Exact real arithmetic for interval number systems

Petr Kůrka

Center for Theoretical Study
Academy of Sciences and Charles University in Prague
Jilská 1, CZ-11000 Praha 1, Czechia

Abstract

An interval number system is given by an initial interval cover of the extended real line and by a finite system of nonnegative Möbius transformations. Each sequence of transformations applied to an initial interval determines a sequence of nested intervals whose intersection contains a unique real number. We adapt in this setting exact real algorithms which compute arithmetical operations to arbitrary precision.

Keywords: Möbius transformation, exact real arithmetic.

1. Introduction

In an influential manuscript, Gosper [3] shows that arithmetical operations can be performed with redundant continued fractions. Based on these ideas, Vuillemin [11], Kornerup and Matula [4] or Potts [10] developped **exact real arithmetical algorithms** which work with arbitrary precision. Using the methods of symbolic dynamics, exact real arithmetic has been generalized in the theory of **Möbius number systems** introduced in Kůrka [5] and developed in Kůrka and Kazda [7]. Möbius number systems represent real numbers by infinite words from a one-sided **expansion subshift**. The letters of the alphabet stand for real orientation-preserving Möbius transformations and the concatenation of letters corresponds to the composition of transformations. A finite word of the expansion subshift represents an interval of real numbers. An infinite word represents a sequence of nested intervals of its prefixes, whose intersection contains a unique real number.

In Kůrka [6] we have characterized Möbius number systems whose expansion subshifts are of finite type or sofic. In these systems, the intervals of finite words are obtained by a particularly simple procedure. Intervals are represented by (2×2)-matrices, and the interval of a word is obtained by matrix multiplication from the interval of its immediate prefix. In the present paper we generalize this approach and develop a theory of **interval number systems**. They are given by finite interval covers $\mathcal{W} = \{W_b : b \in B\}$ of the extended real line $\overline{\mathbb{R}}$ and by finite sets of nonnegative (2×2)-matrices $\mathcal{F} = \{F_a : a \in A\}$. Each finite word $u \in B \times A^n$ determines an interval $W_u = W_{u_0} F_{u_1} \cdots F_{u_n}$ and an infinite word

$u \in B \times A^{\mathbb{N}}$ determines a sequence of nested intervals $\overline{W_{u_{[0,n)}}}$ whose intersection contains a unique real number assigned to u. We characterize transformations which can form interval number systems and show that there exists a class of **uniform** interval number systems, in which the length of intervals decreases uniformly and geometrically with the length of the words.

We modify the exact real algorithms so that they work with intervals instead of transformations. This is faciliated by a calculus of intervals which is based on matrix multiplication and results in a particularly simple tests used by the algorithms.

$$\begin{array}{ccccccccccc}
\frac{0}{1} & & & & & & \frac{1}{0} & & \frac{0}{1} & & & & & & \frac{1}{0} \\
\frac{0}{1} & & & \frac{1}{1} & & & \frac{1}{0} & & \frac{0}{2} & & & \frac{1}{1} & & & \frac{2}{0} \\
\frac{0}{1} & & \frac{1}{2} & & \frac{1}{1} & & \frac{2}{1} & & \frac{1}{0} & & \frac{0}{4} & & \frac{1}{3} & & \frac{2}{2} & & \frac{3}{1} & & \frac{4}{0} \\
\frac{0}{1} & \frac{1}{3} & \frac{1}{2} & \frac{2}{3} & \frac{1}{1} & \frac{3}{2} & \frac{2}{1} & \frac{3}{1} & \frac{1}{0} & & \frac{0}{8} & \frac{1}{7} & \frac{2}{6} & \frac{3}{5} & \frac{4}{4} & \frac{5}{3} & \frac{6}{2} & \frac{7}{1} & \frac{8}{0}
\end{array}$$

Figure 1: The Stern-Brocot tree (left) and its uniformization(right).

2. The Stern-Brocot tree

A simple number system is based on the Stern-Brocot tree (see Niqui [9] or Kůrka [6]). We start with the interval $W_\lambda = (0, \infty) = (\frac{0}{1}, \frac{1}{0})$. If $u \in \{0,1\}^*$ is a binary word and $W_u = (\frac{a_0}{a_1}, \frac{b_0}{b_1})$, then $W_{u0} = (\frac{a_0}{a_1}, \frac{a_0+b_0}{a_1+b_1})$, $W_{u1} = (\frac{a_0+b_0}{a_1+b_1}, \frac{b_0}{b_1})$, so we have $W_0 = (\frac{0}{1}, \frac{1}{1})$, $W_1 = (\frac{1}{1}, \frac{1}{0})$, $W_{00} = (\frac{0}{1}, \frac{1}{2})$, $W_{000} = (\frac{0}{1}, \frac{1}{3})$, etc. (see Figure 1 left). For an infinite binary word $u \in \{0,1\}^{\mathbb{N}}$ we get a sequence of nested intervals $W_{u_{[0,n)}}$ and the intersection of its closures contains a unique point $\Phi(u) \in [0, \infty]$. We get a continuous mapping $\Phi : \{0,1\}^{\mathbb{N}} \to [0, \infty]$ given by $\{\Phi(u)\} = \bigcap_{n>0} \overline{W_{u_{[0,n)}}}$. The map Φ can be described with the help of continued fractions. Each $u \in \{0,1\}^{\mathbb{N}}$ can be written in a unique way as $u = 1^{a_0}0^{a_1}1^{a_2}\ldots$, where $a_0 \geq 0$ and $a_n > 0$ for $n > 0$. Then u is the expansion of the continued fraction $[a_0, a_1, a_2, \ldots]$ (see Kůrka [6] for a proof), i.e.,

$$\Phi(u) = [a_0, a_1, a_2, \ldots] = a_0 + 1/(a_1 + 1/(a_2 + \cdots .$$

If we regard intervals $I = (\frac{a_0}{a_1}, \frac{b_0}{b_1})$ as (2×2)-matrices, then the recursive formula of the Stern-Brocot tree can be written as matrix multiplication $W_{u0} = W_u \cdot (\frac{1}{0}, \frac{1}{1})$, $W_{u1} = W_u \cdot (\frac{1}{1}, \frac{0}{1})$. This approach can be generalized. We obtain a faster convergence if we modify the recursive formula to $W_{u0} = W_u \cdot (\frac{2}{0}, \frac{1}{1})$, $W_{u1} = W_u \cdot (\frac{1}{1}, \frac{0}{2})$. The resulting number system can be seen in Figure 1 right. We get a continuous map $\Phi : \{0,1\}^{\mathbb{N}} \to [0, \infty]$ given by $\Phi(u) = \frac{\varphi(u)}{1-\varphi(u)}$, where $\varphi(u) = \sum_{i=0}^{\infty} u_i \cdot 2^{-i-1}$.

3. The projective line

The **extended real line** $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ can be conceived as the projective space $\mathrm{PL}(\mathbb{R}, 1)$, i.e., the space of one-dimensional subspaces of the two-dimensional vector space $\mathbb{R}^2$. A one-dimensional subspace of $\mathbb{R}^2$ is determined by any its nonzero vector $x = (x_0, x_1) \in \mathbb{R}^2 \setminus \{(0,0)\}$. If $x_1 \neq 0$ then x represents the real number $x_0/x_1 \in \mathbb{R}$ and vectors $(x_0, 0)$ represent ∞. We say that x is a **homogeneous coordinate** of x_0/x_1 or ∞. We regard $x \in \mathbb{R}^2 \setminus \{(0,0)\}$ as a column vector and write it usually as a fraction $x = \frac{x_0}{x_1}$. Two points $x, y \in \mathbb{R}^2 \setminus \{\frac{0}{0}\}$ represent the same projective point if $\det(x, y) = x_0 y_1 - x_1 y_0 = 0$.

The **stereographic projection** $\mathbf{d} : \overline{\mathbb{R}} \to \mathbb{T}$ maps $\overline{\mathbb{R}}$ to the unit circle $\mathbb{T} = \{z \in \mathbb{R}^2 : \ z_0^2 + z_1^2 = 1\}$. For $x \in \mathbb{R}$ we have $\mathbf{d}(x) = (\frac{2x}{x^2+1}, \frac{x^2-1}{x^2+1})$. In homogenous coordinates we get

$$\mathbf{d}\left(\frac{x_0}{x_1}\right) = \left(\frac{2x_0x_1}{x_0^2+x_1^2}, \frac{x_0^2-x_1^2}{x_0^2+x_1^2}\right), \ \mathbf{d}^{-1}(z_0, z_1) = \frac{z_0}{1-z_1}.$$

The angle $0 \leq \varphi(x, y) \leq \pi$ between two nonzero vectors $x, y \in \mathbb{R}^2$ can be computed by the cosine rule as $\varphi(x, y) = \arccos \frac{x \cdot y}{||x|| \cdot ||y||}$, where $x \cdot y = \sum_i x_i y_i$ is the scalar product and $||x|| = \sqrt{x \cdot x}$ is the Euclidean norm. The angle between $-x$ and y is then $\pi - \varphi(x, y) = \arccos \frac{-x \cdot y}{||x|| \cdot ||y||}$. Taking the smaller of these two angles we define the **angle metric** in $\overline{\mathbb{R}}$ by

$$d_a(x, y) = \arccos \frac{|x \cdot y|}{||x|| \cdot ||y||} = \arctan \frac{|\det(x, y)|}{|x \cdot y|} \in [0, \tfrac{\pi}{2}].$$

For $x, y \in \mathbb{R}$ we get $d_a(x, y) = \arccos \frac{|xy+1|}{\sqrt{(x^2+1)(y^2+1)}} = \arctan \frac{|x-y|}{|xy+1|}$.

