THE SHARKOVSKY THEOREM: A NATURAL DIRECT PROOF

KEITH BURNS AND BORIS HASSELBLATT

ABSTRACT. We give a proof of the Sharkovsky Theorem that is self-contained, short, direct and naturally adapted to the doubling structure of the Sharkovsky ordering.

1. INTRODUCTION

In this note f is a continuous function from an interval into $\mathbb{R}$; although this is usually assumed in the literature, the interval need not be closed or bounded. f^n denotes the n-fold composition of f with itself. A point p is a periodic point for f if $f^n(p) = p$ for some $n > 0$. By the *period* of a periodic point p we mean the smallest positive integer m such that $f^m(p) = p$. In other words, the period of p is the number of distinct points in the *orbit* or *cycle* $\mathcal{O} := \{f^k(p) \mid \in \mathbb{N}\}$. If p has period m, then $f^n(p) = p$ if and only if n is a multiple of p.[1] A *fixed point* is a periodic point of period 1. If f has a periodic point of period m, then m is called a *period for* (or *of*) f.

1.1. The Sharkovsky Theorem.

The Sharkovsky Theorem involves the following ordering of the set $\mathbb{N}$ of positive integers, which is known as the Sharkovsky ordering:

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

This is a total ordering; we write $m \rhd l$ or $l \lhd m$ whenever m is to the left of l. Note that $m \rhd l \Leftrightarrow 2m \rhd 2l$ because the odd numbers greater than 1 appear in decreasing order from the left end of the list, the number 1 appears at the right end, and the rest of $\mathbb{N}$ is included by successively doubling these end pieces, and inserting these doubled strings inward:

$$\text{odds}, 2\cdot\text{odds}, 2^2\cdot\text{odds}, 2^3\cdot\text{odds}, \dots, 2^3\cdot 1, 2^2\cdot 1, 2\cdot 1, 1.$$

Sharkovsky showed that this ordering describes which numbers can be periods for a continous map of an interval.

Date: May 14, 2008.

[1]Dynamicists usually refer to m as the *least* period and, unlike in the present paper, call any n for which $f^n(x) = x$ *a* period.

Theorem 1.1 (Sharkovsky Forcing Theorem [S1])**.** *If m is a period for f and $m \rhd l$, then l is also a period for f.*

This shows that the set of periods of a continuous interval map is a *tail* of the Sharkovsky order. A tail is a set $\mathcal{T} \subset \mathbb{N}$ such that $s \rhd t$ for all $s \notin \mathcal{T}$ and all $t \in \mathcal{T}$. There are three types of tail: $\{m\} \cup \{l \in \mathbb{N} \mid l \lhd m\}$ for some $m \in \mathbb{N}$, the set $\{\dots, 16, 8, 4, 2, 1\}$ of all powers of 2, and $\varnothing$.

The following complementary result is sometimes called the converse to the Sharkovsky Theorem, but is proved in Sharkovsky's original papers.

Theorem 1.2 (Sharkovsky Realization Theorem [S1])**.** *Every tail of the Sharkovsky order is the set of periods for a continous map of an interval into itself.*

The Sharkovsky Theorem is the union of Theorem 1.1 and Theorem 1.2: A subset of $\mathbb{N}$ is the set of periods for a continuous map of an interval to itself if and only if the set is a tail of the Sharkovsky order. We reproduce a proof of the Realization Theorem in Section 5 at the end of this note.

Our aim is to present, with all details, a direct proof of the Forcing Theorem that is conceptually simple, short (Subsections 4.1 and 4.3) and involves no artificial case distinctions.

The standard proof of the Sharkovsky Forcing Theorem begins by studying orbits of odd period with the property that their period comes earlier in the Sharkovsky sequence than any other periods for that map. It shows that such an orbit is of a special type, known as a Štefan cycle, and then that such a cycle forces the presence of periodic orbits with Sharkovsky-lesser periods. The second stage of the proof then considers various cases in which the period that comes earliest in the Sharkovsky order is even.

We extract the essence of the first stage of the standard proof to produce an argument that does not need Štefan cycles, and we replace the second stage of the standard proof by a simple and natural induction.

Our main idea is to select a salient sequence of orbit points and to prove that this sequence "spirals out" in essentially the same way as the Štefan cycles considered in the standard proof.

1.2. **History.** A capsule history of the Sharkovsky Theorem is in [M], and [ALM] provides much context. The first result in this direction was obtained by Coppel [C] in the 1950s: every point converges to a fixed point under iteration of a continuous map of a closed interval if the map has no periodic points of period 2; it is an easy corollary that a continuous map must have 2 as a period if it has any periodic points that are not fixed. This amounts to 2 being the penultimate number in the Sharkovsky ordering.

Sharkovsky obtained the results described above and reproved Coppel's theorem in a series of papers published in the 1960s [S1]. He also worked on other aspects of one-dimensional dynamics (see, for instance, [S2]). Sharkovsky appears to have been unaware of Coppel's paper. His work did not become known outside eastern Europe until the second half of the 1970s. In 1975 this Monthly published a famous paper *Period three implies chaos* [LY] by Li and Yorke with the result that the presence of a periodic point of period 3 implies the presence of periodic points of all other periods. This amounts to 3 being the initial number in the Sharkovsky order. Some time after its publication, Yorke attended a conference in East Berlin, and during a river cruise a Ukrainian participant approached him. Although they had no language in common, Sharkovsky (for it was he) managed to convey, with translation by Lasota and Mira, that unbeknownst to Li and Yorke (and perhaps all of western mathematics) he had proved his results about periodic points of interval mappings well before [LY], even though he did not at the time care to say what that result was.

