Geometry of Linear Neural Networks that are Equivariant / Invariant under Permutation Groups

Kathlén Kohn

joint work with

Anna-Laura Sattelberger

Vahid Shahverdi

Algebra & Geometry $\Rightarrow$ Neural Network Theory

Algebra & Geometry $\Rightarrow$ Neural Network Theory

Geometric questions:

1. How does the network architecture affect the geometry of the function space?
2. How does the geometry of the function space impact the training of the network?

Algebra & Geometry ⇒ Neural Network Theory

Geometric questions:

1. How does the network architecture affect the geometry of the function space?
2. How does the geometry of the function space impact the training of the network?
   a) static optimization properties (only depend on the loss)
   b) dynamic optimization properties (depend on the loss and choice of optimization algorithm)

Algebra & Geometry $\Rightarrow$ Neural Network Theory

Geometric questions:

1. How does the network architecture affect the geometry of the function space?
2. How does the geometry of the function space impact the training of the network?
   a) static optimization properties (only depend on the loss)
   b) dynamic optimization properties (depend on the loss and choice of optimization algorithm)

Algebra & Geometry $\Rightarrow$ Neural Network Theory

Algebraic settings:

network architecture		
activation	**network structure**	**loss**

Algebra & Geometry $\Rightarrow$ Neural Network Theory

Algebraic settings:

network architecture		
activation	**network structure**	**loss**
identity		
ReLU		
polynomial		

Algebra & Geometry $\Rightarrow$ Neural Network Theory

Algebraic settings:

network architecture		
activation	**network structure**	**loss**
identity	fully-connected	
ReLU	convolutional	
polynomial	group equivariant	

Algebra & Geometry $\Rightarrow$ Neural Network Theory

Algebraic settings:

network architecture			
activation	**network structure**	**loss**	
identity	fully-connected	squared-error loss	= Euclidean dist
ReLU	convolutional	Wasserstein distance	= polyhedral dist.
polynomial	group equivariant	cross-entropy	$\cong$ KL divergence

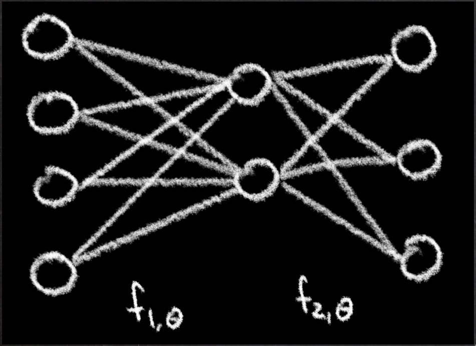
Algebra & Geometry ⇒ Neural Network Theory

Algebraic settings:

network architecture		loss	
activation	**network structure**	**loss**	
identity	fully-connected	squared-error loss	= Euclidean dist
ReLU	convolutional	Wasserstein distance	= polyhedral dist.
polynomial	group equivariant	cross-entropy	≅ KL divergence

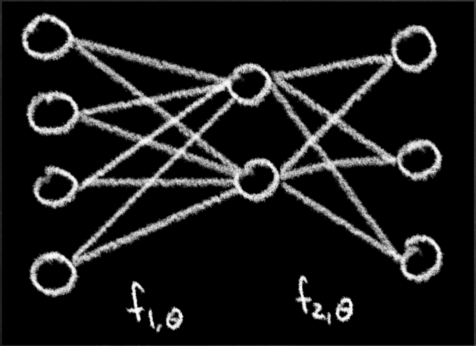
Linear Fully-Connected Networks

activation = identity & network structure = fully-connected



Linear Fully-Connected Networks

activation = identity & network structure = fully-connected

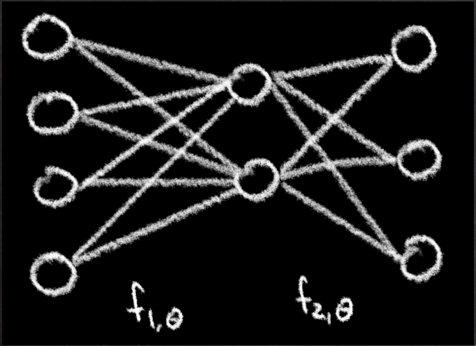


This network parametrizes linear maps:

$$\mu : \mathbb{R}^{2\times 4} \times \mathbb{R}^{3\times 2} \longrightarrow \mathbb{R}^{3\times 4},$$
$$(W_1, W_2) \longmapsto W_2 W_1.$$

Linear Fully-Connected Networks

activation $=$ identity & network structure $=$ fully-connected



This network parametrizes linear maps:

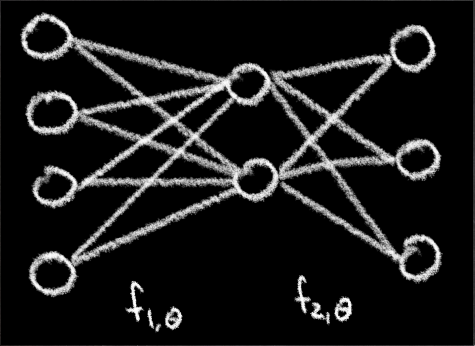
$$\mu : \mathbb{R}^{2\times 4} \times \mathbb{R}^{3\times 2} \longrightarrow \mathbb{R}^{3\times 4}, \quad (W_1, W_2) \longmapsto W_2 W_1.$$

Its function space is

$$\mathcal{M}_2 = \{W \in \mathbb{R}^{3\times 4} \mid \operatorname{rank}(W) \leq 2\}.$$

Linear Fully-Connected Networks

activation = identity & network structure = fully-connected



This network parametrizes linear maps:

$$\mu : \mathbb{R}^{2\times 4} \times \mathbb{R}^{3\times 2} \longrightarrow \mathbb{R}^{3\times 4}, \quad (W_1, W_2) \longmapsto W_2 W_1.$$

Its function space is

$$\mathcal{M}_2 = \{W \in \mathbb{R}^{3\times 4} \mid \mathrm{rank}(W) \leq 2\}.$$

In general: $\mu : \mathbb{R}^{k_1\times k_0} \times \mathbb{R}^{k_2\times k_1} \times \ldots \times \mathbb{R}^{k_L\times k_{L-1}} \longrightarrow \mathbb{R}^{k_L\times k_0},$

$$(W_1, W_2, \ldots, W_L) \longmapsto W_L \cdots W_2 W_1.$$

Its function space $\mathcal{M}_r = \mathrm{im}(\mu) = \{W \in \mathbb{R}^{k_L\times k_0} \mid \mathrm{rank}(W) \leq r\}$, where $r := \min(k_0, \ldots, k_L)$, is an algebraic variety.

Linear Group-Equivariant Networks

Running Example

Consider an autoencoder $\mu : \mathbb{R}^{2\times 9} \times \mathbb{R}^{9\times 2} \longrightarrow \mathbb{R}^{9\times 9}$, $(W_1, W_2) \longmapsto W_2 W_1$

Linear Group-Equivariant Networks

Running Example

Consider an autoencoder $\mu : \mathbb{R}^{2\times 9} \times \mathbb{R}^{9\times 2} \longrightarrow \mathbb{R}^{9\times 9}$, $(W_1, W_2) \longmapsto W_2 W_1$ with function space $\mathcal{M}_2 = \mathrm{im}(\mu) = \{W \in \mathbb{R}^{9\times 9} \mid \mathrm{rank}(W) \leq 2\}$.

Linear Group-Equivariant Networks

Running Example

Consider an autoencoder $\mu : \mathbb{R}^{2\times 9} \times \mathbb{R}^{9\times 2} \longrightarrow \mathbb{R}^{9\times 9}$, $(W_1, W_2) \longmapsto W_2 W_1$ with function space $\mathcal{M}_2 = \mathrm{im}(\mu) = \{W \in \mathbb{R}^{9\times 9} \mid \mathrm{rank}(W) \leq 2\}$.

Its inputs and outputs are 3×3 images:

a_{11}	a_{12}	a_{13}
a_{21}	a_{22}	a_{23}
a_{31}	a_{32}	a_{33}

$\in \mathbb{R}^9$.

Linear Group-Equivariant Networks

Running Example

Consider an autoencoder $\mu : \mathbb{R}^{2\times 9} \times \mathbb{R}^{9\times 2} \longrightarrow \mathbb{R}^{9\times 9}$, $(W_1, W_2) \longmapsto W_2 W_1$ with function space $\mathcal{M}_2 = \mathrm{im}(\mu) = \{W \in \mathbb{R}^{9\times 9} \mid \mathrm{rank}(W) \leq 2\}$.

Its inputs and outputs are 3×3 images: $\begin{array}{|c|c|c|}\hline a_{11} & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array} \in \mathbb{R}^9$.

