Geometry of Linear Neural Networks: Equivariance and Invariance under Permutation Groups

Kathlén Kohn, *Anna-Laura Sattelberger*, Vahid Shahverdi
arxiv:2309.13736 [cs.LG]

Applied CATS seminar

October 17, 2023
KTH Royal Institute of Technology

Motivation

Two questions

1. How do the network's properties affect the geometry of its **function space**?
   How to characterize **equivariance** or **invariance**?
2. How to **parameterize** equivariant and invariant networks?
   Which implications does if have for **network design**?

Training neural networks

Neural networks

A neural network F of depth L is a parameterized family of functions $(f_{L,\theta}, \ldots, f_{1,\theta})$

$$F\colon \mathbb{R}^N \longrightarrow \mathcal{F}, \quad F(\theta) = f_{L,\theta} \circ \cdots \circ f_{1,\theta} =: f_\theta .$$

Each layer $f_{k,\theta}\colon \mathbb{R}^{d_{k-1}} \to \mathbb{R}^{d_k}$ is a composition activation $\circ$ affine-linear.

Training a network

Given training data $\mathcal{D} = \{(\widehat{x}_i, \widehat{y}_i)_{i=1,\ldots,S}\} \subset \mathbb{R}^{d_0} \times \mathbb{R}^{d_L}$, the aim is to minimize the loss

$$\mathcal{L}\colon \mathbb{R}^N \xrightarrow{F} \mathcal{F} \xrightarrow{\ell_\mathcal{D}} \mathbb{R} .$$

Example: For $\ell_\mathcal{D}$ the squared error loss, this gives $\min_{\theta \in \mathbb{R}^N} \sum_{i=1}^{S} (f_\theta(\widehat{x}_i) - \widehat{y}_i)^2$.
On **function space**: $\min_{M \in \mathcal{F}} \|M \cdot \widehat{X} - \widehat{Y}\|^2_{\text{Frob}}$.

Critical points of $\mathcal{L}$

- ◇ **pure**: critical point of $\ell_\mathcal{D}$
- ◇ **spurious**: induced by parameterization

Linear convolutional networks (LCNs)

- ◇ linear: identity as activation function
- ◇ convolutional layers with filter $w \in \mathbb{R}^k$ and stride $s \in \mathbb{N}$:

$$\alpha_{w,s}\colon \mathbb{R}^d \to \mathbb{R}^{d'}, \quad (\alpha_{w,s}(x))_i = \sum_{j=0}^{k-1} w_j x_{is+j}\,.$$

Geometry of linear convolutional networks [1]

Function space $\mathcal{F}_{(\mathbf{d},\mathbf{s})}$ of LCN: semi-algebraic set, Euclidean-closed

Theorem [2]

Let $(\mathbf{d},\mathbf{s})$ be an LCN architecture with all strides > 1 and $N \geq 1 + \sum_i d_i s_i$. For almost all data $\mathcal{D} \in (\mathbb{R}^{d_0} \times \mathbb{R}^{d_L})^N$, every critical point θ_c of $\mathcal{L}$ satisfies one of the following:

1. $F(\theta_c) = 0$, or
2. θ_c is a regular point of F and $F(\theta_c)$ is a **smooth, interior point** of $\mathcal{F}_{(\mathbf{d},\mathbf{s})}$.
   In particular, $F(\theta_c)$ is a critical point of $\ell_{\mathcal{D}}|_{\mathrm{Reg}(\mathcal{F}^{\circ}_{(\mathbf{d},\mathbf{s})})}$.

This is known to be false for...

- ◇ linear fully-connected networks
- ◇ stride-one LCNs

[1] K. Kohn, T. Merkh, G. Montúfar, M. Trager. Geometry of Linear Convolutional Networks. *SIAM J. Appl. Algebra Geom.*, 6(3):368–406, 2022.
[2] K. Kohn, G. Montúfar, V. Shahverdi, M. Trager. Function Space and Critical Points of Linear Convolutional Networks. Preprint arXiv:2304.0572, 2023.

