因子分析モデルにおける構造正則化

廣瀬　慧 [1,2]

[1] 九州大学 マス・フォア・インダストリ研究所
[2] 理化学研究所革新知能統合研究センター

E-mail: hirose@imi.kyushu-u.ac.jp.

概 要

Group Lasso や Fused Lasso をはじめとする構造正則化は，モデルの構造を抽出する有効な手法として広く用いられている．本稿では，因子分析モデルにおける因子負荷行列の構造正則化について考える．とくに，因子回転でよく用いられる Quartimin 基準 (Carroll 1953) に着目する．Hirose and Terada (2016) は，Quartimin 基準に基づく構造正則化が，正則化パラメータ ρ を $\rho \to \infty$ としたとき，完全単純構造とよばれる望ましい性質を持ち，さらに，その完全単純構造推定は，k-means による変数クラスタリングの一般化になることを示した．しかし，Hirose and Terada (2016) は尤度関数に基づくアプローチであり，証明が少し煩雑で分かりにくい．そこで，本稿では，本質的に分かりやすく証明をするために，2 乗損失に基づく Quartimin 正則化を行う．

1　はじめに

Lasso (Least absolute shrinkage and selection operator, Tibshirani 1996) をはじめとする L_1 正則化法は，パラメータのいくつかを正確に 0 とスパース推定することによって，モデルの構造を抽出する手法である．Lasso は有用な手法ではあるが，実際問題，スパースな構造とは別の構造を推定する必要がある場合が多い．例えば，Group Lasso (Yuan and Lin 2006) と Fused Lasso (Tibshirani et al. 2005) は，Lasso とは違った構造を抽出する代表的な手法である．Group Lasso は，パラメータベクトルの成分をいくつか取り出したベクトルを 0 ベクトルと推定できる手法であり，Fused Lasso は，パラメータベクトルのいくつかの成分を同じ値に推定する方法である．Group Lasso や Fused Lasso のようなモデルの構造を抽出する正則化法を構造正則化と呼び，近年，様々な正則化項が提案されている (たとえば，Jacob et al. 2009; Park et al. 2015; Xin et al. 2016)．

モデルの構造推定が重要である統計モデルは数多くあるが，その中でも因子分析モデルは，推定された因子を解釈する上で重要となる．しかし，因子分析モデルにおいて，これまで構造正則化はあまり用いられていない．その理由の一つとして，Varimax 回転 (Kaiser 1958), Promax 回転 (Hendrickson and White 1964) をはじめとする因子回転が伝統的に

用いられており，これらの因子回転は因子を解釈する上で十分実用的であるためだと考えられる．因子回転の詳細については，Browne (2001) を参照されたい．

一方で，因子分析における因子負荷行列の正則化は，因子回転の一般化であることが知られている (Hirose and Yamamoto 2015)．Hirose and Yamamoto (2015) は，Lasso タイプの因子回転である Component loss criterion (Jennrich 2004, 2006) に着目し，MCP (Minimax Concave Penalty; Zhang 2010) などを用いた L_1 型正則化法を提案した．その結果，因子回転よりもスパースな構造を推定でき，かつ高次元データにも適用できることが示された．このように，正則化法を用いることにより，因子回転ではできなかった推定ができることがある．

本稿では，Component loss criterion とは違った因子回転基準として，Quartimin 基準 (Carroll 1953) に着目する．Hirose and Terada (2016) は，この正則化項に基づいて正則化パラメータ ρ を $\rho \to \infty$ とすると，推定された因子負荷行列は完全単純構造 (Perfect simple structure あるいは Perfect cluster configuration, たとえば Browne 2001; Bernaards and Jennrich 2003 参照) を持つことを示した．ここで，因子負荷行列の各行の非 0 要素が最大で 1 つであるとき，因子負荷行列は完全単純構造を持つという．さらに，その完全単純構造の推定は，k-means による変数クラスタリングの一般化であることも示される．

しかしながら，Hirose and Terada (2016) は，対数尤度関数に基づく正則化を考えており，k-means との関係性を理解しにくい．そこで，本稿では，より関係性を理解しやすくするため，2 乗誤差に基づく損失関数を用いた Quartimin 正則化について述べる．

