PACO: Partial-order-Augmented Continuous Optimization for Differentiable Causal Discovery

Xiaoxuan Li
UNSW Sydney
z5261241@ad.unsw.edu.au

Junda Wu
Adobe Research
jundaw@adobe.com

Julian McAuley
University of California San Diego
jmcauley@ucsd.edu

Lina Yao
UNSW Sydney
lina.yao@unsw.edu.au

Tong Yu*
Adobe Research
tyu@adobe.com

Abstract

Differentiable structure learning transforms DAG discovery from a combinatorial search into a continuous optimization problem, enabling gradient-based recovery of causal structures. In many applications, researchers possess partial-order priors over the variables (e.g., from biological signaling cascades or temporal ordering), which substantially narrow the space of plausible causal graphs. Existing approaches integrate such priors by decomposing the prior order graph into a collection of paths and evaluating prior-aware acyclicity constraints path by path; this is correct but its constraint-evaluation cost scales linearly with the number of induced maximal paths, a bottleneck for multi-chain priors. We propose PACO, Partial-order-Augmented Continuous Optimization, which encodes the same partial-order compatibility through a single masked-reachability matrix, that masks the matrix exponential of the learned graph by the strict transitive closure of the prior. The constraint and its Frechet-adjoint gradient are computed in a single $2d \times 2d$ block matrix exponential whose dominant cost is independent of the number of induced prior paths. We prove that PACO has the same exact zero-level feasible set as the path-decomposition formulation. Across synthetic and real-data experiments under matched priors and dataset seeds, PACO matches path-decomposition on structural recovery and substantially reduces wall-clock cost in multi-chain prior settings.

1 Introduction

Causal discovery aims to recover a directed acyclic graph (DAG) from data. In many scientific settings, data are not the only source of information: researchers often possess partial ordering knowledge, such as that a treatment or earlier cohort measurement precedes a later outcome [Bang et al., 2026], an upstream molecular event precedes a downstream signaling response [Sachs et al., 2005], or a temporal process constrains the direction of influence [Spirtes and Zhang, 2016]. Such knowledge is rarely a complete causal graph. More often, it is a partial order: a sparse set of precedence relations that should be respected by any plausible learned graph.

Differentiable structure learning methods such as NOTEARS [Zheng et al., 2018] convert DAG learning into continuous optimization by replacing the combinatorial acyclicity constraint with a smooth equality constraint. This line of work has enabled the use of standard gradient-based optimization tools for linear and nonlinear causal discovery. However, incorporating partial-order priors into this framework remains subtle. A prior edge $i \prec j$ does not simply mean that the learned

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

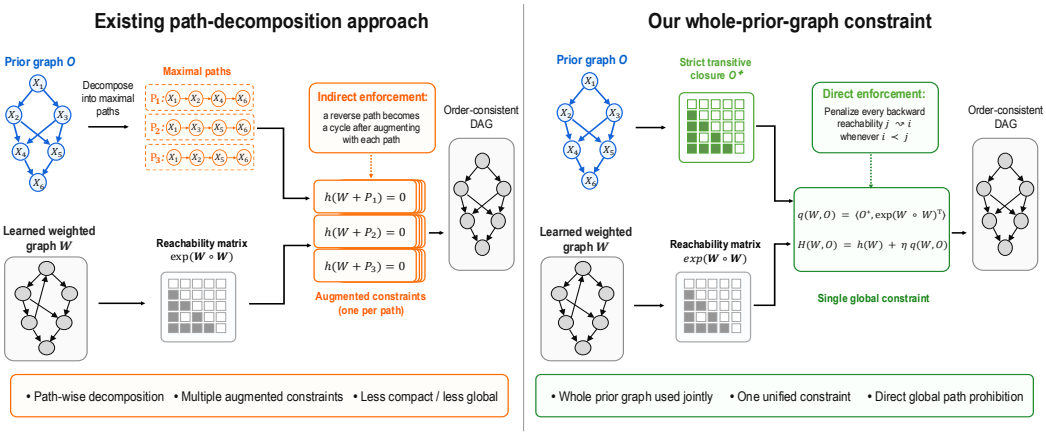

Figure 1: Overview of PACO. **Left**: existing path-decomposition methods [Ban et al., 2024] decompose O into maximal paths $\mathcal{P}(O^-)$ (P_1, P_2, P_3 depicted; the full set may contain more) and impose one augmented acyclicity constraint per path, depicted here as $h(W + P_p) = 0$ for visual clarity (the full path-augmented adjacency $\mathcal{A}(W, p)$ is defined in Section 2). **Right**: PACO forms a single global constraint $H_{\mathrm{PACO}} = h(W) + \eta\, q_{\mathrm{PACO}}(W; O)$, masking the reachability matrix $\exp(W \circ W)$ by the strict transitive closure O^+ to penalize every learned reverse path $j \rightsquigarrow i$ whenever $i \prec j$. Both admit the same family of feasible DAGs (Section 3).

graph must include $i \to j$; rather, it means that the learned graph should not contain edges or directed paths that make j precede i. Encoding this is non-trivial because partial-order compatibility is a path-level property rather than an edge-level one: a simple edge mask cannot rule out indirect reverse paths, while merging the prior graph with the learned graph inside a standard acyclicity constraint can be overly restrictive, rejecting graphs that do not violate any supplied order. Recent path-decomposition methods, which we refer to as PPO baselines following Ban et al. [2024], address this issue by mapping a partial order into a collection of path-wise graph constraints (Figure 1, left). This yields a principled way to use partial orders, but each constraint evaluation requires one acyclicity computation per induced path, making the cost scale linearly with the number of maximal paths $|\mathcal{P}(O^-)|$. Path count depends on the topology of the prior: a total order has only one maximal path, whereas a tiered partial order can induce many while leaving some variable pairs unordered. In either case, ordering information alone does not determine which causal edges exist. When the prior has many interacting precedence relations, the induced path count grows quickly and the per-step cost becomes a bottleneck; an efficient differentiable implementation in this regime has remained underexplored.

To efficiently handle the multi-path regime, we propose PACO, Partial-order-Augmented Continuous Optimization, a masked-reachability formulation for differentiable causal discovery whose constraint value and Frechet-adjoint gradient have cost independent of the number of induced prior paths (Figure 1, right). PACO starts from the observation that the matrix exponential used in continuous DAG learning contains reachability information about the learned graph. Let W be the learned weighted adjacency matrix and let O^+ denote the strict transitive closure of the prior partial order. PACO penalizes the reverse paths forbidden by the prior:

$$q_{\mathrm{PACO}}(W; O) = \left\langle (O^+)^\top, \exp(W \circ W) \right\rangle .$$

This term is zero exactly when, for every prior order $i \prec j$, the learned graph contains no directed path from j back to i. Combined with the standard NOTEARS acyclicity constraint $h(W)$, PACO therefore encodes graph-space partial-order compatibility through a single masked-reachability term. Ban et al. [2024] already formulated partial-order compatibility through a direct reachability constraint, but adopted path decomposition after identifying computational difficulties with the direct formulation. PACO revisits this approach using masked matrix-exponential reachability, preserving the exact feasible set. The constraint value and its gradient are computed jointly via a single Frechet adjoint, with dominant matrix-function cost independent of the number of induced maximal prior paths. Path-decomposition baselines, by contrast, sum acyclicity constraints over augmented matrices of

the form $W + P_p$ for each maximal path p; both formulations characterize the same partial-order feasible set (Section 3). A subsequent work [Ban et al., 2025] addresses path-*existence* constraints; their order-guided initialization subroutine uses a truncated polynomial reachability sum, whereas PACO targets path-absence directly with a matrix exponential and Frechet-adjoint gradient.

Given this feasible-set equivalence, the central empirical question is whether PACO can preserve PPO's structural recovery accuracy in practice while delivering the predicted efficiency gains in multi-path settings. We evaluate three aspects. *Accuracy*: under matched graph families, sample sizes, partial-order regimes, and dataset seeds, we compare PACO with path-decomposition baselines in structural Hamming distance (SHD), true positive rate (TPR), false discovery rate (FDR), and F1 score. *Efficiency*: we evaluate whether PACO reduces wall-clock cost when the number of induced prior paths is large. *Incomplete priors*: we examine how PACO's performance changes as the supplied prior covers less of the true ancestor structure; we follow Ban et al. [2024] in assuming consistent priors (Section 3.6). We study these aspects through graph-level correctness results, complexity analysis, and PPO-style synthetic benchmarks.

Our contributions are:

- We encode partial-order compatibility through a masked matrix-exponential reachability constraint whose value and gradient are computed jointly via a single Frechet adjoint, with dominant matrix-function cost independent of the number of induced maximal prior paths.
- We establish that this formulation preserves exactly the feasible DAGs admitted by path decomposition (Theorem 1).
- We evaluate accuracy, runtime, path-count scaling, and behavior under incomplete priors against NOTEARS and partial-order path-decomposition baselines.

2 Preliminaries

Standing causal assumptions. Following the NOTEARS family [Zheng et al., 2018, 2020, Bello et al., 2022], we assume the causal Markov condition and causal sufficiency (no hidden confounders) for causal interpretation. Faithfulness is not required for PACO's zero-set equivalence, gradient derivation, or complexity analysis; causal recovery depends on the underlying backbone's statistical assumptions.

Weighted DAG learning. Let $X \in \mathbb{R}^{n\times d}$ denote n observations of d variables. We represent a candidate causal graph by a weighted adjacency matrix $W \in \mathbb{R}^{d\times d}$, where W_{ij} is nonzero when variable i is used as a parent of variable j. The strict support graph $\mathrm{supp}(W) = \{(i,j) : W_{ij} \neq 0\}$ is used in all theoretical (zero-set) statements. Empirical metrics use a thresholded support $\mathrm{supp}_\omega(W) = \{(i,j) : |W_{ij}| > \omega\}$ with $\omega = 0.3$, applied uniformly to PACO and all baselines, so paired comparisons and the PACO–PPO equivalence are stated under matched conventions. For linear structural equation models, a common objective is

$$\ell(W; X) = \frac{1}{2n}\|X - XW\|_F^2,$$

augmented with sparsity regularization. The learned graph must be acyclic and have zero diagonal.

The NOTEARS acyclicity characterization. For a nonnegative matrix $A \in \mathbb{R}^{d\times d}_{\geq 0}$, the diagonal entries of A^k count weighted closed walks of length k, so $\mathrm{tr}(\exp(A)) - d$ is zero iff the directed graph represented by A contains no directed cycle. NOTEARS applies this result to $A = W \circ W$, yielding a smooth equality constraint for DAG learning:

$$h(W) = \mathrm{tr}\left(\exp(W \circ W)\right) - d.$$

The elementwise square $W \circ W$ replaces signed weights with nonnegative magnitudes, so the walk-counting argument applies and $h(W) = 0$ characterizes acyclicity.

Walks, paths, and the matrix exponential. The same series expansion underlies a complementary fact about off-diagonal entries of $\exp(M)$, used to characterize the reachability constraint introduced in Section 3. For nonnegative $M \in \mathbb{R}^{d\times d}_{\geq 0}$, the expansion $[\exp(M)]_{ab} = \sum_{k\geq 0}[M^k]_{ab}/k!$ has only

nonnegative terms, with $[M^k]_{ab}$ counting weighted directed walks of length k from a to b in the support graph of M. Hence $[\exp(M)]_{ab} > 0$ iff there exists a directed walk from a to b, equivalently a directed path.

Partial-order priors. A partial-order prior is a binary relation over variables indicating that one variable should precede another. We encode the prior as an acyclic directed graph G_o with binary adjacency matrix $O \in \{0,1\}^{d\times d}$, where $O_{ij} = 1$ means that i is known or assumed to precede j. Importantly, this relation does not require the learned graph to include $i \to j$ as a causal edge. Instead, it forbids learned structures that imply the reverse precedence, namely a directed path from j to i.

Compatibility with a prior order. We define the strict transitive closure $O^+ \in \{0,1\}^{d\times d}$ by $O^+_{ij} = 1$ iff $i \neq j$ and G_o contains a directed path from i to j; thus O^+ collects all precedence relations implied by G_o, excluding self-relations on the diagonal. Since G_o is assumed acyclic, O^+ is irreflexive and represents a strict partial order. A learned graph G_W is compatible with the prior order when, for every (i,j) with $O^+_{ij} = 1$, G_W contains no directed path from j to i; equivalently, the learned graph must not make a later variable an ancestor of an earlier variable under the supplied prior. This allows G_W to add edges among variables not ordered by the prior, while ruling out edges and paths that reverse known precedence relations. The strict (off-diagonal) form of O^+ is essential for the masked-reachability penalty introduced in Section 3: a reflexive closure with $O^+_{ii} = 1$ would force the penalty to include $[\exp(W \circ W)]_{ii} \geq 1$ for every i, making it impossible to drive to zero for any W.

Transitive reduction. The transitive reduction $O^- \in \{0,1\}^{d\times d}$ is the unique minimal subgraph of G_o whose transitive closure equals O^+; equivalently, O^- retains (i,j) only when G_o has no longer directed path from i to j. From O^- one can extract a set of maximal directed paths $\mathcal{P}(O^-)$ that together cover every relation in O^+. PACO operates on O^+ (masked reachability), whereas path-decomposition baselines operate on $\mathcal{P}(O^-)$ (one constraint per maximal path).

Path-decomposition baselines. Path-decomposition approaches, which we refer to as PPO baselines following Ban et al. [2024], first transform the prior graph into a set of paths $\mathcal{P}(O^-)$. For each path p, a binary path matrix $P_p \in \{0,1\}^{d\times d}$ encoding the edges of p is combined with the learned adjacency matrix into a path-augmented adjacency $\mathcal{A}(W,p) = W + \tau P_p - W \circ P_p$, where $\tau > 0$ is an augmentation weight: $\mathcal{A}(W,p)$ overwrites W at the entries of P_p with the value τ and leaves other entries unchanged [Ban et al., 2024, Eq. 9]. Specialized to the matrix-exponential acyclicity used in this paper, the path-decomposition constraint takes the form

$$h_{\mathrm{PPO}}(W;O) = \sum_{p\in\mathcal{P}(O^-)} \left[\mathrm{tr}\left(\exp(\mathcal{A}(W,p)\circ\mathcal{A}(W,p))\right) - d\right].$$

This construction is prior-aware and path-specific, but the number of matrix exponentials per evaluation scales with $|\mathcal{P}(O^-)|$. The exact value of $\tau > 0$ does not affect the zero-set characterization analyzed in Section 3; we use $\tau = 3$ throughout, matching the PPO reference implementation.

3 Methodology

3.1 PACO: masked reachability for partial orders

Using the notation from Section 2, $W \in \mathbb{R}^{d\times d}$ is the learned adjacency, $O \in \{0,1\}^{d\times d}$ the prior, and $O^+ \in \{0,1\}^{d\times d}$ its strict transitive closure with $O^+_{ij} = 1 \iff i \prec j$. We assume throughout that the prior graph G_o encoded by O is acyclic.

PACO encodes the reverse-path prohibition implied by $i \prec j$ (Section 2) through the matrix exponential of $W \circ W$: by the walk-counting fact, $[\exp(W \circ W)]_{ji} > 0$ exactly when $G(W)$ contains a path $j \rightsquigarrow i$, so a violation of $i \prec j$ is detected by this entry.

We define the masked reachability penalty

$$q_{\mathrm{PACO}}(W;O) = \left\langle (O^+)^\top, \exp(W \circ W)\right\rangle = \sum_{(i,j):O^+_{ij}=1} [\exp(W\circ W)]_{ji}.$$

This penalty is inspired by Equation 8b of Ban et al. [2024], which uses the truncated series $\sum_{l=1}^{d}(W \circ W)^l$ in place of $\exp(W \circ W)$. Their Remark 2 identifies the cost of computing $|O^+|$ separate gradient terms as a computational limitation. We address this limitation in Section 3.3: with the matrix-exponential form and a single Frechet-adjoint evaluation, the gradient cost is $\mathcal{O}(C_{\exp}(d) + C_{\mathrm{F}}(d) + |O^+|)$, with the dominant matrix-function terms independent of $|O^+|$ (see Section 3.5). PACO therefore provides an efficient implementation of the direct reachability constraint.

