ISSN: 1803-3814

MENDEL Soft Computing

THE COMPUTATIONAL POWER OF NEURAL NETWORKS AND REPRESENTATIONS OF NUMBERS IN NON-INTEGER BASES

Jiří Šíma

Institute of Computer Science
The Czech Academy of Sciences
P.O. Box 5, 182 07 Prague 8
Czech Republic
sima@cs.cas.cz

Abstract: *We briefly survey the basic concepts and results concerning the computational power of neural networks which basically depends on the information content of weight parameters. In particular, recurrent neural networks with integer, rational, and arbitrary real weights are classified within the Chomsky and finer complexity hierarchies. Then we refine the analysis between integer and rational weights by investigating an intermediate model of integer-weight neural networks with an extra analog rational-weight neuron (1ANN). We show a representation theorem which characterizes the classification problems solvable by 1ANNs, by using so-called cut languages. Our analysis reveals an interesting link to an active research field on non-standard positional numeral systems with non-integer bases. Within this framework, we introduce a new concept of quasi-periodic numbers which is used to classify the computational power of 1ANNs within the Chomsky hierarchy.*

Keywords: *neural network, Chomsky hierarchy, β-expansion, cut language*

1 The Neural Network Model

(Artificial) neural networks (NNs) are biologically inspired computational devices that are alternative to conventional computers, especially in the area of machine learning, with a plethora of successful commercial applications in artificial intelligence [11, 13]. In order to analyze the computability and complexity aspects of practical neurocomputing systems, idealized formal models of NNs are introduced which abstract away from implementation issues, e.g. analog numerical parameters are assumed to be true real numbers. The limits and potential of particular NNs for general-purpose computation have been studied by classifying them within the Chomsky hierarchy (e.g. finite or pushdown automata, Turing machines) and/or more refined complexity classes (e.g. polynomial time) [38]. Moreover, NNs may serve as reference models for analyzing alternative computational resources (other than time or memory space) such as analog state [30], continuous time [37], energy [34], temporal coding [21], etc.

We first specify a common formal model of recurrent NNs which is used in this paper. The network consists of s computational units called *neurons*, indexed as $V = \{1, \dots, s\}$, which are connected into a directed graph representing an *architecture* of N. Each edge (i, j) leading from neuron i to j is labeled with a real *weight* $w(i, j) = w_{ji} \in \mathbb{R}$ and each unit $j \in V$ is associated with a real *bias* $w(0, j) = w_{j0} \in \mathbb{R}$ which can be viewed as the weight from an additional formal neuron 0. The absence of a connection within the architecture corresponds to a zero weight between the respective neurons, and vice versa. The *computational dynamics* of N determines for each unit $j \in V$ its real *state (output)* $y_j^{(t)} \in [0, 1]$ at discrete time instants $t = 0, 1, 2, \dots$, which establishes the global *network state* $\mathbf{y}^{(t)} = (y_1^{(t)}, \dots, y_s^{(t)}) \in [0, 1]^s$ at each discrete time instant $t \geq 0$.

At the beginning of a computation, the neural network N is placed in an *initial state* $\mathbf{y}^{(0)}$ which may also include an external input. At discrete time instant $t \geq 0$, an *excitation* of each neuron $j \in V$ is evaluated as

$$\xi_j^{(t)} = \sum_{i=0}^{s} w_{ji} y_i^{(t)} \tag{1}$$

which includes the bias w_{j0} due to $y_0^{(t)} = 1$ is assumed for every $t \geq 0$. At the next instant $t + 1$, the neurons $j \in \alpha_{t+1}$ from a selected subset $\alpha_{t+1} \subseteq V$ (e.g., corresponding to a layer) compute their new outputs $y_j^{(t+1)}$ in parallel by applying an *activation function* $\sigma : \mathbb{R} \longrightarrow [0, 1]$ to $\xi_j^{(t)}$, whereas the remaining units $j \notin \alpha_{t+1}$ do not update their states, that is,

$$y_j^{(t+1)} = \begin{cases} \sigma\left(\xi_j^{(t)}\right) & \text{for } j \in \alpha_{t+1} \\ y_j^{(t)} & \text{for } j \in V \setminus \alpha_{t+1} \, . \end{cases} \tag{2}$$

We employ the *saturated-linear* activation function,

$$\sigma(\xi) = \begin{cases} 1 & \text{for } \xi \geq 1 \\ \xi & \text{for } 0 < \xi < 1 \\ 0 & \text{for } \xi \leq 0\,, \end{cases} \tag{3}$$

although some of the following results are valid for more general classes of activation functions [17, 29, 33, 41] including the logistic function [16]. In this way, the new network state $\mathbf{y}^{(t+1)}$ at time $t+1$ is determined.