Besides introducing the idea of chaos to a wide audience, Li and Yorke's paper was to lead to global recognition of Sharkovsky's work[2]. Within a few years of [LY] new proofs of the Sharkovsky Forcing Theorem appeared, one due to Štefan [Š], and a later one, which is now viewed as the "standard" proof, due to Block, Guckenheimer, Misiurewicz and Young [BGMY][3], Burkart [B], Ho and Morris [HM] and Straffin [St]. Nitecki's paper [N] provides a lovely survey from that time. Alsedà, Llibre and Misiurewicz improved this standard proof [ALM] and also gave a beautiful proof of the realization theorem, which we reproduce in Section 5.

The result has also been popular with contributors to the Monthly. We mention here a short proof of one step in the standard proof [BB] and several papers by Du [D]. Reading the papers by Du inspired the work that resulted in this article.

1.3. **Related work.** There is a wealth of literature related to periodic points for 1-dimensional dynamical systems. [ALM] is a good source of pertinent information. There is a characterization of the exact structure of a periodic orbit whose period is Sharkovsky-maximal for a specific map. There is also work on generalizations to other permutation patterns (how particular types of periodic points force the presence of others, and how intertwined periodic orbits do so), to different one-dimensional spaces (shaped like "Y", "X" or "∗", say) and to multivalued maps.

[2]It should not be forgotten that Li and Yorke's work contains more than a special case of Sharkovsky's: "chaos" is not just "periods of all orders".

[3]This citation is often pronounced "bigamy".

2. Cycles, intervals and covering relations

Definition 2.1. We say that an interval I *covers* an interval J and write $I \to J$ if $J \subset f(I)$. An interval whose endpoints are in a cycle $\mathcal{O}$ of f is called an $\mathcal{O}$*-interval.* If it contains only two points of $\mathcal{O}$ then it is called a *basic* $\mathcal{O}$-interval, and these two points are said to be *adjacent.*

One of the basic ingredients is that, by the Intermediate Value Theorem, $I \to J$ whenever f maps the endpoints of an interval I to opposite sides of an interval J. Thus knowledge about a cycle $\mathcal{O}$ engenders knowledge about how $\mathcal{O}$-intervals are moved around by the map.

The other basic idea is that this knowledge of how intervals are moved around in turn produces information about the presence of other periodic points. This is the content of the next three lemmas. They hold for any closed bounded intervals, although we apply them to $\mathcal{O}$-intervals.

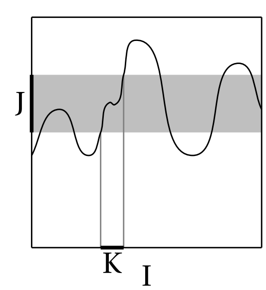
Lemma 2.2. *If* $[a_1, a_2] \to [a_1, a_2]$*, then* f *has a fixed point in* $[a_1, a_2]$.

Proof. Take $b_1, b_2 \in [a_1, a_2]$ with $f(b_i) = a_i$. Then $f(b_1) - b_1 \le 0 \le f(b_2) - b_2$. By the Intermediate-Value Theorem $f(x) - x = 0$ for some x between b_1 and b_2. □

Lemma 2.3 (Itinerary Lemma)**.** *If* $J_0 \dots, J_{n-1}$ *are closed bounded intervals and* $J_0 \underleftarrow{\to \cdots \to} J_{n-1}$ *(this is called a* loop *or* n-loop of intervals*) then there is a fixed point* p *of* f^n *such that* $f^i(p) \in J_i$ *for* $0 \le i < n$.

We say that a point p *follows the loop* if it satisfies the conclusion of the lemma. Note that the period of p must be a factor of n if it follows a loop of length n.

Proof. We write $I \rightarrowtail J$ if $f(I) = J$. If $I \to J$, there is an interval $K \subset I$ such that $K \rightarrowtail J$ because the intersection of the graph of f with the rectangle $I \times J$ contains an arc that joins the top and bottom sides of the rectangle. We can choose K to be the projection to I of such an arc.



Thus, there is an interval $K_{n-1} \subset J_{n-1}$ such that $K_{n-1} \rightarrowtail J_0$. Then $J_{n-2} \to K_{n-1}$, and there is an interval $K_{n-2} \subset J_{n-2}$ such that $K_{n-2} \rightarrowtail K_{n-1}$. Inductively, there are intervals $K_i \subset J_i$, $0 \le i < n$, such that

$$K_0 \rightarrowtail K_1 \rightarrowtail \cdots \rightarrowtail K_{n-1} \rightarrowtail J_0.$$

Any $x \in K_0$ satisfies $f^i(x) \in K_i \subset J_i$ for $0 \le i < n$ and $f^n(x) \in J_0$. Since $K_0 \subset J_0 = f^n(K_0)$, Lemma 2.2 implies that f^n has a fixed point in K_0. $\square$

We wish to ensure that the period of the point p found in Lemma 2.3 is n and not a proper divisor of n, such as for the 2-loop $[-1,0] \rightleftarrows [0,1]$ of $f(x) = -2x$, which is followed only by the fixed point 0.

