_1, \mu_2, \dots, \mu_n) \in \mathbb{R}^n$$

is a real non-zero vector in the kernel of any of the complex matrices $\mathcal{A}(t)$. If, moreover, M and $z_G(t)$ are the same for the two sets of masses, μ is also in the kernel of any of the matrices $\mathcal{A}_0(t)$. It will be important to remember that $\mathcal{A}_0$ and $\mathcal{A}$ are antisymmetric (${}^t\!\mathcal{A}_0 = -\mathcal{A}_0$, ${}^t\!\mathcal{A} = -\mathcal{A}$). This will cause the parity of n to play a role. We start with the obvious

Proposition 1. — *If $n=2$, no solution is perverse. In other words, any planar solution of the 2-body problem determines the masses.*

Proof. — If $n=2$, the matrix $\mathcal{A}(t)$ is of maximal rank whenever it is defined, that is provided $z_{12}(t)\neq 0$. □

As soon as $n\geqslant 3$, perverse solutions do exist, as shown by the following "trivial" examples (thanks to Reinhart Schäfke for proposing immediately the example of an equilateral triangle rotating around a fourth body):

Example 1. — $X(t) = \big(re^{i\omega t}, re^{i\omega t + \frac{2i\pi}{n-1}}, \dots, re^{i\omega t+(n-2)\frac{2i\pi}{n-1}}, 0\big)$ is a relative equilibrium solution with n masses $(m_1, m_1, \dots, m_1, m_0)$ if and only if the following "Kepler-like" condition is satisfied:

$$r^3\omega^2 = \frac{U_n}{I_n} = m_0 + \frac{m_1}{n-1}\sum_{1\leqslant j<k\leqslant n-1} \frac{1}{|z_{jk}|}.$$

In the above formula,

$$U_n = m_1 m_0(n-1) + m_1^2 \sum_{1\leqslant j<k\leqslant n-1} \frac{1}{|z_{jk}|} \quad \text{and} \quad I_n = m_1(n-1)$$

stand respectively for the potential and the moment of inertia with respect to the center of mass, of the configuration normalized by $|z_{in}| = 1$ if $1 \leqslant i \leqslant n-1$. This leaves a one parameter family of admissible sets of masses. Moreover, for the regular $(n-1)$-gon inscribed in the unit circle, we have

$$\sum_{1\leqslant j<k\leqslant n-1} \frac{1}{|z_{jk}|} = \frac{n-1}{2}\Big(\frac{1}{2\sin\frac{\pi}{n-1}}+\frac{1}{2\sin\frac{2\pi}{n-1}}+\cdots+\frac{1}{2\sin\frac{(n-2)\pi}{n-1}}\Big) = (n-1)^2(\delta_{n-1}+1),$$

where we have set

$$\delta_n = -1 + \frac{1}{4n}\sum_{l=1}^{n-1}\frac{1}{\sin\frac{\pi l}{n}}.$$

Hence,

$$r^3\omega^2 = m_0 + (n-1)m_1(\delta_{n-1}+1) = M + (n-1)m_1\delta_{n-1}.$$

Provided δ_{n-1} is different from 0, the right hand side of the above formula is a linear form in the masses which is linearly independent of the total mass $M = m_0 + m_1(n-1)$.

But δ_{n-1} is strictly negative if $n-1 \leqslant 472$ and strictly positive if $n-1 \geqslant 473$ (see [**MS**]; the first occurence of the magic number 472 seems to be in [**M**]). It follows that M may be chosen as a natural parameter of the set of admissible masses. In particular, these examples are *perverse* but not *truly perverse*.

Remark. — For non-newtonian potentials of the form $1/r^{2\beta}$, $\beta \neq 1/2$, the analogue of δ_n becomes

$$\delta_n = -1 + \frac{1}{2^{2\beta+1}n}\sum_{l=1}^{n-1}\frac{1}{(\sin\frac{\pi l}{n})^{2\beta}},$$

and may become zero for some value of β (see [**BCS**]).

Example 2. — Similar to Example 1 are the relative equilibrium solutions whose configuration is made of one central mass m_0 and k regular homothetic n-gons, the masses in the j-th polygon being all equal to m_j, for $j = 1, \dots, k$. In this case, the equations insuring relative equilibrium motion may be put in the form (see [**BE**] or [**BCS**]):

$$\rho_j^3\omega^2 = m_0 + \sum_{s=1}^{k} m_s H_n(\rho_s/\rho_j), \quad j = 1, \dots, k,$$

where ρ_j is the radius of the $j-$th polygon and

$$H_n(x) = \sum_{l=1}^{n^*(x)} \frac{1 - x\cos\frac{2\pi l}{n}}{(1+x^2-2x\cos\frac{2\pi l}{n})^{3/2}}, \quad n^*(x) = \begin{cases} n & \text{if } x \neq 1, \\ n-1 & \text{if } x = 1. \end{cases}$$

In the "generic" case, such solutions will be perverse and not truly perverse. But, as soon as $k \geqslant 3$, one gets truly perverse solutions for special choices of the radii ρ_j and the integer n (see the last section).

When $n = 3$, the situation is still easy to deal with, thanks to Albouy and Moeckel [**AM**].

Proposition 2. — *The perverse solutions of the planar* 3*-body problem are exactly the collinear homographic solutions. The center of mass is the same for all admissible sets of masses, but not the total mass, which is a natural parameter for such sets. In particular, truly perverse solutions do not exist.*

Proof. — If $n = 3$, the matrix $\mathcal{A}(t)$ is of rank 2 as soon as the configuration is not a triple collision. The existence of a fixed non-zero real vector μ in the kernel of $\mathcal{A}(t)$ implies immediately that the three bodies stay collinear, with a fixed configuration up to similarity. This implies that the motion is homographic. Moreover, the center of mass is dynamically defined as the unique common focus of the similar conics described by the bodies in a galilean frame where the center of mass corresponding to one admissible choice of masses is fixed.

Conversely, each collinear homographic solution of the 3-body problem is perverse: this is a direct consequence of Theorem 2 and Proposition 4 of [**AM**] which, together, say that the set of masses for which a given configuration of three bodies is central is of dimension 2 and may be parametrized by the "multiplier" λ (which is determined by the equation $\ddot{X} = -\lambda X$ as soon as the homographic solution X is given) and the total mass M. To finish the proof, it remains to recall that the center of mass of such a 3-body configuration does not depend on the choice of masses for which it is central (see [**AM**] where this observation is attributed to C. Marchal). □

The case $n = 4$. — The determinant of the antisymmetric 4×4 matrix $\mathcal{A}$ is equal to the square of the Pfaffian (if we extend the notation K_4 of [**AM**] to the complex domain, $P = K_4/2$)

$$P(z_1, z_2, z_3, z_4) = a_{12}a_{34} - a_{13}a_{24} + a_{14}a_{23}.$$

Hence, if a solution of the 4-body problem admits two different sets of masses, its configuration must satisfy $P\big(z_1(t), z_2(t), z_3(t), z_4(t)\big) = 0$ at each instant t. As in [**AM**], but in the complex setting, let us use the following notations :

$$A = z_{12}z_{34}, \; B = z_{13}z_{24}, \; C = z_{14}z_{23}.$$

The above condition becomes

$$P = \frac{A}{|A|^3} - \frac{B}{|B|^3} + \frac{C}{|C|^3} \equiv 0.$$

On the other hand, as $\mathcal{A}_0$ represents the bivector $(1,1,1,1)\wedge(z_1,z_2,z_3,z_4)$, it is of rank 2, that is

$$A-B+C\equiv 0.$$

Together, the two identities above imply that A,B,C cannot be $\mathbb{R}$-dependent, i.e. that the three vectors in $\mathbb{R}^2$ represented by the complex numbers A,B,C can never be collinear; indeed, if m_1,m_2,m_3,m_4 lie in this order on a line, A,B,C are real and positive; then $B=A+C$ and $B^{-2}=A^{-2}+C^{-2}$, which is impossible. But then $A-B$ and $A/|A|^3-B/|B|^3$, being respectively equal to $-C$ and $-C/|C|^3$, must be collinear and this can happen only if $|A|=|B|$, which implies immediately that $|A|=|B|=|C|$ (this remark has already been used in [**V**] and [**AM**]). We have proved the

Lemma 1. — *For any perverse solution of the planar 4-body problem, the configuration is such that at any time*

$$(*)\qquad |z_{12}||z_{34}|=|z_{13}||z_{24}|=|z_{14}||z_{23}|.$$

Configurations which satisfy $(*)$ do exist — for example, an equilateral triangle with the fourth mass at the center, a rhombus with small angle $\pi/6$, an isosceles triangle with two angles equal to $\pi/6$ and fourth mass at the middle point of the base — but, as we have just seen, they cannot be collinear.

Definition. — *A 4-body configuration is called* stricly convex *(resp.* strictly non-convex*) if none of the bodies (resp. if one of the bodies) belongs to the interior of the convex hull of the three others.*

A planar 4-body configuration is either strictly convex, or stricly non-convex, or partially collinear (i.e. such that at least three bodies are collinear).

If, for a given t, the configuration $\big(z_1(t),z_2(t),z_3(t),z_4(t)\big)$ is strictly convex (resp. strictly non-convex) and the real vector $(\mu_1,\mu_2,\mu_3,\mu_4)$ belongs to the kernel of $\mathcal{A}(t)$, each μ_i is different from zero. This is because if, for example, $\mu_1=0,\mu_3\neq 0,\mu_4\neq 0$, the bodies 2,3,4 are such that $\mu_3a_{23}(t)+\mu_4a_{24}(t)=0$ and hence collinear. And if only one of the μ_j is different from zero, say μ_4, then all a_{i4} must be zero, which means total collision. Moreover, strict convexity is equivalent to three μ_i being of the same sign and strict non-convexity to only two μ_i being of the same sign. For example, 1 lies in the interior of the triangle defined by 2,3,4 if and only if μ_2,μ_3 and μ_4 are of the same sign. As the μ_i are independent of t, the nature (strictly convex, strictly non-convex, or partially collinear) of the configuration of a perverse solution does not change along the motion. The possibility of collinearities is exluded by the following lemma.

Lemma 2. — *In a perverse solution of the planar 4-body problem, three of the bodies can never become collinear. In other words, either the configuration stays strictly convex for all t, or it stays strictly non-convex for all t.*

Proof. — Let us suppose now that, for example, 2,3,4 are collinear at some instant t. Then $\mu_1 = 0$, otherwise one would deduce from the equation $a_{21}\mu_1 + a_{23}\mu_3 + a_{24}\mu_4 = 0$ that all four bodies are collinear at this instant and we have already excluded this possibility. This implies that (μ_2, μ_3, μ_4) belong, for any t to the kernel of the antisymmetric matrix $(a_{ij}(t))_{2\leqslant i,j\leqslant 4}$, which means that it is proportional to $(a_{34}(t), a_{42}(t), a_{23}(t))$. As in the proof for the case $n = 3$, one concludes that the configuration of the three last bodies remains similar to a given collinear configuration. This in turn implies that the whole configuration remains self-similar: indeed, the relations

$$\frac{|z_{13}|}{|z_{12}|} = \frac{|z_{43}|}{|z_{42}|}, \quad \frac{|z_{13}|}{|z_{14}|} = \frac{|z_{23}|}{|z_{24}|},$$

say that the the fourth body lies at the intersection of two circles centered on the line which contains the three first.

Finally, the solution should be homographic, but this is impossible because it follows immediately from Dziobek's equations in terms of triangle areas [**D**] that a configuration of four bodies with three bodies collinear is never a central configuration (thanks to Alain Albouy for reminding me of this fact). This proves the lemma. □

There exists at least one perverse — but not truly perverse — solution with non-convex configuration: it is our "trivial" example of three equal masses in an equilateral triangle uniformly rotating around the fourth, located at their center of mass. This is the sole homographic perverse solution because in [**MB**], McMillan and Bartky prove that this is the only configuration which is central for more than one set of (non-homothetic) masses. No other example, in particular no convex example, is known.

Question. — *Is the MacMillan and Bartky example the only perverse solution of the planar* 4*-body problem? In other words, do non-homographic perverse solutions of the planar* 4*-body problem exist?*

As a consolation for this disappointing situation, we now prove the

Proposition 3. — *If $n \leqslant 4$, the planar n-body problem does not possess any truly perverse solution.*

Proof. — If $(m_1, \ldots, m_n)$ and $(m'_1, \ldots, m'_n)$ are admissible masses for a truly perverse solution $X(t)$, their differences $\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$ belong, at any time, to the kernel of both matrices $\mathcal{A}_0(t)$ and $\mathcal{A}(t)$, that is

$$\mu_2 z_{12} + \mu_3 z_{13} + \mu_4 z_{14} = 0, \quad \mu_2 \frac{z_{12}}{|z_{12}|^3} + \mu_3 \frac{z_{13}}{|z_{13}|^3} + \mu_4 \frac{z_{14}}{|z_{14}|^3} = 0, \quad \text{etc.}$$

As none of the real numbers $\mu_1, \mu_2, \mu_3, \mu_4$ is equal to zero (because three bodies are never collinear) this implies, in the same way as above, that

$$\begin{cases} |z_{12}| = |z_{13}| = |z_{14}|, \\ |z_{21}| = |z_{23}| = |z_{24}|, \\ |z_{31}| = |z_{32}| = |z_{34}|, \\ |z_{41}| = |z_{42}| = |z_{43}|. \end{cases}$$

Hence, the configuration should be a regular tetrahedron. As it is planar, this is impossible. □

What about 5 bodies?— The homographic perverse solutions include on the one hand all the collinear ones (same reasoning as in the case of three bodies, using [**AM**]), on the other hand the "trivial" example of four equal masses on a square uniformly rotating around the fifth one located at the center of mass. None of these is truly perverse.

Only in the case of choreographies — whose definition was recalled at the beginning of the paper — are we able to say more.

Proposition 4 **(see [C]).** — *For $n \leqslant 5$, the planar n-body problem does not possess any perverse choreography.*

This is done by interverting the roles of the z_{ij} (resp. the a_{ij}) and the masses, that is replacing the equations $\mathcal{A}_0 m = 0$ (resp. $\mathcal{A} m = 0$) by equations which involve the circulant $n \times n$ matrix defined by the n masses. One then uses the spectral structure of such matrices.

More bodies: truly perverse solutions of the planar n-body problem do exist. — It is shown in [**BCS**] that relative equilibria of a central mass and at least three homothetic regular n-gons, with equal masses on each of them, may be truly perverse if n is well chosen. The simplest such example seems to be 3 regular 456-gons, that is 1369 bodies. Finally, we ask the

Question. — *Do non-homographic perverse solutions of the planar n-body problem exist?*

This is probably a difficult question, as are all the questions where one is asked to understand the structure of the solutions of the n-body problem whose configuration remains all the time in a given subset of the configuration space. A famous example of such a question is the Saari conjecture which states that a solution with constant moment of inertia with respect to the center of mass should be rigid (and hence a relative equilibrium by [**AC**]). The only available method seems to be taking enough time derivatives of the constraints and hoping for some new exploitable constraints to emerge.

Acknowledgements. — The author thanks Alain Albouy for the references concerning central configurations and Andrea Venturelli for an early check via Maple of the sign of δ_n. He is also grateful to the three referees for their extremely careful scrutiny of the paper. Their suggestions have led to significative improvements in the exposition.

References

[AC] Albouy A. and Chenciner A., *le problème des n corps et les distances mutuelles*, Invent. Math. 131, p. 151–184 (1998).

[AM] Albouy A. and Moeckel R., *The inverse problem for collinear central configurations*, Celestial Mechanics and Dynamical Astronomy 77, p. 77–91 (2000).

[BCS] Bang D., Chenciner A. and Simó C., *Truly perverse relative equilibria of the planar n-body problem*, in preparation

[BE] Bang D. and Elmabsout B., *Configurations polygonales en équilibre relatif*, C.R. Acad. Sci. Paris, t. 329, Série IIb, p. 243–248 (2001).

[C] Chenciner A., *Are there perverse choreographies?*, to appear in the Proceedings of the HAMSYS conference held in Guanajuato in march 2001.

[CGMS] Chenciner A., Gerver J., Montgomery R. and Simó C., *Simple choreographies of N bodies: a preliminary study*,in *Geometry, Mechanics and Dynamics, Ed. P.Newton, P. Holmes & A. Weinstein*, p. 287–308, Springer (2002).

[D] Dziobek O., *Ueber einen merkwürdigen Fall des Vielkörperproblems*, Astron. Nacht. 152, p. 33–46 (1900).

[M] Moeckel R., *On central configurations*, Math. Zeit., 205, p. 499–517 (1990).

[MB] MacMillan W.B. and Bartky W., *Permanent configurations in the problem of four bodies*, Transactions of the AMS, vol 34, p. 838–875 (1932) (see page 872).

[MS] Moeckel R. and Simó C., *Bifurcation of Spatial Central Configurations from Planar Ones, SIAM J. on Math. Analysis, 26, p. 978–998 (1995)*

[S1] Simó C., *New families of Solutions in N-Body Problems, Proceedings of the ECM 2000, Barcelona (July, 10–14), Ed. C. Casacuberta, R. M. Miró, J. Verdera and S. Xambó,* Progress in Mathematics *201*, p. 101–115, Birkhäuser-Verlag, Basel

[S2] Simó C., *Periodic orbits of the planar N-body problem with equal masses and all bodies on the same path*, in *The Restless Universe, Ed. B. Steves and A. Maciejewski*, p. 265–284, Institute of Physics Pub., Bristol (2001)

[V] Venturelli A., *Une caractérisation variationnelle des solutions de Lagrange du problème plan des trois corps*, C.R. Acad. Sci. Paris, t. 332, Série I, p. 641–644, (2001).

A. Chenciner, Astronomie et Systèmes Dynamiques, IMCCE, UMR 8028 du CNRS, 77, avenue Denfert-Rochereau, 75014 Paris, France et Département de Mathématiques, Université Paris VII-Denis Diderot, 16, rue Clisson, 75013 Paris, France

Astérisque
286, 2003, p. 257–308

CHAOS VERSUS RENORMALIZATION AT QUADRATIC S-UNIMODAL MISIUREWICZ BIFURCATIONS

by

Eduardo Colli & Vilton Pinheiro

Abstract. — We study C^3 families of unimodal maps of the interval with negative Schwarzian derivative and quadratic critical point, transversally unfolding Misiurewicz bifurcations, and for these families we prove that existence of an absolutely continuous invariant probability measure ("chaos") and existence of a renormalization are prevalent in measure along the parameter. Moreover, the method also shows that existence of a renormalization is dense and chaos occurs with positive measure.

1. Introduction

The *quadratic family*

$$f_a : \begin{array}{ccc}[0,1] & \longrightarrow & [0,1] \\ x & \longmapsto & 4ax(1-x)\end{array}, \quad a \in [0,1],$$

is the simplest model that shows the complexity arising in nonlinear dynamical systems. For a fixed value of the parameter a, supposed to vary along the interval $[0,1]$, one is interested to follow the behavior of iterates x_0, $x_1 = f_a(x_0)$, $x_2 = f_a(x_1)$, $\ldots$, in other words of orbits

$$\mathcal{O}(x_0) = \{f_a^n(x_0)\}_{n\geqslant 0}$$

starting at a point x_0. The set $\omega(x_0)$ of accumulation points of $\mathcal{O}(x_0)$ gives a clue of the asymptotic behavior of the orbit of x_0, and is called the *ω-limit set of* x_0. It turns out ([**7**]) that "typical" starting points $x_0 \in [0,1]$ have equal ω-limit sets. This could be stated as follows: *for each $a \in [0,1]$, there is a set $A = A_a$ such that $\omega(x_0) = A$ for Lebesgue almost every $x_0 \in [0,1]$.* Moreover, there are only three types of sets which A_a could be: (i) a *periodic orbit*, i.e. a set $\{p_0, p_1, \ldots, p_{k-1}\}$ such that $f_a(p_0) = p_1$, $f_a(p_1) = p_2$, $\ldots$, $f_a(p_{k-1}) = p_0$; (ii) a *periodic collection of pairwise disjoint intervals*

2000 ***Mathematics Subject Classification.*** — 37E05, 26A18, 37E20.
Key words and phrases. — Unimodal maps, chaos, absolutely continuous invariant measure, renormalization, generalized renormalization, bifurcation.

$\{I_0, I_1, \ldots, I_{k-1}\}$ where $f_a(I_0) = I_1$, $f_a(I_1) = I_2$, $\ldots$, $f_a(I_k) = I_0$; or (iii) a *Cantor set* (i.e. a perfect and totally disconnected compact set) of zero Lebesgue measure.

The striking alternation of behavior of f_a has been revealed and proved along the last three decades. Among others, we know that: parameters for which the typical ω-limit set is a periodic orbit are dense (and contain intervals, implying also positive Lebesgue measure) ([**3**], [**8**]); parameters for which the typical ω-limit set is a collection of intervals have positive measure (following [**4**]); and parameters for which the typical ω-limit set is a Cantor set have zero Lebesgue measure ([**10**]).

Among parameters with a cycle of intervals as its typical ω-limit set, with total Lebesgue measure ([**9**], [**12**]) we find those for which there is an absolutely continuous (with respect to Lebesgue) f_a-invariant probability measure. In this case f_a is said to be *chaotic*, although more intuitive and not exactly equivalent definitions of "chaos" are available. This definition supplies at least some statistical properties for the mean growth of derivatives along orbits and imply some dynamical structure on the configuration space.

On the other hand, parameters where the typical ω-limit set is a non-hyperbolic periodic orbit are rare in measure. In other words, hyperbolicity is prevalent in measure for these parameters. Putting things altogether, we conclude that for almost all $a \in [0, 1]$, the dynamics of f_a is *either hyperbolic or chaotic.*

A largely used concept in one-dimensional dynamics is the idea of renormalization. We say that f_a is *renormalizable* if there is a collection of pairwise disjoint intervals $\{I_0, I_1, \ldots, I_{k-1}\}$ properly contained in $[0, 1]$ such that (i) the critical point $\frac{1}{2}$ of f_a belongs to, say, I_{k-1}; (ii) $f_a(I_{k-1}) \subset I_0$ and $f_a(\partial I_{k-1}) \subset \partial I_0$; (iii) $f_a : I_i \to I_{i+1}$ is a diffeomorphism for all $i = 0, \ldots, k-2$. In particular, if we call $I = I_{k-1}$, then the function $f_a^k|I$ resembles in many ways the general aspect of a quadratic function in $[0, 1]$, since $f_a^k(I) \subset I$, $f_a^k(\partial I) \subset \partial I$ and $f_a^k|I$ has a single (quadratic) critical point (equal to $\frac{1}{2}$). By an affine rescaling a new function $g : [0, 1] \to [0, 1]$ could be defined, but in general we may not expect g to be quadratic.

