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

Outline of Talk

1. The Neural Network Model
2. A Brief Overview of the Computational Power of Neural Networks
3. Integer-Weight Neural Networks with an Extra Rational-Weight Analog Neuron (1ANN)
4. Representations of Numbers in Non-Integer Bases
5. Quasi-Periodic Numbers
6. Classifying the 1ANNs Within the Chomsky Hierarchy
7. Conclusions

(Artificial) Neural Networks (NNs)

1. mathematical models of biological neural networks

- simulating and understanding the brain (The Human Brain Project)
- modeling cognitive functions

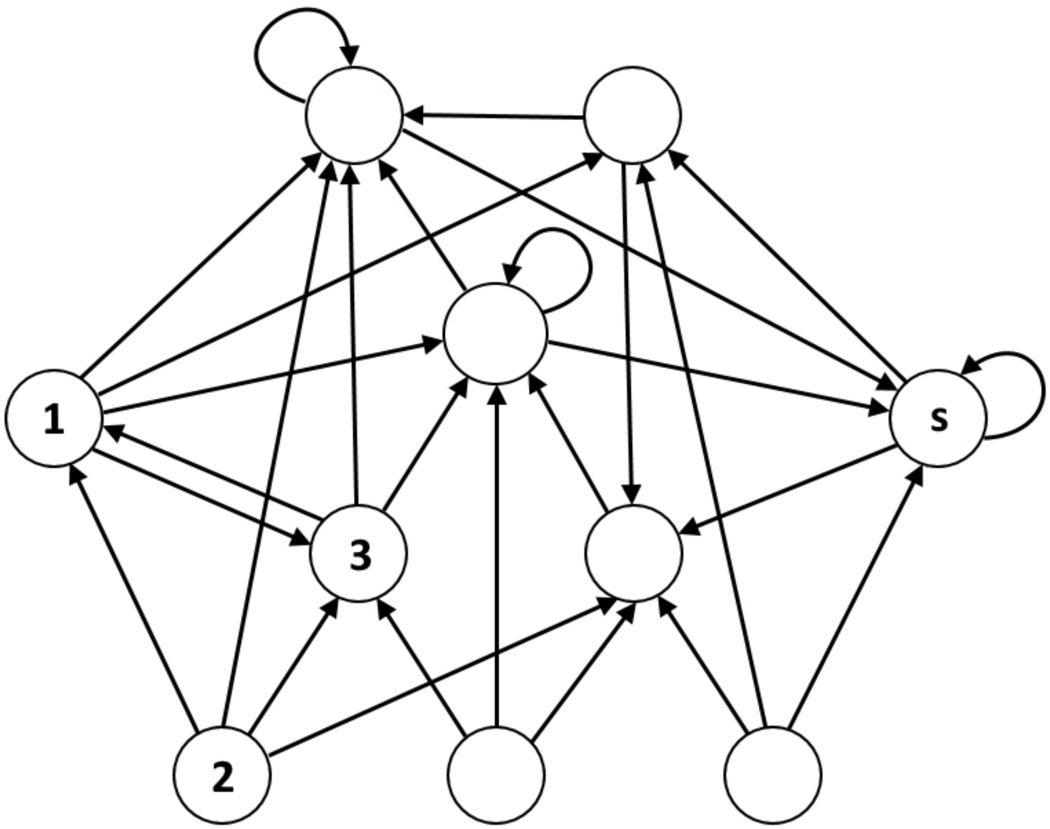
2. computing devices alternative to conventional computers

already first computer designers sought their inspiration in the human brain
(e.g., neurocomputer due to `Minsky, 1951`)

- common tools in machine learning or data mining (learning from training data)
- professional software implementations (e.g. `Matlab`, `Statistica` modules)
- successful commercial applications in AI (e.g. deep learning):
  pattern recognition, control, prediction, decision-making, signal analysis, fault detection, diagnostics, etc.

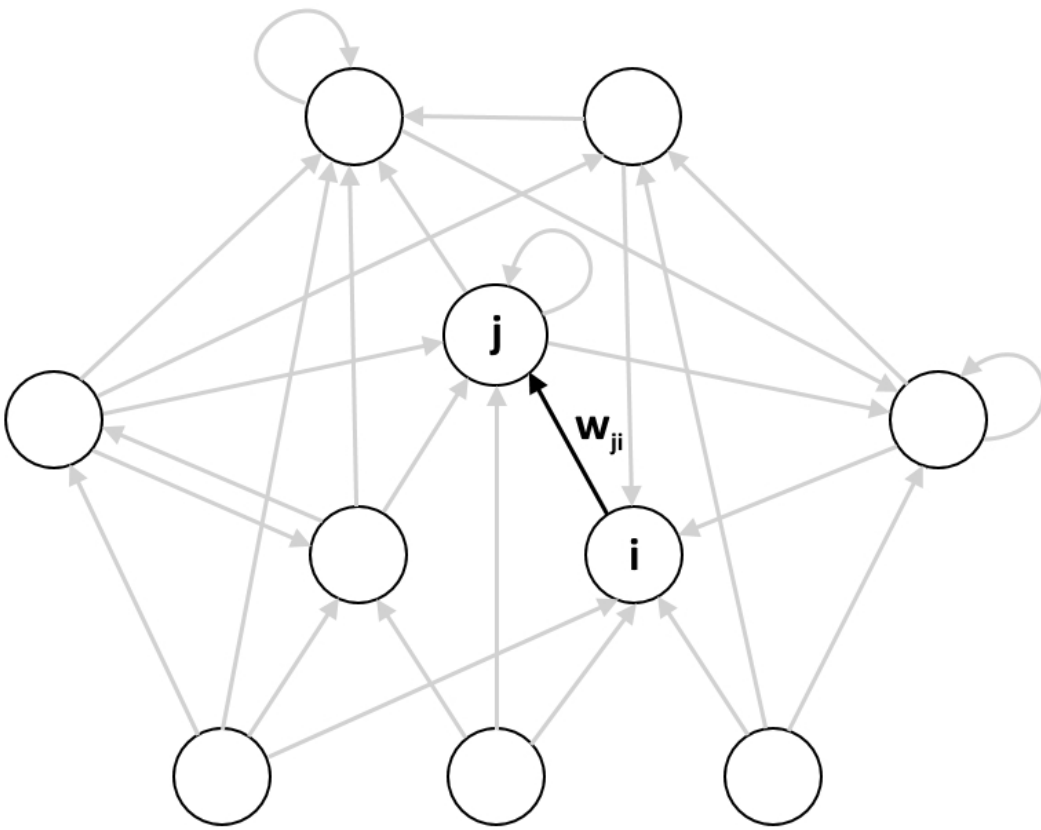
The Neural Network Model – Architecture

s computational units (neurons), indexed as $V = \{1, \ldots, s\}$, connected into a directed graph (V, A) where $A \subseteq V \times V$