The computational power of NNs has been studied analogously to the traditional models of computation so that the networks are exploited as acceptors of formal languages $L \subseteq \Sigma^*$ over a finite alphabet Σ (e.g., the binary alphabet $\Sigma = \{0,1\}$). A language L corresponds to a single-class classification (decision) problem which is solved by a neural network N, in the sense that N accepts language L containing all positive input instances of the problem that belong to the class. For the finite NNs the following input/output protocol has been used [2, 6, 14, 15, 29, 30, 31, 32, 38, 40]. An input word (string) $\mathbf{x} = x_1 \ldots x_n \in \Sigma^n$ of arbitrary length $n \geq 0$ is sequentially presented to the network, symbol after symbol, via so-called *input neurons* from $X \subset V$, at *macroscopic time* instants. For example, each symbol $x \in \Sigma$ can be encoded by one input neuron $j(x) \in X$. Thus, an input symbol $x_i \in \Sigma$ is read at macroscopic time instant $i = 1, \ldots, n$, when the states of input neurons are externally set (and clamped) regardless of any influence of N, as

$$y_j^{(d(i-1)+k)} = \begin{cases} 1 & \text{for } j = j(x_i) \\ 0 & \text{for } j \in X \setminus \{j(x_i)\} \end{cases} \qquad k = 0, \ldots, d-1 \tag{4}$$

where an integer parameter $d \geq 1$ is the *time overhead* for processing a single input symbol which coincides with the macroscopic time step. Finally, N has two *output neurons*, $Y = \{\text{out}, \text{val}\} \subset V$ where $\text{out} \in Y$ signals in computational time $T(n) \geq dn$ whether the input word $\mathbf{x} \in \Sigma^n$ that has been read, belongs to the underlying language L, while $\text{val} \in Y$ indicates the end of computation when the result is produced, that is,

$$y_{\text{out}}^{(T(n))} = \begin{cases} 1 & \text{for } \mathbf{x} \in L \\ 0 & \text{for } \mathbf{x} \notin L\,, \end{cases} \qquad y_{\text{val}}^{(t)} = \begin{cases} 1 & \text{for } t = T(n) \\ 0 & \text{for } t \neq T(n)\,. \end{cases} \tag{5}$$

We say that a language $L \subseteq \Sigma^*$ is *accepted (recognized)* by N, which is denoted by $L = L(N)$, if for any input word $\mathbf{x} \in \Sigma^*$, $\mathbf{x}$ is accepted by N iff $\mathbf{x} \in L$.

2 Neural Networks and the Chomsky Hierarchy

The computational power of NNs with the saturated-linear activation function (3), which represents a usual formal model as introduced in Sect. 1, depends on the descriptive complexity of their weight parameters [30, 38]. NNs with *integer* weights $w_{ji} \in \mathbb{Z}$, corresponding to binary-state networks with 2^s global states, coincide with finite automata which accept precisely the regular languages [2, 22, 40]. For example, size-optimal implementations of a given (deterministic) finite automaton with m states by using a NN with $\Theta(\sqrt{m})$ neurons have been elaborated [14, 15, 34]. Furthermore, *rational* weights $w_{ji} \in \mathbb{Q}$ make the analog-state NNs computationally equivalent to Turing machines which accept the recursively enumerable languages [15, 32]. Thus (by a real-time simulation [32]) polynomial-time computations of such NNs are characterized by the fundamental complexity class P which is composed of the problems that are considered efficiently solvable by conventional computers.

Moreover, NNs with arbitrary *real* weights $w_{ji} \in \mathbb{R}$ (in fact, only one suitable irrational weight suffices) can even derive "super-Turing" computational capabilities [30, 31]. It is because such a NN model may include the infinite amount of information encodable in arbitrarily precise real numbers, which does not fit into the classical definition of algorithm whose description must be finite. In particular, any input/output mapping (including algorithmically undecidable problems) can be computed by such NNs within exponential time while polynomial-time computations correspond to the nonuniform complexity class P/poly. The class P/poly is composed of problems that are solvable in polynomial time (P) by an algorithm (Turing machine) which receives an extra input string of polynomial length (poly) as an external *advise* whose value depends only on the original input length n (i.e. the same string is provided for all inputs of length n). The advice function which assigns a polynomial-length string to each input length n need not be algorithmically computable, which makes the complexity class P/poly nonuniform, including algorithmically undecidable problems. In addition, a proper infinite hierarchy of nonuniform complexity classes between P and P/poly has been established for polynomial-time computations of NNs with increasing Kolmogorov complexity of real weights [6] where the Kolmogorov complexity of a real number is defined as the length of the shortest computer program (in a fixed programming language) that produces this number as output. This proves that with the more information content in real weights a NN can solve strictly more problems within polynomial time.