Renormalization is a kind of reduction tool. For example, the behavior of typical orbits is completely determined by the restriction $f_a^k|I$, since we know (see [**13**] and references therein) that for Lebesgue almost every $x \in [0, 1]$ there is $n = n(x)$ such that $f_a^n(x) \in I$. All subsequent iterates must remain inside the cycle from this iterate on, because of the invariance properties stated above. This suggests that no complete knowledge of the quadratic family could be achieved without the understanding of a larger class of functions which contains in particular the ones generated *via* renormalization. For this class, it would be desirable some qualitative dynamical similarity with quadratic functions, not only for technical reasons (proves with recursive arguments) but also for the sake of some universality in the conclusions.

In [**3**] and [**8**] (denseness of hyperbolicity), [**9**] joint with [**12**] (measure prevalence of chaos) and [**10**] (rareness of Cantor ω-limit sets), this larger class of functions to which the quadratic functions belong (and which is invariant under renormalization) is composed by all analytic functions f which are holomorphically extendible to a neighborhood U of $[0,1]$ in the complex plane, such that $f(U)$ contains the closure of U and f is a double branched covering between U and $f(U)$. Recently ([**1**]) there have been considered the case of real analytic functions, but even so some main arguments are based on constructions developed in the complex plane.

Among the results mentioned for the quadratic family, the positive measure of chaotic parameters, proved for the first time in [**4**], is the only one which has been stated for C^2 families (see for example [**16**] or [**13**] and references therein). The present work is an attempt to provide techniques restricted to the real setting, weakening smoothness considerably, in order to state results that go in the same direction as the ones of the previous paragraph. Unfortunately the extent of the conclusions cannot be as complete as the ones already proved for the quadratic family. The main reason is that our statements are of a local nature, that is, they are valid only for parameters in small intervals around some bifurcation values. This does not allow us to go beyond the first renormalization, where full families appear.

Here we deal with C^3 *unimodal* interval maps f, that is those with a single turning point c, with the (classical) additional hypothesis that the Schwarzian derivative

$$Sf(x) = \frac{f'''(x)}{f'(x)} - \frac{3}{2}\left(\frac{f''(x)}{f'(x)}\right)^2,$$

defined for all $x \neq c$, is non-positive. These functions will be called *S-unimodal.* From this hypothesis some *a priori* conclusions can be derived. For example, there is at most one periodic attractor and if it does exist then it must attract the critical orbit $\mathcal{O}(c)$ ([**15**]). Moreover, distortion of derivatives for powers of f can be uniformly controlled (see statements in [**13**]). This comes from two facts: first, if a diffeomorphism defined in an interval I has non-positive Schwarzian derivative, the ratio between its derivatives evaluated at two points can be bounded by a constant which depends only on the proportion between their mutual distance and their distance to the boundary of I, *but not on the diffeomorphism.* Second, powers of f have also non-positive Schwarzian derivatives, hence distortion bounds may be obtained whenever $f^n|I$ is a diffeomorphism for some I, independently of n.

To make clear the results we want to state below, it is convenient to relate renormalization with the classification of functions into three types we have made above, which are still valid for the larger class we are considering now (see [**7**]). First, we observe that if f is renormalizable then there is an interval $I^{(1)}$ containing the critical point and a number k_1 such that $f^{k_1}|I^{(1)}$ is a unimodal function. It may be that this function is also renormalizable, and in this case we say that f is (at least) twice renormalizable. We can take the maximum chain of renormalization intervals ordered by (proper) inclusion

$$[0,1] = I^{(0)} \supset I^{(1)} \supset I^{(2)} \dots .$$

If this chain has size $N+1$ then we say that f is N times renormalizable, and if its size is not finite ($N=\infty$) then we say that f is *infinitely renormalizable*. The case where the size is equal to 1 is called *non-renormalizable*.

It turns out that f is infinitely renormalizable if and only if typical points have a Cantor set as its ω-limit set ([**13**]). If f is N times renormalizable, its typical ω-limit set is determined by the N-th renormalization $g = f^{k_N}|I^{(N)}$. If g has an attracting fixed point, the ω-limit set is a periodic orbit, otherwise a collection of intervals. Here we are using the fact that if g had an attracting point of period greater or equal than two then g would be renormalizable, characterizing a contradiction.

We say that f is *Misiurewicz* if the critical point c is not recurrent, i.e. $c \notin \omega(c)$. It may happen that $\omega(c)$ is an attracting periodic orbit. If not, then $f(c)$ belongs to a hyperbolic invariant compact set $\Lambda = \Lambda_f$. From hyperbolic theory, we know that for g sufficiently near f (in the C^1 topology), there is a g-hyperbolic invariant compact set Λ_g such that $f|\Lambda_f$ and $g|\Lambda_g$ are conjugated by $h_g : \Lambda \to \Lambda_g$. The function $g \mapsto \Lambda_g$ is in fact C^1 and is called the *hyperbolic continuation* of Λ. Now we embed f in a C^3 family $(f_a)_a$, where $f_0 = f$, and call w the point belonging to Λ such that $w = f(c)$. As a varies, w has its continuation $w_a = h_{f_a}(w)$ and the critical point c has its continuation c_a, which is well defined by the Implicit Function Theorem, using that c is quadratic. We will say that $(f_a)_a$ is *transversal at* $a=0$ if

$$\frac{d}{da}\left(f_a(c_a) - w_a\right) \neq 0.$$

Without loss of generality, we assume $c_a \equiv c$ and $\dfrac{d}{da}(f_a(c) - w_a) > 0$.

Theorem 1.1. — *Let $f : [0,1] \to [0,1]$ be a C^3 S-unimodal non-renormalizable Misiurewicz function, without periodic attractors. Let $(f_a)_a$ be a C^3 family with $f_0 = f$, transversal at $a = 0$. Then there is $\varepsilon > 0$ such that*

(1) *for almost all $a \in [-\varepsilon, \varepsilon]$, f_a is chaotic or renormalizable;*

(2) *parameters for which f_a is renormalizable constitute a countable union of closed intervals which is dense in $[-\varepsilon, \varepsilon]$;*

(3) *parameters for which f_a is at the same time non-renormalizable and Misiurewicz have zero Lebesgue measure in $[-\varepsilon, \varepsilon]$.*

All items of Theorem 1.1 are new for non-analytic families (the third item is analogous to the statements in [**14**])

As a corollary of the method, we are also able to show that parameters for which f_a is chaotic have positive Lebesgue measure in $[-\varepsilon, \varepsilon]$, assertion which has already been proved, even in more generality, for C^2 families (see [**16**] and [**13**], Chap.V, Section 6; in fact, they prove that the relative measure goes to one at the bifurcation value). The techniques, however, go in a totally different direction, since they work with exclusion of “bad” parameters (which in general include everyone for which there is a renormalization), showing then that the remaining ones have positive measure and reasonably good expansion properties (an absolutely continuous invariant probability measure, for instance). These methods however may exclude also some positive measure set of "good" parameters, for which one could also prove the existence of stochastic dynamics. Here, on the other hand, we show that chaos is prevalent in non-renormalizable dynamics *and* non-renormalizable dynamics occurs with positive measure in the parameter.

Our methods could also be useful to obtain precise estimates of the measure of chaotic parameters and even an upper bound for the Hausdorff dimension of non-renormalizable non-chaotic parameters, provided enough control was achieved in configuration space (see [**5**], for attempts in this direction for C^2 families).

After suitable changes in the conclusion, we could drop the assumption that the bifurcating map is "non-renormalizable" in Theorem 1.1 by writing, instead, that f is finitely renormalizable. In this case, f would be N times renormalizable ($N \geqslant 1$), Misiurewicz and without periodic attractors. Then for the transversal family $(f_a)_a$ we would have two possibilities: (i) f_a is at least N times renormalizable for all $a \in [-\varepsilon, \varepsilon]$, for $\varepsilon > 0$ small; (ii) f_a is at least N times renormalizable for $a \in [-\varepsilon, 0]$ and at least $N-1$ times renormalizable for $a \in (0, \varepsilon]$. The first statement might be rephrased, respectively, into: (i) almost every $a \in [-\varepsilon, \varepsilon]$ is chaotic or $N+1$ times renormalizable; (ii) almost all $a \in [-\varepsilon, 0]$ is chaotic and $N+1$ times renormalizable and almost all $a \in [0, \varepsilon]$ is chaotic or N times renormalizable. The proof would run on in the same way, with minor adaptations.

The proof of Theorem 1.1 uses a result proved in [**2**]. Some "starting conditions" must be satisfied for the functions f_a, $a \in [-\varepsilon, \varepsilon]$, allowing an inductive argument to work. This will be better explained in the next section.

2. Mounting the proof

Let $f : [0,1] \to [0,1]$ be an S-unimodal C^3 function and c its critical point. Assume that f is Misiurewicz, i.e. the critical point c is not recurrent, and f does not have a periodic attractor. The following definitions and Proposition 2.2 can be found in [**11**] (in fact without the Misiurewicz hypothesis).

Let $x \in [0,1]$ and $\tau(x) \neq x$ be such that

$$f(x) = f(\tau(x)),$$

and let $V_x = (x, \tau(x))$.

Definition 2.1. — A point $x \in [0,1]$ is *nice* if $f^n(x) \notin V_x$ for all $n \geqslant 1$. In this case V_x is a *nice interval*.

For example, every periodic orbit contains a nice point, for instance the one maximizing the value of f. Moreover, as f does not have a periodic attractor then there are periodic points arbitrarily near c, assuring arbitrarily small nice intervals.

Let $\mathcal{U}_x \subset [0,1]$ be the set of points that visit V_x at least once (including the points of V_x), and

$$\Lambda_x = [0,1] \setminus \mathcal{U}_x.$$

The following Proposition is proved in [**11**].

Proposition 2.2

(1) *If I is a connected component of $\mathcal{U}_x$ then there is n such that $f^n : I \to V_x$ is monotone and onto. This function is called the* transfer map *from I to V_x.*

(2) *In this case, the intervals of the collection*

$$\{I, f(I), \dots, f^n(I) = V_x\}$$

are pairwise disjoint.

(3) *The set Λ_x is invariant and hyperbolic (hence with zero measure), and if $w \in \Lambda_x$ is such that $f^n(w) \notin \overline{V}_x$, $\forall n \geqslant 1$, then Λ_x accumulates from both sides on w (for short, $w \in \Lambda_x \setminus \partial\Lambda_x$, where $\partial\Lambda_x$ denotes the set of points of Λ_x which belong to the boundary of a connected component of $(0,1) \setminus \Lambda_x$).*

Proof. — See [**11**]. □

As f is Misiurewicz, there is a neighbourhood V of c such that $f^n(c) \notin V$, $\forall n \geqslant 1$. Take a hyperbolic periodic nice point y in V (all periodic points must be hyperbolic under the hypotheses, since $Sf \leqslant 0$ implies that nonhyperbolic periodic points must be attractors). Then $\overline{V}_y \subset V$ and, as $f^n(c) \notin \overline{V}_y$, $\forall n \geqslant 1$, it follows from Proposition 2.2 that $f(c) \in \Lambda_y \setminus \partial\Lambda_y$. In other words, $f(c)$ is accumulated from both sides by arbitrarily small connected components of $\mathcal{U}_y$.

Now we define a new nice point as follows. Take $z \in V_y \cap [0,c)$ such that $f(z) \in \partial I$, for some connected component I of $\mathcal{U}_y$. As $f(c) \in \Lambda_y \setminus \partial\Lambda_y$, z can be chosen arbitrarily near c, so that

$$\frac{|V_z|}{|V_y|}$$

can be as small as desired. With a minor modification in context, the following Proposition is also stated in [**11**].

Proposition 2.3. — *Let I be a connected component of $\mathcal{U}_z$ and, by Proposition 2.2, let n be such that $f^n : I \to V_z$ is monotone and onto. Then there is $\widehat{I} \supset I$ such that $f^n : \widehat{I} \to V_y$ is monotone and onto.*

Proof. — Let T be the maximal interval containing I such that $f^n|T$ is monotone and $f^n(T) \subset V_y$. It is easy to see by Proposition 2.2, item 2, that $I \subset \operatorname{int}(T)$. Supposing by contradiction that $f^n(T) \neq V_y$, there is at least one connected component L of $T \setminus I$ such that $\overline{f^n(L)} \subset V_y$. By the maximality of T, there is $j < n$ such that $c \in \partial f^j(L)$. Again by Proposition 2.2, $f^j(I) \cap V_z = \varnothing$, hence $z \in f^j(L)$ (or $\tau(z) \in f^j(L)$). But $\overline{f^n(L)} \subset V_y$ implies $f^{n-j}(z) \in V_y$, contradiction, since $f(z) \in \Lambda_y$. □

Let $(f_a)_a$ be a C^3 family of S-unimodal functions with $f_0 = f$, transversal at $a = 0$, where a varies in the range $[-\varepsilon, \varepsilon]$, for some $\varepsilon > 0$. As y is a hyperbolic periodic point, it has a continuation y_a defined for small values of a. Also z has a continuation z_a, since it is a preimage of y. Moreover the hyperbolic sets Λ_y and Λ_z have continuations $\Lambda_{y,a}$ and $\Lambda_{z,a}$ and the whole "hyperbolic structure" is preserved. This could be stated as follows: for each sufficiently small a there is a homeomorphism

$$h_a : [0,1] \smallsetminus V_z \longrightarrow [0,1] \smallsetminus V_{z_a}$$

such that

$$f_a^n \circ h_a(x) = h_a \circ f_0^n(x),$$

whenever $\{x, f_0(x), \ldots, f_0^n(x)\} \subset [0,1] \smallsetminus V_z$. In particular, Proposition 2.3 remains valid (if adapted to the continuations) for $a \in [-\varepsilon, \varepsilon]$.

Lemma 2.4. — *Let $f_a^n : I \to V_{z_a}$ be the transfer map of some preimage I of V_{z_a}, and let $f_a^n : \widehat{I} \to V_{y_a}$ be its extension. If $I \cap [f_a(z_a), 1] \neq \varnothing$ then $\widehat{I} \subset [f_a(z_a), 1]$.*

Proof. — Otherwise $f_a(z_a) \in \operatorname{int}(\widehat{I})$ and $f_a^n(f_a(z_a)) \in V_{y_a}$, contradiction, since by the choice of z the orbit of z_a never intersects V_{y_a}. ☐

Now we fix some notation, which the reader can follow with the help of Figure 1 (depicted for $a > 0$). Let $\widetilde{w} > w$ be a point of Λ_y, for $a = 0$, and $\widetilde{w_a}$ its continuation. Since

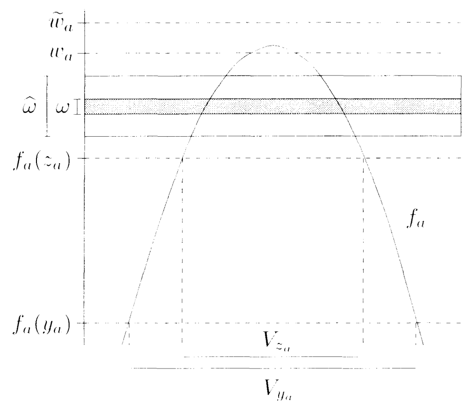


FIGURE 1. Mounting the proof

Λ_y accumulates from both sides in w, we may suppose that $|\widetilde{w}_a - w_a| \ll |w_a - f_a(z_a)|$. By requiring $\varepsilon > 0$ small enough we also beg that $f_a(c) < \widetilde{w}_a$, for all $a \in [-\varepsilon, \varepsilon]$. Let $\mathcal{W} = \mathcal{W}_a$ be the collection of preimages of V_{z_a} intersecting $[f_a(z_a), \widetilde{w}_a]$. For each $\omega = \omega_a \in \mathcal{W}$ let $W : \omega \to V_{z_a}$ be its transfer map, and let $\widehat{\omega} = \widehat{\omega}_a$ be its extension domain relative to V_{y_a}. Although hidden in the notation, we look at W as a function of both parameter and space, defined in the domain

$$\{(a, x)\ ;\ x \in \omega_a, a \in [-\varepsilon, \varepsilon]\}.$$

We will adopt capital letters to indicate two-variable dependence in other situations. For example, we write $F(a, x) = f_a(x)$, so that partial derivatives are denoted by F_a, F_x, F_{xx}, F_{xa}, etc. In this notation, compositions are denoted with respect to the second variable (configuration space), for example $W \circ F$ means the function $(W \circ F)(a, x) = W(a, F(a, x))$. The powers F^k are inductively defined as $F^k(a, x) = F(a, F^{k-1}(a, x))$ and we write F_x^k, F_a^k, etc. for their derivatives. The notation $(F_x)^k$, in turn, means the k-th power of the x-derivative of F. We sometimes treat these functions as functions of one-variable (the x-variable), writing expressions as $F(x)$, meaning $F(a, x)$, or $F|I$, meaning $f_a|I$, where I is an interval, whenever it is clear that the parameter is fixed.

In Section 5 we will show that the transversality of the family $(f_a)_a$ at $a = 0$ implies that the critical value $f_a(c)$ transversally crosses the hyperbolic set Λ_{z_a} not only at w_a, for $a = 0$, but also at nearby points for small parameter values. In particular, if we fix some preimage ω_0 of V_{z_a}, whose continuation is $\omega_{0,a}$, the set of parameters

$$J_0 = \{a \in [-\varepsilon, \varepsilon] : f_a(c) \in \omega_{0,a}\}$$

is an interval, for ε small (see Figure 2, where J_0 occurs for $a < 0$). Moreover, we will

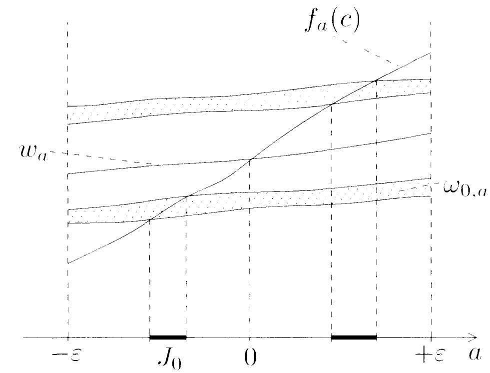


FIGURE 2. Evolution of the critical value

show that the set

$$\Gamma = \{a \in [-\varepsilon, \varepsilon] : f_a(c) \in \Lambda_{z_a}\}$$

has zero Lebesgue measure. Hence all of our assertions will be made for a fixed preimage ω_0 to which the critical value belongs, for parameters in the corresponding interval J_0.

We now focus our attention on the first return map $\Phi = \Phi_a$ of V_{z_a}, for parameters $a \in J_0$, for a fixed $\omega_0 = \omega_{0,a}$ (see Figure 3). The connected component of $\mathrm{dom}(\Phi)$ containing the critical point c is called the *central interval* and will be denoted by $\gamma_0 = \gamma_{0,a}$ (note that $\Phi(\partial\gamma_0) \subset \partial V_{z_a}$). The restriction

$$H \equiv \Phi|\gamma_0 = W_0 \circ F|\gamma_0$$

is called the *central branch*, where $W_0 : \omega_0 \to V_{z_a}$ is the transfer map associated to ω_0. The remaining connected components of $\mathrm{dom}(\Phi)$, together with the central interval, cover V_{z_a} up to measure zero. They form a collection which will be denoted by $\mathcal{P}$, where for each $\pi \in \mathcal{P}$ we have $F(\pi) = \omega$, for some $\omega \in \mathcal{W}$. In other words,

$$P \equiv \Phi|\pi = W \circ F|\pi : \pi \longrightarrow V_{z_a}$$

is a diffeomorphism, where $W : \omega \to V_{z_a}$ is the transfer map associated to ω. Each $\pi \in \mathcal{P}$ is called a *regular interval.*

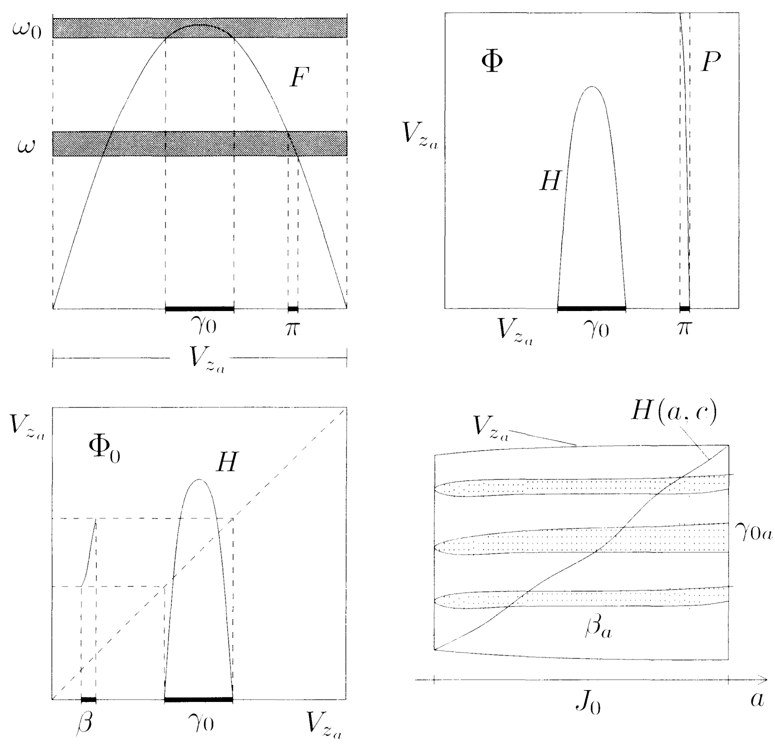


FIGURE 3. Return functions

A further refinement is made, obtaining from Φ a new map Φ_0, defined in V_{z_a} (up to measure zero). This map coincides with Φ in the central interval, and outside it corresponds to the first entry map into γ_0. The domain of Φ_0 is composed by the central image together with a collection $\mathcal{B}$ of intervals called the *preimages of the central interval*. To each preimage $\beta \in \mathcal{B}$ we define the diffeomorphism $B \equiv \Phi_0|\beta : \beta \to \gamma_0$, assigning $\pi_1, \pi_2, \ldots, \pi_n$ such that $\beta \subset \pi_1$, $P_1(\beta) \subset \pi_2, \ldots, (P_m \circ \cdots \circ P_1)(\beta) \subset \pi_{m+1}, \ldots, (P_n \circ \cdots \circ P_1)(\beta) = \gamma_0$, where $P_m : \pi_m \to V_{z_a}$ is the restriction of Φ to π_m, $m = 1, \ldots, n$. We also define

$$\mathcal{U}(\beta) = (P_n \circ \cdots \circ P_1)^{-1}(V_{z_a}),$$

which in particular coincides with π_1 in the case $n = 1$.

