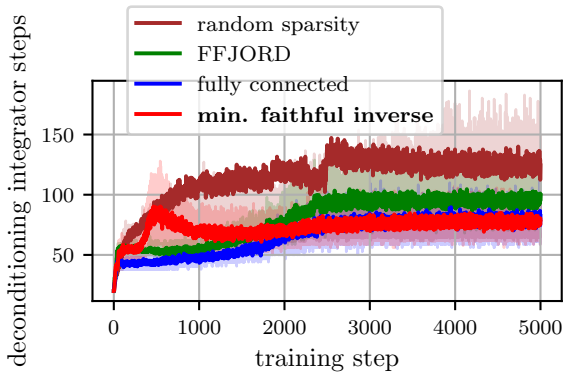
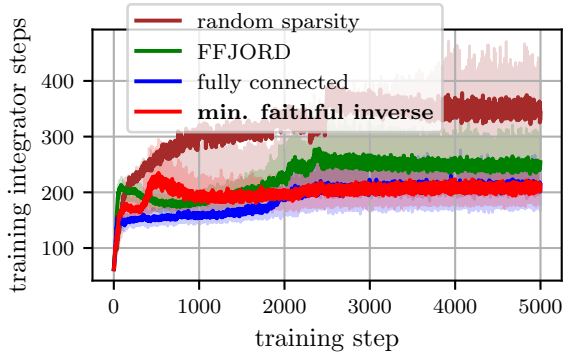
A Stability



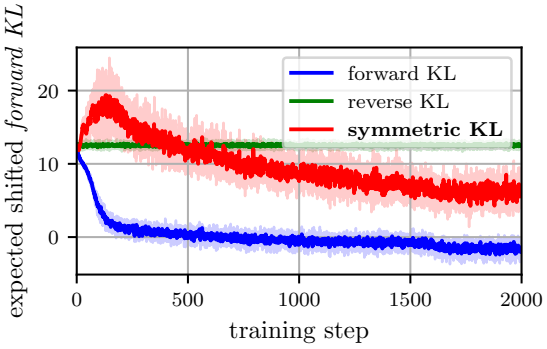
(a) Deconditioning.



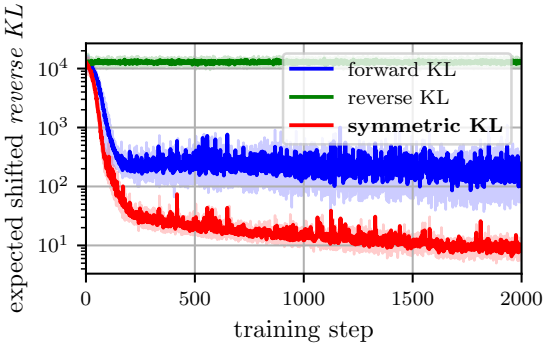
(b) Backpropagation (both)

Figure 10: Effect of sparsity patterns on the numerical stability and computation time during training. More iterations are worse. The faithfully inverted flow conditions the flow better than both the fully connected variant and the FFJORD baseline, both for deconditioning and the joint dynamics in training. It has some difficulty in the beginning to do stable conditioning, an issue that warrants further investigation. We again plot the median and a confidence band between the 16th and 84th percentiles over 10 runs.

B Loss components



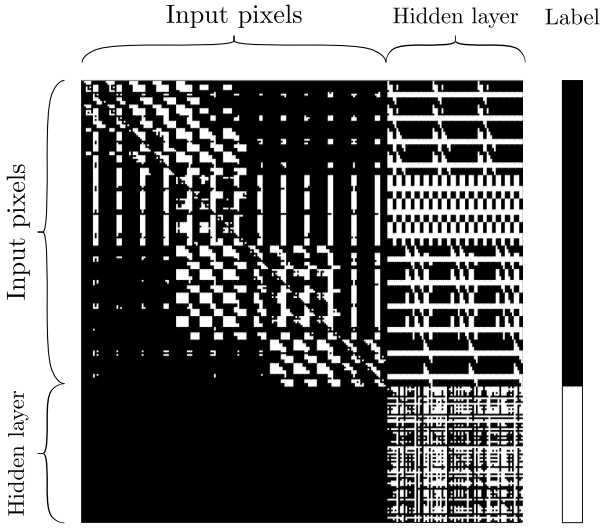
(a) Effect on forward KL of different optimization objectives.



(b) Effect on reverse KL of different optimization objectives.

Figure 11: Continuing Figure 6, correlations between different KL variants used as training loss and test metric.

C Inverted MNIST classifier



(a) Faithful inverse structure of the image classifier.

(b) Posterior samples of input images.

Figure 12: Building on Section 4.4, we consider the stochastic inversion of an entire image classifier. Its architecture consists of a convolution to a hidden layer with a ReLU activation and Gaussian dropout, followed by a linear mapping and softmax function.

D Gaussian state space model

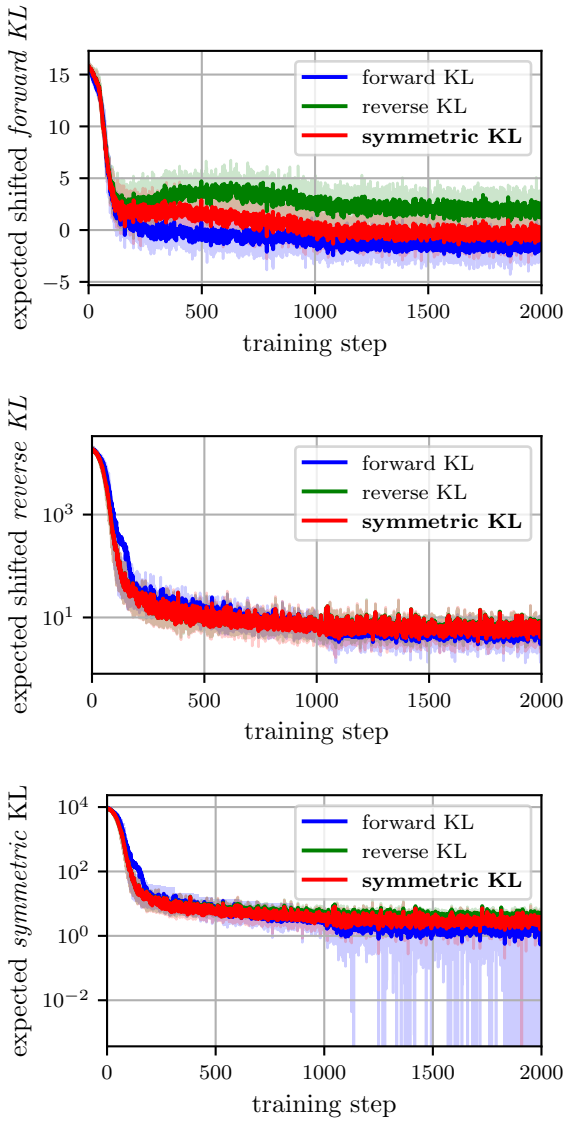


Figure 13: A simple Gaussian one dimensional state space model with 5 state transitions. Optimization using the reverse KL loss gives better performance than in the arithmetic circuit model in Figure 6, but is still worse than using either the forward or symmetric KL.