Disentangling by Subspace Diffusion

David Pfau, Irina Higgins, Aleksandar Botev and Sébastien Racanière
{pfau, irinah, botev, sracaniere}@google.com

DeepMind

Background and Motivation

Can we learn to **disentangle** independent factors of variation in the world, e.g. pose, illumination, etc [Bengio et al 2013]?

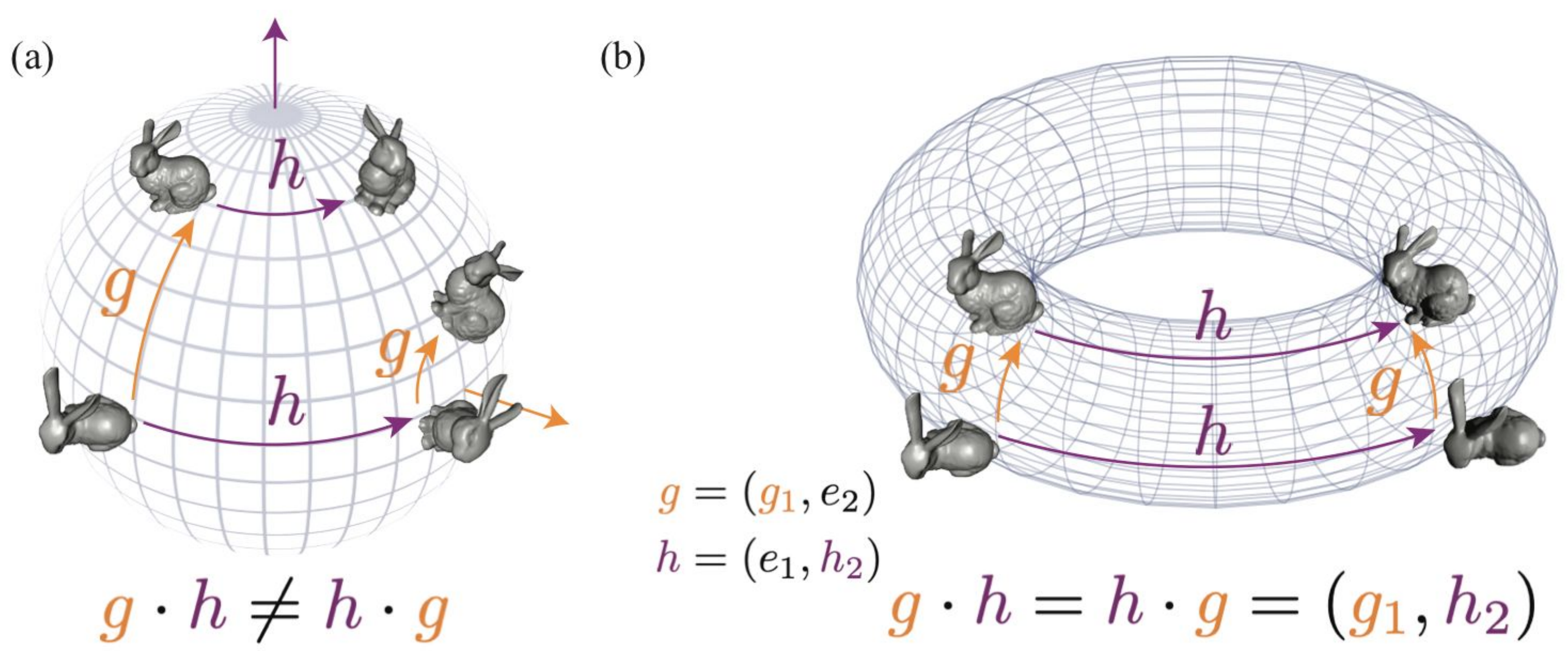
Probabilistic disentangling [Locatello et al. 2019]:
*Latent vectors are sampled from a **product of independent distributions**. A representation is disentangled if it correctly recovers the **statistically independent** latent factors.*
Pessimistic result - disentangled directions are not identifiable without some prior knowledge [Hyvarinen and Pajunen 1999]

Symmetry-based or ***Geometric*** disentangling [Higgins et al. 2018]:
*Latent vectors are generated by a **product of group actions**. A representation is disentangled if the representation space can be partitioned into subspaces that are **invariant to all group actions** except one.*
Optimistic definition - not yet known under what conditions it is possible, largely novel research direction.

