Chapter 2

Structural stability

2.1 Definitions and one-dimensional examples

A very important notion, both from a theoretical point of view and for applications, is that of stability: the qualitative behavior should not change under small perturbations.

Definition 2.1.1: A C^r map f is C^m *structurally stable* (with $1 \le m \le r \le \infty$) if there exists a neighbourhood U of f in the C^m topology such that every $g \in U$ is topologically conjugated to f.

Remark 2.1.1. The reason that for structural stability we just ask the existence of a *topological* conjugacy with close maps is because we are interested only in the *qualitative* properties of the dynamics. For instance, the maps $f(x) = \frac{1}{2}x$ and $g(x) = \frac{1}{3}x$ have the same qualitative dynamics over $\mathbb{R}$ (and indeed are topologically conjugated; see below) but they cannot be C^1-conjugated. Indeed, assume there is a C^1-diffeomorphism $h: \mathbb{R} \to \mathbb{R}$ such that $h \circ g = f \circ h$. Then we must have $h(0) = 0$ (because the origin is the unique fixed point of both f and g) and

$$\frac{1}{3}h'(0) = h'\bigl(g(0)\bigr)g'(0) = (h \circ g)'(0) = (f \circ h)'(0) = f'\bigl(h(0)\bigr)h'(0) = \frac{1}{2}h'(0);$$

but this implies $h'(0) = 0$, which is impossible.

Let us begin with examples of non-structurally stable maps.

EXAMPLE 2.1.1. For $\varepsilon \in \mathbb{R}$ let $F_\varepsilon: \mathbb{R} \to \mathbb{R}$ given by $F_\varepsilon(x) = x - x^2 + \varepsilon$. We have $\|F_\varepsilon - F_0\|_r = |\varepsilon|$ for all $r \ge 0$, and hence $F_\varepsilon \to F_0$ in the C^r topology. But F_ε has two distinct fixed points for $\varepsilon > 0$, only one for $\varepsilon = 0$, and none for $\varepsilon < 0$; therefore F_0 cannot be topologically conjugated to F_ε for $\varepsilon \ne 0$, and hence F_0 is not C^1-structurally stable.

EXAMPLE 2.1.2. The rotations $R_\alpha: S^1 \to S^1$ are not structurally stable. Indeed, R_α is periodic if α is rational, and it has no periodic points if α is irrational, and so a rational rotation cannot ever be topologically conjugated to an irrational rotation, no matter how close they are.

Exercise 2.1.1. For $\lambda \in \mathbb{R}$ let $T_\lambda: \mathbb{R} \to \mathbb{R}$ given by $T_\lambda(x) = x^3 - \lambda x$. Prove that:

(i) $\mathrm{Fix}(T_\lambda) = \{-\sqrt{1+\lambda}, 0, \sqrt{1+\lambda}\}$ for $-1 < \lambda$;
(ii) when $-1 < \lambda \le 1$ the open interval $(-\sqrt{1+\lambda}, \sqrt{1+\lambda})$ is attracted by the origin;
(iii) when $\lambda > 1$ the point $\sqrt{\lambda - 1} \in (-\sqrt{1+\lambda}, \sqrt{1+\lambda})$ is periodic of exact period 2;
(iv) T_1 is not C^1-structurally stable.

And now a couple of examples of structurally stable maps.

Proposition 2.1.1: *The map $L: \mathbb{R} \to \mathbb{R}$ given by $L(x) = \frac{1}{2}x$ is C^1-structurally stable.*

Proof: We shall show that every $g \in C^1(\mathbb{R}, \mathbb{R})$ such that $\|g - L\|_1 < 1/2$ is topologically conjugated to L. The first remark is that

$$0 < \frac{1}{2} - \|g - L\|_1 \le g'(x) \le \frac{1}{2} + \|g - L\|_1 < 1 \qquad (2.1.1)$$

for all $x \in \mathbb{R}$; therefore g is a contraction (Corollary 1.2.3), has a unique fixed point $p_0 \in \mathbb{R}$ and all g-orbits converge exponentially to p_0. Up to replacing g by the map $x \mapsto g(x + p_0) - p_0$, which is conjugated to g, we can also assume $p_0 = 0$.

It is clear that for every $x \neq 0$ there exists a *unique* $k_0 \in \mathbb{Z}$ such that $L^{k_0}(x) \in [-10,-5) \cup (5,10]$; let us prove a similar property for g.

The inequalities (2.1.1) imply that g is strictly increasing and $g - \mathrm{id}$ is strictly decreasing; in particular, $x < 0$ implies $0 > g(x) > x$ and $x > 0$ implies $0 < g(x) < x$. We then claim that for every $x \neq 0$ there exists a *unique* $h_0 \in \mathbb{Z}$ such that $g^{h_0}(x) \in \bigl[-10, g(-10)\bigr) \cup \bigl(g(10), 10\bigr]$. Indeed, take $x > 0$; since $g^h(x) \to 0^+$ there exists a minimum $h_0 \geq 0$ such that $g^h(x) \leq g(10)$ for all $h \geq h_0 + 1$. Then

$$g(10) < g^{h_0}(x) \leq 10 < g^{h_0-1}(x) < g^{h_0-2}(x) < \cdots$$

as required. A similar argument works for $x < 0$.

To build a topological conjugation $h\colon \mathbb{R} \to \mathbb{R}$ between L and g let us begin by requiring it to be a linear homeomorphism of $[-10,-5] \cup [5,10]$ with $[-10, g(-10)] \cup [g(10), 10]$ fixing ± 10, and hence sending ± 5 in $g(\pm 10)$, so that $g \circ h(\pm 10) = h \circ f(\pm 10)$. For $x \notin [-10,-5] \cup [5,10] \cup \{0\}$ take $k_0 \in \mathbb{Z}$ such that $L^{k_0}(x) \in [-10,-5) \cup (5,10]$ and set

$$h(x) = g^{-k_0} \circ h \circ L^{k_0}(x).$$

Finally, set $h(0) = 0$. We must show that $h\colon \mathbb{R} \to \mathbb{R}$ is a homeomorphism conjugating L and g.

First of all, h is invertible. Indeed, let us define $h_1\colon \mathbb{R} \to \mathbb{R}$ as follows: put $h_1(0) = 0$, and, for $y \neq 0$, take $h_0 \in \mathbb{Z}$ such that $g^{h_0}(y) \in \bigl[-10, g(-10)\bigr) \cup \bigl(g(10), 10\bigr]$ and put $h_1(y) = L^{-h_0} \circ h^{-1} \circ g^{h_0}(y)$. Then it is easy to check that $h \circ h_1 = h_1 \circ h = \mathrm{id}$.

Now let us show that h is continuous. If $x_0 \neq 0$ and $L^{k_0}(x_0) \neq \pm 10$ then there is a $\delta > 0$ such that $L^{k_0}(x) \in (-10,5) \cup (5,10)$ as soon as $|x - x_0| < \delta$, and hence h is continuous in x_0. If $L^{k_0}(x) = 10$ and $x \to x_0^-$ then $h(x) = g^{-k_0} \circ h \circ L^{k_0}(x) \to g^{-k_0}(10) = h(x_0)$; if instead $x \to x_0^+$ then

$$h(x) = g^{-k_0-1} \circ h \circ L^{k_0+1}(x) \to g^{-k_0-1} \circ h \circ L(10) = g^{-k_0-1}\bigl(h(5)\bigr) = g^{-k_0-1}\bigl(g(10)\bigr) = g^{-k_0}(10) = h(x_0),$$

and so h is continuous in x_0. So $h\colon \mathbb{R}^* \to \mathbb{R}^*$ is a homeomorphism and strictly increasing, because it is so on $[-10,5) \cup (5,10]$; therefore it must be continuous in zero too.

Finally, for $x \in \mathbb{R}^*$ we have

$$g \circ h(x) = g^{-k_0+1} \circ h \circ L^{k_0}(x) = g^{-k_0+1} \circ h \circ L^{k_0-1}\bigl(L(x)\bigr) = h \circ L(x),$$

and $g \circ h = h \circ f$ as required. □

The previous proof used, without naming it, the notion of fundamental domain.

Definition 2.1.2: Let $f\colon X \to X$ a continuous self-map of a topological space X. A *fundamental domain* for f is an open subset $D \subset X$ such that every orbit of f intersect D in *at most* one point and intersect $\overline{D}$ in *at least* one point.

Exercise 2.1.2. Let $f\colon X \to X$ and $g\colon Y \to Y$ be two continuous self-maps. Assume that there are a fundamental domain $D_f \subset X$ for f, a fundamental domain $D_g \subset Y$ for g and a homeomorphism $h\colon \overline{D_f} \to \overline{D_g}$ such that $g \circ h = h \circ f$ on $f^{-1}(\overline{D_f}) \cap \overline{D_f}$. Show that f and g are topologically conjugated.

A variation on the previous argument allows us to prove the following

Proposition 2.1.2: *If $\mu > 2 + \sqrt{5}$ then F_μ is C^2-structurally stable.*

Proof: We shall use the notations introduced in Section 1.6. In particular, the hypothesis $\mu > 2 + \sqrt{5}$ implies $|F'_\mu(x)| > 1$ for all $x \in I_0 \cup I_1$.

