# AI Pictures at a Mathematical Exhibition: How Applied Harmonic Analysis meets Machine Learning

**Radu Balan**

Department of Mathematics and Norbert Wiener Center for Harmonic Analysis and Applications
University of Maryland, College Park, MD

June 29, 2023
University of Torino, Turin, Italy

Norbert Wiener Center
for Harmonic Analysis and Applications

# Acknowledgments

This material is based upon work partially supported by the National Science Foundation under grant no. DMS-2108900 and Simons Foundation. "Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation."

**Papers available online** at:
https://www.math.umd.edu/ rvbalan/

# Table of Contents:

# Table of Contents

# High-Level Overview

In this series of lectures, we discuss a few harmonic analysis techniques and problems applied to machine learning.

1. NN: Neural networks (NN) and their universal approximation property.
2. Lipschitz analysis: we provide rationals for studying Lipschitz properties of NNs, and then we perform a Lipschitz analysis of these networks. We focus on two aspects of this analysis: stochastic modelng of local vs. global analysis, and a scattering network inspired Lipschitz analysis of convolutive networks.
3. Invariance and Equivariance: We highlight the duality between invariance and covariance/equivariance, with focus on G-invariant representations.
4. Applications to data analysis and modeling: We present applications on a variety of problems: classification and regression on graphs; generative models for data sets; neural network based modeling of time-evolution of dynamical systems; discrete optimizatons.

# Table of Contents

# Table of Contents

# Acknowledgments

This material is based upon work partially supported by the National Science Foundation under grant no. DMS-2108900 and Simons Foundation. "Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation."

**Joint works with**:

Naveed Haghani (UMD,APL-JHU) Daniel Levy (UMD)

Maneesh Singh (Verisk, Comcast) Efstratos Tsoukanis (UMD)

R. Balan, N. Haghani, M. Singh, "Permutation Invariant Representations with Applications to Graph Deep Learning", arXiv preprint: 2203.07546 [math.FA], [cs.LG]

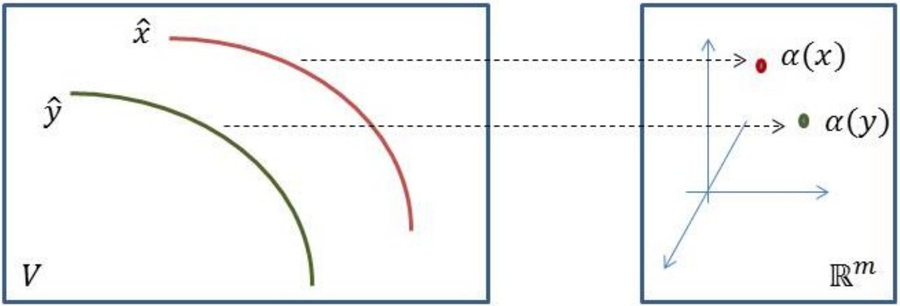
R. Balan, E. Tsoukanis, "Relationships between the Phase Retrieval Problem and Permutation Invariant Embeddings", arXiv preprint: 2306.13111 [math.FA]

# High-Level View

Two related problems with many variations:

Given a (discrete) group $G$ acting on a normed space $V$:

1. Construct a (bi)Lipschitz Euclidean embedding of the quotient space $V/G$, $\alpha : \hat{V} \to \mathbb{R}^m$. Classification of cosets.
2. Construct the projection onto cosets, $\pi : V \to [y] = \hat{y} = \{g.y \ , \ g \in G\}$.

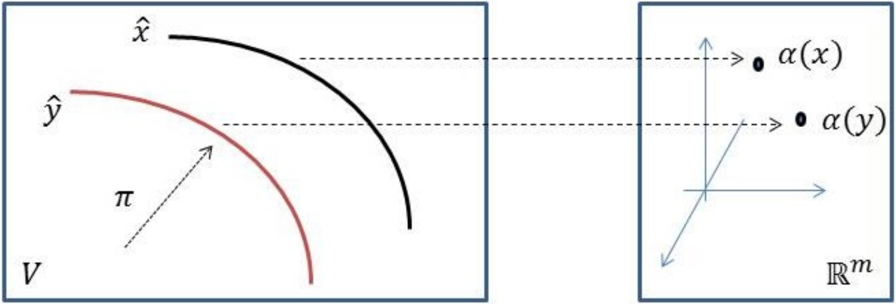


# Overview

Two related problems with many variations:

Given a (discrete) group $G$ acting on a normed space $V$:

1. Construct a (bi)Lipschitz Euclidean embedding of the quotient space $V/G$, $\alpha : \hat{V} \to \mathbb{R}^m$. Classification of cosets.
2. Construct projections onto cosets, $\pi : V \to [y] = \hat{y} = \{g.y \; , \; g \in G\}$. Optimizations within cosets.



# Table of Contents:

# A. Similarity of Matrices

Consider two symmetric matrices $A, B \in \mathrm{Sym}(n)$. When are they equivalent modulo an orthonormal change of coordinates?
Specificaly, is there an orthogonal matrix $U \in O(n)$ so that $B = UAU^T$ ?

An elementary derivation in linear algebra shows that $A \overset{O(n)}{\sim} B$ if and only if $A$ and $B$ have the same set of eigenvalues with exactly same multiplicities.

But what about other groups $G$? For instance what about the group of permutation matrices $\mathcal{S}_n$?
Find necessary and sufficient conditions so that $A \overset{\mathcal{S}_n}{\sim} B$.
Recall:

$$\mathcal{S}_n = \{P \in O(n) : P_{i,j} \in \{0,1\}\} = O(n) \cap \{W \in [0,1]^{n\times n} : W1 = 1, W^T 1 = 1\}$$

# A. The Graph Isomorphism Problem

Consider two graphs $G = (\mathcal{V}, \mathcal{E})$ and $\tilde{G} = (\tilde{\mathcal{V}}, \tilde{\mathcal{E}})$ with $n$ nodes. The graph isomorphism problem is the computational problem of determining whether these graphs are identical after a relabeling of nodes.

If $A$ and $\tilde{A}$ denote their adjacency matrices, these graphs are isomorphic if and only if $\tilde{A} = \Pi A \Pi^T$ for some permutation matrix $\Pi \in \mathcal{S}_n$.

Current state-of-the-art (Wikipedia): Babai (2015,2017) presented a quasi-polynomial algorithm with running time $2^{O((\log n)^c)}$, for some fixed $c > 0$. Helfgott (2017) claims that one can take $c = 3$.

Similar problem can be stated for weighted graphs: $A, \tilde{A} \in \mathrm{Sym}(n)$ with nonnegative entries, isomorphic if and only if $\tilde{A} = \Pi A \Pi^T$ for some $\Pi \in \mathcal{S}_n$.

# B. Graph Alignment Problems

Consider two $n \times n$ symmetric matrices $A, B$. In the alignment problem for quadratic forms one seeks an orthogonal matrix $U \in O(n)$ that minimizes

$$\|UAU^T - B\|_F^2 := trace((UAU^T - B)^2) = \|A\|_F^2 + \|B\|_F^2 - 2trace(UAU^T B).$$

The solution is well-known and depends on the eigendecomposition of matrices $A, B$: if $A = U_1 D_1 U_1^T$, $B = U_2 D_2 U_2^T$ then

$$U_{opt} = U_2 U_1^T \quad , \quad \|U_{opt} A U_{opt}^T - B\|_F^2 = \sum_{k=1}^{n} |\lambda_k - \mu_k|^2,$$

where $D_1 = diag(\lambda_k)$ and $D_2 = diag(\mu_k)$ are diagonal matrices with eigenvalues ordered monotonically.

# B. Quadratic Assignment Problem

The challenging case is when $U$ is constrained to the permutation group as is the case in the *graph matching problem*. In this case, the optimization problem becomes

$$\min_{U \in \mathcal{S}_n} \|UAU^T - B\|_F$$

turns into a QAP:

$$\max_{U \in \mathcal{S}_n} trace(UAU^T B).$$

This is equivalent to computing the natural distance $\mathbf{d}(\hat{A}, \hat{B}) = \min_{P,Q \in \mathcal{S}_n} \|PAP^T - QBQ^T\|_F$ between the equivalence classes $\hat{A}, \hat{B} \in \widehat{\mathrm{Sym}(n)}$ induced by the group action $\mathcal{S}_n \times \mathrm{Sym}(n) \to \mathrm{Sym}(n)$, $(\Pi, A) \mapsto \Pi A \Pi^T$.

# C. Graph Learning Problems

Given a data graph (e.g., social network, transportation network, citation network, chemical network, protein network, biological networks):

- Graph adjacency or weight matrix, $A \in \mathbb{R}^{n \times n}$;
- Data matrix, $X \in \mathbb{R}^{n \times r}$, where each row corresponds to a feature vector per node.

