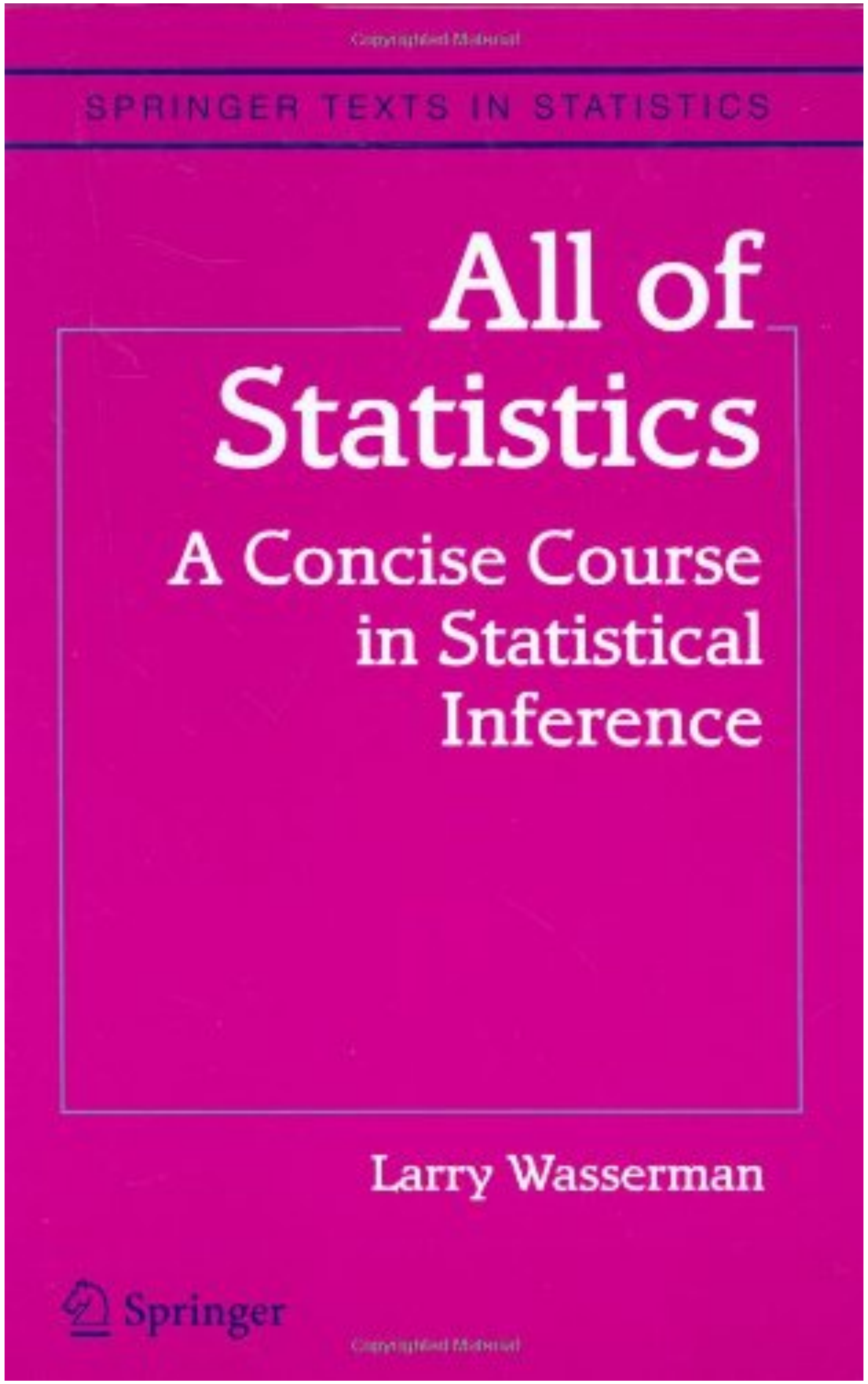
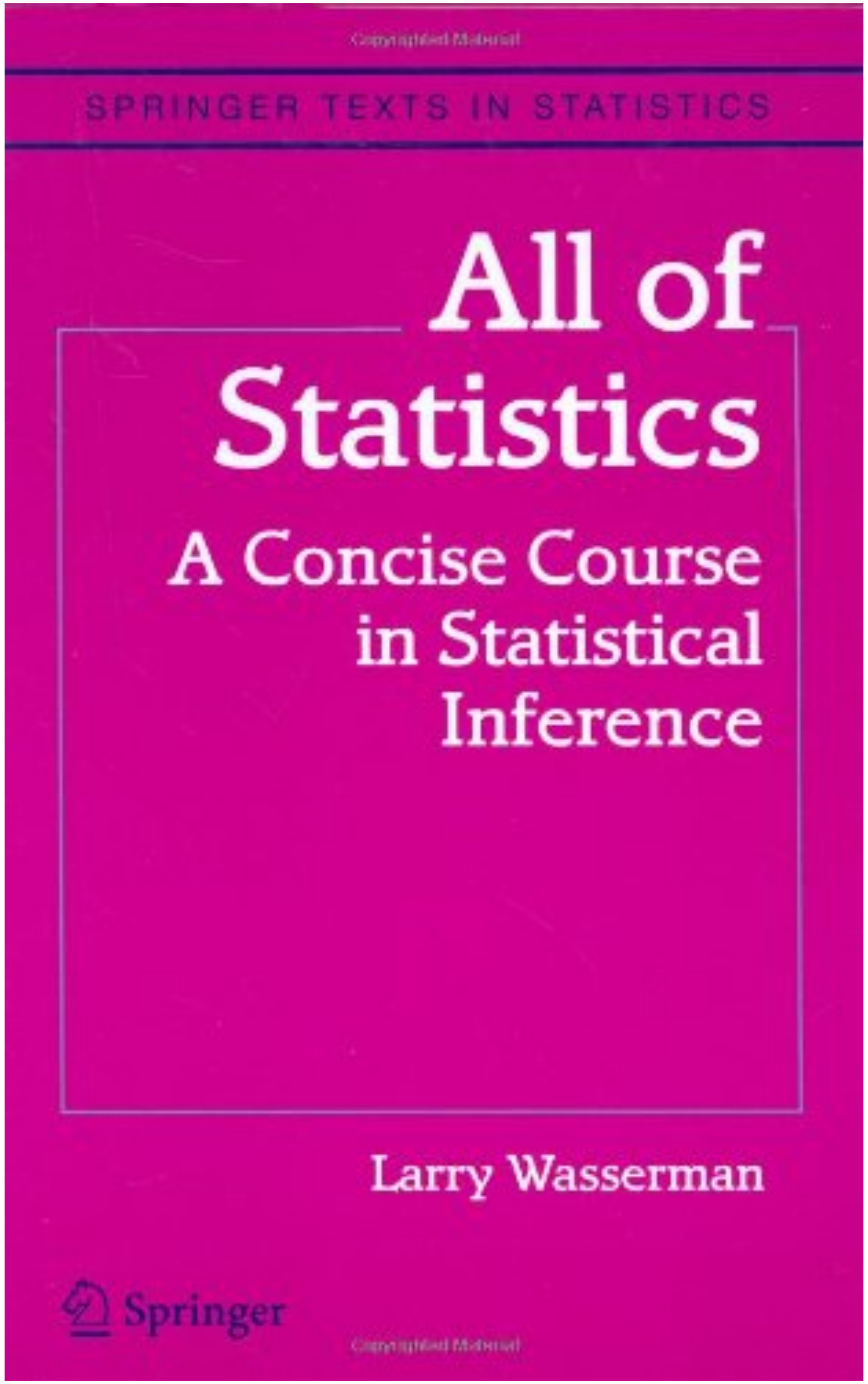
All of© Causal Discovery

Frederick Eberhardt
fde@caltech.edu

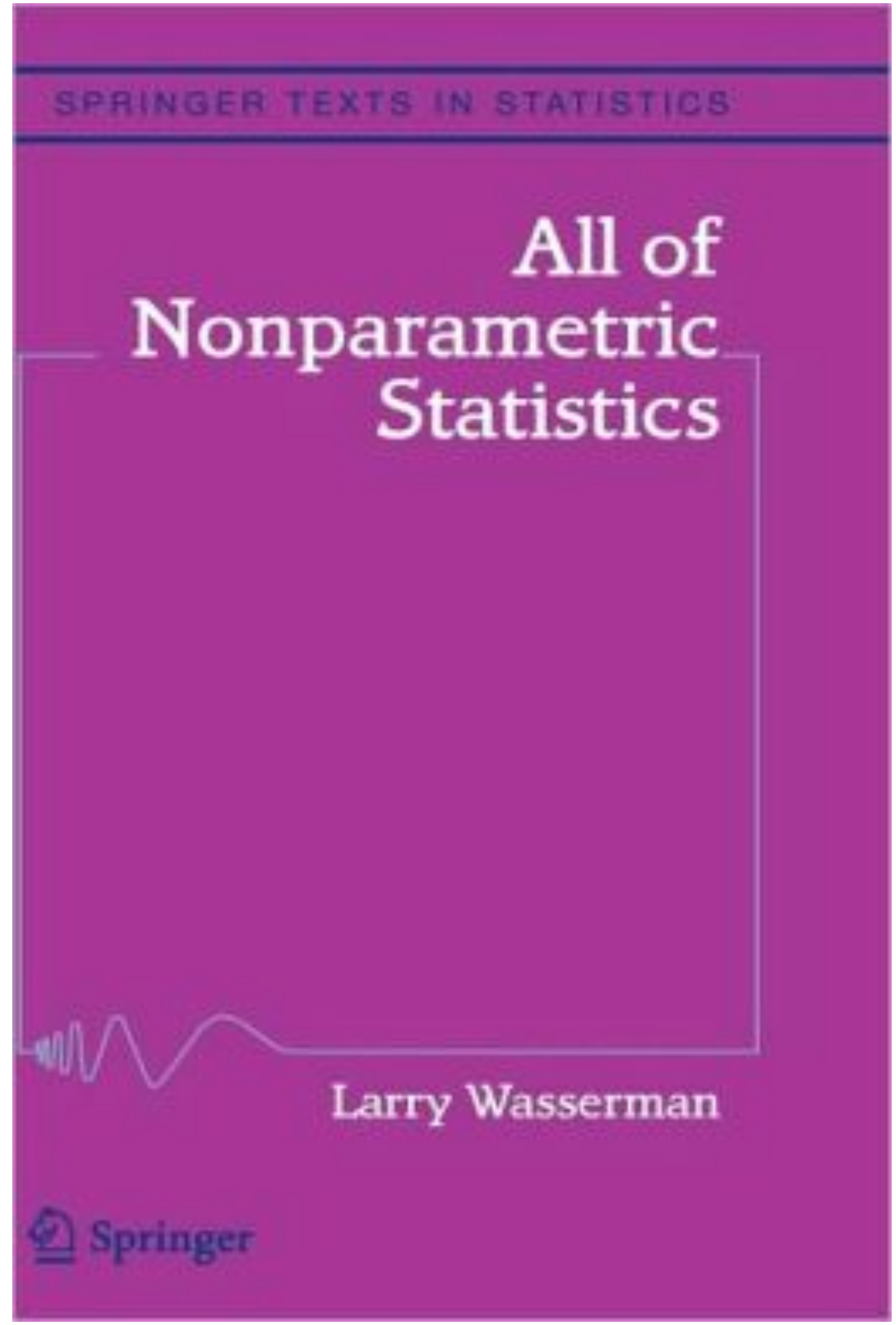
(only tiny bits of this are my work)


SPRINGER TEXTS IN STATISTICS
All of Statistics
A Concise Course in Statistical Inference
Larry Wasserman
Springer