Take $g \in C^2(\mathbb{R}, \mathbb{R})$. Since $F''_\mu \equiv -2\mu$, there is $\varepsilon_1 > 0$ such that $\|g - F_\mu\|_2 < \varepsilon_1$ then $g''(x) < 0$ for all $x \in \mathbb{R}$; in particular, g is concave, and so it can have at most two fixed points. Actually, if $\|g - F_\mu\|_0$ is small enough then g has exactly two fixed points α and β, as close as we want to 0 and p_μ. Indeed, choose $x_1 < 0 < x_2 < p_\mu$ and let $\delta = \min\{|F_\mu(x_1) - x_1|, |F_\mu(x_2) - x_2|\}$. If $\|g - F_\mu\|_0 < \delta/2$ then we have $g(x_1) - x_1 < 0$ and $g(x_2) - x_2 > 0$, and hence g must have a fixed point $\alpha \in (x_1, x_2)$. In a similar way one proves that g has a fixed point β close to p_μ.

Being g concave, it has a unique critical point c; up to take $\|g-F_\mu\|_1$ small enough we can assume that c is close to $1/2$ and $\alpha < c < \beta$, and so there must exist $\alpha < \beta' < c < \beta < \alpha'$ such that $g(\alpha') = \alpha$ and $g(\beta') = \beta$. Finally, up to decreasing $\|g - F_\mu|_1$ again, we can also assume that there are $\alpha < a_0 < c < a_1 < p_\mu$ such that $g(a_0) = g(a_1) = \alpha'$ and $|g'| > 1$ on $[\alpha, a_0] \cup [a_1, \alpha']$.

So g has on $[\alpha, \alpha']$ the same qualitative properties F_μ has on $[0, 1]$. Furthermore, it is easy to see that $g^k(x) \to -\infty$ for all $x \in (-\infty, \alpha) \cup (\alpha', +\infty)$. Moreover, arguing as we did for F_μ in Theorems 1.6.3 and 1.7.6, we can show that all points in $[\alpha, \alpha']$ have orbits tending to $-\infty$ except for the orbits belonging to a g-invariant Cantor set Λ_g on which g is topologically conjugated to a 2-shift. In particular, we have a topological conjugation $h\colon \Lambda \to \Lambda_g$ between $F_\mu|_\Lambda$ and $g|_{\Lambda_g}$.

To extend h to the rest of $\mathbb{R}$ we again use a fundamental domain. Choose $x_0 < \min\{g^2(c), F_\mu^2(1/2)\} < 0$; it is then not difficult to see that $D = \big(F_\mu(x_0), x_0\big)$ and $D_g = \big(g(x_0), x_0\big)$ are fundamental domains for F_μ on $(-\infty, 0)$ and for g on $(-\infty, \alpha)$ respectively. Let $h\colon \overline{D} \to \overline{D_g}$ be a linear increasing homeomorphism; then, using the technique delineated in the previous proof (and in Exercise 2.1.2) we can extend h to a homeomorphism $h\colon (-\infty, 0) \to (-\infty, \alpha)$ conjugating F_μ and g. We now extend h to $(1, +\infty)$ by taking as $h(x)$ the unique solution $y \in (\alpha', +\infty)$ of the equation $g(y) = h \circ F_\mu(x)$.

Now we put $h(1/2) = c$ and we extend h to A_0 by taking, for $x \in A_0 \cap (0, 1/2)$, as $h(x)$ the unique solution $y \in (a_0, c)$ of the equation $g(y) = h \circ F_\mu(x)$, and, for $x \in A_0 \cap (1/2, 1)$, as $h(x)$ the unique solution $y \in (c, a_1)$ of the same equation.

Arguing by induction it is now easy to extend h to all the A_n, and we end up with a homeomorphism $h\colon \mathbb{R}\setminus\Lambda \to \mathbb{R}\setminus\Lambda_g$ conjugating F_μ with g. In this way we get an invertible map $h\colon \mathbb{R} \to \mathbb{R}$ conjugating F_μ with g; we must only show that h and its inverse are continuous at Λ, respectively Λ_g.

By construction, h sends $I_0 \setminus \Lambda$ onto $[\alpha, a_0] \setminus \Lambda_g$, and $I_1 \setminus \Lambda$ onto $[a_1, \alpha'] \setminus \Lambda_g$. Furthermore, again by construction, h sends $I_0 \cap \Lambda$ onto $[\alpha, a_0] \cap \Lambda_g$, and $I_1 \cap \Lambda$ onto $[a_1, \alpha'] \cap \Lambda_g$. Therefore $h(I_0) = [\alpha, a_0]$ and $h(I_1) = [a_1, \alpha']$.

Take now $x_0 \in \Lambda$ with $S(x_0) = \mathbf{s}$, so that $x_0 = \bigcap_{n\ge 0} F_\mu^{-n}(I_{s_n})$, and hence $h(x_0) = \bigcap_{n\ge 0} g^{-n}(I^g_{s_n})$, where $I^g_0 = [\alpha, a_0]$ and $I^g_1 = [a_1, \alpha']$. But we have $h\left(\bigcap_{n=0}^{n_0} F_\mu^{-n}(I_{s_n})\right) = \bigcap_{n=0}^{n_0} g^{-n}(I^g_{s_n})$ for every $n_0 \ge 0$; since the intersections $\bigcap_{n=0}^{n_0} g^{-n}(I^g_{s_n})$ form a fundamental system of neighbourhoods of $h(x_0)$ it follows that h is continuous at x_0, and we are done. □

Exercise 2.1.3. Let $I \subseteq \mathbb{R}$ an interval, $f\colon I \to I$ of class C^1, and $p \in I$ a hyperbolic fixed point of f with $|f'(p)| \ne 0$, 1. Prove that there are a neighbourhood U of p, a neighbourhood V of 0 in $\mathbb{R}$ and a homeomorphism $h\colon U \to V$ such that $h \circ f(x) = f'(p) \cdot h(x)$ for all $x \in U \cap f^{-1}(U)$.

Exercise 2.1.4. Show that hyperbolic fixed points in one variable are locally C^1-structurally stable, in the sense that if $I \subseteq \mathbb{R}$ is an interval, $f\colon I \to I$ is of class C^1, and $p \in I$ a hyperbolic fixed point of f with $|f'(p)| \ne 0$, 1, then there are a neighbourhood U of p and an $\varepsilon > 0$ so that if $g \in C^1(I, I)$ is such that $\|f - g\|_{1,U} < \varepsilon$ then g has a hyperbolic fixed point in U and $g|_U$ is topologically conjugated to $f|_U$.

2.2 Expanding maps of the circle

This section is devoted to a particularly nice example of structurally stable maps: the expanding self-maps of the circle S^1.

Definition 2.2.1: A continuous self-map $f\colon X \to X$ of a metric space X is *expanding* if there are $\mu > 1$ and $\varepsilon_0 > 0$ such that for all $x, y \in X$ such that $d(x, y) < \varepsilon_0$ one has

$$d\big(f(x), f(y)\big) \ge \mu\, d(x, y).$$

Exercise 2.2.1. Prove that every open expanding map of a compact connected metric space is a covering map.

If X actually is a Riemannian manifold M, we have an infinitesimal characterization of expansivity:

Proposition 2.2.1: *Let $f\colon M \to M$ be a C^1 self-map of a Riemannian manifold M. Then:*

(i) *If f is expanding there is $\mu > 1$ so that $\|df_x(v)\| \ge \mu\|v\|$ for every $x \in M$ and $v \in T_xM$.*

(ii) *If M is compact and there is $\mu > 1$ such that $\|df_x(v)\| \ge \mu\|v\|$ for every $x \in M$ and $v \in T_xM$, then f is expanding.*

Proof: (i) Let $\mu > 1$ and $\varepsilon_0 > 0$ be the constants associated to f. Choose $v \in T_xM \setminus \{O\}$, and let $\sigma\colon(-\varepsilon,\varepsilon) \to M$ be a geodesic with $\sigma(0) = x$ and $\sigma'(0) = v$. In particular, for $t > 0$ small we have $d\bigl(x,\sigma(t)\bigr) = t\|v\| < \varepsilon_0$, and thus

$$\mu t\|v\| = \mu\, d\bigl(x,\sigma(t)\bigr) \le d\bigl(f(x), f\bigl(\sigma(t)\bigr)\bigr) \le \int_0^t \bigl\|df_{\sigma(s)}\bigl(\sigma'(s)\bigr)\bigr\|\, ds.$$

Dividing by t and letting $t \to 0$ we get $\mu\|v\| \le \|df_x(v)\|$, as claimed.

