On the Functional Equation $f \circ f = g$

Vlad Vicol* and Apostol Apostolov†

Abstract

The functional equation $f \circ f = g$ for continuous g has no continuous solutions defined on intervals. If f may have finitely many discontinuities, we show that solutions exist on closed, but not on open intervals.

1 Introduction

The functional equation

$$f(f(x)) = g(x) \quad \text{for all} \quad x \in \mathbb{R} \tag{1}$$

with a strictly decreasing function g has no solution f in the class of continuous functions. Indeed, because $g : \mathbb{R} \to \mathbb{R}$ is strictly decreasing, it is injective. This implies injectivity of f, and, by continuity, f is strictly monotone. In both cases, $f \circ f$ is strictly increasing, contradicting the decay of g.

Nevertheless, if we allow f not to be continuous, even for the easiest example of a decreasing function g, say $g = -id$, the problem becomes intricate and very complex (Section 2). For countable many discontinuities we can construct a solution of $f \circ f = -id$, but in the class of functions with **finitely** many points of discontinuity the situation is not quite predictable: for f defined on a closed interval there exist solutions (see Proposition 2), but for f defined on open intervals (e.g. $\mathbb{R}$), no solution exists (see Theorem 1).

All the results we obtain for functions $f, g : I \to I$ for an open interval I extend to functions $\tilde{f}, \tilde{g} : \mathbb{R} \to \mathbb{R}$, since any open interval is homeomorphic to $\mathbb{R}$. For example given homeomorphism $h : (-1,1) \to \mathbb{R}$, $h(x) = \tan(\frac{\pi}{2}x)$, and function $f : (-1,1) \to (-1,1)$ such that $f \circ f = -id_{(-1,1)}$. Then function $\tilde{f} : \mathbb{R} \to \mathbb{R}$, $\tilde{f} = h^{-1} \circ f \circ h$ satisfies $\tilde{f} \circ \tilde{f} = -id_{\mathbb{R}}$.

*International University Bremen; 28759 Bremen, Germany; v.vicol@iu-bremen.de
†International University Bremen; 28759 Bremen, Germany; a.apostolov@iu-bremen.de

2 f∘f = -id and the "odd-even effects" on orbits

We denote by f^n the function $\underbrace{f \circ f \circ ... \circ f}_{\text{n times}}$.

The *orbit* of a point x_0 is the set $\mathrm{orb}(x_0) = \{f^n(x_0) \mid n \in \mathbb{N}\}$. Similarly, the orbit of an interval I is the set of sets $\mathrm{orb}(I) = \{f^n(I) | n \in \mathbb{N}\}$.

Let us first focus on the functional equation $f \circ f = -id$. In this case it is easy to show that f is bijective. Indeed, if $f(x) = f(y)$, then $f(f(x)) = f(f(y))$ and hence $-x = -y$. Surjectivity of f follows from the equation $f(-f(x)) = x$ for all $x \in (-1, 1)$. Further observe that $f(x) = x$ implies $f(f(x)) = f(x) = x$, hence $-x = x$ and $x = 0$. Thus 0 is the only fixed point of f.

Then, for all $x \in (-1, 1)$, $f^4(x) = f(f(f(f(x)))) = -(-x) = x$; hence, the number of elements in any orbit has to divide 4. If $x = f(f(x))$, then $x = -x$ and $x = 0$. If $x = f(x)$, then by uniqueness of the fixed point $x = 0$. Thus the orbit of every point, except zero, has exactly 4 elements. Also the orbit of every point of discontinuity has 4 elements. These observations are central to the construction and the understanding of the functional equation $f \circ f = -id$.

But does it matter whether there are finitely or infinitely many such points? We shall see that this does matter and that in addition to this the domain on which our functions are defined is equally as important.

