Proceedings of Machine Learning Research 157, 2021 ACML 2021

# Sinusoidal Flow: A Fast Invertible Autoregressive Flow (Supplementary Materials)

**Yumou Wei** YUMOUWEI@UMICH.EDU
*University of Michigan — Ann Arbor, MI USA*

**Editors:** Vineeth N Balasubramanian and Ivor Tsang

## Appendix A. Code

Our code is available at https://github.com/weiyumou/ldu-flow.

## Appendix B. Experiment Details

In this section, we describe the architectural and training-specific details about our Sinusoidal Flow applied to the two-dimensional toy dataset from Grathwohl et al. (2019), the five high-dimensional tabular datasets proposed by Papamakarios et al. (2017) and a version of MNIST (Lecun et al., 1998) and CIFAR-10 (Krizhevsky, 2009) pre-processed by Papamakarios et al. (2017). In addition, please refer to our code for more details.

### B.1. Toy and Tabular Datasets

Consistent with existing work, we use the masked linear layers from MADE (Germain et al., 2015) as the conditioners for the shift transformations. Table 1 summarises the experiment details. We use the Adam optimiser (Kingma and Ba, 2015), and either exponential decay or cosine annealing (Loshchilov and Hutter, 2016) for reducing learning rates over time.

Table 1: Architectural and training-specific details about our Sinusoidal Flow applied to the toy and tabular datasets. "Embedding dim" refers to the number of parallel sinusoidal functions bundled in the convex sum inside a sinusoidal transformer, denoted as $K$ in the main text.

| Data Set | 2D TOY | POWER | GAS | HEPMASS | MINIBOONE | BSDS300 |
|---|---|---|---|---|---|---|
| **Architectural** | | | | | | |
| # LDU blocks | 16 | 12 | 12 | 12 | 12 | 12 |
| # D-scale per block | 4 | 4 | 4 | 4 | 4 | 4 |
| Embedding dim ($K$) | 4 | 4 | 4 | 4 | 4 | 4 |
| Hidden size | 100 | 256-256 | 256-256 | 512-512 | 256-256 | 512-512 |
| Dropout | - | - | - | - | 0.3 | 0.1 |
| **Training-specific** | | | | | | |
| # Steps | 50K | 1.2M | 2M | 1M | 125K | 400K |
| Batch size | 128 | 512 | 128 | 128 | 128 | 512 |
| Learning rate | $1 \times 10^{-3}$ | $5 \times 10^{-4}$ | $1 \times 10^{-3}$ | $1 \times 10^{-3}$ | $5 \times 10^{-4}$ | $5 \times 10^{-4}$ |
| LR decay | - | cosine | 0.99 | 0.99 | cosine | cosine |
| Weight decay | - | - | $1 \times 10^{-5}$ | $5 \times 10^{-4}$ | $1 \times 10^{-3}$ | - |

### B.2. MNIST and CIFAR-10

For modelling image data, we use a miniature PixelCNN (van den Oord et al., 2016) as the conditioners for the shift transformations. The experiment details are shown in Table 2. The AdamW optimiser (Loshchilov and Hutter, 2019) appears to help stabilise training for CIFAR-10. We also reduce the learning rate by a factor of 0.9 whenever the validation loss stops improving for two epochs.

Table 2: Architectural and training-specific details about our Sinusoidal Flow applied to MNIST and CIFAR-10.

| DATA SET | MNIST | CIFAR-10 |
|---|---|---|
| **Architectural** | | |
| # LDU BLOCKS | 12 | 12 |
| # D-SCALE PER BLOCK | 4 | 4 |
| EMBEDDING DIM ($K$) | 4 | 4 |
| # FEATURE MAPS | 16 | 16 |
| # RESIDUAL BLOCKS | 2 | 2 |
| **Training-specific** | | |
| # STEPS | 200K | 140K |
| BATCH SIZE | 128 | 128 |
| LEARNING RATE | $1 \times 10^{-3}$ | $5 \times 10^{-4}$ |
| LR DECAY | COSINE | 0.9 |
| WEIGHT DECAY | $1 \times 10^{-2}$ | $1 \times 10^{-1}$ |

## Appendix C. Reconstruction Analysis

To develop further insights into the fast invertibility of our Sinusoidal Flow, we perform an analysis on how well it can reconstruct MNIST and CIFAR-10 images from the mapped $\mathbf{z}$ of some input images. A visually indiscernible reconstruction would require an accurate inversion of our Sinusoidal Flow. The left of Figure 1 shows that the reconstruction errors for both datasets decrease fairly quickly as more iterations are allowed for inversion at each LDU block. And in fact, with as few as 70 iterations allowed, we are able to obtain visually indiscernible reconstructions for both datasets as shown on the right of Figure 1.

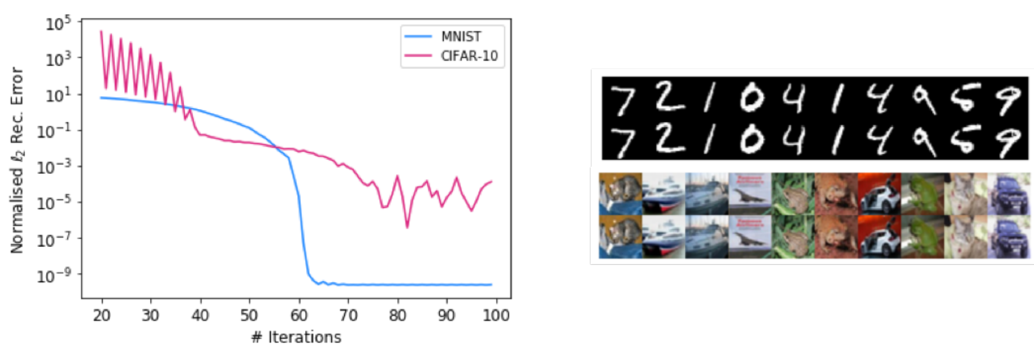


Figure 1: **Left**: $\ell_2$ reconstruction error versus maximum number of iterations allowed. **Right**: original images (odd rows) versus their reconstructions (even rows).

## References

Mathieu Germain, Karol Gregor, Iain Murray, and Hugo Larochelle. Made: Masked autoencoder for distribution estimation. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 881–889, Lille, France, 07–09 Jul 2015. PMLR.

Will Grathwohl, Ricky T. Q. Chen, Jesse Bettencourt, and David Duvenaud. Scalable reversible generative models with free-form continuous dynamics. In *International Conference on Learning Representations*, 2019.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, 2009.

Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Ilya Loshchilov and Frank Hutter. SGDR: stochastic gradient descent with restarts. *CoRR*, abs/1608.03983, 2016.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30, pages 2338–2347. Curran Associates, Inc., 2017.

Aaron van den Oord, Nal Kalchbrenner, Lasse Espeholt, koray kavukcuoglu, Oriol Vinyals, and Alex Graves. Conditional image generation with pixelcnn decoders. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.