

---

# 独立成分分析による周期的な因果 モデルの発見

---

グスタボ・ラセ  
ルダ

機械学習部 コンピュータサイ  
エンス学部 カーネギーメ  
ロン大学 ピッツバーグ, PA  
15213

ピーター・ス  
パートス ジョ

セフ・ラムゼ  
イ

カーネギーメロン大学哲  
学科 15213, PA Pittsburgh

パトリック・O・ホイ  
ヤー

ヘルシンキ大学コンピュ  
ータサイエンス学部 フ  
インランド、ヘルシンキ

## アブス トラク ト

我々は、清水ら(2006)のICAに基づく、因果的に十分な連続値の観測データから線形非ガウス非周期構造方程式モデル(LiNGAM)を発見する手法を一般化する。SEMのグラフが非周期的であるという仮定を緩和することで、より一般的な線形非ガウス(LiNG) SEMの発見という問題を解決することができる。LiNG発見アルゴリズムは、SEMの分布同値クラスを

出力し、それは大標本限界において、母集団分布を表現する。我々は、LiNG発見アルゴリズムを模擬データに適用する。最後に、出力されたクラスのSEMのうち1つだけが「安定」である十分条件を与える。

線形構造方程式モデル(SEM)は、自然科学や社会科学(計量経済学、政治学、生物学など)で広く使われている統計学的因果関係モデルです[1]。

線形 SEM の変数は、誤差項(通常、非観測)と実体変数の2つのセットに分けられる。各実体変数  $x_i$  に対して、左辺に  $x_i$  を、右辺に  $x_i$  の直接原因と対応する誤差項を持つ線形方程式が存在する。

共同独立誤差項を持つ各SEMは、モデルの因果構造と連立方程式の形式を表す有向グラフ(DGと略記)と関連付けることができる。グラフの頂点は実体変数であり、 $x_j$  の構造方程式における  $x_i$  の係数がゼロでないときだけ、 $x_i$  から  $x_j$  への有向エッジが存在する。<sup>1</sup>

<sup>1</sup>従来、非周期的なグラフを持つSEMは、「リ

## 1.1 モデル、イラスト付き

$\mathbf{x}$ を実体変数のランダムなベクトル、 $\mathbf{e}$ を誤差項のベクトル、 $\mathbf{B}$ を実体変数の線形係数の行列とする。すると、次の式が線形SEMモデルを記述する：

$$\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e} \quad (1)$$

例えば、次のように定義されたモデルを考えてみましょう：

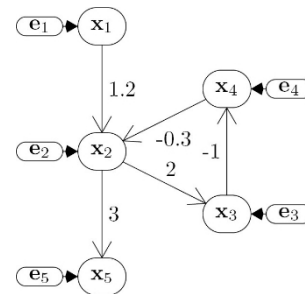
$$\begin{aligned} \mathbf{x}_1 &= \mathbf{e}_1 \\ \mathbf{x}_2 &= 1.2\mathbf{x}_1 - 0.3\mathbf{x}_4 + \mathbf{e}_2 \\ \mathbf{x}_3 &= 2\mathbf{x}_2 + \mathbf{e}_3 \\ \mathbf{x}_4 &= -\mathbf{x}_3 + \mathbf{e}_4 \\ \mathbf{x}_5 &= 3\mathbf{x}_2 + \mathbf{e}_5 \end{aligned} \quad (2)$$

なお、式の左辺にある各変数の係数は1である。

図1：例1

$\mathbf{x}$  は、 $I - \mathbf{B}$  が反転可能である限り、誤差項の線形結合として直接表現することもできる。式1の $\mathbf{x}$ を解くと、 $\mathbf{x} = (\mathbf{I} - \mathbf{B})^{-1} \mathbf{e}$ となる。 $\mathbf{A} = (\mathbf{I} - \mathbf{B})^{-1}$ とすると、 $\mathbf{x} = \mathbf{A}\mathbf{e}$ となる。

cursive"、環状グラフを持つSEMは "non-recursive"[15]である。我々はこの用法を避け、代わりに "acyclic "または "cyclic "を使用する。



成分分析（3.1項参照）、「混合行列」と呼ばれる）。

SEMの誤差項に対する分布は、線形方程式と同様に、実体変数に対する共同分布を伴います。この共同分布は、次に示すように、物理的なプロセスで解釈することができる。

## 1.2 リニアSEMの解釈

これらの方程式（行列式(1)）は、いくつかの異なる解釈を与えることができる。ある種の解釈では、これらは平衡状態における変数 $\mathbf{x}$ の集合が満たす一連の方程式である。さらにいくつかの仮定を置くと、Simultaneous equations（別名「平衡方程式」）の $B$ 行列は、決定論的力学系を記述する一連の力学方程式の係数を表す。

Fisher [5]は、そのような解釈の1つを次のように示した：長さ1の比較的長い観測期間と、長さ $\Delta\theta = 1/n$ のはるかに短い「反応」ラグが存在する。観測変数はベクトル $\bar{\mathbf{x}}[t]$ で、 $t$ で始まる観測期間中の $\mathbf{x}$ の平均と定義される：

$$\bar{\mathbf{x}}[t] \equiv \frac{1}{n} \sum_{k=1}^n \mathbf{x}[t+k\Delta\theta] \quad \text{とする。} \quad (3)$$

基礎となる力学的方程式があるとする：

$$\mathbf{x}[t+k\Delta\theta] = B_{dyn} \mathbf{x}[t+(k-1)\Delta\theta] + \mathbf{e} \quad (4)$$

