IMO SL 2017A6

Let R be a ring. We say that a function $f : R \to R$ is *good* if for any $x, y \in R$,

$$f(f(x)f(y)) + f(x + y) = f(xy). \qquad (*)$$

We are interested in finding all good functions. Unfortunately, this is too hard to solve in general. We present a complete solution for three subcases.

Theorem 0.1. *Let R be a ring such that $2 \in R^\times$ and 3 is not a zero divisor in R. Let (S, ϕ, ι, a) be a 4-tuple with the following data:*

- *S is a ring;*
- *$\phi : R \to S$ is a ring homomorphism;*
- *$\iota : S \to R$ is a group homomorphism with $\phi \circ \iota = \mathrm{id}_S$;*
- *a is an element of $Z(S)$ such that $a^2 = 1$.*

Then $f(x) = \iota(a(1 - \phi(x)))$ works, and all good functions on R arise from some 4-tuple of the above form.

Theorem 0.2. *Let F be a division ring with $\mathrm{char}(F) \neq 2$. Then the only good functions on F are 0, $x \mapsto 1 - x$, and $x \mapsto x - 1$.*

Theorem 0.3. *Let F be a field with $\mathrm{char}(F) = 2$. Then the only good functions on F are 0 and $x \mapsto x + 1$.*

An example of "weird" good function that falls into the description in Theorem 0.1 is

$$R = R'[X], \quad f(P) = (1 - P(1))X^k \;\; \forall P \in R'[X],$$

where R' is an arbitrary ring and $k \in \mathbb{N}$ is arbitrary. Here, Theorem 0.1 holds with $S = R'[X]/\langle X - 1\rangle \cong R'$, $\iota(r) = rX^k$, and $a = 1$. The image of f is the subgroup $\{rX^k : r \in R'\} \subseteq R'[X]$. By changing it to another subgroup and also changing S to another quotient of R, we could generate other good functions. This should explain why the theorem is a reasonable guess in the first place in the general case.

Solution

References:

- `https://artofproblemsolving.com/community/c6h1480146p8693244`.
  Solution in AoPS by **anantmudgal09** (post #75).
- `https://artofproblemsolving.com/community/c6h1480146p29214012`
  Solution in AoPS by **BlazingMuddy** (author of this project, post #176).

The two references prove Theorem 0.2 and Theorem 0.3, respectively. Here, we build up the necessary theory of good functions and use it to prove all three main results. Our proof of Theorem 0.2 and Theorem 0.3 uses this theory to simplify some arguments.

First, it is clear that $x \mapsto 1-x$ is a good function. For any good function $f : R \to R$ and $a \in Z(R)$ such that $a^2 = 1$, it can be checked that $x \mapsto af(x)$ is also a good function. To prove the easy direction of Theorem 0.1, it remains to show:

Lemma 0.4. *Let $\phi : R \to S$ is a ring homomorphism and $\iota : S \to R$ be a group homomorphism with $\phi \circ \iota = \mathrm{id}_S$. For any $f : S \to S$, the function $\iota \circ f \circ \phi$ is good iff f is good.*

Proof. Since $\phi \circ \iota = \mathrm{id}_S$, we know that ι is injective and ϕ is surjective. Thus, $\iota \circ f \circ \phi$ is good iff

$$\begin{aligned} & \forall x, y \in R,\ \iota(f(f(\phi(x))f(\phi(y)))) + \iota(f(\phi(x) + \phi(y))) = \iota(f(\phi(x)\phi(y))) \\ \iff & \forall x, y \in R,\ f(f(\phi(x))f(\phi(y))) + f(\phi(x) + \phi(y)) = f(\phi(x)\phi(y)) \\ \iff & \forall x, y \in S,\ f(f(x)f(y)) + f(x+y) = f(xy), \end{aligned}$$

iff f is good. □

In all three main results, it remains to do the harder direction. We start with the following big observation.

Lemma 0.5. *The period set $I = \{c \in R : \forall x \in R, f(x+c) = f(x)\}$ is a double-sided ideal. The induced function $\tilde{f} : R/I \to R/I$ defined by $[x] \mapsto [f(x)]$ is good with no non-zero period.*

Proof. Fix any $c \in I$. Comparing (*) using $y = c$ and $y = 0$ gives $f(xc) = f(0)$ for all $x \in R$. Then replacing y with yc yields

$$f(x) + f(f(x)f(0)) = f(0) = f(xyc) = f(x+yc) + f(f(x)f(yc)) = f(x+yc) + f(f(x)f(0)),$$

so $f(x+yc) = f(x)$ for all $x, y \in R$. That is, $yc \in I$ for any $y \in R$. Similarly, we also get $cy \in I$ for any $y \in R$.

