Vol.2014-CG-157 No.16
Vol.2014-CVIM-194 No.16
2014/11/21

因果指標と偏正準相関分析

渋谷 崇[1,†1,a)]　山下 裕也[1,†1,b)]　椋田 悠介[1,c)]　原田 達也[1,d)]

概要：本研究では，Granger Causality や Transfer Entropy などの因果指標と相互情報量の関係性について，情報理論や統計学の知識を用いて解釈を行い，統一的に定式化する．特に情報理論に基づく因果指標である Transfer Entropy は，時系列の確率変数がガウス分布に従うときに偏正準相関分析に帰着されることを示す．この偏正準相関分析をカーネル化することで様々な計量を考慮することができる．また，正則化，確率的モデル化などの拡張を行うことで因果指標の安定した計算方法を提案する．

1. はじめに

経済や物理などの分野において時系列予測は重要な課題である．実世界には多種多様な事象が存在し，それらがお互いに影響しあい変化している．特に時系列データの解析では，データの変動を予測・制御するために，データ間の因果関係を発見することが必要とされている．以上の背景に基づいて，本論文では因果指標と呼ばれる時系列データ間の因果性を定量化する手法について検討・提案を行う．特に情報理論に基づく因果指標である Transfer Entropy [1] が，時系列の確率変数がガウス分布に従うときに偏正準相関分析に帰着されることを示す．この偏正準相関分析をカーネル化，正則化，確率的モデル化などの拡張を行うことにより，様々な計量を因果指標に適用可能，少サンプルであっても安定して計算可能，確率的観点からのパラメタ推定などのメリットが生まれる．

2. 従来の因果指標

2.1 Transfer Entropy

Transfer Entropy [1] は一方の事象がもう一方の事象に与えている情報量を算出することで因果性の強さを計るという，情報理論に基づいた指標である．X, Y は時刻 t において $\boldsymbol{x}_t$, $\boldsymbol{y}_t$ という多次元データをもつ時系列の事象とする．これらの時系列データはそれぞれオーダー k, l の定常マルコフ過程と近似できると仮定する．このとき，Y のダイナミクスは遷移確率 $p(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}^{(l)})$ で表現できる．ここで $\boldsymbol{y}_{t-1}^{(l)} = (\boldsymbol{y}_{t-1},\ \ldots,\ \boldsymbol{y}_{t-l})^T$ で，埋め込みベクトルと呼ばれる．過去の状態 $\boldsymbol{y}_{t-1}^{(l)}$ が既知のもとで現在の状態 $\boldsymbol{y}_t$ が与えられたときの平均情報量は以下の式で与えられる．

$$\iint p(\boldsymbol{y}_t, \boldsymbol{y}_{t-1}^{(l)}) \log_2 \frac{1}{p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1}^{(l)})} d\boldsymbol{y}_t d\boldsymbol{y}_{t-1}^{(l)}. \tag{1}$$

一方，真の遷移確率 $p(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}^{(l)})$ は一般に未知で，事前知識を利用して遷移確率を推定する必要がある．$p(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}^{(l)})$ の代わりに推定遷移確率 $q(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}^{(l)})$ を用いた場合の 2 つの確率分布の差を情報量で表現したものが Kullback-Leibler Information である．

$$\iint p(\boldsymbol{y}_t, \boldsymbol{y}_{t-1}^{(l)}) \log_2 \frac{p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1}^{(l)})}{q(\boldsymbol{y}_t|\boldsymbol{y}_{t-1}^{(l)})} d\boldsymbol{y}_t d\boldsymbol{y}_{t-1}^{(l)}. \tag{2}$$

さて，Y の現在の状態 $\boldsymbol{y}_t$ が X の過去の状態に依存しない場合，次式が成り立つ．

$$p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1}^{(l)}, \boldsymbol{x}_{t-1}^{(k)}) = p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1}^{(l)}). \tag{3}$$

もし X から Y に因果的な影響がある場合は上の式は成り立たないことから，上の式の差をとることで X から Y への影響の強さを測ることができる．よって，Kullback-Leibler Information において，$p(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}^{(l)},\ \boldsymbol{x}_{t-1}^{(k)})$ を真の遷移確率, $p(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}^{(l)})$ を推定遷移確率と考える．このように定義された X から Y への情報量の流れを Transfer Entropy（以下 TE と略す）と呼び, 以下の式で与えられる．

$$T_{\boldsymbol{x}\to\boldsymbol{y}} = \iiint p(\boldsymbol{y}_t,\ \boldsymbol{y}_{t-1}^{(l)},\ \boldsymbol{x}_{t-1}^{(k)}) \times \log_2 \frac{p(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}^{(l)},\ \boldsymbol{x}_{t-1}^{(k)})}{p(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}^{(l)})} d\boldsymbol{y}_t d\boldsymbol{y}_{t-1}^{(l)} d\boldsymbol{x}_{t-1}^{(k)}. \tag{4}$$

TE は条件付き相互情報量の一種であり，$\boldsymbol{y}_{t-1}$ が与えられ

[1] 東京大学 大学院情報理工学系研究科
7-3-1 Hongo Bunkyo-ku, Tokyo 113–8656, Japan
[†1] 現在，ソニー株式会社
Presently with Sony Corporation
[a)] takashi@isi.imi.i.u-tokyo.ac.jp
[b)] yamashita@isi.imi.i.u-tokyo.ac.jp
[c)] mukuta@mi.t.u-tokyo.ac.jp
[d)] harada@mi.t.u-tokyo.ac.jp

た下での条件付きエントロピーとして表現できる．

$$T_{\boldsymbol{x}\to\boldsymbol{y}} = \iiint p(\boldsymbol{y}_t, \boldsymbol{y}_{t-1}^{(l)}, \boldsymbol{x}_{t-1}^{(k)}) \times \log_2 \frac{p(\boldsymbol{x}_{t-1}^{(k)}, \boldsymbol{y}_t | \boldsymbol{y}_{t-1}^{(l)})}{p(\boldsymbol{x}_{t-1}^{(k)} | \boldsymbol{y}_{t-1}^{(l)}) p(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}^{(l)})} d\boldsymbol{y}_t d\boldsymbol{y}_{t-1}^{(l)} d\boldsymbol{x}_{t-1}^{(k)} = H_{\boldsymbol{x}_{t-1}^{(k)} | \boldsymbol{y}_{t-1}^{(l)}} + H_{\boldsymbol{y}_t | \boldsymbol{y}_{t-1}^{(l)}} - H_{\boldsymbol{x}_{t-1}^{(k)} \otimes \boldsymbol{y}_t | \boldsymbol{y}_{t-1}^{(l)}}. \quad (5)$$

2.2 Continuous Transfer Entropy

時系列データがガウス分布に従うと仮定した場合の TE（$T_{\boldsymbol{x}\to\boldsymbol{y}}$）は次のように表される．

$$\frac{1}{2}\log_2 \frac{\left|\Sigma_{\{\boldsymbol{y}_{t-1}^{(l)}, \boldsymbol{x}_{t-1}^{(k)}\}\{\boldsymbol{y}_{t-1}^{(l)}, \boldsymbol{x}_{t-1}^{(k)}\}}\right| \left|\Sigma_{\{y_t, \boldsymbol{y}_{t-1}^{(l)}\}\{y_t, \boldsymbol{y}_{t-1}^{(l)}\}}\right|}{\left|\Sigma_{\{y_t, \boldsymbol{y}_{t-1}^{(l)}, \boldsymbol{x}_{t-1}^{(k)}\}\{y_t, \boldsymbol{y}_{t-1}^{(l)}, \boldsymbol{x}_{t-1}^{(k)}\}}\right| \left|\Sigma_{\boldsymbol{y}_{t-1}^{(l)}\boldsymbol{y}_{t-1}^{(l)}}\right|}, \quad (6)$$

ここで，$\Sigma_{\boldsymbol{y}_{t-1}^{(l)}\boldsymbol{y}_{t-1}^{(l)}}$ は確率分布 $p\left(\boldsymbol{y}_{t-1}^{(l)}\right)$ の共分散行列で，$\left|\Sigma\right|$ はその行列式を意味する．これは Continuous Transfer Entropy（以下 CTE と略す）と呼ばれている [2].

