

Continuous Solutions of the Gołab-Schinzel Equation on the Nonnegative Reals and on Related Domains

By

J. Aczél and J. Schwaiger

(Vorgelegt in der Sitzung der math.-nat. Klasse am
16. Dezember 1999 durch das w. M. Ludwig Reich)

Abstract

We determine all continuous solutions $f : \mathbb{R} \rightarrow \mathbb{R}$ of the functional equation

$$f[x + yf(x)] = f(x)f(y)$$

on restricted domains like $\{x \geq 0, y \geq 0\}$ or $\{x > 0, y > 0\}$.

0

The Gołab-Schinzel equation

$$f[x + yf(x)] = f(x)f(y), \quad (1)$$

originating in the theory of continuous groups [3, 4, 6], has a rich literature (see e.g. [1–10]). In most of these works particular affine semigroups on \mathbb{R} are considered, particular in the sense that elements of the form $(0, \beta)$ are permitted. In other words, we have pairs (α, β) ($\alpha, \beta \in \mathbb{R}$) with the composition

$$(\alpha, \beta) * (\alpha', \beta') = (\alpha\alpha', \alpha\beta' + \beta) \quad (\alpha, \alpha', \beta, \beta' \in \mathbb{R})$$

The aim was to determine all one-parameter subsemigroups of this structure. Thus $\alpha = \alpha(U)$, $\alpha' = \alpha(V)$, $\beta = \beta(U)$, $\beta' = \beta(V)$. Under the supposition that β is injective this leads with $x = \beta(U) \in \mathbb{R}$, $y = \beta(V) \in \mathbb{R}$, $f(t) = \alpha[\beta^{-1}(T)]$ to the functional equation (1). For its continuous solutions on $x \in \mathbb{R}$, $y \in \mathbb{R}$ see [9, 2, 3]. If one considers only semigroups of pairs (α, β) ($\alpha, \beta \in \mathbb{R}_+ = [0, \infty)$) with the same composition

$$(\alpha, \beta) * (\alpha', \beta') = (\alpha\alpha', \alpha\beta' + \beta) \quad (\alpha, \alpha', \beta, \beta' \in \mathbb{R}_+)$$

then we get (1) restricted to $x \geq 0$, $y \geq 0$. Also in an application to meteorology (oral communication of Peter Kahlig, Vienna) the validity of (1) can be justified only for nonnegative x, y . We determine the continuous solutions of (1) on this and related domains.

While it would be justified to assume $f(t) \geq 0$ (see Corollary), in our main result we do not make this assumption. If negative values $f(t) < 0$ are permitted then, see (1), $f(t)$ has to be defined also for negative t , even though we suppose only that (1) be satisfied for $x \geq 0$, $y \geq 0$. For one family of solutions Eq. (1), required only for nonnegative x, y , determines $f(t)$ also for negative t . In case of the remaining three classes of solutions the equation

$$f[x + yf(x)] = f(x)f(y) \quad (x \geq 0, y \geq 0) \quad (2)$$

says nothing about the values of f at negative places.

1

Notice that, if $y = 0$ is in the domain of (1), then we have

$$f(x) = f(x)f(0) \text{ i.e. } f(0) = 1 \text{ or } f(x) \equiv 0. \quad (3)$$

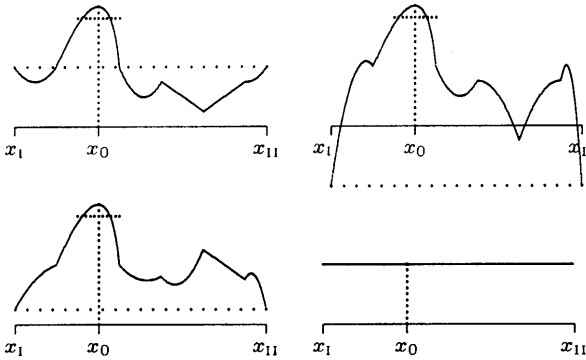
We will need a result about continuous solutions of (2) which assume two (or more) equal nonzero values at different (nonnegative) places.

