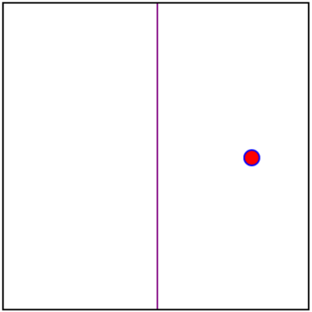
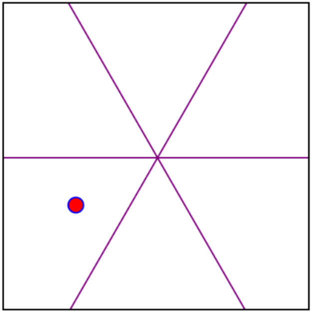
Piecewise linear representation theory

Or: Non-linear, but still rep theory

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



$L_0 = L_{triv}, \mathrm{ord}(a) = 1$

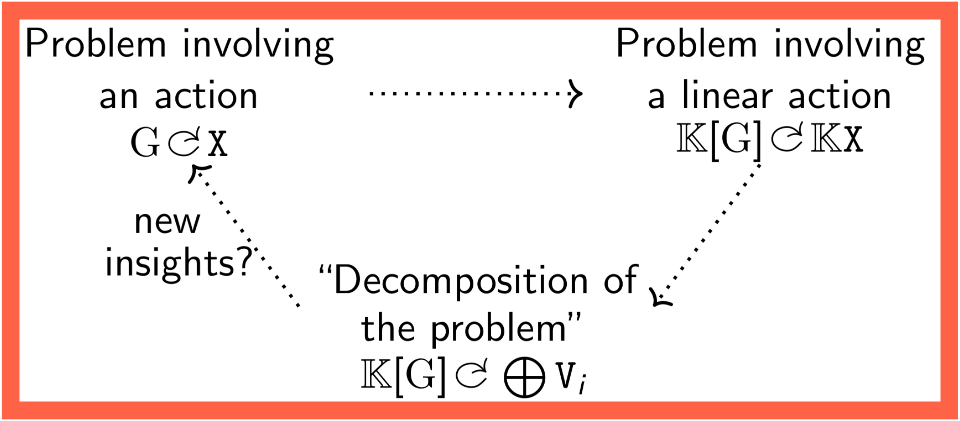


$L_1, \mathrm{ord}(a) = 3$

I report on work of Joel Gibson and Geordie Williamson

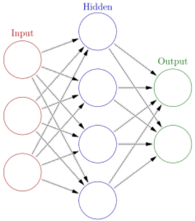
July 2023

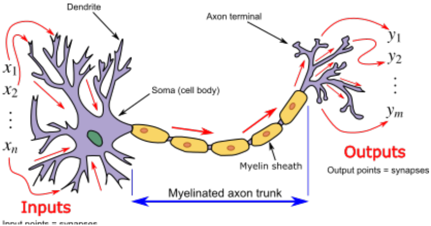
Learning with piecewise linear maps



- Representation theory approach Decompose a problem into simples and take it from there
- Today Representation theory applied to machine learning

Learning with piecewise linear maps





Vanilla neural net:

$$\phi: \mathbb{R}^{n_1} \xrightarrow{\phi_1} \mathbb{R}^{n_2} \xrightarrow{\phi_2} \mathbb{R}^{n_3} \xrightarrow{\phi_3} \mathbb{R}^{n_4} \xrightarrow{\phi_4} \mathbb{R}^{n_5} \xrightarrow{\phi_5} \mathbb{R}^{n_6}$$

"ReLU"

Each "layer" is: $\mathbb{R}^{n_i} \xrightarrow[\text{(or affine linear)}]{\text{linear}} \mathbb{R}^{n_{i+1}} \xrightarrow[\max(0,-)]{\text{coordinatewise}} \mathbb{R}^{n_{i+1}}$

- Neural network "=" a sequence of maps $\mathbb{R}^{n_1} \xrightarrow{\Phi_1} \mathbb{R}^{n_2} \xrightarrow{\Phi_2} \dots \xrightarrow{\Phi_k} \mathbb{R}^{n_{k+1}}$
- Deep = many layers Crucial 1 #layers $\leftrightsquigarrow$ accuracy of the result

Example (picture recognition)

Are we seeing ice cream? Yes, maybe or no?

