Towards a Definition of Disentangled Representations

Authors: Irina Higgins, David Amos, David Pfau, Sebastien Racaniere, Loic Matthey, Danilo Rezende and Alexander Lerchner

Presented by: Sarthak Mittal

October 21, 2022

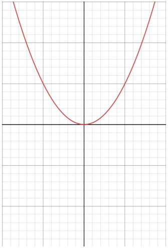
Outline

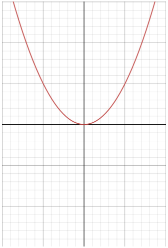
- Motivation
- Group Theory
- Disentangled Representation
- Linear Disentangled Representation
- Conclusion

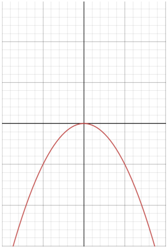
- Group Representation Theory
- Compatibility

Symmetries

A symmetry is a transformation that leaves certain properties of the object invariant.







We exploit different symmetries in the world!

Symmetries

The success of AI systems is quite largely attributed to symmetries

- **CNNs**: Spatial
- **RNNs**: Temporal
- **Transformers**: Permutation

Here, we know what symmetries we are going for.[1]

[1] What if we want to learn that as well?

Noether's Theorem

Informally, every conservation law is grounded in a corresponding symmetry transformation.

Conservation of

- **Energy**: Time Translation Symmetry
- **Momentum**: Space Translation Symmetry
- **Angular Momentum**: Rotational Symmetry

Cool Note: Gell-Mann predicted the existence of a particle Ω^- based on certain symmetries of its quarks. It was actually observed two years later.

Disentanglement

What is disentanglement?

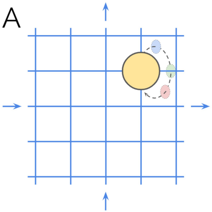
It is learning of a representation that reflects the *symmetries* of the world in the corresponding representations.

Example

Consider an object in a grid. It can

- Move horizontally
- Move vertically
- Move on the color axis

These actions are called the symmetry transformations of this world

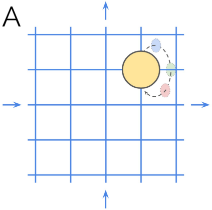


Example

Disentanglement means learning a representation that follows these transformations

State: (x, y, c)

Suppose cyclic translations according to each symmetry leads to cyclic translations in corresponding state.



Disentangled Representation

A vector representation is called a **disentangled representation** *with respect to a particular decomposition of a symmetry group into subgroups, if it decomposes into independent subspaces, where each subspace is affected by the action of a single subgroup, and the actions of all other subgroups leave the subspace unaffected.*

If the actions of all the subgroups are linear, then we get **linear disentangled representation**

Group Theory

A *group* $(G, \circ)$ is a set G with the binary operation $\circ : G \times G \to G$ which satisfies

- Associativity: $\forall x, y, z \in G : x \circ (y \circ z) = (x \circ y) \circ z$
- Identity: $\exists e, \forall x \in G : e \circ x = x \circ e = x$
- Inverse: $\forall x \in G, \exists x^{-1} \in G : x \circ x^{-1} = x^{-1} \circ x = e$

Example of Groups

- $GL(n, \mathbb{R})$: the set of $n \times n$ invertible matrices with the operation as matrix multiplication
- S_n the set of permutations of the numbers $\{1, ..., n\}$ with the operation as function composition
- Countless other examples...

Sub-Groups

Given a group G, one can find a subset H such that

- $\forall x, y \in H : x \circ y \in H$
- $\forall x \in H : x^{-1} \in H$

That is, H is a group that is contained in G under the same operation.

