On the algorithmic complexity of some numeration-related problems

Dávid Bóka
bdavid1001@gmail.com
Eötvös Loránd University (Budapest)
Faculty of Informatics

Péter Burcsi
bupe@inf.elte.hu
Eötvös Loránd University (Budapest)
Faculty of Informatics

M. János Uray
uray.janos@inf.elte.hu
Eötvös Loránd University (Budapest)
Faculty of Informatics

Abstract

In this paper we give explicit upper bounds on the algorithmic complexity of some problems related to matrix-based numeration systems. We consider different types of deterministic algorithms, and compare their worst-case time and space complexity. We also analyze probabilistic algorithms, and examine the question whether they can serve as a reliable alternative.

1 Introduction

We consider the complexity of some algorithms related to a specific type of numeration systems, namely matrix-based numeration systems. Our main results are upper bounds on the running time of two problems: the decision of the finiteness property and finding all periodic orbits in a matrix-based numeration system. We give our results in terms of the parameters of the input: the dimension, the size of the entries in the matrix, the norm of the digits, and the so-called Jordan condition number of the matrix.

We consider several versions of a deterministic method and also analyze the behavior of a randomized algorithm. Our empirical observations suggest that in many cases, following the orbits of randomly chosen points quickly give all cycles. We therefore analyze on various systems whether this observation can be justified and used for bounding the soundness error of a randomized algorithm (i.e. the probability that the algorithm fails to detect the existence of non-trivial cycles).

The paper is built up as follows: Section 2 covers the most important definitions and previous results on matrix-based numeration systems, and introduces the deterministic algorithm. Section 3 calculates the running time of different versions of the algorithm, along with some other properties and results which are used in the calculation. Section 4 is devoted to randomized algorithms. Finally, we give future directions and open problems in Section 5.

2 Definitions and prior results

In this section, we give the main definitions needed to formulate the algorithmic problems considered in the rest of the paper.

Definition 2.1. *Let n be a positive integer, let $M \in \mathbb{Z}^{n\times n}$ with $|\det M| \geq 2$, and $D \subseteq \mathbb{Z}^n$ with $0 \in D$ a complete residue system mod M (i.e. $|D| = |\det M|$ and for all distinct pair of elements $d, d' \in D$, we have $M^{-1}(d - d') \notin \mathbb{Z}^n$). To such a pair (M, D), we associate a discrete dynamical system given by the map $\tau\colon \mathbb{Z}^n \to \mathbb{Z}^n$ defined as*

$$\tau(v) = M^{-1}(v - d) \quad \text{with the unique } d \in D \text{ s.t. } \tau(v) \in \mathbb{Z}^n. \tag{1}$$

The pair (M, D) together with τ is called a matrix-based numeration system*, where M is called the* base *and the elements of D are called the* digits*. We say that the system has the* finiteness property *(it is a* generalized number system*) if for all $v \in \mathbb{Z}^n$, there exists a non-negative integer k such that $\tau^k(v) = 0$ (here, τ^k denotes k-fold iteration).*

Matrix based numeration systems have been considered in several papers, e.g. in [6, 9, 11, 12]. The following theorem also appears in the literature, see the survey paper [2] for details.

Theorem 2.2. *Let (M, D) be a matrix-based numeration system. It has the finiteness property if and only if every non-zero vector $v \in \mathbb{Z}^n \setminus \{0\}$ is uniquely representable as a finite sum*

$$v = \sum_{j=0}^{k} M^j d_j$$

with $k \geq 0$, $d_j \in D$ and $d_k \neq 0$.

The theorem above explains the term *generalized number system*: ordinary (binary, decimal etc.) number systems have similar unique representation for all positive integers.

Example 2.3. *Let $n = 2$, $M = \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}$ and $D = \{(0, 0), (1, 0)\}$. It is proved in an even more general form in [5] that this system has the finiteness property. Identifying $\mathbb{Z}^2$ with the Gaussian integers, multiplication by M represents multiplication by $-1+i$. So Theorem 2.2 states that for all $a, b \in \mathbb{Z}$ (not both zero), we have a unique representation*

$$a + bi = \sum_{j=0}^{k} (-1 + i)^j d_j$$

where $d_j \in \{0, 1\}$ and $d_k = 1$.

No short characterization is known for numeration systems having the finiteness property. We have the following necessary condition.

Theorem 2.4. *[14] If (M, D) has the finiteness property, then M is expansive, i.e. all eigenvalues of M lie outside the closed unit disk.*

In the present paper we analyze the algorithmic complexity of the following problem: given (M, D), decide if the numeration system has the finiteness property. Our analysis also applies to a more general problem: if the matrix is expansive, find all cycles of τ.

On a sufficiently large scale, the dynamics of τ and that of the linear map M^{-1} are similar, so the expansivity of M implies that τ behaves like a contraction. This means that there exists a finite region around the origin such that all orbits of τ eventually lie in this region, and therefore τ has finitely many cycles and every orbit is eventually periodic.

The high-level description of the bounding box algorithms examined in this paper is the following: give an easily calculated bounding box on this finite region, i.e. give a bound B on it in some vector norm, then exhaustively search through all points having size smaller than B and explore their orbits.

