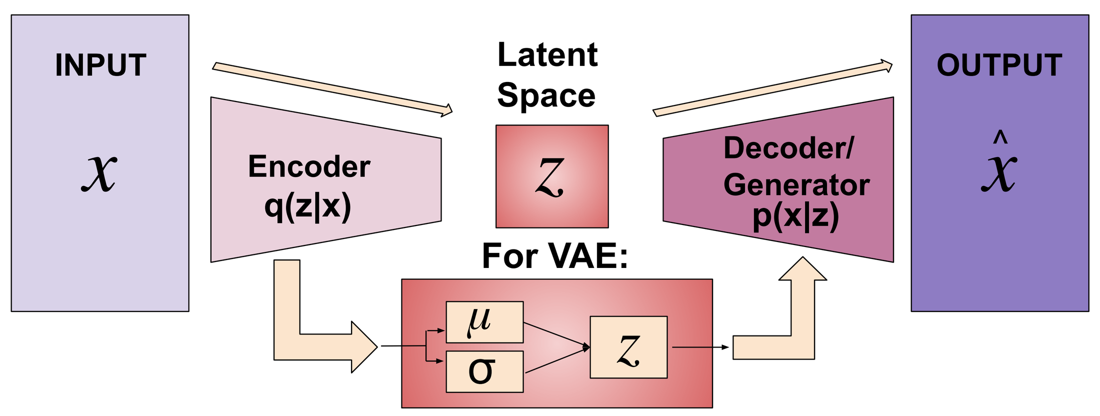
Enhancing VAEs with Flexible Conditional Normalizing Flows

Callihan Bertley [1], Claire Gan [2], Rishi Leburu [3], Malia Walewski [3]

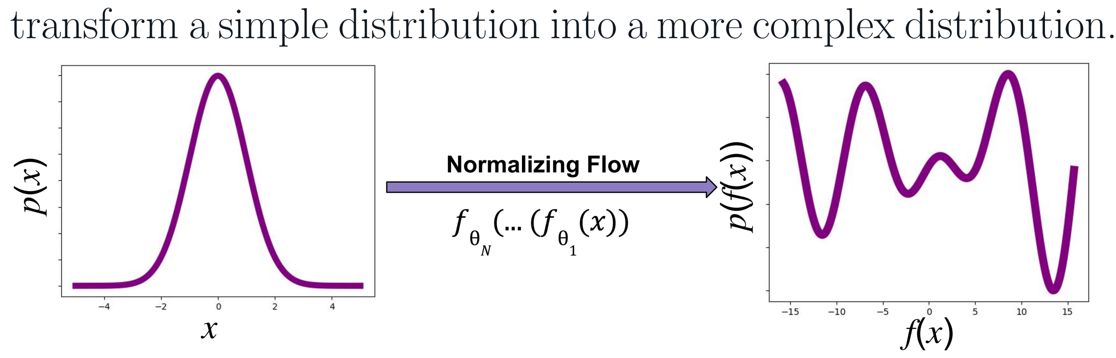
[1] The University of Texas at Austin, [2] The University of Nevada Reno, [3] Emory University

Introduction

A **Variational Autoencoder (VAE)** is a generative model that, using the architecture of neural networks, learns a probabilistic mapping between some input data $\mathbf{x}$ and a latent space $\mathbf{z}$. Thus, output data $\hat{\mathbf{x}}$ that resembles the input $\mathbf{x}$ can be generated by sampling from the latent space and passing samples through the inverse map.

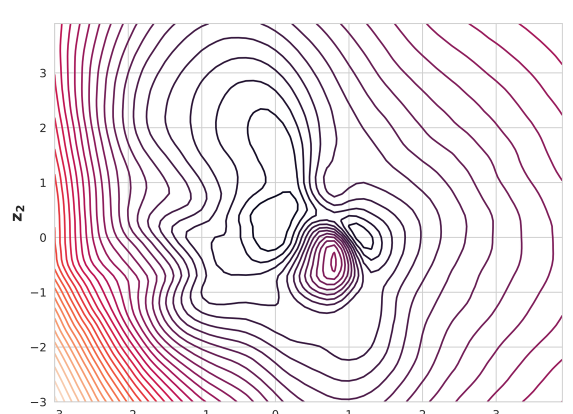


Normalizing flows are a series of invertible mappings that transform a simple distribution into a more complex distribution.