Eg. $SO(n)$ is the group of orthogonal matrices with determinant 1, is a sub-group of $GL(n, \mathbb{R})$

Direct Product of Groups

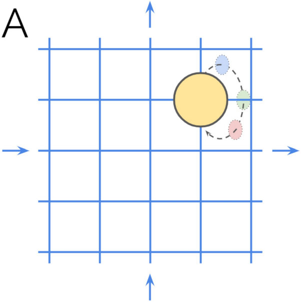
Given two groups $(G, \circ_g)$ and $(H, \circ_h)$, one can create a new group $G \times H$ where

- The underlying set is (g_1, h_1) with $g_1 \in G, h_1 \in H$
- Group operation is defined as: $(g_1, h_1) \circ (g_2, h_2) = (g_1 \circ_g g_2, h_1 \circ_h h_2)$

You can see that $G \times H$ has a sub-group $G \times e_h$.

Informal Teaser

Given a group G, if it admits a decomposition $G_1 \times \ldots \times G_n$, it will play some role in disentanglement.



Group Action

Groups can often be seen as some transformations of a space

- An element of $SO(3)$ can act on 3-D space by rotating it
- An element of S_n can permute the contents of an ordered set.

Group Action

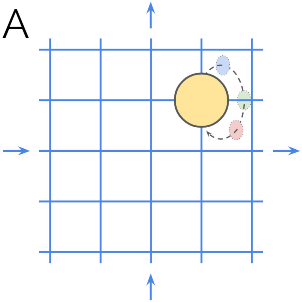
A group action on a set X is a function $\odot : G \times X \to X$ such that

- $e \odot x = x \; \forall x \in X$
- $(g \circ h) \odot x = g \odot (h \odot x) \; \forall g, h \in G, x \in X$

When a group acts on a space with some structure (eg. vector space), then it is required that the action preserves that structure (eg. linearity). *(Why?)*

Informal Teaser

Given a group $G = (G_1 \times ... \times G_n)$ with action $\odot$ on some space X, the action $\odot$ is disentangled if X decomposes as $X_1 \times ... \times X_n$ and there are actions $\odot_i$ such that G_i acts on X_i with action $\odot_i$ while leaving other X_j's unchanged.



Disentangled Group Action

Suppose we have a group action $\odot : G \times X \to X$ and the group G decomposes as $G_1 \times G_2$. The action $\odot$ is called disentangled if there exists a decomposition $X = X_1 \times X_2$ and actions $\odot_i : G_i \times X_i \to X_i$ such that

$$(g_1, g_2) \odot (v_1, v_2) = (g_1 \odot_1 v_1, g_2 \odot_2 v_2)$$

This extends to an arbitrary decomposition $G_1 \times ... \times G_n$.

Recall

- **Group**: An abstract structure
- **Group Action**: A way of a group affecting a space
- **Disentangled Group Action**: A decomposition of space and action that follows the group decomposition

Disentangled Representation

To generate a general scene, we have:

- W: The set of world-states.
- O: The observed space.
- $b : W \to O$: The generative process.

Typically, our systems do not have access to W, and hence we rely on an inference procedure $h : O \to Z$ where Z is the representation.

Disentangled Representation

Lets consider the function $f : W \to Z$ as $h \circ b$.

Assume we are also given a group G of symmetries which acts on W via action $\odot_w$ and on Z via action $\odot_z$.

Equivariant Map

The function f is an equivariant map if given the group G and actions $\odot_w$ and $\odot_z$

$$g \odot_z f(w) = f(g \odot_w w) \qquad \forall g \in G, w \in W$$

Equivariant Map

Remember, to talk about equivariant-ness of a map, we need some ingredients:

- A group G.
- A group action by G on W.
- Another group action by G on Z.[2]

[2] Unclear if it requires an action on Z or there exists argument

Why bother?

Suppose you want to do 3-D object segmentation and the spaces W and Z are standard Euclidean spaces. f is the function that takes a scene and gives its segmentation and let the group G be $SO(3)$.

Why bother?

You want the rotation of your segmented scene to be the same as segmentation of your rotated scene. Is this guaranteed by our typical Neural Nets?

Equivariant Map

In general, there is no guarantee of existence of a compatible action on the Z space.

