ARKIV FÖR MATEMATIK Band 5 nr 26

1.64032 Communicated 12 February 1964 by O. FROSTMAN and L. CARLESON

The functional equation $f^n(x)=g(x)$

By JAMES C. LILLO

1. Introduction and notation

We are interested in studying the real functional equation $f^n(x)=g(x)$ on an interval $[a,b]$ of the real line. In particular we wish to obtain conditions on g which will assure one that solutions f of the given equation possess certain properties. If one insists only that f be a pointwise solution, then the problem for $n=2$ has been solved [3]. If one insists that f be continuous, only very limited results are known [1], [2], [5]. In Theorem 2.1 we obtain results which suggest studying the problem in a certain subclass $M[a,b]$ of the class of continuous functions. In example 1 we show that there exists a continuous function g defined on a closed interval $[a,b]$ for which the equation $f^2(x)=g(x)$ does not possess any continuous solutions f but does have a solution f which possesses the Darboux property. Theorem 2.4 gives sufficient conditions to insure that if g is continuous then any solution f of the equation $f^n(x)=g(x)$, which possesses the Darboux property, will also be continuous. In Theorem 2.5 we consider the special equation $f^n(x)=f^{n+p}(x)$.

To facilitate matters we introduce the following notation. Let $[a,b]$ denote any closed interval of the real line where the endpoints $+\infty$ and $-\infty$ are allowed. The set of all functions defined on $[a,b]$ with values in $[a,b]$ will be denoted by $R[a,b]$. A function is said to possess the Darboux property if it takes connected sets into connected sets. $D[a,b]$ will denote those functions of $R[a,b]$ which possess the Darboux property. $C[a,b]$ will denote those functions of $R[a,b]$ which are continuous on $[a,b]$. We denote by $M[a,b]$ those functions of $C[a,b]$ which are piecewise monotone (written p.m.) on $[a,b]$. Here, f is said to be piecewise monotone on $[a,b]$ if there exists a finite partition $P=[p_0,\dots p_n]$ of $[a,b]$ such that on each subinterval $[p_i,p_{i+1}]$ the function f is strictly monotone (written s.m.). If every partition P^* which possesses this property with respect to f is a refinement of P, then P is said to be the partition asssociated with f and will be denoted by $P(f)$. We define $f^0(x)=x$ and $f^{n+1}(x)=f(f^n(x))$ for $n\geq 0$. Finally, we define the set $S(n,g)=\{f\in R[a,b]\,|\,f^n(x)=g(x)$ for all $x\in[a,b]\}$.

2. The general equation $f^n(x)=g(x)$

It is clear that if $f\in M[a,b]$ then $f^i\in M[a,b]$ for any i. We now establish the converse. If $f\in D[a,b]$ and $f^i\in M[a,b]$ then $f\in M[a,b]$.

Theorem 2.1. *If $g\in M[a,b]$ then $S(n,g)\cap D[a,b]\subset M[a,b]$ and $P(g)$ is a refinement of $P(f)$ for every $f\in S(n,g)\cap D[a,b]$.*

Proof. We first note that if $f \in D[a,b]$ is s.m. on each subinterval $[p_i, p_{i+1}]$ of $P(g)$ then $f \in C[a,b]$ and so $f \in M[a,b]$. Thus, it suffices to show that any $f \in S(n,g) \cap D[a,b]$ is s.m. on every subinterval $[p_i, p_{i+1}]$ of $P(g)$. Assume f is not s.m. on $[p_i, p_{i+1}]$, then since $f \in D[a,b]$ it follows easily that there are at least two points $x, y \in [p_i, p_{i+1}]$ for which $f(x) = f(y)$. But then $g(x) = f^n(x) = f^n(y) = g(y)$ which contradicts the fact that g is s.m. on $[p_i, p_{i+1}]$. This completes the proof of Theorem 2.1.