ここで、 $\mathbf{e}$ は観測期間中一定（ただし、例えば、集団の中の異なる単位で異なる場合がある。異なる観測期間）。

Fisherは、 $\Delta\theta$ が0に近づく極限で、次のことを示した、

というような $B_{equil} = B_{dyn}$ が存在する：

$$\bar{\mathbf{x}}[t] = B_{equil} \bar{\mathbf{x}}[t] + \mathbf{e} \quad (5)$$

すなわち、LiNG発見アルゴリズムが出力する $B$ 行列は、対角がすべて0である。これは、平衡データだけでは $B$ 行列の対角エントリの値を決定することができないためである。

基礎となる力学方程式において、あるインデックス $a$ に対して、 $\mathbf{x}_a[t+(k-1)\Delta\theta]$ が $\mathbf{x}_a[t+k\Delta\theta]$ に影響を与える（すなわち、 $b_{a,a} \neq 0$ ）ことがある。我々の目標は、 $\bar{\mathbf{x}}$ の分布を表し、操作の効果を正しく予測する係数を回収することである。変数 $\mathbf{x}_i$ の分布 $P$ への操作は、 $\mathbf{x}_i$ の力学方程式を新しい力学方程式 $\mathbf{x}_i[t+k\Delta\theta] = \mathbf{e}'_i$ で置き換えることでモデル化され、 $\mathbf{e}'_i$ は分布 $P$ を持つ[10]。

これらの目的のために、以下の議論は、基礎となる力学方程式において $b_{a,a} \neq 1$ である限り、均衡データによる $B_{equil}$ の対角の過小決定が問題でない理由をスケッチする。

$\bar{\mathbf{x}}_a$ の方程式は、次のような形になっています：

$$\bar{\mathbf{x}}_a = b_{a,a} \bar{\mathbf{x}}_a + \sum_{k \neq a, k=1} b_{a,k} \bar{\mathbf{x}}_k + e_a \quad (6)$$

$b_{a,a} \neq 1$ とすると、次のように書き換えることが可能である：

$$\bar{\mathbf{x}}_a - b_{a,a} \bar{\mathbf{x}}_a = \sum_{k \neq a, k=1} b_{a,k} \bar{\mathbf{x}}_k + e_a \quad (7)$$

$$\bar{\mathbf{x}}_a = \frac{1}{1 - b_{a,a}} \sum_{k \neq a, k=1} b_{a,k} \bar{\mathbf{x}}_k + e_a \quad (8)$$

何れ $b_{a,a} = 0$ となり、修正された方程式系がコン

式8を満たすグラフは、自己ループを持たないグラフで表現され、その根底にある力学的性質が異なる

。

$B_{dyn}$  の各固有値のモジュラスが1以下であり、かつ、1に等しい固有値がない場合に限る。

このモデルの基礎となる仮定はかなり強いが、計量経済学では一般的なものであり、Fisher [5]はこの仮定を解除している。

より単純だが、似たような解釈として、観測値 $\mathbf{x}$ が観測期間の平均値ではなく、力学系が収束した状態であるとするものがある。

### 1.2.1 自己ループへの対応

本論文で紹介するLiNG発見アルゴリズム（第4節で説明）は、有向グラフの集合を出力します。

$a$ の式において、 $\mathbf{x}_a[t + (k - 1)\square 0]$ の係数が0になるような式です。

なお、第2式では、誤差項 $e_a$ を $1/(1 - b_{a,a})$ で再スケールリングし、新たな誤差項を形成しています。 $e'_a$ とし、 $(I - B)^{-1}$ が還元型になるようにした場合

を形成する係数は、第1式の $e_a$ に対応する係数に $(1 - b_{a,a})$ を乗じることになり、2つの変化は互いに相殺される。

ここで、元の力学系と $B$ の対角を0にした場合の力学系を考えると（上記のように）、一つの力学系が力学方程式が連立方程式に近づく条件を満たす場合がある。

---

<sup>2</sup>フィッシャーは、セルフループは現実的でないと主張するが、これらの主張は完全に説得力があるわけではない。

$a$ の式の係数の大きさが異なるため、もう一方は極限で連立方程式にならない。両形式がフィッシャーの条件を満たす場合、任意の変数をすべての $t$ について固定された分布に操作する行為によって、2組の力学方程式は等価な極限連立方程式を持つことになります。

### 1.2.2 係数1の自己ループ

残念ながら、 $b_{a,a} = 1$  の場合は、 $1/(1 - b_{a,a})$  が無限大であるため、同じように扱うことができない。 $b_{a,a} = 1$ の場合、自己ループのない等価な形式は存在しないかもしれません（より正確には、自己ループのない対応する方程式は、いくつかの誤差項の分散をゼロに設定する必要があるかもしれません）。 $b_{a,a} = 1$ のケースは、現在のところ解決策を持たない真の問題である。本論文では、係数1の自己ループは存在しないと仮定する。

Dashが指摘するように[4]、連立方程式が基礎となる力学方程式と異なるグラフを持ち、そのため連立方程式を表すグラフは基礎となる力学系を操作した場合の影響を予測できない場合がある。4]では、Dashがこのような例を2つ紹介している。いずれも、実質的には、 $B_{dyn}$  の対角線に1が入っている。

## 2 問題点とその歴史

### 2.1 DG因果関係発見の問題

1.2の解釈を用いると、次のように問題を構成することができる：観測された変数が因果的に十分な集合<sup>3</sup>を形成するLiNG過程の平衡分布のサンプルが与えられたとき、この分布を記述するSEMの集合を、それが空でないという仮定の下で見つけることだ。

### 2.2 リチャードソンのCyclic Causal Discovery (CCD) アルゴリズム

有向無サイクルグラフ（DAG）の（同値クラスを）データから発見するアルゴリズムは数多く提案されているが、一般線形有向グラフ（DG）については、証明的に正しいアルゴリズムが1つだけあった。