Disentangled Representation

Given a group G with decomposition $G = G_1 \times G_2$ which acts on W, the representation space Z is **disentangled** if

- There exists an action by G on Z.
- The map $f : W \to Z$ is equivariant.
- There is a decomposition $Z = Z_1 \times Z_2$ such that the action on Z is disentangled.

Note that they don't require the action on W to be disentangled. (*Didn't feel natural*)

Disentangled Representation: The Recipe

You give me:

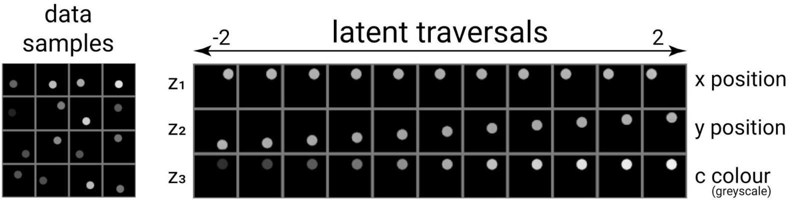
- A group G with a particular decomposition.
- A group action by G on W.
- Inference procedure to learn representations: f

I will find:

- An action by G on Z.
- Whether f is equivariant given the actions.
- Whether the action on Z is disentangled.

If I succeed, the representation is disentangled.

Example of Disentangled Representation



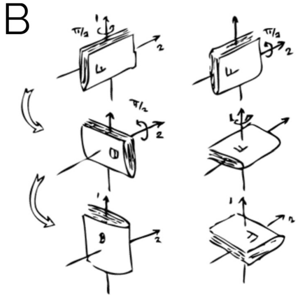
Training CCI-VAE on the data samples leads to a decent notion of disentanglement.

Counter-Intuitive

Consider the group $SO(3)$ which acts on $\mathbb{R}^3$ and rotates it. Maybe we can decompose it into three subgroups -

- Rotate about the X-axis
- Rotate about the y-axis
- Rotate about the Z-axis

Umm no. Can't act independently.



Linear Disentangled Representations

- We have a definition of disentanglement.
- There is no constraint on the disentangled group action.
- What if we want linear transformations on the disentangled subspaces?
- What does linear even mean?

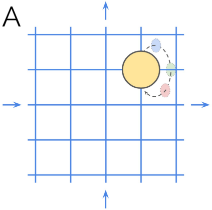
Linear Disentangled Representations

- We can go through group representation theory, but lets skip it.
- Suppose our group acts on a vector space now.
- Now we know what is linear here :)
- If the group action leads to a linear transformation in the vector space, it is linear.
- Works out!

Example of Linear Disentangled Representation

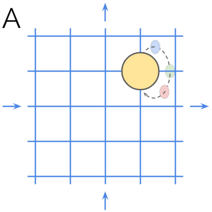
Lets take our favourite example

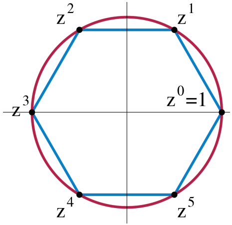
- W: $\{(x_i, y_j, z_k\}_{i,j,k=1}^{N}$
- Z: $\mathbb{C}^3$ (but lets think of it as $\mathbb{R}^6$)
- We have group decomposition $G_x \times G_y \times G_c$
- The action of each G_i on W is translation on its subspace.
- We also have group generators g_x, g_y, g_c



Consider the full inference process as the map

$$f(x, y, c) = (e^{2\pi x i/N}, e^{2\pi y i/N}, e^{2\pi c i/N})$$

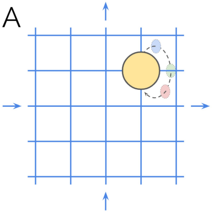




Group Action on Z

Our disentangled group action is

$$g_x \odot (z_x, z_y, z_c) = (e^{2\pi i/N} z_x, z_y, z_c)$$
$$g_y \odot (z_x, z_y, z_c) = (z_x, e^{2\pi i/N} z_y, z_c)$$
$$g_c \odot (z_x, z_y, z_c) = (z_x, z_y, e^{2\pi i/N} z_c)$$



Is f Equivariant?

