THE SEMIGROUP OF CONTINUOUS SELFMAPS OF
I HAS INFINITELY MANY IDEALS

by J. W. BAKER and J. S. PYM

1. Introduction. Let I denote the interval $[0, 1]$ in its usual topology and let $S = S(I)$ be the semigroup of continuous mappings of I into itself with function composition as the semigroup operation. In a survey talk given at Oberwolfach in 1989 (see [5]), K. D. Magill Jr. pointed out that some elementary algebraic properties of S are still unknown. We shall answer one of the questions he raised (which appears as Problem 4.6 in [5] and which he had asked earlier, as long ago as 1975 [3]) by showing that S has infinitely many distinct two-sided ideals. In fact, we shall produce an infinite descending sequence of distinct ideals. As Magill points out, this also solves his Problem 4.5: S has infinitely many distinct congruences. We believe that S must have $\mathfrak{c}$ distinct ideals, but we have been unable to prove this.

Paradoxically, it is the topological simplicity of I which makes our problem difficult. For example, if I is replaced by the circle $\mathbb{T}$, then the map which sends $f \in S(\mathbb{T})$ to its degree is a homomorphism of the composition semigroup $S(\mathbb{T})$ onto the multiplicative semigroup of the ring $\mathbb{Z}$ of integers ([2, Corollary 3.4], for example). Each multiplicative ideal of $\mathbb{Z}$ therefore gives an ideal of $S(\mathbb{T})$. A much deeper study of the ideals of $S(\mathbb{T})$ is to be found in [4]. In addition, the fact that $S(I^n)$ has infinitely many ideals for any $n \geq 2$ is in [3], and is also a consequence of Theorem 11.2 of this paper.

Finally, the referee pointed out to us that the results on $S(I)$ can be used fairly easily to obtain similar results for a large class of semigroups. An account of this is given as the last section of the paper.

2. Ideals of S. We aim to show that S contains a minimal ideal K, a maximal ideal U, an infinite chain $\{J_n\}$ of ideals, and an ideal L satisfying

$$K \subset L \subset J_{n+1} \subset J_n \subset U \quad \text{(for } n = 2, 3, \ldots)$$

with all inclusions strict. These ideals are described below. In fact U is really just the ideal J_1.

First let us introduce some terminology and notation. We let ι denote the identity function in S, $\iota(t) = t$ $(t \in I)$, and $\hat{\iota}$ the function given by the formula

$$\hat{\iota}(t) = 1 - t \quad (t \in I).$$

More generally, for $f \in S$ we define $\hat{f}$ to be $f \circ \hat{\iota}$, so that $\hat{f}$ can be regarded as a reflection of f in the line $t = \frac{1}{2}$. By an *interval* in I is to be understood throughout an interval of the form (a, b) with $0 \leq a < b \leq 1$. For $f \in S$ a *local maximum* (resp. *minimum*) of f is a point $t \in (0, 1)$ such that there is $\epsilon > 0$ so that $(t - \epsilon, t + \epsilon) \subseteq (0, 1)$ and $f(t) \geq f(x)$ (resp. $f(t) \leq f(x)$) for all $x \in (t - \epsilon, t + \epsilon)$; the set of such points is denoted by $M(f)$ (resp. $m(f)$). More generally, an *n-fold local maximum* (resp. *minimum*) of f is an n-tuple $(t_1, t_2, \ldots, t_n)$ such that each t_i is in $M(f)$ (resp. $m(f)$) and $f(t_1) = f(t_i)(1 \leq i \leq n)$; the set of such n-tuples is denoted by $M_n(f)$ (resp. $m_n(f)$). More generally, if (a, b) is an interval in I then $M_n(f)(a, b)$ (resp. $m_n(f)(a, b)$) denotes those elements of $M_n(f)$ (resp. $m_n(f)$) with each coordinate in (a, b).

We now define our ideals.

$K = \{f \in S : f \text{ is constant}\}$.

$L = \{f \in S : \text{every interval in } I \text{ has a subinterval on which } f \text{ is constant}\}$.

