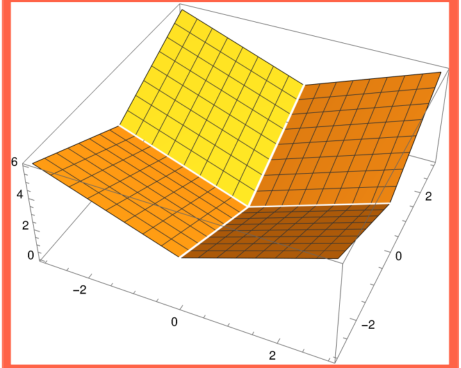
Equivariant neural networks and representation theory

Or: Beyond linearity

~~Accept~~Change what you cannot ~~change~~accept



I report on work of Gibson and Williamson

©Daniel Tubbenhauer

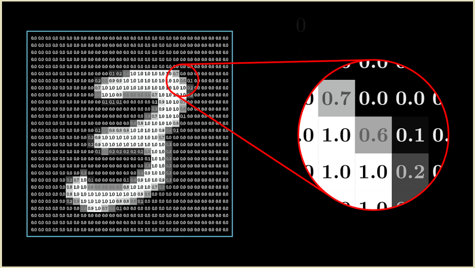
What is a Neural Network (NN)?

- Task: Identify handwritten digits
- We can see this as a function in the following way:
  - Convert the pictures into grayscale values, e.g., 28×28 grid of numbers
  - Flatten the result into a vector, e.g., $28 \times 28 \mapsto$ a vector with $28^2 = 784$ entries
  - The output is a vector with 10 entries
- We thus have a function $\mathbb{R}^{784} \to \mathbb{R}^{10}$

What is a Neural

- Task: Identify
- We can see thi
  - Convert th … 28 grid of numbers
  - Flatten the … ith $28^2 = 784$ entries
  - The outpu
- We thus have a

Input example

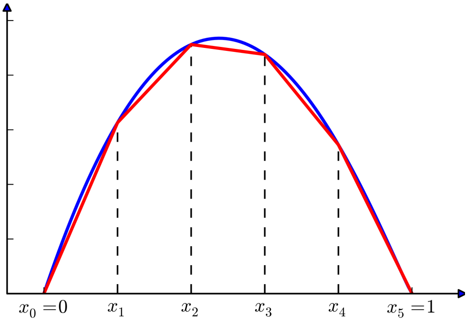


Output example

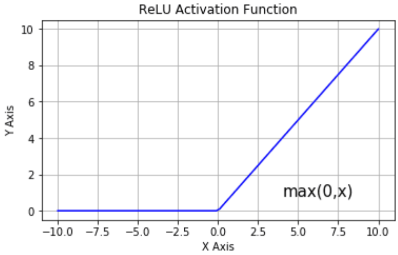
0.1	0
0.0	1
0.0	2
0.6	3
0.0	4
0.0	5
0.1	6
0.0	7
0.0	8
0.2	9

$\Rightarrow$ with 60% certainty this is a 3

What is a Neural Network (NN)?

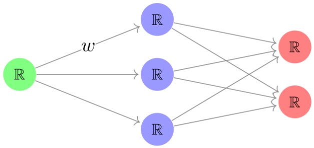


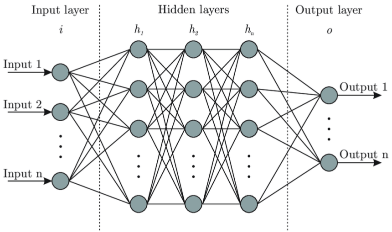
,



- **Idea:** Approximate the unknown function $\mathbb{R}^{784} \to \mathbb{R}^{10}$
- **NN:** A piecewise linear (PL) approximation (matrices + PL maps)
- The **matrices** represent numbers (weights) and offsets (biases; mostly ignored today)
- The **PL maps** are usually ReLUs (rectified linear unit)

What is a Neural Network (NN)?