2.3 Granger Causality

Granger Causality（以下 GC と略す）[3] は，時系列データの予測を行う際の，自身のデータのみを用いた自己回帰モデルと，自身のデータの他にもう 1 つ別のデータを加えた混合回帰モデルとの精度を比較する手法である．

まず，自己回帰モデル（AR モデル）を考える．

$$y_t = a_0 + \boldsymbol{a}^T \boldsymbol{y}_{t-1}^{(l)} + \epsilon_t^{(y)}, \quad (7)$$

ここで，a_0 は定数項，$\boldsymbol{a} \in \boldsymbol{R}^l$ は最小二乗法により算出される回帰係数である．また，$\epsilon_t^{(y)}$ は平均 0，分散 σ_y^2 の正規分布に従う誤差である．

それとは別に次のような混合回帰モデルを考える．

$$y_t = b_0 + \boldsymbol{b}^T \boldsymbol{y}_{t-1}^{(l)} + \boldsymbol{c}^T \boldsymbol{x}_{t-1}^{(k)} + \epsilon_t^{(y|x)}. \quad (8)$$

AR モデルと同様，b_0 は定数項, $\boldsymbol{b} \in \boldsymbol{R}^l$, $\boldsymbol{c} \in \boldsymbol{R}^k$ は最小二乗法により算出される回帰係数である．また，$\epsilon_t^{(y|x)}$ は平均 0，分散 $\sigma_{y|x}^2$ の正規分布に従う誤差である．

回帰モデルのモデリング精度は, 誤差の分散によって評価できる．つまり，σ_y^2 と $\sigma_{y|x}^2$ を比較することで X から Y への因果性を計ることができる．Y のモデリング精度が X を回帰モデルに取り込むことによって向上し，分散が小さくなった ($\sigma_y^2 > \sigma_{y|x}^2$) ならば，$X$ は Y に影響を持つと言える．しかし，σ_y^2 と $\sigma_{y|x}^2$ との比較方法は $\sigma_y^2 - \sigma_{y|x}^2$ や $\sigma_{y|x}^2/\sigma_y^2$ など複数の方法が考えられ，実際に様々な方法が用いられている．

3. 因果指標の関係性と拡張

本章では因果指標の関係性を統一的な解釈を与，因果指標の拡張を考える．

3.1 Transfer Entropy と Granger Causality

ここでは TE と GC の関係性を述べる [4]．X, Y が連続値データであるとする．時系列データ Y が式 (7) の AR モデルに従うと仮定した場合，y_t の条件付き確率分布 $p(y_t \mid \boldsymbol{y}_{t-1}^{(l)})$ は以下の式で与えられる．

$$\frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{\left(y_t - \boldsymbol{a}^T \boldsymbol{y}_{t-1}^{(l)}\right)^2}{2\sigma_y^2}\right). \quad (9)$$

また，AR モデルでは $\boldsymbol{y}_{t-1}^{(l)}$ の確率分布は多次元正規分布となる．いま，$\boldsymbol{y}_{t-1}^{(l)}$ の平均を $\boldsymbol{\mu}_{\boldsymbol{y}_{t-1}^{(l)}}$，分散共分散行列を $\Sigma_{\boldsymbol{y}_{t-1}^{(l)}\boldsymbol{y}_{t-1}^{(l)}}$ とすると，同時確率分布 $p(y_t,\ \boldsymbol{y}_{t-1}^{(l)})$ は以下のような多次元正規分布となる．

$$\frac{1}{\sqrt{(2\pi)^{l+1}\sigma_y^2 \left|\Sigma_{\boldsymbol{y}_{t-1}^{(l)}\boldsymbol{y}_{t-1}^{(l)}}\right|}} \exp\left(-\frac{1}{2}{\boldsymbol{\psi}_t^{(l+1)}}^T X \boldsymbol{\psi}_t^{(l+1)}\right) \quad (10)$$

$$\text{where} \quad \boldsymbol{\psi}_t^{(l+1)} = \begin{bmatrix} y_t - \boldsymbol{a}^T \boldsymbol{\mu}_{\boldsymbol{y}_{t-1}^{(l)}} \\ \boldsymbol{y}_{t-1}^{(l)} - \boldsymbol{\mu}_{\boldsymbol{y}_{t-1}^{(l)}} \end{bmatrix}, \quad (11)$$

$$X = \begin{bmatrix} \frac{1}{\sigma_y^2} & -\frac{\boldsymbol{a}^T}{\sigma_y^2} \\ -\frac{\boldsymbol{a}}{\sigma_y^2} & \frac{\boldsymbol{a}\boldsymbol{a}^T}{\sigma_y^2} + \Sigma^{-1}_{\boldsymbol{y}_{t-1}^{(l)}\boldsymbol{y}_{t-1}^{(l)}} \end{bmatrix}. \quad (12)$$

この分布の分散共分散行列の行列式は $\sigma_y^2 \left|\Sigma_{\boldsymbol{y}_{t-1}^{(l)}\boldsymbol{y}_{t-1}^{(l)}}\right|$ となる．よって，以下の式が成り立つ．

$$\left|\Sigma_{\{y_t, \boldsymbol{y}_{t-1}^{(l)}\}\{y_t, \boldsymbol{y}_{t-1}^{(l)}\}}\right| = \sigma_y^2 \left|\Sigma_{\boldsymbol{y}_{t-1}^{(l)}\boldsymbol{y}_{t-1}^{(l)}}\right|. \quad (13)$$

$\{y_t, \boldsymbol{y}_{t-1}^{(l)}, \boldsymbol{x}_{t-1}^{(k)}\}$ の同時確率分布についても，式 (8) の混合回帰モデルを仮定すると，以下の式が成り立つ．

$$\left|\Sigma_{\{y_t, \boldsymbol{y}_{t-1}^{(l)}, \boldsymbol{x}_{t-1}^{(k)}\}\{y_t, \boldsymbol{y}_{t-1}^{(l)}, \boldsymbol{x}_{t-1}^{(k)}\}}\right| = \sigma_{y|x}^2 \left|\Sigma_{\{\boldsymbol{y}_{t-1}^{(l)}, \boldsymbol{x}_{t-1}^{(k)}\}\{\boldsymbol{y}_{t-1}^{(l)}, \boldsymbol{x}_{t-1}^{(k)}\}}\right|. \quad (14)$$

回帰モデルに従う時系列データの同時確率分布は多次元正規分布になることから，式 (13) および式 (14) を CTE の定義式である式 (6) に代入すると，以下の式が導出される．

$$T_{\boldsymbol{x}\to\boldsymbol{y}} = \frac{1}{2}\log_2 \frac{\sigma_y^2}{\sigma_{y|x}^2}. \quad (15)$$

この指標は 2 つの回帰モデルのモデリング精度を比較するという枠組みの中で，X が Y に与えている平均情報量を算出した指標となっている．

3.2 Transfer Entropy と偏正準相関分析

ここでは，TE が偏正準相関分析に帰着されることを示す [5]．確率変数 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ に対して，偏相関係数の考え方と同様にして $\boldsymbol{z}$ の影響を除いた $\boldsymbol{x}$ と $\boldsymbol{y}$ で正準相関分析を行うのが偏正準相関分析（Partial Canonical Correlation

Analysis，以下 PCCA と略す）[6] である．PCCA は以下の一般化固有値問題を解くことで求められる

$$\Sigma_{\boldsymbol{xy}|\boldsymbol{z}}\Sigma^{-1}_{\boldsymbol{yy}|\boldsymbol{z}}\Sigma_{\boldsymbol{yx}|\boldsymbol{z}}\boldsymbol{a} = \lambda\Sigma_{\boldsymbol{xx}|\boldsymbol{z}}\boldsymbol{a}, \tag{16}$$

$$\Sigma_{\boldsymbol{yx}|\boldsymbol{z}}\Sigma^{-1}_{\boldsymbol{xx}|\boldsymbol{z}}\Sigma_{\boldsymbol{xy}|\boldsymbol{z}}\boldsymbol{b} = \lambda\Sigma_{\boldsymbol{yy}|\boldsymbol{z}}\boldsymbol{b}. \tag{17}$$

ここで，$\Sigma_{\boldsymbol{uv}|\boldsymbol{w}} = \Sigma_{\boldsymbol{uv}} - \Sigma_{\boldsymbol{uw}}\Sigma^{-1}_{\boldsymbol{ww}}\Sigma_{\boldsymbol{wv}}$ であり，$\Sigma_{\boldsymbol{uv}}$ は $\boldsymbol{u}$ と $\boldsymbol{v}$ との共分散行列である．式 (16) と (17) は共通の固有値を持つ．

まず，2 つの時系列の確率変数 $\boldsymbol{x}$ と $\boldsymbol{y}$ の次元をそれぞれ $D_{\boldsymbol{x}}$ と $D_{\boldsymbol{y}}$ する．また 2 つの変数ともにガウス分布に従うとする．この時，各ガウス分布のエントロピーは以下のように表現される．

$$H_{\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}} = \frac{kD_{\boldsymbol{x}}}{2}\log_2(2\pi e) + \frac{1}{2}\log_2\left|\Sigma_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}}\right|, \tag{18}$$

$$H_{\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}} = \frac{D_{\boldsymbol{y}}}{2}\log_2(2\pi e) + \frac{1}{2}\log_2\left|\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}}\right|, \tag{19}$$

$$H_{\boldsymbol{x}^{(k)}_{t-1}\otimes\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}} = \frac{kD_{\boldsymbol{x}}+D_{\boldsymbol{y}}}{2}\log_2(2\pi e) + \frac{1}{2}\log_2\left|\Sigma_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}}\right|\left|\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t|\{\boldsymbol{x}^{(k)}_{t-1},\boldsymbol{y}^{(l)}_{t-1}\}}\right|, \tag{20}$$

