

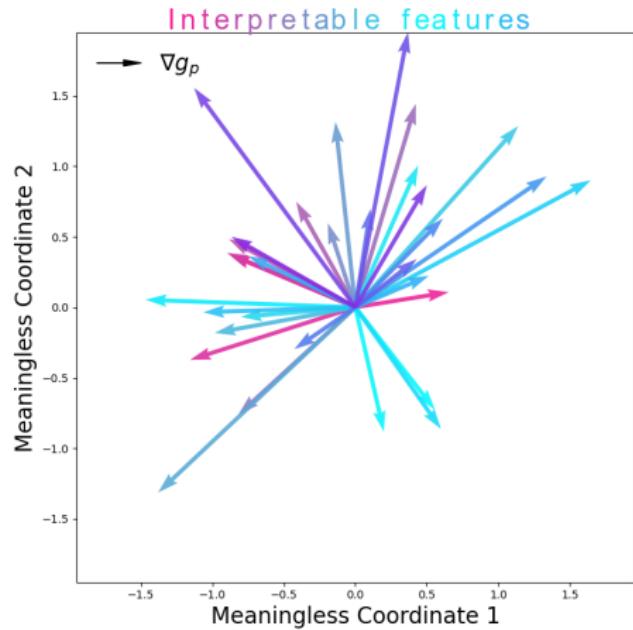
# Isometry pursuit



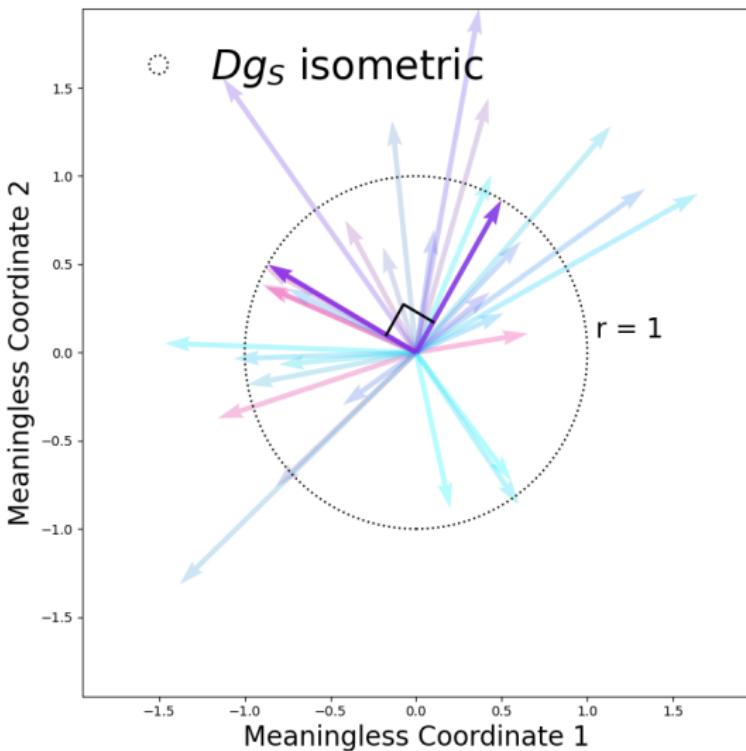
Samson Koelle

April 17, 2025

# The problem



# Isometry selection



# Yet another matrix loss

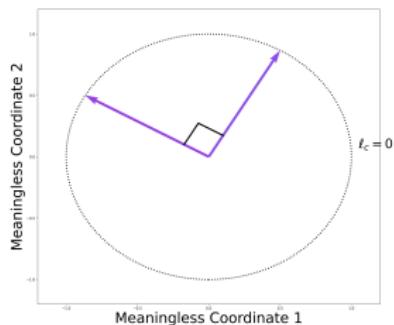
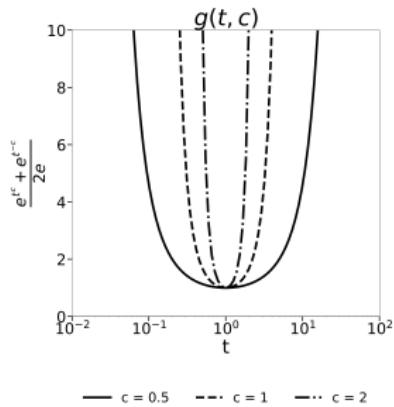
Given a rank  $D$  matrix  $X \in \mathbb{R}^{D \times P}$  with singular values  $\sigma_d \in [D]$ , let

