COMP 760 Week 8: Product Manifolds and Latent Manifolds

By Joey Bose and Prakash Panangaden

Admin stuff week 8

Final Project Presentations

- Google Sheet Signup:
- https://docs.google.com/spreadsheets/d/1FVd1WnqZJ0KcZbtNjfncEJST-PZbwbWW8cE7IGJr_Vw/edit?usp=sharing

Product Manifolds

Product Manifolds

- Consider $\mathscr{M}_1, \ldots, \mathscr{M}_k$ smooth manifolds

- Product manifold is the cartesian product :

$$\mathscr{M} = \mathscr{M}_1 \times \mathscr{M}_2 \times \ldots \times \mathscr{M}_k$$

- A point $p \in \mathscr{M}$ is defined as $p = (p_1, \ldots, p_k) : p_i \in \mathscr{M}_i$

- A tangent vector $v \in \mathscr{T}_p\mathscr{M}$ is $v = (v_1, \ldots, v_k) : v_i \in \mathscr{T}_{p_i}\mathscr{M}_i$

Mila

Product Manifolds Exponential Map and Distance

$$\mathscr{M} = \mathscr{M}_1 \times \mathscr{M}_2 \times \ldots \times \mathscr{M}_k$$

- The exponential map also decomposes across manifolds

$$\exp_p(v) = (\exp_{p_1}(v_1), \ldots, \exp_{p_k}(v_k))$$

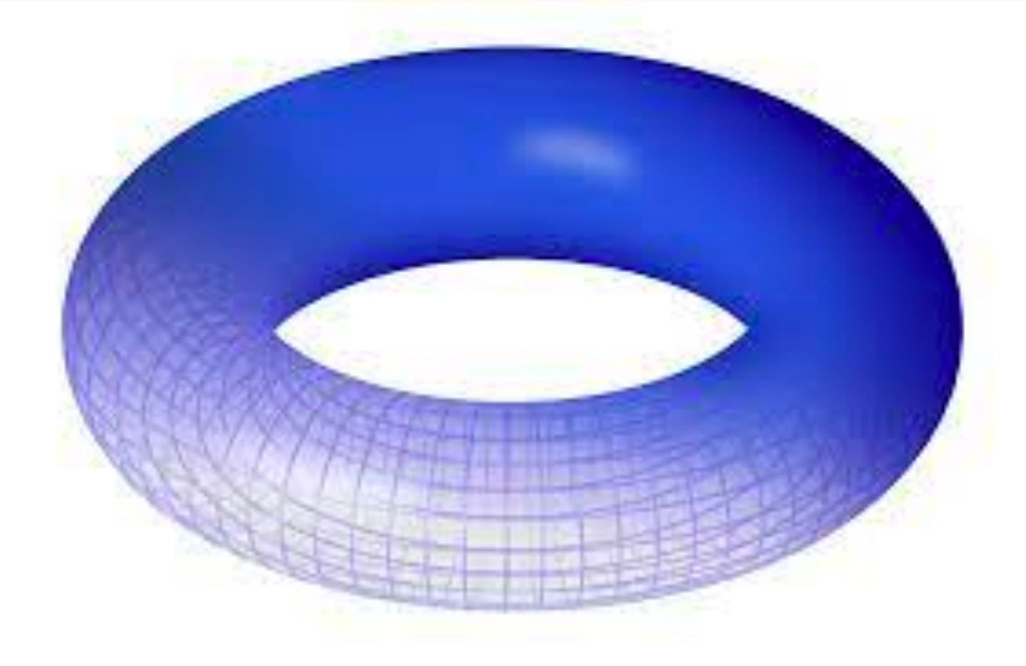
- The distance becomes the sum of distance over each manifold in the product:

$$d_{\mathscr{M}}(x, y) = \sum_i^k d_{\mathscr{M}_i}(x_i, y_i)$$

Mila

Examples

- $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \simeq \mathbb{R}^{n_1+n_2}$
- Torus: $\mathbb{T}^k = \mathbb{S}^1 \times \ldots \times \mathbb{S}^k$



Latent Geometry

What kind of structure can we ask of a latent space?

- Take the latent space to be an explicit manifold (e.g. Spherical, Hyperbolic).
- The latent space can also be an implicit manifold given to us by the decoder.
- We can impose symmetry constraints (e.g. equivariance).
- What if we want a causal representation? Can we make the latent space be a DAG?

Implicit Latent Manifolds

$$x = f(\boxed{z})$$

Generated sample Latent sample

- f can be the decoder in VAE for instance.