ここで，$\Sigma_{\boldsymbol{uu}|\{\boldsymbol{v},\boldsymbol{w}\}} = \Sigma_{\boldsymbol{uu}|\boldsymbol{w}} - \Sigma_{\boldsymbol{uv}|\boldsymbol{w}}\Sigma^{-1}_{\boldsymbol{vv}|\boldsymbol{w}}\Sigma_{\boldsymbol{vu}|\boldsymbol{w}} = \Sigma_{\boldsymbol{uu}|\boldsymbol{v}} - \Sigma_{\boldsymbol{uw}|\boldsymbol{v}}\Sigma^{-1}_{\boldsymbol{ww}|\boldsymbol{v}}\Sigma_{\boldsymbol{wu}|\boldsymbol{v}}$ である．式 (18), (19), と (20) を式 (5) に代入すると，以下の式を得る．

$$T_{\boldsymbol{x}\to\boldsymbol{y}} = \frac{1}{2}\log_2\frac{\left|\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}}\right|}{\left|\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t|\{\boldsymbol{x}^{(k)}_{t-1},\boldsymbol{y}^{(l)}_{t-1}\}}\right|} = \frac{1}{2}\log_2\frac{\left|\Sigma_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}}\right|}{\left|\Sigma_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{x}^{(k)}_{t-1}|\{\boldsymbol{y}_t,\boldsymbol{y}^{(l)}_{t-1}\}}\right|}. \tag{21}$$

ここで，$s_t = \boldsymbol{a}^T\boldsymbol{y}_t$ つまり $\boldsymbol{y}_t$ の線形写像に関して TE を最大化することを考える．この射影は多次元データから情報の流れを多く含む線形部分空間を求めることに対応する．式 (21) を用いることで，$\boldsymbol{x}$ から s_t への TE は次のように書くことができる．

$$T_{\boldsymbol{x}\to s_t} = \frac{1}{2}\log_2\frac{\boldsymbol{a}^T\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}}\boldsymbol{a}}{\boldsymbol{a}^T\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t|\{\boldsymbol{x}^{(k)}_{t-1},\boldsymbol{y}^{(l)}_{t-1}\}}\boldsymbol{a}} = \frac{1}{2}\log_2\frac{1}{1-\rho},$$

$$\rho = \frac{\boldsymbol{a}^T\Sigma_{\boldsymbol{y}_t\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}}\Sigma^{-1}_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}}\Sigma_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}}\boldsymbol{a}}{\boldsymbol{a}^T\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}}\boldsymbol{a}}. \tag{22}$$

$T_{\boldsymbol{x}\to s_t}$ と ρ の最大化は同じであることに注意されたい．ここで $\boldsymbol{a}^T\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}}\boldsymbol{a} = 1$ の拘束条件のもとで式 (22) の分子の最大化を考える．これに対応するラグランジュ関数は以下のように与えられる．

$$J = \boldsymbol{a}^T\Sigma_{\boldsymbol{y}_t\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}}\Sigma^{-1}_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}}\Sigma_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}}\boldsymbol{a} + \lambda\left(1 - \boldsymbol{a}^T\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}}\boldsymbol{a}\right), \tag{23}$$

ここで λ はラグランジュの未定乗数である．$\boldsymbol{a}$ と λ に関して，式 (23) の停留条件により以下の式が導かれる．

$$\Sigma_{\boldsymbol{y}_t\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}}\Sigma^{-1}_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}}\Sigma_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}}\boldsymbol{a} = \lambda\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}}\boldsymbol{a}. \tag{24}$$

これは一般化固有値問題であり，PCCA と等価となる．この問題を解くことによって，D_y 個の固有値と固有値に対応する固有ベクトル $\{\lambda_i, \boldsymbol{a}_i\}_{i=1}^{D_y}$，$(\lambda_1 \geq \cdots \geq \lambda_{D_y} \geq 0)$ を得る．よって TE は次のように表される．

$$T_{\boldsymbol{x}\to\boldsymbol{y}} = \sum_{i=1}^{D_y} T^{\{i\}}_{\boldsymbol{x}\to\boldsymbol{y}} = \sum_{i=1}^{D_y}\frac{1}{2}\log_2\frac{1}{1-\lambda_i}, \tag{25}$$

ここで $T^{\{i\}}_{\boldsymbol{x}\to\boldsymbol{y}}$ は i 番目の部分空間における TE の値である．よって，ガウス分布の仮定の下では，TE は PCCA の固有値を評価することと等価となる．

また，$\boldsymbol{x}^{(k)}_{t-1}$ の線形写像 $\boldsymbol{b}^T\boldsymbol{x}^{(k)}_{t-1}$ も考えられる．この場合，$\boldsymbol{b}^T\Sigma_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}}\boldsymbol{b} = 1$ の拘束条件を用いることで次の式が導かれる．

$$\Sigma_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}}\Sigma^{-1}_{\boldsymbol{y}_t\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}}\Sigma_{\boldsymbol{y}_t\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}}\boldsymbol{b} = \lambda\Sigma_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}}\boldsymbol{b}. \tag{26}$$

式 (24) と (26) は式 (16) と (17) と同様に，共通の固有値を持つ．そのために，行列の次元や要求される固有ベクトルによってどちらの問題を解くのかを選択することができる．固有ベクトル $\boldsymbol{a}$ と $\boldsymbol{b}$ は次の関係で表現される．

$$\boldsymbol{a}_i = \frac{1}{\sqrt{\lambda}}\Sigma^{-1}_{\boldsymbol{y}_t\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}}\Sigma_{\boldsymbol{y}_t\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}}\boldsymbol{b}_i,$$

$$\boldsymbol{b}_i = \frac{1}{\sqrt{\lambda}}\Sigma^{-1}_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}^{(l)}_{t-1}}\Sigma_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{y}_t|\boldsymbol{y}^{(l)}_{t-1}}\boldsymbol{a}_i.$$

3.3 Transfer Entropy と相互情報量

TE と時間遅れのある相互情報量（Time Delayed Mutual Information 以下 TDMI と略す）をシステマティックに評価する [5]．このために，TDMI を $\boldsymbol{x}_{t-k}$ の代わりに $\boldsymbol{x}^{(k)}_{t-1}$ を用いて拡張する．つまり，

$$M_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{y}_t} = \iint p(\boldsymbol{x}^{(k)}_{t-1},\boldsymbol{y}_t)\times\log_2\frac{p(\boldsymbol{x}^{(k)}_{t-1},\boldsymbol{y}_t)}{p(\boldsymbol{x}^{(k)}_{t-1})p(\boldsymbol{y}_t)}d\boldsymbol{x}^{(k)}_{t-1}d\boldsymbol{y}_t = \frac{1}{2}\log_2\frac{\left|\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t}\right|}{\left|\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t|\boldsymbol{x}^{(k)}_{t-1}}\right|} = \frac{1}{2}\log_2\frac{\left|\Sigma_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{x}^{(k)}_{t-1}}\right|}{\left|\Sigma_{\boldsymbol{x}^{(k)}_{t-1}\boldsymbol{x}^{(k)}_{t-1}|\boldsymbol{y}_t}\right|}. \tag{27}$$

TE と PCCA との関係と同様に，$s_t = \boldsymbol{a}^T\boldsymbol{y}_t$ に関して，TDMI を最大化することを考える．この時，式 (27) は次

のようになる．

$$M_{\boldsymbol{x}_{t-1}^{(k)}s_t} = \frac{1}{2}\log_2 \frac{\boldsymbol{a}^T\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t}\boldsymbol{a}}{\boldsymbol{a}^T\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t|\boldsymbol{x}_{t-1}^{(k)}}\boldsymbol{a}}. \tag{28}$$

$\boldsymbol{a}^T\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t}\boldsymbol{a}=1$ という拘束条件のもとで式 (28) を最大化することを考える．ラグランジュの未定乗数 λ を用いることで以下の式が得られる．

$$\Sigma_{\boldsymbol{y}_t\boldsymbol{x}_{t-1}^{(k)}}\Sigma^{-1}_{\boldsymbol{x}_{t-1}^{(k)}\boldsymbol{x}_{t-1}^{(k)}}\Sigma_{\boldsymbol{x}_{t-1}^{(k)}\boldsymbol{y}_t}\boldsymbol{a} = \lambda\Sigma_{\boldsymbol{y}_t\boldsymbol{y}_t}\boldsymbol{a}. \tag{29}$$

これも D_y 個の固有値とこれに対応する固有ベクトル $\{\lambda_i, \boldsymbol{a}_i\}_{i=1}^{D_y}$, $(\lambda_1 \geq \cdots \geq \lambda_{D_y} \geq 0)$ を持つ一般化固有値問題となる．この問題は正準相関分析（Canonical Correlation Analysis，以下 CCA と略す）と等価である．$\boldsymbol{x}_{t-1}^{(k)}$ の線形写像についても考えると次のようになる．

$$\Sigma_{\boldsymbol{x}_{t-1}^{(k)}\boldsymbol{y}_t}\Sigma^{-1}_{\boldsymbol{y}_t\boldsymbol{y}_t}\Sigma_{\boldsymbol{y}_t\boldsymbol{x}_{t-1}^{(k)}}\boldsymbol{b} = \lambda\Sigma_{\boldsymbol{x}_{t-1}^{(k)}\boldsymbol{x}_{t-1}^{(k)}}\boldsymbol{b}. \tag{30}$$

式 (29) と (30) は共通の固有値を持つ．これらの固有値を用いることで，以下の式を得る．

$$M_{\boldsymbol{x}_{t-1}^{(k)}\boldsymbol{y}_t} = \sum_{i=1}^{D_y} M^{\{i\}}_{\boldsymbol{x}_{t-1}^{(k)}\boldsymbol{y}_t} = \sum_{i=1}^{D_y}\frac{1}{2}\log_2\frac{1}{1-\lambda_i}, \tag{31}$$

