A Constraint Optimization Approach to Causal Discovery from Subsampled Time Series Data

Antti Hyttinen
Joint work with Sergey Plis, Matti Järvisalo, Frederick Eberhardt, David Danks

University of Helsinki, Helsinki Institute for Information Technology
Mind Research Network and University of New Mexico
California Institute of Technology
Carnegie Mellon University

AMBN 2017, Kyoto, Japan
6.9.2016
How to discover the causal structure at the **system timescale** from time series data obtained at a coarser **measurement timescale**?

\[ \ldots \ X^{t-4} \ X^{t-2} \ X^t \ \ldots \ \rightarrow \ \ldots \ \]  
\[ \ldots \ Y^{t-4} \ Y^{t-2} \ Y^t \ \ldots \ \]  
\[ \ldots \ Z^{t-4} \ Z^{t-2} \ Z^t \ \ldots \ \]
Problem Statement

How to discover the causal structure at the **system timescale** from time series data obtained at a coarser **measurement timescale**?

\[
\ldots X^{t-4} \quad X^{t-2} \quad X^t \quad \ldots \quad \ldots\]

\[
\ldots Y^{t-4} \quad Y^{t-2} \quad Y^t \quad \ldots \quad \ldots\]

\[
\ldots Z^{t-4} \quad Z^{t-2} \quad Z^t \quad \ldots \quad \ldots\]

- Only every \( u \):th vector of values is observed (**subsampling rate** \( u \))
Problem Statement

How to discover the causal structure at the **system timescale** from time series data obtained at a coarser **measurement timescale**?

\[
\cdots \; X_{t-4} \; X_{t-2} \; X_t \; \cdots \;
\cdots \; Y_{t-4} \; Y_{t-2} \; Y_t \; \cdots \;
\cdots \; Z_{t-4} \; Z_{t-2} \; Z_t \; \cdots \;
\]

- Only every \( u \):th vector of values is observed (**subsampling rate** \( u \))
- Subsampling induces confounding, and unidentifiability
Problem Statement

How to discover the causal structure at the system timescale from time series data obtained at a coarser measurement timescale?

\[ \cdots \ X_{t-4} \ X_{t-2} \ X_t \ \cdots \ \cdots \ X_{t-1} \ X_t \ \cdots \]

\[ \cdots \ Y_{t-4} \ Y_{t-2} \ Y_t \ \cdots \ \longrightarrow \ \cdots \ Y_{t-1} \ Y_t \ \cdots \]

\[ \cdots \ Z_{t-4} \ Z_{t-2} \ Z_t \ \cdots \ \cdots \ Z_{t-1} \ Z_t \ \cdots \]

- Only every \( u \):th vector of values is observed (subsampling rate \( u \))
- Subsampling induces confounding, and unidentifiability
- Applications: e.g. fMRI.
Subsampling needs to be taken into account!

When ignoring subsampling:

- All direct causal relationships misspecified.
- Wrong result for interventions.
- Wrong interventions suggested.

True structure at the system timescale

measurement time scale structure
Subsampling needs to be taken into account!

When ignoring subsampling:
- All direct causal relationships misspecified.
Subsampling needs to be taken into account!

When ignoring subsampling:

- All direct causal relationships misspecified.
- Wrong result for interventions.
Subsampling needs to be taken into account!

When ignoring subsampling:
- All direct causal relationships misspecified.
- Wrong result for interventions.
- Wrong interventions suggested.

True structure at the system timescale

measurement time scale structure
Overview

1. Previous Literature
2. Graphical Representation
3. A Constraint Satisfaction Solution
4. A Constraint Optimization Solution
5. Conclusion
Previous Literature
• Adding instantaneous effects in a linear model (see for example Lütkepohl 2005 or Hyvärinen et al 2010).
• Adding instantaneous effects in a linear model (see for example Lütkepohl 2005 or Hyvärinen et al 2010).