Together with the standard NOTEARS acyclicity function

$$h(W) = \operatorname{tr}\left(\exp(W \circ W)\right) - d,$$

we use the combined exact-feasibility characterization

$$H_{\mathrm{PACO}}(W; O, \eta) = h(W) + \eta q_{\mathrm{PACO}}(W; O), \qquad \eta > 0.$$

Both terms are nonnegative, so any $\eta > 0$ leaves the exact feasible set unchanged: $H_{\mathrm{PACO}}(W; O, \eta) = 0 \iff h(W) = 0$ and $q_{\mathrm{PACO}}(W; O) = 0$. In finite-tolerance optimization, however, η rescales the relative gradient of acyclicity and prior-compatibility violations: small values can under-enforce the prior term, while large values can let the reachability term dominate the data-fit signal. Appendix D reports a sweep over two orders of magnitude in η.

A five-variable example illustrating PPO path decomposition and PACO reverse-reachability detection is provided in Appendix B.

3.2 Zero-set characterization

The following result states the exact graph-space feasible set characterized by PACO.

Theorem 1 (PACO characterizes partial-order feasibility)**.** *Assume O^+ is the strict transitive closure of an acyclic partial-order prior and $\eta > 0$. Then*

$$H_{\mathrm{PACO}}(W; O, \eta) = 0$$

if and only if the learned graph $G(W)$ is acyclic and, for every (i, j) with $O^+_{ij} = 1$, $G(W)$ contains no directed path from j to i.

Proof. Let $M = W \circ W$. Since M is elementwise nonnegative, the NOTEARS matrix-exponential characterization gives

$$h(W) = 0 \quad \iff \quad G(W) \text{ is acyclic.}$$

Moreover, the expansion

$$\exp(M) = I + M + \frac{M^2}{2!} + \frac{M^3}{3!} + \cdots$$

has only nonnegative terms. Thus $[\exp(M)]_{ab} > 0$ if and only if some power M^k has positive (a, b) entry, which is equivalent to the existence of a directed walk, and therefore a directed path, from a to b in $G(W)$.

It follows that

$$q_{\mathrm{PACO}}(W; O) = 0$$

if and only if $[\exp(W \circ W)]_{ji} = 0$ for every prior relation $i \prec j$ in O^+, which is equivalent to the absence of every forbidden reverse path $j \rightsquigarrow i$. Finally, $h(W) \geq 0$, $q_{\mathrm{PACO}}(W; O) \geq 0$, and $\eta > 0$, so $H_{\mathrm{PACO}}(W; O, \eta) = 0$ if and only if both $h(W) = 0$ and $q_{\mathrm{PACO}}(W; O) = 0$. □

Equivalently, PACO and the path-decomposition formulation admit the same family of feasible DAGs under any partial-order prior, differing only in how the constraint is evaluated rather than in which graphs satisfy it.

Applicability beyond linear W. The theorem only requires (i) a backbone acyclicity term h that is nonnegative with zero set equal to acyclic graphs, and (ii) a real $d \times d$ matrix W on which $\exp(W \circ W)$ is evaluated. In NOTEARS-MLP [Zheng et al., 2020], W is the first-layer column-norm matrix $W(\Theta)_{ij} = \|\theta^{(1)}_{ij}\|_2$ formed from the per-variable MLPs; in DAGMA [Bello et al., 2022], h is the log-determinant variant but W is unchanged. In both cases $W \circ W$ is nonnegative and the proof goes through unchanged.

The theorem is an exact zero-set statement; it does not imply identical optimization landscapes. PPO penalizes a violation through closed walks formed with an inserted prior path, so gradient scales depend on path count, overlap, length, and augmentation weight. PACO penalizes reverse reachability directly without inserting prior paths, distributing finite-tolerance gradient signal differently and motivating diagnostics on both recovery and runtime.

3.3 Optimization and gradient

We optimize $\ell(W;X) + \lambda_1\|W\|_1$ subject to $H_{\mathrm{PACO}}(W;O,\eta) = 0$ and $\mathrm{diag}(W) = 0$. Because h and q_{PACO} are nonnegative, the combined $H_{\mathrm{PACO}} = h + \eta q_{\mathrm{PACO}}$ is itself a single nonnegative scalar constraint, and we solve the resulting program with a single augmented-Lagrangian block [Zheng et al., 2018]:

$$\mathcal{L}(W) = \ell(W;X) + \lambda_1\|W\|_1 + \alpha\, H_{\mathrm{PACO}}(W;O,\eta) + \frac{\rho}{2}\, H_{\mathrm{PACO}}(W;O,\eta)^2,$$

with dual update $\alpha \leftarrow \alpha + \rho\, H_{\mathrm{PACO}}$. The standard positive–negative variable splitting handles the ℓ_1 penalty under box constraints.

The gradient of h takes the standard NOTEARS form $\nabla_W h(W) = 2W \circ \exp(W \circ W)^\top$. With $M = W \circ W$, the reachability gradient follows from the chain rule and the Frechet derivative of $\exp$:

$$\nabla_W q_{\mathrm{PACO}}(W;O) = 2W \circ L_{\exp}(M^\top)\big((O^+)^\top\big)\,,$$

where $L_{\exp}(A)(E)$ denotes the Frechet derivative of $\exp$ at A in direction E. The constraint value q_{PACO} and this gradient are computed jointly in a single $2d \times 2d$ matrix exponential via Higham's block-augmentation identity (Appendix C, Algorithm 2).

3.4 Relationship to path-decomposition constraints

The path-decomposition formulation that Ban et al. [2024] adopt in place of their Eq. 8b computes a transitive reduction O^-, extracts maximal paths $\mathcal{P}(O^-)$, and sums path-augmented acyclicity constraints of the form

$$h_{\mathrm{PPO}}(W;O) = \sum_{p \in \mathcal{P}(O^-)} h(\mathcal{A}(W,p)),$$

where $\mathcal{A}(W,p) = W + \tau P_p - W \circ P_p$ is the path-augmented adjacency [Ban et al., 2024, Eq. 9], with P_p the binary path matrix and $\tau > 0$ the augmentation weight (Section 2). Its zero set is the set of DAGs that satisfy the same reverse-path semantics for the prior order.

PACO targets this same exact feasible set by using masked matrix-exponential reachability rather than path-wise augmented acyclicity. Hence, under exact arithmetic and the strict support convention of Section 2,

$$\{W : H_{\mathrm{PACO}}(W;O,\eta) = 0\} = \{W : h_{\mathrm{PPO}}(W;O) = 0\}.$$

The equivalence is only at the level of exact feasibility. The optimization landscapes are different: PPO uses path-specific augmented acyclicity terms, while PACO uses one global reachability term masked by $(O^+)^\top$.

This also avoids the over-restriction of naively folding the prior into a single acyclicity check ($\mathrm{tr}(\exp(W \circ W + \tau O)) - d$): requiring $G(W) \cup O$ to be acyclic is sufficient but not complete — independent prior fragments can couple via learned edges into a union cycle even when no relation is reversed (Ban et al. [2024], Example 3: prior $\{(1,2),(3,4)\}$ + learned $\{(2,3),(4,1)\}$ gives a 4-cycle). PACO checks only the reverse paths in the strict transitive closure O^+.

3.5 Computational complexity

Let $C_{\exp}(d)$ denote the cost of one $d \times d$ matrix exponential and let $C_{\mathrm{F}}(d)$ denote the cost of one Frechet derivative evaluation for the matrix exponential. A path-decomposition baseline evaluates one acyclicity function and gradient per extracted path, yielding cost approximately

$$\mathcal{O}\big(|\mathcal{P}(O^-)|\, C_{\exp}(d)\big)\,.$$

PACO evaluates one matrix exponential for h and q_{PACO}, and one Frechet derivative for the gradient of q_{PACO}:

$$\mathcal{O}\big(C_{\exp}(d) + C_{\mathrm{F}}(d) + |O^+|\big)\,.$$

Table 1: Constraint-evaluation cost for different prior-handling strategies. $C_{\exp}(d)$ and $C_{\mathrm{F}}(d)$ denote the cost of one $d\times d$ matrix exponential and one Frechet derivative evaluation, respectively. Path decomposition scales linearly with $|\mathcal{P}(O^-)|$, the number of induced prior paths, while PACO's dominant cost is independent of $|\mathcal{P}(O^-)|$.

Method	Prior use	Main matrix-function calls	Cost
Unconstrained	none	1 expm	$\mathcal{O}(C_{\exp}(d))$
Path-decomp.	per path	$\lvert\mathcal{P}(O^-)\rvert$ expm	$\mathcal{O}(\lvert\mathcal{P}(O^-)\rvert C_{\exp}(d))$
PACO	reachability mask	1 expm + 1 Frechet	$\mathcal{O}(C_{\exp}(d) + C_{\mathrm{F}}(d) + \lvert O^+\rvert)$

The mask cost $|O^+|$ is at most $\mathcal{O}(d^2)$ and is usually dominated by the matrix-function evaluations. Asymptotically, $C_{\exp}(d)$ and $C_{\mathrm{F}}(d)$ are both $\mathcal{O}(d^3)$ via scaling-and-squaring [Higham, 2008], so PACO is $\mathcal{O}(d^3)$ per outer iteration, the same asymptotic class as unconstrained NOTEARS (Table 1).

PACO is not uniformly faster: when $|\mathcal{P}(O^-)| = 1$ a path-wise method may be comparable or faster due to PACO's Frechet overhead. The advantage concentrates at larger path counts; the dominant cost is independent of $|O^+|$, addressing the gradient-computation concern raised by Ban et al. [2024] for Eq. 8b. Appendix E reports the single-chain overhead, and Section 4 reports the multi-chain speedup.

3.6 Scope and prior-quality assumptions

PACO operates under a consistent-prior assumption shared with Ban et al. [2024]: the supplied prior G_o is acyclic and its relations are a subset of true ancestor relations of the data-generating process. We consider incomplete and inconsistent priors.

Incomplete priors. Missing relations simply do not contribute to O^+, so the constraint becomes weaker but never wrong. PACO degrades gracefully to the unconstrained NOTEARS solution as the prior coverage decreases. Our single-chain experiments span chain coverage from 0% to 100%, and our multi-chain experiments span chain count from 0 to 10 (Section 4).

Inconsistent priors. A wrong relation $i \prec j$ would forbid any learned directed path from j back to i, suppressing a true causal direction. Tolerating such priors requires a softer constraint formulation (e.g., the soft edge-constraint approach of Chen et al. [2025]). We follow Ban et al. [2024] in assuming consistent priors.

4 Experiments

We evaluate whether PACO can match path-decomposition baselines on structural recovery while reducing constraint-evaluation cost, on synthetic settings mirroring the PPO benchmark. Three differentiable backbones are compared: linear NOTEARS [Zheng et al., 2018] matches PPO's primary setting, DAGMA [Bello et al., 2022] tests a log-determinant acyclicity characterization, and NOTEARS-MLP [Zheng et al., 2020] tests nonlinear SEMs. Score-and-search and constraint-based methods are excluded as they do not share the continuous graph-space optimization interface.

4.1 Experimental setup

Synthetic graph families. We follow the synthetic-grid conventions of NOTEARS-style benchmarks [Zheng et al., 2018, 2020, Ban et al., 2024]. We generate random DAGs with $d \in \{20, 30, 50\}$ nodes under Erdos–Renyi (ER) and scale-free (SF) graph models. For each graph family we use expected edge-to-node ratios in $\{2, 4\}$. Linear SEM experiments use Gaussian and exponential noise with sample size $n = 4d$, following Zheng et al. [2018]. Nonlinear NOTEARS-MLP experiments use the `mim` nonlinear SEM with sample size $n = 20d$, following Zheng et al. [2020]. Each configuration is repeated over $K = 10$ independently seeded datasets.

Partial-order priors. We use the same two partial-order regimes as the PPO synthetic benchmark. In the single-chain regime, the prior consists of one chain with length fraction in $\{0, 0.5, 0.75, 1.0\}$, denoted `chain_1_0`, `chain_1_50`, `chain_1_75`, and `chain_1_100`. In the multi-chain regime, the chain length is fixed at $0.5d$ and the number of chains is varied in $\{0, 1, 2, 3, 5, 10\}$. Zero-prior

Table 2: Aggregated synthetic results across graph types, edge densities, noise models, and seeds. Each backbone group reports the unconstrained baseline (top row, no prior) and three matched PPO/PACO pairs; bold marks the strictly better mean within a pair at the displayed precision (ties unbolded; baselines excluded). Obed = fraction of completed paired runs (80 per cell for linear NOTEARS and DAGMA; 38–40 for NOTEARS-MLP, single `mim` noise) in which the learned graph contains no path reversing the supplied prior; '—' indicates no prior. The complete benchmark with per-cell standard deviations is in Table 11.

Model	Method	$d = 20$					$d = 30$					$d = 50$				
		SHD↓	TPR↑	FDR↓	F1↑	Obed↑	SHD↓	TPR↑	FDR↓	F1↑	Obed↑	SHD↓	TPR↑	FDR↓	F1↑	Obed↑
NOTEARS	NOTEARS (no prior)	26.20	0.66	0.13	0.75	—	34.39	0.72	0.11	0.79	—	52.46	0.75	0.10	0.82	—
	PPO-1-chain-50	23.35	0.70	0.11	0.78	1.00	30.82	0.75	0.09	0.82	1.00	**44.46**	0.78	**0.07**	**0.85**	1.00
	PACO-1-chain-50	**23.26**	0.70	0.11	0.78	1.00	**30.52**	0.75	0.09	0.82	1.00	46.24	0.78	0.08	0.84	1.00
	PPO-3-chain-50	21.30	0.72	0.09	0.80	1.00	27.44	0.77	0.07	0.84	1.00	39.56	0.80	0.06	0.86	1.00
	PACO-3-chain-50	**20.39**	**0.73**	**0.08**	**0.81**	1.00	**26.36**	0.77	**0.06**	**0.85**	1.00	**38.65**	0.80	**0.05**	**0.87**	1.00
	PPO-5-chain-50	19.57	0.74	0.07	0.82	1.00	25.30	0.78	0.06	0.85	1.00	38.00	0.81	0.05	0.87	0.988
	PACO-5-chain-50	**19.14**	**0.75**	0.07	**0.83**	1.00	**25.12**	0.78	**0.05**	**0.86**	1.00	**36.90**	0.81	0.05	0.87	**1.00**
DAGMA	DAGMA (no prior)	24.96	0.67	0.12	0.76	—	31.56	0.73	0.09	0.81	—	49.66	0.75	0.07	0.83	—
	PPO-1-chain-50	22.94	**0.70**	0.10	0.78	1.00	32.06	0.74	0.09	0.81	1.00	54.08	0.74	0.09	0.82	1.00
	PACO-1-chain-50	**22.84**	0.69	**0.09**	0.78	1.00	**30.20**	0.74	**0.08**	**0.82**	1.00	**48.67**	**0.75**	**0.07**	**0.83**	1.00
	PPO-3-chain-50	21.48	0.72	0.09	0.80	1.00	28.89	0.75	0.07	0.83	1.00	51.00	0.78	0.08	0.84	0.988
	PACO-3-chain-50	**20.35**	**0.73**	**0.07**	**0.81**	1.00	**27.12**	**0.76**	**0.06**	**0.84**	1.00	**41.49**	**0.79**	**0.05**	**0.86**	**1.00**
	PPO-5-chain-50	20.11	0.74	0.07	0.82	1.00	27.52	0.76	0.06	0.84	1.00	46.12	0.77	0.08	0.84	1.00
	PACO-5-chain-50	**19.45**	0.74	0.07	0.82	1.00	**26.02**	**0.77**	**0.05**	**0.85**	1.00	**39.36**	**0.80**	**0.05**	**0.87**	1.00
NOTEARS-MLP	NOTEARS-MLP (no prior)	17.90	0.81	0.14	0.83	—	25.00	0.77	0.07	0.84	—	43.30	0.73	0.03	0.83	—
	PPO-1-chain-50	17.75	0.82	0.14	0.83	1.00	23.95	0.78	0.06	0.85	1.00	42.65	0.74	0.02	0.83	1.00
	PACO-1-chain-50	**17.48**	0.82	0.14	**0.84**	1.00	**23.93**	0.78	0.06	0.85	1.00	42.65	0.74	0.02	0.83	1.00
	PPO-3-chain-50	**16.50**	0.83	**0.12**	0.85	1.00	23.48	0.78	0.06	0.85	0.975	43.46	0.73	0.04	0.82	0.923
	PACO-3-chain-50	16.93	**0.84**	0.14	0.85	1.00	**23.30**	**0.79**	0.06	0.85	**1.00**	**40.67**	**0.75**	**0.02**	**0.84**	**1.00**
	PPO-5-chain-50	**16.15**	**0.84**	**0.12**	0.85	1.00	23.02	0.79	**0.05**	0.85	1.00	42.32	0.74	0.03	0.83	0.816
	PACO-5-chain-50	16.70	0.83	0.13	0.85	1.00	23.02	0.79	0.06	0.85	1.00	**40.63**	**0.75**	**0.02**	**0.84**	**1.00**

settings are used only for unconstrained baselines; nonzero prior settings are used for PACO and PPO variants.

