Learning Acyclic Directed Mixed Graphs from Observations and Interventions

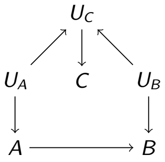
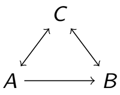
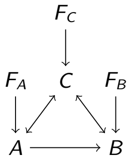
Jose M. Peña

Department of Computer and Information Science
Linköping University, Sweden

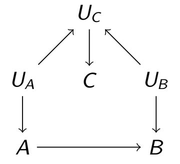
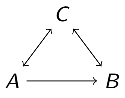
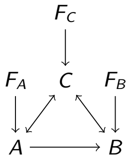
LINKÖPING UNIVERSITY

Causal effect identification in original ADMGs

Causal effect identification in original ADMGs

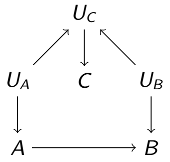
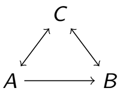
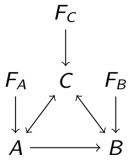
DAG D	Original ADMG G	Original ADMG G'
		

Causal effect identification in original ADMGs

DAG D	Original ADMG G	Original ADMG G'
		

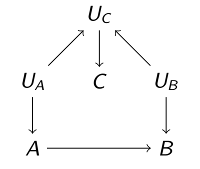
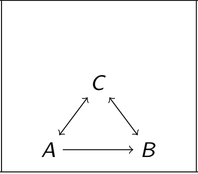
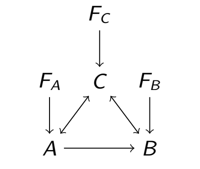
- $X \leftrightarrow Y$ in G iff $U_X \not\perp_D U_Y | \emptyset$.

Causal effect identification in original ADMGs

DAG D	Original ADMG G	Original ADMG G'
		

- $X \leftrightarrow Y$ in G iff $U_X \not\perp_D U_Y | \emptyset$.
- Effect identification in G: $p(B|do(A)) = p(B|A)$.

Causal effect identification in original ADMGs

DAG D	Original ADMG G	Original ADMG G'
		

- $X \leftrightarrow Y$ in G iff $U_X \not\perp_D U_Y | \emptyset$.
- Effect identification in G: $p(B|do(A)) = p(B|A)$.
- Identification is possible due to Pearl's *do*-calculus on G':
  - Rule 1 (insertion/deletion of observations):
    $p(Y|do(X), Z \cup W) = p(Y|do(X), W)$ if $Y \perp_{G'} Z | X \cup W || X$.
  - Rule 2 (intervention/observation exchange):
    $p(Y|do(X), do(Z), W) = p(Y|do(X), Z \cup W)$ if $Y \perp_{G'} F_Z | X \cup W \cup Z || X$.
  - Rule 3 (insertion/deletion of interventions):
    $p(Y|do(X), do(Z), W) = p(Y|do(X), W)$ if $Y \perp_{G'} F_Z | X \cup W || X$.

  Where $\cdot \perp_{G'} \cdot | \cdot || X$ denotes separation in G' after intervention on X.

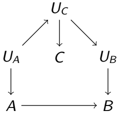
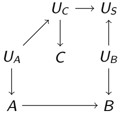
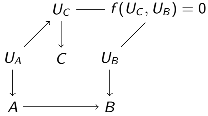
Causal effect identification in **alternative** ADMGs

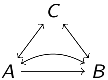
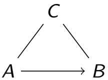
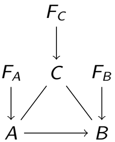
Causal effect identification in **alternative** ADMGs

- Assume **continuous** random variables and **additive** noise.

Causal effect identification in **alternative** ADMGs

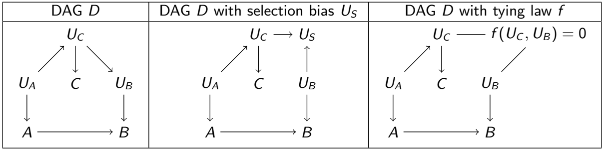
- Assume **continuous** random variables and **additive** noise.

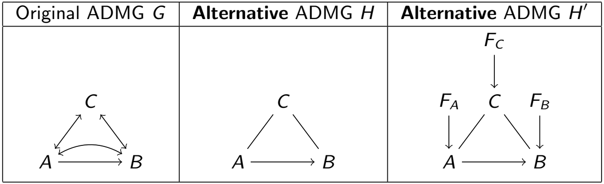
DAG D	DAG D with selection bias U_S	DAG D with tying law f
		

Original ADMG G	**Alternative** ADMG H	**Alternative** ADMG H'
		

Causal effect identification in **alternative** ADMGs

- Assume **continuous** random variables and **additive** noise.