$J_n = \{f \in S : \text{for every interval } (a, b) \text{ in } I, \text{ both } M_n(f)(a, b) \neq \varnothing \text{ and } m_n(f)(a, b) \neq \varnothing\}$, for $n = 1, 2, 3, \dots$.

$U = \{f \in S : \text{on no subinterval of } I \text{ is } f \text{ either strictly increasing or strictly decreasing}\}$.

The defining property of U is the same as requiring that f has local maxima and local minima in each subinterval of I, so that $U = J_1$. Also, as S consists of continuous functions, f is strictly monotone on an interval if and only if it is bijective there.

THEOREM 2.1. *The set K is the minimal ideal of S, U is the unique maximal ideal, each J_n is an ideal and $K \subset L \subset J_{n+1} \subset J_n \subset U$ ($n = 2, 3, 4, \dots$), with all the inclusions strict.*

The facts that K is the minimal ideal and U is the unique maximal ideal are in [3] (Section 3.5, page 241). It is very easy to see that K is the minimal ideal, but we include a brief proof that U is the unique maximal ideal for completeness. That $K \subseteq L \subseteq J_{n+1} \subseteq J_n \subseteq U = J_1$ ($n = 2, 3, \dots$) is obvious. We need to prove that L and the J_n are ideals, and the strictness of the inclusions.

3. The set J_n is an ideal.   Let $f \in J_n$, $g \in S$. Let (a, b) be an interval in I. Clearly if g is constant on (a, b) then so is $f \circ g$, so that $M_n(f \circ g)(a, b)$ and $m_n(f \circ g)(a, b)$ are nonempty. If g is not constant on (a, b) choose an interval $(c, d) \subseteq g((a, b))$. Let $(t_1, t_2, \dots, t_n) \in M_n(f)(c, d)$. Choose $x_i \in (a, b)$ with $g(x_i) = t_i$ ($1 \leq i \leq n$). Then $(x_1, x_2, \dots, x_n) \in M_n(f \circ g)(a, b)$. Similarly for m_n. Now consider $g \circ f$. One possibility is that f is constant on some subinterval (c, d) of (a, b). In this case it is clear that $M_n(g \circ f)(c, d) \neq \varnothing$, $m_n(g \circ f)(c, d) \neq \varnothing$, and so the same is true for (a, b). The only remaining possibility is that f is not constant on any subinterval of (a, b). We assume this and consider three further cases.

Case (i).   There is a subinterval (c, d) of (a, b) such that g is monotonic increasing on $f((c, d))$ (which is necessarily non-trivial). Then clearly if $(t_1, t_2, \dots, t_n) \in M_n(f)(c, d)$ we have also that

$$(t_1, t_2, \dots, t_n) \in M_n(g \circ f)(c, d) \subseteq M_n(g \circ f)(a, b).$$

Likewise for m_n.

Case (ii).   There is a subinterval (c, d) of (a, b) for which g is monotonic decreasing on $f((c, d))$. This is similar to case (i) except that an element of $M_n(f)(c, d)$ will be in $m_n(g \circ f)(a, b)$. Similarly with m_n and M_n interchanged.

Case (iii).   For any subinterval (c, d) of (a, b), g is not monotonic on $f((c, d))$. Let $(t_1, t_2, \dots, t_n) \in M_n(f)(a, b)$. Put $z = f(t_1)$. Choose disjoint open intervals $U_i (1 \leq i \leq n)$ contained in (a, b) such that $f(t_i) \geq f(t)$ for $t \in U_i$ ($1 \leq i \leq n$). Since f is not constant on U_i we have

$$\bigcap_{i=1}^{n} f(U_i) \supseteq (w, z],$$

where $w < z$. Then g is not monotonic on any subinterval of (w, z), and so has local maxima and minima in (w, z). Let u be a local maximum of g in (w, z). For each i,

$1 \le i \le n$, choose $x_i \in U_i$ with $f(x_i) = u$. Then clearly $(x_1, x_2, \dots, x_n) \in M_n(g \circ f)(a, b)$. A similar argument applies to minima.

