Faithful Model Inversion Substantially Improves Auto-encoding Variational Inference

Stefan Webb, Adam Goliński, Robert Zinkov, Yee Whye Teh, Frank Wood

Correspondence to: `stefan.webb@eng.ox.ac.uk`

UNIVERSITY OF OXFORD

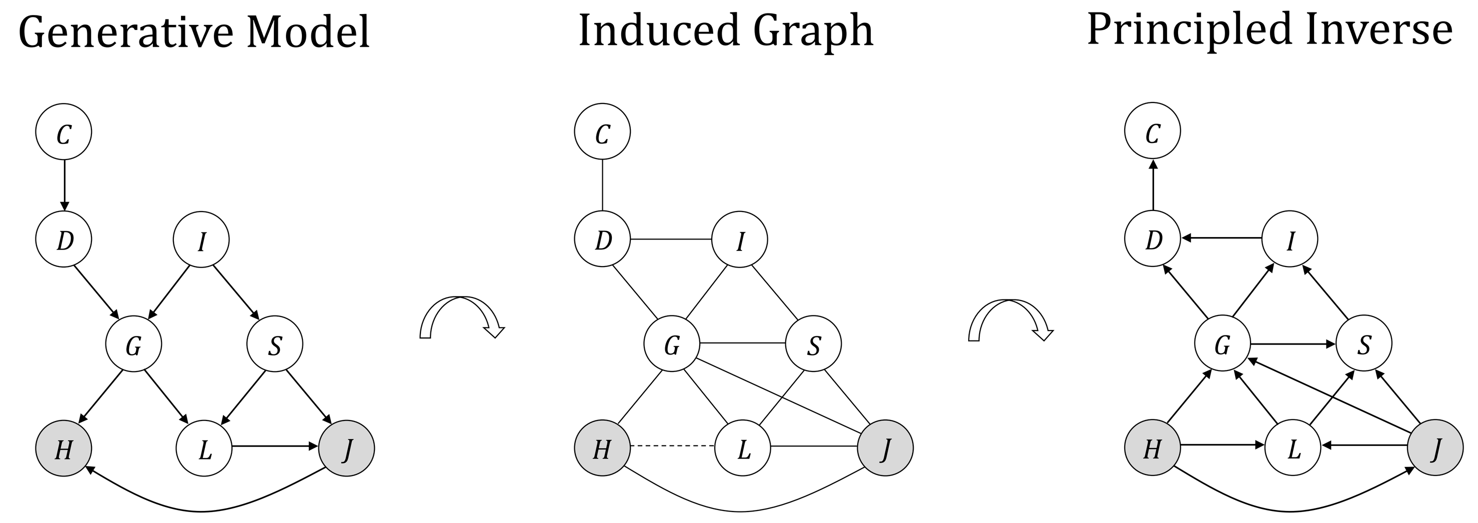
Abstract

- Our chief insight is that *the d-separation properties of the directed graphical model (DGM) structure of the forward model should be used to produce an encoder factorization that is faithful to the posterior*, in the sense that it does not impose independencies not present in the true posterior
- Auto-encoding variational inference requires the design of an encoder for learning either the model and/or the posterior
- Typically, the structure of the encoder is formed in an ad hoc way by simply reversing the edges in the DGM of the generative model
- We introduce a novel algorithm that given the factorization for a generative model, produces a faithful factorization for the posterior that is optimal in a certain technical sense

