CS6360 Assignment 4: Extensions

Pranav K Nayak

1 Extending the Proof

The assumption that the set of transformations being worked with lie on a differentiable manifold (i.e. that they form a Lie group) is surprisingly lenient. Most sets of transformations that are interesting could be wrangled into this framework. For example:

- All transformations that can be represented through an invertible $n \times n$ matrix whose determinant is 1, the group $\mathbb{SL}(\mathbb{R}, n)$
- All rotations (i.e. the group $\mathbb{SO}(n)$) when $n \leq 2$

For those transformations that cannot be represented through matrix multiplication, they allow for the existence of a 'world state space' which has a bijection to the space of observations, and for which these transfomations are representable. As a concrete example, if our observation space is the vector space $\mathbb{R}^n$, then there is no $n \times n$ matrix that can perform translation in this space (this is easy to see by noting that the origin will be a fixed point for any matrix multiplication, meaning that not all points get translated). Performing translation requires that one actually operate in $\mathbb{R}^{n+1}$, multiplying by an $n \times n$ matrix, before projecting the first n dimensions back down into $\mathbb{R}^n$.

Interestingly, this allows us to capture all possible affine transformations, since they will just be a composition of a linear tranformation followed by a translation (this is exactly what is done in linear regression when the bias term is folded into the matrix-vector product, i.e. $\mathbf{Ax} + \mathbf{b} \mapsto \mathbf{A}'\mathbf{x}'$).

They acccount for this process of transforming in a different space by allowing for a world state space, on which the actual transformations are performed, and for which there is a surjective mapping to the space of observations.

The one place where I could see room for change is the requirement that the group of transformations be closed and connected. This doesn't really cover any transformation that belongs to the general group of linear invertible matrices (with determinant $\neq 1$). This assumption is used to ensure that the Lie Algebra - Lie Group correspondence is bijective, allowing the entire group to be captured by the linear structure of the algebra. Were this not the case, then the algebra only captures local structure around the identity.

If we can still show that if we do restrict ourselves to a neighborhood of the identitiy transformation, the deviation for all ”realistic” group elements is below some threshold from its representation in the lie algebra, then we can potentially extend the set of allowed transformations whose representations can be learned.

2 Incorporating Diffusion

The step-wise application of transformations does look similar to the diffusion process. That the transformations can be composed further cements this similarity, since it reminds one of the reparameterization trick.

The obvious question is then to see how well the diffusion process can be performed within this autoencoding + transformation-representation learning setting. The immediate first step, I think, before making any further claims is showing that the addition/removal of Gaussian noise can be considered as belonging to a group. If that is not the case (which I doubt, since the group axioms are extremely general), this avenue is most likely a dead end.

If additive noise does in fact form a group, then the next question is analyzing the structure of the manifold this group lies on. This should not be particularly difficult, since, intuitively, addition of noise is a translation operation, meaning that it lies on a very tight[1] subspace in the augmented space.

If everything falls in place, and Gaussian noise can simply be substituted into all formulae and propositions of the paper as the group G, we still have to deal with the central question: *what now?*.

I can see two directions to take this:

1. Simply replace the translation group of the paper with the ”Gaussian group”, repeat all experiments, and see exactly what the representation space ends up looking like (through some sort of dimensionality reduction).

2. Interleave the two processes by incorporating the disentangling autoencoder into the diffusion model’s architecture, and optimizing for some combined goal.

Even then, the question of what we’re gaining still remains. Since we’re just adding Gaussian noise, there isn’t any ”transformation” happening in the sense of an agent acting on a world. It is not obvious to me that additive noise can ”mean” anything. The only potentially interesting result I can think of is if the representations follow a cyclical trajectory, and if for all inputs the trajectories coincide at the last noise-addition step, leading to some sort of hypersphere. If this is the case, then ideally this hypersphere can be used to interpolate between two different representations.

[1] By *tight*, I refer to the fact that its form is sparse

This idea, however, relies on some heavy assumptions and hopes that things ”work out”.

The diffusion process can be framed as an optimal transport problem (credit to Piyushi), so there could be tools and open problems from OT that could be used/addressed with this framework.

3 Introducing Non-Determinism

The encoder and decoder in the architecture are deterministic, and this is just assumed to be the case, with no argument given for why a VAE-like architecture was not chosen instead. This could also serve as a direction of inquiry.

One immediate benefit is that probing the latent space becomes significantly easier. Another question that arises is how the group of transformations chosen affects the learned distribution, which is (to the best of my knowledge), easier to answer with a VAE.