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

Kathlén Kohn





joint work with

Anna-Laura Sattelberger

Vahid Shahverdi

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?

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)

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)

Algebraic settings:

netwo	rk architecture		
activation	network structure	loss	
i start			
			II - XXIX

Algebraic settings:

network architecture		
activation	network structure	loss
identity		
ReLU		
polynomial		4
		II - XXIX

Algebraic settings:

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



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

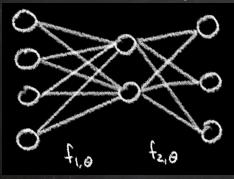
II - XXIX

Algebraic settings:

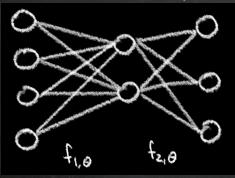
network architecture		rk architecture		
	activation	network structure	loss	
	identity	fully-connected	squared-error loss	= Euclidean dist
	ReLU	convolutional	Wasserstein distance	= polyhedral dist.
	polynomial	group equivariant	cross-entropy	\cong KL divergence

II - XXIX

activation = identity & network structure = fully-connected



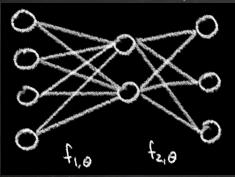
activation = identity & network structure = fully-connected



This network parametrizes linear maps:

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

activation = identity & network structure = fully-connected

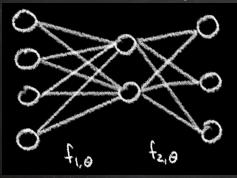


This network parametrizes linear maps:

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

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

activation = identity & network structure = fully-connected



This network parametrizes linear maps:

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

Its function space is $\mathcal{M}_2 = \{ W \in \mathbb{R}^{3 imes 4} \mid \mathrm{rank}(W) \leq 2 \}.$

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

Running Example

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

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 $\mathcal{M}_2 = \operatorname{im}(\mu) = \{W \in \mathbb{R}^{9 \times 9} | \operatorname{rank}(W) \leq 2\}.$

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 $\mathcal{M}_2 = \operatorname{im}(\mu) = \{W \in \mathbb{R}^{9 \times 9} | \operatorname{rank}(W) \leq 2\}.$

Its inputs and outputs are 3×3 images:

a ₁₁	a ₁₂	a ₁₃	
a ₂₁	a ₂₂	a ₂₃	$\in \mathbb{R}^{9}$.
a ₃₁	a ₃₂	a33	

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 $\mathcal{M}_2 = \operatorname{im}(\mu) = \{W \in \mathbb{R}^{9 \times 9} | \operatorname{rank}(W) \leq 2\}.$

Its inputs and outputs are 3×3 images:

a ₁₁	a ₁₂	a ₁₃	
a ₂₁	a ₂₂	a ₂₃	$\in \mathbb{R}^{9}$.
a ₃₁	a ₃₂	a33	

Consider the clockwise rotation by 90°:

$\sigma: \mathbb{R}^9 \longrightarrow \mathbb{R}^9,$			a ₁₃	\mapsto a			a ₁₁
	a ₂₁	a ₂₂	a ₂₃		a ₃₂	a ₂₂	a ₁₂ .
	a ₃₁	a32	a 33		a 33	a ₂₃	a ₁₃

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 $\mathcal{M}_2 = \operatorname{im}(\mu) = \{W \in \mathbb{R}^{9 \times 9} | \operatorname{rank}(W) \leq 2\}.$

Its inputs and outputs are 3×3 images:

a ₁₁	a ₁₂	a ₁₃	
a ₂₁	a ₂₂	a ₂₃	$\in \mathbb{R}^9$.
a ₃₁	a ₃₂	a33	

Consider the clockwise rotation by 90°:

		a ₁₂					a ₁₁
$\sigma: \mathbb{R}^9 \longrightarrow \mathbb{R}^9,$	a ₂₁	a ₂₂	a ₂₃	\mapsto	a ₃₂	a ₂₂	a ₁₂
	a ₃₁	a ₃₂	<i>a</i> 33		a33	a ₂₃	a ₁₃

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

IV - XXIX

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 $\mathcal{M}_2 = \operatorname{im}(\mu) = \{W \in \mathbb{R}^{9 \times 9} | \operatorname{rank}(W) \leq 2\}.$

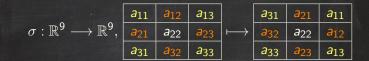
Its inputs and outputs are 3×3 images:

a ₁₁	a ₁₂	a ₁₃	
a ₂₁	a ₂₂	a ₂₃	$\in \mathbb{R}^9$.
a ₃₁	a ₃₂	a33	

Consider the clockwise rotation by 90°:

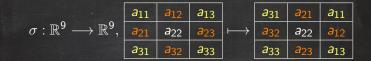
		a ₁₂					a ₁₁
$\sigma: \mathbb{R}^9 \longrightarrow \mathbb{R}^9,$	a ₂₁	a ₂₂	a ₂₃	\mapsto	a ₃₂	a ₂₂	a ₁₂
	a ₃₁	a ₃₂	<i>a</i> 33		a33	a ₂₃	a ₁₃