線形係数、およびそれらの有向グラフに共通する特徴（祖先関係など）。このアルゴリズムは、PAG を構築するために、ゼロ部分相関の一連の統計的テストを実行する。無相関の誤差を持つ線形SEM に含まれるゼロ部分相関のセットは、線形係数のみに依存し、誤差項の分布には依存しない。ある仮定<sup>4</sup>では、大標本限界において、CCD は真のグラフを表現するPAG を出力する。

このアルゴリズムにはいくつかの制限がある。まず、PAG に含まれるDG の集合は大きくなり、それらはすべて同じゼロ部分相関（すなわち、母集団で成立すると判断されるもの）を伴うが、同じ共同分布や同じ共分散さえも伴う必要はない。したがって、PAGが表現する集合には、データにうまく適合しない環状グラフが含まれる場合がある。したがって、誤差がすべてガウスであると仮定しても、CCD が出力するグラフの集合の大きさを小さくすることは可能であるが、実際には困難である場合がある。アルゴリズムの詳細については、[11] を今まで）知られていた、リチャードソンの循環的因果関係 CCD (Discovery) アルゴリズム。

のすべての値について、ゼロ部分相関の同じセットを伴う有向グラフの集合を表す「部分祖先グラフ」(PAG)を出力します。

<sup>3</sup>変数の集合 $V$ がある集団にとって因果的に十分であるのは、集団において $V$ 内の2つ以上の変数の共通の直接原因がすべて $V$ にある場合のみである（この定義に関する微妙な点は[13]を参照のこと）。

参照されたい。

### 3 清水らによるLiNGAM SEM の発見アプローチ

生成過程の構造方程式が線形で非周期グラフで表現できること、誤差項が非ゼロ分散であること、サンプルが独立かつ同一分布であること、誤差項がガウスであること、誤差項が合同で独立であること、などのデータに関する前提条件のもと、ICA (Independent Components Analysis) を用いた「LiNGAM アルゴリズム」[12] は独自の正しいLiNGAM SEMを確実に発見しています<sup>5</sup>。

#### 3.1 独立成分分析 (ICA)

独立成分分析 [3, 8] は、 $\mathbf{x} = \mathbf{A}\mathbf{e}$  ( $\mathbf{e}$  はしばしば「ソース」と呼ばれ、 $\mathbf{s}$  と書かれる) の形の方程式における混合行列 $\mathbf{A}$  を推定するために用いられる統計学的手法であり、 $\mathbf{x}$  は観測され、 $\mathbf{e}$  と $\mathbf{A}$  は観測されない。

ICA アルゴリズムは反転可能な線形変換を求めます。

<sup>4</sup>その前提とは、サンプルが独立かつ同一分布であること、誤差項がゼロ分散でないこと、ゼロ偏相関の統計的検定が一致すること、方程式の線形性、ユニークな縮小形の存在、忠実性（すなわち、真のグラフの自由パラメータのすべての値に対して伴わないゼロ偏相関は集団に存在しない）、誤差項が無関連であることです。

<sup>5</sup>変数の集合が因果的に十分でない場合、誤差項は一般的に共同独立でない。

の暗黙のサンプル $E$ に対応する誤差分布が最大に非ガウスの（したがって、最大に独立的）になるようなデータ $X$ の $W = A^{-1}$ を定義します。行列 $A$ は、観測された分布が独立な成分の線形で反転可能な混合物であり、そのうち少なくとも1つはガウスである限り、スケーリングと並べ替えまで特定することができる[3]。 $A$ を推定するための計算効率のよいアルゴリズムがある[8]。

### 3.2 LiNGAM発見アルゴリズム

線形SEMによって生成されたデータに対してICAアルゴリズムを実行すると、得られる行列 $W_{ICA}$ は、 $I - B$ の行スケーリング、行パーミュートバージョンとなる（ここで $B$ は真のモデルの係数行列）（これはセクション1.1での導出の結果である）。 $W$ 行列が $I - B$ と等しくなるように、適切な順列とスケールを見つけるという問題が残されているのです。

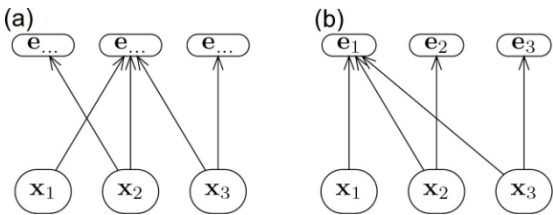


図2: 係数が統計的にゼロと区別できないエッジを除去した後: (a) 生 $W$

のグラフを持つSEM上でICAにより出力された行列を $x_2 \rightarrow x_1 \leftarrow x_3$  (b) 対応する $W$ 行列は、次のようにして得られる。

$W_{ICA}$ の誤差項を並べ替える

ICAで与えられる誤差項の順序はarbitraryなので、アルゴリズムは各誤差項 $e_i$ をそれぞれの

アルゴリズムが正しい順列を見つけた後、正しいスケーリング、すなわち、各行を対角要素で除することによって $W$ を「正規化」し、出力行列の対角がすべて1（すなわち、セクション1で指定したように、各誤差項の係数は1）になるようにする。

ここで、 $W' = \text{normalize}(W)$ 、 $W \equiv \text{RowPermute}(W_{ICA})$ 、 $W_{ICA} = \text{ICA}(X)$ を用いて、 $B = I - W'$ を算出する。

LiNGAMは、因果関係のある矢印の方向を決定するという事実の他に、条件付き独立性ベースの手法[13]に比べて、アルゴリズムの正しさが忠実性の仮定を必要としないという利点もあります。