Methods. For the linear setting, we compare NOTEARS, PACO, and PPO. For the log-determinant acyclicity setting, we compare DAGMA, PACO-DAGMA, and PPO-DAGMA, where PACO and PPO differ in whether the partial-order prior is enforced through masked reachability or path-wise constraints. For nonlinear SEMs, we compare NOTEARS-MLP, PACO-MLP, and PPO-MLP. All constrained PACO variants use the same supplied partial-order matrix as the corresponding PPO variant.

Metrics. We report SHD, TPR, FDR, F1 score, prior-order obedience (the fraction of completed runs in which the learned graph contains no path reversing any supplied prior order), and wall-clock runtime per method call. The primary accuracy comparison is paired by dataset seed and experimental condition. The primary computational comparison is the runtime ratio between PPO-style path decomposition and PACO's masked-reachability constraint. Detailed experimental settings and implementation details are provided in Appendix F.

4.2 Main synthetic results

Table 2 reports the main synthetic comparison. On structural recovery, PACO obtains lower SHD than PPO in most cells, with improvements most pronounced for DAGMA at d=50 (e.g., DAGMA 5-chain: $46.12 \rightarrow 39.36$ at d=50). TPR is usually tied or slightly improved, indicating gains come from cleaner recovery rather than lost true positives. The closeness of PPO and PACO across all 27 multi-chain cells is consistent with Theorem 1: under matched priors and seeds the two methods admit the same family of feasible DAGs, while differences in selected outputs can reflect the optimization process. The unconstrained baseline rows (top of each backbone group) provide an absolute reference: both PPO and PACO generally benefit from prior integration, with the gain over baseline depending on chain count and graph size.

Figure 2 plots runtime at d=50 against the empirical maximal-path count $|\mathcal{P}(O^-)|$. For small path counts, PACO's Frechet overhead can offset its per-path savings, so it is not uniformly faster. As $|\mathcal{P}(O^-)|$ grows, PPO's path-wise acyclicity cost scales linearly while PACO's matrix-function cost

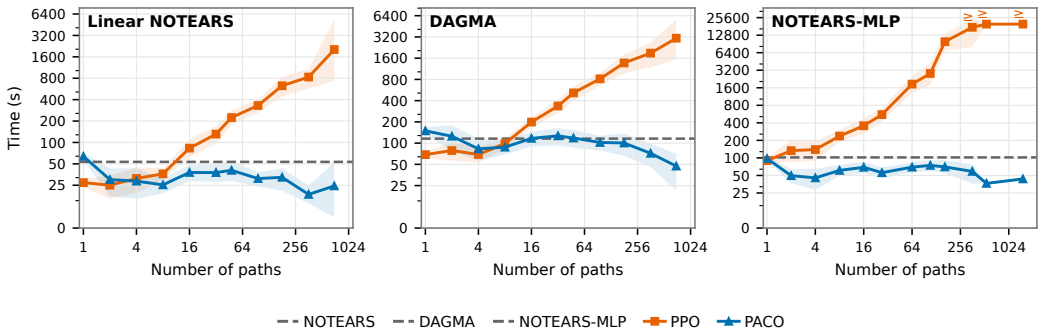

Figure 2: Runtime vs. empirical maximal-path count $|\mathcal{P}(O^-)|$ at d=50, with PACO (blue), PPO (orange), and the unconstrained learner (dashed). Markers and shaded bands show mean and descriptive 95% bootstrap CI per path-count bin, aggregated over multi-chain priors $c \in \{1, 2, 3, 5, 10\}$, both graph models (ER, SF), edge densities, and noise models, with 10 seeds per configuration (400 attempted runs/backbone for linear/DAGMA; 200 for NOTEARS-MLP). The chain-count→path-count mapping is not one-to-one: multi-chain priors generate combinatorial paths through shared closure relations.

remains independent of the path count. At $|\mathcal{P}(O^-)|\approx 96$, the PPO/PACO ratio of mean runtimes is approximately $10.7\times$ for linear NOTEARS and $8.0\times$ for DAGMA; for NOTEARS-MLP it is $37.8\times$ at $|\mathcal{P}(O^-)|\approx 108$.

The Obed column in Table 2 reports prior-order obedience: PACO satisfies the supplied prior in all completed pairs across all 27 multi-chain cells, while PPO drops below 1.00 in 5 cells, with obedience as low as 0.816 at d=50 for NOTEARS-MLP with 5 chains. This may reflect the optimization difficulty of jointly satisfying $|\mathcal{P}(O^-)|$ competing path-augmented constraints under finite tolerance.

Scaling to $d = 100$. We extend the linear NOTEARS runtime comparison to $d = 100$, using the same synthetic settings and prior regimes as above, with $K = 10$ seeds per configuration and matched data and priors. Table 3 groups runtimes by the empirical maximal-path count $|\mathcal{P}(O^-)|$; Figure 3 shows the individual runs.

Table 3: Mean runtime at $d = 100$ for linear NOTEARS, grouped by empirical maximal-path count. The ratio is PPO/PACO mean runtime. Means include all attempts, with time-limited runs contributing their limits; $\geq$ denotes a lower bound.

Path count	PPO mean (s)	PACO mean (s)	Ratio
1	196.6	305.3	0.64
2–100	1,520.4	186.3	8.16
101–1,000	$\geq$14,152.4	247.5	$\geq$57.18
1,001–10,000	$\geq$79,112.3	259.7	$\geq$304.62
$>$ 10,000	$\geq$120,680.6	187.4	$\geq$643.99

As at $d = 50$, PACO is not uniformly faster: at $|\mathcal{P}(O^-)| = 1$, mean runtime is 305.3 seconds for PACO versus 196.6 seconds for PPO. The PPO/PACO ratio of mean runtimes increases to $8.16\times$ for 2–100 paths and at least $643.99\times$ above 10,000 paths. These results extend the computational advantage at high path counts to $d = 100$, while retaining the single-chain overhead observed above.

4.3 Real-data case study

We evaluate PACO on the Sachs dataset [Sachs et al., 2005], following the causal-sufficiency assumption of the benchmark in Ban et al. [2024], although protein signaling may involve unmeasured regulators. Table 4 reports linear NOTEARS, PPO, and PACO under matched single-chain priors at $n \in \{50, 100, 500, 853\}$ (mean $\pm$ std over $K = 10$ seeded prior draws on fixed observations);

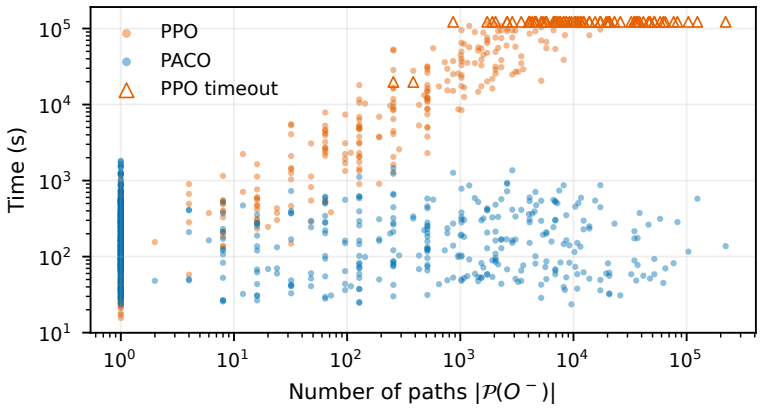

Figure 3: Runtime versus empirical maximal-path count at $d = 100$ across seven nonzero prior regimes (logarithmic axes). Filled points show completed PPO (orange) and PACO (blue) runs; open triangles mark PPO time limits, giving lower bounds on completion time.

we restrict the real-data study to linear NOTEARS for comparison with the benchmark of Ban et al. [2024]. Chain priors follow the PPO reference implementation. At the longest 11-node chain prior, PACO improves over PPO at n=500 (SHD 14→11), n=100 (SHD 13→9, F1 0.48→0.64), n=853 (SHD 11→9, F1 0.44→0.64), and n=50 (SHD 20→12). For the shorter 6- and 8-node chains, the two methods show setting-dependent differences, with PACO obtaining lower mean SHD across the reported sample sizes.

Table 4: Structural discovery on the Sachs protein-signaling dataset. PPO-1-m and PACO-1-m use the same single-chain prior over m nodes; entries are mean ± std over $K = 10$ seeded prior draws on fixed observations; bold values mark the strictly better mean within each matched PPO/PACO pair. Std = 0.00 indicates deterministic results across these repeats (priors over all 11 nodes have no chain-selection randomness).

	Sachs-50				Sachs-100				Sachs-500				Sachs-853			
Method	SHD↓	FDR↓	TPR↑	F1↑	SHD↓	FDR↓	TPR↑	F1↑	SHD↓	FDR↓	TPR↑	F1↑	SHD↓	FDR↓	TPR↑	F1↑
NOTEARS	24.0±0.0	0.75±0.00	0.41±0.00	0.31±0.00	17.0±0.0	0.68±0.00	0.35±0.00	0.33±0.00	15.0±0.0	0.60±0.00	0.35±0.00	0.38±0.00	12.0±0.0	0.57±0.00	0.35±0.00	0.39±0.00
PPO-1-6	21.5±2.2	0.70±0.05	**0.43**±0.08	0.35±0.06	15.8±2.9	0.63±0.06	**0.39**±0.03	0.38±0.03	14.6±1.5	0.56±0.05	**0.39**±0.05	**0.41**±0.04	13.4±1.6	0.57±0.07	**0.41**±0.06	0.42±0.05
PACO-1-6	**20.7**±1.8	**0.67**±0.04	0.42±0.07	**0.36**±0.05	**14.0**±1.1	**0.58**±0.05	0.38±0.04	**0.39**±0.02	**14.1**±0.9	**0.55**±0.06	0.35±0.04	0.40±0.04	**11.7**±0.7	**0.44**±0.09	0.36±0.02	**0.44**±0.04
PPO-1-8	21.1±2.1	0.70±0.04	**0.44**±0.08	0.36±0.05	14.1±2.0	0.58±0.06	0.42±0.07	0.42±0.06	15.5±1.6	0.58±0.04	**0.36**±0.05	0.39±0.04	13.0±1.9	0.54±0.09	0.38±0.06	0.41±0.06
PACO-1-8	**17.5**±2.1	**0.61**±0.07	0.42±0.04	**0.40**±0.04	**12.4**±1.2	**0.44**±0.14	**0.43**±0.05	**0.48**±0.08	**12.9**±0.9	**0.42**±0.10	0.35±0.04	**0.43**±0.05	**11.0**±1.1	**0.35**±0.08	0.38±0.04	**0.47**±0.04
PPO-1-11	20.0±0.0	0.67±0.00	0.41±0.00	0.37±0.00	13.0±0.0	0.50±0.00	0.47±0.00	0.48±0.00	14.0±0.0	0.50±0.00	0.35±0.00	0.41±0.00	11.0±0.0	0.40±0.00	0.35±0.00	0.44±0.00
PACO-1-11	**12.0**±0.0	**0.27**±0.00	**0.47**±0.00	**0.57**±0.00	**9.0**±0.0	**0.00**±0.00	0.47±0.00	**0.64**±0.00	**11.0**±0.0	**0.12**±0.00	**0.41**±0.00	**0.56**±0.00	**9.0**±0.0	**0.00**±0.00	**0.47**±0.00	**0.64**±0.00

5 Limitations

PACO has three main limitations. First, PACO inherits the consistent-prior assumption of Ban et al. [2024] (Section 3.6): G_o must be acyclic with relations forming a subset of true ancestors; wrong-direction priors can suppress true causal directions. Second, PACO assumes causal sufficiency (no latent confounders), shared with the NOTEARS family; extending masked reachability to settings with latent confounders (e.g., maximal ancestral graphs) is left to future work. Third, PACO is not uniformly faster than path decomposition: at $|\mathcal{P}(O^-)| = 1$, the observed runtime overhead is bounded ($\leq 2.51\times$ at $d = 50$; Appendix E), so PACO's advantage concentrates in multi-chain priors.

Acknowledgments and Disclosure of Funding

This research includes computations using the computational cluster Katana supported by Research Technology Services at UNSW Sydney. The authors declare no competing interests.

References

Awad H Al-Mohy and Nicholas J Higham. Computing the fréchet derivative of the matrix exponential, with an application to condition number estimation. *SIAM Journal on Matrix Analysis and Applications*, 30(4):1639–1657, 2009.

Bryan Andrews, Peter Spirtes, and Gregory F. Cooper. On the completeness of causal discovery in the presence of latent confounding with tiered background knowledge. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 4002–4011. PMLR, 26–28 Aug 2020.

Taiyu Ban, Lyuzhou Chen, Xiangyu Wang, Xin Wang, Derui Lyu, and Huanhuan Chen. Differentiable structure learning with partial orders. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang, editors, *Advances in Neural Information Processing Systems 37: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024.

Taiyu Ban, Changxin Rong, Xiangyu Wang, Lyuzhou Chen, Xin Wang, Derui Lyu, Qinrui Zhu, and Huanhuan Chen. Differentiable structure learning with ancestral constraints. In *Forty-second International Conference on Machine Learning*, 2025.

Christine Bang, Janine Witte, Ronja Foraita, and Vanessa Didelez. Improving finite sample performance of causal discovery by exploiting temporal structure. *Statistical Methods in Medical Research*, 35(4):867–886, 2026. ISSN 1477-0334.

Kevin Bello, Bryon Aragam, and Pradeep Ravikumar. DAGMA: learning dags via m-matrices and a log-determinant acyclicity characterization. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022.

Jeroen Berrevoets, Jakob Raymaekers, Mihaela van der Schaar, Tim Verdonck, and Ruicong Yao. Differentiable causal structure learning with identifiability by NOTIME. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pages 3115–3123. PMLR, 2025. URL `https://proceedings.mlr.press/v258/berrevoets25a.html`.

Giorgos Borboudakis and Ioannis Tsamardinos. Incorporating causal prior knowledge as path-constraints in bayesian networks and maximal ancestral graphs. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012*. icml.cc / Omnipress, 2012.

Philippe Brouillard, Sébastien Lachapelle, Alexandre Lacoste, Simon Lacoste-Julien, and Alexandre Drouin. Differentiable causal discovery from interventional data. *Advances in Neural Information Processing Systems*, 33:21865–21877, 2020.

Lyuzhou Chen, Taiyu Ban, Derui Lyu, Yijia Sun, Kangtao Hu, Xiangyu Wang, and Huanhuan Chen. Continuous structure constraint integration for robust causal discovery. In Yingzhen Li, Stephan Mandt, Shipra Agrawal, and Mohammad Emtiyaz Khan, editors, *International Conference on Artificial Intelligence and Statistics, AISTATS 2025, Mai Khao, Thailand, 3-5 May 2025*, Proceedings of Machine Learning Research, pages 3160–3168. PMLR, 2025.