Definition 2.4. We say that a loop $J_0 \to \cdots \to J_{n-1}$ of intervals is *elementary* if every p that follows it has period n.[4]

With this notion, the conclusion of Lemma 2.3 gives us:

Proposition 2.5. *The presence of an elementary loop $J_0 \to \cdots \to J_{n-1}$ implies the existence of a periodic point p with period n that follows the loop.*

This makes it interesting to give convenient criteria for being elementary. The simplest is that any loop of length 1 is elementary (since the period of a point that follows such a loop must be a factor of 1). A criterion with wider utility is:

Lemma 2.6. *A loop $J_0 \to \cdots \to J_{n-1}$ of $\mathcal{O}$-intervals is elementary if it is not followed by a point of $\mathcal{O}$ and the interior* $\mathrm{Int}(J_0)$ *of J_0 is disjoint from each of $J_1, \dots, J_{n-1}$.*

Proof. If $f^n(p) = p$ and p follows the loop, then $p \notin \mathcal{O}$, so $p \in \mathrm{Int}(J_0)$. If $0 < i < n$ then $f^i(p) \notin \mathrm{Int}(J_0)$, so $p \neq f^i(p)$. Thus p has period n. $\square$

3. Warmup

This section contains three examples that illustrate the arguments that will be used in the subsequent proof of the Sharkovsky Forcing Theorem. This section is not part of the proof and can be skipped, but it may serve to emphasize the ideas of the proof.

The first example is the most celebrated special case of the Sharkovsky Theorem: that period 3 implies all periods. The second example applies the same method to a longer cycle and illustrates how our choice of $\mathcal{O}$-intervals differs from that made in the standard proof. The third example illustrates our induction argument, which is built on the doubling structure of the Sharkovsky order.

3.1. **Period 3 implies all periods.** A 3-cycle comes in two versions that are mirror images of one another:

[4]This is a different use of the word "elementary" from the one in [ALM].

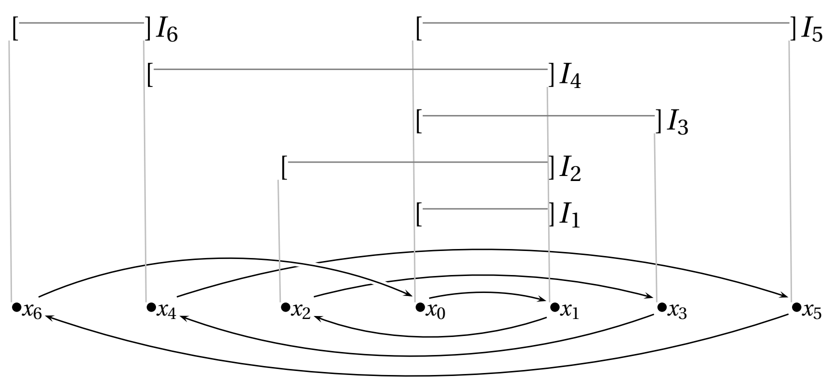
For the cycle shown on the left we denote the left and right intervals between the dots by I_2 and I_1, respectively. For the cycle on the right we make the opposite choice, as shown above. Then $I_1 \to I_1$, $I_1 \to I_2$ and $I_2 \to I_1$. We also write this more graphically as $\circlearrowleft I_1 \rightleftarrows I_2$. Since $I_1 \to I_1$, it follows from Lemma 2.2 that I_1 contains a fixed point of f. The points of $\mathcal{O}$ cannot follow the cycle $I_1 \to I_2 \to I_1$ because they are periodic points with least period 3 whereas a point that follows this cycle must have least period 1 or 2. It follows from Lemma 2.6 that f has an orbit with least period 2. Points of $\mathcal{O}$ cannot stay in the interval I_1 for more than two consecutive iterates of f. Hence the loop

$$I_2 \to \overbrace{I_1 \to I_1 \to \cdots \to I_1}^{l-1 \text{ copies of } I_1} \to I_2$$

is elementary if $l > 3$. By Lemma 2.6, f has a periodic point of least period l for each $l > 3$.

This shows a special case of the Sharkovsky Theorem: the presence of a period-3 orbit causes every positive integer to be a least period.

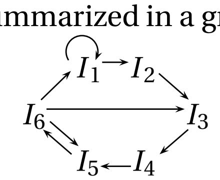
3.2. **A 7-cycle.** Consider a 7-cycle $\mathcal{O}$ and a choice of $\mathcal{O}$-intervals as follows:



With this choice of intervals we get the following covering relations:

(1) $I_1 \to I_1$,
(2) $I_1 \to I_2 \to I_3 \to I_4 \to I_5 \to I_6 \to I_1$, and
(3) $I_6 \to I_5, I_3, I_1$.

This information can be summarized in a graph as follows:



From this graph we read off the following loops.

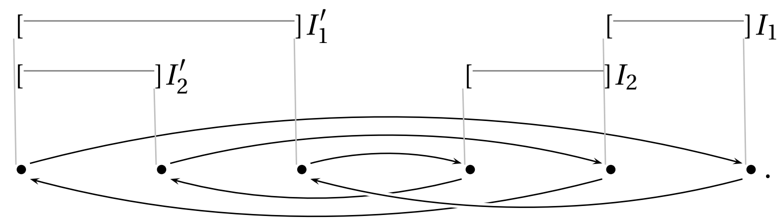
(4) $I_1 \to I_1$,

(5) $I_6 \to I_5 \to I_6$,
(6) $I_6 \to I_3 \to I_4 \to I_5 \to I_6$,
(7) $I_6 \to I_1 \to I_2 \to I_3 \to I_4 \to I_5 \to I_6$,
(8) $I_6 \to I_1 \to I_1 \to \cdots \to I_1 \to I_2 \to I_3 \to I_4 \to I_5 \to I_6$ with 3 or more copies of I_1.