As can be seen, our understanding of the computational power of NNs is satisfactorily fine-grained when changing from rational to arbitrary real weights. In contrast, there is still a gap between integer and rational weights which results in a jump from regular (Type-3) to recursively enumerable (Type-0) languages in the Chomsky hierarchy. However, it is known that already *two* analog-state neurons with rational weights, integrated into an integer-weight NN, can implement two stacks of pushdown automata [32], a model equivalent to Turing machines. Thus, the respective gap in the computational power of NNs is localized between the integer-weight NNs (Chomsky Type-3) and the integer-weight NNs with two extra rational-weight neurons (Chomsky Type-0). Hence, a natural question arises: What is the computational power of integer-weight networks with *one* extra analog neuron having rational weights? In order to answer this question we define an intermediate formal model of NNs with one extra analog rational-weight neuron s (1ANN), which satisfies

$$w_{ji} \in \left\{ \begin{array}{ll} \mathbb{Z} & \text{for } j = 1, \dots, s-1 \\ \mathbb{Q} & \text{for } j = s \end{array} \right. \qquad i = 0, \dots, s\,. \tag{6}$$

We have previously shown that 1ANN is computationally equivalent to a finite automaton with a rational-valued register whose domain is partitioned into a finite number of intervals, each associated with a local state-transition function [35]. We now characterize the class of problems solvable by 1ANNs in the following representation theorem (cf. [36]):

Theorem 1. *A language $L \subset \Sigma^*$ that is accepted by a 1ANN satisfying (6) and $0 < |w_{ss}| < 1$, can be written as*

$$L = h\left(\left(\left(\bigcup_{r=1}^{p} \left(\overline{L_{<c_r}} \cap L_{<c_{r+1}} \right)^R \cdot A_r \right)^{Pref} \cap R_0 \right)^* \cap R \right) \tag{7}$$

in which $(\overline{L_{<c_r}} \cap L_{<c_{r+1}})$ can suitably be replaced by one of the following terms $(L_{>c_r} \cap L_{<c_{r+1}})$, $(L_{>c_r} \cap \overline{L_{>c_{r+1}}})$, $(\overline{L_{<c_r}} \cap \overline{L_{<c_{r+1}}})$, $\overline{L_{>0}}$, or $\overline{L_{<1}}$ (for each value of r, the choice of the term may be different so that the underlying intervals create a disjoint cover of the real line) where

- $A = \left\{ \sum_{i=0}^{s-1} w_{si} y_i \,\middle|\, y_1, \dots, y_{s-1} \in \{0,1\} \right\} \cup \{0,1\} \subset \mathbb{Q}$ *is a finite alphabet of* digits,
- $\{c_1, \dots, c_p\} = \left\{ -\sum_{i=0}^{s-1} \frac{w_{ji}}{w_{js}} y_i \,\middle|\, j \in V \setminus (X \cup \{s\}) \text{ s.t. } w_{js} \neq 0\,, y_1, \dots, y_{s-1} \in \{0,1\} \right\} \cup \{0,1\} \subset \mathbb{Q}$ *is a finite set of* thresholds *such that* $0 = c_1 \leq c_2 \leq \cdots \leq c_p = 1$,
- $L_{<c_r}\,,\ L_{>c_r} \subseteq A^*$ *for $r = 1, \dots, p$, are so-called* cut languages *over alphabet A defined as*

$$L_{<c} = \left\{ a_1 \dots a_n \in A^* \,\middle|\, \sum_{k=1}^{n} a_k \beta^{-k} < c \right\} \tag{8}$$

  (similarly for $L_{>c}$) where $\beta = \frac{1}{w_{ss}} \in \mathbb{Q}$ is called a base (radix),
- $A_1, \dots, A_p$ *is a partition of A,*
- S^{Pref} *denotes the largest prefix-closed subset of $S \cup A \cup \{\varepsilon\}$ (ε denotes the empty string),*
- $R\,, R_0 \subseteq A^*$ *are regular languages,*
- $h : A^* \longrightarrow \Sigma^*$ *is a letter-to-letter morphism.*

Note that the implication in Theorem 1 can be partially reversed (see [36]). The representation formula (7) is in fact based on the cut languages defined by (8), which are combined by usual operations such as complementation, intersection, union, concatenation, Kleene star, reversal, the largest prefix-closed subset, and a letter-to-letter morphism. For example, the classes of regular (Chomsky Type-3) and context sensitive (Chomsky Type-1) languages are closed under these operations. Hence, in these cases it is sufficient to classify the cut languages within the Chomsky hierarchy which is done in Sect. 5.

3 β-Expansions

In the representation Theorem 1, the finite set $A \neq \emptyset$ contains (rational) numbers called *digits* while the (rational) parameter β satisfying $|\beta| > 1$, is no by chance named a *base* or *radix*. It is because within a non-standard positional numeral system whose base and digits need not be integers, a *cut language* $L_{<c}$ introduced in (8) is in fact composed of all finite base-β representations of the numbers that are less than a *threshold* c. Hereafter, we assume that the digits and the base are in general real numbers if not stated otherwise. In particular, a word

(string) of digits $a_1 \ldots a_n \in A^*$ in which the radix point is omitted at the beginning, represents a number in base β using the negative exponents of β as

$$(0\,.\,a_1 \ldots a_n)_\beta = a_1\beta^{-1} + a_2\beta^{-2} + a_3\beta^{-3} + \cdots + a_n\beta^{-n} = \sum_{k=1}^{n} a_k \beta^{-k}\,. \tag{9}$$