4. The set U is an ideal of S. As we remarked earlier, we can regard U as J_1, and so a proof similar to **3** will establish this result. However, a simpler proof is possible. Let $f \in U$, $g \in S$. Now if $g \circ f$ is injective on (a, b) then so is f, and so if $f \in U$ then $g \circ f \in U$. Suppose that $f \circ g$ is injective on $(a, b) \subseteq I$. Then f is injective on $g((a, b))$. As $f \in U$, $g((a, b))$ must be trivial and so g is constant on (a, b). Therefore $f \circ g$ is constant on (a, b). This is a contradiction. Hence $f \circ g \in U$.

5. The set L is an ideal of S. Let $f \in L$ and $g \in S$. Clearly $g \circ f \in L$. Also if (a, b) is a subinterval of I then either g is constant on (a, b) and so $f \circ g$ is constant there, or $g((a, b))$ contains an interval (c, d). Choose a subinterval (p, q) of (c, d) on which f is constant and a subinterval (r, s) of (a, b) with $g((r, s) \subseteq (p, q)$. Then $f \circ g$ is constant on (r, s). Hence $f \circ g \in L$.

6. The set U is the unique maximal ideal. It is clearly enough to show that $S \circ f \circ S = S$ for all $f \in S \setminus U$, since $U \ne S$ is obvious. Let $f \in S \setminus U$. Choose a subinterval of I on which f is strictly monotonic. By replacing f by $\hat{f} \circ \hat{\iota} \in f \circ S$, if necessary, we can assume that f is strictly increasing in an interval $(a, b) \subset I$. Define

$$g(t) = f(a + (b - a)t) \qquad (t \in I).$$

Then g is strictly increasing on I, and so has a continuous inverse g^{-1} mapping $[g(0), g(1)]$ onto I. Let $h \in S$ be any continuous extension of g^{-1} to $[0, 1]$. Then $\iota = h \circ g \in S \circ f \circ S$. Hence $S \circ f \circ S = S$.

7. For $n \ge 1$, J_{n+1} is a proper subset of J_n. This is the most complicated part of the proof. We need to construct a function f in $J_n \setminus J_{n+1}$. To do this we construct a sequence (f_k) of functions in S uniformly convergent to f. We shall define the functions iteratively. In order to describe the construction we first explain the basic idea behind the iterative step. Let $f \in S$ and (a, b) be a subinterval of I on which f is strictly increasing. Let $g \in S$ with $g(0) = 0$, $g(1) = 1$. We make a modification h of f described as *inserting g in f between a and b*; the new function is the following function h:

$$h(t) = \begin{cases} f(t) & (0 \le t \le a, \text{ and } b \le t \le 1) \\ f(a) + (f(b) - f(a))g(\frac{t-a}{b-a}) & (a < t < b). \end{cases}$$

$$f \qquad\qquad g \qquad\qquad h$$
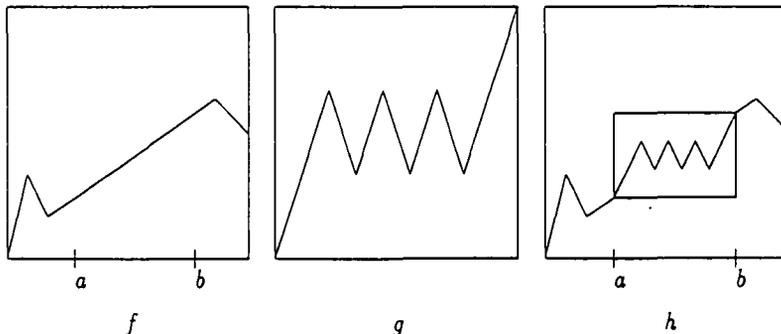
Essentially we have replaced the graph of f between a and b by a suitably scaled copy of the graph of g in such a way as to obtain a continuous function. If f were decreasing rather than increasing on (a, b) we should insert $\hat{g}$ instead of g between a and b, in a similar way. In our present construction we shall use a function g that has n-fold local maxima and minima, which will then be inherited by h.

