日本オペレーションズ・リサーチ学会
2021年 春季研究発表会

Network Lasso における クラスター構造復元の十分条件と非凸拡張

05001372　中央大学　*柳下 翔太郎　YAGISHITA Shotaro
01110520　中央大学　後藤 順哉　GOTOH Jun-ya

1.　はじめに

データ解析において，データのクラスター構造と重心や回帰モデルなどを同時に推定したいという状況がしばしば生じる．近年，Hallac ら [1] が提案した Network Lasso はそれを実現する，スパースモデリングのテクニックを利用した方法である．

$\mathcal{V} = [n] := \{1, \ldots, n\}$ を標本の添字の集合，$\mathcal{E} \subset \{\{i,j\} : i,j \in [n]; i \neq j\}$ を標本間の隣接の有無を表す辺集合，$W = (w_{\{i,j\}})_{\{i,j\}\in\mathcal{E}} \in \mathbb{R}^{|\mathcal{E}|}_{\geq 0}$ を隣接した標本間の類似度を表す重みとし，重み付き無向グラフ $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ を考える．x_i を標本 $i \in \mathcal{V}$ における重心や回帰係数などを表すパラメータ，$f_i : \mathbb{R}^p \to \mathbb{R}$ を標本 i に対する損失関数とし，Network Lasso は以下のように定式化される．

$$\underset{x_1,\ldots,x_n\in\mathbb{R}^p}{\text{minimize}} \quad \sum_{i\in\mathcal{V}} f_i(x_i) + \gamma \sum_{\{i,j\}\in\mathcal{E}} w_{\{i,j\}} \|x_i - x_j\|_2.$$

ただし，$\gamma > 0$ は定数，$\|\cdot\|_2$ は ℓ_2 ノルムである．直感的には，第 1 項は標本ごとの当てはまりを強め，第 2 項は隣接した標本のパラメータ x_i, x_j を重み $w_{\{i,j\}}$ が大きいほど近づける効果を持つ．Group Lasso のスパース性から，最適解 $(x_1^*, ..., x_n^*)$ において多くの $\{i,j\} \in \mathcal{E}$ に対して $x_i^* = x_j^*$ となることを期待し，そうなったとき，i と j が同じクラスターに属していると判定する．たとえば，$a_i \in \mathbb{R}^p$ をデータ点，$f_i(x_i) = \|x_i - a_i\|_2^2$ とすれば，重心とクラスター構造の同時推定，$b_i \in \mathbb{R}$ を目的変数，$a_i \in \mathbb{R}^p$ を説明変数，$f_i(x_i) = (b_i - a_i^\top x_i)^2$ とすれば，回帰モデルとクラスター構造の同時推定に応用できる．$\gamma \to \infty$ とし正則化パスを考えることで，階層クラスタリングにおけるデンドログラムのようなクラスターパスを生成できる（図 1 左）．

本発表ではまず，Network Lasso が潜在的な真のクラスターを復元するための十分条件を与える．次に，Network Lasso の持つ弱点を克服するような定式化を提案し，それに対して交互方向乗数法（ADMM）を適用する．最後に，その定式化の有効性を数値実験により検討する．内容の詳細については [2] を参照されたい．

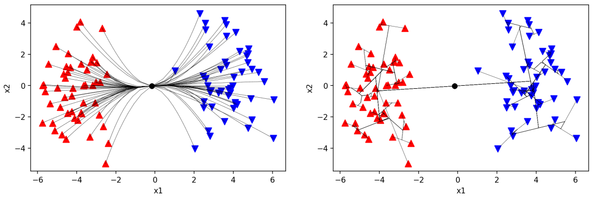

図 1: $\mathbb{R}^2$ 上での Network Lasso によるクラスタリングの失敗例（左）と提案する Network Trimmed Lasso による成功例（右）．三角の点がデータ点 a_i，細い曲線が x_i^* の軌跡，すなわちクラスターパスを表している．

2.　Network Trimmed Lasso

Network Lasso は γ や $f_i, w_{\{i,j\}}$ が適当な条件を満たせば，真のクラスターを復元することを示すことができる．紙面の都合上，詳細については当日報告する．

