Variational Inference using ResNet Flow

Hadi Salman, Naman Gupta, Raghu Aditya Chavali
Advisor: Kayhan Batmanghelich
Carnegie Mellon University

Introduction

- Variational inference involves the need to approximate intractable posterior distributions by a class of known probability distributions, over which we search for the best approximate to the true posterior.
- Our approximations (distributions) are constructed through a normalizing flow by transforming a simple initial density into a more complex one by applying a sequence of invertible transformations until a desired level of complexity is attained.
- We propose a novel method for variational inference using a flow based approach that draws inspiration from modelling the motion of highly viscous fluids.
- Not only we aim to improve the estimate of the posterior distribution but also we aim to solve the core problem of under-estimation of the variance of the posterior distribution based on the approximation.

Related Work

- Rezende & Mohamed (2015):

The family of transformations:

$$f(z_{k-1}) = z_k = z_{k-1} + uh(w^T z_{k-1} + b)$$

Determinant of Jacobian:

$$\det(Jac) = \ |1 + u^T h'(w^T z_{k-1} + b) w|$$

This once calculated can be used to compute the probability density function of the last iterate.

- Kingma et al. (2016):

Inverse Autoregressive flow consists of a chain of T of the following transformations:

$$f(z_{t-1}) = z_t = \mu_t + \sigma_t \odot z_{t-1}$$

The following jacobians are triangular with σ_t on the diagonal:

$$det(Jac) = \prod_{i=1}^{D} \sigma_{t,i}$$

Future work

- Experiments with CIFAR-10 dataset
- Debugging Approximation 3 and 4 and generate results
- Analysing how performance changes with number of flows and step size, Δt.
- Determine the effects of ResNet flow on Gaussian and Uniform distributions.
- Approximate non-Gaussian 2D distribution using ResNet flow and compare with Normalizing flow results.

Our Model

Probability Distribution of z_k is given by:

$$q(z_k) = q(z_{k-1}) \left| det \frac{\partial f}{\partial z_{k-1}} \right|^{-1}$$

We know from the structure of Normalizing flows the following:

$$z_k = f_k \circ f_{k-1} \ldots \circ f_1(z_0)$$

Thus we can write the normalizing flow density:

$$log\, q_k(z_k) = log\, q_0(z_0) - \sum_{k=0}^{K} log \left| det \frac{\partial f}{\partial z_{k-1}} \right|$$

