Structured Conditional Continuous Normalizing Flows for Efficient Amortized Inference in Graphical Models

THE UNIVERSITY OF BRITISH COLUMBIA

Christian Weilbach, Boyan Beronov, William Harvey, Frank Wood

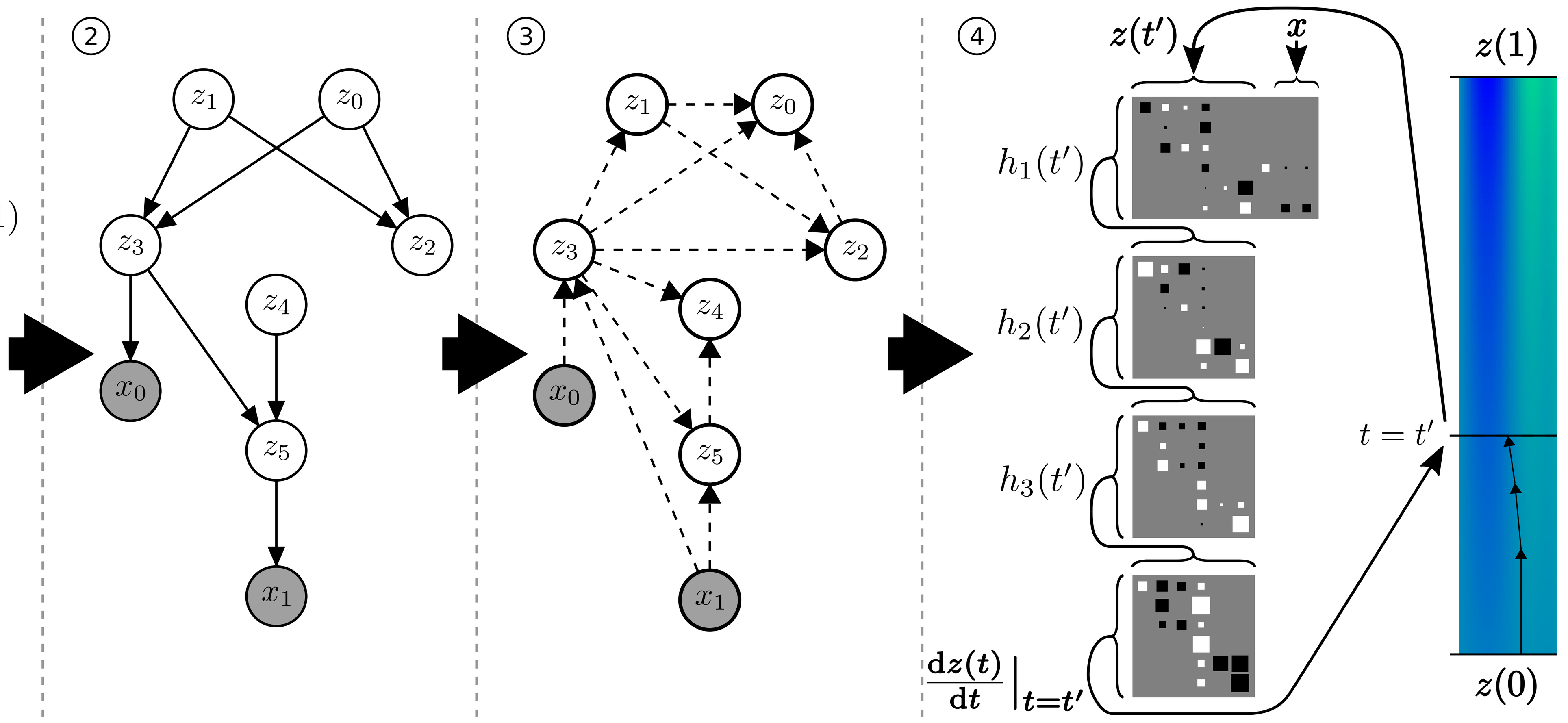
weilbach@cs.ubc.ca

Statistically faithful sparsity and *extended variational objective* lead to ***improved inference amortization*** at ***reduced computational cost***.