LiNGAMアプローチの詳細については、[12]を参照してください。

## 4 LiNG SEMを発見する

以下に説明するLiNG発見アルゴリズムファミリー（「LiNG-D」と略記）の仮定は、LiNGAMの仮定と同じで、SEMが非循環的であるという仮定を、 $B_{dyn}$ の対角線が1を含まないという弱い仮定に置き換えています。このより一般的なケースでは、非周期的なケースと同様に、誤差項（ $e_i$ 's）と観測変数（ $x_i$ 's）のすべての一致を認めることによって、ディベートモデルが生成される。

。言い換えれば、各候補の相関の行permutationに対応する。ア ごとくはゼロレス対角線を持つ。イシエー

実体変数 $x_i$ に正しく一致させる必要があります。これは、 $W_{ICA}$ の行の正しい順列を見つけることを意味します。 $W = I - B$ で、 $B$ の対角はゼロなので、正しいモデルに対応する $W_{ICA}$ の行-順列

は対角にゼロを持っていない（我々はこのような順列を「不一致」(inadmissible) と呼ぶ）ことが分かっています。

仮定として、データはDAGによって生成されたので、 $W_{ICA}$ 、許容されるちょうど1つのローパーミュテーションが存在する[12]。これを視覚化すると、この制約は、すべての $e_i$ がvertical arrowのターゲットとなるように、誤差項を並べ替える方法がちょうど1つあることを意味する。<sup>6</sup>

この例では、図2(b)に見られるように、第1エラー項と第2エラー項を入れ替えることが、アドミシブルな行列を生成する唯一の順列であることがわかります。

---

<sup>6</sup>非周期性のもう一つの帰結は、 $\mathbf{x}$ がDAGに対して位相的にソートされていれば、この表現では右向きの矢印が存在しないことである。

LiNGAMと同様に、出力は許容されるモデルの集合である。LiNGAMでは、非周期性の仮定により、この集合は単一のモデルを含むことが保証される。しかし、真のモデルがサイクルを持つ場合、複数のモデルが許容されることになる。

本節の残りは、ICAが有限のデータを扱うことを前提に、許容されるモデルを見つけるという問題を取り上げる。

## 4.1 制約付きn-Rooksをブルーンして解く

これらのアルゴリズムは、 $W_{ICA}$ のどのエントリーがゼロであるかをテストし（すなわち、ブルーンング）、それに基づいてすべての許容される順列を見つける（すなわち、制約付きn-Rooksを解く、セクション 4.1.2 参照）ことによって候補モデルを生成する。 $W_{ICA}$ の各エントリー $w_{ij}$ に対して、 $w_{ij}$ がゼロかどうかの判断を $w_{ij}$ だけを用いて行う場合、そのアルゴリズムを「ローカル」と呼ぶ。

### 4.1.1 どのエントリーをゼロにするか決める

$W_{ICA}$ のどの項目をゼロにするかは、いくつかの方法がある：



- **閾値設定:**  $W_{ICA}$  のどのエントリーがゼロであるかを推定するための最も単純な方法は、単純に閾値を選択し、 $W_{ICA}$  の閾値より（絶対値で）小さいすべてのエントリーをゼロに設定することです。この方法は、異なる係数が異なる広がりを持つという事実を考慮しておらず、閾値より小さい係数をすべて見逃してしまう。

#### • ブートストラップサンプリングによる非ゼ

**口仮説の検証:**  $W_{ICA}$  のどのエントリーが実際にゼロであるかを推定するためのもう一つの方法は、ブートストラップサンプリングを行うことである。ブートストラップサンプルは、元のデータから置換を伴う再サンプリングによって作成される。そして、各ブートストラップサンプルに対してICAを実行し、各係数 $w_{ij}$ を計算する。そして、各係数について、ノンパラメトリック分位点検定を行って、0が外れ値であるかどうかを決定する。もし外れ値でなければ、その係数は0に設定される（つまり、対応するエッジが刈り込まれる）<sup>8</sup>。

- **スパースICAを使用する:** Zhang and Chan [16]が発表したような、スパース（すなわちプルーニング済み）混合物を返すICAアルゴリズムを使用します。上記の他の方法とは異なり、これはローカルアルゴリズムではありません。

#### 4.1.2 制約付きn-Rooks: 問題とアルゴリズム

どのエントリーがゼロであるかが決まると、アルゴリズムは、ゼロのない対角線を持つ $W_{ICA}$ のすべての行順列を検索する。このような各行順列

は、 $n \times n$ のチェス盤の非ゼロエントリに $n$ 個のルークを配置し、2つのルークが互いに脅かされないようにすることに対応する。そして、すべてのルークが対角線上に配置されるように行が順列化され、対角線にゼロがないことが保証される。

この問題を解決するために、次のルークを置く場所がない探索経路を除去する単純な深さ優先探索を使用する。最悪の場合、すべての順列が許容されるため、検索にかかる時間は $O(n!)$ でなければならない。

<sup>7</sup>ICAを実行するたびに、異なる行permutationの $W_{ICA}$ が返される可能性があるため、これを行う際には注意が必要です。つまり、最初に各ブートストラップ $W_{ICA}$ を元の $W_{ICA}$ と一致するように行permutationする必要があります。

<sup>8</sup>分位数検定の代わりに、推定誤差を用いたヌルハイポセシス（すなわちエッジが存在しない）の下でシミュレーションを行うのが正しい手順である、という意見もあるでしょう。