Algebraic geometry for machine learning

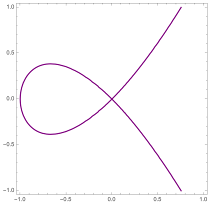
Natural points of entry

◇ algebraic vision [3] ◇ geometry of function spaces

Algebraic varieties

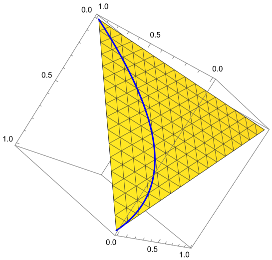
subsets of $\mathbb{C}^n$ obtained as common **zero set of polynomials** $p_1, \ldots, p_N \in \mathbb{C}[x_1, \ldots, x_n]$

Drawing real points of algebraic varieties



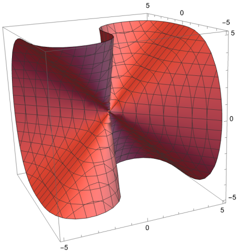
$\mathcal{V}(y^2 - x^2(x+1))$

a nodal curve



$\mathcal{V}(p_0 p_2 - (p_0 + p_1)p_1) \subset \Delta_2$

a discrete statistical model

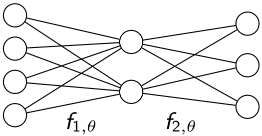


$\mathcal{V}(x^2 y - y^3 - z^3)$

a cubic surface

[3] J. Kileel and K. Kohn. Snapshot of Algebraic Vision. Preprint arXiv:2210.11443, 2022.

Fully connected linear neural networks



Example

$F: \mathbb{R}^{2\times 4} \times \mathbb{R}^{3\times 2} \longrightarrow \mathbb{R}^{3\times 4}, \quad (M_1, M_2) \mapsto M_2 \cdot M_1$

parameter space: $\mathbb{R}^N = \mathbb{R}^{2\times 4} \times \mathbb{R}^{3\times 2}, \; f_{1,\theta} = M_1, \; f_{2,\theta} = M_2$

Its function space $\mathcal{F}$ is the set of real points of the determinantal variety

$$\mathcal{M}_{2,3\times 4}(\mathbb{R}) = \left\{ M \in \mathbb{R}^{3\times 4} \mid \operatorname{rank}(M) \leq 2 \right\}.$$

The determinantal variety $\mathcal{M}_{r,m\times n}$

For $M = (m_{ij})_{i,j} \in \mathbb{C}^{m\times n}$: $\operatorname{rank}(M) \leq r \;\Leftrightarrow\;$ all $(r+1)\times(r+1)$ minors of M vanish.
Define

polynomials in m_{ij}

$$\mathcal{M}_{r,m\times n} = \{ M \mid \operatorname{rank}(M) \leq r \} \subset \mathbb{C}^{m\times n}.$$

Well studied! $\dim(\mathcal{M}_{r,m\times n})$, $\deg(\mathcal{M}_{r,m\times n})$, $\mathcal{M}_{r,m\times n}(\mathbb{R})$, $\mathcal{M}^{\text{reg}}_{r,m\times n}$, ...

Invariant functions

$f_\theta: \ \mathbb{R}^n \to \mathbb{R}^r \to \mathbb{R}^m$ — $r < \min(m, n)$

$G = \langle \sigma_1, \dots, \sigma_g \rangle \leq \mathcal{S}_n$ — a permutation group, acting on $\mathbb{R}^n$ by permuting the entries
induced action on M: permuting its columns

Invariance under $\sigma \in \mathcal{S}_n$: $f_\theta \circ \sigma \ \equiv \ f_\theta$

Decomposing into cycles

The decomposition $\sigma = \pi_1 \circ \cdots \circ \pi_k$ of σ into k disjoint cycles induces a partition $\boxed{\mathcal{P}(\sigma) = \{A_1, \dots, A_k\}}$ of the set $[n] = \{1, \dots, n\}$. $A_1, \dots, A_k \subset [n]$ pairwise disjoint sets

Example: The permutation $\sigma = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 4 & 1 & 2 \end{smallmatrix}\right) = (1\,3\,4)(2\,5) \in \mathcal{S}_5$ induces the partition $\mathcal{P}(\sigma) = \{\{1,3,4\},\{2,5\}\}$ of $[5] = \{1,2,3,4,5\}$. For $\eta = (1\,4\,3)(2\,5) \neq \sigma$: $\mathcal{P}(\eta) = \mathcal{P}(\sigma)$.