Alternatively, we consider the **chord metric** in $\overline{\mathbb{R}}$ which is the distance $\sin \varphi(x, y) = \sin(\pi - \varphi(x, y))$ of $\mathbf{d}(x)$ from $\mathbf{d}(y)$ in $\mathbb{R}^2$:

$$d_c(x, y) = \frac{\sqrt{||x||^2 \cdot ||y||^2 - (x \cdot y)^2}}{||x|| \cdot ||y||} = \frac{|\det(x, y)|}{||x|| \cdot ||y||}.$$

For $x, y \in \mathbb{R}$ we get $d_c(x, y) = |x - y|/\sqrt{(x^2+1)(y^2+1)}$. These two metrics are equivalent. We have $d_c(x, y) \leq d_a(x, y) \leq \frac{\pi}{2} d_c(x, y)$, and $\lim_{y \to x} \frac{d_c(y,x)}{d_a(y,x)} = 1$.

A bijective linear transformation of the vector space $\mathbb{R}^2$ is determined by a (2×2)-matrix $M = (M_{ij})_{i,j=0,1}$ with $\det(M) = M_{00}M_{11} - M_{01}M_{10} \neq 0$ via $(Mx)_i = \sum_{j=0}^{1} M_{ij} x_j$. The M-image of a one-dimensional subspace of $\mathbb{R}^2$ is a one-dimensional subspace of $\mathbb{R}^2$, so M determines a projective isomorphism of the projective space $\mathrm{PL}(\mathbb{R}, 1) = \overline{\mathbb{R}}$ called **fractional linear** or **Möbius transformation**. For $\lambda \neq 0$, λM determines the same transformation as M, so a Möbius transformation is a point of the **three-dimensional projective space** $\mathrm{PL}(\mathbb{R}, 3)$ of one-dimensional linear subspaces of the vector space $\mathbb{R}^4$. A nonzero matrix in such a one-dimensional subspace is a homogenous coordinate of the transfomation.

Definition 1. *A (positively oriented Möbius)* **transformation** *is a self-map of* $\overline{\mathbb{R}}$ *of the form* $M\left(\frac{x_0}{x_1}\right) = \frac{ax_0+bx_1}{cx_0+dx_1}$ *where* $a, b, c, d \in \mathbb{R}$ *and* $\det(M) = ad - bc > 0$. *The space of transformations is* $\mathbb{M}(\mathbb{R}) = \{M \in \mathrm{PL}(\mathbb{R}, 3) : \det(\mathrm{M}) > 0\}$.

We denote by M_{-0} the left column and by M_{-1} the right column of M, so $(M_{-j})_i = M_{ij}$. Often we write M as a pair $M = (M_{-0}, M_{-1}) = (\frac{M_{00}}{M_{10}}, \frac{M_{01}}{M_{11}}) = (\frac{a}{c}, \frac{b}{d})$. Each Möbius transformation is one-to-one and its inverse is a Möbius transformation $(\frac{a}{c}, \frac{b}{d})^{-1} = (\frac{d}{-c}, \frac{-b}{a})$. Möbius transformations form a group.

The **trace** of a transformation $M = (\frac{a}{c}, \frac{b}{d})$ is $\mathrm{tr}(M) = |a + d|/\sqrt{ad - bc}$. If $\mathrm{tr}(M) > 2$ then M has an unstable fixed point $\mathbf{u}(M)$ and a stable fixed point $\mathbf{s}(M)$ such that $\lim_{n\to\infty} F^n(x) = \mathbf{s}(M)$ for each $x \in \overline{\mathbb{R}} \setminus \{\mathbf{u}(M)\}$. We say in this case that M is **hyperbolic**. If $\mathrm{tr}(M) = 2$ then M has a unique fixed point $\mathbf{s}(M)$ such that $\lim_{n\to\infty} F^n(x) = \mathbf{s}(M)$ for each $x \in \overline{\mathbb{R}}$ and we say that M is **parabolic**. If $\mathrm{tr}(M) < 2$ then M has no fixed point in $\overline{\mathbb{R}}$ and we say that M is **elliptic**.

4. Intervals

A **set interval** is a connected subset of $\overline{\mathbb{R}}$. A proper set interval is a nonempty set interval properly included in $\overline{\mathbb{R}}$. For $a, b \in \overline{\mathbb{R}}$, the open interval I° of (a, b) consists of inner points of the counterclockwise arc from its left endpoint a to its right endpoint b. If $a, b \in \mathbb{R}$ then

$$(a,b)^\circ = \begin{cases} \{x \in \mathbb{R} : \ a < x < b\} & \text{if} \quad a < b \\ \{x \in \mathbb{R} : \ a < x \text{ or } x < b\} \cup \{\infty\} & \text{if} \quad b < a \end{cases}$$

For $a = \frac{r\sin\alpha}{r\cos\alpha} \in \overline{\mathbb{R}}$ we get $\mathbf{d}(a) = (\sin 2\alpha, -\cos 2\alpha)$, so the stereographic projection doubles the angles. Matrices with columns $a = \frac{r\sin\alpha}{r\cos\alpha}$, $b = \frac{s\sin\beta}{s\cos\beta}$ where $0 \le \alpha < 2\pi$, $\alpha < \beta < \alpha + \pi$ therefore represent all proper set intervals. Since $\det(a, b) = rs\sin(\alpha - \beta) < 0$, we define intervals as (2×2)-matrices with negative determinant and write them as pairs $I = (\frac{a_0}{a_1}, \frac{b_0}{b_1})$ of their left and right endpoints. A nonzero multiple of I represents the same interval, so intervals are conceived as points of the three-dimensional projective space $\mathrm{PL}(\mathbb{R}, 3)$.

Definition 2. *An interval is a* (2×2)*-matrix with negative determinant. Two intervals are equal if one is a nonzero multiple of the other. The space of intervals is* $\mathbb{I}(\mathbb{R}) = \{I \in \mathrm{PL}(\mathbb{R}, 3) : \det(\mathrm{I}) < 0\}$,

The length of an interval $I = (a, b)$ is the length of the counterclockwise arc from $\mathbf{d}(a)$ to $\mathbf{d}(b)$ divided by 2π. This is the same as the length of the clockwise arc from $a/||a||$ to $b/||b||$ divided by π. We normalize by π to obtain the unit length of the full interval $\overline{\mathbb{R}}$. If $|I| \le \frac{1}{2}$ then $|I| = d_a(a, b)/\pi$:

$$|I| = \frac{1}{\pi}\arccos\frac{a \cdot b}{||a|| \cdot ||b||} = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{a \cdot b}{\det(a, b)}.$$

We can regard an interval $I = (a, b)$ as a basis of the projective space $\mathrm{PL}(\mathbb{R}, 1) = \overline{\mathbb{R}}$. If $y \in \overline{\mathbb{R}}$ and $x = Iy = (y_0 a + y_1 b)$, then $y = I^{-1}x$ is the coordinate of x in the basis I. If $y_0, y_1 \geq 0$ and $y_0 + y_1 = 1$, then x is a **convex combination** of the endpoints of I so x belongs to the closure of I. If both y_0 and y_1 are nonpositive (and $y_0 + y_1 = -1$), then x is another representation of $\frac{-x_0}{-x_1}$ so x belongs to the closure of I as well. For $x \in \mathbb{R}^2$ define $\mathrm{sgn}(x) \in \{-1, 0, 1\}$ as the sign of $x_0 \cdot x_1$, so $\mathrm{sgn}(0) = \mathrm{sgn}(\infty) = 0$.

Definition 3. *The interior and closure of an interval $I \in \mathbb{I}(\mathbb{R})$ are defined by*

$$I^\circ = \{x \in \overline{\mathbb{R}} : \ \mathrm{sgn}(I^{-1}x) > 0\}, \quad \overline{I} = \{x \in \overline{\mathbb{R}} : \ \mathrm{sgn}(I^{-1}x) \geq 0\}.$$

Sometimes we use more conventional notation like $[0, \infty] = \overline{(\frac{0}{1}, \frac{1}{0})}$ or $(-\infty, 0) = (\frac{-1}{0}, \frac{0}{1})^\circ$. If $J \in \mathbb{I}(\mathbb{R})$ is an interval and $M \in \mathbb{M}(\mathbb{R})$ is a transformation, then both matrix products MJ and JM are intervals. The interval MJ is the image of J by M. If $I = JM$, then $M = J^{-1}I$ can be regarded as the coordinate of I in the basis J. If M is a nonnegative matrix, then the columns of I are convex combinations of those of J, so I is a subset of J. The sign of a matrix $X \in \mathrm{PL}(\mathbb{R}, 3)$ is defined similarly as the sign of a point:

$$\mathrm{sgn}(X) = \begin{cases} 1 & \text{if} \quad \exists \lambda \neq 0, \forall i, j, \lambda X_{ij} > 0 \\ 0 & \text{if} \quad \exists \lambda \neq 0, \forall i, j, \lambda X_{ij} \geq 0 \text{ and } \exists i, j, X_{ij} = 0 \\ -1 & \text{otherwise} \end{cases}$$

Proposition 4. *Let $I, J \in \mathbb{I}(\mathbb{R})$ be intervals and let $M \in \mathbb{M}(\mathbb{R})$ be a transformation. Then*
1. $\overline{I} \subseteq \overline{J}$ iff $\mathrm{sgn}(J^{-1}I) \geq 0$.
2. $M(\overline{I}) = \{Mx : \ x \in \overline{I}\} = \overline{MI}$.
3. If $\overline{I} \subseteq \overline{J}$ then $\overline{MI} \subseteq \overline{MJ}$.