Elevator summary:

Fully unsupervised symmetry-based disentangling is possible
***if* we have access to true metric information.**
We develop an algorithm that achieves this:
the *Geometric Manifold Component Estimator* (GEOMANCER)

Analogical Reasoning on Curved Manifolds

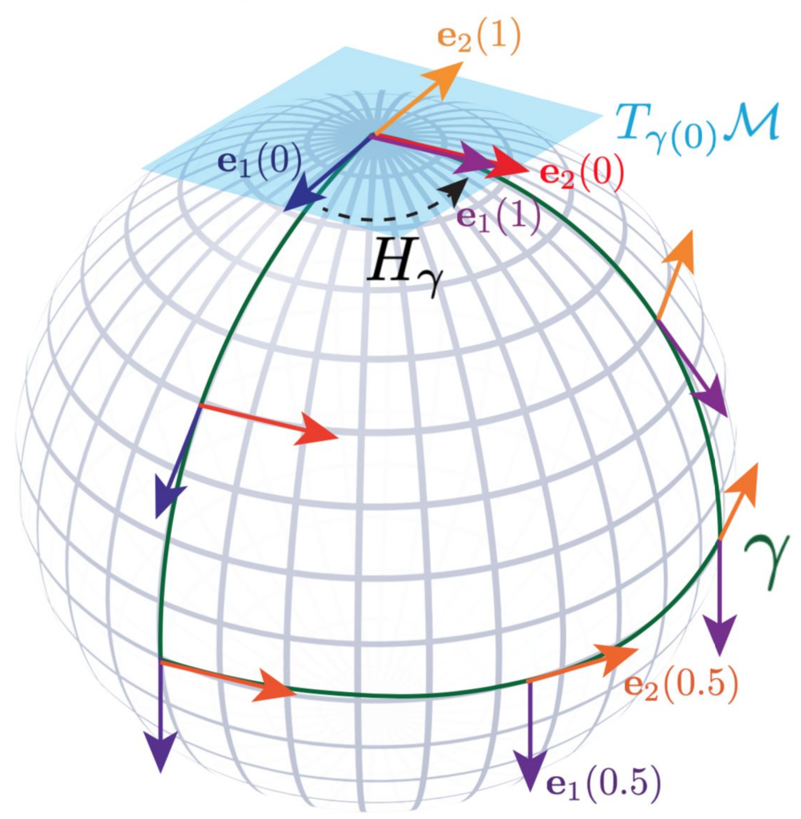


(a) On curved manifolds, analogical reasoning breaks down because group actions do not commute.
(b) On a **product** of curved manifolds, analogical reasoning is possible only along certain dimensions.

The noncommutativity of certain operations is **used as a learning signal** to find disentangled directions on a data manifold.

Code: `tinyurl.com/dm-geomancer`

Holonomy and the de Rham decomposition



Tangent space: local vector space T_xM around point x
Connection: defines how the vector **v** changes when moved in the direction **w**
Parallel transport: sequence of vectors **v**(t) moved along the path γ(t)
Holonomy: Matrix H_γ that gives change to a vector transported around the loop γ

Tangent space decomposes into subspaces left invariant by holonomy → (de Rham 1952) → Manifold factorizes into product of submanifolds
← (trivial)

Method

Subspace Diffusion

Compute subspaces that are ***nearly invariant under random walk diffusion***

Scalar Laplacian [Belkin and Niyogi 2003]

$$\Delta^0[f]_i = \sum f_i - f_j$$

Vector Laplacian [Singer and Wu 2012]

