

# Framework

- Structural causal model (Pearl, 2001)

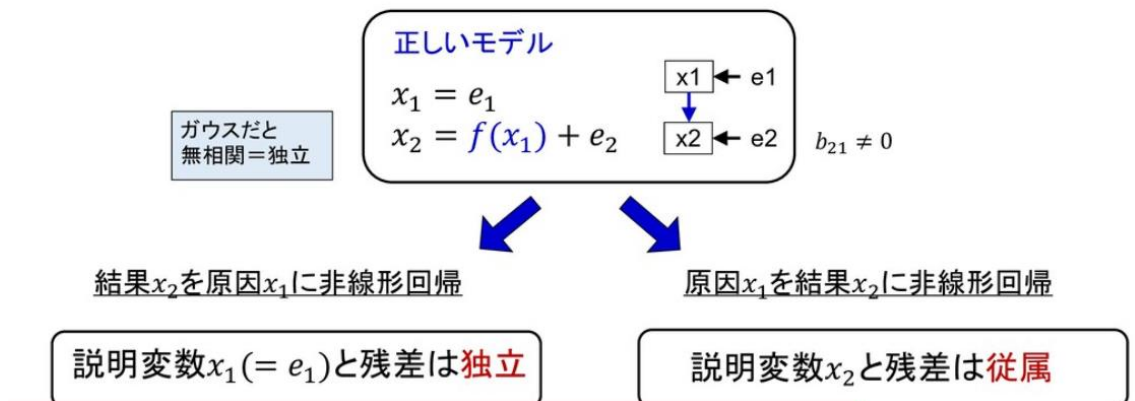
$$x_i = f_i(\text{parents of } x_i, \underbrace{e_i}_{\text{Error variable}})$$

## Other identifiable models

- Nonlinearity + “additive” noise (Hoyer+08NIPS, Zhang+09UAI, Peters+14JMLR)
  - $x_i = f_i(\text{par}(x_i)) + e_i$
  - $x_i = g_i^{-1}(f_i(\text{par}(x_i)) + e_i)$

## 非線形モデル

非線形回帰して説明変数と残差が独立か調べる



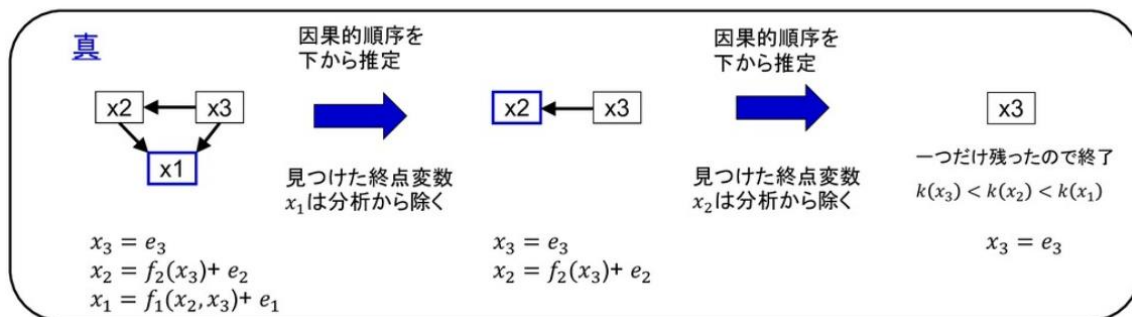
## 因果関数モデルによる因果推定

$Y - f(X)$  が  $X$  と独立となる関数  $f$  が存在し、 $X - g(Y)$  が  $Y$  と独立となる関数  $g$  が存在しないとき、 $X$  は  $Y$  の原因

## 因果的順序を下から推定: 終点変数の同定

- ・ 終点変数: (モデル内に)子を持たない変数
- ・ 終点変数を目的変数にしたときのみ、説明変数と残差が独立になる

$r_1 = x_1 - g_1(x_2, x_3)$ と $x_2, x_3$ は独立だが、 $r_2 = x_2 - g_2(x_1, x_3)$ と $x_1, x_3$ は独立でない...