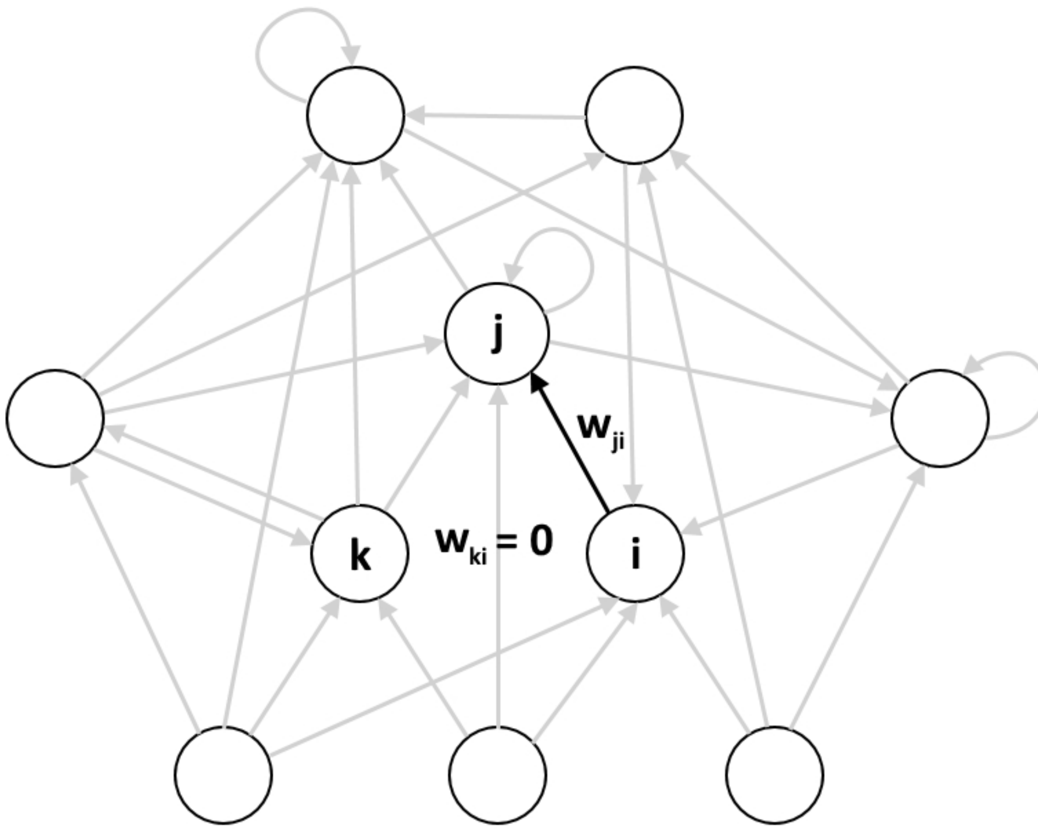
The Neural Network Model – Weights

each edge $(i, j) \in A$ from unit i to j is labeled with a real weight $w_{ji} \in \mathbb{R}$



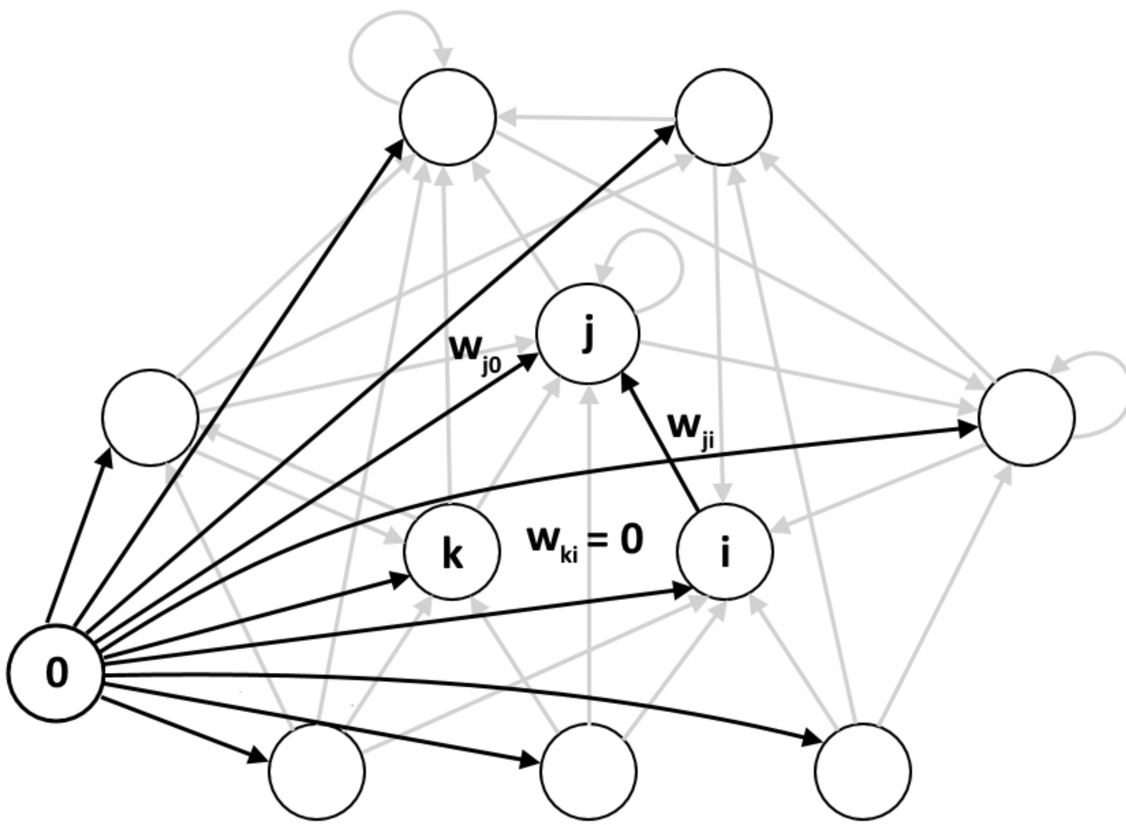
The Neural Network Model – Zero Weights

each edge $(i, j) \in A$ from unit i to j is labeled with a real weight $w_{ji} \in \mathbb{R}$ $(w_{ki} = 0$ iff $(i, k) \notin A)$



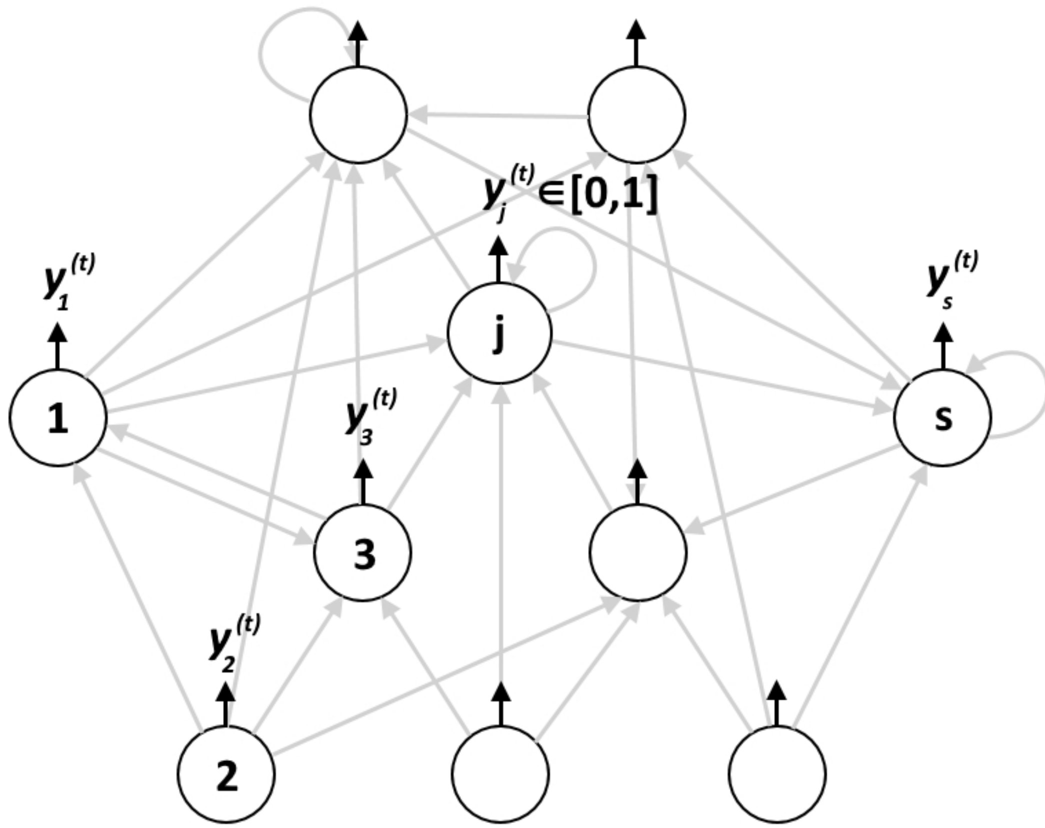
The Neural Network Model – Biases

each neuron $j \in V$ is associated with a real bias $w_{j0} \in \mathbb{R}$
(i.e. a weight of $(0, j) \in A$ from an additional formal neuron $0 \in V$)