Proof: 1. If $\mathrm{sgn}(J^{-1}I) \geq 0$ and $x \in \overline{I}$, then $\mathrm{sgn}(J^{-1}x) = \mathrm{sgn}((J^{-1}I) \cdot (I^{-1}x)) \geq 0$, so $x \in \overline{J}$. To prove the converse, assume by contradiction that $I = (a, b)$, $\overline{I} \subseteq \overline{J}$ and $\mathrm{sgn}(J^{-1}I) < 0$. Since $a, b \in \overline{I} \subseteq \overline{J}$, we have $J^{-1}I = (J^{-1}a, J^{-1}b) = (c, d)$ with $\mathrm{sgn}(c) \geq 0$ and $\mathrm{sgn}(d) \geq 0$. Since $\mathrm{sgn}(J^{-1}I) < 0$, either $c_0, c_1 \geq 0$ and $d_0, d_1 \leq 0$, or $c_0, c_1 \leq 0$ and $d_0, d_1 \geq 0$. In the former case for any $y \in \overline{\mathbb{R}}$ we get $z = \left(\frac{c_0}{c_1}, \frac{d_0}{d_1}\right) \cdot \left(\frac{-d_0}{c_0}, \frac{-d_1}{c_1}\right) \cdot \frac{y_0}{y_1} = \left(\frac{0}{D}, \frac{-D}{0}\right) \cdot \frac{y_0}{y_1} = \frac{-Dy_1}{Dy_0}$, where $D = \det(c, d) > 0$. If $\mathrm{sgn}(y) > 0$, then for $w = (\frac{-d_0}{c_0}, \frac{-d_1}{c_1}) \cdot y$ we have $\mathrm{sgn}(w) \geq 0$, so $Iw \in \overline{I}$. However, $\mathrm{sgn}(J^{-1}Iw) = \mathrm{sgn}(z) < 0$ so $Iw \notin \overline{J}$ and this is a contradiction. If $c_0, c_1 \leq 0$ and $d_0, d_1 \geq 0$, the proof is analogous.
2. We have $y \in \overline{MI}$ iff $\mathrm{sgn}(I^{-1}M^{-1}y) \geq 0$ iff $M^{-1}y \in \overline{I}$ iff $y \in M(\overline{I})$.
3. If $\overline{I} \subseteq \overline{J}$, then $\mathrm{sgn}((MJ)^{-1}MI) = \mathrm{sgn}(J^{-1}I) \geq 0$, so $\overline{MI} \subseteq \overline{MJ}$. □

Sometimes, it is convenient to regard a matrix $M = (a, b) \in \mathbb{M}(\mathbb{R})$ with positive determinant as the interval (b, a). Thus we define the interior and closure of $M \in \mathbb{M}(\mathbb{R})$ by

$$M^\circ = \{x \in \overline{\mathbb{R}} : \ \mathrm{sgn}(M^{-1}x) > 0\}, \ \overline{M} = \{x \in \overline{\mathbb{R}} : \ \mathrm{sgn}(M^{-1}x) \geq 0\}.$$

Definition 5. *A system of intervals* $\{W_b \in \mathbb{I}(\mathbb{R}) :\ b \in B\}$ *is an* **almost-cover** *of* $\overline{\mathbb{R}}$ *if* $\bigcup_{b\in B} \overline{W_b} = \overline{\mathbb{R}}$. *It is a* **cover** *of* $\overline{\mathbb{R}}$, *if* $\bigcup_{b\in B} W_b^\circ = \overline{\mathbb{R}}$. *A system of transformations* $\{F_a \in \mathbb{M}(\mathbb{R}) :\ a \in A\}$ *is an almost-cover of an interval* J *if* $\bigcup_{a\in A} \overline{F_a} = \overline{J}$. *It is a cover of* J, *if* $\bigcup_{a\in A} F_a^\circ = J^\circ$.

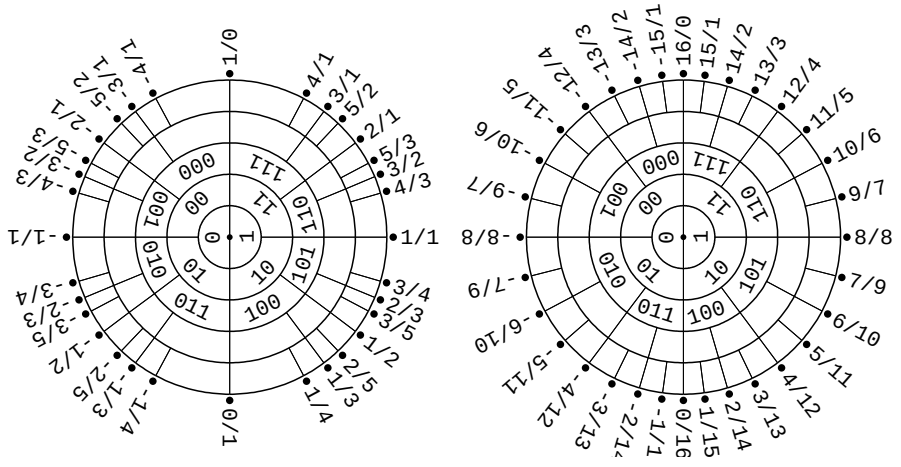

Figure 2: Nonredundant interval number systems: Continued fractions with $\mathcal{F} = ((\frac{1}{0}, \frac{1}{1}), (\frac{1}{1}, \frac{0}{1}))$, $\mathcal{W} = ((\frac{-1}{0}, \frac{0}{1}), (\frac{0}{1}, \frac{1}{0}))$ (left), and $\frac{1}{2}$-uniform system with $\mathcal{F} = ((\frac{2}{0}, \frac{1}{1}), (\frac{1}{1}, \frac{0}{2}))$, $\mathcal{W} = ((\frac{-1}{0}, \frac{0}{1}), (\frac{0}{1}, \frac{1}{0}))$.

5. Interval number systems

Consider a system of intervals $\mathcal{W} = \{W_b :\ b \in B\}$ indexed by an alphabet (finite set) B, and a system of nonnegative transformations $\mathcal{F} = \{F_a :\ a \in A\}$ indexed by an alphabet A. For a finite word $u = u_0u_1 \cdots u_n \in B \times A^*$ of length $|u| = n + 1$ denote by $W_u = W_{u_0}F_{u_1} \cdots F_{u_n}$ (we have $u_0 \in B$ and $u_i \in A$ for $i > 0$). If $m < n$ then $v = u_{[0,m]}$ is a prefix of u, so $\overline{W_u} \subseteq \overline{W_v}$.

Definition 6. *We say that* $(\mathcal{W}, \mathcal{F})$ *is an* **interval number system** *if* $\mathcal{W} = \{W_b :\ b \in B\}$ *is an almost-cover of* $\overline{\mathbb{R}}$ *by intervals* $W_b \in \mathbb{I}(\mathbb{R})$, $\mathcal{F} = \{F_a :\ a \in A\}$ *is an almost-cover of* $[0, \infty]$ *by nonnegative transformations* $F_a \in \mathbb{M}(\mathbb{R})$ *and if* $\lim_{n\to\infty} |W_{u_{[0,n)}}| = 0$ *for each infinite word* $u \in B \times A^{\mathbb{N}}$. *If* $\mathcal{W}$ *is a cover of* $\overline{\mathbb{R}}$ *and* $\mathcal{F}$ *is a cover of* $(0, \infty)$, *we say that the system* $(\mathcal{W}, \mathcal{F})$ *is* **redundant**. *An interval number system determines the* **symbolic map** $\Phi : B \times A^{\mathbb{N}} \to \overline{\mathbb{R}}$ *defined by* $\{\Phi(u)\} = \bigcap_{n>0} \overline{W_{u_{[0,n)}}}$.

Example 1 (Continued fractions, Figures 1 and 2 left). $B = A = \{0, 1\}$, $\mathcal{W} = ((\frac{-1}{0}, \frac{0}{1}), (\frac{0}{1}, \frac{1}{0}))$, $\mathcal{F} = ((\frac{1}{0}, \frac{1}{1}), (\frac{1}{1}, \frac{0}{1}))$.

Example 2 (Uniform $\frac{1}{2}$-system, Figures 1 and 2 right). $B = A = \{0, 1\}$, $\mathcal{W} = ((\frac{-1}{0}, \frac{0}{1}), (\frac{0}{1}, \frac{1}{0}))$, $\mathcal{F} = ((\frac{2}{0}, \frac{1}{1}), (\frac{1}{1}, \frac{0}{2}))$.