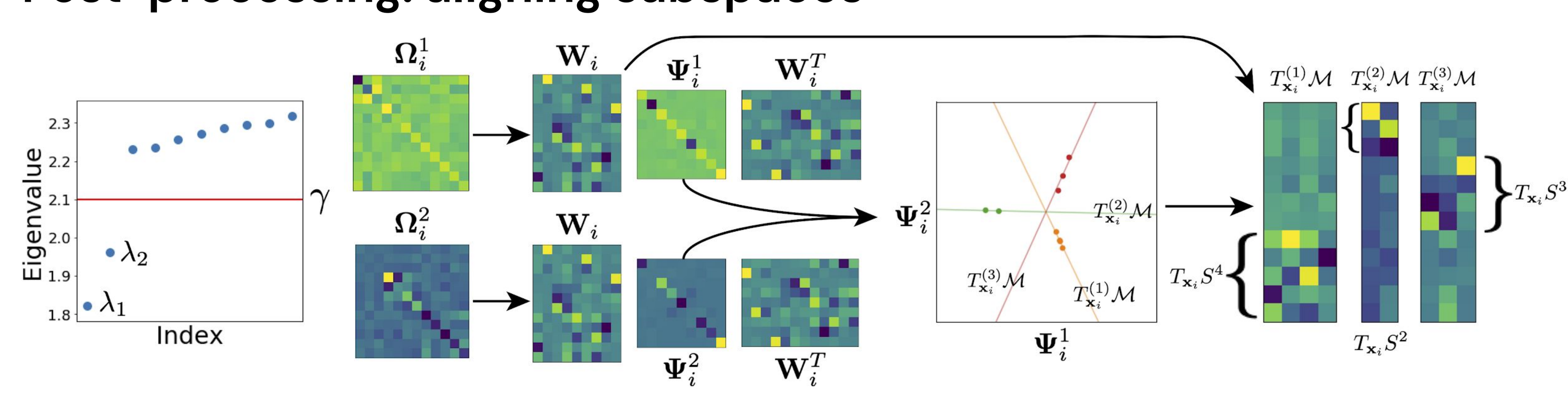
$$\Delta^1[\mathbf{v}]_i = \sum \mathbf{v}_i - \mathbf{Q}_{ij}^T \mathbf{v}_j$$

Matrix Laplacian

$$\Delta^2[\mathbf{\Sigma}]_i = \sum \mathbf{\Sigma}_i - \mathbf{Q}_{ij}^T \mathbf{\Sigma}_j \mathbf{Q}_{ij}$$

Generalizes **Laplacian Eigenmaps** and **Vector Diffusion Maps** to matrices and subspaces, but for an **entirely novel application**

Post-processing: aligning subspaces



Take the eigenvectors of the matrix Laplacian, reshape into set of matrices, simultaneously diagonalize matrices around each point, and partition tangent space around each point into a set of orthogonal subspaces.