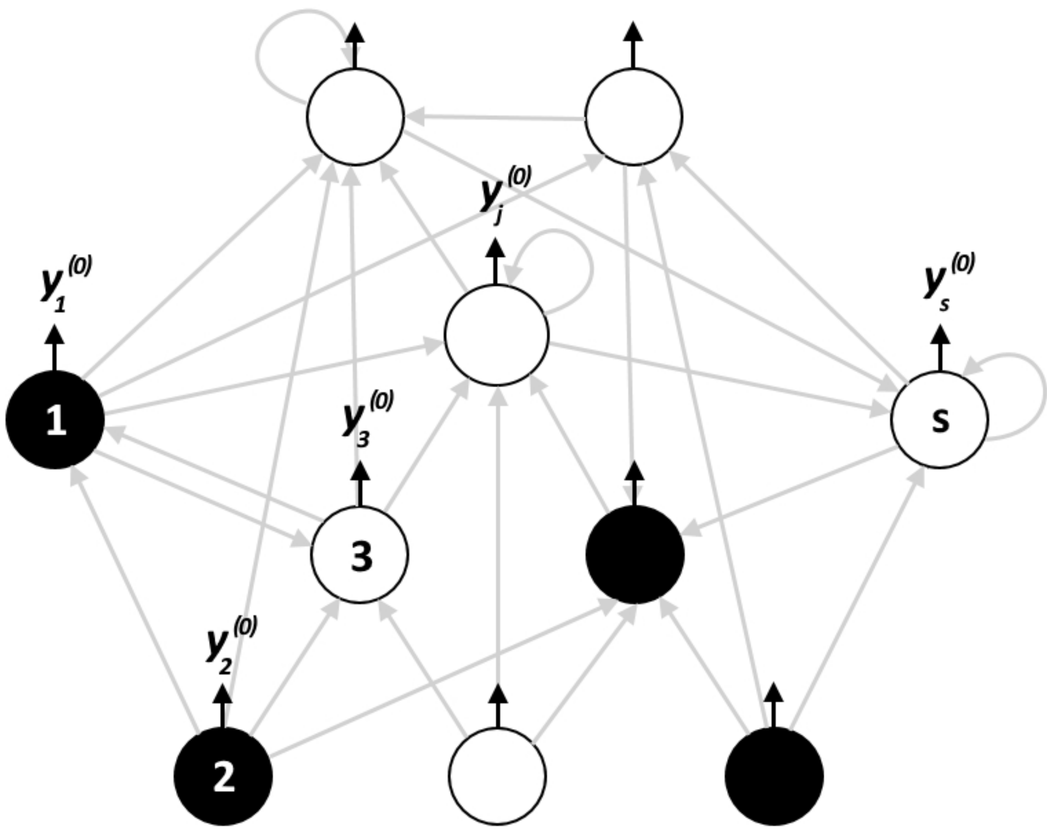
Discrete-Time Computational Dynamics – Network State

the evolution of global network state (output) $\mathbf{y}^{(t)} = (y_1^{(t)}, \ldots, y_s^{(t)}) \in [0,1]^s$
at discrete time instant $t = 0, 1, 2, \ldots$



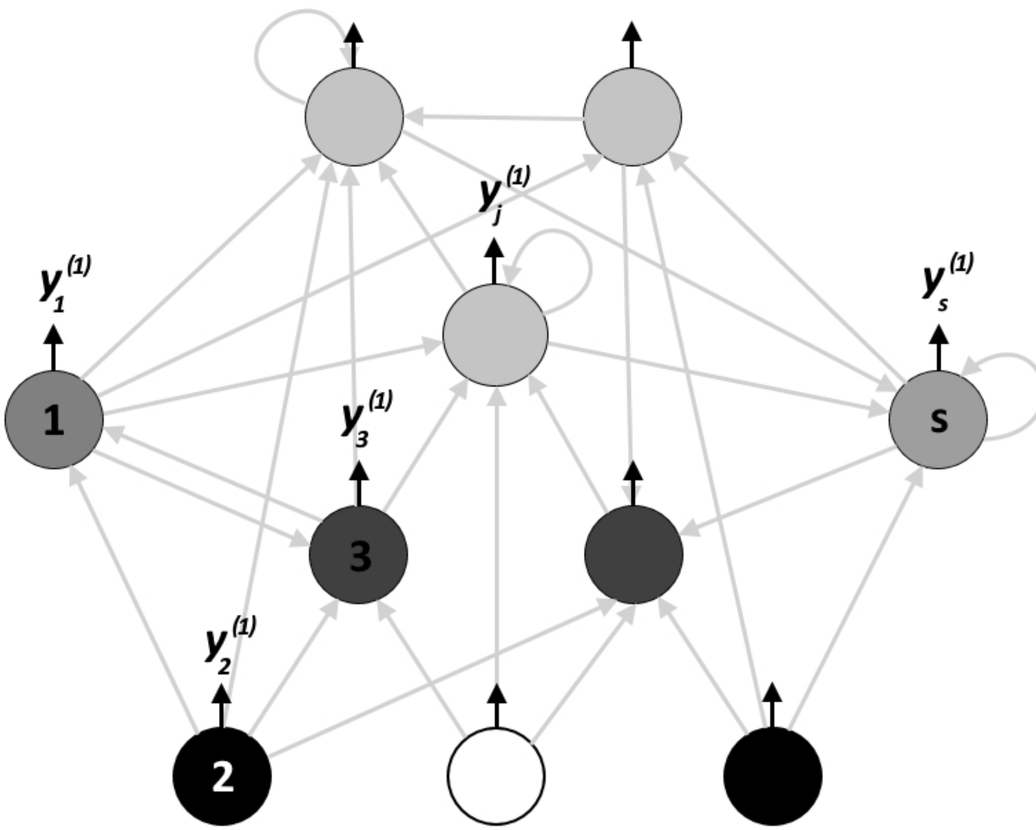
Discrete-Time Computational Dynamics – Initial State

