

重み付き Lasso パス追跡とグラフィカルモデルへの応用

新村 祐紀† 竹内 一郎† 本谷 秀堅† 澤田 好秀†

†名古屋工業大学

1 はじめに

グラフィカルモデルの構造推定は確率変数間の依存関係を理解し, 新たな知識を発見する上で重要な問題である. 特に代表的なものとして, グラフィカルガウシアンモデル (GGM) の構造推定がさかんに行われ, GGM モデルを用いた画像処理技術などへの応用も進んでいる. GGM の構造推定は精度行列 (分散共分散行列の逆行列) の零成分を推定する問題と等価であることが知られており, スパース正則化を用いた精度行列推定手法がいくつも提案されている. 例えば, Graphical Lasso(glasso) [1] は L_1 正則化付き最尤推定問題を解くアルゴリズムで, 疎な精度行列を効率的に推定することができる.

精度行列の L_1 正則化を用いたアプローチは, グラフィカルモデル全体がスパース (疎) であるという事前知識に基づいている. 事前知識が乏しい場合にはこのようなアプローチが適当であるが, より詳細な事前知識が与えられる場合も多い. 例えば, 3次元形状をグラフィカルモデルで表現する場合, ノード間の物理的な近さを事前知識として利用することができる. また, 一群のノードがクラスタを構成することが分かっている場合, クラスタ内の結合を密にするような事前知識の導入が有効であると考えられる. 本稿では, このような事前知識を導入する枠組を制約付き最適化問題として定式化し, その最適化問題を近似的に解くアルゴリズムを提案する.

2 グラフィカルガウシアンモデル (GGM) と構造推定

p 次元の確率変数の GGM を考える. 確率変数を $\mathbf{x} \in \mathbb{R}^p$ とし, これが多変量正規分布 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Theta}^{-1})$ に従うと仮定する. ここで, $\boldsymbol{\mu} \in \mathbb{R}^p$ は平均ベクトル, $\boldsymbol{\Theta} \in \mathbb{R}^{p \times p}$ は精度行列 (分散共分散行列の逆行列) である. GGM において, ノード x_i と x_j の条件付き独立性と精度行列の成分 Θ_{ij} の間には

$$x_i \perp\!\!\!\perp x_j \mid \text{others} \iff \Theta_{ij} = 0, \quad (1)$$

の関係が成り立つことが知られている. ここで $\perp\!\!\!\perp$ は統計的独立を示す. (1) の関係から GGM の構造を求める

*Adaptive Regularization Path Algorithm for Weighted LASSO and Its Application to Graphical Modeling”

†Yuki Shinmura, †Ichiro Takeuchi, †Hidekata Hontani and †Yoshihide Sawada

†Nagoya Insutitute of Technology

ことと疎な精度行列を求めることは等価であることがわかる.

疎な精度行列を求めるアルゴリズムの代表例として glasso アルゴリズムを簡単に紹介する. glasso アルゴリズムでは, 多変量正規分布の対数尤度関数に L_1 正則化項を加えた

$$\max_{\boldsymbol{\Theta} > 0} \log \det \boldsymbol{\Theta} - \text{tr}(\mathbf{S}\boldsymbol{\Theta}) - \rho \sum_{i < j} |\Theta_{ij}|, \quad (2)$$

を最大化する問題を解く. ここで, $\boldsymbol{\Theta} > 0$ は精度行列が正定値行列であることを表している. また, \mathbf{S} は標本分散共分散行列, $\rho \geq 0$ は正則化パラメータである. 目的関数の第 1 項と第 2 項は多変量正規分布の対数尤度関数であり, 第 3 項は $\boldsymbol{\Theta}$ をスパースにする正則化項となっている. 正則化パラメータ ρ を変えると GGM 全体の疎度 (sparsity) を制御することができる. glasso では, 正則化パラメータ ρ を細かく変化させたときの問題 (2) の最適解のパス (正則化パスと呼ばれる) を計算し, さまざまな疎度における精度行列の推定値を得ることができる.

3 Group-wise Sparsity

本研究ではグラフ構造に関する事前知識としてノードのグループの疎度に関する情報が与えられている場合を考察する. 複数のノードから成る G 個のグループを考え, グループの番号を添字 $g \in \{1, \dots, G\}$ で表すこととする. グループ g に含まれるノード番号の集合を $\mathcal{G}_g \subset \{1, \dots, p\}$ と表記し, そのサイズを $|\mathcal{G}_g|$ と表す. また, グループ g 内のエッジの集合を $\mathcal{E}_g := \{(i, j) \mid i, j \in \mathcal{G}_g, i < j\}$ とし, そのサイズを $|\mathcal{E}_g|$ と表す. 以上の表記を用いると, グループ g の疎度は

$$\sigma_g := \frac{\sum_{(i,j) \in \mathcal{E}_g} I\{\Theta_{ij} = 0\}}{|\mathcal{E}_g|} \quad (3)$$

と表される. ここで, $I\{\cdot\}$ は定義関数を表している.