3 Running time of the deterministic algorithm

3.1 Bounds on the expansivity gap

The size of the bounding box depends on (among others) how contractive M^{-1} is. The closer the eigenvalues of M are to the unit disk, the larger the bounding box might be, so to give a bound on the latter, we need to give a lower bound on the expansivity gap of M, i.e. the distance between the eigenvalues and the unit disk. For that, we examine the roots of the characteristic polynomial $\det(M - xI_n)$. We present the following result for the roots of any expansive polynomial:

Lemma 3.1. *Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$ $(n \geq 2, a_i \in \mathbb{Z}, a_n \neq 0)$ be an expansive polynomial, i.e. for all roots x_i (either real or complex) we have: $|x_i| > 1$. Then all $|x_i| > 1 + \varepsilon$ with:*

$$\varepsilon := \frac{1}{2^{\binom{n}{2}}|a_0|^{n-1} + 1}. \tag{2}$$

Proof. It follows from Liouville's inequality [15, Prop. 3.14] that for any algebraic number $\alpha \neq \pm 1$:

$$|\alpha \pm 1| \geq \frac{1}{2^{\deg \alpha - 1} M(\alpha)}, \tag{3}$$

where $M(\alpha)$ is the Mahler measure of α: $M(\alpha) := |a| \prod_{i=1}^{n} \max\{1, |\alpha_i|\}$, where α_i are the conjugates of α and a is the leading coefficient of the minimal polynomial f_α. If f_α is expansive, then $M(\alpha)$ is simply the magnitude of the constant term of f_α.

If α is a real root of f, then (3) directly gives a much better ε than required (since $\deg \alpha \leq n$ and $M(\alpha) \leq |a_0|$ with equality if f is irreducible). If α is a non-real root of f, we will apply (3) to $\alpha\overline{\alpha} \in \mathbb{R}$, which is the root of the following polynomial of degree n^2 [3, p. 159]:

$$\tilde{f}(x) = \mathrm{res}_y\left(f(y), f\left(\frac{x}{y}\right) y^n\right) = a_n^{2n} \prod_{i=1}^{n} \prod_{j=1}^{n} (x - \alpha_i \alpha_j).$$

This polynomial has a factor of degree n with all α_i^2 as roots:

$$f_1(x) = \mathrm{res}_y\left(f(y), x - y^2\right) = a_n^2 \prod_{i=1}^{n} (x - \alpha_i^2),$$

and the remaining is a perfect square, so $\tilde{f}(x) = f_1(x)f_2^2(x)$, where the $\deg(f_2) = \binom{n}{2}$. The $\alpha\overline{\alpha}$ is the root of the latter, so $\deg(\alpha\overline{\alpha}) \leq \binom{n}{2}$, and $M(\alpha\overline{\alpha}) \mid a_0^{n-1}$, which is the constant term of $f_2(x)$ (because the constant term of $\tilde{f}(x)$ and $f_1(x)$ are $(-1)^n a_0^{2n}$ and $(-1)^n a_0^2$, respectively). Applying (3) to $\alpha\overline{\alpha}$ gives:

$$|\alpha|^2 \geq 1 + \frac{1}{2^{\binom{n}{2}-1}|a_0|^{n-1}},$$

from which the statement (2) follows. □

3.2 The size of the bounding box

We give a bound on B, the size of the bounding box, in terms of the parameters of (M, D), including the following matrix property:

Definition 3.2. *For any matrix $A \in \mathbb{R}^{n\times n}$, define its Jordan condition number $K(A)$ as follows: $K(A) := \inf_T \|T\|\|T^{-1}\|$ where $A = TJT^{-1}$ is a Jordan decomposition of A, and $\|\cdot\|$ is any natural matrix norm (i.e. induced by a vector norm).*

The Jordan condition number is defined e.g. in [16, p. 85]. It should not be confused with the regular condition number, which is simply $\|A\|\|A^{-1}\|$, see e.g. in [13, 4].

For integer matrices of fixed size with entries bounded by some constant, their Jordan condition number has a finite constant upper bound (dependent only on the size and the bound on the entries), since there are only finite number of such matrices. However, we have not found such an explicit bound. Therefore, we present our results using this number as a parameter.

Lemma 3.3. *Let (M, D) be a matrix-based numeration system with expansive $M \in \mathbb{Z}^{n\times n}$. Then there exists a bound B such that for any $v \in \mathbb{Z}^n$ which is outside of the box $\|v\|_1 > B$, the function τ defined in (1) brings v into this box in a finite number of steps, i.e. there exists a positive integer k_0 such that for all $k \geq k_0$: $\|\tau^k(v)\|_1 \leq B$. Furthermore, the following value for B will do:*

$$B := en\left(2^{\binom{n}{2}}|D|^{n-1} + 2\right)^n K_1(M^{-1}) \max_{d\in D} \|d\|_1. \tag{4}$$

Proof. Since M is expansive, the spectral radius $\rho(M^{-1})$ (the maximal size of the eigenvalues of M^{-1}) is less than 1. Therefore, according to the well-known properties of the spectral radius, there exists some norm $\|\cdot\|$ such that $\rho(M^{-1}) \leq \|M^{-1}\| < 1$. Let $B' := \max_{d\in D} \|d\|/(1 - \|M^{-1}\|)$, then:

$$\begin{aligned}\|\tau(v)\| = \|M^{-1}(v-d)\| \leq \|M^{-1}\|(\|v\| + \|d\|) \leq \|M^{-1}\| \left(\|v\| + (1 - \|M^{-1}\|)B'\right) \leq \\ \leq \|M^{-1}\|(2 - \|M^{-1}\|)\max(\|v\|, B') = \left(1 - (1 - \|M^{-1}\|)^2\right)\max(\|v\|, B').\end{aligned}$$

Let q be the factor of $\max(\|v\|, B')$, which is between $0 < q < 1$ since $0 < \|M^{-1}\| < 1$. If $\|v\| \geq B'$, then the above means that $\|\tau(v)\| \leq q\|v\|$, otherwise, if $\|v\| \leq B'$, then $\|\tau(v)\| \leq B'$. It follows that for all $v \in \mathbb{Z}^n$, $\|\tau^k(v)\| \leq B'$ for sufficiently large k.

