Discovering Cyclic Causal Models by Independent Components Analysis

Gustavo Lacerda
Peter Spirtes
Joseph Ramsey
Patrik O. Hoyer

Our goal

- Discover the structure and parameters of linear SEMs, without experimental data.

Outline

- Linear SEMs
- LiNGAM method *(Shimizu et al 2006)*: uniquely identifies acyclic linear SEMs, by using ICA
- LiNG-D: our new method, based on LiNGAM, but handles cycles
- Now, the answer given is underdetermined
- But we argue that stability can be a powerful constraint

Linear SEMs

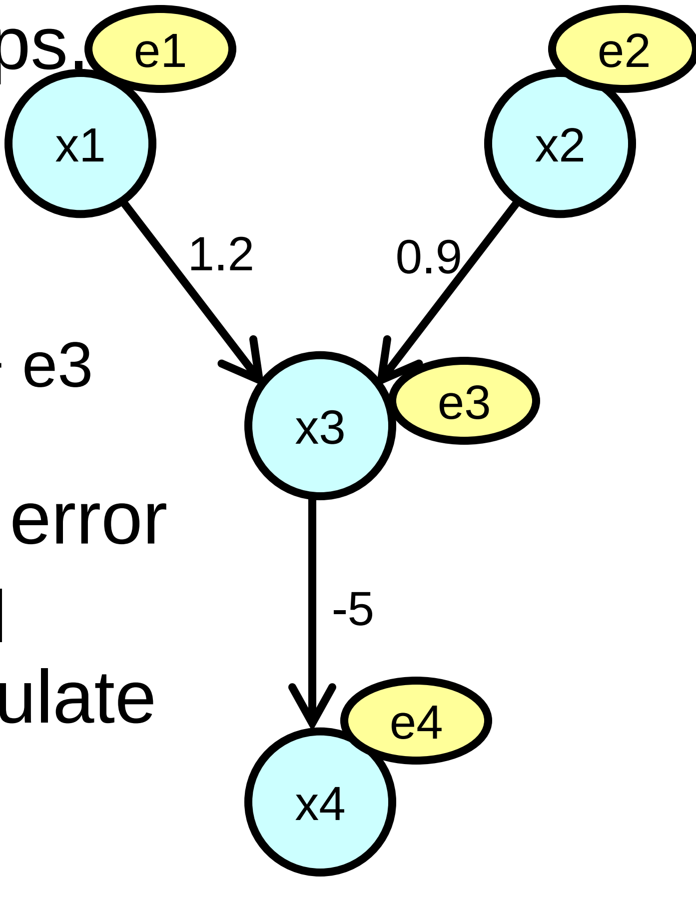
- Directed graphical models

 that represent **causal** relationships.

M:

$$x1 = e1$$
$$x2 = e2$$
$$x3 = 1.2\ x1 + 0.9\ x2 + e3$$
$$x4 = -5\ x3 + e4$$

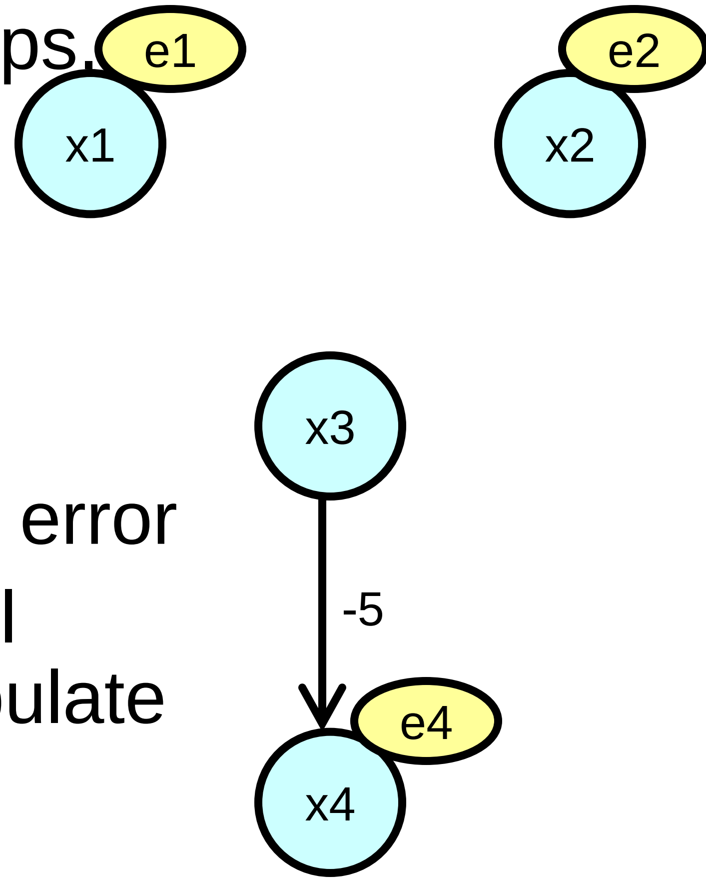
- linear combination of parents + error
- “causal” means that they model what happens when you manipulate
- e.g. manipulating x3…



Linear SEMs

- Directed graphical models

that represent **causal** relationships.



M(do(x3=k)):

$x1 = e1$

$x2 = e2$

$x3 = k$

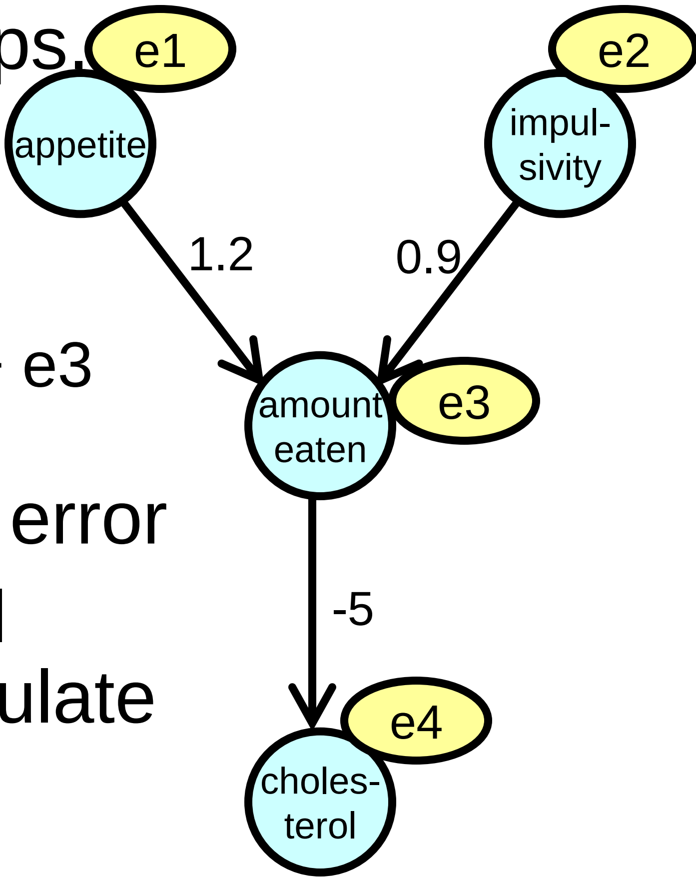
$x4 = -5\, x3 + e4$

- linear combination of parents + error
- "causal" means that they model what happens when you manipulate
- e.g. manipulating x3…

Linear SEMs

- Directed graphical models

that represent **causal** relationships.



