スパース独立成分分析による構造的因果モデルの推定

原田　和治　　総合研究大学院大学　統計科学専攻　5年一貫性博士課程5年，学術振興会特別研究員（DC2）
藤澤　洋徳　　統計数理研究所，総合研究大学院大学，理研AIPセンター

本研究の貢献

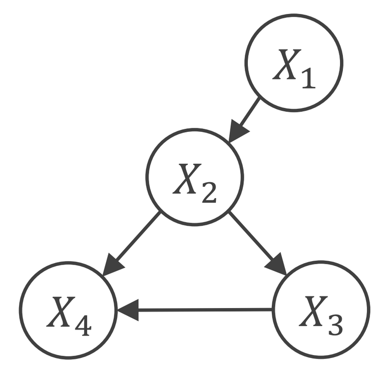
- 線形・非ガウスを仮定した統計的因果探索モデル"LiNGAM"に，新たな推定法を提案した.
- パラメータ推定の一致性について考慮した方法は，スパース独立成分分析に基づくアプローチとしては初めて.
- 交互方向乗数法（ADMM）に基づく推定アルゴリズムを導出した.
  特に，正則行列に対する自然勾配法を本研究の設定に合わせて修正することで，効率的な推定を可能にした.
- チューニングパラメータの決定を含む推定プロセス全体を提案した.

背景・目的

本研究は，因果構造がスパースなデータに対する統計的因果探索に注目する.

LiNGAM (Shimizu+ 2006)は識別可能な構造的因果モデルのひとつで，独立成分分析（ICA）を利用して識別性が証明されている. LiNGAMは以下を仮定する.

- 因果グラフの**非巡回性**
- 変数間の生成構造の**線形性**
- 誤差分布は**非ガウス分布**



LiNGAMの推定法には2つの問題がある.

1. **因果構造（順序）の推定にスパース性を利用していない**
2. **パラメータ推定に非ガウス性を利用していない**

これらを解決する手法として，ICAの罰則付き尤度に基づく方法が知られている（Zhang+ 2006, Zhang+ 2009）が，これらはICAの一致性条件を満たさず，不十分.

本研究の目的は，ICAの罰則付き尤度に基づく方法で前述の問題点を解消し，かつICAの一致性条件をケアした方法の構築である.

提案手法（sICA-LiNGAM）

提案法の目標

① 非ガウス性をパラメータ推定にも利用
② スパース性を因果順序の推定にも利用
（①②：ICAの罰則付き尤度を利用）
③ ICAの一致性条件をケアした安定的な推定法を構築
④ スパース性とパラメータの直交性の両立

推定法（罰則付き最尤法）

- 観測変数：$\mathbf{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_N)^T \in \mathbb{R}^{N\times d}$（中心化・標準化済み）
  白色化：$\mathbf{Z} = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_N)^T \in \mathbb{R}^{N\times d}, \mathbf{Z} = \mathbf{X}\mathbf{V}\mathbf{D}^{-1}$, $\frac{1}{N}\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{D}^2\mathbf{V}^T$（固有値分解）
- 構造方程式：$\mathbf{X} = \mathbf{X}\mathbf{B}^T + \mathbf{N}, \quad \mathbf{N} = \mathbf{X}(\mathbf{I} - \mathbf{B}^T)$
- ICAモデル：$\mathbf{Y} = \mathbf{X}\mathbf{M}^{\mathrm{T}} = \mathbf{Z}\mathbf{V}^T\mathbf{D}\mathbf{M}^T = \mathbf{Z}\mathbf{W}^T$（$\mathbf{W} = \mathbf{M}\boldsymbol{V}\boldsymbol{D}$）
- 対数尤度関数（定数項を除く）： ①
  $\ell(\mathbf{W};\mathbf{X}) = \frac{1}{N}\sum_{i=1}^N \sum_{j=1}^d \log \tilde{p}_j\left(\boldsymbol{w}_j^T \boldsymbol{z}_i\right) + \log|\det \mathbf{W}|$
- スパース罰則：$\mathcal{P}_\gamma(\mathbf{M}) = \mathcal{P}_\gamma\left(\mathbf{W}\mathbf{D}^{-1}\mathbf{V}^T\right)$ ②
- $\mathcal{N}(d)$：**行ノルムが1**であるような正則行列の全体
- 最適化問題（$\lambda > 0, \alpha \in [0,1]$はチューニングパラメータ）：