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



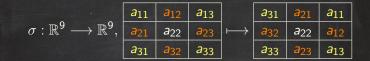
is represented by the permutation matrix

$$P_{\sigma} = egin{bmatrix} 0 & 0 & 0 & 1 & & & & & \ 1 & 0 & 0 & 0 & & 0 & 0 & \ 0 & 1 & 0 & 0 & & & \ 0 & 0 & 1 & 0 & & & \ 0 & 0 & 0 & 0 & 1 & \ 0 & 0 & 1 & 0 & 0 & \ & 0 & 0 & 1 & 0 & \ & 0 & 0 & 1 & 0 & \ & 0 & 0 & 1 & 0 & \ \end{bmatrix}$$



is represented by the permutation matrix

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



is represented by the permutation matrix

$$P_{\sigma} = egin{bmatrix} 0 & 0 & 0 & 1 \ 1 & 0 & 0 & 0 \ 0 & 1 & 0 & 0 \ 0 & 1 & 0 & 0 \ 0 & 0 & 1 & 0 \ 0 & 0 & 1 & 0 & 0 \ 0 & 0 & 0 & 0 & 1 \ 0 & 0 & 0 & 0 & 0 \ 0 & 0 & 1 & 0 & 0 \ 0 & 0 & 0 & 1 & 0 \ 0 & 0 & 0 & 1 \ \end{bmatrix}$$

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

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

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

	α_{1}	α_2	α_3	$lpha_{4}$	β_1	β_2	β_3	eta_{4}	ε_3]
	$lpha_4$	α_1	α_2	α_3	β_4	β_1	β_2	β_3	<i>E</i> 3	
	α_3	α_4	α_1	α_2	β_3	β_4	β_1	β_2	ε3	
	α_2	α_3	α_4	α_1	β_2	β_3	β_4	β_1	ε_3	
W =	γ_1	γ_2	γ_3	γ_4	δ_1	δ_2	δ_3	δ_4	ε_4	
	γ_4	γ_1	γ_2	γ_3	δ_4	δ_1	δ_2	δ_3	ε_4	
	γ_3	γ_4	γ_1	γ_2	δ_3	δ_4	δ_1	δ_2	ε ₄	
	γ_2	γ_3	γ_4	γ_1	δ_2	δ_3	δ_4	δ_1	ε4	
	_ ε ₁	ε_1	ε_1	ε_1	ε2	ε_2	ε_2	ε_2	ε_5	

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

	α_{1}	α_2	α_3	$lpha_{4}$	β_1	β_2	β_3	eta_{4}	ε_3	1
	α_4	α_1	α_2	α_3	β_4	β_1	β_2	β_3	ε_3	
	α_3	α_4	α_1	α_2	β_3	β_4	β_1	β_2	ε_3	
	α_2	α_3	α_4	α_1	β_2	β_3	β_4	β_1	ε_3	
W =	γ_1	γ_2	γ_3	γ_4	δ_1	δ_2	δ_3	δ_4	ε_4	
	γ_4	γ_1	γ_2	γ_3	δ_4	δ_1	δ_2	δ_3	ε_4	
	γ_3	γ_4	γ_1	γ_2	δ_3	δ_4	δ_1	δ_2	ε_4	
	γ_2	γ_3	γ_4	γ_1	δ_2	δ_3	δ_4	δ_1	ε4	
	ε_1	ε_1	ε_1	ε_1	ε2	ε_2	ε_2	ε_2	ε_5	

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

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

	α_{1}	α_2	α_3	$lpha_4$	β_1	β_2	β_3	eta_{4}	ε3]
	α_4	α_1	α_2	α_3	β_4	β_1	β_2	β_3	ε3	
	α_3	α_4	α_1	α_2	β_3	β_4	β_1	β_2	ε3	
	α_2	α_3	α_4	α_1	β_2	β_3	β_4	β_1	ε_3	
$\mathcal{N} =$	γ_1	γ_2	γ_3	γ_4	δ_1	δ_2	δ_3	δ_4	ε4	
	γ_4	γ_1	γ_2	γ_3	δ_4	δ_1	δ_2	δ_3	ε4	
	γ_3	γ_4	γ_1	γ_2	δ_3	δ_4	δ_1	δ_2	ε4	
	γ_2	γ_3	γ_4	γ_1	δ_2	δ_3	δ_4	δ_1	ε4	
	ε_1	ε_1	ε_1	ε_1	ε ₂	ε_2	ε_2	ε_2	ε_5	

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

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



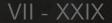
There is no neural network whose function space is $\mathcal{E}^{\sigma} \cap \mathcal{M}_2$!





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 $\mathcal{E}^{\sigma} \cap \mathcal{M}_2$.



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

	α_1	α_1	α_1	α_1	β_1	β_1	β_1	β_1	γ_1
	α_2	α_2	α_2	α_2	β_2	β_2	β_2	β_2	γ_2
	α_3	α_3	α_3	α_3	β_3	β_3	β_3	β_3	γ_3
	$lpha_4$	$lpha_4$	α_4	$lpha_4$	β_4	β_4	eta_4	β_4	γ_4
W =	α_5	α_5	α_5	α_{5}	β_5	β_5	β_5	β_5	γ_5
	α_6	α_6	α_6	α_6	β_6	β_6	β_6	β_6	γ_6
1	α_7	α_7	α_7	α_7	β_7	β_7	β_7	β_7	γ_7
	α_8	α_8	α_8	α_8	β_8	β_8	β_8	β_8	γ_8
1	α 9	lpha9	lpha9	lpha9	β_9	eta_{9}	eta_{9}	eta_{9}	γ_9

VIII - XXIX

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

	α_1	α_1	α_1	α_1	β_1	β_1	β_1	β_1	γ_1 -	
	α_2	α_2	α_2	α_2	β_2	β_2	β_2	β_2	γ_2	
	α_3	α_3	α_3	α_3	β_3	β_3	β_3	β_3	γ_3	
	α_4	α_4	$lpha_4$	$lpha_4$	β_4	β_4	β_4	eta_{4}	γ_4	
W =	α_5	α_5	α_5	α_5	β_5	β_5	β_5	β_5	γ_5	
	α_6	α_6	α_6	α_6	β_6	β_6	β_6	β_6	γ_6	
1	α_7	α_7	α_7	α_7	β_7	β_7	β_7	β_7	γ_7	
	α_8	α_8	$lpha_{8}$	α_8	β_8	β_8	β_8	β_8	γ_8	
	α_9	lpha9	lpha9	lpha9	β_9	eta_{9}	eta_{9}	eta_{9}	γ_9	

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

VIII - XXIX

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

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.

