SIReN-VAE: Leveraging Flows and Amortized Inference for Bayesian Networks

Jacobie Mouton [1] Steve Kroon [1,2]

[1]Computer Science Division, Stellenbosch University [2]National Institute for Theoretical and Computational Sciences (NITheCS) South Africa

Contributions

- This work explores incorporating graphical flows into VAEs.
- This is achieved by extending both the decoder prior and inference network with graphical residual flows—efficiently invertible residual flows that encode conditional independence by masking the weight matrices of the flow's residual blocks.
- Experiments show our model's potential for more efficiently learning in data-sparse settings.

Graphical Residual Flows

Graphical flows add structure to normalizing flows, by encoding non-trivial variable dependencies from a Bayesian network (BN). Graphical Residual Flows (GRFs) [1] incorporate incorporate the graphical structure of a BN into a residual flow by suitably masking the weight matrices of each residual block before applying spectral normalization (indicated by subscript s).

Two GRFs are used in the SIReN-VAE model, each in a different direction. GRF_n (with Jacobian J_n) is **normalizing** and is **used to evaluate the prior in the VAE**.

GRF (normalizing): $f_{t+1}(\mathbf{z}^{(t)}) := \mathbf{z}^{(t)} + (W_2 \odot M_2)_s \cdot h((W_1 \odot M_1)_s \cdot \mathbf{z}^{(t)} + b_1) + b_2$