Neural ice cream network:

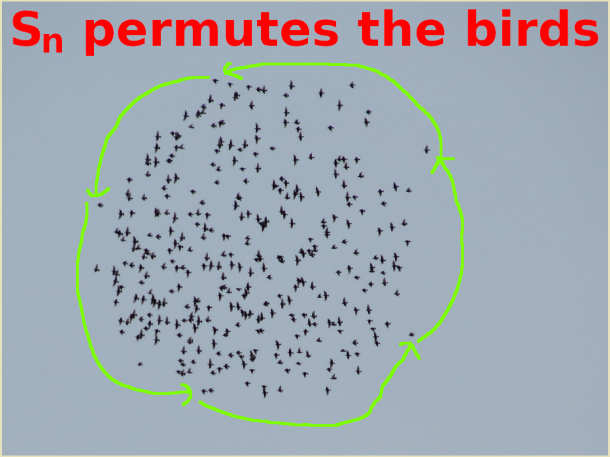
$f\colon \mathbb{R}^{\#pixels} \to \text{layers} \to \mathbb{R}$ and $f(\mathbb{R}^{\#pixels}) \rightsquigarrow$ probability of seeing ice cream

- Neural network "=" a sequence of maps $\mathbb{R}^{n_1} \longrightarrow \mathbb{R}^{n_2} \longrightarrow \dots \longrightarrow \mathbb{R}$
- Deep = many layers … the result

Crucial 1 #layers $\rightsquigarrow$ accuracy of the result

Learning with piecewise linear maps

Example (picture recognition with group action)



If our problem at hand has some group symmetry
then we should be able to use the representation theory approach, right?

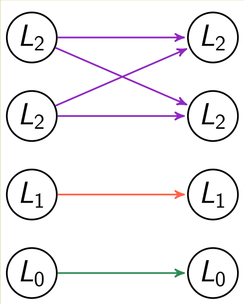
- Deep = many layers Crucial 1 #layers ⇝ accuracy of the result

Why is that potentially amazing? Well:

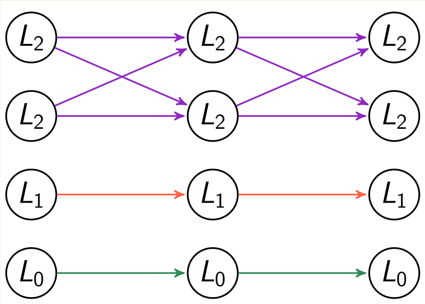
Assume we want to know f^k for linear G-equivariant $f: V \to V$