$t = 0$: initial network state $\mathbf{y}^{(0)} \in \{0,1\}^s$



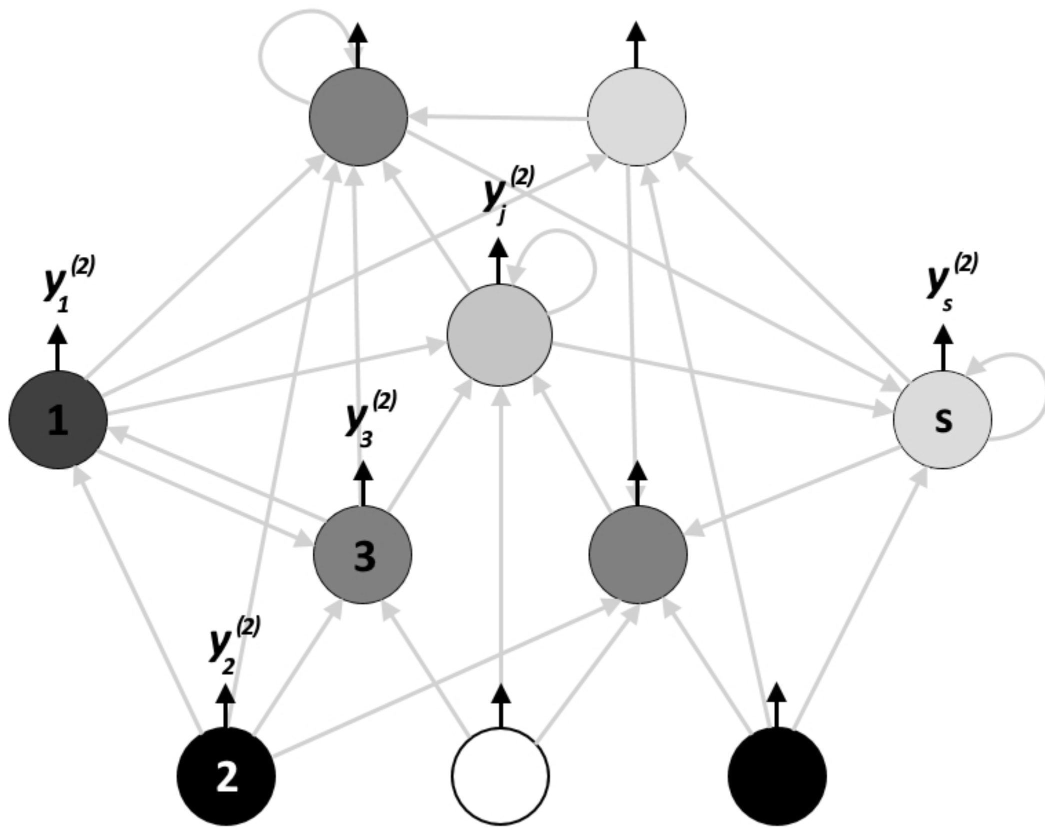
Discrete-Time Computational Dynamics: $t = 1$

$t = 1$: network state $\mathbf{y}^{(1)} \in [\mathbf{0}, \mathbf{1}]^s$



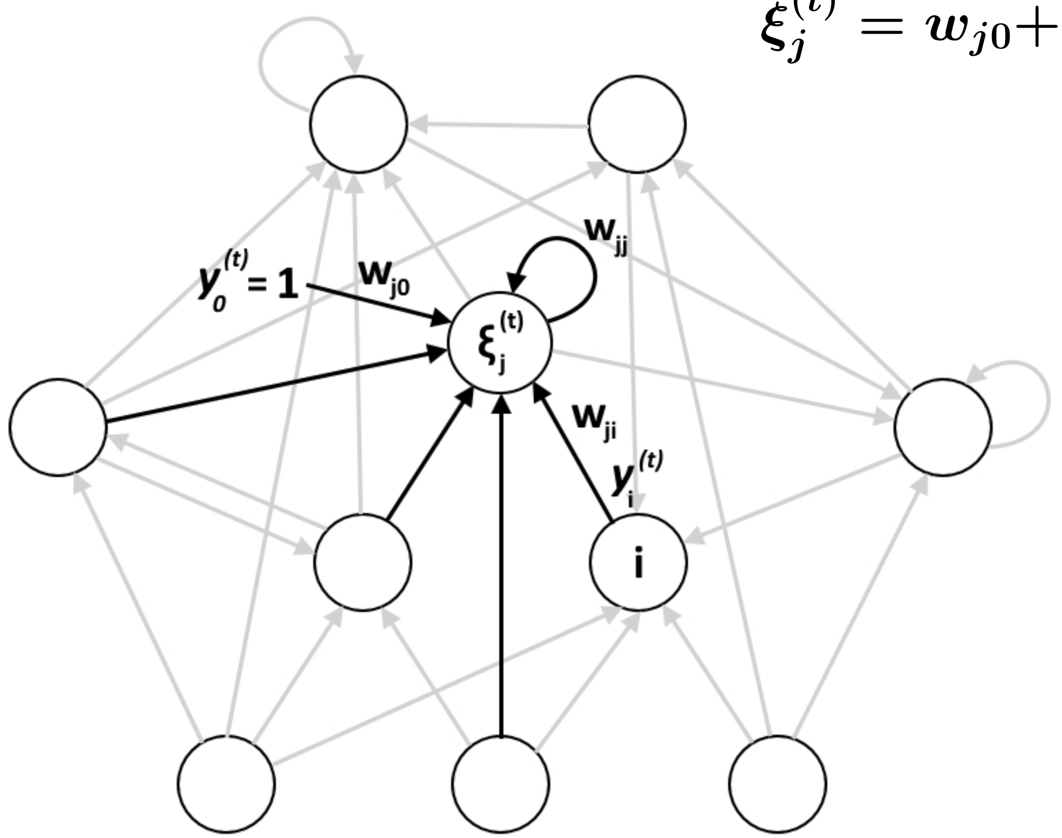
Discrete-Time Computational Dynamics: $t = 2$

$t = 2$: network state $\mathbf{y}^{(2)} \in [0,1]^s$



Discrete-Time Computational Dynamics – Excitations

at discrete time instant $t \geq 0$, an excitation is computed as



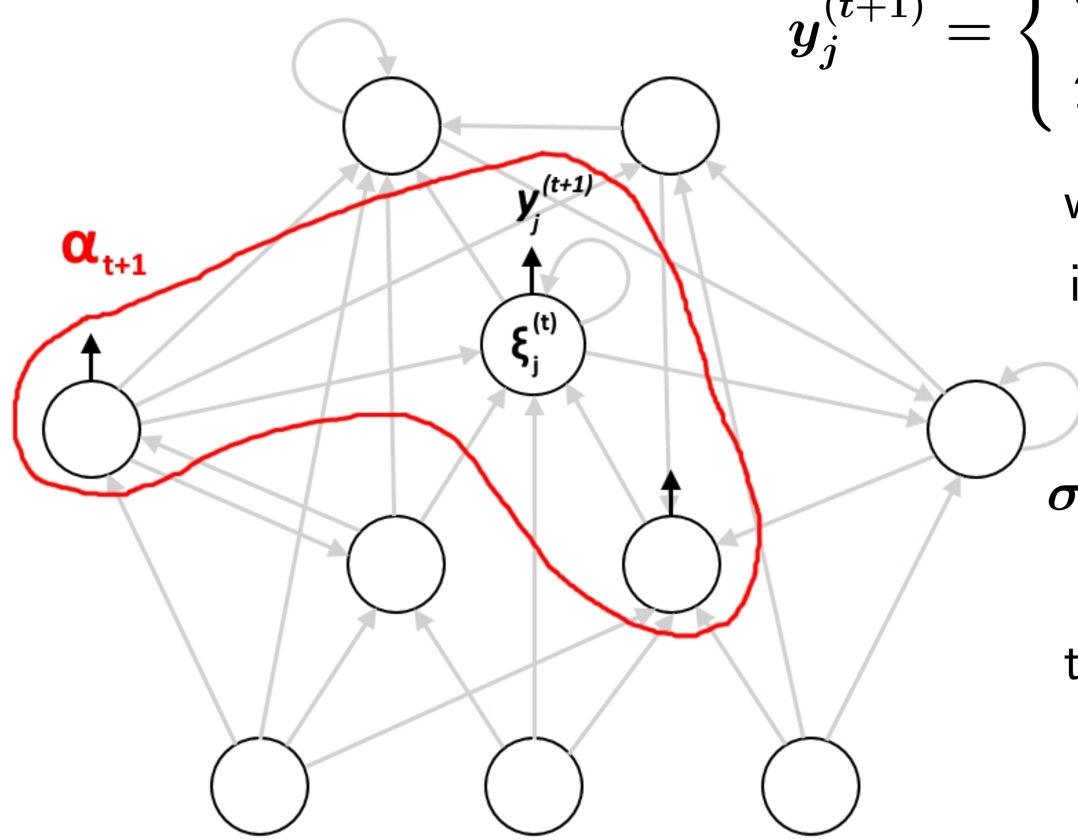
$$\xi_j^{(t)} = w_{j0} + \sum_{i=1}^{s} w_{ji} y_i^{(t)} = \sum_{i=0}^{s} w_{ji} y_i^{(t)}$$

for $j = 1, \ldots, s$

where unit $0 \in V$ has constant output $y_0^{(t)} \equiv 1$ for every $t \geq 0$