Of course all definitions above depend on the parameter a, which is allowed to vary in the interval J_0. Capital letters again are used to denote two-variable functions. The interval $\gamma_0 = \gamma_{0,a}$ is continuously defined for all $a \in J_0$. The same is true for each $\pi \in \mathcal{P}$ and $\beta \in \mathcal{B}$. Figure 3 shows what should be the evolution of the connected components of $\mathrm{dom}\Phi_0$ with respect to the parameter, along the interval J_0. Among others, we will show that $H(a, c)$ transversally crosses these components.

A number of requirements for the map Φ_0, which we call *starting conditions*, must be satisfied, in order to start an induction procedure, developed in [**2**], that proves Theorem 1.1. We separate these requirements into three parts, listed below. We are implicitly assuming non-positive Schwarzian derivative.

Geometry. — There is $\eta > 0$ small such that

$$\frac{|\gamma_{0,a}|}{|V_{z_a}|} < \eta, \qquad \frac{|\beta_a|}{\mathrm{dist}(\beta_a, \gamma_{0,a})} < \eta, \qquad \frac{|\beta_a|}{\mathrm{dist}(\beta_a, \partial V_{z_a})} < \eta,$$

for all $a \in J_0$ and $\beta \in \mathcal{B}$. Moreover, for each $\beta \in \mathcal{B}$, the diffeomorphism $B : \beta \to \gamma_0$ is extendible to a $\eta^{-1}|\beta|$-neighborhood of β, for all $a \in J_0$.

These conditions are uniform in the parameter and have been considered in previous works (see [**6**] and [**7**], for example).

Central branch. — $H_{xx} \neq 0$, $H_a \neq 0$ and there is $\delta_0 > 0$ small such that the quotients

$$|\gamma_0| \cdot \left|\frac{H_{xxx}}{H_{xx}}\right|, \quad |\gamma_0| \cdot \left|\frac{H_{ax}}{H_a}\right|, \quad |J_0| \cdot \left|\frac{H_{aa}}{H_a}\right|, \quad |J_0| \cdot \left|\frac{H_{xxa}}{H_{xx}}\right|,$$

are smaller than δ_0, for all $x \in \gamma_{0,a}$ and $a \in J_0$.

In particular, these conditions imply small distortion of H_{xx} and H_a along $x \in \gamma_{0,a}$ and $a \in J_0$.

Preimages of the central branch. — $|B_x| \geqslant 2$, for all $\beta \in \mathcal{B}$, $x \in \beta_a$, $a \in J_0$. For each $\beta \in \mathcal{B}$, let

$$J(\beta) = \{a \in J_0 \ ; \ \mathrm{Im}\, H \cap \mathcal{U}(\beta) \neq \varnothing \text{ or } |\mathrm{Im}\, H| \geqslant \tfrac{1}{7}|V_{z_a}|\}.$$

Let V be the mean value of $H_a(a, c)$ along $a \in J_0$. Then there is $\delta_1 > 0$ small such that the quotients

$$\left|\frac{B_a}{B_x V}\right|, \quad |\gamma_0| \cdot \left|\frac{B_{xx}}{(B_x)^2}\right|, \quad |\gamma_0| \cdot \left|\frac{B_{xa}}{(B_x)^2 V}\right|,$$
$$|\gamma_0| \cdot \left|\frac{B_{aa}}{(B_x)^2 V^2}\right|, \quad |\gamma_0|^2 \cdot \left|\frac{B_{xxx}}{(B_x)^3}\right|, \quad |\gamma_0|^2 \cdot \left|\frac{B_{xxa}}{(B_x)^3 V}\right|,$$

are smaller than δ_1, for all $x \in \beta_a$, $a \in J(\beta)$ and $\beta \in \mathcal{B}$. The first quotient implies that preimages are transversally crossed by the critical value of H, and the second implies small distortion of derivatives of the functions $B : \beta \to \gamma_0$.

The following Theorem is proved in [**2**], when Φ_0 is C^∞. In Appendix A we show that in fact C^3 is enough.

***Theorem 2.5* (Colli).** — *If* Φ_0 *satisfies the starting conditions* Geometry, Central Branch and Preimages of the Central Branch, *for sufficiently small* $\eta > 0$, $\delta_0 > 0$ *and* $\delta_1 > 0$ *then*

(1) *for almost all* $a \in J_0$, f_a *is chaotic or renormalizable;*

(2) *parameters for which* f_a *is renormalizable constitute a countable union of closed intervals which is dense in* J_0*;*

(3) *parameters for which* f_a *is chaotic have positive Lebesgue measure in* J_0*;*

(4) *parameters for which* f_a *is non-renormalizable and Misiurewicz have zero Lebesgue measure in* J_0.

Therefore we are left to prove that, given $\eta > 0$, $\delta_0 > 0$ and δ_1, there is a choice of V_z and $\varepsilon > 0$ such that for every map Φ_0 as above, constructed for $a \in J_0$, $J_0 \subset [-\varepsilon, \varepsilon]$, the starting conditions are satisfied with the constants η, δ_0 and δ_1.

In the proof we rely mostly on expansion estimates which comes from the Misiurewicz hypothesis. It is known that distortion of derivatives can be obtained using expansion along iterates, and the same will be true for the quotients mentioned above, related to distortion involving both the parameter and the configuration space. The estimates are, however, more delicate, and recovering of bad derivatives must be achieved in unusual manners, mainly when parameter is involved. We call *circular recovering* the ensemble of these techniques, which are developed in Section 4, and their first applications appear already in Section 5, where the first derivative with respect to the parameter appears.

In addition to expansion obtained from the proximity of a Misiurewicz bifurcation, the techniques exposed in Section 4 use also the geometry generated by the dynamics and *a priori* distortion coming from the hypothesis on the Schwarzian derivative.

We believe that this result could be stated without the Misiurewicz hypothesis, but some obstacles should be bypassed. First, a transversality condition should be formulated for germs of families unfolding a general non-renormalizable map. Second, some features of the geometry should be adapted. And third, some expansion would be desirable, unless a completely different approach could control the quotients *without expansion* (more or less like the Schwarzian derivative controls distortion even if little of the dynamics is known).

The Sections are organized as follows. In Section 3 we briefly discuss constants and their hierarchy, and state immediate consequences of non-positiveness of the Schwarzian derivative. The main one is Corollary 3.4, proving the Starting Conditions called "Geometry". We are left to obtain the remaining Starting Conditions, a task which is achieved step by step. In Section 4 we develop the techniques mentioned above which we call "circular recovering". There we deal with the expansion rates of the transfer maps $W : \omega \to V_{z_a}$, for $\omega \in \mathcal{W}$. In fact, more than simply estimating W_x, we also look at derivatives of intermediate iterates, like $F^i|\omega$ if $i < k$ and $W = F^k|\omega$. Moreover, we are able to recover not only "bad derivatives" but also "the square of bad derivatives", which is essential to Section 6. In Section 5 we explore the transversality assumption on the bifurcation and control the quotient W_a/W_x (and also intermediate iterations). This quotient is related with the way pre-images ω of V_{z_a} are crossed by the critical value. We also prove that the set of parameters Γ where the critical value does not belong to any of these pre-images has zero Lebesgue measure. In Section 6 the remaining quotients for the transfer maps W are controlled.

In Section 7 we obtain the Starting Conditions called "Central Branch". Estimates of Sections 5 and 6 are used, since the central branch $H : \gamma_0 \to V_{z_a}$ is the composition $W_0 \circ F|\gamma_0$ (recalling that W_0 is the transfer map of the pre-image ω_0 of V_{z_a} to which the critical value belongs).

In Section 8 we work with regular branches $P : \pi \to V_{z_a}$ and their compositions, which form the maps $B : \beta \to \gamma_0$. Recall that P is the composition $W \circ F|\pi$, for some $W : \omega \to V_{z_a}$, $\omega \in \mathcal{W}$. The goal is to control expansion of compositions, since there are also bad derivatives for some of the P's. But bad derivatives may be recovered as in Section 4, with ideas resembling "circular recovering".

In Section 9 we study the first derivative with respect to the parameter for compositions of regular branches and we achieve control on the first quotient B_a/B_x of the Starting Conditions "Pre-images of the central branch". The remaining quotients are obtained in Section 10.

Everywhere we have to work with mixed derivatives of compositions, using the formula stated in Appendix B. In Appendix A, as we said above, a key lemma in [**2**] is stated for C^3 families, instead of C^∞. The same approach could be useful whenever one has to deal with saddle-nodes and parameter distortion at the same time.

3. Conventions, distortion and geometry

We adopt the following convention on constants. We denote by C_0 a constant greater than 0 which is bigger than any constant used from now on which depends only on functions belonging to a C^3 small neighborhood of f_0. This includes universal constants which do not depend even on these functions. Next, we adopt C_y as the constant which depends also on the choice of y, and C_z as the constant depending on the choice of z. There will be some abuse of notation when we calculate things as "$3C_0^4$" and after all say that it is smaller than C_0. This means that if in some previous Lemma we have estimated something with $\widetilde{C_0}$ and now we are obtaining another estimate $\widehat{C_0} = 3\widetilde{C_0}^4$ then C_0 is greater than both $\widetilde{C_0}$ and $\widehat{C_0}$.

The Greek letter δ will be used as an auxiliary quantifier, appearing always as "given $\delta > 0$ there is...". We will choose δ sufficiently small such that the Starting Conditions are satisfied for given η, δ_0 and δ_1.

Remark that we have the freedom to choose V_z (independently of V_y) in such a way that the ratio $|V_z|/|V_y|$ is small. After the choice of V_z we can also choose ε small. For example, we define

$$r = r(z) = 2\frac{|V_z|}{|V_y|}$$

and choose ε small so that

$$\frac{|V_{z_a}|}{|V_{y_a}|} \leqslant r(z)$$

for all $a \in [-\varepsilon, \varepsilon]$. Moreover, the constant ε has to be chosen small to validate the constants C_0, C_y and C_z.

To be more precise, we will be interested not only on the ratio $|V_{z_a}|/|V_{y_a}|$, but on the size of V_{z_a} compared with both connected components of $V_{y_a} \smallsetminus V_{z_a}$. But the involution function $\tau = \tau_a$ is Lipschitz with constant C_0, for $a \in [-\varepsilon, \varepsilon]$, so that $r(z)$ small also implies that V_{z_a} is uniformly small compared with its adjacent components of $V_{y_a} \smallsetminus V_{z_a}$.

Below we introduce the small constant $\theta > 0$, which will be related to the extendibility of iterations of the map. It will directly depend on $r = r(z)$.

Other constants, $\sigma = \sigma(y) > 1$ and $\lambda = \sqrt{\sigma}$ will depend only on the choice of V_y (with ε small, of course), and will be related to the rate of expansion outside V_{y_a}.

Finally, we use the symbols "$\simeq$", "$\lesssim$" and "$\gtrsim$", in the following sense. For some fixed small constant $\xi > 0$, say $\xi = 10^{-3}$, $C \lesssim D$ whenever $D > 0$ and $C \leqslant (1+\xi)D$. Then $C \gtrsim D$ if and only if $D \lesssim C$ and $C \simeq D$ if and only if $C \lesssim D$ and $D \lesssim C$.

Non-positive Schwarzian derivative has its main consequence in the *Koebe principle*, which is restated in the following form.

Lemma 3.1. — *Given $\theta > 0$, there is $q > 0$ such that if $f : \widehat{I} \to f(\widehat{I})$ is a diffeomorphism, $Sf(x) \leqslant 0$ for all $x \in \widehat{I}$, $I \subset \widehat{I}$ is another interval and $f(I)$ is smaller than q times the size of each connected component of $f(\widehat{I}) \smallsetminus f(I)$ then there is a $\theta^{-1}|I|$-neighborhood $\widetilde{I}$ of I in $\widehat{I}$ such that the derivative of f has small distortion in $\widetilde{I}$, that is*

$$\frac{f'(x)}{f'(y)} \simeq 1.$$

Proof. — See [**13**] for a detailed account. □

Lemma 3.1 has the following important Corollaries, which prove the Geometry of the Starting Conditions. They will be used in many points of this work.

Corollary 3.2. — *Given $\theta > 0$, if $r = r(z)$ is sufficiently small then $|\omega|$ is θ times smaller than the two connected components of $\widehat{\omega} \smallsetminus \omega$.*

Proof. — The transfer map $W : \omega \to V_{z_a}$ is extendible to $W : \widehat{\omega} \to V_{y_a}$. But W has non-positive Schwarzian derivative, since it is a power of f_a and the sign of the Schwarzian derivative is preserved by compositions. Then the Koebe principle can be applied to W. □

Corollary 3.3. — *Given $\theta > 0$, if $r = r(z)$ is sufficiently small then*

$$\frac{|\pi|}{\operatorname{dist}(\pi, \partial V_{z_a})} < \theta,$$

for all $\pi \in \mathcal{P}$.

Proof. — By Lemma 2.4, $\widehat{\omega} \subset [F(z_a), 1]$, $\forall\, \omega \in \mathcal{W}$. Combining with Corollary 3.2, ω is as small as we want compared with $\operatorname{dist}(\omega, F(z_a))$, provided $r(z)$ is small. But for every $\pi \in \mathcal{P}$, $F(\pi) = \omega$, for some $\omega \in \mathcal{W}$. The Lemma follows, since F is approximately quadratic on V_{z_a}. □

Corollary 3.4. — *Given $\eta > 0$, if $r = r(z)$ and ε are sufficiently small then*

$$\frac{|\gamma_0|}{|V_{z_a}|} < \eta, \qquad \frac{|\beta|}{\operatorname{dist}(\beta, \gamma_0)} < \eta, \qquad \frac{|\beta|}{\operatorname{dist}(\beta, \partial V_{z_a})} < \eta,$$

for all $\beta \in \mathcal{B}$. Moreover, for each $\beta \in \mathcal{B}$, the diffeomorphism $B : \beta \to \gamma_0$ is extendible to a $\eta^{-1}|\beta|$-neighborhood of β.

Proof. — The first inequality can be obtained with ε small. The intervals $\omega = \omega_a$ accumulate (uniformly on a) in $w = w_a$. If ε is small then ω_0 must be small for every $J_0 \subset [-\varepsilon, \varepsilon]$ and γ_0 will be small as well, compared with V_{z_a}, whose size is approximately constant. Moreover γ_0 is small compared with each one of the connected components of $V_{z_a} \smallsetminus \gamma_0$.

To prove the remaining assertions, observe that $\mathcal{U}(\beta)$ is into the connected component of $V_{z_a} \smallsetminus \gamma_0$ to which β belongs, and $B : \beta \to \gamma_0$ is extendible to $B : \mathcal{U}(\beta) \to V_{z_a}$. By Lemma 3.1, if ε is small then there is an $\eta^{-1}|\beta|$-neighborhood of β in $\mathcal{U}(\beta)$. In particular the other inequalities are valid and B is extendible to this neighborhood. □

4. Circular recovering

In this Section we deal with expansion of derivatives along the iterates which send an interval $\omega \in \mathcal{W}$ onto V_{z_a}. We use Proposition 4.1 below, proved for example in [**13**], which assures some expansion of derivatives provided some simple information is given about the orbit. In the proof of this Proposition, a loss in expansion at a given iterate is compensated by the iterates following it, which is a kind of *forward recovering* of the derivative. Lemma 4.5 below says that the last loss of expansion in the derivative could also be recovered by the first iterates. This could be called a *backward recovering*. We call *circular recovering* the combined use of these techniques. The same ideas appear in Sections 8 and 9, in a slightly different context. They are in the core of this work and deserve a careful attention.

Proposition 4.1. — *There is $C_y > 0$, $\sigma = \sigma(y) > 1$ and $\varepsilon > 0$ such that if $a \in [-\varepsilon, \varepsilon]$ then $F = F(a, \cdot)$ has the following properties.*

(1) *If $x, \dots, F^{k-1}(x) \notin V_{y_a}$ then $|F_x^k(a, x)| \geqslant C_y^{-1}\sigma^k$.*
(2) *If $x, \dots, F^{k-1}(x) \notin V_{z_a}$ and $F^k(x) \in V_{y_a}$ then $|F_x^k(a, x)| \geqslant C_y^{-1}\sigma^k$.*
(3) *If $x, \dots, F^{k-1}(x) \notin V_{z_a}$ then $|F_x^k(a, x)| \geqslant C_y^{-1}\sigma^k \inf_{i=0,\dots,k-1} |F_x(a, F^i(a, x))|$.*

This constant $\sigma = \sigma(y) > 1$ will be fixed from now on. The first consequence is bounded distortion for iterates outside V_{y_a}.

Lemma 4.2. — *Suppose $\varepsilon > 0$ small and $a \in [-\varepsilon, \varepsilon]$. There is $C_y > 0$ such that if T is an interval satisfying $F^i(T) \cap V_{y_a} = \varnothing$ for all $i = 0, \dots, j-1$ then*

$$\frac{F_x^j(u)}{F_x^j(v)} \leqslant C_y$$

for all $u, v \in T$.

Proof. — Write

$$\big|\log|F_x^j(u)| - \log|F_x^j(v)|\big| = \left|\sum_{i=0}^{j-1} \log|F_x(F^i u)| - \log|F_x(F^i v)|\right|,$$

which is smaller than

$$\widetilde{C}_y \sum_{i=0}^{j-1} |F^i u - F^i v|,$$

where $\widetilde{C}_y = \max\{|\frac{\partial}{\partial x}\log|F_x(a, x)||\ ;\ x \notin V_{y_a}\}$, remarking that $F_x(F^i u)$ and $F_x(F^i v)$ have the same sign for $i = 0, \dots, j-1$. But Proposition 4.1 implies

$$1 \geqslant |F^j u - F^j v| \geqslant C_y^{-1}\sigma^{j-i}|F^i u - F^i v|,$$

proving the Lemma. □

Now we fix some $\omega \in \mathcal{W}$ and $x \in \omega$, and suppose $a \in [-\varepsilon, \varepsilon]$, for ε small. We write $W : \omega \to V_{z_a}$ as $W = F^k|\omega$. The next Lemma says that when the orbit visits the interval V_{y_a} the square of the derivative can be recovered by the next iterates until the next visit of the orbit to V_{y_a}.

Lemma 4.3. — *There is $C_y > 0$ such that if $u = F^l x \in V_{y_a} \setminus V_{z_a}$ and $j \geqslant 2$ is the first integer such that $F^j u \in V_{y_a}$ then*

$$|F_x^{j-1}(Fu)| \cdot |F_x(u)|^2 \geqslant C_y^{-1}.$$

Proof. — Let $T = [Fu, \widetilde{w}_a]$. As F is approximately quadratic and ε is small then

$$|T| \leqslant C_0 |F_x(u)|^2.$$

Hence the Lemma will be proved if we show that

$$|F_x^{j-1}(Fu)| \cdot |T| \geqslant C_y^{-1}.$$

This in turn follows from $|F_x^i(Fu)| \cdot |T| \geqslant C_y^{-1}$, where i is the first integer such that $F^i(T) \cap V_{y_a} \neq \varnothing$, since $i \leqslant j-1$ and $|F_x^{j-1-i}(F^{i+1}u)| \geqslant C_y^{-1}\sigma^{j-1-i}$. Now $F^i(T)$ is an interval intersecting V_{y_a}, but with a point, say $F^i(\widetilde{w}_a)$, outside a neighborhood V containing the closure of V_{y_a} (see definitions of V and V_{y_a} in Section 2). This implies that there is $d > 0$ such that $|F^i(T)| \geqslant d$.

By Lemma 4.2,

$$d \leqslant |F^i(T)| \leqslant C_y |F_x^i(Fu)| \cdot |T|,$$

proving the Lemma. □

The following Lemma is a corollary of the proof of Lemma 4.3. It says that the square of a bad derivative $F_x(u)$, $u = F^l x$, may also be recovered by the *first iterates* of the orbit of x.

Lemma 4.4. — *Let $i_1 \geqslant 1$ be the first integer such that $F^{i_1} x \in V_{y_a}$. There is $C_y > 0$ such that if $u = F^l x$, $l < k$, is such that $u \in V_{y_a} \setminus V_{z_a}$ then*

$$|F_x^{i_1}(x)| \cdot |F_x(u)|^2 \geqslant C_y^{-1}.$$

Proof. — As in the proof of Lemma 4.3, let $T = [Fu, \widetilde{w}_a]$. We want to show that $|F_x^{i_1}(x)| \cdot |T| \geqslant C_y^{-1}$. If i is the first integer such that $F^i(T) \cap V_{y_a} \neq \varnothing$ then $i \leqslant i_1$ (since $T \supset \omega$). Hence it suffices to show that $|F_x^i(x)| \cdot |T| \geqslant C_y^{-1}$. But by the bounded distortion of the derivative of $F^i|T$ and since $x \in T$ we have $d \leqslant |F^i(T)| \leqslant C_y |F_x^i(x)| \cdot |T|$, for some fixed $d > 0$, and the Lemma follows. □

If the square of a bad derivative is recovered by the first iterates then the same happens with the derivative itself. This is the content of the following Corollary. Let $\lambda = \lambda(y) = \sqrt{\sigma(y)}$, where σ is given by Proposition 4.1.

Corollary 4.5. — *Let $i_1 \geqslant 1$ be the first integer such that $F^{i_1}x \in V_{y_a}$. There is $C_y > 0$ such that if $u = F^l x$, $l \leqslant k$, is such that $u \in V_{y_a} \setminus V_{z_a}$ then*

$$|F_x^{i_1}(x)| \cdot |F_x(u)| \geqslant C_y^{-1}\lambda^{i_1}.$$

Proof. — By Lemma 4.4,

$$|F_x^{i_1}(x)|^{1/2} \cdot |F_x(u)| \geqslant C_y^{-1/2}.$$

On the other hand, $x, Fx, \dots, F^{i_1-1}x \notin V_{y_a}$, hence by Proposition 4.1

$$|F_x^{i_1}(x)|^{1/2} \geqslant C_y^{-1/2}(\sigma^{1/2})^{i_1}.$$

The Corollary is proved if we multiply both sides of the first inequality by $|F_x^{i_1}(x)|^{1/2}$ and then use the second inequality. □

The following two Corollaries will be directly applied in the following Sections.