$$\min_{\mathbf{W}\in\mathcal{N}(d),\mathbf{P}\in\mathbb{R}^{d\times d}} -\ell(\mathbf{W};\mathbf{X}) + \lambda\left(\alpha\mathcal{P}_\gamma\left(\mathbf{W}\mathbf{D}^{-1}\mathbf{V}^T\right) + \frac{1-\alpha}{2}\left\|\mathbf{P}^T\mathbf{W} - \mathbf{I}\right\|_F^2\right)$$
$$\text{subject to } \mathbf{P}^T\mathbf{P} = \mathbf{I}$$

③④ 白色化データの下で，パラメータ行列**W**が直交ならば一致性を持つ. **W**が直交に近づくよう，罰則を課した.
λ: 固定，α: CVで選択

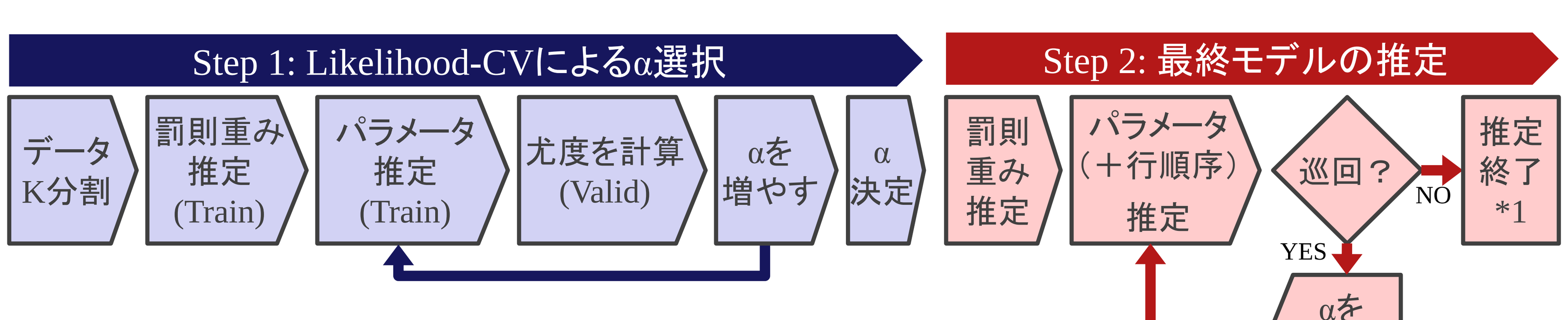
最適化法

Pと**W**を交互に更新する.

Pは**W**の特異値分解により$\mathbf{W} = \tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{V}}^T$とし，$\mathbf{P} = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^T$により更新する.

Wの更新は$\mathcal{L}_{\rho,\lambda}$を拡張ラグランジアンとし，**交互方向乗数法（ADMM）に行ノルムが1である行列の自然勾配法を組み合わせた方法**を用いる.

$$\begin{cases} \mathbf{W}_{t+1} = \mathrm{argmin}_{\mathbf{W}\in\mathcal{N}}\, \mathcal{L}_{\rho,\lambda}(\mathbf{W}, \mathbf{M}_t, \mathbf{U}_t) & \rightarrow \text{自然勾配法}(\mathbf{W}_{t+1} \leftarrow \mathbf{W}_t - \eta\Delta\mathbf{W})\text{で更新} \\ \mathbf{M}_{t+1} = \mathrm{argmin}_{\mathbf{M}\in\mathbb{R}^{d\times d}}\, \mathcal{L}_{\rho,\lambda}(\mathbf{W}_{t+1}, \mathbf{M}, \mathbf{U}_t) & \rightarrow \text{Closed Formの更新式（軟閾値作用素）} \\ \mathbf{U}_{t+1} = \mathbf{U}_t + \rho\left(\mathbf{W}_{t+1}\mathbf{D}^{-1}\mathbf{V}^T - \mathbf{M}_{t+1}\right) & \end{cases}$$



