Proceedings of Machine Learning Research 343:881–894, 2026 Probabilistic Graphical Models (PGM)

Partially Ordering Graphical Models with Latent Confounders by their Equality Constraints

Thijs van Ommen T.VANOMMEN@UU.NL

W. J. Merijn Roelfsema W.J.M.ROELFSEMA@STUDENTS.UU.NL

Department of Information and Computing Sciences
Utrecht University
Utrecht, The Netherlands

Editors: Gustau Camps-Valls, Manuele Leonelli and Gherardo Varando

Abstract

A linear structural equation model can represent causal relations and latent confounders between its variables by means of directed and bidirected edges. The set of distributions in such a graphical model can be difficult to describe: conditional independence constraints do not suffice in this case, so more general equality constraints are needed. We present an algorithm that can check whether two linear structural equation models differ by an equality constraint. Using this algorithm, we empirically study the partially ordered set of these models, for different subclasses of mixed graphs. We show that for some of these classes, a natural generalization of Meek's conjecture does not hold. We also investigate what set of constraints is needed to distinguish the models for various classes of graphs.

Keywords: Structural Equation Model; Mixed Graph; Equality Constraint; Algebraic Equivalence.

1. Introduction

Directed acyclic graphs (DAGs) are commonly used as models for causality, as the directed edges naturally lend themselves to a causal interpretation. But an unobserved variable that is a cause of two or more observed variables (i.e. a latent confounder) may also be responsible for observed correlations between those variables. Trying to explain these correlations using a DAG on just the observed variables will invariably lead to faulty conclusions.

For this reason, more complex graphical models have been introduced to capture both causal relations and latent confounders. But this brings with it new challenges, as in general, the statistical models defined by these graphs can be difficult to describe. Where for DAGs, it is possible to characterize exactly which distributions are and are not in the model by stating a list of conditional independences, this is typically not the case for graphs that also represent latent confounders. There, to characterize the contents of the model, a palette of constraints is needed that includes independences, but also more complex predicates on the observable joint distribution.

While we have a very good understanding of conditional independences, our understanding of those other constraints is still quite limited. A better understanding would be valuable. For example, it could be used to accelerate the estimation of the parameters of these models. It would also enable new advances in algorithms for causal discovery: the task of learning a causal graph from data.

© 2026 T. van Ommen & W.J.M. Roelfsema.

For the case of linear models, Van Ommen (2024) presented an algorithm that can decide, given two graphs, if all constraints imposed by one are also imposed by the other. Specifically, it considers all equality constraints. These are constraints that reduce a model's dimensionality, and thus are most likely to be testable on data in practice. In Section 3 of this work, we extend this algorithm to handle a larger class of graphs.

Using these algorithms, we can compute what the partial order of models looks like; see Figure 1 on page 885 for an illustration. This partial order captures important properties of the search space that causal discovery algorithms need to traverse. The Greedy Equivalence Search algorithm (GES; Meek, 1998) is a good example of this: it starts at the empty graph, proceeds upwards through the partially order set to find a model that contains the true model, then proceeds downwards from there to find the true model. The Meek conjecture, proved by Chickering (2002), guarantees that in the asymptotic case, this greedy algorithm is optimal. In Section 4, we check whether a natural generalization of the Meek conjecture holds for the larger graph classes we study. We also investigate what these partial orders can tell us about what set of constraints is needed to distinguish the models for various classes of graphs.

Ansanelli et al. (2026) study a similar partial order, but for discrete variables and also taking into account inequality constraints. Taeb et al. (2026) use partial orders over graphical models in order to define distances between graphs, though in the case of ADMGs, they consider a notion of model that also reflects interventional information, so that their partial order differs from ours.

2. Preliminaries

2.1. Graph Classes

A *directed acyclic graph (DAG)* is a directed graph with no directed cycles. A DAG naturally represents a Bayesian network, one of the most-used types of graphical models.

If a DAG is interpreted as representing causal relations, each directed edge $v \to w$ signifies that variable X_v is a direct cause of variable X_w (Pearl, 2000). For some sets of variables we may observe in practice, no DAG exists that correctly represents the causal relations between these variables. One reason why this may fail is because of the existence of *latent confounders*: variables that are not observed in our data but that are a cause of two or more observed variables. A solution that has been used to address this issue is to consider *acyclic directed mixed graphs (ADMGs)*, which are graphs that may have bidirected edges in addition to the directed ones. The directed edges may still not form directed cycles, but there is no such restriction on the bidirected edges. A bidirected edge $v \leftrightarrow w$ represents that a latent confounder may exist between X_v and X_w.

The class of ADMGs is very large, and various classes of graphs in between the classes of DAGs and of ADMGs have been studied in the graphical modelling literature. We will give some brief context for these below, but refer to the cited papers for their precise definitions.

Ancestral and *maximal ancestral (MAG)* graphs were introduced by Richardson and Spirtes (2002). (These graph classes also allow undirected edges as a third type of edge, but we do not consider these in this paper.) An ancestral graph is maximal ancestral if every nonadjacency corresponds to a conditional independence relation. MAGs play an important role for structure learning algorithms that test conditional independences in the data, such

as FCI (Spirtes et al., 1999; Zhang, 2008). Instead of returning a more general kind of ADMG, these can always return a MAG that satisfies precisely the found independences.