Corollary 4.6. — *There are $C_y > 0$ and $\lambda = \lambda(y) > 1$ such that the following holds. For all $x \in \omega$, $\omega \in \mathcal{W}$ with transfer map $W = F^k|\omega : \omega \to V_{z_a}$ and $u = F^l x$, for $0 \leqslant l \leqslant k-1$, we have:*

(1) $|F_x^{k-l}(u)| \geqslant C_y^{-1}\lambda^{k-l}$.
(2) *If $u \notin V_{y_a}$ then $|F_x^{k-l-1}(Fu)| \cdot |F_x(u)|^2 \geqslant C_y^{-1}\lambda^{k-l}$.*
(3) *If $u \in V_{y_a}$ then $|F_x^{k-l-1}(Fu)| \cdot |F_x(u)|^2 \geqslant C_y^{-1}\lambda^{s}$, where*

$$s = \#\{l+1 \leqslant i < k \;;\; F^i x \in V_{y_a}\}.$$

Proof. — The first inequality comes directly from Proposition 4.1. It is valid also for λ since $\lambda < \sigma$. The second inequality follows if we use the first and observe that if $u \notin V_{y_a}$ then $|F_x(u)| \geqslant C_0^{-1}|V_{y_a}| \geqslant C_y^{-1}$. For the last inequality we use Lemma 4.3 to assure that the square of the bad derivative is recuperated until the next visit of the orbit to V_{y_a}. From this moment on we use the expansion given by the first inequality, with unknown number of iterates surely greater or equal than s. □

Corollary 4.7. — *There are $C_y > 0$ and $\lambda = \lambda(y) > 1$ such that*

$$|F_x^j(x)| \geqslant C_y^{-1}\lambda^j$$

for all $x \in \omega$, $\omega \in \mathcal{W}$ with transfer map $W = F^k : \omega \to V_{z_a}$ and $1 \leqslant j \leqslant k$.

Proof. — Let $l \leqslant j$ be the last iterate such that $F^l x \in V_{y_a}$ and $1 \leqslant i_1 \leqslant l$ be the first iterate such that $F^{i_1}x \in V_{y_a}$. If $j = l$ then Proposition 4.1 implies the Corollary. Otherwise we write

$$|F_x^j(x)| = |F_x^{j-l-1}(F^{l+1}x)| \cdot |F_x(F^l x)| \cdot |F_x^{l-i_1}(F^{i_1}x)| \cdot |F_x^{i_1}(x)|.$$

As $F^{l+1}x, F^{l+2}x, \dots, F^j x \notin V_{y_a}$, by Proposition 4.1 we have

$$|F_x^{j-l-1}(F^{l+1}x)| \geqslant C_y^{-1}\lambda^{j-l-1}.$$

In addition, $F_x^{i_1}, \dots, F^{l-1}x \notin V_{z_a}$ and $F^l x \in V_{y_a}$, hence again by Proposition 4.1 we have $|F_x^{l-i_1}(F^{i_1}x)| \geqslant C_y^{-1}\lambda^{l-i_1}$. Finally, by Corollary 4.5, $|F_x(F^l x)| \cdot |F_x^{i_1}(x)| \geqslant C_y^{-1}\lambda^{i_1}$. □

5. Exploring transversality

In this Section we combine the estimates of Section 4 with the transversality assumption. For $\omega \in \mathcal{W}$ with transfer map $W = F^k|\omega : \omega \to V_{z_a}$ we may define $x_\omega = x_{\omega,a} = W^{-1}(c)$ as the "center" of $\omega = \omega_a$.

Using the Glossary (at the end of this work), we obtain

$$\frac{d}{da}x_{\omega,a} = -\frac{W_a}{W_x}(a, x_{\omega,a}) = -\sum_{i=1}^{k} \frac{F_a \circ F^{i-1}}{F_x^i}(a, x_{\omega,a}).$$

We want in fact to give estimates on W_a/W_x for every $x \in \omega$ and even estimates on F_a^j/F_x^j, for every $x \in \omega$ and $j = 1, \dots, k$, as in the following Lemma.

Lemma 5.1. — *There is $C_y > 0$ such that*

$$\left|\frac{F_a^j}{F_x^j}\right| \leqslant C_y,$$

for every $x \in \omega$, $\omega \in \mathcal{W}$ with transfer map $W = F^k|\omega : \omega \to V_{z_a}$ and $j = 1, \dots, k$.

Proof. — By the Glossary,

$$\frac{F_a^j}{F_x^j} = \sum_{i=1}^{j} \frac{F_a \circ F^{i-1}}{F_x^i}.$$

But F_a is bounded by C_0 and $|F_x^i| \geqslant C_y^{-1}\lambda^i$, by Corollary 4.7. □

Lemma 5.2. — *Given $\delta > 0$, there are an integer $k = k(\delta, y) \geqslant 1$ and $\mu = \mu(\delta) > 0$ such that if $\omega_1, \omega_2 \in \mathcal{W}$ have transfer maps $W_s = F^{k_s}|\omega_s$, $s = 1, 2$ with $k_1, k_2 \geqslant k$ and moreover $x_s \in \omega_s$, $s = 1, 2$, satisfy $|x_1 - x_2| < \mu$ then*

$$\left|\frac{W_{1,a}}{W_{1,x}}(x_1) - \frac{W_{2,a}}{W_{2,x}}(x_2)\right| < \delta.$$

Proof. — Let $C_y > 0$ and $\lambda = \lambda(y)$ be as in Corollary 4.7 and let

$$C_0 > \max\{|F_a| \,;\, a \in [-\varepsilon, \varepsilon],\, x \in [0, 1]\}.$$

Let $k = k(\delta, y)$ be such that

$$C_y C_0 \frac{\lambda^{-k}}{1 - \lambda^{-1}} < \frac{\delta}{4}.$$

Write

$$\frac{W_{s,a}}{W_{s,x}}(x_s) = \sum_{i=1}^{k_s} \frac{F_a \circ F^{i-1}}{F_x^i}(a, x_s),$$

for $s = 1, 2$. If $k_s > k$ then

$$\left| \sum_{i=k+1}^{k_s} \frac{F_a \circ F^{i-1}}{F_x^i}(a, x_s) \right| \leqslant C_y C_0 \frac{\lambda^{-k}}{1-\lambda^{-1}} < \frac{\delta}{4},$$

for $s = 1, 2$, using Corollary 4.7 and the choice of k. Then we are left to proving that

$$\left| \sum_{i=1}^{k} \frac{F_a \circ F^{i-1}}{F_x^i}(a, x_1) - \sum_{i=1}^{k} \frac{F_a \circ F^{i-1}}{F_x^i}(a, x_2) \right| < \frac{\delta}{2}.$$

But this is true if $|x_1 - x_2| < \mu$, for sufficiently small $\mu > 0$. $\square$

Let us see what are the consequences of Lemma 5.2. Let $\{\omega_N\}_N$ be a sequence converging to w at $a = 0$. In particular the centers $x_N = x_{\omega_N}$ converge to w and their continuations $a \mapsto x_{N,a}$ converge in the C^0 topology to $a \mapsto w_a$, for $a \in [-\varepsilon, \varepsilon]$, $\varepsilon > 0$ small. This is easy to be proved since the rates of expansion outside V_{z_a} are uniform. By Lemma 5.2, $\{a \mapsto x_{N.a}\}_N$ also converges in the C^1 topology. This leads to a formulae on $\dfrac{d}{da} w_a$:

$$\frac{d}{da} w_a = -\sum_{i=1}^{\infty} \frac{F_a \circ F^{i-1}}{F_x^i}(a, w_a).$$

Now let $\nu > 0$ be such that

$$F_a(0, c) - \left. \frac{d}{da} w_a \right|_{a=0} \geqslant 2\nu,$$

by the transversality condition. This implies that if V_z is chosen sufficiently small, in order that every $\omega \in \mathcal{W}$ is forced to be near w, and $a \in [-\varepsilon, \varepsilon]$, for $\varepsilon > 0$ small, then

$$\frac{W_a}{W_x}(a, x) \geqslant \nu - F_a(0, c),$$

for every $x \in \omega = \omega_a$, $\omega \in \mathcal{W}$ and $a \in [-\varepsilon, \varepsilon]$.

Moreover, if $\varepsilon > 0$ is small then for every point $x \in \Lambda_{z_0}$ in $[f_0(z), \widetilde{w}]$, its continuation $x_a = h_a(x)$ has velocity smaller than $\frac{d}{da} f_a(c) - \nu$. This implies two things: (i) to each $x \in \Lambda_z$ corresponds (at most) a single point $a = a(x) \in [-\varepsilon, \varepsilon]$ in the parameter space such that $f_a(c) = x_a$ and (ii) for every $\omega \in \mathcal{W}$ the set $\{a \in [-\varepsilon, \varepsilon] \; ; \; f_a(c) \in \omega = \omega_a\}$ is an interval.

Define

$$\Gamma = \{a \in [-\varepsilon, \varepsilon] \; ; \; f_a(c) \in \Lambda_{z_a}\},$$

which is totally disconnected. We will prove below that $\mathrm{Leb}(\Gamma) = 0$. Each gap of Γ corresponds to the parameters for which $f_a(c)$ belongs to some $\omega = \omega_a \in \mathcal{W}$. The collection of gaps in the complement of Γ will be called $\mathcal{J}_0$, and from the next Section on we shall restrict our attention to a particular element J_0 of this collection, as already described in Section 2.

Lemma 5.3. — $\mathrm{Leb}(\Gamma) = 0$.

Proof. — Without loss of generality and for simplicity we will consider in this proof only the negative range $[-\varepsilon, 0]$ and will assume the following: for $a = -\varepsilon$, $f_a(c)$ belongs to the leftmost boundary point of some $\omega_1 = \omega_{1,a} \in \mathcal{W}$, with transfer map $W_1 = F^{n_1}|\omega_1 : \omega_1 \to V_{z_a}$, and any other $\omega = \omega_a \in \mathcal{W}$ between $\omega_{1,a}$ and w_a has transfer map $W = F^n|\omega : \omega \to V_{z_a}$ with $n > n_1$.

For each interval family $I = (I_a)_a$ let

$$J(I) = \{a \in [-\varepsilon, \varepsilon] \; ; \; f_a(c) \in I_a\}.$$

Let $T^1 = (T^1_a)_a$ be the family of intervals with boundary points $\partial_+ T^1_a = w_a$ and $\partial_- T^1_a$ the rightmost point of $\omega_{1,a}$. It is not difficult to see that the following reasoning is independent of a, so we omit the subindex. Let $n_2 \geqslant 1$ be the first integer such that $f^{n_2}(T^1)$ intersects V_z. Then $n_2 > n_1$ and $f^{n_2}(T^1)$ must contain V_z (in fact V_y), since $f^i(\partial_+ T^1) \notin V_y$, $\forall i \geqslant 0$, and $f^i(\partial_- T^1) \notin V_y$, $\forall i \geqslant n_1 + 1$ (by the definition of V_z and V_y). Therefore there is $\omega_2 \in \mathcal{W}$, $\omega_2 \subset T^1$, with transfer map $W_2 = F^{n_2}|\omega_2 : \omega_2 \to V_z$. Moreover, any other $\omega \in \mathcal{W}$ between ω_1 and ω_2 or else between ω_2 and w has transfer map $W = F^k|\omega : \omega \to V_z$ with $n > n_2$.

By Proposition 4.1, the expansion outside V_z is uniform, up to a constant which depends only on the choice of z. Therefore, analogously to Lemma 4.2, we have bounded distortion for iterates outside V_z, this time with a constant C_z. In this particular case, this means that

$$\frac{F^{n_2}_x(x)}{F^{n_2}_x(y)} \leqslant C_z$$

for every $x, y \in T^1$. Hence

$$\frac{|\omega_2|}{|T^1|} \geqslant C_z^{-1} |V_z|.$$

It is easy to see, because of the bounds on velocities, that

$$\frac{|J(\omega_2)|}{|J(T^1)|} \geqslant C_0^{-1} C_z^{-1} |V_z| \equiv \widetilde{C}_z^{-1}.$$

The interval $J(\omega_2)$ is in the complement of Γ. Hence at this stage $\mathrm{Leb}(\Gamma) \leqslant (1 - \widetilde{C}_z^{-1})|J(T^1)|$. The argument continues by induction in the remaining connected components of $J(T^1) \setminus J(\omega_2)$, and so on, in order that at every stage a definite z-dependent fraction of parameters not belonging to Γ is suppressed from the remaining ones. This proves the Lemma. □

6. Transfer maps

Let $\omega \in \mathcal{W}$ and $W = F^k|\omega : \omega \to V_{z_a}$ its transfer map. We have already established bounds on W_a/W_x in Section 5. In this Section we control the quotients

$$\frac{W_{xx}}{(W_x)^2}, \quad \frac{W_{xa}}{(W_x)^2}, \quad \frac{W_{xxx}}{(W_x)^3}, \quad \frac{W_{aa}}{(W_x)^2}, \quad \frac{W_{xxa}}{(W_x)^3}.$$

Once more we suppose that V_y is already chosen, and then take V_z sufficiently small. We always assume $\omega \in \mathcal{W}$ as above and $a \in [-\varepsilon, \varepsilon]$, for $\varepsilon > 0$ sufficiently small, but constants are independent of these choices. In the Lemmas we omit the argument of functions. It is implicit that they are calculated for $a \in [-\varepsilon, \varepsilon]$ and $x \in \omega$. If we write $F_x \circ F^{i-1}$, for example, it means $F_x(a, F^{i-1}(a, x))$. In this notation, $|F^i - c|$ is the distance from the critical point to the i-th iterate of F.

We start by proving a technical Lemma which is a direct consequence of Corollary 4.6. The goal is to bound the sum

$$S_j = \sum_{i=1}^{j} |(F_x^{k-i} \circ F^i)(F_x \circ F^{i-1})^2|^{-1},$$

where $j \leqslant k$, which appears in all Lemmas of this Section.

Lemma 6.1. — *There is $C_y > 0$ such that $S_j \leqslant C_y$, for all $j \leqslant k$.*

Proof. — This follows from Corollary 4.6. Separate the sum S_j into two sums: the first, containing only those $2 \leqslant i \leqslant j$ such that $F^{i-1} \in V_{y_a}$, is bounded by a y-dependent geometric series, following the third item of the Corollary, and the second, containing only those $1 \leqslant i \leqslant j$ such that $F^{i-1} \notin V_{y_a}$, is also bounded by a y-dependent geometric series, following the second item of the Corollary. □

Lemma 6.2. — *Given $\delta > 0$, if V_z is sufficiently small then*

$$|V_{z_a}| \cdot \left| \frac{W_{xx}}{(W_x)^2} \right| < \delta$$

for all $x \in \omega$, $\omega \in \mathcal{W}$. Moreover, if as above $W = F^k|\omega$ then

$$\frac{|V_{z_a}|}{|F_x^{k-j} \circ F^j|} \cdot \left| \frac{F_{xx}^j}{(F_x^j)^2} \right| < \delta,$$

for all $x \in \omega$ and $j = 1, \ldots, k-1$.

Proof. — Write

$$\frac{1}{F_x^{k-j} \circ F^j} \cdot \frac{F_{xx}^j}{(F_x^j)^2} = \sum_{i=1}^{j} \frac{F_{xx} \circ F^{i-1}}{(F_x^{k-i} \circ F^i)(F_x \circ F^{i-1})^2},$$

for $1 \leqslant j \leqslant k$. As $|F_{xx}|$ is bounded by C_0, the sum is bounded by $C_0 S_j$, where S_j was given above, hence by $C_0 C_y$, by Lemma 6.1. The Lemma is proved if we multiply by $|V_{z_a}|$ and take V_z sufficiently small. □

Lemma 6.3. — *Given $\delta > 0$, if V_z is small enough then*

$$|V_{z_a}|^2 \cdot \left| \frac{W_{xxx}}{(W_x)^3} \right| < \delta,$$

for all $x \in \omega$ and $\omega \in \mathcal{W}$.

Proof. — Write

$$\frac{W_{xxx}}{(W_x)^3} = S_1 + 3S_2,$$

where

$$S_1 = \sum_{i=1}^{k} \frac{F_{xxx} \circ F^{i-1}}{(F_x^{k-i} \circ F^i)^2 (F_x \circ F^{i-1})^3}$$

and

$$S_2 = \sum_{i=2}^{k} \frac{F_{xx} \circ F^{i-1}}{(F_x^{k-i} \circ F^i)(F_x \circ F^{i-1})^2} \cdot \frac{1}{F_x^{k-i+1} \circ F^{i-1}} \frac{F_{xx}^{i-1}}{(F_x^{i-1})^2}.$$

We start by estimating $|V_{z_a}|^2 S_2$. By Lemma 6.2,

$$\frac{|V_{z_a}|}{F_x^{k-i+1} \circ F^{i-1}} \frac{F_{xx}^{i-1}}{(F_x^{i-1})^2}$$

is smaller than δ, for every $i = 2, \ldots, k$, provided V_z is small. Using Lemma 6.1 as in Lemma 6.2 we have

$$|V_{z_a}|^2 |S_2| \leqslant |V_{z_a}| C_0 C_y \delta,$$

which is smaller than $\delta/6$ if V_z is sufficiently small.

Similarly, using the first item of Corollary 4.6, we bound $|V_{z_a}|^2 S_1$ by

$$C_0 C_y |V_{z_a}|^2 \sum_{i=1}^{k} |(F_x^{k-i} \circ F^i)(F_x \circ F^{i-1})^2|^{-1}$$

which is smaller than $\delta/2$, if V_z is sufficiently small, by Lemma 6.1. □

Let $k = k(\omega)$ be the transfer time from ω to V_{z_a}, for $\omega \in \mathcal{W}$. Let

$$N = \min\{k(\omega)\ ;\ \omega \in \mathcal{W}\}.$$

By the definition of $\mathcal{W}$, if V_z is small then any $\omega \in \mathcal{W}$ must be near w_a, hence N is big. In the following Lemma we use the fact that N/λ^N is as small as we wish, provided V_z is sufficiently small.

Lemma 6.4. — *Given $\delta > 0$, if V_z is sufficiently small then*

$$|V_{z_a}| \cdot \left| \frac{W_{xa}}{(W_x)^2} \right| < \delta$$

for all $x \in \omega$, $\omega \in \mathcal{W}$. Moreover

$$\frac{|V_{z_a}|}{|F_x^{k-j} \circ F^j|} \cdot \left| \frac{F_{xa}^j}{(F_x^j)^2} \right| < \delta,$$

for all $j < k$, where $k = k(\omega)$.

Proof. — Write

$$\frac{1}{F_x^{k-j}\circ F^j}\cdot\frac{F^j_{xa}}{(F^j_x)^2}=\sum_{i=1}^{j}\frac{F_{xa}\circ F^{i-1}}{W_x(F_x\circ F^{i-1})}+\sum_{i=2}^{j}\frac{F_{xx}\circ F^{i-1}}{(F_x^{k-i}\circ F^i)(F_x\circ F^{i-1})^2}\cdot\frac{F_a^{i-1}}{F_x^{i-1}}.$$

We have $|F_{xa}|<C_0$, $|W_x|\geqslant C_y^{-1}\lambda^k$ and $|F_x\circ F^{i-1}|\geqslant C_0^{-1}|V_{z_a}|$, therefore

$$|V_{z_a}|\cdot\left|\sum_{i=1}^{j}\frac{F_{xa}\circ F^{i-1}}{W_x(F_x\circ F^{i-1})}\right|<C_0^2C_y\frac{k}{\lambda^k},$$

which is smaller than $\delta/2$ if the choice of V_z implies N sufficiently big.

Moreover, by Lemma 5.1

$$\left|\frac{F_a^{i-1}}{F_x^{i-1}}\right|\leqslant C_y$$

But Lemma 6.1 implies that

$$C_y|V_{z_a}|\cdot\sum_{i=2}^{j}\left|\frac{F_{xx}\circ F^{i-1}}{(F_x^{k-i}\circ F^i)(F_x\circ F^{i-1})^2}\right|<\frac{\delta}{2}$$

if V_z is sufficiently small. □

Lemma 6.5. — *Given $\delta>0$, if V_z is sufficiently small then*

$$|V_{z_a}|\cdot\left|\frac{W_{aa}}{(W_x)^2}\right|<\delta,$$

for every $x\in\omega$, $\omega\in\mathcal{W}$.

Proof. — Write

$$\frac{W_{aa}}{(W_x)^2}=S_1+2S_2+S_3,$$

where

$$S_1=\sum_{i=1}^{k}\frac{F_{aa}\circ F^{i-1}}{W_xF_x^i},$$

$$S_2=\sum_{i=2}^{k}\frac{F_{xa}\circ F^{i-1}}{W_xF_x^i}\cdot\frac{F_a^{i-1}}{F_x^{i-1}}$$

and

$$S_3=\sum_{i=2}^{k}\frac{F_{xx}\circ F^{i-1}}{(F_x^{k-i}\circ F^i)(F_x\circ F^{i-1})^2}\cdot\left(\frac{F_a^{i-1}}{F_x^{i-1}}\right)^2.$$

The proof follows as in Lemma 6.4. Note that here denominators are slightly better, and $|F_x^i|$ can be estimated using Corollary 4.7. □

Lemma 6.6. — *Given $\delta>0$, if V_z is sufficiently small then*

$$|V_{z_a}|^2\cdot\left|\frac{W_{xxa}}{(W_x)^3}\right|<\delta,$$

for all $x\in\omega$, $\omega\in\mathcal{W}$.

Proof. — Write

$$\frac{W_{xxa}}{(W_x)^3} = S_1 + S_2 + S_3 + 2S_4 + 2S_5,$$

where

$$S_1 = \sum_{i=1}^{k} \frac{F_{xxa} \circ F^{i-1}}{W_x(F_x^{k-i} \circ F^i)(F_x \circ F^{i-1})^2},$$

$$S_2 = \sum_{i=2}^{k} \frac{F_{xxx} \circ F^{i-1})(F_x \circ F^{i-1})}{(F_x^{k-i} \circ F^i)^2(F_x \circ F^{i-1})^4} \cdot \frac{F_a^{i-1}}{F_x^{i-1}},$$

$$S_3 = \sum_{i=2}^{k} \frac{F_{xa} \circ F^{i-1}}{W_x F_x^i} \cdot \frac{1}{F_x^{k-i+1} \circ F^{i-1}} \cdot \frac{F_{xx}^{i-1}}{(F_x^{i-1})^2}$$

$$S_4 = \sum_{i=2}^{k} \frac{F_{xx} \circ F^{i-1}}{(F_x^{k-i} \circ F^i)(F_x \circ F^{i-1})^2} \cdot \frac{2}{F_x^{k-i+1} \circ F^{i-1}} \cdot \frac{F_{xa}^{i-1}}{(F_x^{i-1})^2}$$

and

$$S_5 = \sum_{i=2}^{k} \frac{F_{xx} \circ F^{i-1}}{(F_x^{k-i} \circ F^i)(F_x \circ F^{i-1})^2} \cdot \frac{F_a^{i-1}}{F_x^{i-1}} \cdot \frac{1}{F_x^{k-i+1} \circ F^{i-1}} \cdot \frac{F_{xx}^{i-1}}{(F_x^{i-1})^2}.$$

The only "new" term to pay attention is

$$\frac{|V_{z_a}|}{F_x^{k-i+1}} \cdot \frac{F_{xa}^{i-1}}{(F_x^{i-1})^2},$$

but it can be bounded using Lemma 6.4. □

7. Central branch

Now we fix J_0, the parameter interval such that the critical value belongs to $\omega_0 \in \mathcal{W}$. Therefore the central branch $H : \gamma_0 \to V_{z_a}$ of the first return map to V_{z_a} may be written as $H = W_0 \circ F$, where $W_0 : \omega_0 \to V_{z_a}$ is the transfer map associated to ω_0.