Consider the clockwise rotation by 90°:

$$\sigma : \mathbb{R}^9 \longrightarrow \mathbb{R}^9, \quad \begin{array}{|c|c|c|}\hline a_{11} & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array} \longmapsto \begin{array}{|c|c|c|}\hline a_{31} & a_{21} & a_{11} \\ \hline a_{32} & a_{22} & a_{12} \\ \hline a_{33} & a_{23} & a_{13} \\ \hline \end{array}.$$

Linear Group-Equivariant Networks

Running Example

Consider an autoencoder $\mu : \mathbb{R}^{2\times 9} \times \mathbb{R}^{9\times 2} \longrightarrow \mathbb{R}^{9\times 9}$, $(W_1, W_2) \longmapsto W_2 W_1$ with function space $\mathcal{M}_2 = \mathrm{im}(\mu) = \{W \in \mathbb{R}^{9\times 9} \mid \mathrm{rank}(W) \leq 2\}$.

Its inputs and outputs are 3×3 images: $\begin{array}{|c|c|c|} \hline a_{11} & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array} \in \mathbb{R}^9$.

Consider the clockwise rotation by 90°:

$$\sigma : \mathbb{R}^9 \longrightarrow \mathbb{R}^9, \begin{array}{|c|c|c|} \hline a_{11} & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array} \longmapsto \begin{array}{|c|c|c|} \hline a_{31} & a_{21} & a_{11} \\ \hline a_{32} & a_{22} & a_{12} \\ \hline a_{33} & a_{23} & a_{13} \\ \hline \end{array}.$$

Which $W \in \mathcal{M}_2$ are equivariant under σ?

Linear Group-Equivariant Networks

Running Example

Consider an autoencoder $\mu : \mathbb{R}^{2\times 9} \times \mathbb{R}^{9\times 2} \longrightarrow \mathbb{R}^{9\times 9}$, $(W_1, W_2) \longmapsto W_2 W_1$ with function space $\mathcal{M}_2 = \mathrm{im}(\mu) = \{W \in \mathbb{R}^{9\times 9} \mid \mathrm{rank}(W) \leq 2\}$.

Its inputs and outputs are 3×3 images: $\begin{array}{|c|c|c|}\hline a_{11} & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array} \in \mathbb{R}^9$.

Consider the clockwise rotation by $90°$:

$$\sigma : \mathbb{R}^9 \longrightarrow \mathbb{R}^9, \begin{array}{|c|c|c|}\hline a_{11} & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array} \longmapsto \begin{array}{|c|c|c|}\hline a_{31} & a_{21} & a_{11} \\ \hline a_{32} & a_{22} & a_{12} \\ \hline a_{33} & a_{23} & a_{13} \\ \hline \end{array}.$$

Which $W \in \mathcal{M}_2$ are equivariant under σ?
Which are invariant?

example cont'd

$$\sigma : \mathbb{R}^9 \longrightarrow \mathbb{R}^9, \quad \begin{array}{|c|c|c|}\hline a_{11} & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array} \longmapsto \begin{array}{|c|c|c|}\hline a_{31} & a_{21} & a_{11} \\ \hline a_{32} & a_{22} & a_{12} \\ \hline a_{33} & a_{23} & a_{13} \\ \hline \end{array}$$

is represented by the permutation matrix

$$P_\sigma = \left[\begin{array}{c|c|c} \begin{matrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{matrix} & 0 & 0 \\ \hline 0 & \begin{matrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{matrix} & 0 \\ \hline 0 & 0 & 1 \end{array}\right]$$

example cont'd

$$\sigma : \mathbb{R}^9 \longrightarrow \mathbb{R}^9, \quad \begin{array}{|c|c|c|} \hline a_{11} & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array} \longmapsto \begin{array}{|c|c|c|} \hline a_{31} & a_{21} & a_{11} \\ \hline a_{32} & a_{22} & a_{12} \\ \hline a_{33} & a_{23} & a_{13} \\ \hline \end{array}$$

is represented by the permutation matrix

$$P_\sigma = \left[\begin{array}{cccc|cccc|c} 0 & 0 & 0 & 1 & & & & & \\ 1 & 0 & 0 & 0 & & & 0 & & 0 \\ 0 & 1 & 0 & 0 & & & & & \\ 0 & 0 & 1 & 0 & & & & & \\ \hline & & & & 0 & 0 & 0 & 1 & \\ & & 0 & & 1 & 0 & 0 & 0 & 0 \\ & & & & 0 & 1 & 0 & 0 & \\ & & & & 0 & 0 & 1 & 0 & \\ \hline & & 0 & & & & 0 & & 1 \end{array}\right]$$

$W \in \mathbb{R}^{9\times 9}$

is equivariant under σ

$\Leftrightarrow$

$W \cdot P_\sigma = P_\sigma \cdot W.$

example cont'd

$$\sigma : \mathbb{R}^9 \longrightarrow \mathbb{R}^9, \quad \begin{array}{|c|c|c|}\hline a_{11} & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array} \longmapsto \begin{array}{|c|c|c|}\hline a_{31} & a_{21} & a_{11} \\ \hline a_{32} & a_{22} & a_{12} \\ \hline a_{33} & a_{23} & a_{13} \\ \hline \end{array}$$

is represented by the permutation matrix

$$P_\sigma = \left[\begin{array}{cccc|cccc|c} 0 & 0 & 0 & 1 & & & & & \\ 1 & 0 & 0 & 0 & & 0 & & & 0 \\ 0 & 1 & 0 & 0 & & & & & \\ 0 & 0 & 1 & 0 & & & & & \\ \hline & & & & 0 & 0 & 0 & 1 & \\ & 0 & & & 1 & 0 & 0 & 0 & 0 \\ & & & & 0 & 1 & 0 & 0 & \\ & & & & 0 & 0 & 1 & 0 & \\ \hline & 0 & & & & 0 & & & 1 \end{array}\right]$$

$W \in \mathbb{R}^{9\times 9}$
is equivariant under σ
$\Leftrightarrow$
$W \cdot P_\sigma = P_\sigma \cdot W$.

$W \in \mathbb{R}^{9\times 9}$
is invariant under σ
$\Leftrightarrow$
$W \cdot P_\sigma = W$.

example cont'd

$W \in \mathbb{R}^{9\times 9}$ is equivariant under σ iff

$$W = \left[\begin{array}{cccc|cccc|c}
\alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \varepsilon_3 \\
\alpha_4 & \alpha_1 & \alpha_2 & \alpha_3 & \beta_4 & \beta_1 & \beta_2 & \beta_3 & \varepsilon_3 \\
\alpha_3 & \alpha_4 & \alpha_1 & \alpha_2 & \beta_3 & \beta_4 & \beta_1 & \beta_2 & \varepsilon_3 \\
\alpha_2 & \alpha_3 & \alpha_4 & \alpha_1 & \beta_2 & \beta_3 & \beta_4 & \beta_1 & \varepsilon_3 \\
\hline
\gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \delta_1 & \delta_2 & \delta_3 & \delta_4 & \varepsilon_4 \\
\gamma_4 & \gamma_1 & \gamma_2 & \gamma_3 & \delta_4 & \delta_1 & \delta_2 & \delta_3 & \varepsilon_4 \\
\gamma_3 & \gamma_4 & \gamma_1 & \gamma_2 & \delta_3 & \delta_4 & \delta_1 & \delta_2 & \varepsilon_4 \\
\gamma_2 & \gamma_3 & \gamma_4 & \gamma_1 & \delta_2 & \delta_3 & \delta_4 & \delta_1 & \varepsilon_4 \\
\hline
\varepsilon_1 & \varepsilon_1 & \varepsilon_1 & \varepsilon_1 & \varepsilon_2 & \varepsilon_2 & \varepsilon_2 & \varepsilon_2 & \varepsilon_5
\end{array}\right].$$

example cont'd

$W \in \mathbb{R}^{9\times 9}$ is equivariant under σ iff

$$W = \left[\begin{array}{cccc|cccc|c}
\alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \varepsilon_3 \\
\alpha_4 & \alpha_1 & \alpha_2 & \alpha_3 & \beta_4 & \beta_1 & \beta_2 & \beta_3 & \varepsilon_3 \\
\alpha_3 & \alpha_4 & \alpha_1 & \alpha_2 & \beta_3 & \beta_4 & \beta_1 & \beta_2 & \varepsilon_3 \\
\alpha_2 & \alpha_3 & \alpha_4 & \alpha_1 & \beta_2 & \beta_3 & \beta_4 & \beta_1 & \varepsilon_3 \\
\hline
\gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \delta_1 & \delta_2 & \delta_3 & \delta_4 & \varepsilon_4 \\
\gamma_4 & \gamma_1 & \gamma_2 & \gamma_3 & \delta_4 & \delta_1 & \delta_2 & \delta_3 & \varepsilon_4 \\
\gamma_3 & \gamma_4 & \gamma_1 & \gamma_2 & \delta_3 & \delta_4 & \delta_1 & \delta_2 & \varepsilon_4 \\
\gamma_2 & \gamma_3 & \gamma_4 & \gamma_1 & \delta_2 & \delta_3 & \delta_4 & \delta_1 & \varepsilon_4 \\
\hline
\varepsilon_1 & \varepsilon_1 & \varepsilon_1 & \varepsilon_1 & \varepsilon_2 & \varepsilon_2 & \varepsilon_2 & \varepsilon_2 & \varepsilon_5
\end{array}\right].$$