M:

$$x1 = e1$$
$$x2 = e2$$
$$x3 = 1.2\ x1 + 0.9\ x2 + e3$$
$$x4 = -5\ x3 + e4$$

- linear combination of parents + error
- "causal" means that they model what happens when you manipulate
- e.g. manipulating x3…

Linear SEMs

- Directed graphical models

that represent **causal** relationships.

M(do(x3=k)):

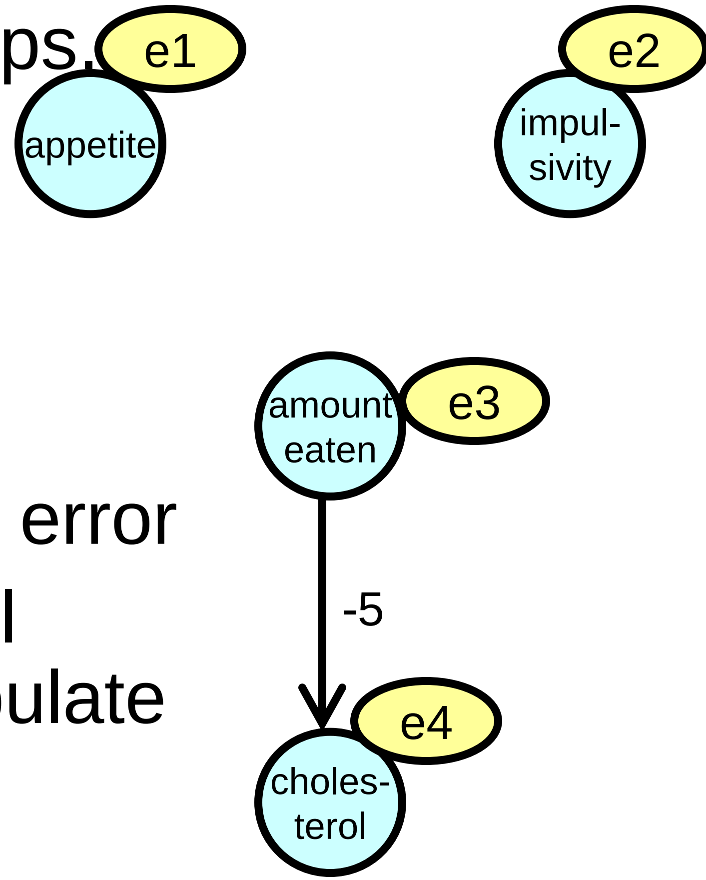
x1 = e1

x2 = e2

x3 = k

x4 = -5 x3 + e4

- linear combination of parents + error
- "causal" means that they model what happens when you manipulate
- e.g. manipulating x3…



Linear SEMs can be cyclic too

- Correspond to dynamical systems:

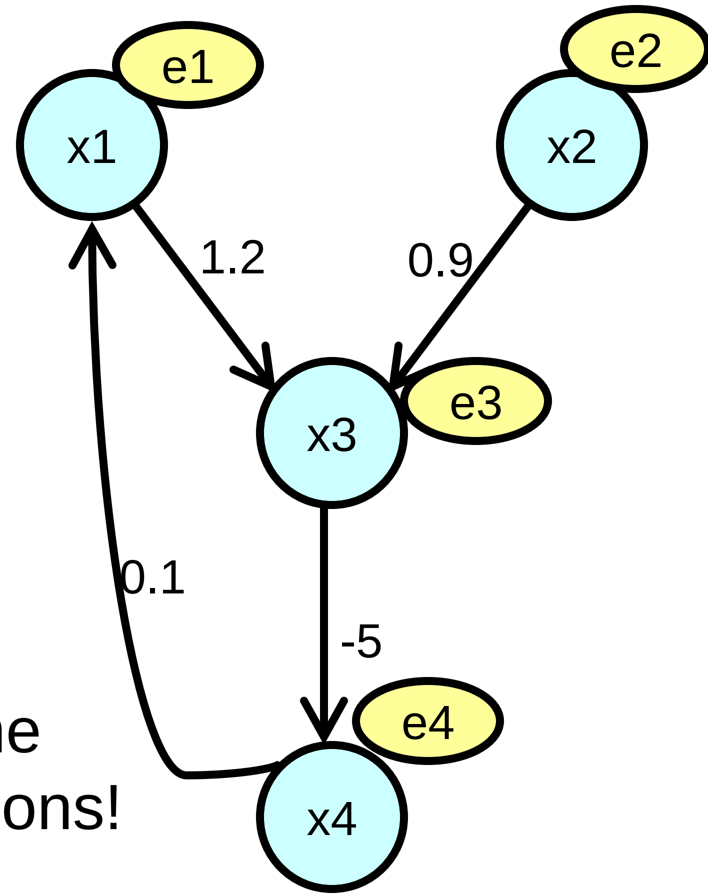
x3[t+1] = 1.2 **x1**[t] + 0.9 **x2**[t] + **e3**
  etc.

- Deterministic

- Reach equilibria (unless unstable)

$x3_{eq} = 1.2\ x1_{eq} + 0.9\ x2_{eq} + e3$

- Equilibrium equations have the same coefficients as the dynamical equations!