本稿の構成は以下の通りである．まず，2 節で正則化因子分析について述べる．次に，3 節で，Quartimin 基準の一般化を行う．4 節では，完全単純構造推定と k-means によるクラスタリングとの比較を行う．5 節にまとめと今後の展望について述べる．

2　正則化因子分析

2.1　定式化

p 次元観測変数ベクトル $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ が，m 次元共通因子ベクトル $\boldsymbol{F} = (F_1, \ldots, F_m)^T$ の一次結合と独自因子ベクトル $\boldsymbol{\varepsilon}$ の和で表される因子分析モデル

$$\boldsymbol{X} = \boldsymbol{\Lambda}\boldsymbol{F} + \boldsymbol{\varepsilon} \tag{2.1}$$

について考える．ここで，$\boldsymbol{\Lambda} = (\lambda_{ik})$ は $p \times m$ 因子負荷行列である．なお，観測変数ベクトル $\boldsymbol{X}$ の平均ベクトルは $\mathbf{0}$ であるとする*．ただし，$\mathbf{0}$ は 0 ベクトルとする．

*実際の解析では中心化したデータを用いる．

共通因子ベクトル $\boldsymbol{F}$ と独自因子ベクトル $\boldsymbol{\varepsilon}$ は，平均ベクトル，共分散行列がそれぞれ $E[\boldsymbol{F}] = \boldsymbol{0}$, $E[\boldsymbol{\varepsilon}] = \boldsymbol{0}$, $V[\boldsymbol{F}] = \boldsymbol{I}$, $V[\boldsymbol{\varepsilon}] = \boldsymbol{\Psi}$ で与えられ，互いに無相関，すなわち $E[\boldsymbol{F}\boldsymbol{\varepsilon}^T] = \boldsymbol{O}$ を仮定する．ここで，$\boldsymbol{\Psi}$ は $p \times p$ 対角行列で，対角成分は独自分散である．また，$\boldsymbol{O}$ は 0 行列であり，$\boldsymbol{A}^T$ は行列 $\boldsymbol{A}$ の転置を表す．このとき，観測変数ベクトル $\boldsymbol{X}$ の分散共分散行列 $\boldsymbol{\Sigma}$ は，

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi} \tag{2.2}$$

で与えられる．

本稿では，データの当てはまりの悪さを表す損失関数 $\ell(\boldsymbol{\Lambda}, \boldsymbol{\Psi})$ にパラメータ $\boldsymbol{\Lambda}$ に関する正則化項 $P(\boldsymbol{\Lambda})$ を加えた以下の関数

$$\ell_\rho(\boldsymbol{\Lambda}, \boldsymbol{\Psi}) = \ell(\boldsymbol{\Lambda}, \boldsymbol{\Psi}) + \rho P(\boldsymbol{\Lambda}) \tag{2.3}$$

の最小化によってパラメータを推定する (Choi et al. 2010; Hirose and Yamamoto 2015)．ここで，$\rho > 0$ は正則化パラメータであり，データの当てはまりの良さとモデルの構造抽出を調整する役割を果たす．適切な ρ の値を選ぶことにより，当てはまりがよく，かつ適切な構造を抽出できると考えられる．

損失関数 $\ell(\boldsymbol{\Lambda}, \boldsymbol{\Psi})$ は，多変量正規分布の対数尤度関数に対応する Discrepancy function (たとえば Lawley and Maxwell 1971; Bai et al. 2012) が用いられることが多い．しかし，因子分析における Discrepancy function は複雑であり，因子回転と k-means との関係性を議論する際，本質を理解しづらい．そこで，次の 2 乗損失関数を考える．

$$\ell(\boldsymbol{\Lambda}, \boldsymbol{\Psi}) = \|\boldsymbol{S} - \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T - \boldsymbol{\Psi}\|_F^2. \tag{2.4}$$

ただし，$\boldsymbol{S}$ は標本分散共分散行列（不偏分散）とする．また，$\|\cdot\|_F$ はフロベニウスノルムを表す．

2.2 正則化法と因子回転

正則化法は，因子回転と密接な関係がある (Hirose and Yamamoto 2015)．本節では，その関係性について解説する．

2.2.1 因子回転

正則化法について述べる前に，まず，(2.4) 式の正則化項なしの損失関数 $\ell(\boldsymbol{\Lambda}, \boldsymbol{\Psi})$ の最小化問題を考える．任意の $m \times m$ 直交行列 $\boldsymbol{T}$ に対し，$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi} = (\boldsymbol{\Lambda}\boldsymbol{T})(\boldsymbol{\Lambda}\boldsymbol{T})^T + \boldsymbol{\Psi}$