Contruct a map $f : (A, X) \to f(A, X)$ that performs:

1. classification: $f(A, X) \in \{1, 2, \cdots, c\}$
2. regression/prediction: $f(A, X) \in \mathbb{R}$.

Key observation: The outcome should be invariant to vertex permutation: $f(PAP^T, PX) = f(A, X)$, for every $P \in \mathcal{S}_n$.

# Invariance vs. Equivariance

Graph learning problems are prime examples of the difference between *invariant* vs. *equivariant* representations.
If the machine learning task is **node** classification or regression:

$$f : (A, X) \mapsto f(A, X) \ \in \ \{1, 2, \cdots, c\}^n \ \text{ or } \ \mathbb{R}^n$$

where $f(A, X)$ is a graph signal, i.e., $f(A, X)_i$ is signal at node $i$, then the nonlinear map $f$ is *equivariant* and must satisfy $f(PAP^T, PX) = Pf(A, X)$, for all $P \in \mathcal{S}_n$.

# Invariance vs. Equivariance

Graph learning problems are prime examples of the difference between *invariant* vs. *equivariant* representations.

If the machine learning task is **node** classification or regression:

$$f : (A, X) \mapsto f(A, X) \in \{1, 2, \cdots, c\}^n \text{ or } \mathbb{R}^n$$

where $f(A, X)$ is a graph signal, i.e., $f(A, X)_i$ is signal at node $i$, then the nonlinear map $f$ is *equivariant* and must satisfy $f(PAP^T, PX) = Pf(A, X)$, for all $P \in \mathcal{S}_n$.

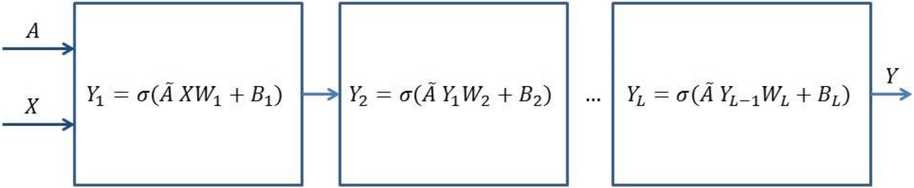
On the other hand, if the machine learning task is **graph** classification or regression,

$$f : (A, X) \mapsto f(A, X) \in \{1, 2, \cdots, c\} \text{ or } \mathbb{R}$$

where $f(A, X)$ is assigned for the entire graph, then the nonlinear map $f$ is *invariant* and must satisfy $f(PAP^T, PX) = f(A, X)$, for all $P \in \mathcal{S}_n$.

# C. Graph Convolution Networks (GCN), Graph Neural Networks (GNN)

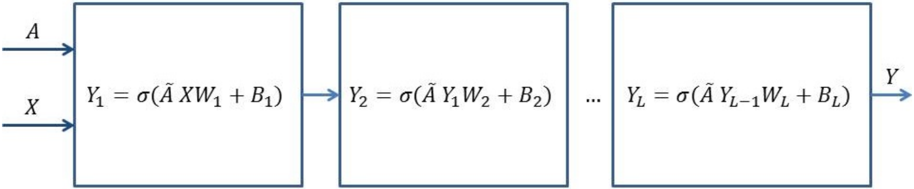
General architecture of a GCN/GNN




GCN (Kipf and Welling ('16)) choses $\tilde{A} = I + A$; GNN (Scarselli et.al. ('08), Bronstein et.al. ('16)) choses $\tilde{A} = p_l(A)$, a polynomial in adjacency matrix. $L$-layer GNN has parameters $(p_1, W_1, B_1, \cdots, p_L, W_L, B_L)$.

# C. Graph Convolution Networks (GCN), Graph Neural Networks (GNN)

General architecture of a GCN/GNN

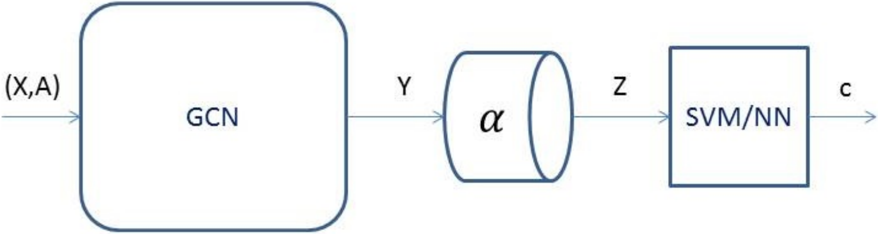



GCN (Kipf and Welling ('16)) choses $\tilde{A} = I + A$; GNN (Scarselli et.al. ('08), Bronstein et.al. ('16)) choses $\tilde{A} = p_l(A)$, a polynomial in adjacency matrix. $L$-layer GNN has parameters $(p_1, W_1, B_1, \cdots, p_L, W_L, B_L)$.

Note the *covariance (or, equivariance) property*: for any $P \in O(n)$ (including $\mathcal{S}_n$), if $(A, X) \mapsto (PAP^T, PX)$ and $B_i \mapsto PB_i$ then $Y \mapsto PY$.

# C. Deep Learning with GCN/GNN

The approach for the two learning tasks (classification or regression) is based on the following scheme (see also Maron et.al. ('19)):




where $\alpha$ is a permutation invariant map (embedding), and SVM/NN is a single-layer or a deep neural network (Support Vector Machine or a Fully Connected Neural Network) trained on invariant representations.

The purpose of this talk is to analyze the $\alpha$ component.

# The metric space $\hat{V}$ when $V = \mathbb{R}^{n\times d}$

Recall the equivalence relation $\sim$ on $V = \mathbb{R}^{n\times d}$ induced by the group of permutation matrices $\mathcal{S}_n$ acting on $V$ by left multiplication: for any $X, X' \in \mathbb{R}^{n\times d}$,

$$X \sim X' \quad \Leftrightarrow \quad X' = PX \ , \ \text{for some } P \in \mathcal{S}_n$$

Let $\widehat{\mathbb{R}^{n\times d}} = \mathbb{R}^{n\times d} / \sim$ be the quotient space endowed with the natural distance induced by Frobenius norm $\|\cdot\|_F$

$$\mathbf{d}(\hat{X}_1, \hat{X}_2) = \min_{P\in S_n} \|X_1 - PX_2\|_F \quad , \quad \hat{X}_1, \hat{X}_2 \in \widehat{\mathbb{R}^{n\times d}}.$$

# The metric space $\hat{V}$ when $V = \mathbb{R}^{n\times d}$

Recall the equivalence relation $\sim$ on $V = \mathbb{R}^{n\times d}$ induced by the group of permutation matrices $\mathcal{S}_n$ acting on $V$ by left multiplication: for any $X, X' \in \mathbb{R}^{n\times d}$,

$$X \sim X' \quad \Leftrightarrow \quad X' = PX \ , \ \text{for some } P \in \mathcal{S}_n$$

Let $\widehat{\mathbb{R}^{n\times d}} = \mathbb{R}^{n\times d}/\sim$ be the quotient space endowed with the natural distance induced by Frobenius norm $\|\cdot\|_F$

$$\mathbf{d}(\hat{X}_1, \hat{X}_2) = \min_{P\in S_n} \|X_1 - PX_2\|_F \quad , \quad \hat{X}_1, \hat{X}_2 \in \widehat{\mathbb{R}^{n\times d}}.$$

The computation of the minimum distance is performed by solving the Linear Assignment Problem (LAP) whose convex relaxation is exact:

$$\max_{P\in\mathcal{S}_n} trace(PX_2X_1^T) = \max_{W\in DS(n)} trace(WX_2X_1^T)$$

where $DS(n) = \{W \in [0,1]^{n\times n} : W1 = 1, W^T1 = 1\}$ is the convex set of doubly stochastic matrices.

# The embedding problem

Problem: Construct a bi-Lipschitz embedding $\hat{\alpha} : \widehat{\mathbb{R}^{n\times d}} \to \mathbb{R}^m$, i.e., an integer $m = m(n,d)$, a map $\alpha : \mathbb{R}^{n\times d} \to \mathbb{R}^m$ with constants $0 < a \leq b < \infty$ so that for any $X, X' \in \mathbb{R}^{n\times d}$,

1. If $X \sim X'$ then $\alpha(X) = \alpha(X')$.
2. If $\alpha(X) = \alpha(X')$ then $X \sim X'$.
3. $a \cdot \mathbf{d}(\hat{X}, \hat{X}') \leq \|\alpha(X) - \alpha(X')\|_2 \leq b \cdot \mathbf{d}(\hat{X}, \hat{X}')$.

where $\mathbf{d}(\hat{X}, \hat{X}') = \min_{P\in\mathcal{S}_n} \|X - PX'\|_F$.