$$\ell_c : \mathbb{R}^{D \times P} \rightarrow \mathbb{R}^+$$

$$X \mapsto \sum_{d=1}^D g(\sigma_d(X), c) - D$$

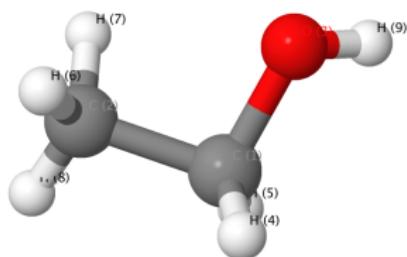
$$g : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$$

$$(t, c) \mapsto \frac{e^{t^c} + e^{t^{-c}}}{2e}.$$



# An application in molecular dynamics simulation (MDS)

- Given a molecule consisting of  $N_a$  atoms, simulate  $T$  timesteps of molecular motion to generate data  $\mathcal{D} \in \mathbb{R}^{T \times 3N_a}$
- These samples concentrate near a low-dimensional manifold which we can estimate using ML



Ethanol ( $N_a = 9$ ) (**Chmiela2018-at**)

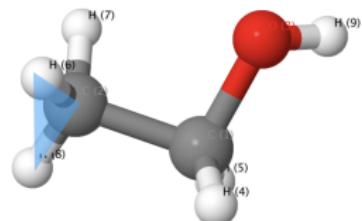
# Representing invariant geometry

- We featurize each configuration using planar angles prior to ML  
**(chen2019modern)**

$$\mathcal{M} \xrightarrow{\alpha} \alpha(\mathcal{M}) \xrightarrow{\phi} \phi(\alpha(\mathcal{M}))$$

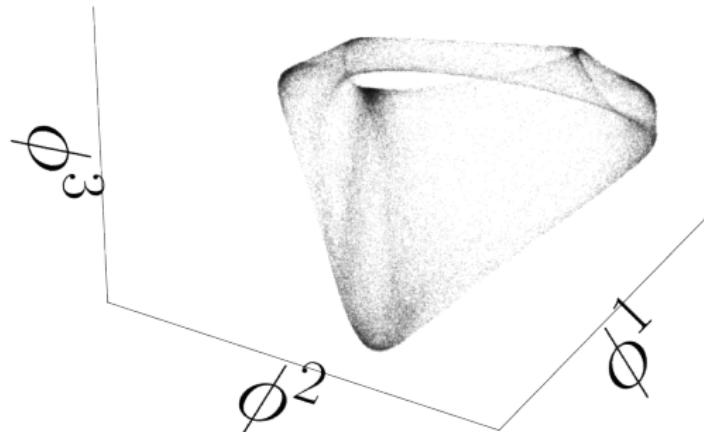
$\mathbb{R}^{3N_a} \quad \cup \quad \mathbb{R}^{3\binom{N_a}{3}} \quad \cup \quad \mathbb{R}^M$

$\alpha$  is the planar angle map



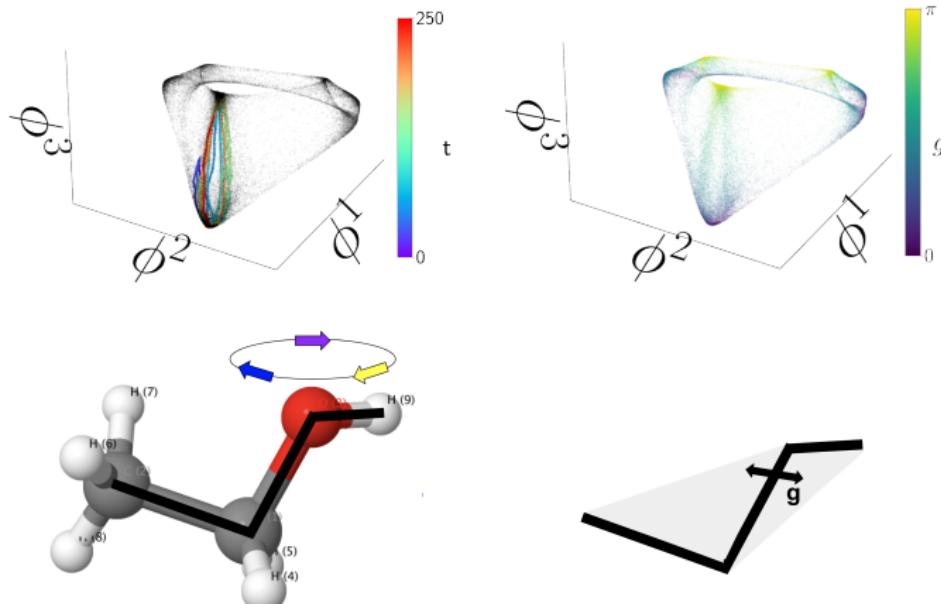
Planar angles are formed by triplets of atoms