## 4.2 非局所的なアルゴリズム

ローカルアルゴリズムは、 $w_{ij}$  の推定値が互いに独立であるという仮定の下で動作します。このため、非局所的な方法を使用する動機となります。

LINGAM（非周期的）アプローチ[12]では、 $W_{ICA}$ の単一の最良の行順列を見つけるための非局所的アルゴリズムが提示されている。これは、対角のエントリーがゼロに近い（ $x \rightarrow 1/x$  / など）ことに大きなペナルティを与える損失関数を最小化する。これは線形代入問題（すなわち、 $e_i$ s と  $x_i$ s の間の最良の一致を見つけること）として書かれ、Hungarian アルゴリズム[9]などで解くことができる。

しかし、一般的なLING 発見では、許容される順列が複数存在する可能性があるため、最良の線形割り当てを求めるアルゴリズムでは不十分です。

一つのアイデアとして、 $k$ 番目のベストアサインアルゴリズムを使用することで

十分なデータがあれば、許容できないモデルに対応するすべての順列はスコアが悪くなり、許容できるモデルとできないモデルの間に明確な分離があるはずである [2] (すなわち、対角線上のペナルティが最も小さい $k$ 番目の順列)。という条件を設定し、得られた分布の比較を行う。ICA統計とそのブートストラップの分布。しかし、これは本論文に不可欠な問題や複雑さをもたらすものである。

上記の非局所法は、閾値法と同様に、 $W_{ICA}$ のエントリーの推定値間の広がりの違いを考慮していない。これを修正するには、単にゼロに近いという理由でペナルティを与えるのではなく、検定が帰無仮説を棄却できなかった対角のエントリーにペナルティを与えるように損失関数を修正する（セクション 4.1.1 のブートストラップサンプルに関する部分で述べたように）ことが簡単であろう。

## 4.3 試運転

実施例1のSEMを用い、対称ガウス2乗分布<sup>9</sup>に従って分布する誤差項を用いて、15000点のサンプルポイントを生成した。

図3は、カットオフを0.05に設定した局所閾値処理アルゴリズムの出力です。

再現性を高めるため、インストラクションを含むコードは、[www.phil.cmu.edu/~tetrad/cd2008.html](http://www.phil.cmu.edu/~tetrad/cd2008.html) から入手できます。

## 5 理論編

### 5.1 DG同値の概念

SEMに関連する有向グラフが、観測データから「等価」または「区別できない」可能性があることには、多くの異なる意味がある、

---

<sup>9</sup>分布は、標準ガウス(0,1)からサンプリングし、それを2乗することで作成した。サンプリングされた値が負の値であった場合は、再び負の値にした。

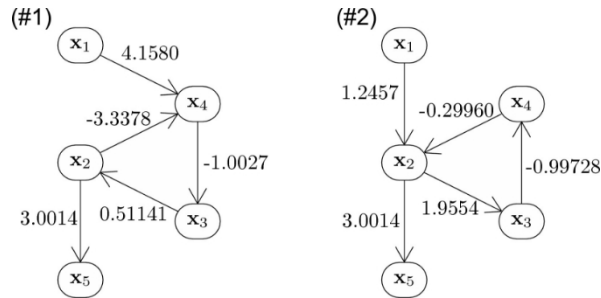


図3 LiNG-Dの出力：候補者#1、候補者#2

は、線形で誤差項間の依存性がないと仮定する：

- DG  $G_1$  と  $G_2$  は、DG  $G_1$  を用いた線形 SEM の自由パラメータ（非ゼロ線形係数、error 項の分布）のすべての値に対して伴うゼロ偏相関のセットが、 $G_2$  を用いた線形 SEM の自由パラメータのすべての値に対して伴うゼロ偏相関のセットと同じである場合にのみゼロ偏相関同値となります。線形モデルについて、これは *d-separation equivalent* と同じです。[13]
- DG  $G_1$  と  $G_2$  は、DG  $G_1$  を持つ線形 SEM の自由パラメータのパラメータ値のすべてのセットに対して、DG  $G_2$  を持つ線形 SEM の自由パラメータのパラメータ値のセットがあり、2つの SEM が実体変数上で同じ共分散行列を伴う場合、およびその逆の場合にのみ共分散同等である。
- DG  $G_1$  と  $G_2$  は、DG  $G_1$  を持つ線形 SEM の自由パラメータのパラメータ値のすべてのセットに対して、DG  $G_2$  を持つ線形 SEM の自由パラメータのパラメータ値のセットがあり、2つの SEM が実体変数に同じ分布を伴う場合、およびその逆の場合にのみ、分布が等価です。これは、SEM間の分布-内包の等価性の概念と混同しないでください：固定パラメータを

持つ2つのSEMは、それらが同じ分布を内包する場合、分布-内包と等価です。

ガウスの場合に関するよく知られた定理[13]と、非ガウスの場合に関する既知の結果のいくつかのつまらない結果[12]から、非周期グラフの異なる意味の等価性の間に以下の関係が存在することがわかる：誤差項がすべてガウスであると仮定した場合、分布同値は共分散同値と等価であり、これはd-分離同値と等価である。誤差項がすべてガウスでない場合

誤差項がガウスであると仮定した場合、分配等価は共分散等価を伴いますが、共分散等価はd-分離等価を伴います（ただし、d-分離等価は伴わない）。

そのため、例えばガウス型の誤差項が与えられた場合、 $A \leftarrow B$ と $A \rightarrow B$ はゼロ偏相関相当、共分散相当、分布相当となります。しかし、非ガウス誤差項が与えられると、 $A \leftarrow B$ と $A \rightarrow B$ はゼロ偏相関等価、共分散等価であるが、分布等価ではない。つまり、ガウス誤差とこのDGの組に対しては、観測データのみに依存するアルゴリズムでは、さらなる仮定をすることなく、正しい因果グラフとして母集団分布に適合するユニークな非周期グラフを確実に抽出することはできないが、すべての（または1つを除くすべての）非ガウス誤差に対しては、母集団分布に適合するユニークな非周期グラフが必ず存在することになる。