Example 3 (Uniform $\frac{2}{3}$-system, Figure 3 left). $B = \{0,1,2,3\}$, $A = \{0,1\}$, $\mathcal{F} = ((\frac{3}{0},\frac{1}{2}),(\frac{2}{1},\frac{0}{3}))$, $\mathcal{W} = ((\frac{-1}{0},\frac{0}{1}),(\frac{-1}{1},\frac{1}{1}),(\frac{0}{1},\frac{1}{0}),(\frac{1}{1},\frac{1}{-1}))$.

Example 4 (Uniform $\frac{2}{4}$-system, Figure 3 right). $B = \{0,1,2,3\}$, $A = \{0,1,2\}$, $\mathcal{F} = ((\frac{2}{0},\frac{1}{1}),(\frac{3}{1},\frac{1}{3}),(\frac{1}{1},\frac{0}{2}))$, $\mathcal{W} = ((\frac{-1}{0},\frac{0}{1}),(\frac{-1}{1},\frac{1}{1}),(\frac{0}{1},\frac{1}{0}),(\frac{1}{1},\frac{1}{-1}))$.

We now generalize these examples.

Definition 7. *Given integers $q \geq 2$, $1 \leq p \leq q-1$, the $\frac{p}{q}$-uniform interval number system has alphabet $A = \{0,1,\ldots,q-p\}$ and transformations $F_i = \left(\frac{q-i}{i}, \frac{q-p-i}{p+i}\right)$, $0 \leq i \leq q-p$. $\mathcal{W}$ is an arbitrary almost cover of $\overline{\mathbb{R}}$.*

Observe that for each $F_i = (\frac{a}{c},\frac{b}{d})$ we have $\det(F_i) = pq$, $a+c = b+d = q$, $a-b = d-c = p$, so $\mathbf{u}(F_i) = -1$ and $\mathbf{s}(F_i) = \frac{q-p-i}{i}$. If $p \geq 2$, and if $\mathcal{W}$ is a cover, then the system is redundant.

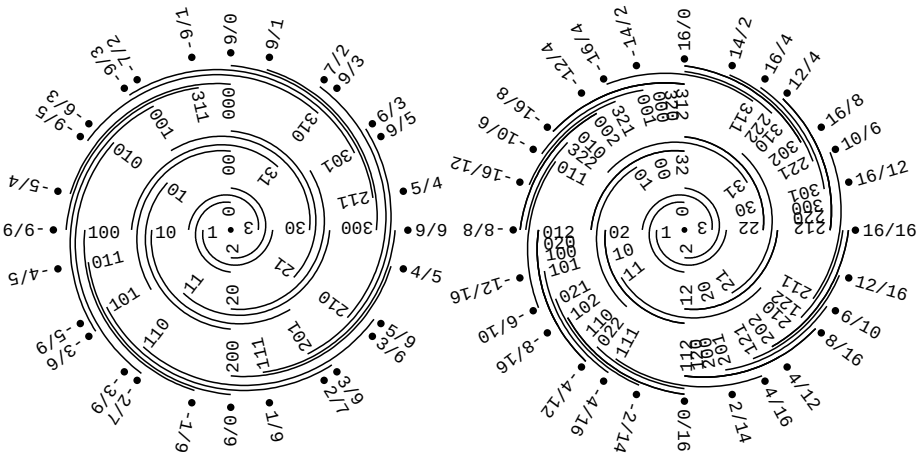

Figure 3: Redundant systems: the homogenous $\frac{2}{3}$-system $\mathcal{F} = ((\frac{3}{0},\frac{1}{2}),(\frac{2}{1},\frac{0}{3}))$ (left) and the $\frac{2}{4}$-system $\mathcal{F} = ((\frac{4}{0},\frac{2}{2}),(\frac{3}{1},\frac{1}{3}),(\frac{2}{2},\frac{0}{4}))$ (right). The initial cover is in both cases $\mathcal{W} = ((\frac{-1}{0},\frac{0}{1}),(\frac{-1}{1},\frac{1}{1}),(\frac{0}{1},\frac{1}{0}),(\frac{1}{1},\frac{1}{-1}))$.

Theorem 8. *For each $\frac{p}{q}$-uniform interval number system there exist positive numbers $0 < C_0 < C_1$ such that $C_0 \cdot (p/q)^{|u|} \leq |W_u| \leq C_1 \cdot (p/q)^{|u|}$ for each $u \in B \times A^*$.*

Proof: Normalize the matrices of the transformations to the unit sums of their columns, so $F_i = \left(\frac{(q-i)/q}{i/q}, \frac{(q-p-i)/q}{(p+i)/q}\right)$. Thus we work with a fixed particular representations of the transformations and also with some fixed representations of the endpoints of the intervals W_b. For $u \in B \times A^{\mathbb{N}}$ denote by $W_{u_{[0,n]}} = (x_n, y_n)$. Then each x_n and y_n is a convex combination of x_0 and y_0, so it lies on the line connecting x_0 with y_0 (see Figure 4 left). By a simple computation

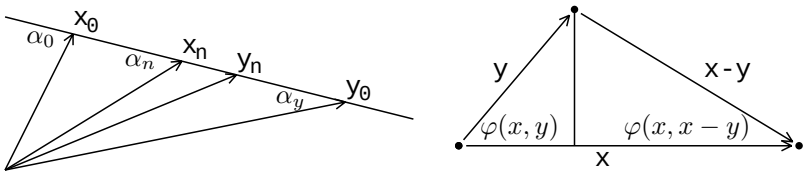

Figure 4: The endpoints of W_u (left), and the chord distance (right).

we get $(y_n - x_n) = (y_0 - x_0)(\frac{p}{q})^n$, so $\lim_{n\to\infty} ||y_n - x_n|| = \lim_{n\to\infty} |W_{u_{[0,n)}}| = 0$. This means that $(\mathcal{W}, \mathcal{F})$ is an interval number system. The chord distance $d_c(x, y)$ of nonzero vectors $x, y \in \mathbb{R}^2$ can be expressed from the angle $\varphi(x, y-x)$ as $d_c(x, y) = \sin\varphi(x, y) = ||x - y|| \sin\varphi(x, y - x)/||y||$ (see Figure 4 right). We get

$$d_c(x_n, y_n) = \left(\frac{p}{q}\right)^n \frac{||x_0 - y_0||}{||y_n||} \sin\alpha_n,$$

where $\alpha_n = \varphi(x_n, y_n - x_n)$. For $\alpha_y = \varphi(y_0, y_0 - x_0)$ we get $0 < \alpha_y \le \alpha_n \le \alpha_0 < \pi$, so $\min\{\sin\alpha_y, \sin\alpha_0\} \le \sin\alpha_n \le 1$,

$$\min\{||x_0||, ||y_0||\} \cos\frac{\varphi(x_0, y_0)}{2} \le ||y_n|| \le \max\{||x_0||, ||y_0||\}.$$

Since $d_c(x_n, y_n) \le |W_{u_{[0,n)}}| \le \frac{\pi}{2} d_c(x_n, y_n)$, we get the result. □

Before we prove a general theorem characterizing transformations of interval number systems we need two lemmas.

Lemma 1. *For $t \in \mathbb{R}$ denote by $v_t = (t, 1) \in \mathbb{R}^2$. Let $r > 0$, $0 < q < 1$, $s < t$, $t + qr < s + r$. Then $\varphi(v_t, v_{t+qr}) \le q \cdot \varphi(v_s, v_{s+r})$ (see Figure 5 left).*

Proof: For a fixed r, the function $f(s) = \varphi(v_s, v_{s+r}) = \arctan(s+r) - \arctan(s)$ has maximum at $s = -r/2$ and is decreasing in the interval $[-r/2, \infty)$. Assume that $||v_t|| \le ||v_{t+qr}||$, so $t \ge -qr/2$ (see Figure 5). Then $\varphi(v_s, v_{s+r}) \ge \varphi(v_t, v_{t+r})$, so

$$\frac{\varphi(v_t, v_{t+qr})}{\varphi(v_s, v_{s+r})} \le \frac{\varphi(v_t, v_{t+qr})}{\varphi(v_t, v_{t+r})}.$$

If $t < u$ then (see Figure 5 right)

$$\begin{aligned} \varphi(v_u, v_{u+qr}) - \varphi(v_u, v_{u+r}) &= -\varphi(v_{u+qr}, v_{u+r}) \ge -\varphi(v_{t+qr}, v_{t+r}) \\ &\ge \varphi(v_t, v_{t+qr}) - \varphi(v_t, v_{t+r}) \end{aligned}$$

Multiplying this inequality by $\varphi(v_t, v_{t+qr}) \ge \varphi(v_u, v_{u+qr})$, we get

$$\begin{aligned} -\varphi(v_t, v_{t+qr}) \cdot \varphi(v_u, v_{u+r}) &\ge -\varphi(v_u, v_{u+qr}) \cdot \varphi(v_t, v_{t+r}) \\ \frac{\varphi(v_t, v_{t+qr})}{\varphi(v_t, v_{t+r})} &\le \frac{\varphi(v_u, v_{u+qr})}{\varphi(v_u, v_{u+r})}. \end{aligned}$$

$$\text{Since} \quad \lim_{u\to\infty} \frac{\varphi(v_u, v_{u+qr})}{\varphi(v_u, v_{u+r})} = \lim_{u\to\infty} \frac{\arctan(u+qr) - \arctan(u)}{\arctan(u+r) - \arctan(u)} = q,$$

we get the result. □

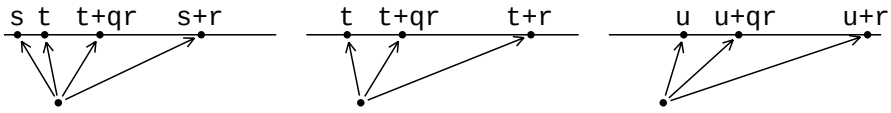

Figure 5: The length of subintervals

Lemma 2. *If F is a nonnegative hyperbolic transformation with $\mathbf{u}(F) \in (-\infty, 0)$, then there exists $0 < q < 1$ such that $|IF| \le q|I|$ for every interval I.*