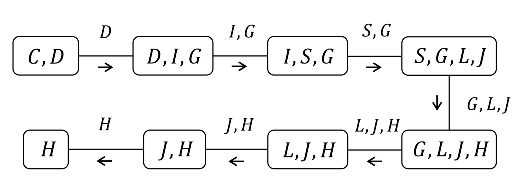
The CoMI Algorithm

INPUT: A graph for the forward model, and set of observed variables

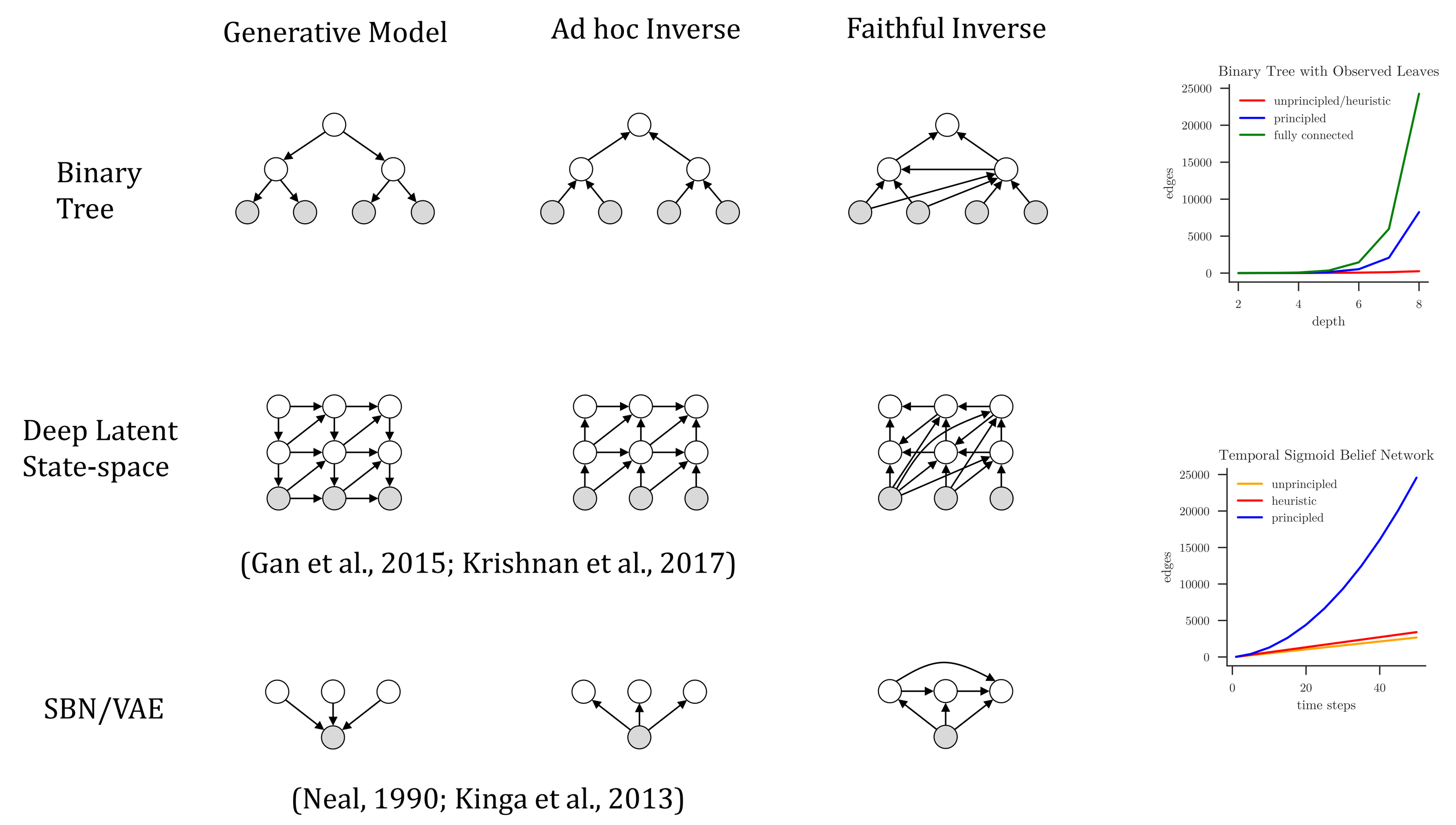
OUTPUT: A compact minimal I-map for the inverse model



- Simulates variable elimination (Koller and Friedman, 2009) on the forward model, eliminating latent variables first, then observed
- Chooses the elimination ordering by combining a topological ordering on the forward model with a min-fill heuristic
- The sepset property of the corresponding clique tree, equivalent to the induced graph, allows to us construct the inverse factorization