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}^{σ} 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, ..., k.

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}^{σ} 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, ..., 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.

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}^{σ} 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, ..., 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 $G \subset S_n$.

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

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}^{σ} 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, ..., 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 $G \subset S_n$.

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

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

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}^{σ} 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, ..., 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 $G \subset S_n$.

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

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

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

IX - XXIX

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}^{σ} 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, ..., 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.

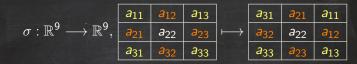
Lemma: Let $G \subset S_n$.

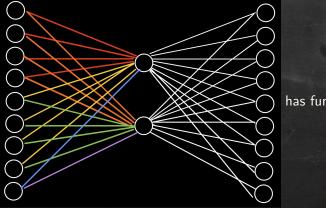
The set of *G*-invariant $W \in \mathbb{R}^{m \times n}$ is \mathcal{I}^{σ} for some $\sigma \in \mathcal{S}_n$.

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

Lemma: $\{(A, B) \in \mathbb{R}^{m \times k} \times \mathbb{R}^{k \times n} | \operatorname{rank}(AB) = k, AB \in \mathcal{I}^{\sigma}\} = \{A \in \mathbb{R}^{m \times k} | \operatorname{rank}(A) = k\} \times \{B \in \mathbb{R}^{k \times n} | \text{ columns indexed by } \pi_i \text{ are equal}\} \Rightarrow \sigma \text{ induces weight sharing on the encoder}!$

IX - XXIX





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

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 $\underline{P_{\sigma} \in \mathbb{R}^{n \times n}}.$

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 $P_{\sigma} \in \mathbb{R}^{n \times n}$.

Idea: Let $T \in GL_n$. W is P_{σ} -equivariant iff $T^{-1}WT$ is $T^{-1}P_{\sigma}T$ -equivariant.

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 $P_{\sigma} \in \mathbb{R}^{n \times n}$.

Idea: Let $T \in GL_n$. W is P_{σ} -equivariant iff $T^{-1}WT$ is $T^{-1}P_{\sigma}T$ -equivariant. This base change also preserves rank!



 $P = P_{\sigma}$

$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$	0	0
0	$\begin{array}{ccccc} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}$	0
	0010	
0	0	1

P-equivariant matrices

Γ	α_1	α_2	α_3	α_4	$ \beta_1 $	β_2	β_3	β_4	ε_3
	α_4	α_1	α_2	α_3	β_4	β_1	β_2	β_3	ε_3
	α_3	α_4	α_1	α_2	β_3	β_4	β_1	β_2	ε_3
	α_2	α_3	α_4	α_1	β_2	β_3	β_4	β_1	ε_3
	γ_1	γ_2	γ_3	γ_4			δ_3		ε_4
	γ_4	γ_1	γ_2	γ_3	1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -		δ_2		ε_4
	γ_3	γ_4	γ_1	γ_2	δ_3	δ_4	δ_1	δ_2	ε_4
_	γ_2	γ_3	γ_4	γ_1	δ_2	δ_3	δ_4	δ_1	ε_4
L	ε_1	ε_1	ε_1	ε_1	ε_2	ε_2	ε_2	ε_2	ε_5

XII - XXIX

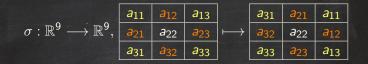


P = diagonalization of P_{σ}

P-equivariant matrices

1000				a ₁₁	0	0	0	a ₁₂	0	0	0	a ₁₃
0 i 0 0	0	0		0	<i>c</i> ₁₁	0	0	0	<i>c</i> ₁₂	0	0	0
$0 \ 0 \ -1 \ 0$	0			0	0	<i>b</i> ₁₁	0	0	0	<i>b</i> ₁₂	0	0
0 0 0 - <i>i</i>				0	0	0	<i>d</i> ₁₁	0	0	0	<i>d</i> ₁₂	0
	1 0 0 0			a ₂₁	0	0	0	a ₂₂	0	0	0	a ₂₃
0	0 i 0 0	0		0	<i>c</i> ₂₁	0	0	0	<i>C</i> ₂₂	0	- 0	0
U	$0 \ 0 \ -1 \ 0$			0	0	<i>b</i> ₂₁	0	0	0	b ₂₂	0	0
	0 0 0 <i>-i</i>		1	0	0	0	<i>d</i> ₂₁	0	0	0	d ₂₂	0
0	0	1		a ₃₁	0	0	0	a ₃₂	0	0	0	a33