The linear space $\mathcal{E}^\sigma$ of σ-equivariant $W \in \mathbb{R}^{9\times 9}$

example cont'd

$W \in \mathbb{R}^{9\times 9}$ is equivariant under σ iff

$$W = \left[\begin{array}{cccc|cccc|c}
\alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \varepsilon_3 \\
\alpha_4 & \alpha_1 & \alpha_2 & \alpha_3 & \beta_4 & \beta_1 & \beta_2 & \beta_3 & \varepsilon_3 \\
\alpha_3 & \alpha_4 & \alpha_1 & \alpha_2 & \beta_3 & \beta_4 & \beta_1 & \beta_2 & \varepsilon_3 \\
\alpha_2 & \alpha_3 & \alpha_4 & \alpha_1 & \beta_2 & \beta_3 & \beta_4 & \beta_1 & \varepsilon_3 \\
\hline
\gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \delta_1 & \delta_2 & \delta_3 & \delta_4 & \varepsilon_4 \\
\gamma_4 & \gamma_1 & \gamma_2 & \gamma_3 & \delta_4 & \delta_1 & \delta_2 & \delta_3 & \varepsilon_4 \\
\gamma_3 & \gamma_4 & \gamma_1 & \gamma_2 & \delta_3 & \delta_4 & \delta_1 & \delta_2 & \varepsilon_4 \\
\gamma_2 & \gamma_3 & \gamma_4 & \gamma_1 & \delta_2 & \delta_3 & \delta_4 & \delta_1 & \varepsilon_4 \\
\hline
\varepsilon_1 & \varepsilon_1 & \varepsilon_1 & \varepsilon_1 & \varepsilon_2 & \varepsilon_2 & \varepsilon_2 & \varepsilon_2 & \varepsilon_5
\end{array}\right].$$

The linear space $\mathcal{E}^\sigma$ of σ-equivariant $W \in \mathbb{R}^{9\times 9}$ intersected with the function space $\mathcal{M}_2 = \{W \in \mathbb{R}^{9\times 9} \mid \mathrm{rank}(W) \leq 2\}$ of our autoencoder is an algebraic variety with

- 10 irreducible components over $\mathbb{C}$
- 4 irreducible components over $\mathbb{R}$

takeaway message

There is no neural network whose function space is $\mathcal{E}^\sigma \cap \mathcal{M}_2$!

takeaway message

There is no neural network whose function space is $\mathcal{E}^\sigma \cap \mathcal{M}_2$!

Any neural network can parametrize at most one of the real irreducible components of $\mathcal{E}^\sigma \cap \mathcal{M}_2$.

example cont'd

$W \in \mathbb{R}^{9\times 9}$ is invariant under σ iff

$$W = \left[\begin{array}{cccc|cccc|c}
\alpha_1 & \alpha_1 & \alpha_1 & \alpha_1 & \beta_1 & \beta_1 & \beta_1 & \beta_1 & \gamma_1 \\
\alpha_2 & \alpha_2 & \alpha_2 & \alpha_2 & \beta_2 & \beta_2 & \beta_2 & \beta_2 & \gamma_2 \\
\alpha_3 & \alpha_3 & \alpha_3 & \alpha_3 & \beta_3 & \beta_3 & \beta_3 & \beta_3 & \gamma_3 \\
\alpha_4 & \alpha_4 & \alpha_4 & \alpha_4 & \beta_4 & \beta_4 & \beta_4 & \beta_4 & \gamma_4 \\
\alpha_5 & \alpha_5 & \alpha_5 & \alpha_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 & \gamma_5 \\
\alpha_6 & \alpha_6 & \alpha_6 & \alpha_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \gamma_6 \\
\alpha_7 & \alpha_7 & \alpha_7 & \alpha_7 & \beta_7 & \beta_7 & \beta_7 & \beta_7 & \gamma_7 \\
\alpha_8 & \alpha_8 & \alpha_8 & \alpha_8 & \beta_8 & \beta_8 & \beta_8 & \beta_8 & \gamma_8 \\
\alpha_9 & \alpha_9 & \alpha_9 & \alpha_9 & \beta_9 & \beta_9 & \beta_9 & \beta_9 & \gamma_9
\end{array}\right].$$

example cont'd

$W \in \mathbb{R}^{9\times 9}$ is invariant under σ iff

$$W = \left[\begin{array}{cccc|cccc|c}
\alpha_1 & \alpha_1 & \alpha_1 & \alpha_1 & \beta_1 & \beta_1 & \beta_1 & \beta_1 & \gamma_1 \\
\alpha_2 & \alpha_2 & \alpha_2 & \alpha_2 & \beta_2 & \beta_2 & \beta_2 & \beta_2 & \gamma_2 \\
\alpha_3 & \alpha_3 & \alpha_3 & \alpha_3 & \beta_3 & \beta_3 & \beta_3 & \beta_3 & \gamma_3 \\
\alpha_4 & \alpha_4 & \alpha_4 & \alpha_4 & \beta_4 & \beta_4 & \beta_4 & \beta_4 & \gamma_4 \\
\alpha_5 & \alpha_5 & \alpha_5 & \alpha_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 & \gamma_5 \\
\alpha_6 & \alpha_6 & \alpha_6 & \alpha_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \gamma_6 \\
\alpha_7 & \alpha_7 & \alpha_7 & \alpha_7 & \beta_7 & \beta_7 & \beta_7 & \beta_7 & \gamma_7 \\
\alpha_8 & \alpha_8 & \alpha_8 & \alpha_8 & \beta_8 & \beta_8 & \beta_8 & \beta_8 & \gamma_8 \\
\alpha_9 & \alpha_9 & \alpha_9 & \alpha_9 & \beta_9 & \beta_9 & \beta_9 & \beta_9 & \gamma_9
\end{array}\right].$$

The linear space $\mathcal{I}^\sigma$ of σ-invariant $W \in \mathbb{R}^{9\times 9}$ intersected with the function space $\mathcal{M}_2 = \{W \in \mathbb{R}^{9\times 9} \mid \mathrm{rank}(W) \leq 2\}$ is an irreducible algebraic variety $\cong \{A \in \mathbb{R}^{9\times 3} \mid \mathrm{rank}(A) \leq 2\}$.

Invariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{m \times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Invariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{m \times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \pi_2 \circ \ldots \circ \pi_k$ into disjoint cycles.

Invariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{m \times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \pi_2 \circ \ldots \circ \pi_k$ into disjoint cycles.

Lemma: The linear space $\mathcal{I}^\sigma$ of σ-invariant $W \in \mathbb{R}^{m \times n}$ consists of all matrices W whose columns indexed by π_i are equal, for all $i = 1, 2, \ldots, k$.

Invariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{m\times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \pi_2 \circ \ldots \circ \pi_k$ into disjoint cycles.

Lemma: The linear space $\mathcal{I}^\sigma$ of σ-invariant $W \in \mathbb{R}^{m\times n}$ consists of all matrices W whose columns indexed by π_i are equal, for all $i = 1, 2, \ldots, k$. Hence, $\mathcal{I}^\sigma \cap \mathcal{M}_r \cong \{W \in \mathbb{R}^{m\times k} \mid \mathrm{rank}(W) \leq r\}$ is an irreducible variety.

Invariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{m \times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \pi_2 \circ \ldots \circ \pi_k$ into disjoint cycles.

Lemma: The linear space $\mathcal{I}^\sigma$ of σ-invariant $W \in \mathbb{R}^{m \times n}$ consists of all matrices W whose columns indexed by π_i are equal, for all $i = 1, 2, \ldots, k$. Hence, $\mathcal{I}^\sigma \cap \mathcal{M}_r \cong \{W \in \mathbb{R}^{m \times k} \mid \mathrm{rank}(W) \leq r\}$ is an irreducible variety.

Lemma: Let $G \subset \mathcal{S}_n$.
The set of G-invariant $W \in \mathbb{R}^{m \times n}$ is $\mathcal{I}^\sigma$ for some $\sigma \in \mathcal{S}_n$.

Invariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{m \times n} \mid \operatorname{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \pi_2 \circ \ldots \circ \pi_k$ into disjoint cycles.

Lemma: The linear space $\mathcal{I}^\sigma$ of σ-invariant $W \in \mathbb{R}^{m \times n}$ consists of all matrices W whose columns indexed by π_i are equal, for all $i = 1, 2, \ldots, k$. Hence, $\mathcal{I}^\sigma \cap \mathcal{M}_r \cong \{W \in \mathbb{R}^{m \times k} \mid \operatorname{rank}(W) \leq r\}$ is an irreducible variety.

Lemma: Let $G \subset \mathcal{S}_n$.
The set of G-invariant $W \in \mathbb{R}^{m \times n}$ is $\mathcal{I}^\sigma$ for some $\sigma \in \mathcal{S}_n$.

What are all ways to parametrize $I^\sigma \cap \mathcal{M}_r$ with autoencoders?

Invariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{m\times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \pi_2 \circ \ldots \circ \pi_k$ into disjoint cycles.

Lemma: The linear space $\mathcal{I}^\sigma$ of σ-invariant $W \in \mathbb{R}^{m\times n}$ consists of all matrices W whose columns indexed by π_i are equal, for all $i = 1, 2, \ldots, k$. Hence, $\mathcal{I}^\sigma \cap \mathcal{M}_r \cong \{W \in \mathbb{R}^{m\times k} \mid \mathrm{rank}(W) \leq r\}$ is an irreducible variety.