が成り立つため，$\ell(\boldsymbol{\Lambda}, \boldsymbol{\Psi}) = \ell(\boldsymbol{\Lambda}\boldsymbol{T}, \boldsymbol{\Psi})$ となり，(2.4) 式を最小化する因子負荷行列 $\boldsymbol{\Lambda}$ は無数に存在する．

そこで，無数の因子負荷行列から最も解釈しやすい因子負荷行列を推定する因子回転を行う．具体的に，まず，(2.4) 式の損失関数 $\ell(\boldsymbol{\Lambda}, \boldsymbol{\Psi})$ を最小にする $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}$ の 1 つを $\hat{\boldsymbol{\Lambda}}_0$, $\hat{\boldsymbol{\Psi}}_0$ とおく．次に，回転基準 $Q(\boldsymbol{\Lambda})$ を最小（もしくは最大）とする回転行列 $\boldsymbol{T}$ を求めるという以下の最適化問題を解く．

$$\min_{\boldsymbol{\Lambda}} Q(\boldsymbol{\Lambda}), \text{ subject to } \boldsymbol{\Lambda} = \hat{\boldsymbol{\Lambda}}_0 \boldsymbol{T}, \ \boldsymbol{T} \in \mathcal{O}(m). \tag{2.5}$$

ただし，$\mathcal{O}(m) = \{\boldsymbol{T} \in \mathbb{R}^{m \times m} \mid \boldsymbol{T}^T \boldsymbol{T} = \boldsymbol{I}_m\}$ とする．

2.3 因子回転の拡張

いま，$\hat{\boldsymbol{\Lambda}}_0$ が，回転の不定性を除いて一意に存在すると仮定すると，問題 (2.5) は次のように表される．

$$\min_{\boldsymbol{\Lambda}, \boldsymbol{\Psi}} Q(\boldsymbol{\Lambda}), \text{ subject to } \ell(\boldsymbol{\Lambda}, \boldsymbol{\Psi}) = \hat{\ell}. \tag{2.6}$$

ただし，$\hat{\ell} = \min_{\boldsymbol{\Lambda}, \boldsymbol{\Psi}} \ell(\boldsymbol{\Lambda}, \boldsymbol{\Psi})$ である．

Hirose and Yamamoto (2015) は，因子回転よりも解釈しやすい因子負荷行列を推定するために，最小 2 乗推定値を求めることを妥協した以下の問題を考えた．

$$\min_{\boldsymbol{\Lambda}, \boldsymbol{\Psi}} Q(\boldsymbol{\Lambda}), \text{ subject to } \ell(\boldsymbol{\Lambda}, \boldsymbol{\Psi}) \leq \ell^*. \tag{2.7}$$

ただし，ℓ^* $(\ell^* \geq \hat{\ell})$ は定数とする．ℓ^* の値はデータへの当てはまりのよさと因子負荷行列の構造抽出のバランスを調整している．$\ell^* = \hat{\ell}$ のときは因子回転と一致し，ℓ^* が大きくなるにつれて，構造抽出が重視される．

問題 (2.7) の解は，次の関数 $\ell_\rho(\boldsymbol{\Lambda}, \boldsymbol{\Psi})$ の最小化によって得られる．

$$\ell_\rho(\boldsymbol{\Lambda}, \boldsymbol{\Psi}) = \ell(\boldsymbol{\Lambda}, \boldsymbol{\Psi}) + \rho Q(\boldsymbol{\Lambda}). \tag{2.8}$$

ただし，$\rho > 0$ は ℓ^* に対応するパラメータである．ここで，(2.8) 式の $Q(\cdot)$ と ρ は，それぞれ (2.3) 式の正則化項と正則化パラメータとみなすことができる．すなわち，因子回転の自然な拡張が，正則化法となっている (Hirose and Yamamoto 2015)．また，正則化法で $\rho \to 0$ としたときの推定値が因子回転に対応する．それゆえ，正則化法は因子回転よりも構造抽出を重視する手法であることがわかる．

