Noname manuscript No.
(will be inserted by the editor)

Addendum to "On recurrences converging to the wrong limit in finite precision and some new examples"

Siegfried M. Rump

Received: date / Accepted: date

Abstract In a recent paper we analyzed Muller's famous recurrence where, for particular initial values, the true real iteration converges to a repellent fixed point, whereas finite precision arithmetic produces a different result, the attracting fixed point. We gave necessary and sufficient conditions for such recurrences to produce only nonzero iterates.

In the above-mentioned paper an example was given where only finitely many terms of the recurrence over $\mathbb{R}$ are well defined, but floating-point evaluation suggests convergence to the attracting fixed point. The input data of that example, however, is not representable in binary floating-point, and the question was posed whether such examples exist with binary representable data. This note answers that question in the affirmative.

Keywords Recurrences · rounding errors · IEEE-754 · exactly representable data · bfloat · half precision (binary16) · single precision (binary32) · double precision (binary64)
Mathematics Subject Classification 65G50 - 11B37

1 Main result

In 1989 Muller [3] presented the recurrence

$$x_0 := 11/2,\ x_1 := 61/11 \quad \text{and} \quad x_{n+1} := 111 - (1130 - 3000/x_{n-1})/x_n. \quad (1)$$

S.M. Rump
Institute for Reliable Computing,
Hamburg University of Technology,
Am Schwarzenberg-Campus 3, 21073 Hamburg, Germany,
and Visiting Professor at Waseda University,
Faculty of Science and Engineering,
3–4–1 Okubo, Shinjuku-ku, Tokyo 169–8555, Japan
E-mail: rump@tuhh.de

The limit of the recurrence over the field of real numbers is 6, whereas in double precision it converges to 100. Subsequently, similar examples were given by Kahan [2] together with some analysis, and again by Muller [4].

In [5] those recurrences were analyzed giving a necessary and sufficient criterion for such a sequence being well defined, i.e., no zero iterate is encountered. More precisely, let

$$x_{n+1} := a + (b + c/x_{n-1})/x_n \quad \text{with} \quad a, b, c \in \mathbb{R} \tag{2}$$

for given initial values $(x_0, x_1) \in \mathbb{R}^2$. Setting $y_{n+1} := x_n y_n$ for $0 \leqslant n \in \mathbb{N}$ and $y_0 := 1$ defines the characteristic polynomial

$$\chi(y) = y^3 - ay^2 - by - c =: (y - \alpha)(y - \beta)(y - \gamma) \tag{3}$$

as in [5, (2.3)]. We restrict our attention to recurrences satisfying

$$|\alpha| > |\beta| > |\gamma| > 0 \quad \text{and} \quad \alpha, \beta, \gamma \in \mathbb{R}. \tag{4}$$

Lemma 1 *[5, Lemma 2.1] Let $x_0, x_1 \in \mathbb{R}$ be given, and let the recurrence* (2) *with characteristic polynomial* (3) *satisfy* (4)*. Then* (2) *is well-defined and $x_i \to \beta$ if, and only if*

$$x_0 \neq \gamma \quad \textit{and} \tag{5}$$

$$x_1 = \beta + \gamma - \beta\gamma/x_0 \quad \textit{and} \tag{6}$$

$$x_0 \neq \gamma - \frac{\gamma^n(\beta - \gamma)}{\beta^n - \gamma^n} \quad \textit{for all } n \geqslant 1. \tag{7}$$

By the lemma the recurrence (x_i) is well-defined and converges to β for (x_0, x_1) on the hyperbola H defined by $x_1 = \beta + \gamma - \beta\gamma/x_0$ except infinitely many discrete points. Moreover, it was shown in [5] that in every ε-neighborhood of initial values (x_0, x_1) with well-defined recurrence converging to β there exists a pair of initial values with not well-defined recurrence.

In [5] we presented the recurrence

$$x_0 := \frac{109225}{43691},\ x_1 := \frac{10923}{4369} \quad \text{and} \quad x_{n+1} := 56.5 + (160 - \frac{737.5}{x_{n-1}})/x_n. \tag{8}$$

Over $\mathbb{R}$ it produces $x_{16} = 0$, but when evaluated in half, single or double precision the floating-point iteration is well-defined and becomes stationary at the attracting fixed point $\alpha = 59$, see [5, Table 2.1].