The representation theory approach plus Schur's lemma give

$f \sim$



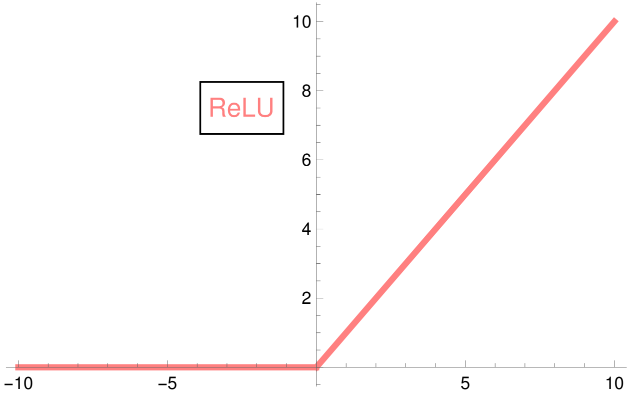
$f^2 \sim$



$\Rightarrow f^k$ is pretty easy to compute

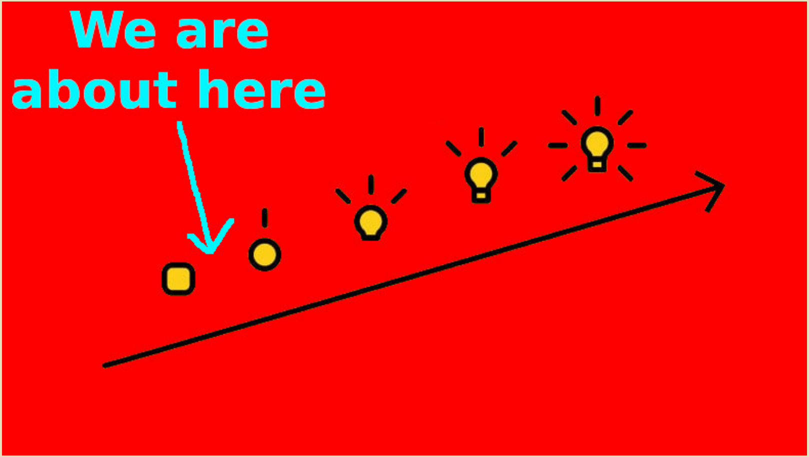
- Neural network = a sequence of maps ...
- Deep = many layers Crucial 1 #layers $\leftrightsquigarrow$ accuracy of the result

Learning with piecewise linear maps



- $ReLU: \mathbb{R} \to \mathbb{R}, x \mapsto \max(x, 0)$ is the most popular activator map in machine learning
- Linear maps are not working Play live at https://playground.tensorflow.org
- Crucial 2 Machine learning likes piecewise linear but non-linear maps

Lear



Crucial 1 #layers ⇝ accuracy of the result

Crucial 2 Machine learning likes piecewise linear but non-linear maps

$\Rightarrow$ study piecewise linear representation theory

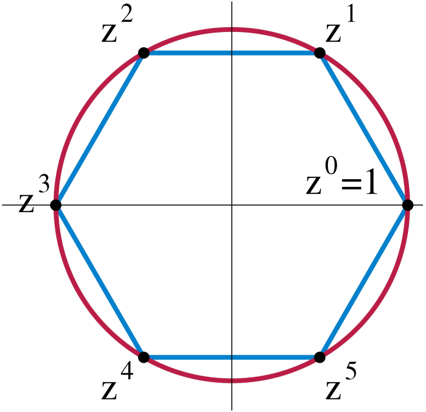
I show you a few teeny-weeny steps in this direction!

rning

org

Let us study cyclic groups

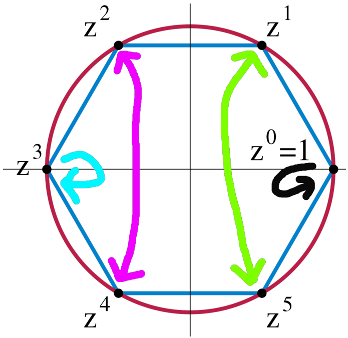
$n = 6$:



- 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?

Let us study cyclic groups

$n = 6$:



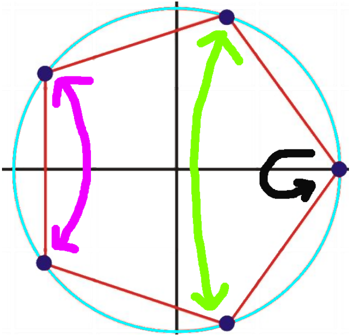
- For $\Theta = 2\pi/n$ observe that

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

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

Let us study cyclic groups

$n = 5$:



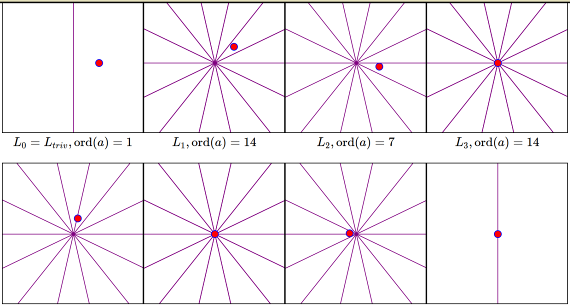
- For $\Theta = 2\pi/n$ observe that

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

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

Let us study cyclic groups

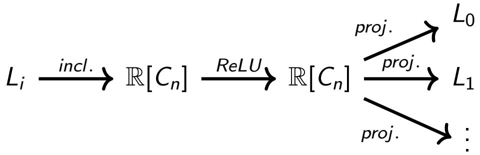
Play live at https://www.dtubbenhauer.com/pl-reps/site/index.html