This B' looks like the required B in the statement, but the problem is that it is a bound on the unknown (and maybe complicated) norm $\|\cdot\|$, not on a simple one like

$\|\cdot\|_1$ as needed. But it is a well-known theorem that any two norms in $\mathbb{R}^n$ are equivalent, i.e. there exists $c, C > 0$ such that $c\|v\|_1 \le \|v\| \le C\|v\|_1$. It follows from this and the definition of B' that $\|v\| \le B' \implies \|v\|_1 \le B$ if B is the following or greater:

$$B := \frac{C}{c}\frac{\max_{d\in D}\|d\|_1}{1-\|M^{-1}\|}. \tag{5}$$

To calculate the constants c and C explicitly, we need to construct a norm that satisfies $\|M^{-1}\| < 1$. Let δ be any positive number in $0 < \delta < 1 - \rho(M^{-1})$. Write M^{-1} in the form $M^{-1} = T^{-1}JT$ with J being its Jordan normal form, and define J_δ to be J except that the off-diagonal 1's are replaced by δ. It can be written as $J_\delta = D_\delta J D_\delta^{-1}$ with a diagonal matrix D_δ, whose structure is illustrated below:

$$J = \left(\begin{array}{cccc|ccc|c}
\frac{1}{\lambda_1} & 1 & & & & & & \\
 & \frac{1}{\lambda_1} & 1 & & & & & \\
 & & \frac{1}{\lambda_1} & 1 & & & & \\
 & & & \frac{1}{\lambda_1} & & & & \\
\hline
 & & & & \frac{1}{\lambda_2} & 1 & & \\
 & & & & & \frac{1}{\lambda_2} & 1 & \\
 & & & & & & \frac{1}{\lambda_2} & \\
\hline
 & & & & & & & \frac{1}{\lambda_3}
\end{array}\right), \quad
J_\delta = \left(\begin{array}{cccc|ccc|c}
\frac{1}{\lambda_1} & \delta & & & & & & \\
 & \frac{1}{\lambda_1} & \delta & & & & & \\
 & & \frac{1}{\lambda_1} & \delta & & & & \\
 & & & \frac{1}{\lambda_1} & & & & \\
\hline
 & & & & \frac{1}{\lambda_2} & \delta & & \\
 & & & & & \frac{1}{\lambda_2} & \delta & \\
 & & & & & & \frac{1}{\lambda_2} & \\
\hline
 & & & & & & & \frac{1}{\lambda_3}
\end{array}\right), \quad
D_\delta = \left(\begin{array}{cccc|ccc|c}
\delta^{-1} & & & & & & & \\
 & \delta^{-2} & & & & & & \\
 & & \delta^{-3} & & & & & \\
 & & & \delta^{-4} & & & & \\
\hline
 & & & & \delta^{-1} & & & \\
 & & & & & \delta^{-2} & & \\
 & & & & & & \delta^{-3} & \\
\hline
 & & & & & & & \delta^{-1}
\end{array}\right).$$

We define our norm as follows: $\|v\|_\delta := \|D_\delta T v\|_1$ with the compatible $\|A\|_\delta := \|D_\delta T A T^{-1} D_\delta^{-1}\|_1$. It meets our requirements:

$$\|M^{-1}\|_\delta = \|D_\delta T M^{-1} T^{-1} D_\delta^{-1}\|_1 = \|J_\delta\|_1 \le \rho(M^{-1}) + \delta < 1.$$

To calculate B from (5), we need the equivalence constants between $\|\cdot\|_\delta$ and $\|\cdot\|_1$. It is easy to show that:

$$\delta^{-1}\|T^{-1}\|_1^{-1}\|v\|_1 \le \|v\|_\delta \le \delta^{-m}\|T\|_1\|v\|_1,$$

where m is the size of the largest Jordan block in J. Putting these constants into (5) and applying $\|M^{-1}\|_\delta \le \rho(M^{-1}) + \delta$ gives a suitable value for B:

$$B := \frac{\|T\|_1\|T^{-1}\|_1 \max\limits_{d\in D}\|d\|_1}{\delta^{m-1}\left(1-\rho(M^{-1})-\delta\right)}.$$

This B works for any T for which $M^{-1} = T^{-1}JT$ holds, which implies that also $\inf_T B$ works. To prove this, it is enough to notice that $\|\tau^k(v)\|_1$ is always an integer, so any B can be rounded down to the nearest integer, and so the infimum simplifies to a minimum. Therefore, $\|T\|_1\|T^{-1}\|_1$ in B can be replaced by its infimum $K_1(M^{-1})$, the Jordan condition number of M^{-1}.

