

# Advanced Machine Learning Group 12 Project Report

Reproducing results from “Variational Inference with Normalizing  
Flows” [Danilo Jimenez, Rezende Shakir Mohamed]



John Ryan, Man Hon Fan, Charalampos Kokkalis, Kieran Gall

University of Oxford

# Contents

# Introduction

The idea of generative models is one of the most promising approaches to Machine Learning tasks. Many such models have been used in recent years to solve the problem of inference of a posterior distribution, given some expert knowledge and assumptions on the prior distribution and using observed data to adapt these. The main obstacle in attempting to generate the true posterior is its computational intractability. Consequently, approximation methods have been employed, aiming to make computation more efficient while at the same time achieving a close estimate of the true posterior.

The paper focusses on the method of variational inference. Essentially, this converts the inference problem to an optimization one where, given a family of posterior approximations, we look for the distribution that is the closest to the true one. To measure this proximity between approximations and the true posterior, we use the Kullback-Leibler divergence. There has been lots of research in the classes of such, efficiently calculatable, approximation families, such as mean-field approximations, but all of these prove to be of limited expressiveness, leading to solutions that are far from the true posterior. In fact, there is strong evidence showing that more expressive classes of posteriors would result in better performance of variational inference, which is what motivates the work of the paper.

To address the above issue, the authors propose constructing the approximations through a normalizing flow, which applies a sequence of invertible transformations to a simple initial density until it achieves a desired level of complexity. The main types of flows considered and used in the experiments are:

1. Invertible, linear-time flows, such as planar and radial flows.
2. Finite, volume-preserving flows, such as the Non-linear Independent Components Estimation (NICE).

The goal of both of the above is to provide a transformation to the original density that ends up being rich and faithful, while at the same time allowing for efficient computation of the determinant. The first kind of flow allows for computing the logdet-Jacobian in linear (in the number of dimensions) time, while the second one gives a Jacobian with a zero upper triangular part and therefore a determinant of 1. Thus, both can be efficiently implemented and used together with a Deep Latent Gaussian Model and an inference network to solve our problem.

In addition to the practical implementation and the experiments, the paper gives proof that infinitesimal flows, which are normalizing flows whose length tends to infinity, can define a class of posteriors that is asymptotically able to recover the true posterior distribution, solving therefore one of the most important problems of variational inference. This is another good indicator that the method should perform well in the finite case.

- TODO: describe in more depth, after we have results The paper’s experiments on the binarized MNIST dataset aim to demonstrate how variational inference with normalizing flows can outperform previous methods, compare the performance of different types of flows, and discuss the relation between the length of the flow and the quality of the resulting approximation.

- TODO: after we finish the extension In our report, we replicate the original contributions of the paper by running the same experiments on MNIST. Additionally, we consider another type of flow, the invertible convolutional flow .... Finally, we provide an open-source implementation of the above in our Github repository: [https://github.com/ATML-Group-12/normalising\\_flows](https://github.com/ATML-Group-12/normalising_flows)

# Chapter 1

TODO: Chapter 1

- our experiments
- results
- compare to paper results

## Chapter 2

TODO: Chapter 2

- propose our extension
- extension experiments
- extension results

### NICE

Using the idea of flows, the authors view the coupling layers in the NICE paper as a flow. A coupling layer  $f$  is a neural network layer with easy to compute inverse and a trivial Jacobian. For an input vector  $\mathbf{z} \in \mathbb{R}^D$ , we have

$$\begin{aligned} f(\mathbf{z}) &= (\mathbf{z}_A, g(\mathbf{z}_B, h(\mathbf{z}_A))) \\ f^{-1}(\mathbf{z}) &= (\mathbf{z}_A, g^{-1}(\mathbf{z}_B, h(\mathbf{z}_A))) \end{aligned}$$

where  $(A, B)$  is a partition of  $\{1, 2, \dots, D\}$ ,  $h$  is an arbitrary function with input size  $|A|$ , and  $g$  is a coupling law, a function that is invertible for the first argument given the second. In the paper,  $h$  is a neural network and  $g(a, b) = a + b$ , so the Jacobian is just the identity matrix. Since the determinant of the Jacobian is 1, the authors of (variation paper) classify it as a volume-preserving flow. The authors of the NICE paper do point out this issue, and introduce a diagonal scaling layer as the last layer of their models. This approach was not used in the original paper. Moreover, the authors introduced variants NICE-ortho and NICE-perm instead of the original directly, where random orthogonal and permutation matrices are applied to  $\mathbf{z}$  before the partition.

We note that the authors of the original paper have not made the choice of  $h$  clear in the experiments. For the replicating experiment 6.1, we have chosen  $h$  to be a ReLU network of 4 layers with hidden dimensions of 2.

# Conclusion

TODO: Conclusion

- critical evaluation of our work
- what we could've done with more time/resources