442 pages



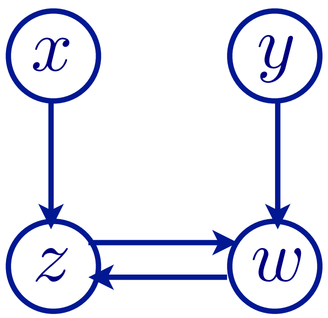
442 pages



268 pages

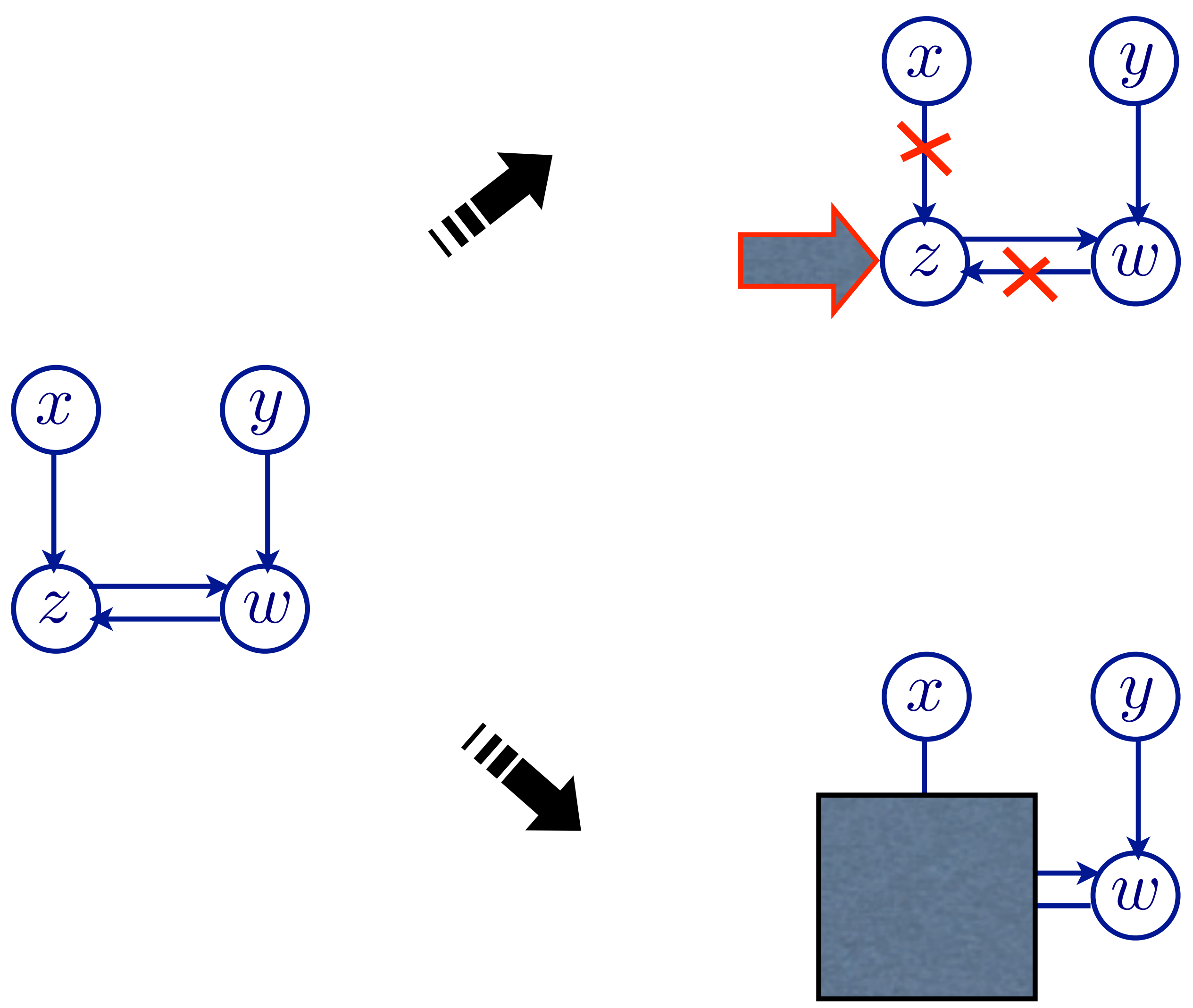
Causal Structure Search

true
(unknown)
model



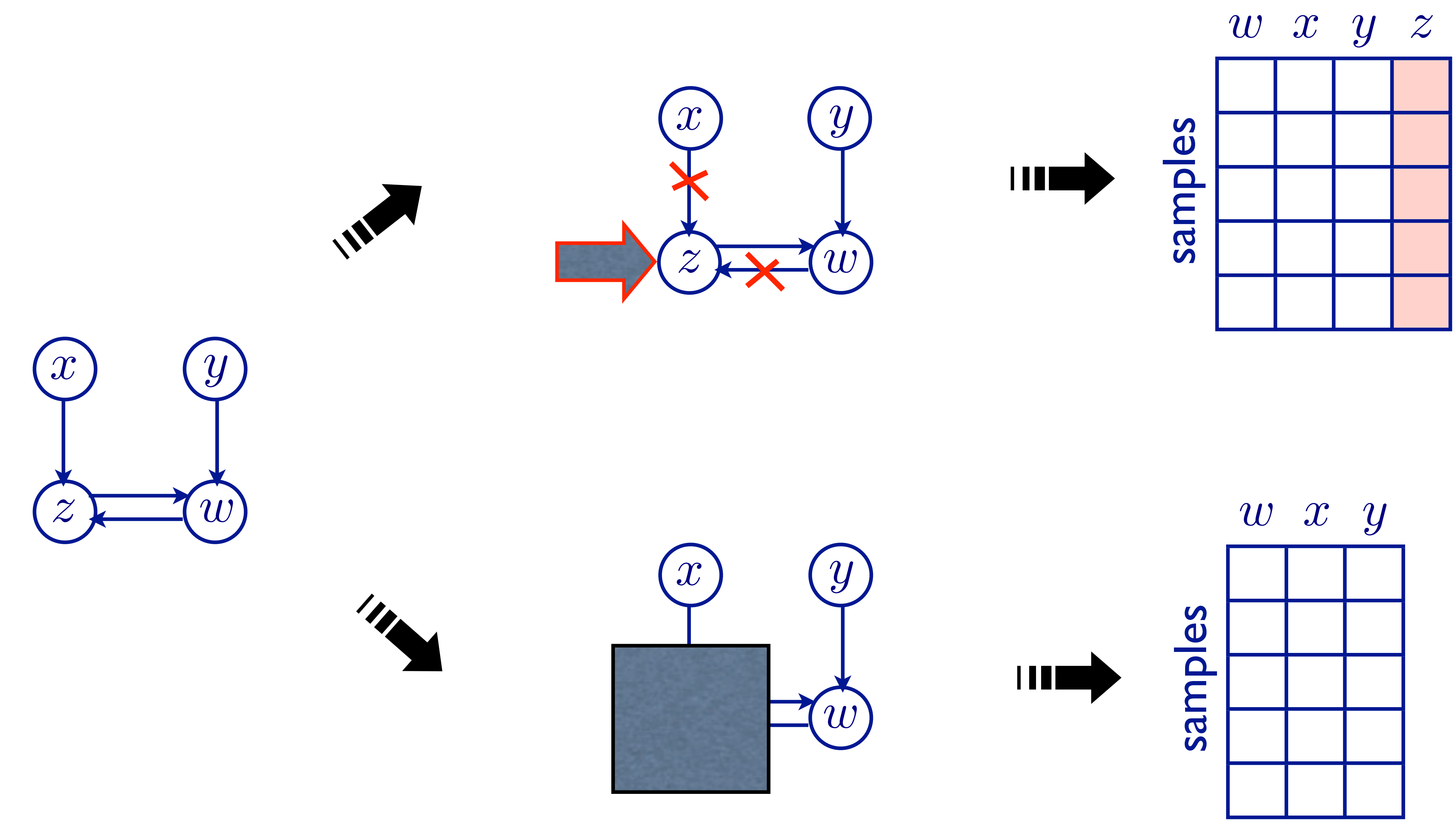
Causal Structure Search

true (unknown) model ⇨ experimental / observational conditions



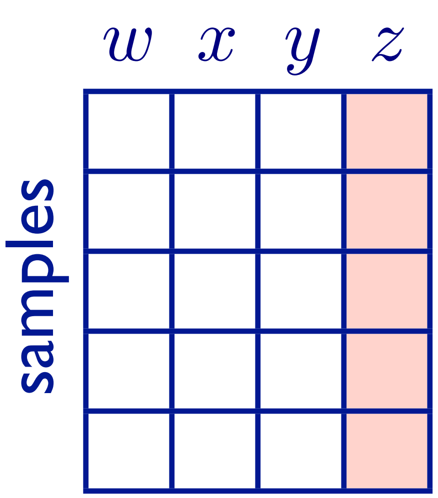
Causal Structure Search

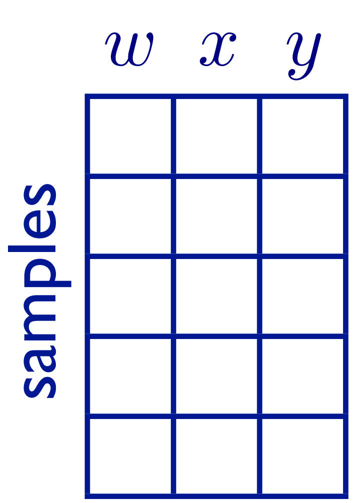
true (unknown) model → experimental / observational conditions → data sample



Search

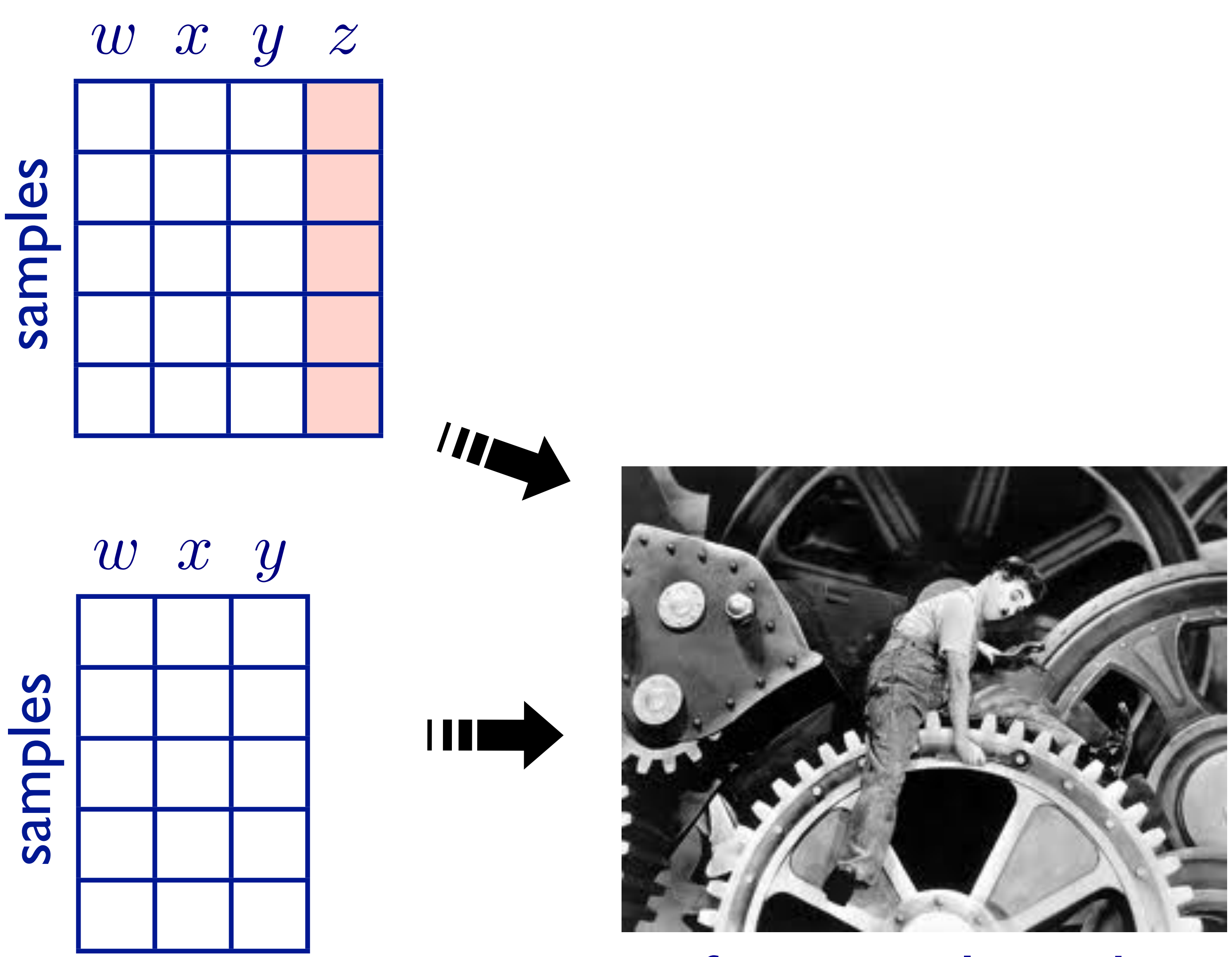
data
sample





Search

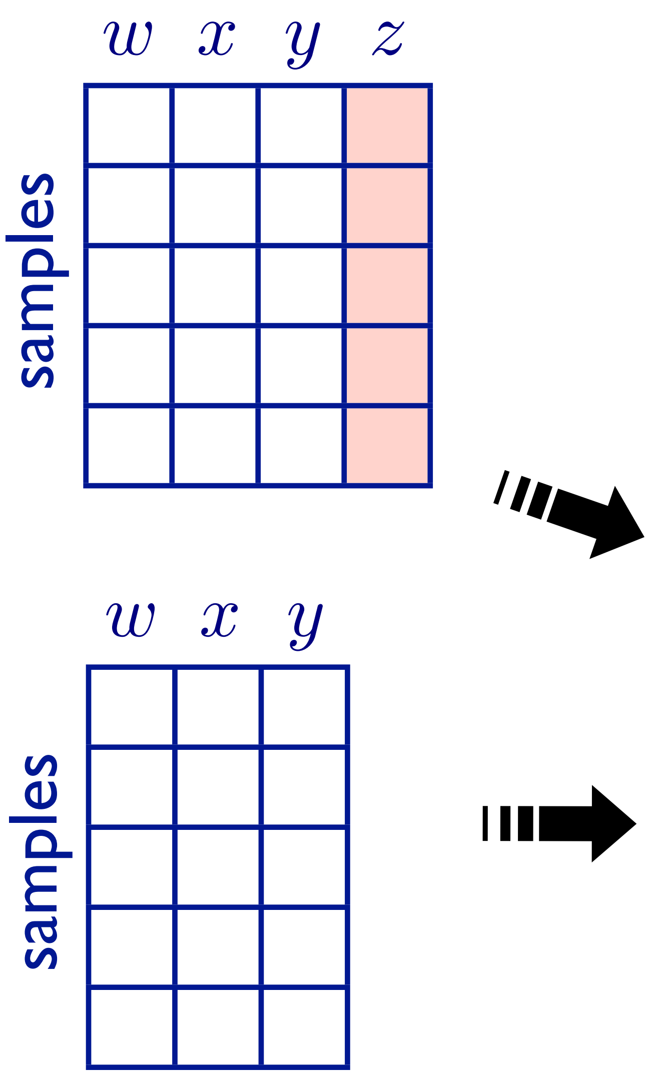
data
sample



inference algorithm

Search

data sample

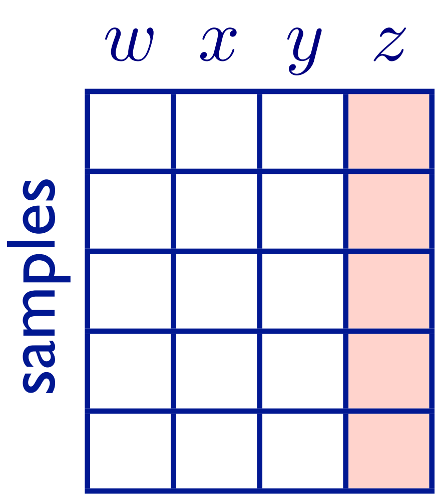


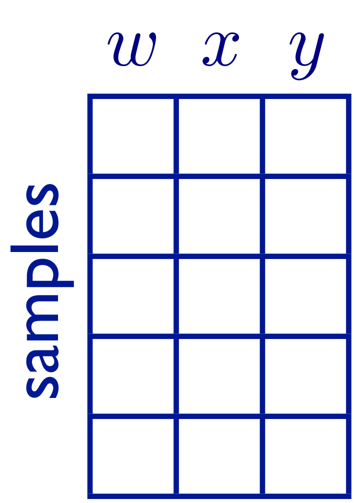
Assumptions, e.g.
- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity
- linearity
- distribution family
- etc.

inference algorithm

Search

data sample



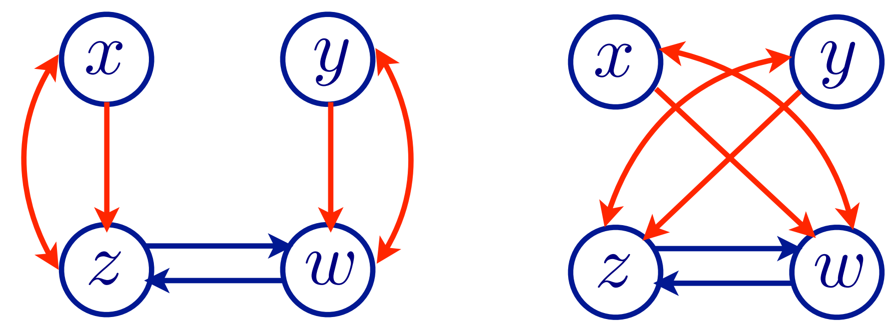


Assumptions, e.g.

- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity
- linearity
- distribution family
- etc.