Discrete-Time Computational Dynamics – Outputs

at the next time instant $t+1$, only the neurons $j \in \alpha_{t+1}$ from a selected subset $\alpha_{t+1} \subseteq V$ update their states:



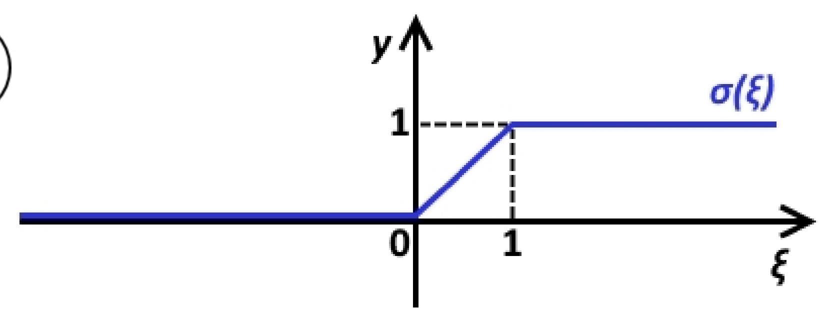
$$y_j^{(t+1)} = \begin{cases} \sigma\left(\xi_j^{(t)}\right) & \text{for } j \in \alpha_{t+1} \\ y_j^{(t)} & \text{for } j \in V \setminus \alpha_{t+1} \end{cases}$$

where $\sigma : \mathbb{R} \longrightarrow [0,1]$

is an activation function, e.g.

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

the saturated-linear function



The Computational Power of NNs – Motivations

- the potential and limits of general-purpose computation with NNs:

  What is ultimately or efficiently computable by particular NN models?

- idealized mathematical models of practical NNs which abstract away from implementation issues, e.g. analog numerical parameters are true real numbers

- methodology: the computational power and efficiency of NNs is investigated by comparing formal NNs to traditional computational models such as finite automata, Turing machines, Boolean circuits, etc.

- NNs may serve as reference models for analyzing alternative computational resources (other than time or memory space) such as analog state, continuous time, energy, temporal coding, etc.

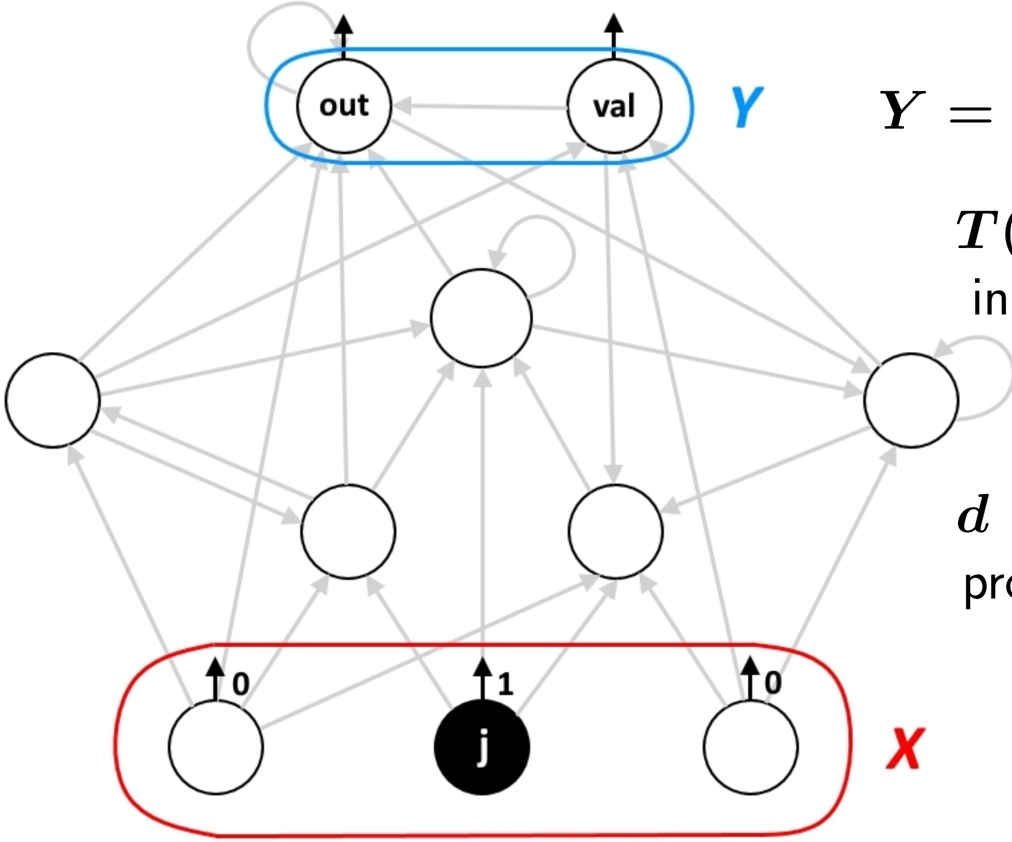
- NNs capture basic characteristics of biological nervous systems (plenty of densely interconnected simple unreliable computational units)

  $\longrightarrow$ computational principles of mental processes

Neural Networks As Formal Language Acceptors

language (problem) $L \subseteq \Sigma^*$ over a finite alphabet Σ

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



$Y = \{\text{out}, \text{val}\}$ output neurons

$T(n)$ is the computational time in terms of input length $n \geq 0$

$d \geq 1$ is the time overhead for processing a single input symbol

$X = \text{enum}(\Sigma) \subseteq V$ input neurons

$$\uparrow \quad y_j^{(d(i-1)+k)} = 1 \quad \text{iff } j = \text{enum}(x_i)$$

$$\mathbf{x} = x_1 x_2 \ldots x_{i-1} \longleftarrow x_i \longleftarrow x_{i+1} x_{i+2} \ldots x_n \in \Sigma^*$$ input word

The Computational Power of NNs – Integer Weights

depends on the information content of weight parameters:

1. **integer** weights: finite automaton (FA) (Minsky, 1967)

$$w_{ji} \in \mathbb{Z} \quad \longrightarrow \quad \text{excitations} \quad \xi_j \in \mathbb{Z} \quad \longrightarrow \quad \text{states} \quad y_j \in \{0,1\}$$

$$\longrightarrow \quad 2^s \text{ global NN states } \mathbf{y} \in \{0,1\}^s \quad \sim \quad \text{FA states}$$

size-optimal implementations:

- $\Theta\left(\sqrt{m}\right)$ neurons for a deterministic FA with m states (Indyk, 1995; Horne, Hush, 1995)
- $\Theta(m)$ neurons for a regular expression of length m (Šíma, Wiedermann 1998)

The Computational Power of NNs – Rational Weights

depends on the information content of weight parameters:

2. **rational** weights: Turing machine (Siegelmann, Sontag, 1995)

- $w_{ji} \in \mathbb{Q}$ are fractions $\frac{p}{q}$ where $p \in \mathbb{Z}$, $q \in \mathbb{N}$
- NNs compute algorithmically solvable problems
- real-time simulation of TMs $\longrightarrow$ polynomial time $\equiv$ complexity class P
- a universal NN with 25 neurons (Indyk, 1995)

  $\longrightarrow$ the halting problem of whether a small NN terminates its computation, is algorithmically undecidable

The Computational Power of NNs – Real Weights

depends on the information content of weight parameters:

3. arbitrary **real** weights: "super-Turing" computation (Siegelmann, Sontag, 1994)

- $w_{ji} \in \mathbb{R}$, e.g. irrational weights $\sqrt{2}$, π
- infinite precision of **ONE** real weight (vs. an algorithm has a finite description) can encode any function f: $\quad 0\,.\,\text{code}(C_1)\,\text{code}(C_2)\,\text{code}(C_3)\ldots$