Wenyu Chen, Mathias Drton, and Y Samuel Wang. On causal discovery with an equal-variance assumption. *Biometrika*, 106(4):973–980, 2019.

Chris Cundy, Aditya Grover, and Stefano Ermon. Bcd nets: Scalable variational approaches for bayesian causal discovery. *Advances in Neural Information Processing Systems*, 34:7095–7110, 2021.

Nicholas J Higham. *Functions of matrices: theory and computation*. SIAM, 2008.

Antti Hyttinen, Frederick Eberhardt, and Matti Järvisalo. Constraint-based causal discovery: Conflict resolution with answer set programming. In *UAI*, pages 340–349, 2014.

Sébastien Lachapelle, Philippe Brouillard, Tristan Deleu, and Simon Lacoste-Julien. Gradient-based neural dag learning. *arXiv preprint arXiv:1906.02226*, 2019.

Pingchuan Ma, Qixin Zhang, Shuai Wang, and Dacheng Tao. Efficient differentiable causal discovery via reliable super-structure learning. *CoRR*, abs/2601.05474, 2026. doi: 10.48550/ARXIV.2601.05474. URL `https://doi.org/10.48550/arXiv.2601.05474`.

Christopher Meek. Causal inference and causal explanation with background knowledge. *CoRR*, abs/1302.4972, 2013. URL `http://arxiv.org/abs/1302.4972`.

Ignavier Ng, AmirEmad Ghassami, and Kun Zhang. On the role of sparsity and dag constraints for learning linear dags. *Advances in Neural Information Processing Systems*, 33:17943–17954, 2020.

Ignavier Ng, Shengyu Zhu, Zhuangyan Fang, Haoyang Li, Zhitang Chen, and Jun Wang. Masked gradient-based causal structure learning. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*, pages 424–432. SIAM, 2022.

Roxana Pamfil, Nisara Sriwattanaworachai, Shaan Desai, Philip Pilgerstorfer, Konstantinos Georgatzis, Paul Beaumont, and Bryon Aragam. Dynotears: Structure learning from time-series data. In *International conference on artificial intelligence and statistics*, pages 1595–1605. Pmlr, 2020.

Karen Sachs, Omar Perez, Dana Pe'er, Douglas A Lauffenburger, and Garry P Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.

Peter Spirtes and Kun Zhang. Causal discovery and inference: concepts and recent methodological advances. In *Applied informatics*, volume 3, page 3. Springer, 2016.

Louis Verny, Nadir Sella, Séverine Affeldt, Param Priya Singh, and Hervé Isambert. Learning causal networks with latent variables from multivariate information in genomic data. *PLoS computational biology*, 13(10):e1005662, 2017.

Matthew J Vowels, Necati Cihan Camgoz, and Richard Bowden. D'ya like dags? a survey on structure learning and causal discovery. *ACM Computing Surveys*, 55(4):1–36, 2022.

Dennis Wei, Tian Gao, and Yue Yu. Dags with no fears: A closer look at continuous optimization for learning bayesian networks. *Advances in Neural Information Processing Systems*, 33:3895–3906, 2020.

Yue Yu, Jie Chen, Tian Gao, and Mo Yu. DAG-GNN: DAG structure learning with graph neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, Proceedings of Machine Learning Research, pages 7154–7163. PMLR, 2019.

Xun Zheng, Bryon Aragam, Pradeep Ravikumar, and Eric P. Xing. Dags with NO TEARS: continuous optimization for structure learning. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 9492–9503, 2018.

Xun Zheng, Chen Dan, Bryon Aragam, Pradeep Ravikumar, and Eric Xing. Learning sparse nonparametric dags. In *International conference on artificial intelligence and statistics*, pages 3414–3425. Pmlr, 2020.

A Related Work

Differentiable causal discovery. NOTEARS [Zheng et al., 2018] reformulates DAG learning as a continuous constrained optimization problem with a smooth matrix-exponential acyclicity constraint. Identifiability in purely observational linear-Gaussian settings requires additional assumptions, such as equal error variances [Chen et al., 2019]. Extensions cover nonlinear structural equation models [Zheng et al., 2020, Lachapelle et al., 2019], likelihood-based variants [Ng et al., 2020], alternative acyclicity functions [Yu et al., 2019, Bello et al., 2022], refinements addressing numerical robustness [Wei et al., 2020], extensions to interventional data [Brouillard et al., 2020], and Bayesian variational approaches [Cundy et al., 2021]; see Vowels et al. [2022] for a survey. PACO inherits this matrix-exponential machinery and repurposes it to encode partial-order reachability restrictions.

Causal discovery with prior knowledge. Prior knowledge has long been used to restrict graph search [Meek, 2013]. Score-and-search and constraint-based methods accept forbidden / required edges, temporal orderings, and path-level ancestor constraints [Borboudakis and Tsamardinos, 2012, Andrews et al., 2020, Hyttinen et al., 2014, Verny et al., 2017]. In differentiable structure learning, edge-level priors are integrated through soft constraints [Chen et al., 2025]. Partial-order priors are harder: they encode a path-level property—no learned graph may make a later variable an ancestor of an earlier one—that edge-level constraints cannot directly express without reducing the prior to many local constraints or losing path-level completeness [Ban et al., 2024].

Partial orders in differentiable structure learning. Ban et al. [2024] formulated partial-order compatibility through a direct reachability constraint, but adopted path decomposition because of concerns about the cost of direct gradient computation. Their path-wise formulation imposes one prior-aware acyclicity constraint per maximal path, so its evaluation cost scales with the path count. PACO revisits the direct formulation using masked matrix-exponential reachability, preserving the same exact feasible set. The constraint value and its gradient are computed jointly via a single Frechet adjoint, with dominant matrix-function cost independent of the number of induced maximal prior paths. A subsequent work [Ban et al., 2025] addresses path-*existence* constraints and reuses the partial-order absence construction via a truncated polynomial $\sum_{k=1}^{d} |W|^k$; PACO targets the absence problem directly, replacing the polynomial sum with the matrix exponential and a single Frechet-adjoint gradient. Since Ban et al. [2025] solves a complementary path-existence problem rather than the partial-order path-absence benchmark we study, we compare against Ban et al. [2024] as the direct absence-constraint baseline. Related orderings appear in dynamic / time-series settings, where temporal precedence is enforced lag-by-lag [Pamfil et al., 2020].

Computational efficiency of acyclicity constraints. Prior efficiency directions include alternative acyclicity characterizations [Bello et al., 2022] and search-space reduction via super-structure learning [Ma et al., 2026]. PACO addresses a third dimension: the path-count scaling of prior-order methods. It replaces path enumeration with a single global reachability term whose gradient is computed via the block-augmentation identity of Al-Mohy and Higham [2009], so its dominant matrix-function cost does not scale linearly with the number of maximal prior paths; the only $|O^+|$-dependent term is a linear-time mask accumulation. Masked gradient updates have also been used to enforce structural constraints [Ng et al., 2022]; PACO's mask is structural and applied to reachability entries rather than to gradient updates.

B Illustrative Example

Consider five variables with prior relations $1 \prec 3$, $2 \prec 3$, and $3 \prec 5$, encoded by $O_{13} = O_{23} = O_{35} = 1$ and all other entries zero. Node 4 is not ordered by the prior. No prior edge is redundant, so $O^- = O$; the strict transitive closure O^+ additionally includes $1 \prec 5$ and $2 \prec 5$. PPO evaluates a path-augmented acyclicity constraint for each maximal prior path, $(1, 3, 5)$ and $(2, 3, 5)$. For PACO, let $E = \exp(W \circ W)$, where, for $i \neq j$, $E_{ji} > 0$ exactly when the learned graph contains a directed path from j to i. The five relations in O^+ therefore select five reverse-reachability entries:

$$q_{\mathrm{PACO}}(W; O) = E_{31} + E_{51} + E_{32} + E_{52} + E_{53}.$$

For example, the learned path $5 \to 4 \to 1$ violates $1 \prec 5$, making $E_{51} > 0$ even without a direct edge $5 \to 1$. PPO detects the same violation because augmenting the learned graph with the prior

Algorithm 1 PACO: Partial-order-Augmented Continuous Optimization

Require: Data X, prior O, hyperparameters $\lambda_1, \omega, \eta, \beta, \rho_{\max}, \epsilon_H$

Verify $G_o = \mathrm{supp}(O)$ is acyclic; if not, abort with an error
Compute the strict transitive closure

$$O^+ \leftarrow \mathbb{I}\Big(\textstyle\sum_{k=1}^{d} O^k > 0\Big)$$

Form the binary mask $B \leftarrow (O^+)^\top$ once
Initialize $W \leftarrow 0$, $\alpha \leftarrow 0$, $\rho \leftarrow 1$, $H_{\mathrm{old}} \leftarrow \infty$
Define $H(W) = h(W) + \eta\, q_{\mathrm{PACO}}(W; O)$
for $t = 1, \ldots, 100$ **do**
  while $\rho < \rho_{\max}$ **do**
    Inner solve. Obtain W_{new} via L-BFGS-B, initialized at W, on
    $\mathcal{L}(W) = \ell(W; X) + \lambda_1 \|W\|_1 + \alpha\, H(W) + \frac{\rho}{2}\, H(W)^2$
    with the analytic gradient $\nabla_W H = \nabla_W h + \eta \nabla_W q_{\mathrm{PACO}}$ (Algorithm 2).
    Compute $H_{\mathrm{new}} \leftarrow H(W_{\mathrm{new}})$
    if $H_{\mathrm{new}} \leq 0.25 H_{\mathrm{old}}$ **then**
      break
    end if
    $\rho \leftarrow \beta\rho$
  end while
  $W \leftarrow W_{\mathrm{new}}$, $H_{\mathrm{old}} \leftarrow H_{\mathrm{new}}$
  Dual update. $\alpha \leftarrow \alpha + \rho H_{\mathrm{new}}$
  if $H_{\mathrm{new}} \leq \epsilon_H$ **or** $\rho \geq \rho_{\max}$ **then**
    break
  end if
end for
return W and $\mathrm{supp}_\omega(W) = \{(i,j) : |W_{ij}| > \omega\}$

path $(1, 3, 5)$ creates the cycle $1 \to 3 \to 5 \to 4 \to 1$ in the graph used for the acyclicity check. This augmentation does not require the learned graph to contain the prior edges. In particular, $1 \prec 3$ forbids a reverse path $3 \rightsquigarrow 1$ but requires neither an edge nor a path from 1 to 3.

C PACO Algorithm

This appendix presents the full algorithmic specification of PACO under a NOTEARS-style augmented-Lagrangian outer loop. The same template applies to the linear, DAGMA, and NOTEARS-MLP backbones; the only backbone-specific changes are the data-fit loss $\ell(W; X)$, the choice of acyclicity term $h(W)$, and the interpretation of W (a real adjacency for linear/DAGMA; the first-layer column-norm matrix $W(\Theta)_{ij} = \|\theta_{ij}^{(1)}\|_2$ for NOTEARS-MLP, see Section 3).

Inputs. Data matrix $X \in \mathbb{R}^{n \times d}$, prior order matrix $O \in \{0,1\}^{d \times d}$ (assumed acyclic), ℓ_1 regularization weight $\lambda_1 > 0$, edge threshold $\omega > 0$, reachability penalty weight $\eta > 0$ (we use $\eta = 1$; see Appendix D), penalty inflation factor $\beta > 1$ (we use $\beta = 10$), maximum penalty $\rho_{\max}$, and convergence tolerance ϵ_H.

Outputs. Learned weighted adjacency $W \in \mathbb{R}^{d \times d}$ and its thresholded support $\mathrm{supp}_\omega(W) = \{(i,j) : |W_{ij}| > \omega\}$.

Computing q_{PACO} and its Gradient

The constraint value $q_{\mathrm{PACO}}(W; O) = \langle B, \exp(W \circ W) \rangle$ and its gradient $\nabla_W q_{\mathrm{PACO}}$ are obtained from a single $2d \times 2d$ matrix exponential, exploiting Higham's block-augmentation identity for the Frechet derivative of $\exp(\cdot)$:

$$\exp\begin{pmatrix} A & E \\ 0 & A \end{pmatrix} = \begin{pmatrix} \exp(A) & L_{\exp}(A)(E) \\ 0 & \exp(A) \end{pmatrix},$$

Algorithm 2 q_{PACO} and gradient via single Frechet adjoint

Require: Weighted adjacency W, mask $B = (O^+)^\top$

$M \leftarrow W \circ W$

Form the $2d \times 2d$ block matrix

$$\widetilde{M} \leftarrow \begin{pmatrix} M^\top & B \\ 0 & M^\top \end{pmatrix}$$

Compute $E \leftarrow \exp(\widetilde{M})$ once via scaling-and-squaring ▷ single matrix exponential

Read off $E_{11} \leftarrow$ upper-left $d \times d$ block, $E_{12} \leftarrow$ upper-right $d \times d$ block

$$E_{11} = \exp(M^\top), \quad E_{12} = L_{\exp}(M^\top)(B)$$

$q_{\mathrm{PACO}}(W; O) \leftarrow \langle B, E_{11}^\top \rangle$ ▷ $\exp(M^\top)^\top = \exp(M)$

$\nabla_W q_{\mathrm{PACO}}(W; O) \leftarrow 2\, W \circ E_{12}$ ▷ chain rule through $M = W \circ W$

return $q_{\mathrm{PACO}}(W; O)$ and $\nabla_W q_{\mathrm{PACO}}(W; O)$

where $L_{\exp}(A)(E)$ denotes the Frechet derivative at A in direction E. Setting $A = M^\top$ with $M = W \circ W$ and $E = B$ yields both $\exp(M^\top)$ (for the constraint value) and $L_{\exp}(M^\top)(B)$ (the Frechet adjoint that drives the gradient) in one call.

Cost summary. Algorithm 2 performs exactly one matrix exponential of size $2d \times 2d$ per outer iteration, jointly returning the constraint value $\exp(M^\top)$ and the Frechet adjoint $L_{\exp}(M^\top)(B)$; the total per-iteration cost is $\mathcal{O}(C_{\exp}(d) + C_{\mathrm{F}}(d) + |O^+|)$ for the value-and-gradient computation, matching the bound stated in Section 3.5. The matrix-function cost is independent of the number of order pairs $|O^+|$ in the prior closure; $|O^+|$ enters only through the linear-time mask construction. This refines the per-pair gradient cost estimated by Ban et al. [2024] (their Remark 2) and is the algorithmic basis for the runtime advantage of PACO over path decomposition reported in Section 4.

Default hyperparameters. Throughout the experiments we use $\lambda_1 = 0.3$, $\omega = 0.3$, $\eta = 1$, $\beta = 10$, $\rho_{\max} = 10^{16}$, $\epsilon_H = 10^{-8}$. The augmented-Lagrangian loop uses a maximum of 100 outer iterations, with the standard NOTEARS positive–negative variable splitting to handle the ℓ_1 penalty under box constraints.

D Sensitivity to the Reachability Penalty Weight η

The combined PACO constraint $H_{\mathrm{PACO}}(W; O, \eta) = h(W) + \eta\, q_{\mathrm{PACO}}(W; O)$ introduces a single hyperparameter $\eta > 0$ that controls the relative scale of the masked-reachability penalty against the standard NOTEARS acyclicity term. As shown in the zero-set characterization (Section 3), the exact feasible set is independent of η for any $\eta > 0$; in finite-tolerance optimization, however, η shifts the relative gradient magnitudes of the two terms and can affect convergence behavior. This appendix reports a sensitivity sweep over $\eta \in \{0.1, 0.3, 1.0, 3.0, 10.0\}$ (two orders of magnitude) under linear NOTEARS, matched to the multi-chain grid of Section 4.

Setup. We use the same data-generation pipeline as the main multi-chain experiments: linear SEM with Gaussian and exponential noise, ER and SF graphs, $K = 10$ seeded datasets per cell, and prior chains `chain_X_50` for $X \in \{1, 3, 5, 10\}$. Because edge density s changes structural-recovery and runtime scales, we report s=2 and s=4 in separate tables; each (η, d, X, s) cell aggregates 2 noise $\times$ 2 graph $\times$ $K = 40$ paired runs. All other PACO hyperparameters ($\lambda_1 = 0.3$, edge threshold $\omega = 0.3$) are held fixed.