Motivation

To approximate the posterior distribution, VAEs default to a Gaussian. However, since the true posterior distribution often deviates from normality, we aim to find a stronger approximation.



$\log p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$ + constant (from MNIST dataset)

Proposed Solution is **Conditional Normalizing Flows:**

- **Inverse Autoregressive Flows (IAF)**
- **Partially Convex Potential Maps (PCP-Map)**

Loss Function

VAEs allow for efficient training of generative models by minimizing the negative evidence lower bound (ELBO).

$$J_{\text{ELBO}} = \underbrace{\mathbb{E}_{\mathbf{z}\sim q_{\psi}(\mathbf{z}|\mathbf{x})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})\right]}_{\text{reconstruction term}} - \underbrace{D_{kl}\left(q_{\psi}(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})\right)}_{\text{regularization term}}$$

Methods

1. Standard VAE

The Standard VAE uses a Gaussian approximation for both the encoder $q_{\psi}(\mathbf{z}|\mathbf{x})$ and prior distribution $p(\mathbf{z})$:

$$q_{\psi}(\mathbf{z}|\mathbf{x}) = \mathcal{N}\left(\mathbf{z}|\boldsymbol{\mu}_{\psi}, \operatorname{diag}\left[\boldsymbol{\sigma}^2_{\psi}(\mathbf{x})\right]\right)$$
$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|0, \mathbf{I}).$$

Note that $\boldsymbol{\psi}$ are the parameters of the approximate encoder and $\boldsymbol{\theta}$ are the parameters of the decoder and intractable distribution, $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$.

2. IAF VAE [1]

The IAF VAE applies a series of invertible, autoregressive transformations to the Gaussian approximated distribution in the encoding process. For each transformation, an autoregressive neural network takes in the input latent sample and the context vector $\mathbf{h}$ and outputs a $\boldsymbol{\mu}_{\psi}$ and $\boldsymbol{\sigma}_{\psi}$ to transform $\mathbf{z}_0$:

$$\mathbf{z}_t = \boldsymbol{\mu}^{\psi}_t(\mathbf{z}_{t-1}, \mathbf{h}) + \boldsymbol{\sigma}^{\psi}_t(\mathbf{z}_{t-1}, \mathbf{h})\mathbf{z}_{t-1}, \quad 1 \le t \le T.$$

The Jacobian matrices $\frac{d\mathbf{z}_t}{d\mathbf{z}_{t-1}}$ are lower triangular, which makes computing the determinant not expensive.

After applying T transformations sequentially:

$$\log\left(\mathbf{q}_{\psi}(\mathbf{z}_T|\mathbf{x})\right) = \log\left(\mathbf{q}_{\psi}(\mathbf{z}_0|\mathbf{x})\right) - \sum_{t=1}^{T} \log\left|\det\left(\frac{d\mathbf{z}_t}{d\mathbf{z}_{t-1}}\right)\right|.$$

3. PCP-Map VAE [3]

The PCP-Map VAE is another way to transform the input, Gaussian distribution to a more complex distribution. This is achieved by parameterizing a transformation as the gradient of a scalar-valued Partially Input Convex Neural Network, $\mathbf{G}_{\psi}$, where:

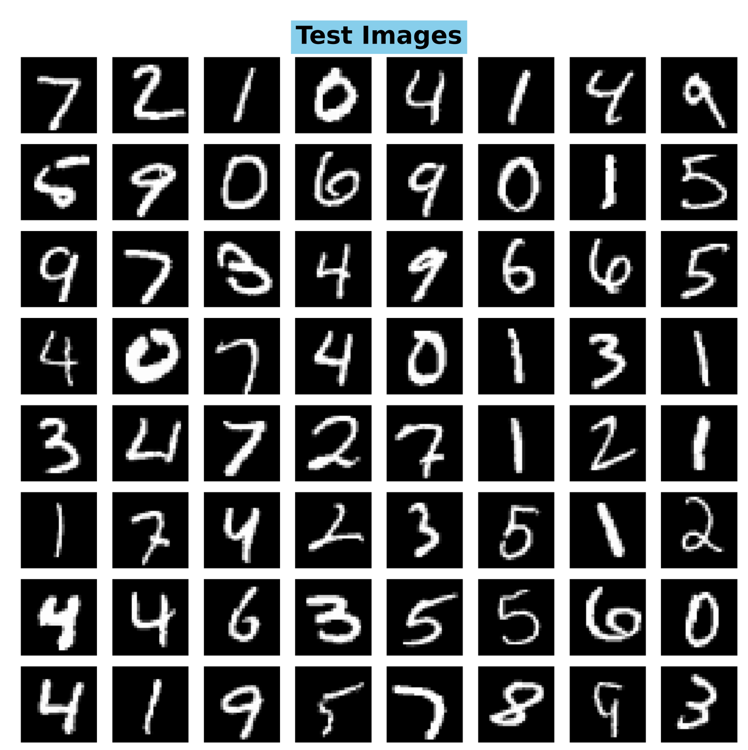
$$\mathbf{G}_{\psi}(\mathbf{z}_0, \mathbf{h}) = \sigma_{softplus}(\gamma_1) * \mathbf{z}_T + (\sigma_{ReLU}(\gamma_2) + \sigma_{softplus}(\gamma_3)) * \frac{1}{2}\|\mathbf{z}_0\|^2,$$