• Continuous time approaches, but some processes are inherently discrete time (e.g. salary payment).
Recently Plis et al. (UAI2015,NIPS2015) considered modeling subsampling directly, assuming on the system timescale level:

- discrete time
- first order Markov: $\mathbf{V}^t \perp \perp \mathbf{V}^{t-k} | \mathbf{V}^{t-1}$
- no instantaneous effects, or unobserved common causes
- nonparametric (continuous or discrete values, SVAR processes, or dynamic BNs)
- Measurements from this at integer intervals (e.g. every second).
Recently Plis et al. (UAI2015,NIPS2015) considered modeling subsampling directly, assuming on the system timescale level:

- discrete time
- first order Markov: $\mathbf{V}^t \perp \mathbf{V}^{t-k} | \mathbf{V}^{t-1}$
- no instantaneous effects, or unobserved common causes
- nonparametric (continuous or discrete values, SVAR processes, or dynamic BNs)
- Measurements from this at integer intervals (e.g. every second).

Corresponding parametric method: Gong et al. (ICML2015) discovered linear models using non-Gaussianity.
Graphical Representation
Rolled Representation

system t.s

... $X^{t-2}$ $X^{t-1}$ $X^t$ ...

... $Y^{t-2}$ $Y^{t-1}$ $Y^t$ ...

... $Z^{t-2}$ $Z^{t-1}$ $Z^t$ ...

unrolling

measurement t.s.
Rolled Representation

system t.s

\[
\begin{align*}
\cdots & \quad X^{t-2} \quad X^{t-1} \quad X^t \quad \cdots \\
\cdots & \quad Y^{t-2} \quad Y^{t-1} \quad Y^t \quad \cdots \\
\cdots & \quad Z^{t-2} \quad Z^{t-1} \quad Z^t \quad \cdots
\end{align*}
\]

marginalization

measurement t.s.

\[
\begin{align*}
\cdots & \quad X^{t-2} \quad X^t \quad \cdots \\
\cdots & \quad Y^{t-2} \quad Y^t \quad \cdots \\
\cdots & \quad Z^{t-2} \quad Z^t \quad \cdots
\end{align*}
\]

unrolling

\[
\begin{align*}
X \quad & \quad \downarrow \quad \uparrow \\
Z \quad & \quad \downarrow \\
Y
\end{align*}
\]
Rolled Representation

system t.s

\[ \cdots \xrightarrow{\downarrow} X^{t-2} \xrightarrow{\downarrow} X^{t-1} \xrightarrow{\downarrow} X^t \xrightarrow{\downarrow} \cdots \]

\[ \cdots \xrightarrow{\downarrow} Y^{t-2} \xrightarrow{\downarrow} Y^{t-1} \xrightarrow{\downarrow} Y^t \xrightarrow{\downarrow} \cdots \]

\[ \cdots \xrightarrow{\downarrow} Z^{t-2} \xrightarrow{\downarrow} Z^{t-1} \xrightarrow{\downarrow} Z^t \xrightarrow{\downarrow} \cdots \]

marginalization

measurement t.s.

\[ \cdots \xrightarrow{\downarrow} X^{t-2} \xrightarrow{\downarrow} X^t \xrightarrow{\downarrow} \cdots \]

\[ \cdots \xrightarrow{\downarrow} Y^{t-2} \xrightarrow{\downarrow} Y^t \xrightarrow{\downarrow} \cdots \]

\[ \cdots \xrightarrow{\downarrow} Z^{t-2} \xrightarrow{\downarrow} Z^t \xrightarrow{\downarrow} \cdots \]

unrolling

\[ \xrightarrow{\leftarrow} X \xrightarrow{\uparrow} \]

\[ Z \xrightarrow{\rightarrow} \]

\[ Y \xrightarrow{\rightarrow} \]

rolling

\[ \xrightarrow{\rightarrow} X \xrightarrow{\downarrow} \]

\[ Z \xrightarrow{\rightarrow} \]

\[ Y \xrightarrow{\rightarrow} \]