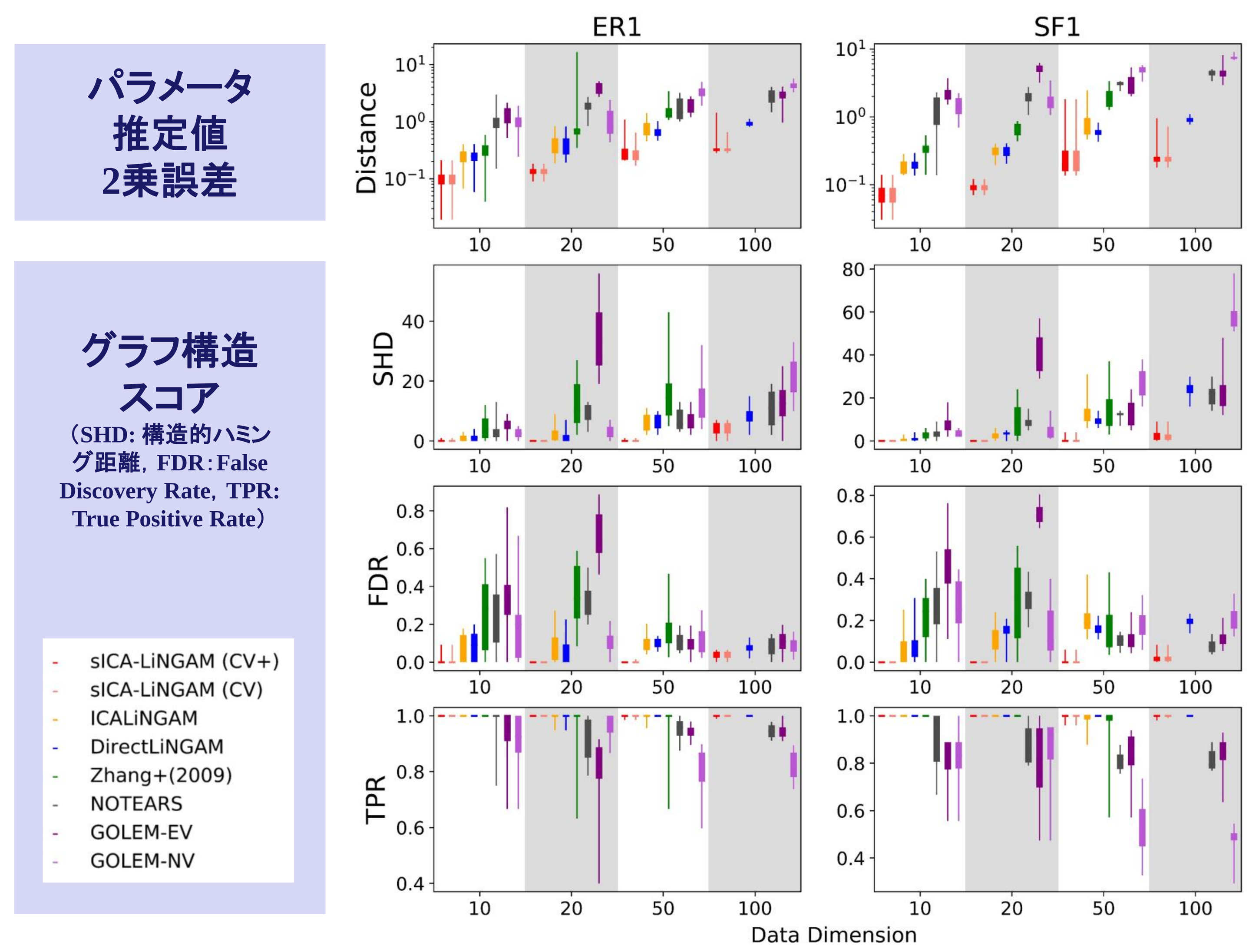
*1 必要に応じ，擬陽性を減らすためのカットオフを行う

数値実験

既存研究（Shimizu+ 2012, Zhang+ 2018）をもとに数値実験を設計した.