Clearly, this is a generalization of integer-base positional numeral systems with integer radix $\beta \in \mathbb{Z}$ and the standard set of integer digits $A = \{0, 1, 2, \ldots, \beta - 1\}$, including the usual *decimal expansions* in which $\beta = 10$ and $A = \{0, 1, 2, \ldots, 9\}$ or *binary expansions* with $\beta = 2$ and $A = \{0, 1\}$. For instance, the fraction $\frac{3}{4}$ cannot only be represented in the usual integer bases as

$$\frac{3}{4} \;=\; (0\,.\,75)_{10} = 7 \cdot 10^{-1} + 5 \cdot 10^{-2} \;=\; (0\,.\,11)_2 = 1 \cdot 2^{-1} + 1 \cdot 2^{-2} \tag{10}$$

but also, for example, in non-integer base $\beta = \frac{5}{2}$ using the non-integer digits from $A = \left\{\frac{5}{16}, \frac{7}{4}\right\}$ as

$$\frac{3}{4} \;=\; \left(0\,.\,\frac{7}{4}\,\frac{5}{16}\right)_{\frac{5}{2}} = \frac{7}{4} \cdot \left(\frac{5}{2}\right)^{-1} + \frac{5}{16} \cdot \left(\frac{5}{2}\right)^{-2}\,. \tag{11}$$

In general, an infinite word of digits $a_1a_2a_3 \ldots \in A^\omega$ represents a number in base β as a convergent series (recall $|\beta| > 1$):

$$(0\,.\,a_1a_2a_3 \ldots)_\beta = a_1\beta^{-1} + a_2\beta^{-2} + a_3\beta^{-3} + \cdots = \sum_{k=1}^{\infty} a_k \beta^{-k}\,, \tag{12}$$

which is widely known as a *β-expansion*. Today's already classical definition of β-expansions was introduced in late 1950s [25, 23] and this concept has up to now been a subject of active research, e.g. [1, 3, 4, 5, 7, 8, 9, 10, 12, 18, 19, 20, 24, 26, 27, 28]. The world of β-expansions is much richer than that of the usual integer-base representations as is briefly illustrated on the uniqueness issue of β-expansions. It is well known that for an *integer base* $\beta > 1$ and the standard digits from $A = \{0, 1, 2, \ldots, \beta - 1\}$, almost any real number from the interval $[0, 1]$ has a *unique* infinite β-expansion. The only exception are those numbers for which also a finite β-expansion (9) exists, that have *two* distinct infinite β-expansions, e.g. the decimal expansions 750^ω and 749^ω of the fraction

$$\frac{3}{4} \;=\; (0\,.\,75)_{10} \;=\; (0\,.\,75000\ldots)_{10} \;=\; (0\,.\,74999\ldots)_{10}\,. \tag{13}$$

In contrast, for any *non-integer base* β, almost every number (for which a base-β representation exists) has infinitely (uncountably) many *distinct* β-expansions [27]. For example, assume $1 < \beta < 2$ and $A = \{0, 1\}$, which ensures that any real number from the interval $D_\beta = (0, \frac{1}{\beta-1})$ has a β-expansion. If $1 < \beta < \varphi$ where $\varphi = (1 + \sqrt{5})/2 \approx 1.618034$ is the *golden ratio*, then *every* number from D_β has an *uncountably* many distinct β-expansions [8]. If $\varphi \le \beta < q$ where $q \approx 1.787232$ is the *Komornik-Loreti constant*, then *countably* many numbers from D_β have a *unique* β-expansion [10]. For example, the infinite word $0^k(10)^\omega$ for any integer $k \ge 0$, represents a unique β-expansion of the number

$$\frac{1}{\beta^{k-1}(\beta^2 - 1)} = \big(0\,.\,\underbrace{0 \ldots 0}_{k \text{ times}}\,10\,10\,10\,10\,10\,\ldots\big)_\beta\,, \tag{14}$$

which gives the unique $\frac{5}{3}$-expansion $00(10)^\omega$ of $\frac{27}{80}$ for $k = 2$. In contrast, there are countably many distinct φ-expansions of the number 1 having the form $(10)^k110^\omega$, $(10)^\omega$, or $(10)^k01^\omega$ for every integer $k \ge 0$. Finally, if $q \le \beta < 2$, then *uncountably* many numbers from D_β have a *unique* β-expansion. These results have partially been generalized to arbitrary bases and digits [4, 18, 19].