\( \text{unrolling} \)

\( \text{rolling} \)
Induced confounding

system t.s.

\[ \ldots \rightarrow X^{t-1} \rightarrow X^{t} \rightarrow \ldots \]
\[ \ldots \rightarrow Y^{t-1} \rightarrow Y^{t} \rightarrow \ldots \]
\[ \ldots \rightarrow Z^{t-1} \rightarrow Z^{t} \rightarrow \ldots \]

unrolling

marginalization

measurement t.s.

\[ \ldots \rightarrow X^{t-2} \rightarrow X^{t} \rightarrow \ldots \]
\[ \ldots \rightarrow Y^{t-2} \rightarrow Y^{t} \rightarrow \ldots \]
\[ \ldots \rightarrow Z^{t-2} \rightarrow Z^{t} \rightarrow \ldots \]

rolling
Induced confounding

system t.s

\[ \ldots X^{t-2} \rightarrow X^{t-1} \rightarrow X^t \rightarrow \ldots \]

\[ \ldots Y^{t-2} \rightarrow Y^{t-1} \rightarrow Y^t \rightarrow \ldots \]

\[ \ldots Z^{t-2} \rightarrow Z^{t-1} \rightarrow Z^t \rightarrow \ldots \]

unrolling

marginalization

measurement t.s.

\[ \ldots X^{t-2} \rightarrow X^t \rightarrow \ldots \]

\[ \ldots Y^{t-2} \rightarrow Y^t \rightarrow \ldots \]

\[ \ldots Z^{t-2} \rightarrow Z^t \rightarrow \ldots \]

rolling

\[ X \]

\[ Z \rightarrow Y \]

\[ X \rightarrow \uparrow \] (Task 1)
Correspondence between System and Measurement T.S.

When subsampling by $u$:

- Measurement time scale edge $Y \rightarrow X$ corresponds to path of length $u$: $Y \rightarrow \cdots \rightarrow X$.

- Measurement time scale edge $X \leftrightarrow Y$ corresponds to paths of length $k < u$: $W \rightarrow \cdots \rightarrow X$ and $W \rightarrow \cdots \rightarrow Y$.
A Constraint Satisfaction Solution
Result: Deciding whether there is a system t.s. structure compatible with the directed edges of a measurement t.s. structure is **NP-complete** for any fixed $u \geq 2$.

Proof: Binary matrix root.
A Constraint Satisfaction Solution

- You write a symbolic encoding.
A Constraint Satisfaction Solution

- You write a symbolic encoding.
- The symbolic encoding gets grounded.
- The encoding gets turned into conjunctive normal form.
- Backtracking DFS by Clingo (Gebser et al. 2011).
A Constraint Satisfaction Solution

- You write a symbolic encoding.
- The symbolic encoding gets grounded.
- The encoding gets turned into conjunctive normal form.
- Backtracking DFS by Clingo (Gebser et al. 2011).
- Exact and complete solution.
- Subsampling rate $u$: fixed or free.

https://srlabs.de/bites/minisat-intro/
node(1..3). % Measurement timescale structure
dgeh(1,2).no_edgeh(1,3).confh(2,3).no_confh(1,2). %and so on

urange(1..5). % Define a range of u:s
1 { u(U): urange(U) } 1. % u(U) is true for only one U

{ edge1(X,Y) } :- node(X), node(Y). %draw G1

% Derive all directed paths up to length U
path(X,Y,1) :- edge1(X,Y).
path(X,Y,L) :- path(X,Z,L-1), edge1(Z,Y), L <= U, u(U).

% Determine measurement t.s. for G1
dgeu(X,Y) :- path(X,Y,L), u(L).
confu(X,Y) :- path(Z,X,L), path(Z,Y,L), node(X;Y;Z),
            X < Y, L < U, u(U).