All the Lemmas below depend on the fact that V_z and ε are sufficiently small, so we omit it in the statements.

The following Lemma shows in particular that $H_a(a,x)$ is nonzero for every $x \in \gamma_0$, $a \in J_0$ and its sign is determined by the sign of $W_{0,x}$.

Lemma 7.1

$$\frac{H_a(a,x)}{W_{0,x}(a,F(a,x))} \geqslant \frac{\nu}{2},$$

for every $x \in \gamma_0$ and $a \in J_0$.

Proof. — Write

$$H_a(a,x) = W_{0,a}(a,F(a,x)) + W_{0,x}(a,F(a,x))F_a(a,x).$$

If V_z and ε are small then $F_a(a,x)$ is very near $F(0,c)$. But in Section 5 we have shown that

$$\frac{W_{0,a}}{W_{0,x}} \geqslant \nu - F_a(0,c),$$

and the estimate follows. □

An analogous statement is valid for H_{xx}.

Lemma 7.2

$$\frac{H_{xx}(a,x)}{W_{0,x}(a,F(a,x))} \simeq F_{xx}(0,c) < 0,$$

for all $x \in \gamma_0$ and $a \in J_0$.

Proof. — As the critical point is quadratic, $(F_x(a,x))^2 < C_0|\omega_0|$ for every $x \in \gamma_0$. Moreover, the function $W_0 : \omega_0 \to V_{z_a}$ has small distortion, by Lemma 3.1 and Proposition 2.3, implying that

$$|W_{0,x}| \simeq \frac{|V_{z_a}|}{|\omega_0|}.$$

Also, by continuity, $F_{xx}(a,x) \simeq F_{xx}(0,c)$. We have

$$\frac{H_{xx}}{W_{0,x}(a,F)} = (F_x)^2 W_{0,x}(a,F)\,\frac{W_{0,xx} \circ F}{(W_{0,x} \circ F)^2} + F_{xx},$$

hence the Lemma is proved if we show that

$$C_0|V_{z_a}|\,\frac{W_{0,xx}}{(W_{0,x})^2}$$

is small for points in ω_0. But this is true by Lemma 6.2, choosing V_z small with respect to V_y. □

At this point we are ready to prove the four starting conditions relative to the central branch. For simplicity, we write from now on $W = W_0$ to designate the transfer map of ω_0.

Lemma 7.3. — *Given $\delta > 0$, V_z is sufficiently small then*

$$|\gamma_0| \cdot \left|\frac{H_{xxx}}{H_{xx}}\right| < \delta,$$

for all $x \in \gamma_0$ and $a \in J_0$.

Proof. — As $H = W \circ F$ we have

$$H_{xxx} = (W_x \circ F)F_{xxx} + (W_{xxx} \circ F)(F_x)^3 + 3(W_{xx} \circ F)F_x F_{xx}.$$

We analyze these three terms, each one divided by H_{xx} and multiplied by $|\gamma_0|$. By Lemma 7.2,

$$H_{xx} \simeq s_0(W_x \circ F),$$

where $s_0 = F_{xx}(0,c)$. Hence

$$|\gamma_0| \cdot \left|\frac{(W_x \circ F)F_{xxx}}{H_{xx}}\right| < \frac{2C_0}{|s_0|}\,|\gamma_0|,$$

which is smaller than $\delta/3$ if V_z is small (since $|\gamma_0| \ll |V_{z_a}|$). The second term can be written as

$$|\gamma_0|(F_x)^3\,\frac{(W_x \circ F)^3}{H_{xx}}\,\frac{W_{xxx} \circ F}{(W_x \circ F)^3}.$$

But $|F_x|^3 \leqslant C_0|\gamma_0|^3$,

$$|W_x \circ F| \simeq \frac{|V_{z_a}|}{|\omega_0|}$$

and $|\gamma_0|^2 \leqslant C_0|\omega_0|$, hence the second term is bounded by

$$\frac{2C_0^3}{|s_0|}\,|V_{z_a}|^2\left|\frac{W_{xxx} \circ F}{(W_x \circ F)^3}\right|$$

which, according to Lemma 6.3, can be smaller than $\delta/3$ if V_z is sufficiently small. Similarly, the last term is bounded by

$$\frac{6C_0^4}{|s_0|}\,|V_{z_a}| \cdot \left|\frac{W_{xx} \circ F}{(W_x \circ F)^2}\right|,$$

which can be made smaller than $\delta/3$ by the choice of V_z, according to Lemma 6.2. □

Lemma 7.4. — *Given $\delta > 0$, if V_z is sufficiently small then*

$$|\gamma_0| \cdot \left|\frac{H_{ax}}{H_a}\right| < \delta$$

for all $x \in \gamma_0$ and $a \in J_0$.

Proof. — Writing $H = W \circ F$ we obtain

$$H_{xa} = (W_x \circ F)F_{xa} + (W_{xa} \circ F)F_x + (W_{xx} \circ F)F_xF_a$$

and then analyze each term when multiplied by $|\gamma_0| \cdot |H_a|^{-1}$. By Lemma 7.1 we have

$$|H_a(a,x)| \geqslant \frac{\nu}{2}|W_x \circ F|,$$

so that

$$|\gamma_0| \cdot \left|\frac{W_x \circ F}{H_a}\right| \cdot |F_{xa}| < \frac{2C_0}{\nu}\,|\gamma_0|\,,$$

which is smaller than $\delta/3$ if V_z (and hence γ_0) is sufficiently small. Also,

$$|\gamma_0| \cdot |F_x|\left|\frac{W_{xa} \circ F}{H_a}\right| < \frac{2}{\nu}\,|\gamma_0| \cdot |F_x| \cdot |W_x \circ F| \cdot \left|\frac{W_{xa} \circ F}{(W_x \circ F)^2}\right|.$$

As in the proof of the previous Lemma, $|F_x| < C_0\,|\gamma_0|$, $|W_x \circ F| \simeq |V_{z_a}|/|\omega_0|$ and $|\gamma_0|^2/|\omega_0| \leqslant C_0$, so that $|\gamma_0| \cdot |F_x| \cdot |W_x \circ F| \leqslant 2C_0^2|V_{z_a}|$. Then Lemma 6.4 implies that

$$\frac{4C_0^2}{\nu}|V_{z_a}| \cdot \left|\frac{W_{xa} \circ F}{(W_x \circ F)^2}\right| < \frac{\delta}{3},$$

provided V_z is small enough.

The same argument combined with Lemma 6.2 is applied to the last term, proving the Lemma. □

Lemma 7.5. — *Given $\delta > 0$, if V_z is sufficiently small then*

$$|J_0| \cdot \left| \frac{H_{aa}}{H_a} \right| < \delta$$

for all $x \in \gamma_0$ and $a \in J_0$.

Proof. — First observe that $|J_0| \leqslant C_0|\omega_0|$ and

$$|H_a| \geqslant \frac{\nu}{2} |W_x \circ F| \simeq \frac{\nu}{2} \cdot \frac{|V_{z_a}|}{|\omega_0|}.$$

Then write

$$H_{aa} = (W_x \circ F)F_{aa} + W_{aa} \circ F + (W_{xa} \circ F)F_a + (W_{xx} \circ F)(F_a)^2$$

and proceed as in the previous Lemmas, using also Lemma 6.5. □

Lemma 7.6. — *Given $\delta > 0$, if V_z is sufficiently small, then*

$$|J_0| \cdot \left| \frac{H_{xxa}}{H_{xx}} \right| < \delta,$$

for all $x \in \gamma_0$ and $a \in J_0$.

Proof. — The proof is similar to the previous Lemmas, after writing

$$\begin{aligned} H_{xxa} &= (W_x \circ F)F_{xxa} + (W_{xxa} \circ F)(F_x)^2 \\ &\quad + (W_{xxx} \circ F)F_a(F_x)^2 + (W_{xx} \circ F)(2F_{xa}F_x + F_aF_{xx}) + (W_{xa} \circ F)F_{xx}. \end{aligned}$$ □

8. Expansion of regular branch compositions

We aim at proving the starting conditions for preimages of the central branch. To any preimage $\beta \in \mathcal{B}$ is assigned a sequence of regular branches

$$\{P_m : \pi_m \longrightarrow V_{z_a}\}_{m=1,\dots,n}$$

such that $B : \beta \to \gamma_0$ is written as $B = P_n \circ \cdots \circ P_1|\beta$. Each P_m in turn is written as $P_m = W_m \circ F|\pi_m$, where $W_m : \omega_m \to V_{z_a}$ is the transfer map of $\omega_m \in \mathcal{W}$, $\omega_m \neq \omega_0$, $m = 1, \dots, n$.

This Section is devoted to estimate the expansion of regular branches and their compositions. The ideas involved here are very similar to the concept of forward recovering, mentioned in Section 4. A kind of backward recovering appears in Section 9, when dealing with the first parameter derivative.

The first estimates give absolute lower bounds for derivatives of regular branches. We will see that expansion may be not sure in some cases. Next we show that every time there is a loss of derivative for some P_m there is an immediate recuperation for P_{m+1}.

From now on we choose a constant $\theta > 0$, and take V_z small so that Corollaries 3.2 and 3.3 are satisfied. This constant will be chosen sufficiently small, according to the needing of various Lemmas until the end of the work. It is implicitly assumed that assertions are valid for every $\beta \in \mathcal{B}$ and constants do not depend on β.

Accordingly to Section 2, the intervals $\omega_0, \omega_1, \ldots, \omega_n$ have extension domains $\widehat{\omega}_0$, $\widehat{\omega}_1, \ldots, \widehat{\omega}_n$ which are mapped onto V_{y_a}. Take ω_m, for some $m = 1, \ldots, n$. We will say that ω_m is subordinated to ω_0 if $\omega_m \subset \widehat{\omega}_0$, that ω_0 is subordinated to ω_m if $\omega_0 \subset \widehat{\omega}_m$ and that ω_0 and ω_m are independent otherwise. By construction it turns out that one and only one of these situations occurs. In particular, this implies by Corollary 3.2 that in the first and third cases

$$\operatorname{dist}(\omega_m, \omega_0) \geqslant \theta^{-1} |\omega_m|,$$

and in the second case

$$\operatorname{dist}(\omega_m, \omega_0) \geqslant \theta^{-1} |\omega_0|.$$

Lemma 8.1. — *If ω_m is subordinated to ω_0, or else ω_m and ω_0 are independent, then*

$$|P_{m,x}(x)| \geqslant C_0^{-1} \theta^{-1},$$

for all $x \in \pi_m$.

Proof. — Write $P_m = W_m \circ F | \pi_m$, hence $|P_{m,x}| = |W_{m,x} \circ F| \cdot |F_x|$. As $W_m : \omega_m \to V_{z_a}$ is extendible to $W_m : \widehat{\omega_m} \to V_{y_a}$, and V_z is chosen small, then by Lemma 3.1

$$|W_{m,x} \circ F| \simeq \frac{|V_{z_a}|}{|\omega_m|}.$$

But the hypotheses imply that

$$|F_x| \geqslant C_0^{-1} \sqrt{\operatorname{dist}(\omega_m, \omega_0)} \geqslant C_0^{-1} \theta^{-1/2} |\omega_m|^{1/2}.$$

Therefore

$$|P_{m,x}| \geqslant C_0^{-1} \theta^{-1/2} \frac{|V_{z_a}|}{|\omega_m|^{1/2}}.$$

On the other hand, if ε is small then $|\widehat{\omega}_m| < 2 |\operatorname{Im} F | V_{z_a}|$, hence

$$|\omega_m|^{1/2} \leqslant \sqrt{2} \theta^{1/2} |\operatorname{Im} F | V_{z_a}|^{1/2} \leqslant C_0 \theta^{1/2} |V_{z_a}|,$$

and the Lemma follows. □

Lemma 8.2. — *If ω_0 is subordinated to ω_m then*

$$|P_{m,x}(x)| \geqslant C_0^{-1} \theta^{-3/2} \frac{|\omega_0|^{1/2}}{|V_{z_a}|},$$

for all $x \in \pi_m$.

Proof. — As in the proof of the previous Lemma,

$$|P_{m,x}| \geqslant C_0^{-1} \frac{|V_{z_a}|}{|\omega_m|} \theta^{-1/2} |\omega_0|^{1/2}.$$

On the other hand, also as in the previous Lemma,

$$|\omega_m| < C_0 \theta |V_{z_a}|^2,$$

and the Lemma follows. □

These two Lemmas suggest that a bad derivative may occur if ω_0 is subordinated to ω_m, as depicted in Figure 4. The problem is overcome with forward recuperation,

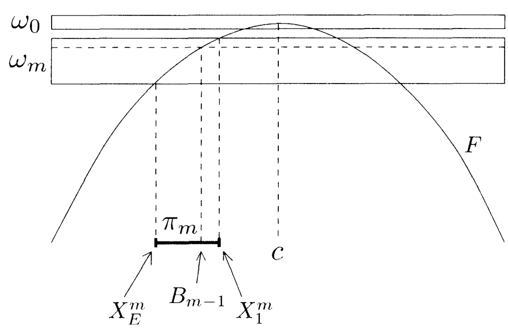


FIGURE 4. Possible bad derivative in the m-th iterate

which we describe now. Let

$$B_m \equiv P_m \circ \cdots \circ P_1 | \beta,$$

hence $B = B_n$. Let B_m denote the point $B_m(x)$, as we did before in other situations, when it is clear that there is no possibility of confusion. In this notation, $B_{m-1} \in \pi_m$, and $F(B_{m-1}) \in \omega_m$, see Figure 4. We call x_I^m and x_E^m, respectively, the innermost and the outermost boundary points of π_m, with respect to the critical point c.

Define

$$\rho_m = \rho_m(x) = \frac{|F(x_I^m) - F(B_{m-1})|}{|\omega_m|}.$$

By the small distortion property of $W_m : \omega_m \to V_{z_a}$, if $\rho_m < 1/3$ then

$$\frac{\operatorname{dist}(B_m, \partial V_{z_a})}{|V_{z_a}|} \simeq \rho_m.$$

We state two technical Lemmas to be used in the sequel.

Lemma 8.3. — $|P_{m,x}(x_E^m)| \gtrsim |P_{m,x}(y)|, \forall\, y \in \pi_m$.

Proof. — As W_m has small distortion, the distortion of P_m is mostly due to $F|\pi_m$. But $|F_x|$ increases as the distance from the critical point increases, since $F_x x \neq 0$. □

Lemma 8.4. — $$\frac{|B_{m-1}-c|}{|x_E^m-c|} > C_0^{-1}\rho_m^{1/2}.$$

Proof. — Since V_{z_a} is small and next to c the function F is nearly quadratic. □

As
$$\frac{1}{|\pi_m|}\int_{\pi_m}|P_{m,x}(y)|dy = \frac{|V_{z_a}|}{|\pi_m|}$$
and V_{z_a} is at least θ^{-1} times greater than π_m, according to Corollary 3.3, it is expected that $|P_{m,x}|$ is big for some points in π_m. In particular, by Lemma 8.3, $|P_{m,x}(x_E^m)| > \frac{3}{4}\theta^{-1}$. The following Lemma gives a lower bound for $|P_{m,x}(B_{m-1})|$ as a function of ρ_m.

Lemma 8.5. — $|P_{m,x}(B_{m-1})| > C_0^{-1}\theta^{-1}\rho_m^{1/2}$.

Proof. — By the remark above $|P_{m,x}(x_E^m)| > \frac{3}{4}\theta^{-1}$, hence
$$|P_{m,x}(B_{m-1})| = |P_{m,x}(x_E^m)|\cdot\frac{|P_{m,x}(B_{m-1})|}{|P_{m,x}(x_E^m)|} > \frac{3}{4}\theta^{-1}\frac{|W_{m,x}(F(B_{m-1}))|}{|W_{m,x}(F(x_E^m))|}\cdot\frac{|F_x(B_{m-1})|}{|F_x(x_E^m)|}.$$
But $W_{m,x}$ is almost constant in ω_m and
$$\frac{|F_x(B_{m-1})|}{|F_x(x_E^m)|}\simeq\frac{|B_{m-1}-c|}{|x_E^m-c|},$$
and the Lemma follows using Lemma 8.4. □

Corollary 8.6. — $|P_{n,x}(B_{n-1})| > C_0^{-1}\theta^{-1}$.

Proof. — As $F(B_{n-1})\in W_n^{-1}(\gamma_0)$ then $\rho_n\simeq 1/2$. □

Corollary 8.7. — *If $\rho_m \geqslant \frac{1}{3}$ then $|P_{m,x}(B_{m-1})| > C_0^{-1}\theta^{-1}$*

Proof. — Directly from Lemma 8.5. □

The following Lemma is the central assertion for forward recovering.

Lemma 8.8. — *If $\rho_m < \frac{1}{3}$ then*
$$|(P_{m+1}\circ P_m)_x(B_{m-1})| \geqslant C_0^{-1}\theta^{-2}.$$

Proof. — First notice that $B_m = P_m(B_{m-1})\in\pi_{m+1}$, and $|\pi_{m+1}| < \theta\operatorname{dist}(\pi_{m+1},\partial V_{z_a})$ by Corollary 3.3. Moreover $\operatorname{dist}(B_m,\partial V_{z_a})\simeq\rho_m|V_{z_a}|$, hence
$$\frac{\operatorname{dist}(y,\partial V_{z_a})}{|V_{z_a}|}\simeq\rho_m,$$
for all $y\in\pi_{m+1}$. This also implies $|\pi_{m+1}|\lesssim\theta\rho_m|V_{z_a}|$. On the other hand, since $\rho_m < 1/3$, $|\pi_{m+1}|\ll\operatorname{dist}(\pi_{m+1},\gamma_0)$, therefore the distortion of P_{m+1} must be small. Then
$$|P_{m+1,x}|\simeq\frac{|V_{z_a}|}{|\pi_{m+1}|}\gtrsim\theta^{-1}\rho_m^{-1}.$$

Combining with Lemma 8.5 we get

$$|(P_{m+1} \circ P_m)_x(B_{m-1})| > C_0^{-1} \rho_m^{-1/2} \theta^{-2},$$

proving the Lemma. $\square$

We obtain some useful Corollaries, but first we introduce the following notation. For $m_0 \leqslant m_1$, let

$$\Delta_{m_0,m_1} = \Delta_{m_0,m_1}(x) = \left| \prod_{m=m_0}^{m_1} P_{m,x}(B_{m-1}) \right|,$$

which is equal to $|(P_{m_1} \circ \cdot \circ P_{m_0})_x(B_{m_0-1})|$, by the Chain Rule. For example, with this notation,

$$|B_x| = \Delta_{1,n}.$$

Let also $\Delta_{m_0,m_1} \equiv 1$ if $m_0 > m_1$.

Corollary 8.9. — *$|B_x| > (C_0^{-1}\theta^{-1})^n \gg 1$. In fact, $\Delta_{m_0,n} > (C_0^{-1}\theta^{-1})^{n-m_0+1}$, for all $m_0 = 1, \ldots, n$.*

Proof. — We prove by (decreasing) induction on m_0, starting from $m_0 = n$. For $m_0 = n$ we use Corollary 8.6. Suppose now that $\Delta_{m,n} > (C_0^{-1}\theta^{-1})^{n-m+1}$, for all $m = m_0 + 1, \ldots, n$. We want to prove that $\Delta_{m_0,n} > (C_0^{-1}\theta^{-1})^{n-m_0+1}$. But if $\rho_{m_0} \geqslant \frac{1}{3}$ then

$$\Delta_{m_0,n} = |P_{m_0,x}(B_{m_0-1})| \cdot \Delta_{m_0+1,n} > (C_0^{-1}\theta^{-1})^{n-m_0+1},$$

by induction and Corollary 8.7. Otherwise $\rho_{m_0} < \frac{1}{3}$, then

$$\Delta_{m_0,n} = |(P_{m_0+1} \circ P_{m_0})_x(B_{m_0-1})| \cdot \Delta_{m_0+2,j} > (C_0^{-1}\theta^{-1})^{n-m_0+1},$$

by induction and Lemma 8.8. $\square$

Corollary 8.10. — *If $|P_{m_1,x}(B_{m_1-1})| > C_0^{-1}\theta^{-1}$ then $\Delta_{m_0,m_1} > (C_0^{-1}\theta^{-1})^{m_1-m_0+1}$, for all $m_0 = 1, \ldots, m_1$.*

Proof. — As in the proof of Corollary 8.9, but now induction starts at m_1. $\square$

Corollary 8.11. — *In general,*

$$\Delta_{m_0,m_1} > (C_0^{-1}\theta^{-1})^{m_1-m_0+1} \frac{|\omega_0|^{1/2}}{|V_{z_a}|},$$

for every $1 \leqslant m_0 \leqslant m_1 \leqslant n$.

Proof. — First we notice that

$$\frac{|\omega_0|^{1/2}}{|V_{z_a}|} < C_0 \theta^{1/2} \ll 1,$$

by Corollary 3.2 and Lemma 2.4, supposing also θ sufficiently small.

Suppose $m_0 < m_1$. If $\rho_{m_1-1} < 1/3$, then

$$|(P_{m_1} \circ P_{m_1-1})_x(B_{m_1-2})| > (C_0^{-1}\theta^{-1})^2,$$

by Lemma 8.8, and by induction, as in the preceding Corollaries,

$$\Delta_{m_0,m_1} > (C_0^{-1}\theta^{-1})^{m_1-m_0+1}.$$

Otherwise $\rho_{m_1-1} \geqslant 1/3$, implying, by Corollaries 8.7 and 8.10 that

$$\Delta_{m_0,m_1-1} > (C_0^{-1}\theta^{-1})^{m_1-m_0}.$$

Then we write

$$\Delta_{m_0,m_1} = |P_{m_1,x}(B_{m_1-1})| \cdot \Delta_{m_0,m_1-1}$$

and use Lemmas 8.1 and 8.2. These Lemmas are directly applied in the case $m_0 = m_1$. □

9. Parameter dependence of regular branches

As remarked in Section 2, for each $\beta \in \mathcal{B}$ the function $B : \beta \to \gamma_0$ is extendible to $B : \mathcal{U}(\beta) \to V_{z_a}$. By Corollary 3.4, as $|\gamma_0|/|V_{z_a}| < \eta$, provided V_z and ε are sufficiently small, we choose $\eta > 0$ so that, by Lemma 3.1, $\mathcal{U}(\beta)$ contains a $\theta^{-1}|\beta|$-neighborhood of β, for some small $\theta > 0$, where the derivative of B has small distortion. Moreover, $\mathcal{U}(\beta)$ is completely inside one of the connected components of $V_{z_a} \setminus \gamma_0$.