To get an upper bound on $\rho(M^{-1})$, we apply Lemma 3.1 for the characteristic polynomial $p_M(x)$ of M. Since the constant term of $p_M(x)$ is $\det M$, and by the definition of the numeration system, $|\det M| = |D|$, we get that the size of the smallest eigenvalue of M is greater than $1 + 1/(N+1)$ with $N := 2^{\binom{n}{2}}|D|^{n-1}$. As M^{-1} has inverse eigenvalues, their maximum, i.e. the spectral radius $\rho(M^{-1})$ is less than $1 - 1/(N+2)$. This can be used to change the denominator of B to $\delta^{m-1}(1/(N+2) - \delta)$.

The last step is to choose an appropriate δ in $0 < \delta < 1/(N+2)$. If $m = 1$ (i.e. M is diagonalizable), then taking the infimum as $\delta \to 0$ gives $B = (N+2)K_1(M^{-1})\max_{d\in D}\|d\|_1$,

which is better than (4). Otherwise, if $m > 1$ (i.e. M is defective), then the denominator of B has maximum at $\delta = \frac{m-1}{m(N+2)}$ with the value $\frac{(1-1/m)^{m-1}}{m(N+2)^m}$. The sequence $(1-1/m)^{m-1}$ can be replaced by its infimum, $1/e$, which gives a slightly greater $B = em(N+2)^m K_1(M^{-1}) \max_{d \in D} \|d\|_1$, which, when substituting $m \leq n$, gives (4). $\square$

3.3 The overall running time

We examine the following two algorithms simultaneously:

- Decision algorithm: returns true if all points reach the origin, otherwise false.
- Classification algorithm: computes all cycles of τ.

The essence of both algorithms is to examine the orbits of all points inside the bounding box and find their cycle. The first algorithm stops as soon as any nontrivial cycle (i.e. any cycle not being the origin itself) is found, while the other always has to visit all points in the bounding box. Our complexity calculation however will be the same for the two cases, since in the worst case, the decision algorithm cannot stop earlier either.

From the complexity point of view, these algorithms are a series of iterations with τ, so their running time is calculated by multiplying the number of τ-invocations by the complexity of one such invocation. The preprocessing phase, e.g. for computing B, has negligible running time compared to the exhaustive search.

There are several methods for computing $\tau(v)$ of an integer vector v, see e.g. [10, Section 2.4] for a comprehensive overview. The difficulty lies in determining which digit is congruent to $v \mod M$. For simplicity, we assume that a single arithmetic operation on two integers can be performed in constant time. In typical numeration systems, the values encountered safely fit into a 32-bit machine word, so this is a reasonable assumption. With this, $\tau(v)$ can be computed in $O(n^2)$ time after a single preprocessing phase, for details see [10].

Let $\tau^\infty(v)$ denote the eventual cycle on the τ-orbit of v. There is a trade-off in orbit computation between the time and space complexity of an algorithm. We mention three approaches. The following table summarizes the space and time complexity of these methods in terms of L, the number of points in the bounding box, ℓ, the maximal length of the orbits, and n, the dimension of the space. After that, the methods are described individually.

Method	**Time** (# of τ-calls)	**Space** (# of ints)
1. Store the computed values for all points	L	$O(L)$
2. Store the current orbit only	$L\ell$	$O(\ell n)$
3. Floyd's Tortoise and Hare algorithm	$3L\ell$	$O(n)$

1. **Abundant memory:**

   We first consider the optimistic scenario where the amount of memory enables to store a little information on each point in the bounding box. This allows us to keep track of the already visited points and avoid examining them more than once. The algorithm proceeds by exhaustively traversing the orbits of all points in the

bounding box until an orbit hits an already visited point. Then there are two cases. First, if that point belongs to the same orbit, then a new cycle is found. If it is a nontrivial cycle and the decision algorithm is concerned, then it immediately returns with the result. Otherwise all visited points are marked as 'processed'. The classification algorithm also stores $\tau^\infty(v)$ on those points (by a pointer to the list of already found cycles). In the second case, if the orbit hits a previously processed point, then, without examining it further, all points along the new orbit can also be marked as processed and the value of the hit point (for the classification algorithm) can also be copied to these points.

In this method, an amortized analysis shows that the function τ is invoked L times, but for the price that the required amount of memory is also proportional to L.

2. **Medium amount of memory:**

   For larger values of n and B, storing all visited orbits is not feasible in practice. An other option is that only the current orbit is stored until repetition occurs, which indicates a cycle. In this way, many orbits and cycles are examined multiple times, which increases the running time. In theory, we cannot rule out the very pessimistic possibility that all points within the box constitute a single very long orbit, in which case this algorithm spares no memory compared to the previous one, but uses much more time. Experiments suggest however that this does not happen, and the orbits are much shorter than the number of points.

   We assume that for each point, the orbit can be stored in memory and then the first repetition is detected in constant time. This gives a multiplicative factor of ℓ in the running time.

3. **Strict memory constraints:**

   It is also possible to do all computation using practically constant memory. Floyd's Tortoise and Hare algorithm can detect cycles without storing the entire orbit. It simultaneously keeps track of $\tau^k(v)$ and $\tau^{2k}(v)$, which, for a periodic orbit, must coincide for some k. The details can be found e.g. in [8].

   In this case, τ may be invoked three times more than in the previous algorithm, but the memory requirement is much lower.