# A Universal Embedding

Consider the map

$$\mu : \widehat{\mathbb{R}^{n\times d}} \to \mathcal{P}(\mathbb{R}^d) \quad , \quad \mu(X)(x) = \frac{1}{n}\sum_{k=1}^{n} \delta(x - x_k)$$

where $\mathcal{P}(\mathbb{R}^d)$ denotes the convex set of probability measures over $\mathbb{R}^d$, and $\delta$ denotes the Dirac measure. $x_k$ is the $k^{th}$ row of $X$.
Clearly $\mu(X') = \mu(X)$ iff $X' = PX$ for some $P \in \mathcal{S}_n$.
The Wasserstein-2 distance is isometrically equivalent to **d**:

$$W_2(\mu(X), \mu(Y))^2 := \inf_{q \in J(\mu(X),\mu(Y))} \mathbb{E}_q[\|x - y\|_2^2] = \min_{P \in \mathcal{S}_n} \|Y - PX\|^2$$

By Kantorovich-Rubinstein theorem, the Wasserstein-1 distance (the Earth moving distance) extends to a norm on the space of signed Borel measures.

# A Universal Embedding

Consider the map

$$\mu : \widehat{\mathbb{R}^{n\times d}} \to \mathcal{P}(\mathbb{R}^d) \;\;,\;\; \mu(X)(x) = \frac{1}{n}\sum_{k=1}^{n} \delta(x - x_k)$$

where $\mathcal{P}(\mathbb{R}^d)$ denotes the convex set of probability measures over $\mathbb{R}^d$, and $\delta$ denotes the Dirac measure. $x_k$ is the $k^{th}$ row of $X$.
Clearly $\mu(X') = \mu(X)$ iff $X' = PX$ for some $P \in \mathcal{S}_n$.
The Wasserstein-2 distance is isometrically equivalent to **d**:

$$W_2(\mu(X), \mu(Y))^2 := \inf_{q\in J(\mu(X),\mu(Y))} \mathbb{E}_q[\|x - y\|_2^2] = \min_{P\in\mathcal{S}_n} \|Y - PX\|^2$$

By Kantorovich-Rubinstein theorem, the Wasserstein-1 distance (the Earth moving distance) extends to a norm on the space of signed Borel measures.
Main drawback: $\mathcal{P}(\mathbb{R}^d)$ is infinite dimensional!

# Finite Dimensional Embeddings

Idea: "Project" the measure onto a finite dimensional space. This is accomplished by *kernel methods*:

Fix a family of functions $f_1, \cdots, f_m$ and consider:

$$\mu(X) \mapsto \int_{\mathbb{R}^d} f_j(x) d\mu(X) = \frac{1}{n} \sum_{k=1}^{n} f_j(x_k) \quad , \quad j \in [m]$$

# Finite Dimensional Embeddings

Idea: "Project" the measure onto a finite dimensional space. This is accomplished by *kernel methods*:
Fix a family of functions $f_1, \cdots, f_m$ and consider:

$$\mu(X) \mapsto \int_{\mathbb{R}^d} f_j(x) d\mu(X) = \frac{1}{n} \sum_{k=1}^{n} f_j(x_k) \quad , \quad j \in [m]$$

Possible choices:

1. Polynomial embeddings: $\mathbb{R}[X]^{\mathcal{S}_n}$, ring of invariant polynomials; [Lipman&al.],[Peyré&al.],[Sanay&al.],[Kemper book] ...
2. Gaussian kernels: $f_j(x) = exp(-\|x - a_j\|^2/\sigma_j^2)$ ; [Gilmer&al.],[Zaheer&al.], [Vinyals&al.],...
3. Fourier kernels (cmplx embd): $f_j(x) = exp(2\pi i \langle x, \omega_j \rangle)$; related to Prony method; [Li&Liao] for bi-Lipschitz estimates.

Main drawback: No global bi-Lipschitz embeddings [Cahill&al.'19]. Ok on (some) compacts.

# Polynomial Expansions - Quadratics

In the case $d = 1$ recall Vieta's formulas, Newton-Girard identities

$$P(X) = \prod_{k=1}^{N} (X - x_k) \leftrightarrow (\sum_k x_k, \sum_k x_k^2, ..., \sum_k x_k^n)$$

# Polynomial Expansions - Quadratics

In the case $d = 1$ recall Vieta's formulas, Newton-Girard identities

$$P(X) = \prod_{k=1}^{N} (X - x_k) \leftrightarrow (\sum_k x_k, \sum_k x_k^2, ..., \sum_k x_k^n)$$

For $d > 1$, consider the quadratic $d$-variate polynomial:

$$\begin{aligned} P(Z_1, \cdots, Z_d) &= \prod_{k=1}^{n} \left( (Z_1 - x_{k,1})^2 + \cdots + (Z_d - x_{k,d})^2 \right) \\ &= \sum_{p_1,\ldots,p_d=0}^{2n} a_{p_1,\ldots,p_d} Z_1^{p_1} \cdots Z_d^{p_d} \end{aligned}$$

Encoding complexity:

$$m = \binom{2n+d}{d} \sim (2n)^d.$$

# Polynomial Expansions - Quadratics (2)

A more careful analysis of $P(Z_1,...,Z_d)$ reveals a form:

$$P(Z_1,...,Z_d) = t^n + Q_1(Z_1,...,Z_d)t^{n-1} + \cdots + Q_{n-1}(Z_1,...,Z_d)t + Q_n(Z_1,...,Z_d)$$

where $t = Z_1^2 + \cdots + Z_d^2$ and each $Q_k(Z_1,...,Z_d) \in \mathbb{R}_k[Z_1,...,Z_d]$ is a (non-homogeneous) polynomial of degree $k$. Hence one needs to encode:

$$m = \begin{pmatrix} d+1 \\ 1 \end{pmatrix} + \begin{pmatrix} d+2 \\ 2 \end{pmatrix} + \cdots + \begin{pmatrix} d+n \\ n \end{pmatrix} = \begin{pmatrix} d+n+1 \\ n \end{pmatrix} - 1$$

number of coefficients.

A significant drawback: Inversion is numerically unstable and embedding is not Lipschitz.

# Readout Mapping Approach

## Polynomial Expansion - Linear Forms

A stable (Lipschitz, not bi-Lipschitz!) embedding can be constructed as follows (see also Gobels' algorithm (1996) or [Derksen, Kemper '02]). Consider the $n$ linear forms $\lambda_k(Z_1, ..., Z_d) = x_{k,1}Z_1 + \cdots x_{k,d}Z_d$. Construct the polynomial in variable $t$ with coefficients in $\mathbb{R}[Z_1, ..., Z_d]$:

$$P(t) = \prod_{k=1}^{n}(t-\lambda_k(Z_1, ..., Z_d)) = t^n - e_1(Z_1, .., Z_d)t^{n-1} + \cdots (-1)^n e_n(Z_1, ..., Z_d)$$

$$= t^n + \sum_{\substack{p_0, p_1, \cdots, p_d \geq 0 \\ p_0 + p_1 + \cdots + p_d = n \ , \ p_0 < n}} c_{p_0,p_1,\cdots,p_d} t^{p_0} Z_1^{p_1} \cdots Z_d^{p_d}$$

The elementary symmetric polynomials $(e_1, ..., e_n)$ are in 1-1 correspondence (Newton-Girard theorem) with the moments: $\mu_p = \sum_{k=1}^{n} \lambda_k^p(Z_1, ..., Z_d)$, $1 \leq p \leq n$.

# Polynomial Expansions - Linear Forms (2)

Each $\mu_p$ is a homogeneous polynomial of degree $p$ in $d$ variables. Hence to encode each of them one needs $\binom{d+p-1}{p}$ coefficients. Hence the embedding dimension is

$$m_0 = \binom{d}{1} + \binom{d+1}{2} + \cdots + \binom{d+n-1}{n} = \binom{d+n}{n} - 1$$

# Polynomial Expansions - Linear Forms (2)

Each $\mu_p$ is a homogeneous polynomial of degree $p$ in $d$ variables. Hence to encode each of them one needs $\binom{d+p-1}{p}$ coefficients. Hence the embedding dimension is

$$m_0 = \binom{d}{1} + \binom{d+1}{2} + \cdots + \binom{d+n-1}{n} = \binom{d+n}{n} - 1$$

The map $\alpha_0 : \mathbb{R}^{n\times d} \to \mathbb{R}^{m_0}$, $X \mapsto (c_{p_0,p_1,\cdots,p_d})_{p_0,p_1,\cdots,p_d}$ is injective modulo $\mathcal{S}_n$ but it is not Lipschitz. However a simple modification as suggested by [Cahill et.al.'19] makes it Lipschitz.