XIII - XXIX



P =diagonalization of P_{σ}

-1

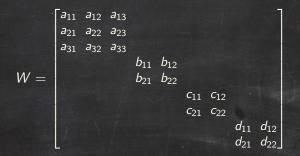
-i

1

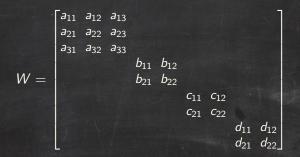
1

P-equivariant matrices

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

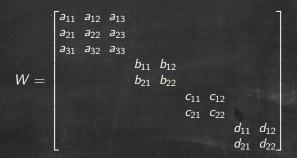


There are 10 ways how W can have rank 2:



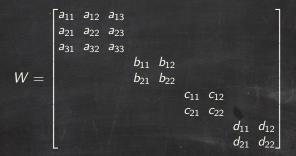
There are 10 ways how W can have rank 2:
One of the diagonal blocks has rank 2;

other blocks are 0



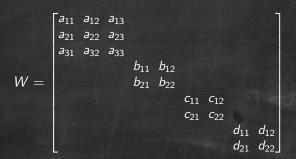
There are 10 ways how W can have rank 2:

 One of the diagonal blocks has rank 2; → 4 components of E^σ ∩ M₂ other blocks are 0



There are 10 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; → 4 components of E^σ ∩ M₂ other blocks are 0
- Two distinct blocks have rank 1; other blocks are 0



There are 10 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; → 4 components of E^σ ∩ M₂ other blocks are 0
- Two distinct blocks have rank 1; other blocks are 0

 \rightsquigarrow 6 components of $\mathcal{E}^{\sigma} \cap \mathcal{M}_2$

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 $P_{\sigma} \in \mathbb{R}^{n \times n}$.

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.

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!

az

74

 γ_1

 $\beta_2 \beta_3 \beta_4$

 $\delta_1 \ \delta_2 \ \delta_3 \ \delta_4$

δ4 δ1 δ2 δ3

 $\delta_3 \ \delta_4 \ \delta_1 \ \delta_2$

82 83 84 81

87 87 87 87

$P_{\sigma} =$	-0001 1000 0100 0010	0	0		α_3	$egin{array}{c} lpha_2 \ lpha_1 \ lpha_4 \ lpha_3 \end{array}$	a a
	0	0 0 0 1 1 0 0 0 0 1 0 0 0 0 1 0	0	W =	$\gamma_4 \\ \gamma_3$		
A Part Part	0	0	1		ε_1	ε_1	ε

ε3 ε3

E3

ε4

εA

ε4

84

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!

	0 0 0 1 1 0 0 0 0 1 0 0	0	0		α_4	$lpha_2 \\ lpha_1 \\ lpha_4$	α_2	α_3		$egin{array}{c} eta_2 \ eta_1 \ eta_4 \ eta_4 \end{array}$			ε ₃ ε ₃ ε ₃
P _	0010	0001		14/		α3			$\frac{\beta_2}{\delta_1}$	$\frac{\beta_3}{\delta_2}$		$\frac{\beta_1}{\delta_4}$	ε <u>3</u> ε4
$P_{\sigma} =$	0	$ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} $ 0	0		$\gamma_1 \\ \gamma_4$			γ_4 γ_3	δ_4	δ_1	δ_2	δ_3	$\varepsilon_4 \\ \varepsilon_4$
					γ_3 γ_2		$\gamma_1 \\ \gamma_4$	γ_2 γ_1		$\delta_4 \\ \delta_3$		$\delta_2 \\ \delta_1$	ε ₄ ε ₄
	L 0	0	1		ε_1	ε_1	ε_1	ε_1	ε2	ε2	ε2	ε2	ε_5

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



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_ =	0 0 0 1 1 0 0 0 0 1 0 0 0 0 1 0	0	0		$\begin{array}{c} \alpha_4 \\ \alpha_3 \end{array}$	$egin{array}{c} lpha_2 \ lpha_1 \ lpha_4 \ lpha_3 \end{array}$	$\begin{array}{c} \alpha_2 \\ \alpha_1 \end{array}$	α_3 α_2	β_3	$\begin{array}{c} \beta_2\\ \beta_1\\ \beta_4\\ \beta_3 \end{array}$	β_1	β_2	ε ₃ - ε ₃ ε ₃ ε ₃
$P_{\sigma} =$	0	0 0 0 1 1 0 0 0 0 1 0 0 0 0 1 0	0	W =	γ_4	γ_2 γ_1 γ_4 γ_3		$\begin{array}{c} \gamma_4 \\ \gamma_3 \\ \gamma_2 \\ \gamma_1 \end{array}$	δ3	$\delta_2 \\ \delta_1 \\ \delta_4 \\ \delta_3$		$\delta_4 \\ \delta_3 \\ \delta_2 \\ \delta_1$	ε4 ε4 ε4 ε4
	L 0	0	1		ε_1	ε_1	ε_1	ε_1	ε2	ε_2	ε_2	ε_2	ε_5

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

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 $\ell_j := \text{length}(\pi_j)$.

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.

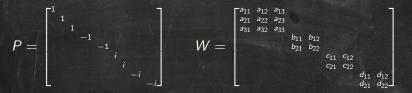
Consider $\mathcal{M}_r = \{ W \in \mathbb{R}^{n \times n} \mid \operatorname{rank}(W) \leq r \}$ and $\sigma \in \mathcal{S}_n$. Decompose $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles and let $\ell_j := \operatorname{length}(\pi_j)$. Diagonalize P_σ and sort the eigenvalues. This yields 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 $d_m := \#\{j \text{ such that } m|\ell_i\}$.