4 Eventually Quasi-Periodic β-Expansions

An important role in our analysis of the computational power of NNs between integer and rational weights (Sect. 2) plays the periodicity of β-expansions. We say that an infinite β-expansions $a_1a_2a_3 \ldots \in A^\omega$ is *eventually periodic* if it can be written as

$$a_1a_2 \ldots a_{k_1}\,(a_{k_1+1}a_{k_1+2} \ldots a_{k_2})^\omega \tag{15}$$

where $a_1a_2 \ldots a_{k_1} \in A^{k_1}$ is a *preperiodic part* of length $k_1 \ge 0$ and $a_{k_1+1}a_{k_1+2} \ldots a_{k_2} \in A^m$ is the infinitely-repeated digit sequence of length $m = k_2 - k_1 > 0$, called the *repetend*. For $k_1 = 0$, the β-expansion is called (purely) *periodic*. Any eventually periodic β-expansion can be evaluated as

$$(0\,.\,a_1a_2 \ldots a_{k_1}\,\overline{a_{k_1+1}a_{k_1+2} \ldots a_{k_2}})_\beta = (0\,.\,a_1a_2 \ldots a_{k_1})_\beta + \beta^{-k_1}\varrho \tag{16}$$

where ϱ is a so-called *periodic point* which can be computed using the formula for the sum of a geometric progression as

$$\varrho = (0\,.\,\overline{a_{k_1+1}a_{k_1+2}\ldots a_{k_2}})_\beta = \frac{\sum_{k=1}^{m} a_{k_1+k}\beta^{-k}}{1-\beta^{-m}}\,. \tag{17}$$

Note that we employ the usual convention to indicate a repeating digits by drawing a horizontal line (a vinculum) above the repetend. For example, for $\beta=\frac{3}{2}$ and $A=\{0,1\}$, the eventually periodic β-expansion $1\,(10)^\omega$ which is composed of the 1-bit preperiodic part 1 and the 2-bit repetend 10, represents the fraction

$$\frac{22}{15} = (0\,.\,1\,\overline{10})_{\frac{3}{2}} = (0\,.\,1)_{\frac{3}{2}} + \left(\frac{3}{2}\right)^{-1}\cdot\varrho \quad \text{with the periodic point} \quad \varrho = \frac{1\cdot\left(\frac{3}{2}\right)^{-1}+0\cdot\left(\frac{3}{2}\right)^{-2}}{1-\left(\frac{3}{2}\right)^{-2}} = \frac{6}{5}\,. \tag{18}$$

We generalize the notion of periodicity for non-integer bases. We say an infinite β-expansion

$$a_1\ldots a_{k_1}\;a_{k_1+1}\ldots a_{k_2}\;a_{k_2+1}\ldots a_{k_3}\;a_{k_3+1}\ldots a_{k_4}\;a_{k_4+1}\ldots a_{k_5}\;\ldots\in A^\omega \tag{19}$$

is *eventually quasi-periodic* if it can split into a preperiodic part $a_1a_2\ldots a_{k_1}\in A^{k_1}$ of length $k_1\ge 0$, followed by an infinite sequence of so-called *quasi-repetends* $a_{k_i+1}\ldots a_{k_{i+1}}\in A^{m_i}$ of minimum length $m_i=k_{i+1}-k_i>0$, that share the same periodic point ϱ, that is,

$$\varrho = \left(0\,.\,\overline{a_{k_i+1}a_{k_i+2}\ldots a_{k_{i+1}}}\right)_\beta = \frac{\sum_{k=1}^{m_i} a_{k_i+k}\beta^{-k}}{1-\beta^{-m_i}} \quad \text{for every } i\ge 1\,, \tag{20}$$

which ensures

$$(0\,.\,a_1\ldots a_{k_1}\;a_{k_1+1}\ldots a_{k_2}\;a_{k_2+1}\ldots a_{k_3}\;a_{k_3+1}\ldots a_{k_4}\;a_{k_4+1}\ldots a_{k_5}\;\ldots)_\beta = (0\,.\,a_1a_2\ldots a_{k_1})_\beta + \beta^{-k_1}\varrho\,, \tag{21}$$

cf. equation (16). For $k_1=0$, the β-expansion is called (purely) *quasi-periodic*. Note that the quasi-repetends of an eventually quasi-periodic β-expansion can be distinct having even different lengths. Clearly, if an eventually quasi-periodic β-expansion is composed of only one quasi-repetend, then the β-expansion is eventually periodic. If it is not the case, then there are uncountably many distinct eventually quasi-periodic β-expansions representing the same number since the quasi-repetends are arbitrarily interchangeable according to (20). Indeed, the value of an eventually quasi-periodic β-expansion (21) does not change if any quasi-repetend is removed or replaced by another quasi-repetend or inserted in between two other quasi-repetends.

