Learning Causal Networks via Additive Faithfulness

Kuang-Yao Lee
Yale School of Public Health

(Joint work with Tianqi Liu, Bing Li, and Hongyu Zhao)

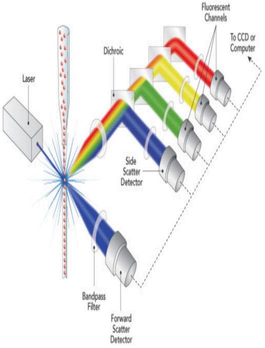
December 12, 2015

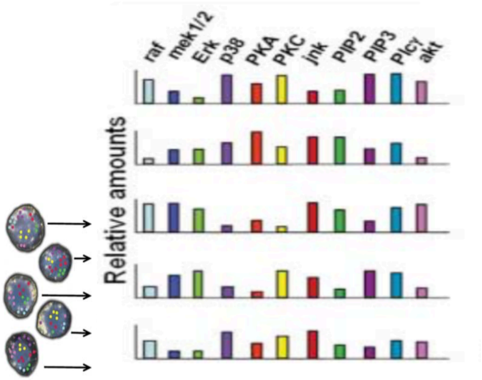
Outline

- Motivating data set
- Additively faithful directed acyclic graph
- Estimation, consistency, and examples

Flow cytometry

- Monitor single cell
- Measure multiple protein levels simultaneously

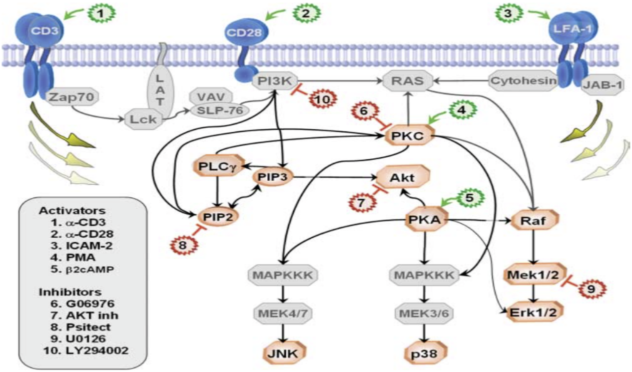




Sachs et al. (2015); `http://www.semrock.com/flow-cytometry.aspx`

Human primary naive $CD4^+$ T cells

- 11 protein levels measured on 7466 cells
- $\boxed{\text{How to recover the signaling pathways?}}$
- A conventionally accepted network:



Sachs et al. (2015)

Directed acyclic graph (DAG)

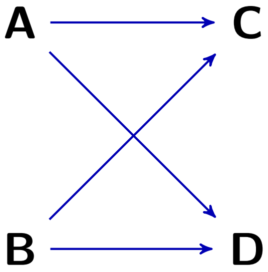
- $\mathsf{G} = \{\mathsf{V}, \mathsf{E}\}$; $\mathsf{V} = \{1, \ldots, p\}$ nodes;
  $\mathsf{E} \subseteq \{(i, j) \in \mathsf{V} \times \mathsf{V},\ i \neq j\}$ directed edges; no cycles
- $X = (X_1, \ldots, X_p)^\top$ is Markovian w.r.t. G if,

  for any $(i, j) \in \mathsf{E}$ and any subset $\mathsf{S} \subseteq V \setminus \{i, j\}$,

  $$\boxed{i \text{ and } j \text{ are } d\text{-separated by } \mathsf{S} \Rightarrow X_i \perp\!\!\!\perp X_j | X_{\mathsf{S}}}$$

Lauritzen (1996)

Markov property



$$
\begin{array}{ccc}
A \perp_d B & & X_A \perp\!\!\!\perp X_B \\
 & \Longrightarrow & \\
C \perp_d D | A, B & & X_C \perp\!\!\!\perp X_D | X_A, X_B
\end{array}
$$

Spirtes et al. (2000)

Identifiability

- For observational data: not possible
  - Equivalent classes, e.g.

$$
\begin{tikzcd} \mathbf{A} \arrow[d] & \\ \mathbf{C} \arrow[r] & \mathbf{B} \end{tikzcd},
\begin{tikzcd} \mathbf{A} & \\ \mathbf{C} \arrow[u] & \mathbf{B} \arrow[l] \end{tikzcd},
\begin{tikzcd} \mathbf{A} & \\ \mathbf{C} \arrow[u] \arrow[r] & \mathbf{B} \end{tikzcd}
$$

  - Same d-separation
- Global search:
  - 11 nodes: 31,603,459,396,418,917,607,425 DAGs
  - p nodes: $\approx p! \cdot 2^{\binom{p}{2}}$

$\boxed{\text{Can we infer DAG from local structures?}}$

Chickering (2003); van der Geer and Buhlmann (2013); Peters and Bühlmann (2014); `http://oeis.org/A003024/list`