The number of points in the bounding box, i.e. L, can be calculated from Lemma 3.3. The box is defined by $\|v\|_1 \leq B$, i.e. its shape is a cross-polytope (e.g. an octahedron for $n = 3$), whose volume is $2^n B^n / n!$, thus:

$$\begin{aligned} L &= O\left(\frac{2^n B^n}{n!}\right) = O\left(\frac{(2e)^n B^n}{n^n}\right) = \\ &= O\left((2e^2)^n 2^{\frac{n^3(n-1)}{2}} |D|^{n^3} K_1 (M^{-1})^n \max_{d \in D} \|d\|_1^n\right) = \\ &= O\left(2^{\frac{n^4}{2}} |D|^{n^3} K_1 (M^{-1})^n \max_{d \in D} \|d\|_1^n\right). \end{aligned}$$

Therefore, the time-complexity of the first algorithm (abundant memory) is the number of τ-invocations multiplied by the complexity of one τ-invocation:

$$T_1 = O(Ln^2) = O\left(2^{\frac{n^4}{2}}|D|^{n^3}K_1(M^{-1})^n \max_{d\in D}\|d\|_1^n\right).$$

Generally, the time-complexity of the second algorithm (medium memory) is $T_2 = O(\ell T_1)$, but in terms of the parameters, we can give no better worst-case bound for ℓ than $\ell \leq L$. The O-notation for the time-complexity of the third algorithm (small memory) is the same, but the constant is three times more. More specifically:

$$T_2, T_3 = O(\ell L n^2) = O(L^2 n^2) = O\left(2^{n^4}|D|^{2n^3}K_1(M^{-1})^{2n} \max_{d\in D}\|d\|_1^{2n}\right).$$

4 Randomized algorithm

The basic idea of a randomized approach is the following: use the bounding box as before but instead of computing the orbit of every point in the box, pick one point at random and check only this orbit, then repeat for another point in the box. If a large number of experiments fail to find a nontrivial cycle, then – with high probability – the system has the finiteness property.

In order to guarantee sufficient accuracy with this method, we have to bound the soundness error, i.e. the probability that a random orbit fails to find a nontrivial cycle. If this can be bounded by a constant $q < 1$, then the algorithm, after k random points, gives incorrect answer with probability $\leq q^k$, which gets sufficiently small quickly (similarly to the case of randomized primality tests). However, our attempts at finding such q have unfortunately been vain. We therefore decided to analyze this probability empirically. As we will later see, there are families of numeration systems where the soundness error is not uniformly bounded away from 1 if we select points from a half-plane according to a uniform distribution.

To illustrate the difficulty of randomly finding a nontrivial cycle, consider the following dynamical system (which is not a numeration system): let $f\colon \mathbb{N} \to \mathbb{N}$ be defined as $f(83) = 83$ and $f(x) = \lfloor x/2 \rfloor$ for $x \neq 83$. The discrete dynamical system obtained by iterating this function has two cycles: the fixed points 0 and 83. A straightforward inductive reasoning shows that picking a uniformly random starting value in the interval $[0, N]$ for $N > 0$ finds the nontrivial cycle with probability less than $1/25$. The intuitive argument is that for a large random x_0, its orbit hits the interval $[51, 100]$ at a random position, and it only detects the cycle if this hit is exactly at 83. This example could be modified to make the probability even smaller. If a numeration system shares the property that nontrivial cycles are similarly hidden, then the soundness error can be similarly large. To investigate this, we present empirical calculations below.

Let N be a non-negative integer and define the sets $S_i(N)$ for $i \in \{1, 2, 3, 4\}$ as follows:

$$\begin{aligned}
S_1(N) &= \{(x, y) \mid x, y \in \mathbb{Z} \text{ and } 0 \leq x, y \leq N\}\\
S_2(N) &= \{(x, y) \mid x, y \in \mathbb{Z} \text{ and } -N \leq x \leq 0, 0 \leq y \leq N\}\\
S_3(N) &= \{(x, y) \mid x, y \in \mathbb{Z} \text{ and } -N \leq x, y \leq 0\}\\
S_4(N) &= \{(x, y) \mid x, y \in \mathbb{Z} \text{ and } 0 \leq x \leq N, -N \leq y \leq 0\}
\end{aligned}$$

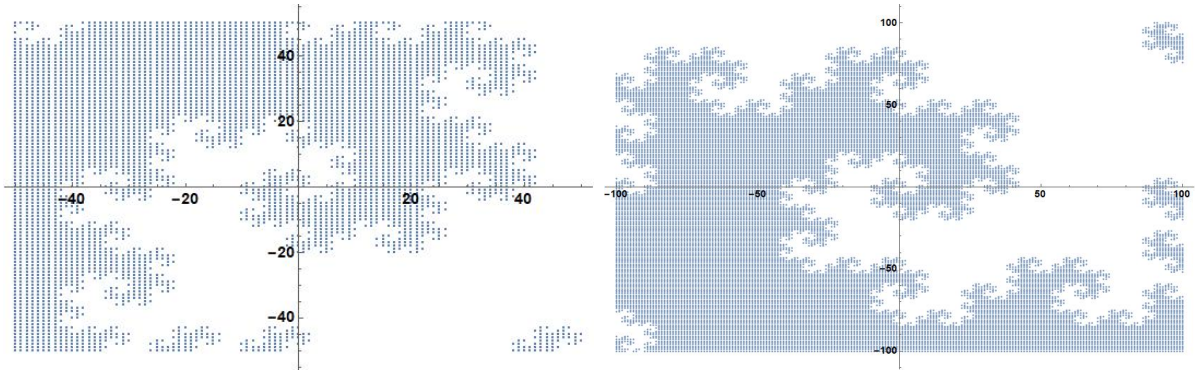
i.e. $S_i(N)$ is a square in the ith quadrant of the Cartesian coordinate system whose side is N and two sides of the square lie on the axes, furthermore let $S(N)$ denote the union of sets $S_i(N)$. We analyzed a large number of binary systems with the canonical digit set and we tried to find a bound for probability as mentioned above. In the following table we present some simulation results for the following matrices:

$$M_1 = \begin{pmatrix} -3 & -7 \\ 1 & 3 \end{pmatrix}, \quad M_2 = \begin{pmatrix} -2 & -4 \\ 2 & 3 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad M_4 = \begin{pmatrix} -10 & -98 \\ 1 & 10 \end{pmatrix}.$$