Equality constraints other than conditional independences exist. These motivate the definition of *arid* and *maximal arid (MArG)* graphs by Shpitser et al. (2018). For the case of discrete variables, it was shown by Evans (2018) that the relevant equality constraints are *nested Markov constraints* (Shpitser et al., 2014; Richardson et al., 2023). Intuitively speaking, a graphical model imposes a nested Markov constraint if we consider an intervention on one or more variables in the graph, compute the interventional distribution from the observational distribution, and find that this interventional distribution must satisfy a conditional independence. MArGs are to nested Markov models as MAGs are to Markov models: for every nonadjacency there is a nested Markov constraint, and a structure learning algorithm based on testing nested Markov constraints could limit its search to MArGs.

In an ADMG (but not in any of the subclasses discussed above), a bidirected edge can occur between two variables that also have a directed edge between them. This configuration is called a *bow*. A *bow-free acyclic path diagram (BAP)* is an ADMG without bows.

2.2. Algebraic Models

There are different ways to associate a statistical model to an ADMG G. In this paper, we consider *linear structural equation models (LSEMs)*. An LSEM defines a joint distribution on the observed variables $(X_v)_{v\in V}$ through

$$X_v = \lambda_{0v} + \sum_{p\in V\setminus\{v\}} \lambda_{pv} X_p + \epsilon_v.$$

The parameters λ_{pv} are only nonzero if $p \to v$ is in the ADMG G. We assume the ϵ_v have expectation 0. We do not assume all pairs are independent, but only assume independence for pairs ϵ_v and ϵ_w for which $v \leftrightarrow w$ is not in G. This is how these models represent the possibility of latent confounders. The λ_{0v} parameters allow each X_v to be shifted arbitrarily; because such shifts give no information about the underlying G, we do not consider these parameters further.

If the ϵ_v are assumed to be Gaussian, their distribution is fully parameterized by their variances and covariances. Let ω_{vw} be the covariance between ϵ_v and ϵ_w (with as special case that ω_{vv} is the variance of ϵ_v). We gather these parameters into two matrices $\Lambda = (\lambda_{vw})_{vw}$ and $\Omega = (\omega_{vw})_{vw}$. The covariance matrix Σ of the observed variables $(X_v)_v$ is then computed from Λ and Ω by the *parametrization map*

$$\phi(\Lambda, \Omega) = (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1}, \tag{1}$$

where $^{-T}$ denotes the transposed inverse; see e.g. (Foygel et al., 2012).

We note that if the ϵ_v deviate substantially from Gaussianity, then there are signals in the observed joint distribution that are relevant for reconstructing G, but that are not captured by the covariance matrix Σ. If the ϵ_v are approximately Gaussian, then the strongest signals about G are captured by Σ.

Let $\mathcal{M}(G)$ be the set of all Σ that can come out of (1) for different choices of Λ and Ω that are permitted by G. For many graphical models, we can characterize which Σ are in $\mathcal{M}(G)$ by stating, for instance, that X_a must be independent of X_c given X_b. Any Σ that satisfies

all stated constraints is then in $\mathcal{M}(G)$. Such a *Markov model* description does not suffice for many ADMGs. Additional types of constraints are needed to give a full description of $\mathcal{M}(G)$: both equality constraints of the form $\sigma_{ab}\sigma_{cd} = \sigma_{ac}\sigma_{bd}$, and inequality constraints involving $\geq$ or $\neq$ instead of $=$. In this paper, we focus on just equality constraints and do not consider the even more complex inequality constraints.

Because the equality constraints in LSEMs can always be written as $f(\Sigma) = 0$ for some polynomial f, they are also known as *algebraic* constraints. If we extend $\mathcal{M}(G)$ to include all Σ that satisfy the equality constraints, regardless of whether they violate any inequality constraints, we obtain the *algebraic model* $\overline{\mathcal{M}}(G)$.

Two ADMGs G_1, G_2 may represent the same algebraic model: $\overline{\mathcal{M}}(G_1) = \overline{\mathcal{M}}(G_2)$. In this case we say the graphs are *algebraically equivalent* (Van Ommen and Mooij, 2017). If G_1 and G_2 are DAGs or MAGs, they are algebraically equivalent iff they are Markov equivalent. For more general types of ADMGs, algebraic equivalence implies Markov equivalence but not the other way around: there are pairs of ADMG models that impose the same conditional independences and hence are Markov equivalent, but that can be distinguished using other algebraic constraints and hence are not algebraically equivalent. Two examples are the graphs in Figure 3(b) and (c) on page 891: their algebraic models impose no independence constraints, yet are different from the saturated model, and also from each other.

2.3. Partially Ordered Sets

A *partially ordered set*, or *poset*, is a set equipped with a reflexive, antisymmetric, transitive relation (Davey and Priestley, 2012). An example is a collection of sets equipped with the inclusion relation $\subseteq$. Note that some pairs may be incomparable: we have neither $\{1,2\} \subseteq \{2,3\}$ nor $\{1,2\} \supseteq \{2,3\}$. We write $A \subsetneq B$ for ($A \subseteq B$ and $A \neq B$).