Since I is a double-sided ideal, there is an induced function $\tilde{f} : R/I \to R/I$ such that $\tilde{f}([x]) = [f(x)]$ for any $x \in R$. Here, $[x]$ denotes the projection of $x \in R$ to R/I. Now it is easy to check that $\tilde{f}$ is good.

Finally, we check that $\tilde{f}$ has no non-zero period. For any $x, y \in R$, if $[f(x)] = [f(y)]$, then we get

$$f(0) - f(x) = f(f(0)f(x)) = f(f(0)f(y)) = f(0) - f(y) \implies f(x) = f(y).$$

Now for any period $c \in R/I$, lift it arbitrarily to $\tilde{c} \in R$. Then $\tilde{f}([x] + c) = \tilde{f}([x]) \iff f(x+\tilde{c}) = f(x)$ for any $x \in R$. The latter yields $\tilde{c} \in I$ and thus $c = 0$. □

From now on, we say that $f : R \to R$ is *reduced good* if f is good and has no non-zero periods. The above lemma means that in order to study good functions, we should study the reduced good ones. We continue with easier observations.

First, plugging $x = y = 0$ into (*) yields $f(f(0)^2) = 0$. Next, we prove:

Lemma 0.6. *For any $a, b \in R$, if $f(a) = f(b)$, then $f(a+1) = f(b+1)$ and $f(-a) = f(-b)$.*

Proof. Plugging $y = 1$ into (*) yields $f(x+1) = f(x) - f(f(x)f(1))$ for any $x \in R$. This clearly implies $f(a+1) = f(b+1)$. Next, plugging $y = -1$ into (*) yields $f(-x) = f(x-1) - f(f(x)f(-1))$ for any $x \in R$. Since $f(a) = f(b)$ and $f(a+1) = f(b+1)$, we get $f(-(a+1)) = f(-(b+1))$. Repeating the previous process once gives $f(-a) = f(-b)$. □

Lemma 0.7. *If f is reduced good, then $f(C) = 0 \iff C = 1$.*

Proof. Since $f^{-1}(0)$ is non-empty, it suffices to show that for any $C \in R$, $f(C) = 0$ implies $C = 1$. We start with $f(Cx) = f(0) + f(C + x)$ for all $x \in R$, obtained from (*). Then we get $f(C + 1) = -f(0)$ and $f(-C^2) = 2f(0)$. Plugging $x = 0$ and $y = -C^2$ into (*) yields $f(2f(0)^2) = -f(0)$. By Lemma 0.6, since $f(C) = f(f(0)^2) = 0$, we get $f(2C) = -f(0)$.

Now plugging $x = y = C$ into (*) yields $f(C^2) = 0$. Plugging $x = C$ and $y = C^2$ yields $f(0) + f(C^2 + C) = f(C^3)$. On the other hand, since $f(C) = 0$, we get $f(C^2 + C) = f(2C + 1) + f(0)$. Since $f(C + 1) = f(2C) = -f(0)$, Lemma 0.6 gives

$$f(2C + 1) = f(C + 2) = f(2C) - f(0) = -2f(0) \implies f(C^2 + C) = -f(0) \implies f(C^3) = 0.$$

Now we are ready for the final step. For any $x \in R$, we write $f(C^4x)$ in two ways, using the fact that $f(C) = f(C^2) = f(C^3) = 0$.

$$f(C^4x) = f(C^2 + C^2x) + f(0) = f(C^2(x + 1)) + f(0) = f(C^2 + x + 1) + 2f(0),$$
$$f(C^4x) = f(C^3 + Cx) + f(0) = f(C(C^2 + x)) + f(0) = f(C^2 + x + C) + 2f(0).$$

Replacing x with $x - C^2$ gives $f(x + C) = f(x + 1)$ for any $x \in R$. Since f is reduced good, this forces $C = 1$, as desired. □

The proof alone implies that for any $f : R \to R$ good and $C \in f^{-1}(0)$, $C - 1$ is a period of f. In particular, $f(1) = 0$ holds and $f(0)^2 - 1$ is a period of f. Plugging $y = 1$ yields

$$f(x + 1) + f(0) = f(x) \quad \forall x \in R. \tag{1}$$

For any ring R, let $Z(R) = \{c \in R : cx = xc \, \forall x \in R\}$ denote the centre of R.