# Polynomial Lipschitz embedding

Denote by $L_0$ the Lipschitz constant of $\alpha_0$ when restricted to the closed unit ball $B_1(\mathbb{R}^{n\times d}) : \{X \in \mathbb{R}^{n\times d}\ ,\ \|X\| \leq 1\}$ of $\mathbb{R}^{n\times d}$, i.e. $\|\alpha_0(X) - \alpha_0(Y)\| \leq L_0\|X - Y\|$ for any $X, Y \in \mathbb{R}^{n\times d}$ with $\|X\|, \|Y\| \leq 1$. Let $\varphi_0 : \mathbb{R} \to [0,1]$, $\varphi_0(x) = min(1, \frac{1}{x})$ be a Lipschitz monotone decreasing function with Lipschitz constant 1.

**Theorem**

*The map:*

$$\alpha_1 : \mathbb{R}^{n\times d} \to \mathbb{R}^m \ ,\ \alpha_1(X) = \begin{pmatrix} \alpha_0\left(\varphi_0(\|X\|)X\right) \\ \|X\| \end{pmatrix},$$

*with* $m = \begin{pmatrix} n+d \\ d \end{pmatrix} = m_0 + 1$ *lifts to an injective and globally Lipschitz map* $\hat{\alpha}_1 : \widehat{\mathbb{R}^{n\times d}} \to \mathbb{R}^m$ *with Lipschitz constant* $Lip(\hat{\alpha}_1) \leq \sqrt{1 + L_0^2}$.

# Minimality

For $d = 1$, $m = n$ which is minimal.

For $d = 2$, $m = \frac{n^2+3n}{2}$. Is this minimal?

# Algebraic Embedding

## Encoding using Complex Roots

Idea: Consider the case $d = 2$. Then each $x_1, \cdots, x_n \in \mathbb{R}^2$ can be replaced by $n$ complex numbers $z_1, \cdots, z_n \in \mathbb{C}$, $z_k = x_{k,1} + i x_{k,2}$.
Consider the complex polynomial:

$$Q(z) = \prod_{k=1}^{n} (z - z_k) = z^n + \sum_{k=1}^{n} \sigma_k z^{n-k}$$

which requires $n$ complex numbers, or $2n$ real numbers.

# Algebraic Embedding

## Encoding using Complex Roots

Idea: Consider the case $d = 2$. Then each $x_1, \cdots, x_n \in \mathbb{R}^2$ can be replaced by $n$ complex numbers $z_1, \cdots, z_n \in \mathbb{C}$, $z_k = x_{k,1} + ix_{k,2}$.
Consider the complex polynomial:

$$Q(z) = \prod_{k=1}^{n}(z - z_k) = z^n + \sum_{k=1}^{n} \sigma_k z^{n-k}$$

which requires $n$ complex numbers, or $2n$ real numbers.

Open problem: Can this construction be extended to $d \geq 3$?
Remark: A drawback of polynomial (algebraic) embeddings: [Cahill'19] showed that polynomial embeddings of translation invariant spaces cannot be bi-Lipschitz.

# The Max Pool approach

The idea is provided by the following observation.
Let $\downarrow: \mathbb{R}^n \to \mathbb{R}^n$ denote the *sorting map* $x \mapsto \downarrow x = \Pi x$, $\Pi \in \mathcal{S}_n$, so that

$$(\Pi x)_1 \geq (\Pi x)_2 \geq \cdots \geq (\Pi x)_n.$$

# The Max Pool approach

The idea is provided by the following observation.
Let $\downarrow : \mathbb{R}^n \to \mathbb{R}^n$ denote the *sorting map* $x \mapsto \downarrow x = \Pi x$, $\Pi \in \mathcal{S}_n$, so that

$$(\Pi x)_1 \geq (\Pi x)_2 \geq \cdots \geq (\Pi x)_n.$$

## Lemma

$\downarrow : \widehat{\mathbb{R}^n} \to \mathbb{R}^n$ *is an isometry (hence bi-Lipschitz):*

$$\| \downarrow (x) - \downarrow (y) \| = \min_{P \in \mathcal{S}_n} \|x - Py\| \ , \quad \text{for all} \quad x, y \in \mathbb{R}^n.$$

Proof is based on the rearrangement inequality (see Wikipedia, or Hardy-Littlewood-Pólya "Inequalities" §10.2).

Our main goal is to extend this construction from $\mathbb{R}^n$ to $\mathbb{R}^{n \times d}$

# The Encoder $\beta_A$

## Notations

Recall the equivalence relation, for $X, Y \in \mathbb{R}^{n\times d}$,

$$X \sim Y \quad \Leftrightarrow \quad \exists \Pi \in \mathcal{S}_n \ , \ Y = \Pi X$$

that induces a quotient space $\widehat{\mathbb{R}^{n\times d}} = \mathbb{R}^{n\times d}/\sim$ and the natural distance

$$\mathbf{d} : \widehat{\mathbb{R}^{n\times d}} \times \widehat{\mathbb{R}^{n\times d}} \to \mathbb{R} \ \ , \ \ \mathbf{d}([X],[Y]) = \min_{\Pi\in\mathcal{S}_n} \|X - \Pi Y\|_F$$

In the following we construct an Euclidean embedding of the form

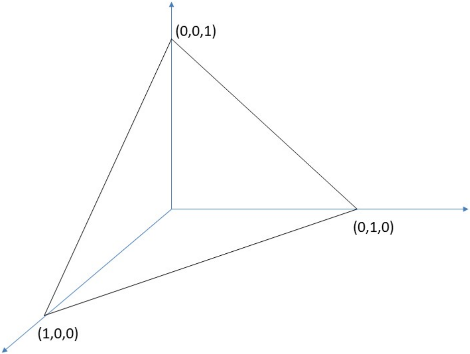
$$\beta_A : \mathbb{R}^{n\times d} \to \mathbb{R}^{n\times D} \ \ , \ \ \beta_A(X) = \downarrow (XA)$$

where $\downarrow(\cdot)$ sorts decreasingly each column of $\cdot$, independently.
The matrix $A \in \mathbb{R}^{d\times D}$ is called the *key* of encoder $\beta_A$.
The key is called *universal* if $\widehat{\beta_A} : \widehat{\mathbb{R}^{n\times d}} \to \mathbb{R}^{n\times D}$ is injective.

# Intuition behind universality of keys

Consider the case
$n = 2$ , $d = 3$

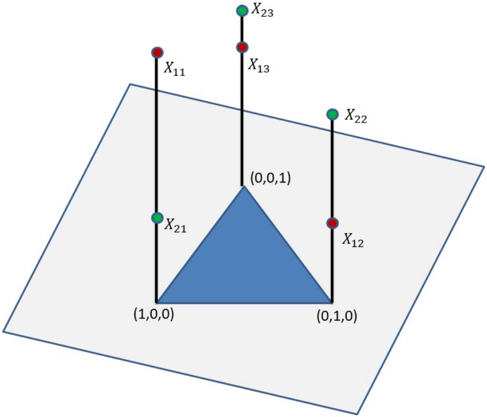
$$X = \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \end{bmatrix}$$



# Intuition behind universality of keys

Consider the case
$n = 2$ , $d = 3$

$$X = \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \end{bmatrix}$$



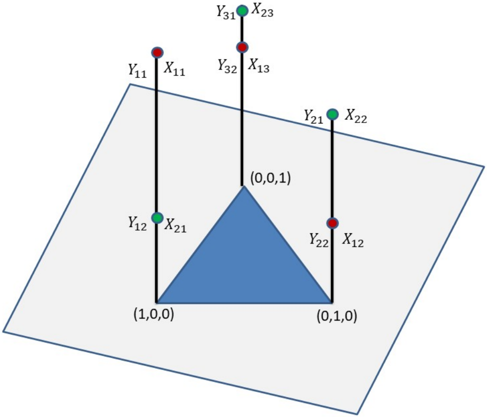
# Intuition behind universality of keys

$$X = \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \end{bmatrix}$$

$$Y = \downarrow X$$

$$Y = \begin{bmatrix} Y_{11} & Y_{12} & Y_{13} \\ Y_{21} & Y_{22} & Y_{23} \end{bmatrix}$$

Information lost!