Implicit Latent Manifolds

- Using an application of Taylors approximation to:

$$||f(z+\Delta z_1) - f(z+\Delta z_2)||^2$$

- We get the following metric at point z

$$G = J_z^T J_z \longleftarrow \text{Jacobian at } z$$

Group Actions

1. We have a set $\mathscr{X}$ and $f : \mathscr{X} \rightarrow \mathbb{R}$

2. Group G acts on $\mathscr{X}$

$$T_g : \mathscr{X} \rightarrow \mathscr{X} \quad \forall g \in G$$

$$\forall g1, g2 \in G, \; T_{g2g1} : T_{g2} \circ T_{g1}$$

If $\mathscr{X}$ is a (finite) Vector Space then $T_g \in GL(n)$

3. Extending the action to functions

$$\mathbb{T}_g : f \rightarrow f' \qquad f'(T_g(x)) = f(x)$$

Groups

1. $e \in G$

Identity

2. $(a \circ b) \circ c = a \circ (b \circ c)$

Associativity

3. $\forall a \in G \quad \exists b \in G$

$a \circ b = e$

Unique Inverses

Mila

Equivariance

$$\begin{tikzcd}
L_{(V_1)}(\mathscr{X}_1) \arrow[r, "\mathbb{T}_g"] \arrow[d, "\phi"'] & L_{(V_1)}(\mathscr{X}_1) \arrow[d, "\phi"] \\
L_{(V_2)}(\mathscr{X}_2) \arrow[r, "\mathbb{T}'_g"] & L_{(V_2)}(\mathscr{X}_2)
\end{tikzcd}$$

Equivariance

- Let X and Y be two sets with an action of a group G. A map $\phi : X \to Y$ is called G-equivariant, if it respects the action, i.e., $g \cdot \phi(x) = \phi(g \cdot x), \forall g \in G$ and $x \in X$.

- A map $\chi : X \to Y$ is called G-invariant, if $\chi(x) = \chi(g \cdot x), \forall g \in G$ and $x \in X$.

Mila

Equivariance in Latent Space

- The latent space itself can be equivariant or invariant to some pre-defined group.
- Partitions of the latent dimensions could be equivariant or invariant to different groups.
- The latent space could decompose as a product of groups.