Faithfulness

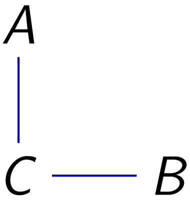
Conditional indepedence also implies *d*-separation

$$
\begin{tikzcd}
\mathbf{A} \arrow[r] \arrow[dr] & \mathbf{C} \\
\mathbf{B} \arrow[r] \arrow[ur] & \mathbf{D}
\end{tikzcd}
$$

$$
\begin{array}{c} A \perp_d B \\ C \perp_d D | A, B \end{array} \iff \begin{array}{c} X_A \perp\!\!\!\perp X_B \\ X_C \perp\!\!\!\perp X_D | X_A, X_B \end{array}
$$

Spirtes et al. (2000)

Characterizing conditional independence



- Under Gaussianity:

$$X_A \perp\!\!\!\perp X_B | X_C \;\Leftrightarrow \mathrm{cor}(X_A, X_B | X_C) = 0$$

  - partial correlation test (PC-algorithm)
- Fully non-parametric? curse of dimensionality!

New criteria like $\perp\!\!\!\perp$ but without multivariate kernels?

Kalisch and Bühlmann (2007); Hoyer et al. (2009); Mooij et al. (2009); Peters et al. (2014)

Outline

- Motivating data set
- Additively faithful directed acyclic graph
- Estimation, consistency, and examples

Additive reproducing kernel Hilbert space

- $\kappa(\cdot,\cdot)$: a positive definite kernel
- $\mathscr{A}_j$, *reproducing kernel Hilbert space* (RKHS) of X_j: space spanned by $\{\kappa(\cdot,x_j^1),\ldots,\kappa(\cdot,x_j^n)\}$

Definition ($\bigoplus\mathscr{A}_j$, direct sum)

$$\bigoplus\mathscr{A}_j \triangleq \sum_{j=1}^{p}\mathscr{A}_j = \left\{\sum_{j=1}^{p} f_j : f_j \in \mathscr{A}_j\right\},$$

with inner product

$$\langle\sum_{j=1}^{p} f_j, \sum_{j=1}^{p} g_j\rangle \triangleq \sum_{j=1}^{p}\langle f_j, g_j\rangle$$

Additive conditional independence (ACI)

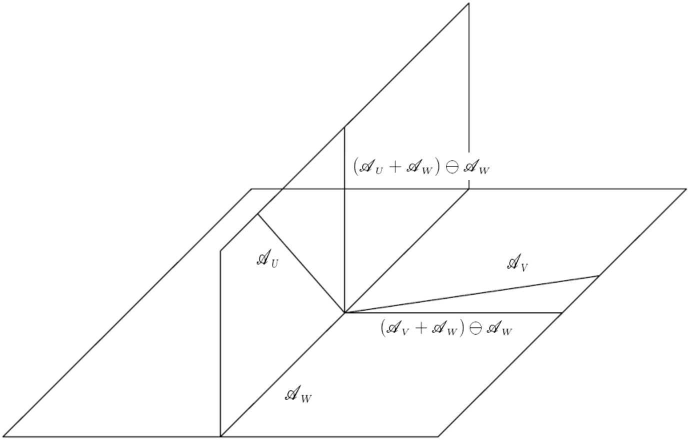
Definition ($U \perp\!\!\!\perp_{\mathscr{A}} V | W$)

(U, V, W), subvectors of X;
U and V are additively independent given W if

$$(\mathscr{A}_U + \mathscr{A}_W) \cap \mathscr{A}_W^{\perp} \perp (\mathscr{A}_V + \mathscr{A}_W) \cap \mathscr{A}_W^{\perp},$$

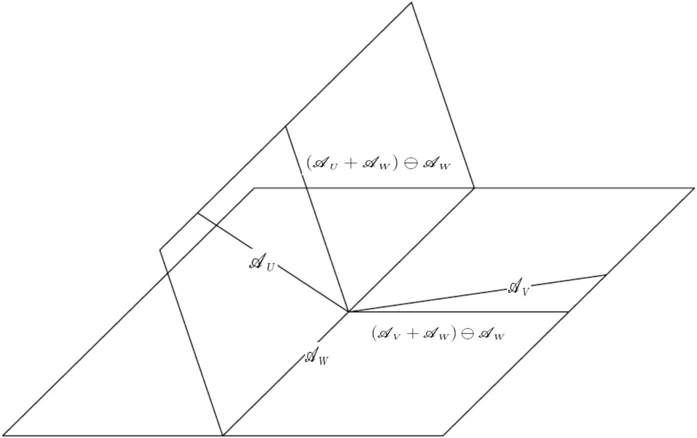
where $\perp$ is in terms of L_2-inner product.