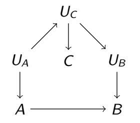


- $X - Y$ in H iff $U_X \not\perp_D U_Y | U \setminus \{U_A, U_B\}$.

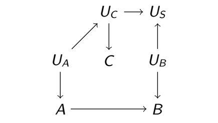
Causal effect identification in **alternative** ADMGs

- Assume **continuous** random variables and **additive** noise.

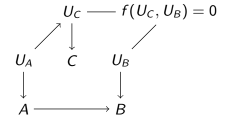
DAG D



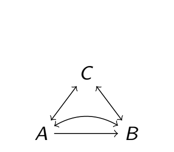
DAG D with selection bias U_S



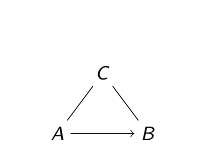
DAG D with tying law f



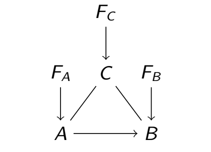
Original ADMG G



Alternative ADMG H



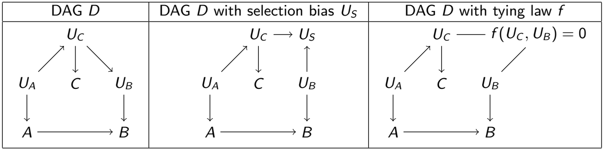
Alternative ADMG H'

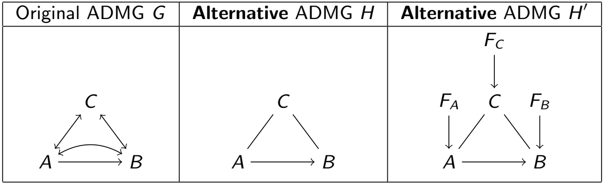


- $X - Y$ in H iff $U_X \not\perp_D U_Y | U \setminus \{U_A, U_B\}$.
- Effect identification in G: Impossible.

Causal effect identification in **alternative** ADMGs

- Assume **continuous** random variables and **additive** noise.

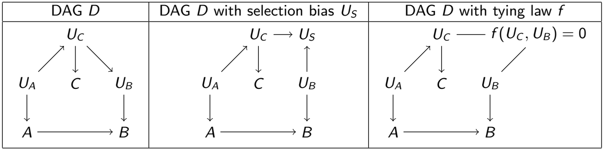


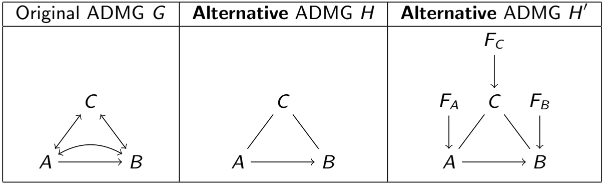


- $X - Y$ in H iff $U_X \not\perp_D U_Y | U \setminus \{U_A, U_B\}$.
- Effect identification in G: Impossible.
- Effect identification in H: $p(B|do(A)) = \int p(B|A, C)p(C)\, dC$.

Causal effect identification in **alternative** ADMGs

- Assume **continuous** random variables and **additive** noise.

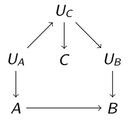
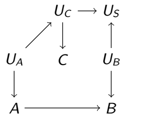
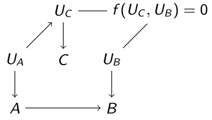


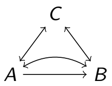
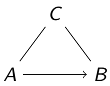
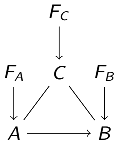


- $X - Y$ in H iff $U_X \not\perp_D U_Y | U \setminus \{U_A, U_B\}$.
- Effect identification in G: Impossible.
- Effect identification in H: $p(B|do(A)) = \int p(B|A, C)p(C)\, dC$.
- Identification is possible due to *do*-calculus on H', but we need to define

Causal effect identification in **alternative** ADMGs

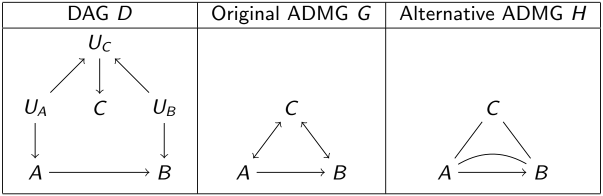
- Assume **continuous** random variables and **additive** noise.