Proof: Let $\mathbf{u}(F) = \frac{u_0}{-u_1}$ with $u_0, u_1 > 0$, and normalize $F = (\frac{a}{c}, \frac{b}{d})$ so that $a, b, c, d \ge 0$ and $d + bu_1/u_0 = 1$. Set $q = d - cu_0/u_1$. By the assumption we have $(au_0 - bu_1)u_1 = (du_1 - cu_0)u_0$, so

$$\begin{aligned} a - bu_1/u_0 &= d - cu_0/u_1 = q \\ a + cu_0/u_1 &= d + bu_1/u_0 = 1 \end{aligned}$$

Subtracting these equations we get $1 - q = bu_1/cu_0 + cu_0/u_1 > 0$ so $q < 1$. Since $0 < ad - bc = ad - (d-q)(a-q) = q(a+d-q)$, we get $q > 0$. Let $I = (x, y)$ and $IF = (z, w) = (ax + cy, bx + dy)$. Then

$$\begin{aligned} u_0 z &= a \cdot u_0 x + (cu_0/u_1) \cdot u_1 y \\ u_1 w &= (bu_1/u_0) \cdot u_0 x + d \cdot u_1 y \\ u_1 w - u_0 z &= (u_1 d - u_0 c)y - (u_0 a - u_1 b)x = q \cdot (u_1 y - u_0 x) \end{aligned}$$

Thus $u_0 z$, $u_1 w$ are convex combinations of $u_0 x$, $u_1 y$, so we have a situation in Figure 5 left with $s = u_0 x$, $t = u_0 z$, $t + qr = u_1 w$, $s + r = u_1 y$. By Lemma 1 we have $|IF| = \varphi(u_0 z, u_1 w) \le q \cdot \varphi(u_0 x, u_1 y) = q \cdot |I|$. □

Theorem 9. *Let $\mathcal{W}$ be an almost cover of $\overline{\mathbb{R}}$ by intervals, and let $\mathcal{F}$ be an almost-cover of $[0, \infty]$ by nonnegative transformations. Then $(\mathcal{W}, \mathcal{F})$ is an interval number system iff every F_a is either parabolic with fixed point $\mathbf{s}(F_a) \in \{\frac{0}{1}, \frac{1}{0}\}$, or hyperbolic with unstable fixed point $\mathbf{u}(F_a) \in (-\infty, 0)$.*

Proof: Assume that $\lim_{n\to\infty} |W_{u_{[0,n)}}| = 0$ for every $u \in B \times A^{\mathbb{N}}$. Since each F_a is nonnegative, $F_a([0, \infty]) \subseteq [0, \infty]$. It follows that F_a has a fixed point in $[0, \infty]$, so it cannot be elliptic. If F_a is parabolic, then its fixed point cannot be in $(0, \infty)$, since $[0, \infty]$ would not be invariant. Thus $\mathbf{s}(F_a) = \frac{0}{1}$ or $\mathbf{s}(F_a) = \frac{1}{0}$. Let F_a be hyperbolic. Since $[0, \infty]$ is invariant, it must contain the stable fixed point $\mathbf{s}(F_a)$, but $(0, \infty)$ cannot contain the unstable fixed point $\mathbf{u}(F_a)$. Assume by contradiction that $\mathbf{u}(F_a) = \frac{1}{0}$, so $F_a(x) = (x + c)/q$ for some $c \ge 0$, $q > 1$. For $I = (\frac{0}{1}, \frac{1}{0})$ we get $IF^n = \left(\frac{c(1+q+\cdots+q^{n-1})}{q^n}, \frac{1}{0}\right)$, so $\bigcap_n \overline{IF^n} = \overline{(\frac{c}{q-1}, \frac{1}{0})}$ which

is not a singleton. If $\mathbf{u}(F_a) = \frac{0}{1}$, the proof is analogous. Thus $\mathbf{u}(F_a) \in (-\infty, 0)$. Conversely assume that each F_a is either parabolic with fixed point $\frac{0}{1}$ or $\frac{1}{0}$ or hyperbolic with unstable fixed point $\mathbf{u}(F_a) \in (-\infty, 0)$ and let $u \in B \times A^{\mathbb{N}}$. If F_{u_i} is hyperbolic for an infinite number of i, then $\lim_{n\to\infty} |W_{u_{[0,n)}}| = 0$ by Lemma 2. The composition of two parabolic transformations with fixed points 0 and ∞ is $(\frac{1}{0}, \frac{a}{1}) \cdot (\frac{1}{b}, \frac{0}{1}) = (\frac{1+ab}{b}, \frac{a}{1})$ which is a hyperbolic transformation with unstable fixed point in $(-\infty, 0)$. If there exists an infinite number of n such that F_{u_n} is parabolic with fixed point $\frac{0}{1}$ and $F_{u_{n+1}}$ is parabolic with fixed point $\frac{1}{0}$, then $\lim_{n\to\infty} |W_{u_{[0,n)}}| = 0$ by Lemma 2. Thus the only remaining case is that for each $n \geq n_0$, F_{u_n} is parabolic with the same fixed point, say $\frac{1}{0}$, so $F_{u_n} = (\frac{1}{0}, \frac{a}{1})$. If $I = (x, y)$ then $IF^n = (x, nax + y)$, so $\bigcap_{n>0} \overline{IF^n} = \{x\}$. □

6. Expansion of rational numbers

Arithmetical algorithms can work with interval number systems whose intervals and transformations have integer entries. Denote by $\mathbb{Z}$ the set of integers, $\overline{\mathbb{Q}} = \{x = \frac{x_0}{x_1} \neq \frac{0}{0} : x_0, x_1 \in \mathbb{Z}\}$ the extended set of rational numbers and

$$\begin{aligned} \mathbb{M}(\mathbb{Z}) &= \{(\tfrac{a_0}{a_1}, \tfrac{b_0}{b_1}) \in \mathbb{Z}^4 : a_0 b_1 - a_1 b_0 > 0\}, \\ \mathbb{I}(\mathbb{Z}) &= \{(\tfrac{a_0}{a_1}, \tfrac{b_0}{b_1}) \in \mathbb{Z}^4 : a_0 b_1 - a_1 b_0 < 0\}. \end{aligned}$$

A **labelled graph** over an alphabet A is a structure $\mathcal{G} = (V, E, s, t, \ell)$, where V is the set of vertices, E is the set of edges, $s, t : E \to V$ are the source and target maps, and $\ell : E \to A$ is a labeling function. A (finite or infinite) path is a sequence of edges $e = e_0 e_1 \cdots$ such that $t(e_i) = s(e_{i+1})$. The label of a path is the concatenation of the labels of its edges.

Given an interval number system $(\mathcal{W}, \mathcal{F})$, we say that $u \in B \times A^{\mathbb{N}}$ is an **expansion** of $x \in \overline{\mathbb{R}}$, if $\Phi(u) = x$. For rational numbers we obtain a simple expansion algorithm based on the search of a path in the expansion graph.

Definition 10. *Given an interval number system $(\mathcal{W}, \mathcal{F})$ with integer matrices, the vertices of the* **expansion graph** *are (x, n) where $x \in \overline{\mathbb{Q}}$ and $n \in \{0, 1\}$. The labelled edges are*

$$\begin{aligned} (x, 0) &\xrightarrow{b} (W_b^{-1} x, 1) \quad \textit{if} \quad \operatorname{sgn}(W_b^{-1} x) \geq 0, \\ (x, 1) &\xrightarrow{a} (F_a^{-1} x, 1) \quad \textit{if} \quad \operatorname{sgn}(F_a^{-1} x) \geq 0. \end{aligned}$$

Proposition 11. *An infinite word $u \in B \times A^{\mathbb{N}}$ is the label of a path with source $(x, 0)$ iff $\Phi(u) = x$.*

Proof: For a finite word $u \in B \times A^*$ we prove by induction that $x \in \overline{W_u}$ iff u is the label of a path with source $(x, 0)$. □

7. The unary algorithm

Assume that we want to compute a transformation $M \in \mathbb{M}(\mathbb{Z})$ in an interval number system $(\mathcal{W}, \mathcal{F})$ over $B \times A$, whose transformations and intervals have integer entries. Given an input $u \in B \times A^{\mathbb{N}}$ we construct an output $v \in B \times A^{\mathbb{N}}$ with $\overline{MW_{u_{[0,n)}}} \subseteq \overline{W_{v_{[0,m)}}}$. Here $u_{[0,n)}$ is the part of the input read and $v_{[0,m)}$ is the part of the output constructed at time $n + m$. The algorithm can be described as a search of a path in the **unary graph**:

Definition 12. *Given an interval number system $(\mathcal{W}, \mathcal{F})$ with integer matrices, the vertices of the* **unary graph** *are (X, n, m), where $X \in \mathbb{M}(\mathbb{Z}) \cup \mathbb{I}(\mathbb{Z})$, $n, m \in \{0, 1\}$. The labelled edges are*

$$\begin{aligned}
(X,0,0) &\xrightarrow{b/\lambda} (XW_b,1,0),\\
(X,1,m) &\xrightarrow{a/\lambda} (XF_a,1,m),\\
(X,1,0) &\xrightarrow{\lambda/b} (W_b^{-1}X,1,1) \quad \textit{if} \quad \operatorname{sgn}(W_b^{-1}X) \geq 0,\\
(X,1,1) &\xrightarrow{\lambda/a} (F_a^{-1}X,1,1) \quad \textit{if} \quad \operatorname{sgn}(F_a^{-1}X) \geq 0.
\end{aligned}$$

Edges with labels b/λ or a/λ are called **absorption edges** and those with labels λ/b or λ/a are called **emission edges**. The label of a path is the concatenation of the labels of its edges. The following proposition is easily proved by induction.