Proposition 1. *There exists a function $f : (-1, 1) \to (-1, 1)$ with countably many points of discontinuity such that $f(f(x)) = -x$ for all $x \in (-1, 1)$.*

Proof. Let us construct f inductively:
Step 1: Split the open interval $(-1, 1)$ into 5 parts

$$I_1^1 = (-1, -\tfrac{2}{3}],\ I_2^1 = (-\tfrac{2}{3}, -\tfrac{1}{3}],\ I_3^1 = (-\tfrac{1}{3}, \tfrac{1}{3}),\ I_4^1 = [\tfrac{1}{3}, \tfrac{2}{3}),\ I_5^1 = [\tfrac{2}{3}, 1).$$

Map f acts as follows (homeomorphically): $I_5^1 \to I_2^1 \to I_1^1 \to I_4^1 \to I_5^1$.
Step k: Split the middle interval $I_3^{k-1} = (-\frac{1}{3^{k-1}}, \frac{1}{3^{k-1}})$ into 5 parts:

$$\begin{array}{lll} I_1^k = (-\frac{1}{3^{k-1}}, -\frac{1}{3^k}], & I_2^k = (-\frac{2}{3^k}, -\frac{1}{3^k}], & I_3^k = (-\frac{1}{3^k}, \frac{1}{3^k}), \\ I_4^k = [\frac{1}{3^k}, \frac{2}{3^k}), & I_5^k = [\frac{2}{3^k}, \frac{1}{3^{k-1}}). & \end{array}$$

Map f acts as follows: $I_5^k \to I_2^k \to I_1^k \to I_4^k \to I_5^k$.
Also set $f(0) = 0$. Then the constructed function is obviously well defined, injective, and surjective (since $\lim_{k\to\infty} \left(\frac{1}{3}\right)^k = 0$). Moreover, it satisfies $f(f(x)) = -x$, hence the construction is done. ■

Proposition 2. *There exists a functions $f : [-1, 1] \to [-1, 1]$ with finitely many points of discontinuity such that $f(f(x)) = -x$ for all $x \in [-1, 1]$.*

Proof. Let's construct f. Set $f(0) = 0$ and let f map $1/2 \to -1 \to -1/2 \to 1 \to 1/2$. The remaining open intervals are mapped as in below, i.e., $f((-1/2, 0)) = (-1, -1/2)$, $f((0, 1/2)) = (1/2, 1)$ (using symmetry over the vertical axis), and $f((1/2, 1)) = (-1/2, 0)$, $f((-1, -1/2)) = (0, 1/2)$ (using symmetry over the origin). Compare with Figure 2, where the orbit of a point x is depicted using symmetry. ■

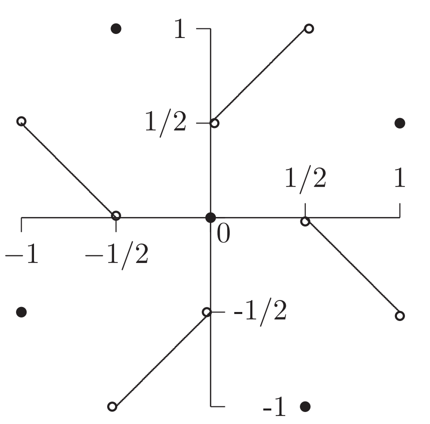


Figure 1: The graph of f

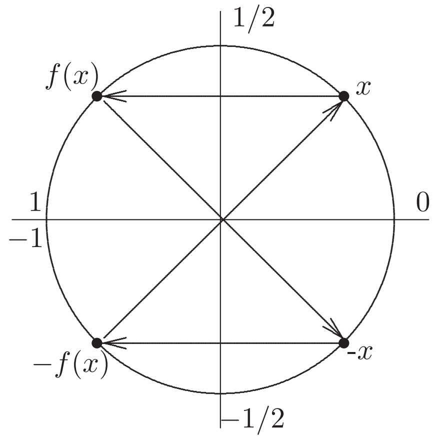


Figure 2: The orbit of x

Note that solving the equation $f(f(x)) = -x$ is equivalent to the fact that the graph of f is symmetric under a plane rotation of $90°$ (see Figure 1).