In this paper, we consider various posets whose elements are the distinct algebraic models defined by all n-node graphs in one of the classes discussed in Section 2.1. Figure 1 illustrates this for DAGs with $n = 3$. For a DAG, the algebraic model equals the Markov model, so this figure can also be understood as the poset of Markov models.

For $n = 3$, each ADMG is algebraically equivalent to a DAG, so the poset of ADMGs still consists of the same eleven algebraic models. As we will see, for larger n, things get substantially more complicated, especially for the larger classes of graphs.

The number of distinct algebraic models is limited by the number of such graphs, hence finite. So the posets we consider in this paper will be *finite* posets.

Each of these posets contains two distinguished elements: the *trivial* model, corresponding to the graph with no edges, and the *saturated* model containing all distributions, corresponding to e.g. any complete DAG. For each model M in the poset, it holds that M contains the trivial model and is contained in the saturated model. This means that the poset is *bounded*.

An element B of a poset is said to *cover* A if $A \subsetneq B$ and there exists no C with $A \subsetneq C \subsetneq B$. An *atom* is an element that covers the trivial model. Dually, a *coatom* is an element covered by the saturated model.

A poset is called *graded* (Grätzer, 2003) if there exists a rank function from the elements of the poset to the integers such that for all pairs $M \subseteq N$, we have $\text{rank}(N) = \text{rank}(M) + 1$ iff N covers M. For $n = 3$, our posets are graded: the rank function is displayed in

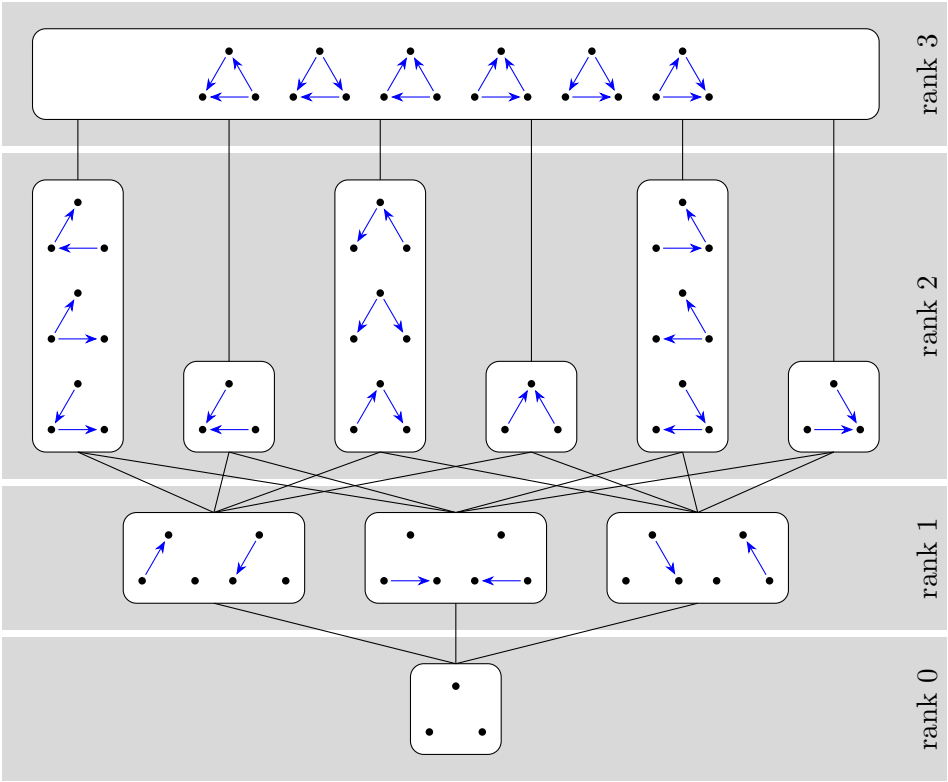


Figure 1: Each DAG on three variables defines one of eleven algebraic models. These models are represented here by the corresponding sets of DAGs. This diagram displays the inclusion relations that exist between these models. Relations implied by transitivity are not shown.

Figure 1. We see that this rank equals the number of edges in the graphs. It also coincides with the *dimension* (Cox et al., 2015) of the algebraic model minus n. A model's dimension $\dim(\mathcal{M}(G))$ can be thought of as the number of free parameters in the model: one λ_{vw} or ω_{vw} for each edge in G, and one ω_{vv} for each variable.

Also for larger classes of graphs we may expect these posets to be graded with a rank function counting the number of edges in the graphs. For ADMGs with bows we would need to be more precise: some ADMGs with bows are algebraically equivalent to graphs with fewer edges. By (Van Ommen and Mooij, 2017, Corollary 3), we can take $\text{rank}(M)$ to be the *smallest* number of edges among ADMGs that have M as their algebraic model. This still satisfies $\text{rank}(M) = \dim(M) - n$. For algebraic models, we know that $M \subsetneq N$ implies $\dim(M) < \dim(N)$, so if two comparable models differ by dimension one, then the larger covers the smaller. The only thing left to show to establish that the poset is graded by rank is that whenever two comparable models differ by rank two or more, neither covers the other, i.e. an intermediate model exists. This is not immediate, and it is one of the questions we will investigate empirically in Section 4.