Program

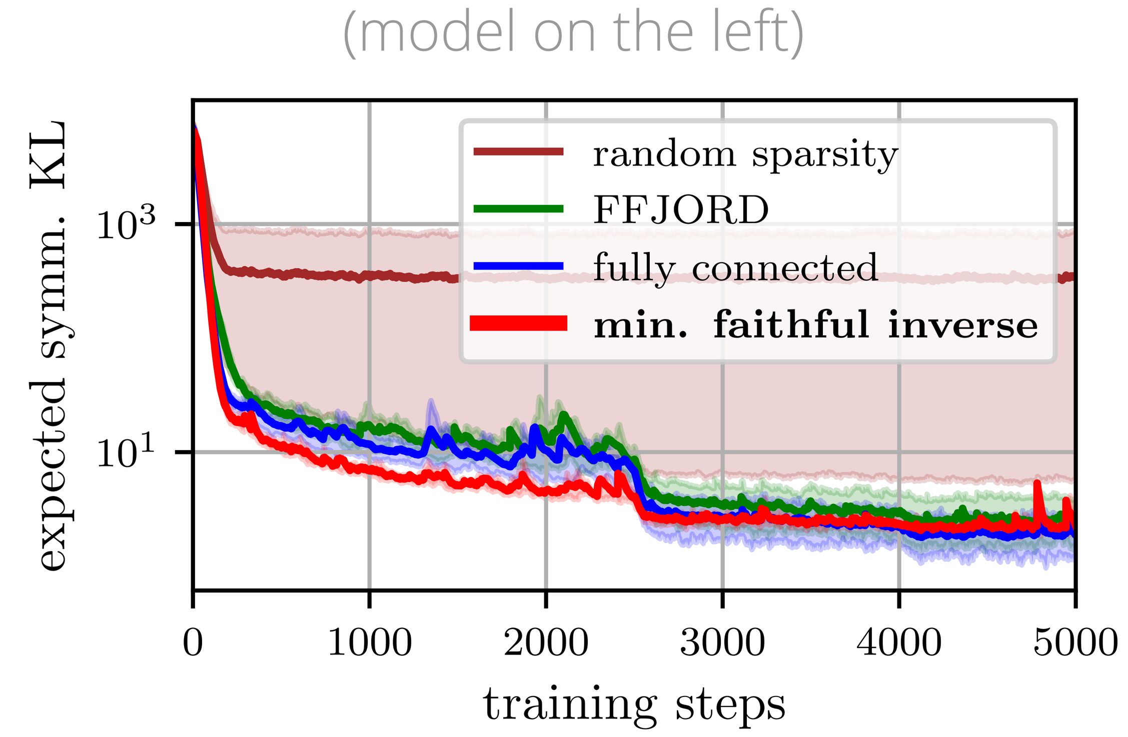
①

$z_0 \sim \mathcal{L}(5;1)$
$z_1 \sim \mathcal{L}(-2;1)$
$z_2 \sim \mathcal{N}(\tanh(z_0 + z_1 - 2.8); 0.1)$
$z_3 \sim \mathcal{N}(z_0 \cdot z_1; 0.1)$
$z_4 \sim \mathcal{N}(7;2)$
$z_5 \sim \mathcal{N}(\tanh(z_3 + z_4); 0.1)$
$x_0 \sim \mathcal{N}(z_3; 0.1)$
$x_1 \sim \mathcal{N}(z_5; 0.1)$