(ii) By the inverse function theorem, f is a local diffeomorphism. By compactness we can choose $\delta_0 > 0$ such that every ball of radius δ_0 is sent diffeomorphically onto its image, and $\delta_1 > 0$ such that every connected component of the inverse image of a δ_1-ball has a diameter less than δ_0. Finally, let $0 < \varepsilon < \delta_0$ be such that $d(x,y) \le \varepsilon$ implies $d\bigl(f(x), f(y)\bigr) < \delta_1/2$. Let $\gamma\colon[0,1] \to M$ be a smooth curve connecting $f(x)$ to $f(y)$ inside a δ_1-ball. Then we can lift the curve γ to a curve $\tilde\gamma$ such that $\tilde\gamma(0) = x$, $\tilde\gamma(1) = y$ and $f \circ \tilde\gamma = \gamma$. Then

$$\mathrm{Length}(\gamma) = \int_0^1 \bigl\|df_{\tilde\gamma(t)}\bigl(\dot{\tilde\gamma}(t)\bigr)\bigr\|\, dt \ge \mu \int_0^1 \|\dot{\tilde\gamma}(t)\|\, dt = \mu\,\mathrm{Length}(\tilde\gamma) \ge \mu\, d(x,y).$$

Since $d\bigl(f(x), f(y)\bigr)$ is the infimum of the length of curves connecting $f(y)$ to $f(y)$ inside a δ_1-ball, we get $d\bigl(f(x), f(y)\bigr) \ge \mu\, d(x,y)$, and we are done. □

Corollary 2.2.2: *If $f\colon M \to M$ is a C^1 expanding map of a compact Riemannian manifold M, then any $g \in C^1(M,M)$ sufficiently C^1-close to f is still expanding.*

Proof: If g is sufficiently C^1-close to f there is, by Proposition 2.2.1.(i), a $\mu' > 1$ such that $\|df_x(v)\| \ge \mu'\|v\|$ for all $x \in M$ and $v \in T_xM$, and hence g is expanding by Proposition 2.2.1.(ii). □

Now we introduce the standard example of expanding map.

Definition 2.2.2: Let $\pi\colon\mathbb{R} \to S^1 = \mathbb{R}/\mathbb{Z}$ be the usual covering map, and endow S^1 with the distance induced by π, that is

$$d\bigl(\pi(s), \pi(t)\bigr) = |s - t| \pmod 1.$$

For $m \in \mathbb{Z}^*$ let then $E_m\colon S^1 \to S^1$ be given by $E_m(x) = mx \pmod 1$.

Exercise 2.2.2. Given $m \in \mathbb{Z}^*$ with $|m| \ge 2$, prove that the map E_m is expanding, chaotic, and has $|m^k - 1|$ periodic points of period k.

Our aim is to prove that the maps E_m (and, more generally, all expanding self-maps of class C^1 of S^1) are C^1-structurally stable. To achieve this we need the notion of degree of a continuous self-map of S^1.

Lemma 2.2.3: *Let $f\colon S^1 \to S^1$ be a continuous self-map of S^1, and $F\colon\mathbb{R} \to \mathbb{R}$ a lift of f to the universal covering $\pi\colon\mathbb{R} \to S^1$. Then the number $F(x+1) - F(x)$ is an integer independent of x and of the chosen lift.*

Proof: We have $\pi\bigl(F(x+1)\bigr) = f\bigl(\pi(x+1)\bigr) = f\bigl(\pi(x)\bigr) = \pi\bigl(F(x)\bigr)$, and so $F(x+1) - F(x)$ is an integer; since it depends continuously on x, it is constant. If $\tilde F$ is another lift, then we also see that $\tilde F - F$ is an integer constant, and thus $\tilde F(x+1) - \tilde F(x) = F(x+1) - F(x)$. □

Definition 2.2.3: If $f\colon S^1 \to S^1$ is a continuous self-map of S^1 and $F\colon\mathbb{R} \to \mathbb{R}$ is any lift of f, the integer $\deg(f) = F(x+1) - F(x)$ is the *degree* of f.

Remark 2.2.1. It is not difficult to see that if $f_*\colon\pi_1(S^1) \to \pi_1(S^1)$ is the endomorphism induced by $f\colon S^1 \to S^1$ on the fundamental group, one has $f_*(\gamma) = \deg(f)\gamma$ for all $\gamma \in \pi_1(S^1)$.

Example 2.2.1. Since a lift of $E_m\colon S^1 \to S^1$ is the map $\tilde E_m\colon\mathbb{R} \to \mathbb{R}$ given by $\tilde E_m(x) = mx$, it immediately follows that $\deg(E_m) = m$.

Lemma 2.2.4: *The degree is continuous and hence locally constant in the C^0 topology.*

Proof: Let f and g be two continuous self-maps of S^1 such that $d\big(f(x), g(x)\big) < 1/4$ for all $x \in S^1$; it suffices to prove that $\deg(f) = \deg(g)$. Let F and G be lifts of f and g, with $|F(0) - G(0)| < 1/4$, and set $\varphi = G - F$. Then

$$G(t+1) - \varphi(t+1) = F(t+1) = F(t) + \deg(f) = G(t) - \varphi(t) + \deg(f)$$
$$= G(t+1) + \deg(f) - \deg(g) - \varphi(t).$$

So $\deg(g) - \deg(f) \equiv \varphi(t+1) - \varphi(t)$. But $d_0(f,g) < 1/4$ and $|F(0) - G(0)| < 1/4$ imply $|\varphi(t)| < 1/4$ for all $t \in \mathbb{R}$; thus $|\varphi(t+1) - \varphi(t)| < 1/2$, that is $\deg(f) = \deg(g)$. □

The degree of a continuous expanding map of the circle is necessarily greater than one in absolute value. More precisely:

Lemma 2.2.5: *Let $f: S^1 \to S^1$ be a continuous expanding map, of constants $\varepsilon_0 > 0$ and $\mu > 1$, and let $F: \mathbb{R} \to \mathbb{R}$ be any lift of f. Then:*

(i) *F is expanding, of constants $\varepsilon_1 \le \varepsilon_0$ and μ;*
(ii) *F is strictly monotone;*
(iii) *we have $|F(s) - F(t)| \ge \mu|s-t|$ for all $s, t \in \mathbb{R}$;*
(iv) *$|\deg(f)| \ge 2$.*

Proof: (i) Choose $\varepsilon_1 \le \varepsilon_0$ so that every interval $I \subset S^1$ of length ε_1 is well-covered by π (that is, $\pi^{-1}(I)$ is the disjoint union of intervals of length ε_1 homeomorphically sent onto I). In particular, then, π restricted to any interval in $\mathbb{R}$ of length ε_1 is a homeomorphism preserving the distance. Hence if $s, t \in \mathbb{R}$ are such that $|s-t| < \varepsilon_1$ we have $d\big(\pi(s), \pi(t)\big) = |s-t| < \varepsilon_0$, and so

$$|F(s) - F(t)| \ge d\Big(f\big(\pi(s)\big), f\big(\pi(t)\big)\Big) \ge \mu\, d\big(\pi(s), \pi(t)\big) = \mu|s-t|.$$

(ii) If F is not strictly monotone we can find $t_1 < t_2 < t_3$ such that $F(t_1) \le F(t_2) \ge F(t_3)$ or $F(t_1) \ge F(t_2) \le F(t_3)$. Clearly, we can assume that t_2 is the absolute maximum (or minimum) of F in $[t_1, t_3]$. But then we can find $t_1' < t_3'$ with $t_1' \le t_2 \le t_3'$, as close as we want to t_2 and such that $F(t_1') = F(t_3')$, and thus F is not expanding.

(iii) Given $s, t \in \mathbb{R}$, subdivide the interval from s to t in a finite number of subintervals of length at most ε_1. By (i), the length of the image of each subinterval is at least μ times the length of the subinterval; moreover, by (ii), the images intersect only in the end points. Therefore the length of the image of the interval from s to t (that is $|F(s) - F(t)|$) is at least μ times the distance from s to t.

(iv) $|\deg f| = |F(t+1) - F(t)| \ge \mu|(t+1) - t| = \mu > 1$. □

Lemma 2.2.6: *Let $f: S^1 \to S^1$ be continuous of degree m, with $m \ne 1$. Then f has a lift with a fixed point $p \in [-1/2, 1/2]$.*

Proof: Let F be a lift of f, and set $H(t) = F(t) - t$. Since

$$H(1/2) - H(-1/2) = F(1/2) - F(-1/2) - 1 = m - 1,$$

there is at least one integer k between $H(-1/2)$ and $H(1/2)$, and so there is $p \in [-1/2, 1/2]$ so that $H(p) = k$. Replacing F by $F - k$ we get a lift with $F(p) = p$. □

Theorem 2.2.7: *Every expanding map f of the circle of degree m is topologically conjugate to the map E_m.*

Proof: Let F denote a lift of f with a fixed point $p \in [-1/2, 1/2]$ as in the previous lemma, and let $\tilde{E}_m$ denote the lift of E_m such that $\tilde{E}_m(0) = 0$, that is, $\tilde{E}_m(t) = mt$.

We shall use a technique of proof known as *coding*. Let us first assume m positive. Set $\Delta_0^0 = [0,1]$, and $\Delta_j^i = [i/m^j, (i+1)/m^j]$ for $j \in \mathbb{N}$, $0 \le i \le m^j - 1$. Then

$$\tilde{E}_m\left(\frac{i}{m^j}\right) = \frac{[i]_{m^{j-1}}}{m^{j-1}} \pmod 1, \tag{2.2.1}$$

where $[i]_{m^{j-1}}$ denotes the unique integer i' between 0 and $m^{j-1}-1$ such that $i' \equiv i \pmod{m^{j-1}}$. The set $\pi(\Sigma) = \{\pi(i/m^j)\}_{j\in\mathbb{N}, i=0,\dots,m^j-1}$, where $\pi\colon \mathbb{R} \to S^1$ is the universal covering, is dense in S^1; we shall define our homemomorphism $h\colon S^1 \to S^1$ on this set and then extend it to S^1 by continuity.