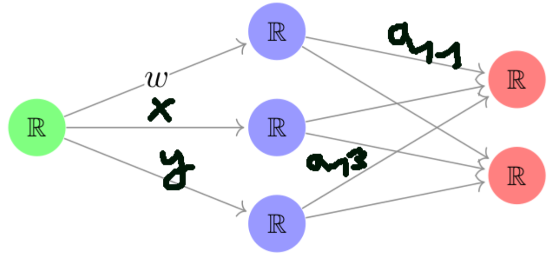




- NN: A directed graph as above
- The task of an NN is to approximate an unknown function
- It consists of neurons = entries of vectors, and weights = entries of matrices

What is a Neural Network (NN)?

Example



Here we have two matrices: a 3×1 matrix and a 2×3 matrix

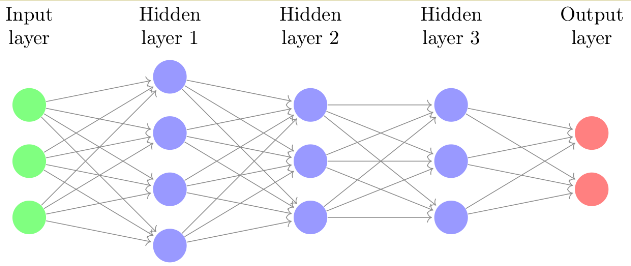
$$\begin{pmatrix} w \\ x \\ y \end{pmatrix} \text{ and } \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

plus two bias terms as in $y = \text{matrix} \cdot x + \text{bias}$

- NN: A
- The task of an NN is to approximate an unknown function
- It consists of neurons = entries of vectors, and weights = entries of matrices

What is a Neural Network (NN)?

Example



Here we have four matrices (plus four biases), whose composition gives a map:

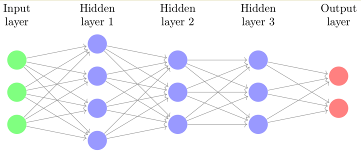
$$\mathbb{R}^3 \to \mathbb{R}^4 \to \mathbb{R}^3 \to \mathbb{R}^3 \to \mathbb{R}^2$$

- ▶ The task of an NN is to approximate an unknown function
- ▶ It consists of neurons = entries of vectors, and weights = entries of matrices

What is

Actually...

we need nonlinear maps as well, say ReLU applied componentwise

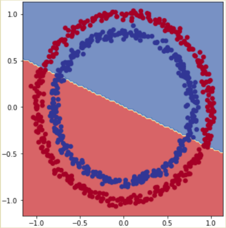


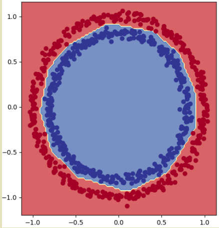
Here we have four matrices, whose composition gives a map:

$$\mathbb{R}^3 \xrightarrow{\text{ReLU}\circ\text{matrix}} \mathbb{R}^4 \xrightarrow{\text{ReLU}\circ\text{matrix}} \mathbb{R}^3 \xrightarrow{\text{ReLU}\circ\text{matrix}} \mathbb{R}^3 \xrightarrow{\text{ReLU}\circ\text{matrix}} \mathbb{R}^2.$$

Life is not linear!