Region	$S_1(N)$ (%)			$S_2(N)$ (%)			$S_3(N)$ (%)			$S_4(N)$ (%)			$S(N)$ (%)		
N	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100
M_1	52.4	51.0	50.5	52.4	51.0	50.5	52.4	51.0	50.5	52.4	51.0	50.5	51.2	50.5	50.3
M_2	74.2	58.4	18.9	26.1	80.8	58.9	24.7	41.9	81.4	72.3	18.5	41.7	49.2	49.9	50.2
M_3	100	100	100	82.5	83.0	83.2	0.2	0	0	4.8	2.0	1.0	46.6	46.2	46.0
M_4	100	100	100	93.2	94.2	94.6	0.2	0	0	4.8	2.0	1.0	49.4	49.0	48.9

Table 1: Frequencies of points with orbit reaching $(0, 0)$.

Observe in column $S(N)$ of Table 1 that the frequency of points with orbit reaching $(0, 0)$ is approximately 50%, so if we pick k points randomly and their orbits end up at the origin, then the probability that the system is not a numeration system is bounded above by approximately $\frac{1}{2^k}$. On the other hand, if we investigate the orbits of elements of sets $S_i(N)$ separately, then we can find some interesting frequency values. For example in the second row of the Table 1, an oscillation is observed in the probabilities which is illustrated in Figure 1 and Table 2.



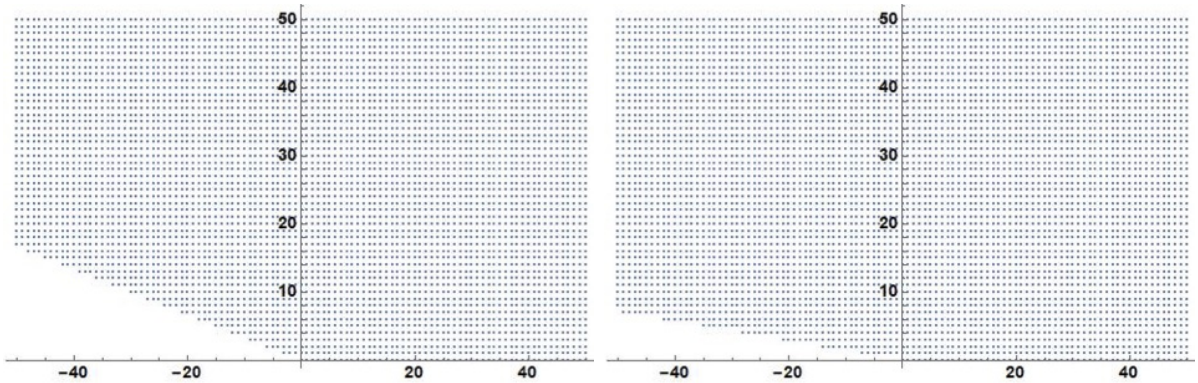
(a) Region $S(50)$ (b) Region $S(100)$

Figure 1: Points with orbit reaching $(0, 0)$ in base $\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$

Region \ N	10	20	50	100	150	200
$S_1(N)$	29.75%	74.15%	58.36%	18.88%	25.14%	41.15%
$S_2(N)$	28.10%	26.10%	80.78%	58.95%	28.99%	18.75%
$S_3(N)$	74.38%	24.72%	41.91%	81.41%	75.12%	58.78%
$S_4(N)$	77.69%	72.34%	18.53%	41.71%	71.42%	81.43%

Table 2: Frequencies of points with orbit reaching $(0, 0)$ in base $\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$

The last two rows of Table 1 contain high frequency values in columns $S_1(N)$ and $S_2(N)$, so in the region $S_1(N) \cup S_2(N)$, the probability that we can detect a nontrivial cycle is small. From now on, we will analyze the points of the first two quadrants of the coordinate system, and let $U(N) = S_1(N) \cup S_2(N)$. In this section we will prove, that for all $\varepsilon > 0$, there exists a matrix M such that the probability of finding a point reaching a nontrivial cycle is less than ε. Let p be a non-negative integer, let $M(p) = \begin{pmatrix} -p & -p^2+2 \\ 1 & p \end{pmatrix} \in \mathbb{Z}^{2\times 2}$ be a family of expansive matrices and let D denote the canonical digit set i.e. $D = \{(0,0),(1,0)\}$. Figure 2 illustrates the points with orbit ending up at $(0,0)$ in the systems $(M(3), D)$ and $(M(7), D)$.