  ($\text{code}(C_n)$ encodes the circuit C_n computing f for inputs of length n)

  $\longrightarrow$ exponential time $\equiv$ any I/O mapping
  (including algorithmically undecidable problems)

- polynomial time $\equiv$ nonuniform complexity class P/poly:

  problems solvable by a polynomial-time (P) algorithm that for input $x \in \Sigma^*$ of length $n = |x|$, receives an external advise: a string $s(n) \in \Sigma^*$ of polynomial length $|s(n)| = O(n^c)$ (poly), which depends only on n

The Computational Power of NNs – Rough Overview

depends on the information content of weight parameters:

1. **integer** weights: finite automaton
2. **rational** weights: Turing machine

   polynomial time $\equiv$ complexity class P
3. arbitrary **real** weights: "super-Turing" computation

   polynomial time $\equiv$ nonuniform complexity class P/poly

   exponential time $\equiv$ any I/O mapping

Neural Networks Between Rational and Real Weights

1. **integer** weights: finite automaton

2. **rational** weights: Turing machine

   polynomial time $\equiv$ **P**

   polynomial time & increasing **Kolmogorov complexity** of real weights:

   the length of the shortest program (in a fixed programming language) that produces a real weight,

   e.g. $K\left(\text{“}\sqrt{2}\text{”}\right) = O(1)$, $K(\text{“random strings”}) = n + O(1)$

   $\equiv$ a proper hierarchy of nonuniform complexity classes **between P and P/poly**

   (Balcázar, Gavaldà, Siegelmann, 1997)

3. arbitrary **real** weights: “super-Turing” computation

   polynomial time $\equiv$ **P/poly**

Neural Networks Between Integer and Rational Weights

1. **integer** weights: finite automata $\equiv$ **regular (Type-3) languages**

a gap between integer and rational weights w.r.t. the Chomsky hierarchy:

pushdown automata $\equiv$ context-free (Type-2) languages

linear-bounded automata (NSPACE(O(n))) $\equiv$ context-sensitive (Type-1) languages

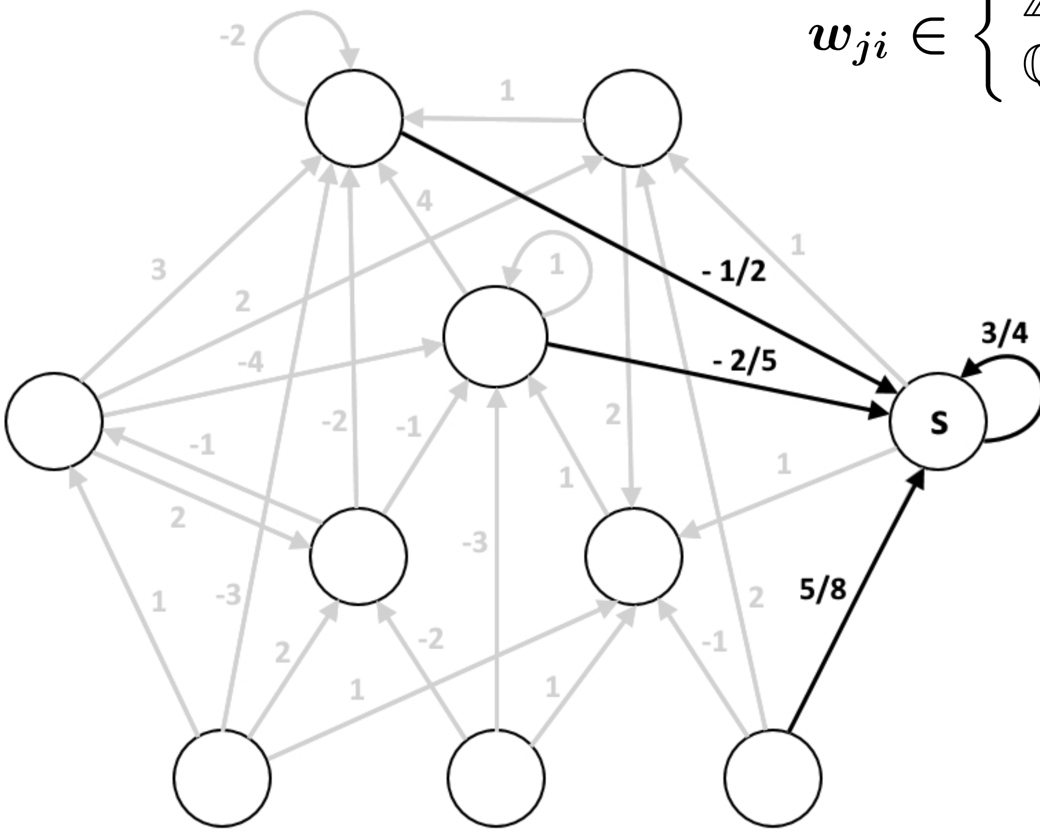
2. **rational** weights: Turing machines $\equiv$ **recursively enumerable (Type-0) lang.**

TWO analog neurons with **rational weights** + a few **integer-weight** neurons can implement a 2-stack pushdown automaton $\equiv$ Turing machine

$\longrightarrow$ **What is the computational power of ONE extra analog neuron ?**

A Neural Network with an Extra Analog Neuron (1ANN)

all the weights to neurons are integers except for ONE neuron s with rational weights:



$$w_{ji} \in \begin{cases} \mathbb{Z} & j = 1, \ldots, s-1 \\ \mathbb{Q} & j = s \end{cases}$$

$$i \in \{0, \ldots, s\}$$

The Representation Theorem for 1ANNs (Šíma, IJCNN 2017)

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

$$L = h\left(\left(\left(\left(\bigcup_{r=0}^{p} \left(\overline{L_{<c_r}} \cap L_{<c_{r+1}}\right)^R \cdot A_r\right)^{Pref} \cap R_0\right)^* \cap R\right)\right)$$

(options: $\overline{L_{>0}}$, $L_{>c_r} \cap L_{<c_{r+1}}$, $L_{>c_r} \cap \overline{L_{>c_{r+1}}}$, $\overline{L_{<c_r}} \cap \overline{L_{<c_{r+1}}}$, $\overline{L_{<1}}$)

where

- $A = \left\{ \sum_{i=0}^{s-1} w_{si} y_i \,\middle|\, y_1, \ldots, y_{s-1} \in \{0,1\} \right\} \cup \{0,1\} \subset \mathbb{Q}$ is a finite alphabet of (rational) digits
- $h : A^* \longrightarrow \Sigma^*$ is a letter-to-letter morphism
- $R\,, R_0 \subseteq A^*$ are regular languages
- S^{Pref} denotes the largest prefix-closed subset of $S \cup A \cup \{\varepsilon\}$
- $A_1, \ldots, A_p$ is a partition of a finite alphabet A
- K^R denotes the reversal of language K

The Representation Theorem for 1ANNs (Šíma, IJCNN 2017)

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

$$L = h\left(\left(\left(\bigcup_{r=0}^{p} \left(\overline{L_{<c_r}} \cap L_{<c_{r+1}}\right)^R \cdot A_r\right)^{Pref} \cap R_0\right)^* \cap R\right)$$

where (continued)

- $L_{<c_r}\,,\ L_{>c_r} \subseteq A^*$ are so-called **cut languages** over digit alphabet A,

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

- $0 = c_1 \leq c_2 \leq \cdots \leq c_p = 1$ are (rational) thresholds such that

$$C = \{c_1, \ldots, c_p\} = \left\{ -\sum_{i=0}^{s-1} \frac{w_{ji}}{w_{js}} y_i \;\middle|\; j \in V \setminus (X \cup \{s\}) \text{ s.t. } w_{js} \neq 0\,, \; y_1, \ldots, y_{s-1} \in \{0,1\} \right\} \cup \{0,1\} \subset \mathbb{Q}$$