$I_1 \to I_1$ is elementary because it has length 1, and the remaining loops are elementary by Lemma 2.6 because the interior of I_6 is disjoint from the interiors of the other intervals and the loops cannot be followed by a point of $\mathcal{O}$ for reasons familiar from the previous example. The lengths of these loops are 1, 2, 4, 6 and anything larger than 7, which proves that this cycle forces every period $l \lhd 7$.

The standard proof uses a different choice of $\mathcal{O}$-intervals to study this example: the interval I_i for each i with $2 \le i \le 5$ is replaced by the interval between x_i and x_{i-2}. With this alternate choice one still obtains the covering relations (1)–(3). Our choice of $\mathcal{O}$-intervals adapts better to the more general situation studied in Subsection 4.1.

3.3. **A 6-cycle.** Consider the 6-cycle



The crucial feature here is that the 3 points in the left half are mapped to the 3 points in the right half and vice versa.

Therefore, the 3 points in the right half form a cycle for the second iterate f^2, and specifically, it is a cycle of the form $\bullet \to \bullet \to \bullet$, just like the first example of this section. As before, we then get the covering relations $I_1 \to I_1$, $I_1 \to I_2$ and $I_2 \to I_1$, and so we can conclude as before that f^2 has elementary loops of all lengths.

Here is a way of producing elementary loops for f itself from these: For every elementary k-loop for f^2 made using the covering relations $I_1 \to I_1$, $I_1 \to I_2$ and $I_2 \to I_1$, replace each occurrence of "$I_1 \to$" by "$I_1 \to I_1' \to$" and each occurrence of "$I_2 \to$" by "$I_2 \to I_2' \to$". This produces a legitimate $2k$-loop for f, which is elementary for the following reason. If a point p follows the $2k$-loop under f, then p follows the original elementary k-loop under f^2 and hence has period k for f^2. On the other hand, the iterates of p under f alternate sides since the $2k$-loop for f alternates between primed and unprimed intervals. Therefore the period of p for f is $2k$. Since k was arbitrary, this shows that this 6-cycle forces all even periods (as well as period 1, which we obtain from the unnamed interval in the center, which covers itself under f).

4. Proof the Sharkovsky Forcing Theorem

Let $\mathcal{O}$ be a cycle of f with length m. We show that there are orbits of period l for every $l \lhd m$. This is accomplished by finding elementary l-loops of $\mathcal{O}$-intervals for all $l \lhd m$ and then applying Proposition 2.5 to deduce the existence of cycles of those lengths. In doing so we work directly with the cycle $\mathcal{O}$ itself; we do *not* need to assume that m is the period of f that comes earliest in the Sharkovsky order.

If $\mathcal{O}$ is a nontrivial cycle, *i.e.*, if $m \geq 2$, let p be the rightmost of those points in $\mathcal{O}$ for which $f(p) > p$, and let q be the point of $\mathcal{O}$ to the immediate right of p. Then $f(p) \geq q$ and $f(q) \leq p$. We set $I = [p, q]$. These choices have as an immediate consequence that $I \to I$.

We fix a point $c \in \operatorname{Int}(I)$. The midpoint of I is a natural choice, but any point of $\operatorname{Int}(I)$ will do.

For $x \in \mathcal{O}$ we denote by $\mathcal{O}_x$ the set of points of $\mathcal{O}$ in the closed interval bounded by x and c.

Definition 4.1. We say that $x \in \mathcal{O}$ *switches sides* if x and $f(x)$ are on opposite sides of c.

The endpoints of the interval I, namely p and q, switch sides.

Remark 4.2. If all points of $\mathcal{O}$ switch sides, then f is a bijection between $\mathcal{O}_L$ and $\mathcal{O}_R$, where $L := \min \mathcal{O}$ and $R := \max \mathcal{O}$. In particular m is even.

The proof begins with the case in which not every point switches sides. This is the content of the next proposition, which then provides the base for an inductive argument.

4.1. The main case: not all points of $\mathcal{O}$ switch sides.

Proposition 4.3. *If an m-cycle $\mathcal{O}$ with $m \geq 2$ contains a point that does not switch sides, then there is an elementary l-loop of $\mathcal{O}$-intervals for each $l \lhd m$.*

Proof. We begin by constructing a sequence $x_0, x_1, \ldots, x_k$ of points in $\mathcal{O}$ that "spirals out as fast as possible".

Choose x_0 and x_1 to be the endpoints of I, labeled in such a way that $f(x_1) \neq x_0$. Such a labeling is possible, since otherwise $\mathcal{O} = \{x_0, x_1\}$ and all points of $\mathcal{O}$ would switch sides, contrary to the hypothesis of the proposition:

$\bullet\ \ p\ \ q$	$p\ \ q\ \ \bullet$	$\bullet\ \ p\ \ q\ \ \bullet$
$x_1 = q$	$x_1 = p$	$x_1 = p$ or q

We extend the sequence inductively as follows. If $i \geq 1$ and all points of $\mathcal{O}_{x_i}$ switch sides, then x_{i+1} is the point of $f(\mathcal{O}_{x_i})$ that is furthest from c; otherwise x_{i+1} is not defined.

Consecutive terms of this sequence are on opposite sides of c. Consequently, $x_0, x_2, \dots, x_{2i}, \dots$ are all on one side of c and $x_1, x_3, \dots, x_{2i+1}, \dots$ are all on the other side of c.