Some Examples



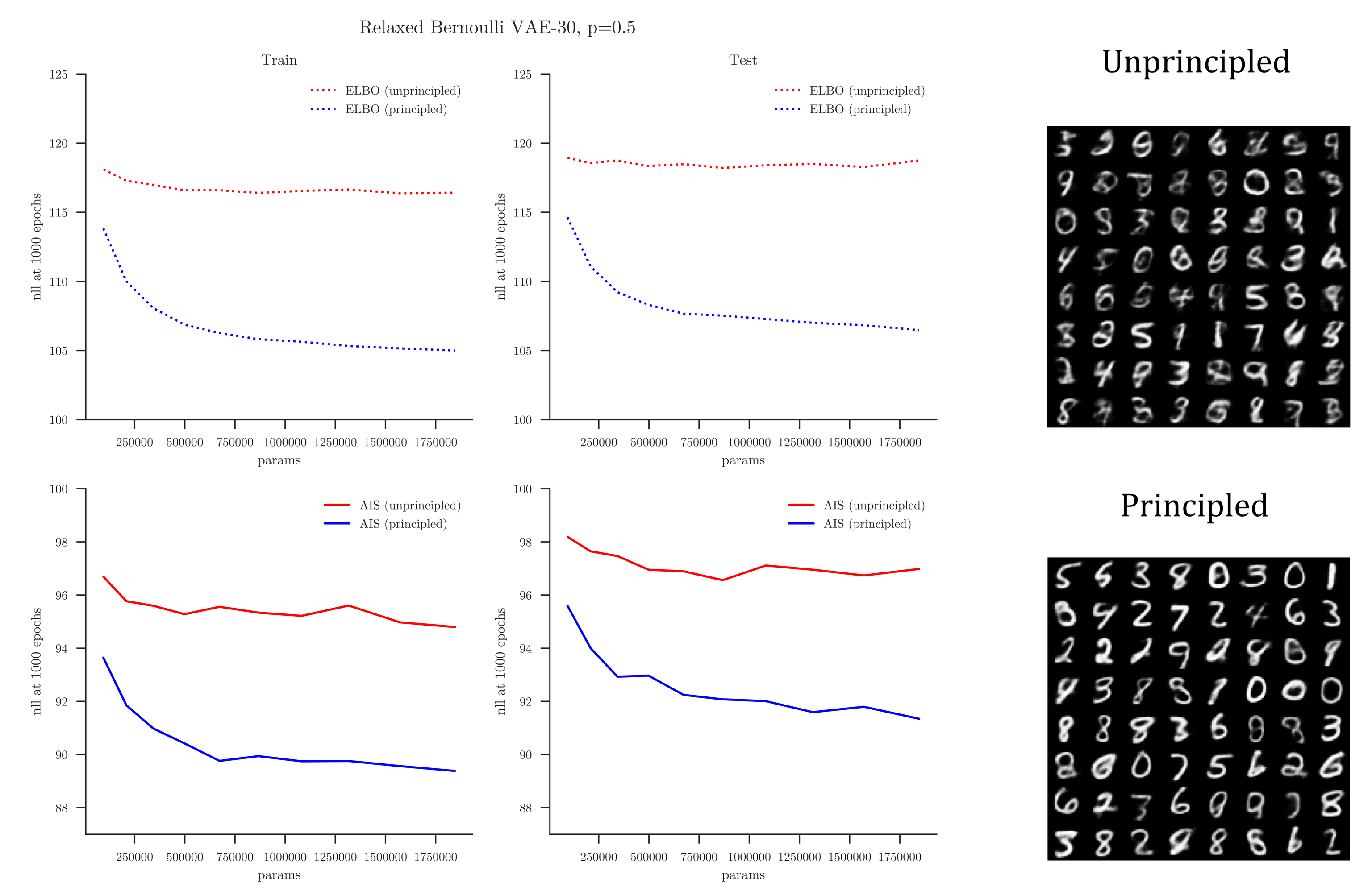
SW and AG gratefully acknowledge support from the EPSRC AIMS CDT through grant EP/L015987/2. RZ acknowledges support under DARPA D3M, under Cooperative Agreement FA8750-17-2-0093. FW was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1, DARPA PPAML through the U.S. AFRL under Cooperative Agreement FA8750-14-2-0006, an Intel Big Data Center grant, and DARPA D3M, under Cooperative Agreement FA8750-17-2-0093.`

