Arithmetic operations for floating-point intervals

Denotations

$\mathbb{R}$: the set of real numbers,
$\mathbb{F} \subseteq \mathbb{R}$: the set of floating-point numbers of a given format and encoding,
$\mathbb{IR}$: the set of nonempty, closed, and bounded real intervals,
$\mathbb{IF}$: the subset of $\mathbb{IR}$ with floating-point bounds,
$(\mathbb{IR})$: the set of closed bounded and unbounded real intervals, $\varnothing \in (\mathbb{IR})$.
$(\mathbb{IF})$: the subset of $(\mathbb{IR})$ where all finite bounds are elements of $\mathbb{F}$,
$\overset{\nabla}{+}$, $\overset{\nabla}{-}$, $\overset{\nabla}{*}$, $\overset{\nabla}{/}$: the operations for elements of $\mathbb{F}$ with rounding downwards,
$\overset{\triangle}{+}$, $\overset{\triangle}{-}$, $\overset{\triangle}{*}$, $\overset{\triangle}{/}$: the operations for elements of $\mathbb{F}$ with rounding upwards,
$+, -, *, /$: the operations for elements of $(\mathbb{IF})$,
$\boldsymbol{a} = [a_1, a_2]$, $\boldsymbol{b} = [b_1, b_2] \in \mathbb{IF}$.

For intervals of $\mathbb{IF}$ the arithmetic operations addition, subtraction, multiplication, and division are well established:

Addition $\quad [a_1, a_2] + [b_1, b_2] = [a_1 \overset{\nabla}{+} b_1, a_2 \overset{\triangle}{+} b_2]$.

Subtraction $\quad [a_1, a_2] - [b_1, b_2] = [a_1 \overset{\nabla}{-} b_2, a_2 \overset{\triangle}{-} b_1]$.

Multiplication $[a_1, a_2] * [b_1, b_2]$	$[b_1, b_2]$ $b_2 \le 0$	$[b_1, b_2]$ $b_1 < 0 < b_2$	$[b_1, b_2]$ $b_1 \ge 0$
$[a_1, a_2], a_2 \le 0$	$[a_2 \overset{\nabla}{*} b_2, a_1 \overset{\triangle}{*} b_1]$	$[a_1 \overset{\nabla}{*} b_2, a_1 \overset{\triangle}{*} b_1]$	$[a_1 \overset{\nabla}{*} b_2, a_2 \overset{\triangle}{*} b_1]$
$a_1 < 0 < a_2$	$[a_2 \overset{\nabla}{*} b_1, a_1 \overset{\triangle}{*} b_1]$	$[min(a_1 \overset{\nabla}{*} b_2, a_2 \overset{\nabla}{*} b_1),$ $max(a_1 \overset{\triangle}{*} b_1, a_2 \overset{\triangle}{*} b_2)]$	$[a_1 \overset{\nabla}{*} b_2, a_2 \overset{\triangle}{*} b_2]$
$[a_1, a_2], a_1 \ge 0$	$[a_2 \overset{\nabla}{*} b_1, a_1 \overset{\triangle}{*} b_2]$	$[a_2 \overset{\nabla}{*} b_1, a_2 \overset{\triangle}{*} b_2]$	$[a_1 \overset{\nabla}{*} b_1, a_2 \overset{\triangle}{*} b_2]$

Division, $0 \notin \boldsymbol{b}$ $[a_1, a_2]/[b_1, b_2]$	$[b_1, b_2]$ $b_2 < 0$	$[b_1, b_2]$ $b_1 > 0$
$[a_1, a_2], a_2 \le 0$	$[a_2 \overset{\nabla}{/} b_1, a_1 \overset{\triangle}{/} b_2]$	$[a_1 \overset{\nabla}{/} b_1, a_2 \overset{\triangle}{/} b_2]$
$[a_1, a_2], a_1 < 0 < a_2$	$[a_2 \overset{\nabla}{/} b_2, a_1 \overset{\triangle}{/} b_2]$	$[a_1 \overset{\nabla}{/} b_1, a_2 \overset{\triangle}{/} b_1]$
$[a_1, a_2], a_1 \ge 0$	$[a_2 \overset{\nabla}{/} b_2, a_1 \overset{\triangle}{/} b_1]$	$[a_1 \overset{\nabla}{/} b_2, a_2 \overset{\triangle}{/} b_1]$

In real analysis division by zero is not defined. In interval arithmetic, however, the interval in the denominator of a quotient may contain zero. So this case has to be considered also.

In interval arithmetic the result of an operation is a set. Thus, the result of division by the interval $\boldsymbol{b} = [0, 0]$ can only be the empty set $\varnothing$. This means, the element 0 in the denominator of an interval division does not contribute to the solution set. So it can be excluded without changing the solution set.