Our goal

- Given the equilibrium data **x**, recover the structure of the process, i.e. values of B that entail the observed distribution of equilibria

$$\mathbf{x} = B\,\mathbf{x} + \mathbf{e}$$

Linear SEMs

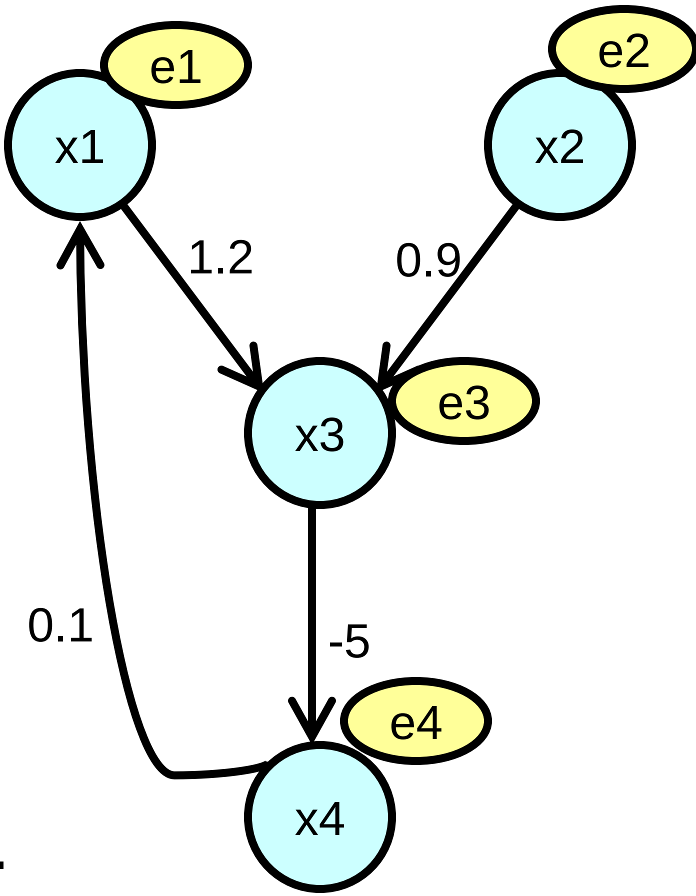
- **Claim:** joint distribution of **e** + the equations → joint distribution of **x.**
- $\mathbf{x} = B\,\mathbf{x} + \mathbf{e}$
- Solving for **x**, we get:

  $$\mathbf{x} = (I - B)^{-1}\,\mathbf{e}$$
- QED.

- Cycles → infinite sums
- Let $A = (I - B)^{-1}$

  then $\mathbf{x} = A\,\mathbf{e}$

  (A is called the “mixing matrix”).



Our assumptions

- Data are equilibria of a linear SEM
- The sample data is i.i.d.
- The error terms have a positive variance
- At most one error term is Gaussian
- Weak Causal Markov: if X, Y causally unconnected, then they are independent
- Causal sufficiency (no hidden confounders)
- "causal faithfulness": the effect of e_i on x_i is not zero.

Causal Faithfulness and violations

- **Definition:** In the equilibrium, the effect of e_i on x_i (reduced-form coefficient) is not 0.

- Acyclic case: never violated

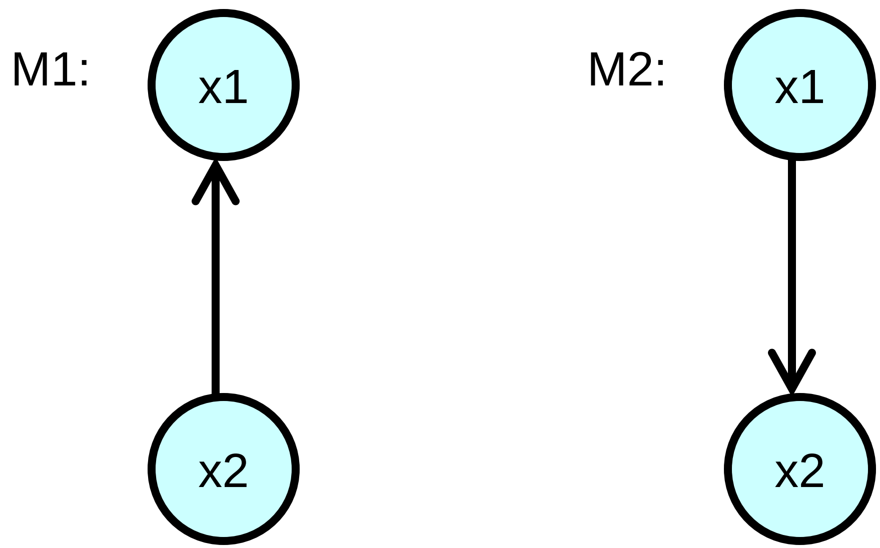
- Cyclic case, violations:
  - e.g. 1: x1[t+1] = 1 * x1[t] + …
  - e.g. 2: a polynomial function of the cycle-products of the SEMs is equal to 1