Characterizing invariance $\quad MP_\sigma \overset{!}{=} M$

Let $\sigma \in \mathcal{S}_n$ and $\mathcal{P}(\sigma) = \{A_1, \dots, A_k\}$ its induced partition. A matrix $M = (m_1 | \cdots | m_n)$ is invariant under $\sigma = \pi_1 \circ \cdots \circ \pi_k$ if and only if for each i, the columns $\{m_j\}_{j \in A_i}$ **coincide**.

$\Rightarrow$ If M is invariant under σ, its rank is at most k.

Example: rotation-invariance for $p \times p$ pictures

Setup: $n = p^2$ an even square number, $f_\theta : \mathbb{R}^n \to \mathbb{R}^n$ linear

$\sigma \in \mathcal{S}_n$: rotating an $p \times p$ picture clockwise by 90 degrees:

$$\sigma : \mathbb{R}^{p\times p} \to \mathbb{R}^{p\times p}, \quad \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{pmatrix} \mapsto \begin{pmatrix} a_{p1} & a_{p-1,1} & \dots & a_{11} \\ a_{p2} & a_{p-1,2} & \dots & a_{12} \\ \vdots & \vdots & \ddots & \vdots \\ a_{pp} & a_{p,p-1} & \dots & a_{1p} \end{pmatrix}$$

Identify $\mathbb{R}^{p\times p} \cong \mathbb{R}^n$ via $A \mapsto (a_{1,1}, a_{1,p}, a_{p,p}, a_{p,1}, a_{1,2}, a_{2,p}, a_{p,p-1}, a_{p-1,1}, \dots, a_{1,p-1}, a_{p-1,p}, a_{p,2}, a_{2,1}, a_{2,2}, a_{2,p-1}, a_{p-1,p-1}, a_{p-1,2}, \dots, a_{\frac{p}{2},\frac{p}{2}}, a_{\frac{p}{2},\frac{p}{2}+1}, a_{\frac{p}{2}+1,\frac{p}{2}}, a_{\frac{p}{2}+1,\frac{p}{2}+1})^\top$.

Under this identification, σ acts on $\mathbb{R}^n$ by the $n \times n$ block matrix

$$\begin{pmatrix} 0 & 0 & 0 & 1 & & & & & \\ 1 & 0 & 0 & 0 & & & & & \\ 0 & 1 & 0 & 0 & & & & & \\ 0 & 0 & 1 & 0 & & & & & \\ & & & & \dots & & & & \\ & & & & & 0 & 0 & 0 & 1 \\ & & & & & 1 & 0 & 0 & 0 \\ & & & & & 0 & 1 & 0 & 0 \\ & & & & & 0 & 0 & 1 & 0 \end{pmatrix}.$$

N.B.: σ-invariance of f_θ implies that columns 1–4, 5–8, ..., $(n-3)$–n of M **coincide**.

Properties of $\mathcal{I}^G_{r,m\times n} \subset \mathcal{M}_{r,m\times n}$

$G = \langle \sigma_1, \ldots, \sigma_g \rangle \leq \mathcal{S}_n$	a permutation group
$\sigma_i = \pi_{i,1} \circ \cdots \circ \pi_{i,k_i},\ i = 1, \ldots, g$	decomposition into pairwise disjoint cycles π_i

Reduction to cyclic case

There exists $\sigma \in \mathcal{S}_n$ such that $\boxed{\mathcal{I}^G_{r,m\times n} = \mathcal{I}^\sigma_{r,m\times n}}$. Any σ for which $\mathcal{P}(\sigma)$ is the **finest common coarsening** of $\mathcal{P}(\sigma_1), \ldots, \mathcal{P}(\sigma_g)$ does the job!

Proposition

Let $G = \langle \sigma \rangle \leq \mathcal{S}_n$ be cyclic, and $\sigma = \pi_1 \circ \cdots \circ \pi_k$ its decomposition into pairwise disjoint cycles π_i. The variety $\mathcal{I}^\sigma_{r,m\times n}$ is isomorphic to the determinantal variety $\mathcal{M}_{\min(r,k),m\times k}$ via a linear isomorphism $\psi_{\mathcal{P}(\sigma)} \colon \mathcal{I}^\sigma_{r,m\times n} \to \mathcal{M}_{\min(r,k),m\times k}$. deleting repeated columns

Via that, we can determine $\dim(\mathcal{I}^\sigma_{r,m\times n})$, $\deg(\mathcal{I}^\sigma_{r,m\times n})$, and $\mathrm{Sing}(\mathcal{I}^\sigma_{r,m\times n})$.