We shall see later that there are $g \in M[a,b]$ such that $D[a,b] \cap S(2,g)$ is empty while $R[a,b] \cap S(2,g)$ is not empty. We shall also see, by means of an example, that there are $g \in C[a,b]$ for which $S(2,g) \cap C[a,b]$ is empty but $S(2,g) \cap D[a,b]$ is not empty. To facilitate the construction of this example we now obtain several results which are needed here and later in the development. The first result is closely related [2] to the case $g(x) = f(x) = f^n(x)$.

Theorem 2.2. *If $f \in D[a,b]$, $f(p) = p$ and S is a nondegenerate maximal connected set containing p, such that $f^n(x) = x$ for $x \in S$, then $S = [c,d]$ and (a) $f|S$ is a homeomorphism of S onto S, (b) $f(x) = x$ on S or $f(c) = d$, $f(d) = c$ and $f|[c,d]$ is a reflection of $[c,d]$ about $p \in (c,d)$.*

Proof. Since $f^n = g$ is s.m. on S and $f \in D[a,b]$, it follows, as in Theorem 2.1, that f^i is s.m. on S, $i = 1, 2, \ldots, n-1$. It then follows that f^i is continuous and s.m. on the closure $\bar{S} = [c,d]$ of S, $i = 1, \ldots, n-1$. Consider first the case where f is increasing on $[c,d]$. If $p \neq d$ then there exists $q \in (p,d)$ such that $f^i(q) \in (p,d)$ for $i = 0, \ldots, n$. Either $f^{i+1}(q) > f^i(q)$, $f(q) = q$ or $f^{i+1}(q) < f^i(q)$ for $i = 0, \ldots, n$ since f is s.m. in $[p,d]$. But $f^n(q) = q$ and so $f(q) = q$. It now follows that $f(x) \equiv x$ on $[p,d]$. If $p \neq c$ a similar treatment shows that $f(x) \equiv x$ on $[c,p]$. Thus, if f is increasing on $[c,d]$ then $f(x) \equiv x$ on $[c,d]$. Let f be decreasing on $[c,d]$ and assume that $p \neq c$. Then f is s.m. on $[p, f(c)]$. If f is increasing on $[p, f(c)]$ there is a point $w \in (p, f(c))$ such that $f^i(w) \in (p, f(c))$ for $i = 1, 2, \ldots, n$. Let $q \in (c,p)$ be such that $f(q) = w$, then $f^n(q) \neq q$. Thus, f is decreasing in $[p, f(c)] \cup [c,p]$. If $f^2(c) \neq c$ then either $f^2(c) \in (c, f(c))$ or there is a $\mu \in (c,p)$ such that $f^{n-j}(\mu) \in [p, f(c)] \cup [p,c) j = 1, \ldots, n-1$ and $f^n(\mu) = c$. In the first case, $f^i(c) \in (c, f(c))$ for all $i \geqslant 2$ and so $f^n(c) \neq c$. In the second case, we have $\mu \in S$ for which $f^n(\mu) \neq \mu$ which is impossible. Thus, $f^2(c) = c$, n is even, and f^2 is an increasing function on $[c,p]$. Thus, f is a reflection of $[c, f(c)]$ about p. In the same way one may show that f is a reflection of $[f(d), d]$ about p. Since S is maximal $f(d) = c$ and $f(c) = d$. This completes the proof of Theorem 2.2.

Corollary 2.1. *If f satisfies the hypothesis of Theorem 2.2, $f \in C[a,b]$ and $S = [c,d] \neq [a,b]$ is a ray, then $f \equiv x$ on S.*

Proof. Either $c \in (a,b)$ and $d = b = +\infty$ or $d \in (a,b)$ and $c = a = -\infty$. Since $f \in C[a,b]$ it is clear that in both cases we may not have $f(c) = d$ and $f(d) = c$ and the result follows.

If $g \in R[a,b]$ we define $\gamma(g) = \{x \mid x \in [a,b] \text{ and } g(x) = x\}$. $\gamma(g)$ is called the set of fixed points of g. If $f \in S(n,g)$ then one may say a great deal about $f|\gamma(g)$. One of these results is contained in the following theorem.