Lemma 4.4. *x_{i+2} is further from c than x_i if both points are defined.*

Proof. For $i = 0$ this follows from the fact that $x_2 = f(x_1) \neq x_0$.

If $i \geq 1$ and x_{i+1} and x_{i+2} are both defined, all points of $\mathcal{O}_{x_i}$ and $\mathcal{O}_{x_{i+1}}$ switch sides and we have the inclusions $f(\mathcal{O}_{x_i}) \subset \mathcal{O}_{x_{i+1}}$ and $f(\mathcal{O}_{x_{i+1}}) \subset \mathcal{O}_{x_{i+2}}$, whence $f^2(\mathcal{O}_{x_i}) \subset \mathcal{O}_{x_{i+2}}$. Since f is one-to-one on $\mathcal{O}$ this shows that $\mathcal{O}_{x_{i+2}}$ has at least as many points as $\mathcal{O}_{x_i}$. Consequently, x_{i+2} is at least as far from c as x_i.

On the other hand, we cannot have $x_{i+2} = x_i$ for then we would have $f(\mathcal{O}_{x_i} \cup \mathcal{O}_{x_{i+1}}) \subset \mathcal{O}_{x_i} \cup \mathcal{O}_{x_{i+1}}$ but $\mathcal{O}_{x_i} \cup \mathcal{O}_{x_{i+1}} \neq \mathcal{O}$ because all points of $\mathcal{O}_{x_i} \cup \mathcal{O}_{x_{i+1}}$ switch sides (since x_{i+1} and x_{i+2} are both defined). This is impossible since f is a *cyclic* permutation of $\mathcal{O}$. □

Corollary 4.5. *The points $x_0, x_1, \dots$ are all distinct, and this sequence terminates with x_k for some $k < m$.*

We now construct a sequence of $\mathcal{O}$-intervals for which we have covering relations that will produce loops of length 1, length l for every even $l \leq k$ and length l for every $l \geq k+2$. We will then use Lemma 2.6 to verify that these loops are elementary. Since $k < m$, this set of lengths includes 1, every even $l < m$ and every $l > m$, that is, every $l \triangleleft m$.

A simpleminded choice of intervals that produces the desired covering relations (but does not allow application of Lemma 2.6) is as follows. For $1 \leq i < k$ let J_i be the shortest $\mathcal{O}$-interval that contains $\mathcal{O}_{x_i}$ and both endpoints of the interval I. It follows from Lemma 4.4 that $J_{k-1} \supset J_{k-3} \supset J_{k-5} \supset \cdots$ and $J_{k-2} \supset J_{k-4} \supset \cdots$. Let J_k be the shortest $\mathcal{O}$-interval that contains $\mathcal{O}_{x_k}$. With these choices we have:

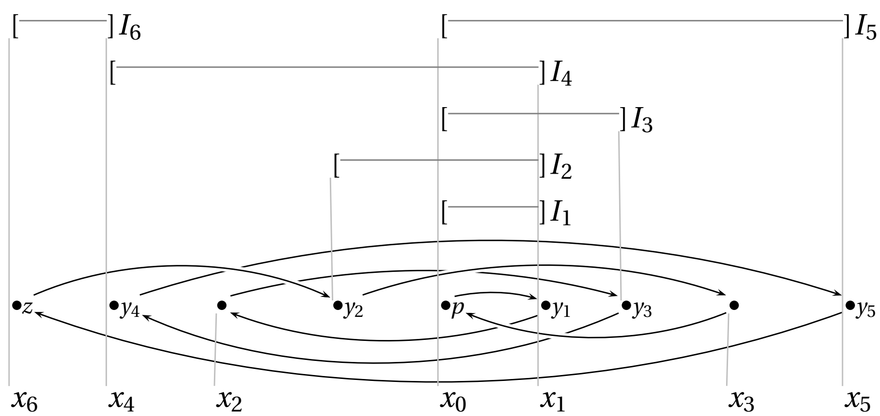
(1) $J_1 \to J_1$;
(2) $J_i \to J_{i+1}$ for $1 \leq i < k$;
(3) $J_k \to J_{k-1}$.

We obtain (1) because $J_1 = I \to I$ as noted above. We obtain (2) because J_i contains the endpoints of I and the point $y_i \in \mathcal{O}_{x_i}$ for which $f(y_{x_i}) = x_{i+1}$; this ensures that $f(J_i)$ contains both the interval I and the point x_{i+1}. Since $\mathcal{O}_{x_k} \supset \mathcal{O}_{x_{k-2}}$ by Lemma 4.4, J_k contains y_{k-2}, which switches sides and maps to x_{k-1}. But J_k also contains a point $z \in \mathcal{O}$ that does not switch sides. It follows that $f(J_k)$ contains both x_{k-1} and $f(z)$, which is on the other side of c from x_{k-1}. This gives us (3).

We now shrink the intervals J_i to intervals I_i that have all of the desired properties. For $1 \le i < k$ let I_i be the shortest $\mathcal{O}$-interval that contains y_i and both endpoints of I. Let I_k be the $\mathcal{O}$-interval bounded by y_{k-2} and the point z chosen in the previous paragraph. It is obvious from the definitions that $I_i \subset J_i$ for $1 \le i \le k$ and $I_1 = J_1 = I$. The arguments in the previous paragraph apply word-for-word to show that $I_1 \to J_1$, $I_i \to J_{i+1}$ for $1 \le i < k$, and $I_k \to J_{k-1}$. Since $J_{k-1} \supset I$ and $J_{k-1} \supset J_{k-3} \supset \cdots$ we obtain the covering relations:

(4) $I_1 \to I_1$,
(5) $I_1 \to I_2 \to \cdots \to I_k \to I_1$, and
(6) $I_k \to I_{k-1}, I_{k-3}, \ldots$.