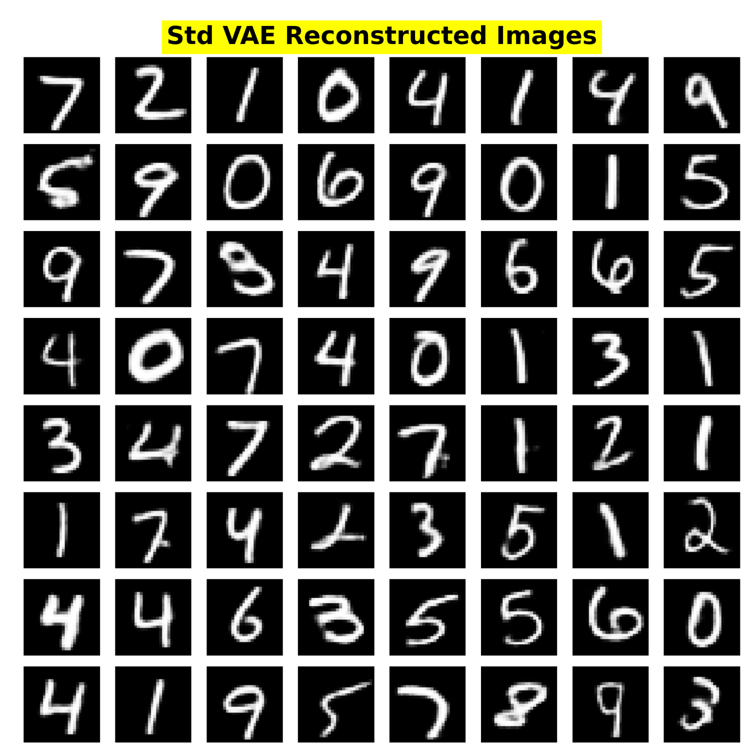
and for $0 \le k < T$,

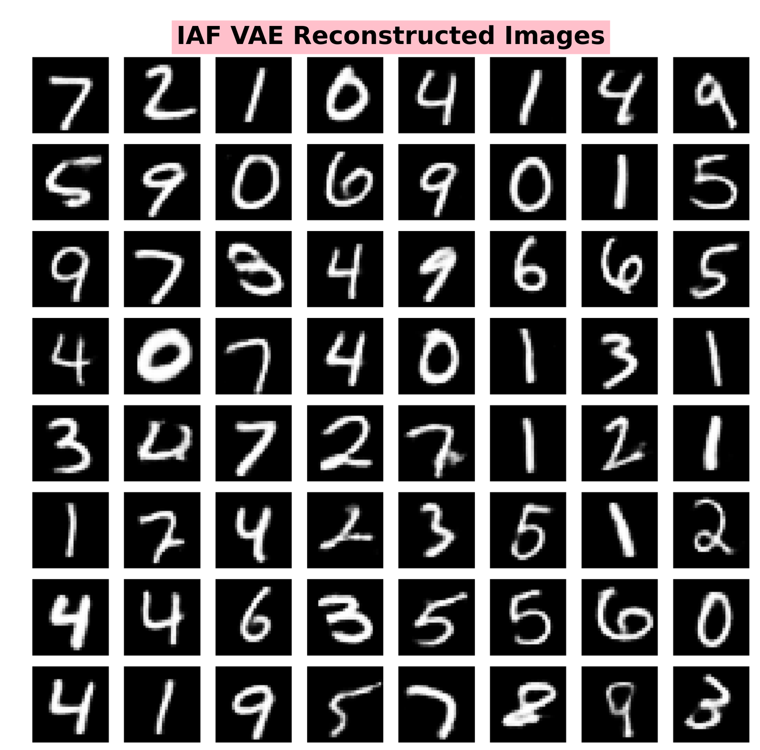
$$\mathbf{v}_{k+1} = \sigma^{(v)}\left(\mathbf{L}^{(v)}_k \mathbf{v}_k + \mathbf{b}^{(v)}_k\right),$$
$$\mathbf{z}_{k+1} = \sigma^{(z)}\left(\left(\mathbf{L}^{(z)}_k\right)\left(\mathbf{z}_k \odot \left(\mathbf{L}^{(zv)}_k \mathbf{v}_k + \mathbf{b}^{(zv)}_k\right)\right) + \mathbf{L}^{(z)}_k\left(\mathbf{z}_0 \odot \left(\mathbf{L}^{(zv)}_k \mathbf{v}_k + \mathbf{b}^{(zv)}_k\right)\right) + \mathbf{L}^{(vz)}_k \mathbf{v}_k + \mathbf{b}^{(z)}_k\right).$$

where $\mathbf{v}_0 = \mathbf{h}$, and $\gamma_1, \gamma_2, \gamma_3$ are some learnable parameters of $\mathbf{G}_{\psi}$. The functions σ, σ_{ReLU} and $\sigma_{softplus}$ are activation functions to ensure convexity.

