日本データベース学会和文論文誌
Vol. 20-J, Article No. 5, 2022 年 3 月

潜在変数を用いた Gaussian Graphical Model の柔軟な疎構造推定

小山 和輝 [1]　切通 恵介 [2]　大川内 智海 [3]
泉谷 知範 [4]

変数間の依存関係を表現する Gaussian Graphical Model に対して，Graphical LASSO はスパースなグラフ構造を抽出する．本論文では，これに潜在構造正則化学習の枠組みを取り入れて，変数に内在する構造を反映させた新しい正則化手法を提案する．提案手法では精度行列を分解した潜在変数に対してグループ構造を設定し，スパース性を誘導する多変量 Student's t 分布を事前分布とした確率モデルで，各グループの重要度とスパースな精度行列を同時推定する．本論文では提案手法を人工データと実データに適用し，スパース度合いや異常スコアなどを用いて有効性を検証した．その結果，提案手法は入力構造に沿って変数間の依存関係をスパースに抽出することが確認された．

1　はじめに

高次元データやグラフからの知識発見は，多くの社会科学現象のデータマイニングにおける重要な課題である．特に，システムが複数の要因で構成されている場合には，変数間に相互依存性があると仮定し，データからその依存関係を抽出してシステムの理解に繋げたいとすることは自然な着想である [1]．変数間の依存関係をグラフ化する最も直感的な方法のひとつとして，2 つの変数間の依存関係にのみ着目し，他の全ての変数が与えられた条件下で 2 つの変数が条件付き独立であることと 2 つの変数間に辺がないことを等価として，条件付き独立性に基づいたグラフを構築する Pairwise Markov Graph がある [2–4]．特に，データが多変量正規分布に従うと仮定する場合には Gaussian Graphical Model と呼ばれ，解析的な簡便さや性質の良さもあり，様々な分野で広く応用されている [5].

このような Gaussian Graphical Model の枠組みに対して，Banerjee ら及び Friedman らによって，スパースなグラフ構造を抽出する Graphical LASSO が体系的に提案された [6, 7]．Graphical LASSO では，データの従う多変量正規分布の精度行列に対して，スパース性を誘導するラプラス分布を事前分布として仮定することにより，変数間の依存関係を反映したスパースなグラフ構造を得ることができる．しかし，Graphical LASSO は全ての依存関係が同一の事前分布に従っていることを仮定しており，いくつかの問題設定においては必ずしも適応的な方法ではないと考えられる．例えば，変数間の個別ないし複数の依存関係に対して何らかの関連が認められる場合や，個々の依存関係への解析者の興味関心が異なる場合，あるいはデータに関する事前知識やノウハウを反映した推定を行いたい場合などが挙げられる．このように，変数や依存関係の中に何らかの構造を当てはめたい場合，単純な L_1 正則化を用いた Graphical LASSO では，全ての依存関係に同等の罰則が課されるのみであるため，現象をうまく捉えられない可能性がある [8–10].

本論文では，潜在構造正則化学習のひとつである Latent Group LASSO [11–14] の枠組みに着目し，Gaussian Graphical Model に内在する構造化されたスパースな依存関係を抽出する手法を提案する．提案手法における構造化されたスパースな関係性とは，1 つ以上の依存関係に対して重複を許容して複数のグループを定義し，グループ構造に沿ったスパースな推論を行うことを意味する．Graphical LASSO に対して，このような構造を導入する研究は様々あるが [15–19]，提案手法は個々のグループに対して潜在変数と確率モデルを設定する点で大きく異なる．特に Tao らは重複を許容したグループノルムに基づく手法を提案しているが [18]，提案手法では確率モデルに基づいた最適化の過程で，個々のグループの重要度も自動調整し，不要なグループに適応的に強い罰則を与えるという点で，より柔軟性が高い手法であると考えられる．

2　関連研究

2.1　Group LASSO

構造正則化学習を行う最もシンプルかつ重要な手法として，Yuan らによって提案された Group LASSO がある [10]．Group LASSO は LASSO [8] に似たスパースな解を導出するが, LASSO とは異なり事前に与えられた離散構造に沿ったスパースな解が誘導される．

ここで，パラメータ $\boldsymbol{\omega} = [\omega_1, \omega_2, \ldots, \omega_P]^\top$ のインデックス集合を $\mathcal{I} = \{1, 2, \ldots, P\}$ とし，変数間の離散構造を集合で表現する．すなわち，例えば $\{i, j, k\}$ は i 番目，j 番目，k 番目のパラメータで構成されるグループ構造を意味する．これに従うと，$2^{\mathcal{I}}$ を $\mathcal{I}$ の冪集合として，変数間の離散構造全体の集合は $\mathcal{G} \subseteq 2^{\mathcal{I}}$ と表現できる．このとき，Group LASSO の正則化項 $\Omega(\boldsymbol{\omega})$ は以下のように与えられる．

$$\Omega(\boldsymbol{\omega}) = \sum_{G \in \mathcal{G}} \|\boldsymbol{\omega}_G\|_2 \tag{1}$$

ただし，$\boldsymbol{\omega}_G \in \mathbb{R}^P$ は $\boldsymbol{\omega}$ のうち $\mathcal{I} \backslash G$ に属するインデックスをゼロとしたベクトルを意味する．

2.2　Latent Group LASSO

Group LASSO に対して，Latent Group LASSO は潜在変数を用いた構造正則化学習であり，G を台にもつ独立な潜在ベクトル $\boldsymbol{\nu}_G \in \mathbb{R}^P$ の和で $\boldsymbol{\omega}$ が与えられる [11–14]．そのため Group LASSO とは異なり，個々のグループ G に対して（線形和で $\boldsymbol{\omega}$ を定義していることを除いて）独立に潜在変数が導入されており，潜在変数の各成分が複数のグループ間で共有されていない

[1] 非会員　NTT コミュニケーションズ株式会社
kazuki.koyama@ntt.com
[2] 非会員　NTT コミュニケーションズ株式会社
k.kiritoshi@ntt.com
[3] 非会員　NTT コミュニケーションズ株式会社
t.okawachi@ntt.com
[4] 非会員　NTT コミュニケーションズ株式会社
tomonori.izumitani@ntt.com

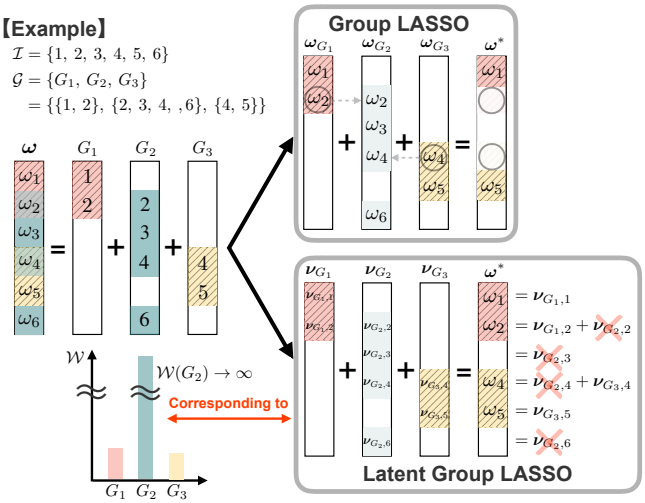

図 1　Group LASSO と Latent Group LASSO の推定解の違い．ここでは 6 変数と重複 3 グループの例を考えている．G_2 のみが不要なグループと判断されたとき，Group LASSO の推定解では ω_2 と ω_4 が 0 に縮退するのに対して，Latent Group LASSO では潜在変数 $\boldsymbol{\nu}_{G_2}$ がゼロベクトルとなるのみで，最終的な推定解でも ω_2 と ω_4 は 0 に縮退しない．このとき対応する重要度 $\mathcal{W}(G_2)$ は $\mathcal{W}(G_1)$ や $\mathcal{W}(G_3)$ に対して十分大きな値をとる．