Theorem 1. *There exists* ***no*** *function $f : (-1, 1) \to (-1, 1)$ with finitely many points of discontinuity such that $f(f(x)) = -x$ for all $x \in (-1, 1)$.*

Proof. Suppose there exists f which satisfies (1). Then, as showed above, f is bijective and 0 is the only fixed point. We also proved that the orbit of every point, except zero, contains exactly 4 elements.
Consider the set $A = \{a_1, a_2, ..., a_n\}$ of points of discontinuity of f, and define

$$A^* := \mathrm{orb}(A) \cup \{0\}.$$

Then A^* contains all the discontinuity points of f and their orbits. This means that A^* is fully invariant under f, i.e.

$$f(y) \in A^* \text{ iff } y \in A^*.$$

For further use let's set $A^* = \{b_1, b_2, \ldots, b_m\}$, where without loss of generality we can assume that $b_1 < b_2 < \ldots < b_m$. Since the orbit of any point in A^* has 4 elements, with the exception of the fixed point 0, we deduce

$$m = 4k + 1 \quad \text{for some } k \in \mathbb{N}. \tag{2}$$

Then define $I_1 = (-1, b_1)$ and for $2 \leq j \leq m$ let $I_j = (b_{j-1}, b_j)$; $I_{m+1} = (b_m, 1)$.
Using Lemma (1) below, for $I = (-1, 1)$ and $M = A^*$, we get that the f-orbit of each I_j has 4 elements, and hence $m + 1 = 4p$ for some $p \in \mathbb{N}$. But this contradicts (2): $4p - 1 = m = 4k + 1$, and therefore the existence of f with finitely many points of discontinuity. ■
Here we can see the combinatorial aspect of the problem. We show that the orbits of the points of discontinuity are even (4), and hence the number of intervals in the induced partition should be odd (4k+1). But it turns out that the orbit of the intervals is also even (4) and hence the contradiction is reached.

3 f∘f=g, g continuous

We shall now pose a more general problem: given a strictly decreasing continuous function g, does there exist a function f with **finitely many points of discontinuity** such that $f \circ f = g$? The answer of this question is our main result in this article.

Main Theorem: *Let $g : \mathbb{R} \to \mathbb{R}$ be a strictly decreasing and continuous function. There exists **no** function $f : \mathbb{R} \to \mathbb{R}$ with only finitely many points of discontinuity such that $f(f(x)) = g(x)$ for all $x \in \mathbb{R}$.*

Before we can prove this theorem we will first have to provide the following lemmas:

Lemma 1. *Let $I \subseteq \mathbb{R}$ be an open interval and $f : I \to I$ be a bijective function. Consider a set of points $M \subset I$ with $|M| = m$ for some $m \in \mathbf{N}$, such that $x \in M$ iff $f(x) \in M$. Let $\{I_j\}_{j=1}^{m+1}$ be the partition of I into open intervals induced by M. Then:*

i. If f is continuous on each I_j, then $f(I_j) = I_k$ for some $1 \leq k \leq m+1$.

ii. If in addition to i., $f \circ f$ is strictly decreasing, then $f^4(I_j) = I_j$ for all $1 \leq j \leq m + 1$.

iii. In addition to i., ii., if there exists a point $x_0 \in I$ such that $f(f(x_0)) = x_0$, then x_0 is unique. Further, if $x_0 \in M$, then the orbit of each I_j has 4 elements.

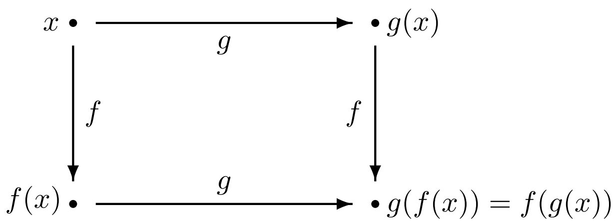
Proof. Fix some I_j in the partition. Using a canonical numbering of the I_j's we may assume that for $p < j$, I_p is to the left of I_j and vice versa.
f maps I_j to $f(I_j)$ continuously and bijective; hence $f(I_j) = (c, d)$ for some $c, d \in I$. Since $I_j \cap M = \emptyset$ and M is fully invariant under f, we conclude that $f(I_j) \cap f(M) = \emptyset$, i.e., $(c, d) \cap M = \emptyset$. This implies that $(c, d) \subseteq I_k$