環状グラフとガウス誤差の場合に関する定理はあるが、非ガウス誤差の環状グラフに関する分布同値に関する定理は知られていない。すべてのガウス誤差を持つ巡回グラフの場合、分布同値は共分散同値と等価であり、これはd-分離同値を内包する（が内包しない）[14]。最大1つの誤差項が非ガウスである環状グラフの場合、分布同値は共分散同値と等価であり、その結果、d-分離同値と等価になる（しかし、等価ではない）。しかし、最大で1つのガウス誤差項が与えられたとき、非周期グラフと周期グラフの重要な違いは、異なる2つの非周期グラフは分布等価ではないが、分布等価な異なる周期グラフは存在するという点である。

したがって、観測データのみに依存するアルゴリズムでは、さらなる仮定なしに、データに適

合する固有の環状グラフを正しい因果グラフとして確実に選択することはできない。例えば、図3の2つの環状グラフは分布的に等価である。

## 5.2 LiNG-Dの出力は正しく、可能な限り微細なものです

**定理1** *LiNG-Dの出力は、分布-内容同値クラスを構成するSEMの集合である。*

**証明する：** まず、LiNG-Dのアウトプットに含まれる任意の2つのSEMが、同じ分布を伴うことを示す。

ICAで出力される重み行列は、スケーリングと行の並べ替えまでしか決定しない。直感的には、誤差項を並べ替えても、混合度は変わらない。さて、より正式には

LiNG- が出力するモデル候補を  $M_1$  と  $M_2$  とする。

D.そして、 $W_1$  と  $W_2$  は、 $W_{ICA}: W_1 = P_1 W_{ICA}, W_2 = P_2 W_{ICA}$  の行変態である。

同様に、誤差項についても  $e_1 = p_1 e, e_2 = p_2 e$

そして、 $M_1$  が意味するサンプル  $X$  のリストは、 $A E_{11} = (W_1)^{-1} E_1 = (P_1 W_{ICA})^{-1} (P_1 E) = W^{-1} P_1^{-1} E$  です。

$W_{ICAE}$  です。

同じ論法で、 $M_2$  が意味するサンプル  $X$  のリストは  $W^{-1} E$  でも ある。したがって、LiNG-Dが出力する2つのSEM mod-elは同じ分布を意味することになる。

さて、LiNG-Dが分布  $P$  を内包するSEMを1つ出力する場合、 $P$  を内包するすべてのSEMを出力することが示される必要がある。

LiNG-Dが出力するある  $T$  と同じ分布を表すSEM  $S$  があるとする。このとき、 $S$  と  $T$  の縮小形係数行列  $A_S$  と  $A_T$  は、列の並べ替えとスケールリングまで同じである。したがって、 $I - B_S$  と  $I - B_T$  も、スケールリングと行順列まで同じである ( $I - B = A^{-1}$  によって)。係数1の自己ループがないという仮定により、 $I - B_T$  も  $I - B_S$  も対角線上にゼロはない。 $I - B_T$  は、対角線上にゼロを持たない  $W_{ICA}$  のスケールリングされた行変態なので、 $I - B_S$  も同様です。したがって、 $S$  もLiNG-Dによって出力されます。

**定理2** 連立方程式が線形で有向グラフで表現でき、誤差項が非ゼロ分散で、サンプルが独立同分布で、誤差項が1つ以上ガウスで、誤差項が共同独立であれば、大標本限界においてLiNG-Dは母分布に関係するすべてのSEMを出力します。

の尾を引く (ソース)。Richardsonはサイクルを反転させる操作によって、内包されるゼロ部分相関の集合を保存することを証明したが、分布の等価性については考慮していない[11]。

### 5.3 安定性の前提を加える

力学系では、「安定した」モデルとは、以下のようなものです。

**証明する:** ICAは、[3]に記載された仮定の下で、 $A$  と  $W$  の点的整合性のある推定値を与えます。これは、 $W$  行列のエントリーがゼロかどうかの点的整合性のある検定があることを意味し、したがって、(定義により) 大標本限界において、ゼロ係数の検定のタイプIとタイプIIの両方の誤差の限界はゼロである。 $W$  行列のゼロが正しい場合、LiNG-Dアルゴリズムのローカルバージョンの出力は、連立方程式が母集団分布を記述するという意味で正しいです。

一般に、各候補モデル  $B' = I - W'$  は、 $W_{ICA}$  の行順列の構造を持つ。この構造は、 $W'$  の行を順列化したときに何が起こるかを分析することで生成できる。 $B'$  (従って  $W'$ ) の辺は列から行に読み替えられることを思い出してほしい。したがって、 $W'$  の行順列は、矢印の頭 (ターゲット) の位置は変えるが、矢印の頭 (ターゲット) は変えない。

となり、1回限りのノイズの影響が消える。例えば、サイクル積（サイクルのエッジの係数の積）が1以上のサイクルを1つ持つモデルは不安定であり、サイクル積が-1～1のサイクルを1つ持つモデルは安定である。一方、サイクル積2の正帰還ループをサイクル積-1.5の負帰還ループで打ち消すと、有効サイクル積が0.5となり、モデルは安定となる。