## 独立性を測る尺度

- 厳密にはあらゆる非線形関数 $f, g$ の非線形相関ゼロを確認せねばならないが、現実には無理。

独立性判定: 非線形相関係数

$$C_{fg}(x_1, r_2) = \frac{\sum_{i=1}^n f(x_1)g(r_2)}{\sqrt{\sum_{i=1}^n f(x_1)^2} \sqrt{\sum_{i=1}^n g(r_2)^2}} = 0 \quad (\text{あらゆる非線形関数 } f, g \text{ について})$$

- 以下の非線形相関により近似的に変数とその残差の独立性を検定評価

$$C_{fg}(x_1, r_2) \cong |\text{corr}\{x_1, g(r_2)\}| + |\text{corr}\{g(x_1), r_2\}| \quad (g = \tanh)$$

モデルの前提がデータで成立していれば、非線形相関の近似誤差を除いて、この方法は必ず正しい唯一の因果構造を同定することが証明されている。

[https://ibismml.org/archive/ibismml001/2010\\_IBISML01\\_Washio.pdf](https://ibismml.org/archive/ibismml001/2010_IBISML01_Washio.pdf)

[https://www.sas.com/content/dam/SAS/ja\\_jp/doc/event/sas-user-groups/usergroups2015-g-01.pdf](https://www.sas.com/content/dam/SAS/ja_jp/doc/event/sas-user-groups/usergroups2015-g-01.pdf)

ここで、 $g$  は非線形かつ非二次関数となる関数、たとえば  $g(\cdot) = \tanh(\cdot)$  である。変数  $x_j$  とその残差  $r_i^{(j)}$  が独立であれば、式 (5) の統計量  $T$  は 0 となる。厳密には、独立であることは統計量  $T$  が 0 になることよりもはるかに厳しい条件である。しかし、独立成分分析に関する文献 [Hyvärinen 01] にあるように、式 (5) のような非線形相関の評価を用いることは、実際上は、多くの場合において十分に良い方法といえる。必要であれば、独立性を評価するために式 (5) の統計量  $T$  を用いる代わりに、より洗練されたノンパラメトリックな独立性指標 [Bach 02, Gretton 05] を用いることができる。

<http://www.ar.sanken.osaka-u.ac.jp/~inazumi/Inazumi10JSAI.pdf>

<https://www.mathsoc.jp/publication/tushin/1604/1604washio.pdf>

2変数  $x, y$  間の独立性を評価するために、しばしば以下の非線形相関係数が使われます.

$$C_{f,g}(x, y) = \frac{E_{x,y}(f(x)g(y))}{E_x(f(x)^2)E_y(g(y)^2)}$$

任意の非線形関数  $f, g$  について  $C_{f,g}(x, y)$  が十分ゼロに近ければ,  $x$  と  $y$  は独立であると判断できます. しかしながら, あらゆる  $f, g$  の組み合わせについて  $C_{f,g}(x, y)$  を計算することは現実的ではありません. そこで, 与えられた  $x, y$  のデータについて  $C_{f,g}(x, y)$  が最大になる  $f$  と  $g$  を前述したカーネル法を用いて推定し, その  $C_{f,g}(x, y)$  が十分ゼロに近いことを以って独立性を判断する手法がよく使われています.

これに対し、本研究提案者の研究グループは、非線形系では非線形回帰の残差が系内の因果の向きで異なることを発見し、理論的解析と実データ検証を進めた[9]。それによれば、ノイズがガウスか非ガウスか、変数とノイズが独立かなどに係わらず、変数間の因果の向き沿った非線形回帰残差が逆向きの残差より小さい。すなわち我々の理論では、図 1 (b)に示すように対象系が非線形ならば、ノイズの性質や独立性、交絡変数の有無などに依らず、非線形回帰残差から変数間の因果関係を一意に推定できる。さらに、このような非線形回帰は高速に計算可能である。

以上の着想により、上記欧米研究の困難に直面することなく、大規模対象系の多数の観測変数間の因果関係を、高精度、高速に推定可能な原理と手法を確立できると予想された。このように汎用でスケーラブルな手法の研究は世界的にも行われておらず、本提案研究の成果は統計的因果推論の学術研究、実用化研究において、大きなブレークスルーをもたらし、日本の当該研究分野での世界的地位を確固たるものにする予想された。