Example ($m = 2,\ n = 5,\ r = 1$)

Let $\sigma = (1\,3\,4)(2\,5) \in \mathcal{S}_5$ and hence $k = 2$. Any invariant matrix $M \in \mathcal{M}_{2\times 5}(\mathbb{R})$ is of the form $\left(\begin{smallmatrix} a & c & a & a & c \\ b & d & b & b & d \end{smallmatrix}\right)$ for some $a, b, c, d \in \mathbb{R}$. The rank constraint $r = 1$ imposes that $(c, d) = \lambda \cdot (a, b)^\top$ for some $\lambda \in \mathbb{R}$, where we assume that $(a, b) \neq (0, 0)$. Then

$$\boxed{\psi_{\mathcal{P}(\sigma)} \colon \begin{pmatrix} a & \lambda a & a & a & \lambda a \\ b & \lambda b & b & b & \lambda b \end{pmatrix} \mapsto \begin{pmatrix} a & \lambda a \\ b & \lambda b \end{pmatrix}.}$$

Parameterizing invariance and network design

$\mathcal{S}_n \ni \sigma = \pi_1 \circ \cdots \circ \pi_k, \; \mathcal{P}(\sigma) = \{A_1, \ldots, A_k\}$

Invariance of $M \in \mathcal{M}_{m\times n}$: forces columns $\{m_j\}_{j\in A_i}$ to coincide. For each i, remember representative m_{A_i} and denote $\boxed{M_1 := (m_{A_1} \mid \cdots \mid m_{A_k}) \in \mathcal{M}_{m\times k}\,.}$

Parameterization

Any σ-invariant $M \in \mathcal{M}_{m\times n}$ of rank k factorizes as $M = M_1 \cdot (e_{i_1}|\cdots|e_{i_n})$, $e_{i_j} \in \mathbb{R}^k$.
i-th standard unit vector in column j for all $j \in A_i$

Fibers of multiplication map

Let $r \leq \min(m,n)$. Denote by $\mu\colon \mathcal{M}_{m\times r} \times \mathcal{M}_{r\times n}, \; (A,B) \mapsto A\cdot B$. If $\operatorname{rank}(M) = r$ and $M = \mu(A,B)$ for some A, B, then the fiber of μ over M is

$$\mu^{-1}(M) = \left\{ \left(AT^{-1}, TB\right) \mid T \in \mathrm{GL}_n(\mathbb{C}) \right\} \subset \mathcal{M}_{m\times r} \times \mathcal{M}_{r\times n}\,.$$

Factorizations of invariant maps

Let M be invariant under σ and of rank k. **Any** factorization $M = A\cdot B$ is of the form

$$\boxed{(A,B) \in \left\{ \left(M_1 \cdot T^{-1}, T\cdot (e_{i_1}|\cdots|e_{i_n})\right) \mid T \in \mathrm{GL}_n \right\}\,.}$$

This parameterization imposes a **weight sharing property** on the encoder factor!

Example

Let $m = n = 5$, $r = 2$ and $\sigma = (1\,3\,4)(2\,5) \in \mathcal{S}_5$. If a matrix $M = AB \in \mathcal{I}^{\sigma}_{2,5\times 5}$ is invariant under σ, the encoder factor B has to fulfill the following weight sharing property:

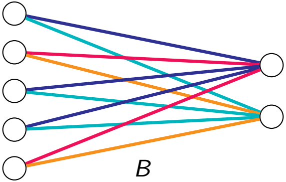


Figure: The σ-weight sharing property imposed on the encoder.

Learning invariant functions

Let $\sigma = \pi_1 \circ \cdots \circ \pi_k \in \mathcal{S}_n$ and let $r \leq k$. Consider the linear autoencoder $\mathbb{R}^n \to \mathbb{R}^r \to \mathbb{R}^n$ with fully-connected dense decoder $\mathbb{R}^r \to \mathbb{R}^n$ and encoder $\mathbb{R}^n \to \mathbb{R}^r$ with σ-weight sharing. Its function space is $\mathcal{I}^{\sigma}_{r,n\times n}(\mathbb{R})$.