ACI



$(\mathscr{A}_U + \mathscr{A}_W) \ominus \mathscr{A}_W$
$\mathscr{A}_U$
$\mathscr{A}_V$
$(\mathscr{A}_V + \mathscr{A}_W) \ominus \mathscr{A}_W$
$\mathscr{A}_W$

non-ACI


$(\mathscr{A}_U + \mathscr{A}_W) \ominus \mathscr{A}_W$
$\mathscr{A}_U$
$\mathscr{A}_V$
$(\mathscr{A}_V + \mathscr{A}_W) \ominus \mathscr{A}_W$
$\mathscr{A}_W$

Semi-graphoid

Theorem (Li, Chun, and Zhao, 2014)

"$\perp\!\!\!\perp_{\mathscr{A}}$" is a semi-graphoid.

Semi-graphoid axioms

1. symmetry:
$$\begin{tikzcd} \mathbf{A} \arrow[d, dash] & & & \mathbf{B} \arrow[d, dash] & \\ \mathbf{C} \arrow[r, dash] & \mathbf{B} & \Longrightarrow & \mathbf{C} \arrow[r, dash] & \mathbf{A} \end{tikzcd}$$

2. decomposition:
$$\begin{tikzcd} \mathbf{A} \arrow[d, dash] & & & \mathbf{A} \arrow[d, dash] & \\ \mathbf{C} \arrow[r, dash] & \mathbf{B,D} & \Longrightarrow & \mathbf{C} \arrow[r, dash] & \mathbf{B} \end{tikzcd}$$

3. weak union:
$$\begin{tikzcd} \mathbf{A} \arrow[d, dash] & & & \mathbf{A} \arrow[d, dash] & \\ \mathbf{C} \arrow[r, dash] & \mathbf{B,D} & \Longrightarrow & \mathbf{B,C} \arrow[r, dash] & \mathbf{D} \end{tikzcd}$$

4. contraction:
$$\begin{tikzcd} \mathbf{A} \arrow[d, dash] & & & \mathbf{A} \arrow[d, dash] & & & \mathbf{A} \arrow[d, dash] & \\ \mathbf{B,C} \arrow[r, dash] & \mathbf{D} & + & \mathbf{C} \arrow[r, dash] & \mathbf{B} & \Longrightarrow & \mathbf{C} \arrow[r, dash] & \mathbf{B,D} \end{tikzcd}$$

Pearl and Verma (1987); Pearl, Geiger, and Verma (1989)

Additive faithfulness DAG (AFDAG)

Definition

$X = (X_1, \ldots, X_p)^\top$ is additively faithful w.r.t. G if, for any $(i,j) \in \mathrm{E}$ and any subset $\mathrm{S} \subseteq V \setminus \{i,j\}$,

$$i \text{ and } j \text{ are } d\text{-separated by } \mathrm{S} \quad \Leftrightarrow \quad X_i \perp\!\!\!\perp_{\mathscr{A}} X_j | X_{\mathrm{S}}.$$

Proposition

Suppose

a. *X follows a multivariate Gaussian copula distribution with transforming functions $(f^1, \ldots, f^p)$;*

b. $\mathscr{A}_i = \mathrm{Span}\{f^i\}$.

Then $\boxed{\textit{faithfulness} \Leftrightarrow \textit{additive faithfulness.}}$

Li, Chun, and Zhao (2014)

Outline

- Motivating data set
- Additively faithful directed acyclic graph
- Estimation, consistency, and examples

Additive covariance operator

Definition (Σ_{XX})

$\Sigma_{XX} : \oplus \mathscr{A}_j \to \oplus \mathscr{A}_j$, for each $f = f_1 + \cdots + f_p$, satisfies

$$\Sigma_{XX} f = \begin{pmatrix} \Sigma_{X_1 X_1} & \cdots & \Sigma_{X_1 X_p} \\ \vdots & \ddots & \vdots \\ \Sigma_{X_p X_1} & \cdots & \Sigma_{X_p X_p} \end{pmatrix} \begin{pmatrix} f_1 \\ \vdots \\ f_p \end{pmatrix};$$

with pairwise covariance operator $\Sigma_{X_i X_j}$ induced by

$$\langle g, \Sigma_{X_i X_j} f \rangle = \mathrm{cov}[f(X_j), g(X_i)],$$

for any $f \in \mathscr{A}_j$, and $g \in \mathscr{A}_i$.