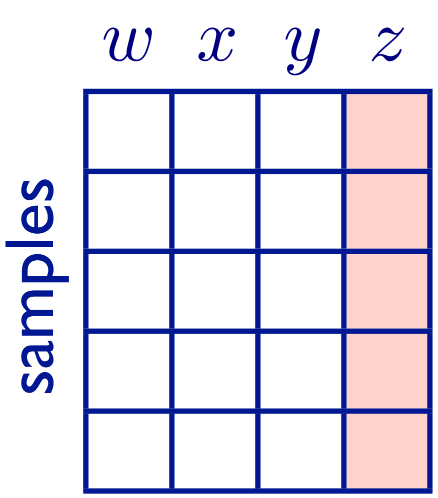
inference algorithm

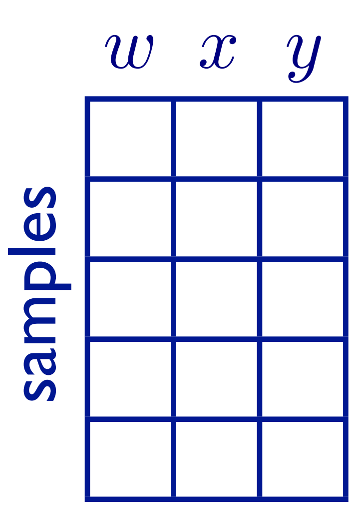
equivalence classes



Search

data sample

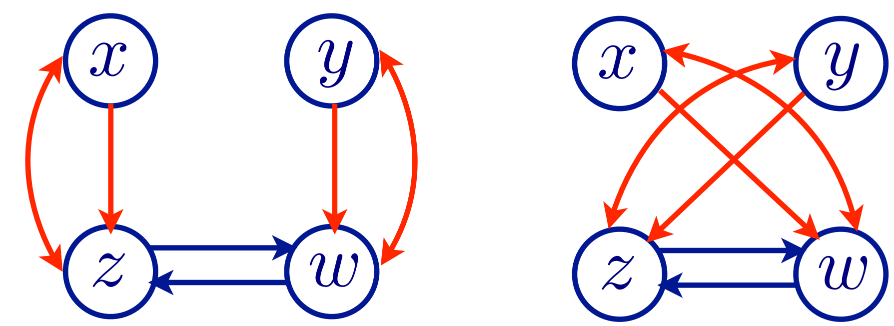




Assumptions, e.g.
- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity
- linearity
- distribution family
- etc.

inference algorithm

equivalence classes



model specifications

	w	x	y	z
w	0	0	?	a
x	0	0	0	0
y	0	0	0	0
z	b	?	?	0

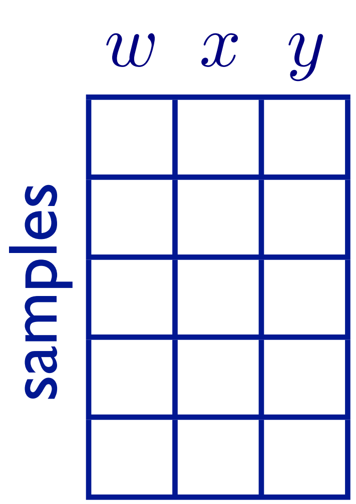
direct edges

	w	x	y	z
w	0	0	?	?
x	0	0	0	?
y	?	0	0	0
z	?	?	0	0

confounders

Example

data
sample

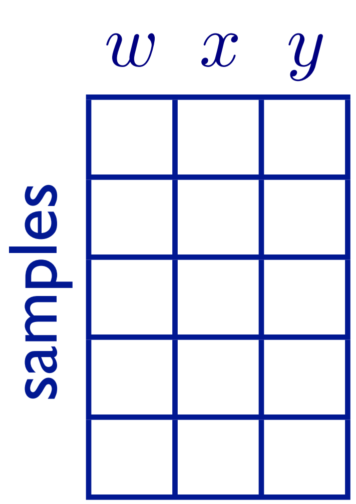


PC-algorithm

Example

Assumptions
- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity

data
sample

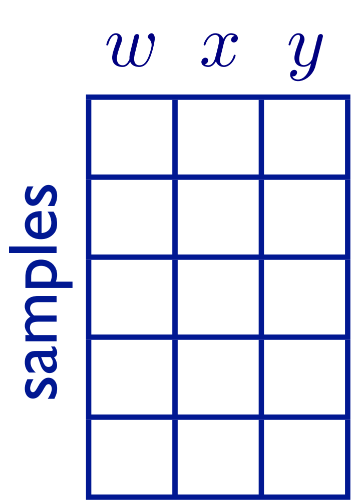


PC-algorithm

Example

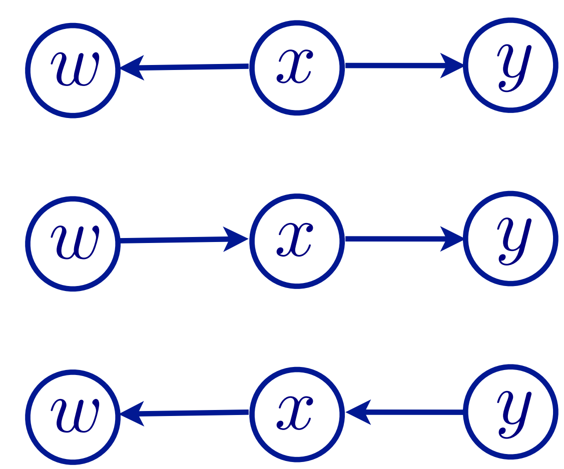
Assumptions
- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity

data
sample



PC-algorithm

Markov equivalence
class

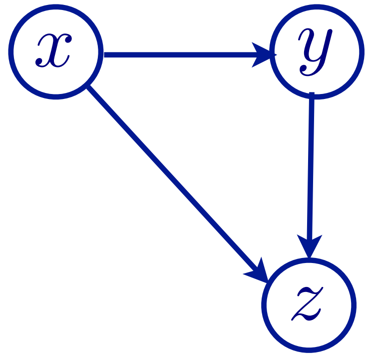


Assumptions

- **causal Markov**: permits inference from probabilistic dependence to causal connection

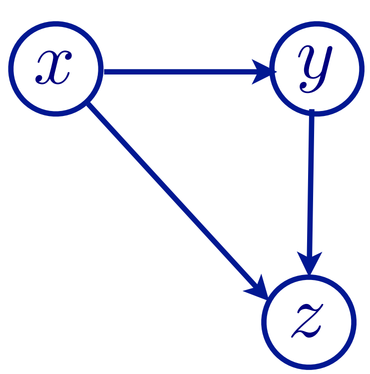
Assumptions

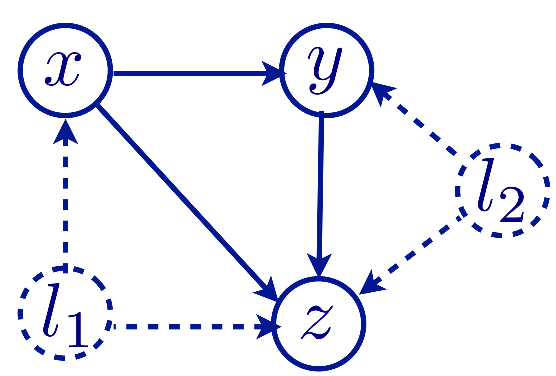
- **causal Markov**: permits inference from probabilistic dependence to causal connection
- **causal faithfulness**: permits inference from probabilistic independence to causal separation



Assumptions

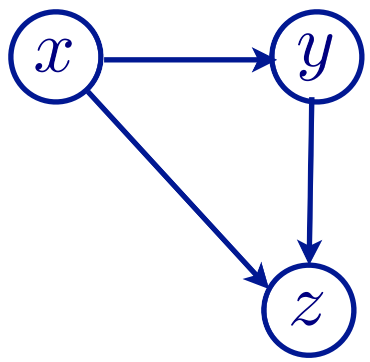
- **causal Markov**: permits inference from probabilistic dependence to causal connection
- **causal faithfulness**: permits inference from probabilistic independence to causal separation
- **causal sufficiency**: there are no unmeasured common causes

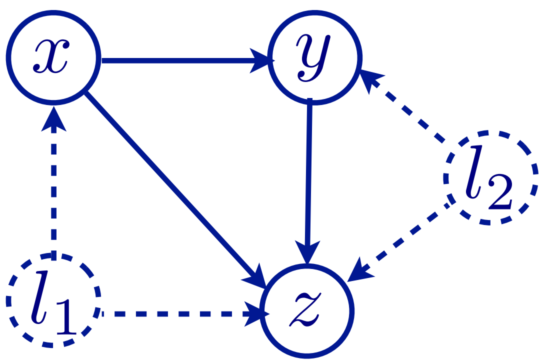


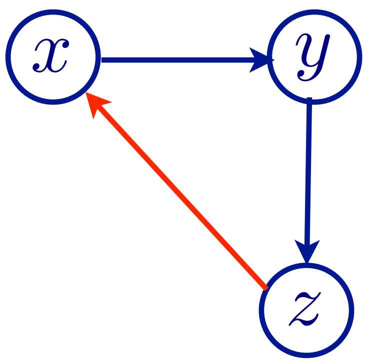


Assumptions

- **causal Markov**: permits inference from probabilistic dependence to causal connection
- **causal faithfulness**: permits inference from probabilistic independence to causal separation
- **causal sufficiency**: there are no unmeasured common causes
- **acylicity**: no variable is an (indirect) cause of itself



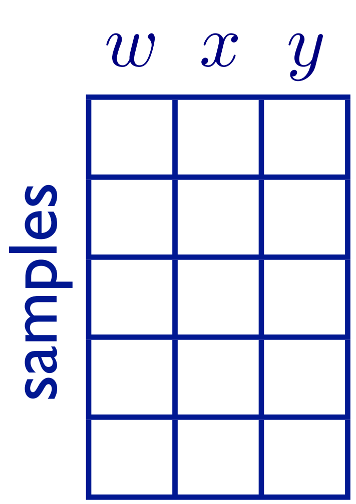




Search for the Markov equivalence class

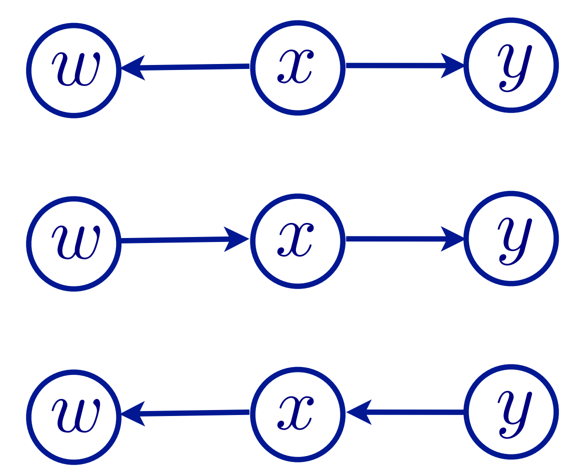
Assumptions
- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity

data sample



- independence based algorithms (e.g. PC and its variants)

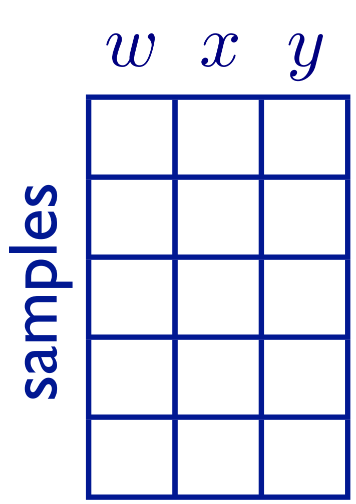
Markov equivalence class



Search for the Markov equivalence class

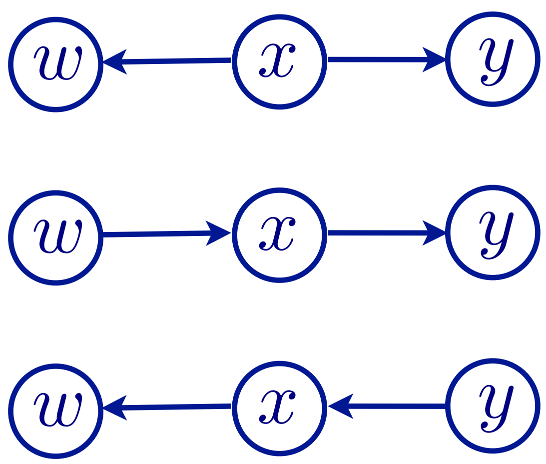
Assumptions
- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity

data sample



- independence based algorithms (e.g. PC and its variants)
- greedy Bayesian algorithms (e.g. GES)

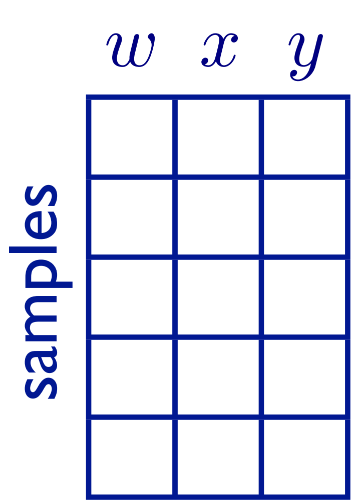
Markov equivalence class



Search for the Markov equivalence class

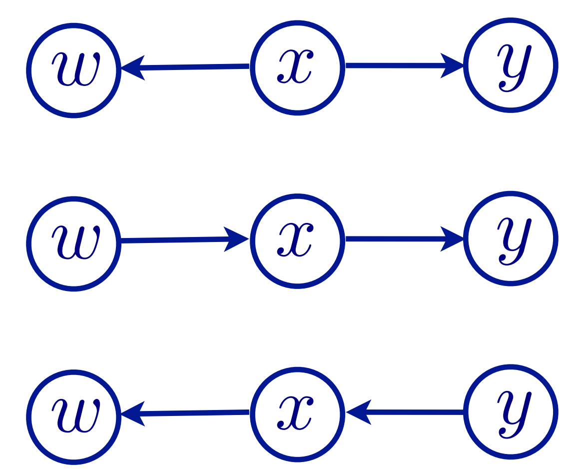
Assumptions
- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity

data
sample



- independence based algorithms (e.g. PC and its variants)
- greedy Bayesian algorithms (e.g. GES)
- exact Bayesian algorithms

Markov equivalence
class

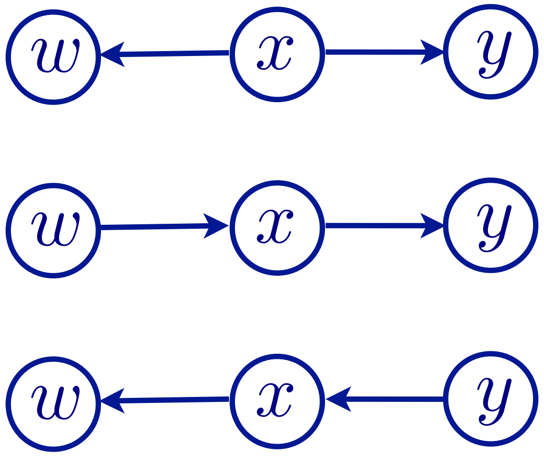


Aim of the game

Assumptions

- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity

Markov equivalence class



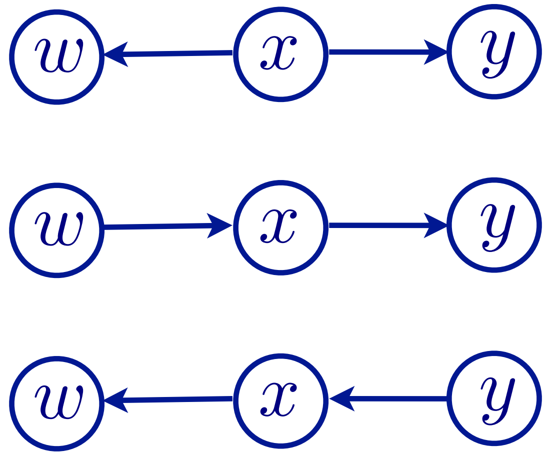
Aim of the game

reduce

Assumptions

- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity

Markov equivalence class



Aim of the game

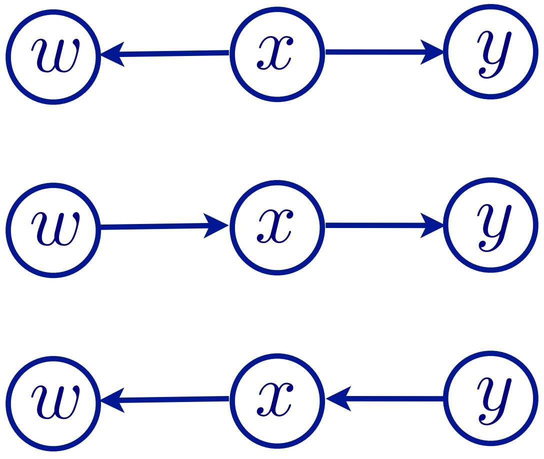
weaken

reduce

Assumptions

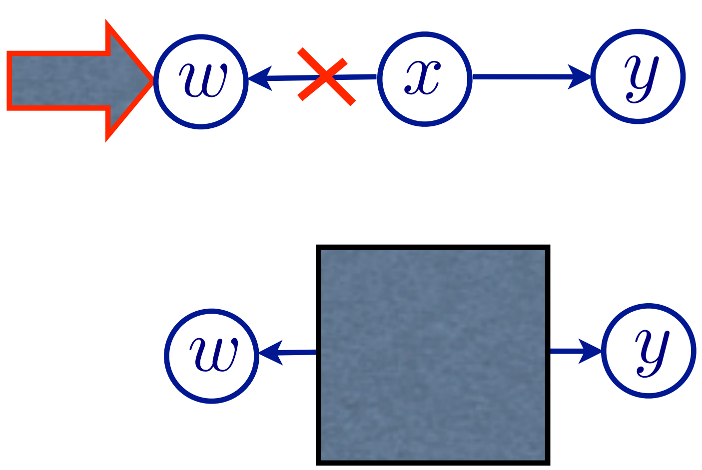
- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity

Markov equivalence class



Aim of the game

experimental /
observational conditions



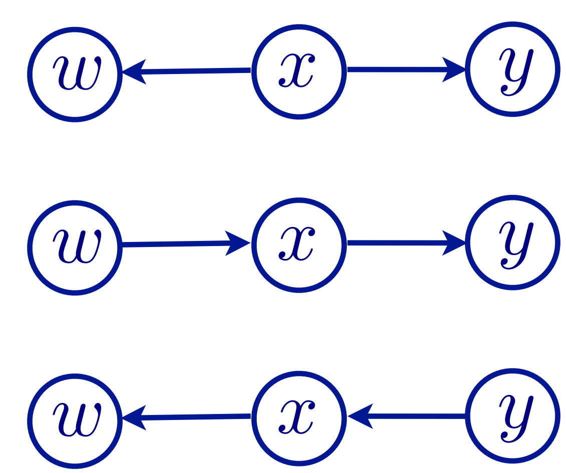
weaken

Assumptions

- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity

reduce

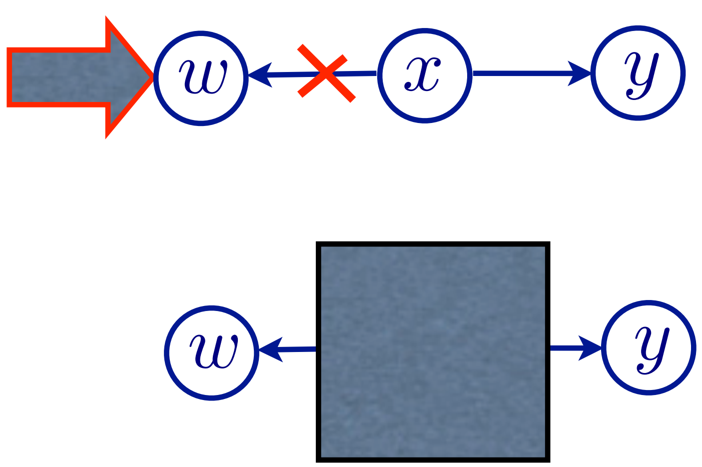
Markov equivalence class



Aim of the game

generalize

experimental /
observational conditions



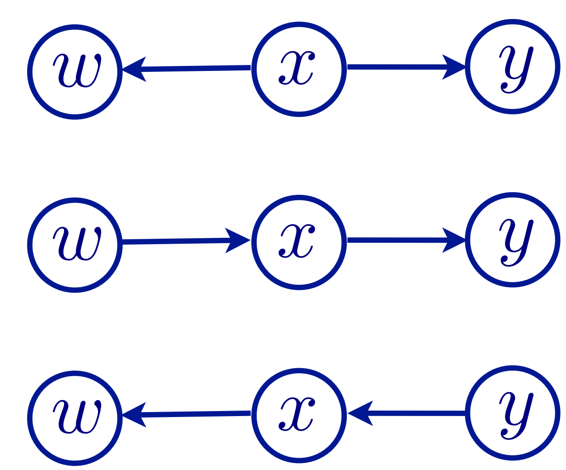
weaken

Assumptions

- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity

reduce

Markov equivalence
class

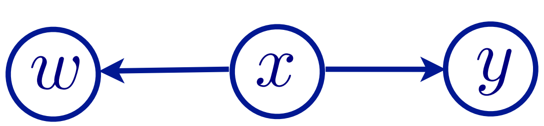


Limitations

ger **e**

*For **linear Gaussian** and for **multinomial** causal relations, an algorithm that identifies the Markov equivalence class of the true model is complete.*
(Pearl & Geiger 1988, Meek 1995)

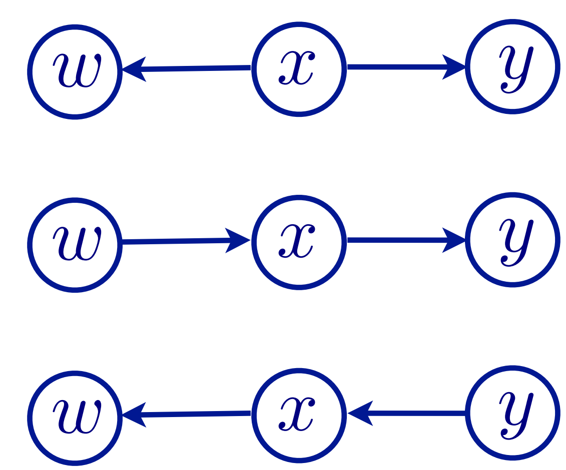
observational condition



Assumptions

- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity

Markov equivalence class



Moving forward

1. **Weaken the assumptions** (and increase the equivalence class)
   a. allow for unmeasured common causes
   b. allow for cycles
   c. all of the above

Moving forward

1. **Weaken the assumptions** (and increase the equivalence class)
    a. allow for unmeasured common causes
    b. allow for cycles
    c. all of the above
2. **Exclude the limitations** (and reduce the equivalence class)
    a. restrict to non-Gaussian error distributions
    b. restrict to non-linear causal relations

Moving forward

1. **Weaken the assumptions** (and increase the equivalence class)
   a. allow for unmeasured common causes
   b. allow for cycles
   c. all of the above
2. **Exclude the limitations** (and reduce the equivalence class)
   a. restrict to non-Gaussian error distributions
   b. restrict to non-linear causal relations
3. **Include for more general data collection set-ups** (and see how assumptions can be adjusted and what equivalence class results)
   a. experimental evidence
   b. multiple (overlapping) data sets

Linear non-Gaussian method (LiNGaM) (Shimizu et al., 2006)

- Linear causal relations:

$$x_i = \sum_{x_j \in \mathbf{Pa}(x_i)} \beta_{ij} x_j + \epsilon_j$$

- Assumptions:
  - causal Markov
  - causal sufficiency
  - acyclicity

Linear non-Gaussian method (LiNGaM) (Shimizu et al., 2006)

- Linear causal relations:

$$x_i = \sum_{x_j \in \mathbf{Pa}(x_i)} \beta_{ij} x_j + \epsilon_j$$

- Assumptions:
  - causal Markov
  - causal sufficiency
  - acyclicity

▸ If $\epsilon_j \sim$ non-Gaussian, then the true graph is **uniquely identifiable** from the joint distribution.

Linear non-Gaussian method (LiNGaM) (Shimizu et al., 2006)

- Linear causal relations:

$$x_i = \sum_{x_j \in \mathbf{Pa}(x_i)} \beta_{ij} x_j + \epsilon_j$$

- Assumptions:
  - causal Markov
  - causal sufficiency
  - acyclicity

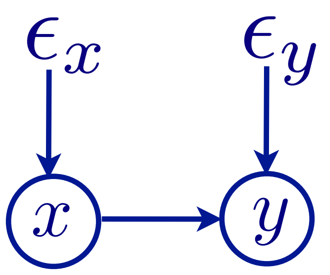
▶ If $\epsilon_j \sim$ non-Gaussian, then the true graph is **uniquely identifiable** from the joint distribution.

- (faithfulness not required!)

Two variable case

True model

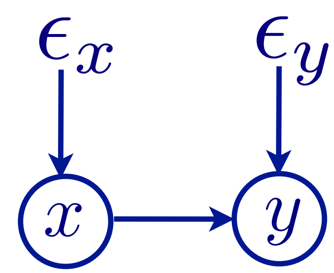
$$y = \beta x + \epsilon_y$$



Two variable case

True model

$$y = \beta x + \epsilon_y$$

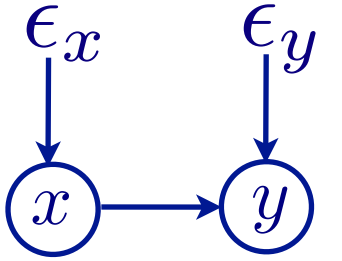


$$x \perp\!\!\!\perp \epsilon_y$$

Two variable case

True model

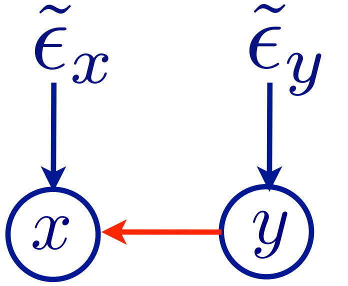
$$y = \beta x + \epsilon_y$$



$$x \perp\!\!\!\perp \epsilon_y$$

Backwards model

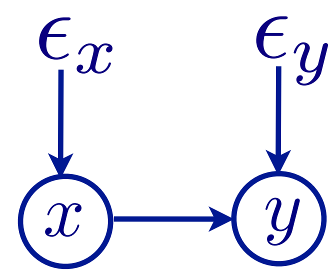
$$x = \theta y + \tilde{\epsilon}_x$$



Two variable case

True model

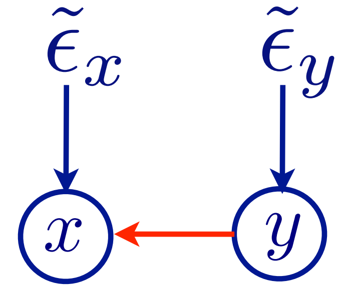
$$y = \beta x + \epsilon_y$$



$$x \perp\!\!\!\perp \epsilon_y$$

Backwards model

$$x = \theta y + \tilde{\epsilon}_x$$

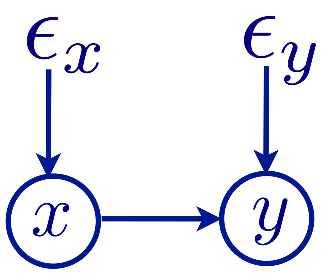


$$y \perp\!\!\!\perp \tilde{\epsilon}_x$$

Two variable case

True model

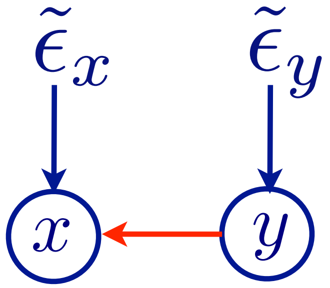
$$y = \beta x + \epsilon_y$$



$$x \perp\!\!\!\perp \epsilon_y$$

Backwards model

$$x = \theta y + \tilde{\epsilon}_x$$



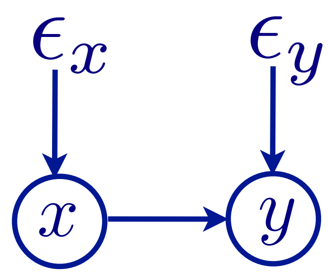
$$y \perp\!\!\!\perp \tilde{\epsilon}_x$$

$$\begin{aligned} \tilde{\epsilon}_x &= x - \theta y \\ &= x - \theta(\beta x + \epsilon_y) \\ &= (1 - \theta\beta)x - \theta\epsilon_y \end{aligned}$$

Two variable case

True model

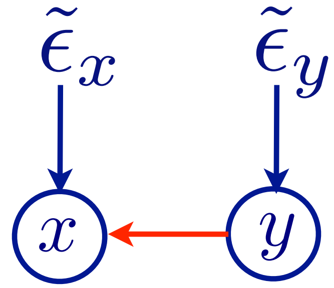
$$y = \beta x + \epsilon_y$$



$$x \perp\!\!\!\perp \epsilon_y$$

Backwards model

$$x = \theta y + \tilde{\epsilon}_x$$



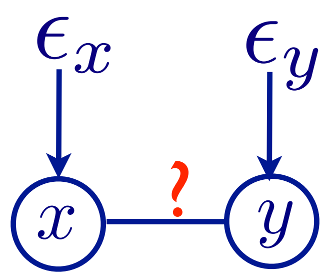
$$y \perp\!\!\!\perp \tilde{\epsilon}_x$$

$$\begin{aligned} \tilde{\epsilon}_x &= x - \theta y \\ &= x - \theta(\beta x + \epsilon_y) \\ &= (1 - \theta\beta)x - \theta\epsilon_y \end{aligned}$$

Why Normals are unusual

Forwards model $y = \beta x + \epsilon_y$

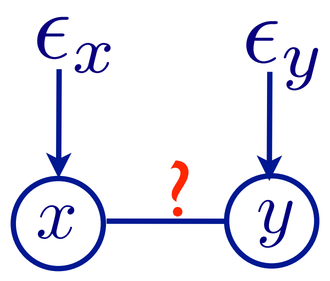
For backwards model $\tilde{\epsilon}_x = (1 - \theta\beta)x - \theta\epsilon_y$



Why Normals are unusual

Forwards model $y = \beta x + \epsilon_y$

For backwards model $\tilde{\epsilon}_x = (1 - \theta\beta)x - \theta\epsilon_y$



Theorem 1 (Darmois-Skitovich) *Let $X_1, \ldots, X_n$ be independent, non-degenerate random variables. If for two linear combinations*

$$
\begin{aligned}
l_1 &= a_1 X_1 + \ldots + a_n X_n, \quad a_i \neq 0 \\
l_2 &= b_1 X_1 + \ldots + b_n X_n, \quad b_i \neq 0
\end{aligned}
$$

are independent, then each X_i is normally distributed.

algorithm/ assumption
Markov
faithfulness
causal sufficiency
acyclicity
parametric assumption
output

algorithm/ assumption	PC / GES
Markov	✓
faithfulness	✓
causal sufficiency	✓
acyclicity	✓
parametric assumption	✗
output	Markov equivalence

algorithm/ assumption	PC / GES	FCI
Markov	✓	✓
faithfulness	✓	✓
causal sufficiency	✓	✗
acyclicity	✓	✓
parametric assumption	✗	✗
output	Markov equivalence	PAG

algorithm/ assumption	PC / GES	FCI	CCD
Markov	✓	✓	✓
faithfulness	✓	✓	✓
causal sufficiency	✓	✗	✓
acyclicity	✓	✓	✗
parametric assumption	✗	✗	✗
output	Markov equivalence	PAG	PAG

algorithm/ assumption	PC / GES	FCI	CCD	LiNGaM
Markov	✓	✓	✓	✓
faithfulness	✓	✓	✓	✗
causal sufficiency	✓	✗	✓	✓
acyclicity	✓	✓	✗	✓
parametric assumption	✗	✗	✗	linear non-Gaussian
output	Markov equivalence	PAG	PAG	unique DAG

algorithm/ assumption	PC / GES	FCI	CCD	LiNGaM	lvLiNGaM
Markov	✓	✓	✓	✓	✓
faithfulness	✓	✓	✓	✗	✓
causal sufficiency	✓	✗	✓	✓	✗
acyclicity	✓	✓	✗	✓	✓
parametric assumption	✗	✗	✗	linear non-Gaussian	linear non-Gaussian
output	Markov equivalence	PAG	PAG	unique DAG	set of DAGs

algorithm/ assumption	PC / GES	FCI	CCD	LiNGaM	lvLiNGaM	cyclic LiNGaM
Markov	✓	✓	✓	✓	✓	✓
faithfulness	✓	✓	✓	✗	✓	~
causal sufficiency	✓	✗	✓	✓	✗	✓
acyclicity	✓	✓	✗	✓	✓	✗
parametric assumption	✗	✗	✗	linear non-Gaussian	linear non-Gaussian	linear non-Gaussian
output	Markov equivalence	PAG	PAG	unique DAG	set of DAGs	set of graphs

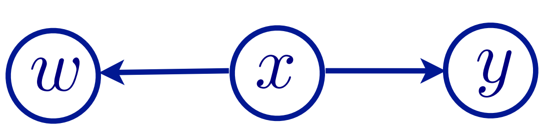
Limitations

ger ... e

*For **linear** **Gaussian** and for **multinomial** causal relations, an algorithm that identifies the Markov equivalence class of the true model is complete.*

(Pearl & Geiger 1988, Meek 1995)