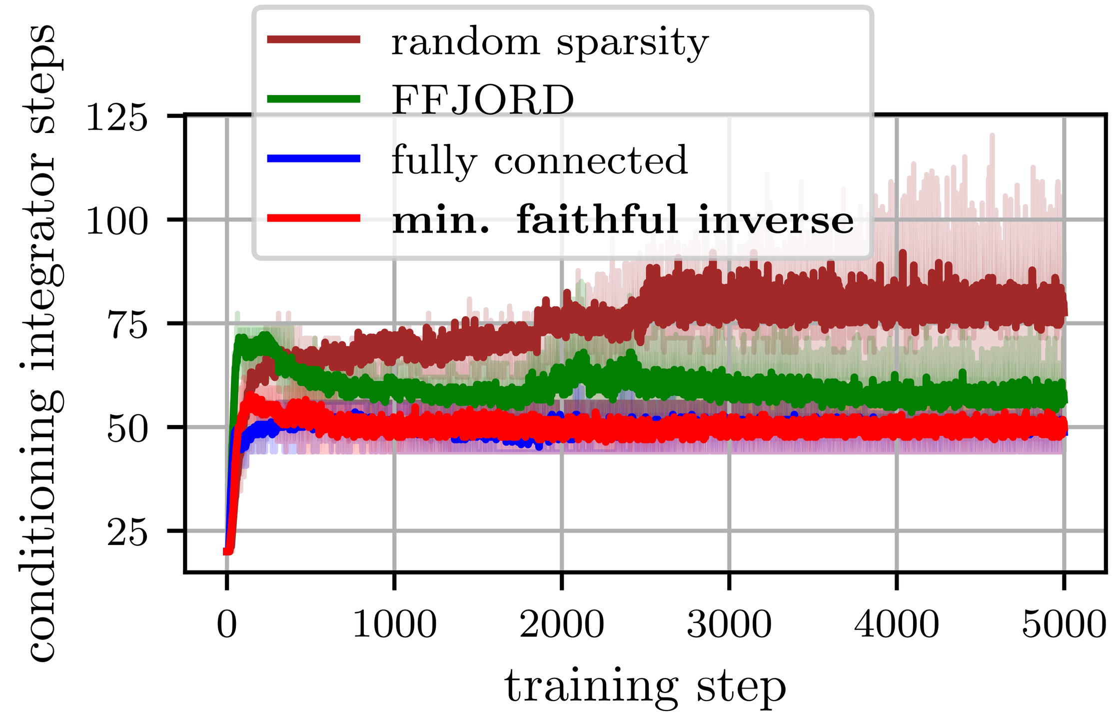


Effects of Sparsity

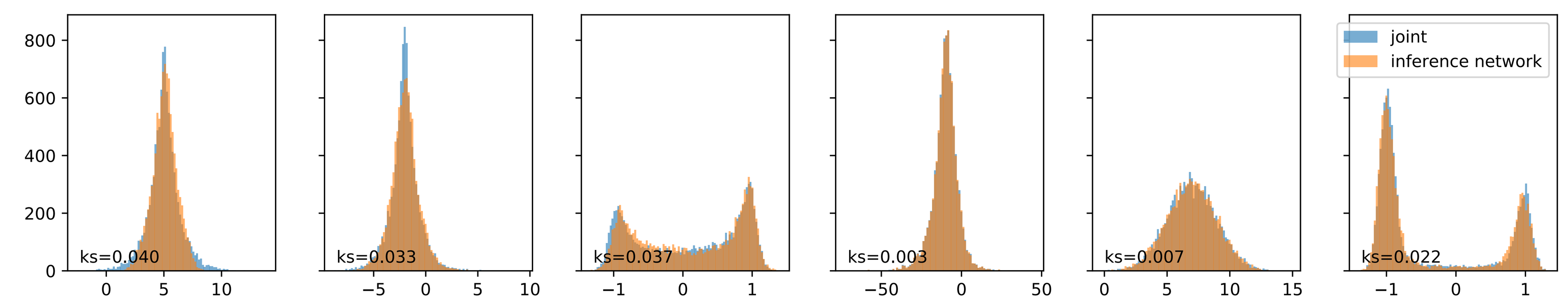
Comparable loss with < 50% parameters, compared to FFJORD [2] (model on the left)



Faithful sparsity is numerically stable (no increase in integration steps)



Comparison between marginals of *joint model* and *expected posterior* for latents z_0 to z_5



Sparse Neural ODE

Automated compiler pipeline for probabilistic programs (*contributions* [0])

① *Compile* into graphical model
② Faithfully ***invert*** graphical model using NaMI algorithm [1]
③ *Map* inverse graph onto *sparsity* of FFJORD [2] neural network layers
④ *Train* conditional continous [3] normalizing flow with *symmetric KL*

Continuous Normalizing Flow

Learn neural network f_Φ for volume preserving, invertible particle flow

$$\frac{\mathrm{d}}{\mathrm{d}t} z_t = f_\Phi(z_t, t; x)$$

$$\frac{\mathrm{d}}{\mathrm{d}t} \ln q_\Phi(z_t, t) = -\nabla_z \cdot f_\Phi(z_t, t; x)$$