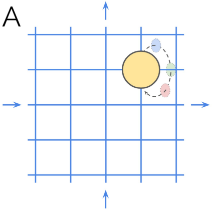
$$
\begin{aligned}
g &= (g_x, g_y, g_c) \\
w &= (x, y, c) \\
f(g \odot_w w) &= f((x+1, y+1, c+1) \mod N) \\
g \odot_z f(w) &= g \odot_z (e^{2\pi x i/N}, e^{2\pi y i/N}, e^{2\pi c i/N}) \\
&= (e^{2\pi (x+1) i/N}, e^{2\pi (y+1) i/N}, e^{2\pi (c+1) i/N}) \\
&= f(g \odot_w w)
\end{aligned}
$$

Hence, we have an equivariant map.

Linear Disentangled Representation

- We have found an action on Z.
- f is an equivariant map.
- $Z = Z_x \times Z_y \times Z_c$ where Z_i is isomorphic to $\mathbb{C}$.
- Rotation in $\mathbb{R}^6$ is linear.

We have thus found a disentangled linear representation!



Conclusion

- First mathematical formulation of *disentanglement*
- Heavily dependent on the choice of G[3]
- Unclear how to find an action on Z, especially one that makes some general function f equivariant
- Beneficial to construct datasets and benchmarks for quantifying disentanglement

Questions?

[3] Not necessarily a bad thing

Appendix

- Group Representation Theory
- Compatibility

Group Representation Theory

Group representation theory relates abstract groups to linear transformations of vector spaces.

A group representation $\rho : G \to GL(V)$ satisfies

- $\rho(g \circ h) = \rho(g)\rho(h)$
- $\rho(e) = 1_V$

(I would assume the latter follows from the former?)

Group Representation as Group Action

Given a representation ρ, we can define the action $\odot$ as

$$\begin{aligned} \odot : G \times V &\to V \\ (g, v) &\to \rho(g)v \end{aligned}$$

Creating new group representations

Given two representations $\rho_1 : G \to GL(V)$ and $\rho_2 : G \to GL(W)$, we can define a new representation $\rho_1 \oplus \rho_2 : G \to GL(V \oplus W)$ as

$$(\rho_1 \oplus \rho_2)(g)(v, w) = (\rho_1(g)(v), \rho_2(g)(w))$$

One can relate this to

$$(\rho_1 \oplus \rho_2)(g) = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

Reducing representations

Given a representation $\rho : G \to GL(V)$, we can get a subrepresentation $W \leq V$ such that

$$\rho(g)(w) \in W \qquad \forall g \in G, w \in W \tag{1}$$

Eg. $\{0\}$ or V itself. If others exist, V is *reducible*.

Factorizing Representations

If V is reducible, we can have $V = W_1 \oplus W_2 \oplus \ldots \oplus W_n$.

One can see, then, that the representation ρ will be a respective block-diagonal matrix.

Cool Thing: *We always end up with the same set of irreducible representations no matter how we do the decomposition.*[4]

[4] under certain conditions; and up to isomorphisms, order and change of basis

Linear Disentangled Group Action

$\rho : G \to GL(V)$ is linearly disentangled wrt $G = G_1 \times G_2$ if there is a decomposition $V = V_1 \times V_2$ and representations $\rho_i : G_i \to GL(V_i)$ such that

$$\rho(g_1, g_2)(v_1, v_2) = (\rho_1(g_1)(v_1), \rho_2(g_2)(v_2))$$

This is the same as disentangled group action, just with ρ being linear.

Compatibility

Modularity: Single latent dimension encodes at most a single data generative factor. Works with the definition; think data generative factor as disentangled action by a symmetry group.

Compatibility

Compactness: Each data generative factor is encoded by a single latent dimension. Not necessary; think along SO(3).

Compatibility

Explicitness: If all data generative factors can be decoded from the latent representation using a linear transformation. General agreement, with linearity not necessary unless *linear disentangled representation*.