Example 4.6. Here are the intervals I_1, I_2, ... for a typical orbit.

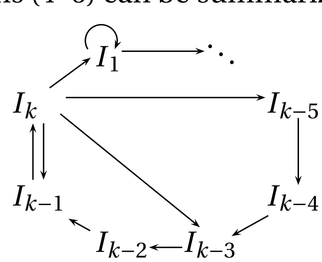


The advantage of the I-intervals over the J-intervals is that we can apply Lemma 2.6 to loops involving the I-intervals:

Lemma 4.7. $I_i \cap \operatorname{Int}(I_k) = \varnothing$ *for* $1 \le i < k$.

Proof. The point z is further from c than x_{k-2} because z does not switch sides and all points of $\mathcal{O}_{x_{k-2}}$ do switch sides. Consequently $\operatorname{Int}(I_k)$ lies on the opposite side of x_{k-2} from c. On the other hand, $J_{k-2} \cup J_{k-1}$ lies on the same side of x_{k-2} as c and $I_i \subset J_{k-2} \cup J_{k-1}$ for $1 \le i < k$. □

The covering relations (4–6) can be summarized in a graph as follows:



From this graph we read off the following loops:

(7) $I_1 \to I_1$;
(8) $I_k \to I_{k-(l-1)} \to I_{k-(l-2)} \to \cdots \to I_{k-2} \to I_{k-1} \to I_k$ for even $l \le k$;
(9) $I_k \to I_1 \to I_1 \to \cdots \to I_1 \to I_2 \to \cdots \to I_{k-1} \to I_k$ with j occurrences of I_1.

The loop in (7) is elementary because it has length 1. Lemma 4.7 and Lemma 2.6 will tell us that the loops in (8) and (9) are elementary once we show that they cannot be followed by a point of $\mathcal{O}$. This is the case for the loops in (8) because they have length $l \le k < m$. The loops in (9) are not followed by a point of $\mathcal{O}$ in the following cases.

- $j = 1$: since this loop has length $k < m$.
- $j = 2$ and $k < m - 1$: since this loop has length $k + 1 < m$.
- $j > 2$: since these loops have at least 3 repetitions of I_1.

The one exceptional case is that of $j = 2$ and $k = m - 1$ when the loop in (9) has length m and is not needed to produce a periodic point.

The loop in (7) has length 1. The loops in (8) have all even lengths up to k. The elementary loops in (9) have all lengths $l \ge k$, except possibly m itself.

Note that if $l \lhd m$ then either $l = 1$, $l > m$ or $l < m$ and l is even. Thus there are elementary loops of length l for every $l \lhd m$. $\square$

4.2. **Digression.** As we just noted in the proof, if $k < m - 1$ in Corollary 4.5, the orbit $\mathcal{O}$ will force periods $l = 1$, even $l \le k$ and every $l \ge k$, and this includes some periods that precede m in the Sharkovsky order.

An extreme situation of this type is given by any cycle in which the point q chosen at the beginning of Section 4 is $R = \max \mathcal{O}$ and $f(q) = L = \min \mathcal{O}$, *i.e.*, a cycle of the form $\bullet \cdots \bullet \to \bullet$. It has $k = 2$ and thus forces period 3 and hence all periods.

On the other hand, if $k = m-1$, there will be only one point of $\mathcal{O}$, namely x_{m-1}, that does not switch sides. The point x_{m-1} must be either the leftmost or rightmost point of $\mathcal{O}$ and the sequence $x_0, x_1, \ldots$ must spiral outwards clockwise or counterclockwise as shown:

$$x_{m-1} \ \cdots \ x_4 \ x_2 \ x_0 \ x_1 \ x_3 \ \cdots \ x_{m-2} \quad \text{or} \quad x_{m-2} \ \cdots \ x_3 \ x_1 \ x_0 \ x_2 \ x_4 \ \cdots \ x_{m-1}\,.$$

Furthermore we must have $f(x_i) = x_{i+1}$ for $0 \le i < m - 1$. These orbits are called *Štefan cycles*. They are central to the standard proof of the Sharkovsky Theorem. Our proof is more direct because we do not need these cycles, but they inspired our construction of the sequence x_i in the proof of Proposition 4.3.

4.3. **The general case: inductive argument.** The Sharkovsky Forcing Theorem 1.1 follows immediately from Proposition 2.5 once we establish that an m-cycle $\mathcal{O}$ has an elementary l-loop of $\mathcal{O}$-intervals for each $l \lhd m$.

This fact will be proved by induction on m. In order to carry out this induction we first need to remark on a feature of the loops obtained in Proposition 4.3 that we did not previously comment on. This is that every covering relation in those loops was obtained from information on how points of the given cycle move around. This means that all covering relations produced by these arguments have the following property.

Definition 4.8. A covering relation $I \to J$ of $\mathcal{O}$-intervals is said to be $\mathcal{O}$-*forced* if J lies in the closed interval whose endpoints are the leftmost and rightmost points of $f(I \cap \mathcal{O})$. A loop of $\mathcal{O}$-intervals is said to be $\mathcal{O}$-*forced* if every arrow in it arises from an $\mathcal{O}$-forced covering relation.