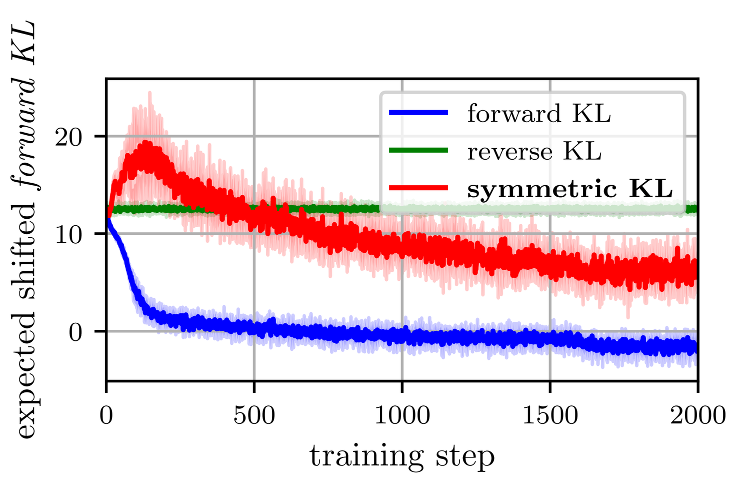
Sparsity Structure

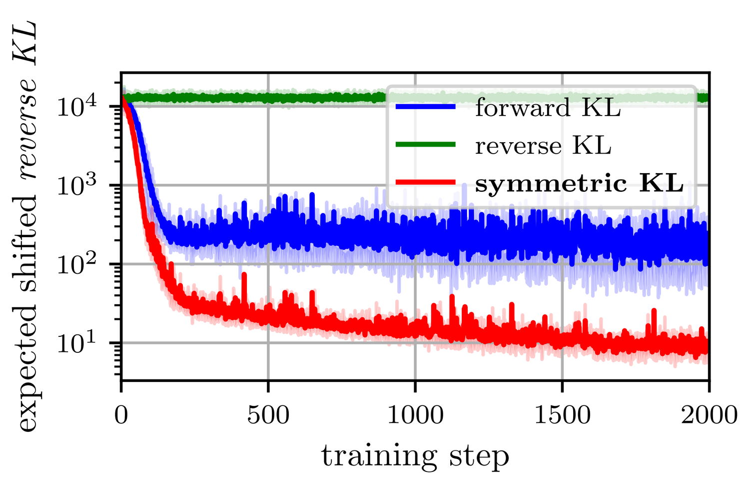
Encode sparsity H of inverse graph into each layer of f_Φ:

$$f_\Phi(z, t; x) = \big(h_{\Phi_L}(\cdot, t) \circ \cdots \circ h_{\Phi_1}(\cdot, t)\big)(z \oplus x)$$

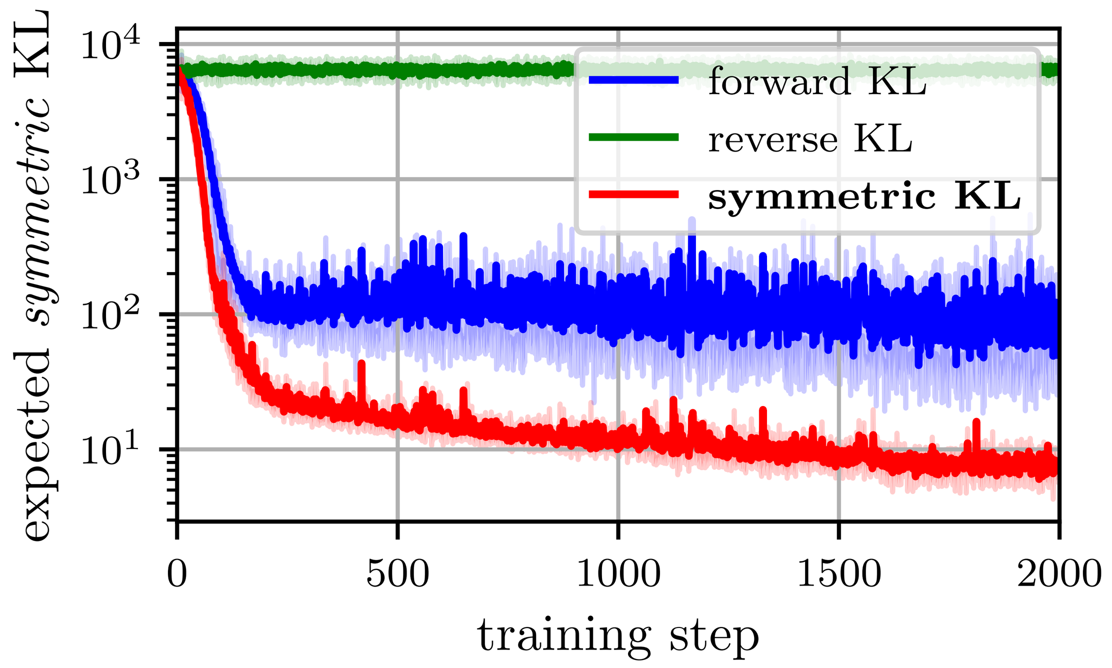
$$h_{\Phi_l}(\hat{z}, t) = \sigma\{(W_l \odot H)\hat{z} \odot \eta_{l,1}(t)\} + b_l \odot \eta_{l,2}(t)$$