- ランダムグラフ（ER），スケールフリーグラフ（SF）を生成
- 非ガウス分布（指数分布，ラプラス分布，一様分布から乱択）に従うノイズを発生させ，グラフの隣接行列を用いて観測変数を生成
- データの次元は$d = 10,20,50,100$，サンプルサイズは$N = 1{,}000$
- 各設定10個ずつデータセットを作成し，各評価指標のBox plotで評価

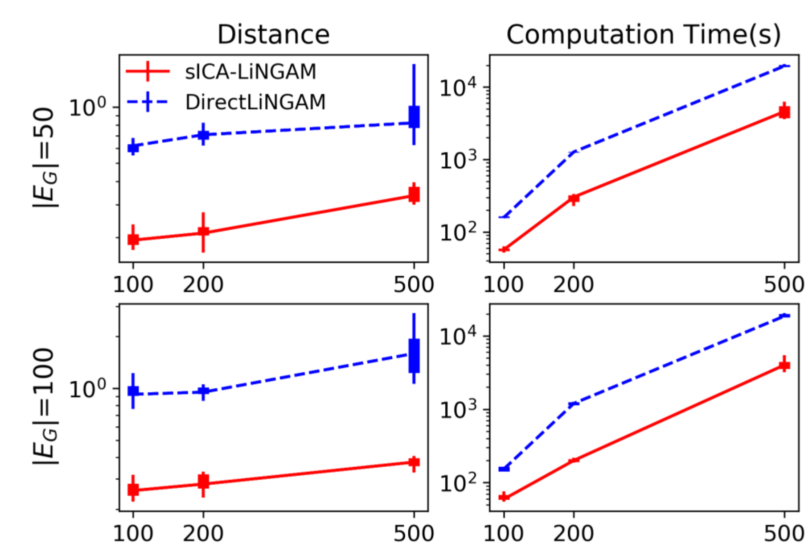
提案手法（sICA-LiNGAM）は，一貫して比較手法よりも高い性能を示した



高次元への拡張性

$N = 1{,}000$のまま，$d = 500$まで次元を拡大し，パラメータ推定性能と計算時間を評価した. 簡単のため，分布はラプラス分布のみ，辺の数は50または100とした. また，提案手法のαは固定とした.

提案手法は高次元でも安定した推定性能を示した. 計算速度は既存手法と同等だった.

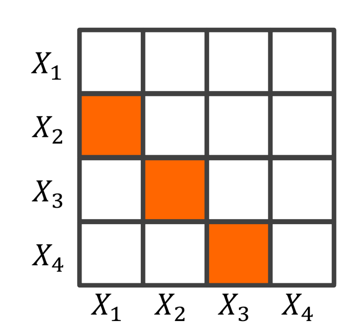


実データへの適用

Shimizu+ (2006)を参考に，時系列の復元性能をパラメータ行列**B**のヒートマップで視覚的に評価した.

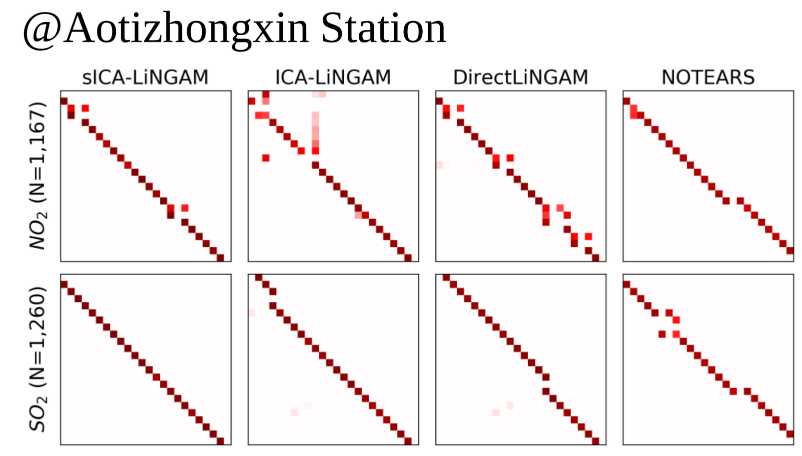
4変数AR(1)モデル

$X_1 = \epsilon_1$
$X_2 = \beta_{21}X_1 + \epsilon_2$
$X_3 = \beta_{32}X_2 + \epsilon_3$
$X_4 = \beta_{43}X_3 + \epsilon_4$



Beijing MultiSite Air-Quality Data (Zhang+ 2017) の1時間ごとの観測値を24時間ごとに区切り，24次元のデータセットを作成した. 各手法により時系列構造の復元を試みた.

@Aotizhongxin Station



提案手法は精度よく時系列を復元した.

参考文献

Harada, Kazuharu, and Hironori Fujisawa. "Estimation of Structural Causal Model via Sparsely Mixing Independent Component Analysis." arXiv preprint arXiv:2009.03077 (2020).