Results.

- **Structural recovery is insensitive to η.** Within every (d, X, s) cell mean SHD spans at most 3.60 across the full $100\times$ range of η (Tables 5 and 6); within-row variation is well under one standard deviation. Mean TPR and FDR vary by at most 0.016 and 0.018, respectively.
- **Prior-order obedience is uniform at 1.00.** PACO satisfies the supplied prior in all 40 paired runs at every η, d, X, and s tested.
- **Runtime grows monotonically at large η in low-path-count regimes.** Single-chain runtime at $d = 50$, $s = 2$ rises from 30.1 s ($\eta = 1$) to 44.9 s ($\eta = 10$), and at $s = 4$ from

98.5 s to 127.5 s (Tables 7 and 8); the trend is consistent across single-chain rows but the magnitudes lie within one standard deviation per cell. In high-chain-count regimes the trend can reverse: at $d = 50$, $s = 2$ with 10 chains, runtime decreases from 10.22 s ($\eta = 0.1$) to 8.76 s ($\eta = 10$).

These observations support a default of $\eta = 1$, which sits in the middle of the relatively stable SHD region and avoids the single-chain runtime increase at the upper end of the sweep.

Table 5: SHD (lower is better) at edge density s=2, mean±std over 40 paired runs per cell (2 noise $\times$ 2 graph $\times$ K=10 seeds).

d	chains	$\eta = 0.1$	$\eta = 0.3$	$\eta = 1.0$	$\eta = 3.0$	$\eta = 10.0$
20	1	10.65±5.62	11.00±6.27	10.88±6.12	10.90±5.93	11.18±6.25
	3	9.93±5.50	10.18±5.86	9.70±4.93	9.55±4.78	9.30±4.65
	5	8.72±4.44	9.07±4.98	8.62±4.23	8.62±4.23	8.60±4.21
	10	8.68±4.33	8.70±4.33	8.72±4.31	8.32±3.89	8.32±3.90
30	1	12.78±6.40	13.40±7.61	12.90±6.49	12.45±5.86	12.40±6.56
	3	11.45±5.05	11.07±4.72	10.93±4.82	10.72±4.31	10.20±4.12
	5	10.28±4.85	10.28±4.74	10.03±4.06	10.00±4.06	10.00±4.08
	10	10.03±4.02	10.00±3.95	10.05±3.94	10.03±3.95	10.05±3.94
50	1	18.80±8.01	18.70±7.90	19.32±8.62	19.23±9.11	18.88±9.15
	3	16.55±6.85	16.57±6.94	16.07±5.98	16.32±6.15	16.12±5.81
	5	16.68±7.16	17.25±7.67	16.40±6.57	16.43±6.45	16.20±6.78
	10	15.65±5.65	15.70±5.68	15.65±5.60	15.65±5.65	15.68±5.60

Table 6: SHD at edge density s=4, mean±std over 40 paired runs per cell.

d	chains	$\eta = 0.1$	$\eta = 0.3$	$\eta = 1.0$	$\eta = 3.0$	$\eta = 10.0$
20	1	35.48±11.63	36.40±12.87	35.65±13.36	36.27±13.14	35.80±12.23
	3	32.17±10.65	32.27±10.97	31.07±9.26	31.00±9.30	30.77±8.64
	5	30.10±7.77	30.10±7.72	29.65±7.41	29.27±7.53	29.05±7.79
	10	28.95±8.22	28.73±7.81	28.68±7.73	28.73±7.71	28.80±7.81
30	1	47.67±16.13	48.48±16.66	48.15±16.36	50.38±17.32	50.27±17.34
	3	42.52±11.08	42.33±10.95	41.80±9.10	42.17±10.17	42.02±9.52
	5	39.52±8.23	39.58±8.36	40.23±8.42	40.05±8.48	40.02±8.41
	10	39.70±8.63	39.23±7.93	39.25±7.94	39.20±8.05	38.98±7.94
50	1	70.22±23.01	71.65±26.16	73.15±28.40	72.12±26.21	72.40±24.89
	3	60.77±16.99	61.25±18.17	61.23±17.26	59.67±15.91	57.65±12.76
	5	57.67±11.70	57.17±11.51	57.40±11.54	58.05±10.96	56.80±10.68
	10	53.90±9.81	53.85±9.77	53.85±9.74	53.77±9.81	53.83±9.86

Table 7: Wall-clock runtime per PACO call (seconds) at edge density s=2, mean±std over 40 paired runs per cell.

d	chains	$\eta = 0.1$	$\eta = 0.3$	$\eta = 1.0$	$\eta = 3.0$	$\eta = 10.0$
20	1	5.59±10.24	5.35±10.88	5.25±9.29	6.43±11.28	8.00±12.14
	3	2.81±2.60	2.47±2.09	2.48±2.19	2.77±2.28	3.17±2.34
	5	2.44±2.34	2.15±1.94	2.07±1.98	2.21±2.32	2.33±1.94
	10	2.11±1.60	1.85±1.33	1.77±1.44	1.72±1.23	1.74±1.25
30	1	8.83±9.55	8.46±8.88	8.74±8.93	11.61±14.81	13.72±12.99
	3	4.75±4.00	4.19±3.42	4.19±3.13	4.86±3.66	6.01±4.26
	5	3.69±2.99	3.31±2.54	3.27±2.55	3.54±2.57	3.96±2.99
	10	3.40±2.79	2.96±2.35	2.81±2.15	2.82±2.14	2.81±2.22
50	1	28.98±21.92	26.79±20.25	30.12±23.50	35.84±25.63	44.94±31.92
	3	13.89±8.06	13.13±7.63	13.82±7.91	16.75±10.38	20.17±9.80
	5	11.46±7.17	10.41±6.12	10.57±6.06	11.36±6.12	12.56±6.37
	10	10.22±7.18	9.25±6.27	8.87±5.66	8.82±5.45	8.76±6.09

SHD sensitivity to η. The convergence of the augmented-Lagrangian outer loop drives both $h(W)$ and $q_{\mathrm{PACO}}(W; O)$ to numerical zero (absolute final values $\lesssim 1.2 \times 10^{-13}$ across all η tested), so the inner solver reaches a numerically feasible point for every tested η. What η controls is the gradient-scale balance during the early outer iterations; in this multi-chain regime the inner subproblem has enough optimization budget that the final feasible point is reached for every η in $[0.1, 10]$. The

Table 8: Wall-clock runtime per PACO call (seconds) at edge density s=4, mean±std over 40 paired runs per cell.

d	chains	$\eta = 0.1$	$\eta = 0.3$	$\eta = 1.0$	$\eta = 3.0$	$\eta = 10.0$
20	1	13.97±17.10	13.36±14.69	13.07±13.64	14.97±16.12	18.37±16.95
	3	8.75±6.90	7.94±5.94	8.57±7.87	8.34±5.79	9.27±5.97
	5	7.61±5.22	7.13±5.85	8.46±12.96	6.68±4.60	7.03±4.48
	10	7.68±5.62	6.54±4.45	6.17±4.33	6.00±4.07	6.04±4.17
30	1	21.84±16.46	21.23±17.79	23.21±17.49	30.70±25.43	34.64±22.67
	3	14.61±8.47	14.43±9.46	15.48±11.13	15.89±8.28	18.86±10.12
	5	12.66±7.58	11.38±5.89	11.59±6.45	11.71±6.28	12.75±7.30
	10	12.06±7.20	11.11±6.06	10.49±5.91	10.29±5.40	9.92±5.03
50	1	81.17±65.79	82.41±68.06	98.50±75.72	106.14±74.05	127.46±82.64
	3	53.22±34.18	52.11±34.66	54.75±35.37	57.37±34.85	61.60±29.89
	5	47.08±29.06	44.74±26.40	44.53±23.46	46.90±24.81	49.57±24.12
	10	43.39±27.81	40.84±24.92	40.29±24.83	39.56±23.49	37.29±21.60

relatively stable SHD pattern is consistent with the zero-set characterization in Section 3 and suggests that the augmented-Lagrangian implementation does not require careful η tuning in the tested settings.

Runtime trade-off. The mild runtime increase at large η in single-chain settings may reflect two effects in the augmented-Lagrangian outer loop: (i) when η is large the reachability term can dominate the gradient early on, slowing the descent on the data-fit objective and prolonging the inner L-BFGS solve; (ii) larger η values can require more outer iterations to balance the dual updates of the two penalties. Conversely, in high-path-count settings the prior already strongly constrains the feasible region, so a larger η can shorten the feasible-set search. Across the range explored, the largest within-cell ratio of maximum to minimum mean runtime is approximately $1.68\times$.

Default and recommendation. We adopt $\eta = 1$ throughout the main experiments, motivated by (a) relatively stable SHD, (b) moderate runtime in the sweep, and (c) symmetric scale between h and q_{PACO} in the absence of dataset-specific calibration. For users running PACO at very high $|\mathcal{P}(O^-)|$ where wall-clock matters, $\eta \in [0.3, 1]$ often gives low runtime with comparable structural recovery.

E Single-Chain Runtime Comparison

The complexity analysis in Section 3 predicts that, in single-chain prior settings ($|\mathcal{P}(O^-)| = 1$), the constant-factor overhead of PACO's $2d \times 2d$ Frechet block exponential can outweigh the savings of avoiding per-path acyclicity evaluations. This appendix verifies that prediction and quantifies the magnitude across the three backbones used in the main experiments.

Setup. We report wall-clock runtime per call (seconds) under single-chain priors `chain_1_p` with $p \in \{50, 75, 100\}$ encoding the fraction of d nodes covered. Edge density s affects runtime, so we report s=2 and s=4 in separate tables. Each (s, d, chain length) cell aggregates K=10 seeds across 2 noise × 2 graph configurations for linear NOTEARS and DAGMA (n=40 per cell), and across 1 noise × 2 graph configurations for NOTEARS-MLP (n=20 per cell). Cells report PPO and PACO mean±std and the per-cell ratio of means.

Backbone-dependent overhead. For Linear NOTEARS and DAGMA, each gradient step involves matrix-function evaluations, and PACO can incur a constant-factor overhead, with observed PACO/PPO runtime ratios in 0.89–3.23 across the grid. The overhead generally shrinks as the prior chain grows (`chain_1_100` cells), as a longer prior can more strongly restrict the search. The maximum observed ratio at d=50 is $2.51\times$ (Linear NOTEARS, s=4, `chain_1_50`); at s=2 it does not exceed $1.83\times$.

NOTEARS-MLP: overhead absorbed by neural inner loop. For NOTEARS-MLP the per-step cost includes per-variable MLP backward passes as well as matrix-exponential evaluation. The Frechet overhead is small in relative terms (observed runtime ratio ≤ 1.28 at `chain_1_50`), and at the longest single-chain prior (`chain_1_100`) PACO is uniformly faster than PPO across both

Table 9: Single-chain runtime (seconds) at edge density s=2. PPO and PACO entries are mean±std; the Ratio column is PACO/PPO (values > 1 mean PACO is slower in absolute terms, the constant-factor Frechet overhead predicted in Section 3).

		$d = 20$			$d = 30$			$d = 50$		
Backbone	Chain length	PPO	PACO	Ratio	PPO	PACO	Ratio	PPO	PACO	Ratio
Linear NOTEARS	`chain_1_50`	1.73±2.18	4.27±5.18	2.47	3.58±3.92	7.90±9.07	2.21	16.06±16.19	29.37±28.26	1.83
	`chain_1_75`	1.13±0.83	2.66±2.21	2.34	2.31±1.74	4.40±3.33	1.91	9.88±7.53	16.55±9.56	1.68
	`chain_1_100`	0.88±0.77	1.63±1.45	1.84	1.48±0.89	2.25±1.50	1.53	8.94±6.70	7.94±4.33	0.89
DAGMA	`chain_1_50`	5.62±4.51	10.17±9.29	1.81	12.31±11.56	21.44±20.64	1.74	40.65±27.88	70.18±48.35	1.73
	`chain_1_75`	4.78±3.70	7.73±7.29	1.62	9.61±8.44	14.87±13.37	1.55	32.44±19.88	51.82±32.07	1.60
	`chain_1_100`	3.71±2.72	4.70±3.94	1.27	8.03±6.84	8.57±6.75	1.07	27.55±16.57	31.98±20.05	1.16
NOTEARS-MLP	`chain_1_50`	17.22±6.88	20.80±9.39	1.21	30.97±13.77	34.03±15.92	1.10	89.61±52.91	94.33±54.75	1.05
	`chain_1_75`	14.88±5.42	17.19±8.13	1.16	27.29±13.46	27.08±13.96	0.99	94.75±92.74	78.69±51.90	0.83
	`chain_1_100`	12.27±3.93	10.50±4.91	0.86	24.28±10.36	15.31±6.86	0.63	78.30±59.71	41.49±24.72	0.53

Table 10: Single-chain runtime (seconds) at edge density s=4. Same conventions as Table 9.

		$d = 20$			$d = 30$			$d = 50$		
Backbone	Chain length	PPO	PACO	Ratio	PPO	PACO	Ratio	PPO	PACO	Ratio
Linear NOTEARS	`chain_1_50`	3.01±2.00	9.73±8.23	3.23	8.74±7.45	24.90±24.29	2.85	40.72±34.12	102.02±77.72	2.51
	`chain_1_75`	2.65±2.44	7.26±6.57	2.74	6.70±5.71	16.61±15.12	2.48	33.02±28.31	65.92±50.12	2.00
	`chain_1_100`	1.91±1.22	4.74±3.02	2.49	5.42±4.99	11.14±8.64	2.06	25.30±16.61	41.04±25.98	1.62
DAGMA	`chain_1_50`	10.92±7.72	22.97±17.84	2.10	26.73±17.11	63.79±44.97	2.39	105.51±69.83	252.63±187.06	2.39
	`chain_1_75`	8.51±5.48	18.67±16.42	2.19	21.69±12.19	48.52±29.62	2.24	89.71±61.76	208.50±148.44	2.32
	`chain_1_100`	6.93±4.62	12.84±8.83	1.85	18.44±11.25	34.33±18.71	1.86	69.95±40.67	155.49±112.75	2.22
NOTEARS-MLP	`chain_1_50`	19.63±6.49	25.06±9.72	1.28	32.70±12.02	36.87±14.58	1.13	93.03±35.74	104.93±47.29	1.13
	`chain_1_75`	18.21±5.81	23.48±8.90	1.29	30.08±10.51	31.68±12.71	1.05	77.20±25.62	88.24±39.65	1.14
	`chain_1_100`	14.53±3.40	13.99±5.07	0.96	23.86±8.69	16.57±6.56	0.69	63.73±25.24	39.86±19.17	0.63

densities and all d after aggregation over graph models, with the largest speedup ($1/0.53 \approx 1.89\times$) at s=2, d=50. In this regime the chain prior constrains a long sequence of variable orders.

Summary. The empirical results confirm Section 3's prediction: PACO's runtime advantage is concentrated in regimes where the prior induces multiple paths, and in single-chain regimes there can be a backbone-dependent constant-factor overhead. The observed overhead is at most $2.51\times$ at d=50 for linear backbones and substantially smaller for NOTEARS-MLP, where PACO is often faster. At high path counts, Section 4 reports approximately 8–38$\times$ speedups.

F Complete Synthetic Results

This appendix reports the complete synthetic benchmark for the PACO, PPO, and unconstrained baselines. Table 11 aggregates structural recovery (SHD, TPR, FDR, F1) and prior-order obedience across all single-chain length fractions (`chain_1_50/75/100`) and all multi-chain counts (`chain_2_50`, `chain_3_50`, `chain_5_50`, `chain_10_50`) for each backbone. The figures that follow break down the same data by noise model and edge density; each figure fixes one method family, one partial-order regime, and one graph family.