Lemma: Let $G \subset \mathcal{S}_n$.
The set of G-invariant $W \in \mathbb{R}^{m\times n}$ is $\mathcal{I}^\sigma$ for some $\sigma \in \mathcal{S}_n$.

What are all ways to parametrize $I^\sigma \cap \mathcal{M}_r$ with autoencoders?

Lemma: $\{(A, B) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{k\times n} \mid \mathrm{rank}(AB) = k, AB \in \mathcal{I}^\sigma\} =$

Invariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{m\times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \pi_2 \circ \ldots \circ \pi_k$ into disjoint cycles.

Lemma: The linear space $\mathcal{I}^\sigma$ of σ-invariant $W \in \mathbb{R}^{m\times n}$ consists of all matrices W whose columns indexed by π_i are equal, for all $i = 1, 2, \ldots, k$. Hence, $\mathcal{I}^\sigma \cap \mathcal{M}_r \cong \{W \in \mathbb{R}^{m\times k} \mid \mathrm{rank}(W) \leq r\}$ is an irreducible variety.

Lemma: Let $G \subset \mathcal{S}_n$.
The set of G-invariant $W \in \mathbb{R}^{m\times n}$ is $\mathcal{I}^\sigma$ for some $\sigma \in \mathcal{S}_n$.

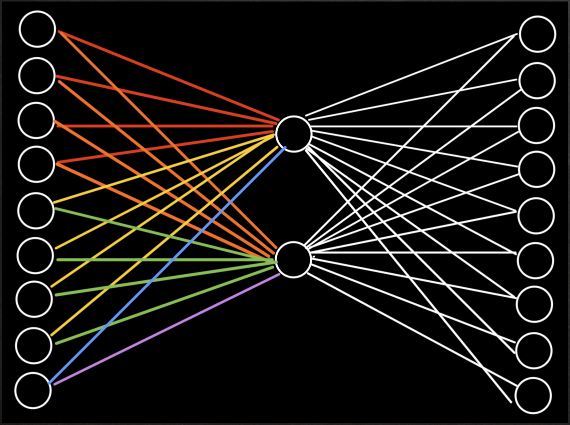
What are all ways to parametrize $I^\sigma \cap \mathcal{M}_r$ with autoencoders?

Lemma: $\{(A, B) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{k\times n} \mid \mathrm{rank}(AB) = k, AB \in \mathcal{I}^\sigma\} =$
$\{A \in \mathbb{R}^{m\times k} \mid \mathrm{rank}(A) = k\} \times \{B \in \mathbb{R}^{k\times n} \mid \text{columns indexed by } \pi_i \text{ are equal}\}$

$\Rightarrow$ σ induces weight sharing on the encoder!

running example

$$\sigma : \mathbb{R}^9 \longrightarrow \mathbb{R}^9, \quad \begin{array}{|c|c|c|}\hline a_{11} & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array} \longmapsto \begin{array}{|c|c|c|}\hline a_{31} & a_{21} & a_{11} \\ \hline a_{32} & a_{22} & a_{12} \\ \hline a_{33} & a_{23} & a_{13} \\ \hline \end{array}$$



has function space $\mathcal{I}^\sigma \cap \mathcal{M}_2$

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \operatorname{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$ represented by $P_\sigma \in \mathbb{R}^{n\times n}$.

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$ represented by $P_\sigma \in \mathbb{R}^{n\times n}$.

Idea: Let $T \in \mathrm{GL}_n$.
W is P_σ-equivariant iff $T^{-1}WT$ is $T^{-1}P_\sigma T$-equivariant.

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n \times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$ represented by $P_\sigma \in \mathbb{R}^{n \times n}$.

Idea: Let $T \in \mathrm{GL}_n$.
W is P_σ-equivariant iff $T^{-1}WT$ is $T^{-1}P_\sigma T$-equivariant.

This base change also preserves rank!

running example

$$\sigma : \mathbb{R}^9 \longrightarrow \mathbb{R}^9, \quad \begin{array}{|c|c|c|}\hline a_{11} & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array} \longmapsto \begin{array}{|c|c|c|}\hline a_{31} & a_{21} & a_{11} \\ \hline a_{32} & a_{22} & a_{12} \\ \hline a_{33} & a_{23} & a_{13} \\ \hline \end{array}$$

$P = P_\sigma$

$$\left[\begin{array}{cccc|cccc|c} 0 & 0 & 0 & 1 & & & & & \\ 1 & 0 & 0 & 0 & & & & & \\ 0 & 1 & 0 & 0 & & 0 & & & 0 \\ 0 & 0 & 1 & 0 & & & & & \\ \hline & & & & 0 & 0 & 0 & 1 & \\ & & & & 1 & 0 & 0 & 0 & \\ & & 0 & & 0 & 1 & 0 & 0 & 0 \\ & & & & 0 & 0 & 1 & 0 & \\ \hline & & 0 & & & & 0 & & 1 \end{array}\right]$$

P-equivariant matrices

$$\left[\begin{array}{cccc|cccc|c} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \varepsilon_3 \\ \alpha_4 & \alpha_1 & \alpha_2 & \alpha_3 & \beta_4 & \beta_1 & \beta_2 & \beta_3 & \varepsilon_3 \\ \alpha_3 & \alpha_4 & \alpha_1 & \alpha_2 & \beta_3 & \beta_4 & \beta_1 & \beta_2 & \varepsilon_3 \\ \alpha_2 & \alpha_3 & \alpha_4 & \alpha_1 & \beta_2 & \beta_3 & \beta_4 & \beta_1 & \varepsilon_3 \\ \hline \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \delta_1 & \delta_2 & \delta_3 & \delta_4 & \varepsilon_4 \\ \gamma_4 & \gamma_1 & \gamma_2 & \gamma_3 & \delta_4 & \delta_1 & \delta_2 & \delta_3 & \varepsilon_4 \\ \gamma_3 & \gamma_4 & \gamma_1 & \gamma_2 & \delta_3 & \delta_4 & \delta_1 & \delta_2 & \varepsilon_4 \\ \gamma_2 & \gamma_3 & \gamma_4 & \gamma_1 & \delta_2 & \delta_3 & \delta_4 & \delta_1 & \varepsilon_4 \\ \hline \varepsilon_1 & \varepsilon_1 & \varepsilon_1 & \varepsilon_1 & \varepsilon_2 & \varepsilon_2 & \varepsilon_2 & \varepsilon_2 & \varepsilon_5 \end{array}\right]$$

running example

$$\sigma : \mathbb{R}^9 \longrightarrow \mathbb{R}^9, \quad \begin{array}{|c|c|c|}\hline a_{11} & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array} \longmapsto \begin{array}{|c|c|c|}\hline a_{31} & a_{21} & a_{11} \\ \hline a_{32} & a_{22} & a_{12} \\ \hline a_{33} & a_{23} & a_{13} \\ \hline \end{array}$$

P = diagonalization of P_σ

$$\left[\begin{array}{cccc|cccc|c} 1 & 0 & 0 & 0 & & & & & \\ 0 & i & 0 & 0 & & 0 & & & 0 \\ 0 & 0 & -1 & 0 & & & & & \\ 0 & 0 & 0 & -i & & & & & \\ \hline & & & & 1 & 0 & 0 & 0 & \\ & 0 & & & 0 & i & 0 & 0 & 0 \\ & & & & 0 & 0 & -1 & 0 & \\ & & & & 0 & 0 & 0 & -i & \\ \hline & 0 & & & & 0 & & & 1 \end{array}\right]$$

P-equivariant matrices

$$\left[\begin{array}{cccc|cccc|c} a_{11} & 0 & 0 & 0 & a_{12} & 0 & 0 & 0 & a_{13} \\ 0 & c_{11} & 0 & 0 & 0 & c_{12} & 0 & 0 & 0 \\ 0 & 0 & b_{11} & 0 & 0 & 0 & b_{12} & 0 & 0 \\ 0 & 0 & 0 & d_{11} & 0 & 0 & 0 & d_{12} & 0 \\ \hline a_{21} & 0 & 0 & 0 & a_{22} & 0 & 0 & 0 & a_{23} \\ 0 & c_{21} & 0 & 0 & 0 & c_{22} & 0 & 0 & 0 \\ 0 & 0 & b_{21} & 0 & 0 & 0 & b_{22} & 0 & 0 \\ 0 & 0 & 0 & d_{21} & 0 & 0 & 0 & d_{22} & 0 \\ \hline a_{31} & 0 & 0 & 0 & a_{32} & 0 & 0 & 0 & a_{33} \end{array}\right]$$

running example

$$\sigma : \mathbb{R}^9 \longrightarrow \mathbb{R}^9, \quad \begin{array}{|c|c|c|}\hline a_{11} & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array} \longmapsto \begin{array}{|c|c|c|}\hline a_{31} & a_{21} & a_{11} \\ \hline a_{32} & a_{22} & a_{12} \\ \hline a_{33} & a_{23} & a_{13} \\ \hline \end{array}$$