The function f_k constructed at the kth stage will have the following properties.

(i) It will be piecewise linear, so that the graph will consist of a finite number of straight lines between *vertices*.

(ii) None of the lines will be horizontal. Some of the vertices (but not necessarily all) will be at local maxima and minima of f_k; let us call these vertices the *peak* points and *trough* points of f_k. As we travel from left to right the peak and trough points will alternate, and we can speak of *adjacent* peak and trough points.

(iii) Each local maximum and each local minimum of f_k will be taken precisely n times.

The function g is to have precisely n peak and n trough points in its graph (excluding $(0, 0)$ and $(1, 1)$). It is defined by the following rules:

(i) g is linear between $\frac{k}{2n+1}$ and $\frac{k+1}{2n+1}$ $(k = 0, 1, 2, \ldots, 2n)$;

(ii) $g(0) = 0, g(1) = 1$;

(iii) $g(\frac{2k+1}{2n+1}) = \frac{2}{3}, g(\frac{2k+2}{2n+1}) = \frac{1}{3}(k = 0, 1, 2, \ldots, n-1)$.

Observe that g takes the same value, $\frac{2}{3}$ (resp. $\frac{1}{3}$), at each of its n peak (resp. trough) points.

We now proceed to detail the construction. Choose a sequence (p_k, q_k) of subintervals of I such that every subinterval of I contains some $(p_k q_k)$ and such that $q_k - p_k \to 0$ as $k \to \infty$. For example we can enumerate the set of all intervals (a, b) with $a, b \in [0, 1] \cap \mathbb{Q}$ as a sequence (p_k, r_k) and then define $q_k = \min\{r_k, p_k + 1/k\}$. Now define $f_0 = \iota$ (the identity function). Suppose that f_{k-1}, with properties (i), (ii) and (iii) described above, has been constructed. Choose a subinterval (a_k, b_k) of (p_k, q_k) such that f_{k-1} is linear on a neighbourhood of $[a_k, b_k]$, $|f_{k-1}(b_k) - f_{k-1}(a_k)| < 2^{-k}$ and there is no vertex of f_{k-1} whose y-coordinate is in the closed interval from $f_{k-1}(a_k)$ to $f_{k-1}(b_k)$. This is possible because f_{k-1} has only a finite number of vertices. We then obtain f_k by inserting either g or $\hat{g}$ into f_{k-1} according as f_{k-1} is increasing or decreasing on (a_k, b_k). Observe the following facts. Firstly any vertex of f_{k-1} is also a vertex of f_k, and so f_k agrees with f_{k-1} at these points. Also any peak (resp. trough) point of f_{k-1} is a peak (resp. trough) point of f_k. The values of f_k at those of its peak and trough points which are not vertices of f_{k-1} are different from the values of f_{k-1} at any of its peak and trough points. Next

if u and w are trough points and v is a peak point of f_{k-1} with $u < v < w$, and if
$f_{k-1}(u) \leqslant f_{k-1}(t) \leqslant f_{k-1}(v)$ and $f_{k-1}(w) \leqslant f_{k-1}(s) \leqslant f_{k-1}(v)$ for $t \in (u, v)$, $\quad\quad$ (*)
$s \in (v, w)$, then these same inequalities hold with f_{k-1} replaced by f_k.