Hardware and runtime environment. For the $K = 10$ synthetic benchmark, linear NOTEARS and DAGMA experiments use four CPU cores and 8 GB memory; NOTEARS-MLP experiments use four CPU cores, 16 GB memory, and one NVIDIA L40S GPU. Paired PPO and PACO runs execute sequentially on the same node, with their order alternated across dataset seeds. Computation uses double precision, with BLAS/OpenMP thread counts set to one. CPU matrix functions use NumPy/SciPy, while MLP matrix exponentials use PyTorch on GPU. The main-grid MLP runs use PyTorch 2.11.0 with CUDA 13.0. The additional runs use Python 3.11, NumPy 2.3.5, and SciPy 1.17.1. For the $d = 100$ experiments, PPO runs have time limits of up to 34 hours; timed-out runs contribute their respective limits as lower bounds on runtime.

Table 11: Complete synthetic benchmark across all single-chain length fractions and multi-chain counts. Cells report mean±std across completed matched pairs in the cell. Each backbone block lists the unconstrained baseline (top row, no prior) and matched PPO/PACO pairs; bold marks the strictly better mean within a pair at the displayed precision (ties unbolded; baselines excluded from comparison). Baseline Obed shown as "—" (no prior to violate). Aggregated over $K = 10$ seeds across noise (`gauss`,`exp`) and graph (`ER`,`SF`) configurations (linear/DAGMA: 80 pairs per cell; NOTEARS-MLP: normally 40, with 39 and 38 for 3 and 5 chains at $d = 50$, single `mim` noise model).

Method-Prior	$d = 20$					$d = 30$					$d = 50$				
	SHD↓	TPR↑	FDR↓	F1↑	Obed↑	SHD↓	TPR↑	FDR↓	F1↑	Obed↑	SHD↓	TPR↑	FDR↓	F1↑	Obed↑
Linear NOTEARS															
NOTEARS (no prior)	26.20±18.01	0.66±0.15	0.13±0.11	0.75±0.13	—	34.39±24.16	0.72±0.12	0.11±0.10	0.79±0.11	—	52.46±38.05	0.75±0.10	0.10±0.09	0.82±0.09	—
PPO-1-chain-50	23.35±16.39	0.70±0.13	0.11±0.10	0.78±0.11	1.00	30.82±22.49	0.75±0.11	0.09±0.08	0.82±0.10	1.00	**44.46±31.21**	0.78±0.09	**0.07±0.07**	**0.85±0.08**	1.00
PACO-1-chain-50	**23.26±16.19**	0.70±0.13	0.11±0.09	0.78±0.11	1.00	**30.52±21.62**	0.75±0.11	0.09±0.08	0.82±0.09	1.00	46.24±34.18	0.78±0.09	0.08±0.08	0.84±0.08	1.00
PPO-1-chain-75	21.82±14.58	0.72±0.12	0.09±0.08	0.80±0.10	1.00	28.31±19.34	0.76±0.10	0.08±0.06	0.83±0.08	1.00	**39.36±25.17**	0.80±0.07	0.06±0.05	0.86±0.06	0.975
PACO-1-chain-75	**21.16±14.57**	0.72±0.12	0.09±0.07	0.80±0.10	1.00	**28.27±19.17**	0.76±0.10	0.08±0.07	0.83±0.08	1.00	39.46±26.13	0.80±0.07	**0.05±0.05**	0.86±0.06	**1.00**
PPO-1-chain-100	18.60±11.97	0.75±0.10	0.06±0.05	0.83±0.08	1.00	24.45±15.86	0.79±0.09	0.05±0.04	0.86±0.06	1.00	34.90±20.80	0.82±0.06	0.04±0.03	0.88±0.04	0.963
PACO-1-chain-100	**18.44±11.89**	0.75±0.10	0.06±0.05	0.83±0.07	1.00	**24.41±15.88**	0.79±0.09	0.05±0.04	0.86±0.06	1.00	**34.61±20.66**	0.82±0.06	0.04±0.03	0.88±0.04	**1.00**
PPO-2-chain-50	**21.54±13.98**	0.72±0.11	0.10±0.08	0.80±0.10	1.00	28.98±20.22	0.76±0.10	0.08±0.07	0.83±0.08	1.00	41.52±27.87	0.79±0.08	0.06±0.05	0.86±0.06	1.00
PACO-2-chain-50	21.71±14.58	0.72±0.12	0.10±0.08	0.80±0.10	1.00	**28.12±19.05**	0.76±0.10	**0.07±0.07**	0.83±0.08	1.00	**40.67±26.83**	0.79±0.08	0.06±0.05	0.86±0.06	1.00
PPO-3-chain-50	21.30±14.27	0.72±0.12	0.09±0.08	0.80±0.10	1.00	27.44±18.34	0.77±0.10	0.07±0.06	0.84±0.08	1.00	39.56±25.04	0.80±0.07	0.06±0.05	0.86±0.06	1.00
PACO-3-chain-50	**20.39±13.04**	**0.73±0.11**	**0.08±0.07**	**0.81±0.09**	1.00	**26.36±17.14**	0.77±0.09	**0.06±0.05**	**0.85±0.07**	1.00	**38.65±26.09**	0.80±0.07	**0.05±0.05**	**0.87±0.06**	1.00
PPO-5-chain-50	19.57±12.84	0.74±0.11	0.07±0.06	0.82±0.08	1.00	25.30±16.66	0.78±0.09	0.06±0.04	0.85±0.07	1.00	38.00±23.71	0.81±0.07	0.05±0.04	0.87±0.05	0.988
PACO-5-chain-50	**19.14±12.16**	**0.75±0.10**	0.07±0.05	**0.83±0.08**	1.00	**25.12±16.55**	0.78±0.09	**0.05±0.04**	**0.86±0.07**	1.00	**36.90±22.64**	0.81±0.06	0.05±0.04	0.87±0.05	**1.00**
PPO-10-chain-50	18.93±11.95	0.75±0.10	0.07±0.06	0.83±0.08	1.00	24.91±16.33	0.78±0.09	0.05±0.04	0.86±0.06	1.00	35.15±21.10	0.82±0.06	0.04±0.03	0.88±0.04	0.963
PACO-10-chain-50	**18.70±11.81**	0.75±0.10	0.07±0.05	0.83±0.07	1.00	**24.65±15.96**	**0.79±0.09**	0.05±0.04	0.86±0.06	1.00	**34.75±20.78**	0.82±0.06	0.04±0.03	0.88±0.04	**1.00**
DAGMA															
DAGMA (no prior)	24.96±16.30	0.67±0.14	0.12±0.10	0.76±0.12	—	31.56±21.68	0.73±0.11	0.09±0.07	0.81±0.09	—	49.66±35.68	0.75±0.11	0.07±0.07	0.83±0.09	—
PPO-1-chain-50	22.94±15.48	**0.70±0.12**	0.10±0.08	0.78±0.10	1.00	32.06±22.86	0.74±0.12	0.09±0.08	0.81±0.10	1.00	54.08±39.26	0.74±0.11	0.09±0.08	0.82±0.10	1.00
PACO-1-chain-50	**22.84±14.89**	0.69±0.12	**0.09±0.07**	0.78±0.10	1.00	**30.20±21.40**	0.74±0.11	**0.08±0.07**	**0.82±0.09**	1.00	**48.67±33.97**	**0.75±0.10**	**0.07±0.06**	**0.83±0.08**	1.00
PPO-1-chain-75	22.15±15.07	0.71±0.12	0.09±0.08	0.80±0.10	1.00	31.73±20.31	0.74±0.10	0.09±0.07	0.81±0.08	1.00	48.55±31.89	0.76±0.09	0.07±0.06	0.83±0.07	1.00
PACO-1-chain-75	**20.82±13.27**	**0.72±0.11**	**0.08±0.06**	**0.81±0.09**	1.00	**28.07±18.82**	**0.76±0.10**	**0.07±0.05**	**0.83±0.08**	1.00	**42.15±27.71**	**0.78±0.08**	**0.05±0.05**	**0.85±0.07**	1.00
PPO-1-chain-100	21.20±14.61	0.73±0.11	0.09±0.07	0.81±0.09	1.00	30.26±20.38	0.75±0.11	0.08±0.06	0.82±0.08	1.00	42.84±26.80	0.78±0.08	0.06±0.04	0.85±0.06	1.00
PACO-1-chain-100	**19.16±12.41**	**0.75±0.10**	**0.07±0.05**	**0.83±0.08**	1.00	**25.43±16.33**	**0.78±0.09**	**0.05±0.04**	**0.85±0.06**	1.00	**37.16±22.66**	**0.80±0.07**	**0.04±0.03**	**0.87±0.05**	1.00
PPO-2-chain-50	21.61±14.39	0.72±0.12	0.09±0.07	0.80±0.10	1.00	29.89±19.24	0.74±0.10	0.07±0.06	0.82±0.08	1.00	45.71±29.81	0.77±0.08	0.06±0.05	0.84±0.07	1.00
PACO-2-chain-50	**20.81±13.09**	0.72±0.11	**0.08±0.06**	**0.81±0.08**	1.00	**27.93±18.30**	**0.76±0.10**	0.07±0.06	**0.84±0.07**	1.00	**43.54±29.71**	**0.78±0.08**	0.06±0.06	**0.85±0.07**	1.00
PPO-3-chain-50	21.48±14.73	0.72±0.12	0.09±0.07	0.80±0.10	1.00	28.89±18.92	0.75±0.10	0.07±0.05	0.83±0.08	1.00	51.00±62.97	0.78±0.09	0.08±0.11	0.84±0.10	0.988
PACO-3-chain-50	**20.35±13.08**	**0.73±0.11**	**0.07±0.06**	**0.81±0.08**	1.00	**27.12±17.31**	**0.76±0.09**	**0.06±0.05**	**0.84±0.07**	1.00	**41.49±27.04**	**0.79±0.08**	**0.05±0.04**	**0.86±0.06**	**1.00**
PPO-5-chain-50	20.11±13.04	0.74±0.11	0.07±0.06	0.82±0.09	1.00	27.52±18.19	0.76±0.10	0.06±0.05	0.84±0.07	1.00	46.12±34.32	0.77±0.11	0.08±0.11	0.84±0.11	1.00
PACO-5-chain-50	**19.45±12.44**	0.74±0.10	0.07±0.05	0.82±0.08	1.00	**26.02±17.23**	**0.77±0.09**	**0.05±0.04**	**0.85±0.07**	1.00	**39.36±25.35**	**0.80±0.07**	**0.05±0.04**	**0.87±0.06**	1.00
PPO-10-chain-50	20.93±14.03	0.73±0.11	0.08±0.07	0.81±0.09	1.00	28.57±18.53	0.76±0.10	0.07±0.05	0.83±0.07	1.00	41.94±26.88	0.78±0.08	0.05±0.04	0.86±0.06	1.00
PACO-10-chain-50	**18.96±12.29**	**0.75±0.10**	**0.06±0.05**	**0.83±0.08**	1.00	**25.41±16.48**	**0.78±0.09**	**0.05±0.04**	**0.85±0.06**	1.00	**37.65±23.23**	**0.80±0.07**	**0.04±0.03**	**0.87±0.05**	1.00
NOTEARS-MLP															
NOTEARS-MLP (no prior)	17.90±7.12	0.81±0.08	0.14±0.07	0.83±0.04	—	25.00±14.76	0.77±0.14	0.07±0.05	0.84±0.08	—	43.30±31.05	0.73±0.17	0.03±0.02	0.83±0.11	—
PPO-1-chain-50	17.75±6.83	0.82±0.08	0.14±0.08	0.83±0.04	1.00	23.95±14.35	0.78±0.13	0.06±0.05	0.85±0.08	1.00	42.65±30.94	0.74±0.17	0.02±0.02	0.83±0.11	1.00
PACO-1-chain-50	**17.48±6.32**	0.82±0.07	0.14±0.09	**0.84±0.04**	1.00	**23.93±13.92**	0.78±0.13	0.06±0.05	0.85±0.08	1.00	42.65±30.99	0.74±0.17	0.02±0.02	0.83±0.11	1.00
PPO-1-chain-75	**16.75±5.97**	0.83±0.07	0.13±0.07	0.84±0.04	1.00	**23.68±14.13**	0.78±0.13	0.06±0.04	0.85±0.08	1.00	42.23±31.29	0.74±0.17	0.02±0.02	0.83±0.11	1.00
PACO-1-chain-75	16.77±6.43	0.83±0.07	0.13±0.08	**0.85±0.04**	1.00	23.82±13.64	**0.79±0.13**	0.06±0.05	0.85±0.07	1.00	**42.08±31.32**	0.74±0.17	0.02±0.02	0.83±0.11	1.00
PPO-1-chain-100	**15.95±5.88**	0.83±0.08	**0.11±0.07**	0.85±0.04	1.00	**23.18±13.89**	0.79±0.13	**0.05±0.04**	0.85±0.08	1.00	42.10±31.67	0.74±0.17	0.02±0.02	0.83±0.11	1.00
PACO-1-chain-100	16.23±6.01	**0.84±0.07**	0.12±0.08	0.85±0.04	1.00	23.27±12.92	0.79±0.13	0.06±0.05	0.85±0.07	1.00	**41.40±31.12**	**0.75±0.17**	0.02±0.02	**0.84±0.11**	1.00
PPO-2-chain-50	**16.75±5.84**	0.83±0.07	**0.13±0.08**	0.84±0.04	1.00	23.85±13.69	0.78±0.13	0.06±0.04	0.84±0.07	1.00	43.48±30.60	0.73±0.17	0.03±0.03	0.82±0.11	0.975
PACO-2-chain-50	16.85±5.61	**0.84±0.07**	0.14±0.08	0.84±0.03	1.00	**23.77±13.17**	**0.79±0.13**	0.06±0.05	**0.85±0.07**	1.00	**42.05±31.24**	**0.74±0.17**	**0.02±0.02**	**0.83±0.11**	**1.00**
PPO-3-chain-50	**16.50±6.12**	0.83±0.07	**0.12±0.09**	0.85±0.04	1.00	23.48±13.67	0.78±0.14	0.06±0.04	0.85±0.08	0.975	43.46±29.69	0.73±0.17	0.04±0.04	0.82±0.11	0.923
PACO-3-chain-50	16.93±6.12	**0.84±0.07**	0.14±0.08	0.85±0.04	1.00	**23.30±13.49**	**0.79±0.13**	0.06±0.04	0.85±0.07	**1.00**	**40.67±29.97**	**0.75±0.17**	**0.02±0.02**	**0.84±0.11**	**1.00**
PPO-5-chain-50	**16.15±6.19**	**0.84±0.07**	**0.12±0.08**	0.85±0.04	1.00	23.02±13.13	0.79±0.13	**0.05±0.05**	0.85±0.07	1.00	42.32±30.08	0.74±0.17	0.03±0.02	0.83±0.11	0.816
PACO-5-chain-50	16.70±6.35	0.83±0.07	0.13±0.08	0.85±0.04	1.00	23.02±13.05	0.79±0.13	0.06±0.05	0.85±0.07	1.00	**40.63±29.59**	**0.75±0.17**	**0.02±0.02**	**0.84±0.11**	**1.00**
PPO-10-chain-50	16.15±5.62	0.83±0.08	0.12±0.08	0.85±0.04	1.00	**23.10±14.09**	0.79±0.13	**0.05±0.04**	0.85±0.08	1.00	42.12±31.45	0.74±0.17	0.02±0.02	0.83±0.11	1.00
PACO-10-chain-50	16.15±5.90	**0.84±0.07**	0.12±0.07	0.85±0.03	1.00	23.30±12.91	0.79±0.13	0.06±0.05	0.85±0.07	1.00	**41.27±31.25**	**0.75±0.17**	0.02±0.02	**0.84±0.11**	1.00

F.1 Linear NOTEARS Results

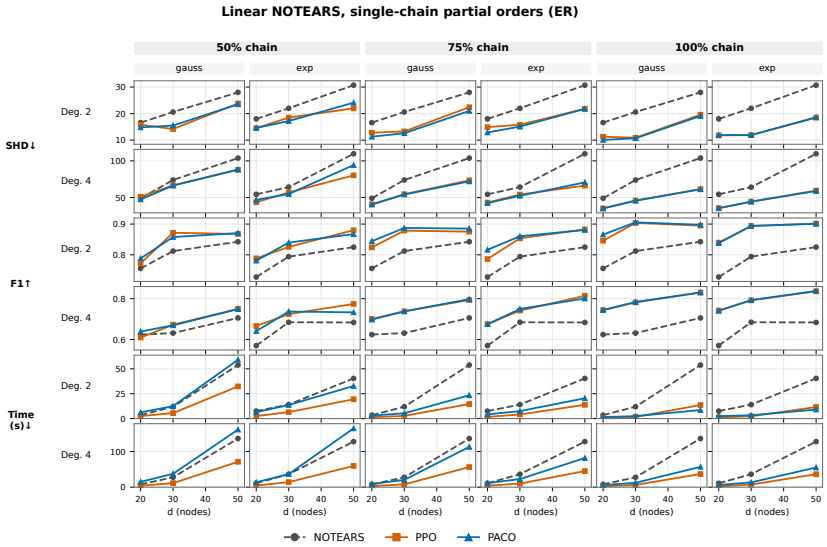

Figure 4: Linear NOTEARS on ER graphs, single-chain priors. Each panel: metric vs. $d \in \{20, 30, 50\}$ for unconstrained NOTEARS (gray dashed), PPO (orange), PACO (blue); points are means over $K=10$ seeds. SHD and F1 differences between PPO and PACO vary across settings; PACO generally incurs a constant-factor Frechet overhead in wall-clock time.