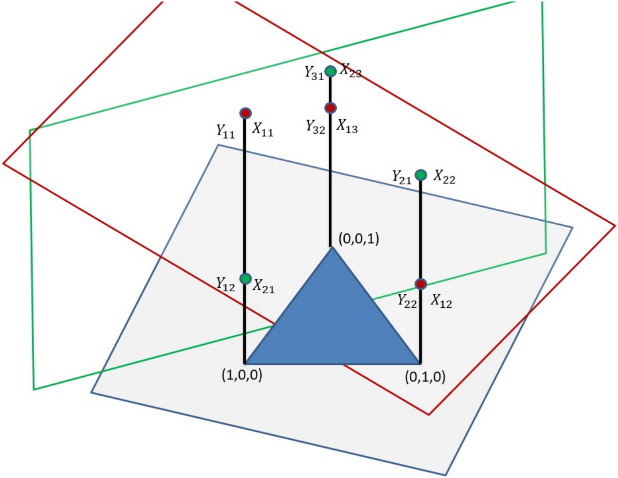


# Intuition behind universality of keys

$$X = \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \end{bmatrix}$$

$$Y = \downarrow X$$

$$Y = \begin{bmatrix} Y_{11} & Y_{12} & Y_{13} \\ Y_{21} & Y_{22} & Y_{23} \end{bmatrix}$$

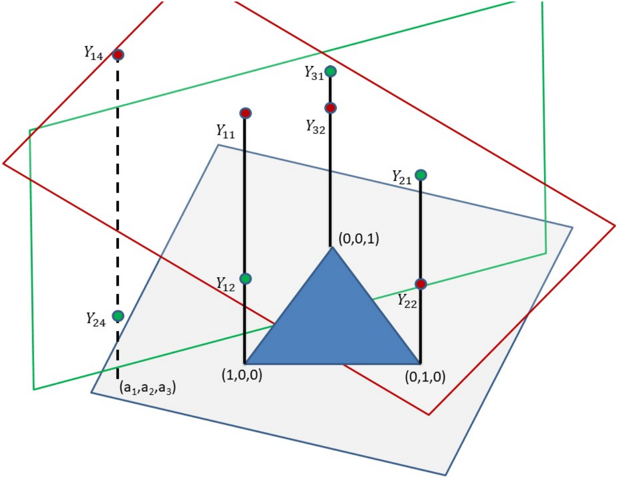


# Intuition for this encoder

$$X = \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \end{bmatrix}$$

$$Y = \downarrow \begin{bmatrix} X & Xa \end{bmatrix}$$

$$Y = \begin{bmatrix} Y_{11} & Y_{12} & Y_{13} & Y_{14} \\ Y_{21} & Y_{22} & Y_{23} & Y_{24} \end{bmatrix}$$



# Three results (1)

## Existence of Universal Keys

### Theorem

*Consider the metric space* $(\widehat{\mathbb{R}^{n\times d}}, \mathbf{d})$. *Set* $D = 1 + (d-1)n!$ *and let* $A \in \mathbb{R}^{d\times D}$ *be a matrix whose columns form a full spark frame. Then the key* $A$ *is universal and the induced map* $\hat{\beta}_A : \widehat{\mathbb{R}^{n\times d}} \to \mathbb{R}^{n\times D}$, $\hat{\beta}_A([X]) = \downarrow(XA)$ *is injective. Furthermore,* $\hat{\beta}_A$ *is bi-Lipschitz with constants* $a_0 = \min_{J\subset[D],|J|=d} s_d(A[J])$ *and* $b_0 = s_1(A)$, *where* $s_1(A)$ *denotes the largest singular value of* $A$, $A[J]$ *denotes the submatrix of* $A$ *formed by columns indexed by* $J$, *and* $s_d(A[J])$ *denotes the* $d^{th}$ *singular value (in this case, the smallest) of* $A[J]$. *Specifically, for any* $X, Y \in \mathbb{R}^{n\times d}$,

$$a_0\mathbf{d}([X],[Y]) \le \|\beta_A(X) - \beta_A(Y)\| \le b_0\mathbf{d}([X],[Y]) \qquad (3.1)$$

*where all norms are Frobenius norms.*

# Three results (2)

## Bi-Lipschitz Property of Universal Keys

### Theorem

*Assume the key* $A \in \mathbb{R}^{d \times D}$ *is universal, i.e., the induced map* $\hat{\beta}_A : \widehat{\mathbb{R}^{n \times d}} \to \mathbb{R}^{n \times D}$, $[X] \mapsto \beta_A(X) = \downarrow(XA)$ *is injective. Then* $\hat{\beta}_A$ *is bi-Lipschitz, that is, there are constants* $a_0 > 0$ *and* $b_0 > 0$ *so that for all* $X, Y \in \mathbb{R}^{n \times d}$,

$$a_0\, \mathbf{d}([X],[Y]) \leq \|\beta_A(X) - \beta_A(Y)\| \leq b_0\, \mathbf{d}([X],[Y]) \qquad (3.2)$$

*where all are Frobenius norms. Furthermore, an estimate for* $b_0$ *is provided by the largest singular value of A,* $b_0 = s_1(A)$.

# Three results (3)

## Dimension Reduction

### Theorem

*Assume $A \in \mathbb{R}^{d \times D}$ is a universal key for $\widehat{\mathbb{R}^{n \times d}}$ with $D \geq 2d$. Then, for $m \geq 2nd$, a generic linear operator $B : \mathbb{R}^{n \times D} \to \mathbb{R}^m$ with respect to Zariski topology on $\mathbb{R}^{n \times D \times m}$, the map*

$$\hat{\beta}_{A,B} : \widehat{\mathbb{R}^{n \times d}} \to \mathbb{R}^{2nd} \; , \; \hat{\beta}_{A,B}(\hat{X}) = B\left(\hat{\beta}_A(\hat{X})\right) \tag{3.3}$$

*is bi-Lipschitz. In particular, almost every full-rank linear operator $B : \mathbb{R}^{n \times D} \to \mathbb{R}^{2nd}$ produces such a bi-Lipschitz map.*

This result is compatible with a Whitney embedding theorem with the important caveat that the Whitney embedding result applies to smooth manifolds, whereas $\widehat{\mathbb{R}^{n \times d}}$ is not a manifold.

# Highlights of proofs

## First result: Universal keys

The upper bound is imediate. For lower bound, fix $X, Y \in \mathbb{R}^{n \times d}$:

$$\|\beta_A(X) - \beta_A(Y)\|_2^2 = \sum_{k=1}^{D} \| \downarrow (Xa_k) - \downarrow (Ya_k)\|_2^2 = \sum_{k=1}^{D} \|P_k X a_k - Q_k Y a_k\|_2^2$$

$$\stackrel{\Pi_k := Q_k^T P_k}{=} \sum_{k=1}^{D} \|(\Pi_k X - Y) a_k\|_2^2$$

# Highlights of proofs

## First result: Universal keys

The upper bound is imediate. For lower bound, fix $X, Y \in \mathbb{R}^{n \times d}$:

$$\|\beta_A(X) - \beta_A(Y)\|_2^2 = \sum_{k=1}^{D} \| \downarrow (Xa_k) - \downarrow (Ya_k)\|_2^2 = \sum_{k=1}^{D} \|P_k X a_k - Q_k Y a_k\|_2^2$$

$$\overset{\Pi_k := Q_k^T P_k}{=} \sum_{k=1}^{D} \|(\Pi_k X - Y) a_k\|_2^2 \geq \sum_{j=1}^{d} \|(\Pi_{k_j} X - Y) a_{k_j}\|_2^2$$

so that $\Pi_{k_1} = \cdots = \Pi_{k_d} = \Pi_0$ (pigeonhole principle: needs $D > (d-1)n!$).

# Highlights of proofs

## First result: Universal keys

The upper bound is imediate. For lower bound, fix $X, Y \in \mathbb{R}^{n\times d}$:

$$\|\beta_A(X)-\beta_A(Y)\|_2^2=\sum_{k=1}^{D}\|\downarrow(Xa_k)-\downarrow(Ya_k)\|_2^2=\sum_{k=1}^{D}\|P_kXa_k-Q_kYa_k\|_2^2$$

$$\stackrel{\Pi_k:=Q_k^TP_k}{=}\sum_{k=1}^{D}\|(\Pi_kX-Y)a_k\|_2^2\geq\sum_{j=1}^{d}\|(\Pi_{k_j}X-Y)a_{k_j}\|_2^2$$

so that $\Pi_{k_1}=\cdots=\Pi_{k_d}=\Pi_0$ (pigeonhole principle: needs $D>(d-1)n!$). Then:

$$\|\beta_A(X)-\beta_A(Y)\|_2^2\geq\sum_{j=1}^{d}\|(\Pi_0X-Y)a_{k_j}\|_2^2\stackrel{full\ \ spark}{\geq}s_d(A[J])^2\|\Pi_0X-Y\|^2$$

$$\geq s_d(A[J])^2\min_{\Pi\in\mathcal{S}_n}\|\Pi X-Y\|^2=s_d(A[J])^2\mathbf{d}([X],[Y])^2$$

# Highlights of proofs

## Second result: Bi-Lipschitz Property

The proof resembles the treatment of phase retrieval problem:

1. Homogeneity and compactness reduce the problem to local analysis.
2. The encoder is "locally" linearized. The failure of local lower Lipschitz bound implies a certain behavior for a Quadratically Constrained Ratio of Quadratics (QCRQ).
3. QCRQ has a minimizer:inf $\Rightarrow$ min. [Teboulle&al.]
   This step took most of time and lots of (self)convincing !
4. Contradiction to injectivity assumption.