ことが特徴である．このような潜在変数を用いて，Latent Group LASSO の正則化項は以下のように与えられる．

$$\Omega(\boldsymbol{\omega}) = \sum_{G \in \mathcal{G}} \|\boldsymbol{\nu}_G\|_2 \mathcal{W}(G) \quad \text{s.t.} \quad \sum_{G \in \mathcal{G}} \boldsymbol{\nu}_G = \boldsymbol{\omega} \tag{2}$$

ここで，$\mathcal{W}(G)$ は個々の $\boldsymbol{\nu}_G$ の重要度を調整する関数であり，実数の正値全体の集合 $\mathbb{R}_+$ を用いて $\mathcal{W} : 2^{\mathcal{I}} \to \mathbb{R}_+$ で定義されている．正則化項 (2) の形式から，$\mathcal{W}(G)$ の値が小さいほど，対応する $\boldsymbol{\nu}_G$ は非ゼロベクトルとなりやすくなり，G はより重要なグループであるとみなすことができる．

潜在変数を導入することで，単純にモデルの表現力が増すだけでなく，Group LASSO の欠点を補うことができる．図 1 で例示するように，Group LASSO は各グループでパラメータを共有しているため，ゼロベクトルに縮退するグループの影響で最終的な解の非ゼロパターンは個々のグループの非ゼロパターンの積集合となる．一方，Latent Group LASSO では，各グループに個別の潜在変数を導入することでパラメータの共有を回避しているため，最終的な解の非ゼロパターンは個々のグループの非ゼロパターンの和集合となる．従って，Latent Group LASSO は重要と判断された構造に属する全てのパラメータが最終的な解に反映されるため，解析者にとってより自然な解を与えることができる．

2.3　Gaussian Graphical Model

あらかじめ平均 0 に中心化されたデータ $\boldsymbol{x} \in \mathbb{R}^M$ に対して，Gaussian Graphical Model では M 個の変数を頂点とするグラフ構造を考える．このモデルにおいて x_i と x_j の間に辺がないことは，これら 2 つの変数以外の全ての変数が与えられた場合に x_i と x_j が条件付き独立であることを意味する [5]．従って，Gaussian Graphical Model は次の M 次元多変量正規分布に従って変数間の辺の存在が定義される [5]．

$$\mathcal{N}\left(\boldsymbol{x} \,|\boldsymbol{0}, \boldsymbol{\Lambda}^{-1}\right) = \frac{(\det \boldsymbol{\Lambda})^{\frac{1}{2}}}{(2\pi)^{\frac{M}{2}}} \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{\Lambda} \boldsymbol{x}\right) \tag{3}$$

ここで，$\boldsymbol{\Lambda} \in \mathbb{R}^{M \times M}$ は分散共分散行列の逆行列であり，一般に分布の精度行列と呼ばれている．また $\boldsymbol{x}_{-\{i,j\}}$ を x_i と x_j 以外の全ての変数とすると，多変量正規分布の仮定から x_i と x_j の依存関係を表現する条件付き確率は次のように書き下せる．

$$\begin{aligned} &p(x_i, x_j | \boldsymbol{x}_{-\{i,j\}}) \\ &\quad \propto \exp\left(-\frac{1}{2}\left(\Lambda_{ii}x_i^2 + 2\Lambda_{ij}x_i x_j + \Lambda_{jj}x_j^2\right)\right) \end{aligned} \tag{4}$$

従って，もし $\Lambda_{ij} = 0$ であるならば，x_i と x_j が条件付き独立となることがわかる．

$$\Lambda_{ij} = 0 \Rightarrow x_i \perp\!\!\!\perp x_j \,|\, \boldsymbol{x}_{-\{i,j\}} \tag{5}$$

よって，Gaussian Graphical Model に基づいたグラフ構造の学習は，データから確率分布を学習する問題に帰着される [20]．

2.4　Graphical LASSO

Graphical LASSO の目的は，Gaussian Graphical Model の枠組みに基づいて変数間の依存関係を表現するスパースな精度行列 $\boldsymbol{\Lambda}$ を抽出することである．抽出された $\boldsymbol{\Lambda}$ は，x_i と x_j に本質的な依存関係が存在する場合には $\Lambda_{ij} \neq 0$ となり，依存関係が存在しない場合や，ノイズなどの影響による見かけの関係性に対しては $\Lambda_{ij} = 0$ となることが期待される．

あらかじめ平均 0 と標準偏差 1 に正規化された観測データを $\mathcal{D} = \{\boldsymbol{x}^{(n)} | \boldsymbol{x}^{(n)} \in \mathbb{R}^M, n = 1, \ldots, N\}$ とする．このとき $\mathcal{D}$ を適当に並べたデータ行列を $\boldsymbol{X} = [\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}] \in \mathbb{R}^{M \times N}$ とすれば，標本分散共分散行列は $\boldsymbol{\Upsilon} = \boldsymbol{X}\boldsymbol{X}^\top / N$ で与えられる．Graphical LASSO では，次のようなスパースな解を誘導するラプラス分布を精度行列 $\boldsymbol{\Lambda}$ の事前分布として仮定し，最大事後確率推定に基づいて $\boldsymbol{\Lambda}$ を求める [6,7,21]．

$$p(\boldsymbol{\Lambda}) = \frac{\rho}{2} \exp\left(-\rho \|\boldsymbol{\Lambda}\|_1\right) \tag{6}$$

ここで，$1/\rho$ はラプラス分布の尺度パラメータで，ρ はスパース推定における正則化係数に対応する．従って，(3) と (6) から，Graphical LASSO では次の最適化問題を考える．

$$\boldsymbol{\Lambda}^* = \arg\max_{\boldsymbol{\Lambda}} \left\{ \log\left(p(\boldsymbol{\Lambda}) \prod_{n=1}^{N} \mathcal{N}\left(\boldsymbol{x}^{(n)} | \boldsymbol{0}, \boldsymbol{\Lambda}^{-1}\right)\right)\right\} \tag{7}$$

さらに (7) の最適化問題は，定数項が最適解に影響を与えないことに注意すると，以下の最適化問題と等価となる．

$$\boldsymbol{\Lambda}^* = \arg\max_{\boldsymbol{\Lambda}} f(\boldsymbol{\Lambda}; \boldsymbol{\Upsilon}, \rho) \tag{8}$$

$$f(\boldsymbol{\Lambda}; \boldsymbol{\Upsilon}, \rho) \equiv \log\det \boldsymbol{\Lambda} - \mathrm{tr}\,(\boldsymbol{\Upsilon}\boldsymbol{\Lambda}) - \rho\|\boldsymbol{\Lambda}\|_1 \tag{9}$$

Friedman らは (8) が凸最適化問題であることに着目し，この問題を効率的に解く次のような劣勾配アルゴリズムを提案している [7]．すなわち，特定の変数 x_i に注目し，それが最後の行となるようにデータ行列 $\boldsymbol{X}$ を並び替えても一般性は失われないこと

を利用すれば，次のように $\mathbf{\Lambda}, \mathbf{\Sigma} \equiv \mathbf{\Lambda}^{-1}, \mathbf{\Upsilon}$ を x_i に関連する要素が最後の行と列となるように分割することができる．

$$\mathbf{\Lambda}, \mathbf{\Sigma}, \mathbf{\Upsilon} = \begin{bmatrix} \boldsymbol{L}_i & \boldsymbol{l}_i \\ \boldsymbol{l}_i^\top & \lambda_i \end{bmatrix}, \begin{bmatrix} \boldsymbol{S}_i & \boldsymbol{s}_i \\ \boldsymbol{s}_i^\top & \sigma_i \end{bmatrix}, \begin{bmatrix} \boldsymbol{U}_i & \boldsymbol{u}_i \\ \boldsymbol{u}_i^\top & v_i \end{bmatrix} \tag{10}$$

この分割に基づけば，x_i にのみ注目した $\partial f / \partial \mathbf{\Lambda} = 0$ の条件は，次のように表現される．

$$\sigma_i = v_i + \rho \tag{11}$$
$$\boldsymbol{s}_i = \boldsymbol{u}_i + \rho \,\mathrm{sign}\,(\boldsymbol{l}_i) \tag{12}$$