2.4 正則化項

およそ 15 年前，Jennrich (2004, 2006) は，因子回転の基準として，パラメータの各成分に対して独立に制約を置く Component loss criterion とよばれる基準

$$P(\mathbf{\Lambda}) = \sum_{i=1}^{p}\sum_{k=1}^{m} P_0(|\lambda_{ik}|) \tag{2.9}$$

を提案した．ただし，$P_0(\cdot)$ は，component loss function とよばれるもので，この関数を変えることにより様々な正則化項が得られる．たとえば，$P_0(x) = x$ のとき，L_1 正則化法でよく用いられる Lasso (Tibshirani 1996)

$$P(\mathbf{\Lambda}) = \sum_{i=1}^{p}\sum_{k=1}^{m} |\lambda_{ik}|, \tag{2.10}$$

となり，より複雑な関数を仮定することにより，以下の MCP (Minimax Concave Penalty, Zhang 2010)

$$\begin{aligned}\rho P(\mathbf{\Lambda};\rho;\gamma) &= \sum_{i=1}^{p}\sum_{k=1}^{m} \rho \int_0^{|\lambda_{ik}|} \left(1 - \frac{x}{\rho\gamma}\right)_+ dx \\ &= \sum_{i=1}^{p}\sum_{k=1}^{m} \left\{ \rho\left(|\lambda_{ik}| - \frac{\lambda_{ik}^2}{2\rho\gamma}\right) I(|\lambda_{ik}| < \rho\gamma) + \frac{\rho^2\gamma}{2} I(|\lambda_{ik}| \geq \rho\gamma) \right\}\end{aligned}$$

が得られる．ここで，$\gamma > 0$ は MCP のチューニングパラメータで，γ の値が小さいほどスパースになりやすい．なお，$\gamma \to \infty$ のとき，MCP は Lasso と一致する．Lasso や MCP を用いることにより，因子負荷量のいくつかは正確に 0 と推定されるため，解釈しやすい構造が得られる．

しかしながら，Lasso タイプの正則化法によって得られた因子負荷行列は，従来の因子回転でよく用いられる Varimax 基準や Quartimin 基準によって得られる因子負荷行列と大きく異なる．実際，Lasso を用いると，$\mathbf{\Lambda}$ の成分のうち一部を「ぴったり」0 にしようとするため，他の非ゼロのパラメータに影響を及ぼすことがある (Hirose and Yamamoto 2014)．一方で，従来の因子回転は，ぴったり 0 にするのでなく，0 と非 0 のコントラストを明確にする．それゆえ，Lasso よりも従来の因子回転のほうが頑健な結果が得られる．

3 Quartimin 基準

本稿では，直交回転・斜交回転でよく用いられる Quartimin 基準 (Carroll 1953)

$$Q_{\mathrm{QMIN}}(\mathbf{\Lambda}) = \sum_{k=1}^{m}\sum_{l\neq k}^{m}\sum_{i=1}^{p} \lambda_{ik}^2\lambda_{il}^2 \tag{3.1}$$

	F_1	F_2	F_3
X_1	∗	0	0
X_2	∗	0	0
X_3	∗	0	0
X_4	∗	0	0
X_5	0	∗	0
X_6	0	∗	0
X_7	0	∗	0
X_8	0	∗	0
X_9	0	0	∗
X_{10}	0	0	∗
X_{11}	0	0	∗

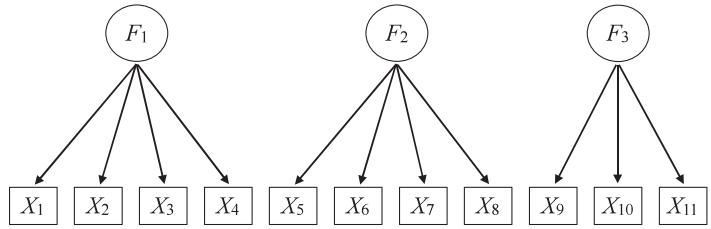

図 3.1: Perfect simple structure を持つ因子負荷行列（左）と対応するパス図（右）．左図の $*$ は非ゼロ要素を表す．また，それぞれの因子がクラスターと対応しており，この場合は，1〜4 変数，5〜8 変数，9〜11 変数それぞれがクラスターになっている．