P = diagonalization of P_σ

$$\begin{bmatrix} 1 &&&&&&&& \\ & 1 &&&&&&& \\ && 1 &&&&&& \\ &&& -1 &&&&& \\ &&&& -1 &&&& \\ &&&&& i &&& \\ &&&&&& i && \\ &&&&&&& -i & \\ &&&&&&&& -i \end{bmatrix}$$

P-equivariant matrices

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} &&&&&& \\ a_{21} & a_{22} & a_{23} &&&&&& \\ a_{31} & a_{32} & a_{33} &&&&&& \\ &&& b_{11} & b_{12} &&&& \\ &&& b_{21} & b_{22} &&&& \\ &&&&& c_{11} & c_{12} && \\ &&&&& c_{21} & c_{22} && \\ &&&&&&& d_{11} & d_{12} \\ &&&&&&& d_{21} & d_{22} \end{bmatrix}$$

running example

$$W = \begin{bmatrix} a_{11} & a_{12} & a_{13} & & & & & & \\ a_{21} & a_{22} & a_{23} & & & & & & \\ a_{31} & a_{32} & a_{33} & & & & & & \\ & & & b_{11} & b_{12} & & & & \\ & & & b_{21} & b_{22} & & & & \\ & & & & & c_{11} & c_{12} & & \\ & & & & & c_{21} & c_{22} & & \\ & & & & & & & d_{11} & d_{12} \\ & & & & & & & d_{21} & d_{22} \end{bmatrix}$$

There are 10 ways how W can have rank 2:

running example

$$W = \begin{bmatrix} a_{11} & a_{12} & a_{13} & & & & & & \\ a_{21} & a_{22} & a_{23} & & & & & & \\ a_{31} & a_{32} & a_{33} & & & & & & \\ & & & b_{11} & b_{12} & & & & \\ & & & b_{21} & b_{22} & & & & \\ & & & & & c_{11} & c_{12} & & \\ & & & & & c_{21} & c_{22} & & \\ & & & & & & & d_{11} & d_{12} \\ & & & & & & & d_{21} & d_{22} \end{bmatrix}$$

There are 10 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; other blocks are 0

running example

$$W = \begin{bmatrix} a_{11} & a_{12} & a_{13} & & & & & & \\ a_{21} & a_{22} & a_{23} & & & & & & \\ a_{31} & a_{32} & a_{33} & & & & & & \\ & & & b_{11} & b_{12} & & & & \\ & & & b_{21} & b_{22} & & & & \\ & & & & & c_{11} & c_{12} & & \\ & & & & & c_{21} & c_{22} & & \\ & & & & & & & d_{11} & d_{12} \\ & & & & & & & d_{21} & d_{22} \end{bmatrix}$$

There are 10 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; other blocks are 0 $\rightsquigarrow$ 4 components of $\mathcal{E}^\sigma \cap \mathcal{M}_2$

running example

$$W = \begin{bmatrix} a_{11} & a_{12} & a_{13} & & & & & & \\ a_{21} & a_{22} & a_{23} & & & & & & \\ a_{31} & a_{32} & a_{33} & & & & & & \\ & & & b_{11} & b_{12} & & & & \\ & & & b_{21} & b_{22} & & & & \\ & & & & & c_{11} & c_{12} & & \\ & & & & & c_{21} & c_{22} & & \\ & & & & & & & d_{11} & d_{12} \\ & & & & & & & d_{21} & d_{22} \end{bmatrix}$$

There are 10 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; other blocks are 0 $\rightsquigarrow$ 4 components of $\mathcal{E}^\sigma \cap \mathcal{M}_2$
- Two distinct blocks have rank 1; other blocks are 0

running example

$$W = \begin{bmatrix} a_{11} & a_{12} & a_{13} & & & & & & \\ a_{21} & a_{22} & a_{23} & & & & & & \\ a_{31} & a_{32} & a_{33} & & & & & & \\ & & & b_{11} & b_{12} & & & & \\ & & & b_{21} & b_{22} & & & & \\ & & & & & c_{11} & c_{12} & & \\ & & & & & c_{21} & c_{22} & & \\ & & & & & & & d_{11} & d_{12} \\ & & & & & & & d_{21} & d_{22} \end{bmatrix}$$

There are 10 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; other blocks are 0 $\rightsquigarrow$ 4 components of $\mathcal{E}^\sigma \cap \mathcal{M}_2$
- Two distinct blocks have rank 1; other blocks are 0 $\rightsquigarrow$ 6 components of $\mathcal{E}^\sigma \cap \mathcal{M}_2$

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$ represented by $P_\sigma \in \mathbb{R}^{n\times n}$.

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n \times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$ represented by $P_\sigma \in \mathbb{R}^{n \times n}$.

The decomposition $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles chops P_σ into blocks.

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \operatorname{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$ represented by $P_\sigma \in \mathbb{R}^{n\times n}$.

The decomposition $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles chops P_σ into blocks.
Chop W into blocks following the same pattern!

$$P_\sigma = \left[\begin{array}{cccc|cccc|c} 0&0&0&1 & & & & & \\ 1&0&0&0 & &0& & & 0\\ 0&1&0&0 & & & & & \\ 0&0&1&0 & & & & & \\ \hline & & & & 0&0&0&1 & \\ &0& & & 1&0&0&0 & 0\\ & & & & 0&1&0&0 & \\ & & & & 0&0&1&0 & \\ \hline &0& & & &0& & & 1 \end{array}\right] \qquad W = \left[\begin{array}{cccc|cccc|c} \alpha_1&\alpha_2&\alpha_3&\alpha_4 & \beta_1&\beta_2&\beta_3&\beta_4 & \varepsilon_3\\ \alpha_4&\alpha_1&\alpha_2&\alpha_3 & \beta_4&\beta_1&\beta_2&\beta_3 & \varepsilon_3\\ \alpha_3&\alpha_4&\alpha_1&\alpha_2 & \beta_3&\beta_4&\beta_1&\beta_2 & \varepsilon_3\\ \alpha_2&\alpha_3&\alpha_4&\alpha_1 & \beta_2&\beta_3&\beta_4&\beta_1 & \varepsilon_3\\ \hline \gamma_1&\gamma_2&\gamma_3&\gamma_4 & \delta_1&\delta_2&\delta_3&\delta_4 & \varepsilon_4\\ \gamma_4&\gamma_1&\gamma_2&\gamma_3 & \delta_4&\delta_1&\delta_2&\delta_3 & \varepsilon_4\\ \gamma_3&\gamma_4&\gamma_1&\gamma_2 & \delta_3&\delta_4&\delta_1&\delta_2 & \varepsilon_4\\ \gamma_2&\gamma_3&\gamma_4&\gamma_1 & \delta_2&\delta_3&\delta_4&\delta_1 & \varepsilon_4\\ \hline \varepsilon_1&\varepsilon_1&\varepsilon_1&\varepsilon_1 & \varepsilon_2&\varepsilon_2&\varepsilon_2&\varepsilon_2 & \varepsilon_5 \end{array}\right]$$

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \operatorname{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$ represented by $P_\sigma \in \mathbb{R}^{n\times n}$.

The decomposition $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles chops P_σ into blocks. Chop W into blocks following the same pattern!

$$P_\sigma = \left[\begin{array}{cccc|cccc|c} 0 & 0 & 0 & 1 & & & & & \\ 1 & 0 & 0 & 0 & & 0 & & & 0 \\ 0 & 1 & 0 & 0 & & & & & \\ 0 & 0 & 1 & 0 & & & & & \\ \hline & & & & 0 & 0 & 0 & 1 & \\ & 0 & & & 1 & 0 & 0 & 0 & 0 \\ & & & & 0 & 1 & 0 & 0 & \\ & & & & 0 & 0 & 1 & 0 & \\ \hline & 0 & & & & 0 & & & 1 \end{array}\right] \qquad W = \left[\begin{array}{cccc|cccc|c} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \varepsilon_3 \\ \alpha_4 & \alpha_1 & \alpha_2 & \alpha_3 & \beta_4 & \beta_1 & \beta_2 & \beta_3 & \varepsilon_3 \\ \alpha_3 & \alpha_4 & \alpha_1 & \alpha_2 & \beta_3 & \beta_4 & \beta_1 & \beta_2 & \varepsilon_3 \\ \alpha_2 & \alpha_3 & \alpha_4 & \alpha_1 & \beta_2 & \beta_3 & \beta_4 & \beta_1 & \varepsilon_3 \\ \hline \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \delta_1 & \delta_2 & \delta_3 & \delta_4 & \varepsilon_4 \\ \gamma_4 & \gamma_1 & \gamma_2 & \gamma_3 & \delta_4 & \delta_1 & \delta_2 & \delta_3 & \varepsilon_4 \\ \gamma_3 & \gamma_4 & \gamma_1 & \gamma_2 & \delta_3 & \delta_4 & \delta_1 & \delta_2 & \varepsilon_4 \\ \gamma_2 & \gamma_3 & \gamma_4 & \gamma_1 & \delta_2 & \delta_3 & \delta_4 & \delta_1 & \varepsilon_4 \\ \hline \varepsilon_1 & \varepsilon_1 & \varepsilon_1 & \varepsilon_1 & \varepsilon_2 & \varepsilon_2 & \varepsilon_2 & \varepsilon_2 & \varepsilon_5 \end{array}\right]$$

Lemma: A matrix $W \in \mathbb{R}^{n\times n}$ is σ-equivariant iff each block is a (possily non-square) circulant matrix.