れ、一般性を失うことなく  $C$  の値域は有限であるとし、 $N$  は期待値がゼロであるとする。

以上より、 $\phi$  は  $C$  から  $E$  を回帰するときの最小二乗誤差を最小化する関数であり、すなわち  $\phi$  は  $C$  がある値  $c$  を取るときの  $E$  の値の条件付き期待値  $\phi(c) = E[E|c]$  である。また、 $\phi$  を  $E$  から  $C$  を回帰するときの最小二乗誤差を最小化する関数が存在し、すなわち  $\phi$  は  $E$  がある値  $e$  を取るときの  $C$  の値の条件付き期待値  $\phi(e) = E[C|e]$  である。

以上の問題設定において、我々は以下の不等式が常に成立することを数学的に証明した。

$$E[(E - \phi(C))^2] < E[(C - \phi(E))^2]$$

これは特に  $C$  と  $E$  の値域がそれぞれ  $[0, 1]$  の間に規格化されている場合には、次のように書き直すことができる。

$$E[\text{Var}[E|C]] < E[\text{Var}[C|E]]$$

この不等式は  $C, E$  が線形関係である場合には等号となり成立しないが、任意の非線形関係である場合には必ず成立する。従って、どちらが原因変数  $C$  か結果変数  $E$  であるかが不明であるが関係が非線形であることが知られている 2 変数  $X, Y$  のデータが与えられた時に、 $X$  から  $Y$  への回帰誤差と  $Y$  から  $X$  への回帰誤差を比較して、誤差が小さい回帰向きに従って何れが原因変数と結果変数であるかを決定すれば良い。

<https://kaken.nii.ac.jp/file/KAKENHI-PROJECT-17K00305/17K00305seika.pdf>

## 回帰統計値 [編集]

この手順の根底にある仮定は、このモデルが線形関数、すなわち一次のテイラー級数

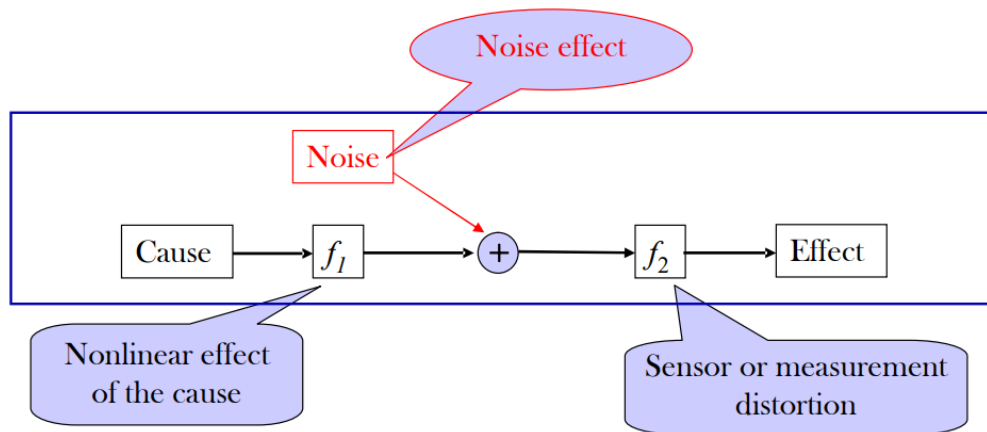
$$f(x_i, \beta) \approx f(x_i, 0) + \sum_j J_{ij} \beta_j$$

によって近似できるというものである。上式において  $J_{ij} = \frac{\partial f(x_i, \beta)}{\partial \beta_j}$ 。最小二乗推定量は以下の式で与えられる。