に基づく以下の関数

$$\ell_\rho(\mathbf{\Lambda}, \mathbf{\Psi}) = \ell(\mathbf{\Lambda}, \mathbf{\Psi}) + \rho Q_{\mathrm{QMIN}}(\mathbf{\Lambda}) \tag{3.2}$$

の最小化問題によってパラメータを推定することを考える．いま，(3.1) 式が最小となるのは，因子負荷行列の各行の非ゼロ要素が 1 以下のとき，すなわち，完全単純構造のときに限る．それゆえ，$\rho \to \infty$ のとき，得られる因子負荷行列は完全単純構造となる．

完全単純構造は，因子分析において望ましい性質のひとつである (Browne 2001)．実際，完全単純構造であるとき，どの因子がどの変数に影響しているのか明確に分かれているため，因子の解釈が容易になる．また，完全単純構造は変数クラスタリングと対応している (Browne 2001)．たとえば，図 3.1 は $p = 11, m = 3$ のときの完全単純構造を表している．実際，1〜4 変数，5〜8 変数，9〜11 変数それぞれが 1 つのクラスターになっているため，観測される 11 変数を 3 つのクラスターに分けたとみなすことができる．

なお，Hirose and Terada (2016) は，(3.1) 式に絶対値に基づく項を追加した Prenet (Product elastic net) 正則化

$$Q_{\mathrm{Prenet}}(\mathbf{\Lambda}) = \sum_{k=1}^{m} \sum_{l \neq k}^{m} \sum_{i=1}^{p} (\alpha \lambda_{ik}^2 \lambda_{il}^2 + (1-\alpha)|\lambda_{ik}||\lambda_{il}|), \quad (0 \leq \alpha < 1) \tag{3.3}$$

を提案した．絶対値の項を含む Prenet 正則化を用いることにより，$\rho \to \infty$ でなくても，十分大きい有限の ρ に対し，完全単純構造を推定することができる．

4 k-means との関係性

先に述べたように，ρ を十分大きくしたときに得られる完全単純構造は，変数クラスタリングと対応する．それでは，そのクラスタリングは，従来のクラスタリング手法とどの

ような関係性があるだろうか．本節ではこのことを考察する．

いま，$\boldsymbol{X}_n$ を $n \times p$ データ行列とし，$\boldsymbol{X}_n = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_p^*)$ で表されるとする．ここで，$\boldsymbol{x}_i^*$ は $\boldsymbol{X}_n$ の第 i 列ベクトルとする．このとき，$\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_p^*$ を k-means によってクラスタリングすることを考える．C_k $(k = 1, \ldots, m)$ を第 k クラスターに属する変数の番号を集めた集合とする．たとえば，図 3.1 に対しては，$C_1 = \{1, 2, 3, 4\}$, $C_2 = \{5, 6, 7, 8\}$, $C_3 = \{9, 10, 11\}$ である．このとき，k-means の目的関数を $(N-1)^{-1}$ 倍した関数は

$$\frac{1}{N-1} \sum_{k=1}^{m} \sum_{i \in C_k} \|\boldsymbol{x}_i^* - \boldsymbol{\mu}_k\|^2 = \sum_{i=1}^{p} s_{ii} - \sum_{k=1}^{m} \frac{1}{p_k} \sum_{i \in C_k} \sum_{j \in C_k} s_{ij} \tag{4.1}$$

で与えられる．ただし，$p_k = \#\{C_k\}$, $\boldsymbol{\mu}_k = \frac{1}{p_k} \sum_{i \in C_k} \boldsymbol{x}_i^*$ $(k = 1, \ldots, m)$ とする．また，s_{ij} は，標本分散共分散行列 $\boldsymbol{S}$ の第 (i, j) 成分であり，データが中心化されているため，$s_{ij} = \boldsymbol{x}_i^{*T} \boldsymbol{x}_j^* / (N-1)$ で表される．

(4.1) 式の k-means による変数クラスタリングは以下の問題と同等になる (Ding et al. 2005).

$$\min_{\boldsymbol{\Lambda}} \|\boldsymbol{S} - \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T\|_F^2, \quad \text{subject to } \lambda_{ik} = \begin{cases} 1/\sqrt{p_k} & i \in C_k, \\ 0 & i \notin C_k. \end{cases} \tag{4.2}$$