Lemma 1. *If a continuous solution of (2) assumes the same nonzero value at more than one (nonnegative) place then it is constant on $\mathbb{R}_+ := [0, \infty)$.*

Proof: (The proof is more difficult than for $(x, y) \in \mathbb{R}^2$.) If there exist $x_I < x_{II}$ in \mathbb{R}_+ such that $f(x_I) = f(x_{II})$ then f has a maximum or minimum $f(x_0)$ on the open interval $]x_I, x_{II}[$.

Case 1: $f(x_0) > 0$. Then there exist x_1, x_2 ($x_1 < x_0 < x_2$) as close as we want such that $f(x_1) = f(x_2) > 0$. Put into (2) $y = (t - x_1)/f(x_1)$ ($t \geq x_1$) in order to obtain

$$\begin{aligned}
 f[t + (x_2 - x_1)] &= f\left[x_2 + \frac{t - x_1}{f(x_1)}f(x_1)\right] = f\left[x_2 + \frac{t - x_1}{f(x_1)}f(x_2)\right] = \\
 &= f(x_2)f\left(\frac{t - x_1}{f(x_1)}\right) = f(x_1)f\left(\frac{t - x_1}{f(x_1)}\right) = f\left[x_1 + \frac{t - x_1}{f(x_1)}f(x_1)\right] = f(t).
 \end{aligned}
 \tag{4}$$



Thus f is periodic on $[x_1, \infty[$, in particular on $[x_0, \infty[$, with arbitrarily small periods, thus, being continuous, is *constant*: $f(x) = b > 0$ ($x \geq x_0$). Now choose in (2) $x \geq x_0, 0 \leq y \leq x_0$ in order to get, since $x + yb \geq x \geq x_0$,

$$b = f(x + yb) = bf(y),$$

thus $f(y) = 1$ for $y \leq x_0$. But $f(x) = b$ for $x \geq x_0$ and f is continuous, so

$$f(x) = 1 \text{ (constant) on } \mathbb{R}_+.$$

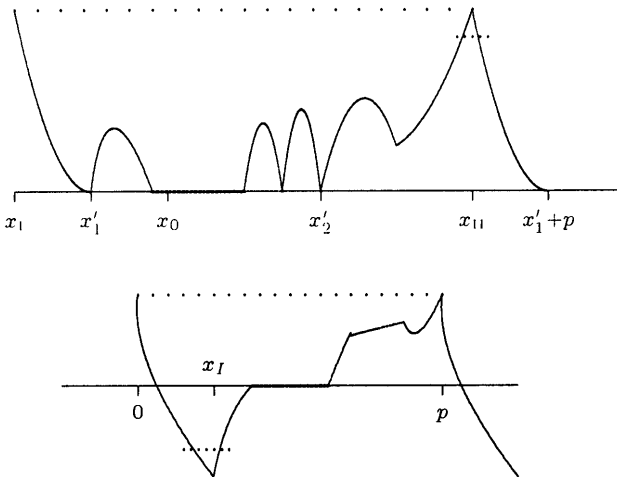
Case 2: $f(x_0) < 0$. Then there exist x_1, x_2 ($0 < x_1 < x_0 < x_2$), as close as we want, with $f(x_1) = f(x_2) < 0$. We set in (2) $y = (t - x_1)/f(x_1)$, now with $t \leq x_1$. Thus the continuous f is periodic with arbitrarily small periods, and so constant on $[0, x_1]$, in particular on a right neighbourhood $[0, \varepsilon[$ of 0, that is, $f(x) = b < 0$ for $x \in [0, \varepsilon]$. But, by (3), either $f(0) = 1 > 0$ or $f(x) \equiv 0$. Both *contradict* $f(x) = b < 0$ ($x \in [0, \varepsilon]$).

Case 3: $f(x_0) = 0$ but there is no proper neighbourhood of x_0 on which $f = 0$. In this case too, there exist $x_1 < x_0 < x_2$ such that $f(x_1) = f(x_2) \underset{<}{>} 0$. According to whether $>$ or $<$ applies, one proceeds as in case 1 or 2, respectively.