ここで $M^{\{i\}}_{\boldsymbol{x}_{t-1}^{(k)}\boldsymbol{y}_t}$ は i 番目の部分空間の TDMI の値である．このようにガウス分布の仮定のもとでは，TDMI は CCA の固有値の評価と同等となる．

この時点で，TE と TDMI を一般化固有値問題という同じ視点から評価できるようになる．TDMI と TE はほぼ同じような構造を持つことが分かるが，$\boldsymbol{y}_{t-1}^{(l)}$ を事前情報として用いるかが異なる点となる．より詳しく言うと，確率密度分布としてガウス分布を仮定すると，TDMI は $\boldsymbol{x}_{t-1}^{(k)}$ と $\boldsymbol{y}_t$ の間のクロス相関を $\boldsymbol{y}_{t-1}^{(l)}$ を考慮せずに測っていることになる．一方，TE は $\boldsymbol{y}_{t-1}^{(l)}$ が与えられた下での $\boldsymbol{x}_{t-1}^{(k)}$ と $\boldsymbol{y}_t$ との偏相関を測っていることになる．つまり，$\boldsymbol{y}_{t-1}^{(l)}$ の影響が大きければ，TE と TDMI との差が大きくなる．

4. カーネル化

これまで議論してきたように，多次元時系列の因果指標の算出は CCA の一種である PCCA に帰着される．ここではカーネル PCCA を用いた因果指標を提案する．

主成分分析（PCA）や CCA にカーネルトリックを適用したカーネル PCA やカーネル CCA では，正定値カーネル関数 $k_{\boldsymbol{x}}(\boldsymbol{x}_n, \boldsymbol{x}_m)$ による N 個のデータ $\boldsymbol{x}_n$ の写像

$$\Phi : \boldsymbol{x}_n \to k_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{x}_n) \tag{32}$$

の線形結合作用素 $\sum_{n=1}^{N}\alpha_n k_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{x}_n)$ で構成されるベクトル空間, 再生核ヒルベルト空間において通常の PCA や CCA を行う. カーネル PCCA も同様に, 再生核ヒルベルト空間での PCCA として定義される.

ここで, 行列 $K_{\boldsymbol{x}}$ を第 (n,m) 要素が,

$$K_{\boldsymbol{x},nm} = k_{\boldsymbol{x}}(\boldsymbol{x}_n, \boldsymbol{x}_m) \tag{33}$$

で定められる $N\times N$ の対称行列として定義する. この行列はグラム行列と呼ばれる. グラム行列を用いると, データ $\boldsymbol{x}_n$ の再生核ヒルベルト空間における共分散作用素 $\hat{\Sigma}_{\boldsymbol{x}\boldsymbol{x}}$ は次の式で表される.

$$\hat{\Sigma}_{\boldsymbol{x}\boldsymbol{x}} = \tilde{K}_{\boldsymbol{x}}^2, \tag{34}$$

ただし $\tilde{K}_{\boldsymbol{x}} = K_{\boldsymbol{x}} - \mathbf{1}_N K_{\boldsymbol{x}} - K_{\boldsymbol{x}}\mathbf{1}_N + \mathbf{1}_N K_{\boldsymbol{x}}\mathbf{1}_N$ である．$\mathbf{1}_N$ は全ての要素が $1/N$ という値を取る $N\times N$ 行列である. データ $\boldsymbol{y}_n$ についても正定値カーネル関数 $k_{\boldsymbol{y}}(\boldsymbol{y}_n, \boldsymbol{y}_m)$ によって写像した再生核ヒルベルト空間を用いて考える. 行列 $K_{\boldsymbol{y}}$ を第 (n,m) 要素が,

$$K_{\boldsymbol{y},nm} = k_{\boldsymbol{y}}(\boldsymbol{y}_n, \boldsymbol{y}_m) \tag{35}$$

の対称行列とすると, 共分散作用素 $\hat{\Sigma}_{\boldsymbol{y}\boldsymbol{y}}$ および $\hat{\Sigma}_{\boldsymbol{x}\boldsymbol{y}}$ は同様に次の式で表される.

$$\hat{\Sigma}_{\boldsymbol{y}\boldsymbol{y}} = \tilde{K}_{\boldsymbol{y}}^2, \tag{36}$$

$$\hat{\Sigma}_{\boldsymbol{x}\boldsymbol{y}} = \tilde{K}_{\boldsymbol{x}}\tilde{K}_{\boldsymbol{y}}, \tag{37}$$

ただし $\tilde{K}_{\boldsymbol{y}} = K_{\boldsymbol{y}} - \mathbf{1}_N K_{\boldsymbol{y}} - K_{\boldsymbol{y}}\mathbf{1}_N + \mathbf{1}_N K_{\boldsymbol{y}}\mathbf{1}_N$ である．$k_{\boldsymbol{x}}(\boldsymbol{x}_n, \boldsymbol{x}_m)$ と $k_{\boldsymbol{y}}(\boldsymbol{y}_n, \boldsymbol{y}_m)$ は異なる正定値カーネル関数であってもかまわない. このとき, $\boldsymbol{x}_n$ が与えられたときの $\boldsymbol{y}_n$ の条件付き共分散作用素は,

$$\hat{\Sigma}_{\boldsymbol{y}\boldsymbol{y}|\boldsymbol{x}} = \hat{\Sigma}_{\boldsymbol{y}\boldsymbol{y}} - \hat{\Sigma}_{\boldsymbol{y}\boldsymbol{x}}\hat{\Sigma}^{-1}_{\boldsymbol{x}\boldsymbol{x}}\hat{\Sigma}_{\boldsymbol{x}\boldsymbol{y}} \tag{38}$$

$$= \left(\tilde{K}_{\boldsymbol{y}}^2 + \epsilon I_N\right) - \tilde{K}_{\boldsymbol{y}}\tilde{K}_{\boldsymbol{x}}\left(\tilde{K}_{\boldsymbol{x}}^2 + \epsilon I_N\right)^{-1}\tilde{K}_{\boldsymbol{x}}\tilde{K}_{\boldsymbol{y}} \tag{39}$$

となる. $\hat{\Sigma}_{\boldsymbol{x}\boldsymbol{x}}$ や $\hat{\Sigma}_{\boldsymbol{y}\boldsymbol{y}}$ は逆作用素を考える必要がある場合があるが, 一般にこれらの作用素は非可逆である. そこで $\hat{\Sigma}_{\boldsymbol{x}\boldsymbol{x}}$ や $\hat{\Sigma}_{\boldsymbol{y}\boldsymbol{y}}$ に対して正則化を行う.

カーネル PCCA は次の一般化固有値問題で定義される.

$$\hat{\Sigma}_{\boldsymbol{y}_t\boldsymbol{x}_{t-1}^{(k)}|\boldsymbol{y}_{t-1}^{(l)}}\hat{\Sigma}^{-1}_{\boldsymbol{x}_{t-1}^{(k)}\boldsymbol{x}_{t-1}^{(k)}|\boldsymbol{y}_{t-1}^{(l)}}\hat{\Sigma}_{\boldsymbol{x}_{t-1}^{(k)}\boldsymbol{y}_t|\boldsymbol{y}_{t-1}^{(l)}}\boldsymbol{\alpha} = \lambda\hat{\Sigma}_{\boldsymbol{y}_t\boldsymbol{y}_t|\boldsymbol{y}_{t-1}^{(l)}}\boldsymbol{\alpha}, \tag{40}$$

ただし

$$\hat{\Sigma}_{\boldsymbol{y}_t\boldsymbol{y}_t|\boldsymbol{y}_{t-1}^{(l)}} = \left(\tilde{K}_{\boldsymbol{y}_t}^2 + \epsilon I_N\right) - \tilde{K}_{\boldsymbol{y}_t}\tilde{K}_{\boldsymbol{y}_{t-1}^{(l)}}\left(\tilde{K}_{\boldsymbol{y}_{t-1}^{(l)}}^2 + \epsilon I_N\right)^{-1}\tilde{K}_{\boldsymbol{y}_{t-1}^{(l)}}\tilde{K}_{\boldsymbol{y}_t}, \tag{41}$$

$$\hat{\Sigma}_{\boldsymbol{x}_{t-1}^{(k)}\boldsymbol{x}_{t-1}^{(k)}|\boldsymbol{y}_{t-1}^{(l)}} = \left(\tilde{K}_{\boldsymbol{x}_{t-1}^{(k)}}^2 + \epsilon I_N\right) - \tilde{K}_{\boldsymbol{x}_{t-1}^{(k)}}\tilde{K}_{\boldsymbol{y}_{t-1}^{(l)}}\left(\tilde{K}_{\boldsymbol{y}_{t-1}^{(l)}}^2 + \epsilon I_N\right)^{-1}\tilde{K}_{\boldsymbol{y}_{t-1}^{(l)}}\tilde{K}_{\boldsymbol{x}_{t-1}^{(k)}}, \tag{42}$$

$$\hat{\Sigma}_{\boldsymbol{y}_t\boldsymbol{x}_{t-1}^{(k)}|\boldsymbol{y}_{t-1}^{(l)}} = \tilde{K}_{\boldsymbol{y}_t}\tilde{K}_{\boldsymbol{x}_{t-1}^{(k)}} - \tilde{K}_{\boldsymbol{y}_t}\tilde{K}_{\boldsymbol{y}_{t-1}^{(l)}}\left(\tilde{K}^2_{\boldsymbol{y}_{t-1}^{(l)}} + \epsilon I_N\right)^{-1}\tilde{K}_{\boldsymbol{y}_{t-1}^{(l)}}\tilde{K}_{\boldsymbol{x}_{t-1}^{(k)}}, \quad (43)$$

$$\hat{\Sigma}_{\boldsymbol{x}_{t-1}^{(k)}\boldsymbol{y}_t|\boldsymbol{y}_{t-1}^{(l)}} = \tilde{K}_{\boldsymbol{x}_{t-1}^{(k)}}\tilde{K}_{\boldsymbol{y}_t} - \tilde{K}_{\boldsymbol{x}_{t-1}^{(k)}}\tilde{K}_{\boldsymbol{y}_{t-1}^{(l)}}\left(\tilde{K}^2_{\boldsymbol{y}_{t-1}^{(l)}} + \epsilon I_N\right)^{-1}\tilde{K}_{\boldsymbol{y}_{t-1}^{(l)}}\tilde{K}_{\boldsymbol{y}_t}. \quad (44)$$