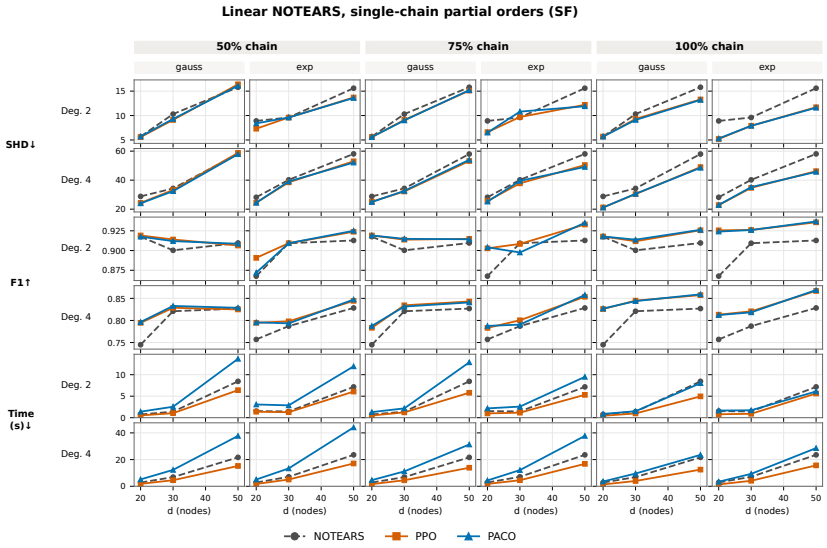

Figure 5: Linear NOTEARS on SF graphs, single-chain priors. Each panel: metric vs. $d \in \{20, 30, 50\}$ for unconstrained NOTEARS (gray dashed), PPO (orange), PACO (blue); points are means over $K=10$ seeds. As on ER, PPO and PACO have similar structural-recovery trends; prior-induced gains over the unconstrained baseline vary with chain length.

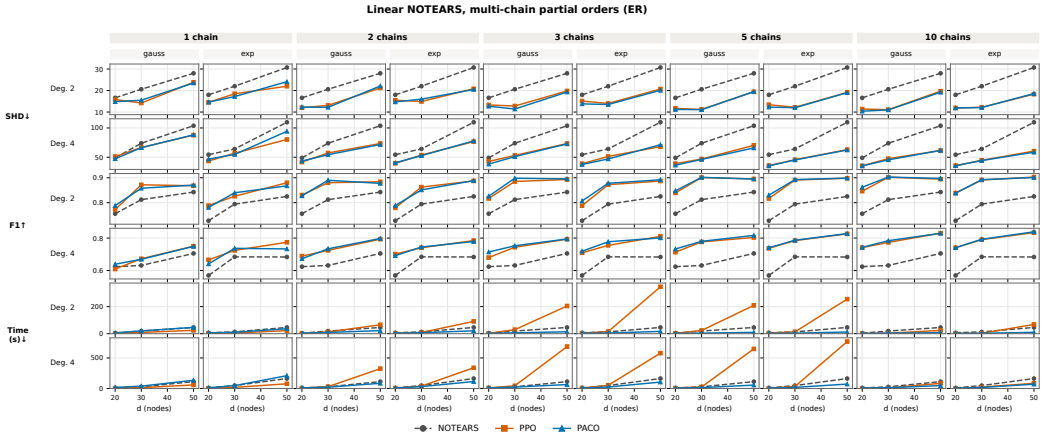

Figure 6: Linear NOTEARS on ER graphs, multi-chain priors $c \in \{1, 2, 3, 5, 10\}$ at chain length $0.5d$. Each panel: metric vs. $d \in \{20, 30, 50\}$ for unconstrained NOTEARS (gray dashed), PPO (orange), PACO (blue); points are means over $K{=}10$ seeds. SHD and F1 differences between PPO and PACO vary across settings; PACO can be faster in multi-chain settings, reflecting the path-count-dependent cost of path decomposition.

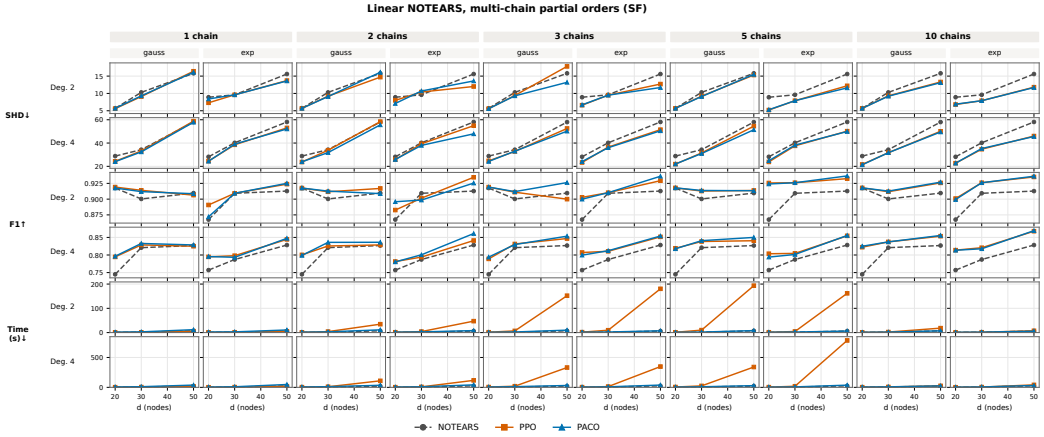

Figure 7: Linear NOTEARS on SF graphs, multi-chain priors $c \in \{1, 2, 3, 5, 10\}$ at chain length $0.5d$. Each panel: metric vs. $d \in \{20, 30, 50\}$ for unconstrained NOTEARS (gray dashed), PPO (orange), PACO (blue); points are means over $K{=}10$ seeds. Recovery differences and PACO runtime gains vary across settings, as on ER; the runtime gap depends on the induced $|\mathcal{P}(O^-)|$ for fixed c.

F.2 DAGMA Results

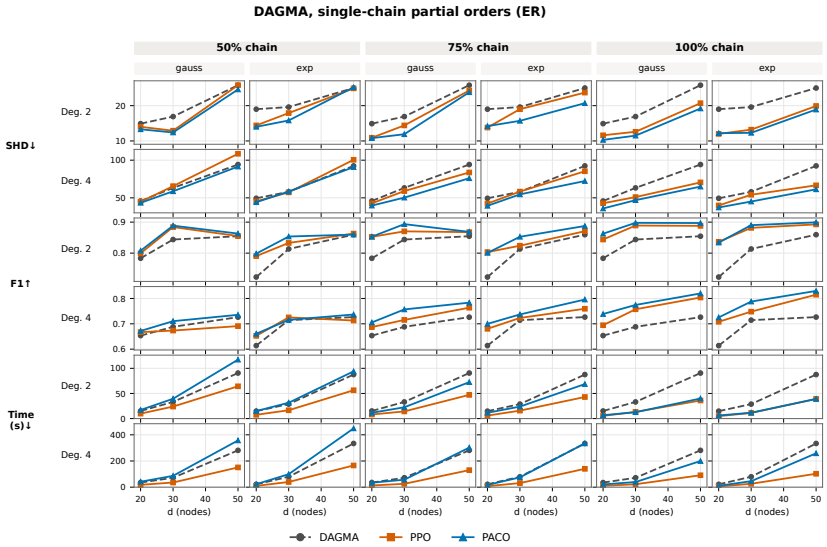

Figure 8: DAGMA on ER graphs, single-chain priors. Each panel: metric vs. $d \in \{20, 30, 50\}$ for unconstrained DAGMA (gray dashed), PPO-DAGMA (orange), PACO-DAGMA (blue); points are means over $K=10$ seeds. PACO-DAGMA often improves SHD over PPO-DAGMA; runtime shows the same general Frechet overhead pattern as linear NOTEARS, consistent with transfer across the log-determinant acyclicity formulation.

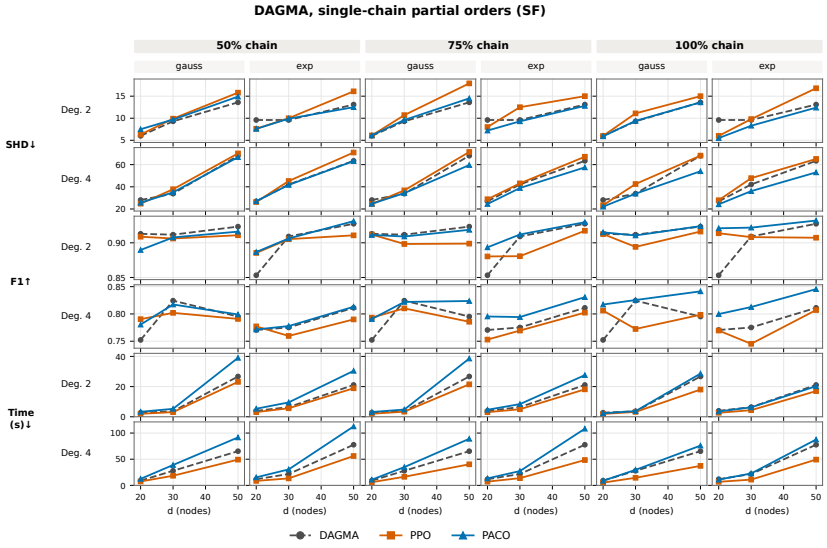

Figure 9: DAGMA on SF graphs, single-chain priors. Each panel: metric vs. $d \in \{20, 30, 50\}$ for unconstrained DAGMA (gray dashed), PPO-DAGMA (orange), PACO-DAGMA (blue); points are means over $K=10$ seeds. Same backbone-transfer pattern as on ER.

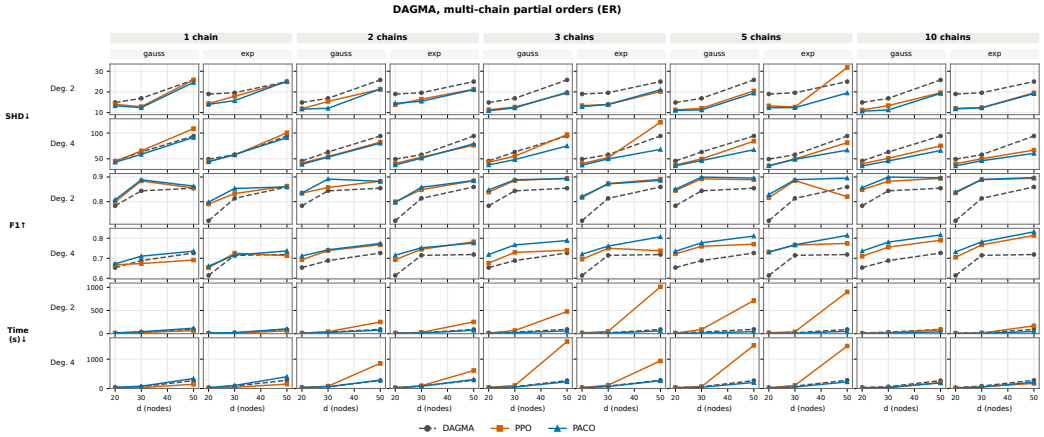

Figure 10: DAGMA on ER graphs, multi-chain priors $c \in \{1, 2, 3, 5, 10\}$ at chain length $0.5d$. Each panel: metric vs. $d \in \{20, 30, 50\}$ for unconstrained DAGMA (gray dashed), PPO-DAGMA (orange), PACO-DAGMA (blue); points are means over $K=10$ seeds. PACO-DAGMA often improves SHD over PPO-DAGMA and can be faster in multi-chain settings, paralleling the linear-backbone result.

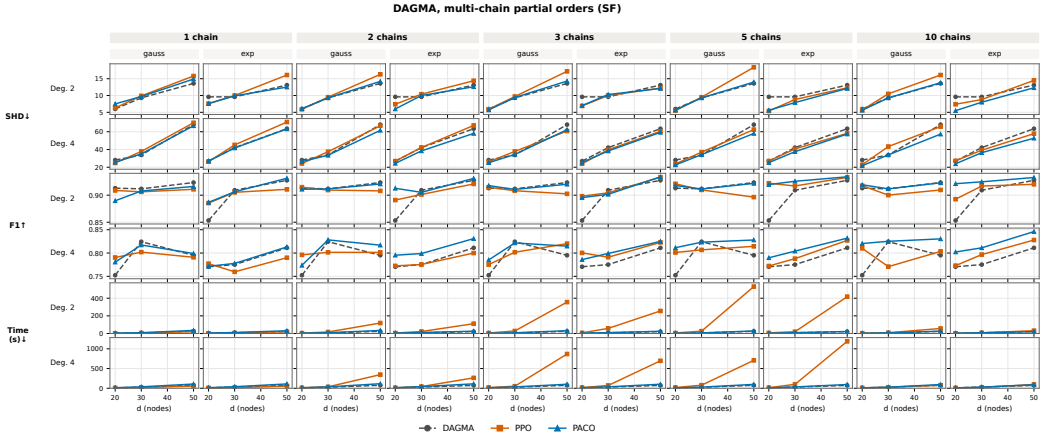

Figure 11: DAGMA on SF graphs, multi-chain priors $c \in \{1, 2, 3, 5, 10\}$ at chain length $0.5d$. Each panel: metric vs. $d \in \{20, 30, 50\}$ for unconstrained DAGMA (gray dashed), PPO-DAGMA (orange), PACO-DAGMA (blue); points are means over $K=10$ seeds. PACO-DAGMA's runtime advantage depends on c and the induced $|\mathcal{P}(O^-)|$; its crossover relative to ER varies across settings.

F.3 Nonlinear NOTEARS-MLP Results

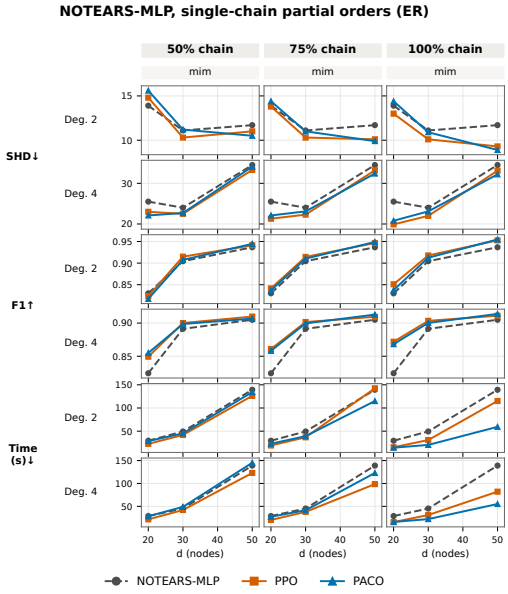

Figure 12: NOTEARS-MLP on ER graphs, single-chain priors (single `mim` noise). Each panel: metric vs. $d \in \{20, 30, 50\}$ for unconstrained NOTEARS-MLP (gray dashed), PPO-MLP (orange), PACO-MLP (blue); points are means over $K{=}10$ seeds. Frechet overhead is relatively small, and PACO-MLP is faster at the 100%-chain prior in five of the six plotted settings.