3. Algorithm for Determining ADMG Model Inclusion

Van Ommen (2024) presented algorithms that, given two BAPs G and G', can decide whether $\overline{\mathcal{M}}(G) \subseteq \overline{\mathcal{M}}(G')$, and whether $\overline{\mathcal{M}}(G) = \overline{\mathcal{M}}(G')$, i.e. the two are algebraically equivalent. These algorithms essentially work by sampling a random element Σ from $\mathcal{M}(G)$ and checking whether $\Sigma \in \overline{\mathcal{M}}(G')$. One thing that may go wrong with this approach, is that Σ may happen to be in $\overline{\mathcal{M}}(G')$ even if $\overline{\mathcal{M}}(G) \not\subseteq \overline{\mathcal{M}}(G')$. It can be shown that this will 'rarely' happen: the probability of such an error can be made arbitrarily small by adjusting the parameter p of the algorithm. There is no probability of error in the other direction: if the algorithm finds a $\Sigma \notin \overline{\mathcal{M}}(G')$, it will report that $\overline{\mathcal{M}}(G) \not\subseteq \overline{\mathcal{M}}(G')$, and such an answer is always correct.

The algorithms presented there are limited to BAPs, because for more general ADMGs it is hard to test whether $\Sigma \in \overline{\mathcal{M}}(G')$. The algorithms essentially do this test by identifying parameters (Λ', Ω') compatible with G' such that $\phi(\Lambda', \Omega') = \Sigma$. This identification procedure requires solving a sequence of linear systems. If G' is a BAP, it is shown that the systems encountered this way are 'rarely' singular, and so the algorithm can answer positively and with high probability be correct. If G' is not a BAP, then singular systems may be more common; in fact, there exist combinations of G and G' for which the identification algorithm will always encounter a singular system, yet $\overline{\mathcal{M}}(G) \not\subseteq \overline{\mathcal{M}}(G')$.

In this paper, we extend the algorithm to work for non-BAP ADMGs G'. The new version is presented in Algorithm 1. We can test whether $\overline{\mathcal{M}}(G) = \overline{\mathcal{M}}(G')$ by running Algorithm 1 a second time with G and G' interchanged. Like the algorithms of Van Ommen (2024), we use computations over the finite field $\mathbb{F}_p$ (i.e. we use exact integer arithmetic modulo a large prime p) to effectively bound the error probabilities. The extension requires two main changes to the algorithm:

- more powerful identification algorithms are needed;
- singular systems must be handled more carefully.

3.1. Identification Algorithms

The algorithms of Van Ommen (2024) use the HTC identification criterion of Foygel et al. (2012) — or rather, a special case of this algorithm that suffices to identify the parameters of BAPs. HTC is also applicable to many graphs with bows, and even to graphs with directed cycles. Yet ADMGs remain for which it is mathematically possible to identify the parameters, but that are not recognized as such by HTC. More general criteria have therefore been proposed, though no complete criterion is known.

For our empirical results, we have used two identification algorithms in addition to HTC: EID and TSID, both by Weihs et al. (2017). Where HTC always attempts to identify the parameters $\lambda_{\cdot v}$ of all directed edges into a single node v at the same time, the *edgewise identification (EID)* algorithm can also identify smaller subsets of edges at a time, and sometimes make new progress this way. The *trek-separation identification (TSID)* algorithm can identify a single directed edge by looking for t-separations (Sullivant et al., 2010).

Given a graph G', these criteria and algorithms establish a sequence of *identification steps*, each filling in the values for one or more entries of Λ'. Once Λ' is known, Ω' can be identified as in line 6 of Algorithm 1.

Algorithm 1: Determine whether one algebraic model is contained in another.

Input: Two ADMGs G and G', identification steps Q that identify G', and a prime p
Output: Is $\overline{\mathcal{M}}(G) \subseteq \overline{\mathcal{M}}(G')$?

1 Sample Λ and Ω for G uniformly at random from $\mathbb{F}_p$
2 Let $\Sigma = (I - \Lambda)^{-T}\Omega(I - \Lambda)^{-1}$
3 Identify Λ' from Σ for G' using identification steps Q
4 **if** an identification step encountered a linear system with no solutions **then**
5 | **return** `inconclusive` *// No evidence either way*
6 Let $\Omega' = \Lambda'^{T}\Sigma\Lambda'$
7 **if** $\Omega'_{vw} = 0$ for all $\{v, w\}$ with $v \leftrightarrow w$ not in G' **then**
8 | **return** `true` *// Evidence that* $\overline{\mathcal{M}}(G) \subseteq \overline{\mathcal{M}}(G')$
9 **else if** an identification step encountered a linear system with multiple solutions **then**
10 | **return** `inconclusive` *// No evidence either way*
11 **else**
12 | **return** `false` *// Definitely* $\overline{\mathcal{M}}(G) \not\subseteq \overline{\mathcal{M}}(G')$

We assume here that the identification steps compute each entry of Λ' (and of Ω') as a *rational* function of Σ. This is true for HTC and EID, which solve a linear system in each step, and for TSID, which takes the quotient of two determinants. To our knowledge, no example of a generically but not rationally identifiable LSEM model is known, which would suggest that this assumption is without loss of generality, and Algorithm 1 will work correctly also when paired with any future identification algorithms.