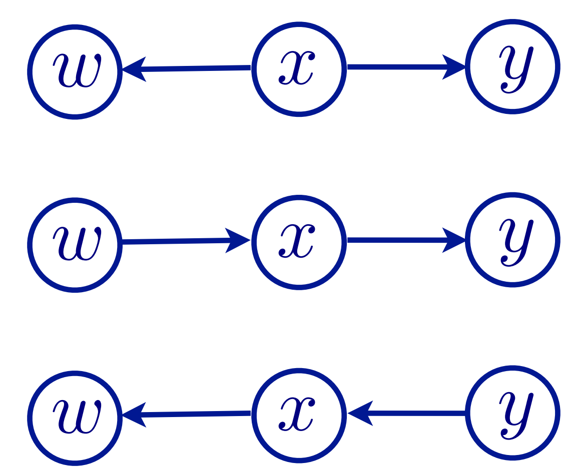
observational condition



Assumptions

- causal Markov
- causal faithfulness
- causal sufficiency
- acyclicity

Markov equivalence class

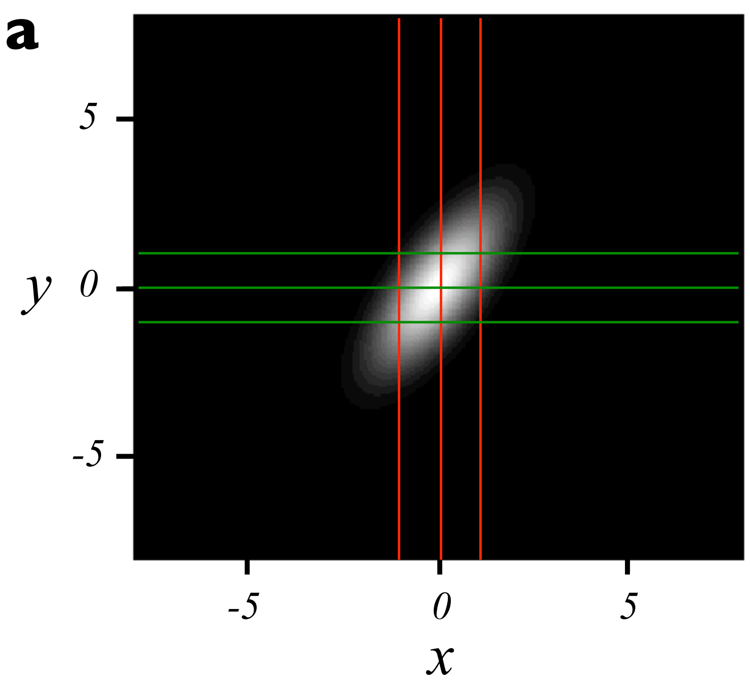


Bivariate Linear Gaussian case

True model

$$x = \epsilon_x$$
$$y = x + \epsilon_y$$

$$\epsilon_x, \epsilon_y \sim \text{indep. Gaussian}$$



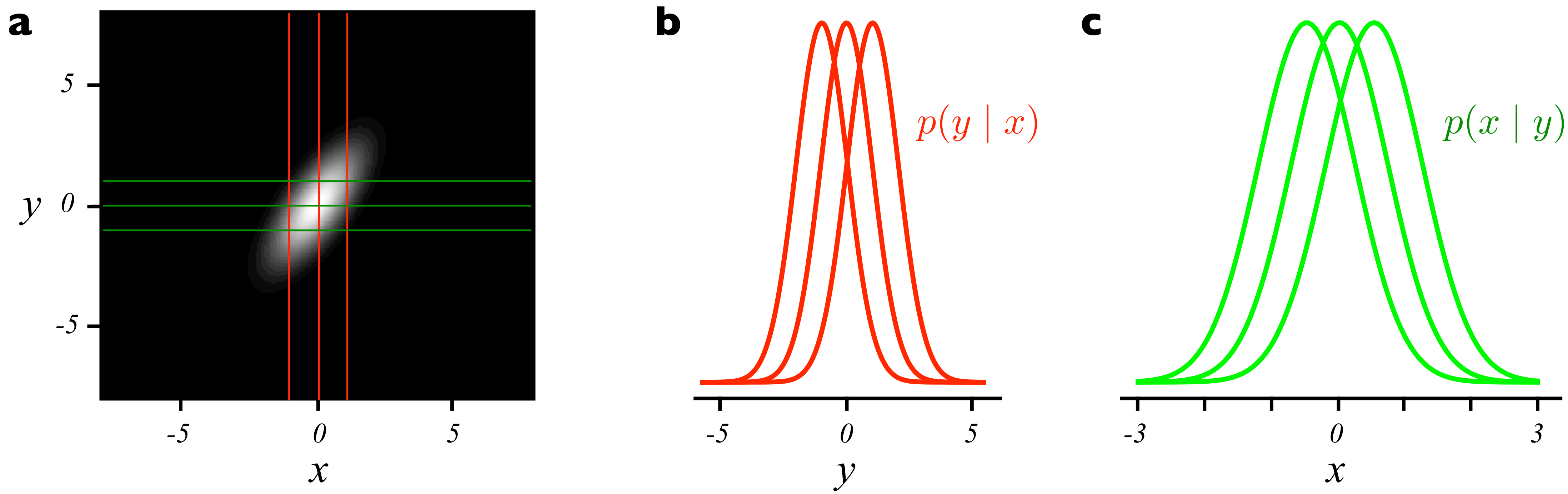
(graphics from Hoyer et al. 2009)

Bivariate Linear Gaussian case

True model

$$x = \epsilon_x$$
$$y = x + \epsilon_y$$

$$\epsilon_x, \epsilon_y \sim \text{indep. Gaussian}$$



Forwards (true) model

Backwards model

(graphics from Hoyer et al. 2009)

Continuous additive noise models

$$x_j = f_j(pa(x_j)) + \epsilon_j$$

Continuous additive noise models

$$x_j = f_j(pa(x_j)) + \epsilon_j$$

- If $f_j(.)$ is linear, then non-Gaussian errors are required for identifiability

Continuous additive noise models

$$x_j = f_j(pa(x_j)) + \epsilon_j$$

- If $f_j(.)$ is linear, then non-Gaussian errors are required for identifiability

➡ What if the errors are Gaussian, but $f_j(.)$ is non-linear?

Continuous additive noise models

$$x_j = f_j(pa(x_j)) + \epsilon_j$$

- If $f_j(.)$ is linear, then non-Gaussian errors are required for identifiability

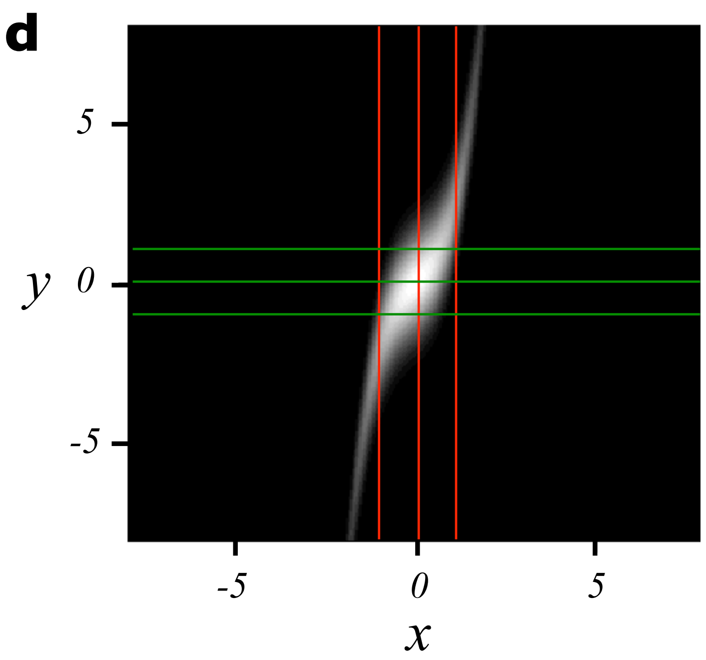
➡ What if the errors are Gaussian, but $f_j(.)$ is non-linear?

➡ More generally, under what circumstances is the graphical structure identifiable?

Bivariate non-linear Gaussian additive noise model

True model

$$x = \epsilon_x \qquad \epsilon_x, \epsilon_y \sim \text{indep. Gaussian}$$
$$y = x + x^3 + \epsilon_y$$



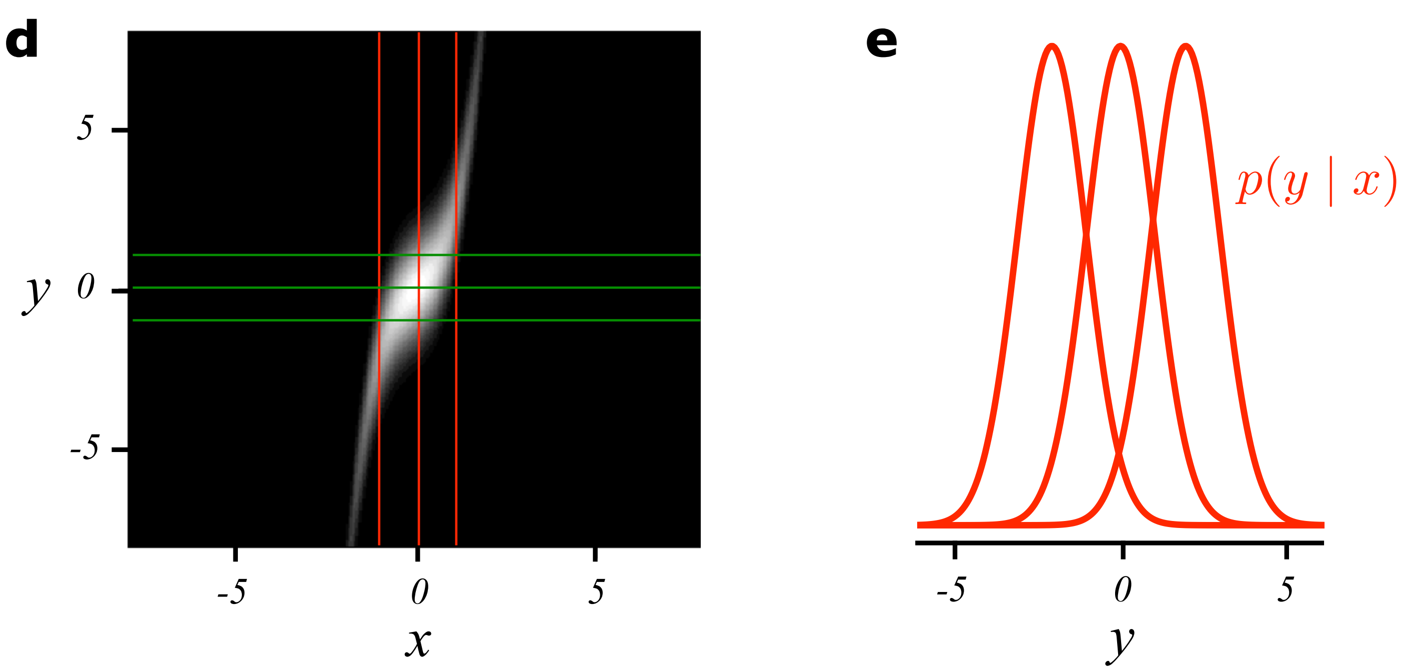
(graphics from Hoyer et al. 2009)

Bivariate non-linear Gaussian additive noise model

True model

$$x = \epsilon_x$$
$$y = x + x^3 + \epsilon_y$$

$\epsilon_x, \epsilon_y \sim$ indep. Gaussian



Forwards (true) model

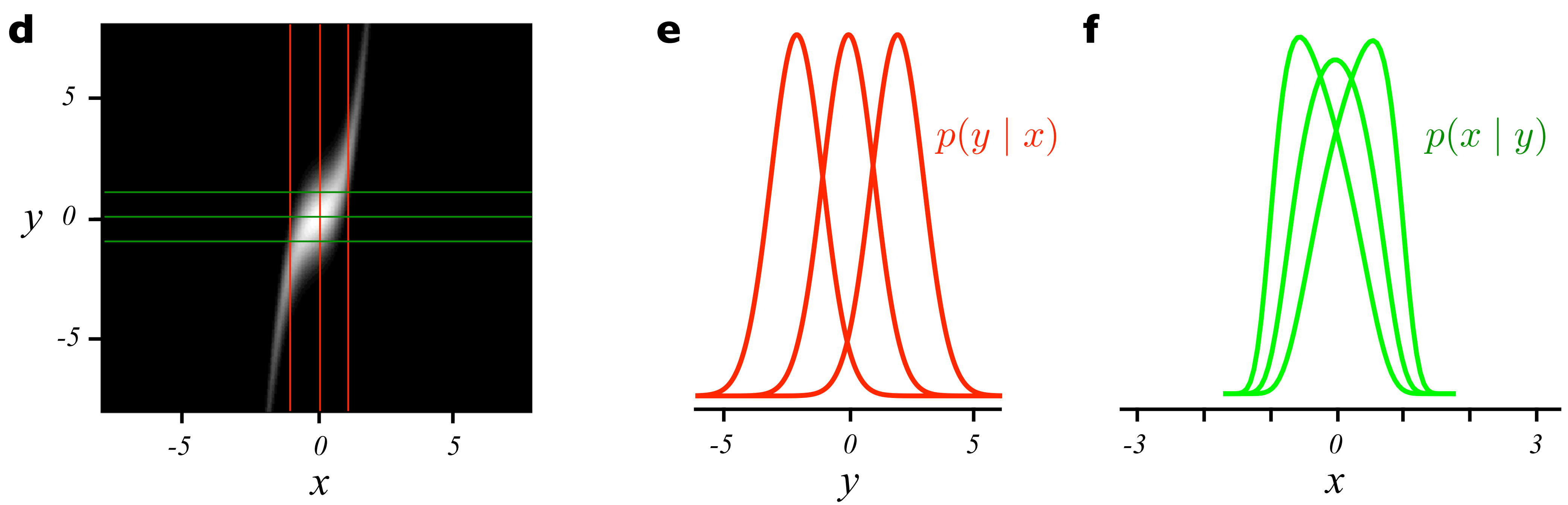
(graphics from Hoyer et al. 2009)

Bivariate non-linear Gaussian additive noise model

True model

$$x = \epsilon_x$$
$$y = x + x^3 + \epsilon_y$$

$$\epsilon_x, \epsilon_y \sim \text{indep. Gaussian}$$



Forwards (true) model

Backwards model

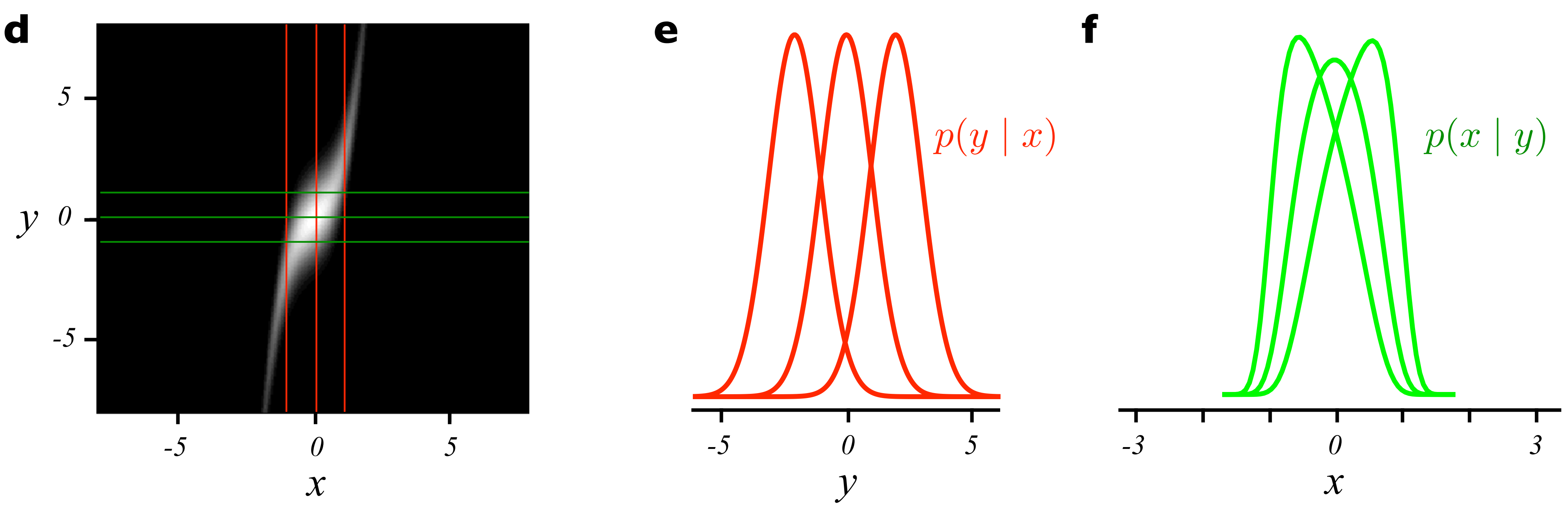
(graphics from Hoyer et al. 2009)

Bivariate non-linear Gaussian additive noise model

True model

$$x = \epsilon_x$$
$$y = x + x^3 + \epsilon_y$$

$\epsilon_x, \epsilon_y \sim$ indep. Gaussian



Forwards (true) model

Backwards model

$$x = g(y) + \tilde{\epsilon}_x$$
$$y \not\perp\!\!\!\perp \tilde{\epsilon}_x$$

(graphics from Hoyer et al. 2009)

General non-linear additive noise models

Hoyer et al. condition (HetalC): *Technical condition on the relation between the function, the noise distribution and the parent distribution that, if satisfied, permits a backward model.*

General non-linear additive noise models

Hoyer et al. condition (HetalC): *Technical condition on the relation between the function, the noise distribution and the parent distribution that, if satisfied, permits a backward model.*

- If the error terms are **Gaussian**, then the **only** functional form that **satisfies** HetalC is **linearity**, otherwise the model is **identifiable**.