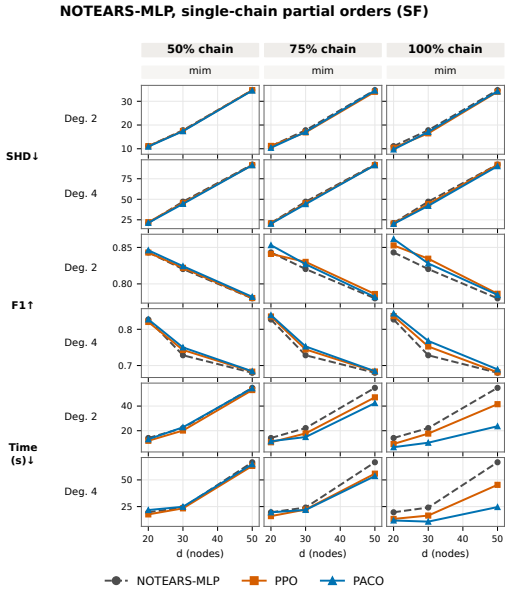

Figure 13: NOTEARS-MLP on SF graphs, single-chain priors (single `mim` noise). Each panel: metric vs. $d \in \{20, 30, 50\}$ for unconstrained NOTEARS-MLP (gray dashed), PPO-MLP (orange), PACO-MLP (blue); points are means over $K{=}10$ seeds. As on ER, PACO-MLP becomes faster than PPO-MLP at the 100%-chain prior.

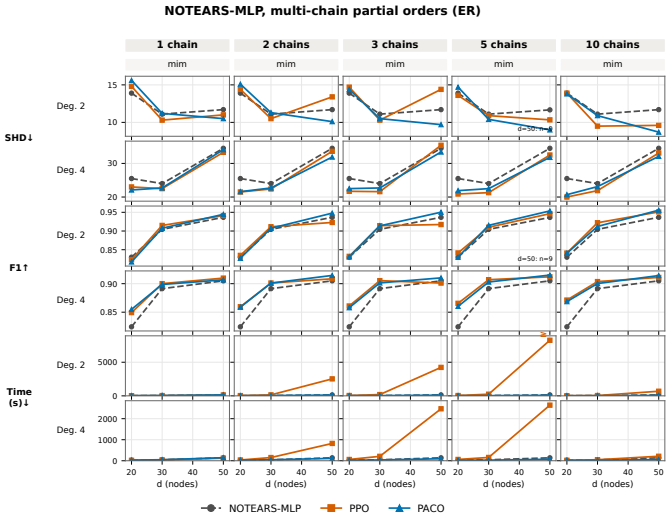

Figure 14: NOTEARS-MLP on ER graphs, multi-chain priors $c \in \{1, 2, 3, 5, 10\}$ at chain length $0.5d$ (single `mim` noise). Each panel: metric vs. $d \in \{20, 30, 50\}$ for unconstrained NOTEARS-MLP (gray dashed), PPO-MLP (orange), PACO-MLP (blue); points summarize experiments with $K{=}10$ seeds; sample counts and timeout handling are described below. PACO-MLP shows a large runtime advantage at large c.

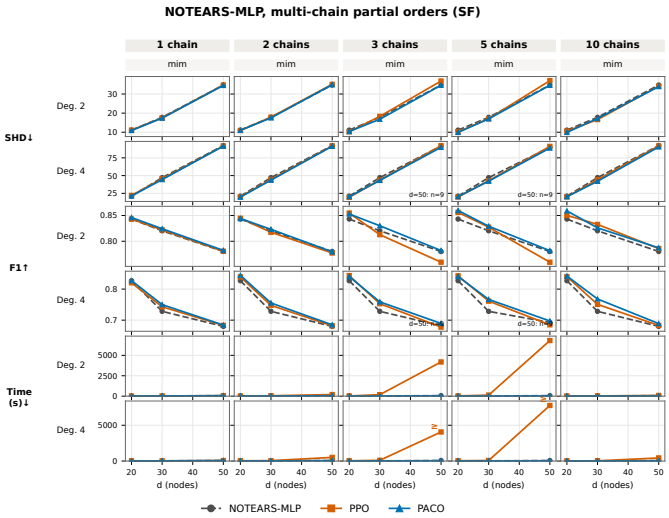

Figure 15: NOTEARS-MLP on SF graphs, multi-chain priors $c \in \{1, 2, 3, 5, 10\}$ at chain length $0.5d$ (single `mim` noise). Each panel: metric vs. $d \in \{20, 30, 50\}$ for unconstrained NOTEARS-MLP (gray dashed), PPO-MLP (orange), PACO-MLP (blue); points summarize experiments with $K{=}10$ seeds; sample counts and timeout handling are described below. PACO-MLP's reported runtime lead at large c is often wider on SF than on ER.

Recovery metrics and obedience use completed matched pairs, with counts in Table 11; the NOTEARS-MLP results for one chain of length $0.5d$ come from the experiments varying chain count. Runtime summaries retain all attempted pairs; the three PPO-MLP timeouts contribute their 5.5-hour limits to lower-bound means marked $\geq$, without imputed recovery metrics. Main-figure bootstrap intervals resample datasets within graph/noise/density settings, retaining paired methods and priors together; plotted path counts are bin medians.

G Sensitivity to Variable Scaling

We compare linear NOTEARS, PPO, PACO, and NOTIME [Berrevoets et al., 2025] under variable rescaling, using $K = 10$ independently seeded linear SEMs at each $d \in \{20, 50\}$. ER graphs have d edges, coefficients in $[-2, -0.5] \cup [0.5, 2]$, unit-scale exponential noise, and $n = 1000$ observations, differing from the main grid.

Paired scale transformations. Each dataset has three views: original, centered and standardized to unit sample variance, and centered and rescaled to reverse the rank order of marginal standard deviations. All views share the same underlying graph and observations. PPO and PACO share one-chain or five-chain priors, each chain covering $0.5d$ variables, fixed across views. NOTEARS and NOTIME use no prior.

Settings. The least-squares methods use $\lambda_1 = \omega = 0.3$, with $\eta = 1$ for PACO. NOTIME uses ICA initialization and dimension-specific calibration on three additional SEMs, disjoint from testing, across all three views. Selection minimizes mean SHD among candidates that converge and return DAGs in all nine calibration cases, breaking ties by FDR and edge count. Its dependence-penalty weight, regularization, and threshold $(\gamma, \lambda_1, \omega)$ are $(10, 0.001, 0.3)$ for $d = 20$ and $(10^{13}, 0.01, 0.3)$ for $d = 50$, fixed across test views.

Table 12: SHD under variable rescaling (lower is better), mean $\pm$ standard deviation over $K = 10$ paired datasets per cell.

d	Method	Original	Standardized	Reverse scale
20	NOTEARS	4.5 ± 1.8	13.8 ± 2.8	23.3 ± 3.3
20	NOTIME	5.6 ± 4.9	7.4 ± 6.4	15.3 ± 4.8
20	PPO (1 chain)	4.0 ± 1.9	13.0 ± 1.9	22.7 ± 3.7
20	PACO (1 chain)	4.1 ± 2.2	13.1 ± 2.1	22.5 ± 2.6
20	PPO (5 chains)	2.8 ± 1.3	10.6 ± 2.2	23.4 ± 4.7
20	PACO (5 chains)	2.6 ± 1.4	10.3 ± 2.2	17.2 ± 1.2
50	NOTEARS	6.4 ± 1.5	36.1 ± 4.6	55.1 ± 4.6
50	NOTIME	57.8 ± 16.6	53.4 ± 18.6	59.0 ± 17.6
50	PPO (1 chain)	5.7 ± 1.3	32.8 ± 3.9	54.8 ± 6.7
50	PACO (1 chain)	6.1 ± 1.7	33.0 ± 5.4	53.6 ± 6.3
50	PPO (5 chains)	5.1 ± 1.5	29.0 ± 2.9	54.2 ± 8.2
50	PACO (5 chains)	5.0 ± 1.6	25.0 ± 3.1	40.8 ± 4.0

Results. Rescaling increases mean SHD for all least-squares methods, although five-chain PACO improves over matched PPO in all six dimension–view combinations. Its recovery still degrades: from original to reverse-scale data, mean TPR falls from 0.875 to 0.180 at $d = 20$ and from 0.906 to 0.224 at $d = 50$. NOTIME improves SHD over NOTEARS on transformed data at $d = 20$, but has higher SHD and substantial variability at $d = 50$.

Five learned graphs are cyclic: three PPO outputs at $d = 20$ under reverse scaling (one with one chain and two with five chains), one one-chain PPO output at $d = 50$ under reverse scaling, and one one-chain PACO output at $d = 50$ under standardization. Three NOTIME fits at $d = 50$, one per view, did not meet the convergence criterion. All are retained in the metrics. Informative priors help, but PACO retains its least-squares backbone's scale sensitivity and statistical limitations.

H Experiments with Incorrect Partial-Order Priors

We compare an original prior O with $O^\top$, reversing all supplied relations while preserving acyclicity and maximal-path count. This severe stress test reverses all supplied relations; relations between truly incomparable variables remain non-conflicting after reversal. Thus, full transposition does not imply that every supplied relation contradicts a true ancestor relation.

Paired experiment. We use Gaussian-noise linear ER SEMs with $d \in \{20, 30, 50\}$, $s = 2$, $n = 4d$, and $K = 10$ datasets per dimension. PPO and PACO share each dataset's compatible `chain_5_50` prior and its exact transpose. At $d = 20, 30, 50$, respectively, 44.6%, 39.2%, and 26.3% of transposed closure relations conflict with true ancestor relations on average; the remainder concern incomparable variables. Both methods use $\lambda_1 = \omega = 0.3$, with $\eta = 1$ for PACO.

Table 13: Original versus transposed priors: SHD is mean $\pm$ standard deviation; TPR and FDR are means over $K = 10$ datasets per configuration. Obed and DAG are the fractions obeying the supplied prior and returning acyclic graphs. All 120 outputs are included.

d	Prior	Method	SHD ↓	TPR ↑	FDR ↓	Obed	DAG
20	Original	PPO	11.7 ± 4.8	0.782	0.093	1.00	1.00
20	Original	PACO	11.2 ± 3.9	0.792	0.088	1.00	1.00
20	Transposed	PPO	55.7 ± 6.6	0.150	0.803	0.00	0.60
20	Transposed	PACO	50.0 ± 3.1	0.027	0.941	1.00	1.00
30	Original	PPO	11.2 ± 3.2	0.850	0.041	1.00	1.00
30	Original	PACO	11.1 ± 3.3	0.852	0.041	1.00	1.00
30	Transposed	PPO	91.2 ± 9.0	0.222	0.785	0.00	0.20
30	Transposed	PACO	76.1 ± 5.0	0.017	0.966	1.00	1.00
50	Original	PPO	19.4 ± 4.9	0.836	0.036	1.00	1.00
50	Original	PACO	19.5 ± 5.0	0.836	0.037	1.00	1.00
50	Transposed	PPO	119.0 ± 18.7	0.404	0.606	0.00	0.50
50	Transposed	PACO	134.3 ± 6.2	0.018	0.969	1.00	1.00

Results. Transposition substantially degrades both methods. PACO always obeys the supplied prior and returns a DAG, yet mean TPR falls below 0.03 and FDR exceeds 0.93. Enforcing the reversed relations suppresses true causal directions, even when the returned graph satisfies every supplied constraint. PPO obeys none of the transposed priors and sometimes returns cyclic outputs.

Extension across graph and noise settings. We extend the transposed-prior experiment to $d = 50$, Gaussian/exponential noise, ER/SF graphs, and $s \in \{2, 4\}$, retaining $n = 4d$ and $K = 10$ (Table 14). The Gaussian ER $s = 2$ transposed-prior results at $d = 50$ in Table 13 are also included here.

Table 14: Transposed priors at $d = 50$, $K = 10$ per configuration. Metrics follow Table 13; non-DAG outputs are retained.

Noise/graph/s	Method	SHD ↓	TPR ↑	FDR ↓	Obed	DAG
Gauss./ER/2	PPO	119.0 ± 18.7	0.404	0.606	0.00	0.50
Gauss./ER/2	PACO	134.3 ± 6.2	0.018	0.969	1.00	1.00
Gauss./ER/4	PPO	269.4 ± 37.2	0.092	0.839	0.00	0.10
Gauss./ER/4	PACO	242.0 ± 5.8	0.024	0.931	1.00	1.00
Gauss./SF/2	PPO	117.7 ± 27.4	0.193	0.713	0.00	0.50
Gauss./SF/2	PACO	105.8 ± 4.2	0.025	0.915	1.00	1.00
Gauss./SF/4	PPO	194.6 ± 5.9	0.073	0.656	0.00	0.30
Gauss./SF/4	PACO	198.0 ± 5.3	0.029	0.840	1.00	1.00
Exp./ER/2	PPO	128.6 ± 16.0	0.327	0.691	0.00	0.50
Exp./ER/2	PACO	135.0 ± 7.1	0.019	0.967	1.00	1.00
Exp./ER/4	PPO	269.1 ± 45.2	0.210	0.770	0.00	0.30
Exp./ER/4	PACO	236.7 ± 5.6	0.009	0.966	1.00	1.00
Exp./SF/2	PPO	104.5 ± 20.2	0.228	0.665	0.00	0.50
Exp./SF/2	PACO	106.8 ± 4.5	0.026	0.919	1.00	1.00
Exp./SF/4	PPO	195.4 ± 10.5	0.058	0.669	0.00	0.30
Exp./SF/4	PACO	197.2 ± 5.1	0.030	0.839	1.00	1.00

Across the 80 pairs, PACO always returns a DAG obeying the prior despite poor recovery; PPO obeys none and returns 30 DAGs. Prior obedience therefore does not imply accurate recovery. PACO does not detect or correct incorrect priors; these results concern fully transposed priors.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: All the main claims are supported by our theoretical and experimental results.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discussed the limitations of our work in Section 5

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Theorem 1 (Section 3) states its assumptions explicitly (O^+ is the strict transitive closure of an acyclic partial-order prior, $\eta > 0$) and is followed by a complete proof in the main body.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Section 4 fully specifies the synthetic grid, sample sizes, $K{=}10$ seeded repeats, prior regimes, and Sachs setup; Appendix C provides the PACO algorithm with default hyperparameters; all priors and data are seeded deterministically.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: All datasets are publicly available; code is withheld for policy reasons. The paper provides the full PACO algorithm (Appendix C), default hyperparameters, and the seeded data/prior-generation scheme (Section 4) required to reproduce the main results.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 4 specifies sample sizes, $K{=}10$ seeded repeats, and prior regimes (no train/test split is used). Appendix C lists hyperparameter defaults and the L-BFGS-B inner-solver setup; defaults follow NOTEARS-family conventions, and the choice of $\eta = 1$ is supported by the sweep in Appendix D.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Sachs results and appendix result tables report mean$\pm$std; the main synthetic table reports means only for compactness, and Figure 2 reports descriptive 95% bootstrap confidence intervals.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.

- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Hardware (CPU and GPU) and per-experiment wall-clock runtime are reported in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We are guided by the NeurIPS Code of Ethics when conducting this research.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

Justification: We didn't include a dedicated discussion of societal impact, the question is not applicable to this work.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper doesn't introduce any dataset or model, the question is not applicable to this work.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We appropriately credit all used assets, specify their licenses, and adhere to their terms of use.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.

- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: This paper does not release any new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The work doesn't involve any crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs are not used as a component of the core methodology, experiments, or evaluation in this research.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.