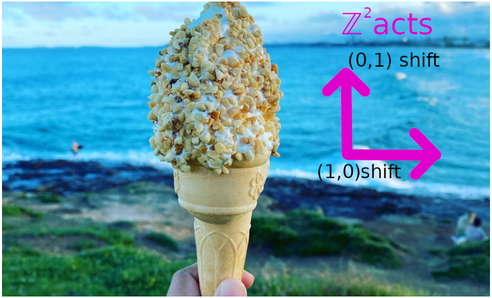
ReLU is crucial and brings in nonlinearity

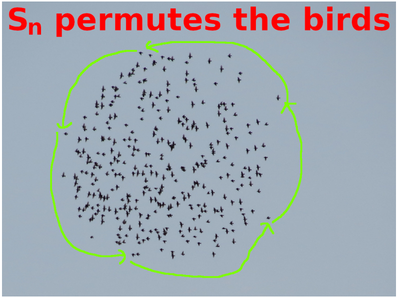
no ReLU: 

, with ReLU: 

- NN:
- The
- It c … f matrices

What is an Equivariant Neural Network (ENN)?

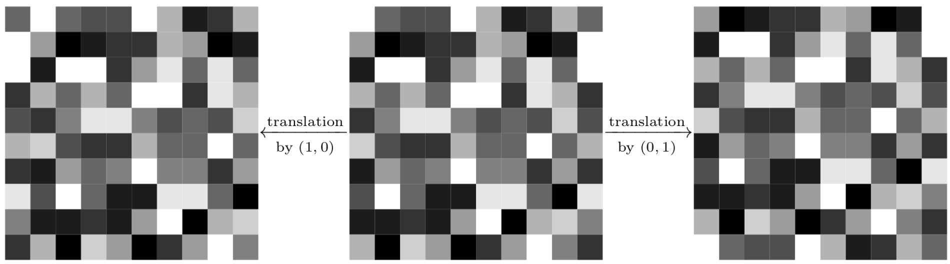




- **Task** Learn a map that is equivariant with respect to some symmetry
- **Examples**
  - Image recognition is often invariant under translation; for example, whether ice cream is present in an image should not depend on the position of the ice cream in the image
  - Point cloud problems are invariant under the symmetric group

What is an Equivariant Neural Network (ENN)?

Toy example - part 1



Periodic images, say n-by-n pixels,
have an action of $C_n^2 = (\mathbb{Z}/n\mathbb{Z})^2$ by translation

In math, a periodic image is an element of $V = \mathrm{Fun}(C_n^2, \mathbb{R})$ (maps $C_n^2 \to \mathbb{R}$)
The standard action on V is the shift above: $((a,b) \circlearrowright f)(x,y) = f(x+a, y+b)$

position of the ice cream in the image

- Point cloud problems are invariant under the symmetric group

Toy example - part 2

0.0	0.6	0.7	0.7	0.0	0.3	0.9	0.8	0.3	0.6
0.4	1.0	0.9	0.8	0.8	0.3	0.4	1.0	0.9	0.0
0.9	0.0	0.0	0.8	0.4	0.1	0.6	0.1	0.6	0.0
0.3	0.6	0.3	0.6	0.0	0.0	0.8	0.1	0.3	0.8
0.8	0.5	0.1	0.1	0.5	0.7	0.6	0.7	0.2	0.7
0.2	0.7	0.8	0.8	0.3	0.6	0.6	0.0	0.6	0.3
0.9	0.4	0.6	0.5	0.0	0.5	0.5	0.8	0.4	0.8
0.7	0.0	0.6	0.9	0.9	0.1	0.1	0.6	1.0	0.1
0.9	0.9	0.8	0.9	0.4	0.0	1.0	0.3	0.2	0.5
0.3	1.0	0.2	0.4	1.0	0.8	0.4	0.0	0.5	0.8

$\xrightarrow{\text{action by } c}$

0.475	0.675	0.6	0.475	0.7	0.5	0.475	0.55	0.7	0.275
0.475	0.475	0.625	0.8	0.375	0.4	0.7	0.55	0.475	0.475
0.175	0.625	0.5	0.45	0.425	0.325	0.35	0.575	0.325	0.575
0.775	0.275	0.325	0.3	0.375	0.4	0.325	0.475	0.425	0.325
0.425	0.55	0.425	0.5	0.275	0.425	0.7	0.225	0.575	0.525
0.675	0.475	0.55	0.425	0.475	0.525	0.425	0.675	0.225	0.575
0.525	0.55	0.575	0.575	0.55	0.3	0.5	0.375	0.8	0.425
0.475	0.65	0.575	0.725	0.35	0.375	0.55	0.55	0.325	0.75
0.6	0.675	0.65	0.625	0.7	0.575	0.2	0.45	0.575	0.5
0.675	0.5	0.725	0.7	0.4	0.425	0.675	0.5	0.325	0.475