Proposition 13. *If $(M, 0, 0) \xrightarrow{u/v} (X, 1, 1)$ is a path in the unary graph, then $X = W_v^{-1} M W_u$ and $\overline{MW_u} \subseteq \overline{W_v}$.*

For each vertex of the unary graph there exists several outgoing absorption edges. For some vertices there exist outgoing emission edges as well. To get a deterministic algorithm, we consider selectors, which select one of the outgoing edges.

Definition 14. *The* **Euclidean norm** *of $X \in \mathbb{I}(\mathbb{Z}) \cup \mathbb{M}(\mathbb{Z})$ is $||X|| = \sqrt{\sum_{ij} X_{ij}^2}$. Its* **admissible sets** *are defined by*

$$\begin{aligned}
\mathcal{A}(X) &= \{a \in A : \operatorname{sgn}(F_a^{-1}X) \geq 0\},\\
\mathcal{B}(X) &= \{b \in B : \operatorname{sgn}(W_b^{-1}X) \geq 0\}.
\end{aligned}$$

Selectors are functions $s_A : \mathbb{M}(\mathbb{Z}) \cup \mathbb{I}(\mathbb{Z}) \to A \cup \{\lambda\}$, $s_B : \mathbb{M}(\mathbb{Z}) \cup \mathbb{I}(\mathbb{Z}) \to B \cup \{\lambda\}$, such that if $s_A(X) \neq \lambda$ then $s_A(X) \in \mathcal{A}(X)$ and if $s_B(X) \neq \lambda$ then $s_B(X) \in \mathcal{B}(X)$. The value λ of a selector signifies that an absorption is selected. To make the algorithm faster, we use a threshold parameter of a selector. The **Lebesgue number** of an open cover $\mathcal{W} = \{W_b : b \in B\}$ is the largest value $\mathcal{L}(\mathcal{W})$ such that each interval I with $|I| < \mathcal{L}(\mathcal{W})$ is included in some W_b. If $X \in \mathbb{M}(\mathbb{R}) \cup \mathbb{I}(\mathbb{R})$ and $|X| < \mathcal{L}(\mathcal{W})$ then $\mathcal{B}(X) \neq 0$. If $|X| < \mathcal{L}(\mathcal{F})$ then $\mathcal{A}(X) \neq 0$. The **least norm selector** s_A with parameter τ selects $a \in \mathcal{A}(X)$

threshold parameter: $\tau < \min\{\mathcal{L}(\mathcal{W}), \mathcal{L}(\mathcal{F})\}$;
input: $X \in \mathbb{I}(\mathbb{Z}) \cup \mathbb{M}(\mathbb{Z})$; output: $s_A \in \mathcal{A} \cup \{\lambda\}$;
begin
  if $|X| > \tau$ then begin $s_A := \lambda$; exit; end
  $r := 0$;
  for $a \in \mathcal{A}(X)$ do
    if $r = 0$ or $||F_a^{-1}X|| \leq r$ then begin $s_A := a$; $r := ||F_a^{-1}X||$; end;
  end;

Table 1: The least norm selector s_a for $(\mathcal{W}, \mathcal{F})$

input: $M \in \mathbb{M}(\mathbb{Z})$, $u \in B \times A^{\mathbb{N}}$; output: $v \in B \times A^{\mathbb{N}}$;
variables $X \in \mathbb{M}(\mathbb{Z}) \cup \mathbb{I}(\mathbb{Z})$, $s \in A \cup B \cup \{\lambda\}$, $n, m \in \mathbb{N}$;
begin
  $X := MW_{u_0}$; $n := 1$; $m := 0$;
  repeat
    if $m = 0$ then $s := s_B(X)$ else $s := s_A(X)$;
    if $s = \lambda$ then begin $X := XF_{u_n}$; $n := n + 1$; end;
    else begin
      $v_m := s$
      if $m = 0$ then $X := W_{v_0}^{-1}X$ else $X := F_{v_m}^{-1}X$;
      $m := m + 1$; end;
    end;
  end;

Table 2: The unary algorithm with selectors s_B, s_A.

with the smallest norm of $F_a^{-1}X$ provided $|X| < \tau$ and λ otherwise (see Table 1). If τ is chosen sufficiently small, then the overlaps of $\mathcal{W}$ are used effectively. An algorithm which computes a transformation with the use of selectors is in Table 2.

Theorem 15. *If $(\mathcal{W}, \mathcal{F})$ is redundant interval number system, then for each $M \in \mathbb{M}(\mathbb{Z})$, $u \in B \times A^{\mathbb{N}}$ the unary algorithm computes (in infinite time) $v \in B \times A^{\mathbb{N}}$ with $\Phi(v) = M(\Phi(u))$.*

Proof: For each m there exists n_m such that the algorithm computes a path with source $(M, 0, 0)$ and label $u_{[0,n_m]}/v_{[0,m]}$. Since $\overline{MW_{u_{[0,n_m]}}} \subseteq \overline{W_{v_{[0,m)}}}$ and both $\Phi(v)$ and $M\Phi(u)$ belong to $\overline{W_{v_{[0,m]}}}$, we get $\Phi(v) = M\Phi(u)$. □

8. Singular transformations and intervals

Besides orientation-preserving transformations with positive determinant, we consider orientation-reversing transformations with negative determinant, **singular** transformations with zero determinant but positive norm, and the **zero** transformation $M = (\frac{0}{0}, \frac{0}{0})$. If $\det(M) \neq 0$, then M is **regular**. Singular

transformations $\{M \in \mathrm{PL}(\mathbb{R},3) :\ \det(\mathrm{M}) = 0\}$ form a quadric in the three-dimensional projective space. A singular transformation has an **unstable point** $\mathbf{u}(M) \in \overline{\mathbb{R}}$ and a **stable point** $\mathbf{s}(M) \in \overline{\mathbb{R}}$ such that $M(x) = \mathbf{s}(M)$ for each $x \neq \mathbf{u}(M)$. If $\mathbf{s}(M) = s$ and $\mathbf{u}(M) = u$, then $M = (\frac{s_0 u_1}{s_1 u_1}, \frac{-s_0 u_0}{-s_1 u_0})$. For $x = \mathbf{u}(M)$ we have $Mx = \frac{0}{0}$. We interpret this fact in the sense that $M(x)$ is arbitrary. Each transformation $M = (\frac{a}{c}, \frac{b}{d})$ defines a closed graph (relation)

$$\widetilde{M} = \{(x,y) \in \overline{\mathbb{R}}^2 :\ (ax_0 + bx_1)y_1 = (cx_0 + dx_1)y_0\}.$$

If M is singular then $\widetilde{M} = (\overline{\mathbb{R}} \times \{\mathbf{s}(M)\}) \cup (\{\mathbf{u}(M)\} \times \overline{\mathbb{R}})$. If M is zero, then $\widetilde{M} = \overline{\mathbb{R}} \times \overline{\mathbb{R}}$. The operation of inversion $(\frac{a}{c}, \frac{b}{d})^{-1} = (\frac{d}{-c}, \frac{-b}{a})$ is used to singular transformations as well. If M is singular, then $\mathbf{s}(M^{-1}) = \mathbf{u}(M)$, $\mathbf{u}(M^{-1}) = \mathbf{s}(M)$, and MM^{-1} is the zero transformation. By a simple verification we get

Proposition 16. *If M is a singular transformation and F is a regular transformation, then both MF and FM are singular and*

$$\mathbf{s}(MF) = \mathbf{s}(M),\ \mathbf{s}(FM) = F(\mathbf{s}(M)),\ \mathbf{u}(FM) = \mathbf{u}(M),\ \mathbf{u}(MF) = F^{-1}(\mathbf{u}(M).$$

The image $M(\overline{I})$ of a regular interval I is defined by $M(\overline{I}) = \{y \in \overline{\mathbb{R}} :\ \exists x \in \overline{I} :\ (x,y) \in \widetilde{M}\}$. For $I = (a,b)$ we get

$$M(\overline{I}) = \begin{cases} \overline{(M(a), M(b))} & \text{if} \quad \det(M) > 0 \\ \overline{(M(b), M(a))} & \text{if} \quad \det(M) < 0 \\ \{\mathbf{s}(M)\} & \text{if} \quad \det(M) = 0,\ \mathbf{u}(M) \notin \overline{I} \\ \overline{\mathbb{R}} & \text{if} \quad \det(M) = 0,\ \mathbf{u}(M) \in \overline{I} \\ \overline{\mathbb{R}} & \text{if} \quad M = 0 \end{cases}$$

For a singular interval $I = (a,b) \in \mathrm{PL}(\mathbb{R},3)$ with $\det(a,b) = 0$ we define I° and $\overline{I}$ by the same formulas which we have used for regular intervals, adopting the convention $\mathrm{sgn}(\frac{0}{0}) = 0$.