ただし (11) は，$\mathbf{\Lambda}$ が正定値行列であり，その対角成分は正値になることを利用している．ところで，(12) の非対角成分に関する条件は，$\boldsymbol{L}_i$ と $\boldsymbol{S}_i$ を固定し，$\boldsymbol{\omega}_i \equiv \boldsymbol{S}_i^{-1}\boldsymbol{s}_i$ と $\boldsymbol{S}_i\boldsymbol{l}_i + \lambda_i \boldsymbol{s}_i = \mathbf{0}$ であることを用いると，次の方程式と等価である．

$$\mathbf{0} = \frac{\partial}{\partial \boldsymbol{\omega}_i} \left\{ \frac{1}{2} \boldsymbol{\omega}_i^\top \boldsymbol{S}_i \boldsymbol{\omega}_i - \boldsymbol{u}_i^\top \boldsymbol{\omega}_i + \rho \left\| \boldsymbol{\omega}_i \right\|_1 \right\} \tag{13}$$

したがって，Graphical LASSO における最適な $\boldsymbol{\omega}_i^*$ は (13) を満たし，これは次の最適化問題と等価となる．

$$\boldsymbol{\omega}_i^* = \arg\min_{\boldsymbol{\omega}_i} \left\{ \frac{1}{2} \left\| \boldsymbol{S}_i^{-\frac{1}{2}} \boldsymbol{u}_i - \boldsymbol{S}_i^{\frac{1}{2}} \boldsymbol{\omega}_i \right\|^2 + \rho \left\| \boldsymbol{\omega}_i \right\|_1 \right\} \tag{14}$$

(14) は L_1 制約をもつ通常の二次計画問題であり，通常の LASSO などで良く知られた劣勾配アルゴリズムなどで効果的に解くことができる．よって，最適な $\boldsymbol{\omega}_i^*$ は (14) から容易に求まり，それに基づいて最適な $\boldsymbol{s}_i^*$ や $\boldsymbol{l}_i^*$ も決定される．数値計算的には，(11) と (14) を $i = 1, \ldots, M$ に関して収束するまで繰り返すことで，スパースな精度行列の最適解 $\mathbf{\Lambda}^*$ を得る．

3 提案手法

3.1 定式化

提案手法である Latent Structured Graphical LASSO の問題設定を与える．提案手法は潜在構造正則化学習を Graphical LASSO の枠組みに適用し，Gaussian Graphical Model に内在する構造に沿ったスパースな依存関係の抽出を目標としている．

まず非対角成分の定式化に関して，提案手法では (14) の着想に基づいた最適化問題を考える．Graphical LASSO との最大の違いは，精度行列 $\mathbf{\Lambda}$ の各要素に重複を許したグループ構造 $\mathcal{G}$ を導入し，$\mathcal{G}$ に沿ったスパースな精度行列の最適解 $\mathbf{\Lambda}^*$ を，Latent Group LASSO を適用した最適化問題に基づいて得ることである．精度行列は対称行列なので Λ_{ij} と Λ_{ji} は同一のパラメータとみなせることに注意すると，提案手法では非対角成分に関する $M(M-1)/2$ 個のパラメータに対してグループ構造 $\mathcal{G}$ を設定することとなる．入力構造と期待される推定解の挙動についての詳細は 3.2 節を参照されたい．

表記の簡単化のために，(14) の最適化問題において目的変数に対応する部分を $\boldsymbol{y}_i \equiv \boldsymbol{S}_i^{-\frac{1}{2}} \boldsymbol{u}_i$，説明変数に対応する部分を $\boldsymbol{Z}_i \equiv \boldsymbol{S}_i^{\frac{1}{2}}$ とする．Graphical LASSO では (14) を $i = 1, \ldots, M$ について解が収束するまで繰り返すが，提案手法では全ての非対角成分に対してグループ構造 $\mathcal{G}$ を設定したいので，これらを一括して最適化する方法を採用する．このため各 $i = 1, \ldots, M$ についてのパラメータ，目的変数，説明変数を，次のように適当に並べて最適化問題を構成する．

$$\boldsymbol{\omega} = \left[\boldsymbol{\omega}_1^\top, \ldots, \boldsymbol{\omega}_M^\top \right]^\top \tag{15}$$
$$\boldsymbol{y} = \left[\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_M^\top \right]^\top \tag{16}$$
$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{Z}_M \end{bmatrix} \tag{17}$$

ここで，$\boldsymbol{Z}$ は $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M$ についてのブロック対角行列である．(15), (16), (17) のような結合の仕方は、精度行列の対称性などの影響で同じ目的変数と説明変数の組みを複数回含んでいるため，このまま定式化すれば最適化問題は不必要に冗長となるが，ここでは表記の簡単化を優先して冗長性を許容することとする．冗長性を取り除いて定式化しても等価な最適解を得ることは可能である．以上の準備から，提案手法では (14) の最適化問題を一括し，潜在構造正則化学習に基づく正則化項を追加して，次の最適化問題を解くことを考える．

$$\boldsymbol{\omega}^* = \arg\min_{\boldsymbol{\omega} = \sum_{G \in \mathcal{G}} \boldsymbol{\nu}_G} \left\{ \frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\omega} \right\|^2 + \sum_{G \in \mathcal{G}} \left\| \boldsymbol{\nu}_G \right\|_2 \mathcal{W}(G)^{\frac{1}{2}} \right\} \tag{18}$$

(18) を解いて獲得した $\boldsymbol{\omega}^*$ から，$\boldsymbol{l}_i^* = \lambda_i \boldsymbol{\omega}_i^*$ によって，所望の構造に沿ったスパースな精度行列の最適解 $\mathbf{\Lambda}^*$ の非対角成分 $\boldsymbol{l}^* = \left[\boldsymbol{l}_1^\top, \ldots, \boldsymbol{l}_M^\top \right]^\top$ が得られる．

対角成分に関しては，提案手法は (18) の形式で全ての変数を結合した最適化問題を考えているため，(11) のような単純な更新則は適用することができない．よって，提案手法では非対角成分を $\boldsymbol{l}^*$ に固定した上で，対角成分 $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_M]^\top$ を次のように尤度が最大化するように最適化する方法を採用する．

$$\boldsymbol{\lambda}^* = \arg\max_{\boldsymbol{\lambda}} \left\{ \log \prod_{n=1}^{N} \mathcal{N}\left(\boldsymbol{x}^{(n)} \middle| \mathbf{0}, \mathbf{\Lambda}^{-1}(\boldsymbol{\lambda}, \boldsymbol{l}^*) \right) \right\} \tag{19}$$

ここで，$\mathbf{\Lambda}^{-1}(\boldsymbol{\lambda}, \boldsymbol{l}^*)$ は対角成分が $\boldsymbol{\lambda}$ で非対角成分が $\boldsymbol{l}^*$ であるような精度行列とする．このように対角成分を更新しても，(19) は観測データ $\mathcal{D}$ に基づく最尤推定であるため，精度行列 $\mathbf{\Lambda}$ が満たすべき正則性や正定値性などを保つよう働く．数値計算的には最適解が収束するまで (18) と (19) を繰り返し，所望の構造に沿ったスパースな精度行列の最終的な最適解 $\mathbf{\Lambda}^*$ を得る．

3.2 入力構造

提案手法では最適化の過程で不要と判断されるグループの重要度は $\mathcal{W}(G) \to \infty$ と無限大に発散し，対応する潜在変数は $\boldsymbol{\nu}_G \to 0$ と 0 に縮退していくので，基本的には $\mathcal{G}$ に解析者が考え得る全てのグループを任意に含めることができる．計算時間などを考慮するならば，何らかの基準に基づいて $\mathcal{G}$ を設定することも可能である．例えば，4.2 節の数値実験でみるように，国や地域などの観測データ特有の特徴に基づいて $\mathcal{G}$ を決めることもできる．あるいは，データに対する事前知識がない場合は，Graphical LASSO などの他の手法で得られた結果に基づいてグループ分けしてもよい．いずれにせよ，提案手法は $\mathcal{W}(G)$ を介して各グ