Clearly $\|f_k - f_{k-1}\| < 2^{-k}$ (where the norm is the supremum norm on $C(I)$) so that the sequence (f_k) is Cauchy and therefore converges uniformly to some $f \in S$. From (*) above we see if u, v and v, w are adjacent trough and peak points of f_k for some k then, for every $l \geqslant k$, $f_l(v) \geqslant f_l(t)$ for all $t \in (u, w)$. Thus v is a maximum for f_l on an interval (u, w) which does not vary with l, so that v is also a local maximum for f. Notice too that, for the peak point v of f_k, $f_l(v) = f_k(v)$ for all $l \geqslant k$, so that $f(v) = f_k(v)$. Naturally the

same conclusions hold for trough points. Because every interval (a, b) contains some interval (p_k, q_k), and by construction $M_n(f_k)(p_k, q_k) \neq \varnothing$, we see that $M_n(f)(a, b) \neq \varnothing$. Similarly, $m_n(f)(a, b) \neq \varnothing$. We have proved that $f \in J_n$.

Next, we claim that a point $(0, 1)$ is a local maximum or minimum of f if and only if it is a local maximum or minimum of f_k for some k. We have already observed that every local maximum or minimum of f_k is also one for f and that f agrees with f_k at that point. Suppose that t is not such a point of $(0, 1)$. Let $\varepsilon > 0$ with $(t - \varepsilon, t + \varepsilon) \subset I$. We can find l so that (p_l, q_l), and so (a_l, b_l), is contained in $(t - \varepsilon, t)$. Therefore there are peak points of f_l in $(t - \varepsilon, t)$. Similarly we can find m so that $(t, t + \varepsilon)$ contains peak points of f_m. Put $k = \max\{l, m\}$. Then there exist peak points of f_k in $(t - \varepsilon, t + \varepsilon)$ either side of t. Now t is not a peak or trough point of f_k. So we can find adjacent peak and trough points in $(t - \varepsilon, t + \varepsilon)$ with one on each side of t. If we let $\varepsilon \to 0$ we can obtain sequences $u_l \nearrow t$, $v_l \searrow t$ where each pair (u_l, v_l) is an adjacent pair of peak and trough points of some f_k. By taking a subsequence we can assume that all u_l's are peak points and v_l's are trough points, or vice-versa. Let us assume the former is true. Also, since the values of f_k at the peak and trough points not belonging to f_{k-1} are not the values of f_{k-1} at its trough points, we can assume, again by taking subsequences, that $f(u_l) > f(u_{l+1})$, $f(v_{l+1}) > f(v_l)$. (Recall that the value of f at u_l is the value of the functions f_k which have u_l as peak point). We then have

$$f(u_l) > f(u_{l+1}) \geqslant f(t) \geqslant f(v_{l+1}) > f(v_l),$$

by $(*)$ above and so t is not a local maximum of f.

Now consider the local maxima and minima of f. These occur as sets of n of each introduced at each step in the inductive process. At each stage the values taken at the new local maxima and minima are different from those taken at all previous local maxima and minima. So f takes each of its maximum or minimum values precisely n times. Hence $f \notin J_{n+1}$.

That completes the proof of **7**.

8. The set J_2 is a proper subset of U. Since $U = J_1$, this is a special case of **7**. Here there seems to be no simpler proof (in the way that **4** is simpler than **3**).

9. The ideal L is not equal to K. We construct $f \in L \backslash K$ in a similar way to the construction in $J_n \backslash J_{n+1}$. The difference is in the choice of the function g. We define g by
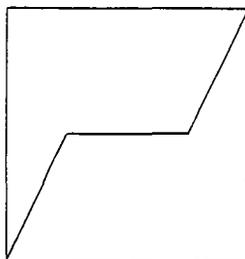
g

the formula

$$g(t) = \begin{cases} \frac{3}{2}t & (0 \leqslant t \leqslant \frac{1}{3}) \\ \frac{1}{2} & (\frac{1}{3} < t < \frac{2}{3}) \\ \frac{3}{2}t - \frac{1}{2} & (\frac{2}{3} \leqslant t \leqslant 1) \end{cases}$$

At each stage we obtain a monotonic increasing, piecewise linear function which is strictly increasing on a finite number of intervals and constant on another finite set of intervals. We begin as before with $f_0 = \iota$. We suppose we have constructed f_{k-1}. Consider the interval (p_k, q_k). If f_{k-1} is constant on some interval contained in (p_k, q_k) we just put $f_k = f_{k-1}$. If not we choose a subinterval (a_k, b_k) on which f_{k-1} is strictly increasing and $|f_{k-1}(a_k) - f_{k-1}(b_k)| < 2^{-k}$, and so that $[a_k, b_k] \subseteq (p_k, q_k)$ and insert g between a_k and b_k. Then we observe that if f_{k-1} is constant on an interval so is f_k. We put $f = \lim_k f_k$. Then clearly $f \in L$. Since $f(0) = 0 \neq 1 = f(1)$, we have $f \notin K$.