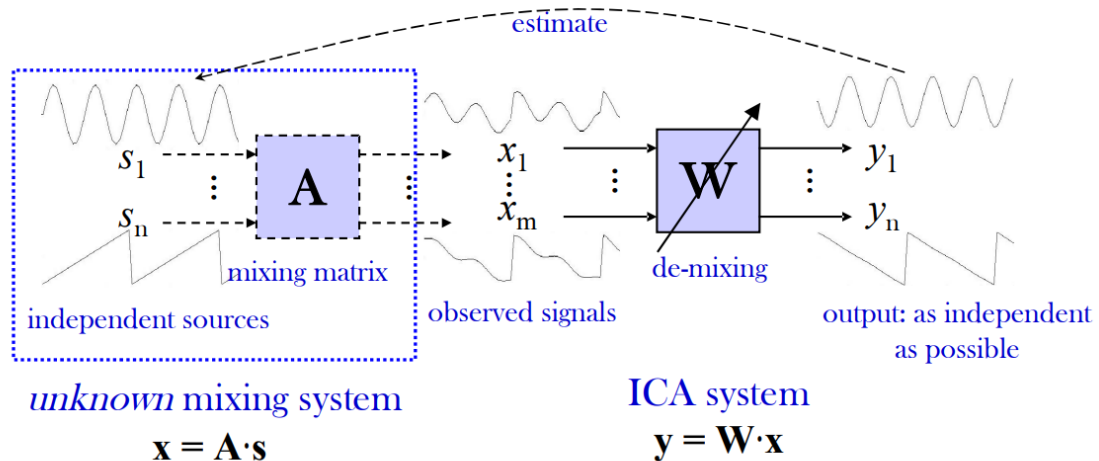
$$\hat{\beta} \approx (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \mathbf{y}.$$

非線形回帰統計値は計算され、線形回帰統計値と同じように使われるが、式では  $\mathbf{X}$  の位置に  $\mathbf{J}$  を使用している。線形近似はこの統計値へ偏り（バイアス）を導入する。したがって、非線形モデルから導き出された統計値を解釈するためには普段よりもより注意が必要である。

<https://ja.wikipedia.org/wiki/%E9%9D%9E%E7%B7%9A%E5%BD%A2%E5%9B%9E%E5%B8%B0>



## Linear ICA: separable under weak assumptions

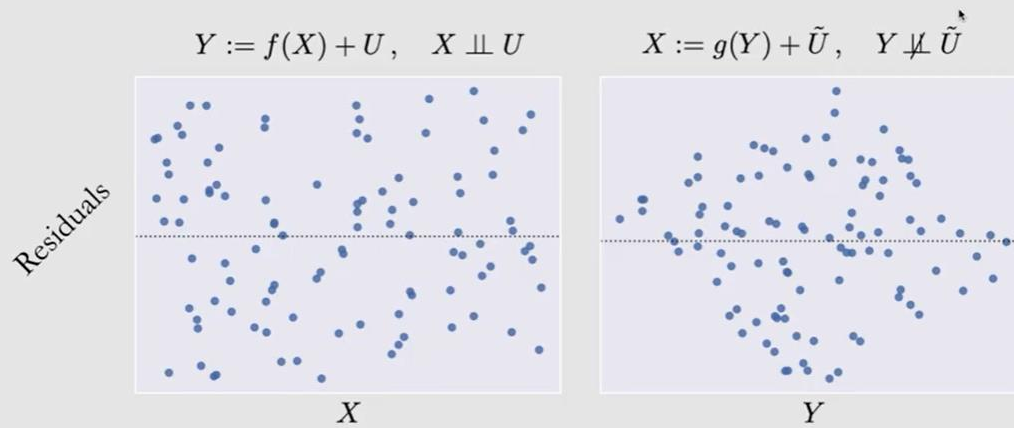


## Nonlinear ICA: $\mathbf{A}$ and $\mathbf{W}$ become invertible nonlinear mappings

- not separable:  $y_i$  may be totally different from  $s_i$



## Identifiability in Linear Non-Gaussian Setting: Residuals



### 着目点① – 非線形な手法

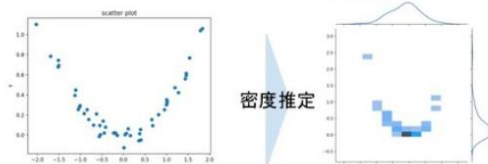
相互情報量をベースとするもの

離散

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

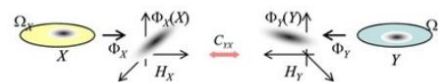
連続

$$I(X; Y) = \int_Y \int_X p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy$$