# More detailed proof of the bi-Lipschitz result (1)

1. Reduction to local lower Lipschitz bound.
Assume $\inf_{X \not\sim Y} \|\beta_A(X) - \beta_A(Y)\|_2 / \mathbf{d}([X],[Y]) = 0$. By *homogeneity* and *compactness*, extract/construct sequences $(X_j)_j$ and $(Y_j)_j$ so that: (i) $X_j \to Z$; (ii) $Y_j \to Z$; (iii) $\|Y_j\| \leq \|X_j\| = \|Z\| = 1$; (iv) $\mathbf{d}([X_j],[Z]) = \|X_j - Z\|$; (v) $\mathbf{d}([X_j],[Y_j]) = \|X_j - Y_j\|$; (vi) $\mathbf{d}([Y_j],[Z]) = \|Y_j - z\|$.

2. Local linearization.
Let $H = \{P \in \mathcal{S}_n \ ; \ PZ = Z\}$ denote the stabilizer of $Z$. Let $U_j = X_j - Z$ and $V_j = Y_j - Z$. Then:

$$\lim_{j \to \infty} \frac{\sum_{k=1}^{D} \min_{Q \in H} \|QU_j a_k - V_j a_k\|^2}{\|U_j - V_j\|^2} = 0\ , \quad \|U_j - V_j\| \leq \|U_j - PV_j\|, \ \forall P \in H.$$

# More detailed proof of the bi-Lipschitz result (2)

3. QCQP

Last limit implies:

$$\inf_{\substack{(u,v)\in\mathbb{R}^{n\times d}:\\ U\neq QV,\forall V\in H}} \max_{P\in H} \frac{\sum_{k=1}^{D}\|(U-\Pi_k V)a_k\|_2^2}{\|U-PV\|^2} = 0$$

where $\Pi_k$ achieves alignment between $U_j a_k$ and $V_j a_k$.

Since these groups are finite, we obtain that the infimum is achieved!

Hence:

4. Injectivity no-go

There are $U, V \in \mathbb{R}^{n\times d}$ so that $Z+U \not\sim Z+V$ and yet $(Z+U)a_k = \Pi_k(Z+V)a_k$ for all $k\in[D]$. This shows $\beta_A(Z+U) = \beta_A(Z+V)$ which contradicts injectivity!

Q.E.D.

# Highlights of proofs

## Third result: Dimension Reduction

The proof follows the approach in [Cahill&al.], [Dufresne]:

$$0 = B(\beta_A(X)) - B(\beta_A(Y)) \Rightarrow \beta_A(X) - \beta_A(Y) \in \ker(B)$$

Need to show: $\beta_A(X) - \beta_A(Y) = 0$, or, $Ran(\Delta) \cap \ker(B) = \{0\}$, where

$$\Delta : \mathbb{R}^{n\times d} \times \mathbb{R}^{n\times d} \to \mathbb{R}^{n\times D} \ , \ \Delta(X,Y) = \beta_A(X) - \beta_A(Y).$$

In the polynomial case, [Cahill&al.] exploit arguments from algebraic geometry. Here the problem is simpler since $Ran(\Delta)$ is included in a finite union of linear subspaces of dimension at most $2nd$.
By a dimension argument it follows that the target space for $B$ must be of dimension at least $2nd$ to obtain an injective embedding. In this case, generically, $Ran(\Delta)$ and $\ker(B)$ intersect transversally.

# Towards universal keys

The arXiv preprint provides necessary and sufficient conditions for a key to be universal.

Open Problem: Given $(n, d)$ find the smallest dimension $D$ so that there exists a universal key $A \in \mathbb{R}^{d \times D}$ for $\mathbb{R}^{n \times d}$.

So far we obtained (joint with Daniel Levy (UMD) ):

| n | d | D-d |
|---|---|---|
| 2 | 2 | 1 |
| 3 | 2 | 2 |
| 4 | 2 | 2 |
| 5 | 2 | 3 |
| 6 | 2 | $\geq 4$ |

Open Problem: If a universal key exists for a triple $(n, d, D)$ then is it true that universal keys are generic in $\mathbb{R}^{d \times D}$ ?

# Related results

A sequence of preprints came out almost simultaneously:

1. R. Balan, N. Haghani, M.Singh, Permutation Invariant Representations with Applications to Graph Deep Learning, arXiv:2203.07546 (2022)
2. N. Dym, S. J. Gortler, Low Dimensional Invariant Embeddings for Universal Geometric Learning, arXiv:2205.02956 (2022)
3. J. Cahill, J. W. Iverson, D. G. Mixon, D. Packer, Group-invariant max filtering, arXiv:2205.14039 (2022)

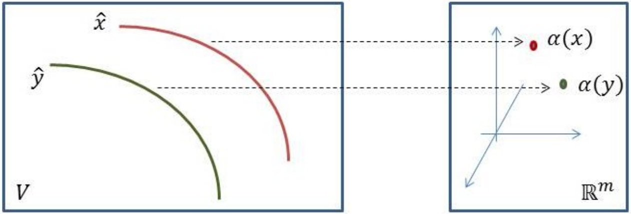
all of them based on sorting in one way or another. [Dym and Gortler] shows that the key size should be significantly smaller than $n!$. [Cahill et.al.'22] introduced the concept of *max filter* which is a special case of a more general G-invariant representation discussed next.

# Table of Contents

# High-Level View

Recall the framework for Euclidean embeddings of metric spaces induced by orthogonal representations of (finite) groups $G$ acting on a linear space $V$.

Metric space $(\hat{V}, \mathbf{d})$ where: $\hat{V} = V/G$ is the set of orbits, $[x] = \{U_g \ , \ g \in Gx\}$, for $x \in V$; and $\mathbf{d}(\hat{x}, \hat{y}) = \min_{u \in \hat{x}, v \in \hat{y}} \|u - v\|_V$.



# The Program

Given a (discrete) group $G$ acting unitarily on a normed space $V$, we formulate four general problems

1. Construct injective embeddings of the quotient space $V/G$, $\alpha : \hat{V} \to \mathbb{R}^m$. The injectivity problem.
2. Construct/Obtain bi-Lipschitz properties for the Euclidean embedding $\alpha : \hat{V} \to \mathbb{R}^m$. The stability problem.
3. Develop algorithms for inversion $\alpha^{-1} : \mathbb{R}^m \to \hat{V}$. The recovery problem.
4. Analyze specific cases. Applications.

# The Program

Given a discrete group $G$ acting unitarily on a normed space $V$, we formulate four general problems

1. Construct injective embeddings of the quotient space $V/G$, $\alpha : \hat{V} \to \mathbb{R}^m$. The injectivity problem.
2. Construct/Obtain bi-Lipschitz properties for the Euclidean embedding $\alpha : \hat{V} \to \mathbb{R}^m$. The stability problem.
3. Develop algorithms for inversion $\alpha^{-1} : \mathbb{R}^m \to \hat{V}$. The recovery problem.
4. Analyze specific cases. Applications.

# The Program

Given a discrete group $G$ acting unitarily on a normed space $V$, we formulate four general problems

1. Construct injective embeddings of the quotient space $V/G$, $\alpha : \hat{V} \to \mathbb{R}^m$. The injectivity problem.
2. Construct/Obtain bi-Lipschitz properties for the Euclidean embedding $\alpha : \hat{V} \to \mathbb{R}^m$. The stability problem.
3. Develop algorithms for inversion $\alpha^{-1} : \mathbb{R}^m \to \hat{V}$. The recovery problem.
4. Analyze specific cases. Applications.

Today we focus on the first two problems: injectivity and bi-Lipschitz stability.