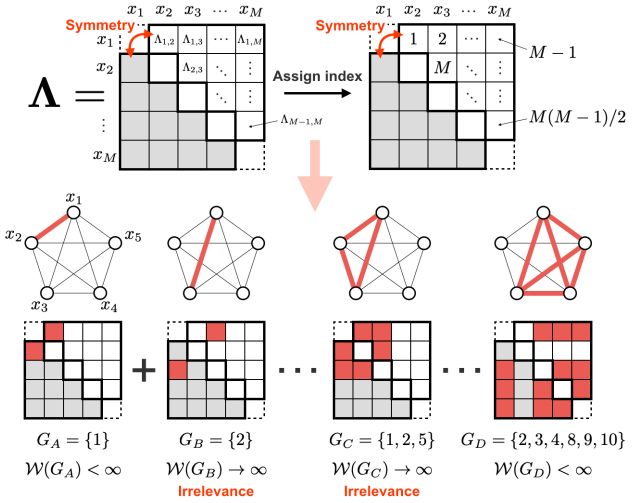

図 2　精度行列の非対角成分に導入するグループ構造の例．表記の簡単化のため図のようにインデックスを割り当てた後に，ここでは特に $M = 5$ の場合を考えることとする．このようなインデックスの割り当てに対して，提案手法では G_A や G_B のように単一の要素をひとつのグループとしたり，G_C や G_D のようにデータに対する事前知識や何らかの基準に基づいて複数の要素をひとつのグループに含めることができる．これらのグループ構造は重複を許して任意に柔軟に設定することができる．

ループ G の重要度を比較し，$\mathcal{G}$ に応じて適応的に重要な構造を抽出することができる．

特に，提案手法は $M(M-1)/2$ 個の非対角パラメータに対して，それぞれ単一の要素を単一のグループとするグループ構造 $\mathcal{G} = \{\{1\}, \ldots, \{M(M-1)/2\}\}$ を設定し，全ての重要度 ${}^{\forall}\mathcal{W}(G)$ $(G \in \mathcal{G})$ を同じ値に固定すれば，提案手法の最適化問題 (18) は Graphical LASSO の最適化問題 (14) を $\rho = \mathcal{W}(G)^{\frac{1}{2}}$ としたものと一致する．最適化手法の違いによる誤差はあるものの，このようなグループ構造を入力した提案手法は Graphical LASSO と同様の精度行列を抽出することが確認されている．よって，提案手法は Graphical LASSO を内包する純粋な拡張であるとみなすこともでき，入力するグループ構造 $\mathcal{G}$ に $\{\{1\}, \ldots, \{M(M-1)/2\}\}$ を含めることで，Graphical LASSO との比較も最適化問題に内包することができる．これらの提案手法の挙動については，図 2 で模式的に示している．

3.3　最適化

非対角成分の最適化については，Shervashidze らが提案した手法を応用して最適化を行う [14]．全般的な提案手法の最適化のフローは，アルゴリズム 1 にまとめている．

(18) の具体的な最適化では，線形モデルの K 分割マルチタスク学習に確率モデルを導入することで，潜在変数 $\boldsymbol{\nu}_G$ 及び対応する重要度 $\mathcal{W}(G)$ の推定を行う．ここでは，データから説明変数 $\boldsymbol{Z}^{(k)} \in \mathbb{R}^{N_k \times P}$ と目的変数 $\boldsymbol{y}^{(k)} \in \mathbb{R}^{N_k}$ を用意したとする．ただし，$k \in \{1, 2, \ldots, K\}$ であり，簡単化のため本研究では分割サンプル数 N_k は各 k に対して同一とする．

各分割において，目的変数と説明変数の間には標準偏差 σ の

Algorithm 1: Latent Structured Graphical LASSO

Input: 観測データ $\{\mathcal{D}^{(k)}\}_k$，グループ構造 $\mathcal{G}$，事前分布 q_G
Output: 最適精度行列 $\{\boldsymbol{\Lambda}^{(k)^*}\}_k$

$\{\boldsymbol{\Upsilon}^{(k)}\}_k \leftarrow \boldsymbol{X}^{(k)}\boldsymbol{X}^{(k)^\top}/N_k$
$\{\boldsymbol{\Lambda}^{(k)}\}_k, \{\boldsymbol{\Sigma}^{(k)}\}_k \leftarrow \{\boldsymbol{\Upsilon}^{(k)}\}_k, \{\boldsymbol{\Upsilon}^{(k)^{-1}}\}_k$
while $\{\boldsymbol{\Lambda}^{(k)}\}_k$ 未収束 **do**
　for $i = 1$ **to** M **do**
　　$\{\boldsymbol{y}_i^{(k)}\}_k, \{\boldsymbol{Z}_i^{(k)}\}_k \leftarrow \{\boldsymbol{S}_i^{(k)^{-\frac{1}{2}}}\boldsymbol{u}_i^{(k)}\}_k, \{\boldsymbol{S}_i^{(k)^{\frac{1}{2}}}\}_k$
　end for
　周辺尤度 $p\left(\{\boldsymbol{y}^{(k)}\}_k | \mathcal{W}\right)$ 最大化で
　$\{\boldsymbol{\nu}_G^{(k)}\}_{k,G}, \{\mathcal{W}(G)\}_G, \sigma$ 更新
　$\{\boldsymbol{\omega}^{(k)}\}_k \leftarrow \{\sum_G \boldsymbol{\nu}_G^{(k)}\}_k$
　for $i = 1$ **to** M **do**
　　$\{\boldsymbol{l}_i^{(k)}\}_k \leftarrow \{-\lambda_i^{(k)}\boldsymbol{\omega}_i^{(k)}\}_k$
　end for
　$\{\boldsymbol{\lambda}^{(k)}\}_k \leftarrow \left\{\arg\max_{\boldsymbol{\lambda}^{(k)}} \mathcal{N}\left(\mathcal{D}^{(k)} \middle| \boldsymbol{0}, \boldsymbol{\Lambda}^{(k)^{-1}}\left(\boldsymbol{\lambda}^{(k)}, \boldsymbol{l}^{(k)}\right)\right)\right\}_k$
　$\{\boldsymbol{l}^{(k)}\}_k, \{\boldsymbol{\lambda}^{(k)}\}_k$ から $\{\boldsymbol{\Lambda}^{(k)}\}_k, \{\boldsymbol{\Sigma}^{(k)}\}_k$ 更新
end while

ガウス雑音を加えた，以下のような関係が仮定される．

$$\boldsymbol{y}^{(k)} \sim \mathcal{N}(\boldsymbol{Z}^{(k)}\boldsymbol{\omega}^{(k)}, \sigma^2\boldsymbol{I}) \tag{20}$$

ここで $\boldsymbol{I} \in \mathbb{R}^{N_k \times N_k}$ は単位行列である．潜在変数 $\boldsymbol{\nu}_G^{(k)}$ は各 k とグループ G に対して独立に導入され，$\boldsymbol{\omega}^{(k)}$ は $\{\boldsymbol{\nu}_G^{(k)}\}_{G \in \mathcal{G}}$ の線形和で与えられるとする．すなわち，(20) は以下のように置き換わる．

$$\boldsymbol{y}^{(k)} \sim \mathcal{N}\left(\boldsymbol{Z}^{(k)} \sum_{G \in \mathcal{G}} \boldsymbol{\nu}_G^{(k)}, \sigma^2\boldsymbol{I}\right) \tag{21}$$

個別の $\boldsymbol{\nu}_G^{(k)}$ の確率密度関数については，スパースな解を誘導する性質を持った裾の重い分布で，特に G の基数 $|G|$ と重要度 $\mathcal{W}(G)$ について以下の形式で表現できる分布が仮定されている．

$$p(\boldsymbol{\nu}_G^{(k)} | \mathcal{W}(G)) = q_G(\|\boldsymbol{\nu}_G^{(k)}\|_2 \mathcal{W}(G)^{\frac{1}{2}})\mathcal{W}(G)^{\frac{|G|}{2}} \tag{22}$$