連続な場合は適当な分割で離散化した上で計算をする、など

データ点を別の空間に射影した上で共分散  
(のようなもの)を評価する



$$C_{XY} := \mathbb{E}_{x,y}[(\phi(x) - \mu_x) \otimes (\psi(y) - \mu_y)],$$

$$\text{HSIC}(p_{xy}, \mathcal{F}, \mathcal{G}) := \|C_{XY}\|_{\text{HS}}^2,$$

内積計算がカーネル関数の評価で行える空間で  
共分散作用素のノルムを独立性の指標とする  
( $\Rightarrow$  HSIC)

### Noise models [edit]

Incorporate an independent noise term in the model to compare the evidences of the two directions.

Here are some of the noise models for the hypothesis  $Y \rightarrow X$  with the noise  $E$ :

- Additive noise:<sup>[9]</sup>  $Y = F(X) + E$
- Linear noise:<sup>[10]</sup>  $Y = pX + qE$
- Post-nonlinear:<sup>[11]</sup>  $Y = G(F(X) + E)$
- Heteroskedastic noise:  $Y = F(X) + E \cdot G(X)$
- Functional noise:<sup>[12]</sup>  $Y = F(X, E)$

[https://en.wikipedia.org/wiki/Causal\\_inference](https://en.wikipedia.org/wiki/Causal_inference)

## 背景: Granger 因果性

- 変数 A の予測誤差が, 予測モデルに変数 B を入れると減少するとき「B から A の Granger 因果性が存在する」という

– 通常 VAR モデルの OLS 残差を比較検討する (Causality test)

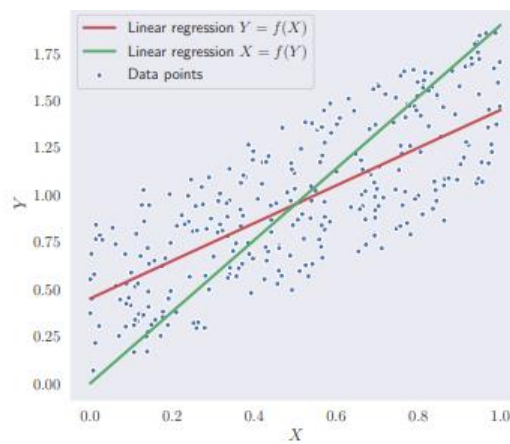
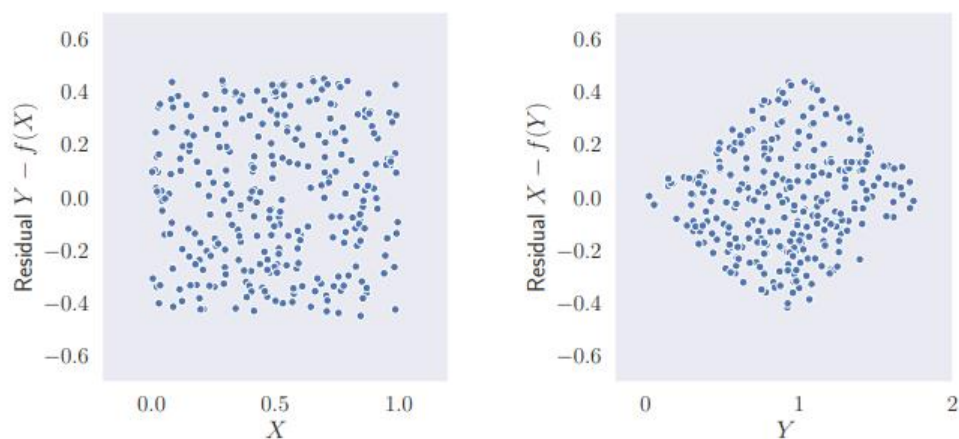


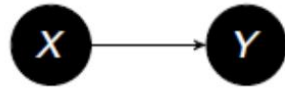
Figure 2.5



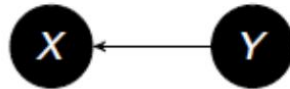


If two variables  $X$  and  $Y$  are statistically dependent then either

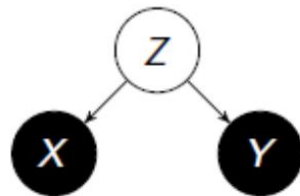
1



2



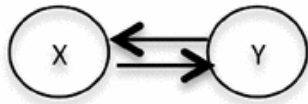
3



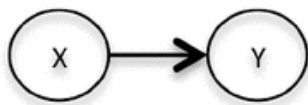
THE DEATH OF SHERLOCK HOLMES.