To define h we shall actually define a lift $H\colon [0,1] \to [p, p+1]$ proceeding by induction on j. For $j = 0$ we set $a_0^0 = p$, $a_0^1 = p+1$, $H(0) = H(1) = p$ and $\Gamma_0^0 = [p, p+1]$. For $j = 1$, since $F(p) = p$, $F(p+1) = p+m$ and F is strictly monotone, there are uniquely defined points $a_1^0 = p < a_1^1 < \cdots < a_1^{m-1} < p+1 = a_1^m$ such that $F(a_1^i) = p+i$, for $i = 0,\dots,m$. Set then $H(i/m) = a_1^i$ and $\Gamma_1^i = [a_1^i, a_1^{i+1}]$ for $i = 0,\dots,m-1$. Clearly, we have

$$f\big(\pi(\Gamma_1^i)\big) = \pi\big(F(\Gamma_1^i)\big) = \pi([p+i, p+i+1]) = S^1,$$

and f restricted to $\pi(\Gamma_1^i)$ is injective but for the identification at the ends.

Assume, by induction, we have defined points $a_{j-1}^0 = p < \cdots < a_{j-1}^{m^{j-1}} = p+1$. For $i = 0,\dots,m^{j-1}-1$ there are uniquely defined points

$$a_j^{mi} = a_{j-1}^i < a_j^{mi+1} < \cdots < a_j^{m(i+1)} = a_{j-1}^{i+1}$$

such that

$$F(a_j^{mi+l}) = a_{j-1}^{[mi+l]_{m^{j-1}}} \pmod 1. \tag{2.2.2}$$

Set then $H(i/m^j) = a_j^i$ and $\Gamma_j^i = [a_j^i, a_j^{i+1}]$ for $i = 0,\dots,m^j-1$. Clearly, $f^j\big(\pi(\Gamma_j^i)\big) = S^1$, and f^j restricted to $\pi(\Gamma_j^i)$ is injective but for the identification at the ends.

In other words, we have done the following: we started subdividing $[p, p+m]$ in m interval of length 1, and we subdivided $[p, p+1]$ using the inverse images via F of those intervals. Then we subdivided each $[p+i, p+i+1]$ as we did in $[p, p+1]$ working modulo 1; the inverse images via F provide a subdivision of the first subdivision of $[p, p+1]$, and so on.

In this way the map $H\colon \Sigma \to [p, p+1]$ given by $H(i/m^j) = a_j^i$ is strictly monotone. Since Σ is dense in $[0,1]$, we can extend H to a strictly monotone map $H\colon [0,1] \to [p, p+1]$ by setting

$$H(t) = \sup\{H(s) \mid s \in \Sigma,\ s < t\}.$$

The map H induces an invertible map $h\colon S^1 \to S^1$ such that $f \circ h = h \circ E_m$, thanks to (2.2.1) and (2.2.2). To end the proof we need to show that H is continuous.

Now H is not continuous only if we have $\sup\{H(s) \mid s \in \Sigma,\ s < t\} < H(t)$ for some $t \in \Sigma$; so to avoid this it suffices to prove that the set $\{a_j^i\}$ is dense in $[p, p+1]$. Since, as already remarked, $f^j\big(\pi(\Gamma_j^i)\big) = S^1$, and f^j restricted to $\pi(\Gamma_j^i)$ is injective but for the identification at the ends, Lemma 2.2.5 implies that the length of each Γ_j^i does not exceed μ^{-j} — and this is enough to prove that $\{a_j^i\}$ is dense in $[p, p+1]$.

Finally, a very similar agument works if m is negative; the only difference is that the relative order of the a_j^i will depend on the parity of j, exactly as it happens for numbers of the form i/m^j with m negative. □

Corollary 2.2.8: *Every C^1 expanding map of the circle is C^1-structurally stable.*

Proof: By Corollary 2.2.2, every $g \in C^1(S^1, S^1)$ which is C^1-close to an expanding map f of the circle is still expanding and, by Lemma 2.2.4, has the same degree. Therefore we can apply Theorem 2.2.7 to infer that they are both topologically conjugated to the same E_m, where $m = \deg(f)$, and we are done. □

Exercise 2.2.3. Prove that every expanding map of the circle of degree m is semiconjugate to the shift on $\Omega^+_{|m|}$. *Hint:* use the $|m|$-ary representation.

2.3 Recurrence and Smale's horseshoe

As we have seen, a characteristic of chaotic dynamical systems is recurrence: points almost go back to themselves. This happens, for instance, if there is a dense orbit, that is for topologically transitive systems. In this section we shall explore the notion of recurrence a bit further.

Definition 2.3.1: Let (X, f) be a dynamical system. A point $y \in X$ is a *ω-limit point* of a point $x \in X$ if there is a sequence $k_j \to +\infty$ such that $f^{k_j}(x) \to y$. The set of all ω-limit points of x is denoted by $\omega(x)$ and called its *ω-limit set.* We shall denote by $\omega(f)$ the closure of the union of all ω-limit sets as x varies in X. If f is invertible, the *α-limit set* $\alpha(x)$ of $x \in X$ is the ω-limit set of x with respect to f^{-1}. We denote by $\alpha(f)$ the closure of the union of all α-limit sets as x varies in X.

Definition 2.3.2: A point $x \in X$ is *recurrent* if $x \in \omega(x)$, that is if $f^{k_j}(x) \to x$ for some sequence $k_j \to +\infty$. If f is invertible, the point x is *negatively recurrent* if $x \in \alpha(x)$, and *fully recurrent* if it is both recurrent and negatively recurrent.

Exercise 2.3.1. Using the full shift σ_2 in Ω_2, show that there are recurrent points which are not negatively recurrent, and that the set of recurrent points is not closed.

A better notion of recurrence is:

Definition 2.3.3: Let (X, f) be a dynamical system. A point $x \in X$ is *non-wandering* if for any open neighbourhood U of x there is $N > 0$ such that $f^N(U) \cap U \neq \varnothing$. The set of all non-wandering point is denoted by $NW(f)$. A point which is not non-wandering is called *wandering.*

Remark 2.3.1. In a Hausdorff space, a pre-periodic point which is not periodic cannot be recurrent or non-wandering.

Proposition 2.3.1: *Let (X, f) be a dynamical system on a Hausdorff space X. Then:*

(i) *If $x \in X$ is non-wandering then for every neighbourhood U of x and $N_0 \in \mathbb{N}$ there is $N > N_0$ such that $f^N(U) \cap U \neq \varnothing$.*
(ii) *$NW(f)$ is closed.*
(iii) *$NW(f)$ is f-invariant. If f is open, $NW(f)$ is completely f-invariant.*
(iv) *If f is invertible, then $NW(f^{-1}) = NW(f)$.*
(v) *We have $\omega(f) \cup \alpha(f) \subseteq NW(f)$.*
(vi) *If X is compact then $NW(f) \neq \varnothing$.*
(vii) *If X has no isolated points and f is topologically transitive then $NW(f) = X$.*

Proof: (i) Assume there are a neighbourhood U of x and $N_0 \in \mathbb{N}$ so that $f^N(U) \cap U = \varnothing$ for all $N > N_0$. In particular, x is not periodic, and it cannot be pre-periodic because it is non-wandering by assumption. Hence, being X Hausdorff, for $j = 0, \ldots, N_0$ we can find a neighbourhood V_j of $f^j(x)$ such that $V_h \cap V_k = \varnothing$ if $h \neq k$. Set $V = U \cap \bigcap_{j=0}^{N_0} f^{-j}(V_j)$; then V is a neighbourhood of x such that $f^N(V) \cap V = \varnothing$ for all $N > 0$, and x is wandering.

(ii) It suffices to show that the complement is open. Let $x \notin NW(f)$; then there is an open neighbourhood U of x such that $f^k(U) \cap U = \varnothing$ for all $k > 0$. But then $U \cap NW(f) = \varnothing$, and $U \subseteq X \setminus NW(f)$, as desired.

(iii) Take $x \in NW(f)$ and let U be a neighbourhood of $f(x)$. Then since $V = f^{-1}(U)$ is a neighbourhood of x, there is $N > 0$ such that $f^N(V) \cap V \neq \varnothing$. Applying f we find $\varnothing \neq f\big(f^N(V) \cap V\big) \subseteq f^N(U) \cap U$, and thus $f(x) \in NW(f)$. If f is open, a similar argument shows $f^{-1}\big(NW(f)\big) \subseteq NW(f)$.

(iv) Take $x \in NW(f)$, and U a neighbourhood of x. Then there is $N > 0$ such that $f^N(U) \cap U \neq \varnothing$; applying f^{-N} we get $f^{-N}(U) \cap U \neq \varnothing$. Then $NW(f) \subseteq NW(f^{-1})$, and the other inclusion is completely analogous.

(v) Let $x \in X$, and $y = \lim_{j \to \infty} f^{k_j}(x) \in \omega(x)$; we can assume that k_j is increasing. Let U be an open neighbourhood of y; then $f^{k_j}(x) \in U$ eventually, and so $f^{k_{j+1} - k_j}(U) \cap U \neq \varnothing$ eventually. The same argument, if f is invertible, shows that $\alpha(x) \subseteq NW(f^{-1})$; the assertion then follows from (ii) and (iv).

(vi) If X is compact then $\omega(x) \neq \varnothing$ for any $x \in X$.

(vii) It follows from Proposition 1.4.3 □

Periodic points are the easiest example of non-wandering points, but not every dynamical system, even on compact spaces, has periodic points. The next best example of recurrence is found in f-invariant minimal subsets (that is, closed f-invariant subsets $Y \subseteq X$ such that every orbit is dense in Y); and they always exist in compact spaces.