% Check consistency
:- edgeh(X,Y), not dgeu(X,Y). :- no_edgeh(X,Y), edgeu(X,Y).
:- confh(X,Y), not confu(X,Y). :- no_confh(X,Y), confu(X,Y).
Scalability of Enumerating 1000 Solutions

( fixed subsampling rate 2, SAT is our approach, MSL is the previous state of art by Plis et al. (2015) )
Identifiability: Underdetermination

Measurement timescale structure:

\[ \begin{align*}
X & \rightarrow Y \\
Z & \rightarrow W
\end{align*} \]

could be produced by system timescale structures:

\[ \begin{align*}
X & \rightarrow Y \\
Z & \rightarrow W
\end{align*} \]

or a four cycle in either direction and symmetrically!

\[ u = 1, 2, 3, \ldots \quad u = 2, 4, 6, \ldots \quad u = 3, 6, 9, \ldots \]
But measurement timescale structure:

uniquely identifies system timescale structure

and the subsampling rate $u = 2$. 
A Constraint Optimization Solution
Task 2: Finding Structures Compatible with Data

... $X^{t-4}$ $X^{t-2}$ $X^t$ ...

... $Y^{t-4}$ $Y^{t-2}$ $Y^t$ ...

... $Z^{t-4}$ $Z^{t-2}$ $Z^t$ ...

data measurement t.s. system t.s.

$X \rightarrow Z \rightarrow Y$ $X \rightarrow Z \rightarrow Y$
Task 2: Finding Structures Compatible with Data

\[ \cdots X^{t-4} \quad X^{t-2} \quad X^t \quad \cdots \]

\[ \cdots Y^{t-4} \quad Y^{t-2} \quad Y^t \quad \cdots \rightarrow \]

\[ \cdots Z^{t-4} \quad Z^{t-2} \quad Z^t \quad \cdots \]

- Measurement t.s. structure can be consistently estimated from data under faithfulness: e.g.
  \[ X \to Z \iff X^{t-u} \perp Z^t \mid V^{t-u} \setminus X^{t-u} \]
  \[ X \leftrightarrow Z \iff X^t \not\perp Y^t \mid V^{t-u} \]
Task 2: Finding Structures Compatible with Data

\[
\cdots X^{t-4} \ X^{t-2} \ X^t \ \cdots \\
\cdots Y^{t-4} \ Y^{t-2} \ Y^t \ \cdots \\
\cdots Z^{t-4} \ Z^{t-2} \ Z^t \ \cdots
\]

\[
\begin{array}{llll}
\rightarrow & X & \downarrow & Z \\
& Z & \rightarrow & Y \\
& Y & \leftarrow & X
\end{array}
\]

\[
\begin{array}{llll}
\rightarrow & X & \downarrow & Z \\
& Z & \rightarrow & Y \\
& Y & \leftarrow & X
\end{array}
\]

data measurement t.s. system t.s.

- Measurement t.s. structure can be consistently estimated from data under faithfulness: e.g.
  \[
  X \rightarrow Z \quad \iff \quad X^{t-u} \not\perp \perp Z^t \mid V^{t-u} \setminus X^{t-u}
  \]
  \[
  X \leftrightarrow Z \quad \iff \quad X^t \not\perp \perp Y^t \mid V^{t-u}
  \]

- Due to finite samplesize, the constraint satisfaction approach will often return **UNSATISFIABLE**.
Task 2: Finding Structures Compatible with Data

\[ \cdots X^{t-4} \quad X^{t-2} \quad X^t \quad \cdots \]

\[ \cdots Y^{t-4} \quad Y^{t-2} \quad Y^t \quad \cdots \]

\[ \cdots Z^{t-4} \quad Z^{t-2} \quad Z^t \quad \cdots \]

\[ \text{data} \quad \text{measurement t.s.} \quad \text{system t.s.} \]

- Measurement t.s. structure can be consistently estimated from data under faithfulness: e.g.
  \[ X \rightarrow Z \iff X^{t-u} \not\perp\!\!\!\perp Z^t \mid V^{t-u} \backslash X^{t-u} \]
  \[ X \leftrightarrow Z \iff X^t \not\perp\!\!\!\perp Y^t \mid V^{t-u} \]

- Due to finite samplesize, the constraint satisfaction approach will often return **UNSATISFIABLE**.