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \operatorname{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$ represented by $P_\sigma \in \mathbb{R}^{n\times n}$.

The decomposition $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles chops P_σ into blocks.
Chop W into blocks following the same pattern!

$$P_\sigma = \left[\begin{array}{cccc|cccc|c} 0 & 0 & 0 & 1 & & & & & \\ 1 & 0 & 0 & 0 & & 0 & & & 0 \\ 0 & 1 & 0 & 0 & & & & & \\ 0 & 0 & 1 & 0 & & & & & \\ \hline & & & & 0 & 0 & 0 & 1 & \\ & 0 & & & 1 & 0 & 0 & 0 & 0 \\ & & & & 0 & 1 & 0 & 0 & \\ & & & & 0 & 0 & 1 & 0 & \\ \hline & 0 & & & & 0 & & & 1 \end{array}\right] \qquad W = \left[\begin{array}{cccc|cccc|c} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \varepsilon_3 \\ \alpha_4 & \alpha_1 & \alpha_2 & \alpha_3 & \beta_4 & \beta_1 & \beta_2 & \beta_3 & \varepsilon_3 \\ \alpha_3 & \alpha_4 & \alpha_1 & \alpha_2 & \beta_3 & \beta_4 & \beta_1 & \beta_2 & \varepsilon_3 \\ \alpha_2 & \alpha_3 & \alpha_4 & \alpha_1 & \beta_2 & \beta_3 & \beta_4 & \beta_1 & \varepsilon_3 \\ \hline \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \delta_1 & \delta_2 & \delta_3 & \delta_4 & \varepsilon_4 \\ \gamma_4 & \gamma_1 & \gamma_2 & \gamma_3 & \delta_4 & \delta_1 & \delta_2 & \delta_3 & \varepsilon_4 \\ \gamma_3 & \gamma_4 & \gamma_1 & \gamma_2 & \delta_3 & \delta_4 & \delta_1 & \delta_2 & \varepsilon_4 \\ \gamma_2 & \gamma_3 & \gamma_4 & \gamma_1 & \delta_2 & \delta_3 & \delta_4 & \delta_1 & \varepsilon_4 \\ \hline \varepsilon_1 & \varepsilon_1 & \varepsilon_1 & \varepsilon_1 & \varepsilon_2 & \varepsilon_2 & \varepsilon_2 & \varepsilon_2 & \varepsilon_5 \end{array}\right]$$

Lemma: A matrix $W \in \mathbb{R}^{n\times n}$ is σ-equivariant iff
each block is a (possily non-square) circulant matrix.

$$\begin{bmatrix} a & a & a \\ a & a & a \\ a & a & a \\ a & a & a \end{bmatrix}, \quad \begin{bmatrix} a & a & a \\ a & a & a \end{bmatrix}, \quad \begin{bmatrix} a & b & a & b \\ b & a & b & a \end{bmatrix}, \quad \ldots$$

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n \times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles and let $\ell_j :=$ length(π_j).

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n \times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles and let $\ell_j :=$ length(π_j).

Diagonalize P_σ and sort the eigenvalues. This yields the diagonal matrix P.

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \operatorname{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles and let $\ell_j :=$ length(π_j).

Diagonalize P_σ and sort the eigenvalues. This yields the diagonal matrix P.

Lemma: A matrix $W \in \mathbb{R}^{n\times n}$ is P-equivariant iff its block diagonal with $\#(\mathbb{Z}/m\mathbb{Z})^\times$ many blocks of size $d_m \times d_m$, where $d_m := \#\{j$ such that $m|\ell_j\}$.

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \operatorname{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles and let $\ell_j :=$ length(π_j).

Diagonalize P_σ and sort the eigenvalues. This yields the diagonal matrix P.

Lemma: A matrix $W \in \mathbb{R}^{n\times n}$ is P-equivariant iff its block diagonal with $\#(\mathbb{Z}/m\mathbb{Z})^\times$ many blocks of size $d_m \times d_m$, where $d_m := \#\{j \text{ such that } m|\ell_j\}$.

$$P = \begin{bmatrix} 1 &&&&&&&& \\ & 1 &&&&&&& \\ && 1 &&&&&& \\ &&& -1 &&&&& \\ &&&& -1 &&&& \\ &&&&& i &&& \\ &&&&&& i && \\ &&&&&&& -i & \\ &&&&&&&& -i \end{bmatrix} \qquad W = \begin{bmatrix} a_{11} & a_{12} & a_{13} &&&&&& \\ a_{21} & a_{22} & a_{23} &&&&&& \\ a_{31} & a_{32} & a_{33} &&&&&& \\ &&& b_{11} & b_{12} &&&& \\ &&& b_{21} & b_{22} &&&& \\ &&&&& c_{11} & c_{12} && \\ &&&&& c_{21} & c_{22} && \\ &&&&&&& d_{11} & d_{12} \\ &&&&&&& d_{21} & d_{22} \end{bmatrix}$$

$\ell_1 = 4,\ \ell_2 = 4,\ \ell_3 = 1$

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \mathrm{rank}(W) \le r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles and let $\ell_j :=$ length(π_j).

Diagonalize P_σ and sort the eigenvalues. This yields the diagonal matrix P.

Lemma: A matrix $W \in \mathbb{R}^{n\times n}$ is P-equivariant iff its block diagonal with $\#(\mathbb{Z}/m\mathbb{Z})^\times$ many blocks of size $d_m \times d_m$, where $d_m := \#\{j \text{ such that } m|\ell_j\}$.

$$P = \begin{bmatrix} 1 &&&&&&&& \\ & 1 &&&&&&& \\ && 1 &&&&&& \\ &&& -1 &&&&& \\ &&&& -1 &&&& \\ &&&&& i &&& \\ &&&&&& i && \\ &&&&&&& -i & \\ &&&&&&&& -i \end{bmatrix} \qquad W = \begin{bmatrix} a_{11} & a_{12} & a_{13} &&&&&& \\ a_{21} & a_{22} & a_{23} &&&&&& \\ a_{31} & a_{32} & a_{33} &&&&&& \\ &&& b_{11} & b_{12} &&&& \\ &&& b_{21} & b_{22} &&&& \\ &&&&& c_{11} & c_{12} && \\ &&&&& c_{21} & c_{22} && \\ &&&&&&& d_{11} & d_{12} \\ &&&&&&& d_{21} & d_{22} \end{bmatrix}$$

$\ell_1 = 4,\ \ell_2 = 4,\ \ell_3 = 1$

$d_1 = 3,\ d_2 = 2,\ d_3 = 0,\ d_4 = 2,\ d_5 = 0,\ \ldots$

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n \times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles and let $\ell_j :=$ length(π_j).

Diagonalize P_σ and sort the eigenvalues. This yields the diagonal matrix P.

Lemma: A matrix $W \in \mathbb{R}^{n \times n}$ is P-equivariant iff its block diagonal with $\#(\mathbb{Z}/m\mathbb{Z})^\times$ many blocks of size $d_m \times d_m$, where $d_m := \#\{j \text{ such that } m|\ell_j\}$.

$$P = \begin{bmatrix} 1 &&&&&&&& \\ & 1 &&&&&&& \\ && 1 &&&&&& \\ &&& -1 &&&&& \\ &&&& -1 &&&& \\ &&&&& i &&& \\ &&&&&& i && \\ &&&&&&& -i & \\ &&&&&&&& -i \end{bmatrix} \qquad W = \begin{bmatrix} a_{11} & a_{12} & a_{13} &&&&&& \\ a_{21} & a_{22} & a_{23} &&&&&& \\ a_{31} & a_{32} & a_{33} &&&&&& \\ &&& b_{11} & b_{12} &&&& \\ &&& b_{21} & b_{22} &&&& \\ &&&&& c_{11} & c_{12} && \\ &&&&& c_{21} & c_{22} && \\ &&&&&&& d_{11} & d_{12} \\ &&&&&&& d_{21} & d_{22} \end{bmatrix}$$

$\ell_1 = 4,\ \ell_2 = 4,\ \ell_3 = 1$

$d_1 = 3,\ d_2 = 2,\ d_3 = 0,\ d_4 = 2,\ d_5 = 0,\ \ldots$

$\#(\mathbb{Z}/1\mathbb{Z})^\times = 1,\ \#(\mathbb{Z}/2\mathbb{Z})^\times = 1,\ \#(\mathbb{Z}/4\mathbb{Z})^\times = 2$

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles and let $\ell_j := \text{length}(\pi_j)$.

Diagonalize P_σ and sort the eigenvalues. This yields the diagonal matrix P.

Lemma: A matrix $W \in \mathbb{R}^{n\times n}$ is P-equivariant iff its block diagonal with $\#(\mathbb{Z}/m\mathbb{Z})^\times$ many blocks of size $d_m \times d_m$, where $d_m := \#\{j \text{ such that } m|\ell_j\}$.