Proposition 2.3.2: *Let (X, f) be a dynamical system on a compact Hausdorff space X. Then there exists an f-invariant minimal subset of X.*

Proof: Let $\mathcal{C}$ be the collection of all not empty closed f-invariant subsets of X, partially ordered by inclusion. The intersection of any family of closed f-invariant subsets is still closed and f-invariant; therefore any chain in $\mathcal{C}$ admits a lower bound (not empty by the compactness of X). We can therefore apply Zorn's lemma to get a minimal element Y of $\mathcal{C}$. Then Y is closed, f-invariant and contains no proper f-invariant closed subsets; since the closure of an orbit is such a subset, every orbit must be dense in Y, and thus $f|_Y$ is minimal. □

Exercise 2.3.2. Prove the previous proposition without using Zorn's lemma.

Exercise 2.3.3. Let $f\colon X \to X$ be a homeomorphism of a compact connected metric space X with no isolated points. Assume that periodic points of f are dense and that $f^k \neq \mathrm{id}$ for all $k > 0$. Prove that f has a nonperiodic recurrent point.

Now we want to study the nonwandering set of *the* classical example of hyperbolic dynamical system: *Smale's horseshoe.*

Let X denote a unit square in the plane, and let $\tilde{X}$ be the union of X and two half-disks on opposite sides of the square, H^+ and H^- (see Figure 1). Then let $f\colon \tilde{X} \to \tilde{X}$ be defined as follows: first of all, shrink X in the horizontal direction by a factor $0 < \lambda < 1/4$, shrinking at the same time isotropically the two half-disks by the same factor; then stretch the central rectangle in the vertical direction by a factor $\mu > 4$, acting on the two half-disks by a translation only; finally, bend the whole thing in a horseshoe shape and put it inside the original $\tilde{X}$ as shown in Figure 1. We would like to study the dynamics of f.

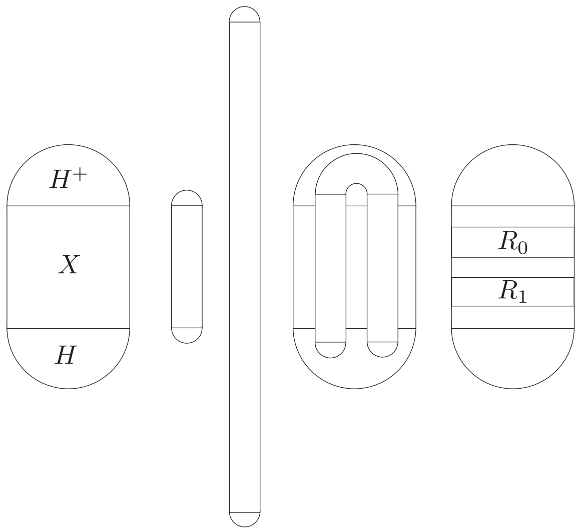

Figura 1 The horseshoe.

First of all, $f(H^-) \subset H^-$; since f acts on H^- by the composition of a linear contraction, a translation and a rotation, $f|_{H^-}$ is a contraction, and so $(f|_{H^-})^k$ converges to the unique fixed point p_0 of f in H^- by Theorem 1.2.1. Since $f(H^+) \subset H^-$, the orbit of every point of H^+ converges to p_0 too. This means that the orbit of $x \in \tilde{X}$ does not converge to p_0 if and only if $f^k(x) \in X$ for all $k \geq 0$; therefore we have described the dynamics of all points of $\tilde{X}$ but for the points of the closed set $\Lambda^- = \bigcap_{k\in\mathbb{N}} f^{-k}(X)$.

Now, it is clear that $f(\Lambda^-) \subseteq \Lambda^- \cap f(X) \subset \Lambda^-$; therefore every ω-limit set of a point of Λ^- is contained in the closed set $\Lambda = \bigcap_{k\in\mathbb{Z}} f^k(X)$. The set Λ is clearly completely f-invariant, $f|_\Lambda$ is a homeomorphism, and the non-trivial dynamics of f is concentrated on Λ — in the sense that every orbit not converging to p_0 is eventually in a neighbourhood of Λ, and the orbit of every $x \in \Lambda$ is contained in Λ. Our aim is then to describe the topological structure of Λ, and to give a model for the dynamics of $f|_\Lambda$.

Proposition 2.3.3: *Λ is the product of two Cantor sets (and hence a Cantor set itself), and there is a homeomorphism $h: \Omega_2 \to \Lambda$ such that $f|_\Lambda \circ h = h \circ \sigma_2$, that is $f|_\Lambda$ is conjugated to the full left 2-shift. Furthermore, $NW(f) = \Lambda \cup \{p_0\}$.*

Proof: It is easy to see that $f^{-1}(X) \cap X$ is the union of two horizontal rectangles R_0 and R_1 (see Figure 1). By induction it is then clear that $\bigcap_{k=0}^n f^{-k}(X)$ is the union of 2^n rectangles R_ω, where $s = (s_0, \ldots, s_{n-1}) \in \mathbb{Z}_2^n$ and $R_\omega = \bigcap_{j=0}^{n-1} f^{-j}(R_{s_j})$. In other words, s_j is such that $f^j(R_\omega) \subseteq R_{s_j}$ for all $j = 0, \ldots, n-1$.

Now, the height of every rectangle in $\bigcap_{k=0}^n f^{-k}(X)$ is μ^{-n}. This means that for every semi-infinite sequence $(s_0, s_1, \ldots) \in \mathbb{Z}_2^{\mathbb{N}}$ the intersection $\bigcap_{j\in\mathbb{N}} f^{-j}(R_{s_j})$ is a single horizontal segment. As a consequence, it is not difficult to see (exercise) that Λ^- is the product of a horizontal segment with a Cantor set.

A completely analogous argument shows that $\Lambda^+ = \bigcap_{j\in\mathbb{N}} f^j(X)$ is the product of a Cantor set with a vertical segment. Then $\Lambda = \Lambda^+ \cap \Lambda^-$ is the product of two Cantor sets. Furthermore, we can define a map $h: \Omega_2 \to \Lambda$ by setting $h(\mathbf{s}) = \bigcap_{j\in\mathbb{Z}} f^j(R_{s_{-j}})$. The previous discussion shows that h is bijective; furthermore

$$f\big(h(\mathbf{s})\big) = f\left(\bigcap_{j\in\mathbb{Z}} f^j(R_{s_{-j}})\right) = \bigcap_{j\in\mathbb{Z}} f^{j+1}(R_{s_{-j}}) = \bigcap_{j\in\mathbb{Z}} f^j(R_{s_{-(j-1)}}) = h\big(\sigma_2(\mathbf{s})\big).$$

The continuity of h and h^{-1} follows from the fact that $\mathbf{s}$ and $\mathbf{s}'$ belong to a symmetric cylinder of rank r if and only if $h(\mathbf{s})$ and $h(\mathbf{s}')$ belong to a rectangle of sides λ^r and μ^{-r} obtained as the intersection $\bigcap_{j=-r}^r f^j(R_{s_{-j}})$.

Finally, we are left to compute $NW(f)$. Clearly, p_0 is the only nonwandering point outside Λ^-. Since $NW(f)$ is f-invariant and $NW(f) \cap X \subseteq \Lambda^-$, we have $NW(f) \cap X \subseteq \Lambda$. To prove the converse it suffices to notice that all points of the full 2-shift are nonwandering, because, by Proposition 1.7.5, the full 2-shift is topologically mixing. □

Corollary 2.3.4: *$f|_\Lambda$ is chaotic, topologically mixing and it has 2^k periodic points of period k for all $k \geq 1$.*

Proof: It follows from Propositions 2.3.3 and 1.7.5. □

2.4 Stability of hyperbolic toral automorphisms

Our aim is to prove structural stability of hyperbolic toral automorphisms using a clever applicatino of the contraction principle. We shall need the following

Exercise 2.4.1. Prove that $C^0(\mathbb{T}^n, \mathbb{T}^n)$ is locally connected by arcs, and shows that this implies that the homotopy classes (i.e., the arc components) are open, and thus two C^0-close maps are necessarily homotopic.

Proposition 2.4.1: *Let $F_L: \mathbb{T}^2 \to \mathbb{T}^2$ be a hyperbolic automorphism of $\mathbb{T}^2$, and $g: \mathbb{T}^2 \to \mathbb{T}^2$ any homeomorphism in the same homotopy class of F_L. Then F_L is semiconjugate to g via a uniquely defined semiconjugacy h homotopic to the identity.*

Proof: We want a continuous surjective map $h: \mathbb{T}^2 \to \mathbb{T}^2$ homotopic to the identity such that $h \circ g = F_L \circ h$, that is

$$h = F_L^{-1} \circ h \circ g. \tag{2.4.1}$$

Now, any map of the torus into itself can be lifted to the universal cover $\mathbb{R}^2$; furthermore, a map $\hat{f}: \mathbb{R}^2 \to \mathbb{R}^2$ is a lift if and only if there is an endomorphism $A: \mathbb{Z}^2 \to \mathbb{Z}^2$ such that $\hat{f}(x+m) = \hat{f}(x) + A(m)$. In particular, the lifts of two homotopic maps (and thus the lift of two C^0-close maps) must have the same A (why?), and the lift of a map homotopic to the identity is of the form $\hat{h} = \mathrm{id} + \tilde{h}$ with $\tilde{h}$ doubly periodic, that is $\tilde{h}(x+m) = \tilde{h}(x)$ for all $x \in \mathbb{R}^2$ and $m \in \mathbb{Z}^2$.