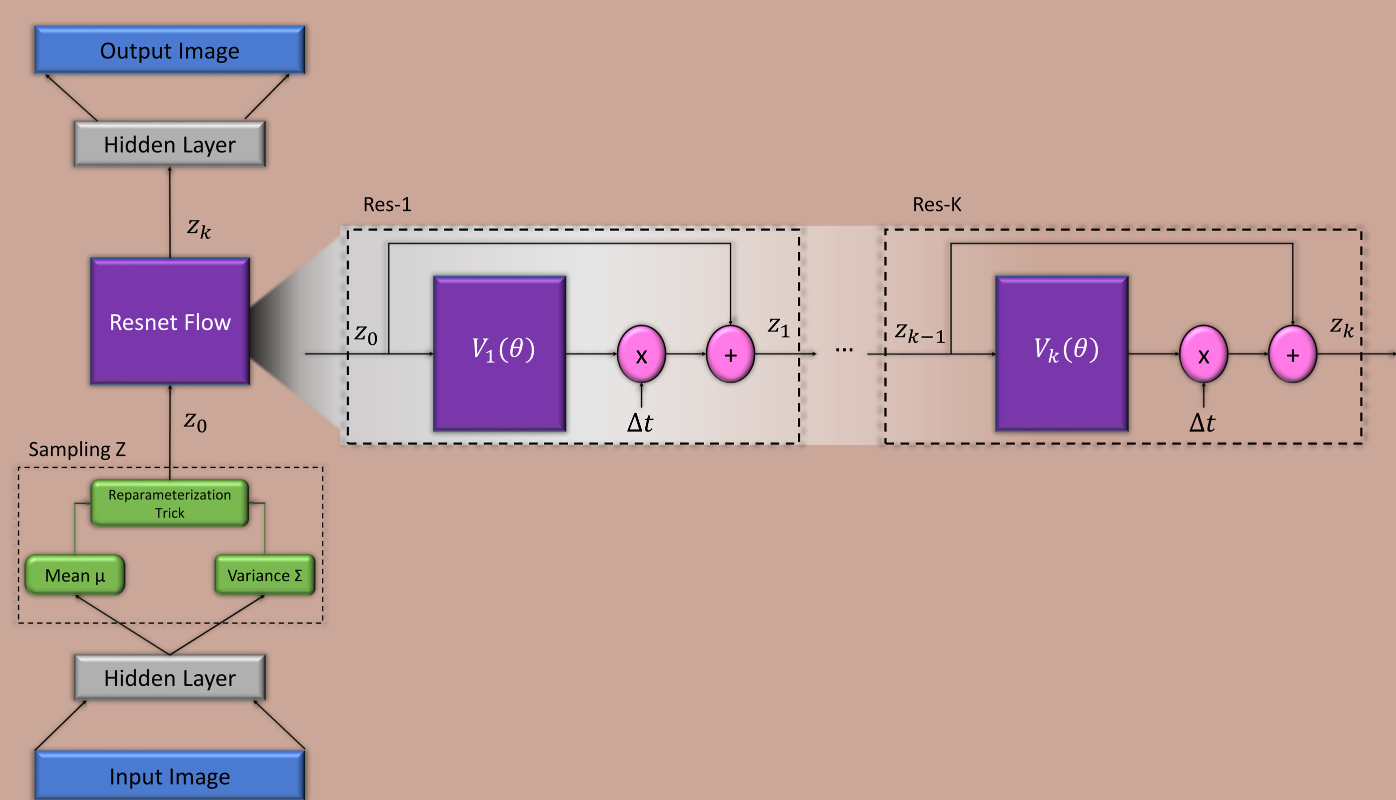


Figure 1: Architecture of our Model with incorporated ResNet flow

From Figure 1, We consider a family of transformations of the form:

$$z_k = f(z_{k-1}) = z_{k-1} + \Delta t\, z_{k-1} V_k$$

The Jacobian can be written in the form:

$$Jac = \frac{\partial f}{\partial z_{k-1}} = \mathrm{I} + \frac{\partial V_k}{\partial z_{k-1}} \Delta t$$

The whole idea is how to approximate the logdet of the jacobian that will be used in the loss function of the VI. Below are few approximations that we are investigating

- **Approximation 1:** Using Taylor Expansion, we approximate the log of the determinant of the jacobian as follows (as directly used in the code) up to first order:

$$log\,(det(Jac)) = \Delta t\, Tr \left[\frac{\partial V_k}{\partial z_{k-1}}\right]$$

- **Approximation 2:** second order Taylor series

$$log\,(det(Jac)) = \Delta t\, Tr \left[\frac{\partial V_k}{\partial z_{k-1}}\right] - \Delta t^2\, Tr \left[\frac{\partial V_k}{\partial z_{k-1}} \frac{\partial V_k}{\partial z_{k-1}}^T\right]$$

- **Approximation 3:** Another approximation that we used was inspired from McLaurin et. Al. (2015) based on early stopping:

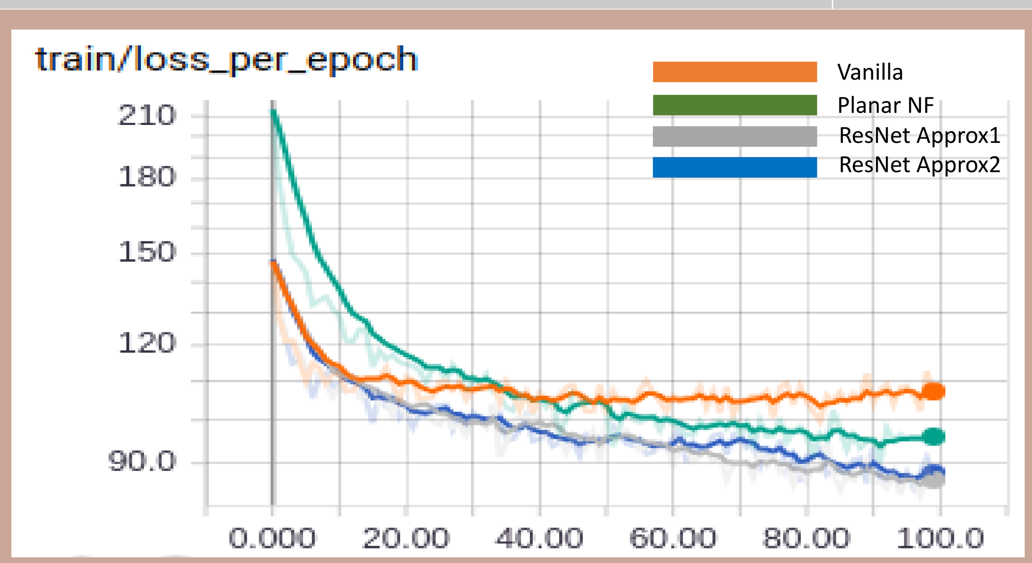
$$r_0 \sim N(0, I_D)$$
$$r_1 = r_0 - \Delta t\, r_0^T \frac{\partial f}{\partial z_{k-1}}$$
$$r_2 = r_0 - \Delta t\, r_1^T \frac{\partial f}{\partial z_{k-1}}$$
$$log\,(det(\mathrm{Jac})) = \mathrm{r}_0^{\mathrm{T}} (-2\mathrm{r}_0 + 3\mathrm{r}_1 - \mathrm{r}_2)$$