なお，q_G は G の構造に対して等方的で，そのスケールは $\mathcal{W}(G)$ の逆数で与えられるような関数である．以上より，$\boldsymbol{\omega}^{(k)}$ の確率密度関数については，(22) を用いて以下のように表現される．

$$p(\boldsymbol{\omega}^{(k)} | \mathcal{W}) = \prod_{G \in \mathcal{G}} p(\boldsymbol{\nu}_G^{(k)} | \mathcal{W}(G)) \tag{23}$$

図 3 は潜在構造正則化学習で導入されている確率モデルのグラフィカルモデルである．この確率モデルにおいて，対数尤度関数の (22)(23) の部分は (2) の正則化項に対応している．すなわち，$\boldsymbol{\omega}^{(k)}$ の事前分布として (22)(23) を仮定し，最大事後確率推定を行うことは，(2) の正則化項を用いて潜在構造正則化学習を行うことに対応している．よって，周辺尤度

$$p(\{\boldsymbol{y}^{(k)}\}_k | \mathcal{W}) = \prod_{k=1}^{K} \int p(\boldsymbol{y}^{(k)} | \boldsymbol{Z}^{(k)}\boldsymbol{\omega}^{(k)}, \sigma^2\boldsymbol{I}) p(\boldsymbol{\omega}^{(k)} | \mathcal{W}) \mathrm{d}\boldsymbol{\omega}^{(k)} \tag{24}$$

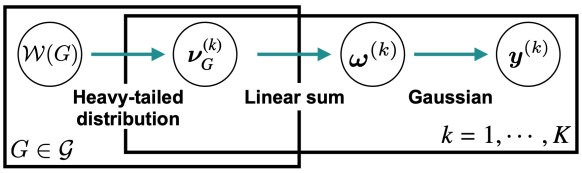

図 3　Latent Structured Graphical LASSO のグラフィカルモデル.

を最大化することで，構造 G に対応する重要度 $\mathcal{W}(G)$ と潜在変数 $\boldsymbol{\nu}_G^{(k)}$ の最適解を決定する．

数値計算では，Palmer らによって提案された非ガウス潜在変数モデルの変分 EM アルゴリズムによって周辺尤度を最大化している [22]．Palmer らによると (22) で登場する $q_G(u)$ について，$\log q_G(u)$ を凸共役な関数 $\phi_G(s)$ で次のように表現できる．

$$\log q_G(u) = \sup_{s \geq 0} \left[-\frac{u^2}{2s} - \phi_G(s) \right] \tag{25}$$

この等式を利用すれば，(22) の右辺は次のようにガウス分布を用いた形式に置き換えることができる．

$$\begin{aligned} & q_G(\|\boldsymbol{\nu}_G^{(k)}\|_2 \mathcal{W}(G)^{\frac{1}{2}}) \\ &= \sup_{\xi_G^{(k)} \geq 0} \mathcal{N}\left(\boldsymbol{\nu}_G^{(k)} \middle| \boldsymbol{0}, \frac{\xi_G^{(k)}}{\mathcal{W}(G)} I\right) \left(\frac{2\pi\xi_G^{(k)}}{\mathcal{W}(G)}\right)^{\frac{|G|}{2}} \mathrm{e}^{-\phi_G(\xi_G^{(k)})} \end{aligned} \tag{26}$$

これにより，(24) の対数周辺尤度は次のように変分下限で評価することができる．

$$\begin{aligned} & \log p(\{\boldsymbol{y}^{(k)}\}_k | \mathcal{W}) \\ &= \sum_{k=1}^{K} \log \int p(\boldsymbol{y}^{(k)} | \boldsymbol{Z}^{(k)} \boldsymbol{\omega}^{(k)}, \sigma^2 \boldsymbol{I}) p(\boldsymbol{\omega}^{(k)} | \mathcal{W}) \mathrm{d}\boldsymbol{\omega}^{(k)} \\ &\geq \sum_{k=1}^{K} \sup_{\xi_G^{(k)} \geq 0} \left\{ \log \mathcal{N}(\boldsymbol{y}^{(k)} | \boldsymbol{0}, \boldsymbol{Z}^{(k)} \boldsymbol{M} \boldsymbol{\Xi}^{(k)} \boldsymbol{M}^\top \boldsymbol{Z}^{(k)^\top} + \sigma^2 \boldsymbol{I}) \right. \\ &\quad \left. + \sum_{G \in \mathcal{G}} \left[\frac{|G|}{2} \log \mathcal{W}(G) + \frac{|G|}{2} \log \left(\frac{2\pi\xi_G^{(k)}}{\mathcal{W}(G)} \right) - \phi_G(\xi_G^{(k)}) \right] \right\} \end{aligned} \tag{27}$$

ここで，$\{\boldsymbol{\nu}_G^{(k)}\}_{G \in \mathcal{G}}$ を適当に並べたベクトルを $\boldsymbol{\nu}^{(k)} \in \mathbb{R}^{\sum_{G \in \mathcal{G}} |G|}$ とすれば，$\boldsymbol{M}$ は $\boldsymbol{\omega}^{(k)} = \boldsymbol{M}\boldsymbol{\nu}^{(k)}$ を満たす行列であり，$\boldsymbol{\Xi}^{(k)}$ は $\boldsymbol{\nu}^{(k)}$ の並べ順で対応する $\xi_G^{(k)}/\mathcal{W}(G)$ を $|G|$ 回繰り返し対角に配置した対角行列である．また，(27) の不等式は上限と積分の交換によるものである．よって，(24) に対して対数周辺尤度を最大化する代わりに，(27) の変分下限の最大化を考えることができ，特に (27) の最大化では関連するパラメータに対して閉じた更新則が得られるので，煩雑な数値計算を回避することができる．

ところで，この確率モデルにおいては $p(\boldsymbol{\nu}_G^{(k)} | \mathcal{W}(G))$ の標準偏差が σ より小さいとき，$\mathcal{W}(G)$ を分離して適切に評価できず，値が過小評価になる傾向が確認されている．このことは，本来重要ではないグループ，すなわち $\mathcal{W}(G) \to \infty$ となるべきグループ G が，過大に重要であると評価されゼロベクトルとなりづらいことを意味する．これを回避するために，Shervashidze らはハイパーパラメータ $\beta \in \mathbb{R}_+$ を導入し，$\mathcal{W}(G)$ にも冪関数の分布 $p(\mathcal{W}(G)) \propto \mathcal{W}(G)^\beta$ を仮定したモデルを考えている [14]．

$$p(\boldsymbol{\omega}^{(k)} | \mathcal{W}) = \prod_{G \in \mathcal{G}} p(\boldsymbol{\nu}_G^{(k)} | \mathcal{W}(G)) p(\mathcal{W}(G)) \tag{28}$$

4　評価実験

評価実験全体を通して，潜在変数 $\boldsymbol{\nu}_G$ の確率密度関数 (22) には，以下のような多変量に拡張した Student's t-分布を用いた．

$$\begin{aligned} p(\boldsymbol{\nu}_G^{(k)} | \mathcal{W}(G), \alpha) = & \frac{\Gamma(\alpha + \frac{|G|}{2})}{\Gamma(\alpha)} \left(\frac{\mathcal{W}(G)}{2\pi} \right)^{\frac{|G|}{2}} \\ & \times \left(1 + \frac{\|\boldsymbol{\nu}_G^{(k)}\|_2^2 \mathcal{W}(G)}{2} \right)^{-\alpha - \frac{|G|}{2}} \end{aligned} \tag{29}$$

ここで，$\alpha \in \mathbb{R}_+$ は分布の形状を決めるハイパーパラメータで，α が小さいほど分布の裾は重くなる．ただし，$\alpha \leq 1$ では (29) の分散を定義することができない．本研究では簡単化のために，全て $\alpha = 1.5$ 及び $K = 100$ で実験を行うこととした．