Every lift of F_L is clearly of the form $L + c$ for a constant $c \in \mathbb{Z}^2$. Since we have $h \circ g = F_L \circ h$ if and only if $\hat{h} \circ \hat{g} = L \circ \hat{h} + c'$ for some constant $c' \in \mathbb{Z}^2$ (where $\hat{h}$ is a lift of h and $\hat{g}$ is a lift of g), up to a suitable choice of the lift of F_L we see that (2.4.1) has a unique solution homotopic to the identity if and only if

$$\hat{h} \circ \hat{g} = L \circ \hat{h} \tag{2.4.2}$$

has a unique solution of the form $\hat{h} = \mathrm{id} + \tilde{h}$ with $\tilde{h}$ doubly periodic. So it suffices to look for a unique solution of the latter equation.

Since g is homotopic to F_L, every lift of g is of the form $L + \tilde{g}$, where $\tilde{g}$ is doubly periodic. Then (2.4.2) becomes

$$\tilde{h} = L^{-1}\tilde{g} + L^{-1}\tilde{h} \circ (L + \tilde{g}). \tag{2.4.3}$$

So $\tilde{h}$ is the fixed point of an operator on a suitable space of maps. Unfortunately, this operator is not a contraction; but we can bypass this problem by using the structure of L.

Let e_1 and e_2 be two unit eigenvectors for L, with $Le_j = \lambda_j e_j$ and $|\lambda_1| = |\lambda_2^{-1}| > 1$. We can then write $\tilde{g} = \tilde{g}_1 e_1 + \tilde{g}_2 e_2$, and similarly for $\tilde{h}$. So (2.4.3) is equivalent to

$$\begin{cases} \tilde{h}_1 = \lambda_1^{-1}\tilde{g}_1 + \lambda_1^{-1}\tilde{h}_1 \circ (L + \tilde{g}) = \mathcal{F}_1(\tilde{h}_1), \\ \tilde{h}_2 = \lambda_2^{-1}\tilde{g}_2 + \lambda_2^{-1}\tilde{h}_2 \circ (L + \tilde{g}). \end{cases} \tag{2.4.4}$$

We consider $\mathcal{F}_1$ as operator on the Banach space E of doubly periodic functions on $\mathbb{R}^2$, endowed with the $\|\cdot\|_0$ norm. Now,

$$\begin{aligned} \|\mathcal{F}_1(\tilde{h}) - \mathcal{F}_1(\tilde{h}')\|_0 &= |\lambda_1|^{-1} \sup_{x\in\mathbb{R}^2} \big|\tilde{h}\big(Lx + \tilde{g}(x)\big) - \tilde{h}'\big(Lx + \tilde{g}(x)\big)\big| \\ &\le |\lambda_1|^{-1} \sup_{x\in\mathbb{R}^2} |\tilde{h}(x) - \tilde{h}'(x)| = |\lambda_1|^{-1}\|\tilde{h} - \tilde{h}'\|_0; \end{aligned}$$

therefore $\mathcal{F}_1$ is a contraction, and thus it has a unique fixed point $\tilde{h}_1$. We can also estimate the norm of $\tilde{h}_1$ as follows:

$$\|\tilde{h}_1\|_0 \le \sum_{j=0}^{\infty} \|\mathcal{F}_1^{j+1}(0) - \mathcal{F}_1^{j}(0)\|_0 \le \frac{1}{|\lambda_1| - 1}\|\tilde{g}_1\|_0.$$

In particular, $\|\tilde{h}_1\|_0$ is small if $\|\tilde{g}_1\|_0$ is small, that is if g is C^0-close to F_L.

To get $\tilde{h}_2$ we use a similar argument. The map $L+\tilde{g}$ is a homeomorphism, because g is; let $S = (L+\tilde{g})^{-1}$. Then the second equation in (2.4.4) becomes

$$\tilde{h}_2 = \lambda_2\tilde{h}_2 \circ S - \tilde{g}_2 \circ S = \mathcal{F}_2(\tilde{h}_2).$$

Then the same computations show that $\mathcal{F}_2$ is a contraction with a unique fixed point $\tilde{h}_2$ satisfying

$$\|\tilde{h}_2\|_0 \le \frac{1}{(1 - |\lambda_2|)}\|\tilde{g}_2\|_0.$$

We are left to proving that the map h is surjective. Since g is surjective, we know that $h(\mathbb{T}^2)$ is a closed connected completely F_L-invariant subset of $\mathbb{T}^2$. Since F_L is topologically transitive, if $h(\mathbb{T}^2)$ contains an open set it must be equal to $\mathbb{T}^2$, as desired. If, by contradiction, $h(\mathbb{T}^2)$ had empty interior, it must be either a fixed point or an invariant circle. But then h cannot induce the identity in homotopy, and thus it could not be homotopic to the identity, contradiction. □

This is not enough for structural stability: we need an actual conjugation. But first we show that if g is sufficiently close to F_L we can reverse the argument.

Proposition 2.4.2: *Let $F_L\colon \mathbb{T}^2 \to \mathbb{T}^2$ be a hyperbolic automorphism. Then for every C^1 map $g\colon \mathbb{T}^2 \to \mathbb{T}^2$ sufficiently C^1-close to F_L there exists a unique continuous map $h\colon \mathbb{T}^2 \to \mathbb{T}^2$ homotopic to the identity such that*

$$g \circ h = h \circ F_L. \tag{2.4.5}$$

Proof: Using the notations introduced in the previous proof, it is easy to see that solving (2.4.5) is equivalent to solving the equation $\mathcal{L}(\tilde{h}) = \mathcal{T}(\tilde{h})$ on the space of doubly periodic mappings, where the operators $\mathcal{L}$ and $\mathcal{T}$ are given by $\mathcal{L}(\tilde{h}) = \tilde{h} \circ L - L \circ \tilde{h}$ and $\mathcal{T}(\tilde{h}) = \tilde{g} \circ (\mathrm{id} + \tilde{h})$.

For transforming this equation in a fixed point equation, we notice that $\mathcal{L}$ is invertible. Indeed, we can write $\mathcal{L}(\tilde{h}) = \mathcal{L}_1(\tilde{h}_1)e_1 + \mathcal{L}_2(\tilde{h}_2)e_2$, where $\mathcal{L}_j(\tilde{h}_j) = \tilde{h}_j \circ L - \lambda_j \tilde{h}_j$, and then it is easy to write the inverses of the $\mathcal{L}_j$'s:

$$\mathcal{L}_1^{-1}(\tilde{h}_1) = -\sum_{n=0}^{\infty} \lambda_1^{-(n+1)} \tilde{h}_1 \circ L^n,$$
$$\mathcal{L}_2^{-1}(\tilde{h}_2) = \sum_{n=0}^{\infty} \lambda_2^n \tilde{h}_2 \circ L^{-(n+1)}.$$

Thus our equation is equivalent to the fixed-point equation $\tilde{h} = (\mathcal{L}^{-1}\mathcal{T})\tilde{h}$. Now $\tilde{g}$ is a C^1 map; therefore

$$\|\mathcal{T}(\tilde{h}) - \mathcal{T}(\tilde{h}')\|_0 = \sup_{x\in\mathbb{R}^2} \left|\tilde{g}\big(x + \tilde{h}(x)\big) - \tilde{g}\big(x + \tilde{h}'(x)\big)\right| \le \|d\tilde{g}\|_0 \|\tilde{h} - \tilde{h}'\|_0,$$

and thus

$$\|(\mathcal{L}^{-1}\mathcal{T})(\tilde{h}) - (\mathcal{L}^{-1}\mathcal{T})(\tilde{h}')\|_0 \le \|\mathcal{L}^{-1}\| \|d\tilde{g}\|_0 \|\tilde{h} - \tilde{h}'\|_0.$$

If g is sufficiently C^1-close to F_L we have $\|d\tilde{g}\|_0 < \|\mathcal{L}^{-1}\|^{-1}$, and hence the operator $\mathcal{L}^{-1}\mathcal{T}$ is a contraction, and thus it has a unique fixed point. □

Remark 2.4.1. Every continuous map $h: \mathbb{T}^2 \to \mathbb{T}^2$ homotopic to the identity is necessarily surjective. A way to see this is the following: a non-surjective map has degree zero, whereas a map homotopic to the identity has the same degree as the identity, that is 1. Thus the map provided by the previous proposition is a semiconjugation.

Theorem 2.4.3: *Any hyperbolic automorphism of the 2-torus is C^1-structurally stable.*

Proof: Let g be a C^1 map C^1-close to F_L; in particular, it is homotopic to F_L. Let h' be the semiconjugation provided by Proposition 2.4.1, and h'' the semiconjugation provided by Proposition 2.4.2. Then $F_L \circ h' = h' \circ g$, $g \circ h'' = h'' \circ F_L$ and so

$$F_L \circ (h' \circ h'') = (h' \circ g) \circ h'' = (h' \circ h'') \circ F_L.$$

In other words, $h' \circ h''$ commmutes with F_L, and it is homotopic to the identity. If we prove that it is the identity, it will follow that h'' is injective, and hence invertible; then $h' = (h'')^{-1}$, and we are done.