Theorem 2.3. *If $g \in R[a,b]$ and $f \in S(n,g)$ then $f|\gamma(g)$ defines a one to one map of $\gamma(g)$ onto $\gamma(g)$.*

Proof. Assume $x \in \gamma(g)$, but that $y = f(x) \notin \gamma(g)$. Then $g(y) \neq y$ and $f^{n+1}(x) = f(f^n(x)) = f(x) = y = f(f^n(x)) = g(y) \neq y$. Thus, $f^i(\gamma(g)) \subset \gamma(g)$ for any i. Let $x \in \gamma(g)$, then $x = g(x) =$

$f(f^{n-1}(x)) \subset f(\gamma(g))$ and so $\gamma(g) \subset f(\gamma(g))$. Thus, f defines a map of $\gamma(g)$ onto itself. Since for any $x, y \in \gamma(g) x \neq y$ implies $f^n(x) \neq f^n(y)$, it follows that the map is one to one.

Corollary 2.2. *If $g \in R[-\infty, \infty]$, $\gamma(g)$ is a ray, and $f \in S(n,g) \cap C[-\infty, \infty]$, then $f(x) \equiv x$ for $x \in \gamma(g)$.*

Proof. Since $g \in C[-\infty, \infty]$, $\gamma(g)$ is a closed interval and f defines a homeomorphism of $\gamma(g)$ onto itself. Because $f \in C[-\infty, \infty]$ the finite endpoint of $\gamma(g)$ must be mapped onto a finite point so it must be mapped onto itself since $f|\gamma(g)$ is a homeomorphism. Our result now follows from Corollary 2.1.

It is possible to obtain information concerning the existence of solutions $f \in R[a,b]$ for $f^n(x) = g(x)$ by studying the sets $\gamma(g^i)$. Thus, for example, the fact that the function $g(x) = -x$, $x \in [0, -1]$, and $g(x) = -x^2$, $x \in [0,1]$, possesses only one cycle of order 2, namely $[1, -1]$, implies that $S(2,g)$ is empty. In fact, Isaacs [5] has stated necessary and sufficient conditions for $S(2,g)$ to be non empty in terms of the cycles of g. Unfortunately, these results give no information about $S(2,g) \cap D[a,b]$ except, of course, in the case where $S(2,g)$ is empty.

We now display a function $g \in C[-\infty, \infty]$ for which $S(2,g) \cap C[-\infty, \infty]$ is empty but $S(2,g) \cap D[-\infty, \infty]$ is not empty.

Example 1. We first define the functions h, f, g.

We define h on $[0, 1]$: $h(1/n) = (-1)^n$ $n = 1, 2, \ldots$;

$$h'(x) = (-1)^n\, 2n(n+1) x \in (1/n+1,\ 1/n),\ n = 1, 2 \ldots.$$

We define f on $[-\infty, \infty]$: $f(x) = x$, $x \leqslant 0$; $f(x) = 0$, $x \geqslant 2$ and $0 \leqslant x \leqslant 1$;

$$f(x) = x - 1,\ 1 < x \leqslant \tfrac{5}{4};\ f(x) = (x - \tfrac{1}{4}) h(4x-5)/2 + \tfrac{1}{2},\ \tfrac{5}{4} < x \leqslant \tfrac{3}{2};$$

$$f(x) = -\tfrac{1}{8} + \tfrac{1}{4}(x - \tfrac{3}{2}),\ \tfrac{3}{2} \leqslant x \leqslant 2.$$

We define $g(x) = f^2(x)$ for $x \in [-\infty, \infty]$. Clearly $f \in D[-\infty, \infty]$, $g \in C[-\infty, \infty]$, and it remains only to prove that $S(2,g) \cap C[-\infty, \infty]$ is empty. Assume $f \in S(2,g) \cap C[-\infty, \infty]$. Then by Corollary 2.2 $f(x) \equiv x$ for $-\infty \leqslant x \leqslant 0$. Then for all x, such that $f(x) \leqslant 0$, we have $g(x) = f(f(x)) = f(x)$. Thus, $f(x) \geqslant 0$ in $[0,1]$. We assert that there exists $\delta > 0$ such that $f(x) \equiv 0$ for $x \in [0,\delta]$. Assume $f(x) \not\equiv 0$ on $[0,1]$ and define $\sigma = \max_{[0,1]} f(x)$. Since $g(x) = f^2(x) = 0$ for $x \in [0,1]$ it is clear that $f(x) = 0$ for $x \in [0,\sigma]$. Thus, if $g(x) = f(f(x)) > 0$ then $f(x) > \delta$. Since $f(\delta) = 0$ and $g(x) < x$ for all $x > 0$, it is clear that $f(x) < x$ for all $x > 0$.