The input data x_0 and x_1 are not representable in binary in any precision, and it was asked in [5, p. 364] whether there are similar examples with all data representable in some binary format. To answer that in the affirmative we use the following lemma.

Lemma 2 *For given $a, b, c \in \mathbb{C}$, $c \neq 0$, let β and γ be any roots of $x^3 - ax^2 - bx - c = 0$. Let $n \in \mathbb{N}$ with $n \geqslant 3$ be given and assume $\beta^j \neq \gamma^j$ for $j \in \{1, \ldots, n\}$. Then*

$$x_0 = \gamma - \frac{\gamma^n(\beta - \gamma)}{\beta^n - \gamma^n}, \quad x_1 := \beta + \gamma - \beta\gamma/x_0$$

and $x_{k+1} := a + (b + c/x_{k-1})/x_k$ for $k \geqslant 1$ imply

$$x_k = \frac{\beta\gamma(\beta^{n-k-1} - \gamma^{n-k-1})}{\beta^{n-k} - \gamma^{n-k}} \qquad \text{for } 0 \leqslant k \leqslant n-1. \tag{9}$$

Remark Note that $\beta\gamma \neq 0$ because $c \neq 0$, and that (9) implies $x_0 x_1 \neq 0$ and $x_{n-1} = 0$.

Proof A computation shows that (9) is true for $k = 0$, and similarly the assumption $x_1 = \beta + \gamma - \beta\gamma/x_0$ implies (9) for $k = 1$. Abbreviate $\delta_j := \beta^j - \gamma^j$ and note that $\delta_j \neq 0$ for $j \in \{1, \ldots, n\}$. We have to prove $x_k = \frac{\beta\gamma\delta_{n-k-1}}{\delta_{n-k}}$. The definition of the recurrence implies

$$\begin{aligned} x_{k+1} &= a + \left(b + \frac{c\delta_{n-k+1}}{\beta\gamma\delta_{n-k}}\right) \frac{\delta_{n-k}}{\beta\gamma\delta_{n-k-1}} \\ &= \frac{a\beta^2\gamma^2\delta_{n-k-1} + b\beta\gamma\delta_{n-k} + c\delta_{n-k+1}}{\beta^2\gamma^2\delta_{n-k-1}} \\ &= \frac{\beta^{n-k+1}(a\gamma^2 + b\gamma + c) - \gamma^{n-k+1}(a\beta^2 + b\beta + c)}{\beta^2\gamma^2\delta_{n-k-1}} \\ &= \frac{\beta^{n-k+1}\gamma^3 - \gamma^{n-k+1}\beta^3}{\beta^2\gamma^2\delta_{n-k-1}} = \frac{\beta\gamma\delta_{n-k-2}}{\delta_{n-k-1}} \end{aligned}$$

and proves the result. □

Let $x_{n+1} = a + (b + c/x_{n-1})/x_n$ for given $a, b, c, x_0, x_1 \in \mathbb{R}$. Then, for $\varphi \in \mathbb{R}$,

$$X_{n+1} := A + (B + C/X_{n-1})/X_n$$

with

$$A := \varphi a, \; B := \varphi^2 b, \; C := \varphi^3 c, \; X_0 := \varphi x_0, \; X_1 := \varphi x_1 \tag{10}$$

satisfies $X_k = \varphi x_k$ for $k \geqslant 0$. Hence, a recurrence with rational a, b, c, x_0, x_1 can be transformed into a similar one with integer quantities. Using Lemma 2 a desired example with integer data may be constructed as follows:

- Choose some integer $n \geqslant 2$.
- Choose $p, q \in \mathbb{Q}$, $q \neq 0$, and denote the roots of $x^2 + px + q$ by β, γ.
- Make sure that $\beta^j \neq \gamma^j$ for $j \in \{1, \ldots, n\}$.
- Choose $\alpha \in \mathbb{Q}$ with $|\alpha| > \max(|\beta|, |\gamma|)$.
- Let $x^3 - ax^2 - bx - c = (x - \alpha)(x^2 + px + q)$.
- Define $x_{n-1} := 0$ and $x_{n-2} := \frac{\beta\gamma}{\beta+\gamma} = -q/p$.

– Compute x_0, x_1 recursively by $x_{k-1} = c(x_k x_{k+1} - ax_k - b)^{-1}$.