3.2. Dealing with Singular Systems

The identification steps used in Algorithm 1 may encounter singular systems of linear equations. A system may be either overdetermined (having no solution) or underdetermined (having multiple solutions: infinitely many over the real numbers). If an overdetermined system is encountered, the computation cannot proceed, and `inconclusive` is returned.

In the case of an underdetermined system, it is possible to pick one of the many solutions and continue the computation. If the computation completes and finds parameters Λ' and Ω' for G' that reproduce Σ, then this establishes that $\Sigma \in \mathcal{M}(G')$, and the algorithm can answer `true`. If the computation completes but the parameters do not satisfy the requirements, then the algorithm must answer `inconclusive`: it is possible that $\Sigma \in \mathcal{M}(G')$ and that we would have learned this if we had picked a different solution.

If G is the empty graph or a graph with few edges, then Σ generated in line 2 of Algorithm 1 will have many zeroes. If simultaneously G' is a more complex ADMG where identification is nontrivial, it is quite likely that an identification step will need to solve an equation of the form $0 \cdot \lambda_{vw} = 0$. These are the cases in which it helps most to pick a solution and continue the computation. Without this special case, the algorithm would often answer `inconclusive` for inputs where 'obviously' $\overline{\mathcal{M}}(G) \subseteq \overline{\mathcal{M}}(G')$.

We found empirically that we more often get a conclusive answer if we prefer to pick zero as a solution, compared to when picking a solution uniformly at random. In the general case of linear systems with multiple variables, we did this as follows: We put the augmented

matrix in reduced row-echelon form. Then each variable whose column does not contain a pivot is assigned the value zero. A unique solution exists for the remaining variables.

3.3. Error Bounds

Algorithm 1 returns one of three possible results: `true`, `false`, or `inconclusive`. The latter possibility is necessary because there exist inputs for which the identification steps will always encounter a singular linear system, for any randomly sampled Σ, and so the algorithm is truly unable to provide a conclusive answer in this case. But an inconclusive result can also appear by chance: a singular system was encountered for the randomly sampled Σ, but a different random sample would not have encountered a singular system and led to a conclusive answer. We bound the probability of this type of error in equation (2) below.

The other type of error is that Algorithm 1 may give a false positive answer: it may return `true` even though $\overline{\mathcal{M}}(G) \not\subseteq \overline{\mathcal{M}}(G')$. This probability is bounded in equation (3). It is important to note that the algorithm never produces false negative results.

The proofs of equations (2) and (3) are based on the proofs by Van Ommen (2024), and rely similarly on the Schwartz–Zippel lemma (Schwartz, 1980). This lemma bounds the number of zeroes of a polynomial in terms of the degree of that polynomial. Thus, the degrees of relevant polynomials will figure in the bounds.

We write ℓ_G for the largest number of edges in any directed path in G. Then in line 2 of the algorithm, each entry of Σ is computed as a polynomial with degree at most $2\ell_G + 1$.

In line 3, the algorithm proceeds to identify Λ' as a rational function of Σ. A rational function has a polynomial numerator and denominator. We have that $c\phi(\Lambda, \Omega) = \phi(\Lambda, c\Omega)$, so a rescaling of Σ should not change Λ'. This implies that the numerator and denominator polynomials, expressed in terms of Σ, must have the same degree. For each identification step i, let d_i be this degree for the rational functions encountered there. Note that exponential growth can occur here, because each identification step may use previously computed values of Λ'.

Finally, in line 6, Ω' is computed through polynomial functions of Λ' and Σ. In terms of Σ, these are rational functions with numerator of degree at most $2 \max_i d_i + 1$; expressed in terms of the original (Λ, Ω), the degrees are at most $(2\ell_G + 1)(2 \max_i d_i + 1)$.

We see that `inconclusive` may be returned if *any* of the identification steps in line 3 had a zero denominator, and `true` may be returned if *all* numerators in line 6 are zero. Applying the Schwartz–Zippel lemma (Schwartz, 1980), we find

$$P\left(\text{algorithm returns } \texttt{inconclusive} \middle| \exists \Sigma \in \mathcal{M}(G) \text{ s.t. all systems nonsingular}\right) \leq \frac{2\ell_G + 1}{p} \sum_i d_i \quad (2)$$

and

$$P\left(\text{algorithm returns } \texttt{true} \mid \overline{\mathcal{M}}(G) \not\subseteq \overline{\mathcal{M}}(G')\right) \leq \frac{2\ell_G + 1}{p} \left(2 \max_i d_i + 1\right). \quad (3)$$

Though these bounds are far from tight, they demonstrate that Algorithm 1 is highly accurate when parameter p is taken to be a suitably large prime.

Table 1: Numbers of graphs, algebraic models, and coatoms for ADMGs and the subclasses of ADMGs studied in this paper, for $n = 4$ and $n = 5$. The columns labelled '/iso' list the numbers of algebraic models and coatoms up to isomorphism.