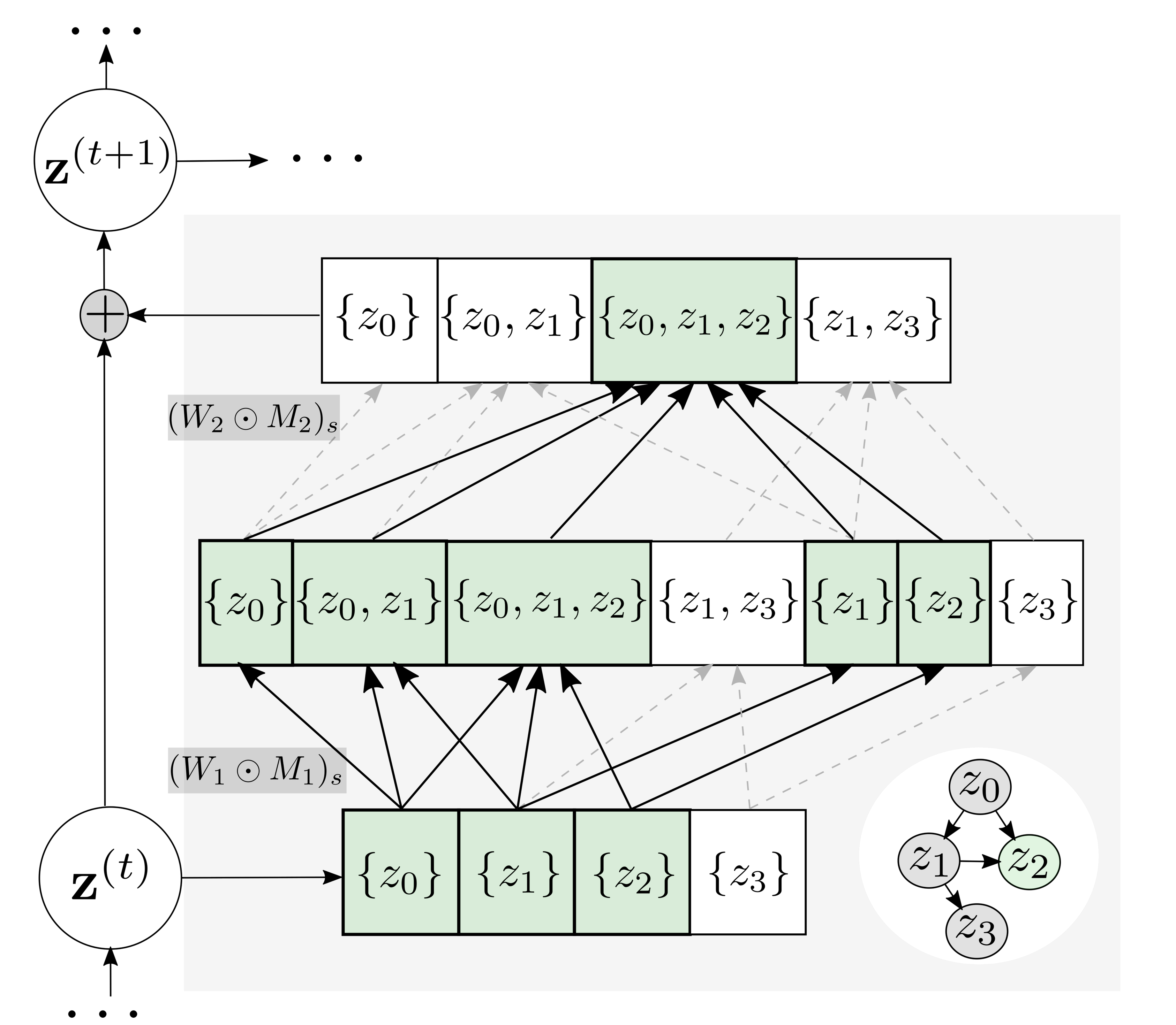

Figure 1. The update to $\mathbf{z}$ at flow step t. Edges removed by the masks M_1 and M_2 are not shown. The remaining edges encode the graphical structure of the given BN. **Mask construction:** Each output node is assigned the set consisting of the associated variable and its parents in the BN. Each hidden node is assigned a set consisting of either a single variable or a variable and its parents. Edges are retained when the set in the next layer is a superset of the set in the previous layer. For example, the transformation applied to variable z_2 is only conditioned on its parents within the residual block (shown here with solid arrows).

GRF_g (with Jacobian J_g) is **generative** and **used to approximate the posterior in the VAE** by conditioning on an observation. The structures of GRF_n and GRF_g are related by faithful inversion of the BN [2].

GRF (generating): $f_{t+1}(\mathbf{z}^{(t)}; \mathbf{x}) := \mathbf{z}^{(t)} + (W_2 \odot M_2)_s \cdot h((W_1 \odot M_1)_s(\mathbf{z}^{(t)} \oplus \mathbf{x}) + b_1) + b_2$

Acknowledgements

We would like to thank DeepMind for financially supporting J. Mouton through a scholarship.

References

[1] J. Mouton and S. Kroon. Graphical residual flows. In *ICLR Workshop on Deep Generative Models for Highly Structured Data*, 2022.

[2] S. Webb, A. Golinski, R. Zinkov, N. Siddharth, T. Rainforth, Y. W. Teh, and F. Wood. Faithful inversion of generative models for effective amortized inference. In *NeurIPS*, 2018.