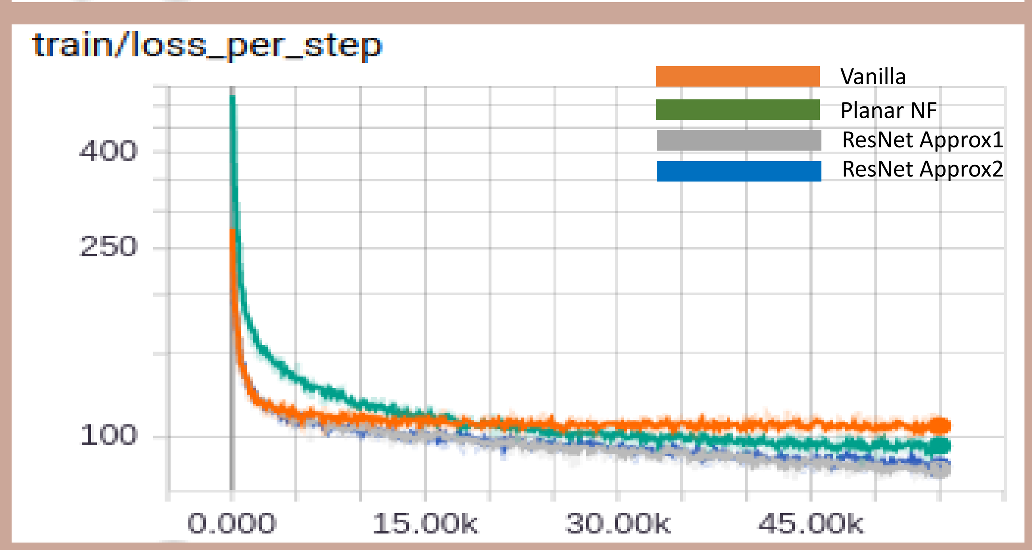
- **Approximation 4:** Another idea that stems from early stopping approximation:

$$log(det(Jac)) = 0$$
$$For\ m\ iterations:$$
$$r \sim N(0, I_D)$$
$$log(det(Jac)) += r^T (Jac) r$$
$$log(det(Jac)) = \frac{log(det(Jac))}{m} \Delta t$$

Results

Model	-ln p(x)
Vanilla VAE (diagonal Covariance)	104.9
Planar Normalizing flows (NF)	96.06
Inverse Autoregressive Flows (IAF)	81.5
Resnet Flow (Approximation 1)	86.32
Resnet Flow (Approximation 2)	86.81





Vanilla

Planar NF

ResNet Approx 1

ResNet Approx 2

References

Rezende, Danilo Jimenez and Mohamed, Shakir. Variational inference with normalizing flows. arXiv preprint arXiv:1505.05770, 2015.

Kingma, Diederik P, Salimans, Tim, Jozefowicz, Rafal, Chen, Xi, Sutskever, Ilya, and Welling, Max. Improved variational inference with inverse autoregressive flow. In Advances in Neural Information Processing Systems , pp. 4743–4751, 2016

Altieri, Nicholas and Duvenaud, David. Variational inference with gradient flows. Advances in Approximate Bayesian Inference, NIPS Workshop, 2015.

McLaurin et. Al. Early stopping is Nonparametric Variational Inference, arXiv preprint arXiv:1504.01344, 2015