Outline

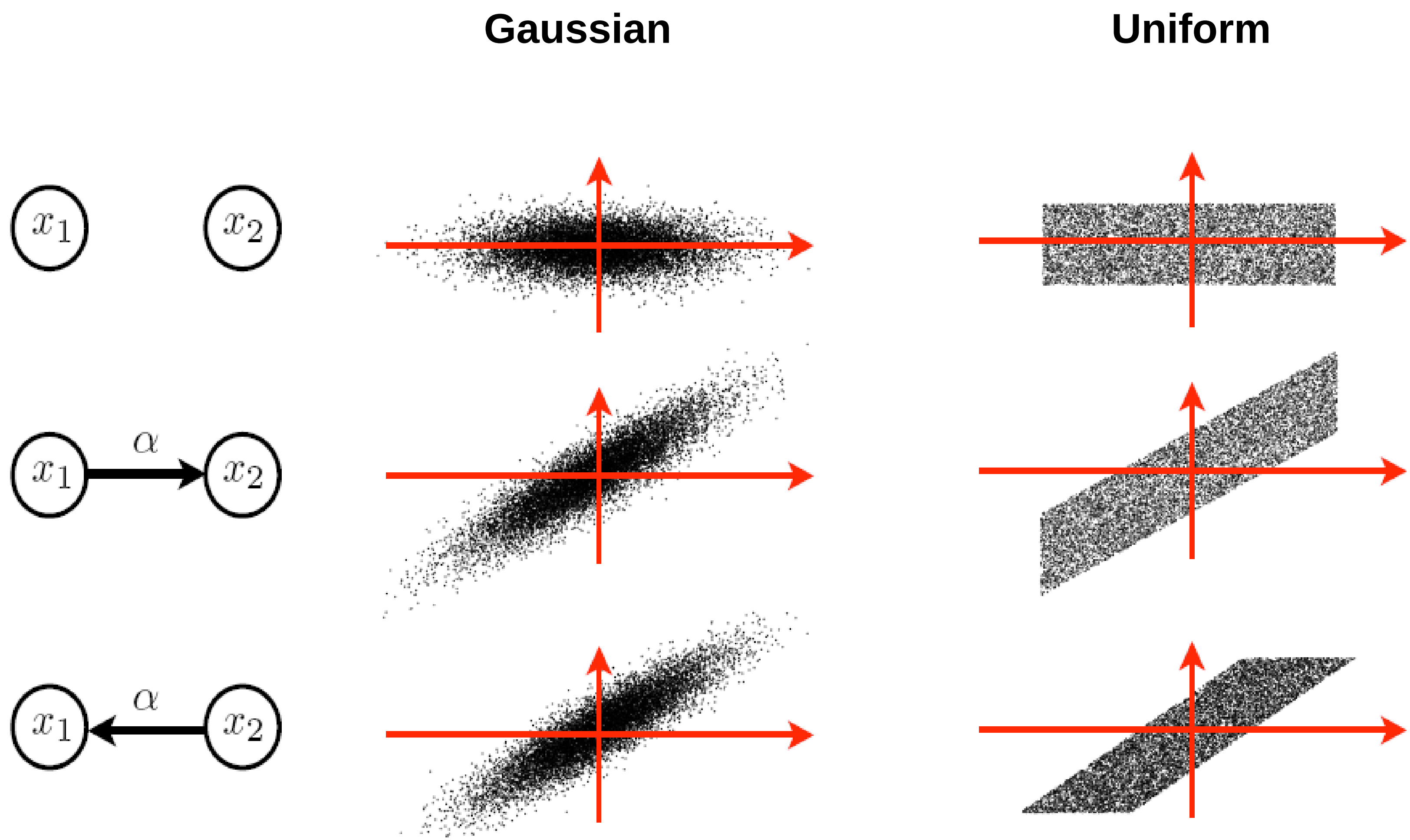
- Linear SEMs 
- LiNGAM method *(Shimizu et al 2006)*: uniquely identifies B for acyclic linear SEMs, by using ICA
- LiNG-D: our new method, based on LiNGAM, but handles cycles
- Now, B is not unique
- But we show that stability can be a powerful constraint

How much can be identified from observational data alone?

- When Gaussian, d-separation equivalence class
- e.g. can't tell the difference between:



Why not?

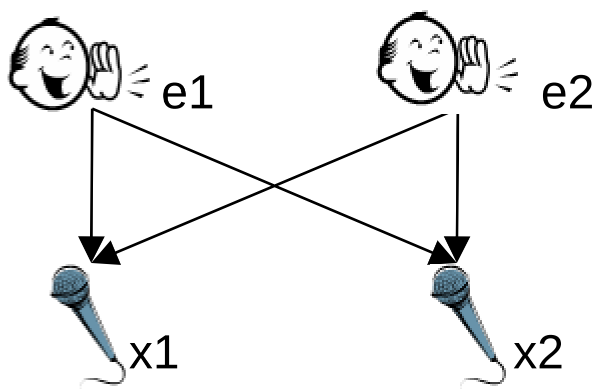


Images by Patrik Hoyer et al, used with permission

from “Estimation of causal effects using linear non-Gaussian causal models with hidden variables”

Independent Components Analysis (ICA)

- Cocktail party problem



$\mathbf{x} = A\,\mathbf{e}$

- You want to get back the original signals, i.e. we need the inverse of the mixing matrix

Let $W = A^{-1}$
Then $\mathbf{e} = W\,\mathbf{x}$
and $W = I - B$

- in general case, there would be infinitely many solutions for A.
- BUT, assuming independence and non-Gaussianity, it is possible to estimate A and **e** from just **x**. This is what ICA does.

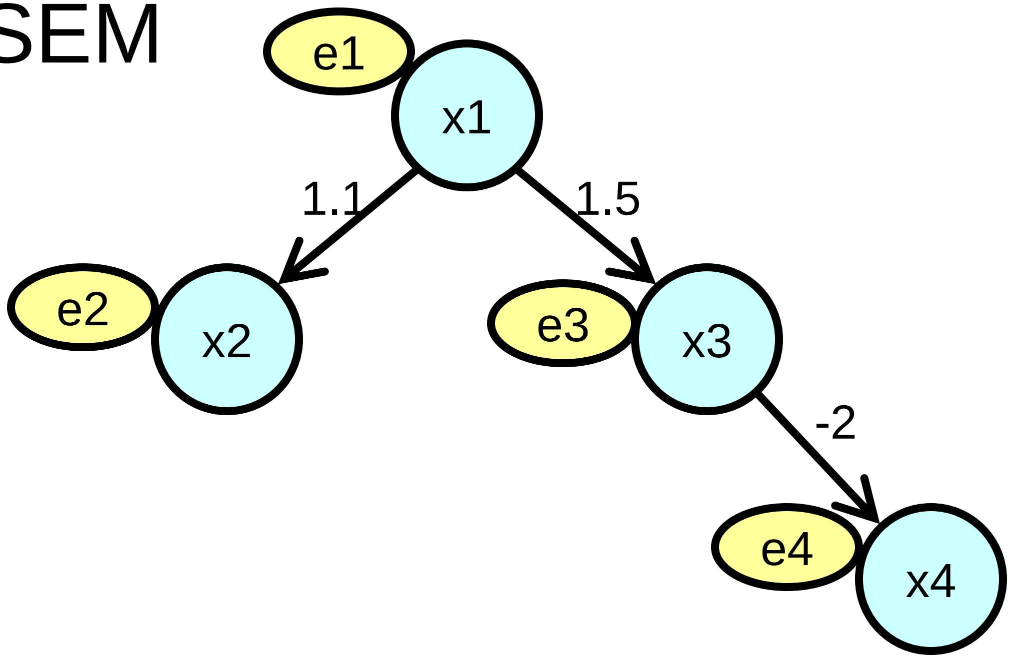
Independent Components Analysis (ICA)

- sources in cocktail-party problem = error terms in SEMs
- Underdetermination: ICA returns W up permutation of error terms

The LiNGAM method

(Shimizu et al, 2006)