4.1　トイデータ

まず最初に，簡単なトイデータを用いて提案手法の基本的な性質を確認する．この実験で用いる真の精度行列は，対角成分は常に 1，非対角成分は 50% の確率で 0，50% の確率で $[-0.5, 0.5]$ の範囲の一様分布に従う値として，あらかじめランダムに生成したものを用いた（生成した真の精度行列の値は図 4 中の Ground Truth を参照）．このように生成した精度行列をもち，平均ベクトルが 0 ベクトルとなる多変量正規分布に従ってトイデータを生成し，さらに平均 0，標準偏差 0.5 のガウスノイズを印加して観測データとした．ただし，この実験ではサンプルサイズは $N = 10000$，データの次元は $M = 5$ としてデータを生成し，対称性を考慮した精度行列の非対角成分のパラメータ数は $M(M-1)/2 = 10$ である．ここでは比較手法として，最尤推定（MLE）と Graphical LASSO（GL）の 2 つの手法を採用した．Graphical LASSO については正則化係数を 0 から 1 まで 0.01 刻みで変化させ，Ground Truth に最も近いグラフ構造をもつ正則化係数を採用している．

図 4 はグループ構造 $\mathcal{G} = \{\{1\}, \{2\}, \ldots \{10\}\}$ に対して，提案手法で $\beta = 0.8, 1.0, 1.2$ として推定された精度行列の推定解である．ここでは比較のため，精度行列から計算された偏相関行列を示している．比較手法である Graphical LASSO ではスパースな精度行列が得られているものの，全ての成分に等価なペナルティが課されるモデルのため，濃淡が薄く全体的にぼやけたような精度行列が推定される．これに対して提案手法では，例えば $\beta = 0.8$ の場合に着目すると，Graphical LASSO と同じく 0 に縮退すべき要素は十分縮退している一方，縮退すべきでない要素は明瞭な非ゼロ値を取ることがわかる．さらに $\beta = 1.0$ や $\beta = 1.2$ も確認すると，新たに 0 に縮退する成分が現れたとしても，他の成分の縮小推定度合いは抑制されていることも見て取れる．これは提案手法のモデルにおける重要度 $\mathcal{W}$ がデータから適応的に調整されるためである．

次に $M = 20$ とし，先ほどと同様の手法で生成した精度行列と観測データに対して，推定されたグラフ構造の性能をいくつ

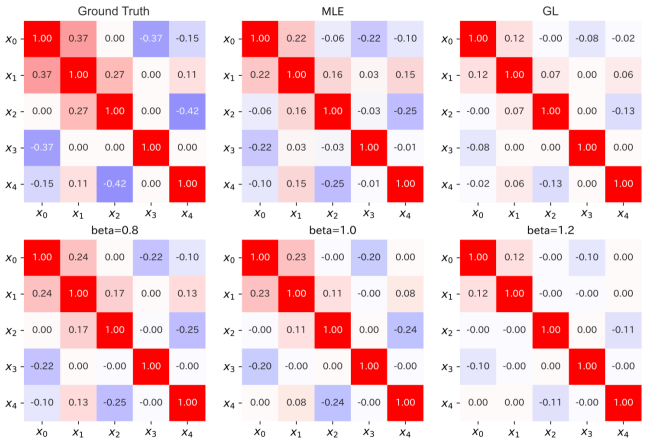

図 4　4.1 節の方法で生成した真の精度行列（Ground Truth）に対する，提案手法と比較手法の精度行列の推定解．ここでは，精度行列から計算された偏相関行列を示している．上段は左から，データ生成に用いた Ground Truth，最尤推定解，Graphical LASSO の推定解，下段は左から，提案手法における $\beta = 0.8$, $\beta = 1.0, \beta = 1.2$ の場合の推定解を示している．

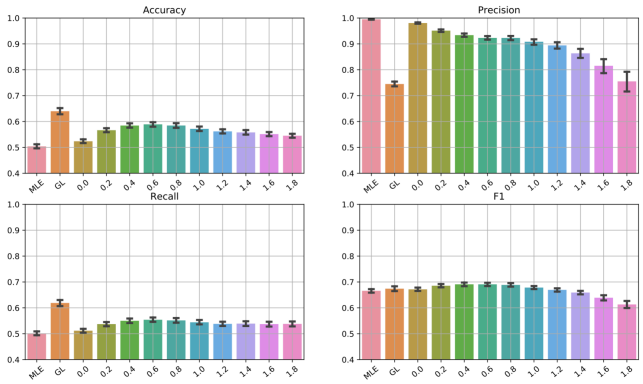

図 5　提案手法と比較手法で抽出されたグラフ構造の Ground Truth に対するスコア．ここでは最尤推定解，Graphical LASSO の推定解，提案手法の $\beta = 0.0, 0.2, \ldots, 1.8$ の場合の推定解について，正解率（Accuracy），適合率（Precision），再現性（Recall），及びそれらの F1 スコア（F1）を示している．

かの指標で比較した．ここでは，推定された精度行列の絶対値をグラフ構造における辺の存在確率とみなし，Ground Truth に応じたグラフ構造との比較を行っている．図 5 は実験結果のスコアで，$\beta = 0.0, 0.2, \ldots, 1.8$ とした提案手法と，比較手法である最尤推定解と Graphical LASSO の推定解について，正解率（Accuracy），適合率（Precision），再現性（Recall），及びそれらの F1 スコア（F1）を示したものである．結果を俯瞰すると，提案手法は Graphical LASSO よりも適合率（Precision）の面でかなり有利なモデルであり，特に $\beta = 0.2, 0.4, 0.6, 0.8, 1.0$ の提案手法については F1 スコアによる総合評価でも高いスコアが読み取れる．これらの実験により，提案手法は従来の Graphical LASSO よりも学習モデルとしての表現力が高く，データから所望の精度行列を柔軟に適応的に抽出できることが確認される．

4.2　Actual Spot Rates

次に実データを用いて，提案手法で入力するグループ構造 $\mathcal{G}$ の変化が推定に与える影響を検討した．ここでは 1986 年 10 月 9 日から 1996 年 9 月 8 日までの 567 日間のドルに対する諸通貨スポット価格データを用いる．ここで用いた通貨は，AUD（オーストラリア），BEF（ベルギー），CAD（カナダ），CHF（スイス），DEM（ドイツ），ESP（スペイン），FRF（フランス），GBP（イギリス），JPY（日本），NLG（オランダ），NZD（ニュージーランド），SEK（スウェーデン）である [20]．

データを平均 0，標準偏差 1 となるように正規化した後に，正則化係数を（a’）$\rho = 0.3$，（a”）$\rho = 0.6$，（a”’）$\rho = 0.9$，とした Graphical LASSO の推定解と，（b’）$\beta = 1.5$ でグループ構造 $\mathcal{G}_b$ を 4.1 節のように 1 つの成分を 1 つのグループとした提案手法の推定解を図 6 に示した．（c）に関しては $\beta = 2.1$ であるが，後述のようにいくつかの成分をひとまとめにしたグループ構造 $\mathcal{G}_c$ をもつ．図 6 の各グラフにおいて，辺の太さは精度行列の絶対値の大きさに対応しており，さらに正値は赤色で負値は青色で表現している．また，頂点の色は Frey らの Affinity Propagation によるクラスタリング結果を意味している [23]．