So the general rule for computing the set $\boldsymbol{a}/\boldsymbol{b}$ with $0 \in \boldsymbol{b}$ is to remove its zero from the interval $\boldsymbol{b}$ and perform the division with the remaining set. Whenever the zero in

$\boldsymbol{b}$ coincides with a bound of the interval $\boldsymbol{b}$ the result of the division can directly be obtained from the above table for division with $0 \notin \boldsymbol{b}$ by the limit process $b_1 \to 0$ or $b_2 \to 0$ respectively. The results are shown in the following table.

Division, $0 \in \boldsymbol{b}$ $[a_1, a_2]/[b_1, b_2]$	$\boldsymbol{b} =$ $[0, 0]$	$[b_1, b_2]$ $b_1 < b_2 = 0$	$[b_1, b_2]$ $0 = b_1 < b_2$
$[a_1, a_2] = [0, 0]$	$\varnothing$	$[0, 0]$	$[0, 0]$
$a_1 < a_2 \leq 0$	$\varnothing$	$[a_2 \triangledown b_1, +\infty)$	$(-\infty, a_2 \vartriangle b_2]$
$[a_1, a_2], a_1 < 0 < a_2$	$\varnothing$	$(-\infty, +\infty)$	$(-\infty, +\infty)$
$0 \leq a_1 < a_2$	$\varnothing$	$(-\infty, a_1 \vartriangle b_1]$	$[a_1 \triangledown b_2, +\infty)$

Here, the parentheses indicate that the bounds $-\infty$ and $+\infty$ are not elements of the interval.

Whenever zero is an interior point of the denominator the following consideration leads to the correct answer.

A basic concept of mathematics is that of a function or mapping. A function consists of a pair (f, D_f). It maps each element x of its domain of definition D_f on a single element y of the range R_f of f, $f : D_f \to R_f$.

A rational function $y = f(x)$ where the denominator is zero for $x = c$ is not defined for $x = c$; i.e., c is not an element of the domain of definition D_f. Since the function $f(x)$ is not defined at $x = c$ it does not have any value or property there. In this strict mathematical sense, division by an interval $[b_1, b_2]$ with $b_1 < 0 < b_2$ is not well posed. The interval $[b_1, b_2]$ overflows the range of definition of the function $f(x)$. For division the set $b_1 < 0 < b_2$ devolves into the two distinct sets $[b_1, 0)$[1] and $(0, b_2]$ and division by the set $b_1 < 0 < b_2$ actually means two divisions. The results of the two divisions are already shown in the table for divison with $0 \in \boldsymbol{b}$. It is highly desirable to perform the two divisions consecutively.

In the user's program, however, the two divisions appear as a single operation, as division by an interval $[b_1, b_2]$ with $b_1 < 0 < b_2$. So an arithmetic operation in the user's program delivers two distinct results. This is an unusual phenomenon in digital computing, but it can be handled.

A solution to the problem would be for the computer to provide a flag for *distinct intervals*. The situation occurs if the divisor is an interval that contains zero as an interior point. In this case the flag would be raised and signaled to the user. The user may then apply a routine of his choice to deal with the situation as is appropriate for his application.

This routine could be: return the entire set of real numbers $(-\infty, +\infty)$ as result and continue the computation, or continue the computation with one of the sets and ignore the other one, or put one of the sets on a list and continue the computation

[1]Since division by zero does not contribute to the solution set it does not matter whether a parenthesis or bracket is used here.

with the other one, or modify the operands and recompute, or stop computing, or some other action.[2]

Thus only four kinds of result come from division by an interval of $\mathbb{IF}$ that contains zero:

$$\varnothing, \quad (-\infty, a], \quad [b, +\infty), \quad \text{and} \quad (-\infty, +\infty).$$

We call such elements extended intervals.

The union of the set of closed and bounded intervals of $\mathbb{IR}$ with the set of extended real intervals is denoted by $(\mathbb{IR})$. $\varnothing \in (\mathbb{IR})$. $(\mathbb{IR})$ is the set of closed real intervals. (A subset of $\mathbb{R}$ is called closed if its complement is open). $(\mathbb{IF})$ is the subset of closed real intervals where all finite bounds are elements of $\mathbb{F}$. The elements of the set $(\mathbb{IF})$ are themselves simply called floating-point intervals.

Intervals of $\mathbb{IF}$ and of $(\mathbb{IF})$ are sets of real numbers. $-\infty$ and $+\infty$ are not elements of these intervals. Arithmetic operations for extended intervals of $(\mathbb{IF})$ are now to be defined.