Therefore to end the proof it suffices to show that a continuous map $h: \mathbb{T}^2 \to \mathbb{T}^2$ satisfying the fixed-point equation $h = F_L^{-1} \circ h \circ F_L$ and homotopic to the identity is the identity. Proceeding as in the previous proofs, we can write a lift of h in the form $\mathrm{id} + \tilde{h}$, where $\tilde{h}$ is doubly periodic and satisfies $\tilde{h} = L^{-1} \circ \tilde{h} \circ L$. Writing $\tilde{h} = \tilde{h}_1 e_1 + \tilde{h}_2 e_2$, we see that $\tilde{h}_j$ is a fixed point of the operator $\mathcal{L}_j$ given by

$$\mathcal{L}_1(h) = \frac{1}{\lambda_1} h \circ L, \qquad \mathcal{L}_2(h) = \lambda_2 h \circ L^{-1}.$$

It is absolutely clear that these operators are contractions; furthermore, the zero function is a fixed point of both. Then the uniqueness of the fixed point implies $\tilde{h} \equiv 0$, and thus $h = \mathrm{id}$. □

Exercise 2.4.2. Prove that any hyperbolic automorphism of the n-torus is C^1 strongly structurally stable.

Exercise 2.4.3. Let $f: \tilde{X} \to \tilde{X}$ be the horseshoe map described in Section 2.3, and let $g: \tilde{X} \to \mathbb{R}^2$ be any C^1 map sufficiently C^1 close to f. Prove that there is an injective continuous map $h = h_g: \Lambda \to \tilde{X}$ such that $g \circ h_g = h_g \circ f$. Deduce that $\Lambda_g = h_g(\Lambda)$ is a closed g-invariant set and that $g|_{\Lambda_g}$ is topologically conjugate to the full left 2-shift σ_2.

2.5 Sarkovskii's Theorem

We end this chapter with an interesting result of a different nature, and tipically one-dimensional. We have seen functions $f\colon\mathbb{R}\to\mathbb{R}$ with a single fixed point and no other periodic points; for instance, $f(x)=\frac{1}{2}x$. On the other hand, it is easy to see that if f has a periodic point of period 2, then it must have a fixed point: indeed, if $a_1<a_2$ are such that $f(a_1)=a_2$ and $f(a_2)=a_1$, then $f(a_1)-a_1<0$ and $f(a_2)-a_2>0$, and hence $f-\mathrm{id}$ must have a zero in the interval (a_1,a_2).

What happens if f has a periodic point of (exact) period 3? The surprising answer is that then f must necessarily have periodic points of any (exact) period! To prove this we shall need a definition and two lemmas.

Definition 2.5.1: Let $f\colon I\to\mathbb{R}$ be a continuous function, where I is a closed interval in $\mathbb{R}$. If $J\subseteq\mathbb{R}$ is another closed interval, we shall say that I *covers* J (and we shall write $I\to J$) if $f(I)\supseteq J$.

Lemma 2.5.1: *Let $f\colon J\to\mathbb{R}$ be a continuous function defined on an interval $J\subseteq\mathbb{R}$. If $I_0\subseteq J$ and $I_1\subset\mathbb{R}$ are closed bounded intervals so that $I_0\subseteq I_1$ and $I_0\to I_1$ then f has a fixed point in I_0.*

Proof: Write $I_0=[a,b]$ and $I_1=[c,d]$ with $c\le a<b\le d$. Since I_0 covers I_1 we must have either

$$f(a)\le c\le a<b\le d\le f(b)\qquad\text{or}\qquad f(b)\le c\le a<b\le d\le f(a).$$

In the first case we have $f(a)-a<0$ and $f(b)-b>0$, in the second case $f(a)-a>0$ and $f(b)-b<0$; in both cases $f-\mathrm{id}$ must have a zero in (a,b). □

Lemma 2.5.2: *Let $f\colon J\to\mathbb{R}$ be a continuous function defined on an interval $J\subseteq\mathbb{R}$. Let $I_0,\dots,I_n\subseteq J$ be a sequence of closed bounded intervals such that $I_0\to I_1\to\cdots\to I_n$. Then there exists a closed interval $A_0\subseteq I_0$ so that $f^j(A_0)\subseteq I_j$ for $j=0,\dots,n-1$ and $f^n(A_0)=I_n$.*

Proof: By induction on n. For $n=1$, let $f^{-1}(I_1)\cap I_0=\bigcup_\lambda A_\lambda$ be the decomposition in connected components (i.e., closed intervals). If $I_1=[a,b]$, let $\Lambda_b=\{\lambda\mid b\in f(A_\lambda)\}$ and $a'=\inf\bigcup_{\lambda\in\Lambda_b}f(A_\lambda)$. Using the compactness it is easy to see that there exists $\lambda\in\Lambda_b$ such that $f(A_\lambda)=[a',b]$, and it is not difficult to check (exercise) that necessarily $a'=a$, so that $A_\lambda=A_0$ is as desired.

Assume the assertion is true for $n-1$. Then there exists a closed interval $A_1\subseteq I_1$ such that $f^j(A_1)\subseteq I_{j+1}$ for $j=0,\dots,n-2$ and $f^{n-1}(A_1)=I_n$. Since I_0 covers A_1, we can find a closed interval $A_0\subseteq I_0$ such that $f(A_0)=A_1$, nad clearly A_0 is as desired. □

Corollary 2.5.3: *Let $f\colon J\to\mathbb{R}$ be a continuous function defined on an interval $J\subseteq\mathbb{R}$. Let $I_0,\dots,I_n\subseteq J$ be a sequence of closed bounded intervals such that $I_0\to I_1\to\cdots\to I_n\to I_0$. Then there exists a point $p_0\in I_0$ of period $n+1$ such that $f^j(p_0)\in I_j$ for $j=0,\dots,n$.*

Proof: It suffices to apply Lemma 2.5.1 to the interval A_0 given by Lemma 2.5.2. □

Then:

Theorem 2.5.4: *Let $I\subseteq\mathbb{R}$ an interval, and $f\colon I\to I$ continuous. Assume that f has a point of exact period 3. Then f has periodic points of any exact period.*

Proof: Let a, b, $c\in I$ be such that $f(a)=b$, $f(b)=c$ and $f(c)=a$. Assume that $a<b<c$; the other case $(a>b>c)$ will be analogous.

Put $I_0=[a,b]$ and $I_1=[b,c]$; we clearly have

$$I_0\rightleftarrows I_1\rightleftarrows I_1.\tag{2.5.1}$$

In particular, I_1 covers itself under f and I_0 covers itself under f^2; therefore Lemma 2.5.1 yields a fixed point of f in I_1 and a fixed point of f^2 in I_0. Since $f(I_0)\cap I_0=\{b\}$ and $f(b)\ne b$, the fixed point of f^2 in I_0 is a periodic point of f of exact period 2.

To get a fixed point of exact period $n\ge3$, we first remark that (2.5.1) yields a sequence

$$I_0\to I_1\to I_1\to\cdots\to I_1\to I_0$$

of length n (that is, with n arrows). Lemma 2.5.2 yields an interval $A_0\subseteq I_0$ such that $f^j(A_0)\subseteq I_1$ for $j=1,\dots,n-1$ and $f^n(A_0)=I_0$, and Lemma 2.5.1 yields a periodic point $p_0\in A_0$ of period n. If we had $f^j(p_0)=p_0$ for some $1\le j\le n-1$, the orbit of p_0 should be completely contained in I_1; in particular, $p_0\in I_0\cap I_1$, that is $p_0=b$. But $f^2(b)=a\notin I_1$, contradiction. □

This is just the beginning.

Definition 2.5.2: The *Sarkovskii order* $\rhd$ on $\mathbb{N}^*$ is defined as follows: writing $h_1 = 2^{l_1}p_1$ and $h_2 = 2^{l_2}q_2$ with p_1, p_2 odd numbers, one has

$$h_1 \rhd h_2 \qquad \text{if and only if} \qquad \begin{cases} l_1 < l_2 & \text{if } p_1,\, p_2 > 1, \text{ or} \\ p_1 < p_2 & \text{if } p_1,\, p_2 > 1 \text{ and } l_1 = l_2, \text{ or} \\ p_1 > p_2 = 1 & \text{if } p_1 > 1 \text{ and } p_2 = 1, \text{ or} \\ l_1 > l_2 & \text{if } p_1 = p_2 = 1. \end{cases}$$

In other words, the Sarkovskii order is

$$3 \rhd 5 \rhd 7 \rhd \cdots \rhd 2\cdot 3 \rhd 2\cdot 5 \rhd 2\cdot 7 \rhd \cdots \rhd 2^2\cdot 3 \rhd 2^2\cdot 5 \rhd \cdots \rhd 2^3\cdot 3 \rhd 2^3\cdot 5 \rhd \cdots\cdots \rhd 2^3 \rhd 2^2 \rhd 2 \rhd 1.$$

Then:

Theorem 2.5.5: (Sarkovskii) *Let $I \subseteq \mathbb{R}$ an interval, and $f\colon I \to I$ continuous. Assume that f has a point of exact period h. Then f has periodic points of exact period k for all $k \lhd h$.*

Proof: It suffices to prove the assertion assuming that f has no periodic points of exact period greater than h in the Sarkovskii order.