結果を見てみると 4.1 節の実験と同様，（a’）（a”）（a”’）の Graphical LASSO では辺の存在の有無が曖昧になりがちであるが，（b’）の提案手法では辺の存在する要素では非常に太い辺で示され，より明瞭に依存関係が抽出されていることが確認できる．この（b’）の結果における各グループの重要度 $\mathcal{W}(G)$ の値を，より重要な順，すなわち $\mathcal{W}(G)$ の値が小さい順に上位 20 グループを並べた結果が（b”）である．縦軸は各グループ G を表現し，特にここでは精度行列の各要素をそれぞれひとつのグループとしたグループ構造 $\mathcal{G}_b$ を考えているので，$\mathcal{G}_b$ は全ての 2 通貨の組み合わせに対応している．横軸は重要度 $\mathcal{W}(G)$ を対数スケールでプロットしている．その結果，不要と判断されたグループ G に対応する重要度 $\mathcal{W}(G)$ は，対数スケールの意味で相対的に発散傾向を示していることが確認でき，このことからも $\mathcal{W}(G)$ は最適化の過程で各グループの縮退加減を適応的に調整していることがわかる．提案手法の定式化を考慮すれば，最終的な $\mathcal{W}(G)$ の収束値は各グループの重要度を定量的に評価する有用な指標であると考えることもできる．

ところで，（b’）の頂点の色を見てみると，Affinity Propagation では{BEF, CHF, DEM, FRF, NLG}と{CAD, ESP, GBP, SEK}の間に何らかのクラスターが存在している可能性が示唆されている．そこで（c）では，これを事前知識と考え，グループ構造 $\mathcal{G}_c$ に対して（b’）で使用したグループ構造 $\mathcal{G}_b$ に{BEF, CHF, DEM, FRF, NLG}と{CAD, ESP, GBP, SEK}を追加して次のように与えた．

$$\mathcal{G}_c = \mathcal{G}_b \cup \{\{\text{BEF, CHF, DEM, FRF, NLG}\}, \{\text{CAD, ESP, GBP, SEK}\}\} \tag{30}$$

このように追加したグループについても，提案手法ではデータから不適と判断されれば $\mathcal{W}(G) \to \infty$ 及び $\nu_G \to 0$ となるため，$\mathcal{G}_c$ の中でその重要性が比較検討されることとなる．この実験の結果，（c）では（b’）と同様に辺の存在の有無が明瞭でありつつも，Affinity Propagation のクラスタリング結果とも齟齬のないグ

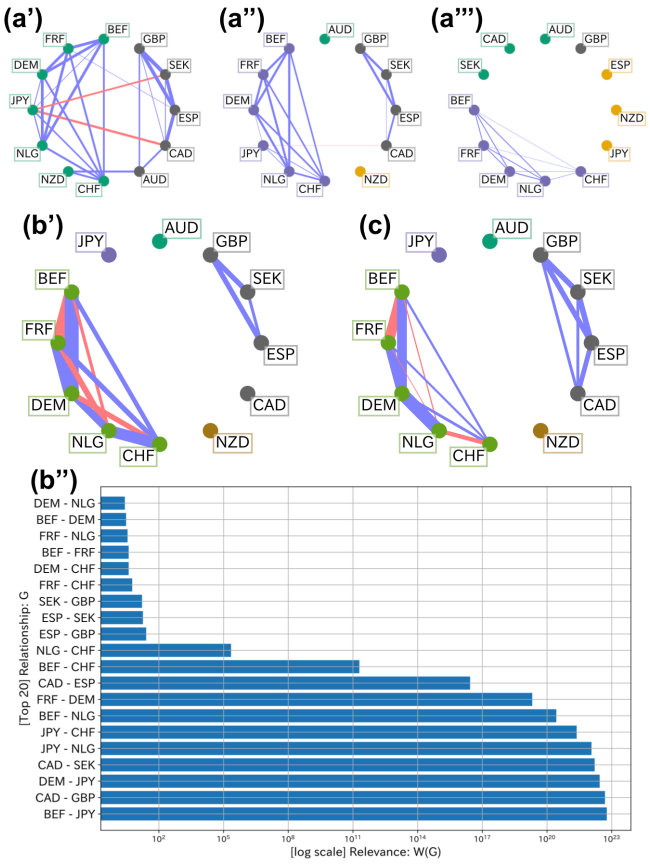

図 6　スポット価格データから推定されたグラフ構造とクラスタリングの結果．上段は正則化係数が（a'）$\rho = 0.3$，（a''）$\rho = 0.6$，（a'''）$\rho = 0.9$ の場合の Graphical LASSO 推定解，中段は（b'）$\beta = 1.5$ で各成分をそれぞれひとつのグループとしたグループ構造 $\mathcal{G}_b$ の場合の提案手法の推定解と，（c）$\beta = 2.1$ でクラスタリング結果を事前情報として追加したグループ構造 $\mathcal{G}_c$ の場合の提案手法の推定解である．（b'）の結果については，重要度 $\mathcal{W}(G)$ から上位 20 グループの結果を（b''）に示している．

ラフが得られていることがわかる．このように提案手法では，入力するグループ構造 $\mathcal{G}$ を通して，事前知識や任意の基準を推定に反映することができ，入力した $\mathcal{G}$ に沿ったスパースな精度行列やグラフを得ることができる．

4.3　Credit Card Fraud Detection

最後に，クレジットカードの不正利用が含まれるデータセットに対して，抽出した精度行列を用いた異常検知タスクを通した性能比較を行った．このデータセットには，ある 2 日間に発生した 284,807 件の取引が含まれており，このうち 492 件の取引は不正であったとわかっている．データは主成分分析によって匿名化されており，含まれる変数は第 28 主成分までの値（V1〜V28）と，各取引の経過時刻（Time），各取引での金額（Amount）である [24]．

ここでは Time を除いた V1〜V28 と Amount の計 29 個の変数を使用することとした．データは 492 件の不正は全てテストデータに含まれるようにした上で，訓練用 235,907 件，テスト用

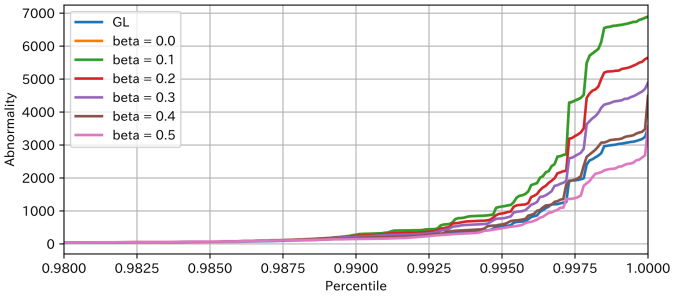

図 7　各手法の異常スコア $\|\boldsymbol{d}\|_2$ に対するパーセンタイル値．尤度交差検証でチューニングした Graphical LASSO（GL）と，$\beta = 0.0, 0.1, \ldots, 0.5$ とした提案手法の結果を示している．

49,200 件にランダムに分割している．従って，今回使用したデータではテストデータに 1% の不正な取引が含まれていることになる．このデータを平均 0 標準偏差 1 に正規化した後，正則化係数を尤度交差検証によりチューニングした Graphical LASSO（GL）と，いくつかの β に対する提案手法に対して，それぞれ推定された精度行列に基づいた異常スコアを比較した．なお，ここでは変数 x_i に対する，各テストサンプルの i 番目の変数についての異常スコア d_i を次のように定義した．

$$d_i \propto -\log p(x'_i | \boldsymbol{x}'_{-i}, \mathcal{D}_{\text{train}}), \tag{31}$$

ただし，訓練データから推定された精度行列 $\boldsymbol{\Lambda}^*$ に対して，テストデータは $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda}^{*-1})$ に従うと仮定し，$\mathcal{D}_{\text{train}}$ は訓練データセット，$\boldsymbol{x}'_{-i}$ はテストサンプル $\boldsymbol{x}'$ から i 番目の変数を除いたベクトルを意味する．また，各テストサンプルごとの総合的な異常スコアは，(31) を並べたベクトルに対して $\|d\|_2$ とした．