Improved Objective: Symmetric KL



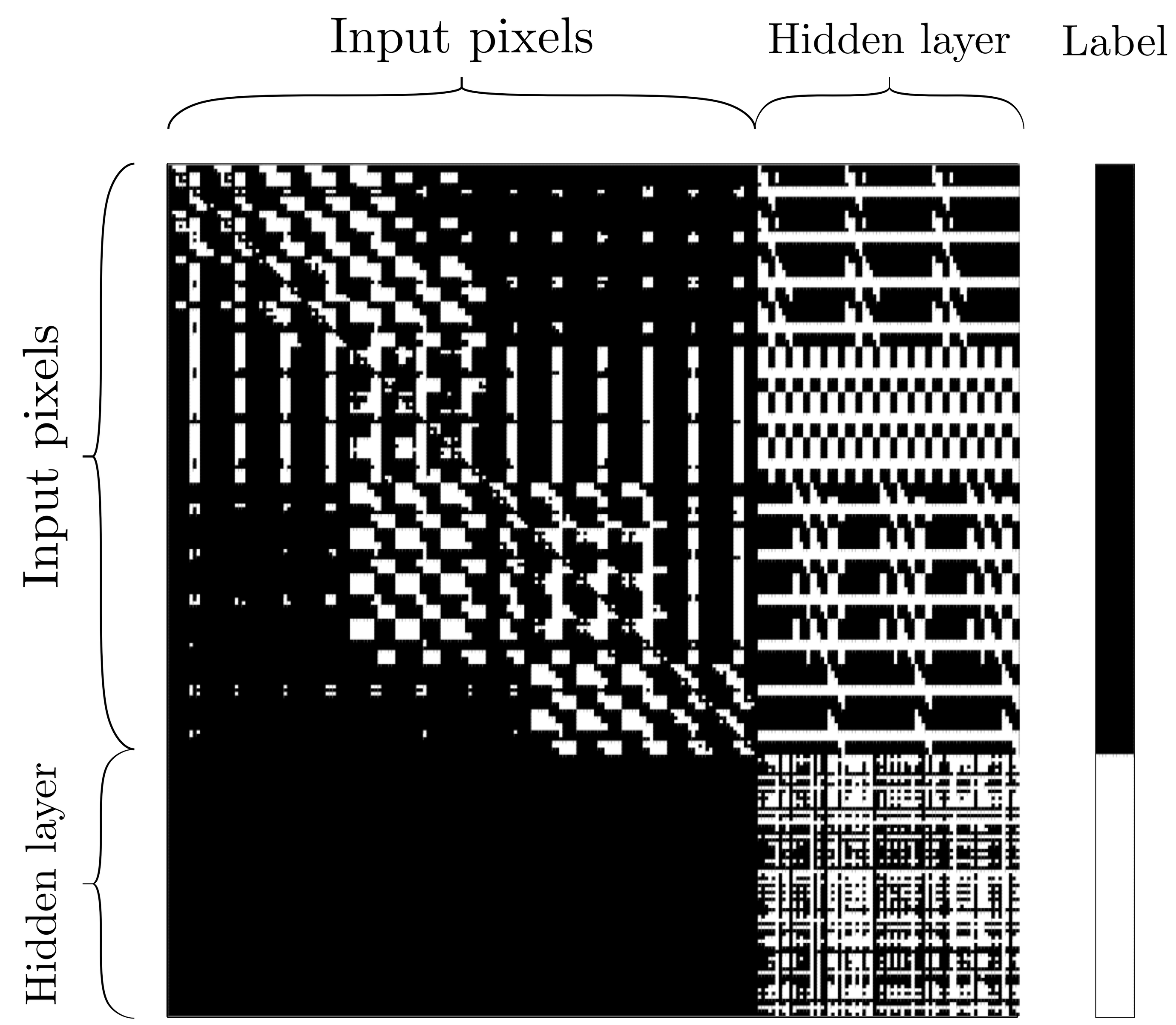


In comparison to *forward* or *backward* KL (above), optimizing expected *symmetric* KL loss (below) improves amortized sample efficiency (model above).

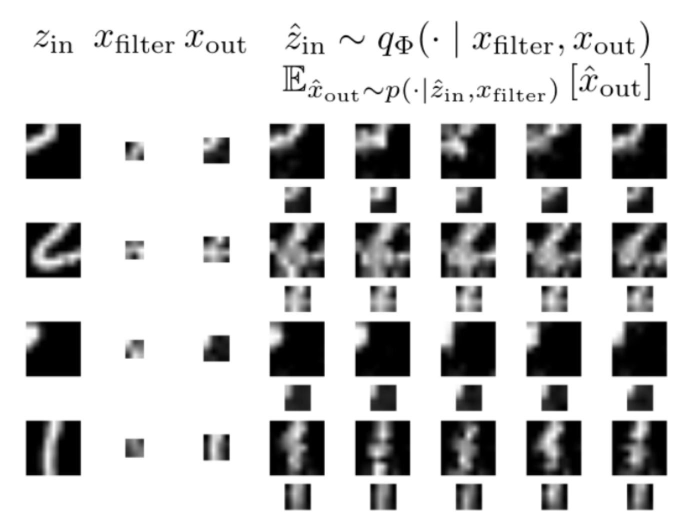


Generative Model from Stochastic Inverse of MLP Classifier (MNIST)

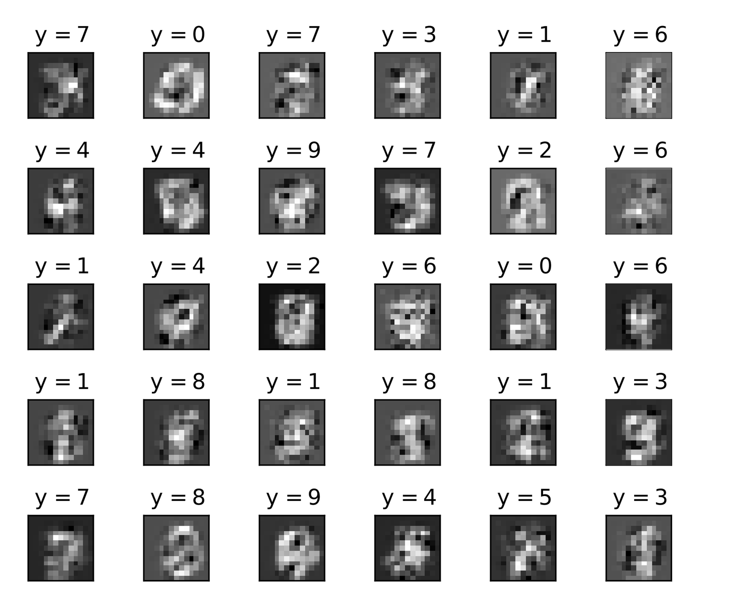
Minimally faithful inverse sparsity for MNIST generator



Stochastic inverse of a convolution layer & output reconstructions



Stochastic inverse of full classifier



References

[0] C. Weilbach et al. Structured Conditional Continuous Normalizing Flows for Efficient Amortized Inference in Graphical Models. AISTATS 2020.
[1] S. Webb et al. Faithful inversion of generative models for effective amortized inference. NeurIPS 2018.
[2] W. Grathwohl et al. FFJORD: free-form continuous dynamics for scalable reversible generative models. ICLR 2019.
[3] X. Chen et al. Particle Flow Bayes' Rule. ICML 2019.

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada CIFAR AI Chairs Program, Compute Canada, Intel, and DARPA under its D3M and LWLL programs.