$\boldsymbol{y}_t$, $\boldsymbol{y}_{t-1}^{(l)}$, $\boldsymbol{x}_{t-1}^{(k)}$ の全てに対して線形なカーネル関数 $k(\boldsymbol{x}, \boldsymbol{x'}) = \boldsymbol{x}^T\boldsymbol{x'}$ を用いた場合は, 通常のデータ空間での PCCA と等価である. この一般化固有値問題を解いて得られる固有値 λ_i を用いて, 変数 $\boldsymbol{x}$ の部分空間から $\boldsymbol{y}$ の部分空間への非線形な TE は次のようになる.

$$T^{\{i\}}_{\boldsymbol{x}\to\boldsymbol{y}} = \frac{1}{2}\log_2\frac{1}{1-\lambda_i}. \quad (45)$$

このとき非ゼロの固有値から算出された $T^{\{i\}}_{\boldsymbol{x}\to\boldsymbol{y}}$ の総和は 2 つの変数 $\boldsymbol{x}$ 全体から $\boldsymbol{y}$ 全体への非線形な TE となる.

$$T_{\boldsymbol{x}\to\boldsymbol{y}} = \sum_i T^{\{i\}}_{\boldsymbol{x}\to\boldsymbol{y}}. \quad (46)$$

5. 正則化

多次元データに対して回帰分析や CCA を適用する時の問題点として，データの各次元の間の相関が強い場合や，(データの次元数)>(サンプル数) である場合に，計算に使う分散共分散行列がランク落ち，あるいはそれに近い状態になり逆行列計算が不可能または不安定になることがあげられる．そこで，多次元 TE に対し正則化を行う [7]．本章では $\mathbf{X}_t^{(m)}$ を，ある時刻 t_s から t_e までの埋め込みベクトルを並べた行列とする．ただし，$\mathbf{X}_t = \mathbf{X}_t^{(1)}$ である．

$$\mathbf{X}_t^{(m)} = \left(\begin{array}{cccc} \boldsymbol{x}_{t_s}^{(m)} & \boldsymbol{x}_{t_s+1}^{(m)} & \cdots & \boldsymbol{x}_{t_e}^{(m)} \end{array}\right)^T.$$

5.1 偏正準相関分析の正則化

多次元 TE（あるいは GC）は PCCA で表現されるため，多次元の TE の正則化を行うには，PCCA に対して正則化を行えば良いことになる．

まずは，正則化した場合の偏共分散行列について考える．X'，Z をそれぞれ $N \times D_{x'}$，$N \times D_z$ 行列とし，係数行列 A を $D_z \times D_{x'}$ 行列，残差 $\epsilon_{x'}$ を $N \times D_{x'}$ 行列とした場合の線形回帰式は，以下のようになる．

$$\mathbf{X}' = \mathbf{Z}\mathbf{A} + \epsilon_x. \quad (47)$$

次に正則化を行った回帰の場合の偏共分散行列を求める．正則化とは誤差関数に正則化項を付加することにより，係数が必要以上に大きくなる事などを防ぐ効果を得るものである．また，この場合は $\mathbf{Z}$ の分散共分散行列の逆行列計算を安定化する効果がある．正則化項を付加した誤差関数は以下のようになる．

$$J = \frac{1}{N}||\mathbf{X}' - \mathbf{Z}\mathbf{A}||^2 + \eta_z\mathbf{A}^T\mathbf{A}, \quad (48)$$

ここで η_z は正則化のパラメータである．この様な，係数の 2 乗に比例する正則化項を付加した回帰分析のことを，リッジ回帰と呼ぶ.

$\mathbf{X}' = [\mathbf{X}\ \mathbf{Y}]$ として $\mathbf{A}$ の推定値 $\check{\mathbf{A}}$ を求める．$\mathbf{A}$ に関する J の導関数を 0 とすると，

$$\check{\mathbf{A}} = (\Sigma_{zz} + \eta_z\mathbf{I})^{-1}[\Sigma_{zx}\ \Sigma_{zy}] \quad (49)$$

$$= \tilde{\Sigma}_{zz}^{-1}[\Sigma_{zx}\ \Sigma_{zy}]. \quad (50)$$

ただし，式 (50) で，$\tilde{\Sigma}_{zz} = \Sigma_{zz} + \eta_z\mathbf{I}$ と置いた．得られた $\check{\mathbf{A}}$ による $\mathbf{X}$，$\mathbf{Y}$ の予測値と実際の値との間の誤差の分散共分散行列の事を偏共分散行列と呼び，正則化した場合の偏共分散行列は以下のようになる．

$$\Sigma_{\{xy\}\{xy\}|z} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} - 2\begin{bmatrix} \Sigma_{xz} \\ \Sigma_{yz} \end{bmatrix}\check{\mathbf{A}} + \check{\mathbf{A}}^T\Sigma_{zz}\check{\mathbf{A}} \quad (51)$$

$$= \begin{bmatrix} \Sigma_{xx|\tilde{z}} & \Sigma_{xy|\tilde{z}} \\ \Sigma_{yx|\tilde{z}} & \Sigma_{yy|\tilde{z}} \end{bmatrix}. \quad (52)$$

ここで,

$$\begin{aligned}
\Sigma_{xx|\tilde{z}} &= \Sigma_{xx} - 2\Sigma_{xz}\tilde{\Sigma}_{zz}^{-1}\Sigma_{zx} + \Sigma_{xz}\tilde{\Sigma}_{zz}^{-1}\Sigma_{zz}\tilde{\Sigma}_{zz}^{-1}\Sigma_{zx},\\
\Sigma_{xy|\tilde{z}} &= \Sigma_{xy} - 2\Sigma_{xz}\tilde{\Sigma}_{zz}^{-1}\Sigma_{zy} + \Sigma_{xz}\tilde{\Sigma}_{zz}^{-1}\Sigma_{zz}\tilde{\Sigma}_{zz}^{-1}\Sigma_{zy},\\
\Sigma_{yx|\tilde{z}} &= \Sigma_{yx} - 2\Sigma_{yz}\tilde{\Sigma}_{zz}^{-1}\Sigma_{zx} + \Sigma_{yz}\tilde{\Sigma}_{zz}^{-1}\Sigma_{zz}\tilde{\Sigma}_{zz}^{-1}\Sigma_{zx},\\
\Sigma_{yy|\tilde{z}} &= \Sigma_{yy} - 2\Sigma_{yz}\tilde{\Sigma}_{zz}^{-1}\Sigma_{zy} + \Sigma_{yz}\tilde{\Sigma}_{zz}^{-1}\Sigma_{zz}\tilde{\Sigma}_{zz}^{-1}\Sigma_{zy}.
\end{aligned}$$

ここまで，リッジ回帰を利用することにより，変量 $\mathbf{Z}$ における共線性などによる分散共分散行列 Σ_{zz} の逆行列計算の不安定性を緩和し，その場合の偏共分散行列を導出した．しかし，変量 $\mathbf{X}$，$\mathbf{Y}$ に関する正則化はできていないため，Σ_{xx}，Σ_{yy} がランク落ちしていた場合，PCCA を解くことはまだ出来ない．この様な場合にも使える CCA の正則化手法として，Canonical Ridge という手法がある．Canonical Ridge は以下の様な，相関係数を最大化する射影を見つける問題として表現される．

$$\rho = \frac{\boldsymbol{a}^T\Sigma_{xy}\boldsymbol{b}}{\sqrt{\boldsymbol{a}^T(\Sigma_{xx} + \eta_x\mathbf{I})\boldsymbol{a}}\sqrt{\boldsymbol{b}^T(\Sigma_{yy} + \eta_y\mathbf{I})\boldsymbol{b}}}, \quad (53)$$