	$n = 4$					$n = 5$				
Class	graphs	models	/iso	coatoms	/iso	graphs	models	/iso	coatoms	/iso
ADMG	34,752	389	34	135	10	29,983,744	168,579	1,668	43,565	409
BAP	3,608	368	31	120	8	765,664	87,219	882	8,720	87
arid	3,560	368	31	120	8	719,344	76,119	787	3,860	44
MArG	3,524	332	29	84	6	687,424	56,559	612	1,640	22
ancestral	2,504	260	26	36	4	336,064	28,399	354	260	7
MAG	2,492	248	25	24	3	328,924	24,259	313	80	4
DAG	543	185	20	24	3	29,281	8,782	142	80	4

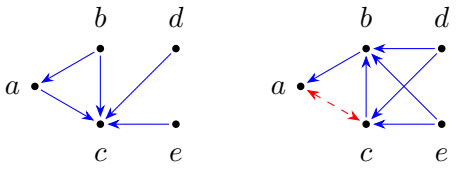
4. Empirical Results on the Partial Orders

Using Algorithm 1, we can construct posets and check their properties. We did this for $n = 4$ and $n = 5$. Due to the numbers of graphs, this is computationally expensive even with the help of efficient algorithms. Therefore we first grouped together all graphs for which algebraic equivalence follows from Van Ommen and Mooij (2017, Theorem 2) into *clusters*. This gave us 194,574 clusters for $n = 5$. Also, we say two algebraic models are isomorphic if their variables can be renamed to make them equal. For example, in Figure 1, the three algebraic models of rank 1 are isomorphic to each other, while among models of rank 2, there are two sets of three isomorphic models each. For each set of isomorphic clusters, we only took a G from one representative, comparing it to G's from all clusters; the skipped results can then be reconstructed from results for other pairs (G, G').

For each such pair of clusters of graphs, if the second cluster contained a BAP G', we used Algorithm 2 of Van Ommen (2024) as it always gives a conclusive answer. Otherwise we ran our Algorithm 1 with each graph from the second cluster until receiving a conclusive result. We used $p = 2^{31} - 1$ for both algorithms. For $n = 4$, 21 clusters required our new algorithm; this led to inconclusive results in 7% of these cases. For $n = 5$, our new algorithm was used for 105,950 clusters, and 8.6% of the cases were inconclusive; 1225 clusters were not identifiable with HTC/EID/TSID and thus the algorithm could not be used. We also took into account that if G is a subgraph of G', then $\overline{\mathcal{M}}(G) \subseteq \overline{\mathcal{M}}(G')$.

In inconclusive cases, it might be possible to still get a conclusive result by using a different identification algorithm (e.g. EID instead of HTC), by choosing different identification steps where the identification algorithm had a choice of options, or even by changing the preference of where to put 0's in case of an underdetermined linear system (see Section 3.2). This would add to the already large computational cost, while the benefit might be limited.

Results about the numbers of unique algebraic models are displayed in Table 1. The counts for ADMGs could be overestimates due to inconclusive results from our algorithm: sets of ADMGs are counted as not algebraically equivalent unless we could establish that one's model included the other and vice versa.



(a) DAG $G^{(5)}$ (b) MArG $G^{(7)}$

Figure 2: Counterexample demonstrating that the MArG and larger posets are not graded.

The equivalence classes for $n = 4$ were described in (Van Ommen and Mooij, 2017, Appendix B). This is the first time that similar results for $n = 5$ were gathered.

4.1. Gradedness and Meek's Conjecture

At the end of Section 2.3, we raised the question whether the posets we study are graded: in particular, if two models M and N satisfy $M \subsetneq N$, is there always a sequence of models between them such that $M = M_1 \subsetneq M_2 \subsetneq \dots M_k = N$ and such that adjacent models' ranks differ by 1? This would imply that to know all relations in the partial ordering, it suffices to know the relations between pairs of graphs that differ by rank 1, as the other relations then follow from transitivity.

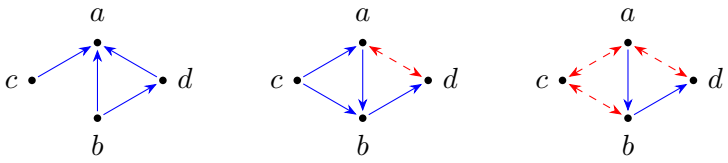
This question is relevant for causal discovery because it is a necessary condition for Meek's conjecture (Meek, 1998). For DAGs, this conjecture states that for two models with $M \subsetneq N$, a sequence of intermediate models as above exists, and further that for each adjacent pair of models M_i, M_j, there exists DAGs G_i, G_j with $\mathcal{M}(G_i) = M_i$, $\mathcal{M}(G_j) = M_j$ such that G_i can be obtained from G_j by deleting a single edge. This conjecture, proved by Chickering (2002), establishes the optimality of GES for DAGs. We ask whether the same condition holds for larger classes of graphs.

Both Claassen and Bucur (2022) and Hu and Evans (2024) present a version of GES generalized to MAGs, including efficient methods to traverse the poset of MAGs. The latter paper mentions the same generalization of Meek's conjecture that we discuss here, and their algorithm becomes more efficient if this conjecture is assumed true.