# Invariant Representations

Let $V$ be a $d$-dimensional Hilbert space and $G$ a finite group of size $N = |G|$ acting unitarily on $V$, $\{U_g, g \in G\}$.
The quotient space $\widehat{V} = V/G$ is the set of orbits $[x] = \{U_g x\ ,\ g \in G\}$ induced by the group action, where for $x, y \in V$, $x \sim y$ iff $y = U_g x$ for some $g \in G$. $(\widehat{V}, \mathbf{d})$ becomes a metric space with the natural distance

$$\mathbf{d}([x],[y]) = \min_{g \in G} \|x - U_g y\|$$

How to construct an invariant representation?
The standard method in the computational invariant theory: Find generators of the ring of invariant polynomials in $d$ variables. This method goes back to Cayley, Hilbert, Noether .... However this approach has a drawback: it cannot produce bi-Lipschitz embeddings [1], unless special cases.

[1] J. Cahill, A. Contreras, A.C. Hip, Complete Set of translation Invariant Measurements with Lipschitz Bounds, ACHA 2020

# Sorting based Representations

Different approaches were considered recently [2],[3],[4] based on sorting. A unified framework for these approaches is presented here.

Fix a generator $w \in V$ (call it, window or template) and consider the nonlinear map induced by sorting its coorbit:

$$\phi_w : V \to \mathbb{R}^N \quad , \quad \phi_w(x) = \downarrow ((\langle x, U_g w \rangle)_{g \in G}) .$$

where $\downarrow(y) = (y_{\pi(i)})_{i \in [N]}$ is the non-increasing sorting operator: $y_{\pi(1)} \geq \cdots \geq y_{\pi(N)}$.

[2] R. Balan, N. Haghani, M.Singh, Permutation Invariant Representations with Applications to Graph Deep Learning, arXiv:2203.07546 (2022)

[3] N. Dym, S. J. Gortler, Low Dimensional Invariant Embeddings for Universal Geometric Learning, arXiv:2205.02956 (2022)

[4] J. Cahill, J. W. Iverson, D. G. Mixon, D. Packer, Group-invariant max filtering, arXiv:2205.14039 (2022)

# Representations based on sorting (2)

$$\phi_w : V \to \mathbb{R}^N \quad , \quad \phi_w(x) = \downarrow \left( (\langle x, U_g w \rangle)_{g \in G} \right).$$

Remarks:

1. $\phi_w(U_g x) = \phi_w(x)$ for every $g \in G$ and $x \in V$. Thus $\phi_w$ lifts to the quotient space $\widehat{V}$.
2. Invariant polynomials, and more generally, invariant functions obtained by the averaging operator (the Reynolds operator), can be obtained as:

$$K \mapsto F_K(x) = \frac{1}{|G|} \sum_{g \in G} K(\langle U_g x, w \rangle) = \frac{1}{|G|} \sum_{g \in G} K(\phi_w(x))$$

# Invariant Coorbit Representations

For a collection $\mathbf{w} = (w_1, \cdots, w_p) \in V^p$ let

$$\Phi_{\mathbf{w}} : V \to \mathbb{R}^{N \times p} \quad , \quad \Phi_{\mathbf{w}}(x) = \left[ \phi_{w_1}(x) | \cdots | \phi_{w_p}(x) \right] .$$

For a subset $S \subset [N] \times [p]$ of cardinal $m = |S|$, let

$$\Phi_{\mathbf{w},S} : V \to l^2(S) \sim \mathbb{R}^m \ , \ \Phi_{\mathbf{w},S}(x) = (\Phi_{\mathbf{w}}(x))|_S$$

be the restriction of $\Phi_{\mathbf{w}}$ to $S$. For a linear operator $\mathcal{L} : l^2(S) \to \mathbb{R}^m$, let

$$\Psi_{\mathbf{w},S,\mathcal{L}} : V \to \mathbb{R}^m \quad , \quad \Psi_{\mathbf{w},\mathcal{L}}(x) = \mathcal{L}(\Phi_{\mathbf{w},S}(x))$$

be the "projection" of $\Phi_{\mathbf{w},S}$ through $\mathcal{L}$ into $\mathbb{R}^m$.
Problems: Construct $(\mathbf{w}, S)$ so that $\Phi_{\mathbf{w},S}$ is a bi-Lipschitz embedding of $\widehat{V}$. Construct $(\mathbf{w}, S, \mathcal{L})$ so that $\Psi_{\mathbf{w},S,\mathcal{L}}$ is bi-Lipschitz.

# Invariant Coorbit Representations

Special cases:
1. If $G = S_n$ and $V = \mathbb{R}^{n \times d}$ with action $(P, X) \mapsto PX$, then [5] introduced the embedding $\beta_A(X) = \downarrow(XA)$, for *key* $A \in \mathbb{R}^{d \times D}$ and sorting operator acting independently in each column.
Equivalent recasting: Let $w_1 = \delta_1 \cdot a_1^T, \ldots, w_D = \delta_1 \cdot a_D^T$, where $\delta_1 = (1, 0, \cdots, 0)^T$ and $A = [a_1 | \cdots | a_D]$. Then note $\phi_{w_1}(X) = \downarrow(Xa_1) \otimes 1_{(n-1)!}$. Thus $\Phi_{\mathbf{w}}(X) = \beta_A(X) \otimes 1_{(n-1)!}$. Thus $\beta_A(X) = \Phi_{\mathbf{w},S}(X)$ for an appropriate subset $S \subset [n!] \times [D]$ of size $nD$.
2. The *max filter* introduced in [6] for some template $w \in V$ is defined by $\langle\langle \cdot, w \rangle\rangle : V \to \mathbb{R}$, $\langle\langle x, w \rangle\rangle = \max_{g \in G} \langle x, U_g w \rangle$. Equivalent recasting: $\langle\langle x, w \rangle\rangle = \Phi_{w,S}(X)$, for $S = \{1\}$.

[5] R. Balan, N. Haghani, M.Singh, Permutation Invariant Representations with Applications to Graph Deep Learning, arXiv:2203.07546 (2022)
[6] J. Cahill, J. W. Iverson, D. G. Mixon, D. Packer, Group-invariant max filtering, arXiv:2205.14039 (2022)

# Sufficient conditions for an injective embedding

## Theorem

*Consider $G$ finite group of size $N$ acting unitarily on the d-dimensional $V$. Let $\mathbf{w} \in V^p$, $S \subset [N] \times [p]$, $S_k$ the $k^{th}$ slice, and linear map $\mathcal{L} : l^2(S) \to \mathbb{R}^m$. Denote $\gamma_2 = \min_{g \in G, g \neq 1} \min_{\lambda \in \mathbb{R}} rank(\lambda I_d - U_g)$, $\gamma_3 = \max_{g \in G, g \neq 1} \min_{\lambda \in \mathbb{R}} rank(\lambda I_d - U_g)$. Then for almost every $\mathbf{w}$ and $\mathcal{L}$ the maps $\Phi_{\mathbf{w},S}$ or $\Psi_{\mathbf{w},S,\mathcal{L}}$ are injective on $\widehat{V}$ in any of the following cases:*

1. *(Max filter, Cahill et.al. 2022) If $p \geq 2d$ and $S_{max} = \{(1,1), \cdots, (1,p)\}$ then the max filterbank $\Phi_{\mathbf{w},S_{max}}$ is injective for a.e. $\mathbf{w} \in V^p$.*
2. *(variation of previous result) If $p \geq 2d$ and $|S_k| \geq 1$ for all $k \in [p]$ then $\Phi_{\mathbf{w},S}$ is injective for a.e. $\mathbf{w} \in V^p$.*
3. [a] *If $G$ is a reflection group and $p \geq d$ then the max filterbank $\Phi_{\mathbf{w},S_{max}}$ is injective for a.e. $\mathbf{w} \in V^p$.*

---

[a] D. Mixon, Y. Qaddura, Injectivity, stability, and positive definiteness of max filtering, arXiv:2212.11156

# Sufficient conditions for injective embedding (cont)

**Theorem**

4. *If* $p \geq 2d - \gamma_2$, $|S| \geq 2d$, *and for each* $k$, $|S_k| \in \{1,2\}$ *then* $\Phi_{\mathbf{w},S}$ *is injective for a.e.* $\mathbf{w} \in V^p$.
5. *If* $2d - \gamma_3 \leq p \leq 2d$, $|S_1| = \cdots = |S_{2d-p}| = N$, *and* $|S_{2d-p+1}| = \cdots = |S_p| = 1$ *then* $\Phi_{\mathbf{w},S}$ *is injective for a.e.* $\mathbf{w} \in V^p$.
6. *If* $\Phi_{\mathbf{w},S}$ *is injective and* $m \geq 2d$ *then the map* $\Psi_{\mathbf{w},S,\mathcal{L}}$ *is injective for a.e. linear map* $\mathcal{L} : l^2(S) \to \mathbb{R}^m$.

Remark:
This result can be extended to the case when $S$ has an irregular structure. However this requires some involved spectral conditions.