for some $1 \leq k \leq m+1$. If $(c,d) \subsetneq I_k$, then c (or d) $\in I_k$. But using f^{-1}, we get $f^{-1}(c) \in I_j$ and thus $c = f(f^{-1}(c)) \in f(I_j) = (c,d)$, a contradiction. Thus we've showed that $f(I_j) = I_k$ for some $1 \leq k \leq m+1$, and statement (i) is proven.

Using that f^2 is strictly decreasing, we immediately get that f^4 is strictly increasing. There exist *exactly* $j-1$ intervals to the left of I_j. Then by applying f^4 and (i), there will be *exactly* $j-1$ intervals to the left of $f^4(I_j)$. Thus we conclude that $f^4(I_j) = I_j$ and (ii) is proven.

By (ii), we get that the orbit of each I_j has either 1, 2, or 4 elements (since it has to divide 4 a orbit of 3 is impossible). If $f \circ f$ has a fixed point, then it is unique since the function $h_1 = f^2 - \mathrm{id}$ is strictly decreasing. If we would have that $f(f(I_j)) = I_j$ then, using the continuity of h_1 on each I_k and the Intermediate Value Theorem, we would get that there exists $x \in I_j$ such that $h_1(x) = 0$, i.e. $x = x_0 \in M$, a contradiction. Using that f^2 is injective and $f^2(x_0) = x_0$, we get that $f(x_0) = f(f^2(x_0)) = f^2(f(x_0))$ and hence $f(x_0) = x_0$. If there exists another fixed point x of f then $f(x) = x = f(f(x))$ and hence x_0 is the only fixed point of both f and $f \circ f$. Using the same argument as above, the function $h_2(x) = f(x) - x$ is continuous on each I_j and hence we cannot have $f(I_j) = I_j$. Thus the orbit of each interval can only have 4 elements, and (iii) is proven. ■

Lemma 2. *Let I and J be open intervals, such that $J \subseteq I$. Take two functions $f, g : I \to J$, such that f and g commute ($f \circ g = g \circ f$) and g is continuous and bijective. Then, f is continuous at x_0 iff f is continuous at $g(x_0)$.*



Proof. Due to symmetry it is enough to prove only one implication. One only needs to switch the mapping g with g^{-1}. Note that the bijectivity and continuity of g implies the bijectivity and continuity of g^{-1}.
Let f be continuous at $x_0 \in I$ and set $y_0 = g(x_0) \in J$. Take any sequence $(y_n) \subset J$, such that $y_n \to y_0$. g^{-1} is bijective, thus for any $n \in \mathbb{N}$ there exists exactly one $x_n \in I$ such that $g(x_n) = y_n$. Further, $x_n \to x_0$, since g^{-1} is continuous. Using the fact that f is continuous at x_0 we conclude

$$f(y_n) = f(g(x_n)) = g(f(x_n)) \to g(f(x_0)) = f(g(x_0)) = f(y_0)$$

Thus f is continuous at $g(x_0) = y_0$, and the lemma is proven. ■

Lemma 3. *Let $I \subset \mathbb{R}$ be an open interval and $f, g : I \to I$ such that $f \circ f = g$, g is continuous and bijective. Then if f has only finitely many points of discontinuity, the orbit of each non-fixed point has exactly 4 elements.*

Proof. Let A be the set of points of discontinuity of f. Since g and f obviously commute, we can apply Lemma 2 and get that the g-image of a discontinuity point of f is a discontinuity point of f, i.e. $g(A) \subseteq A$. Since A is finite and g injective we conclude that $g(A) = A$. Take any $a \in A$. Let $l \in \mathbb{N}_0$ be the number of elements in A that are strictly smaller than a. Since $g \circ g$ is strictly increasing and $g(g(A)) = A$, there will be *exactly* l elements in A that are strictly less than $g(g(a))$, hence $g(g(a)) = a$.
$f(a) \neq a$ and $g(a) = f(f(a)) \neq a$ since a is not a fixed point of f or g, but $f^4(a) = a$. ■