We define $\sigma(n) = 1 + \frac{1}{4} + \frac{1}{4}(1/n)$. Then if n is odd we have $g(x) > g(\sigma(n))$ for all $0 \leqslant x < \sigma(n)$. Thus, for n odd $f(\sigma(n)) = g(\sigma(n))$, and it follows that $f(x) \equiv g(x)$ whenever $f(x)$ or $g(x)$ is negative. Thus, $f(\sigma(n) < 0$ for n odd. But for n even $f(f(\sigma(n))) = g'\sigma(n)) > 0$ and so $f(\sigma(n)) > \delta$. Since $\lim_{n\to\infty} \sigma(n) = 1 + \frac{1}{4}$ it follows that f is discontinuous at $x = 1 + \frac{1}{4}$. This completes Example 1.

Consideration of the above example suggests the restrictions on $g \in C[a,b]$ which will insure that the solutions of $f^n(x) = g(x)$ also belong to $C[a,b]$. This result is contained in the following theorem.

Theorem 2.4. *If $g \in C[a,b]$ and if either (a) or (b) below are satisfied then $S[n,g] \cap D[a,b] = S[n,g] \cap C[a,b]$.*

(a) *Range of $g = [a,b]$.*

(b) *g is not constant on any non degenerate interval.*

Proof. Assume (a) is satisfied and $f \in S(n,g) \cap D[a,b]$. Then the range of $f = [a,b]$. Let $h(x)$ denote $f^{n-1}(x)$. Then range $h = [a,b]$ and $h \in D[a,b]$. Let f be discontinuous at z. Thus, there exists a sequence $\{x_1\}$ tending to z such that no subsequence of $\{f(x_i)\}$ converges to $f(z)$. One may also assume that $|x_i - z| > |x_{i+1} - z|$ for all i and that the sign of $(x_i - z)$ is independent of i, say negative. We now define a sequence $\{y_j\}$ converging to a point y such that $h(y_j) = z$ for j odd, and for j even $\{h(y_j)\}$ is a subsequence of $\{x_i\}$. Since $[a,b]$ = range of h there exist y_1 and y_2 such that $h(y_1) = z$ and $h(y_2) = x_1$. If y_1 and y_2 are both finite define $\sigma = (y_1 + y_2)/2$. If either y_1 or y_2 is infinite, let σ be any point in (y_1, y_2) for which $|y_1 - \sigma| \geqslant 1$ and $|y_2 - \sigma| \geqslant 1$. If $h(\sigma) = z$ set $y_3 = \sigma$ and let y_4 be any point in $[y_3, y_2]$ for which $h(y_4) = x_2$. If $h(\sigma) > z$ let y_3 be any point in $[\sigma, y_2]$ for which $h(y_3) = z$, and y_4 be any point in $[y_3, y_2]$ for which $h(y_4) = x_2$. If $h(\sigma) < z$ let $y_3 = y_1$. Since $h(\sigma) < x_k$ for some $k \geqslant 2$ let $y_4 \in [y_3, \sigma]$ be any point for which $h(y_4) = x_k$. Using y_3, y_4 in place of y_1, y_2 and x_2 or x_k in place of x_1 we repeat the procedure. In this way we obtain a sequence $\{y_j\}$ with the stated properties. But then $f(z) = \lim_{k\to\infty} f(h(y_{2k+1})) = g(y) = \lim_{k\to\infty} g(y_{2k}) \neq f(z)$. Thus $f(x) \in C[a,b]$.