# The learned manifold



Diffusion Maps embedding of  $T = 50000$  ethanol configurations into  $M = 3$  dimensions shows a 2 dimensional toroidal molecular manifold

# Visually interpreting embeddings

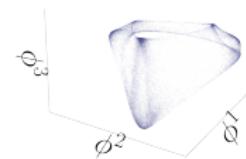
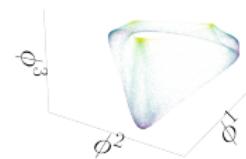
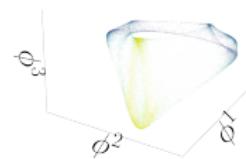
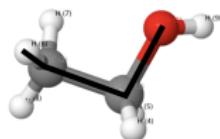


Movement in the ethanol simulation corresponds to hydroxyl rotation

# Interpretability is important

- Mechanistic understanding
- Generative control
- Statistical efficiency

# Scaling interpretability

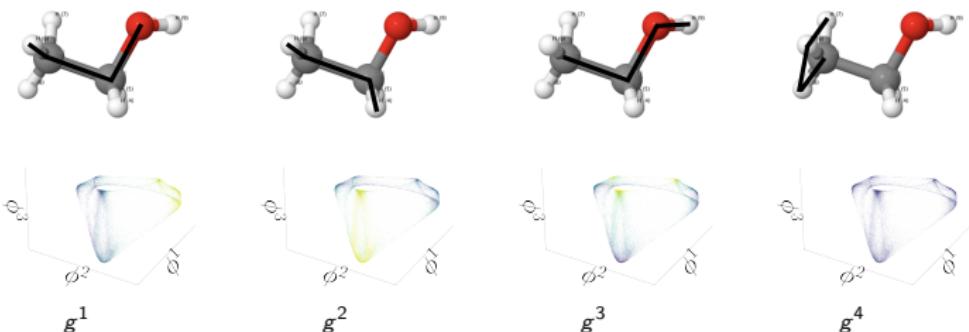


4 of the 756 bond torsions in ethanol

# Explanations

## Definition (Explanation)

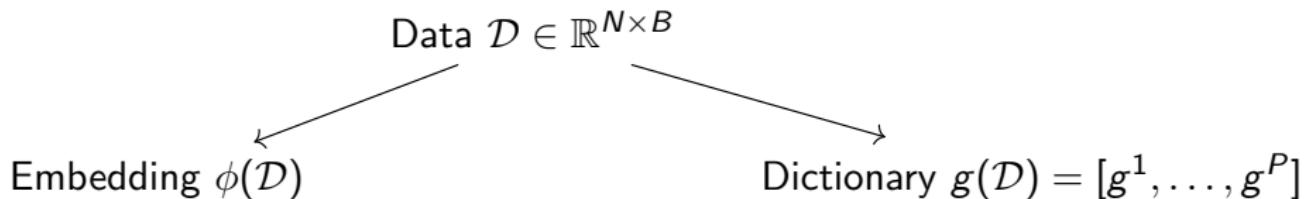
$D$  functions  $g^p$  form a *explanation* if their gradients form rank  $D$  linear maps from  $\mathbb{R}^D \rightarrow \mathbb{R}^D$ .



	$g^1$	$g^2$	$g^3$	$g^4$
$g^1$	x	x	✓	x
$g^2$	x	x	✓	x
$g^3$	✓	✓	x	x
$g^4$	x	x	x	x

Are these explanations?

# Interpretability examples



Data	Featurization	Embedding	Dictionary	Explanation
Configurations	Planar angles	UMAP	Dihedral angles	Methyl rotation
Sentences	Tokens	LLM embeddings	Sparse autoencoder	Time of day
Cells	Gene expression	UMAP	Gene sets	Cell cycle
Galaxies	Spectra	Diffusion maps	Stellar features	Age

Examples of explanations in various scientific domains

# Problem statement

- Our goal: find  $D$  function subset of  $g$  whose gradients w.r.t.  $\mathcal{M}$  form a rank  $D$  matrix almost everywhere

## Proposition (Cut-locus theorem (**Sheng2009-ij**))

*Every manifold  $\mathcal{M}$  has an explanation defined everywhere except for a singular set.*

- **Koelle2022-obj:** Provide a sparse regression estimator for selecting explanations w.r.t.  $\phi$
- **Koelle2024-no:** Select explanations without  $\phi$