Structured Invertible Residual Network VAE

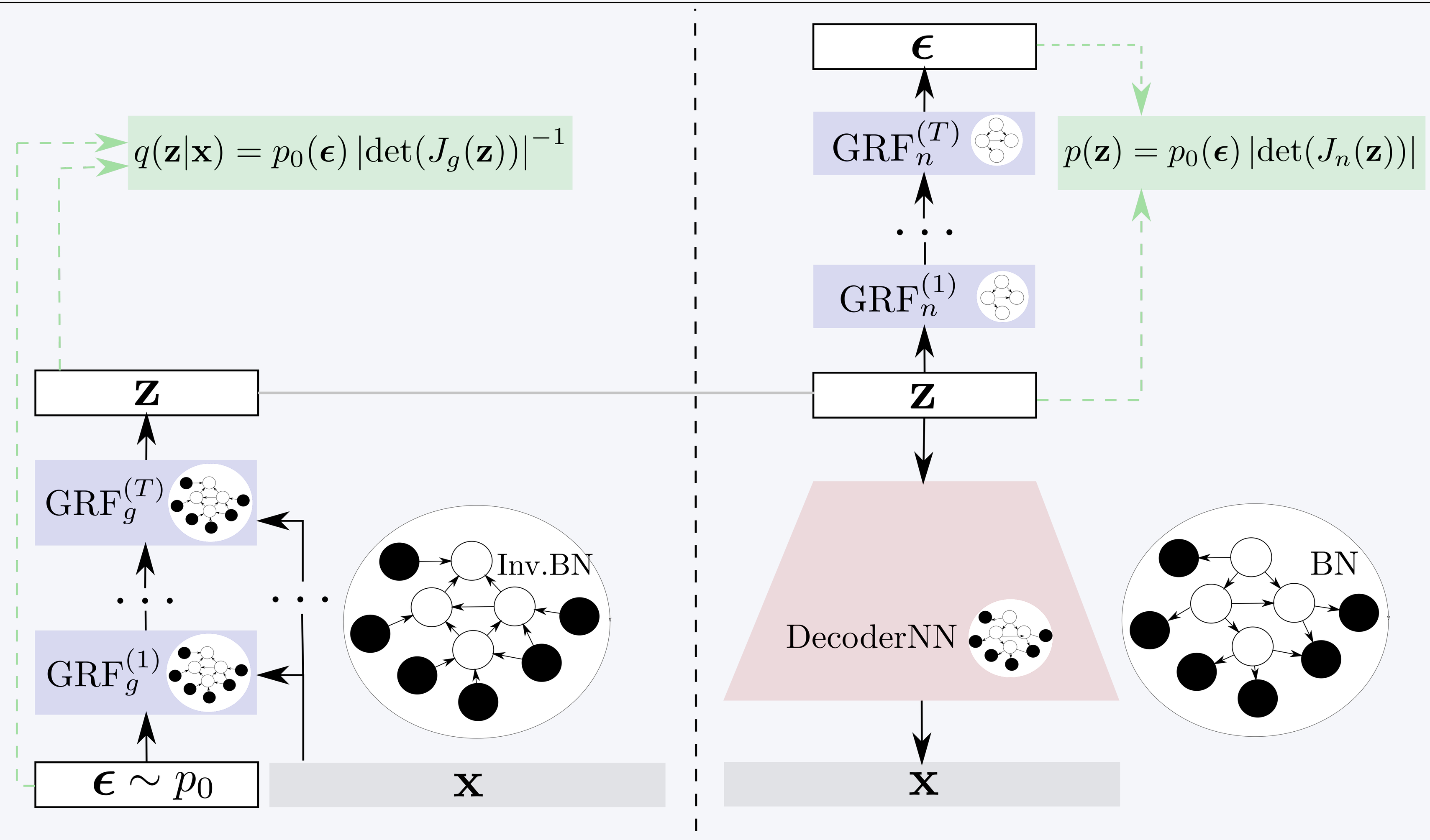

Figure 2. SIReN-VAE encodes the BN's graphical structure into the decoder (right) via masking of the normalizing GRF and decoder neural network weights. The inverted BN structure is similarly encoded in the inference network's generating GRF (left).

We assume access to a BN graph specifying the dependencies between observed $\mathbf{x}$ and latent $\mathbf{z}$ variables. Our goal is to incorporate this dependency information into the VAE's encoder and decoder. This means that the VAE's likelihood $p(\mathbf{x}|\mathbf{z})$ and prior $p(\mathbf{z})$ should factorize as specified by the BN. Approximating the posterior $p(\mathbf{z}|\mathbf{x})$ while taking the knowledge from the BN into account requires suitably inverting the BN such that one obtains edges from $\mathbf{x}$ to $\mathbf{z}$ [2].

Benefit of Incorporating Graphical Structures in Data-sparse Settings

Table 1. Negative Log-likelihood (NLL) and reconstruction error (RE) results. The number of observed (D) and latent (K) variables, as well as the number of edges (E) in the datasets' associated BNs are provided. Subscript ind indicates a model that encodes conditional independence between all latent variables in the decoder, FC indicates a fully-connected structure and $true$ indicates a model that encodes the BN faithfully.

					$2 \times \lvert\mathcal{G}\rvert$ training samples		$100 \times \lvert\mathcal{G}\rvert$ training samples	
	D	K	E	Model	NLL	RE	NLL	RE
ECoil70	29	15	59	VAE	$42.99_{\pm 1.50}$	$6.21_{\pm .32}$	$36.95_{\pm .01}$	$4.68_{\pm .01}$
				SIReN-VAE$_{\text{ind}}$	$43.77_{\pm 0.16}$	$6.32_{\pm .08}$	$35.84_{\pm .23}$	$4.07_{\pm .10}$
				SIReN-VAE$_{\text{FC}}$	$42.91_{\pm 1.04}$	$6.03_{\pm .29}$	$\mathbf{35.77}_{\pm \mathbf{.23}}$	$4.05_{\pm .09}$
				SIReN-VAE$_{\text{true}}$	$\mathbf{38.98}_{\pm \mathbf{0.81}}$	$\mathbf{4.95}_{\pm \mathbf{0.22}}$	$36.22_{\pm .19}$	$\mathbf{3.88}_{\pm \mathbf{.10}}$
Arth150	67	40	150	VAE	$74.13_{\pm 3.27}$	$6.13_{\pm .70}$	$38.92_{\pm .06}$	$4.50_{\pm .02}$
				SIReN-VAE$_{\text{ind}}$	$42.60_{\pm 0.38}$	$4.85_{\pm .02}$	$\mathbf{38.84}_{\pm \mathbf{.11}}$	$4.45_{\pm .02}$
				SIReN-VAE$_{\text{FC}}$	$42.15_{\pm 0.00}$	$4.82_{\pm .00}$	$39.06_{\pm .07}$	$4.49_{\pm .01}$
				SIReN-VAE$_{\text{true}}$	$\mathbf{42.06}_{\pm \mathbf{0.40}}$	$\mathbf{4.80}_{\pm \mathbf{.01}}$	$38.86_{\pm .22}$	$\mathbf{4.42}_{\pm \mathbf{.04}}$
Magic-Irri	5	59	102	VAE	$15.70_{\pm 1.89}$	$11.95_{\pm 2.58}$	$10.18_{\pm .21}$	$\mathbf{7.30}_{\pm \mathbf{.60}}$
				SIReN-VAE$_{\text{ind}}$	$14.66_{\pm 1.08}$	$16.66_{\pm 1.40}$	$10.03_{\pm .00}$	$9.12_{\pm .01}$
				SIReN-VAE$_{\text{FC}}$	$12.41_{\pm 3.34}$	$12.05_{\pm 3.99}$	$10.03_{\pm .00}$	$9.13_{\pm .01}$
				SIReN-VAE$_{\text{true}}$	$\mathbf{10.83}_{\pm \mathbf{0.64}}$	$\mathbf{11.13}_{\pm \mathbf{1.95}}$	$\mathbf{10.02}_{\pm \mathbf{.01}}$	$8.91_{\pm .29}$