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 $\ell_j := \operatorname{length}(\pi_j)$. 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 $d_m := \#\{j \text{ such that } m|\ell_i\}$.



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

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 $\ell_j := \operatorname{length}(\pi_j)$. 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 $d_m := \#\{j \text{ such that } m|\ell_i\}$.



 $\ell_1 = 4, \ \ell_2 = 4, \ \ell_3 = 1$ $d_1 = 3, \ d_2 = 2, \ d_3 = 0, \ d_4 = 2, \ d_5 = 0, \ \dots$

XVII - XXIX

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 $\ell_j := \operatorname{length}(\pi_j)$. 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 $d_m := \#\{j \text{ such that } m|\ell_i\}$.



Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n \times n} \mid \operatorname{rank}(W) \leq r\}$ and $\sigma \in S_n$. Decompose $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles and let $\ell_j := \operatorname{length}(\pi_j)$. 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

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



Consider $\mathcal{M}_r = \{W \in \mathbb{R}^{n \times n} \mid \operatorname{rank}(W) \leq r\}$ and $\sigma \in S_n$. Decompose $\sigma = \pi_1 \circ \ldots \circ \pi_k$ into disjoint cycles and let $\ell_j := \operatorname{length}(\pi_j)$. 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

 $\sum_{m\in\mathbb{Z}_{>0}}\sum_{u\in(\mathbb{Z}/m\mathbb{Z})^{ imes}}r_{m,u}=r, \hspace{1em} ext{where } 0\leq r_{m,u}\leq d_m.$

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

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

VIII - X

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!

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

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 $\mathbb R$

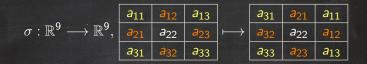
Consider $\mathcal{M}_r = \{ W \in \mathbb{R}^{n \times n} \mid \operatorname{rank}(W) \leq r \}$ and $\sigma \in \mathcal{S}_n$.

To diagonalize P_{σ} , we need a complex base change!

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

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
$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}$$
$$\sim \frac{1}{2} \begin{bmatrix} 1 & \sqrt{2} & 1 & 0 \\ 1 & 0 & -1 & -\sqrt{2} \\ 1 & -\sqrt{2} & 1 & 0 \\ 1 & 0 & -1 & \sqrt{2} \end{bmatrix} \in O_4(\mathbb{R})$$
$$\times |X - XX|X$$



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

P-equivariant matrices

$$\begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & -1 & & & \\ & & -1 & & & \\ & & & -1 & & \\ & & & -1 & & \\ & & & & -1 & \\ & & & & 0 & 1 \\ & & & & & 0 & 1 \\ & & & & & 0 & 1 \\ & & & & & -1 & 0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & & & \\ a_{21} & a_{22} & a_{23} & & & \\ a_{31} & a_{32} & a_{33} & & & \\ & & & b_{11} & b_{12} & & \\ & & & b_{21} & b_{22} & & \\ & & & & c_1 & -c_2 & d_1 & -d_2 \\ & & & & c_2 & c_1 & d_2 & d_1 \\ & & & & e_1 & -e_2 & f_1 & -f_2 \\ & & & & e_2 & e_1 & f_2 & f_1 \end{bmatrix}$$

	a ₁₁	a ₁₂	a ₁₃]
	a ₂₁	a ₂₂	a ₂₃						
	a ₃₁	a ₃₂	a33						
				<i>b</i> ₁₁	<i>b</i> ₁₂				
W =				b ₂₁	b ₂₂				
						<i>c</i> ₁	$-c_{2}$	d_1	$-d_2$
						<i>c</i> ₂	<i>c</i> ₁	<i>d</i> ₂	d_1
						e_1	$-e_2$	f_1	$-f_2$
						e ₂	e_1	f_2	f_1

There are 4 ways how W can have rank 2:

XXI - XXIX

	a ₁₁	a ₁₂	a ₁₃						1
	a ₂₁	a ₂₂	a ₂₃						
	a ₃₁	a ₃₂	a33						
				b_{11}	<i>b</i> ₁₂				
W =				b ₂₁	b ₂₂				
						<i>c</i> ₁	$-c_{2}$	d_1	$-d_2$
						<i>c</i> ₂	<i>c</i> ₁	<i>d</i> ₂	d_1
						e_1	$-e_{2}$	f_1	$-f_2$
						e ₂	e_1	f_2	f_1

There are 4 ways how W can have rank 2:
One of the diagonal blocks has rank 2; other blocks are 0

	a ₁₁	a ₁₂	a ₁₃						1	
	a ₂₁	a ₂₂	a ₂₃							
	a ₃₁	a ₃₂	a33							
				<i>b</i> ₁₁	<i>b</i> ₁₂					
W =				<i>b</i> ₂₁	b ₂₂					
						<i>c</i> ₁	$-c_{2}$	d_1	$-d_2$	
						<i>c</i> ₂	<i>c</i> ₁	<i>d</i> ₂	d_1	
						e_1	$-e_{2}$	f_1	$-f_2$	
						e ₂	e_1	f_2	f_1	

There are 4 ways how W can have rank 2:

 One of the diagonal blocks has rank 2; → 3 components of E^σ ∩ M₂ other blocks are 0

XXI - XXIX

	a ₁₁	a ₁₂	a ₁₃						1
	a ₂₁	a ₂₂	a ₂₃						
	a ₃₁	a ₃₂	a33						
				<i>b</i> ₁₁	<i>b</i> ₁₂				
W =				b ₂₁	b ₂₂				
						<i>c</i> ₁	$-c_{2}$	d_1	$-d_2$
						<i>c</i> ₂	<i>c</i> ₁	<i>d</i> ₂	d_1
						e_1	$-e_{2}$	f_1	$-f_2$
						e_2	e_1	f_2	f_1