Assume now that (b) is satisfied. Since g is not constant on any interval and $f^n(x) = g(x)$ we have that $f^i(x)$, $i = 1, \ldots, n$ is not constant on any interval. Let f be discontinuous at z and set $r = f(z)$. Thus, there exists a $\sigma > 0$ such that either for any $w \in [r, r+\sigma]$ or for any $w \in [r, r-\sigma]$ there is a sequence $\{x_i\} \to z$ such that $f(x_i) = w$. Since h is not constant in any interval we may choose in $[r, r+\sigma]$ or in $[r, r-\sigma]$, whichever is necessary, a w such that $h(w) \neq h(r)$. But then $h(w) = \lim g(x_i) = g(z) = h(r)$ which is not possible. Thus, f is continuous.

Corollary 2.3. *Let $g(x)$ be a real analytic function in $(-\infty, \infty)$ which is not the constant function.* Let f be defined on $(-\infty, \infty)$ into $(-\infty, \infty)$, possess the Darboux property there, and satisfy $f^n(x) = g(x)$. Then f is continuous in $(-\infty, \infty)$.

Proof. Clearly g satisfies condition (b). However, g need not belong to $C[-\infty, \infty]$ nor f to $D[-\infty, \infty]$. It is, however, easily verified that Theorem 2.4 is valid for the open interval $(-\infty, \infty)$.

We now consider the equation $f^n(x) = f^m(x)$, $m = n + p$, for $f \in D[a,b]$. Let R^i denote the range of f^i and $R^0 = [a,b]$. R^i is connected and so is an interval. Also $R^{i+1} \subset R^i$ for all i, and since $f^n(x) = f^m(x)$, it is clear that $R^n = R^{n+1} = R^{n+i}$ for $i = 1, 2, \ldots$. If R^n is a point c then $f^n(x) = f^{n+1}(x) = C$ and we have one of the exceptional cases of Theorem 2.4. If $n = 1$, we have the case studied by Ewing and Utz [2]. The following theorem extends their results.

Theorem 2.5. *A necessary and sufficient condition that $f \in D[a,b]$ satisfy $f^n(x) = f^{n+p}(x)$ for all $x \in [a,b]$, where p and n are minimal, is that either (a) or (b) is satisfied:*

(a) $p = 1$, there exists a sequence of intervals $R^i, i = 1, \ldots, n$ such that $f | R^i = R^{i+1}$, $f | R^n = R^n$, $f(x) = x$ for $x \in R^n$, $R^i \neq R^{i+1}$ for $i = 0, \ldots, n-1$ and $R^0 = [a,b]$.

(b) $p = 2$, the sequence R^i is as above except that $f | R^n$ is a reflection.

Proof. Assume $f \in D[a,b]$ and satisfies the equation $f^n(x) = f^{n+p}(x)$ for all $x \in [a,b]$. Set $R^\circ = [a,b]$ and R^i = range of f^i. Then if in Theorem 2.2 we set $S = R^n$, and assume S non-degenerate our result follows. Conversely, if f satisfies either (a) or (b) it is clear that $f^{n+p}(x) = f^n(x)$ for all $x \in [a,b]$.

ACKNOWLEDGEMENT

This research was sponsored by National Science Foundation Grant G18915.

Mathematics Dept., Purdue University, Lafayette, Indiana, U.S.A.

REFERENCES

1. BODEWADT, U. T., Zur Iteration reeller Funktionen. *Math. Z. 49*, 497–516 (1944).
2. EWING, G. M., and UTZ, W. R., Continuous solutions of $f^n(x) = f(x)$. *Can. J. Math. 5*, 101–3 (1953).
3. ISACCS, R., Iterates of fractional order. *Can. J. Math. 2*, 409–16 (1950).
4. LOJASIEWICS, S., Solution générale de l'équation fonctionelle $f(f(\ldots f(x)\ldots)) = g(x)$. *Ann. Soc. Polon. Math. 24*, 88–91 (1952).
5. MASSERA, J. L., and PETRACCA, A., On the functional equation $f(f(x)) = 1/x$. *Revista Union Mat. Argentina 11*, 206–211 (1946).

Tryckt den 3 september 1964

Uppsala 1964. Almqvist & Wiksells Boktryckeri AB