Obviously all data are rational, and using (10) we may produce integer data. By construction, the recurrence (2) with initial values x_0, x_1 produces $x_{n-1} = 0$ over $\mathbb{R}$. If in some finite precision one of the x_k for $2 \leqslant k \leqslant n-2$ is not representable, likely the floating-point approximation of x_{n-1} will be nonzero and the recurrence will converge to the attracting fixed point α.

Lemma 3 *For given $a, b, c \in \mathbb{R}$ assume the roots α, β, γ of $x^3 - ax^2 - bx - c = 0$ satisfy $|\alpha| > |\beta| > |\gamma| > 0$. For given $x_0 \in \mathbb{R}, x_0 \neq \gamma$ let $x_1 := \beta + \gamma - \beta\gamma/x_0$ and assume $x_0 x_1 \neq 0$. Finally assume*

$$x_0 = \gamma - \frac{\gamma^n(\beta - \gamma)}{\beta^n - \gamma^n}$$

for some integer $n \geqslant 2$. Then in every ε-neighborhood of (x_0, x_1) there exist (x_0', x_1') and (x_0'', x_1'') for which the recurrence $x_{k+1} := a + (b + c/x_{k-1})/x_k$ is well defined for all k, such that for initial values (x_0', x_1') it converges to the repelling fixed point β, whereas for initial values (x_0'', x_1'') it converges to the attracting fixed point α.

Proof By [5, Lemma 2.1], for each pair of initial values (x_0, x_1) on the hyperbola $x_1 := \beta + \gamma - \beta\gamma/x_0$ the recurrence converges to the repelling fixed point β provided it is well defined, i.e., $x_0 \neq \gamma - \dfrac{\gamma^n(\beta - \gamma)}{\beta^n - \gamma^n}$ for all $n \in \mathbb{N}$. Thus, the set of exceptional pairs (x_0, x_1) for which the recurrence is not well defined is countable, implying existence of initial values (x_0', x_1') with the desired property. The existence of a pair (x_0'', x_1'') follows by [5, Corollary 2.4]. □

Based on the previous considerations it is not difficult to construct examples with the desired property, for instance

$$x_{n+1} := 6496 - (4205 \cdot 2^{10} + 609725 \cdot 2^{15}/x_{n-1})/x_n \text{ for } x_0 := -1305,\ x_1 := -1440.$$

The roots of the characteristic polynomial are

$$\alpha = 4640 \quad \text{and} \quad \beta, \gamma = 928 \pm 928\sqrt{6} \approx [-1345.13,\ 3201.13].$$

The data x_0, x_1, a, b, c are exactly representable in 20-bit binary format. The left two columns of Table 1 show the result in IEEE-754 [1] single (binary32) and double (binary64) precision. As can be seen, both in single and double precision the recurrence is defined and converges to the attracting fixed point $\alpha = 4640$. However, at the 8-th iterate it becomes visible that something happened during the iteration. The second example was constructed by Paul Zimmermann [7] from INRIA using Sage [6]:

$$x_{n+1} := -256 + (131072/x_{n-1})/x_n \quad \text{for} \quad x_0 := 3,\ x_1 := 170.$$

The roots of the characteristic polynomial are approximately $-253.97, -23.76$ and 21.72, and the data x_0, x_1, a, b, c are representable in 7 bits. The results of

Table 1 Results for $x_{n+1} := 6496 - (4205 \cdot 2^{10} + 609725 \cdot 2^{15}/x_{n-1})/x_n$ with initial values $x_0 := -1305,\ x_1 := -1440$.

n	single	double	over $\mathbb{R}$
0	-1305.0000000000000000	-1305.0000000000000000	-1305
1	-1440.0000000000000000	-1440.0000000000000000	-1440
2	-1145.6791992187500000	-1145.6790123456794390	-92800/81
3	-1855.9990234375000000	-1855.9999999999981810	-1856
4	-580.0024414062500000	-580.0000000000027285	-580
5	-4639.9638671875000000	-4639.9999999999672582	-4640
6	-0.0195312500000000	-0.0000000000109139	0
7	4780.7998046875000000	3680.0000000000000000	
8	213975808.0000000000000000	497456029492482816.00000000	
9	6495.9604492187500000	6495.9999999999799911	
10	5833.1245117187500000	5833.1428571428486975	
...	...	...	
46	4640.0009765625000000	4640.0009773540996321	
47	4640.0004882812500000	4640.0006742744462827	
48	4640.0004882812500000	4640.0004651804893001	
49	4640.0000000000000000	4640.0003209270334992	
50	4640.0000000000000000	4640.0002214068863395	
...	...	...	
102	4640.0000000000000000	4640.0000000000009095	
103	4640.0000000000000000	4640.0000000000000000	
104	4640.0000000000000000	4640.0000000000000000	

the floating-point iteration in bfloat (8 bits), half (11 bits), single and double is displayed in the left four columns of Table 2.