,

zooming into the node at 2-2:

0.0	0.6	0.7
0.4	1.0	0.9
0.9	0.0	0.0

$\xrightarrow{\text{action by } c}$

0.475	0.675	0.6
0.475	0.475	0.625
0.175	0.625	0.5

where $0.475 = 1/4(0.6 + 0.9 + 0.0 + 0.4)$.

What is a C_n^2-equivariant map $V \to V$?

For example, convolution, say by $c = \frac{1}{4}((1,0) + (0,1) + (-1,0) + (0,-1))$

What is an Equi

- Task Learn … e symmetry
- Examples
  - Image re… for example, whether … epend on the position
  - Point cl… c group

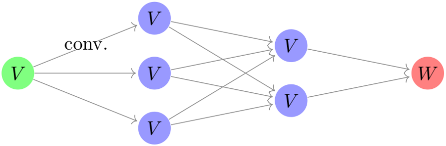
By the way

c is an example of what convolution usually does

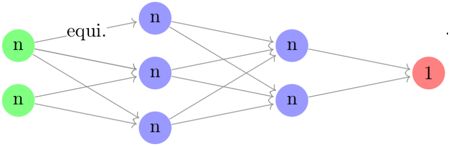
Operation	Filter	Convolved Image
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	
Gaussian blur (approximation)	$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	

What is an Equivariant Neural Network (ENN)?

C_n^2 equivariant
convolutional NN
$(W = V \text{ or } \mathbb{R})$:



S_n equivariant
point cloud NN
$(\mathtt{n} = \mathrm{Fun}(\{1, \dots, n\}, \mathbb{R}))$:



- Enter, rep theory All involved spaces are reps, all maps equivariant
- Rep theory questions
  - Can we decompose the network into smaller ("simple") bits?
  - The space of all weights is much smaller than for a vanilla NN: can we use rep theory to study the involved maps?

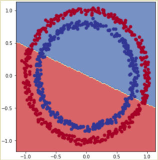
1+2 motivate the study of PL rep theory

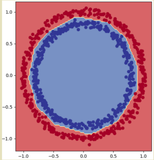
Fact 1

NN and ENN need nonlinear maps

Life is not linear!

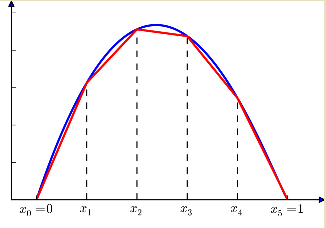
ReLU is crucial and brings in nonlinearity

no ReLU: 

, with ReLU: 

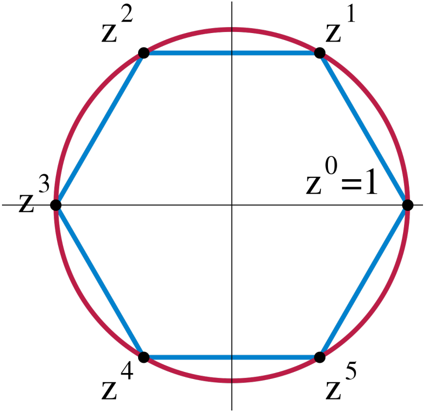
Fact 2

Every NN and ENN can be approximated using PL activation functions



PL rep theory (of cyclic groups)

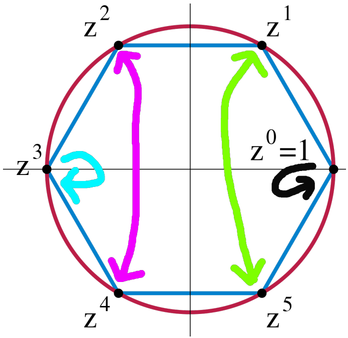
$n = 6$:



- Let $C_n = \mathbb{Z}/n\mathbb{Z} = \langle a | a^n = 1 \rangle$
- The simple complex C_n-reps are given by the nth roots of unity L_{z^k}
- What about the simple real C_n-reps?