Results For $n = 4$, we found no counterexamples. But for $n = 5$ it turns out that the posets encompassing non-ancestral graphs are not graded. See Figure 2 for a pair of graphs $G^{(5)}$, $G^{(7)}$ that differ by grade 2 and have $\overline{\mathcal{M}}(G^{(5)}) \subsetneq \overline{\mathcal{M}}(G^{(7)})$, but there is no intermediate MArG $G^{(6)}$, or in fact any ADMG, with $\overline{\mathcal{M}}(G^{(5)}) \subsetneq \overline{\mathcal{M}}(G^{(6)}) \subsetneq \overline{\mathcal{M}}(G^{(7)})$.

The other part of Meek's conjecture asks whether for pairs of models $M_i \subsetneq M_j$ that differ by rank 1, there exist corresponding graphs G_i, G_j that differ by removing one edge. We also checked this on our computed posets.

Results For the posets of DAGs, of MAGs, and of ancestral graphs, no counterexamples were found for $n = 4$ or $n = 5$. But when including nonancestral graphs, again counterexamples were found. Two are displayed in Figure 3. Algorithm 1 can certify that $\overline{\mathcal{M}}(G_0) \subseteq \overline{\mathcal{M}}(G_1)$ and $\overline{\mathcal{M}}(G_0) \subseteq \overline{\mathcal{M}}(G_2)$, and we also checked this algebraically. There are many graphs G_0' algebraically equivalent to G_0; G_1 has four more algebraically equivalent



(a) DAG G_0 (b) MArG G_1 (c) Arid graph G_2

Figure 3: Two counterexamples to the one-step case of the generalized Meek conjecture.

ADMGs G_1'; and G_2 has none. But no algebraically equivalent graphs exist such that G_1' or G_2 can be obtained from G_0' by adding a single edge.

4.2. Coatomicity and Relevant Constraints

The statistical model corresponding to a DAG can be succinctly described as a Markov model: by specifying a set of (conditional) independence constraints that hold for all distributions in the model, we can precisely characterize which distributions are and are not in the model. When moving to more general ADMGs, conditional independences no longer suffice to make such a characterization. As we saw in Section 2.2, other algebraic constraints are needed. But it is unclear what set of polynomials needs to be considered as potential constraints in order to characterize the contents of $\overline{\mathcal{M}}(G)$ for the various classes of graphs.

In the case of DAGs, we observe that there is a one-to-one correspondence between the set of conditional independence constraints ($\{a \perp\!\!\!\perp b, a \perp\!\!\!\perp b \mid c, \ldots, a \perp\!\!\!\perp c, \ldots\}$) over n variables and the coatoms of the DAG poset for that n. (Recall from Section 2.3 that a coatom is a model in the poset that has only the saturated element above it.) We may hope that also for larger classes of graphs, a one-to-one correspondence exists between the coatoms and the relevant equality constraints.

For general algebraic models M, if M is one rank below the saturated model, it imposes exactly one equality constraint (it is defined by a principal ideal; see (Cox et al., 2015)). To be able to distinguish M from the saturated model, this constraint is *necessary*. This is in contrast to models that impose more than one constraint: as multiples of one polynomial constraint can be added to another without changing the model, we cannot conclude for any of these constraints that they are necessary.

The question that remains is: are these constraints also *sufficient* to characterize all algebraic models of a particular class? One way to formalize this question, in a way that can be answered based on posets of models, is by coatomicity.

A poset is *coatomistic* (Grätzer, 2003) if $M \neq N$ implies $\bar{A}(M) \neq \bar{A}(N)$. Here, $\bar{A}(M)$ denotes the set of coatoms N for which $M \subseteq N$. Similarly, $A(M)$ denotes the set of atoms N with $N \subseteq M$. Coatomicity of the poset of models would imply that any pair of models can be distinguished based on the coatom constraints.

We can hope for a stronger condition, for which we need a new definition.

Definition 1 *A poset is* strongly atomistic *if for all elements M, N of the poset, $M \subseteq N \iff A(M) \subseteq A(N)$. Dually, a poset is* strongly coatomistic *if for all M, N, $M \subseteq N \iff \bar{A}(M) \supseteq \bar{A}(N)$.*

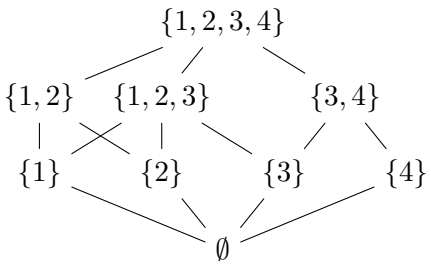


Figure 4: A (finite, bounded, graded) poset that is atomistic but not strongly atomistic: the elements $\{1, 2\}$ and $\{1, 2, 3\}$ are not comparable but their sets of atoms are. Each element is marked by $A(\cdot)$, the set of atoms below it.

This implies that the entire partial order relation can be reconstructed if we know each element's coatoms $\bar{A}(\cdot)$. As the names suggest, strong (co)atomicity is strictly stronger than (co)atomicity. This is demonstrated by the example in Figure 4.

Note that strong coatomicity still would not guarantee that the algebraic model is equal to the set of all distributions that satisfy these constraints. But it is a necessary condition.

Results We checked coatomicity and strong coatomicity empirically for each of the graph classes in Section 2.1 for $n = 4, 5$, and found no counterexamples.