Induced filtration of $\mathcal{M}_{r,m\times n}$

Observation

If $M \in \mathcal{M}_{m\times n}$ is invariant under $\sigma \in \mathcal{S}_n$, then it is also invariant under every permutation $\eta \in \mathcal{S}_n$ whose associated partition $\mathcal{P}(\eta)$ of $[n]$ is a refinement $\mathcal{P}(\eta) \prec \mathcal{P}(\sigma)$ of $\mathcal{P}(\sigma)$.

This induces the filtration $\mathcal{I}^\bullet_{m\times n}$ of $\mathcal{M}_{m\times n}$. (and $\mathcal{I}^\bullet_{r,m\times n}$ of $\mathcal{M}_{r,m\times n}$)

indexed by partitions $\mathcal{P}([n]) = \mathcal{S}_n/\sim$ of the set $[n]$

Together with refinements of partitions, the set $\mathcal{P}([n])$ of partitions of $[n]$ is a poset.

As categories:

	$\underline{\text{Part}}^{\prec}_{[n]}$	$\underline{\text{Subv}}^{\subset}_{\mathcal{M}_{r,m\times n}}$
Ob	partitions of the set $[n]$	subvarieties of $\mathcal{M}_{r,m\times n}$
Mor	morphism from $\mathcal{P}_1$ to $\mathcal{P}_2$ iff $\mathcal{P}_1 \prec \mathcal{P}_2$	morphism from U_1 to U_2 iff $U_1 \subset U_2$

The filtration $\mathcal{I}^\bullet_{r,m\times n}$ is the functor $\underline{\text{Part}}^{\prec}_{[n]} \longrightarrow \underline{\text{Subv}}^{\subset}_{\mathcal{M}_{r,m\times n}}, \ \mathcal{P} \mapsto \mathcal{I}^{\mathcal{P}}_{r,m\times n}$.

Dually: The opposite category $\underline{\text{Part}}^{\prec,\text{op}}_{[n]}$ of $\underline{\text{Part}}^{\prec}_{[n]}$ is $\underline{\text{Part}}^{\succ}_{[n]}$, i.e., partitions of $[n]$ with coarsenings $\succ$ of partitions as morphisms. In this formulation, the finest common coarsening of partitions $\mathcal{P}_1, \dots, \mathcal{P}_k$ then is their inverse limit.

Euclidean distance (ED) degree

Motivation: complexity during and after training

1. For an arbitrary learned function, find a nearest invariant function.
2. Training invariant networks: determine pure critical points for Euclidean loss.

Definition

The **Euclidean distance degree** of an algebraic variety $\mathcal{X}$ in $\mathbb{R}^N$ is the number of complex critical points of the squared Euclidean distance from $\mathcal{X}$ to a general point outside the variety. It is denoted by $\deg_{\mathrm{ED}}(\mathcal{X})$.

Examples: $\deg_{\mathrm{ED}}(\text{circle}) = 2$, $\deg_{\mathrm{ED}}(\text{ellipse}) = 4$.

ED degree of $\mathcal{M}_{r,m\times n}(\mathbb{R})$ and $\mathcal{I}^{\sigma}_{r,m\times n}(\mathbb{R})$

Let $\sigma = \pi_1 \circ \cdots \circ \pi_k \in \mathcal{S}_n$ and $r \leq \min(m,n)$. Then

- $\deg_{\mathrm{ED}}(\mathcal{M}_{r,m\times n}(\mathbb{R})) = \binom{\min(m,n)}{r}$, proof via Eckart–Young theorem
- $\deg_{\mathrm{ED}}\left(\mathcal{I}^{\sigma}_{r,m\times n}(\mathbb{R})\right) = \deg_{\mathrm{ED}}\left(\mathcal{M}_{\min(r,k),m\times k}(\mathbb{R})\right) = \binom{\min(m,k)}{\min(r,k)}$.

[4] J. Draisma, E. Horobeţ, G. Ottaviani, B. Sturmfels, R. R. Thomas. The Euclidean Distance Degree of an Algebraic Variety. *Found. Comp. Math.*, 16:99–149, 2016.
[5] K. Kozhasov, A. Muniz, Y. Qi, L. Sodomaco. On the minimal algebraic complexity of the rank-one approximation problem for general linear products. Preprint arXiv:2309.15105, 2023.