安定性の一般的な表現方法として、 $\lim_{t \rightarrow \infty} B^t = 0$  があります、

これは数学的に次のように等価です：すべての固有値の

この固有値の基準は簡単に計算できる。

異なる変数間の係数のみを考えると、自己ループを仮定しない限り、SEMの安定性を測定することは不可能である。そこで、本節では、真のモデルには自己ループがないものと仮定する。

LiNG-Dが出力するSEMの多くが不安定であることはよくあることです。多くの場面で、変数は平衡状態にあると仮定されるため、不安定なモデルを除外することが許されることが多い。

本節では、母集団分布を生成するSEMが周期が不連続なグラフを持つ場合、LiNG-Dが出力する候補SEMのうち、最大1つは安定であることを証明する。

**定理3**  $\pi / \square 1$  であるようなサイクル積 $\pi$ を持つ単純サイクルの形のSEMは不安定である。

**証明する：** サイクルの長さを $k$ とする。すると、 $B^k = \pi I$  となる。すると、すべての整数 $i$  について、 $B^{ik} = \pi^i I$  となる。つまり、 $\pi / \square 1$  であれば、 $i$  が大きくなっても $B^{ik}$  のエントリは $B$  のエントリより小さくならない。したがって、 $B^t$  は $t \rightarrow \infty$  のように0に収束することはない。

**補論1：** 単純なサイクルの形をしたSEMにおい

て、サイクル積 $\square 1$  であることは、固有値 $\square 1$ （モジュラスにおいて）であることと同じであり、不安定であることと同じである。

**定理4** 係数行列 $B$  とグラフ $G$  を持ち、分布 $Q$  を内包するdis-joint cycles を持つSEM  $M$  と、グラフ $G_0$ 、係数行列 $B_0$  を持つSEM  $M_0 \models M$  があり、これは $M$  のadmiss-sible permutation であり $Q$  も内包するとする。  $\square$

もまた、不連続なサイクルを含み、そのうちの少なくとも1つは、 $G$ のサイクル $C$ の反転であり、そのサイクル積は $C$ のサイクル積の逆数である。

**証明する：**紙面の都合上、ここでは証明の概略を述べるにとどめる。ここで、 $a \rightarrow b$ は、 $a$ から $b$ に写されることを意味する。（いくつかの環状部分変異は、それ自身に写された単一の対象を含む、つまり、トリビアルであるかもしれない）。したがって、 $B$ の単一の許容される環状行順列変化について定理を証明することで十分である。 $m \rightarrow n \rightarrow a$ が許容される場合、 $G$ は $a \leftarrow b \leftarrow \dots m \leftarrow n \leftarrow a$ に等しいサイクル $C$ を含み、 $G_0$ は $a \rightarrow b \rightarrow \dots m \rightarrow n \rightarrow a$ に等しい逆サイクル $C$ を含む。さらに $G_0$ に触れる2サイクルが含まれる場合、 $G$ も同様である。

$B_C$ 、サイクル $C$ のエッジの係数を含む $B$ の部分行列を考える。

$$B_C = \begin{pmatrix} 0 & \dots & 0 & b_{k,1} \\ b_{0,1,2} & \dots & 0 & \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

なお、サイクルプロダクト $\pi_C = b_{k,1}^{Qk-1} b_{l,i+1}$  .

$$W_C = I - B_C .$$

反転」は、ローパーミュテーションで、その中では下段に "ローテート" されます：

$$\begin{pmatrix} -b_{1,2} & \dots & 0 \end{pmatrix}$$

$$\text{ロー・パーミュート}(W_C) = \begin{pmatrix} 0 & -b_{2,3} & \dots & 0 \end{pmatrix}$$

の場合、サイクルの固有値)。  $G$ のサイクルが不安定であるとする。すると、それは（モジュラスで）固有値 $\square 1$ を持つ。しかし、これは $B_G$ の固有値でもあるので、 $G$ が不安定であることがわかる。他の方向も同様である。

**定理6** 真のSEMが安定で、サイクルが不連続なグラフを持つ場合、LiNG-Dの出力に含まれる他のSEMが安定であることはない。

**証明する：**真のSEMが安定で、サイクルが不連続であるグラフを持つとする。定理2により、LiNG-Dの出力は、真のSEMに代わる立派な分布同化等価物であるので、他の立派な候補がすべて不安定であることを示せば十分である。

定理5により、 $G$ のすべてのサイクルは安定である。定理4により、 $H$ は $G$ に対して少なくとも1つのサイクル $C$ を反転させ、この反転されたサイクルは、以下のサイクル積の逆であるcycle積を持つ。

C.Collary1により、反転した行列は安定ではない。したがって、定理5により、 $H$ は不安定である。

したがって、 $G$ に代わる安定した許容できる唯一の選択肢は $G$ そのものです。

真のモデルのサイクルが不連続であれば、真のモデルがstableであるという仮定のもと、LiNG discovery algorithmを用いて完全に特定できる（LiNG discovery algorithmの出力中の最大1つのSEMが安定である）ことがわかります。

例えば、図3に示す2つのモデル候補を考えてみましょう。真のモデルが安定していると仮定すれば、候補の2番を選択することになる。我々のシミュレーションでは

が安定したモデルを使用した場合、これは確かに正しい答えとなります（図1参照）。

$$\begin{pmatrix} 0 & 0 & \cdots & 1 \\ 1 & 0 & \cdots & -b_{k1} \end{pmatrix}$$

対角線がすべて1になるように正規化すると、 $W_C'$  が得られます  $B_C' = I - W_C'$  を計算すると、サイクル-は次のようになります。  

$$\text{product}_b \pi_C' = \prod_{i=0}^Q \frac{1}{b_{i,i+1}} = 1/\pi_C.$$

ここで、サイクルが不連続であるSEMについて、その安定性はサイクルの安定性にのみ依存することを示すことにする。

**定理5** サイクルが不連続であるSEMは、不安定なサイクルを持たない場合にのみ安定である。

**証明する：**  $G$  をサイクルが不連続なSEMとする。このとき、 $B_G$  は、各対角ブロックがサイクルであるブロック-三角行列として書くことができる。ブロック三角行列の固有値の集合は、対角線上のブロックの固有値の集合の和である（この場合は