PL rep theory (of cyclic groups)

$n = 6$:



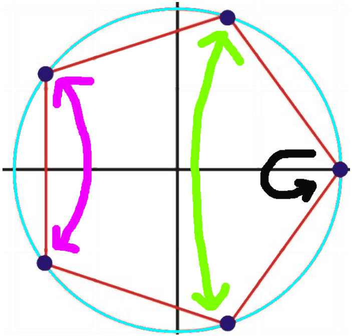
- For $\Theta = 2\pi/n$ observe that

$$\begin{pmatrix} \exp(ik\Theta) & 0 \\ 0 & \overline{\exp(ik\Theta)} \end{pmatrix} \sim_{\mathbb{C}} \begin{pmatrix} \cos(k\Theta) & -\sin(k\Theta) \\ \sin(k\Theta) & \cos(k\Theta) \end{pmatrix}$$

- $\Rightarrow$ the simple real C_n-reps are $L_0 = L_{z^0}$, $L_1 = L_{z^1} \oplus \overline{L_{z^1}}$, etc.

PL rep theory (of cyclic groups)

$n = 5$:



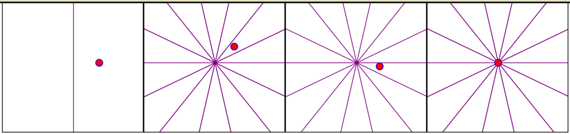
- For $\Theta = 2\pi/n$ observe that

$$\begin{pmatrix} \exp(ik\Theta) & 0 \\ 0 & \overline{\exp(ik\Theta)} \end{pmatrix} \sim_{\mathbb{C}} \begin{pmatrix} \cos(k\Theta) & -\sin(k\Theta) \\ \sin(k\Theta) & \cos(k\Theta) \end{pmatrix}$$

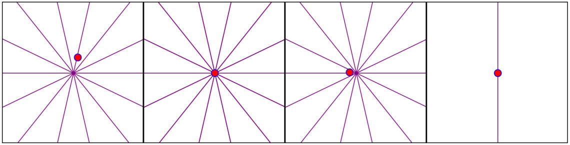
- $\Rightarrow$ the simple real C_n-reps are $L_0 = L_{z^0}$, $L_1 = L_{z^1} \oplus \overline{L_{z^1}}$, etc.

PL rep theory (of cyclic groups)

Play live at https://www.dtubbenhauer.com/pl-reps/site/index.html



$L_0 = L_{triv}, \mathrm{ord}(a) = 1$ | $L_1, \mathrm{ord}(a) = 14$ | $L_2, \mathrm{ord}(a) = 7$ | $L_3, \mathrm{ord}(a) = 14$



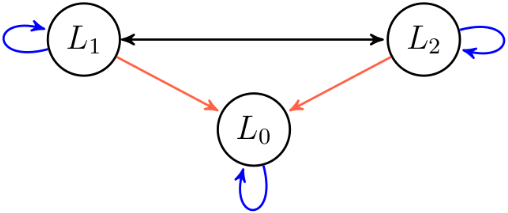
$L_4, \mathrm{ord}(a) = 7$ | $L_5, \mathrm{ord}(a) = 14$ | $L_6, \mathrm{ord}(a) = 7$ | $L_7 = L_{sign}, \mathrm{ord}(a) = 2$