General non-linear additive noise models

Hoyer et al. condition (HetalC): *Technical condition on the relation between the function, the noise distribution and the parent distribution that, if satisfied, permits a backward model.*

- If the error terms are **Gaussian**, then the **only** functional form that **satisfies** HetalC is **linearity**, otherwise the model is **identifiable**.
- If the errors are **non-Gaussian**, then there are (rather contrived) functions that satisfy HetalC, but **in general identifiability is guaranteed**.

General non-linear additive noise models

Hoyer et al. condition (HetalC): *Technical condition on the relation between the function, the noise distribution and the parent distribution that, if satisfied, permits a backward model.*

- If the error terms are **Gaussian**, then the **only** functional form that **satisfies** HetalC is **linearity**, otherwise the model is **identifiable**.
- If the errors are **non-Gaussian**, then there are (rather contrived) functions that satisfy HetalC, but **in general identifiability is guaranteed**.
  - this generalizes to multiple variables (assuming minimality*)!

General non-linear additive noise models

Hoyer et al. condition (HetalC): *Technical condition on the relation between the function, the noise distribution and the parent distribution that, if satisfied, permits a backward model.*

- If the error terms are **Gaussian**, then the **only** functional form that **satisfies** HetalC is **linearity**, otherwise the model is **identifiable**.
- If the errors are **non-Gaussian**, then there are (rather contrived) functions that satisfy HetalC, but **in general identifiability is guaranteed**.
  - this generalizes to multiple variables (assuming minimality*)!
  - extension to discrete additive noise models

General non-linear additive noise models

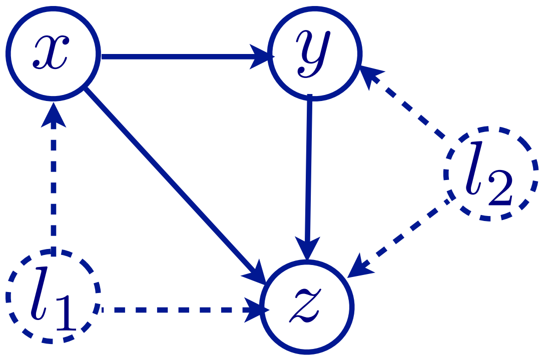
Hoyer et al. condition (HetalC): *Technical condition on the relation between the function, the noise distribution and the parent distribution that, if satisfied, permits a backward model.*

- If the error terms are **Gaussian**, then the **only** functional form that **satisfies** HetalC is **linearity**, otherwise the model is **identifiable**.
- If the errors are **non-Gaussian**, then there are (rather contrived) functions that satisfy HetalC, but **in general identifiability is guaranteed**.
  - this generalizes to multiple variables (assuming minimality*)!
  - extension to discrete additive noise models
- If the function is **linear**, but the error terms **non-Gaussian**, then one can't fit a linear backwards model (Lingam), but there are cases where **one can fit a non-linear backwards model**

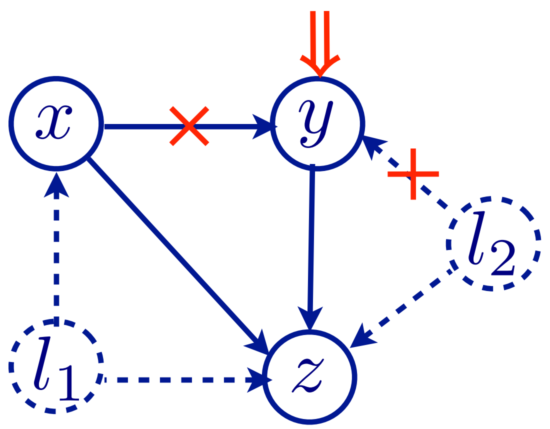
algorithm/ assumptions	PC / GES	FCI	CCD	LiNGaM	lvLiNGaM	cyclic LiNGaM
Markov	✓	✓	✓	✓	✓	✓
faithfulness	✓	✓	✓	✗	✓	~
causal sufficiency	✓	✗	✓	✓	✗	✓
acyclicity	✓	✓	✗	✓	✓	✗
parametric assumption	✗	✗	✗	linear non-Gaussian	linear non-Gaussian	linear non-Gaussian
output	Markov equivalence	PAG	PAG	unique DAG	set of DAGs	set of graphs

algorithm/ assumptions	PC / GES	FCI	CCD	LiNGaM	lvLiNGaM	cyclic LiNGaM	non-linear additive noise
Markov	✓	✓	✓	✓	✓	✓	✓
faithfulness	✓	✓	✓	✗	✓	~	minimality
causal sufficiency	✓	✗	✓	✓	✗	✓	✓
acyclicity	✓	✓	✗	✓	✓	✗	✓
parametric assumption	✗	✗	✗	linear non-Gaussian	linear non-Gaussian	linear non-Gaussian	non-linear additive noise
output	Markov equivalence	PAG	PAG	unique DAG	set of DAGs	set of graphs	unique DAG

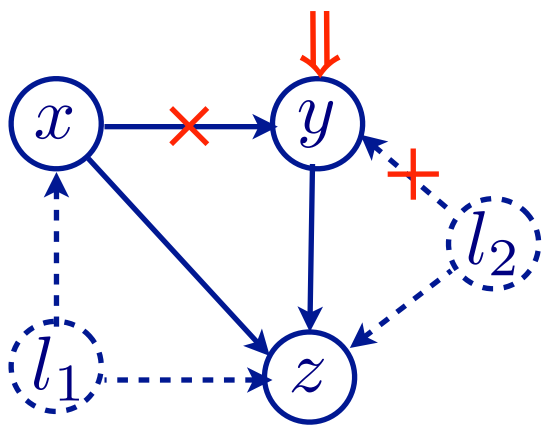
Experiments



Experiments



Experiments



- which experiments to perform?
- how to integrate the results from experimental data?
- what search space assumptions are still required?

SAT-based causal discovery

graphical constraints

SAT-based causal discovery

graphical constraints

- x is a cause of y

SAT-based causal discovery

graphical constraints

- x is a cause of y
- there is a path from x to y (via z)

SAT-based causal discovery

graphical constraints

- x is a cause of y
- there is a path from x to y (via z)
- x and y are independent

SAT-based causal discovery

graphical constraints

- x is a cause of y
- there is a path from x to y (via z)
- x and y are independent
- x and y are correlated conditional on **C** in an experiment where x was subject to intervention

SAT-based causal discovery

graphical constraints

- x is a cause of y
- there is a path from x to y (via z)
- x and y are independent
- x and y are correlated conditional on **C** in an experiment where x was subject to intervention

propositional constraints (in CNF) on true graph

$$(A \vee B \vee C) \wedge (D \vee E)$$

$$A$$

$$B \wedge E$$

$$\cdots$$

SAT-based causal discovery

propositional constraints (in CNF) on true graph

$$(A \lor B \lor C) \land (D \lor E)$$

$$A$$

$$B \land E$$

$$\ldots$$

graphical constraints

- x is a cause of y
- there is a path from x to y (via z)
- x and y are independent
- x and y are correlated conditional on **C** in an experiment where x was subject to intervention

SAT-solver

SAT-based causal discovery

graphical constraints

- x is a cause of y
- there is a path from x to y (via z)
- x and y are independent
- x and y are correlated conditional on **C** in an experiment where x was subject to intervention

propositional constraints (in CNF) on true graph

$(A \vee B \vee C) \wedge (D \vee E)$

A

$B \wedge E$

$\ldots$

SAT-solver

	w	x	y	z
w	0	0	?	1
x	0	0	0	0
y	0	0	0	0
z	?	?	?	0

direct edges

	w	x	y	z
w	0	0	?	?
x	0	0	0	?
y	?	0	0	0
z	?	?	0	0

confounders

SAT-based causal discovery

graphical constraints

- x is a cause of y
- there is a path from x to y (via z)
- x and y are independent
- x and y are correlated conditional on **C** in an experiment where x was subject to intervention

propositional constraints (in CNF) on true graph

$$(A \vee B \vee C) \wedge (D \vee E)$$

$$A$$

$$B \wedge E$$

$$\ldots$$

SAT-solver

	w	x	y	z
w	0	0	?	1
x	0	0	0	0
y	0	0	0	0
z	?	?	?	0

direct edges

	w	x	y	z
w	0	0	?	?
x	0	0	0	?
y	?	0	0	0
z	?	?	0	0

confounders

SATisfiability solver

- finds a truth value assignment for a Boolean formula in Conjunctive Normal Form (CNF)

SATisfiability solver

- finds a truth value assignment for a Boolean formula in Conjunctive Normal Form (CNF)
- a Boolean term X is a **backbone variable** if X takes the same value (T or F) in all satisfying truth value assignments of a given formula

Encoding a dependence: track the endpoints of paths

$$[x \not\perp y \mid \mathbf{C} \mid\mid \mathbf{J}]$$

$$\Leftrightarrow$$

Encoding a dependence: track the endpoints of paths

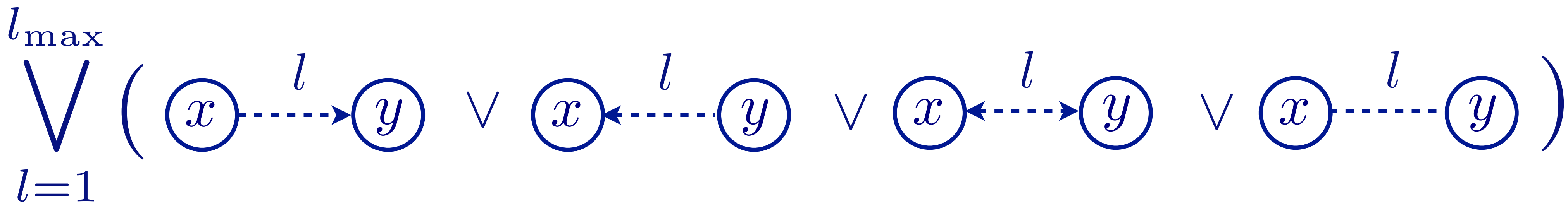
$$[x \not\perp y \mid \mathbf{C} \mid\mid \mathbf{J}]$$