(a) In system $(M(3), D)$ (b) In system $(M(7), D)$

Figure 2: Points with orbit reaching $(0,0)$

Lemma 4.1. *Let $p = 2m$ or $p = 2m+1$ be a non-negative integer. In the system $(M(p), D)$ the map τ can be applied as follows:*

$$\tau^2\left(\begin{bmatrix} 2l \\ 2k \end{bmatrix}\right) = \begin{bmatrix} l \\ k \end{bmatrix} \tag{6}$$

$$\tau^2\left(\begin{bmatrix} 2l+1 \\ 2k \end{bmatrix}\right) = \begin{bmatrix} l \\ k \end{bmatrix} \tag{7}$$

$$\tau^2\left(\begin{bmatrix} 2l \\ 2k+1 \end{bmatrix}\right) = \begin{bmatrix} m+l \\ k \end{bmatrix} \tag{8}$$

$$\tau^2\left(\begin{bmatrix} 2l+1 \\ 2k+1 \end{bmatrix}\right) = \begin{cases} \begin{bmatrix} m+l \\ k \end{bmatrix} & \text{if } p = 2m \\ \begin{bmatrix} m+l+1 \\ k \end{bmatrix} & \text{if } p = 2m+1 \end{cases} \tag{9}$$

Proof. We prove property (9) only, because the proof of (6)-(8) is essentially the same as (9). If $p = 2m$, then

$$\tau^2\left(\begin{bmatrix} 2l+1 \\ 2k+1 \end{bmatrix}\right) = \tau\left(\begin{bmatrix} 1+2k-2lm-2m^2-4km^2 \\ l+m+2km \end{bmatrix}\right) =$$

$$= -\frac{1}{2}\begin{bmatrix} 2m & 4m^2-2 \\ -1 & -2m \end{bmatrix} \cdot \begin{bmatrix} 1+2k-2lm-2m^2-4km^2-d \\ l+m+2km \end{bmatrix} \overset{d=1}{=} \begin{bmatrix} m+l \\ k \end{bmatrix}.$$

If $p = 2m + 1$, then

$$\tau^2\left(\begin{bmatrix}2l+1\\2k+1\end{bmatrix}\right) = \tau\left(\begin{bmatrix}k-l-3m-4km-2lm-2m^2-4km^2\\1+k+l+m+2km\end{bmatrix}\right) =$$
$$= -\frac{1}{2}\begin{bmatrix}2m+1 & 4m^2+4m-1\\-1 & -2m-1\end{bmatrix}\cdot\begin{bmatrix}k-l-3m-4km-2lm-2m^2-4km^2-d\\1+k+l+m+2km\end{bmatrix}$$
$$\overset{d=1}{=}\begin{bmatrix}m+l+1\\k\end{bmatrix}.$$

□

We consider a system $(M(p), D)$. Let $k \in \mathbb{N}$ be a non-negative integer and define the sets H_k^+ and H_k^- by:

$$H_k^+ = \{(i,k) \mid i \geq -p \cdot k\},$$
$$H_k^- = \{(i,k) \mid i < -p \cdot k\},$$

i.e. these sets split the line $y = k$ into two parts for all $k \in \mathbb{N}$.

Corollary 4.2. *The orbit of each $x \in H_0^+$ eventually reaches $(0,0)$.*

Proof. $H_0^+ = \{(i,0) \mid i \geq 0\}$. If we use the properties (6) and (7) of Lemma 4.1 then we can describe the orbits of elements as

$$\tau^2\left(\begin{bmatrix}i\\0\end{bmatrix}\right) = \begin{bmatrix}\lfloor\frac{i}{2}\rfloor\\0\end{bmatrix}$$

This means that if $x \in H_0^+$ then

$$\|x\|_2 > \left\|\tau^2(x)\right\|_2$$

thus the Euclidean norm of elements of sequence $x, \tau^2(x), \tau^4(x), \ldots$ is non-negative and strictly monotonically decreasing, so in a finite number of iterations, the orbit of x reaches $(0,0)$. □

Corollary 4.3. *The orbit of each element of H_0^- eventually reaches $(-1,0)$, and this point is a periodic point.*

Proof. Using the property (7) of Lemma 4.1, it can be easily seen that

$$\tau^2\left(\begin{bmatrix}-1\\0\end{bmatrix}\right) = \begin{bmatrix}-1\\0\end{bmatrix}$$

i.e. $(-1,0)$ is a periodic point. The proof proceeds analogously to that of Corollary 4.2. □

Lemma 4.4. *For all $k \in \mathbb{N}$: $x \in H_k^+$ if and only if the orbit of x ends up in $(0,0)$.*

Proof. Assume that $p = 2m$ and $x = (2l+1, 2k+1)$ is contained in H^+_{2k+1}. We can give $\tau^2(x)$ using the property (9) of Lemma 4.1 as $(m+l, k)$. If the definition of H^+_{2k+1} is applied then we get $2l+1 \geq -p(2k+1)$ and $2k+1 \geq 0$, which implies $k \geq 0$. Furthermore, the left hand side of $2l+1 \geq -p(2k+1)$ is odd, but the right hand side is even because of $p = 2m$, so $2l+1 > -p(2k+1)$ must hold and therefore

$$\begin{aligned} 2l+1 > -p(2k+1) \;&\Rightarrow\; 2l \geq -p(2k+1) \;\Rightarrow \\ \Rightarrow\; 2l \geq -2m(2k+1) \;&\Rightarrow\; m+l \geq -2mk = -pk. \end{aligned}$$