さまざまな応用問題において, ノードのグループの疎度に関する事前知識を導入すると有用である. 例えば, 3次元形状モデリングにグラフィカルモデルを用いる場合, “バランスのよい” 構造を推定することが要求される. ここで, “バランスの悪い” 構造とは, 例えば, 3次元形状の一部だけが密に連結し, 他の部分にほとんど連結がないようなものである. このような場合, 近隣のノード

ドの集合をグループとみなし、各グループの疎度がほぼ一定であるというような事前知識を与えれば、“バランスのよい”構造の推定が可能となる。

グループの疎度に関する事前知識を制約条件として定式化すると、精度行列の推定問題は、

$$\max_{\Theta > 0} \log \det \Theta - \text{tr}(S\Theta) \quad (4a)$$

$$\text{s.t.} \quad \sigma_g = su_g, \quad g = 1, \dots, G \quad (4b)$$

と定式化される。ここで、 s はグラフ全体の疎度を、 u_g はグループ g の相対的な疎度を表すパラメータで、一意性を保つため、 $\sum_{g=1}^G u_g = G$ とする (すなわち、 $u_g = 1$ であれば、グループ g は平均的な疎度を持つ)。制約条件 (4b) はグループ g の疎度 σ_g が su_g であることを課している。

例えば、3次元形状モデリングの例のように局所的なノードのグループを考え、 $u_1 = \dots = u_G = 1$ として最適化問題 (4) を解けば、“バランスのよい”グラフィカルモデルを推定することができる。

残念ながら、最適化問題 (4) は離散非凸最適化問題であり最適解を得ることは難しい。また、グループに重複がある場合、すべてのグループに対する制約条件を満たすような解 (実行可能解) が存在するとは限らない。そのため、次節では、制約条件 (4b) を近似的に満たすようなアルゴリズムを提案する。

4 適応的正則化パス追跡アルゴリズム

本稿では、*glasso* と同様に、正則化パス追跡アルゴリズムを考える。以下で導入する正則化パス追跡アルゴリズムの目的は、最適化問題 (4) の疎度パラメータ s を細かく変化させたときの最適解のパスを追跡することである。前述のように厳密な最適解を得ることは困難であるため、近似的な解のパスを考えることとする。提案アルゴリズムを導出するため、最適化問題:

$$\max_{\Theta > 0} \log \det \Theta - \text{tr}(S\Theta) - \sum_{i < j} \rho_{ij} |\Theta_{ij}| \quad (5)$$

を考える。問題 (5) は、*glasso* の問題 (2) の一般化となっていることに留意されたい。(2) では共通の正則化パラメータ ρ が用いられるのに対し、(5) では精度行列の各 (i, j) 要素に個別の正則化パラメータ ρ_{ij} が用いられる。 $\{\rho_{ij}\}_{i < j}$ を固定すると、最適化問題 (5) は *glasso* とほぼ同様のアルゴリズムで解くことができる。

glasso では $\rho = \infty$ から開始し、 ρ を徐々に減らしたときの最適解パスを計算していく。提案アルゴリズムでも同様に、 $\rho_{ij} = \infty, \forall (i, j)$ から開始し、 ρ_{ij} を徐々に変化させていく。しかし、すべての ρ_{ij} を均一に減らすのではなく、(4b) のグループ疎度に関する制約を満たす方向に ρ_{ij} を適応的に変化させる。おおまかには、グループ g の

疎度 σ_g が su_g より疎である場合には、 $(i, j) \in \mathcal{E}_g$ に対応する正則化パラメータ ρ_{ij} を減少し、 σ_g が su_g より密である場合には、 $(i, j) \in \mathcal{E}_g$ に対応する正則化パラメータ ρ_{ij} を増加させる。提案アルゴリズムでは、各グループの疎度の変化を検出しながら、 $\binom{p}{2}$ 個の正則化パラメータ $\{\rho_{ij}\}_{i < j}$ の更新方向を適応的に決めていくものとなっている。以下の提案アルゴリズムの骨子を示す:

Algorithm 1 適応的正則化パスアルゴリズム

Input 共分散行列 S , グループ $\{\mathcal{G}_g\}_{g=1}^G$, グループの相対的疎度 $\{u_g\}_{g=1}^G$, ステップ幅 $\Delta s, \Delta \rho$

output 各グループの (近似的な) 疎度が $u_g, (1 - \Delta s)u_g, (1 - 2\Delta s)u_g, \dots, 0, g = 1, \dots, G$, であるような精度行列のパス

Initialize $s \leftarrow 1, \rho_{ij} \leftarrow \infty, \forall (i, j)$

while $s \geq 0$ **do**

repeat

 // 各グループの疎度 σ_g を計算

 Compute $\{\sigma_g\}_{g=1}^G$ in (3)

 // 正則化パラメータ行列の更新

$\rho_{ij} \leftarrow \rho_{ij} - \Delta \rho \frac{I((i,j) \in \mathcal{E}_g)(\sigma_g - su_g)}{\sum_{i < j} I((i,j) \in \mathcal{E}_g)}$

 // 最適化問題 (5) を解く

 Solve (5) using current $\{\rho_{ij}\}_{i < j}$

until グループ疎度の誤差: $\sum_{g=1}^G (\sigma_g - su_g)^2$ が変化しなくなるまで

$s \leftarrow s - \Delta s$;

end while

5 数値実験

紙面の都合上、数値実験結果については割愛するが、発表にて、3次元画像形状モデリングタスクへの応用例を示す。

参考文献

- [1] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso", *Biostatistics*, **9-3**, pp. 421–441 (2008).
- [2] S. Allasonniere, P. Jolivet, and C. Giraud, "Detecting Long Distance Conditional Correlations Between Anatomical Regions Using Gaussian Graphical Models", *Mathematical Foundation of Computational Anatomy (MFCA) workshop of the MICCAI 2011 conference* (2011)