- $\beta = \frac{1}{w_{ss}} \in \mathbb{Q}$ is called a (rational) base (radix) $\longrightarrow$ $|\beta| > 1$

Representations of Numbers in Non-Integer Bases

non-standard positional numeral systems: a base and/or digits need not be integers

- $\beta \in \mathbb{R}$ is a real base (radix) such that $|\beta| > 1$
- $\emptyset \neq A \subset \mathbb{R}$ is a finite set of real digits

a word (string) composed of digits $a_1 \dots a_n \in A^*$ (the radix point omitted), called a finite β-expansion, represents a number in base β as

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

Examples:

1. $\beta = 10$, $A = \{0, 1, 2, \dots, 9\}$
   decimal expansion 75 represents $\frac{3}{4} = (0\,.\,75)_{10} = 7 \cdot 10^{-1} + 5 \cdot 10^{-2}$
2. $\beta = 2$, $A = \{0, 1\}$
   binary expansion 11 represents $\frac{3}{4} = (0\,.\,11)_2 = 1 \cdot 2^{-1} + 1 \cdot 2^{-2}$
3. $\beta = \frac{5}{2}$, $A = \left\{\frac{5}{16}, \frac{7}{4}\right\}$
   $\frac{5}{2}$-expansion $\frac{7}{4}\ \frac{5}{16}$ represents $\frac{3}{4} = \left(0\,.\,\frac{7}{4}\ \frac{5}{16}\right)_{\frac{5}{2}} = \frac{7}{4} \cdot \left(\frac{5}{2}\right)^{-1} + \frac{5}{16} \cdot \left(\frac{5}{2}\right)^{-2}$

Finite β-Expansions & Cut Languages

a cut language $L_{<c}$ contains all the finite β-expansions $a_1 \ldots a_n \in A^*$ of numbers that are less than a threshold $c \in \mathbb{R}$ (similarly for $L_{>c}$):

$$L_{<c} = \left\{ a_1 \ldots a_n \in A^* \,\middle|\, (0 \,.\, a_1 \ldots a_n)_\beta = \sum_{k=1}^{n} a_k \beta^{-k} < c \right\}$$

$\beta \in \mathbb{Q}$, $A \subset \mathbb{Q}$: $L_{<c}$ is composed of finite β-expansions of a Dedekind cut

(Infinite) β-Expansions (Rényi, 1957; Parry, 1960)

an infinite word composed of digits $a_1 a_2 a_3 \cdots \in A^\omega$ is a β-expansion of number

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

which is a convergent power series due to $|\beta| > 1$

Example: $\beta = \frac{3}{2}$, $A = \{0, 1\}$

$\frac{3}{2}$-expansion $000(10)^\omega = 000\,10\,10\,10\,10\,10\,\ldots \in \{0,1\}^\omega$ represents the number

$$(0\,.\,000\,10\,10\,10\,10\,10\,\ldots)_{\frac{3}{2}} = \left(\frac{3}{2}\right)^{-4} + \left(\frac{3}{2}\right)^{-6} + \left(\frac{3}{2}\right)^{-8} + \cdots$$

$$= \sum_{k=2}^{\infty} \left(\frac{3}{2}\right)^{-2k} = \sum_{k=2}^{\infty} \left(\frac{4}{9}\right)^{k} = \frac{\frac{16}{81}}{1 - \frac{4}{9}} = \frac{16}{45}$$

Uniqueness of β-Expansions for Integer Base β

for an integer base $\beta > 0$ and the standard digits, $A = \{0, 1, \ldots, \beta - 1\}$,

almost any number from the interval $(0, 1)$ has a **unique** β-expansion,

e.g. the decimal expansion $70710678118\ldots \in \{0, 1, 2, \ldots, 9\}^\omega$ of

$$\frac{\sqrt{2}}{2} = (0\,.\,70710678118\ldots)_{10}$$

except for those with a finite β-expansion, which have **two distinct** β-expansions,

e.g. two decimal expansions $750^\omega = 75000\ldots$, $749^\omega = 74999\ldots$ of

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

Uniqueness of β-Expansions for Non-Integer Base β

for a non-integer base β, almost every number has **infinitely (uncountably) many distinct** β-expansions (Sidorov, 2003)

Example: $1 < \beta < 2, \quad A = \{0, 1\}, \quad D_\beta = \left(0\,,\, \frac{1}{\beta - 1}\right)$

- $1 < \beta < \varphi$ where $\varphi = (1 + \sqrt{5})/2 \approx 1.618034$ is the golden ratio:
  every $x \in D_\beta$ has uncountably many distinct β-expansions (Erdös et al.,1990)

- $\varphi \leq \beta < q$ where $q \approx 1.787232$ is the Komornik-Loreti constant:
  countably many $x \in D_\beta$ have unique β-expansions (Glendinning,Sidorov,2001)

  e.g. $\beta = \frac{5}{3} = 1\,.\,\overline{6} = 1\,.\,666\ldots \quad \longrightarrow \quad D_{\frac{5}{3}} = \left(0\,,\, \frac{3}{2}\right)$

  the infinite word $0^k(10)^\omega$ $(k \geq 0)$ represents a unique $\frac{5}{3}$-expansion of

  $$\left(0\,.\,\underbrace{0\,\ldots\,0}_{k\text{ times}}\,10\,10\,10\,10\,10\,\ldots\right)_{\frac{5}{3}} = \left(\tfrac{3}{5}\right)^{k-1} \cdot \frac{9}{16}$$

  vs. $\beta = \varphi = (1 + \sqrt{5})/2 \approx 1.618034 \quad \longrightarrow \quad D_\varphi = (0\,,\,\varphi)$

  countably many distinct φ-expansions $(10)^k 110^\omega$, $(10)^\omega$, $(10)^k 01^\omega$ $(k \geq 0)$
  of the number 1, e.g. $1 = \left(0\,.\,\overline{10}\right)_\varphi = (0\,.\,10\,10\,10\,10\,10\,\ldots)_\varphi$

- $q \leq \beta < 2$: uncountably many $x \in D_\beta$ have unique β-expansions

Eventually Periodic β-Expansions

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

- $a_1a_2\dots a_{k_1} \in A^{k_1}$ is a preperiodic part of length $k_1 \geq 0$ (purely periodic β-expansions meet $k_1 = 0$)
- $a_{k_1+1}a_{k_1+2}\dots a_{k_2} \in A^m$ is a repetend of length $m = k_2 - k_1 > 0$ whose minimum is the period of β-expansion
- $(0\,.\,a_1a_2\dots a_{k_1}\,\overline{a_{k_1+1}a_{k_1+2}\dots a_{k_2}})_\beta = (0\,.\,a_1a_2\dots a_{k_1})_\beta + \beta^{-k_1}\varrho$

  where $\varrho = (0\,.\,\overline{a_{k_1+1}a_{k_1+2}\dots a_{k_2}})_\beta = \dfrac{\sum_{k=1}^{m} a_{k_1+k}\beta^{-k}}{1-\beta^{-m}}$ is a periodic point

Example: $\beta = \frac{3}{2}$, $A = \{0,1\}$, $1\,(10)^\omega = 1\,10\,10\,10\,10\,10\,\dots$

$$\frac{22}{15} = (0\,.\,1\,\overline{10})_{\frac{3}{2}} = (0\,.\,1)_{\frac{3}{2}} + \left(\frac{3}{2}\right)^{-1}\cdot(0\,.\,\overline{10})_{\frac{3}{2}} = \left(\frac{3}{2}\right)^{-1} + \left(\frac{3}{2}\right)^{-1}\cdot\varrho$$

where $\varrho = (0\,.\,\overline{10})_{\frac{3}{2}} = \displaystyle\sum_{k=0}^{\infty}\left(\frac{3}{2}\right)^{-2k-1} = \frac{1\cdot\left(\frac{3}{2}\right)^{-1} + 0\cdot\left(\frac{3}{2}\right)^{-2}}{1-\left(\frac{3}{2}\right)^{-2}} = \frac{6}{5}$