Theorem: The irreducible components of $\mathcal{E}^\sigma \cap \mathcal{M}_r$ over $\mathbb{C}$ are in 1-to-1 correspondence with the integer solutions $(r_{m,u})$ of

$$\sum_{m\in\mathbb{Z}_{>0}} \sum_{u\in(\mathbb{Z}/m\mathbb{Z})^\times} r_{m,u} = r, \qquad \text{where } 0 \leq r_{m,u} \leq d_m.$$

Equivariance

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \operatorname{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

Decompose $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles and let $\ell_j := \text{length}(\pi_j)$.

Diagonalize P_σ and sort the eigenvalues. This yields the diagonal matrix P.

Lemma: A matrix $W \in \mathbb{R}^{n\times n}$ is P-equivariant iff its block diagonal with $\#(\mathbb{Z}/m\mathbb{Z})^\times$ many blocks of size $d_m \times d_m$, where $d_m := \#\{j \text{ such that } m|\ell_j\}$.

Theorem: The irreducible components of $\mathcal{E}^\sigma \cap \mathcal{M}_r$ over $\mathbb{C}$ are in 1-to-1 correspondence with the integer solutions $(r_{m,u})$ of

$$\sum_{m\in\mathbb{Z}_{>0}} \sum_{u\in(\mathbb{Z}/m\mathbb{Z})^\times} r_{m,u} = r, \quad \text{where } 0 \leq r_{m,u} \leq d_m.$$

The component indexed by $(r_{m,u})$ is

$$\cong \prod_{m\in\mathbb{Z}_{>0}} \prod_{u\in(\mathbb{Z}/m\mathbb{Z})^\times} \{A \in \mathbb{C}^{d_m\times d_m} \mid \operatorname{rank}(A) \leq r_{m,u}\}.$$

Equivariance over $\mathbb{R}$

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

To diagonalize P_σ, we need a complex base change!

Equivariance over $\mathbb{R}$

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \operatorname{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

To diagonalize P_σ, we need a complex base change!

Remedy: Replace complex conjugated pairs of eigenvectors by their real and imaginary parts.

Equivariance over $\mathbb{R}$

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

To diagonalize P_σ, we need a complex base change!

Remedy: Replace complex conjugated pairs of eigenvectors by their real and imaginary parts.

This new basis can be scaled to become orthonormal!

Example: to diagonalize $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$, use base change $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}$

Equivariance over $\mathbb{R}$

Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n\times n} \mid \mathrm{rank}(W) \leq r\}$ and $\sigma \in \mathcal{S}_n$.

To diagonalize P_σ, we need a complex base change!

Remedy: Replace complex conjugated pairs of eigenvectors by their real and imaginary parts.

This new basis can be scaled to become orthonormal!

Example: to diagonalize $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$, use base change $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}$

$$\rightsquigarrow \frac{1}{2}\begin{bmatrix} 1 & \sqrt{2} & 1 & 0 \\ 1 & 0 & -1 & -\sqrt{2} \\ 1 & -\sqrt{2} & 1 & 0 \\ 1 & 0 & -1 & \sqrt{2} \end{bmatrix} \in O_4(\mathbb{R})$$

running example

$$\sigma : \mathbb{R}^9 \longrightarrow \mathbb{R}^9, \quad \begin{array}{|c|c|c|}\hline a_{11} & a_{12} & a_{13} \\ \hline a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ \hline \end{array} \longmapsto \begin{array}{|c|c|c|}\hline a_{31} & a_{21} & a_{11} \\ \hline a_{32} & a_{22} & a_{12} \\ \hline a_{33} & a_{23} & a_{13} \\ \hline \end{array}$$

$P = P_\sigma$ after $O_9(\mathbb{R})$-base change

$$\begin{bmatrix} 1 &&&&&&&& \\ & 1 &&&&&&& \\ && 1 &&&&&& \\ &&& -1 &&&&& \\ &&&& -1 &&&& \\ &&&&& 0 & 1 && \\ &&&&& -1 & 0 && \\ &&&&&&& 0 & 1 \\ &&&&&&& -1 & 0 \end{bmatrix}$$

P-equivariant matrices

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} &&&&&& \\ a_{21} & a_{22} & a_{23} &&&&&& \\ a_{31} & a_{32} & a_{33} &&&&&& \\ &&& b_{11} & b_{12} &&&& \\ &&& b_{21} & b_{22} &&&& \\ &&&&& c_1 & -c_2 & d_1 & -d_2 \\ &&&&& c_2 & c_1 & d_2 & d_1 \\ &&&&& e_1 & -e_2 & f_1 & -f_2 \\ &&&&& e_2 & e_1 & f_2 & f_1 \end{bmatrix}$$

running example

$$W = \begin{bmatrix} a_{11} & a_{12} & a_{13} & & & & & & \\ a_{21} & a_{22} & a_{23} & & & & & & \\ a_{31} & a_{32} & a_{33} & & & & & & \\ & & & b_{11} & b_{12} & & & & \\ & & & b_{21} & b_{22} & & & & \\ & & & & & c_1 & -c_2 & d_1 & -d_2 \\ & & & & & c_2 & c_1 & d_2 & d_1 \\ & & & & & e_1 & -e_2 & f_1 & -f_2 \\ & & & & & e_2 & e_1 & f_2 & f_1 \end{bmatrix}$$

There are 4 ways how W can have rank 2:

running example

$$W = \begin{bmatrix} a_{11} & a_{12} & a_{13} & & & & & & \\ a_{21} & a_{22} & a_{23} & & & & & & \\ a_{31} & a_{32} & a_{33} & & & & & & \\ & & & b_{11} & b_{12} & & & & \\ & & & b_{21} & b_{22} & & & & \\ & & & & & c_1 & -c_2 & d_1 & -d_2 \\ & & & & & c_2 & c_1 & d_2 & d_1 \\ & & & & & e_1 & -e_2 & f_1 & -f_2 \\ & & & & & e_2 & e_1 & f_2 & f_1 \end{bmatrix}$$

There are 4 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; other blocks are 0

running example

$$W = \begin{bmatrix} a_{11} & a_{12} & a_{13} & & & & & & \\ a_{21} & a_{22} & a_{23} & & & & & & \\ a_{31} & a_{32} & a_{33} & & & & & & \\ & & & b_{11} & b_{12} & & & & \\ & & & b_{21} & b_{22} & & & & \\ & & & & & c_1 & -c_2 & d_1 & -d_2 \\ & & & & & c_2 & c_1 & d_2 & d_1 \\ & & & & & e_1 & -e_2 & f_1 & -f_2 \\ & & & & & e_2 & e_1 & f_2 & f_1 \end{bmatrix}$$

There are 4 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; other blocks are 0 $\rightsquigarrow$ 3 components of $\mathcal{E}^\sigma \cap \mathcal{M}_2$

running example

$$W = \begin{bmatrix} a_{11} & a_{12} & a_{13} & & & & & & \\ a_{21} & a_{22} & a_{23} & & & & & & \\ a_{31} & a_{32} & a_{33} & & & & & & \\ & & & b_{11} & b_{12} & & & & \\ & & & b_{21} & b_{22} & & & & \\ & & & & & c_1 & -c_2 & d_1 & -d_2 \\ & & & & & c_2 & c_1 & d_2 & d_1 \\ & & & & & e_1 & -e_2 & f_1 & -f_2 \\ & & & & & e_2 & e_1 & f_2 & f_1 \end{bmatrix}$$

There are 4 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; other blocks are 0 $\rightsquigarrow$ 3 components of $\mathcal{E}^\sigma \cap \mathcal{M}_2$
- Two first 2 blocks have rank 1; last block is 0

running example

$$W = \begin{bmatrix} a_{11} & a_{12} & a_{13} & & & & & & \\ a_{21} & a_{22} & a_{23} & & & & & & \\ a_{31} & a_{32} & a_{33} & & & & & & \\ & & & b_{11} & b_{12} & & & & \\ & & & b_{21} & b_{22} & & & & \\ & & & & & c_1 & -c_2 & d_1 & -d_2 \\ & & & & & c_2 & c_1 & d_2 & d_1 \\ & & & & & e_1 & -e_2 & f_1 & -f_2 \\ & & & & & e_2 & e_1 & f_2 & f_1 \end{bmatrix}$$

There are 4 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; other blocks are 0 $\rightsquigarrow$ 3 components of $\mathcal{E}^\sigma \cap \mathcal{M}_2$
- Two first 2 blocks have rank 1; last block is 0 $\rightsquigarrow$ 1 component of $\mathcal{E}^\sigma \cap \mathcal{M}_2$

Equivariance over $\mathbb{R}$

In general: After the $O_n(\mathbb{R})$-base change, the σ-equivariant matrices become block diagonal:

- at most 2 blocks are arbitrary (corresponding to eigenvalues ± 1 of P_σ);
- all other blocks are $2m \times 2m$ matrices consisting of m^2 matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$.

Equivariance over $\mathbb{R}$

In general: After the $O_n(\mathbb{R})$-base change, the σ-equivariant matrices become block diagonal:

- at most 2 blocks are arbitrary (corresponding to eigenvalues ± 1 of P_σ);
- all other blocks are $2m \times 2m$ matrices consisting of m^2 matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$.