$$\Longleftrightarrow$$

Encoding a dependence: track the endpoints of paths

$$[x \not\perp y \mid \mathbf{C} \mid\mid \mathbf{J}]$$

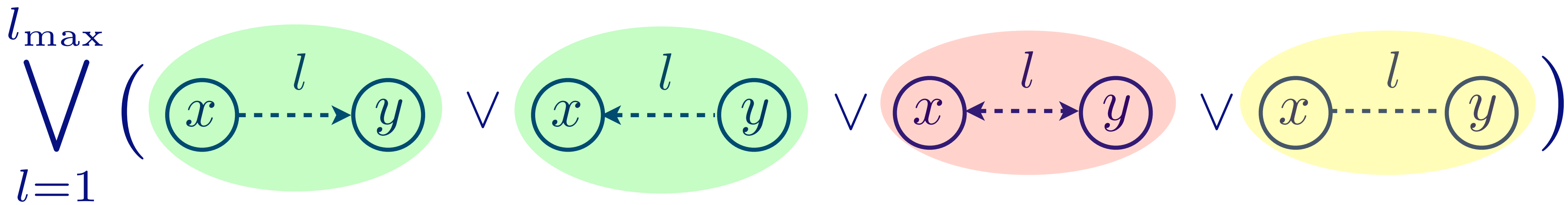
$$\Leftrightarrow$$



Encoding a dependence: track the endpoints of paths

$$[x \not\perp y \mid \mathbf{C} \mid\mid \mathbf{J}]$$

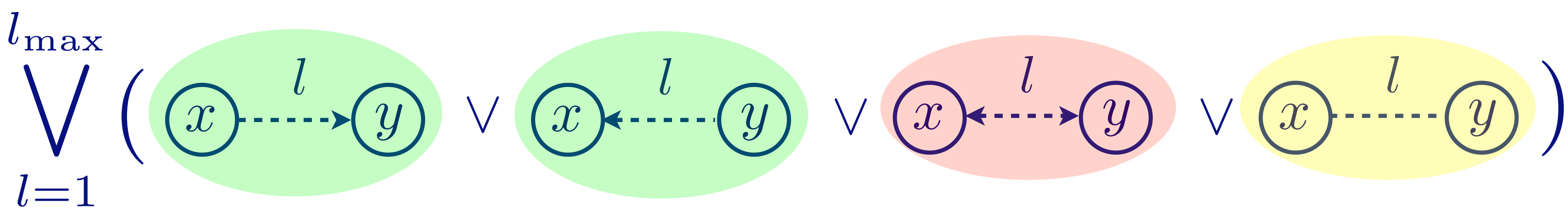
$$\Longleftrightarrow$$

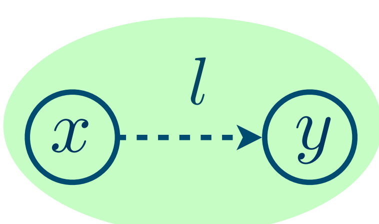


Encoding a dependence: track the endpoints of paths

$$[x \not\perp y \mid \mathbf{C} \mid\mid \mathbf{J}]$$

$$\Leftrightarrow$$

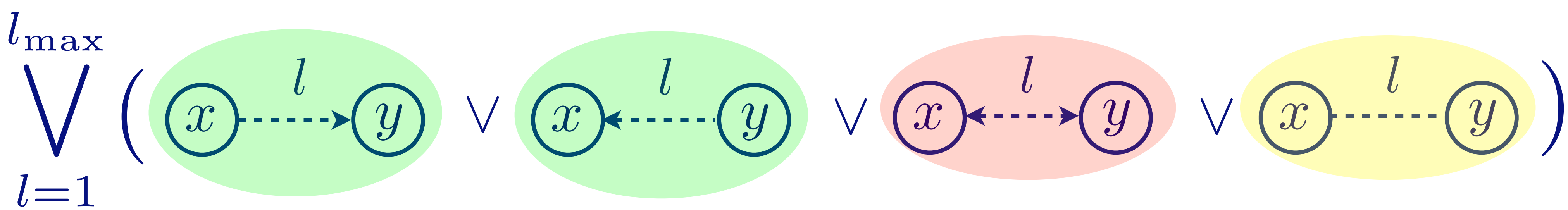


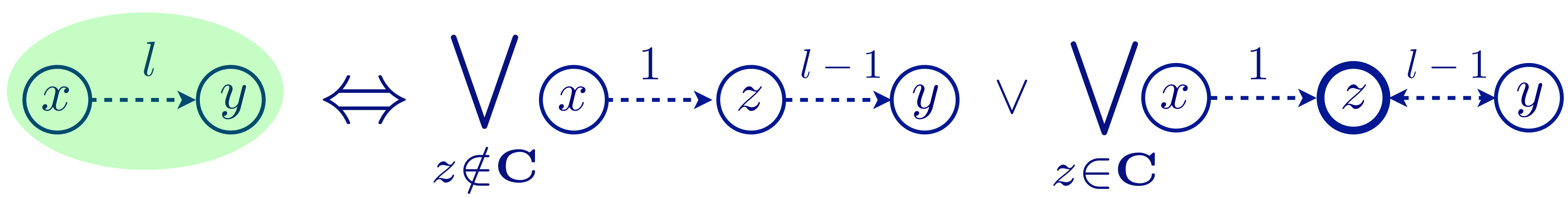


Encoding a dependence: track the endpoints of paths

$$[x \not\perp y \mid \mathbf{C} \mid\mid \mathbf{J}]$$

$$\Leftrightarrow$$

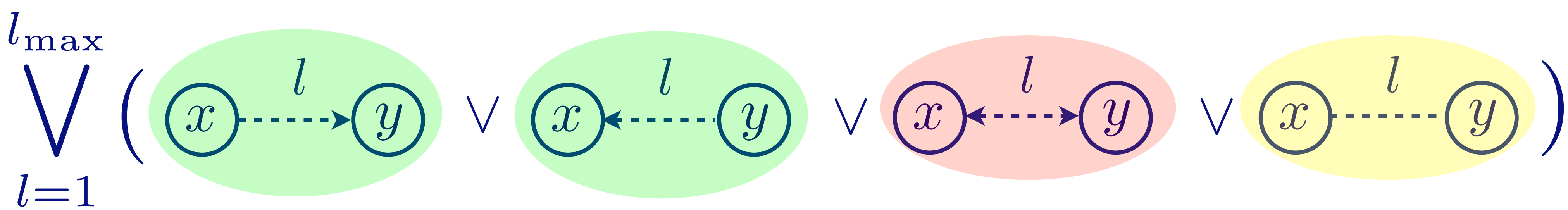


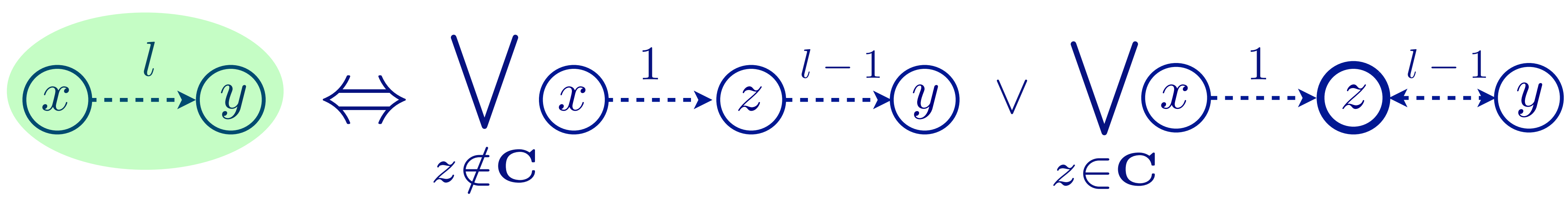


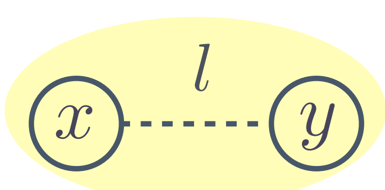
Encoding a dependence: track the endpoints of paths

$$[x \not\perp y \mid \mathbf{C} \mid\mid \mathbf{J}]$$

$$\Leftrightarrow$$

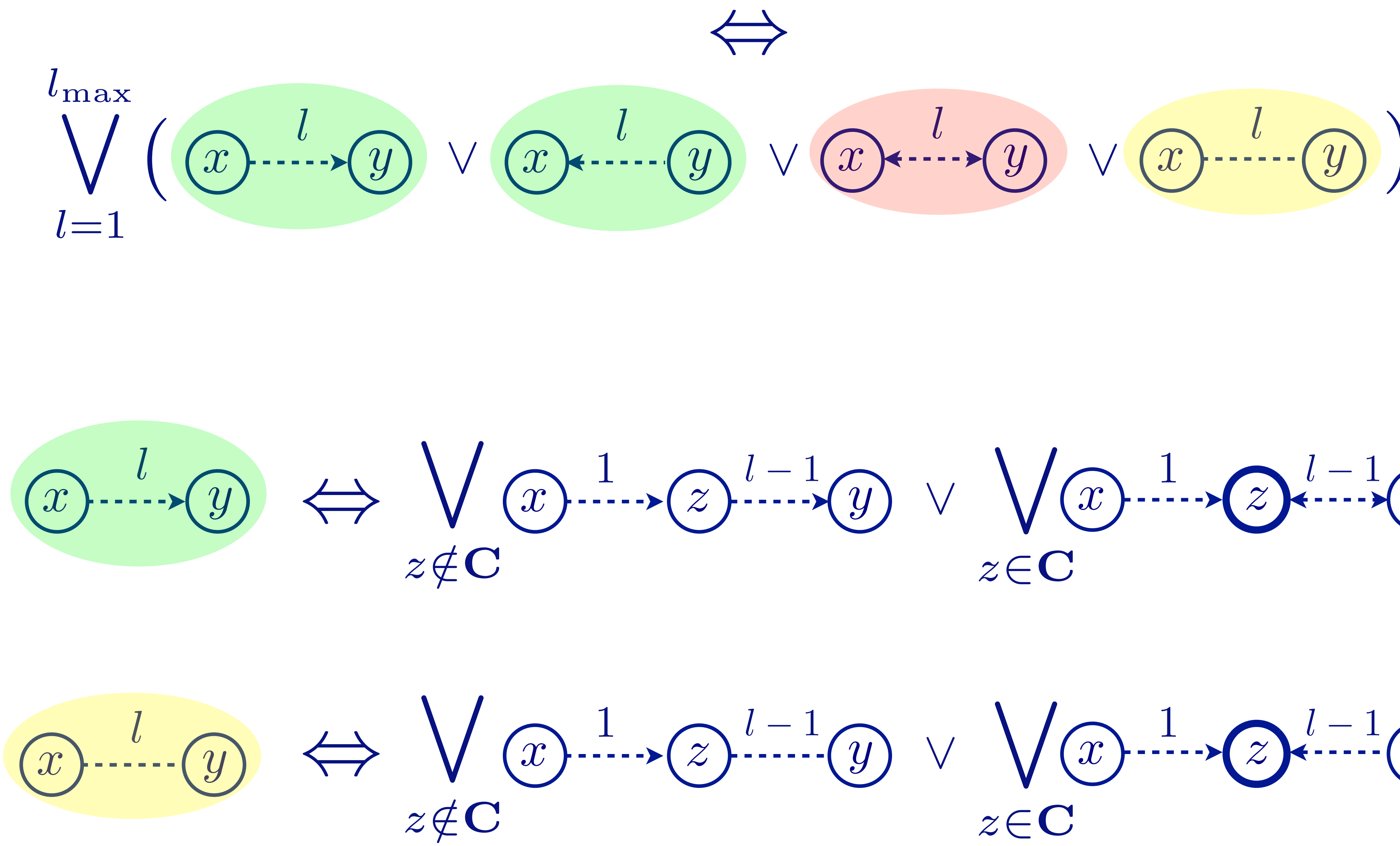




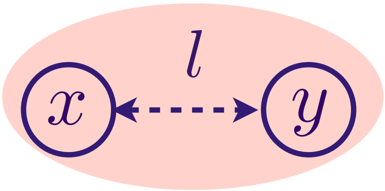


Encoding a dependence: track the endpoints of paths

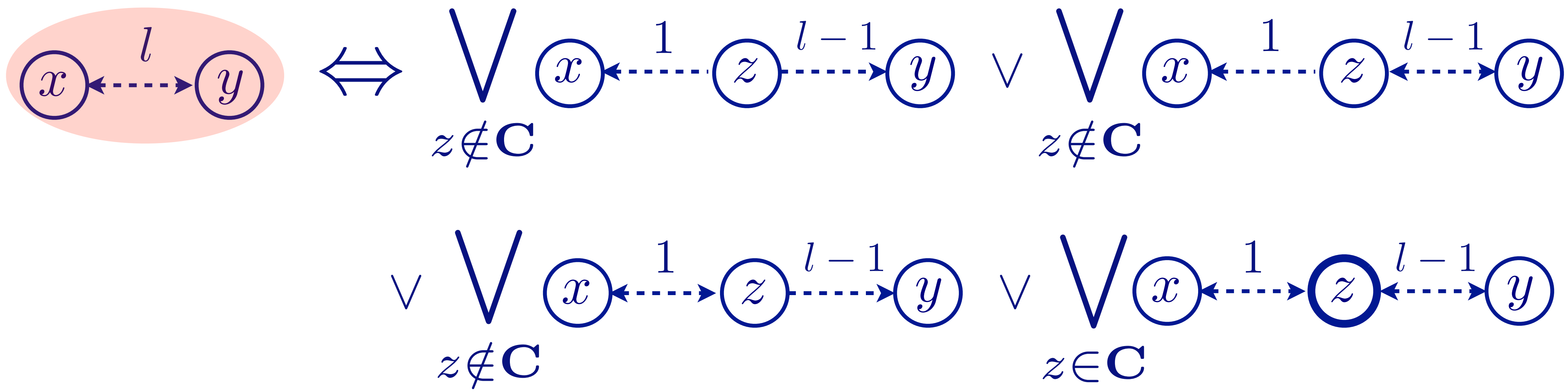
$$[x \not\perp y \mid \mathbf{C} \mid\mid \mathbf{J}]$$



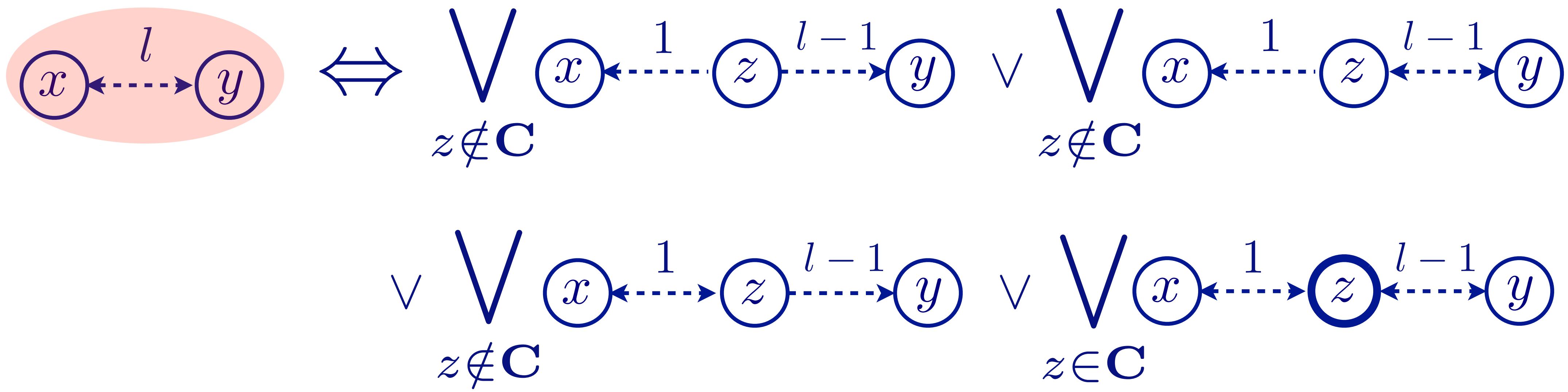
Encoding continued

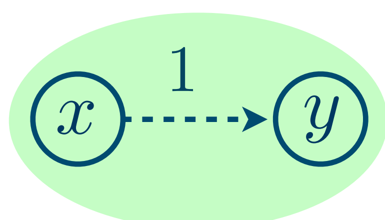


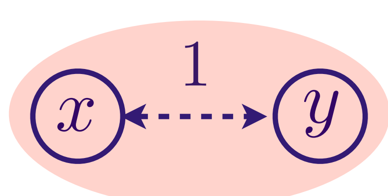
Encoding continued

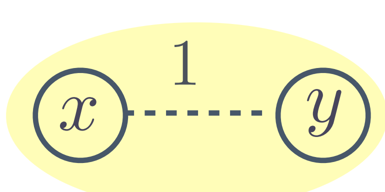


Encoding continued

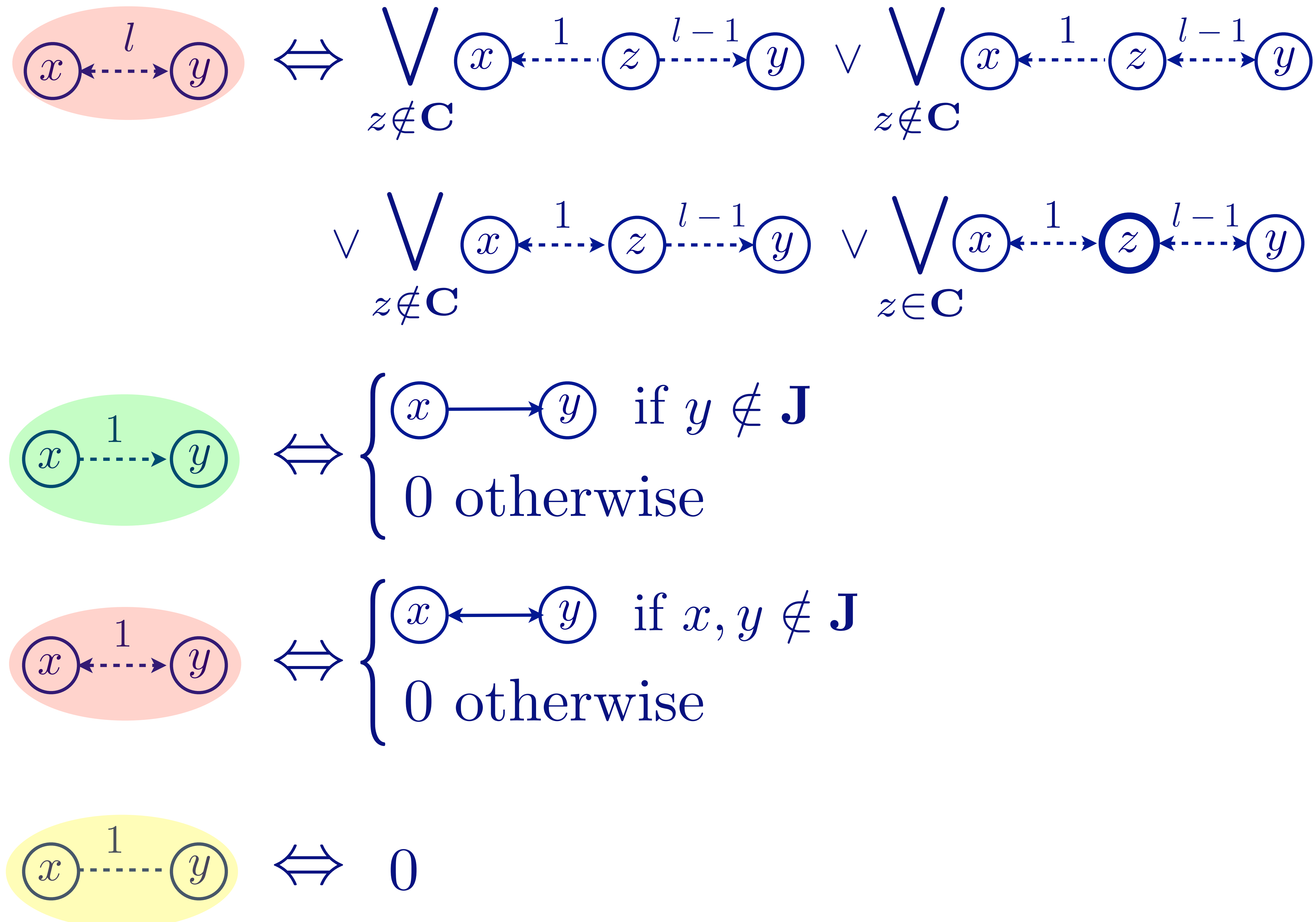








Encoding continued



$$[x \not\perp y \,|\, \mathbf{C} \,||\, \mathbf{J}] \quad \Leftrightarrow \quad \bigvee_{l=1}^{l_{\max}} \left([x \underset{\mathbf{C},\mathbf{J}}{\overset{l}{-\!\cdots\!>}} y] \vee [y \underset{\mathbf{C},\mathbf{J}}{\overset{l}{-\!\cdots\!>}} x] \vee [x \underset{\mathbf{C},\mathbf{J}}{\overset{l}{<\!\cdots\!>}} y] \vee [x \underset{\mathbf{C},\mathbf{J}}{\overset{l}{-\!\cdots\!-}} y] \right)$$