We illustrate the concept of eventually quasi-periodic β-expansions on a numerical example. Assume the base $\beta=\frac{5}{2}$ and the set of digits $A=\left\{0\,,\,\frac{1}{2}\,,\,\frac{7}{4}\right\}$. It can be shown that the words $\frac{7}{4}\,\left(\frac{1}{2}\right)^n 0\in A^{n+2}$ of length $n+2$, for every integer $n\ge 0$, where $\left(\frac{1}{2}\right)^n$ denotes the symbol $\frac{1}{2}\in A$ repeated n times, share the same periodic point

$$\begin{aligned}\varrho \;&=\; \left(0\,.\,\overline{\tfrac{7}{4}\;0}\right)_{\frac{5}{2}} = \left(0\,.\,\overline{\tfrac{7}{4}\;\tfrac{1}{2}\;0}\right)_{\frac{5}{2}} = \left(0\,.\,\overline{\tfrac{7}{4}\;\tfrac{1}{2}\;\tfrac{1}{2}\;0}\right)_{\frac{5}{2}} = \left(0\,.\,\overline{\tfrac{7}{4}\;\tfrac{1}{2}\;\tfrac{1}{2}\;\tfrac{1}{2}\;0}\right)_{\frac{5}{2}} = \left(0\,.\,\overline{\tfrac{7}{4}\;\tfrac{1}{2}\;\tfrac{1}{2}\;\tfrac{1}{2}\;\tfrac{1}{2}\;0}\right)_{\frac{5}{2}} \\ &=\; \cdots = \left(0\,.\,\overline{\tfrac{7}{4}\;\underbrace{\tfrac{1}{2}\cdots\tfrac{1}{2}}_{n\text{ times}}\;0}\right)_{\frac{5}{2}} = \cdots = \frac{\frac{7}{4}\cdot\left(\frac{5}{2}\right)^{-1}+\sum_{i=2}^{n+1}\frac{1}{2}\cdot\left(\frac{5}{2}\right)^{-i}+0\cdot\left(\frac{5}{2}\right)^{-n-2}}{1-\left(\frac{5}{2}\right)^{-n-2}} = \frac{3}{4}\end{aligned} \tag{22}$$

according to (20). Thus, for any infinite sequence of nonnegative integers $n_1,n_2,n_3,n_4,\ldots$, we obtain a quasi-periodic $\frac{5}{2}$-expansion $\frac{7}{4}\,\left(\frac{1}{2}\right)^{n_1}0\,\frac{7}{4}\,\left(\frac{1}{2}\right)^{n_2}0\,\frac{7}{4}\,\left(\frac{1}{2}\right)^{n_3}0\,\frac{7}{4}\,\left(\frac{1}{2}\right)^{n_4}0\ldots\in A^\omega$ of $\frac{3}{4}$ because

$$\frac{3}{4} = \left(0\,.\;\frac{7}{4}\;\underbrace{\frac{1}{2}\cdots\frac{1}{2}}_{n_1\text{ times}}\;0\;\;\frac{7}{4}\;\underbrace{\frac{1}{2}\cdots\frac{1}{2}}_{n_2\text{ times}}\;0\;\;\frac{7}{4}\;\underbrace{\frac{1}{2}\cdots\frac{1}{2}}_{n_3\text{ times}}\;0\;\;\frac{7}{4}\;\underbrace{\frac{1}{2}\cdots\frac{1}{2}}_{n_4\text{ times}}\;0\;\cdots\right)_{\frac{5}{2}}\,. \tag{23}$$

Hence, the fraction $\frac{3}{4}$ has uncountably many distinct quasi-periodic $\frac{5}{2}$-expansions using the digits $0\,,\,\frac{1}{2}\,,\,\frac{7}{4}$.

Finally, we introduce the notion of quasi-periodic numbers. We say that a real number $c\in\mathbb{R}$ is *β-quasi-periodic within A* if every infinite β-expansion of c is eventually quasi-periodic. Note that a number that has no β-expansion at all is formally considered to be quasi-periodic, e.g. the numbers from the complement of the Cantor set are 3-quasi-periodic within $\{0,2\}$. It can be shown that the fraction $\frac{3}{4}$ is $\frac{5}{2}$-quasi-periodic within $A=\left\{0\,,\,\frac{1}{2}\,,\,\frac{7}{4}\right\}$ as there are no other $\frac{5}{2}$-expansions of $\frac{3}{4}$ than that presented in the previous example (23). In contrast, the number $c=\frac{40}{57}=(0\,.\,0\,\overline{011})_{\frac{3}{2}}$, although defined by the eventually periodic $\frac{3}{2}$-expansion $0(011)^\omega$ over alphabet $\{0,1\}$, *is not* $\frac{3}{2}$-quasi-periodic within $\{0,1\}$ since the so-called *greedy* (i.e. lexicographically maximal) $\frac{3}{2}$-expansion $100000001\ldots\in\{0,1\}^\omega$ of $\frac{40}{57}$ proves to be not eventually periodic.

5 Neural Networks Between Integer and Rational Weights

In the following three theorems [39], we first classify the cut languages from the representation Theorem 1 within the Chomsky hierarchy using the quasi-periodicity of their thresholds which is defined in Sect. 4.