We have also defined the parameter interval

$$J(\beta) = \{a \in J_0 \ ; \ \operatorname{Im} H \cap \mathcal{U}(\beta) \neq \varnothing \text{ or } |\operatorname{Im} H| \geqslant \frac{1}{7}|V_{z_a}|\}.$$

Observe that, according to the notation of the previous Section, $\beta \subset \mathcal{U}(\beta) \subset \pi_1$, so that $\operatorname{Im} H \cap \mathcal{U}(\beta) \neq \varnothing$ implies $\operatorname{Im} H \cap \pi_1 \neq \varnothing$. On the other hand, $|\operatorname{Im} H| \geqslant \frac{1}{7}|V_{z_a}|$ implies $|\operatorname{Im} F \cap \omega_0| \geqslant \frac{1}{8}|\omega_0|$, by the small distortion of $W_0 : \omega_0 \to V_{z_a}$.

We now define $Y = Y(a, \beta)$, for $a \in J_0$, as the distance between $F(\beta)$ and $F(c)$ (see Figure 5). As $\beta \in \pi_1$ then $F(\beta) \in \omega_1$. We also define $Z = Z(a) = |\operatorname{Im} F|V_{z_a}|$, $X = X(a, \beta) = Z - Y$ and $\tau = X/Z$.

The underlining idea in this Section is to better control the derivative of $P_1 : \pi_1 \to V_{z_a}$. As $P_{1,x} = (W_{1,x} \circ F)F_x$, expansion depends on the relative position of $F(\beta)$ with respect to the critical value, controlled by τ.

Roughly speaking, we deal with the following situations. Fixing β and taking $a \in J(\beta)$ we may have $|\operatorname{Im} H \cap V_{z_a}| \geqslant \frac{1}{7}|V_{z_a}|$. In this case, since $|\operatorname{Im} F \cap \omega_0|$ is relatively large with respect to $|\omega_0|$ the derivative of F outside γ_0 is always bounded by something of the order of $|\omega_0|^{1/2}$, which will be enough to our purposes. Otherwise $a \in J(\beta)$ implies $\operatorname{Im} H \cap \mathcal{U}(\beta) \neq \varnothing$. In this case we may have τ small or not. If τ is small it means that β is near ∂V_{z_a}, and the derivative of F is not so small. But if τ is near 1 this means that β (and also π_1), is near the critical point. The consequence is that $\operatorname{Im} H$ occupies approximately one half of V_{z_a}, which in turn implies that $\operatorname{Im} F$

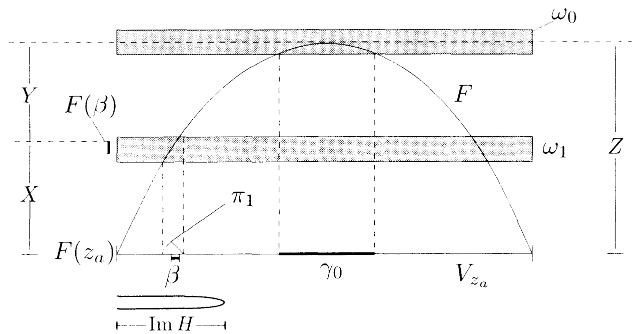


FIGURE 5. Placement of $F(\beta)$ with respect to the critical value

occupies approximately one half of ω_0. Once again derivatives outside γ_0 must be at least of the order of $|\omega_0|^{1/2}$.

The next two Lemmas quantify these arguments.

Lemma 9.1. — *If* $a \in J(\beta)$ *then* $|\gamma_0| > C_0^{-1}\tau^{1/2}|\omega_0|^{1/2}$.

Proof. — If $|\operatorname{Im} H| \geqslant \frac{1}{7}|V_{z_a}|$ then $|\operatorname{Im} F \cap \omega_0| \geqslant \frac{1}{8}|\omega_0|$, hence $|\gamma_0| > C_0^{-1}|\omega_0|^{1/2}$. Now it is enough to verify the inequality when $\operatorname{Im} H \cap \mathcal{U}(\beta) \neq \varnothing$ but $|\operatorname{Im} H| < \frac{1}{7}|V_{z_a}|$. By Lemma 2.4 and Corollary 3.2 $|\omega_1| < \theta X$, and also $\operatorname{dist}(\omega_1, F(z_a)) \simeq \tau|\operatorname{Im} F|V_{z_a}|$. As $|F_x| < C_0|V_{z_a}|$ in V_{z_a} then

(1) $$\operatorname{dist}(\pi_1, \partial V_{z_a}) \geqslant C_0^{-1}|V_{z_a}|^{-1}\tau|\operatorname{Im} F|V_{z_a}|.$$

The assumption $\operatorname{Im} H \cap \mathcal{U}(\beta) \neq \varnothing$ implies $|\operatorname{Im} H| \gtrsim \operatorname{dist}(\pi_1, \partial V_{z_a})$, hence by the small distortion of $W_0 : \omega_0 \to V_{z_a}$

$$|\gamma_0| \geqslant C_0^{-1}\left[\operatorname{dist}(\pi_1, \partial V_{z_a}) \cdot \frac{|\omega_0|}{|V_{z_a}|}\right]^{1/2}.$$

Using Equation (1) we obtain the Lemma, taking into account that

$$\frac{|\operatorname{Im} F|V_{z_a}|^{1/2}}{|V_{z_a}|} \geqslant C_0^{-1}. \qquad \square$$

Lemma 9.2. — *If* $\operatorname{dist}(\pi_1, \gamma_0) \geqslant \frac{1}{4}|V_{z_a}|$ *then* $|P_{1,x}| \geqslant C_0^{-1}\tau^{-1}\theta^{-1}$.

Proof. — We write $|P_{1,x}| = |W_{1,x} \circ F| \cdot |F_x|$. By small distortion properties, $|W_{1,x}| \simeq |V_{z_a}|/|\omega_1|$. Moreover

$$|\omega_1| < C_0\tau\theta|V_{z_a}|^2,$$

since $|\omega_1| \leqslant \theta X = \tau\theta Z$ and $Z = |\operatorname{Im} F|V_{z_a}| \leqslant C_0|V_{z_a}|^2$. On the other hand, $\operatorname{dist}(\pi_1, \gamma_0) \geqslant \frac{1}{4}|V_{z_a}|$ implies $|F_x| \geqslant C_0^{-1}|V_{z_a}|$, and the Lemma follows. $\square$

The goal of this Section is to show that

$$\frac{B_{m,a}}{B_{m,x}H_a}$$

is as small as desired, for all $a \in J(\beta)$ and $x \in \beta$, provided V_z and ε are sufficiently small. Here $B_m = P_m \circ P_{m-1} \circ \cdots \circ P_1$ and H_a is the mean value of $H(a,x)$, for $x \in \gamma_0$ and $a \in J_0$, based on the statements of Section 7. For $m = n$ this gives the first quotient of the starting conditions Preimages of the central branch. The cases $m < n$ will be used for the other quotients. We write

$$\frac{B_{m,a}}{B_{m,x}} = \sum_{t=1}^{m} \frac{P_{t,a} \circ B_{t-1}}{B_{t,x}},$$

therefore

$$\frac{B_{m,a}}{B_{m,x}H_a} = \frac{P_{1,a}}{P_{1,x}H_a} + \sum_{t=2}^{m} \frac{1}{B_{t-1,x}} \cdot \frac{P_{t,a} \circ B_{t-1}}{(P_{t,x} \circ B_{t-1})H_a}. \tag{2}$$

This last equation motivates the following Lemmas.

Lemma 9.3. — *For $m = 1, \ldots, n$,*

$$\left| \frac{P_{m,a} \circ B_{m-1}}{(P_{m,x} \circ B_{m-1})H_a} \right| < C_y \frac{|\omega_0|}{|V_{z_a}|} \cdot \frac{1}{|B_{m-1} - c|}.$$

Proof. — Write

$$\frac{P_{m,a} \circ B_{m-1}}{P_{m,x} \circ B_{m-1}} = \frac{F_a \circ B_{m-1}}{F_x \circ B_{m-1}} + \frac{1}{F_x \circ B_{m-1}} \cdot \frac{W_{m,a} \circ F \circ B_{m-1}}{W_{m,x} \circ F \circ B_{m-1}}.$$

We know that $|F_a| \leqslant C_0$, $|F_x \circ B_{m-1}| \geqslant C_0^{-1}|B_{m-1} - c|$ and $|W_{m,a}|/|W_{m,x}| \leqslant C_y$ (by Lemma 5.1). The Lemma follows using Lemma 7.1. □

Lemma 9.4. — *If* $\operatorname{dist}(\pi_1, \gamma_0) \geqslant \frac{1}{4}|V_{z_a}|$ *then*

$$\left| \frac{P_{1,a}}{P_{1,x}H_a} \right| < C_y\theta.$$

Proof. — The quotient is evaluated at $B_0 = B_0(x) = x$. By the hypothesis, $|B_0 - c| \gtrsim \frac{1}{4}|V_{z_a}|$. By Lemma 9.3,

$$\left| \frac{P_{1,a}}{P_{1,x}H_a} \right| < C_y \frac{|\omega_0|}{|V_{z_a}|^2} < C_y\theta.$$ □

Lemma 9.5. — *If* $\operatorname{dist}(\pi_1, \gamma_0) < \frac{1}{4}|V_{z_a}|$ *and $a \in J(\beta)$ then*

$$\left| \frac{P_{m,a} \circ B_{m-1}}{(P_{m,x} \circ B_{m-1})H_a} \right| < C_y \frac{|\gamma_0|}{|V_{z_a}|},$$

for all $m = 1, \ldots, n$.

Proof. — If $a \in J(\beta)$ then $|\operatorname{Im} H| \geqslant \frac{1}{7}|V_{z_a}|$ or $\operatorname{Im} H \cap \mathcal{U}(\beta) \neq \varnothing$. In the latter case, $\operatorname{Im} H \cap \pi_1 \neq \varnothing$ and, by the hypotheses, $|\operatorname{Im} H| \gtrsim \frac{3}{8}|V_{z_a}|$. In any case, $|\operatorname{Im} H| \geqslant \frac{1}{7}|V_{z_a}|$, therefore $|\operatorname{Im} F \cap \omega_0| > \frac{1}{8}|\omega_0|$. This implies $|B_{m-1} - c| > C_0^{-1}|\omega_0|^{1/2}$, for all $m = 1, \ldots, n$. By Lemma 9.3 the quotient of the statement is bounded by $C_y|\omega_0|^{1/2}/|V_{z_a}|$. By Lemma 9.1 and the hypothesis, implying τ bounded away from zero by C_0^{-1}, it follows

$$|\omega_0|^{1/2} < C_0\,|\gamma_0|$$

and the Lemma. □

The next Lemma is somehow analogous to the idea of backward recovering of Section 4.

Lemma 9.6. — *If $a \in J(\beta)$ and $\operatorname{dist}(\pi_1, \gamma_0) \geqslant \frac{1}{4}|V_{z_a}|$ then*

$$|P_{1,x}|^{-1} \cdot \left|\frac{P_{m,a} \circ B_{m-1}}{(P_{m,x} \circ B_{m-1})H_a}\right| < C_y\theta \frac{|\omega_0|^{1/2}}{|V_{z_a}|},$$

for all $m = 1, \ldots, n$.

Proof. — By Lemma 9.1, $|B_{m-1} - c| > C_0^{-1}\tau^{1/2}|\omega_0|^{1/2}$. Putting into Lemma 9.3 and using Lemma 9.2 the Lemma follows. Observe also that $\tau \leqslant 1$. □

Lemma 9.7. — *Given $\delta > 0$, if V_z and ε are sufficiently small then*

$$\left|\frac{B_{m,a}}{B_{m,x}H_a}\right| < \delta,$$

for all $\beta \in \mathcal{B}$, $x \in \beta$, $a \in J(\beta)$ and $m = 1, \ldots, n$, where $B = P_n \circ \cdots \circ P_1|\beta : \beta \to \gamma_0$ and $B_m = P_m \circ \cdots \circ P_1|\beta$. The value H_a indicates the mean value of $H(a, x)$ for $x \in \gamma_0$ and $a \in J_0$.

Proof. — We evaluate term by term the R.H.S of Equation (2) supposing always that $a \in J(\beta)$. We have to consider two separate cases: A) $\operatorname{dist}(\pi_1, \gamma_0) \geqslant \frac{1}{4}|V_{z_a}|$ and B) $\operatorname{dist}(\pi_1, \gamma_0) < \frac{1}{4}|V_{z_a}|$. The first term,

$$\frac{P_{1,a}}{P_{1,x}H_a},$$

is bounded by $C_y\theta$ if $\operatorname{dist}(\pi_1, \gamma_0) \geqslant \frac{1}{4}|V_{z_a}|$, by Lemma 9.4, and bounded by $C_y|\gamma_0|/|V_{z_a}|$ if $\operatorname{dist}(\pi_1, \gamma_0) < \frac{1}{4}|V_{z_a}|$, by Lemma 9.5. In both cases the first term is bounded by $\delta/2$, provided θ is small, and this is guaranteed if V_z is small enough. We are left with the remaining terms, from $t = 2$ to $t = m$.

In Case B, where $\operatorname{dist}(\pi_1, \gamma_0) < \frac{1}{4}|V_{z_a}|$, Lemma 9.5 implies

$$\left|\frac{P_{t,a}}{P_{t,x}H_a}\right| < C_y \frac{|\omega_0|^{1/2}}{|V_{z_a}|}.$$

On the other hand,

$$|B_{t-1,x}|^{-1} = \Delta_{1,t-1}^{-1} \leqslant (C_0\theta)^{t-1}\frac{|V_{z_a}|}{|\omega_0|^{1/2}},$$

by Corollary 8.11. Hence we are left with

$$C_y \sum_{t=2}^{m} (C_0\theta)^{t-1},$$

which is smaller than $\delta/2$ if θ is sufficiently small.

In Case A, where $\operatorname{dist}(\pi_1, \gamma_0) \geqslant \frac{1}{4}|V_{z_a}|$, we bound the t-th term by

$$\Delta_{2,t-1}^{-1}|P_{1,x}|^{-1} \cdot \left|\frac{P_{t,a} \circ B_{t-1}}{(P_{t,x} \circ B_{t-1})H_a}\right|,$$

which is smaller than $C_y\theta(C_0\theta)^{t-2}$, by Corollary 8.11 and Lemma 9.6. Therefore the sum is smaller than $\delta/2$ if θ is small enough. □

10. Other derivatives

We keep the same notation introduced in the preceding Sections. The goal is to bound derivatives of $P : \pi \to V_{z_a}$, for all $\pi \in \mathcal{P}$, and take their compositions to bound derivatives of $B : \beta \to \gamma_0$, for all $\beta \in \mathcal{B}$. This will complete the proof of the starting conditions Preimages of the central branch.

Lemma 10.1. — *If V_z is small then*

$$|\gamma_0| \cdot \left|\frac{P_{xx}}{(P_x)^2}\right| < \frac{|\gamma_0|}{|V_{z_a}|} + C_0|P_x|^{-1}.$$

Proof. — Write $P = W \circ F$ and

$$\frac{P_{xx}}{(P_x)^2} = \frac{F_{xx}}{P_x F_x} + \frac{W_{xx}}{(W_x)^2}.$$

We have

$$|\gamma_0| \cdot \left|\frac{W_{xx}}{(W_x)^2}\right| \ll \frac{|\gamma_0|}{|V_{z_a}|},$$

for V_z sufficiently small, by Lemma 6.2. This controls the second term. For the first, we have $|F_{xx}| < C_0$ and

$$\frac{|\gamma_0|}{|F_x|} < C_0. \qquad \square$$

Lemma 10.2. — *If V_z is small then*

$$|\gamma_0| \cdot \left|\frac{B_{m,xx}}{(B_{m,x})^2}\right| < \frac{|\gamma_0|}{|V_{z_a}|} + C_0 \sum_{t=1}^{m} \Delta_{t,m}^{-1},$$

for all $m = 1, \ldots, n$.

Proof. — Write

$$\frac{B_{m,xx}}{(B_{m,x})^2} = \sum_{t=1}^{m} \frac{1}{(P_m \circ \cdots \circ P_{t+1})_x \circ B_t} \cdot \frac{P_{t,xx} \circ B_{t-1}}{(P_{t,x} \circ B_{t-1})^2}.$$

Then

$$|\gamma_0| \cdot \left| \frac{B_{m,xx}}{(B_{m,x})^2} \right| \leqslant \sum_{t=1}^{m} \Delta_{t+1,m}^{-1} \left(\frac{|\gamma_0|}{|V_{z_a}|} + C_0 |P_{t,x} \circ B_{t-1}|^{-1} \right),$$

by Lemma 10.1. In the statement we separate the term $\Delta_{m+1,m}^{-1} \equiv 1$. □

Lemma 10.3. — *Given $\delta > 0$, if V_z is sufficiently small then*

$$|\gamma_0| \cdot \left| \frac{B_{xx}}{(B_x)^2} \right| < \delta,$$

for all $x \in \beta$, $\beta \in \mathcal{B}$ and $a \in J_0$.

Proof. — Put $m = n$ in Lemma 10.2. Then

$$\Delta_{t+1,n}^{-1} \leqslant (C_0\theta)^{n-t}, \quad \Delta_{t,n}^{-1} \leqslant (C_0\theta)^{n-t+1},$$

by Corollary 8.9. As V_z small implies θ small and $|\gamma_0|/|V_{z_a}|$ small, then the Lemma follows. □

Lemma 10.4. — *If V_z is sufficiently small then*

$$|\gamma_0|^2 \cdot \left| \frac{P_{xxx}}{(P_x)^3} \right| < \frac{|\gamma_0|^2}{|V_{z_a}|^2} + |P_x|^{-1},$$

for all $x \in \pi$, $\pi \in \mathcal{P}$ and $a \in J_0$.

Proof. — Write $P = W \circ F$ and

$$\frac{P_{xxx}}{(P_x)^3} = \frac{F_{xxx}}{(F_x)^2 W_x P_x} + \frac{W_{xxx}}{(W_x)^3} + 3\frac{W_{xx}}{(W_x)^2} \cdot \frac{F_{xx}}{F_x P_x},$$

where W_x, W_{xx}, etc, mean $W_x \circ F$, $W_{xx} \circ F$, etc. The second term, multiplied by $|\gamma_0|^2$, is smaller than $|\gamma_0|^2/|V_{z_a}|^2$, by Lemma 6.3, if V_z is small. For the first term, $|\gamma_0|^2/(F_x)^2 < C_0$ and $|W_x| \gg C_0$ (by the choice of V_z) implies that it is bounded by $\frac{1}{2}|P_x|^{-1}$. For the third term, we have $|\gamma_0|/|F_x| < C_0$ and

$$|\gamma_0| \cdot \left| \frac{W_{xx}}{(W_x)^2} \right| \ll \frac{|\gamma_0|}{|V_{z_a}|} \ll C_0,$$

by Lemma 6.2, hence it is bounded by $\frac{1}{2}|P_x|^{-1}$, and the Lemma follows. □

Lemma 10.5. — *Given $\delta > 0$, if V_z is sufficiently small then*

$$|\gamma_0|^2 \cdot \left| \frac{B_{xxx}}{(B_x)^3} \right| < \delta,$$

for all $x \in \beta$, $\beta \in \mathcal{B}$ and $a \in J_0$.

Proof. — We write
$$\frac{B_{xxx}}{(B_x)^3} = S_1 + 3S_2,$$
where
$$S_1 = \sum_{t=1}^{n} \frac{P_{t,xxx}}{(P_{n,x}\dots P_{t+1,x})^2 (P_{t,x})^3}$$
and
$$S_2 = \sum_{t=2}^{n} \frac{P_{t,xx}}{P_{n,x}\dots P_{t+1,x}(P_{t,x})^2} \cdot \frac{1}{P_{n,x}\dots P_{t,x}} \frac{B_{t-1,xx}}{(B_{t-1,x})^2}.$$
For simplicity, we are omitting arguments, writing $P_{t,x}$ instead of $P_{t,x} \circ B_{t-1}$, etc. Using Lemma 10.4 and $\Delta_{n+1,n} \equiv 1$ we obtain
$$|\gamma_0|^2 |S_1| \leqslant \frac{|\gamma_0|^2}{|V_{z_a}|^2} + |P_{n,x}|^{-1} + \sum_{t=1}^{n-1} \Delta_{t+1,n}^{-2} + \sum_{t=1}^{n-1} \Delta_{t+1,n}^{-1} \Delta_{t,n}^{-1},$$
which is smaller than $\delta/2$ if θ is sufficiently small, by Corollary 8.9. On the other hand $|\gamma_0|^2|S_2|$ is bounded by
$$\sum_{t=2}^{n} \Delta_{t+1,n}^{-1} \Delta_{t,n}^{-1} \left(1 + C_0 |P_{t,x}|^{-1}\right) \cdot \left(1 + C_0 \sum_{s=1}^{t-1} \Delta_{s,t-1}^{-1}\right),$$
using Lemmas 10.1 and 10.2 and $|\gamma_0| \ll |V_{z_a}|$. It is straightforward to see that this sum is smaller than $\delta/6$, if θ is small, using Corollary 8.9. □

Lemma 10.6. — *If V_z is sufficiently small then*
$$|\gamma_0| \cdot \left| \frac{P_{xa}}{(P_x)^2 H_a} \right| < |P_x|^{-1} + \frac{|\omega_0|}{|V_{z_a}|^2},$$
for all $x \in \pi$, $\pi \in \mathcal{P}$ and $a \in J_0$.

Proof. — Write $P = W \circ F$ and
$$\frac{P_{xa}}{(P_x)^2} = \frac{F_{xa}}{F_x P_x} + \frac{1}{F_x} \frac{W_{xa}}{(W_x)^2} + \frac{W_{xx}}{(W_x)^2} \cdot \frac{F_a}{F_x}.$$
We analyze each term multiplied by $|\gamma_0|/|H_a|$, which is smaller than $C_0|\gamma_0|/|W_{0,x}|$. As $|\gamma_0|/|F_x| < C_0$, the first term can be bounded by
$$\frac{C_0}{|W_{0,x}|} \cdot |P_x|^{-1} < |P_x|^{-1},$$
if V_z is small. For the second term we still use that
$$|V_{z_a}| \frac{W_{xa}}{(W_x)^2}$$
is much smaller than C_0, by Lemma 6.4, and $|W_{0,x}| \simeq |V_{z_a}|/|\omega_0|$. The same in the third term, but now using Lemma 6.2. □

Lemma 10.7. — *If V_z is sufficiently small then*

$$|\gamma_0| \cdot \left|\frac{B_{m,xa}}{(B_{m,x})^2 H_a}\right| < m\Delta_{1,m}^{-1} + C_0 \left(\frac{|\omega_0|^{1/2}}{|V_{z_a}|} + \sum_{t=2}^{m} \Delta_{t,m}^{-1}\right),$$

for all $m = 1, \ldots, n$, $x \in \beta$, $\beta \in \mathcal{B}$ and $a \in J(\beta)$.