η_x，η_y は正則化のパラメータ $(\eta_x, \eta_y \geq 0)$ である．

偏共分散行列を，正則化を行った場合のものに変更した PCCA に対して，式 (53) と同じように記述すると以下の様になる．

$$\rho = \frac{\boldsymbol{a}^T\Sigma_{xy|\tilde{z}}\boldsymbol{b}}{\sqrt{\boldsymbol{a}^T(\Sigma_{xx|\tilde{z}} + \eta_x\mathbf{I})\boldsymbol{a}}\sqrt{\boldsymbol{b}^T(\Sigma_{yy|\tilde{z}} + \eta_y\mathbf{I})\boldsymbol{b}}}. \quad (54)$$

ラグランジュの乗数 λ，ν を用いて，次の最大化問題を考える．

$$J = \boldsymbol{a}^T\Sigma_{xy|\tilde{z}}\boldsymbol{b} - \frac{\sqrt{\lambda}}{2}\left(\boldsymbol{a}^T(\Sigma_{xx|\tilde{z}} + \eta_x\mathbf{I})\boldsymbol{a} - 1\right) - \frac{\sqrt{\nu}}{2}\left(\boldsymbol{b}^T(\Sigma_{yy|\tilde{z}} + \eta_y\mathbf{I})\boldsymbol{b} - 1\right). \quad (55)$$

$\boldsymbol{a}$, $\boldsymbol{b}$ それぞれに関する J の導関数を 0 とすると,

$$\frac{\partial J}{\partial \boldsymbol{a}} = \Sigma_{xy|\tilde{z}}\boldsymbol{b} - \sqrt{\lambda}(\Sigma_{xx|\tilde{z}} + \eta_x\mathbf{I})\boldsymbol{a}, \quad (56)$$

$$\frac{\partial J}{\partial \boldsymbol{b}} = \Sigma_{yx|\tilde{z}}\boldsymbol{a} - \sqrt{\nu}(\Sigma_{yy|\tilde{z}} + \eta_y\mathbf{I})\boldsymbol{b}, \quad (57)$$

となり，これらの式に左からそれぞれ $\boldsymbol{a}^T$，$\boldsymbol{b}^T$ を掛けて式 (55) に代入すると，$\sqrt{\lambda} = \sqrt{\nu} = \boldsymbol{a}^T\Sigma_{xy|\tilde{z}}\boldsymbol{b}$ が成り立つ．$\boldsymbol{a}$, $\boldsymbol{b}$ のいずれかを消去すると，次の一般化固有値問題を得る．

$$\Sigma_{xy|\tilde{z}}\tilde{\Sigma}^{-1}_{yy|\tilde{z}}\Sigma_{yx|\tilde{z}}\boldsymbol{a} = \lambda\tilde{\Sigma}_{xx|\tilde{z}}\boldsymbol{a}, \quad (58)$$

$$\Sigma_{yx|\tilde{z}}\tilde{\Sigma}^{-1}_{xx|\tilde{z}}\Sigma_{xy|\tilde{z}}\boldsymbol{b} = \lambda\tilde{\Sigma}_{yy|\tilde{z}}\boldsymbol{b}. \quad (59)$$

ただし，$\tilde{\Sigma}_{xx|\tilde{z}} = \Sigma_{xx|\tilde{z}} + \eta_x\mathbf{I}$, $\tilde{\Sigma}_{yy|\tilde{z}} = \Sigma_{yy|\tilde{z}} + \eta_y\mathbf{I}$.

5.2 多次元 Transfer Entropy の正則化

ここからは，先程導出した正則化 PCCA を，実際に多次元の TE に用いた時の定式化を行う．前節まで扱ってきた PCCA は，$\mathbf{Z}$ の情報を取り除いた状態で，$\mathbf{X}$ と $\mathbf{Y}$ の相関を最大化するという問題であった．データにガウス分布を仮定したときの多次元 TE では，変量 $\mathbf{X}$ から $\mathbf{Y}$ への因果を見るときは，$\mathbf{Y}^{(n)}_{t-1}$ の情報を取り除いた状態で，$\mathbf{X}^{(m)}_{t-1}$ と $\mathbf{Y}_t$ の相関を最大化する問題を解くことになる．

以上のことから正則化を行った PCCA に対して,

$$\mathbf{X} \to \mathbf{X}^{(m)}_{t-1}, \mathbf{Y} \to \mathbf{Y}_t, \mathbf{Z} \to \mathbf{Y}^{(n)}_{t-1},$$

と置き換えると，多次元 TE を正則化したものは，以下の一般化固有値問題で表現される．

$$\Sigma_{y_t x^{(m)}_{t-1}|\tilde{y}^{(n)}_{t-1}}\tilde{\Sigma}^{-1}_{x^{(m)}_{t-1}x^{(m)}_{t-1}|\tilde{y}^{(n)}_{t-1}}\Sigma_{x^{(m)}_{t-1}y_t|\tilde{y}^{(n)}_{t-1}}\boldsymbol{w}_{y_t} = \lambda\tilde{\Sigma}_{y_t y_t|\tilde{y}^{(n)}_{t-1}}\boldsymbol{w}_{y_t}, \quad (60)$$

$$\Sigma_{x^{(m)}_{t-1}y_t|\tilde{y}^{(n)}_{t-1}}\tilde{\Sigma}^{-1}_{y_t y_t|\tilde{y}^{(n)}_{t-1}}\Sigma_{y_t x^{(m)}_{t-1}|\tilde{y}^{(n)}_{t-1}}\boldsymbol{w}_{x^{(m)}_{t-1}} = \lambda\tilde{\Sigma}_{x^{(m)}_{t-1}x^{(m)}_{t-1}|\tilde{y}^{(n)}_{t-1}}\boldsymbol{w}_{x^{(m)}_{t-1}}. \quad (61)$$

ただし，次の様に変数を置き換えた．

$$\tilde{\Sigma}_{x^{(m)}_{t-1}x^{(m)}_{t-1}|\tilde{y}^{(n)}_{t-1}} = \Sigma_{x^{(m)}_{t-1}x^{(m)}_{t-1}|\tilde{y}^{(n)}_{t-1}} + \eta_{x^{(m)}_{t-1}}\mathbf{I}, \quad (62)$$

$$\tilde{\Sigma}_{y_t y_t|\tilde{y}^{(n)}_{t-1}} = \Sigma_{y_t y_t|\tilde{y}^{(n)}_{t-1}} + \eta_{y_t}\mathbf{I}, \quad (63)$$

$$\Sigma_{x^{(m)}_{t-1}x^{(m)}_{t-1}|\tilde{y}^{(n)}_{t-1}} = \Sigma_{x^{(m)}_{t-1}x^{(m)}_{t-1}} - \Sigma_{x^{(m)}_{t-1}y^{(n)}_{t-1}}(\Sigma_{y^{(n)}_{t-1}y^{(n)}_{t-1}} + \eta_{y^{(n)}_{t-1}}\mathbf{I})^{-1}\Sigma_{y^{(n)}_{t-1}x^{(m)}_{t-1}}, \quad (64)$$

$$\Sigma_{y_t y_t|\tilde{y}^{(n)}_{t-1}} = \Sigma_{y_t y_t} - \Sigma_{y_t y^{(n)}_{t-1}}(\Sigma_{y^{(n)}_{t-1}y^{(n)}_{t-1}} + \eta_{y^{(n)}_{t-1}}\mathbf{I})^{-1}\Sigma_{y^{(n)}_{t-1}y_t}, \quad (65)$$

$$\Sigma_{x^{(m)}_{t-1}y_t|\tilde{y}^{(n)}_{t-1}} = \Sigma_{x^{(m)}_{t-1}y_t} - \Sigma_{x^{(m)}_{t-1}y^{(n)}_{t-1}}(\Sigma_{y^{(n)}_{t-1}y^{(n)}_{t-1}} + \eta_{y^{(n)}_{t-1}}\mathbf{I})^{-1}\Sigma_{y^{(n)}_{t-1}y_t}, \quad (66)$$

$$\Sigma_{y_t x^{(m)}_{t-1}|\tilde{y}^{(n)}_{t-1}} = \Sigma_{y_t x^{(m)}_{t-1}} - \Sigma_{y_t y^{(n)}_{t-1}}(\Sigma_{y^{(n)}_{t-1}y^{(n)}_{t-1}} + \eta_{y^{(n)}_{t-1}}\mathbf{I})^{-1}\Sigma_{y^{(n)}_{t-1}x^{(m)}_{t-1}}. \quad (67)$$

$\eta_{x^{(m)}_{t-1}}$，$\eta_{y^{(n)}_{t-1}}$，η_{y_t} はそれぞれ正則化のパラメータである（$\eta_{x^{(m)}_{t-1}}, \eta_{y^{(n)}_{t-1}}, \eta_{y_t} \geq 0$）．式 (60)，あるいは式 (61) の一般化固有値問題から固有値 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M \geq 0\,(M = \min\{m \times D_x, D_y\})$ が得られる．

第 i 正準空間同士の TE，x 全体から y 全体への TE はそれぞれ式 (45)，式 (46) と同様に表現できる．

6. 確率的解釈とベイズモデル

この章では因果指標計算に用いる PCCA と等価な確率的生成モデルを与える．これにより確率的観点からのモデル拡張が可能になる．その拡張としてモデルパラメタに事前分布を仮定しパラメタを分布推定するベイズ推定のモデルを提案する [8].