Proposition 17. *Let I be a singular interval.*
1. If $\mathrm{sgn}(\mathbf{u}(I)) < 0$ then $I^\circ = \emptyset$, $\overline{I} = \{\mathbf{s}(I)\}$, $|I| = 0$.
2. If $\mathrm{sgn}(\mathbf{u}(I)) = 0$ then $I^\circ = \emptyset$, $\overline{I} = \overline{\mathbb{R}}$, $|I|$ is not defined.
3. If $\mathrm{sgn}(\mathbf{u}(I)) > 0$ then $I^\circ = \overline{\mathbb{R}} \setminus \{\mathbf{s}(I)\}$, $\overline{I} = \overline{\mathbb{R}}$, $|I| = 1$.
4. If J is a regular interval, then $\overline{I} \subseteq \overline{J}$ iff $\mathrm{sgn}(J^{-1}I) \geq 0$.

Proof: 1,2,3: If $I = (\frac{a}{b}, \frac{\lambda a}{\lambda b})$, then $\mathrm{sgn}(\mathbf{u}(I)) = -\mathrm{sgn}(\lambda)$, $I^{-1}x = \frac{-\lambda(ax_1 - bx_0)}{ax_1 - bx_0}$, and the statements follow. If $I = (\frac{\lambda c}{\lambda d}, \frac{c}{d})$, the proof is similar.
4. $\overline{I} \subseteq \overline{J}$ iff $\mathrm{sgn}(\mathbf{u}(I)) < 0$ and $\mathbf{s}(I) \in \overline{J}$ iff $\mathrm{sgn}(J^{-1}I) \geq 0$. □

Proposition 18. *Let M be a transformation, let I be an interval and assume that either M or I is regular. Then $M(\overline{I}) = \overline{MI}$.*

Proof: Let M be regular and I singular. If $\mathrm{sgn}(\mathbf{u}(MI)) = \mathrm{sgn}(\mathbf{u}(I)) < 0$, then $\overline{I} = \{\mathbf{s}(I)\}$ and $M(\overline{I}) = \{M(\mathbf{s}(I))\} = \{\mathbf{s}(MI)\} = \overline{MI}$. If $\mathrm{sgn}(\mathbf{u}(I)) \geq 0$, then $M(\overline{I}) = \overline{\mathbb{R}} = \overline{MI}$.
If M is singular and I is regular then $\overline{MI} = \{\mathbf{s}(MI)\}$ iff $\mathrm{sgn}(I^{-1}(\mathbf{u}(M)))\} = \mathrm{sgn}(\mathbf{u}(MI))\} < 0$ iff $\mathbf{u}(M) \notin \overline{I}$ iff $M(\overline{I}) = \{\mathbf{s}(M)\}$. Since $\mathbf{s}(MI) = \mathbf{s}(M)$, we get $M(\overline{I}) = \overline{MI}$ in this case. If $\mathrm{sgn}(\mathbf{u}(MI)) \geq 0$, then $M(\overline{I}) = \overline{\mathbb{R}} = \overline{MI}$. $\square$

9. Tensors

Binary arithmetical operations like addition or multiplication are obtained from bilinear functions $T : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$. While a linear function $M : \mathbb{R}^2 \to \mathbb{R}^2$ is a 1-contravariant and 1-covariant tensor, a bilinear function is a 1-contravariant and 2-covariant tensor given by $T(x,y)_k = \sum_{i=0}^{1}\sum_{j=0}^{1} T_{kij}x_iy_j$ (see e.g., Bishop and Goldberg [1]). Such a tensor determines a function $T : \overline{\mathbb{R}} \times \overline{\mathbb{R}} \to \overline{\mathbb{R}} \cup \{\frac{0}{0}\}$ defined by

$$T(x,y) = \frac{T_{000}x_0y_0 + T_{001}x_0y_1 + T_{010}x_1y_0 + T_{011}x_1y_1}{T_{100}x_0y_0 + T_{101}x_0y_1 + T_{110}x_1y_0 + T_{111}x_1y_1}.$$

A nonzero multiple of a tensor defines the same function on $\overline{\mathbb{R}} \times \overline{\mathbb{R}}$, so tensors are conceived as points of the projective space $\mathrm{PL}(\mathbb{R}, 7)$. They are sometimes written as (4×2)-matrices $T = (\frac{T_{000}}{T_{100}}, \frac{T_{001}}{T_{101}}, \frac{T_{010}}{T_{110}}, \frac{T_{011}}{T_{111}})$. For a tensor T and vectors $x, y, z \in \overline{\mathbb{R}}$ we have matrices zT, $T_\star x$, $T^\star y$ obtained from T by fixing a variable:

$$(T_\star x)_{kj} = \sum_i T_{kij}x_i,\ (T^\star y)_{ki} = \sum_j T_{kij}y_j,\ (zT)_{ij} = \sum_k z_k T_{kij}.$$

For matrices I, J, K we define tensors T_*I, T^*J and KT by

$$(T_*I)_{kij} = \sum_p T_{kpj}I_{pi},\ (T^*J)_{kij} = \sum_q T_{kiq}J_{qj},\ (KT)_{kij} = \sum_r K_{kr}T_{rij}.$$

Then $(T_*I)_\star x = T_\star(Ix)$, $(T^*J)^\star y = T^\star(Jy)$. The operations with the first and second argument of a tensor commute, so we adopt notations

$$\begin{aligned} T(x,y) &= (T_\star x)y = (T^\star y)x, \\ T(x,J) &= (T_\star x)J = (T^*J)_\star x, \\ T(I,y) &= (T^\star y)I = (T_*I)^\star y, \\ T(I,J) &= (T_*I)^*J = (T^*J)_*I. \end{aligned}$$

The multiplication from the left commutes with the multiplication from the right, so e.g., $K(T_*I) = (KT)_*I = KT_*I$. For a matrix I denote its left and right columns by I_{-0}, I_{-1}, so $(I_{-j})_i = I_{ij}$. Similarly for a tensor T we denote by T_{k--}, T_{-i-}, T_{--j} the **marginal matrices** obtained from T by fixing a coordinate, and T_{-ij}, T_{k-j}, T_{ki-} **marginal vectors** obtained by fixing two coordinates. A simple algebra shows that the tensor $T(I,J)$ consists of T-images of the endpoints of I and J:

Proposition 19. *For a tensor T and matrices I, J we have*

$$T(I,J)_{-i-} = T(I_{-i}, J),\ T(I,J)_{--j} = T(I, J_{-j}),\ T(I,J)_{-ij} = T(I_{-i}, J_{-j}).$$

The image of intervals I, J by a tensor T is defined by

$$\begin{aligned} T(\overline{I}, \overline{J}) &= \{z \in \overline{\mathbb{R}} :\ \exists x \in \overline{I}, \exists y \in \overline{J}, (x,y,z) \in \widetilde{T}\}, \text{ where} \\ \widetilde{T} &= \{(x,y,z) \in \overline{\mathbb{R}}^3 :\ T(x,y)_0 z_1 = T(x,y)_1 z_0\}. \end{aligned}$$

The sign of a tensor is defined similarly as the sign of a matrix: it is nonnegative if there exists nonzero λ such that all λT_{kij} are nonnegative.

Definition 20. *We say that T is a* **regular tensor***, if for each $x, y, z \in \overline{\mathbb{R}}$, the matrices zT, $T_\star x$, $T^\star y$ are nonzero.*

A tensor is regular iff its pairs of marginal matrices are linearly independent, i.e., if $T_{0--} \neq T_{1--}$, $T_{-0-} \neq T_{-1-}$ and $T_{--0} \neq T_{--1}$ are different points of the projective space $\mathrm{PL}(\mathbb{R}, 3)$. Examples of regular tensors are $(\frac{1}{0}, \frac{0}{0}, \frac{0}{0}, \frac{0}{1})$ (multiplication), $(\frac{0}{0}, \frac{1}{0}, \frac{1}{0}, \frac{0}{1})$ (addition), or $(\frac{0}{0}, \frac{1}{0}, \frac{0}{1}, \frac{0}{0})$ (division).