$$[x \underset{\mathbf{C},\mathbf{J}}{\overset{l}{-\!\cdots\!>}} y] \quad \Leftrightarrow \quad \bigvee_{z \notin \mathbf{C}} \left([x \underset{\mathbf{C},\mathbf{J}}{\overset{1}{-\!\cdots\!>}} z] \wedge [z \underset{\mathbf{C},\mathbf{J}}{\overset{l-1}{-\!\cdots\!>}} y] \right) \vee \bigvee_{z \in \mathbf{C}} \left([x \underset{\mathbf{C},\mathbf{J}}{\overset{1}{-\!\cdots\!>}} z] \wedge [z \underset{\mathbf{C},\mathbf{J}}{\overset{l-1}{<\!\cdots\!>}} y] \right)$$

$$\begin{aligned}[x \underset{\mathbf{C},\mathbf{J}}{\overset{l}{<\!\cdots\!>}} y] \quad \Leftrightarrow \quad & \bigvee_{z \notin \mathbf{C}} \left([z \underset{\mathbf{C},\mathbf{J}}{\overset{1}{-\!\cdots\!>}} x] \wedge [z \underset{\mathbf{C},\mathbf{J}}{\overset{l-1}{-\!\cdots\!>}} y] \right) \vee \bigvee_{z \notin \mathbf{C}} \left([z \underset{\mathbf{C},\mathbf{J}}{\overset{1}{-\!\cdots\!>}} x] \wedge [z \underset{\mathbf{C},\mathbf{J}}{\overset{l-1}{<\!\cdots\!>}} y] \right) \vee \\ & \bigvee_{z \notin \mathbf{C}} \left([x \underset{\mathbf{C},\mathbf{J}}{\overset{1}{<\!\cdots\!>}} z] \wedge [z \underset{\mathbf{C},\mathbf{J}}{\overset{l-1}{-\!\cdots\!>}} y] \right) \vee \bigvee_{z \in \mathbf{C}} \left([x \underset{\mathbf{C},\mathbf{J}}{\overset{1}{<\!\cdots\!>}} z] \wedge [z \underset{\mathbf{C},\mathbf{J}}{\overset{l-1}{<\!\cdots\!>}} y] \right)\end{aligned}$$

$$[x \underset{\mathbf{C},\mathbf{J}}{\overset{l}{-\!\cdots\!-}} y] \quad \Leftrightarrow \quad \bigvee_{z \notin \mathbf{C}} \left([x \underset{\mathbf{C},\mathbf{J}}{\overset{1}{-\!\cdots\!>}} z] \wedge [z \underset{\mathbf{C},\mathbf{J}}{\overset{l-1}{-\!\cdots\!-}} y] \right) \vee \bigvee_{z \in \mathbf{C}} \left([x \underset{\mathbf{C},\mathbf{J}}{\overset{1}{-\!\cdots\!>}} z] \wedge [y \underset{\mathbf{C},\mathbf{J}}{\overset{l-1}{-\!\cdots\!>}} z] \right)$$

$$[x \underset{\mathbf{C},\mathbf{J}}{\overset{1}{-\!\cdots\!>}} y] \quad \Leftrightarrow \quad \begin{cases} [x \to y] & \text{if } y \notin \mathbf{J} \\ 0 & \text{otherwise} \end{cases}$$

$$[x \underset{\mathbf{C},\mathbf{J}}{\overset{1}{<\!\cdots\!>}} y] \quad \Leftrightarrow \quad \begin{cases} [x \leftrightarrow y] & \text{if } x \notin \mathbf{J} \text{ and } y \notin \mathbf{J} \\ 0 & \text{otherwise} \end{cases}$$

$$[x \underset{\mathbf{C},\mathbf{J}}{\overset{1}{-\!\cdots\!-}} y] \quad \Leftrightarrow \quad 0$$

$$[x \not\perp y \,|\, \mathbf{C} \,||\, \mathbf{J}] \;\Leftrightarrow\; \bigvee_{l=1}^{l_{\max}} \left([x \overset{l}{\underset{\mathbf{C},\mathbf{J}}{-\cdots>}} y] \vee [y \overset{l}{\underset{\mathbf{C},\mathbf{J}}{-\cdots>}} x] \vee [x \overset{l}{\underset{\mathbf{C},\mathbf{J}}{<\cdots>}} y] \vee [x \overset{l}{\underset{\mathbf{C},\mathbf{J}}{-\cdots-}} y] \right)$$

$$[x \overset{l}{\underset{\mathbf{C},\mathbf{J}}{-\cdots>}} y] \;\Leftrightarrow\; \bigvee_{z \notin \mathbf{C}} \left([x \overset{1}{\underset{\mathbf{C},\mathbf{J}}{-\cdots>}} z] \wedge [z \overset{l-1}{\underset{\mathbf{C},\mathbf{J}}{-\cdots>}} y] \right) \vee \bigvee_{z \in \mathbf{C}} \left([x \overset{1}{\underset{\mathbf{C},\mathbf{J}}{-\cdots>}} z] \wedge [z \overset{l-1}{\underset{\mathbf{C},\mathbf{J}}{<\cdots>}} y] \right)$$

$$[x \overset{l}{\underset{\mathbf{C},\mathbf{J}}{<\cdots>}} y] \;\Leftrightarrow\; \bigvee_{z \notin \mathbf{C}} \left([z \overset{1}{\underset{\mathbf{C},\mathbf{J}}{-\cdots>}} x] \wedge [z \overset{l-1}{\underset{\mathbf{C},\mathbf{J}}{-\cdots>}} y] \right) \vee \bigvee_{z \notin \mathbf{C}} \left([z \overset{1}{\underset{\mathbf{C},\mathbf{J}}{-\cdots>}} x] \wedge [z \overset{l-1}{\underset{\mathbf{C},\mathbf{J}}{<\cdots>}} y] \right) \vee$$

$$\bigvee_{z \notin \mathbf{C}} \left([x \overset{1}{\underset{\mathbf{C},\mathbf{J}}{<\cdots>}} z] \wedge [z \overset{l-1}{\underset{\mathbf{C},\mathbf{J}}{-\cdots>}} y] \right) \vee \bigvee_{z \in \mathbf{C}} \left([x \overset{1}{\underset{\mathbf{C},\mathbf{J}}{<\cdots>}} z] \wedge [z \overset{l-1}{\underset{\mathbf{C},\mathbf{J}}{<\cdots>}} y] \right)$$

$$[x \overset{l}{\underset{\mathbf{C},\mathbf{J}}{-\cdots-}} y] \;\Leftrightarrow\; \bigvee_{z \notin \mathbf{C}} \left([x \overset{1}{\underset{\mathbf{C},\mathbf{J}}{-\cdots>}} z] \wedge [z \overset{l-1}{\underset{\mathbf{C},\mathbf{J}}{-\cdots-}} y] \right) \vee \bigvee_{z \in \mathbf{C}} \left([x \overset{1}{\underset{\mathbf{C},\mathbf{J}}{-\cdots>}} z] \wedge [y \overset{l-1}{\underset{\mathbf{C},\mathbf{J}}{-\cdots>}} z] \right)$$

$$[x \overset{1}{\underset{\mathbf{C},\mathbf{J}}{-\cdots>}} y] \;\Leftrightarrow\; \begin{cases} [x \rightarrow y] & \text{if } y \notin \mathbf{J} \\ 0 & \text{otherwise} \end{cases}$$

$$[x \overset{1}{\underset{\mathbf{C},\mathbf{J}}{<\cdots>}} y] \;\Leftrightarrow\; \begin{cases} [x \leftrightarrow y] & \text{if } x \notin \mathbf{J} \text{ and } y \notin \mathbf{J} \\ 0 & \text{otherwise} \end{cases}$$

$$[x \overset{1}{\underset{\mathbf{C},\mathbf{J}}{-\cdots-}} y] \;\Leftrightarrow\; 0$$

longest path that needs to be considered: $l\_max = 2n-4$ where $n = |\mathbf{V}|$

Algorithm

Proceed in order of conditioning set size

- heuristically **find unknown independence / dependence relations** and determine them.
- **Encode the relations** into the working formula F, including definitions as needed.
- **Determine the "backbone"** of F using the SAT-solver, i.e. for each pair of variables (x,y) in **V** and for each edge type determine whether it is
  - **present** in all causal structures consistent with the input.
  - **absent** in all causal structures consistent with the input.
  - **unknown**, i.e. present in some, and absent in other causal structures consistent with the input.

Algorithm

any **background knowledge** representable using encoding can be included

Proceed in order of conditioning set size

- heuristically **find unknown independence / dependence relations** and determine them.
- **Encode the relations** into the working formula F, including definitions as needed.
- **Determine the "backbone"** of F using the SAT-solver, i.e. for each pair of variables (x,y) in **V** and for each edge type determine whether it is
  - **present** in all causal structures consistent with the input.
  - **absent** in all causal structures consistent with the input.
  - **unknown**, i.e. present in some, and absent in other causal structures consistent with the input.

Algorithm

Proceed in order of conditioning set size

- heuristically **find unknown independence / dependence relations** and determine them.
- **Encode the relations** into the working formula F, including definitions as needed.
- **Determine the "backbone"** of F using the SAT-solver, i.e. for each pair of variables (x,y) in **V** and for each edge type determine whether it is
  - **present** in all causal structures consistent with the input.
  - **absent** in all causal structures consistent with the input.
  - **unknown**, i.e. present in some, and absent in other causal structures consistent with the input.

any **background knowledge** representable using encoding can be included

independence constraints can be treated separately from dependence constraints

Algorithm

Proceed in order of conditioning set size

- heuristically **find unknown independence / dependence relations** and determine them.
- **Encode the relations** into the working formula F, including definitions as needed.
- **Determine the "backbone"** of F using the SAT-solver, i.e. for each pair of variables (x,y) in **V** and for each edge type determine whether it is
  - **present** in all causal structures consistent with the input.
  - **absent** in all causal structures consistent with the input.
  - **unknown**, i.e. present in some, and absent in other causal structures consistent with the input.

any **background knowledge** representable using encoding can be included

independence constraints can be treated separately from dependence constraints

you can compute the backbone over any graphical feature that you are interested in

assumption/ algorithm	PC / GES	FCI	CCD	LiNGaM	lvLiNGaM	cyclic LiNGaM	non-linear additive noise
Markov	✓	✓	✓	✓	✓	✓	✓
faithfulness	✓	✓	✓	✗	✓	~	minimality
causal sufficiency	✓	✗	✓	✓	✗	✓	✓
acyclicity	✓	✓	✗	✓	✓	✗	✓
parametric assumption	✗	✗	✗	linear non-Gaussian	linear non-Gaussian	linear non-Gaussian	non-linear additive noise
output	Markov equivalence	PAG	PAG	unique DAG	set of DAGs	set of graphs	unique DAG

assumption/ algorithm	PC / GES	FCI	CCD	LiNGaM	lvLiNGaM	cyclic LiNGaM	non-linear additive noise	SAT
Markov	✓	✓	✓	✓	✓	✓	✓	✓
faithfulness	✓	✓	✓	✗	✓	~	minimality	✓
causal sufficiency	✓	✗	✓	✓	✗	✓	✓	✗
acyclicity	✓	✓	✗	✓	✓	✗	✓	✗*
parametric assumption	✗	✗	✗	linear non-Gaussian	linear non-Gaussian	linear non-Gaussian	non-linear additive noise	✗
output	Markov equivalence	PAG	PAG	unique DAG	set of DAGs	set of graphs	unique DAG	query based

assumption/ algorithm	PC / GES	FCI	CCD	LiNGaM	lvLiNGaM	cyclic LiNGaM	non-linear additive noise	SAT
Markov	✓	✓	✓	✓	✓	✓	✓	✓
faithfulness	✓	✓	✓	✗	✓	~	minimality	✓
causal sufficiency	✓	✗	✓	✓	✗	✓	✓	✗
acyclicity	✓	✓	✗	✓	✓	✗	✓	✗*
parametric assumption	✗	✗	✗	linear non-Gaussian	linear non-Gaussian	linear non-Gaussian	non-linear additive noise	✗
output	Markov equivalence	PAG	PAG	unique DAG	set of DAGs	set of graphs	unique DAG	query based

assumption/ algorithm	PC / GES	FCI	CCD	LiNGaM	lvLiNGaM	cyclic LiNGaM	non-linear additive noise	SAT
Markov	✓	✓	✓	✓	✓	✓	✓	✓
faithfulness	✓	✓	✓	✗	✓	~	minimality	✓
causal sufficiency	✓	✗	✓	✓	✗	✓	✓	✗
acyclicity	✓	✓	✗	✓	✓	✗	✓	✗*
parametric assumption	✗	✗	✗	linear non-Gaussian	linear non-Gaussian	linear non-Gaussian	non-linear additive noise	✗
output	Markov equivalence	PAG	PAG	unique DAG	set of DAGs	set of graphs	unique DAG	query based
application	wide use	some?	none	fMRI	requires too much data	fMRI	starting	in development

References

- Limitations
  - Geiger & Pearl, **On the logic of influence diagrams**, UAI 1988.
  - Meek, **Strong completeness and faithfulness in Bayesian networks**, UAI 1995.
- LiNGaM
  - Shimizu et al, **A linear non-Gaussian acyclic model for causal discovery**, JMLR, 2006.
  - Hoyer et al., **Estimation of causal effects using linear non-Gaussian causal models with hidden variables**, IJAR 2008.
  - Lacerda et al., **Discoverying cyclic causal models by Independent Component Ananlysis**, UAI 2008.
- additive noise models
  - Hoyer et al., **Nonlinear causal discovery with additive noise models**, NIPS 2009.
  - Mooij et al., **Regression by dependence minimization and its application to causal inference**, ICML 2009.
  - Peters et al., **Causal inference on discrete data using additive noise models**, IEEE..., 2011.
  - Peters et al., **Identifiability of causal graphs using functional models**, UAI 2011.
- SAT-based approaches
  - Triantafillou et al., **Learning causal structure from overlapping variable sets**, AISTATS 2010.
  - Claassen & Heskes, **A logical characterization of constraint-based causal discovery**, UAI 2011.
  - Hyttinen et al., **Discovering cyclic causal models with latent variables: A SAT-based approach**, UAI 2013.

Thank you!