Equivariant linear autoencoders

$f_\theta : \mathbb{R}^n \longrightarrow \mathbb{R}^r \longrightarrow \mathbb{R}^n \qquad r < n$

$G = \langle \sigma \rangle \leq \mathcal{S}_n$ a **cyclic** permutation group generated by a single $\sigma \in \mathcal{S}_n$

Equivariance under σ: $f_\theta \circ \sigma \equiv \sigma \circ f_\theta$.

For matrices: M equivariant if $MP_\sigma = P_\sigma M$. commutator of P_σ

In- and output

- $n = p^2$: $p \times p$ image with real pixels
- $n = p^3$: cubic 3D scenery

Characterizing $\mathcal{E}^\sigma_{r,n\times n}$

- dim: ✓
- deg: ✓
- Sing: ✓
- ED degree: under construction!

Exploiting similarity transforms of the form

$$P_\sigma = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \overset{\sim_{T_1}}{\mapsto} \left(\begin{array}{ccc|cc} 0 & 0 & 1 & & \\ 1 & 0 & 0 & \multicolumn{2}{c}{0} \\ 0 & 1 & 0 & & \\ \hline & 0 & & 0 & 1 \\ & & & 1 & 0 \end{array}\right) \overset{\sim_{T_2}}{\mapsto} \begin{pmatrix} 1 & & & & \\ & \zeta_3 & & & \\ & & \zeta_3^2 & & \\ & & & 1 & \\ & & & & -1 \end{pmatrix}.$$

For a subvariety $\mathcal{X} \subset \mathcal{M}_{m\times n}$ and fixed $T \in \mathrm{GL}_n(\mathbb{C})$, we denote by $\mathcal{X}^{\cdot T}$ the image of $\mathcal{X}$ under the linear isomorphism

$$\cdot T\colon\ \mathcal{M}_{m\times n} \longrightarrow \mathcal{M}_{m\times n}, \quad M \mapsto MT.$$

Lemma

Let $\mathcal{X} \subset \mathcal{M}_{m\times n}$ be a subvariety and let $T \in \mathrm{GL}_n(\mathbb{C})$. Then:

- $\boxed{\dim(\mathcal{X}^{\cdot T}) = \dim \mathcal{X},}$
- $\boxed{\deg(\mathcal{X}^{\cdot T}) = \deg \mathcal{X},}$
- $\boxed{\mathrm{Sing}(\mathcal{X}^{\cdot T}) = \mathrm{Sing}(\mathcal{X})^{\cdot T},}$
- $\boxed{(\mathcal{X}^{\cdot T}) \cap \mathcal{M}_{r,m\times n} = (\mathcal{X} \cap \mathcal{M}_{r,m\times n})^{\cdot T}}$ for any $r \leq \min(m,n)$.

Notation: For $T \in \mathrm{GL}_n(\mathbb{C})$ and $M \in \mathcal{M}_{n\times n}$, denote $\boxed{M^{\sim T} := T^{-1}MT}$.

Observation

A matrix M commutes with a matrix P if and only if $P^{\sim T}$ commutes with $M^{\sim T}$.

Characterizing equivariance

Proposition

There is a one-to-one correspondence between the **irreducible components** of $\mathcal{E}^{\sigma}_{r,n\times n}$ and the integer solution vectors $\mathbf{r} = (r_{l,m})$ of

$$\boxed{\sum_{l\geq 1} \sum_{m\in(\mathbb{Z}/l\mathbb{Z})^{\times}} r_{l,m} \;=\; r\,,}$$

where $0 \leq r_{l,m} \leq d_l$. d_l the dimension of the eigenspace of P_{σ} of the eigenvalue $\zeta_k = e^{2\pi i/l}$

Properties of $\mathcal{E}^{\sigma}_{r,n\times n}$

$$\dim\left(\mathcal{E}^{\sigma}_{r,n\times n}(\mathbb{C})\right) \;=\; \max_{\mathbf{r}=(r_{l,m})}\left\{\sum_{l\geq 1}\sum_{m\in(\mathbb{Z}/l\mathbb{Z})^{\times}} (2d_k - r_{l,m})\cdot r_{l,m}\right\},$$

$$\deg\left(\mathcal{E}^{\sigma,(\mathbf{r})}_{r,n\times n}(\mathbb{C})\right) \;=\; \prod_{l\geq 1}\prod_{m\in(\mathbb{Z}/l\mathbb{Z})^{\times}}\prod_{i=0}^{d_k-r_{l,m}-1}\frac{(d_k+i)!\cdot i!}{(r_{l,m}+i)!\cdot(d_k-r_{l,m}+i)!},$$

$$\operatorname{Sing}\left(\mathcal{E}^{\sigma}_{r,n\times n}(\mathbb{K})\right) \;=\; \mathcal{E}^{\sigma}_{r-1,n\times n}(\mathbb{K})\,. \qquad \mathbb{K}\in\{\mathbb{R},\mathbb{C}\}$$

Conclusion

Key points: algebraic geometry helps for...