図 7 は，49,200 個のテストデータそれぞれから計算された $\|d\|_2$ を小さい順に並べて，横軸のパーセンタイルに沿って縦軸に $\|d\|_2$ をプロットしたものである．テストデータに含まれる不正な取引の割合は 1% であるので，99 パーセンタイルまではどの手法でも $\|d\|_2$ はほぼ横ばいであるが，その後はいずれの手法も劇的に値が増加している．しかし，その伸び方は Graphical LASSO よりも提案手法の方が激しく，提案手法によって推定された精度行列は異常検知タスクにおいてもより明瞭に異常を発見しやすいことが示唆されている．途中まで Graphical LASSO よりも下回っていた $\beta = 0.5$ の提案手法についても，100 パーセンタイル付近では上回る結果となっている．

さらに，表 1 に示すように，いくつかの手法で精度行列のスパース度合いと，テストデータに対する異常スコアの ROC 曲線下面積（AUC）及び適合率-再現性曲線下面積（AUPRC）を比較した．ここでは，比較手法として最尤推定解（MLE），Oracle Approximating Shrinkage（OAS），Basic Shrinkage（BS）を追加している [25, 26]．また，スパース度合いを比較する頑健な指標として Hurley らによって提案されたジニ指数を用いた [27]．具体的には，精度行列 $\boldsymbol{\Lambda}$ に対するスパース度合い $Sp(\boldsymbol{\Lambda})$ は次のように定義している．

$$Sp(\boldsymbol{\Lambda}) = 1 - \frac{1}{2M^2 \|\boldsymbol{\Lambda}\|_1} \sum_{p=1}^{M^2} |r_p| \left(M^2 - p + \frac{1}{2} \right) \tag{32}$$

表 1　各手法から推定された精度行列のスパース度合いと，テストデータに対する AUC 及び AUPRC.

	SPARSITY	AUC	AUPRC
MLE	0.9149	0.9494	**0.7784**
OAS	0.9152	0.9494	**0.7784**
BS	0.9326	0.9502	0.7651
GL	0.9175	0.9495	**0.7784**
$\beta = 0.0$	0.9184	0.9494	0.7782
$\beta = 0.1$	0.9190	0.9494	0.7782
$\beta = 0.2$	**0.9377**	**0.9505**	0.7752

ここで，r_p は精度行列 $\mathbf{\Lambda}$ の各成分を小さい順で並べ替えた（$r_1 \leq r_2 \leq \ldots \leq r_{M^2}$）ものである．従って，定義から $0 \leq S_p \leq 1$ であり，その値が大きいほど精度行列のスパース度合いが高いといえる．結果的に表 1 から，提案手法で抽出した精度行列はスパース度合いが非常に高く，異常検知タスクにおいても高い性能を示すものであることが示唆される．

5　まとめ

本研究では，Graphical LASSO の枠組みに潜在構造正則化学習を応用した新しい正則化手法を提案した．4.2 節にみるように，提案手法では事前知識や任意の基準などをグループ構造という形で反映することができ，入力したグループ構造に沿ったスパースな精度行列の抽出を可能としている．さらに 4.3 節にみるように，提案手法で得られた精度行列を用いて，一般的な異常検知タスクなどにおいても従来法よりも高い性能が得られることも確認された．これらは，従来の Graphical LASSO が全ての要素に対して正則化係数に応じた等価なペナルティを与える一方で，提案手法は個々の変数グループに対してデータから適応的に調整されたペナルティを与えるためである．

今後の展望としては，事前に入力するグループ構造 $\mathcal{G}$ それ自体も，モデルの尤度比較を通して最適なものを決定し，よりデータに応じた学習の枠組みを検討している．また，一般的にデータが単峰性の多変量正規分布に従っているとは言い切れないため，従来の Graphical LASSO でも研究されているように，多峰性の混合正規分布への拡張なども行いたい [28].

参考文献

[1] Tsuyoshi Idé and Hisashi Kashima. Eigenspace-based anomaly detection in computer systems. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’04, page 440–449, New York, NY, USA, 2004. Association for Computing Machinery.
[2] R. Kinderman and S.L. Snell. *Markov random fields and their applications*. American mathematical society, 1980.
[3] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.
[4] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. Adaptive computation and machine learning. MIT Press, 2009.
[5] Steffen L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.
[6] Onureena Banerjee, Laurent El Ghaoui, Alexandre d’Aspremont, and Georges Natsoulis. Convex optimization techniques for fitting sparse gaussian graphical models. In William W. Cohen and Andrew W. Moore, editors, *ICML*, volume 148 of *ACM International Conference Proceeding Series*, pages 89–96. ACM, 2006.
[7] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.
[8] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B)*, 58:267–288, 1996.
[9] Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.
[10] Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.
[11] Laurent Jacob, Guillaume Obozinski, and Jean-Philippe Vert. Group lasso with overlap and graph lasso. In Andrea Pohoreckyj Danyluk, Léon Bottou, and Michael L. Littman, editors, *ICML*, volume 382 of *ACM International Conference Proceeding Series*, pages 433–440. ACM, 2009.
[12] Francis R. Bach, Rodolphe Jenatton, Julien Mairal, and Guillaume Obozinski. Optimization with sparsity-inducing penalties. *Foundations and Trends in Machine Learning*, 4(1):1–106, 2012.
[13] Francis Bach, Rodolphe Jenatton, Julien Mairal, and Guillaume Obozinski. Structured sparsity through convex optimization. *Statistical Science*, 27(4):450–468, 2012.
[14] Nino Shervashidze and Francis R. Bach. Learning the structure for structured sparsity. *IEEE Trans. Signal Processing*, 63(18):4894–4902, 2015.
[15] Benjamin M. Marlin and Kevin P. Murphy. Sparse gaussian graphical models with unknown block structure. In Andrea Pohoreckyj Danyluk, Léon Bottou, and Michael L. Littman, editors, *ICML*, volume 382 of *ACM International Conference Proceeding Series*, pages 705–712. ACM, 2009.
[16] Patrick Danaher, Pei Wang, and Daniela M. Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society Series B*, 76(2):373–397, 2014.
[17] Alex Gibberd and J.D.B. Nelson. Regularized estimation of piecewise constant gaussian graphical models: The group-fused graphical lasso. *Journal of Computational and Graphical Statistics*, 2015.
[18] Shaozhe Tao, Yifan Sun, and Daniel Boley. Inverse covariance estimation with structured groups. In *IJCAI*, pages 2836–2842, 2017.
[19] Veronica Tozzo, Federico Tomasi, Margherita Squillario, and Annalisa Barla. Group induced graphical lasso allows for discovery of molecular pathways-pathways interactions. *arXiv preprint arXiv:1811.09673*, 2018.
[20] Tsuyoshi Idé, Aurelie C. Lozano, Naoki Abe, and Yan Liu. Proximity-based anomaly detection using sparse structure learning. In *SDM*, pages 97–108. SIAM, 2009.
[21] Trevor Park and George Casella. The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.
[22] Jason Palmer, David Wipf, Kenneth Kreutz-Delgado, and Bhaskar Rao. Variational em algorithms for non-gaussian latent variable models. *Advances in neural information processing systems*, 18:1059, 2006.
[23] Brendan J J. Frey and Delbert Dueck. Clustering by passing messages between data points. *Science*, 2007.
[24] Andrea Dal Pozzolo, Olivier Caelen, Reid A. Johnson, and Gianluca Bontempi. Calibrating probability with undersampling for unbalanced classification. In *SSCI*, pages 159–166. IEEE, 2015.
[25] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365 – 411, 2004.
[26] Yilun Chen, Ami Wiesel, Yonina C. Eldar, and Alfred O. Hero III. Shrinkage algorithms for mmse covariance estimation. *IEEE Trans. Signal Processing*, 58(10):5016–5029, 2010.
[27] Niall Hurley and Scott Rickard. Comparing measures of sparsity. *IEEE Transactions on Information Theory*, 55(10):4723–4741, 2009.
[28] Tsuyoshi Idé, Ankush Khandelwal, and Jayant Kalagnanam. Sparse gaussian markov random field mixtures for anomaly detection. In Francesco Bonchi, Josep Domingo-Ferrer, Ricardo Baeza-Yates, Zhi-Hua Zhou, and Xindong Wu, editors, *ICDM*, pages 955–960. IEEE Computer Society, 2016.