Lemma 0.8. *Let $f : R \to R$ be an injective good function. Then $f(0) \in Z(R)$, $f(0)^2 = 1$, and $f(x) = f(0)(1 - x)$ for any $x \in R$.*

Proof. Since $f(1) = f(f(0)^2) = 0$, injectivity yields $f(0)^2 = 1$. For any $x \in R$, we have $f(f(0)f(x)) = f(0) - f(x)$. So $f(0) - f(x) \in f(R)$, and applying that to the same equation yields

$$f(1 - f(0)f(x)) = f(f(0)(f(0) - f(x))) = f(x) \implies 1 - f(0)f(x) = x \iff f(x) = f(0)(1 - x).$$

Finally, note that $f(f(0)f(x)) = f(f(x)f(0)) = f(0) - f(x)$, so $f(0)f(x) = f(x)f(0)$ for any $x \in R$. Thus $f(1 - x) = f(0)x = xf(0)$ for any $x \in R$, proving that $f(0) \in Z(R)$. □

Lemma 0.9. *Let R be a ring such that 2 is not a zero divisor in R. Let $f : R \to R$ be a reduced good function. Then $f(0) \in Z(R)$, $f(0)^2 = 1$, and $f(x) = f(0)(1 - x)$ for any $x \in R$.*

Proof. By Lemma 0.8, it suffices to show that f is injective. Indeed, consider any $a, b \in R$ and suppose that $f(a) = f(b)$. Then $f(a)$ and $f(b)$ commute, so $f(ab) = f(ba) = f(a + b) + f(f(a)f(b))$. By Lemma 0.6, we get $f(-a) = f(-b)$ and $f(-ab) = f(-ba)$. Hence, we get

$$f(a - b) = f(-ab) - f(f(a)f(-b)) = f(-ba) - f(f(b)f(-a)) = f(b - a).$$

It now remains to show more generally that $f(-c) = f(c)$ implies $c = 0$. Indeed, plugging $x = -1$ and $y = c$ into (*) yields

$$f(f(-1)f(c)) + f(c - 1) = f(-c) \implies f(f(-1)f(c)) + f(c) + f(0) = f(c) \implies f(f(-1)f(c) - 1) = 0.$$

By Lemma 0.7, this yields $f(-1)f(c) = 2$. However, $f(-1) = f(0) + f(0) = 2f(0)$. Since 2 is not a zero divisor in R, we get $f(0)f(c) = 1$. Since $f(0)^2 = 1$, this yields $f(c) = f(0) \iff f(c+1) = 0 \iff c = 0$. □

At this point, Theorem 0.2 has already been proved. Indeed, if F is a division ring, then the only double-sided ideals of F are (0) and F. So, the good functions on F are either zero or reduced good. If $\text{char}(F) \neq 2$, then 2 is not a zero divisor in F and $f(0)^2 = 1 \iff f(0) = \pm 1$.

Proof of Theorem 0.1

Again, let I be the set of periods of f. We take $S = R/I$ and $\phi : R \to R/I$ to be the natural quotient map. From now on, for any $x \in R$, we denote $[x] = \phi(x)$.

By Lemma 0.5, the induced map $\tilde{f} : R/I \to R/I$ is reduced good. Since $2 \in R^\times$, it is guaranteed that 2 is not a zero divisor in R/I. Thus, by Lemma 0.9, there exists $a \in Z(R/I)$ with $a^2 = 1$ such that $\tilde{f}(x) = a(1-x)$ for any $x \in R/I$. Equivalently, $[f(x)] = a(1-[x])$ for any $x \in R$. It remains to find the desired $\iota : S \to R$.

For any $x, y \in R$, we have $[f(x)f(y)] = [a(1-x)a(1-y)] = [(1-x)(1-y)]$. Thus, $f(f(x)f(y)) = f((1-x)(1-y))$, and the original functional equation changes to

$$f((1-x)(1-y)) + f(x+y) = f(xy).$$

We start by proving that the function $g(x) = f(1-x)$ is a group homomorphism. That is, g is additive.