Now we can proceed with the proof of our Main Theorem:

Proof of Main Theorem. Assume that there exists an f satisfying the conditions of the theorem.
Obviously, $g(\mathbb{R}) \subseteq \mathbb{R}$, and for all $n \in \mathbb{N}$, $g^{n+1}(\mathbb{R}) \subseteq g^n(\mathbb{R})$. Since g is continuous and strictly monotone, $g^k(\mathbb{R})$ is an open interval in $\mathbb{R}$. Let

$$I = \bigcap_{k=0}^{\infty} g^k(\mathbb{R}).$$

First, let's show that I is not empty. Since g is strictly decreasing, $h(x) = g(x) - x$ has $\lim_{x\to\infty} h(x) = -\infty$ and $\lim_{x\to-\infty} h(x) = \infty$. Since h is continuous and strictly decreasing, the equation $h(x) = 0$ has exactly one solution, say x_0. Note that this implies $f(x_0) = f(g(x_0)) = g(f(x_0))$ and thus $f(x_0) = x_0$. $g(x_0) = x_0$ implies $g^k(x_0) = x_0$ for all $k \in \mathbb{N}$, hence $x_0 \in I$. Since $I \neq \emptyset$ is a countable intersection of open intervals, we can say that it is an interval. Further $g(I) = I$, that is: g is bijective on I. It is easy to see that I cannot be a semi-open interval due to the monotonicity of g. Hence I is either closed or open.

If I is closed, then $I = [a, b]$ for some $a < b$. But since g is strictly decreasing, we get $g(a) = b$, $g(b) = a$. So since the endpoints would not intervene in our reasoning, we consider I as being open.
As one can easily see, if $x \notin I$, then $g(x) \notin I$. Now define A as the set of points at which f is discontinuous. Since it is finite we can represent it as $A = \{a_1, a_2, \ldots, a_n\}$.

Take any $x_0 \in \mathbb{R} \setminus A$, and let $y_0 = g(x_0)$. Since A is finite and g is a continuous local bijection, we can find a neighborhood U of y_0 and a neighborhood V of x_0 such that $U \cap A = \emptyset$, $V \cap A = \emptyset$ and $g(V) = U$. Obviously, f and g commute. Hence we can apply Lemma 2 for $g : I \to U$, $U, V \subseteq I$, and get $y_0 = g(x_0) \in \mathbb{R} \setminus A$. Conversely, $g(x_0) \in \mathbb{R} \setminus A$ implies $x_0 \in \mathbb{R} \setminus A$. Hence we find that A is fully invariant under f (and g). But from this we conclude that $g(A) \subseteq A$, and since g is injective $g(A) = A$. Hence $A \subset I$.

By applying Lemma 3 to $f, g : I \to I$, we conclude that the orbit of each a_i has 4 elements. Then the set $A_{ext} = \bigcup_{k=1}^{4} \{f^k(a_i) \,|\, i = 1, \ldots, n\} \cup \{x_0\}$ contains all points of discontinuity of f and their orbits. By lemma 2 and $g(x_0) = x_0$, the cardinality of A_{ext} is $4p + 1$ for some $p \in \mathbb{N}$. The partition $I_1 = (a, b_1)$, $I_2 = (b_1, b_2)$, ..., $I_{4p+1} = (b_{4p}, b_{4p+1})$, $I_{4p+2} = (b_{4p+1}, b)$, induced on I by A_{ext}, should then be made of 4p+2 open, disjoint intervals (where we denote the b_i as being the elements of A_{ext} such that $b_i < b_j \Leftrightarrow i < j$). Now we can apply Lemma 1 to I with $M = A_{ext}$ and find that the orbit of each I_j has 4 elements, and hence the number of intervals should be divisible by 4. But this contradicts the fact that the number of elements in the partition is $4p + 2$. Hence we got a contradiction to the existence of f. ■

4 f∘f=g, g discontinuous

We now consider the equation $f \circ f = g$ in the case when $g : \mathbb{R} \to \mathbb{R}$ is *not* continuous, but both f and $g : \mathbb{R} \to \mathbb{R}$ are in the class of functions with finitely many points of discontinuity. Is it possible to find solutions of equation (1)?