DAG D	DAG D with selection bias U_S	DAG D with tying law f
		

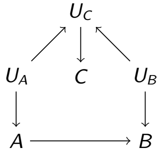
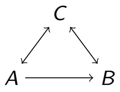
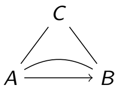
Original ADMG G	**Alternative** ADMG H	**Alternative** ADMG H'
		

- $X - Y$ in H iff $U_X \not\perp_D U_Y | U \setminus \{U_A, U_B\}$.
- Effect identification in G: Impossible.
- Effect identification in H: $p(B|do(A)) = \int p(B|A,C)p(C)\, dC$.
- Identification is possible due to *do*-calculus on H', but we need to define
  - interventions in alternative ADMGs, and
  - a separation criterion for alternative ADMGs.

Causal effect identification in alternative ADMGs

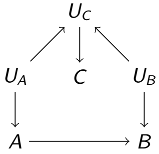
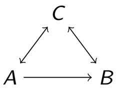
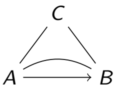


Causal effect identification in alternative ADMGs

DAG D	Original ADMG G	Alternative ADMG H
		

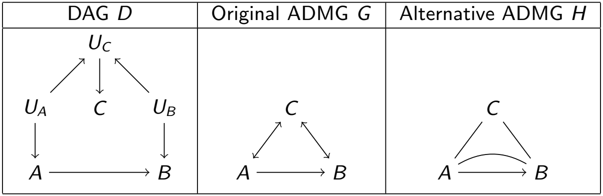
- Effect identification in G: $p(B|do(A)) = p(B|A)$.

Causal effect identification in alternative ADMGs

DAG D	Original ADMG G	Alternative ADMG H
		

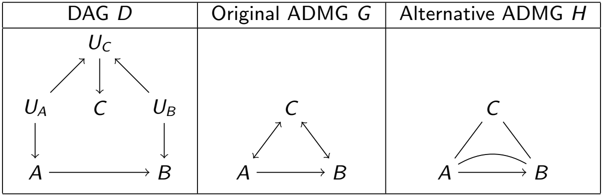
- Effect identification in G: $p(B|do(A)) = p(B|A)$.
- Effect identification in H: Impossible.

Causal effect identification in alternative ADMGs



- Effect identification in G: $p(B|do(A)) = p(B|A)$.
- Effect identification in H: Impossible.
- Thus, original and alternative ADMGs are complementary and we propose unifying them into what we simply call ADMGs.

Causal effect identification in alternative ADMGs



- Effect identification in G: $p(B|do(A)) = p(B|A)$.
- Effect identification in H: Impossible.
- Thus, original and alternative ADMGs are complementary and we propose unifying them into what we simply call ADMGs.
- To do so, we define
  - interventions in ADMGs, and
  - a separation criterion for ADMGs.

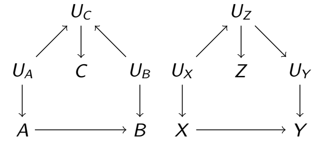
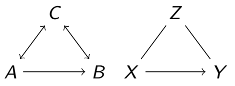
Causal effect identification in ADMGs

Causal effect identification in ADMGs

- Assume continuous random variables and additive noise.

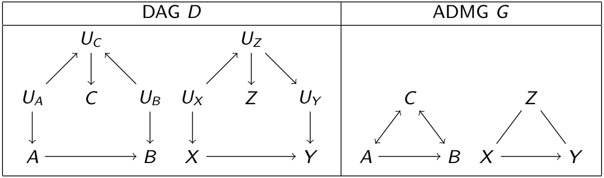
Causal effect identification in ADMGs

- Assume continuous random variables and additive noise.

DAG D	ADMG G
	

Causal effect identification in ADMGs

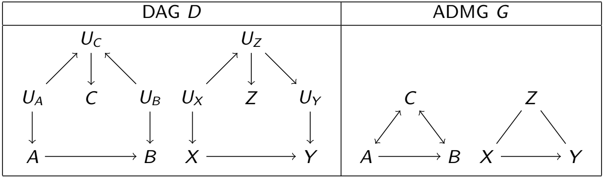
- Assume continuous random variables and additive noise.



- Effect identification in G:
  $p(B, Y|do(A, X)) = p(B|A) \int p(Y|X, Z)p(Z)\,dZ$.

Causal effect identification in ADMGs

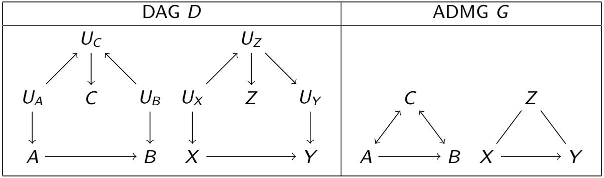
- Assume continuous random variables and additive noise.