Theorem 2. *A cut language $L_{<c}$ is regular iff c is β-quasi-periodic within A.*

Theorem 3. *If c is not β-quasi-periodic within A, then the cut language $L_{<c}$ is not context-free.*

Theorem 4. *Let $\beta \in \mathbb{Q}$ and $A \subset \mathbb{Q}$. Every cut language $L_{<c}$ with threshold $c \in \mathbb{Q}$ is context-sensitive.*

Hence, we obtain a dichotomy that any cut language $L_{<c}$ which is anyway context-sensitive (Chomsky Type-1) for its rational parameters (base, digits, threshold), is either regular (Chomsky Type-3) or not context-free (Chomsky Type-2), depending on whether c is or is not β-quasi-periodic within A, respectively.

Further we apply this classification of cut languages to the representation Theorem 1, which gives the following three theorems [36] mainly by the closure properties of regular and context-sensitive languages regarding the operations that appear in equation (7).

Theorem 5. *Let N be a 1ANN and assume $0 < |w_{ss}| < 1$. Define $\beta \in \mathbb{Q}$, $A \subset \mathbb{Q}$, and $c_1, \ldots, c_p \in \mathbb{Q}$ as in Theorem 1 using the weights of N. If thresholds $c_1, \ldots, c_p$ are β-quasi-periodic within A, then N accepts a regular language.*

Theorem 6. *There is a language accepted by a 1ANN, which is not context-free.*

Theorem 7. *Any language accepted by a 1ANN is context-sensitive.*

Thus, we have refined the analysis of the computational power of NNs between integer and rational weights within the Chomsky hierarchy, which is schematically summarized in the following diagram:

$$\text{integer-weight NNs} \equiv \text{1ANNs with quasi-periodic parameters} \equiv \text{regular languages (Type-3)}$$

$$\text{1ANNs} \not\subset \text{context-free languages (Type-2)}$$

$$\text{1ANNs} \subset \text{context-sensitive languages (Type-1)}$$

$$\text{rational-weight NNs} \equiv \text{recursively enumerable languages (Type-0)}$$

6 Conclusion

In this paper we have briefly surveyed the results on the computational power of NNs which basically depends on the information content of their weights. Then we have characterized the class of languages accepted by integer-weight NNs with an extra rational-weight neuron, using the cut languages. We have shown an interesting link to active research on β-expansions in non-integer bases. We have introduced a new notion of quasi-periodic numbers which is used to refine the analysis of the computational power of NNs between integer and rational weights within the Chomsky hierarchy.

Nevertheless, there are still some important open problems. For example, the question of whether the condition in Theorem 5 is also necessary remains open. Another challenge for further research is to generalize the results to other domains of the feedback weight w_{ss} associated with analog unit s, such as $w_{ss} \in \mathbb{R}$ or $|w_{ss}| > 1$. Moreover, it would be interesting to study the possibility of having a proper hierarchy of 1ANN classes of increasing complexity, e.g. in terms of the maximum length of quasi-repetends of rational 1ANN parameters, similarly to the Kolmogorov-weight NN hierarchy [6].

Acknowledgement: This paper is an extended abstract of an invited talk which is based on author's previous work [35, 39, 36]. The research was done with institutional support RVO: 67985807 and partially supported by the grant of the Czech Science Foundation No. P202/12/G061.

References

[1] Allouche, J.P., Clarke, M., Sidorov, N.: Periodic unique beta-expansions: The Sharkovskiĭ ordering. Ergodic Theory and Dynamical Systems **29**(4), 1055–1074 (2009)

[2] Alon, N., Dewdney, A.K., Ott, T.J.: Efficient simulation of finite automata by neural nets. Journal of the ACM **38**(2), 495–514 (1991)

[3] Baker, S.: Generalized golden ratios over integer alphabet. Integers **14**(A15) (2014)

[4] Baker, S.: On small bases which admit countably many expansions. Journal of Number Theory **147**, 515–532 (2015)

[5] Baker, S., Masáková, Z., Pelantová, E., Vávra, T.: On periodic representations in non-Pisot bases. arXiv:1604.03354v1 [math.NT] (2016)

[6] Balcázar, J.L., Gavaldà, R., Siegelmann, H.T.: Computational power of neural networks: A characterization in terms of Kolmogorov complexity. IEEE Transactions on Information Theory **43**(4), 1175–1183 (1997)

[7] Dombek, D., Masáková, Z., Pelantová, E.: Number representation using generalized $(-\beta)$-transformation. Theoretical Computer Science **412**(48), 6653–6665 (2011)

[8] Erdös, P., Joó, I., Komornik, V.: Characterization of the unique expansions $1 = \sum_{i=1}^{\infty} q^{-n_i}$ and related problems. Bulletin de la Société Mathématique de France **118**(3), 377–390 (1990)

[9] Frougny, C., Lai, A.C.: Negative bases and automata. Discrete Mathematics & Theoretical Computer Science **13**(1), 75–94 (2011)

[10] Glendinning, P., Sidorov, N.: Unique representations of real numbers in non-integer bases. Mathematical Research Letters **8**(4), 535–543 (2001)