Results







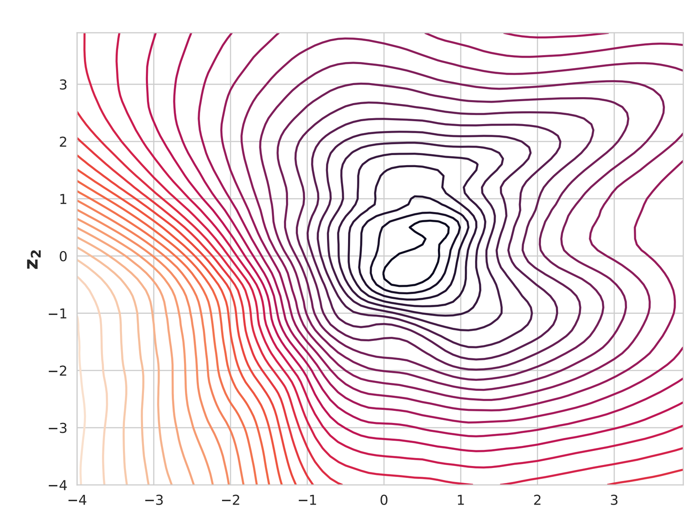


Fig. 4: $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$ + constant

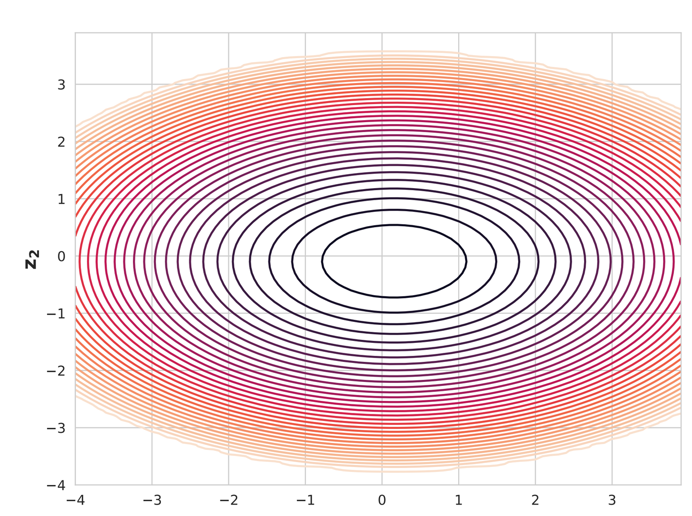


Fig. 5: Standard $q_{\psi}(\mathbf{z}|\mathbf{x})$

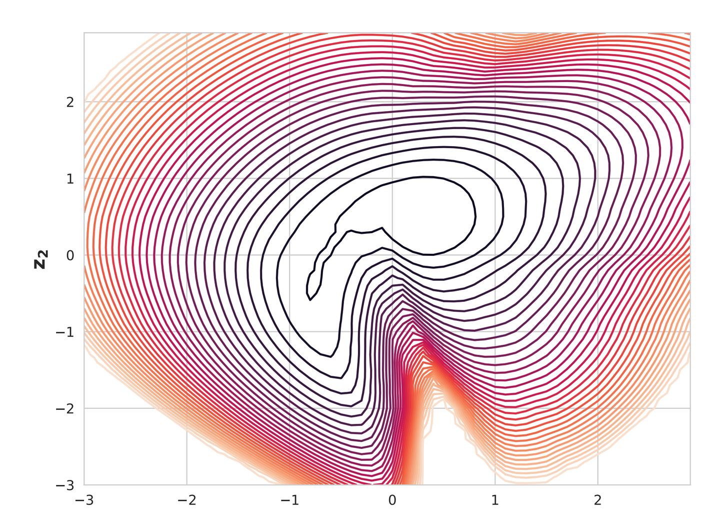
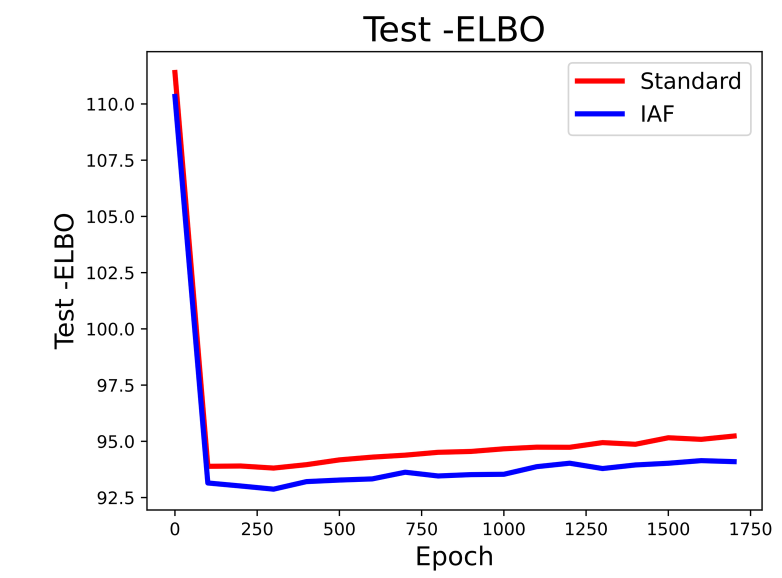


Fig. 6: IAF $q_{\psi}(\mathbf{z}|\mathbf{x})$

Conclusion



Key Takeaways:

- The IAF VAE demonstrates a lower test loss compared to the standard VAE, indicating better performance in terms of reconstruction quality and latent space representation.
- While the IAF VAE performs better than the standard, it comes with increased computational complexity. This trade-off between performance and computational cost should be considered based on the specific application requirements.

Future Work

- Test and evaluate the performance of the PCP-Map in comparison to the Standard VAE model and IAF VAE.
- Explore other Conditional Normalizing Flow models, such as the COT-Flow.

Acknowledgements

This work is supported in part by the US National Science Foundation awards DMS-2349535 and DMS-2038118. We would like to thank Dr. Deepanshu Verma for his mentorship during this program and the rest of the REU mentors for their guidance and support.

References

[1] Diederik P. Kingma et al. *Improving Variational Inference with Inverse Autoregressive Flow.* 2017. arXiv: 1606.04934 [cs.LG].

[2] Lars Ruthotto and Eldad Haber. "An introduction to deep generative modeling". In: *GAMM-Mitt.* 44.2 (2021), Paper No. e202100008, 24. ISSN: 0936-7195,1522-2608. DOI: 10.1002/gamm.202100008. URL: https://doi.org/10.1002/gamm.202100008.

[3] Zheyu Oliver Wang et al. *Efficient Neural Network Approaches for Conditional Optimal Transport with Applications in Bayesian Inference.* 2023. arXiv: 2310.16975 [stat.ML]. URL: https://arxiv.org/abs/2310.16975.