There are 4 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; other blocks are 0
- Two first 2 blocks have rank 1; last block is 0

 \rightsquigarrow 3 components of $\mathcal{E}^{\sigma} \cap \mathcal{M}_2$

running example

	a ₁₁	a ₁₂	a ₁₃							
	a ₂₁	a ₂₂	a ₂₃							
	a ₃₁	a ₃₂	a33							
				b_{11}	<i>b</i> ₁₂					
W =				<i>b</i> ₂₁	b ₂₂					
						<i>c</i> ₁	$-c_{2}$	d_1	$-d_2$	
						<i>c</i> ₂	<i>c</i> ₁	<i>d</i> ₂	d_1	
						e_1	-e ₂	f_1	$-f_2$	
						en	e1	fz	f_1	

There are 4 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; other blocks are 0
- Two first 2 blocks have rank 1; last block is 0

 \rightsquigarrow 3 components 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 $2d \times 2d$ matrices consisting of d^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 2d × 2d matrices consisting of d² 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 2d × 2d matrices consisting of d² matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$.

Blocks of the latter kind have even rank!

Definition:

For $z = a + ib \in \mathbb{C}$, define $\mathcal{R}(z) := \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. For $M \in \mathbb{C}^{d \times e}$, let $\mathcal{R}(M) \in \mathbb{R}^{2d \times 2e}$ be obtained by replacing each entry m_{ij} of M by $\mathcal{R}(m_{ij})$. We call $\mathcal{R}(M)$ the realization of M.

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 $2d \times 2d$ matrices consisting of d^2 matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. $\mathcal{R}(\mathbb{C}^{d \times d})$

Blocks of the latter kind have even rank! $\operatorname{rank}(\mathcal{R}(M)) = 2 \cdot \operatorname{rank}(M)$

Definition:

For $z = a + ib \in \mathbb{C}$, define $\mathcal{R}(z) := \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. For $M \in \mathbb{C}^{d \times e}$, let $\mathcal{R}(M) \in \mathbb{R}^{2d \times 2e}$ be obtained by replacing each entry m_{ij} of M by $\mathcal{R}(m_{ij})$. We call $\mathcal{R}(M)$ the realization of M.

Equivariance over \mathbb{R}

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

$$r_{1,1} + r_{2,1} + \sum_{m>2} \sum_{\substack{u \in (\mathbb{Z}/m\mathbb{Z})^{ imes}, \\ rac{1}{2} < rac{u}{m} < 1}} 2 \cdot r_{m,u} = r, \quad ext{ where } 0 \leq r_{m,u} \leq d_m.$$

Equivariance over \mathbb{R}

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

$$r_{1,1} + r_{2,1} + \sum_{m>2} \sum_{\substack{u \in (\mathbb{Z}/m\mathbb{Z})^{\times}, \\ \frac{1}{2} < \frac{u}{m} < 1}} 2 \cdot r_{m,u} = r, \quad \text{where } 0 \le r_{m,u} \le d_m.$$

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

 $\cong \{A \in \mathbb{R}^{d_1 \times d_1} \mid \operatorname{rank}(A) \leq r_{1,1}\} \times \{A \in \mathbb{R}^{d_2 \times d_2} \mid \operatorname{rank}(A) \leq r_{2,1}\} \\ \times \prod_{m>2} \prod_{u \in (\mathbb{Z}/m\mathbb{Z})^{\times}, \\ \frac{1}{2} < \frac{u}{m} < 1} \mathcal{R}(\{A \in \mathbb{C}^{d_m \times d_m} \mid \operatorname{rank}(A) \leq r_{m,u}\}).$

Which of these 4 components is best ??

	a ₁₁	a ₁₂	a ₁₃							
	a ₂₁	a ₂₂	a ₂₃							
	a ₃₁	a ₃₂	a33							
				<i>b</i> ₁₁	<i>b</i> ₁₂					
W =				b ₂₁	b ₂₂					
						<i>c</i> ₁	$-c_{2}$	d_1	$-d_2$	
						<i>c</i> ₂	<i>c</i> ₁	<i>d</i> ₂	d_1	
						e_1	$-e_{2}$	f_1	$-f_{2}$	
						e ₂	e_1	f_2	f_1	

There are 4 ways how W can have rank 2:

- One of the diagonal blocks has rank 2; other blocks are 0
- Two first 2 blocks have rank 1; last block is 0

 \rightsquigarrow 3 components of $\mathcal{E}^{\sigma} \cap \mathcal{M}_2$

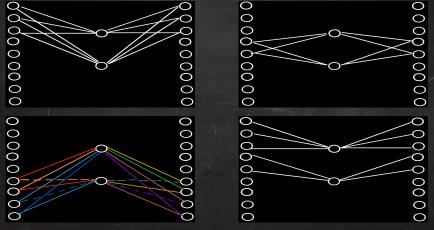
 $\rightsquigarrow 1 \text{ component of } \mathcal{E}^{\sigma} \cap \mathcal{M}_2$

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

Parametrizing equivariant functions with autoencoders There is no neural network whose function space is $\mathcal{E}^{\sigma} \cap \mathcal{M}_2$! But: Each irreducible component of $\mathcal{E}^{\sigma} \cap \mathcal{M}_2$ is the function space of an autoencoder.