Blocks of the latter kind have even rank!

Equivariance over $\mathbb{R}$

In general: After the $O_n(\mathbb{R})$-base change, the σ-equivariant matrices become block diagonal:

- at most 2 blocks are arbitrary (corresponding to eigenvalues ± 1 of P_σ);
- all other blocks are $2m \times 2m$ matrices consisting of m^2 matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$.

Blocks of the latter kind have even rank!
The variety of such blocks of bounded rank is irreducible.

Equivariance over $\mathbb{R}$

In general: After the $O_n(\mathbb{R})$-base change, the σ-equivariant matrices become block diagonal:

- at most 2 blocks are arbitrary (corresponding to eigenvalues ± 1 of P_σ);
- all other blocks are $2m \times 2m$ matrices consisting of m^2 matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$.

Blocks of the latter kind have even rank!
The variety of such blocks of bounded rank is irreducible.

$\Rightarrow$ we can list all irreducible components of $\mathcal{E}^\sigma \cap \mathcal{M}_r$,

Equivariance over $\mathbb{R}$

In general: After the $O_n(\mathbb{R})$-base change, the σ-equivariant matrices become block diagonal:

- at most 2 blocks are arbitrary (corresponding to eigenvalues ± 1 of P_σ);
- all other blocks are $2m \times 2m$ matrices consisting of m^2 matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$.

Blocks of the latter kind have even rank!
The variety of such blocks of bounded rank is irreducible.

$\Rightarrow$ we can list all irreducible components of $\mathcal{E}^\sigma \cap \mathcal{M}_r$,
parametrize them via autoencoders,

Equivariance over $\mathbb{R}$

In general: After the $O_n(\mathbb{R})$-base change, the σ-equivariant matrices become block diagonal:

- at most 2 blocks are arbitrary (corresponding to eigenvalues ± 1 of P_σ);
- all other blocks are $2m \times 2m$ matrices consisting of m^2 matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$.

Blocks of the latter kind have even rank!
The variety of such blocks of bounded rank is irreducible.

$\Rightarrow$ we can list all irreducible components of $\mathcal{E}^\sigma \cap \mathcal{M}_r$,
parametrize them via autoencoders,
understand their algebraic properties such as dimension, degree, ...

Which of these 4 components is best ??

$$W = \begin{bmatrix} a_{11} & a_{12} & a_{13} & & & & & & \\ a_{21} & a_{22} & a_{23} & & & & & & \\ a_{31} & a_{32} & a_{33} & & & & & & \\ & & & b_{11} & b_{12} & & & & \\ & & & b_{21} & b_{22} & & & & \\ & & & & & c_1 & -c_2 & d_1 & -d_2 \\ & & & & & c_2 & c_1 & d_2 & d_1 \\ & & & & & e_1 & -e_2 & f_1 & -f_2 \\ & & & & & e_2 & e_1 & f_2 & f_1 \end{bmatrix}$$

There are 4 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; other blocks are 0 $\rightsquigarrow$ 3 components of $\mathcal{E}^\sigma \cap \mathcal{M}_2$
- Two first 2 blocks have rank 1; last block is 0 $\rightsquigarrow$ 1 component of $\mathcal{E}^\sigma \cap \mathcal{M}_2$

Euclidean distance degree

Lemma: Given a sufficient amount of training data that is sufficiently generic, training a network with function space $\mathcal{M}$ using the squared-error loss means to solve an optimization problem of the form

$$\min_{W \in \mathcal{M}} \|W - U\|_F^2, \qquad \text{where } U \text{ is a generic matrix encoding the data.}$$

Euclidean distance degree

Lemma: Given a sufficient amount of training data that is sufficiently generic, training a network with function space $\mathcal{M}$ using the squared-error loss means to solve an optimization problem of the form

$$\min_{W \in \mathcal{M}} \|W - U\|_F^2, \quad \text{where } U \text{ is a generic matrix encoding the data.}$$

The number of complex critical points of that problem measures its algebraic complexity,

Euclidean distance degree

Lemma: Given a sufficient amount of training data that is sufficiently generic, training a network with function space $\mathcal{M}$ using the squared-error loss means to solve an optimization problem of the form

$$\min_{W \in \mathcal{M}} \|W - U\|_F^2, \qquad \text{where } U \text{ is a generic matrix encoding the data.}$$

The number of complex critical points of that problem measures its algebraic complexity, called its Euclidean distance degree. Notation: $\mathrm{EDdeg}(\mathcal{M})$.

Euclidean distance degree

Lemma: Given a sufficient amount of training data that is sufficiently generic, training a network with function space $\mathcal{M}$ using the squared-error loss means to solve an optimization problem of the form

$$\min_{W \in \mathcal{M}} \|W - U\|_F^2, \qquad \text{where } U \text{ is a generic matrix encoding the data.}$$

The number of complex critical points of that problem measures its algebraic complexity, called its Euclidean distance degree. Notation: $\mathrm{EDdeg}(\mathcal{M})$.

We can compute EDdeg of each real irreducible component of $\mathcal{E}^\sigma \cap \mathcal{M}_r$!

Euclidean distance degree

Lemma: Given a sufficient amount of training data that is sufficiently generic, training a network with function space $\mathcal{M}$ using the squared-error loss means to solve an optimization problem of the form

$$\min_{W \in \mathcal{M}} \|W - U\|_F^2, \qquad \text{where } U \text{ is a generic matrix encoding the data.}$$

The number of complex critical points of that problem measures its algebraic complexity, called its Euclidean distance degree. Notation: $\mathrm{EDdeg}(\mathcal{M})$.

We can compute EDdeg of each real irreducible component of $\mathcal{E}^\sigma \cap \mathcal{M}_r$!

1. EDdeg stays invariant under orthogonal base changes.

Euclidean distance degree

Lemma: Given a sufficient amount of training data that is sufficiently generic, training a network with function space $\mathcal{M}$ using the squared-error loss means to solve an optimization problem of the form

$$\min_{W \in \mathcal{M}} \|W - U\|_F^2, \qquad \text{where } U \text{ is a generic matrix encoding the data.}$$

The number of complex critical points of that problem measures its algebraic complexity, called its Euclidean distance degree. Notation: $\mathrm{EDdeg}(\mathcal{M})$.

We can compute EDdeg of each real irreducible component of $\mathcal{E}^\sigma \cap \mathcal{M}_r$!

1. EDdeg stays invariant under orthogonal base changes.
2. $\mathrm{EDdeg}(\mathcal{X} \times \mathcal{Y}) = \mathrm{EDdeg}(\mathcal{X}) \cdot \mathrm{EDdeg}(\mathcal{Y})$

Euclidean distance degree

Lemma: Given a sufficient amount of training data that is sufficiently generic, training a network with function space $\mathcal{M}$ using the squared-error loss means to solve an optimization problem of the form

$$\min_{W \in \mathcal{M}} \|W - U\|_F^2, \qquad \text{where } U \text{ is a generic matrix encoding the data.}$$

The number of complex critical points of that problem measures its algebraic complexity, called its Euclidean distance degree. Notation: $\mathrm{EDdeg}(\mathcal{M})$.

We can compute EDdeg of each real irreducible component of $\mathcal{E}^\sigma \cap \mathcal{M}_r$!

1. EDdeg stays invariant under orthogonal base changes.
2. $\mathrm{EDdeg}(\mathcal{X} \times \mathcal{Y}) = \mathrm{EDdeg}(\mathcal{X}) \cdot \mathrm{EDdeg}(\mathcal{Y})$
3. $\mathrm{EDdeg}(\mathcal{X}_{s,d}) = \binom{d}{s}$, where $\mathcal{X}_{s,d}$ is either the space of $d \times d$ matrices of rank $\leq s$ or of $2d \times 2d$ matrices of rank $\leq 2s$ that consist of $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$–submatrices

Which of these 4 components is best ??

$$W = \begin{bmatrix} a_{11} & a_{12} & a_{13} & & & & & & \\ a_{21} & a_{22} & a_{23} & & & & & & \\ a_{31} & a_{32} & a_{33} & & & & & & \\ & & & b_{11} & b_{12} & & & & \\ & & & b_{21} & b_{22} & & & & \\ & & & & & c_1 & -c_2 & d_1 & -d_2 \\ & & & & & c_2 & c_1 & d_2 & d_1 \\ & & & & & e_1 & -e_2 & f_1 & -f_2 \\ & & & & & e_2 & e_1 & f_2 & f_1 \end{bmatrix}$$

There are 4 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; $\rightsquigarrow$ 3 components of $\mathcal{E}^\sigma \cap \mathcal{M}_2$
  with EDdeg 3, 1, and 2, respectively.
- Two first 2 blocks have rank 1; $\rightsquigarrow$ 1 component of $\mathcal{E}^\sigma \cap \mathcal{M}_2$
  with EDdeg $3 \cdot 2 = 6$.

Data science requires us to rethink the schism between mathematical disciplines!

differential geometry ⇒

algebraic geometry ⇒

data science ⇒

Paul Breiding, Kathlén Kohn and Bernd Sturmfels

Metric Algebraic Geometry

Springer Nature