Proof. — Write

$$\frac{B_{m,xa}}{(B_{m,x})^2} = S_1 + S_2,$$

where

$$S_1 = \sum_{t=1}^{m} \frac{P_{t,xa}}{P_{t,x} B_{m,x}}$$

and

$$S_2 = \sum_{t=2}^{m} \frac{1}{P_{m,x} \ldots P_{t+1,x}} \cdot \frac{P_{t,xx}}{(P_{t,x})^2} \cdot \frac{B_{t-1,a}}{B_{t-1,x}}.$$

Then, by Lemma 10.6,

$$\frac{|\gamma_0|}{|H_a|}|S_1| \leqslant \sum_{t=1}^{m} \frac{P_{t,x}}{B_{m,x}} \cdot \left(|P_{t,x}|^{-1} + \frac{|\omega_0|}{|V_{z_a}|^2},\right)$$

which is smaller than

$$m\Delta_{1,m}^{-1} + \frac{|\omega_0|}{|V_{z_a}|^2} \sum_{t=1}^{m} \Delta_{t+1,m}^{-1} \Delta_{1,t-1}^{-1}.$$

Using Corollary 8.11, we simplify

$$\frac{|\omega_0|}{|V_{z_a}|^2} \sum_{t=1}^{m} \Delta_{t+1,m}^{-1} \Delta_{1,t-1}^{-1} \leqslant \frac{|\omega_0|^{1/2}}{|V_{z_a}|} + \sum_{t=2}^{m} \Delta_{t,m}^{-1}.$$

On the other hand, as in Lemma 10.2,

$$\frac{|\gamma_0|}{|H_a|}|S_2| \leqslant \frac{|\gamma_0|}{|V_{z_a}|} + C_0 \sum_{t=2}^{m} \Delta_{t,m}^{-1},$$

using Lemmas 10.1 and 9.7. □

Lemma 10.8. — *Given $\delta > 0$, if V_z is sufficiently small then*

$$|\gamma_0| \cdot \left|\frac{B_{xa}}{(B_x)^2 H_a}\right| < \delta,$$

for all $x \in \beta$, $\beta \in \mathcal{B}$ and $a \in J(\beta)$.

Proof. — It is enough to apply Corollary 8.9 in Lemma 10.7, with θ small. □

Lemma 10.9. — *If V_z is sufficiently small then*

$$|\gamma_0|^2 \cdot \left|\frac{P_{xxa}}{(P_x)^3 H_a}\right| < \frac{|\omega_0|}{|V_{z_a}|^2} + |P_x|^{-1},$$

for all $x \in \pi$, $\pi \in \mathcal{P}$ and $a \in J_0$.

Proof. — We have

$$\frac{P_{xxa}}{(P_x)^3} = Q_1 + Q_2 + Q_3 + Q_4 + 4Q_5 + 2Q_6,$$

where

$$Q_1 = \frac{F_{xxa}}{P_x W_x (F_x)^2}, \quad Q_2 = \frac{W_{xxa}}{F_x (W_x)^3}, \quad Q_3 = \frac{W_{xxx}}{(W_x)^3} \cdot \frac{F_a}{F_x},$$
$$Q_4 = \frac{1}{P_x} \frac{W_{xa}}{(W_x)^2} \frac{F_{xx}}{(F_x)^2}, \quad Q_5 = \frac{1}{W_x} \frac{W_{xx}}{(W_x)^2} \frac{F_{xa}}{(F_x)^2}, \quad Q_6 = \frac{1}{W_x} \frac{W_{xx}}{(W_x)^2} \frac{F_{xx}}{(F_x)^2} \frac{F_a}{F_x}.$$

We multiply each of these terms by $|\gamma_0|^2/|H_a|$, which is smaller than $C_0|\gamma_0|^2|\omega_0|/|V_{z_a}|$, by Lemma 7.1, using then the following estimates, which are valid for V_z sufficiently small: $|F_{xxa}|$, $|F_{xx}|$, $|F_a|$, $|F_{xa}| < C_0$, $|\gamma_0| < C_0|F_x|$, $C_0|W_x|^{-1} < 1$, $C_0|\omega_0| < |V_{z_a}|$, $C_0|\gamma_0| < |V_{z_a}|$ and Lemmas 6.2, 6.4, 6.6, 6.3, with $\delta = C_0^{-1}$ or $\delta = 1$. □

Lemma 10.10. — *Given $\delta > 0$, if V_z is sufficiently small then*

$$|\gamma_0|^2 \cdot \left| \frac{B_{xxa}}{(B_x)^3 H_a} \right| < \delta,$$

for all $x \in \beta$, $\beta \in \mathcal{B}$ and $a \in J(\beta)$.

Proof. — Write

$$\frac{|\gamma_0|^2}{H_a} \frac{B_{xxa}}{(B_x)^3} \leqslant |S_1| + |S_2| + |S_3| + 4|S_4| + 2|S_5|,$$

where

$$S_1 = \sum_{t=1}^{n} \Delta_{t+1,n}^{-2} \Delta_{1,t-1}^{-1} \cdot |\gamma_0|^2 \frac{P_{t,xxa}}{(P_{t,x})^3 H_a},$$
$$S_2 = \sum_{t=2}^{n} \Delta_{t+1,n}^{-2} \cdot |\gamma_0|^2 \frac{P_{t,xxx}}{(P_{t,x})^3} \cdot \frac{B_{t-1,a}}{B_{t-1,x} H_a},$$
$$S_3 = \sum_{t=2}^{n} \Delta_{1,n}^{-1} \Delta_{t+1,n}^{-1} \cdot |\gamma_0| \frac{P_{t,xa}}{(P_{t,x})^2 H_a} \cdot |\gamma_0| \frac{B_{t-1,xx}}{(B_{t-1,x})^2},$$
$$S_4 = \sum_{t=2}^{n} \Delta_{t+1,n}^{-1} \Delta_{t,n}^{-1} \cdot |\gamma_0| \frac{P_{t,xx}}{(P_{t,x})^2} \cdot |\gamma_0| \frac{B_{t-1,xa}}{(B_{t-1,x})^2 H_a},$$
$$S_5 = \sum_{t=2}^{n} \Delta_{t+1,n}^{-1} \Delta_{t,n}^{-1} \cdot |\gamma_0| \frac{P_{t,xx}}{(P_{t,x})^2} \cdot |\gamma_0| \frac{B_{t-1,xx}}{(B_{t-1,x})^2} \cdot \frac{B_{t-1,a}}{B_{t-1,x} H_a}.$$

By Lemma 10.9,

$$|S_1| \leqslant \sum_{t=1}^{n} \Delta_{t+1,n}^{-2} \Delta_{1,t-1}^{-1} \left(\frac{|\omega_0|}{|V_{z_a}|^2} + |P_{t,x}|^{-1} \right),$$

which is smaller than $\delta/5$ if θ is small, by Corollaries 8.9 and 8.11. With the other sums we proceed in the same way, using the same Corollaries and also Lemmas 9.7, 10.2, 10.7, 10.4, 10.6 and 10.1. □

Lemma 10.11. — *If V_z is sufficiently small then*

$$|\gamma_0| \cdot \left| \frac{P_{aa}}{(P_x)^2 (H_a)^2} \right| < C_0 \frac{|\omega_0|^2}{|V_{z_a}|^2} |P_x|^{-1} + \frac{|\omega_0|^2}{|V_{z_a}|^3} |x - c|^{-1},$$

for all $x \in \pi$, $\pi \in \mathcal{P}$ and $a \in J_0$.

Proof. — Write

$$\frac{P_{aa}}{(P_x)^2} = \frac{F_{aa}}{P_x F_x} + \frac{W_{aa}}{(F_x)^2 (W_x)^2} + \frac{2 F_a W_{xa}}{(F_x)^2 (W_x)^2} + \left(\frac{F_a}{F_x}\right)^2 \frac{W_{xx}}{(W_x)^2}.$$

Then we use $|\gamma_0| < C_0 |F_x|$, $|F_x|^{-1} < C_0 |x - c|^{-1}$, $|H_a|^{-1} < C_0 |\omega_0| / |V_{z_a}|$ and Lemmas 6.5, 6.4 and 6.2, with $\delta = C_0^{-1}$. □

Lemma 10.12. — *Given $\delta > 0$, if V_z is sufficiently small then*

$$|\gamma_0| \cdot \left| \frac{B_{aa}}{(B_x)^2 (H_a)^2} \right| < \delta,$$

for all $x \in \beta$, $\beta \in \mathcal{B}$ and $a \in J(\beta)$.

Proof. — Write

$$\frac{|\gamma_0|^2}{H_a^2} \frac{B_{aa}}{(B_x)^2} \leqslant |S_1| + 2|S_2| + |S_3|,$$

where

$$S_1 = \sum_{t=1}^{n} \Delta_{1,t-1}^{-2} \Delta_{t+1,n}^{-1} \cdot |\gamma_0| \frac{P_{t,aa}}{(P_{t,x})^2 (H_a)^2},$$

$$S_2 = \sum_{t=2}^{n} \Delta_{t+1,n}^{-1} \Delta_{1,t-1}^{-1} \cdot |\gamma_0| \frac{P_{t,xa}}{(P_{t,x})^2 H_a} \cdot \frac{B_{t-1,a}}{B_{t-1,x} H_a},$$

$$S_3 = \sum_{t=2}^{n} \Delta_{t+1,n}^{-1} \cdot |\gamma_0| \frac{P_{t,xx}}{(P_{t,x})^2} \cdot \left(\frac{B_{t-1,a}}{B_{t-1,x} H_a} \right)^2.$$

Using Lemma 10.6, Lemma 9.7 (with $\delta = 1$) and Corollaries 8.9 and 8.11 we get

$$|S_2| \leqslant (n-1)(C_0\theta)^n + \frac{|\omega_0|^{1/2}}{|V_{z_a}|} (n-1)(C_0\theta)^{n-1},$$

which is smaller than $\delta/6$ if θ is small. It is also easy to see that $|S_3|$ is smaller than $\delta/3$ if θ is small, using Lemma 10.1 and Corollary 8.9. The difficult part is contained in S_1. By Lemma 10.11 we have

$$|S_1| \leqslant \sum_{t=1}^{n} \Delta_{1,t-1}^{-2} \Delta_{t+1,n}^{-1} \left(C_0 \frac{|\omega_0|^2}{|V_{z_a}|^2} |P_{t,x}|^{-1} + \frac{|\omega_0|^2}{|V_{z_a}|^3} |B_{t-1} - c|^{-1} \right).$$

If we separate into two sums, the first one is bounded by

$$\frac{|\omega_0|^2}{|V_{z_a}|^2} \Delta_{1,n}^{-1} \sum_{t=1}^{n} \Delta_{1,t-1}^{-1},$$

where it is implicit that $\Delta_{1,0} \equiv 1$. By Corollary 8.11, $\sum_{t=1}^{n} \Delta_{1,t-1}^{-1} \leqslant 2|V_{z_a}|/|\omega_0|^{1/2}$, so that the first sum is bounded by

$$|\omega_0| \cdot \frac{|\omega_0|^{1/2}}{|V_{z_a}|}(C_0\theta)^n,$$

which is smaller than $\delta/6$ if θ is small. For the second sum,

$$S = \frac{|\omega_0|^2}{|V_{z_a}|^3} \sum_{t=1}^{n} \Delta_{1,t-1}^{-2} \Delta_{t+1,n}^{-1} \cdot |B_{t-1} - c|^{-1},$$

we use the ideas of Section 9.

We have two cases: A) $\operatorname{dist}(\pi_1, \gamma_0) \geqslant \frac{1}{4}|V_{z_a}|$ and B) $\operatorname{dist}(\pi_1, \gamma_0) < \frac{1}{4}|V_{z_a}|$. In Case B, $|\gamma_0| > C_0^{-1}|\omega_0|^{1/2}$ (see Lemma 9.5, for example), which implies, by Corollaries 8.9 and 8.11,

$$S \leqslant C_0 \frac{|\omega_0|}{|V_{z_a}|^2} < \frac{\delta}{6},$$

if V_z is sufficiently small. In Case A we have

$$|B_{t-1} - c|^{-1} \leqslant 2|\gamma_0|^{-1} < C_0 \tau^{-1/2} |\omega_0|^{-1/2},$$

for all $t = 1, \ldots, n$, by Lemma 9.1, $|B_0 - c|^{-1} < 5|V_{z_a}|^{-1}$ by hypothesis and $|P_{1,x}|^{-1}$, by Lemma 9.2, hence

$$S \leqslant \frac{|\omega_0|^2}{|V_{z_a}|^3} \left\{ C_0 \tau^{1/2} \theta |\omega_0|^{-1/2} \sum_{t=2}^{n} \Delta_{1,t-1}^{-1} \Delta_{2,t-1}^{-1} \Delta_{t+1,n}^{-1} + \Delta_{2,n}^{-1} \cdot 5|V_{z_a}|^{-1} \right\},$$

which is smaller than

$$5 \frac{|\omega_0|^2}{|V_{z_a}|^4} \Delta_{2,n}^{-1} + C_0 \frac{|\omega_0|^{1/2}}{|V_{z_a}|} \theta \sum_{t=2}^{n} (C_0\theta)^{n-t},$$

by Corollaries 8.9 and 8.11. If V_z is small the Lemma is proved. □

A. Appendix

As remarked at the end of Section 2, Theorem 2.5 is proved in [**2**] assuming C^∞ differentiability. This hypothesis is used only for estimates of derivatives near saddle-node bifurcations, where a map is considered as a time-one map of a flow. Here we are able to reduce the needed differentiability to 3, obtaining the same bounds (Lemmas S.7 and S.8 of [**2**]) without any embedding into a flow. In addition, as the arguments are direct, they allow much more control on constants.

The proof of Theorem 2 is made in [**2**] by induction, starting from the map Φ_0 defined in Section 2. The central interval γ_0 together with the preimages of the central branch β belonging to the collection $\mathcal{B}_0$ form the set of connected components of the domain of Φ_0, contained in $\gamma_{-1} \equiv V_{z_a}$.

The central branch $H_0 = \Phi_0|\gamma_0$ is unimodal and $H_0(\partial\gamma_0) \subset \partial\gamma_{-1}$. We also have the diffeomorphic branches $B = \Phi_0|\beta : \beta \to \gamma_0$. The map Φ_{n+1}, for $n \geqslant 0$, is defined by induction with domain in γ_n, with a central branch $H_{n+1} = \Phi_{n+1}|\gamma_{n+1} : \gamma_{n+1} \to \gamma_n$, $H_{n+1}(\partial\gamma_{n+1}) \subset \partial\gamma_n$ and with diffeomorphic branches $B = \Phi_{n+1}|\beta : \beta \to \gamma_{n+1}$, for β in the collection $\mathcal{B}_{n+1}$. The map H_{n+1} is the critical component of the Φ_n-first entry map into γ_n *after escaping* from this same γ_n (and not the Φ_n-first return map to γ_n, as usual). The maps $B : \beta \to \gamma_{n+1}$ are the branches of the Φ_n-first entry map into γ_{n+1}.

At all stages of the induction, the maps Φ_n are shown to satisfy the same three sets of conditions Geometry, Central Branch and Preimages of the Central Branch of Section 2, with small and uniform constants $\eta > 0$, $\delta_0 > 0$ and $\delta_1 > 0$. One of the main steps in the proof is the analysis of H_n-iterates near the creation of a saddle-node fixed point for H_n. In [**2**], this analysis is resumed in Lemmas S.7 and S.8.

The function H_n is a two-variable function $H_n = H_n(a, x)$, defined for $x \in \gamma_n = \gamma_{n,a}$ and $a \in J$, where J is some interval. As a varies along J, $H_n(a, c)$ crosses $\gamma_{n-1} = \gamma_{n-1,a}$. For simplicity we assume $c = 0$ and $H_n(0, 0) = 0$. The starting conditions named Central Branch imply that there are non-zero constants S_n and V_n such that

$$1 - 2\delta_0 \leqslant \frac{H_{n,xx}(a, x)}{2S_n} \leqslant 1 + 2\delta_0, \quad 1 - 2\delta_0 \leqslant \frac{H_{n,a}(a, x)}{V_n} \leqslant 1 + 2\delta_0,$$

for all $x \in \gamma_{n,a}$ and $a \in J$, if δ_0 is sufficiently small.

The sign of $S_n \cdot V_n$ is always the same as the sign of $S_0 \cdot V_0$, which we suppose to be negative, without loss of generality. If we do a linear coordinate change $x \mapsto -S_n x$, $a \mapsto -S_n V_n a$ we normalize H_n so that

$$|H_{n,xx}(a, x) + 2| \leqslant 4\delta_0, \quad |H_{n,a}(a, x) - 1| \leqslant 2\delta_0.$$

The starting conditions Central Branch are kept unaffected by linear changes of coordinates. Integrating these two last inequalities we have

$$|H_{n,x}(a, x) + 2x| \leqslant 4\delta_0|x|,$$

$$|H_n(a, x) - (H_n(a, 0) - x^2)| \leqslant 2\delta_0 x^2$$

and

$$|H_n(a, 0) - a| \leqslant 2\delta_0|a|.$$

In fact, we are only concerned here with negative values of the parameter, where the saddle-node appears. Let a_s be the least (and unique) value for which there is a (unique) solution for the equation

$$H_n(a, x) = x$$

and let x_s be such that $H_n(a_s, x_s)$.

By solving the equation $H_{n,x}(x_s) = 1$ we get

$$-\frac{1}{2}\frac{1}{1 - 2\delta_0} \leqslant x_s \leqslant -\frac{1}{2}\frac{1}{1 + 2\delta_0}.$$

On the other hand, $a_s \in [a_s^1, a_s^2]$, where a_s^1 and a_s^2 are, respectively, the first parameter values for which $(1-2\delta_0)(a-x^2) = x$ and $(1+2\delta_0)(a-x^2) = x$ have a solution. Hence

$$-\frac{1}{4}\frac{1}{(1-2\delta_0)^2} \leqslant a_s \leqslant -\frac{1}{4}\frac{1}{(1+2\delta_0)^2}.$$

These values are very near $a = -\frac{1}{4}$ and $x = -\frac{1}{2}$, which are the bifurcation values for $(a,x) \mapsto a - x^2$. Now we normalize H_n again by linear changes of coordinates both in a and x so that $a_s = -\frac{1}{4}$ and $x_s = -\frac{1}{2}$. The values of $H_{n,a}$ and $H_{n,xx}$ do change, but are still very near 1 and -2, if δ_0 is sufficiently small.

For the sake of simplicity, we write $H = H_n$, in these coordinates. For every such H the starting conditions Central Branch are satisfied, H_{xx} is near -2, H_a is near 1, and the values of the saddle-node bifurcation are given by $(a_s, x_s) = (-\frac{1}{4}, -\frac{1}{2})$. The constant δ_0 regulates the proximity to the function $(a,x) \mapsto a - x^2$. We call $\mathcal{H} = \mathcal{H}_{\delta_0}$ the set of functions satisfying these conditions. Since here we are only interested in a bounded region of the plane and the parameter space near the saddle-node bifurcation, we can fix the domain of each $H \in \mathcal{H}$ as

$$\{(a,x)\ ;\ (a,x) \in [-10,10] \times [-10,10]\}.$$

Every constant appearing in the estimates will be uniform among the functions $H \in \mathcal{H}$, provided δ_0 is sufficiently small.

Let $a_0 < a_s$ be such that $|H(a,0)| \geqslant 2$ for every $a \leqslant a_0$. This is the lowest parameter value we are interested in, since all iterates outside the critical region $H^{-1}([H^2(0), H(0)])$ have some expansion (approximately greater or equal than 4) and can be treated by other methods. For $H(a,x) = a - x^2$, we have $a_0 = -2$ and in the remaining cases there is an error of the order of δ_0 about -2.

For $a > a_0$ we are concerned with iterates $x, Hx, \ldots, H^j x$, where $|x| < 2$, $x \notin H^{-1}([H^2(0), H(0)])$ and $|H^j(x)| \geqslant 2$. For each a we associate the number $l = l(a)$ which gives the maximal j. In other words,

$$l = l(a) = \min\{j \geqslant 1\ ;\ |H^j(H(0))| \geqslant 2\}.$$

Now for $x \in (-2,2) \setminus H^{-1}([H^2(0), H(0)])$ and j as above we denote

$$F_S(a,x) = H^j(a,x).$$

We aim at proving the following Lemma (corresponding to Lemma S.8 in [**2**]).

Lemma A.1

There is $C > 0$ such that for all $H \in \mathcal{H}$, $x \in (-2,2) \setminus H^{-1}([H^2(0), H(0)])$ and $a_0 < a < a_s = -\frac{1}{4}$, we have

$$|F_{S,x}|^{-1},\ \left|\frac{F_{S,xx}}{(F_{S,x})^2}\right|,\ \left|\frac{F_{S,xxx}}{(F_{S,x})^3}\right| \leqslant C,$$

$$|F_{S,x}|,\ \left|\frac{F_{S,xx}}{F_{S,x}}\right| \leqslant Cl^2,$$

$$|F_{S,a}|,\ \left|\frac{F_{S,xa}}{F_{S,x}}\right|,\ \left|\frac{F_{S,xxa}}{(F_{S,x})^2}\right| \leqslant Cl^3,$$
$$|F_{S,aa}| \leqslant Cl^6,$$

where $l = l(a)$ as above. Moreover, if $x \in [H^2(0), H(0)]$ then

$$|F_{S,a}| \geqslant C^{-1}l^3$$

and

$$|F_{S,x}|,\ |F_{S,xx}| \leqslant C.$$

This Lemma will be proved in the following way. We will fix $a_1 < a_s$ and define

$$l_0 = \max_{H\in\mathcal{H}} \max_{a_0\leqslant a\leqslant a_1} l(a).$$

Also we let $x_l < -\frac{1}{2} < x_r$ be such that $x_l > -2$, $x_r < H(0)$ and some conditions stated below are satisfied. The order of choice is this one: first x_l and x_r, then a_1 and finally δ_0. If δ_0 is sufficiently small then the constant C will be uniform for all $H \in \mathcal{H}$.