Thus, we can (explicitly) decompose $\mathbb{R}[C_n] \cong L_0 \oplus L_1 \oplus \dots$ and compute

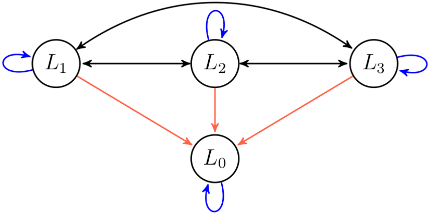
$$\begin{tikzcd} & & & L_0 \\ L_i \arrow[r, "incl."] & \mathbb{R}[C_n] \arrow[r, "ReLU"] & \mathbb{R}[C_n] \arrow[ur, "proj."] \arrow[r, "proj."] \arrow[dr, "proj."'] & L_1 \\ & & & \vdots \end{tikzcd}$$

PL rep theory (of cyclic groups)

$n = 5$:



$n = 7$:

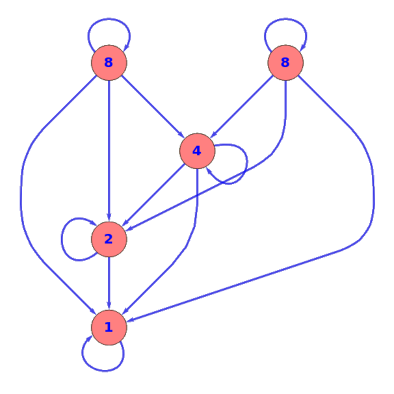


- Interaction graph Γ vertices = simples, edges = nonzero maps $L_i \to L_j$
- This is a measurement of difficulty: a lot of ingoing arrows = hard

PL rep theory (of c

$n = 5:$

It is easy to produce these with a machine



Let us play online

$n = 7$

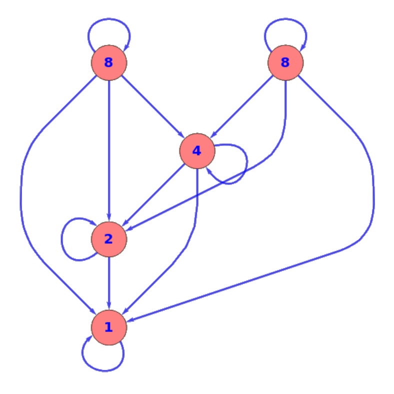
L_0

- Interaction graph Γ vertices = simples, edges = nonzero maps $L_i \to L_j$
- This is a measurement of difficulty: a lot of ingoing arrows = hard

$n = 5:$

$n = $

It is easy to produce these with a machine



Let us play online

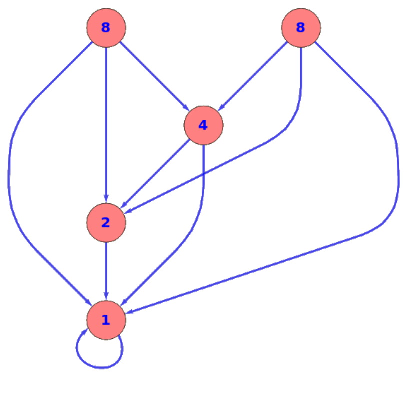
$ord(L_i)$ = order of the action of a
$ord'(L_i)$ = same but divide by two if $ord(L_i)$ is even

Theorem (for ReLU)
Every vertex has a loop and there is a non-loop edge from L to K in Γ if and only if $ord(K)$ divides $ord'(L)$

- Int
- Thi

n

It is easy to produce these with a machine – for any map



This is now the absolute value

L_0

$ord(L_i)$ = order of the action of a

$ord'(L_i)$ = same but divide by two if $ord(L_i)$ is even

Theorem (for Abs)

There is a non-loop edge from L to K in Γ

if and only if $ord(K)$ divides $ord'(L)$

- Interaction ... aps $L_i \to L_j$
- This is a m ... = hard

PL rep theory (of cyclic groups)