5. Discussion and Conclusion

We presented an algorithm for determining whether the algebraic model of one ADMG includes another, extending the algorithms of Van Ommen (2024). Using this algorithm, we found that when considering MArGs or larger classes of ADMGs, the poset is not graded and Meek's conjecture does not hold. This suggests that a GES-like algorithm for this setting would need to take larger steps through the search space, sometimes adding, removing, or otherwise changing multiple edges at a time.

We also tested whether the posets satisfy (strong) coatomicity, two properties that suggest that all relevant constraints needed for describing classes of models can be found among the coatoms of the posets. We found no counterexamples against these properties.

Our results pertain to linear Gaussian models. Under different assumptions, different posets may be found, with possibly different properties. For instance, with non-Gaussian distributions, many new models become distinguishable (Wang and Drton, 2023). On the other hand, the posets for discrete and nonparametric models may very well be the same as the poset we found for MArGs (Van Ommen, 2024).

Logical next steps for future work include extending Algorithm 1 even further, specifically to also deal with cyclic graphs. The empirical results from our paper suggest that the posets of ancestral and maximal ancestral graphs are graded and satisfy Meek's conjecture, and that posets for all classes of graphs are (strongly) coatomistic. Mathematical proofs are needed to establish whether these properties truly hold. The empirical evidence for (strong) coatomicity does also suggest that the constraints imposed by coatom models play a central role among equality constraints and may be worth further investigation.

References

Marina Maciel Ansanelli, Elie Wolfe, and Robert W. Spekkens. The observational partial order of causal structures with latent variables. *Journal of Causal Inference*, 14(1), August 2026.

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3:507–554, 2002.

Tom Claassen and Ioan Gabriel Bucur. Greedy equivalence search in the presence of latent confounders. In James Cussens and Kun Zhang, editors, *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, volume 180 of *Proceedings of Machine Learning Research*, pages 443–452. PMLR, 2022.

David A. Cox, John Little, and Donal O'Shea. *Ideals, Varieties, and Algorithms*. Springer, New York, fourth edition, 2015.

B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, Cambridge, second edition, 2012.

Robin J. Evans. Margins of discrete Bayesian networks. *The Annals of Statistics*, 46(6A): 2623–2656, 2018.

Rina Foygel, Jan Draisma, and Mathias Drton. Half-trek criterion for generic identifiability of linear structural equation models. *The Annals of Statistics*, 40(3):1682–1713, 2012.

George Grätzer. *General Lattice Theory*. Birkhäuser Verlag, Basel, second edition, 2003.

Zhonyi Hu and Robin J. Evans. A fast score-based search algorithm for maximal ancestral graphs using entropy. *arXiv preprint arXiv:2402.04777v1*, 2024.

Christopher Meek. *Graphical Models: Selecting causal and statistical models*. Phd thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania, 1998.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2000.

Thomas S. Richardson and Peter Spirtes. Ancestral graph Markov models. *The Annals of Statistics*, 30(4):962–1030, 2002.

Thomas S. Richardson, Robin J. Evans, James M. Robins, and Ilya Shpitser. Nested Markov properties for acyclic directed mixed graphs. *The Annals of Statistics*, 51(1): 334–361, 2023.

J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the Association for Computing Machinery*, 27(4):701–717, 1980.

Ilya Shpitser, Robin J. Evans, Thomas S. Richardson, and James M. Robins. Introduction to nested Markov models. *Behaviormetrika*, 41(1):3–39, 2014.

Ilya Shpitser, Robin J. Evans, and Thomas S. Richardson. Acyclic linear SEMs obey the nested Markov property. In *Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI 2018)*, 2018.

Peter Spirtes, Christopher Meek, and Thomas Richardson. An algorithm for causal inference in the presence of latent variables and selection bias. In C. Glymour and G.F. Cooper, editors, *Computation, Causation and Discovery*, chapter 6, pages 211–252. MIT Press, Cambridge, MA, 1999.

Seth Sullivant, Kelli Talaska, and Jan Draisma. Trek separation for Gaussian graphical models. *The Annals of Statistics*, 38(3):1665–1685, 2010.

Armeen Taeb, F. Richard Guo, and Leonard Henckel. Model-oriented graph distances via partially ordered sets. *arXiv preprint arXiv:2511.10625v3*, 2026.

Thijs van Ommen. Efficiently deciding algebraic equivalence of bow-free acyclic path diagrams. In *Proceedings of the 40th Annual Conference on Uncertainty in Artificial Intelligence (UAI 2024)*, 2024.

Thijs van Ommen and Joris M. Mooij. Algebraic equivalence of linear structural equation models. In *Proceedings of the 33rd Annual Conference on Uncertainty in Artificial Intelligence (UAI 2017)*, 2017.

Y. Samuel Wang and Mathias Drton. Causal discovery with unobserved confounding and non-Gaussian data. *Journal of Machine Learning Research*, 24:1–61, 2023.

Luca Weihs, Bill Robinson, Emilie Dufresne, Jennifer Kenkel, Kaie Kubjas, McGee Reginald, Nhan Nguyen, Elina Robeva, and Mathias Drton. Determinantal generalizations of instrumental variables. *Journal of Causal Inference*, 6, December 2017. doi: 10.1515/jci-2017-0009.

Jiji Zhang. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16–17):1873–1896, 2008.