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

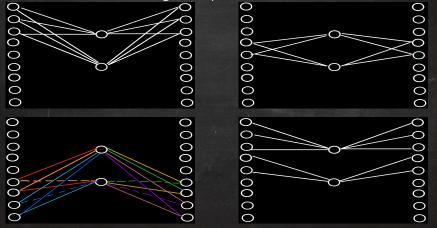
But: Each irreducible component of $\mathcal{E}^{\sigma} \cap \mathcal{M}_2$ is the function space of an autoencoder. In the block-diagonal representation:



XXV - XXIX

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

But: Each irreducible component of $\mathcal{E}^{\sigma} \cap \mathcal{M}_2$ is the function space of an autoencoder. In the block-diagonal representation:



This works in general for $\mathcal{E}^{\sigma} \cap \mathcal{M}_r \subset \mathbb{R}^{n \times n}$

Consider a function space $\mathcal{M} \subset \mathbb{R}^{m \times n}$. Given training data $X \in \mathbb{R}^{n \times d}$ and $Y \in \mathbb{R}^{m \times d}$, the squared-error loss is

 $\mathcal{M} \to \mathbb{R}, \quad \mathcal{W} \mapsto \|\mathcal{W} X - Y\|_F^2.$

Consider a function space $\mathcal{M} \subset \mathbb{R}^{m \times n}$. Given training data $X \in \mathbb{R}^{n \times d}$ and $Y \in \mathbb{R}^{m \times d}$, the squared-error loss is

$$\mathcal{M} \to \mathbb{R}, \quad \mathcal{W} \mapsto \|\mathcal{W}X - Y\|_F^2.$$

Lemma: If $rank(XX^{\top}) = n$ (which holds for a sufficient amount of training data that is sufficiently generic), minimizing the squared-error loss is equivalent to minimizing the weighted Euclidean distance

$$\min_{W \in \mathcal{M}} \|W - U\|_{XX^{\top}}^2, \quad \text{where } U = YX^{\top}(XX^{\top})^{-1}.$$

Consider a function space $\mathcal{M} \subset \mathbb{R}^{m \times n}$. Given training data $X \in \mathbb{R}^{n \times d}$ and $Y \in \mathbb{R}^{m \times d}$, the squared-error loss is

$$\mathcal{M} \to \mathbb{R}, \quad \mathcal{W} \mapsto \|\mathcal{W} X - Y\|_F^2.$$

Lemma: If $rank(XX^{\top}) = n$ (which holds for a sufficient amount of training data that is sufficiently generic), minimizing the squared-error loss is equivalent to minimizing the weighted Euclidean distance

$$\min_{W \in \mathcal{M}} \|W - U\|_{XX^{\top}}^2, \quad \text{where } U = YX^{\top}(XX^{\top})^{-1}.$$

recall: $\|A\|_{XX^{ op}}^2 = \|A(XX^{ op})^{1/2}\|_F^2$.

Consider a function space $\mathcal{M} \subset \mathbb{R}^{m \times n}$. Given training data $X \in \mathbb{R}^{n \times d}$ and $Y \in \mathbb{R}^{m \times d}$, the squared-error loss is

$$\mathcal{M} \to \mathbb{R}, \quad \mathcal{W} \mapsto \|\mathcal{W}X - Y\|_F^2.$$

Lemma: If $rank(XX^{\top}) = n$ (which holds for a sufficient amount of training data that is sufficiently generic), minimizing the squared-error loss is equivalent to minimizing the weighted Euclidean distance

$$\min_{W \in \mathcal{M}} \|W - U\|_{XX^{\top}}^{2}, \quad \text{where } U = YX^{\top}(XX^{\top})^{-1}.$$

recall: $||A||^2_{XX^{\top}} = ||A(XX^{\top})^{1/2}||^2_F$.

Ί_ Χ

Now let's assume that XX^{\top} is (close to) a multiple of the identity, and \mathcal{M} is an irreducible component of $\mathcal{E}^{\sigma} \cap \mathcal{M}_{r}$.

Orthogonal base changes do not affect the standard Euclidean distance! Hence, our task is

$$\min_{\tilde{V}\in\tilde{\mathcal{M}}}\|\tilde{W}-\tilde{U}\|_{F}^{2},\tag{1}$$

where $\tilde{\mathcal{M}}$ and \tilde{U} are obtained from \mathcal{M} and U by applying our orthogonal base change.

Orthogonal base changes do not affect the standard Euclidean distance! Hence, our task is

$$\min_{\tilde{\mathcal{N}}\in\tilde{\mathcal{M}}}\|\tilde{\mathcal{W}}-\tilde{\mathcal{U}}\|_{F}^{2},\tag{1}$$

where $\tilde{\mathcal{M}}$ and \tilde{U} are obtained from \mathcal{M} and U by applying our orthogonal base change.

The matrices in $\tilde{\mathcal{M}}$ are block diagonal. By orthogonally projecting \tilde{U} , we can assume that \tilde{U} has the same block-diagonal structure.

Orthogonal base changes do not affect the standard Euclidean distance! Hence, our task is

$$\min_{\tilde{\mathcal{N}}\in\tilde{\mathcal{M}}}\|\tilde{\mathcal{W}}-\tilde{\mathcal{U}}\|_{F}^{2},\tag{1}$$

where $\tilde{\mathcal{M}}$ and \tilde{U} are obtained from \mathcal{M} and U by applying our orthogonal base change.