Iterates done for $a_0 \leqslant a \leqslant a_1$ and outside $[x_l, x_r]$ for $a > a_1$ are in (uniformly) finite number, so that they contribute only with constants to the Lemma. The main problem lies on the "unbounded" part $[a_1, a_s] \times [x_l, x_r]$, which is solved if we prove the following Lemma (Lemma S.7 in [**2**]).

Lemma A.2. — *There is $C > 0$ such that for $a > a_1$*

(1) $C^{-1} \leqslant |H^j_x| \leqslant Cl^2$, *for all $x \in [x_l, x_r]$;*
(2) $C^{-1} \leqslant |H^j_x| \leqslant C$, *for all $x \in [Hx_r, x_r]$;*
(3) $|H^j_a| \leqslant Cl^3$, *for all $x \in [x_l, x_r]$;*
(4) $|H^j_a| \geqslant C^{-1}l^3$, *for all $x \in [Hx_r, x_r]$;*
(5) $|H^j_{aa}| \leqslant Cl^6$, *for all $x \in [x_l, x_r]$;*
(6) $|H^j_{xx}| \leqslant C|H^j_x|^2$, *for all $x \in [x_l, x_r]$;*
(7) $|H^j_{xxx}| \leqslant C|H^j_x|^3$, *for all $x \in [x_l, x_r]$;*
(8) $|H^j_{xa}| \leqslant C|H^j_x|l^3$, *for all $x \in [x_l, x_r]$;*
(9) $|H^j_{xxa}| \leqslant C|H^j_x|^2l^3$, *for all $x \in [x_l, x_r]$.*

Establishing the relation between a and l is one of the main steps in the proof of Lemma A.2. We change coordinates again by $x \mapsto x + \frac{1}{2}$ and $a \mapsto a + \frac{1}{4}$, so that now the saddle-node occurs for $(a, x) = (0, 0)$. We also regard a_0, a_1, x_l, x_r in the new coordinates. Then we suppose that x_l and x_r are chosen so that

$$\frac{2}{3} \leqslant H_x \leqslant \frac{3}{2},$$

for all $x \in [H^3x_l, x_r]$. Moreover we take a_1 such that if $a \geqslant a_1$ and $H^j(a, x_r) < x_l$ then $j \geqslant 10$ (actually these choices are somewhat arbitrary).

The following Lemma compares H with purely quadratic functions. It is a direct consequence of the assumed proximity to $(a, x) \mapsto a - x^2$.

Lemma A.3. — *If $H \in \mathcal{H}$, δ_0 is sufficiently small, $a_1 \leqslant a < 0$ and $x \in [H^2 x_l, x_r]$ then*

$$\frac{5}{4}(a - x^2) \leqslant H(a, x) - x \leqslant \frac{3}{4}(a - x^2).$$

A *fundamental domain* for H is any interval of the form $[Hx, x]$. The smaller fundamental domain in $[x_l, x_r]$ has size equal to $\min |H(a,x) - x|$, which is greater or equal than $\frac{3}{4}a$, by Lemma A.3.

Consider now a fundamental domain $[H^{i+1}x_r, H^i x_r]$, $i \geqslant 0$. Let m be the first integer such that $H^m x_r < x_l$ (m differs from l by a finite amount). The power H^{m-i} maps $[H^{i+1}x_r, H^i x_r]$ diffeomorphically onto $[H^{m+1}x_r, H^m x_r]$. Note that H^{m-i} is extendible to the adjacent fundamental domains, so that the image extends to $[H^{m+2}x_r, H^{m-1}x_r]$. As the Schwarzian derivative of H is non-positive, the distortion of the power map derivative is bounded. In other words, there is $C_1 > 0$ such that

$$\frac{H_x^{m-i}(x_1)}{H_x^{m-i}(x_2)} \leqslant C_1,$$

for every $x_1, x_2 \in [H^{i+1}x_r, H^i x_r]$ and $0 \leqslant i \leqslant m - 1$. In particular, by the Mean Value Theorem and the estimate on the least size of a fundamental domain, there is $C > 0$ such that

$$C^{-1} < |H_x^j| < Ca^{-1},$$

for all $x \in [x_l, x_r]$, where j here is the first integer such that $H^j x < x_l$.

Another consequence is that if $x \in [Hx_r, x_r]$ then $|H_x^j| < C$. This proves the first two items of Lemma A.2, provided we have Lemma A.7 below, relating a and l. To prove this Lemma, however, we need three others.

Lemma A.4. — *If i_0 is such that $|Hx - x|$ is not monotone in $[H^{i_0+1}x_r, H^{i_0}x_r]$ then*

$$|H^{i_0+1}x_r - H^{i_0}x_r| \leqslant \frac{25}{8}a.$$

Proof. — Let x_c be the unique point where $\min |Hx - x|$ is attained. Then x_c belongs to $[H^{i_0+1}x_r, H^{i_0}x_r]$ and, by Lemma A.3, $|Hx_c - x_c| \leqslant \frac{5}{4}a$. But $[H^{i_0+1}x_r, H^{i_0}x_r] \subset [Hx_c, H^{-1}x_c]$, hence

$$|H^{i_0+1}x_r - H^{i_0}x_r| \leqslant |Hx_c - x_c| + |H^{-1}x_c - x_c|.$$

By the choice of x_l and x_r,

$$|H^{-1}x_c - x_c| \leqslant \frac{3}{2}|Hx_c - x_c|,$$

and the Lemma is proved. □

Lemma A.5. — *If $|Hx - x|$ is monotone in $[H^{i+1}x_r, H^i x_r]$ then*

$$\frac{2}{3} \leqslant \int_{H^{i+1}x_r}^{H^i x_r} \frac{1}{|Hx - x|}dx \leqslant \frac{3}{2}.$$

Proof. — By the Mean Value Theorem, there is x_i in $[H^{i+1}x_r, H^i x_r]$ such that

$$\int_{H^{i+1}x_r}^{H^i x_r} \frac{1}{|Hx-x|}dx = \frac{|H^{i+1}x_r - H^i x_r|}{|Hx_i - x_i|}.$$

As the maximum and the minimum values of $|Hx-x|$ are attained at the boundary of $[H^{i+1}x_r, H^i x_r]$, we compare them with $|H^{i+1}x_r - H^i x_r|$ using the supposition $\frac{2}{3} \leqslant |H_x| \leqslant \frac{3}{2}$, and the Lemma is proved. □

Lemma A.6. — *If i_0 is such that $|Hx-x|$ is not monotone in $[H^{i_0+1}x_r, H^{i_0}x_r]$ then*

$$\int_{H^{i_0+1}x_r}^{H^{i_0}x_r} \frac{1}{|Hx-x|}dx \leqslant 5.$$

Proof. — This is a consequence of Lemma A.4, since

$$\int_{H^{i_0+1}x_r}^{H^{i_0}x_r} \frac{1}{|Hx-x|}dx \leqslant |H^{i_0+1}x_r - H^{i_0}x_r| \max |Hx-x|^{-1} \leqslant \frac{25}{8}a \cdot \frac{4}{3}a^{-1}. \qquad \square$$

Lemma A.7. — *There is $C>0$ such that*

$$C^{-1}a^{-1/2} \leqslant l(a) \leqslant Ca^{-1/2},$$

for all $a_0 \leqslant a < 0$.

Proof. — It is enough to prove the same statement for m instead of l and for $a \geqslant a_1$. By Lemmas A.5 and A.6, we have

$$\frac{2}{3}\left(\int_{H^m x_r}^{x_r} \frac{1}{|Hx-x|}dx - 5\right) \leqslant m-1 \leqslant \frac{3}{2}\int_{H^m x_r}^{x_r} \frac{1}{|Hx-x|}dx.$$

Applying Lemma A.3 to the left inequality, we get

$$\begin{aligned} m &\geqslant -4 + \frac{2}{3}\cdot\frac{4}{5}\int_{x_l}^{x_r} \frac{1}{|a|+x^2}dx \\ &\geqslant -4 + \frac{1}{2}|a|^{-1/2}\left(\arctan\frac{x_r}{\sqrt{|a_1|}} - \arctan\frac{x_l}{\sqrt{|a_1|}}\right) \\ &\geqslant C^{-1}|a|^{-1/2}, \end{aligned}$$

where C is fixed after the choice of x_l, x_r and a_1. On the other hand,

$$\int_{H^m x_r}^{x_l} \frac{1}{|Hx-x|}dx \leqslant \int_{Hx_l}^{x_l} \frac{1}{|Hx-x|}dx \leqslant 1,$$

since the maximum of $|Hx-x|^{-1}$ in $[Hx_l, x_l]$ is attained in x_l. Then

$$\begin{aligned} m &\leqslant \frac{5}{2} + \frac{3}{2}\int_{x_l}^{x_r} \frac{1}{|Hx-x|}dx \\ &\leqslant \frac{3}{2}\cdot\frac{4}{3}|a|^{-1/2}\left(\arctan\frac{x_r}{\sqrt{|a|}} - \arctan\frac{x_l}{\sqrt{|a|}}\right) \\ &\leqslant \frac{5}{2} + 2\pi|a|^{-1/2} \leqslant C|a|^{-1/2}. \end{aligned}$$ □

Now we are able to prove the remaining assertions of Lemma A.2.

Lemma A.8. — *$|H_a^j| \leqslant Cl^3$, for all $x \in [x_l, x_r]$.*

Proof. — For $x \in [x_l, x_r]$, write

$$H_a^j = H_x^j \sum_{i=1}^{j} \frac{H_a \circ H^{i-1}}{H_x^i} \leqslant 2\sum_{i=1}^{j} |H_x^{j-i}\circ H^i|.$$

But this last sum is bounded by $Cjl^2 \leqslant Cl^3$. □

Lemma A.9. — *If $x \in [Hx_r, x_r]$ then $|H_a^j| \geqslant C^{-1}l^3$.*

Proof. — As in the previous Lemma,

$$H_a^j \geqslant \frac{3}{4}\sum_{i=1}^{j} H_x^{j-i},$$

since $H_a \simeq 1$. By bounded distortion,

$$H_x^{j-i} \geqslant C^{-1}|H^{i+1}x_r - H^i x_r|^{-1},$$

hence, similarly to the proof of Lemma A.7,

$$\begin{aligned} |H_a^j| &\geqslant C^{-1}\sum_{i=1}^{m} \frac{|H^{i+1}x_r - H^i x_r|}{|H^{i+1}x_r - H^i x_r|^2} \\ &\geqslant C^{-1}\int_{x_l}^{x_r} \frac{1}{(|a|+x^2)^2}dx \geqslant C^{-1}|a|^{-3/2} \geqslant C^{-1}l^3. \end{aligned}$$ □

Lemma A.10. — *$|H_{xx}^j| \leqslant C|H_x^j|^2$, for all $x \in [x_l, x_r]$.*

Proof. — Writing

$$\frac{H_{xx}^j}{(H_x^j)^2} = \sum_{i=1}^{j} \frac{H_{xx}\circ H^{i-1}}{(H_x^{j-i+1}\circ H^{i-1})(H_x \circ H^{i-1})}$$

we get

$$|H_{xx}^j| \leqslant C(H_x^j)^2 \sum_{i=1}^{j} \frac{1}{|H_x^{j-i+1}\circ H^{i-1}|}.$$

Using bounded distortion,

$$|H_x^{j-i+1}|^{-1} \leqslant C|H^i x - H^{i-1}x|,$$

for $i = 1, \dots, j$. As the sum of the sizes of the fundamental domains is bounded, the Lemma follows. □

Lemma A.11. — *$|H_{aa}^j| \leqslant Cl^6$, for all $x \in [x_l, x_r]$.*

Proof. — Write

$$\frac{H_{aa}^j}{(H_x^j)^2} = S_1 + 2S_2 + S_3,$$

where

$$S_1 = \sum_{i=1}^{j} \frac{H_{aa} \circ H^{i-1}}{H_x^i H_x^j}$$

$$S_2 = \sum_{i=2}^{j} \frac{H_{xa} \circ H^{i-1}}{H_x^j (H_x \circ H^{i-1})} \cdot \frac{H_a^{i-1}}{H_x^{i-1}}$$

$$S_3 = \sum_{i=2}^{j} \frac{H_{xx} \circ H^{i-1}}{(H_x^{j-i+1} \circ H^{i-1})(H_x \circ H^{i-1})} \cdot \left(\frac{H_a^{i-1}}{H_x^{i-1}}\right)^2 .$$

Then, as in Lemma A.8,

$$(H_x^j)^2 |S_1| \leqslant C \sum_{i=1}^{j} |H_x^{j-i}| \leqslant Cl^3.$$

In addition,

$$(H_x^j)^2 |S_2| \leqslant C \sum_{i=2}^{j} \sum_{t=1}^{i-1} |H_x^{j-t} \circ H^t| \leqslant Cl^4,$$

as in the proof of Lemma A.8. Finally,

$$\begin{aligned}(H_x^j)^2 |S_3| &\leqslant C \sum_{i=2}^{j} |H_x^{j-i+1} \circ H^{i-1}|^{-1} \left(\sum_{t=1}^{i-1} |H_x^{j-t} \circ H^t| \right)^2 \\ &\leqslant Cl^6 \sum_{i=2}^{j} |H_x^{j-i+1} \circ H^{i-1}|^{-1} \leqslant Cl^6,\end{aligned}$$

where the last inequality is similar to the proof of Lemma A.10. □

Lemma A.12. — *$|H_{xxx}^j| \leqslant C|H_x^j|^3$, for all $x \in [x_l, x_r]$.*

Proof. — Similar to Lemma A.10. It is enough to bound $H_{xxx}/(H_x)^3$ by

$$C\left(\sum_{i=1}^{j} |H^i x_r - H^{i-1} x_r| \right)^2 .$$

□

Lemma A.13. — *$|H_{xa}^j| \leqslant C|H_x^j| l^3$, for all $x \in [x_l, x_r]$.*

Proof. — Write $H^j_{xa}/H^j_x = S_1 + S_2$, where

$$S_1 = \sum_{i=1}^{j} \frac{H_{xa} \circ H^{i-1}}{(H_x \circ H^{i-1})}$$

and

$$S_2 = H^j_x \sum_{i=2}^{j} \frac{H_{xx} \circ H^{i-1}}{(H^{j-i+1}_x \circ H^{i-1})(H_x \circ H^{i-1})} \cdot \frac{H^{i-1}_a}{H^{i-1}_x}.$$

The techniques employed in the previous Lemmas lead to $|S_1| \leqslant Cl$ and $|S_2| \leqslant Cl^3$. □

Lemma A.14. — $|H^j_{xxa}| \leqslant C|H^j_x|^2 l^3$, *for all* $x \in [x_l, x_r]$.

Proof. — Write

$$\frac{H^j_{xxa}}{(H^j_x)^2} = H^j_x(S_1 + S_2 + S_3 + 4S_4 + 2S_5),$$

where

$$S_1 = \frac{1}{H^j_x} \sum_{i=1}^{j} \frac{H_{xxa} \circ H^{i-1}}{(H^{j-i+1}_x \circ H^{i-1})(H_x \circ H^{i-1})},$$

$$S_2 = \sum_{i=2}^{j} \frac{H_{xxx} \circ H^{i-1}}{(H^{j-i+1}_x \circ H^{i-1})^2(H_x \circ H^{i-1})} \cdot \frac{H^{i-1}_a}{H^{i-1}_x},$$

$$S_3 = \sum_{i=2}^{j} \frac{H_{xa} \circ H^{i-1}}{H^j_x H^i_x} \cdot \frac{1}{H^{j-i+1}_x \circ H^{i-1}} \frac{H^{i-1}_{xx}}{(H^{i-1}_x)^2},$$

$$S_4 = \sum_{i=2}^{j} \frac{H_{xx} \circ H^{i-1}}{(H^{j-i+1}_x \circ H^{i-1})^2(H_x \circ H^{i-1})} \cdot \frac{H^{i-1}_{xa}}{(H^{i-1}_x)^2},$$

$$S_5 = \sum_{i=2}^{j} \frac{H_{xx} \circ H^{i-1}}{(H^{j-i+1}_x \circ H^{i-1})^2(H_x \circ H^{i-1})} \cdot \frac{H^{i-1}_a}{H^{i-1}_x} \cdot \frac{H^{i-1}_{xx}}{(H^{i-1}_x)^2},$$

then proceed as in the previous Lemmas. □

B. Glossary

The formula below appear in many places of this work. They give mixed derivatives of compositions of parameter dependent diffeomorphisms. Let $\{F_i\}_{i=1,\ldots,j}$, $F_i = F_i(a, x)$, be a sequence of diffeomorphisms and $G = G_j$ its composition $G_j = F_j \circ \cdots \circ F_1$. Consider also the partial compositions $G_i = F_i \circ \cdots \circ F_1$ and $Q_i = F_j \circ \cdots \circ F_i$. To simplify the notation, we omit the points where the functions are evaluated.

$$\frac{G_a}{G_x} = \sum_{i=1}^{j} \frac{F_{i,a}}{G_{i,x}} \tag{3}$$

$$\frac{G_{xx}}{(G_x)^2} = \sum_{i=1}^{j} \frac{F_{i,xx}}{Q_{i+1,x}(F_{i,x})^2} \tag{4}$$

$$\frac{G_{xa}}{(G_x)^2} = \sum_{i=1}^{j} \frac{F_{i,xa}}{F_{i,x}G_{j,x}} + \sum_{i=2}^{j} \frac{F_{i,xx}}{Q_{i+1,x}(F_{i,x})^2} \cdot \frac{G_{i-1,a}}{G_{i-1,x}} \tag{5}$$

$$\frac{G_{aa}}{(G_x)^2} = \sum_{i=1}^{j} \frac{F_{i,aa}}{G_{i,x}G_{j,x}} + 2\sum_{i=2}^{j} \frac{F_{i,xa}}{F_{i,x}G_{j,x}} \cdot \frac{G_{i-1,a}}{G_{i-1,x}} + \sum_{i=2}^{j} \frac{F_{i,xx}}{Q_{i+1,x}(F_{i,x})^2} \cdot \left(\frac{G_{i-1,a}}{G_{i-1,x}}\right)^2 \tag{6}$$

$$\frac{G_{xxx}}{(G_x)^3} = \sum_{i=1}^{j} \frac{F_{i,xxx}F_{i,x}}{(Q_{i+1,x})^2(F_{i,x})^4} + 3\sum_{i=2}^{j} \frac{F_{i,xx}}{Q_{i+1,x}(F_{i,x})^2} \cdot \frac{1}{Q_{i,x}} \frac{G_{i-1,xx}}{(G_{i-1,x})^2} \tag{7}$$

$$\begin{aligned} \frac{G_{xxa}}{(G_x)^3} &= \sum_{i=1}^{j} \frac{F_{i,xxa}}{G_{j,x}Q_{i+1,x}(F_{i,x})^2} \\ &+ \sum_{i=2}^{j} \frac{F_{i,xxx}F_{i,x}}{(Q_{i+1,x})^2(F_{i,x})^4} \cdot \frac{G_{i-1,a}}{G_{i-1,x}} + \sum_{i=2}^{j} \frac{F_{i,xa}}{F_{i,x}G_{j,x}} \cdot \frac{1}{Q_{i,x}} \frac{G_{i-1,xx}}{(G_{i-1,x})^2} \\ &+ 2\sum_{i=2}^{j} \frac{F_{i,xx}}{Q_{i+1,x}(F_{i,x})^2} \cdot \left[\frac{2}{Q_{i,x}} \frac{G_{i-1,xa}}{(G_{i-1,x})^2} + \frac{G_{i-1,a}}{G_{i-1,x}} \cdot \frac{1}{Q_{i,x}} \frac{G_{i-1,xx}}{(G_{i-1,x})^2}\right] \end{aligned} \tag{8}$$

References

[1] A. Ávila, M. Lyubich, W. de Melo. Regular or stochastic dynamics in real analytic families of unimodal maps. Preprint.

[2] E. Colli. A starting condition approach to parameter distortion in generalized renormalization. *Qualitative Theory of Dynamical Systems* **2** (2001), 221–288.

[3] J. Graczyk, G. Świątek. Generic hyperbolicity in the quadratic family. *Ann. of Math.* **146** (1997), 1–52.

[4] M. Jakobson. Absolutely continuous invariant measures for one parameter families of one dimensional maps. *Comm. Math. Phys.* **81** (1981), 39–88.

[5] M. Jakobson. Piecewise smooth maps with absolutely continuous invariant measures and uniformly scaled Markov partitions. *Proceedings of Symposia in Pure Mathematics* **69** (2001), 825–881.

[6] M. Jakobson, G. Świątek. Metric properties of non-renormalizable S-unimodal maps. Part I: Induced expansion and invariant measures, *Ergod. Th. & Dynam. Sys.* **14** (1994), 721–755.

[7] M. Lyubich. Combinatorics, geometry and attractors of quasi-quadratic maps. *Ann. of Math.* **140** (1994), 347–404.

[8] M. Lyubich. Dynamics of quadratic polynomials, I, II. *Acta Math.* **178** (1997), 185–247, 247–297.

[9] M. Lyubich. Dynamics of quadratic polynomials, III. Parapuzzle and SBR measures. *Astérisque* **261** (2000), 173–200.

[10] M. LYUBICH. Almost every real quadratic map is either regular or stochastic. *Ann. of Math.* **156** (2002), no. 1, 1–78.

[11] M. MARTENS. Distortion results and invariant Cantor sets of unimodal maps. *Ergod. Th. & Dynam. Sys.* **14** (1994), 331–349.

[12] M. MARTENS, T. NOWICKI. Invariant measures for typical quadratic maps. *Astérisque* **261** (2000), 239–252.

[13] W. DE MELO, S. VAN STRIEN. *One-dimensional dynamics.* Springer Verlag, Berlin, Heidelberg, New York, 1993.

[14] D. SANDS. Misiurewicz maps are rare. *Comm. Math. Phys.* **197** (1998), 109–129.

[15] D. SINGER. Stable orbits and bifurcation of maps of the interval. *SIAM J. Appl. Math* **35** (1978), 260–267.

[16] PH. THIEULLEN, C. TRESSER, L.S. YOUNG. Positive Lyapunov exponent for generic one-parameter families of unimodal maps. *J. Anal. Math.* **64** (1994), 121–172.

E. COLLI, Departamento de Matemática Aplicada, Instituto de Matemática e Estatística, Universidade de São Paulo, R. do Matão, 1010, 05508-090 São Paulo (SP), Brasil
E-mail : `colli@ime.usp.br` • *Url :* `http://www.ime.usp.br/~colli/`

V. PINHEIRO, Instituto de Matemática, Universidade Federal da Bahia, Av. Ademar de Barros s/n, 40170-110 Salvador (BA) Brasil • *E-mail :* `viltonj@ufba.br` • *E-mail :* `viltonj@ufba.br`