# Injectivity – sketch of proof

The proof provides a semi-algebraic characterization of the set of "bad" windows, i.e., windows $\mathbf{w}$ that fail to separate, say $\mathcal{F}$.

$$\mathcal{F} \subset \bigcup_{\mathbf{g},\mathbf{h} \in G^{p*}} \mathcal{F}_{\mathbf{g},\mathbf{h}} \ , \quad G^{p*} = \{(g_i^k)_{(i,k)\in S} \ , \ \forall k, (g_i^k)_{i\in S_k} \in G^{|S_k|} \text{are distinct}\}$$

$$\mathcal{F}_{\mathbf{g},\mathbf{h}} = \bigcup_{(x,y)\in\Gamma} \otimes_{k=1}^{p} \{U_{g_1^k}x - U_{h_1^k}y, \ldots, U_{g_{m_k}^k}x - U_{h_{m_k}^k}y\}^{\perp}$$

where $\Gamma = \{(x,y) \in V^2 \ : \ x \not\sim y, \|x\|^2 + \|y\|^2 = 1\}$, $m_k = |S_k|$. Using the "lift-and-project" technique, we realize each $\mathcal{F}_{\mathbf{g},\mathbf{h}}$ as finite unions of projection onto second term of total manifolds of certain real-analytic vector bundles. The vector bundles have as base manifolds subsets of $\Gamma$ where dimension of the orthogonal complement of constant. In turn those subsets are controled by spectral properties of $U_g$'s.

# Injectivity – sketch of proof

The base manifolds of these vector bundles are themselves total spaces of a different vector bundles living over Grassmanian manifolds. For instance, for $|S_k| = m_k = 2$, First construct the bundle $(Gr(1,\mathbb{R}^2),\pi,E)$ over $Gr(1,\mathbb{R}^2) = \mathbb{RP}^1 \sim [0,\pi)$ with total space

$$E = \{(\theta,x,y) \in [0,\pi)\times V^2 \; ; \; cos(\theta)(U_{g_1}x - U_{h_1}y) + sin(\theta)(U_{g_2}x - U_{h_2}y) = 0\}$$

Most of fibers are $d$-dimensional except what $tan(\theta)$ is an eigenvalue of some unitary $U_g$. Those two cases induce a disjoint partition $\Gamma = (\Gamma \setminus \Pi_2(E)) \cup (\Gamma \cap \Pi_2(E))$ so that

$$\begin{array}{rcl} (x,y) \in \Gamma_1 := \Gamma \setminus \Pi_2(E) & \rightarrow & \dim\{U_{g_1}x - U_{h_1}y), (U_{g_2}x - U_{h_2}y\}^{\perp} = d-2 \\ (x,y) \in \Gamma_2 := \Gamma \cap \Pi_2(E) & \rightarrow & \dim\{U_{g_1}x - U_{h_1}y), (U_{g_2}x - U_{h_2}y\}^{\perp} = d-1 \end{array}$$

from where the dimension estimates arise.

# Main Result

## Theorem

*Consider $G$ finite group of size $N$ acting unitarily on the $d$-dimensional $V$. Let $\mathbf{w} \in V^p$, $S \subset [N] \times [p]$ and $\mathcal{L} : l^2(S) \to \mathbb{R}^m$. Let*

$$B = \max_{\substack{\sigma_1, \cdots, \sigma_p \subset G, \\ |\sigma_k| = |S_k|, \forall k}} \lambda_{max}\left(\sum_{k=1}^{p} \sum_{g \in \sigma_k} U_g w_k w_k^T U_g^T\right)$$

*where $S_k = \{i \in [N], (i,k) \in S\}$ for each $k \in [p]$.*

1. $\Phi_{\mathbf{w},S} : (\widehat{V}, \mathbf{d}) \to l^2(S)$ *is Lipschitz with constant upper bounded by* $\sqrt{B}$.
2. *If* $S = [N] \times [p]$ *and* $\Phi_{\mathbf{w},S} : (\widehat{V}, \mathbf{d}) \to (\mathbb{R}^m, \|\cdot\|_2)$ *is injective then it is also bi-Lipschitz;*
3. *If* $S = [N] \times [p]$ *and* $\Phi_{\mathbf{w},S} : (\widehat{V}, \mathbf{d}) \to (\mathbb{R}^m, \|\cdot\|_2)$ *is injective then for a generic* $\mathcal{L}$ *with* $m \geq 2d$, *the map* $\Psi_{\mathbf{w},S,\mathcal{L}} : (\widehat{V}, \mathbf{d}) \to (\mathbb{R}^m, \|\cdot\|_2)$ *is injective and bi-Lipschitz.*

# Sketch of Proof

1. The upper Lipschitz bound is not too hard. A quick way to obtain it is by the Fundamental Theorem of Calculus: Fix $x, y \in V$ and choose them so that $\mathbf{d}([x],[y]) = \|x-y\|$. The function $f:[0,1] \to l^2(S)$, $f(t) = \Phi_{\mathbf{w},S}((1-t)x+ty)$ is Lipschitz because the sorting operator $\downarrow$ is Lipschitz. The upper Lipschitz constant is computable from FTC and Lebesgue's differentiation theorem:

$$\|f(1)-f(0)\|_2 = \| \int_0^1 (Jf)|_{(1-t)x+ty}(y-x)dt\| \leq sup_z \|J\Phi_{\mathbf{w},S}(z)\|_\infty \mathbf{d}([x],[y])$$

But wherever $\Phi$ is differentiable, $J\Phi_{\mathbf{w},S}(z) = \left[ \left( U_{g(\pi_k(i)} w_k \right)^T \right]_{(i,k)\in S}$ where $\pi_k$ is the permutation that sorts $\phi_{w_k}(z)$. From here one obtains the upper bound.

The same goes for $\Psi_{\mathbf{w},S,\mathcal{L}}$.

# Sketch of Proof (2)

2. The lower Lipschitz bound is more challanging. The proof follows the recipe from (Balan et.al. 2022) and is by contradiction.

Assume the lower bound is $A = 0$.

Step 1. A compactness argument together with the homogeneity of map $\Phi$ implies the local lower Lipschitz constant must vanish: there is $z \in S_1(V)$:

$$\lim_{r \downarrow 0} \inf_{\substack{x \not\sim y \\ \mathbf{d}([x],[z])<r, \mathbf{d}([y],[z])<r}} \frac{\|\Phi_{\mathbf{w},S}(x) - \Phi_{\mathbf{w},S}(y)\|_2}{\mathbf{d}([x],[y])} = 0.$$

Step 2. Construct sequences $(x_n)_n$ and $(y_n)_n$ so that: (i) $\|x_n\| = 1$, (ii) $\|y_n\| \leq 1$; (iii) $\mathbf{d}([x_n],[y_n]) = \|x_n - y_n\|$; (iv) $\mathbf{d}([x_n],[z]) = \|x_n - z\|$; and $x_n \to z$, $[y_n] \to [z]$, and $y_n \to y_\infty$.

Step 3. Let $H = \{g \in G : U_g z = z\}$ denote the stabilizer of $z$. Let $\Delta_0 = \min_{g \in G \setminus H} \|z - U_g z\| > 0$. Assume $n$ large enough so that $u_n = x_n - z$, $v_n = y_n - z$ satisfy $\|u_n\|, \|v_n\| < \frac{1}{4}\Delta_0$. This forces $y_\infty = z$.

# Sketch of Proof (2)

Step 4. By finiteness of $G$, we extract subsequences so that $(\Phi_{\mathbf{w},S}(x_n))_{i,k} = \langle x_n, U_{g(1,i,k)} w_k \rangle$ and $(\Phi_{\mathbf{w},S}(y_n))_{i,k} = \langle y_n, U_{g(2,i,k)} w_k \rangle$ (note the group elements are independent on n !). It follows:

$$\lim_{n\to\infty} \frac{1}{\|u_n - v_n\|^2} \sum_{(i,k)\in S} |\langle w_k, U^T_{g(1,i,k)} u_n - U^T_{g(2,i,k)} v_n \rangle|^2 = 0$$

Step 5. Using an argument about ratios of quadratics, it follows that one is able to produce $u, v$ so that $u \not\sim v$ and $\langle w_k, U^T_{g(1,i,k)} u - U^T_{g(2,i,k)} v \rangle = 0$ for all $(i,k) \in S$. Then for $s > 0$ small enough, $x = z + su$ and $y = z + sv$ we have $\mathbf{d}([x],[y]) > 0$ and yet $\Phi_{\mathbf{w},S}(x) = \Phi_{\mathbf{w},S}(y)$. Contradiction!

# Table of Contents

Thank you!

Questions?