しかし、一般的には、安定なモジュールが複数存在する可能性があり、確実に正しいものを選択することはできない。周期が不連続でない場合、安定な候補が複数存在する例を見つけるのは簡単である。

不連続なサイクルという条件は十分だが、そうではない。SEMを思いつくのは簡単で、そこで

は、交差するCy-clesにもかかわらず、分布-結合の同値クラスでちょうど1つの安定したSEMを持っています。

## 6 ディスカッション

我々は、LiNGAM SEMを発見する清水氏のアプローチを紹介し、それを一般的なLiNG SEMを発見する方法へと一般化した。これは、d-分離等価クラス全体ではなく、SEMの分布-内包等価クラスのみを出力することにより、環状線形SEMの発見に関する最先端技術を改善するものである。



また、忠実性の保証を緩和することによっても。  
また、安定性が強力な制約となり、候補を1つのSEMに絞り込むことができることも示した。

今後の研究課題として、いくつかの疑問が残されています：

- LiNG-Dアルゴリズムは、すべての許容される順列を生成する。n-Rooksの最悪の場合の時間複雑度は高いが、ランダムなインスタンスに対して深さ優先の探索よりも良いことができるのか？ すべての候補を通過することなく、効率的に安定モデルを探索するアルゴリズムはあるのか？ サイクルが不連続である場合、各サイクルに対して正しい順列を順次見つけることが可能であるが、一般にそのようなトリックは知られていない。
- 事前情報をどのようにアルゴリズムに組み込むことができるのか？
- 因果的充足性の仮定を緩和できるようにするには、アルゴリズムをどのように修正すればよいか。非周期的な場合については、[7]を参照のこと。
- 非ガウシアンとガウシアン（あるいは最もガウシアン）の誤差項の混合を許容するために、どのようにアルゴリズムを修正することができるでしょうか？ Hoyer et al [6]は、非周期的な場合についてこの問題を解決している。
- この方法をどのようにすれば主流の力学系研究に取り入れることができるのか？ このアルゴリズムはノイズの多いダイナミクスやノイズの多い観測を扱うことができるのか？ 非線

形ダイナミクスを扱えるようにできるか？ 係数1の自己ループについてはどうだろうか？ 非平衡時系列データを利用した手法とどのように統合できるか？

## 謝辞

Anupam Gupta、Michael Dinitz、Cosma Shaliziに感謝したい。GLは、NSF Award No.REC-0537198の一部支援を受けた。

## 参考文献

- [1] K.Bollen (1989) - *Structural Equations with Latent Variables*, John Wiley & Sons, New York.
- [2] C.R. Chegireddy, H. W. Hamacher (1987) - Algorithms for finding K-best perfect matchings *Discrete Applied Mathematics*, **18**:155-165.
- [3] P.Comon (1994) - Independent component analysis - a new concept? *信号処理*, **36**:287- 314.

- [4] D.Dash (2005) - Restructuring Dynamic Causal Systems in Equilibrium. 第10回人工知能と統計に関する国際ワークショップ(AIStats 2005) 予稿集
- [5] F.Fisher (1970) - 連立方程式モデルに対する対応原理.*Econometrica*, **38**(1):73-92.
- [6] P.O. Hoyer, A. Hyvärinen, R. Scheines, P. Spirtes, J. Ramsey, G. Lacerda, S. Shimizu (2008) - 任意の分布を持つ線形非周期モデルの因果関係発見。第24回人工知能における不確実性会議予稿集。
- [7] P.O. Hoyer, S. Shimizu, and A. J. Kerminen (2006) - Estimation of linear, non-gaussian causal models in presence of confounding latent variables *Proc. 第3回ヨーロッパワークショップ (PGM'06)*、pp.155-162。
- [8] A. Hyvärinen, J. Karhunen, E. Oja (2001) - *Independent Component Analysis*. Wiley Interscience.
- [9] H.W. Kuhn (1955) - 配置問題に対するハンガリー法。 *Naval Research Logistics Quarterly*, **2**:83-97, 1955.
- [10] J.Pearl (2000) - *Causality: モデル、推論、推察*ケンブリッジ大学出版局
- [11] T.Richardson (1996) - A Polynomial-Time Algorithm for Deciding Markov Equivalence of Directed Cyclic Graphical Models. 第12回人工知能における不確実性会議議事録
- [12] S.Shimizu, P. Hoyer, A. Hyvärinen, A. Kerminen (2006) - A linear, non-gaussian acyclic model for causal discovery. *機械学習研究* **7**:2003-2030.
- [13] P.Spirtes, C. Glymour, R. Scheines (1993) - *Causation, Prediction, Search*. Springer-Verlag Lecture Notes in Statistics 81.
- [14] P.Spirtes (1995) - Directed Cyclic Graphical Representation of Feedback Models, *Proceedings of Eleventh Conference on Uncertainty in Artificial Intelligence* (人工知能の不確実性に関する第11回会議) .
- [15] R.Strotz and H. Wold (1960) - Recursive versus nonrecursive systems: an attempt at synthesis.*Econometrica* **28**:417-427.
- [16] K.Zhang, L-W.Chan (2006) - ICA with Sparse Connections *Intelligent Data Engineering and Automated Learning IDEAL 2006*.