6.1 生成モデル

影響を除きたい第三変数 $\boldsymbol{x}_n$，潜在変数 $\boldsymbol{z}_n$ から同時に線形回帰する以下の式に表される生成モデルを考える．

$$\begin{aligned}\boldsymbol{z}_n &\sim \mathcal{N}(\mathbf{0}, I_{d_z}),\\ \boldsymbol{y}^m_n &\sim \mathcal{N}\left(W^m_x\boldsymbol{x}_n + W^m_z\boldsymbol{z}_n + \boldsymbol{\mu}^m, \Psi^m\right).\end{aligned} \quad (68)$$

このモデルの最尤推定解 $\arg\max \log p(\boldsymbol{y}|\boldsymbol{x}; W_x, W_z, \Psi)$ が PCCA を用いて計算出来ることを示す．そのために，提案モデルが確率的正準相関分析の式

$$\begin{aligned}\boldsymbol{z}_n &\sim \mathcal{N}(\mathbf{0}, I_{d_z}),\\ \boldsymbol{y}^m_n &\sim \mathcal{N}\left(W^m_z\boldsymbol{z}_n + \boldsymbol{\mu}^m, \Psi^m\right).\end{aligned} \quad (69)$$

に帰着されることを示す．

今，対数尤度を L と置き，潜在変数 $\boldsymbol{z}$ を積分消去した時の $\boldsymbol{y}$ の共分散を

$$C = \begin{pmatrix}\Psi^1 & 0\\ 0 & \Psi^2\end{pmatrix} + \begin{pmatrix}W^1_z\\ W^2_z\end{pmatrix}\begin{pmatrix}{W^1_z}^T {W^2_z}^T\end{pmatrix}$$

と置くと，

$$\frac{\partial L}{\partial \boldsymbol{\mu}} = -\sum_{n=1}^{N} C^{-1}\left(\begin{pmatrix}\boldsymbol{\mu}^1\\ \boldsymbol{\mu}^2\end{pmatrix} - \begin{pmatrix}\boldsymbol{y}^1_n\\ \boldsymbol{y}^2_n\end{pmatrix} + \begin{pmatrix}W^1_x\\ W^2_x\end{pmatrix}\boldsymbol{x}_n\right)$$

が成り立つ．C は正定値なので，対数尤度を最大にする $\boldsymbol{\mu}$ は偏微分が $\mathbf{0}$ になる点である．従って

$$\boldsymbol{\mu}^m = \overline{\boldsymbol{y}}^m - W^m_x\overline{\boldsymbol{x}} \quad (70)$$

となる．各サンプルからサンプル平均を引いたものを

$\widetilde{\boldsymbol{y}}_n^1 = \boldsymbol{y}_n^1 - \overline{\boldsymbol{y}}^1$ のように書くと，式 (70) を代入し，

$$\frac{\partial L}{\partial W_x} = \sum_{n=1}^{N} C^{-1} \left(\begin{pmatrix} \widetilde{\boldsymbol{y}}_n^1 \\ \widetilde{\boldsymbol{y}}_n^2 \end{pmatrix} \widetilde{\boldsymbol{x}}_n^T - \begin{pmatrix} W_x^1 \\ W_x^2 \end{pmatrix} \widetilde{\boldsymbol{x}}_n \widetilde{\boldsymbol{x}}_n^T \right)$$

となる．この式はサンプルが入力空間を張る時 L が W_x の負定値二次形式になることが示せるので，偏微分が 0 となる W_x で L は最大化される．従って，

$$W_x^m = \Sigma_{mx} \Sigma_{xx}^{-1} \tag{71}$$

となる．この式を生成モデルの式 (68) に代入すると，確率的正準相関分析の式 (69) に入力として $\boldsymbol{y}'^m_n = \widetilde{\boldsymbol{y}}_n^m - \Sigma_{mx}\Sigma_{xx}^{-1}\widetilde{\boldsymbol{x}}_n$ を代入したものと等価になる．これらの共分散は

$$\begin{aligned} \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{y}'^{m_1}_n \boldsymbol{y}'^{m_2 T}_n &= \Sigma_{m_1 m_2} - \Sigma_{m_1 x}\Sigma_{xx}^{-1}\Sigma_{x m_2} \\ &= \Sigma_{m_1 m_2 | x} \end{aligned} \tag{72}$$

と，偏共分散になるため, 最尤解は確率的正準相関分析の最尤解で共分散を偏共分散で置き換えたもの，すなわち PCCA に帰着されることが示された．式で表現すると，最尤解は以下に表される．

$$\begin{aligned} W_x^m &= \Sigma_{mx}\Sigma_{xx}^{-1}, \\ W_z^m &= \Sigma_{mm|x} U_{d_z}^m M_m, \\ \Psi^m &= \Sigma_{mm|x} - W_z^m W_z^{m T}, \\ \boldsymbol{\mu}^m &= \overline{\boldsymbol{y}}^m - W_x^m \overline{\boldsymbol{x}}, \end{aligned} \tag{73}$$

ここで $U_{d_z}^m$ は d 列に d 番めの固有ベクトルを並べた行列，P_d は対応する固有値を d 個対角に並べた対角行列，M_m は $M_1 M_2^T = P_{d_z}$ を満たしスペクトルノルムが 1 より小さい任意の行列である．

以下の議論では簡潔性のためデータは平均が 0 になっていると仮定し，$\boldsymbol{\mu}$ の推論は行わない．

6.2 ベイズモデル

この節では，Wang [9] の手法にしたがい，生成モデルのパラメタ Ψ, W に事前分布を仮定し，パラメタの事後分布をベイズ推定するモデルを考える．これによりロバストな推定，モデル選択が行われる．W に列毎に ARD 事前分布 [10] を仮定し，Ψ に逆 Wishart 分布を仮定する．生成モデルは以下の式に表される．

$$\begin{aligned} \alpha_k^m &\sim \mathrm{Gamma}(a_0, b_0), \\ W_{:,k}^m &\sim \mathcal{N}(\boldsymbol{0}, (\alpha_k^m)^{-1} I_{d_m}), \\ \Psi^m &\sim \mathcal{IW}(\nu_0^m, K_0^m), \\ z_n &\sim \mathcal{N}(\boldsymbol{0}, I_{d_z}), \\ \boldsymbol{y}_n^m &\sim \mathcal{N}(W_x^m \boldsymbol{x}_n + W_z^m \boldsymbol{z}_n, \Psi^m), \end{aligned} \tag{74}$$

ここで第三変数の事前分布 $p(x)$ は各サンプルで $p(x) > 0$ であれば推論に影響しない．ここで $\mathrm{Gamma}(a, b)$ はガンマ分布，$\mathcal{IW}(\nu, K)$ は逆 Wishart 分布である．$W^m = \begin{pmatrix} W_x^m & W_z^m \end{pmatrix}$ であり，$W_{:,k}^m$ は W^m の k 列目を表す．ハイパーパラメタ a_0，b_0，ν_0^m，K_0^m は事前分布のすそが広くなるように小さいものがこのましい．しかしながら逆 Wishart 分布の定義から $\nu_0^m > d_m - 1$ となる．ARD 事前分布により重要度の低い列が $\boldsymbol{0}$ に向かうため，十分大きな d_z をとれば適切な潜在変数 $\boldsymbol{z}$ の次元が推定できる．

推論は変分ベイズ法により行う．近似した事後分布を

$$q(Z, \Theta) = q(Z) \prod_{m=1}^{2} \left(q(\Psi^m) q(\boldsymbol{\alpha}^m) \prod_{j=1}^{d_m} q(\boldsymbol{w}_j^m) \right) \tag{75}$$

と置き，各分布を逐次的に更新する．ここで w_j^m は W^m の j 番めの行である．各事後分布は

$$\begin{aligned} q(\boldsymbol{z}_n) &= \mathcal{N}(\boldsymbol{\mu}_{z_n}, \Sigma_{z_n}), \\ q(\Psi^m) &= \mathcal{IW}(\nu_m, K_m), \\ q(\boldsymbol{w}_j^m) &= \mathcal{N}(\boldsymbol{\mu}_{w_j^m}, \Sigma_{w_j^m}), \\ q(\boldsymbol{\alpha}^m) &= \prod_k \mathrm{Gamma}(a_m, b_{mk}) \end{aligned} \tag{76}$$