実際，(4.2) 式の目的関数を展開すると，$\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} = \boldsymbol{I}_m$ を満たすため，

$$\begin{aligned} \|\boldsymbol{S} - \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T\|_F^2 &= \|\boldsymbol{S}\|_F^2 - 2\mathrm{tr}(\boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T) + \|\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T\|_F^2 \\ &= \|\boldsymbol{S}\|_F^2 - 2\sum_{k=1}^{m} \frac{1}{p_k} \sum_{i \in C_k} \sum_{j \in C_k} s_{ij} + m \\ &= -2\sum_{k=1}^{m} \frac{1}{p_k} \sum_{i \in C_k} \sum_{j \in C_k} s_{ij} + C \end{aligned}$$

となり，定数項を除いて (4.1) と同等の関数となっている．ここで，C は $\boldsymbol{\Lambda}$ に依存しない定数とする．

(4.2) 式の最適化問題は，因子負荷行列に強い制約

$$\lambda_{ik} = \begin{cases} 1/\sqrt{p_k} & i \in C_k, \\ 0 & i \notin C_k. \end{cases} \tag{4.3}$$

をおいた因子分析モデルの推定であるとみなせる．そのため，因子分析モデルと k-means は密接に結びついていることがわかる．しかしながら，(4.2) 式の最適化問題は，独自分散 $\boldsymbol{\Psi}$ に依存しないなど，因子分析モデルとはかなりかけ離れているものになっている．そこで，この問題を因子分析モデルの推定に近づけることを考える．

まず，(4.3) 式が成り立てば，$\lambda_{ik}^2 \lambda_{il}^2 = 0$ $(k \neq l)$ が成り立つ．そこで，(4.2) 式の問題をゆるめ，さらに目的関数に $\boldsymbol{\Psi}$ を加えた以下の最適化問題を考える．

$$\min_{\boldsymbol{\Lambda}, \boldsymbol{\Psi}} \|\boldsymbol{S} - \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T - \boldsymbol{\Psi}\|^2 \text{ subject to } \lambda_{ik}^2 \lambda_{il}^2 = 0 \ (k \neq l). \tag{4.4}$$

この最適化問題の解は，(3.2) 式の最小化問題の $\rho \to \infty$ としたときの推定値と一致する．そのため，Quartimin (もしくは Prenet) 基準に基づく正則化法は，k-means を緩和した問題と対応している．

それでは，k-means に対応する問題 (4.2) と，Qaurtimin 基準で $\rho \to \infty$ とした問題 (4.4) は，どのような違いがあるのだろうか．以下，その違いをまとめた．

- 問題 (4.4) の条件では，各クラスターに対して因子負荷量の値がすべて同じ値にならなくても良い．さらに，因子負荷量が負の値をとっても良い．これは，負の相関を持つ 2 つの変数が同じクラスターに属するのを許すことを意味する．以上より，問題 (4.4) の因子分析のほうが，問題 (4.2) の k-means よりも変数とクラスターとの関係性をより豊かに表現することができる．

- 問題 (4.4) では，$\boldsymbol{\Psi}$ が正定値であると仮定して $\boldsymbol{\Psi}$ を推定する．それゆえ，各変数に対して共通因子で説明できない部分を独自因子として表現している．一方，k-means に対応する問題 (4.2) は，そのような独自因子を仮定していない．

これらの違いを踏まえて，実際のデータ解析を行う上で，因子分析による完全構造推定が，k-means による変数クラスタリングよりもどのような点で優れているのだろうか．筆者は，以下の 2 つが因子分析の強みだと考えている．

1. 変数の数が多い場合，どのクラスターにも属さないような変数がある状況が考えられる．たとえば，画像のピクセルを変数だと思ったピクセルクラスタリング (Hirose and Terada 2016) が挙げられる．このとき，分散が小さかったり他の画素とは相関関係のない画素が存在する．その場合，通常の k-means ではすべての画素を平等にクラスタリングするが，因子分析では，相関関係のない画素は独自因子として表現される．

2. 因子分析では，負の相関関係がある変数に対して同じクラスターだとみなされるため，たとえばアンケートデータに対し，「商品 A は好きであるが商品 B は好きでない」などの負の相関関係のある項目を，同じクラスターとみなすことができる．