Proposition 3. *Let $I \subset \mathbb{R}$ be an open interval. There exist functions $f, g : I \to I$ with finitely many points of discontinuity such that g is strictly decreasing and $f(f(x)) = g(x)$ for all $x \in I$.*

Proof. Here we do not start by picking a function g and then construct f, but construct f such that f and $f \circ f$ respect the conditions of the theorem. To make calculations easy, we consider $I = (-16, 16)$, instead of $(-1, 1)$.

$$\text{Let } f : (-16, 16) \to \mathbb{R}, \quad f(x) = \begin{cases} \frac{x}{2} + 2 : x \in (-16, -8) \\ -\frac{x}{2} + 10 : x \in (-8, 0) \\ -\frac{x}{2} - 10 : x \in (0, 8) \\ \frac{x}{2} - 2 : x \in (8, 16) \\ -2 : x = -8 \\ -10 : x = 0 \\ 2 : x = 8 \end{cases}$$

It is trivial to check that f has 3 points of discontinuity and that $g := f \circ f$ is strictly decreasing with also 3 points of discontinuity. To get a better picture of what happens, see Figure 3 and Figure 4. Again we observe that the graph of f is symmetric (except for one point) under a plane rotation of $180°$ (also see Figure 1). ■

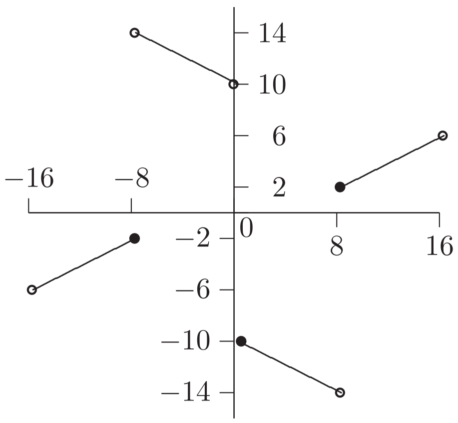


Figure 3: The graph of f

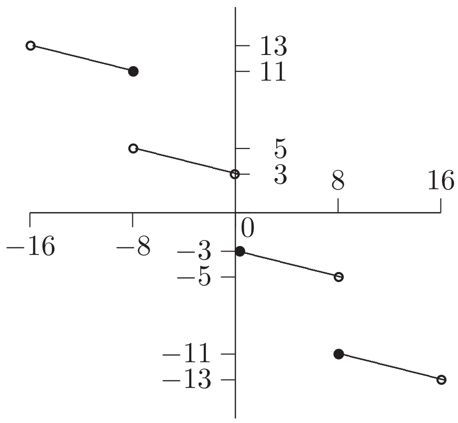


Figure 4: The graph of g

To sum up, we have seen that the aforementioned equation $f(f(x)) = g(x)$ has solutions in the class of *arbitrary* functions (e.g. when $g(x) = -x$). As we have illustrated in our main result, there exist no solutions of (1) when $g(x)$ belongs to the class of continuous, strictly decreasing functions. Surprisingly, when tempted to generalize further to the case when $g(x)$ has finitely many discontinuous points, we obtain an example which shows that (1) actually possesses a solution.
Some other types of functional equations and their relations to dynamical systems can be found in [1] and [2].

ACKNOWLEDGEMENTS: We would like to thank Alexei Belov and Götz Pfander who introduced us to the functional equation $f \circ f = -id$. We are in particular thankful to Alexei Belov who has encouraged us to seek generalizations of this well understood functional equation.

References

[1] Katok, A., Hasselblatt, B.: *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press 1995

[2] Kuczma, M., Choczewski, B., Ger, R.: *Iterative Functional Equations*, Cambridge University Press 1990