Case 4: The maximal or minimal value of f on $]x_I, x_{II}[$ is 0 and the value remains 0 on a proper subinterval of $]x_I, x_{II}[$. Let x'_I be the smallest

zero of f on the right of x_I and x'_2 the greatest zero of f left from x_{II} . Since $f(x_I) = f(x_{II}) \neq 0$, one obtains, as in (4), $f(t + p) = f(t)$ ($p = x_{II} - x_I$) for all $t \geq x_I$ if $f(x_I) > 0$ and for all $t \in [0, x_I]$ if $f(x_I) < 0$.

Subcase 4.1: If $f(x_I) > 0$ then we make use of $f(x'_1 + p) = f(x'_1) = f(x'_2) = 0$ and of the fact that $f \neq 0$ on $]x'_2, x'_1 + p[$ (by the definition of x'_1 and x'_2). So, since $x'_2 < x_{II} < x'_1 + p$, there exists on $]x'_2, x'_1 + p[$ a *positive* maximal value of f and we obtain $f(x) \equiv 1$ (constant) as in case 1.



Subcase 4.2: If $f(x_I) < 0$ we have to be careful because $x'_2 - p$ may be negative. However now $f(t + p) = f(t)$ for all $t \in [0, x_I]$, in particular, $f(p) = f(0) = 1$ (by (3); $f(x) \neq 0$ since $f(x_I) < 0$). So we have points where f has the same *positive* value and we can go on as in cases 1, 3 or in subcase 4.1.

We exhausted all possible cases (and ourselves), so Lemma 1 is proved. □

2

One sees directly from (2) that $f(x) \equiv 0$ and $f(x) \equiv 1$ are the only constant solutions. As to nonconstant solutions, we have the following.

Lemma 2. *If f is a nonconstant solution of equation (2) then $(f(x) - 1)/x$ is constant for all $x > 0$ with $f(x) \neq 0$.*

Proof: By contradiction (as in [9, 2]): If there were $x > 0, y > 0$ with $x \neq y, f(x)f(y) \neq 0$ and with

$$\frac{f(x) - 1}{x} \neq \frac{f(y) - 1}{y}$$

then we would have

$$x + yf(x) \neq y + xf(y). \tag{5}$$

We also have, however,

$$f[x + yf(x)] = f(x)f(y) = f[y + xf(y)] \neq 0. \tag{6}$$

In view of (5) and (6), by Lemma 1, the function f would be constant. This contradiction proves Lemma 2. \square

3

We denote the constant in Lemma 2 by c . Thus $(f(x) - 1)/x = c$, i.e., $f(x) = cx + 1$ if $x > 0$ and $f(x) \neq 0$. We saw in (3) that, if $f(x) \neq 0$, then $f(0) = 1$. So, for every $x \geq 0$, either $f(x) = cx + 1$ or $f(x) = 0$ hold. This describes exactly the following continuous functions:

$$f(x) = 0 \quad (x \geq 0), \tag{7}$$

$$f(x) = \begin{cases} cx + 1 & \text{for } 0 \leq x \leq -1/c, \\ 0 & \text{for } x \geq -1/c, \end{cases} \tag{8}$$

where c is a *negative* constant and

$$f(x) \equiv cx + 1 \quad (x \geq 0), \tag{9}$$

where c is an *arbitrary real* constant.

The solutions (7), (8) and the solution (9) for $c \geq 0$ are *nonnegative* valued, so $x + yf(x) \geq 0$ if $x \geq 0, y \geq 0$. Thus, to these solutions, equation (2) does not offer any $f(t)$ values for $t < 0$. However, the solution (9) with $c < 0$ yields negative values for $x > -1/c$. Thus $x + yf(x) < 0$ for large enough y and this determines $f(t)$ for all $t < 0$. Indeed substitute (9) (with $c < 0$) into equation (2):

$$f[x + y(cx + 1)] = (cx + 1)(cy + 1)$$

or, with $x = x_0 > -1/c$ and $t = x_0 + y(cx_0 + 1)$, that is,

$$y = (t - x_0)/(cx_0 + 1) :$$

$$f(t) = (cx_0 + 1)\left(c \frac{t - x_0}{cx_0 + 1} + 1\right) = c(t - x_0) + cx_0 + 1 = ct + 1.$$

So $f(t) = ct + 1$ also for $t < 0$ and we have following.