It is important for our inductive argument to note that all loops obtained in the proof of Proposition 4.3 are $\mathcal{O}$-forced loops. (Indeed, any covering relation derived only from information about the dynamics on $\mathcal{O}$ will be $\mathcal{O}$-forced.)

We now state and prove the needed fact in the form that allows us to prove it by induction.

Proposition 4.9. *An m-cycle $\mathcal{O}$ has an $\mathcal{O}$-forced elementary l-loop of $\mathcal{O}$-intervals for each $l \lhd m$.*

Proof. The proof proceeds by induction on m.

Proposition 4.9 is vacuously true for $m = 1$ since there is no $l \lhd 1$.

Suppose now that Proposition 4.9 is known for all cycles of length less than m. Let $\mathcal{O}$ be an m-cycle. If there is a point that switches sides, then the conclusion of Proposition 4.9 follows by Proposition 4.3 and our observation that Proposition 4.3 produces $\mathcal{O}$-forced loops.

Otherwise, all points switch sides, and Remark 4.2 tells us that m is even and f is a bijection between $\mathcal{O}_L$ and $\mathcal{O}_R$, where $L := \min \mathcal{O}$ and $R := \max \mathcal{O}$.

For the second iterate, f^2, both $\mathcal{O}_L$ and $\mathcal{O}_R$ are cycles of length $m/2$, and by the inductive assumption we can apply Proposition 4.9 to either of these, in particular to $\mathcal{O}_R$. Hence f^2 has an elementary $\mathcal{O}_R$-forced k-loop of $\mathcal{O}_R$-intervals for each $k \lhd m/2$. It remains to deduce from this that f has an elementary $\mathcal{O}$-forced $2k$-loop of $\mathcal{O}$-intervals for each $k \lhd m/2$ as well as an elementary 1-loop. Accordingly, the next proposition concludes the induction. $\square$

Proposition 4.10. *Let $\mathcal{O}$ be a cycle of f all of whose points switch sides, and suppose the cycle $\mathcal{O}_R$ of f^2 gives rise to an elementary $\mathcal{O}_R$-forced k-loop of $\mathcal{O}_R$-intervals for f^2. Then there is an elementary $\mathcal{O}$-forced $2k$-loop of $\mathcal{O}$-intervals for f. In addition, there is an elementary $\mathcal{O}$-forced 1-loop for f.*

Proof. The 1-loop is the middle $\mathcal{O}$-interval, which lies between the rightmost point of $\mathcal{O}_L$ and the leftmost point of $\mathcal{O}_R$. For an elementary k-loop

$$J_0 \to J_1 \to J_2 \to \cdots \to J_{k-1} \to J_0 \tag{1}$$

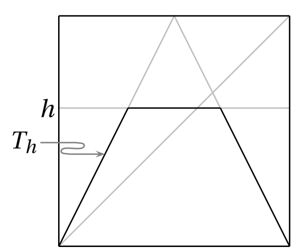
of $\mathcal{O}_R$-intervals for f^2, let J_i' be the shortest closed interval that contains $f(J_i \cap \mathcal{O}) \subset \mathcal{O}_L$. The intervals $J_1', \dots, J_{k-1}'$ lie to the left of c. Because the covering $J_i \to J_{i+1}$ for f^2 is $\mathcal{O}_R$-forced, we also have $J_i' \to J_{i+1}$ for f, and this covering is $\mathcal{O}$-forced. Therefore, we get an $\mathcal{O}$-forced $2k$-loop for f:

$$J_0 \to J_0' \to J_1 \to J_1' \to \cdots \to J_{k-1} \to J_{k-1}' \to J_0\,. \tag{2}$$

A periodic point p for f that follows the loop (2) is a periodic point for f^2 that follows the elementary loop (1) and hence has period k with respect to f^2. Since the intervals in the loop (2) are alternately to the right and the left of the center, so are the iterates of p under f. Hence p has period $2k$ with respect to f, and (2) is elementary. □

5. The Sharkovsky Realization Theorem

An elegant proof of the Sharkovsky Realization Theorem 1.2 is given in [ALM]. They consider the family of truncated tent maps $T_h\colon [0,1] \to [0,1]$, $x \mapsto \min(h, 1-2|x-1/2|)$ for $0 \le h \le 1$.



The truncated tent maps have several key properties.

(a) T_0 has only one periodic point (the fixed point 0) while the tent map T_1 has a 3-cycle $\{2/7, 4/7, 6/7\}$ and hence has all natural numbers as periods by the Sharkovsky Forcing Theorem 1.1.
(b) T_1 has a finite number of periodic points for each period[5].
(c) If $h \le k$, any cycle $\mathcal{O} \subset [0,h)$ of T_h is a cycle for T_k, and any cycle $\mathcal{O} \subset [0,h]$ of T_k is cycle for T_h.

What makes the proof so elegant is that h plays three roles: as a parameter, as the maximum value of T_h, and as a point of an orbit.

For $m \in \mathbb{N}$, let $h(m) := \min\{\max \mathcal{O} \mid \mathcal{O} \text{ is an } m\text{-cycle of } T_1\}$. From this definition and (c) with $k=1$ we obtain:

(d) T_h has an l-cycle $\mathcal{O} \subset [0,h)$ if and only if $h(l) < h$.

[5] Inspection of the graph of T_1^n shows that it has exactly 2^n fixed points.