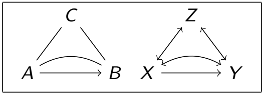
- Effect identification in G:
  $p(B, Y | do(A, X)) = p(B|A) \int p(Y|X, Z)p(Z)\, dZ$.
- Effect identification in original and alternative ADGMs: Impossible.

Causal effect identification in ADMGs

- Assume continuous random variables and additive noise.

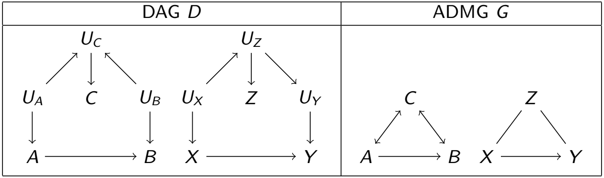


- Effect identification in G:
$p(B, Y|do(A, X)) = p(B|A) \int p(Y|X, Z)p(Z)\, dZ$.
- Effect identification in original and alternative ADGMs: Impossible.
- Note that effect identification is also impossible in the following Markov equivalent ADMG:

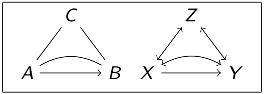


Causal effect identification in ADMGs

- Assume continuous random variables and additive noise.



- Effect identification in G:
  $p(B, Y|do(A, X)) = p(B|A) \int p(Y|X, Z)p(Z)\, dZ$.
- Effect identification in original and alternative ADGMs: Impossible.
- Note that effect identification is also impossible in the following Markov equivalent ADMG:



- This problem also exists for original and alternative ADMGs.

Final Remarks

Final Remarks

- We introduce an assumption-free exhaustive algorithm for learning ADMGs from observations and interventions via answer set programming.

Final Remarks

- We introduce an assumption-free exhaustive algorithm for learning ADMGs from observations and interventions via answer set programming.
- The algorithm can easily be modified to learn original and alternative ADMGs, AMP and MVR CGs, DAGs, UGs and BGs.

Final Remarks

- We introduce an assumption-free exhaustive algorithm for learning ADMGs from observations and interventions via answer set programming.
- The algorithm can easily be modified to learn original and alternative ADMGs, AMP and MVR CGs, DAGs, UGs and BGs.
- Finally, it is worth mentioning that
  - ADMGs can also be seen as (non-causal) statistical models,

Final Remarks

- We introduce an assumption-free exhaustive algorithm for learning ADMGs from observations and interventions via answer set programming.
- The algorithm can easily be modified to learn original and alternative ADMGs, AMP and MVR CGs, DAGs, UGs and BGs.
- Finally, it is worth mentioning that
  - ADMGs can also be seen as (non-causal) statistical models,
  - no existing family of mixed graphical models subsumes ADMGs, and

Final Remarks

- We introduce an assumption-free exhaustive algorithm for learning ADMGs from observations and interventions via answer set programming.
- The algorithm can easily be modified to learn original and alternative ADMGs, AMP and MVR CGs, DAGs, UGs and BGs.
- Finally, it is worth mentioning that
  - ADMGs can also be seen as (non-causal) statistical models,
  - no existing family of mixed graphical models subsumes ADMGs, and
  - the definitions of intervention and separation in ADMGs generalize those for original and alternative ADMGs.

Final Remarks

- We introduce an assumption-free exhaustive algorithm for learning ADMGs from observations and interventions via answer set programming.
- The algorithm can easily be modified to learn original and alternative ADMGs, AMP and MVR CGs, DAGs, UGs and BGs.
- Finally, it is worth mentioning that
  - ADMGs can also be seen as (non-causal) statistical models,
  - no existing family of mixed graphical models subsumes ADMGs, and
  - the definitions of intervention and separation in ADMGs generalize those for original and alternative ADMGs.
- Future work includes
  - deriving local and pairwise Markov properties for ADMGs,
  - characterizing Markov equivalence of ADMGs, and
  - deriving factorization induced by ADMGs.

Final Remarks

- We introduce an assumption-free exhaustive algorithm for learning ADMGs from observations and interventions via answer set programming.
- The algorithm can easily be modified to learn original and alternative ADMGs, AMP and MVR CGs, DAGs, UGs and BGs.
- Finally, it is worth mentioning that
  - ADMGs can also be seen as (non-causal) statistical models,
  - no existing family of mixed graphical models subsumes ADMGs, and
  - the definitions of intervention and separation in ADMGs generalize those for original and alternative ADMGs.
- Future work includes
  - deriving local and pairwise Markov properties for ADMGs,
  - characterizing Markov equivalence of ADMGs, and
  - deriving factorization induced by ADMGs.

Thanks for your attention.