Theorem. *All continuous solutions $f : \mathbb{R} \rightarrow \mathbb{R}$ of equation (2) are given by (7), (8) and by (9) with $c \geq 0$, in which case $f(t)$ can be chosen in a continuous but otherwise arbitrary way for $t < 0$ and, if $c < 0$, by*

$$f(t) = ct + 1 \quad (t \in \mathbb{R}). \quad (10)$$

4. Remarks

1. A similar theorem can be proved if (1) is supposed only for $x > 0$, $y > 0$. Indeed, any solution of (1) for $x > 0$, $y > 0$, except $f(x) \equiv 0$, satisfying (1) for $x > 0$, $y > 0$, satisfies it also for $x = 0$ if $f(0) = 1$ is added to the definition. The new functions are also continuous if (8) or (9) has been so extended. The solution (7) has been an exception also in (3), and should be extended by $f(0) = 0$.

2. If $f \geq 0$ is supposed, then Case 2, Subcase 4.2 and the subcase $f(x_1) = f(x_2) < 0$ of Case 3 can be disposed of. The remaining solutions are indeed nonnegative and for them equation (2) says nothing about values at negative places (in particular (10) does not apply). We have the following.

Corollary. *All continuous nonnegative solutions $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ of equation (2) are given by (7), (8) and by (9) with $c \geq 0$. The same, with $x > 0$, are the general nonnegative continuous solutions of (1) on $\{x > 0, y > 0\}$.*

3. The solution of equation (1) can be obtained similarly for x, y bounded from above.

Acknowledgements

The research of the first author has been supported in part by the Natural Sciences and Engineering Research Council of Canada grant OGP 0002972 and by the Austrian Academy of Sciences. He is grateful for the hospitality of the Institut für Mathematik, Karl-Franzens-Universität Graz.

References

- [1] Aczél, J.: Beiträge zur Theorie der geometrischen Objekte. Acta Math. Acad. Sci. Hung. **8**, 19–52 (1957).
- [2] Aczél, J.: Lectures on functional equations and their applications, pp. 132–135. Academic Press, New York–London, 1966.
- [3] Aczél, J., Dhombres J.: Functional equations in several variables, pp. 311–318. Cambridge Univ. Press, Cambridge–New York, 1989.
- [4] Aczél J., Gołąb, St.: Remarks on one-parameter subsemigroups of the affine group and their homo- and isomorphisms. Aeq. Math. **4**, 1–10 (1970).

- [5] Baron, K.: On the continuous solutions of the Gołab-Schinzel equation. *Aeq. Math.* **38**, 155–162 (1989).
- [6] Brillouët, N., Dhombres, J.: Équations fonctionnelles et recherche de sous-groupes. *Aeq. Math.* **31**, 253–293 (1986).
- [7] Brzdęk, J.: Subgroups of the group Z_n and a generalization of the Gołab-Schinzel equation. *Aeq. Math.* **43**, 59–71 (1992).
- [8] Brzdęk, J.: On some conditional functional equations of Gołab-Schinzel type. *Ann. Math. Siles.* **9**, 65–80 (1995).
- [9] Gołab, St., Schinzel, A.: Sur l'équation fonctionnelle $f[x + yf(x)] = f(x)f(y)$. *Publ. Math. Debrecen* **6**, 113–125 (1960).
- [10] Wołodźko, S.: Solution générale de l'équation fonctionnelle $f[x + yf(x)] = f(x)f(y)$. *Aeq. Math.* **2**, 12–29 (1968).

Authors' addresses: Dr. J. Aczél, Department of Pure Mathematics, University of Waterloo, Waterloo, ON N2L 3G1 Canada; Dr. J. Schwaiger, Institut für Mathematik der Universität Graz, Heinrichstr. 39, A-8010 Graz, Austria.