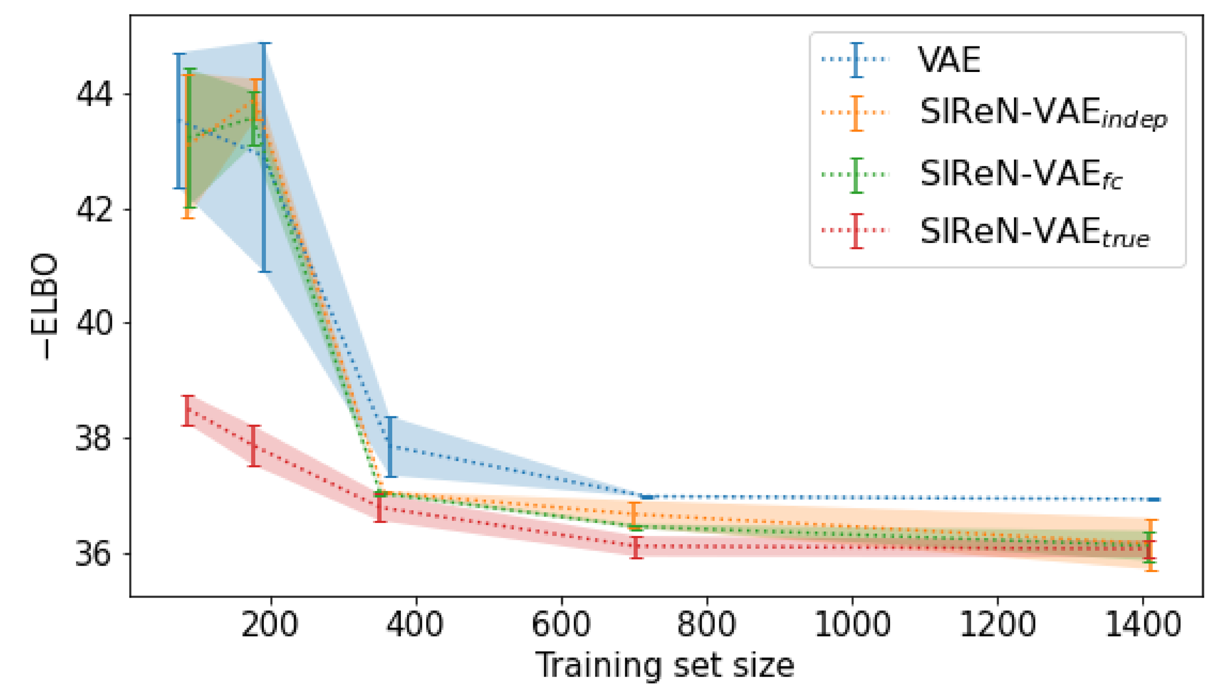

Figure 3. Negative ELBO (lower is better) vs training set size for the ECoil70 dataset. Error bars show one standard deviation from the mean.

Please scan the QR-code for additional results in the paper.

Paper: SIReN-VAE

Paper: Graphical Residual Flows