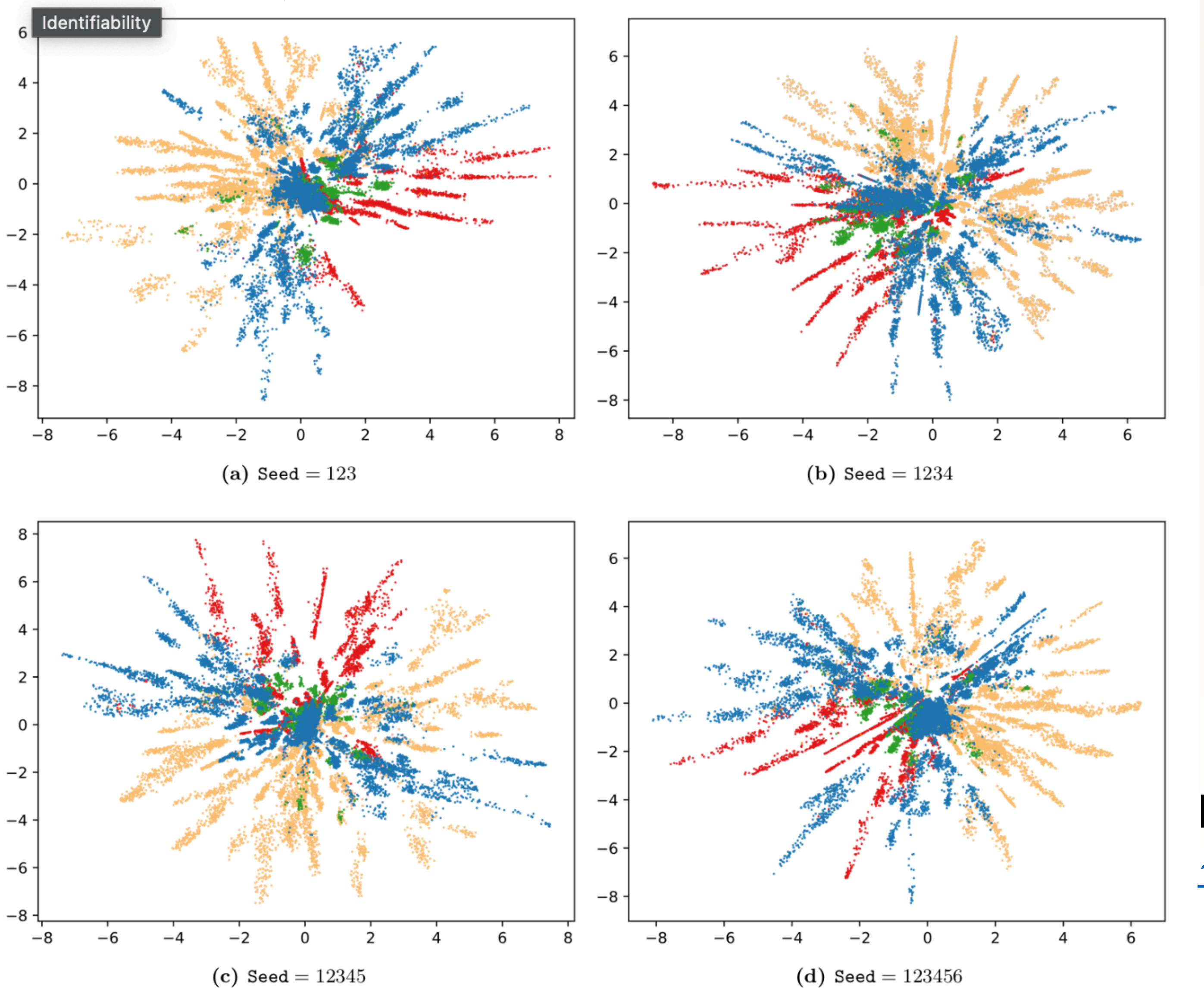
Identifiability in Generative Models

$$x = f(\boxed{z})$$

Generated sample Latent sample

- In practice different training runs could lead to drastically different latent spaces.

Identifiability in Generative Models



4 trainings runs of the same VAE model

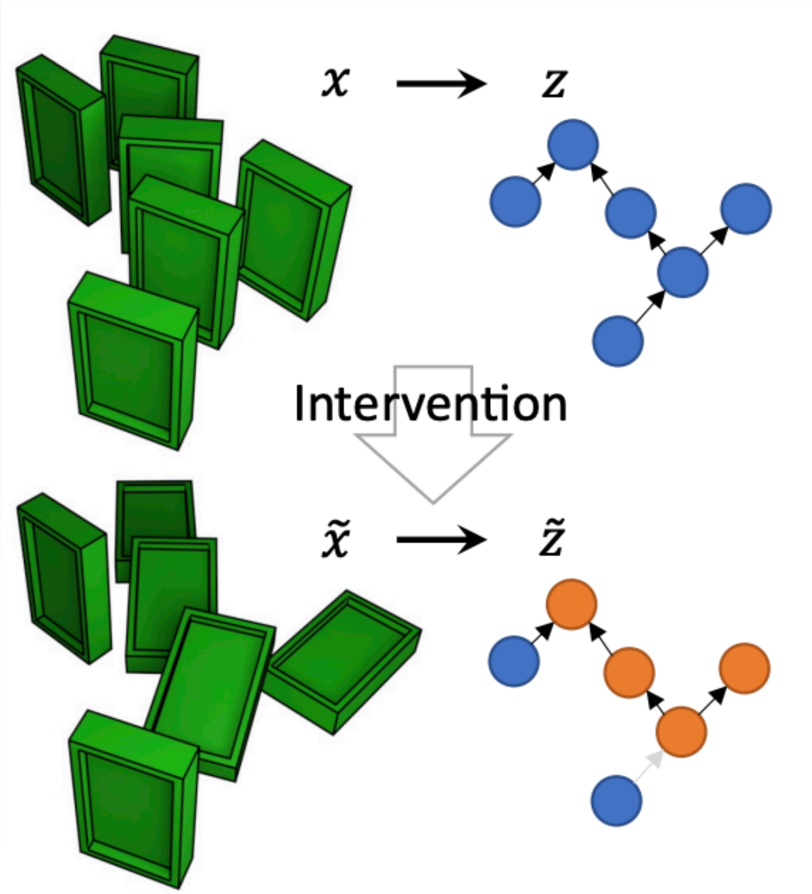
Fig credit: http://www2.compute.dtu.dk/~sohau/weekendwithbernie/

Mila

Identifiability in Generative Models

- If p_θ is a density parameterized by θ then a generative model is identifiable if $\theta \to p_\theta$ is bijective.
- Otherwise two different models might give equally good results but we don't know which to trust.
- Most modern deep generative models are not identifiable.

Rethinking Generative Modeling with Weak Supervision



Assume you have access to both pre and post-interventional data.