The proof of Theorem 2.1 is now complete.

10. One-sided ideals. We remarked in the introduction that we believe that $S(I)$ has $\mathfrak{c}$ distinct two-sided ideals. We should point out that it is easy to find $\mathfrak{c}$ distinct one-sided ideals. Indeed, it is easy to see that for each $a \in I$

$$R_a = \{f \in S(I) : f(I) \subseteq [0, a]\}$$

is a right ideal and that

$$L_a = \{f \in S(I) : f \text{ is constant on } [0, a]\}$$

is a left ideal. These ideals are singly generated (R_a by the function $t \mapsto at(t \in I)$, and L_a by the function g satisfying $g(t) = 0$ $(0 \leqslant t \leqslant a)$, $g(t) = (t - a)/(1 - a)$ $(a \leqslant t \leqslant 1)$). Moreover they are closed in the uniform topology and a fortiori in the topology of pointwise convergence. By way of contrast, K is the only pointwise closed two-sided ideal; this fact can be found in [1], where a full discussion of pointwise closed ideals is presented.

11. A generalization. We are indebted to the referee for the following extension of Theorem 2.1.

LEMMA 11.1. *Let S be a semigroup with identity, and let e be an idempotent of S. Let J and K be ideals of the subsemigroup eSe of S with J a proper subset of K. Then the ideals SJS and SKS are distinct.*

Proof. Let $k \in K \backslash J$. Clearly $k \in SKS$. Suppose that $k \in SJS$. We write $k = xjy$ with $x, y \in S$, $j \in J$. Since $j, k \in eSe$ we can write $k = exjye = exejeye$. Since exe and eye belong to eSe, k is in J. This contradiction shows that $SKS \neq SJS$.

THEOREM 11.2. *Let X be a normal topological space which contains an arc (i.e. a subset homeomorphic to I). Then $S(X)$ contains a countably infinite descending chain of ideals.*

Proof. Clearly we can assume that I is actually a subset of X. Since X is normal, the Tietze extension theorem shows that the identity map of I to itself extends to a continuous

function f mapping X into I. We can regard f as an element of $S(X)$. Then f is idempotent. For $h \in S(I)$ define $\phi(h) \in S(X)$ to be $h \circ f$ (again regarded as an element of $S(X)$). Then $f \circ \phi(h) \circ f = \phi(h)$. It is easy to check that ϕ is a semigroup isomorphism of $S(I)$ onto the subsemigroup $f \circ S(X) \circ f$ of $S(X)$. The theorem now follows immediately by using Lemma 11.1 and Theorem 2.1.

REFERENCES

1. A. M. Forouzanfar, Closed ideals in semigroups of continuous selfmaps, *Semigroup Forum*, to appear.

2. S.-T. Hu, *Homotopy theory* (Academic Press, 1959).

3. K. D. Magill Jr., A survey of semigroups of continuous selfmaps, *Semigroup Forum* **11** (1975/76), 189–282.

4. K. D. Magill Jr., On the ideals of the semigroup of the 1-sphere, *Czech. Math. J.* **39 (114)** 1989, 248–261.

5. K. D. Magill Jr., Congruences on $S(X)$, in *The analytical and topological theory of semigroups—Trends and developments*, ed. K. H. Hofmann et al. (de Gruyter, Berlin, 1990), pp. 379–396.

DEPARTMENT OF PURE MATHEMATICS
THE UNIVERSITY
SHEFFIELD S3 7RH
ENGLAND