1. a thorough study of function spaces of linear neural networks
   fully connected, convolutional
2. understanding the training process
   locating critical points of the loss
3. the design of neural networks
   invariance implies rank constraint & weight sharing property
4. determining the complexity during and post training
   ED degree of real varieties

Current/future work

- ◇ full characterization of equivariance
  non-cyclic permutation groups
- ◇ generalization to other groups
  e.g. non-discrete groups
- ◇ variation of the network architecture
  more layers, non-linear activation functions

Grazie mille/merci beaucoup/tack så mycket för er uppmerksamhet!

Characterizing invariance

$\mathcal{S}_n \ni \sigma = \pi_1 \circ \cdots \circ \pi_k$ decomposition of σ into k pairwise disjoint cycles

$\psi_{\mathcal{P}(\sigma)} \colon \mathcal{I}^{\sigma}_{r,m\times n} \cong \mathcal{M}_{\min(r,k),m\times k}$ linear isomorphism

Properties of $\mathcal{I}^{\sigma}_{r,m\times n}$

$$
\begin{aligned}
\dim\left(\mathcal{I}^{\sigma}_{r,m\times n}\right) &= \min(r,k)\cdot(m+k-\min(r,k)),\\
\deg\left(\mathcal{I}^{\sigma}_{r,m\times n}\right) &= \prod_{i=0}^{k-\min(r,k)-1} \frac{(m+i)!\cdot i!}{(\min(r,k)+i)!\cdot(m-(\min(r,k)+i)!},\\
\operatorname{Sing}(\mathcal{I}^{\sigma}_{r,m\times n}) &= \psi^{-1}_{\mathcal{P}(\sigma)}\left(\mathcal{M}_{\min(r,k)-1,m\times k}\right).
\end{aligned}
$$

Euclidean distance degree

$$
\deg_{\mathrm{ED}}\left(\mathcal{I}^{\sigma}_{r,m\times n}(\mathbb{R})\right) = \binom{\min(m,k)}{\min(r,k)}.
$$

Stepwise diagonalization of permutation matrices: an example

Consider the permutation $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 4 & 1 & 2 \end{pmatrix} = (1\,3\,4)(2\,5) \in \mathcal{S}_5$. Then

$$P_\sigma = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \overset{\sim T_1}{\mapsto} \left(\begin{array}{ccc|cc} 0 & 0 & 1 & & \\ 1 & 0 & 0 & \multicolumn{2}{c}{0} \\ 0 & 1 & 0 & & \\ \hline \multicolumn{3}{c|}{\multirow{2}{*}{0}} & 0 & 1 \\ & & & 1 & 0 \end{array}\right) \overset{\sim T_2}{\mapsto} \begin{pmatrix} 1 & & & & \\ & \zeta_3 & & & \\ & & \zeta_3^2 & & \\ & & & 1 & \\ & & & & -1 \end{pmatrix}$$

with

$$T_1 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad T_2 = \left(\begin{array}{ccc|cc} 1 & 1 & 1 & & \\ 1 & \zeta_3 & \zeta_3^2 & \multicolumn{2}{c}{0} \\ 1 & \zeta_3^2 & \zeta_3 & & \\ \hline \multicolumn{3}{c|}{\multirow{2}{*}{0}} & 1 & 1 \\ & & & 1 & -1 \end{array}\right) \in \mathrm{GL}_5(\mathbb{C}),$$

where ζ_3 denotes the primitive 3rd root of unity $\exp^{2\pi i/3}$.
+ grouping identical eigenvalues (optional step)

N.B.: T_2 is block diagonal with Vandermonde matrix blocks $V(1, \zeta_3, \zeta_3^3)$ and $V(1, -1)$.