- What happens if we generate data from this linear SEM

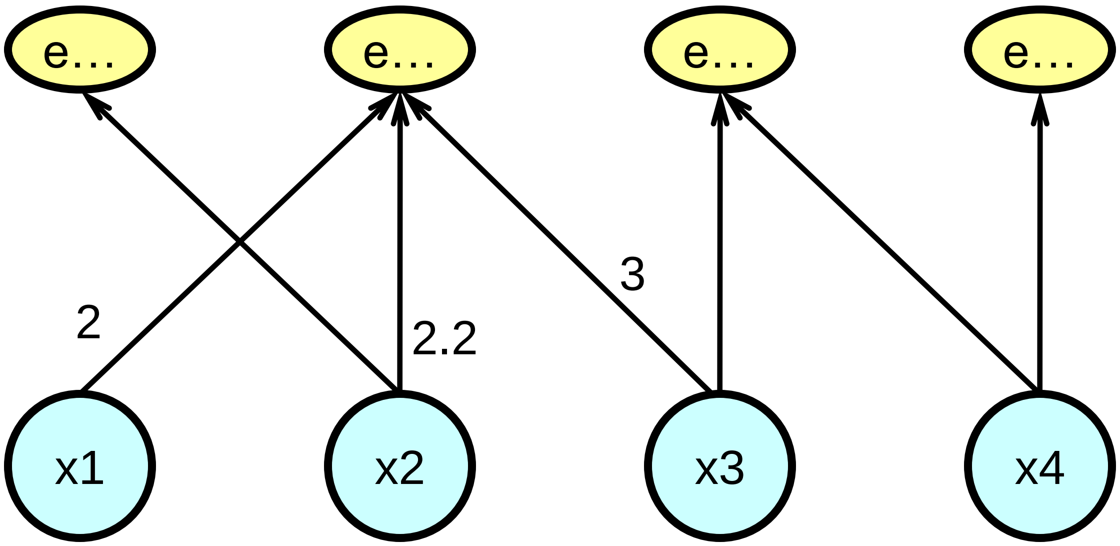


- … and then run ICA?

The LiNGAM method

- ICA returns a *W* matrix such as:

W_{ICA}=

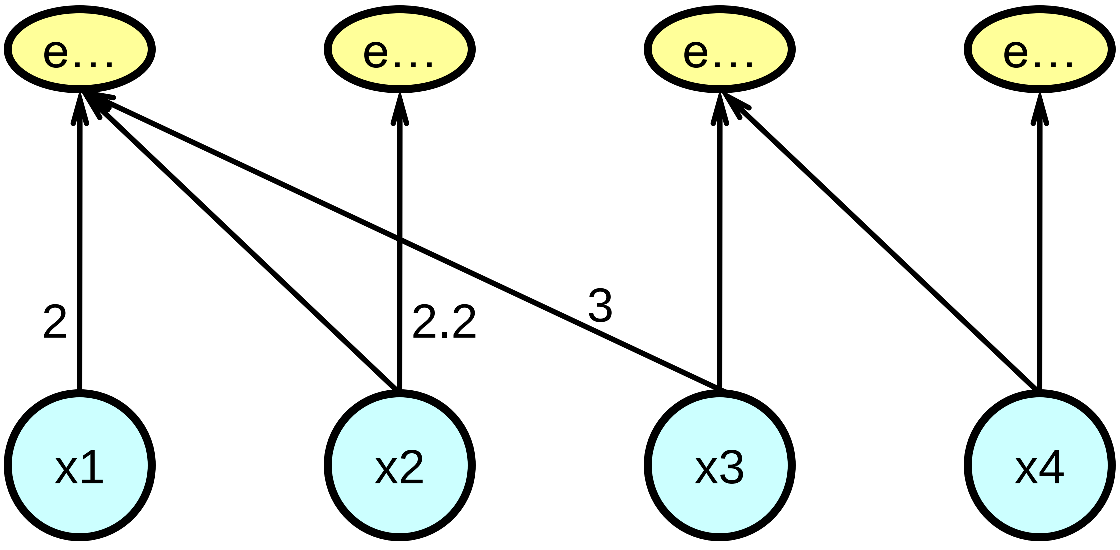


- So **first** we need to find the right permutation of the e's

The LiNGAM method

- So first we need to find the right permutation of the *e*'s

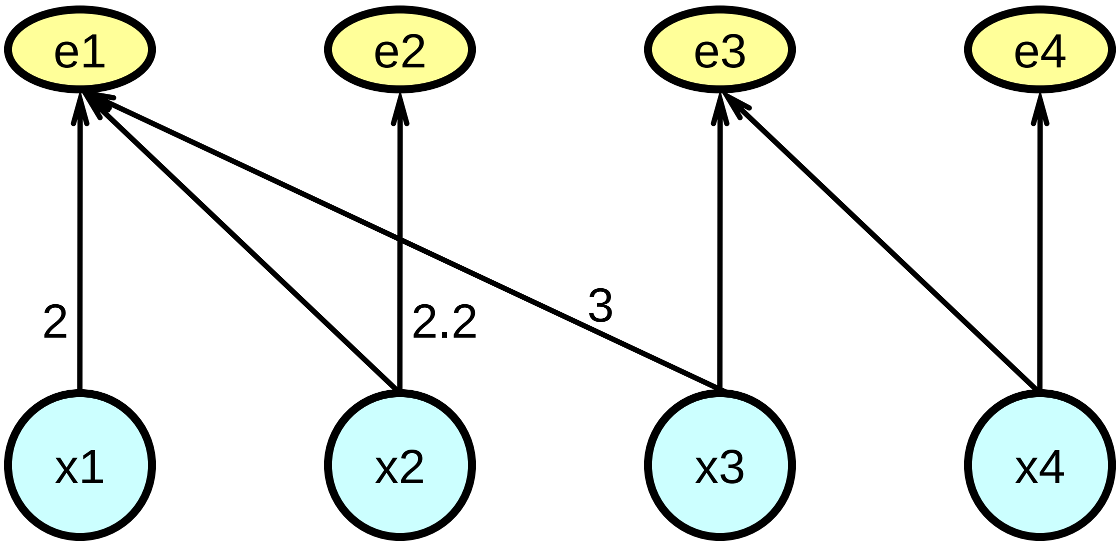
$\tilde{W} =$



The LiNGAM method

- So first we need to find the right permutation of the *e*'s

$\widetilde{W} =$



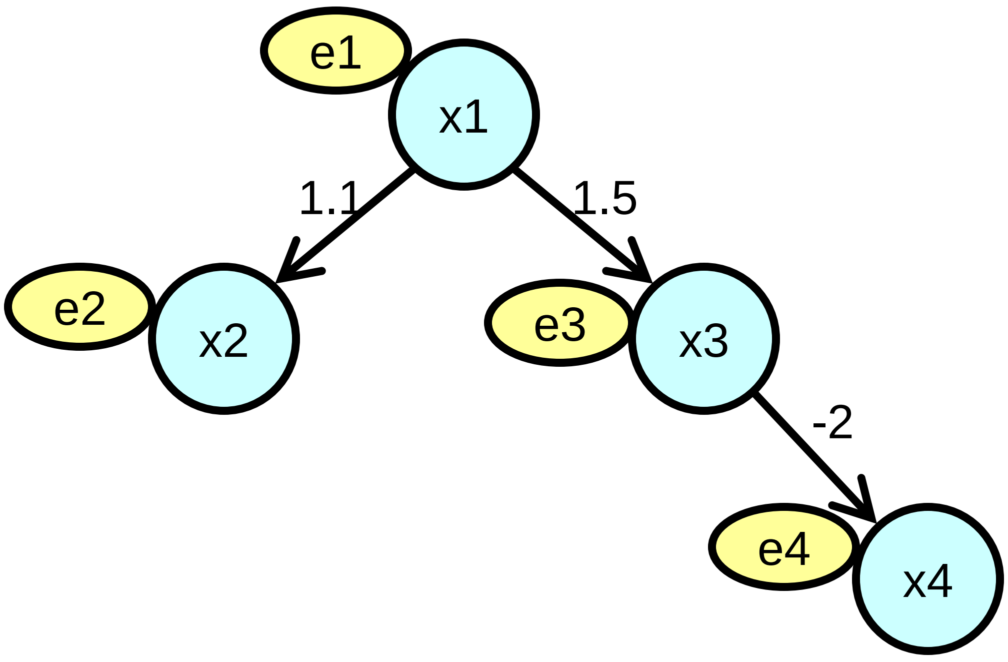
- We label the *e*'s accordingly

The LiNGAM method

- Using the equation $B = I - W'$,

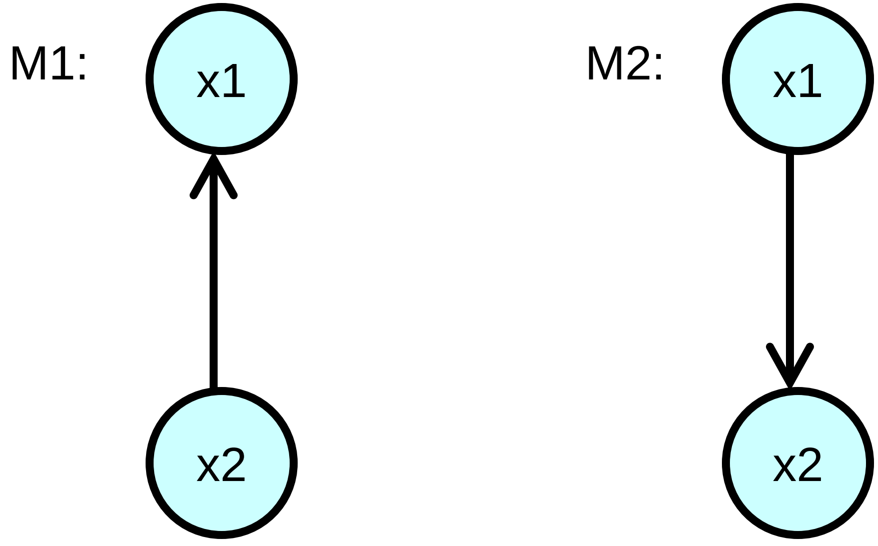
we get back:

B =



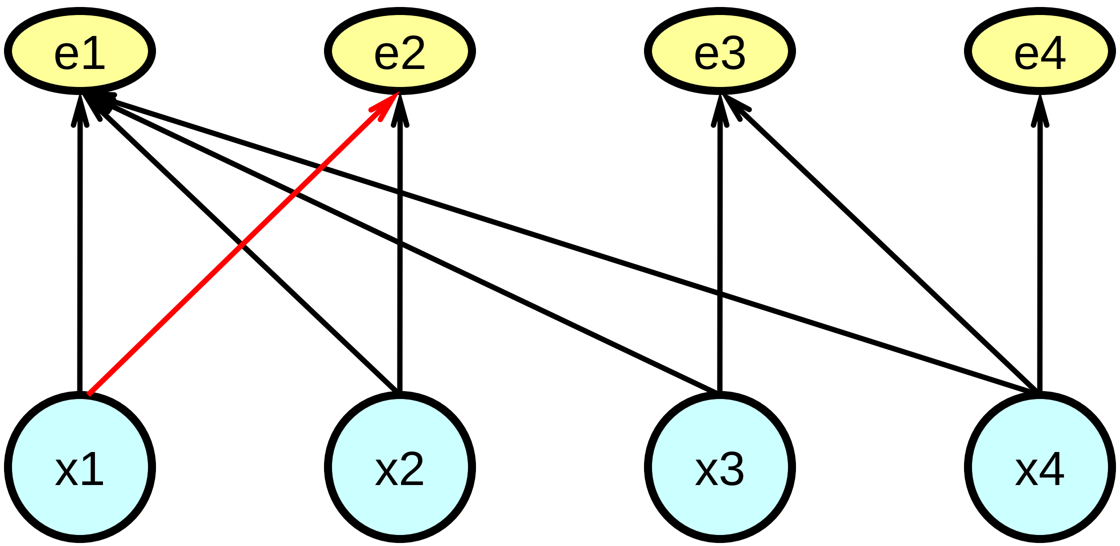
The LiNGAM method

- Discovers the full structure of the DAG
  - causal sufficiency ↔ independence of the error terms
- In particular, now M1 and M2 can be distinguished!



The LiNGAM method – limitation

$W_{ICA}=$



The LiNGAM method – limitation

- LiNGAM cannot discover cyclic models…
- because:
  - since it assumes the data was generated by a DAG,
  - it searches for a single valid permutation, by searching for an ordering

- If we stop imposing an ordering, and search for any number of valid permutations…
- then we can discover cyclic models too.
- That’s exactly what we did!