Calculation complexity is captured in Γ:

Theorem

$\#$ calcs needed to evaluate f^k on L_i = number of length k path ending at L_i in Γ_f

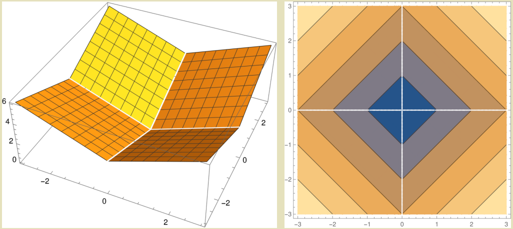
$$\Gamma_{\text{ReLU}} = \begin{tikzcd} L_1 \arrow[loop left] \arrow[d] \\ L_0 \arrow[loop left] \end{tikzcd}$$

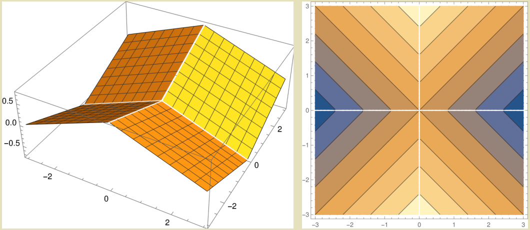
$$\text{ReLU} \leftrightsquigarrow \begin{tikzcd} L_1 \arrow[r, "f_1^1"] \arrow[dr, "f_1^0"] & L_1 \\ L_0 \arrow[r, "f_0^0"'] & L_0 \end{tikzcd}$$

$$\text{ReLU} \circ \text{ReLU} \leftrightsquigarrow \begin{tikzcd} L_1 \arrow[r, "f_1^1"] \arrow[dr, "f_1^0"] & L_1 \arrow[r, "f_1^1"] \arrow[dr, "f_1^0"] & L_1 \\ L_0 \arrow[r, "f_0^0"'] & L_0 \arrow[r, "f_0^0"'] & L_0 \end{tikzcd}$$

- Interaction graph Γ vertices = simples, edges = nonzero maps $L_i \to L_j$
- This is a measurement of difficulty: a lot of ingoing arrows = hard

The piecewise linear but non-linear maps from L_1 to trivial and sign are:

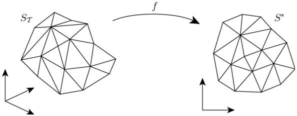




Let us play online

- This is a measurement of difficulty: a lot of ingoing arrows = hard

PL rep theory (of cyclic groups)



Some general facts about PL rep theory

- There are many PL maps, i.e. for G-representations V, W we have

$$\hom_G^{pl}(V, W) = 0 \text{ or } \dim_{\mathbb{R}} \hom_G^{pl}(V, W) > |\mathbb{N}|$$

- Schur's lemma is maximally false, i.e. let L and K be two simple real G-representations with actions maps ρ_L and ρ_K. Then:

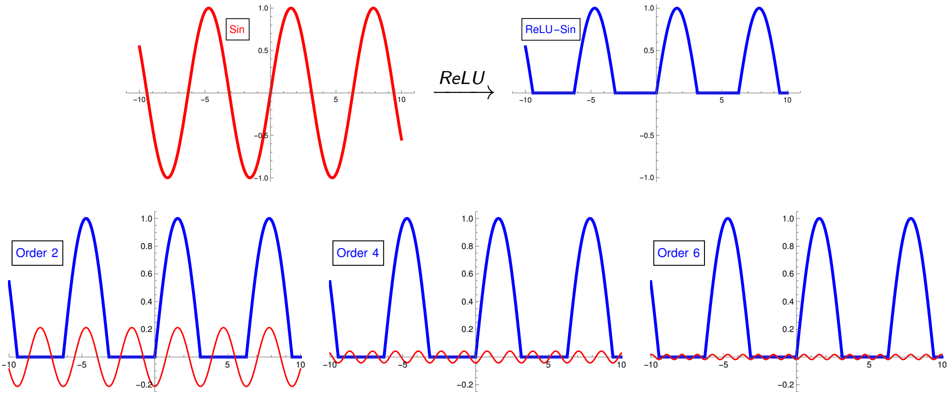
$$\hom_G^{pl}(L, K) \neq 0 \iff \ker(\rho_L) \subset \ker(\rho_K).$$