Take $x \in I$ of exact period h, and let $O^+(x) = \{x_1, \ldots, x_h\}$, with $x_1 < \cdots < x_h$. The function f acts on $O^+(x)$ as a permutation. We clearly have $f(x_h) < x_h$ and $f(x_1) > x_1$; let $1 \le j < h$ be the largest index such that $f(x_j) > x_j$, and set $I_1 = [x_j, x_{j+1}]$. We have $f(x_j) \ge x_{j+1}$ and $f(x_{j+1}) \le x_j$; hence I_1 covers I_1. In particular, f has a fixed point in I_1.

Since $h \ne 2$, we cannot have $f(x_{j+1}) = x_j$ and $f(x_j) = x_{j+1}$; therefore $f(I_1)$ must contain another interval of the form $[x_k, x_{k+1}]$, that we shall call I_2. Analogously, $f(I_2)$ must contain another interval of the same form, that we shall call I_3; and so on. Thus we have obtained a sequence

$$I_1 \rightleftarrows I_1 \to I_2 \to I_3 \to \cdots.$$

Now let us consider several cases.

(i) *h odd.* Since h is odd, at least one x_k must be sent by f on the opposite side with respect to I_1, and at least one x_k must stay on the same side. This means that sooner or later we must have $f(I_k) \supseteq I_1$; let ℓ be the *minimum* integer so that $I_\ell \to I_1$. Therefore we have

$$I_1 \rightleftarrows I_1 \to I_2 \to I_3 \to \cdots \to I_\ell \to I_1.$$

Now, if $\ell < h-1$ then either

$$I_1 \to I_2 \to \cdots \to I_\ell \to I_1 \qquad \text{or} \qquad I_1 \to I_2 \to \cdots \to I_\ell \to I_1 \to I_1$$

yields a periodic point p of odd period $r < h$. Since $I_1 \cap I_2$ has cardinality at most 1, and if there is an intersection point it has period h, the point p must have exact period odd, less than h and greater than 1, that is exact period greater than h in the Sarkovskii order, against the assumption.

So $\ell = h-1$. Since ℓ is minimal, we cannot have $I_r \to I_s$ for some $s > r+1$. Therefore the sequence x_j, $f(x_j)$, $f^2(x_j), \ldots, f^h(x_j)$ bounces at each step from one side to the other of I_1; in particular, I_{h-1} must cover all I_k with k odd. Then a period h' larger (in the standard order) than h is obtained applying Corollary 2.5.3 to a sequence

$$I_1 \to I_2 \to \cdots \to I_{h-1} \to I_1 \to I_1 \to \cdots \to I_1;$$

of length h'; notice that such a periodic point has exact period h' because otherwise one of the points in the orbit would belong to $I_1 \cap I_2$, and they do not travel according to the previous sequence.

To get an even period $2j$ smaller (in the standard order) than h we apply Corollary 2.5.3 to the sequence

$$I_{h-1} \to I_{h-2j} \to I_{h-2j+1} \to \cdots \to I_{h-2} \to I_{h-1}$$

which has length $2j$, and that cannot produce a lower period.

(ii) *h even larger than* 2. If all x_j's stay on the same side of I_1 under the action of f, we necessarily have $f(x_j) = x_h$ and we can take $I_2 = [x_{j-1}, x_j]$. So $f(x_{j-1}) < x_{j-1}$ and I_2 covers I_1; we have gained a periodic point of period 2. If all x_j's are sent by f on the opposite side of I_1, we must have $f([x_1, x_j]) \supseteq [x_{j+1}, x_h]$ and $f([x_{j+1}, x_k]) \supseteq [x_1, x_j]$, and again we get a point of period 2. If there are points staying on the same side and points going on the opposite side, then we may repeat the previous argument obtaining a sequence $I_{h-1} \to I_{h-2} \to I_{h-1}$, and again a point of period 2.

(ii.1) $h = 2^m$ *with* $m \geq 2$. Put $n = 2^\ell$ with $1 \leq \ell < m$ and $g = f^{n/2}$. By assumption, g has a periodic point of exact period $2^{m-\ell+1}$, and hence has a periodic point of exact period 2. Such a point has exact period 2^l for f, and we are done in this case.

(ii.2) $h = 2^m p$ *with* $m \geq 1$ *and* $p \geq 3$ *odd.* Let $g = f^{2^m}$. Then g has a point of exact period p odd; then it has points of exact period q for all odd $q > p$, which means that f has points of exact period $2^m q$ for all odd $q > p$; the period must be exact because, by assumption, f has no periodic points of period $2^l r$ with $l < m$ and r odd.

Let now $g = f^p$. Then g has a point of exact period 2^m, and hence points of exact period 2^l for all $l \leq m$. Therefore f has points of exact period dividing $2^l p$ for $l \leq m$. But since f has no periodic points of period larger (in the Sarkovskii order) of h, the only possibility is that the exact period must divide 2^l — and hence it must be equal to 2^l.

Let us put again $g = f^{2^m}$. Then g has a point x of exact period $2^l q$ for any $l \geq 1$ and $q \geq 3$ odd. Hence x is a point of period $2^{l+m} q$ for f, and hence a point of exact period $2^s r$ for f, where $s \leq l + m$ and r is an odd divisor of q. Since f has no periodic points of period larger (in the Sarkovskii order) of h, we necessarily have $s \geq m$. But then $g^{2^{s-m} r}(x) = x$ implies that $2^l q$ divides $2^{s-m} r$; therefore $r = q$ and $s - m \geq l$. It follows that $s = l + m$, that is x has exact period $2^{l+m} q$ for f, and we are done. □

Remark 2.5.1. As a consequence, if f has a periodic point of exact period which is not a power of 2, then it must have infinite periodic points. In other words, if a continuous function $f: I \to I$ has finitely many periodic points, all periods must be powers of 2.

Remark 2.5.2. This theorem is false for continuous maps of S^1; it suffices to consider rational rotations.

We end this section showing that Sarkovskii's theorem is optimal: if $h \lhd k$ there is a continuous map $f: I \to I$ admitting a periodic point of exact period h and no periodic points of exact period k.

EXAMPLE 2.5.1. Let $f: [1, 5] \to [1, 5]$ be piecewise linear such that

$$f(1) = 3, \quad f(3) = 4, \quad f(4) = 2, \quad f(2) = 5, \quad f(5) = 1.$$

in particular, $f^5(1) = 1$; we shall show that f has no periodic points of exact period 3. First of all,

$$f^3([1, 2]) = [2, 5], \quad f^3([2, 3]) = [3, 5], \quad f^3([4, 5]) = [1, 4],$$

and so $\mathrm{Fix}(f^3) \subset [3, 4]$. But $f: [3, 4] \to [2, 4]$ is decreasing, as well as $f: [2, 4] \to [2, 5]$ and $f: [2, 5] \to [1, 5]$; therefore $f^3: [3, 4] \to [1, 5]$ is decreasing too, and thus it has a unique fixed point. Since f has a fixed point in $[3, 4]$, it follows that $\mathrm{Fix}(f^3) = \mathrm{Fix}(f)$, and f has no periodic points of exact period 3.

Exercise 2.5.1. Given $k \geq 2$ find a piecewise linear continuous function $f: [1, 2k + 3] \to [1, 2k + 3]$ with a periodic point of exact period $2k + 3$ and no periodic point of exact period $2k + 1$.

To deal with the even periods, we need the following notion.

Definition 2.5.3: Let $f: [0, 1] \to [0, 1]$ be a continuous function. The *double* of f is the continuos function $\hat{f}: [0, 1] \to [0, 1]$ given by

$$\hat{f}(x) = \begin{cases} \frac{2}{3} + \frac{1}{3} f(3x) & \text{if } 0 \leq x \leq 1/3, \\ \big(f(1) + 2\big)\big(\frac{2}{3} - x\big) & \text{if } 1/3 \leq x \leq 2/3, \\ x - \frac{2}{3} & \text{if } 2/3 \leq x \leq 1. \end{cases}$$

In particular, $\hat{f}([0, 1/3]) \subseteq [2/3, 1]$, $\hat{f}([2/3, 1]) = [0, 1/3]$ and $|\hat{f}'(x)| > 1$ for all $x \in (1/3, 2/3)$.

Exercise 2.5.2. Let $\hat{f}: [0,1] \to [0,1]$ be the double of a continuous function $f: [0,1] \to [0,1]$.

(i) Prove that $\hat{f}$ has a unique fixed point, which is repelling and belongs to $(1/3, 2/3)$.
(ii) Prove that $\hat{f}$ has no other periodic points in $(1/3, 2/3)$.
(iii) Prove that $x \in [0,1]$ is a periodic point for f of period k if and only if $x/3$ is a periodic point for $\hat{f}$ of period $2k$.
(iv) Prove that all periodic points of $\hat{f}$ in $[0, 1/3] \cup [2/3, 1]$ have even period.
(v) Prove that if f has a periodic point of exact period $2^l q$ and no periodic point of exact period $2^l p$ for some $l \geq 0$ and $p < q$, both odd, then $\hat{f}$ has a periodic point of exact period $2^{l+1} q$ and no periodic point of exact period $2^{l+1} p$.

Exercise 2.5.3. Given $l \geq 0$, find a continuous function $f: [0,1] \to [0,1]$ with a periodic point of exact period 2^l and no periodic point of exact period 2^{l+1}.