[11] Goodfellow, I.J., Bengio, Y., Courville, A.C.: Deep Learning. Adaptive computation and machine learning. MIT Press (2016)

[12] Hare, K.G.: Beta-expansions of Pisot and Salem numbers. In: Proceedings of the Waterloo Workshop in Computer Algebra 2006: Latest Advances in Symbolic Algorithms, pp. 67–84. World Scientific (2007)

[13] Haykin, S.: Neural Networks: A Comprehensive Foundation, 2 edn. Prentice Hall (1998)

[14] Horne, B.G., Hush, D.R.: Bounds on the complexity of recurrent neural network implementations of finite state machines. Neural Networks **9**(2), 243–252 (1996)

[15] Indyk, P.: Optimal simulation of automata by neural nets. In: Proceedings of the STACS 1995 Twelfth Annual Symposium on Theoretical Aspects of Computer Science, *LNCS*, vol. 900, pp. 337–348 (1995)

[16] Kilian, J., Siegelmann, H.T.: The dynamic universality of sigmoidal neural networks. Information and Computation **128**(1), 48–56 (1996)

[17] Koiran, P.: A family of universal recurrent networks. Theoretical Computer Science **168**(2), 473–480 (1996)

[18] Komornik, V., Lai, A.C., Pedicini, M.: Generalized golden ratios of ternary alphabets. Journal of the European Mathematical Society **13**(4), 1113–1146 (2011)

[19] Komornik, V., Pedicini, M.: Critical bases for ternary alphabets. Acta Mathematica Hungarica **152**(1), 25–57 (2017)

[20] Kong, D., Li, W., Dekking, F.M.: Intersections of homogeneous cantor sets and beta-expansions. Nonlinearity **23**(11), 2815–2834 (2010)

[21] Maass, W.: Lower bounds for the computational power of networks of spiking neurons. Neural Computation **8**(1), 1–40 (1996)

[22] Minsky, M.: Computations: Finite and Infinite Machines. Prentice-Hall, Englewood Cliffs (1967)

[23] Parry, W.: On the β-expansions of real numbers. Acta Mathematica Hungarica **11**(3), 401–416 (1960)

[24] Pedicini, M.: Greedy expansions and sets with deleted digits. Theoretical Computer Science **332**(1-3), 313–336 (2005)

[25] Rényi, A.: Representations for real numbers and their ergodic properties. Acta Mathematica Academiae Scientiarum Hungaricae **8**(3-4), 477–493 (1957)

[26] Schmidt, K.: On periodic expansions of Pisot numbers and Salem numbers. Bulletin of the London Mathematical Society **12**(4), 269–278 (1980)

[27] Sidorov, N.: Almost every number has a continuum of β-expansions. The American Mathematical Monthly **110**(9), 838–842 (2003)

[28] Sidorov, N.: Expansions in non-integer bases: Lower, middle and top orders. Journal of Number Theory **129**(4), 741–754 (2009)

[29] Siegelmann, H.T.: Recurrent neural networks and finite automata. Journal of Computational Intelligence **12**(4), 567–574 (1996)

[30] Siegelmann, H.T.: Neural Networks and Analog Computation: Beyond the Turing Limit. Birkhäuser, Boston (1999)

[31] Siegelmann, H.T., Sontag, E.D.: Analog computation via neural networks. Theoretical Computer Science **131**(2), 331–360 (1994)

[32] Siegelmann, H.T., Sontag, E.D.: On the computational power of neural nets. Journal of Computer System Science **50**(1), 132–150 (1995)

[33] Šíma, J.: Analog stable simulation of discrete neural networks. Neural Network World **7**(6), 679–686 (1997)

[34] Šíma, J.: Energy complexity of recurrent neural networks. Neural Computation **26**(5), 953–973 (2014)

[35] Šíma, J.: The power of extra analog neuron. In: Proceedings of the TPNC 2014 Third International Conference on Theory and Practice of Natural Computing, *LNCS*, vol. 8890, pp. 243–254. Springer (2014)

[36] Šíma, J.: Neural networks between integer and rational weights. In: Proceedings of the IJCNN 2017 Thirties International Joint Conference on Neural Networks, pp. 154–161. IEEE (2017)

[37] Šíma, J., Orponen, P.: Continuous-time symmetric Hopfield nets are computationally universal. Neural Computation **15**(3), 693–733 (2003)
[38] Šíma, J., Orponen, P.: General-purpose computation with neural networks: A survey of complexity theoretic results. Neural Computation **15**(12), 2727–2778 (2003)
[39] Šíma, J., Savický, P.: Cut languages in rational bases. In: Proceedings of the LATA 2017 Eleventh International Conference on Language and Automata Theory and Applications, *LNCS*, vol. 10168, pp. 311–322. Springer (2017)
[40] Šíma, J., Wiedermann, J.: Theory of neuromata. Journal of the ACM **45**(1), 155–178 (1998)
[41] Šorel, M., Šíma, J.: Robust RBF finite automata. Neurocomputing **62**, 93–110 (2004)