The first rule is that any operation with the empty set $\varnothing$ has the empty set as its result.

Arithmetic operations for unbounded intervals of $(\mathbb{IF})$ can be performed on the computer by using the above formulas for bounded intervals of $\mathbb{IF}$ if in addition a few formal rules for operations with $-\infty$ and $+\infty$ are applied. These rules are shown in the following tables.

Addition	$-\infty$	b	$+\infty$
$-\infty$	$-\infty$	$-\infty$	
a	$-\infty$		$+\infty$
$+\infty$		$+\infty$	$+\infty$

Subtraction	$-\infty$	b	$+\infty$
$-\infty$		$-\infty$	$-\infty$
a	$+\infty$		$-\infty$
$+\infty$	$+\infty$	$+\infty$	

Multiplication	$-\infty$	$b < 0$	0	$b > 0$	$+\infty$
$-\infty$	$+\infty$	$+\infty$	0	$-\infty$	$-\infty$
$a < 0$	$+\infty$				$-\infty$
0	0				0
$a > 0$	$-\infty$				$+\infty$
$+\infty$	$-\infty$	$-\infty$	0	$+\infty$	$+\infty$

Division	$-\infty$	$+\infty$
a	0	0

These rules are not new in principle. They are well established in real analysis and **IEEE 754 provides them anyway.** The only rule that goes beyond IEEE 754 is

$$0 * (-\infty) = (-\infty) * 0 = 0 * (+\infty) = (+\infty) * 0 = 0.$$

[2]Newton's method reaches its ultimate elegance and strength in the extended interval Newton method. It computes all (single) zeros in a given domain. If a function has several zeros in a given interval its derivative becomes zero in that interval also. Thus Newton's method applied to that interval delivers two distinct sets. This is how the extended interval Newton method separates different zeros from each other. If the method is continued along two separate paths, one for each of the distinct intervals it finally computes all zeros in the given domain. If the method continues with only one of the two distinct sets and ignores the other one it computes an enclosure of only one zero of the given function. If the interval Newton method delivers the empty set, the method has proved that there is no zero in the initial interval.

This rule follows quite naturally from the definition of unbounded intervals. However, it should not be taken as a new mathematical law. It is just a short cut to easily compute the bounds of the result of an operation on unbounded intervals.

In interval arithmetic a real number or an interval over the real numbers is mapped onto the smallest floating-point interval that contains the number or interval respectively. This mapping $\lozenge : (\mathbb{IR}) \to (\mathbb{IF})$ is characterized by the following properties:

(R1): $\lozenge\, \boldsymbol{a} = \boldsymbol{a}$, for all $\boldsymbol{a} \in (\mathbb{IF})$,

(R2): $\boldsymbol{a} \subseteq \boldsymbol{b} \Rightarrow \lozenge\, \boldsymbol{a} \subseteq \lozenge\, \boldsymbol{b}$, for $\boldsymbol{a}, \boldsymbol{b} \in (\mathbb{IR})$,

(R3): $\boldsymbol{a} \subseteq \lozenge\, \boldsymbol{a}$, for all $\boldsymbol{a} \in (\mathbb{IR})$,

(R4): $\lozenge\, (-\boldsymbol{a}) = -\lozenge\, \boldsymbol{a}$, for all $\boldsymbol{a} \in (\mathbb{IR})$.

With the mapping $\lozenge : (\mathbb{IR}) \to (\mathbb{IF})$ and its properties the operations defined in this document have the following property which defines them uniquely:

(RG): $\boldsymbol{a} \mathbin{\diamond\!\!\!\circ}\, \boldsymbol{b} := \lozenge\, (\boldsymbol{a} \circ \boldsymbol{b})$, for all $\boldsymbol{a}, \boldsymbol{b} \in (\mathbb{IF})$ and all $\circ \in \{+, -, *, /\}$.

Here the operations $\boldsymbol{a} \circ \boldsymbol{b}$, are defined as set operations in $\mathbb{R}$, i.e.,

$\boldsymbol{a} \circ \boldsymbol{b} := \{a \circ b \mid a \in \boldsymbol{a} \wedge b \in \boldsymbol{b}\}$, for all $\boldsymbol{a}, \boldsymbol{b} \in (\mathbb{IF})$ and all $\circ \in \{+, -, *, /\}$.

The calculus in $(\mathbb{IF})$ as defined in this document is free of exceptions.

Rationale:

The Kulisch position paper *Complete Interval Arithmetic*,
The Vienna Proposal,
Kulisch, U.: *Computer Arithmetic and Validity—Theory, Implementation, and Applications*, de Gruyter 2008.