(e) The orbit of $h(m)$ is an m-cycle for $T_{h(m)}$, and all other cycles for $T_{h(m)}$ lie in $[0, h(m))$.

From (e) and the Sharkovsky Forcing Theorem 1.1 we see that $T_{h(m)}$ has an l-cycle that lies in $[0, h(m))$ for every $l \lhd m$; it follows from (d) that $h(l) < h(m)$. Since this holds for all m, we obtain:

(f) $h(l) < h(m)$ if and only if $l \lhd m$.

We see from (d), (e) and (f) that for any $m \in \mathbb{N}$ the set of periods of $T_{h(m)}$ is the tail of the Sharkovsky order consisting of m and all $l \lhd m$.

The set of all powers of 2 is the only other tail of the Sharkovsky order (besides $\varnothing$, which is the set of periods of the translation $x \mapsto x+1$ on $\mathbb{R}$). $h(2^\infty) := \sup_k h(2^k) > h(2^k)$ by (f) for all $k \in \mathbb{N}$, so $T_{h(2^\infty)}$ has 2^k-cycles for all k by (d). Suppose $T_{h(2^\infty)}$ has an m-cycle with m not a power of 2. By Theorem 1.1 $T_{h(2^\infty)}$ also has a $2m$-cycle. Since the m-cycle and the $2m$-cycle are disjoint, at least one of them is contained in $[0, h(2^\infty))$, contrary to (d) and (f).

Acknowledgments. We thank the Instituto Superior Técnico for its hospitality and Wah Kwan Ku for pointing out an error in a draft of this paper. We are also grateful for extensive suggestions by the referee and the editor.

References

[ALM] Lluís Alsedà, Jaume Llibre, Michał Misiurewicz: Combinatorial dynamics and entropy in dimension one. Second edition. Advanced Series in Nonlinear Dynamics, 5. World Scientific Publishing Co., Inc., River Edge, NJ, 2000.

[BB] Reid Barton, Keith Burns: *A simple special case of Sharkovskii's theorem.* Amer. Math. Monthly **107** (2000), no. 10, 932–933.

[BGMY] L. Block, J. Guckenheimer, M. Misiurewicz and L. S. Young: *Periodic points and topological entropy of one-dimensional maps,* Global Theory of Dynamical Systems, Z. Nitecki and C. Robinson eds, Lecture Notes in Mathematics vol. 819, Springer Verlag, 1980, 18–34.

[B] U. Burkart: *Interval mapping graphs and periodic points of continuous functions*, J. Combin. Theory Ser. B, **32** (1982), 57–68.

[C] W. A. Coppel: *The solution of equations by iteration.* Proc. Cambridge Philos. Soc. **51** (1955). 41–43.

[D] Bau-Sen Du: *A simple proof of Sharkovsky's theorem*, American Mathematical Monthly **111**, no. 11, (2004), 595–599. *More simple proofs of Sharkovsky's theorem*, preprint. *A simple proof of Sharkovsky's theorem revisited*, American Mathematical Monthly **114**, no. 2, (2007), 152–155.

[HM] Chung Wu Ho, Charles Morris: *A graph-theoretic proof of Sharkovsky's theorem on the periodic points of continuous functions.* Pacific J. Math. 96 (1981), no. 2, 361–370.

[LY] Tien-Yien Li, James A. Yorke: *Period three implies Chaos,* American Mathematical Monthly, **82** no. 10 (1975), 985–992.

[M] Michał Misiurewicz: *Remarks on Sharkovsky's Theorem*, American Mathematical Monthly, **104** no. 10 (1997), 846–847

[N] Zbigniew Nitecki: *Topological dynamics on the interval*, in: Ergodic Theory and Dynamical Systems II, Proceedings Special Year, Maryland 1979–1980, Birkäuser Progress in Mathematics **21** (1982), 1–73.

[S1] Alexander N. Sharkovskiĭ: *Co-existence of cycles of a continuous map of the line into itself*, Ukrainskiĭ Matematicheskiĭ Zhurnal **16** (1964), no. 1, 61–71; English translation: International Journal of Bifurcation and Chaos in Applied Sciences and Engineering **5** (1995), no. 5, 1263–1273; *On cycles and structure of a continuous mapping*, Ukrainskiĭ Matematicheskiĭ Zhurnal **17** (1965), no. 3, 104–111.

[S2] Alexander N. Sharkovskiĭ: *The reducibility of a continuous function of a real variable and the structure of the stationary points of the corresponding iteration process.* Dokl. Akad. Nauk SSSR **139** (1961) 1067–1070; *Fixed points and the center of a continuous mapping of the line into itself.* Dopovidi Akad. Nauk Ukr. RSR **1964** (1964), 865–868; *The set of convergence of one-dimensional iterations.* (Ukrainian) Dopovidi Akad. Nauk Ukr. RSR **1966** 1966 866–870.

[Š] Peter Štefan: *A theorem of Šarkovskii on the existence of periodic orbits of continuous endomorphisms of the real line*, Comm. Math. Phys. **54** (1977), no. 3, 237–248.

[St] P. D. Straffin Jr.: *Periodic points of continuous functions*, Math. Mag. **51** no. 2 (1978), 99–105.

DEPARTMENT OF MATHEMATICS, NORTHWESTERN UNIVERSITY, EVANSTON, IL 60208
E-mail address: `burns@math.northwestern.edu`

DEPARTMENT OF MATHEMATICS, TUFTS UNIVERSITY, MEDFORD, MA 02155
E-mail address: `bhasselb@tufts.edu`