In all used floating-point formats the recurrence converges to the floating-point number nearest to the attracting fixed point α. In bfloat, half and single precision the floating-point iteration camouflages the true behavior of the recurrence, yet another example of the smoothing effect of rounded operations.

2 Acknowledgment

Many thanks to Paul Zimmermann for fruitful discussions, really many nice examples, very careful reading resulting in several suggestions to improve the note. Also many thanks to Florian Bünger, who read the manuscript very carefully and gave several valuable comments.

References

1. IEEE standard for floating-point arithmetic. *IEEE Std 754-2019 (Revision of IEEE 754-2008)*, pages 1–84, 2019.
2. W. Kahan, How futile are mindless assessments of roundoff in floating-point computations, `https://people.eecs.berkeley.edu/~wkahan/Mindless.pdf`, 2006.
3. J.-M. Muller, Arithmétique des ordinateurs, `https://hal-ens-lyon.archives-ouvertes.fr/ensl-00086707`, 1989.
4. J.M. Muller, N. Brunie, F. de Dinechin, C.-P. Jeannerod, M. Joldes, V. Lefevre, G. Melquiond, R. Revol, and S. Torres. Handbook of Floating-Point Arithmetic. Birkhäuser, Boston, 2nd edition, 2018.

Table 2 Results for $x_{n+1} := -256 + (131072/x_{n-1})/x_n$ with $x_0 := 3,\ x_1 := 170$.

n	bfloat	half	single	double	over $\mathbb{R}$
0	3.00	3.000	3.00000000000	3.0000000000000000	3
1	170.00	170.000	170.00000000000	170.0000000000000000	170
2	2.00	1.000	1.00393676758	1.0039215686274474	-256/255
3	130.00	515.000	511.98840332031	512.0000000000027285	512
4	248.00	-1.500	-0.99807739258	-1.0000000000004547	-1
5	-252.00	-425.500	-512.49890136719	-511.9999999998822204	-512
6	-258.00	-50.750	0.24343872070	-0.0000000000575255	0
7	-254.00	-249.875	-1306.57568359375	$4.45019 \cdot 10^{12}$	
8	-254.00	-245.625	-668.08398437500	-768.0000000293351832	
9	-254.00	-253.875	-255.84983825684	-256.0000000000383693	
10	-254.00	-253.875	-255.23318481455	-255.3333333333588939	
11	-254.00	-254.000	-253.99281311035	-253.9947780678856191	
12	-254.00	-254.000	-253.97813415527	-253.9789473684212453	
13	-254.00	-254.000	-253.96815490723	-253.9681697612732023	
14	-254.00	-254.000	-253.96795654297	-253.9679568859273502	
15	-254.00	-254.000	-253.96786499023	-253.9678689491082935	
16	-254.00	-254.000	-253.96786499023	-253.9678665421512846	
...	...	...	...	...	
23	-254.00	-254.000	-253.96786499023	-253.9678657879329933	
24	-254.00	-254.000	-253.96786499023	-253.9678657879329648	
25	-254.00	-254.000	-253.96786499023	-253.9678657879329648	

5. S.M. Rump. On recurrences converging to the wrong limit in finite precision and some new examples. *Electronic Transactions on Numerical Analysis (ETNA)*, 52:358–369, 2020.
6. Paul Zimmermann, Alexandre Casamayou, Nathann Cohen, Guillaume Connan, Thierry Dumont, Laurent Fousse, Francois Maltey, Matthias Meulien, Marc Mezzarobba, Clement Pernet, Nicolas M. Thiery, Erik Bray, John Cremona, Marcelo Forets, Alexandru Ghitza, and Hugh Thomas. *Computational Mathematics with SageMath*. SIAM, December 2018.
7. P. Zimmermann. Private communication, 2020.