The matrices in $\tilde{\mathcal{M}}$ are block diagonal. By orthogonally projecting \tilde{U} , we can assume that \tilde{U} has the same block-diagonal structure. \Rightarrow (1) can be solved on each block independently!

Orthogonal base changes do not affect the standard Euclidean distance! Hence, our task is

$$\min_{\tilde{W}\in\tilde{\mathcal{M}}}\|\tilde{W}-\tilde{U}\|_{F}^{2},\tag{1}$$

where $\tilde{\mathcal{M}}$ and \tilde{U} are obtained from \mathcal{M} and U by applying our orthogonal base change.

The matrices in $\tilde{\mathcal{M}}$ are block diagonal. By orthogonally projecting \tilde{U} , we can assume that \tilde{U} has the same block-diagonal structure. \Rightarrow (1) can be solved on each block independently!

Each block is of the form $\{A \in \mathbb{R}^{d \times d} \mid \operatorname{rank}(A) \leq s\}$ or $\mathcal{R}(\{A \in \mathbb{C}^{d \times d} \mid \operatorname{rank}(A) \leq s\})$

Orthogonal base changes do not affect the standard Euclidean distance! Hence, our task is

$$\min_{\tilde{W}\in\tilde{\mathcal{M}}}\|\tilde{W}-\tilde{U}\|_{F}^{2},\tag{1}$$

/// _ `

where $\tilde{\mathcal{M}}$ and $\tilde{\mathcal{U}}$ are obtained from \mathcal{M} and \mathcal{U} by applying our orthogonal base change.

The matrices in $\tilde{\mathcal{M}}$ are block diagonal. By orthogonally projecting \tilde{U} , we can assume that \tilde{U} has the same block-diagonal structure. \Rightarrow (1) can be solved on each block independently!

Each block is of the form $\{A \in \mathbb{R}^{d \times d} \mid \operatorname{rank}(A) \leq s\}$ or $\mathcal{R}(\{A \in \mathbb{C}^{d \times d} \mid \operatorname{rank}(A) \leq s\}) \cong \{A \in \mathbb{R}^{2d \times d} \mid \operatorname{rank}(A) \leq s\}$ (by deleting every other column).

Orthogonal base changes do not affect the standard Euclidean distance! Hence, our task is

$$\min_{\tilde{W}\in\tilde{\mathcal{M}}}\|\tilde{W}-\tilde{U}\|_{F}^{2},\tag{1}$$

where $\tilde{\mathcal{M}}$ and $\tilde{\mathcal{U}}$ are obtained from \mathcal{M} and \mathcal{U} by applying our orthogonal base change.

The matrices in $\tilde{\mathcal{M}}$ are block diagonal. By orthogonally projecting \tilde{U} , we can assume that \tilde{U} has the same block-diagonal structure. \Rightarrow (1) can be solved on each block independently!

Each block is of the form $\{A \in \mathbb{R}^{d \times d} \mid \operatorname{rank}(A) \leq s\}$ or $\mathcal{R}(\{A \in \mathbb{C}^{d \times d} \mid \operatorname{rank}(A) \leq s\}) \cong \{A \in \mathbb{R}^{2d \times d} \mid \operatorname{rank}(A) \leq s\}$ (by deleting every other column).

Euclidean distance minimization on these blocks typically has a unique local minimum, easily found by SVD (Eckart-Young theorem)!

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

> differential geometry \Rightarrow algebraic geometry \Rightarrow data science \Rightarrow

Rernd Sturmfels

Kathlén Kohn

Paul Breidina

Metric Algebraic Geometry

Historical Snapshot Polars Foci Envoloper

Critical Equations Euclidean Distance Dearee Low-Rank Matrix Approximation Invitation to Polar Degrees

Computations Gröhner Bases Parameter Continuation Theorem Polynomial Homotopy Continuation

Polar Degrees Polar Varieties **Projective Duality**

Chern Classes

Wasserstein Distance Polyhedral Norms Optimal Transport &

Independence Models Wasserstein meets Segre-Veronese

Oberwolfach Seminars

Curvature Plane Curves

Algebraic Varieties Volumes of Tubular Neighborhoods

Reach and Offset Medial Axis and Rottlenecks Offset Hypersurfaces Offset Discriminant

Voronoi Cells Voronoi Basics Algebraic Boundaries Degree Formulas Voronoi meets Eckart-Young

Condition Numbers Errors in Numerical Computations Matrix Inversion and Eckart-Young Condition Number Theorems Distance to the Discriminant

Machine Learning Neural Networks Convolutional Networks Learning Varieties

ISBN 978-3-03

Maximum Likelihood Kullback-Leibler Divergence

Maximum Likelihood Degree Scattering Equations Gaussian Models

Tensors

Tensors and their Rank **Eigenvectors and Singular Vectors** Volumes of Rank-One Varieties

Computer Vision Multiview Varieties

Grassmann Tensors 3D Reconstruction from Unknown Cameras

Volumes of Semialaebraic Sets Calculus and Beyond

D-Modules SDP Hierarchies

Sampling Homology from Finite Samples

Sampling with Density Gugrantees Markov Chains on Varieties Chow goes to Monte Carlo

Open PhD and Postdoc Positions!

- PhD position in Algebraic Geometry & Computer Vision https://kathlenkohn.github.io/phd
- PhD position in Geometric Combinatorics with Katharina Jochemko
- Postdoc position in Algebraic Geometry applied to Machine Learning & Computer Vision https://kathlenkohn.github.io/postdoc
- Researcher position in Graphical Models and Algebraic Statistics with Liam Solus

X _ X