Characterization of ACI

Definition

Additive conditional covariance operator (ACCO):

$$\Sigma_{X_i X_j | X_S} \triangleq \Sigma_{X_i X_j} - \Sigma_{X_i X_S} \Sigma_{X_S X_S}^{\dagger} \Sigma_{X_S X_j},$$

$\dagger$: Moore-Penrose inverse

Theorem

$$X_i \perp\!\!\!\perp_{\mathscr{A}} X_j | X_S \;\Leftrightarrow\; \Sigma_{X_i X_j | X_S} = \mathbf{0}$$

- For non-additive kernels: see Fukumizu et al. (2009)

Re-building the network

$\boxed{i \text{ and } j \text{ disconnected}}$

$$\Updownarrow$$

$\boxed{i \text{ and } j\ d\text{-separated by some } \mathsf{S} \subseteq \mathsf{V} \backslash \{i, j\}}$

$$\Updownarrow$$

$\boxed{X^i \perp\!\!\!\perp_{\mathscr{A}} X^j | X^{\mathsf{S}} \text{ for some } \mathsf{S} \subseteq \mathsf{V} \backslash \{i, j\}}$

$$\Updownarrow$$

$\boxed{\Sigma_{x_i x_j | x_{\mathsf{S}}} = \mathbf{0} \text{ for some } \mathsf{S} \subseteq \mathsf{V} \backslash \{i, j\}}$

Consistency

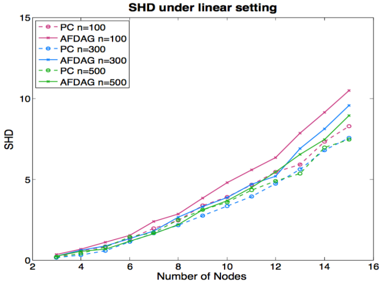
Theorem

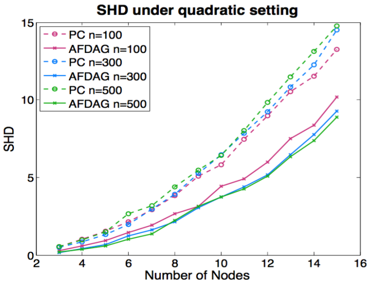
$\mathrm{E}_{\mathrm{CPDAG}}$ *and* $\widehat{\mathrm{E}}_{\mathrm{CPDAG}}$*: true and estimated completed partially directed acyclic graphs (CPDAG). Suppose* $n^{-1/2} \prec \epsilon_n \prec 1$*. Under some regularity conditions,*

1. $\|\hat{\Sigma}_{x_i x_j | x_S}(\epsilon_n) - \Sigma_{x_i x_j | x_S}\| = O_P(\epsilon_n^{1/2} + n^{-1/2}\epsilon_n^{-1})$,
2. $P(\widehat{\mathrm{E}}_{\mathrm{CPDAG}} = \mathrm{E}_{\mathrm{CPDAG}}) \to 1$

- Use ridge inverse (with parameter ϵ_n) to replace Moore-Penrose inverse

Simulations



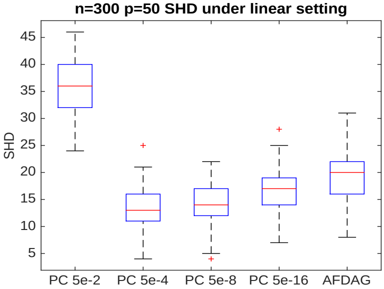


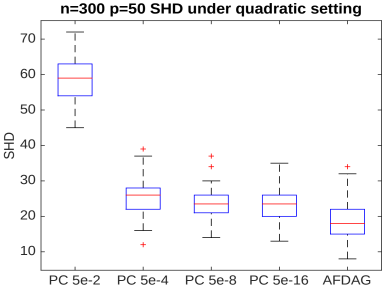
$\star$ SHD: structure Hamming distance

- Given edge E, sequentially generate $\{X_i,\ i = 1, \ldots, p\}$ by
  $X_1 = \varepsilon_1, \quad X_i = f_i(\{X_j : (i,j) \in \mathrm{E}\}, \varepsilon_i), \quad i = 2, \ldots, p, \quad \varepsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$
- Linear : $X_i = \sum_{(i,j)\in \mathrm{E}} \alpha_{i,j} X_j + \varepsilon_i$, Quadratic : $X_i = \sum_{(i,j)\in \mathrm{E}} \alpha_{i,j} (X_j)^2 + \varepsilon_i$.
- $\alpha_{i,j} \sim$ Uniform$(0,1)$

Pearl (2000); Tsamardinos, Brown, and Aliferis (2006)

Simulations (50 nodes)





Pathway analysis

- Randomly draw 2,000 cells as sub-sample
- Repeat analysis 20 times

	AFDAG	PC (5E-02)	PC (5E-04)	PC (5E-08)	PC (5E-16)
mean (std)	$\mathbf{16.95(0.60)}$	22.45 (2.21)	20.20 (2.61)	18.90 (2.25)	18.25(1.29)

Summary

- New principle: additive conditional independence
- AFDAG: a new model for causality learning
- Theoretical and numerical justifications

$\boxed{\text{Other applications of ACI?}}$

THANK YOU!