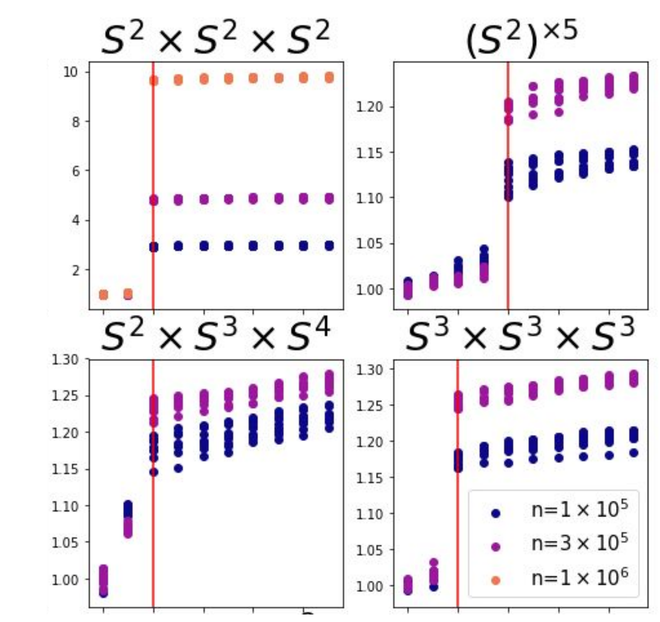
Data: `tinyurl.com/dm-s3o4d`

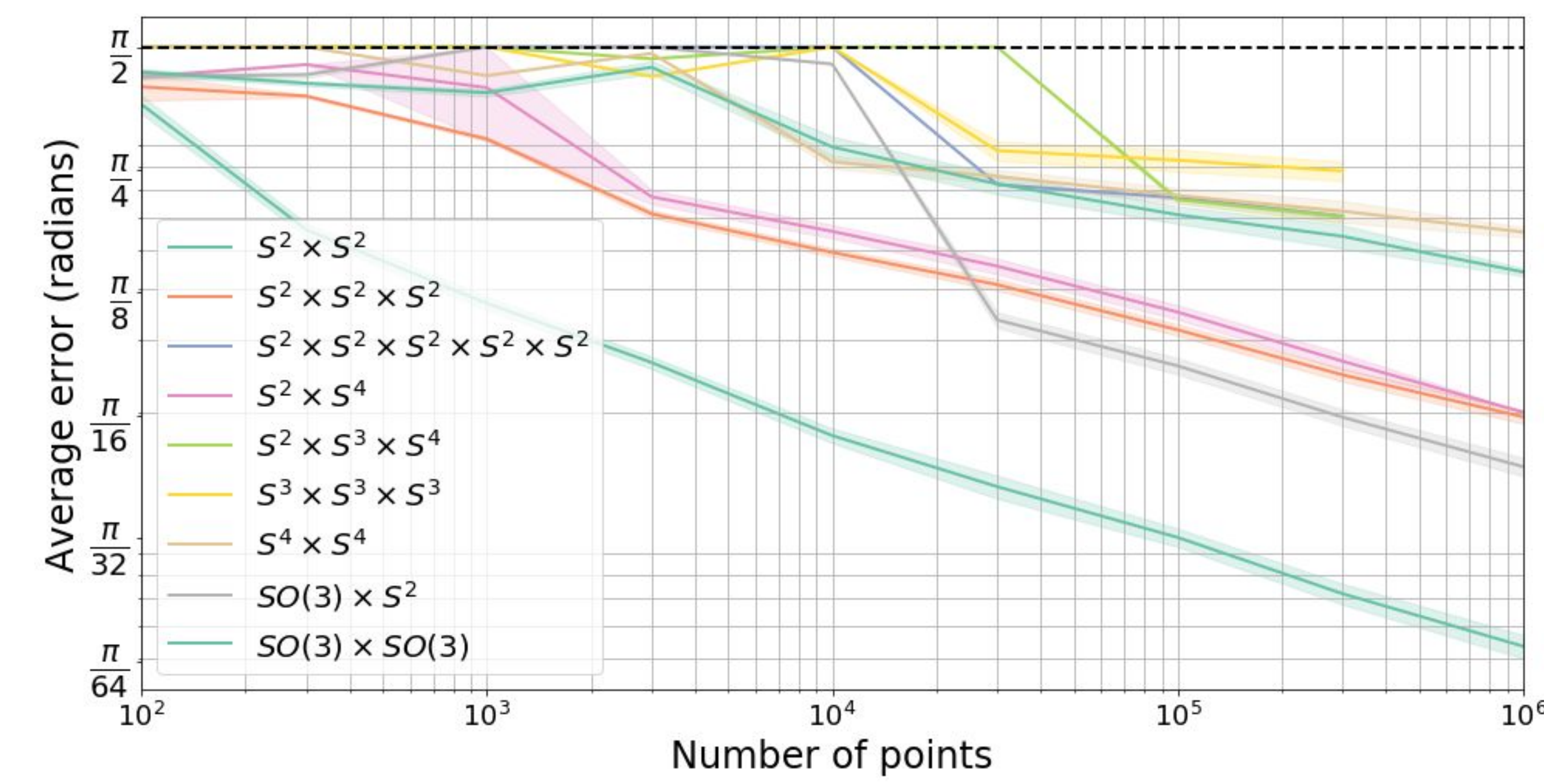
Results

Synthetic Manifolds

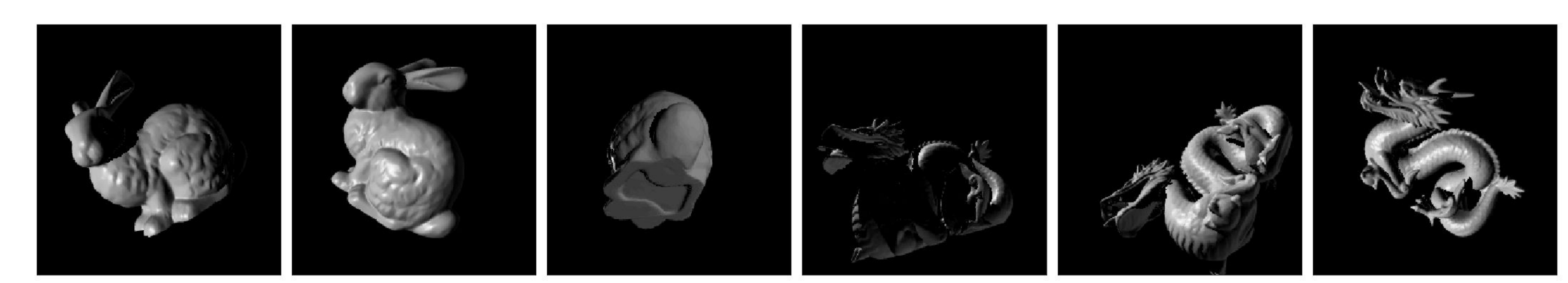
Arbitrary product of up to 5 submanifolds (**spheres** and **rotations**)

- Correctly recovers **# of manifolds**
- Correctly recovers **decomposition of subspaces**
- Performance jumps above chance quickly past a critical threshold in number of training examples





Stanford 3D Objects for Disentangling (S3O4D)



100k renderings for each object from the Stanford 3D Scanning Repository with uniformly sampled illumination (S^2) and pose (**SO(3)**)

- Metric information is ***necessary*** - GEOMANCER fails on pixels
- Metric information is ***sufficient*** - GEOMANCER works on embeddings
- Existing disentangling algorithms are ***insufficient*** - β-VAE fails on pixels

References

Y. Bengio, A. Courville, P. Vincent (2013). "Representation learning: A review and new perspective." IEEE TPAMI, 35(8):1798-1828

F. Locatello et al. (2019). "Challenging common assumptions in the unsupervised learning of disentangled representations.". ICML, PMLR 97:4114-4124

A. Hyvarinen and P. Pajunen (1999). "Nonlinear Independent Component Analysis: Existence and uniqueness results." Neural Networks 12(3):429–439

I. Higgins, D. Amos et al. (2018). "Towards a definition of disentangled representations." arXiv:1812.02230

G. de Rham (1952). "Sur la réductibilité d'un espace de Riemann." Commentarii Mathematici Helvetici 26(1):328-344

M. Belkin and P. Niyogi (2003). "Laplacian eigenmaps for dimensionality reduction and data representation learning." Neural Computation 15(6):1373-1396

A. Singer and H. T. Wu (2012). "Vector diffusion maps and the connection Laplacian." Communications on Pure and Applied Mathematics 65(8):1067-1144

I. Higgins et al. (2017). "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework" ICLR