This means that $\tau^2(x) \in H^+_k$. In the same way as above we can show easily that $x \in H^+_k$ implies $\tau^2(x) \in H^+_{\lfloor \frac{k}{2} \rfloor}$. Therefore there exists an $s \in \mathbb{N}$ such that $\tau^{2s}(x) \in H^+_0$. Applying Corollary 4.2, it follows that for $x \in H^+_k$, the orbit ends in $(0,0)$. A similar reasoning can be used to show that for $x \in H^-_k$, the orbit ends up in $(-1,0)$. □

Theorem 4.5. *For all $\varepsilon > 0$ there exists a natural number p such that in system $(M(p), D)$, the probability that we find a nontrivial cycle in the first two quadrants by following the orbit of a uniformly random point is less than ε.*

Proof. Let $N = rp$, where p is fixed and r is a natural number. The probability can be calculated as follows:

$$\begin{aligned} \lim_{N\to\infty} \mathbb{P}\left(x \in \bigcup_{k=0}^{\infty} H^-_k \cap U(N) \;\middle|\; x \in U(N) \right) &= \lim_{r\to\infty} \frac{\left|\bigcup_{k=0}^{\infty} H^-_k \cap U(rp)\right|}{|U(rp)|} = \\ = \lim_{r\to\infty} \frac{p + 2p + \cdots + rp}{(2rp+1)(rp+1)} &= \lim_{r\to\infty} \frac{rp(r+1)}{2(2rp+1)(rp+1)} = \frac{1}{4p} < \varepsilon. \end{aligned}$$

□

5 Further research

There are several more questions about the algorithmic complexity of numeration systems that one could ask. Proving lower bounds on algorithmic complexity is notoriously difficult, so-called hardness results usually rely on assumptions about computational complexity of other problems, as with NP-completeness interpretations. It would be interesting to show NP-completeness results, but we tried in vain to encode NP-hard problems in the choice of M and in particular, D.

In our ongoing research we attempt to find answers to the following questions.

- It is an interesting question whether the estimate on ε in Lemma 3.1 is sharp. Our conjecture is that it is sharp at least with constant n, i.e. the dependence on $|a_0|$. We are currently investigating this problem and trying to construct a family of polynomials that attain the bound asymptotically.

- An other important question is about Definition 3.2, the Jordan condition number. It would be very useful to find an upper bound on $K_1(M^{-1})$ in Lemma 3.3, in terms of the dimension and the size of the entries of the integer matrix M.

- We would like to give bounds on the length of orbits (ℓ) in order to get better worst-case bounds in Subsection 3.3, or give examples proving that orbits can be really long.
- As mentioned in Section 4, it is an open problem to give (possibly sharp) explicit bounds on the soundness error of a randomized algorithm.
- We also try to generalize the results for shift radix systems, see e.g. [1, 7].

References

[1] S. Akiyama, T. Borbély, H. Brunotte, A. Pethő, J. M. Thuswaldner: Generalized radix representations and dynamical systems. *Acta Mathematica Hungarica* **108**(3) (2005), pp. 207–238

[2] G. Barat, V. Berthé, P. Liardet and J. Thuswaldner: Dynamical directions in numeration. *Annales de l'Institut Fourier* **56**(7) (2006), pp. 1987–2092

[3] H. Cohen: *A Course in Computational Algebraic Number Theory* (Springer-Verlag Berlin Heidelberg, 1996)

[4] R. A. Horn, C. R. Johnson: *Matrix Analysis* (Cambridge University Press, 2012)

[5] I. Kátai, J. Szabó: Canonical number-systems for complex integers. *Acta Scientiarum Mathematicarum* **37**(3-4) (1975), pp. 255–260

[6] I. Kátai: Generalized number systems in Euclidean spaces. *Mathematical and Computer Modelling* **38**(7) (2003), pp. 883–892

[7] P. Kirschenhofer, J. Thuswaldner: Shift Radix Systems – A Survey. *RIMS Kokyuroku Bessatsu* **B46** (2013), pp. 1–59

[8] D. E. Knuth: *The Art of Computer Programming, vol. II: Seminumerical Algorithms* (Addison-Wesley, 1969), p. 7, exercises 6 and 7

[9] A. Kovács: On computation of attractors for invertible expanding linear operators in $\mathbb{Z}^k$. *Publicationes Mathematicae Debrecen* **56**(1-2) (2000), pp. 97–120

[10] A. Kovács: Radix expansion in lattices. *PhD thesis*, Budapest, 2001

[11] A. Kovács: On number expansions in lattices. *Mathematical and Computer* **38** (2003), pp. 909–915

[12] J. Thuswaldner: Attractors of invertible expanding linear operators and number systems in $\mathbb{Z}^2$. *Publicationes Mathematicae Debrecen* **58** (2001), pp. 423–440

[13] L. N. Trefethen, M. Embree: *Spectra and Pseudospectra, The Behavior of Nonnormal Matrices and Operators* (Princeton University Press, 2005)

[14] A. Vince: Replicating tesselations. *SIAM Journal on Discrete Mathematics* **6** (1993), pp. 501–521

[15] M. Waldschmidt: *Diophantine Approximation on Linear Algebraic Groups, Transcendence Properties of the Exponential Function in Several Variables, Grundlehren der Mathematischen Wissenschaften* (Springer-Verlag Berlin Heidelberg, 2000)

[16] D. M. Young: *Iterative Solution of Large Linear Systems* (Academic Press, 1971)