- Find the system t.s. structure such that its measurement t.s. structure is optimally close to the estimated (Task 2).
Specifics:

- Penalize inconsistencies between absences and precences of edges in the measurement t.s.:
  - Either uniform weights, or
  - Log Bayesian probabilities of the corresponding (in)dependence, obtained through Bayesian model selection (see Hyttinen et al. 2014)
  - Objective function is the sum of the penalties
Specifics:

- Penalize inconsistencies between absences and precences of edges in the measurement t.s.:
  - Either uniform weights, or
  - log Bayesian probabilities of the corresponding (in)dependence, obtained through Bayesian model selection (see Hyttinen et al. 2014)
  - Objective function is the sum of the penalities
- Clingo uses Branch-and-Bound search to find the exact weighted Maximum Satisfiability solution.
Specifics:

- Penalize inconsistencies between absences and precences of edges in the measurement t.s.:
  - Either uniform weights, or
  - log Bayesian probabilities of the corresponding (in)dependence, obtained through Bayesian model selection (see Hyttinen et al. 2014)
  - Objective function is the sum of the penalities

- Clingo uses Branch-and-Bound search to find the exact weighted Maximum Satisfiability solution.

- We scale to 11-12 within 10 minutes, depending on the sample size and other specifics
A Constraint Optimization Solution

Specifics:

- Penalize inconsistencies between absences and precences of edges in the measurement t.s.:
  - Either uniform weights, or
  - log Bayesian probabilities of the corresponding (in)dependence, obtained through Bayesian model selection (see Hyttinen et al. 2014)
- Objective function is the sum of the penalities

- Clingo uses Branch-and-Bound search to find the exact weighted Maximum Satisfiability solution.
- We scale to 11-12 within 10 minutes, depending on the sample size and other specifics
- Previous work by Plis et al. 2015.
Accuracy for fixed $u = 2$

( fixed subsampling rate 2, average result of the eq. class, 6 nodes, av. degree 3, 200 samples, 100 data sets, linear models )
Accuracy for $u = 3$
• Hourly measurements of six sensors placed in a house.
• Temperature and humidity recorded.
• Removed trends.
• Handle undetermination: for each edge [Magliacane et al.]
  • run the inference procedure enforcing presence
  • and then enforcing absence
  • difference in objectives gives the support for the edge.
Analysis of Temperature/Humidity data 2

Edges with full lines are found to be present, absent edges are found to be absent, edges with dotted lines are present or absent.
Conclusion
Causal discovery from subsampled time series data:
Causal discovery from subsampled time series data:

- A non-parametric constraint satisfaction approach:
  Much better scalability than previous state-of-the-art.
Conclusion

Causal discovery from subsampled time series data:

- A non-parametric constraint satisfaction approach:
  Much better scalability than previous state-of-the-art.

- A (first) constraint optimization approach:
  More accurate than unweighted or unoptimal solutions.
Causal discovery from subsampled time series data:
- A non-parametric constraint satisfaction approach:
  Much better scalability than previous state-of-the-art.
- A (first) constraint optimization approach:
  More accurate than unweighted or unoptimal solutions.
- Future work: generalizing the model space, e.g. allowing for unobserved confounding time series.
Conclusion

Causal discovery from subsampled time series data:

- A non-parametric constraint satisfaction approach: Much better scalability than previous state-of-the-art.
- A (first) constraint optimization approach: More accurate than unweighted or unoptimal solutions.
- Future work: generalizing the model space, e.g. allowing for unobserved confounding time series.

Thanks!