Network Lasso では，上述の十分条件が成立しない，特に，重み $w_{\{i,j\}}$ が「うまく」与えられていないとき，図 1 左のように，単純な数値例においてさえ厳密に $x_i^* = x_j^*$ とならず，真のクラスターどころかクラスターを得ることさえできない．これを解決するために，以下のような基数制約による定式化を考える．

$$\begin{aligned} &\underset{x_1,\ldots,x_n}{\text{minimize}} \quad \sum_{i\in\mathcal{V}} f_i(x_i) \\ &\text{subject to} \quad \Big|\{\{i,j\} \in \mathcal{E} : \|x_i - x_j\|_2 > 0\}\Big| \leq K. \end{aligned} \tag{1}$$

ただし，K は $K \leq |\mathcal{E}|$ であるような非負整数である．$x_i^* = x_j^*$ となるような制約を直接組み込むことで，クラスターが形成される．さらに，K の減少につれて，$x_i^* = x_j^*$ となる辺が増え，クラスターパスの生成が期待できる．しかし，(1) の基数制約の左辺は $(x_1, ..., x_n)$ の不連続関数であり，一般に，問題 (1) は大域的な解を得ることが難しい．このような事情から，まず基数制約を連続関数による等価な制約に書き換える．$\boldsymbol{z} = (z_{\{i,j\}})_{\{i,j\}\in\mathcal{E}} \in \mathbb{R}^{p|\mathcal{E}|}$ とし，$\|z_{\{i,j\}}\|_2$ の小さいほうから $|\mathcal{E}| - K$ 個の要素の和を

$$T_K(\boldsymbol{z}) = \|z_{(K+1)}\|_2 + \cdots + \|z_{(|\mathcal{E}|)}\|_2$$

とおく．ここで，$\|z_{(i)}\|_2$ は $(\|z_1\|_2, ..., \|z_{|\mathcal{E}|}\|_2) \in \mathbb{R}^{|\mathcal{E}|}$ の大きいほうから i 番目の要素を表す．$\boldsymbol{x} := (x_i)_{i\in\mathcal{V}}$ とし，D を $z_{\{i,j\}} = x_i - x_j$ となるような $p|\mathcal{E}| \times pn$ 行列とすれば，(1) は以下の最適化問題と等価である．

$$\begin{aligned} \underset{x_1,\ldots,x_n}{\text{minimize}} \quad & \sum_{i\in\mathcal{V}} f_i(x_i) \\ \text{subject to} \quad & T_K(D\boldsymbol{x}) = 0. \end{aligned} \tag{2}$$

$T_K(D\boldsymbol{x})$ は非凸であるが，$\boldsymbol{x}$ の連続関数である．さらに，(2) に対して，$\gamma > 0$ として，以下の無制約問題を考える．

$$\underset{x_1,\ldots,x_n}{\text{minimize}} \quad \sum_{i\in\mathcal{V}} f_i(x_i) + \gamma T_K(D\boldsymbol{x}). \tag{3}$$

$T_K(D\boldsymbol{x}) \geq 0$ であり，$T_K(D\boldsymbol{x}) > 0$ と $\left|\left\{\{i,j\} \in \mathcal{E} : \|x_i - x_j\|_2 > 0\right\}\right| > K$ が等価であるので，(3) の第二項は基数制約に対するペナルティ関数の役割を果たす．本稿では，(3) を Network Trimmed Lasso と呼ぶ．次に示すように，適当な仮定の下で γ を十分大きくとれば，Network Trimmed Lasso は (2) と等価であることが示される．

定理 1. *1.* f_i が L_i-平滑であるとし，$\boldsymbol{x}^\gamma := (x_1^\gamma, ..., x_n^\gamma)$ を (3) の最適解とする．ある $C > 0$ が存在し，任意の $i \in \mathcal{V}, \gamma > 0$ に対して $\|x_i^\gamma\|_2 \leq C$ であることを仮定する．このとき，

$$\gamma > \sum_{i\in\mathcal{V}} (\|\nabla f_i(0)\|_2 + 2L_i C)$$

ならば，$\boldsymbol{x}^\gamma$ は (2) の最適解である．

2. f_i に L_i-平滑性に加えて凸性を仮定すれば，*1.* の「最適解」を「局所最適解」に置き換えた主張が成立する．

この定理に動機を得て，基数制約付き問題 (1) の代わりに (3) を解くことを考える．さらに，Network Trimmed Lasso においては，緩い条件の下で方向停留点と局所最適解が一致することを示すことができる．