The LiNG-D method

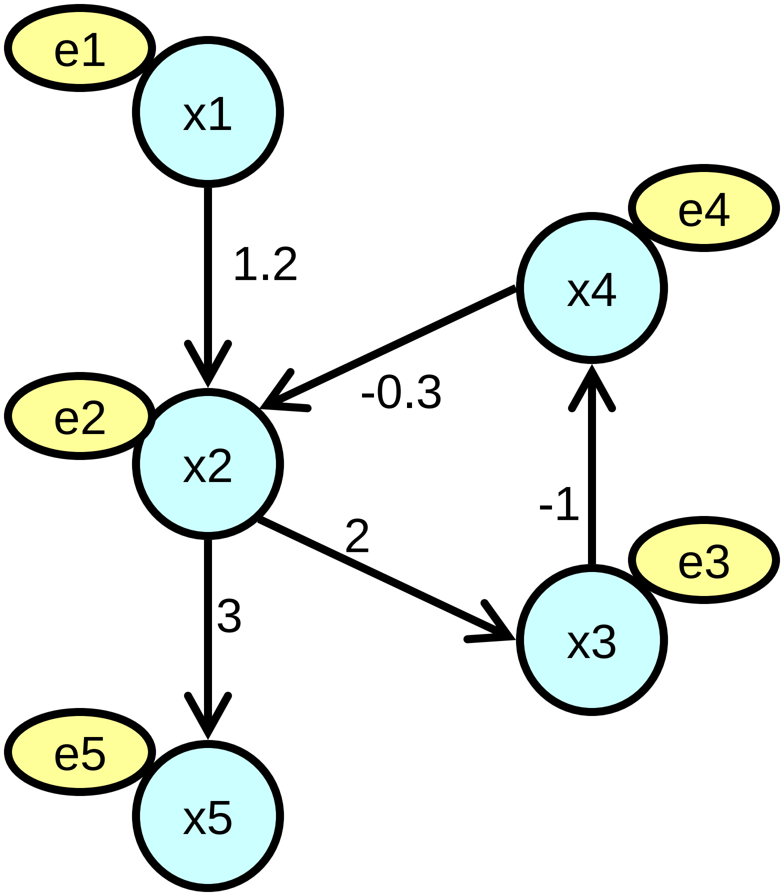
- When the data looks acyclic, it works just like LiNGAM, and returns a single model.
- When the data looks cyclic, more than one permutation is considered valid. Thus, it returns a distribution-equivalent set containing more than one model.
- “distribution-equivalent” means you can’t do better, at least without experimental data or further assumptions.

The LiNG-D method

- Finds (multiple) row-permutations of W_{ICA} that have no zeros in the diagonal
- equivalent to Constrained n-Rooks Problem
- Naïve algorithm: depth-first search
- Better algorithm: set all zeros to 0, all nonzeros to 1, and run *k*-th best assignments algorithm (linear programming) until get a score less than *n*.
- Worst case: *n!* models

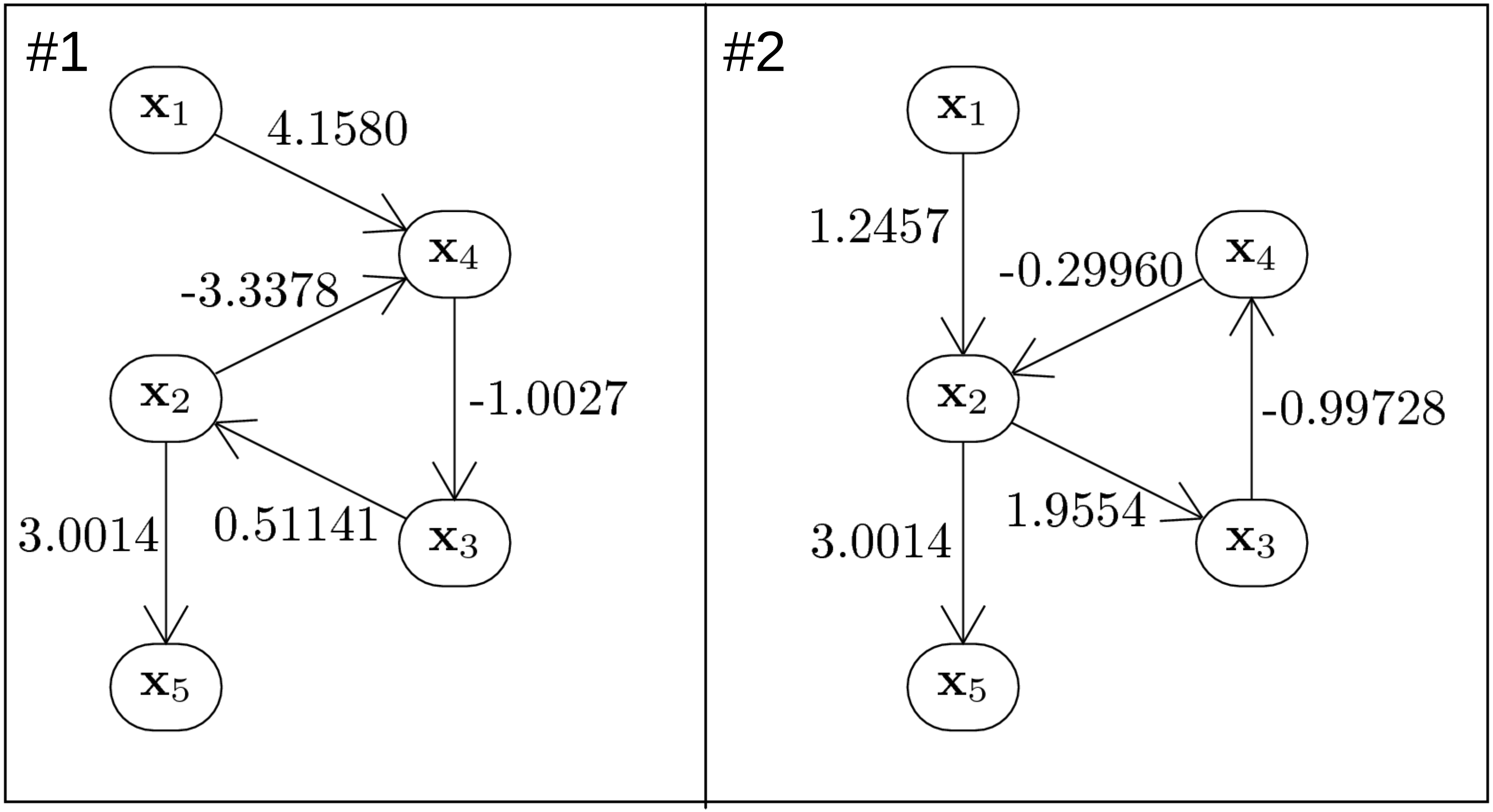
LiNG-D: Demonstration

- Let's simulate using this model:
- Error terms are generated by sampling from a Gaussian and squaring
- 15000 data points
- We prune edges with coefficients < 0.05
- Ready?



LiNG-D: Demonstration

LiNG-D returns a set with 2 models:



LiNG-D + the stability assumption

- Note that only one of these models is stable (assuming no self-loops).
- If our data is a set of equilibria, then the true model must be stable.
- How powerful is this constraint?