5　おわりに

本稿では，Quartimin 基準による因子分析モデルの構造正則化について考察した．Quartimin 基準による正則化項を用いると，$\rho \to \infty$ のときに完全単純構造が得られる．また，この問題は k-means の変数クラスタリングの一般化とみなすことができる．さらに，実用上，k-means よりも優れている点があることを述べた．

因子分析モデルにおいて，変数の数が小さい場合は，わざわざ完全単純構造を推定しなくてもよく，Varimax 回転などの因子回転を用いて因子負荷行列の１つ１つの成分を見て因子を解釈すれば良い．ところが，遺伝子データなど，変数の数が大きい場合は，通常の因子分析では因子の解釈がしにくくなる．その場合，完全単純構造が有用となると考えられる．変数の関係性を解釈しやすく，かつデータへのフィッティングも考慮した因子分析は，今後高次元データのクラスター分析手法として大きな可能性を秘めている筆者は考えている．

謝辞

本研究は JSPS 科研費 19K11862 および JST COI の助成を受けたものです．

参考文献

Bai, J., Li, K. et al. (2012), “Statistical analysis of factor models of high dimension,” *The Annals of Statistics*, 40(1), 436–465.

Bernaards, C. A., and Jennrich, R. I. (2003), “Orthomax rotation and perfect simple structure,” *Psychometrika*, 68(4), 585–588.

Browne, M. W. (2001), “An overview of analytic rotation in exploratory factor analysis,” *Multivariate Behavioral Research*, 36(1), 111–150.

Carroll, J. B. (1953), “An analytical solution for approximating simple structure in factor analysis,” *Psychometrika*, 18(1), 23–38.

Choi, J., Zou, H., and Oehlert, G. (2010), “A penalized maximum likelihood approach to sparse factor analysis,” *Statistics and its Interface*, 3(4), 429–436.

Ding, C. H., He, X., and Simon, H. D. (2005), On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering.,, in *SDM*, Vol. 5, SIAM, pp. 606–610.

Hendrickson, A. E., and White, P. O. (1964), “Promax: A quick method for rotation to oblique simple structure,” *British Journal of Statistical Psychology*, 17(1), 65–70.

Hirose, K., and Terada, Y. (2016), “Simple structure estimation via prenet penalization,” *arXiv.org*, .

Hirose, K., and Yamamoto, M. (2014), “Estimation of an oblique structure via penalized likelihood factor analysis,” *Computational Statistics & Data Analysis*, 79, 120–132.

Hirose, K., and Yamamoto, M. (2015), "Sparse estimation via nonconcave penalized likelihood in factor analysis model," *Statistics and Computing*, 25(5), 863–875.

Jacob, L., Obozinski, G., and Vert, J.-P. (2009), Group lasso with overlap and graph lasso,, in *Proceedings of the 26th annual international conference on machine learning*, ACM, pp. 433–440.

Jennrich, R. I. (2004), Rotation to simple loadings using component loss functions: The orthogonal case,, in *Psychometrika*, University of California, Los Angeles, Los Angeles, United States, Springer-Verlag, pp. 257–273.

Jennrich, R. I. (2006), "Rotation to Simple Loadings Using Component Loss Functions: The Oblique Case," *Psychometrika*, 71(1), 173–191.

Kaiser, H. F. (1958), "The varimax criterion for analytic rotation in factor analysis," *Psychometrika*, 23(3), 187–200.

Lawley, D., and Maxwell, A. (1971), *Factor Analysis as a Statistical Method*, Butterworths mathematical tests American Elsevier Publishing Company.

Park, H., Niida, A., Miyano, S., and Imoto, S. (2015), "Sparse overlapping group lasso for integrative multi-omics analysis," *Journal of Computational Biology*, 22(2), 73–84.

Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B-Methodological*, 58(1), 267–288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005), "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1), 91–108.

Xin, B., Kawahara, Y., Wang, Y., Hu, L., and Gao, W. (2016), "Efficient generalized fused lasso and its applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(4), 60.

Yuan, M., and Lin, Y. (2006), "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1), 49–67.

Zhang, C.-H. (2010), "Nearly unbiased variable selection under minimax concave penalty," *The Annals of Statistics*, 38(2), 894–942.