NeurIPS 2023 supplementary material

Table 1: Walltime results (in minutes, rounded to the nearest minute) of the runtime of different approaches on a single 40GB A100 NVIDIA GPU. The N/A fields indicate that the corresponding method cannot be run within the memory constraints of a single GPU. Our method converges faster while being scalable w.r.t. DIBS, the nonlinear Bayesian causal discovery baseline. Other methods (BGES and DDS), while maybe faster, perform much worse in terms of uncertainty quantification, and are either linear or quasi-Bayesian models.

	d=30	d=50	d=70	d=100
BaDAG (**Ours**)(Bayesian, Nonlinear)	171	238	261	448
DIBS (Bayesian, Nonlinear)	187	350	N/A	N/A
BGES (Quasi-Bayesian, Linear)	2	3	6	11
BCD (Bayesian, Linear)	252	328	418	600
DDS (Quasi-Bayesian, Nonlinear)	92	130	174	N/A

Table 2: Posterior inference results for $d > 100$ variables. BayesDAG performs exceedingly well in terms of $\mathbb{E}$-SHD and slightly better in terms of NLL. With a single 40GB A100 GPU, none of the other Bayesian baselines, apart from our BayesDAG, were capable to be able to run for $d > 100$, demonstrating the superior scalability of our approach. BGES, which is strictly not a Bayesian method, scales reasonably well due to its linearity, but we found it to predict a lot more edges than the ground truth.

		d=150			d=200		
		$\mathbb{E}$-SHD	Edge F1	NLL	$\mathbb{E}$-SHD	Edge F1	NLL
ER	BGES	921.56	0.33	247.92	1495.04	0.27	325.88
	BaDAG	540.40	0.17	269.37	713.38	0.14	348.95
SF	BGES	683.54	0.36	237.71	1128.82	0.29	316.82
	BaDAG	334.73	0.26	231.71	623.42	0.13	330.57