の形になり，パラメタの更新式は

$$\begin{aligned} \Sigma_{z_n} =& \left(I + \sum_m \langle (W_z^m)^T (\Psi^m)^{-1} W_z^m \rangle \right)^{-1}, \\ \boldsymbol{\mu}_{z_n} =& \Sigma_{z_n} \sum_m \Big(\langle (W_z^m)^T \rangle \langle (\Psi^m)^{-1} \rangle \boldsymbol{y}_n^m \\ & - \langle (W_z^m)^T (\Psi^m)^{-1} W_x^m \rangle \boldsymbol{x}_n \Big), \end{aligned}$$

$$\begin{aligned} K_m =& K_0^m + Y^m (Y^m)^T \\ & + \langle W^m \begin{pmatrix} XX^T & XZ^T \\ ZX^T & ZZ^T \end{pmatrix} (W^m)^T \rangle \\ & - Y^m \begin{pmatrix} X^T & \langle Z^T \rangle \end{pmatrix} \langle (W^m)^T \rangle \\ & - \langle W^m \rangle \begin{pmatrix} X \\ \langle Z \rangle \end{pmatrix} Y^m, \\ \nu_m =& \nu_0^m + N, \end{aligned}$$

$$\begin{aligned} \Sigma_{w_j^m} =& \left(\mathrm{diag}\langle \boldsymbol{\alpha}^m \rangle + \langle (\Psi^m)_{j,j}^{-1} \rangle \begin{pmatrix} XX^T & X\langle Z \rangle^T \\ \langle Z \rangle X^T & \langle ZZ^T \rangle \end{pmatrix} \right)^{-1}, \\ \boldsymbol{\mu}_{w_j^m} =& \langle (\Psi^m)_{j,:}^{-1} \rangle Y^m \begin{pmatrix} X^T & Z^T \end{pmatrix} \\ & - \sum_{l \neq j} \langle (\Psi^m)_{j,l}^{-1} \rangle \langle W_{l,:}^m \rangle \begin{pmatrix} XX^T & X\langle Z \rangle^T \\ \langle Z \rangle X^T & \langle ZZ^T \rangle \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} a_m =& a_0 + d_m/2, \\ b_{mk} =& b_0 + \langle ||W_{:,k}^m|| \rangle / 2, \end{aligned} \tag{77}$$

の形になる．ここで，$\mathrm{diag}\langle \boldsymbol{\alpha}^m \rangle$ は k 番めの対角要素が $\langle \boldsymbol{\alpha}_k^m \rangle$ となる対角行列である．

6.3 近似されたベイズモデル

前の節で提案したベイズモデルはノイズ精度行列 Ψ の推論に大きな計算量が必要であり，またサンプル数が小さいときに逆 Wishart 事前分布の影響が大きくなる．そこで Klami ら [11] の手法に従って，ノイズを等方ノイズと非共有の潜在変数からの線形回帰で近似する以下の生成モデルを考える．

$$\begin{aligned}
\boldsymbol{z}_n &\sim \mathcal{N}(\mathbf{0}, I_{d_z}),\\
\boldsymbol{z}_n^m &\sim \mathcal{N}(\mathbf{0}, I_{d_{z_m}}),\\
\boldsymbol{y}_n^m &\sim \mathcal{N}\left(W_x^m \boldsymbol{x}_n + A^m \boldsymbol{z}_n + B^m \boldsymbol{z}_n^m, (\tau^m)^{-1} I_{d_m}\right).
\end{aligned} \tag{78}$$

非共有潜在変数 $\boldsymbol{z}_n^m$ を積分消去すると，前の節の生成モデル式 (74) において $\Psi^m = B^m (B^m)^T + (\tau^m)^{-1} I_{d_m}$ としたものと等価になる．従ってこのモデルは共分散を低ランク近似したものとみなせる．A と B を同時に推論するため，これらをまとめて $W_z = \begin{pmatrix} A^{(1)} & B^{(1)} & 0 \\ A^{(1)} & 0 & B^{(1)} \end{pmatrix}$, $W = \begin{pmatrix} W_x & W_z \end{pmatrix}$ とかき，以下のモデルを考える．

$$\begin{aligned}
\alpha_k^m &\sim \mathrm{Gamma}(a_0, b_0),\\
W_{:,k}^m &\sim \mathcal{N}(\mathbf{0}, (\alpha_k^m)^{-1} I_{d_m}),\\
\tau^m &\sim \mathrm{Gamma}(a_0, b_0),\\
\boldsymbol{z}_n &\sim \mathcal{N}(\mathbf{0}, I_{(d_z + d_{z_1} + d_{z_2})}),\\
\boldsymbol{y}_n^m &\sim \mathcal{N}\left(W_x^m \boldsymbol{x}_n + W_z^m \boldsymbol{z}_n, (\tau^m)^{-1} I_{d_m}\right).
\end{aligned} \tag{79}$$

この表現によりモデルパラメタ数が減り，さらなるロバスト性が期待される．このモデルにおいてもハイパーパラメタ a_0，b_0 は小さいものが好ましい．

近似された事後分布を

$$q(Z, \Theta) = q(Z) \prod_m \left(q\left(\tau^m\right) q\left(\boldsymbol{\alpha}^m\right) q\left(W^m\right) \right) \tag{80}$$

と置くと，変分ベイズ推論により各々の事後分布は

$$\begin{aligned}
q(Z) &= \prod_n \mathcal{N}(\boldsymbol{\mu}_{z_n}, \Sigma_z),\\
q(W^m) &= \prod_d \mathcal{N}(\boldsymbol{\mu}_{W_{d,:}^m}, \Sigma_{W^m}),\\
q(\boldsymbol{\alpha}^m) &= \prod_k \mathrm{Gamma}(a_{\alpha^m}, b_{\alpha_k^m}),\\
q(\tau^m) &= \mathrm{Gamma}(a_{\tau^m}, b_{\tau^m})
\end{aligned} \tag{81}$$

の形になり，パラメタ更新則は以下の通りとなる．

$$\begin{aligned}
\Sigma_{W^m} &= \left(\mathrm{diag}\langle \boldsymbol{\alpha}^m \rangle + \langle \tau^m \rangle \begin{pmatrix} XX^T & X\langle Z \rangle^T \\ \langle Z \rangle X^T & \langle ZZ^T \rangle \end{pmatrix} \right)^{-1},\\
\mu_{W^m} &= Y^m \begin{pmatrix} X^T & \langle Z^T \rangle \end{pmatrix},\\
\Sigma_z &= \left(I + \sum_m \langle \tau^m \rangle \langle (W_z^m)^T W_z^m \rangle \right)^{-1},\\
\langle Z \rangle &= \Sigma_z \left(\sum_m \langle \tau^m \rangle \left(\langle (W_z^m)^T \rangle Y^m - \langle (W_z^m)^T W_x^m \rangle X \right) \right),\\
a_{\alpha^m} &= a_0 + d_m/2,\\
b_{\alpha_k^m} &= b_0 + \langle (W^m)^T W^m \rangle_{k,k}/2,\\
a_{\tau^m} &= a_0 + N d_m/2,\\
b_{\tau^m} &= b_0 + \frac{1}{2} \Bigg(\mathrm{Tr}\left(Y^m (Y^m)^T \right.\\
&\quad \left. - 2 Y^m \begin{pmatrix} X^T & \langle Z^T \rangle \end{pmatrix} \langle (W^m)^T \rangle \right)\\
&\quad + \mathrm{Tr}\left(\langle (W^m)^T W^m \rangle \begin{pmatrix} XX^T & X\langle Z \rangle^T \\ \langle Z \rangle X^T & \langle ZZ^T \rangle \end{pmatrix} \right) \Bigg).
\end{aligned} \tag{82}$$

7. おわりに

本論文では，まず，Transfer Entropy，Continuous Transfer Entropy，Granger Causality といった従来の因果指標を紹介し，これらの関係性を明らかにした．特に時系列確率変数がガウス分布に従うときに，Transfer Entropy が偏正準相関分析に帰着されることを示した．偏正準相関分析をカーネル化することで，様々な計量を考慮可能な因果指標を提案した．また，偏正準相関分析の正則化により，少サンプルでも安定して計算可能な手法を提案した．さらに，偏正準相関分析を確率的に解釈をし，確率的偏正準相関分析という新たなモデルを提案をした．これにより確率的観点からのパラメタ推定やモデル拡張が可能になった．

参考文献

[1] Schreiber, T.: Measuring Information Transfer, *Phys. Rev. Lett.*, Vol. 85, No. 2, pp. 461–464, (2000).
[2] Kaiser, A. and Schreiber, T.: Information transfer in continuous processes, *Physica D Nonlinear Phenomena*, Vol. 166, pp. 43–62 (2002).
[3] Granger, C. W. J.: Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, *Econometrica*, Vol. 37, No. 3, pp. 424–438 (1969).
[4] Shibuya, T., Harada, T. and Kuniyoshi, Y.: Causality Quantification and Its Applications: Structuring and Modeling of Multivariate Time Series, *KDD*, pp. 787–796 (2009).
[5] Shibuya, T., Harada, T. and Kuniyoshi, Y.: Reliable index for measuring information flow, *Phys. Rev. E*, Vol. 84, p. 061109 (2011).
[6] Rao, B. R.: Partial canonical correlations, *Trabajos de estadistica y de investigacion operativa*, Vol. 20, No. 2-3, pp. 211–219 (1969).
[7] Yamashita, Y., Harada, T. and Kuniyoshi, Y.: Causal Flow, *IEEE TMM*, Vol. 14, No. 3, pp. 619–629 (2012).
[8] Mukuta, Y. and Harada, T.: Probabilistic Partial Canonical Correlation Analysis, *ICML*, pp. 1449–1457 (2014).
[9] Wang, C.: Variational Bayesian Approach to Canonical Correlation Analysis, *Trans. Neur. Netw.*, Vol. 18, No. 3, pp. 905–910 (2007).
[10] Neal, R. M.: Bayesian Learning for Neural Networks, PhD Thesis, the University of Toronto (1995).
[11] Klami, A., Virtanen, S. and Kaski, S.: Bayesian Canonical Correlation Analysis, *J. Mach. Learn. Res.*, Vol. 14, No. 1, pp. 965–1003 (2013).