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

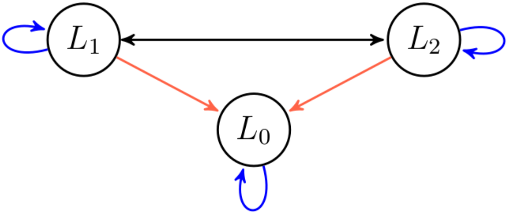
$L_4, \mathrm{ord}(a) = 7$ | $L_5, \mathrm{ord}(a) = 14$ | $L_6, \mathrm{ord}(a) = 7$ | $L_7 = L_{sign}, \mathrm{ord}(a) = 2$

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

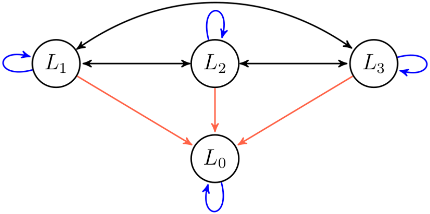


Let us study cyclic groups

$n = 5$:



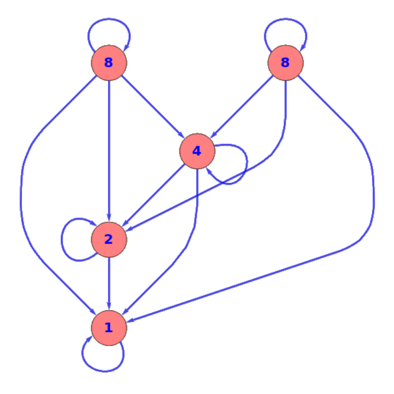
$n = 7$:



- **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

Let us study cyclic

It is easy to produce these with a machine



Let us play online

$n = 5$:

$n =$

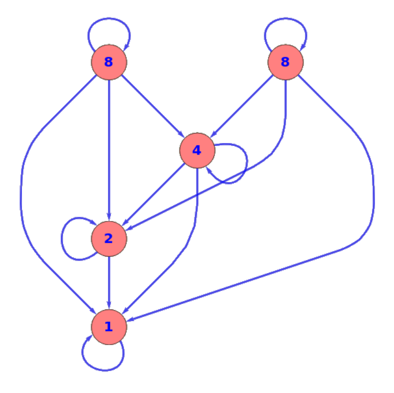
- Interaction graph Γ vertices = simples, edges = nonzero maps $L_i \to L_j$
- This is a measurement of difficulty : a lot of ingoing arrows = hard

Let us study cyclic

$n = 5$:

$n =$

It is easy to produce these with a machine



Let us play online

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 $ReLU$)

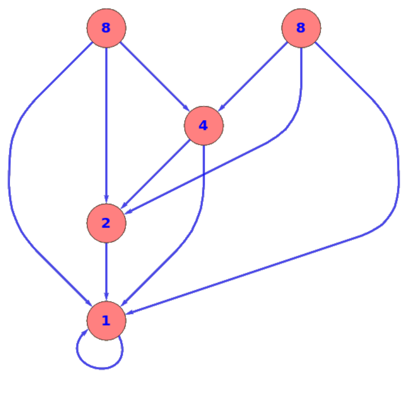
Every vertex has a loop and there is a non-loop edge from L to K in Γ if and only if $ord(K)$ divides $ord'(L)$

- Int
- Thi

L_j

Let us study

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



This is now the absolute value

n

$ord(L_i)$ = 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 $ord'(L)$

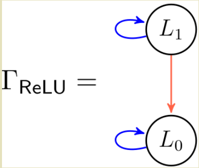
- Interaction ... aps $L_i \to L_j$
- This is a m ... = hard

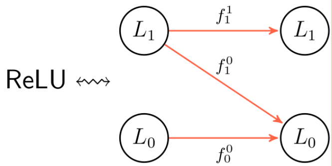
Let us study cyclic groups

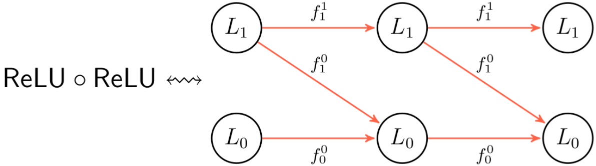
Calculation complexity is captured in Γ:

Theorem

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



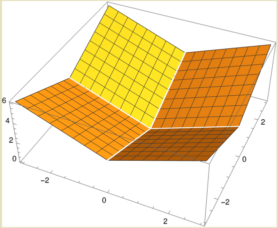


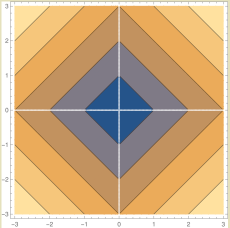


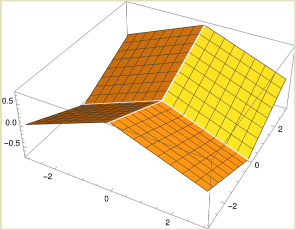
- Interaction graph Γ vertices = simples, edges = nonzero maps $L_i \to L_j$
- This is a measurement of difficulty: a lot of ingoing arrows = hard

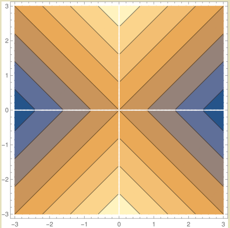
Let us study cyclic groups

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









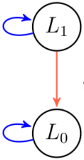
Let us play online

Let us study cyclic groups

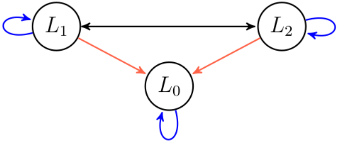
- Finite abelian groups
- Dihedral groups
- Symmetric groups
- Products of these

Finite abelian group = products of cyclic groups

$\Gamma_{\mathsf{ReLU}}(\mathbb{Z}/3\mathbb{Z}) =$

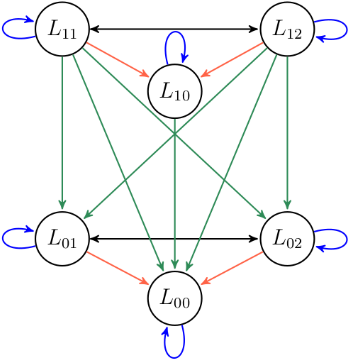


, $\Gamma_{\mathsf{ReLU}}(\mathbb{Z}/5\mathbb{Z}) =$



,

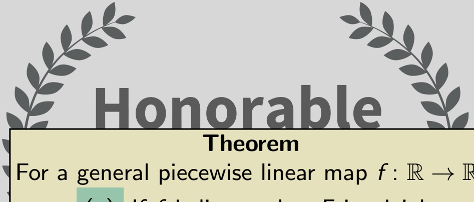
$\Rightarrow \Gamma_{\mathsf{ReLU}}(\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}) =$



.

Its just the product

Let us study cyclic groups



Theorem

For a general piecewise linear map $f\colon \mathbb{R} \to \mathbb{R}$:

(a) If f is linear, then Γ is trivial

(b) If $f = \text{Abs}$, then Γ is as before

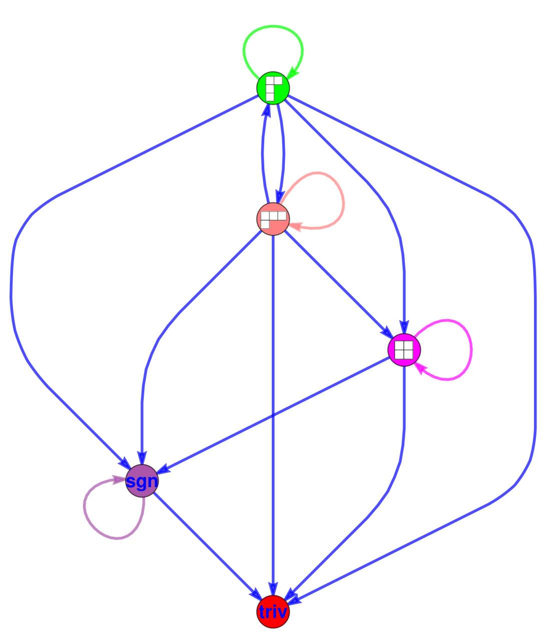
(c) Otherwise Γ is as for *ReLU*

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

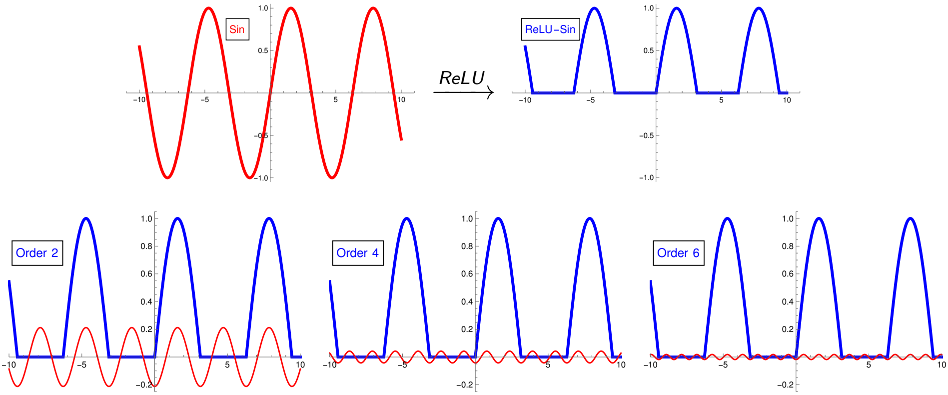
Γ for *ReLU* can thus
be considered as an invariant of the group C_n
or more generally for any finite group one gets an invariant

Let us study c

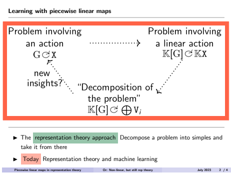
Γ for the symmetric group S_4 (on isotypic components):

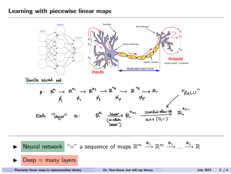


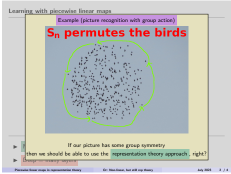
Let us study cyclic groups

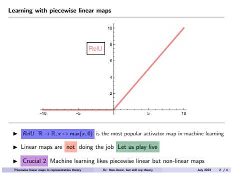


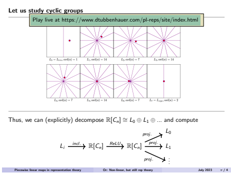
- Linear map representation theory ⇝ Fourier approximation of sin
- Piecewise linear map representation theory ⇝ Fourier approximation of $ReLU \circ \sin$
- Higher frequencies ⇝ simples with a lot of ingoing arrows

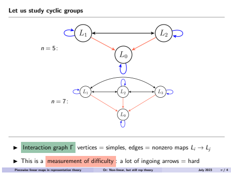


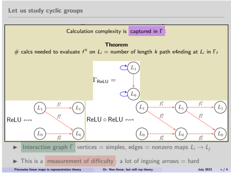


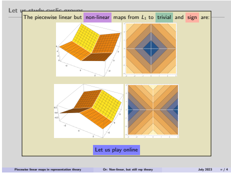


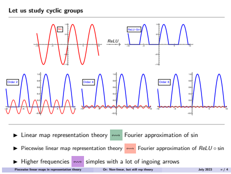




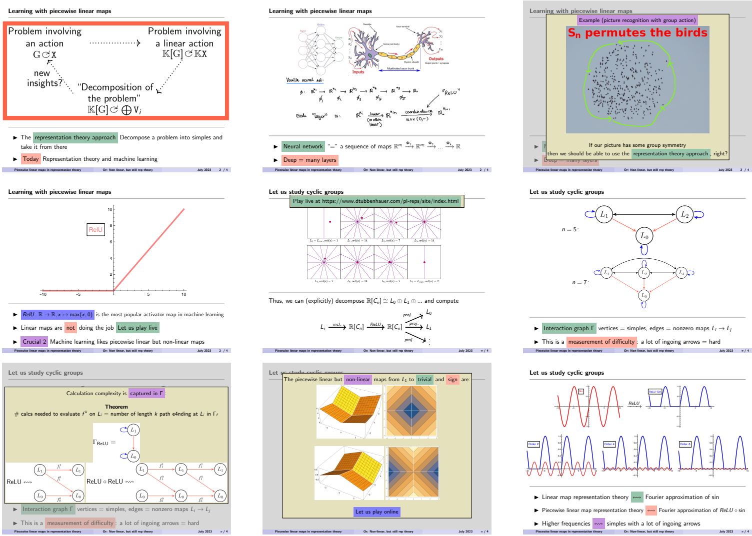








There is still much to do...



Thanks for your attention!