命題 1. $\sum_{i\in\mathcal{V}} f_i(x_i)$ が可微分凸であるとする．このとき，(3) の方向停留点は局所最適解である．

3. アルゴリズム

最適化問題 (3) を解くために，ADMM の適用を考える．まず，(3) を以下の等価な線型等式制約付きの問題に書き換える．

$$\begin{aligned} \underset{\boldsymbol{x},\boldsymbol{z}}{\text{minimize}} \quad & \sum_{i\in\mathcal{V}} f_i(x_i) + \gamma T_K(\boldsymbol{z}) \\ \text{subject to} \quad & \boldsymbol{z} = D\boldsymbol{x}. \end{aligned} \tag{4}$$

双対変数を $\boldsymbol{y} \in \mathbb{R}^{p|\mathcal{E}|}$，$\rho > 0$ を定数として，(4) の拡張ラグランジュ関数を，

$$\begin{aligned} L_\rho(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{y}) = & \sum_{i\in\mathcal{V}} f_i(x_i) + \gamma T_K(\boldsymbol{z}) \\ & + \boldsymbol{y}^\top(\boldsymbol{z} - D\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{z} - D\boldsymbol{x}\|_2^2 \end{aligned}$$

で定義する．ADMM は各反復 t で

$$\boldsymbol{z}^{t+1} \in \underset{\boldsymbol{z}}{\text{argmin}}\, L_\rho(\boldsymbol{x}^t, \boldsymbol{z}, \boldsymbol{y}^t), \tag{5}$$

$$\boldsymbol{x}^{t+1} \in \underset{\boldsymbol{x}}{\text{argmin}}\, L_\rho(\boldsymbol{x}, \boldsymbol{z}^{t+1}, \boldsymbol{y}^t), \tag{6}$$

$$\boldsymbol{y}^{t+1} = \boldsymbol{y}^t + \rho(\boldsymbol{z}^{t+1} - D\boldsymbol{x}^{t+1})$$

のように点列を生成する反復法である．(5) は非凸であるが解析解を導出できる．(6) が解析解を持たない場合には，$L > 0$ として，(6) を以下に置き換える．

$$\begin{aligned} \boldsymbol{x}^{t+1} \in \underset{\boldsymbol{x}}{\text{argmin}} \Big\{ & \nabla f(\boldsymbol{x}^t)^\top(\boldsymbol{x} - \boldsymbol{x}^t) + \frac{L}{2}\|\boldsymbol{x} - \boldsymbol{x}^t\|_2^2 \\ & + (\boldsymbol{y}^t)^\top(\boldsymbol{z}^{t+1} - D\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{z}^{t+1} - D\boldsymbol{x}\|_2^2 \Big\}. \end{aligned}$$

ただし，$f(\boldsymbol{x}) = \sum_{i\in\mathcal{V}} f_i(x_i)$ である．この修正を施したものは，近接 ADMM [3] と呼ばれる．これらのアルゴリズムの収束性に関して，以下が成り立つ．

定理 2. $f_i \in C^1, (i \in \mathcal{V})$ を仮定する．*ADMM*（もしくは，近接 *ADMM*）によって生成された点列 $(\boldsymbol{x}^t, \boldsymbol{z}^t, \boldsymbol{y}^t)$ が $(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{y}^*)$ に収束し，$D\boldsymbol{x}^*$ が *non-ambiguous* であるとする．このとき，$\boldsymbol{x}^*$ は (3) の方向停留点である．

ここで，$\boldsymbol{z} = (z_e)_{e\in\mathcal{E}} \in \mathbb{R}^{p|\mathcal{E}|}$ が non-ambiguous であるとは，$\|z_{(K+1)}\|_2 < \|z_{(K)}\|_2$ が成り立つことをいう．これらのことから，f_i が L_i-平滑かつ凸であり，ADMM（もしくは，近接 ADMM）によって生成された点列が収束し，$D\boldsymbol{x}^*$ が non-ambiguous であれば，基数制約付き問題 (1) の局所最適解を得ることができる．

図 1 右は Trimmed Network Lasso の出力例である．その他の数値実験結果については，当日報告する．

参考文献

[1] D. Hallac, J. Leskovec, and S. Boyd, Network lasso: Clustering and optimization in large graphs, KDD, (2015), 387–396.

[2] S. Yagishita, J. Gotoh, Pursuit of the Cluster Structure of Network Lasso: Recovery Condition and Non-convex Extension, arXiv preprint 2012.07491, (2020).

[3] G. Li and T. K. Pong, Global convergence of splitting methods for nonconvex composite optimization, SIAM J. Optim., 25 (2015), 2434–2460.