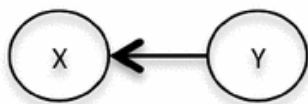
(a)



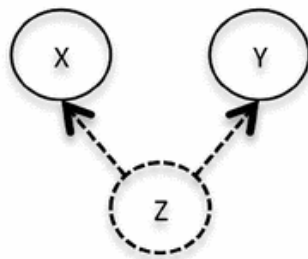
(b)



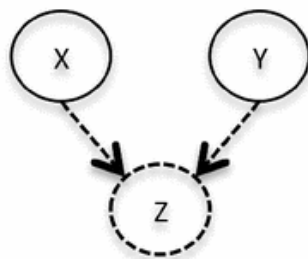
(c)



(d)



(e)



(f)

[Tell cause from effect: models and evaluation](#)

<http://www.ar.sanken.osaka-u.ac.jp/~sshimizu/papers/JJSS08.pdf>  
観測変数を標準化しておくことも、尺度不変になりよいと思われる。

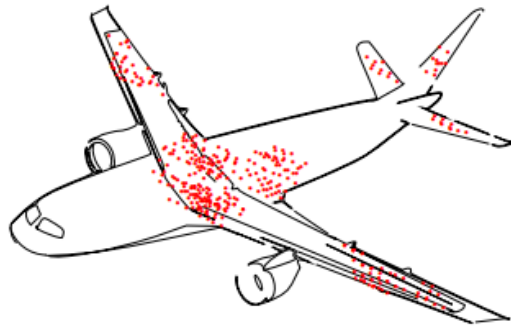


Figure 1.2 – The damaged parts (in red) of returning airplanes show locations where they can support damage and still return home.

## 2. Slope-based IGCI:

$$\hat{V}_{X \rightarrow Y} := \frac{1}{m-1} \sum_{i=1}^{m-1} \log \left| \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \right| \quad (3.5)$$

---

### Algorithm 4 Slope-based IGCI

---

**Input:** Observational data  $X$  and  $Y$ .

**Output:** The estimated result “ $X$  causes  $Y$ ” or “ $Y$  causes  $X$ ”.

Calculate the estimator  $V_{X \rightarrow Y}$  by  $\hat{V}_{X \rightarrow Y} := \frac{1}{m-1} \sum_{i=1}^{m-1} \log \left| \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \right|$ .

**if**  $\hat{V}_{X \rightarrow Y} < 0$  **then**

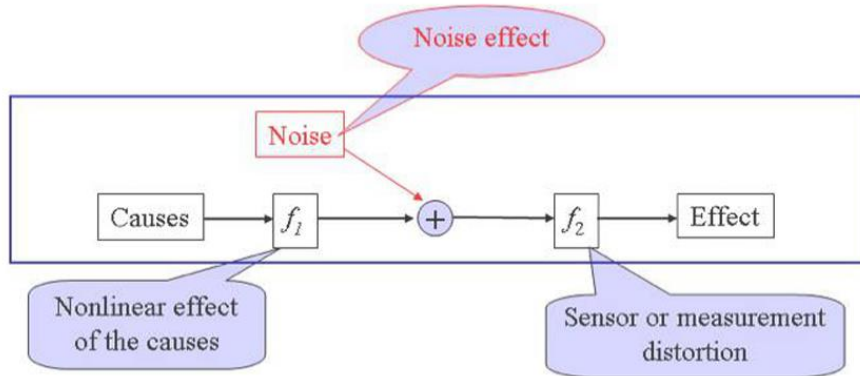
Infer that  $X$  causes  $Y$ .

**else**

Infer that  $Y$  causes  $X$ .

**end if**

---



chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://cra.org/ccc/wp-content/uploads/sites/2/2016/11/Slides\_AAAI\_symposium16.compressed.pdf