LiNG-D + the stability assumption

- **Theorem:** if the true model's cycles don't intersect, then only one model is stable.

- For simple cycle models, cycle-products are inverted: c1 = 1/c2.
- So at least one cycle will be > 1 (in modulus) and thus unstable.
- each cycle works independently, and any valid permutation* will invert at least one cycle, creating an unstable model.

*except for the identity permutation

LiNG-D + the stability assumption

- Real data *(due to Zitian Wang)*: 10-variable model, 28 edges
- LiNG-D tells us that 240 models explain the data equally well. That's too many!
- But only 2 are stable!
- Lesson: stability is a powerful constraint.
- Caveat: this assumes no-self loops (this is a strong assumption!)

What should one use?

	non-Gaussian	Unknown or both or too little data	Gaussian
DAG	LiNGAM unique model	Check out Hoyer et al (2008) (yesterday's poster session)	Constraint-based methods e.g. PC, CPC, SGS (or Geiger and Heckerman 1994 for a Bayesian alternative) d-separation equivalence class
DG	LiNG-D 2 cases acyclic: unique model cyclic: distribution-equivalence class	?	Richardson's CCD very large class: not even covariance equivalent

Take-home message

- LiNGAM exploits non-Gaussianity in acyclic case to find a unique SEM (rather than d-sep equivalence class).
- Similarly, in the cyclic case, LiNG-D narrows the class to a distribution-equivalence class of SEMs.
- Still, there may be multiple SEMs.
- Stability can sometimes be used to rule out a good chunk of those.
- Thank you!

Appendix 1: self-loops

- Equilibrium equations usually correspond with the dynamical equations.

… BUT IF a self-loop has coefficient 1, we will get the wrong structure, and the predicted results of intervention will be wrong!

- self-loop coefficients are underdetermined.

- Our stability results only hold if we assume no self-loops.

Appendix 2: solving n-Rooks efficiently

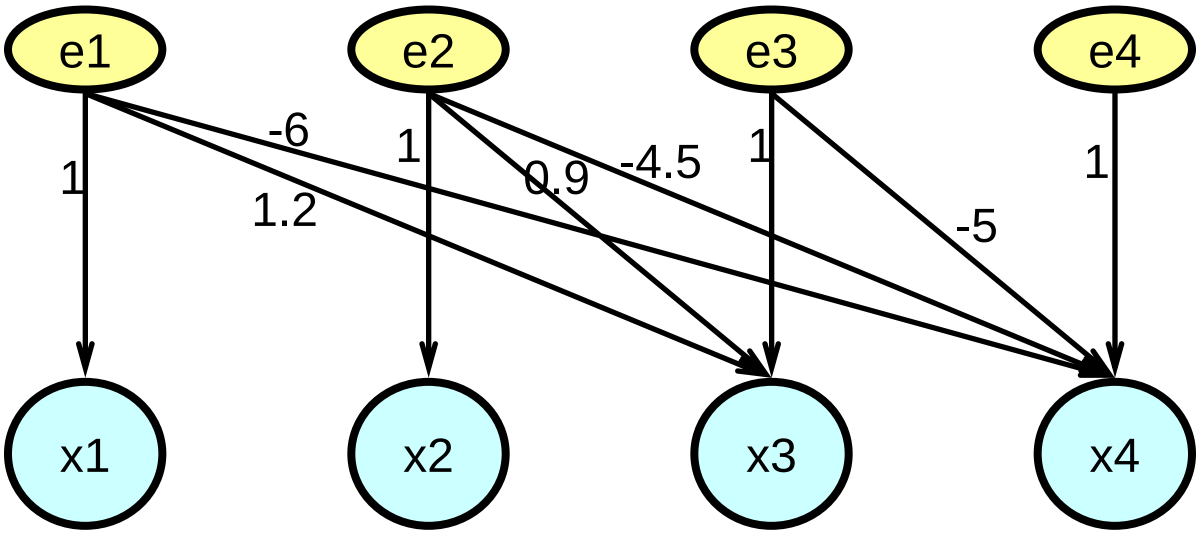
- *W* matrix: use hypothesis tests to turn all zeros to 0, all others to 1.
- Run a *k*-th best assignment algorithm, for increasing *k*, until you reach a suboptimal permutation (all good permutations will score exactly *n*)

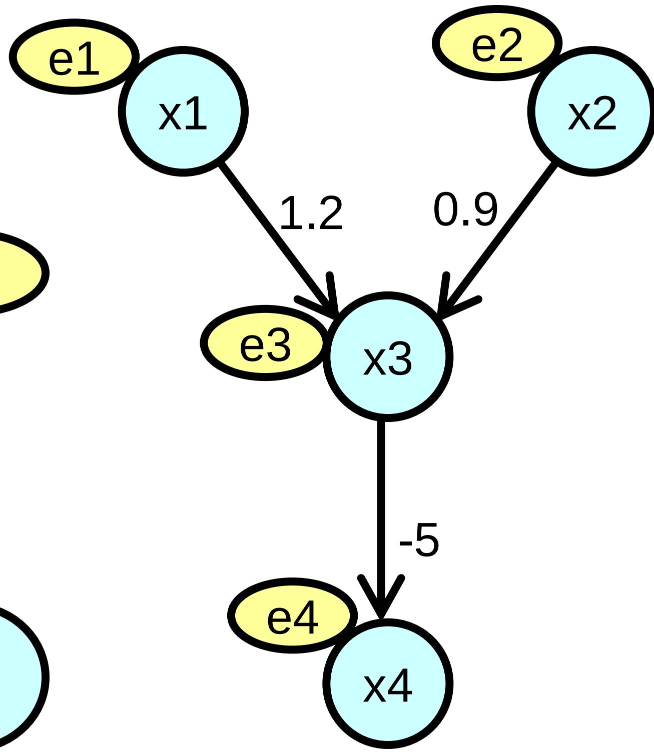
- The time-complexity of this is the same as the *k*-best assignments problem.

Linear Structural Equation Models (SEMs) (with randomness)

- The "mixing matrix" shows how the noise propagates:

Let's make it:





- Done.

Why LiNGAM won't work

- Acyclic: there is a unique permutation of B with no zeros in the diagonal
- LINGAM assigns a score
- Finds an ordering

- In cyclic case, we can't find an ordering, multiple permutations have a zeroless diagonal

- because:
  - since it assumes the data was generated by a DAG,
  - it searches for a single valid permutation