AIMS Autonomous Intelligent Machines & Systems

EPSRC Engineering and Physical Sciences Research Council

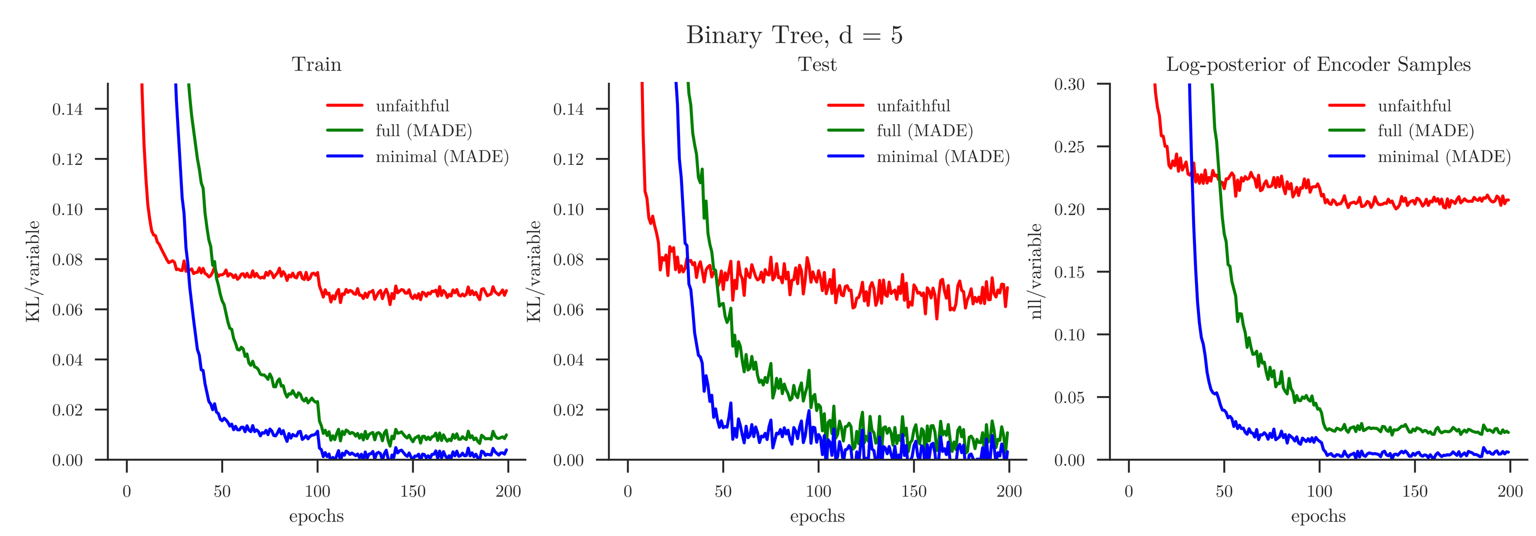
DARPA

Faithful Encoders Appear to Improve Model Learning



- A faithful encoder outperforms an ad hoc one *at every capacity level*
- After 10,000 epochs, reaches 86.5 nats, equivalent to results from an unfaithful encoder with 200 latent units and a multi-sample objective with 50 importance weighted samples (Maddison Et Al., 2016; Jang et al., 2016)

Faithful Encoders Recover the True Posterior



- The unfaithful encoder is unable to learn the true posterior
- A fully connected MADE encoder (Germain et al., 2015) and an encoder based on a novel MADE variant for trees (both faithful, the later minimal) are able to learn the posterior
- The minimally faithful encoder learns quicker than the fully connected one. Why?

Gan, Zhe, Li, Chunyuan, Henao, Ricardo, Carlson, David E, and Carin, Lawrence. Deep temporal sigmoid belief networks for sequence modeling. *In Advances in Neural Information Processing Systems, pp. 2458–2466, 2015.*

Germain, Mathieu, Gregor, Karol, Murray, Iain, and Larochelle, Hugo. Made: masked autoencoder for distribution estimation. *In Proceedings of the 32nd International Conference on Machine Learning (ICML-15), pp. 881–889, 2015.*

Jang, Eric, Gu, Shixiang, and Poole, Ben. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144, 2016.*

Kingma, Diederik P and Welling, Max. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114v10 [stat.ML], 2013.*

Koller, Daphne and Friedman, Nir. *Probabilistic Graphical Models.* MIT Press, 2009. ISBN 9780262013192.

Krishnan, Rahul G, Shalit, Uri, and Sontag, David. Structured inference networks for nonlinear state space models. *In AAAI, pp. 2101–2109, 2017.*

Maddison, Chris J, Mnih, Andriy, and Teh, Yee Whye. The concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712v3 [cs.LG], 2016.*

Neal, Radford M. Learning stochastic feedforward networks. *Department of Computer Science, University of Toronto*, 1990.