Lemma 0.10. *Let R be a ring and $(G, +)$ be an abelian group. Let $g : R \to G$ be a function such that*

$$g(x+y-xy) + g(1-(x+y)) = g(1-xy) \quad \forall x, y \in R. \tag{1.1}$$

Proof. Plugging $x = y = 0$ yields $g(0) = 0$. Next, plugging $y = 1$ and replacing x with $-x$ yields

$$g(x+1) = g(x) + g(1) \quad \forall x \in R. \tag{1.2}$$

Plugging $y = 0$ and then using (1.2) yields

$$g(-x) = -g(x) \quad \forall x \in R. \tag{1.3}$$

After some rearrangement, (1.1) now yields

$$g(xy + x + y) = g(xy) + g(x+y) \quad \forall x, y \in R. \tag{1.4}$$

Plugging $y = 1$ yields

$$g(2x) = 2g(x) \quad \forall x \in R. \tag{1.5}$$

Then plugging $y = 2$ yields

$$g(3x) = 3g(x) \quad \forall x \in R. \tag{1.6}$$

More generally, one can show that $g(nx) = ng(x)$ for any $n \in \mathbb{N}$ and $x \in R$, but the above two are enough.

... $\square$

Now, the function $x \mapsto g(f(0)x)$ is a group homomorphism. One can check that the kernel contains I since $g(0) = 0$. Thus there is an induced group homomorphism $\iota : S \to R$ with $\iota([x]) = g(f(0)x) = f(1 - f(0)x)$ for all $x \in R$. We claim that this ι indeed works.

We first verify the form f takes. We have $[f(0)] = \tilde{f}(0) = a$. Since $f(0)^2 - 1 \in I$, for any $x \in R$, we have

$$\iota(a(1-[x])) = f(1 - f(0) \cdot f(0)(1-x)) = f(1-(1-x)) = f(x).$$

Finally, we check that $\phi \circ \iota = \mathrm{id}_S$. Indeed, for any $x \in R$,

$$[\iota([x])] = [f(1-f(0)x)] = a(1 - [1 - f(0)x]) = a[f(0)x] = a^2[x] = [x].$$

This proves Theorem 0.1.

Proof of Theorem 0.3

Let F be a field of characteristic 2. As argued before, the good functions on F are either zero or reduced good. By Lemma 0.8, it then suffices to show that reduced good functions over F are injective.

Let $f : F \to F$ be a reduced good function. Fix some $a, b \in F$ such that $f(a) = f(b)$. By Lemma 0.6, we get $f(a+1) = f(b+1)$. By Lemma 0.7, we have $a = 0 \iff b = 0$. The remaining case is when a and b are non-zero.

We start with the following observation. Consider (*) with $(x, y) = (a+1, b^{-1}+1)$. Since $f(a+1) = f(b+1)$, applying (1) yields

$$f((a+1)(b^{-1}+1)) = f((a+1)+(b^{-1}+1)) \implies f(ab^{-1}+a+b^{-1}+1) = f(a+b^{-1}).$$

Then, by (0.6), we get

$$f(ab^{-1}+a+b^{-1}) = f(1+a+b^{-1}).$$

By symmetry, we have

$$f(ba^{-1}+b+a^{-1}) = f(1+b+a^{-1}).$$

Now notice the identity

$$(1+a+b^{-1})(1+b+a^{-1}) = (ab^{-1}+a+b^{-1})(ba^{-1}+b+a^{-1}).$$

As a result, by plugging the appropriate values into (*), we get

$$f((1+a+b^{-1})+(1+b+a^{-1})) = f((ab^{-1}+a+b^{-1})+(ba^{-1}+b+a^{-1})).$$

Letting $C = a+b+1$ and $D = a^{-1}+b^{-1}+1$, the above equation is equivalent to saying that

$$f(C+D) = f(CD+1) = f(CD)+1.$$

But then plugging $(x, y) = (C, D)$ into (*) yields

$$f(f(C)f(D)) = 1 \iff f(C)f(D) = 0 \iff f(C) = 0 \vee f(D) = 0 \iff C = 1 \vee D = 1.$$

By definition of C and D, and by $\mathrm{char}(F) = 2$, this is equivalent to

$$a+b+1 = 1 \vee a^{-1}+b^{-1}+1 = 1 \iff a = b \vee a^{-1} = b^{-1}.$$

But $a^{-1} = b^{-1}$ yields $a = b$. So, regardless, we have obtained $a = b$. This proves that f is injective.