Q. Can we recover the causal graph in this setting?

Structural Causal Model

- An SCM C, describing the relation between causal variables $z_1, \ldots, z_n$, with domains $\mathcal{Z}_i$
- Exogenous noise variables $\epsilon_1, \ldots, \epsilon_n$, with domains $\mathcal{E}_i$
- A directed acyclic graph $\mathcal{G}(C)$
- Causal Mechanisms $f_i : \mathcal{E}_i \times \prod_{j \in \mathrm{pa}_i} \mathcal{Z}_j \to \mathcal{Z}_i$
- A unique solution $s : \mathcal{E} \to \mathcal{Z}$ by successively applying the causal mechanisms
- A stochastic intervention $(I, (\tilde{f}_i)_{i \in I})$ that modifies the SCM by replacing for a subset of the causal variables, called the intervention target set $I \subset \{1, \ldots, n\}$

Mila

Latent Causal Model

$$\mathcal{M} = \langle C, \mathcal{X}, g, \mathcal{I}, p_{\mathcal{I}} \rangle$$

- An acyclic faithful SCM C
- An observation space $\mathcal{X}$
- A decoder $g : \mathcal{Z} \rightarrow \mathcal{X}$, that is diffeomorphic on its image
- A set $\mathcal{I}$ of interventions on C
- A probability measure $p_{\mathcal{I}}$ over $\mathcal{I}$

LCM Isomorphism

$$\mathscr{M} = \langle C, \mathscr{X}, g, \mathscr{I}, p_{\mathscr{I}} \rangle$$

- Let $\mathscr{M} = \langle C, \mathscr{X}, g, \mathscr{I}, p_{\mathscr{I}} \rangle$ and $\mathscr{M}' = \langle C', \mathscr{X}, g', \mathscr{I}', p_{\mathscr{I}'} \rangle$ be two LCMs
- An LCM isomorphism is a graph isomorphism $\phi : \mathscr{G}(C) \to \mathscr{G}(C')$
- Along with element wise diffeomorphisms for noise and causal variables
- $\mathscr{M}$ and $\mathscr{M}'$ are equivalent if there exists an LCM iso between them

Mila

Weakly Supervised Generative Process

$$\mathcal{M} = \langle C, \mathcal{X}, g, \mathcal{I}, p_{\mathcal{I}} \rangle$$

$$\epsilon \sim p_{\mathcal{E}}, \qquad z = s_I(\epsilon), \quad x = g(z)$$

$$\mathcal{I} \sim p_{\mathcal{I}}, \quad \forall i \in I,\ \tilde{e}_i \sim p_{\tilde{\mathcal{E}}_i},\ \forall i \notin I,\ \tilde{e}_i = \epsilon_i,\ \forall i \notin I,\ \tilde{e}_i = \epsilon_i, \quad \tilde{z} = \tilde{s}_I(\tilde{\epsilon}), \quad \tilde{x} = g(\tilde{z})$$

Identifiability: Brehmer et. Al 2022

Theorem 1 (Identifiability of $\mathbb{R}$-valued LCMs from weak supervision). *Let $\mathcal{M} = \langle \mathcal{C}, \mathcal{X}, g, \mathcal{I}, p_\mathcal{I} \rangle$ and $\mathcal{M}' = \langle \mathcal{C}', \mathcal{X}, g', \mathcal{I}', p'_{\mathcal{I}'} \rangle$ be LCMs with the following properties:*

- *The LCMs have an identical observation space $\mathcal{X}$.*
- *The SCMs $\mathcal{C}$ and $\mathcal{C}'$ both consist of n real-valued endogeneous causal variables and corresponding exogenous noise variables, i. e. $\mathcal{E}_i = \mathcal{Z}_i = \mathcal{Z}'_i = \mathcal{E}'_i = \mathbb{R}$.*
- *The intervention sets $\mathcal{I}$ and $\mathcal{I}'$ consist of all atomic, perfect interventions, $\mathcal{I} = \{\emptyset, \{z_0\}, \ldots, \{z_n\}\}$ and similar for $\mathcal{I}'$.*
- *The intervention distribution $p_\mathcal{I}$ and $p'_{\mathcal{I}'}$ have full support.*

Then the following two statements are equivalent:

1. *The LCMs entail equal weakly supervised distributions, $p^\mathcal{X}_\mathcal{M}(x, \tilde{x}) = p^\mathcal{X}_{\mathcal{M}'}(x, \tilde{x})$.*
2. *The LCMs are equivalent, $\mathcal{M} \sim \mathcal{M}'$.*

Mila

Implications of Identifiability

- Assume we have access to data pairs $(x, \tilde{x})$
- We can then train an LCM with parameters by Maximum Likelihood!
- Because of identifiability the trained LCM and ground-truth LCM are the same up to relabelling.
- $\mathscr{M}$ and $\mathscr{M}'$ are equivalent if there exists an LCM iso between them

Mila