Proposition 21. *If T is a regular tensor and M is a regular transformation or interval, then MT, T_*M and T^*M are regular tensors.*

Proof: $(T_*M)_\star x = T_\star(Mx)$. □

Theorem 22. *For a regular tensor T and regular intervals I, J, K we have $T(\overline{I}, \overline{J}) \subseteq \overline{K}$ iff $\operatorname{sgn}(K^{-1}T(I,J)) \geq 0$.*

Proof: Assume that $\operatorname{sgn}(K^{-1}T(I,J)) \geq 0$, $x \in \overline{I}$, $y \in \overline{J}$, and $x, y, z \in \widetilde{T}$. For $u = I^{-1}x$ we have $\operatorname{sgn}(u) \geq 0$ and $x = Iu$, so

$$(T_\star x)J = (T_\star(Iu))J = ((T_*I)_\star u)J = ((T_*I)^*J)_\star u = T(I,J)_\star u.$$

Since $y \in \overline{J}$, $(y,z) \in \overline{T_\star x}$, and $T_\star x$ is nonzero, we get by Proposition 18 $z \in (T_\star x)(\overline{J}) = \overline{(T_\star x)J} \subseteq \overline{K}$, so $z \in \overline{K}$.
Conversely if $T(\overline{I}, \overline{J}) \subseteq \overline{K}$, then

$$\overline{(T_\star I_{-i})J} = (T_\star I_{-i})(\overline{J}) \subseteq \overline{K},\ \overline{(T^\star J_{-j})I} = (T^\star J_{-j})(\overline{I}) \subseteq \overline{K}.$$

Since K is regular, we get by Proposition 17

$$\begin{aligned} \operatorname{sgn}(K^{-1}T(I,J)_{-i-}) &= \operatorname{sgn}(K^{-1}(T_\star I_{-i})J) \geq 0, \\ \operatorname{sgn}(K^{-1}T(I,J)_{--j}) &= \operatorname{sgn}(K^{-1}(T^\star J_{-j})I) \geq 0. \end{aligned}$$

It follows $\operatorname{sgn}(K^{-1}T(I,J)) \geq 0$. □

Definition 23. *Given an interval number system $(\mathcal{W}, \mathcal{F})$ we consider labelled* **binary graph** *whose vertices are (X, n, m) where $X \in \mathrm{PL}(\mathbb{R}, 7)$, $n, m \in \{0, 1\}$. The labelled edges are*

$$\begin{aligned}
(X,0,0) &\overset{b/c/\lambda}{\longrightarrow} (X(W_b, W_c), 1, 0),\\
(X,1,m) &\overset{a/\lambda/\lambda}{\longrightarrow} (X_* F_a, 1, m),\\
(X,1,m) &\overset{\lambda/a/\lambda}{\longrightarrow} (X^* F_a, 1, m),\\
(X,1,0) &\overset{\lambda/\lambda/b}{\longrightarrow} (W_b^{-1} X, 1, 1) \quad \textit{if} \quad \mathrm{sgn}(W_b^{-1} X) \geq 0,\\
(X,1,1) &\overset{\lambda/\lambda/a}{\longrightarrow} (F_a^{-1} X, 1, 1) \quad \textit{if} \quad \mathrm{sgn}(F_a^{-1} X) \geq 0.
\end{aligned}$$

Proposition 24. *If $(T,0,0) \overset{u/v/w}{\longrightarrow} (X,1,1)$ is a path in the binary graph, then $X = W_w^{-1} T(W_u, W_v)$ and $T(\overline{W_u}, \overline{W_v}) \subseteq \overline{W_w}$. If $(u, v, w) \in (B \times A^{\mathbb{N}})^3$ is the label of an infinite path with the source $(T, 0, 0)$, then $T(\Phi(u), \Phi(v)) = \Phi(w)$.*

Theorem 22 yields a simple algorithm for the computation of binary operations based on the search of a path in the binary graph. However, the algorithm is not guaranteed to compute an infinite output $w \in B \times A^{\mathbb{N}}$. This happens e.g., if we try to compute indefinite expressions like $0 \cdot \infty$.

10. Rational functions

A rational function $F : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ of degree q is a function of the form

$$F(x) = \frac{F_{00} x_0^q + F_{01} x_0^{q-1} x_1 + \cdots + F_{0q} x_1^q}{F_{10} x_0^q + F_{11} x_0^{q-1} x_1 + \cdots + F_{1q} x_1^q},$$

where $\frac{F_{00}}{F_{10}} \neq \frac{0}{0}$ and $\frac{F_{0q}}{F_{1q}} \neq \frac{0}{0}$. Rational functions are obtained from tensors. A tensor T is **symmetric** if $T_{ijk} = T_{ikj}$ for each i, j, k. For a rational function F of degree 2 there exists a symmetric tensor T such that $F(x) = T(x, x)$. For each interval I we have $F(\overline{I}) = \{F(x) : \ x \in \overline{I}\} \subseteq T(\overline{I}, \overline{I})$, so $F(\overline{I}) \subseteq \overline{J}$ provided $\mathrm{sgn}(J^{-1} T(I, I)) \geq 0$. If T is symmetric, then $T(I, I)$ is symmetric as well, so the algorithm which computes F can be performed with symmetric tensors.

This procedure generalizes to rational functions of higher degrees. A rational function of order q is obtained from a symmetric tensor $T_{k, i_1, \ldots, i_q}$ of order $q+1$. If F is a rational function of order q and M is a transformation, then both compositions $F \circ M$ and $M \circ F$ are rational functions of order q as well. The coefficients of F form a $((q+1) \times 2)$-matrix and the composition MF is the matrix of $M \circ F$. We obtain a simple criterion for the inclusion:

Theorem 25. *Let $F : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ be a rational function and I, J intervals. If $\mathrm{sgn}(J^{-1} F \circ I) \geq 0$, then $F(\overline{I}) \subseteq \overline{J}$.*

The algorithm for the computation of a rational function has the same form as the unary algorithm in Table 2. The only difference is that the products XW_b, XF_a should be replaced by the compositions $X \circ W_b$ and $X \circ F_a$.

system:	1	2	3	4
linear:	0.02	0.50	1.29	0.51
quadratic:	0.10	0.99	2.02	0.99
cubic:	0.28	0.98	1.98	1.00

Table 3: The estimates of the quotient $q = \lim_{k\to\infty} \log_2 ||X_k||/k$ during the computations of fractional linear function $(3x+1)/(x+2)$, quadratic function x^2 and cubic function x^3. The number of the system refers to Examples 1,2,3,4.

11. Numerical results

The time complexity of the unary algorithm depends on the rate of growth of the state matrix X. If its entries are expressed in the positional binary system, then the length of this representation (the bit length of X) is of the order $\log_2 ||X||$. The multiplication of X by a matrix M requires $\log_2 ||X|| \cdot \log_2 ||M||$ elementary operations on their binary representations. Thus there exists a constant $C > 0$ such that each step of the algorithm requires at most $C \cdot \log_2 ||X||$ elementary operations. If X_k are states of a path computed by the unary algorithm, then the time of the computation of n steps is of the order $C \sum_{k=0}^{n-1} \log_2 ||X_k||$.

In Delacourt and Kůrka [2] we have shown that for modular Möbius number systems, whose transformations have unit determinant, the norm of the state matrix remains bounded during the computation, so the unary algorithm has linear time complexity. This result applies to our Example 1 whose transformations have unit determinant as well. This system, however, has the disadvantage of slow convergence and nonuniform length of its intervals.

In general we have $||XF|| \leq ||X|| \cdot ||F||$, so $\log_2 ||X_k|| \leq qk$, where X can be a transformation, tensor or rational function, and $q = \max\{||F_a|| : \ a \in A\}$. It follows that the time complexity of the computation of path of length n is bounded by $Cqn^2/2$. In some systems, however, the quotient q can be smaller, since the entries of X cancel by a common factor which divides $\det(F_a)$ (see Kůrka and Delacourt [8]). The most effective systems seem to be those whose determinant is a power of 2, like the nonredundant uniform $\frac{1}{2}$-system or redundant uniform $\frac{2}{4}$-system. The results of some numerical experiments can be seen in Table 3 which displays the estimates of the quotient $q = \lim_{k\to\infty} \log_2 ||X_k||/k$ for several interval number systems with the linear, quadratic and cubic functions. We can see that for the systems whose transformations have determinant two we get smaller quotient q.

Acknowledgment. The research was supported by the Czech Science Foundation research project GAČR 13-03538S.

References

[1] R. L. Bishop and S.I.Goldberg. *Tensor analysis on manifolds.* Dover publications, New York, 1980.

[2] M. Delacourt and P. Kůrka. Finite state transducesrs for modular Möbius number systems. In B. Rovan, V. Sassone, and P.Widmayer, editors, *MFCS 2012*, volume 7464 of *Lecture Notes in Computer Science*, Berlin, 2012. Springer-Verlag.

[3] R. W. Gosper. Continued fractions arithmetic. *Unpublished manuscript*, 1977. http://www.tweedledum.com/rwg/cfup.htm.

[4] P. Kornerup and D. W. Matula. An algorithm for redundant binary bit-pipelined rational arithmetic. *IEEE Transactions on Computers*, 39(8):1106–1115, August 1990.

[5] P. Kůrka. Möbius number systems with sofic subshifts. *Nonlinearity*, 22:437–456, 2009.

[6] P. Kůrka. Stern-Brocot graph in Möbius number systems. *Nonlinearity*, 25:57–72, 2012.

[7] P. Kůrka and A.Kazda. Möbius number systems based on interval covers. *Nonlinearity*, 23:1031–1046, 2010.

[8] P. Kůrka and M.Delacourt. The unary arithmetical algorithm in bimodular number system. *IEEE Transactions on computers*, 2013. to appear.

[9] M. Niqui. Exact real arithmetic on the Stern-Brocot tree. *J. Discrete Algorithms*, 5(2):356–379, 2007.

[10] P. J. Potts. *Exact real arithmetic using Möbius transformations*. PhD thesis, University of London, Imperial College, London, 1998.

[11] J. E. Vuillemin. Exact real computer arithmetic with continued fractions. *IEEE Transactions on Computers*, 39(8):1087–1105, August 1990.