- Any PL activation map has an interaction graph Γ equal to the one of:

ReLU or Abs or Id

Thus, Γ depends only on G (up to 3 states it takes)

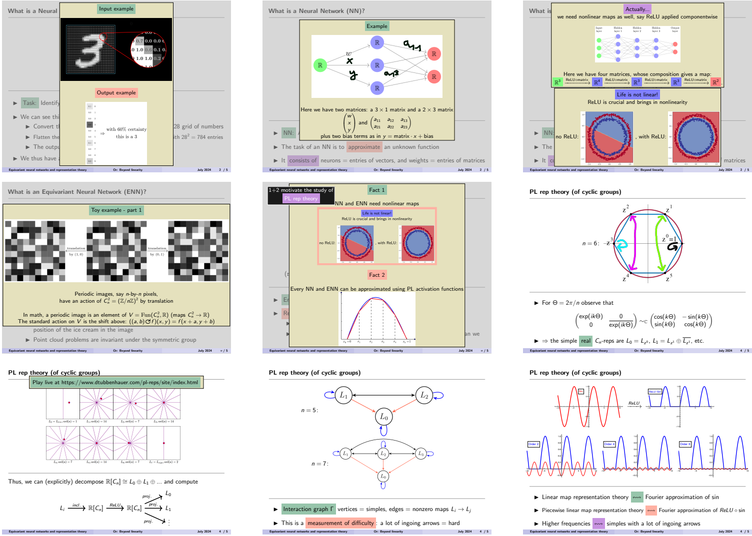
PL rep theory (of cyclic groups)



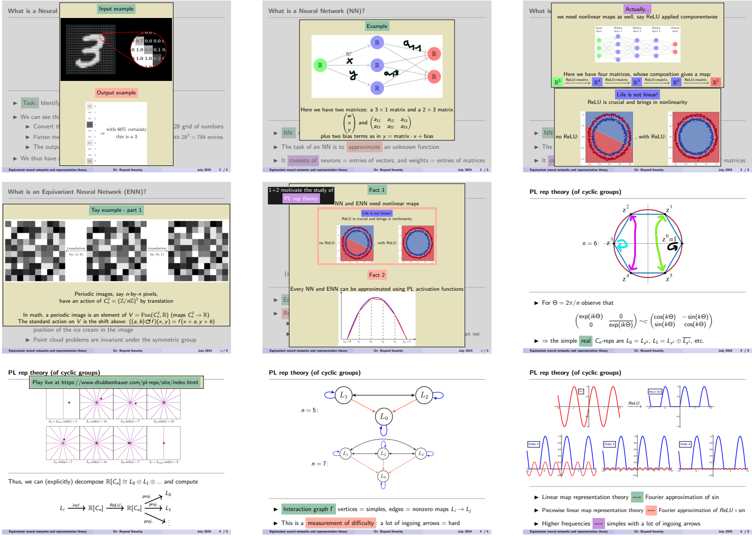
- Linear map representation theory $\leftrightsquigarrow$ Fourier approximation of sin
- Piecewise linear map representation theory $\leftrightsquigarrow$ Fourier approximation of $ReLU \circ \sin$
- Higher frequencies $\leftrightsquigarrow$ simples with a lot of ingoing arrows

PL rep theory (of cyclic groups)

- Finite abelian groups
- Dihedral groups
- Symmetric groups
- Products of these



There is still much to do...



Thanks for your attention!