Eventually Quasi-Periodic β-Expansions

β-expansion $a_1 \ldots a_{k_1}\, a_{k_1+1} \ldots a_{k_2}\, a_{k_2+1} \ldots a_{k_3}\, a_{k_3+1} \ldots a_{k_4} \ldots \in A^\omega$

is eventually quasi-periodic if there is $0 \le k_1 < k_2 < \cdots$ such that

$$\varrho = (0\,.\,\overline{a_{k_1+1} \ldots a_{k_2}})_\beta = (0\,.\,\overline{a_{k_2+1} \ldots a_{k_3}})_\beta = (0\,.\,\overline{a_{k_3+1} \ldots a_{k_4}})_\beta = \cdots$$

- $a_1 a_2 \ldots a_{k_1} \in A^{k_1}$ is a preperiodic part of length k_1 (purely quasi-periodic β-expansions meet $k_1 = 0$)
- $a_{k_i+1} \ldots a_{k_{i+1}} \in A^{m_i}$ is a quasi-repetend of length $m_i = k_{i+1} - k_i > 0$
- $(0\,.\,a_1 a_2 a_2 \ldots)_\beta = (0\,.\,a_1 a_2 \ldots a_{k_1})_\beta + \beta^{-k_1}\varrho$ where for every $i \ge 1$,

$$(0\,.\,\overline{a_{k_i+1} \ldots a_{k_{i+1}}})_\beta = \frac{\sum_{k=1}^{m_i} a_{k_i+k}\beta^{-k}}{1 - \beta^{-m_i}} = \varrho \quad \text{is a periodic point}$$

$\longrightarrow$ quasi-repetends can be mutually replaced with each other arbitrarily

- a generalization of eventually periodic β-expansions:

$$a_{k_1+1} \ldots a_{k_2} = a_{k_2+1} \ldots a_{k_3} = a_{k_3+1} \ldots a_{k_4} = \cdots$$

An Example of Quasi-Periodic β-Expansion

base $\boldsymbol{\beta} = \frac{5}{2}$, digits $\boldsymbol{A} = \left\{0, \frac{1}{2}, \frac{7}{4}\right\}$, periodic point $\boldsymbol{\varrho} = \frac{3}{4}$

$$\left(0\,.\,\overline{\tfrac{7}{4}\;0}\right)_{\frac{5}{2}} = \left(0\,.\,\overline{\tfrac{7}{4}\;\tfrac{1}{2}\;0}\right)_{\frac{5}{2}} = \left(0\,.\,\overline{\tfrac{7}{4}\;\tfrac{1}{2}\;\tfrac{1}{2}\;0}\right)_{\frac{5}{2}} = \left(0\,.\,\overline{\tfrac{7}{4}\;\tfrac{1}{2}\;\tfrac{1}{2}\;\tfrac{1}{2}\;0}\right)_{\frac{5}{2}} = \cdots$$

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

$\longrightarrow$ $\varrho = \frac{3}{4}$ has uncountably many distinct quasi-periodic $\frac{5}{2}$-expansions:

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

where $n_1, n_2, n_3, \ldots$ is any infinite sequence of nonnegative integers

Quasi-Periodic Numbers

$c \in \mathbb{R}$ is β-quasi-periodic within A if every infinite β-expansion of c is eventually quasi-periodic

Examples:

- c from the complement of the Cantor set **is** 3-quasi-periodic within $\{0,2\}$:
  c has no β-expansion at all

- $c = \frac{3}{4}$ **is** $\frac{5}{2}$-quasi-periodic within $A = \{0\,,\,\frac{1}{2}\,,\,\frac{7}{4}\}$:
  all the $\frac{5}{2}$-expansions of $\frac{3}{4}$ using digits from A, are eventually quasi-periodic

- $c = \frac{40}{57} = (0\,.\,0\,\overline{011})_{\frac{3}{2}}$ **is not** $\frac{3}{2}$-quasi-periodic within $A = \{0,1\}$:
  greedy (i.e. lexicographically maximal) $\frac{3}{2}$-expansion $100000001\ldots$ of $\frac{40}{57}$ is not eventually periodic

Cut Languages Within the Chomsky Hierarchy

(Šíma, Savický, LATA 2017)

$$L_{<c} = \left\{ a_1 \dots a_n \in A^* \;\middle|\; (0\,.\,a_1 \dots a_n)_\beta = \sum_{k=1}^{n} a_k \beta^{-k} < c \right\}$$

Theorem 1 *A cut language $L_{<c}$ is regular* **iff** *c is β-quasi-periodic within A.*

Theorem 2 *Let $\beta \in \mathbb{Q}$ and $A \subset \mathbb{Q}$. Every cut language $L_{<c}$ with threshold $c \in \mathbb{Q}$ is context-sensitive.*

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

Corollary 1 *Any cut language $L_{<c}$ is either regular or non-context-free (depending on whether c is a β-quasi-periodic number within A).*

The Computational Power of 1ANNs (Šíma, IJCNN 2017)

applying the results on cut languages to the **representation theorem** for 1ANNs:

$$L = h\left(\left(\left(\bigcup_{r=0}^{p}\left(\overline{L_{<c_r}} \cap L_{<c_{r+1}}\right)^R \cdot A_r\right)^{Pref} \cap R_0\right)^* \cap R\right)$$

Theorem 4 *Let N be a 1ANN and assume $0 < |w_{ss}| < 1$. Define $\beta \in \mathbb{Q}$, $A \subset \mathbb{Q}$, and $C \subset \mathbb{Q}$ as in the representation theorem using the weights of N:*

$$\beta = \frac{1}{w_{ss}}, \quad A = \left\{\sum_{i=0}^{s-1} w_{si} y_i \;\middle|\; y_1, \ldots, y_{s-1} \in \{0,1\}\right\} \cup \{0,1\},$$

$$C = \{c_1, \ldots, c_p\} = \left\{-\sum_{i=0}^{s-1} \frac{w_{ji}}{w_{js}} y_i \;\middle|\; j \in V \setminus (X \cup \{s\}) \text{ s.t. } w_{js} \neq 0, \; y_1, \ldots, y_{s-1} \in \{0,1\}\right\} \cup \{0,1\}.$$

If every $c \in C$ is β-quasi-periodic within A, then N accepts regular language.

Theorem 5 *There is a language accepted by a 1ANN, which is not context-free.*

Theorem 6 *Any language accepted by a 1ANN is context-sensitive.*

NNs Between Integer and Rational Weights & the Chomsky Hierarchy

rational-weight NNs $\equiv$ TMs $\equiv$ recursively enumerable languages **(Type-0)**

1ANNs $\subset$ LBA $\equiv$ context-sensitive languages **(Type-1)**

1ANNs $\not\subset$ PDA $\equiv$ context-free languages **(Type-2)**

integer-weight NNs $\equiv$ "quasi-periodic" 1ANNs $\equiv$ FA $\equiv$ regular languages **(Type-3)**

Conclusions

- we have presented a brief survey of results on the computational power of NNs
- we have characterized the class of languages accepted by 1ANNs—integer-weight NNs with an extra rational-weight analog neuron, using cut languages
- we have shown an interesting link to active research on β-expansions in non-integer bases
- we have introduced the notion of quasi-periodic numbers
- we have refined the analysis of the computational power of NNs between integer and rational weights within the Chomsky hierarchy

Open Problems

- a necessary condition when a 1ANN accepts a regular language
- the analysis for $w_{ss} \in \mathbb{R}$ or $|w_{ss}| > 1$
- a proper hierarchy of 1ANNs, e.g. with increasing quasi-period of weights