平成３０年度 共同利用研究報告書

平成 ３１ 年 ３ 月 ３０ 日

九州大学 マス・フォア・インダストリ研究所長 殿

所属・職名　統計数理研究所・教授

提案者 氏名　二宮 嘉行

下記の通り共同研究の報告をいたします.

記

		※整理番号	２０１８００１３	
1.研究計画題目	混合効果モデルにおけるスパース推定のための条件付き AIC の開発			
2.種目（○で囲む）	a. プロジェクト研究　b. 若手研究　ⓒ. 一般研究			
3.種別（○で囲む）	a. 研究集会Ⅰ　b. 研究集会Ⅱ　c. 短期共同研究　ⓓ. 短期研究員			
4.研究代表者	氏 名	二宮 嘉行		
	所 属 部局名	統計数理研究所 数理・推論研究系	職 名	教授
	連絡先			
	e-mail		TEL	
5.研究実施期間	平成 31 年 1 月 23 日（水曜日）～平成 31 年 1 月 25 日（金曜日） 平成 31 年 3 月 7 日（木曜日）～平成 31 年 3 月 9 日（土曜日）			
6.キーワード（複数可）	混合効果モデル，条件付き AIC，スパース推定，LASSO			
7.参加者数	３人　＊1			

＊1　短期研究員は九大の共同研究者も含める.
Ⅰ，Ⅱ，短期共同研究は事務局から送った参加者データを元に記入.

8.本研究で得られた成果の概要（**成果報告書を別途要添付**　枚数は次頁参照）

混合効果モデル分析は統計学における基本的な手法であるが，比較的最近になってモデル選択のための情報量規準が新たに開発された（Vaida and Blanchard, 2005, Biometrika）. これを条件付き AIC という. 一方で，これは最尤推定法を用いたときの基準であり，いまや標準的な推定法となっている LASSO を代表とするスパース推定法には対応していなかった. そこで，LASSO の自由度を導出するときの基本となる Stein の方法が今の設定でも使えることを確認し，LASSO を用いたときの条件付き AIC を導出した. 数値実験では，古典的な情報量規準 mAIC もスパース推定のもとで導き，比較をおこない，条件付き AIC の優位性を確認した.

1 混合効果モデル

m 個のクラスターからなるデータとして $\boldsymbol{y}$ を考える.
Laird and Ware (1982) において, 混合効果モデルは以下のように定義された.

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \qquad (i = 1, \cdots, m) \tag{1}$$

このとき $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{G})$ はそれぞれのクラスター番号 i に関して互いに独立であり, $\boldsymbol{y}_i$ はクラスター i に対応した n_i 次元ベクトルとする. また, $\boldsymbol{\beta}$(固定効果) は p 次元ベクトル, $\boldsymbol{b}_i$(クラスター i に対するランダム効果) は q 次元ベクトルとし, $\boldsymbol{X}_i(n_i \times p$ 行列), $\boldsymbol{Z}_i(n_i \times q$ 行列) はそれぞれ $\boldsymbol{\beta}, \boldsymbol{b}_i$ に対応する計画行列 (フルランク) とする. 加えて, $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{n_i})$ は $\boldsymbol{b}_i$ と互いに独立で, $\boldsymbol{G}$ は $q \times q$ 正定値行列とする.

ここで $N = \sum_{i=1}^{m} n_i$ とするとき, (1) は Vaida and Blanchard (2005) において次式のように書き直せる.

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{b} \sim N(\boldsymbol{0}, \boldsymbol{G}_0) \tag{2}$$

このとき, それぞれの記号は, $\boldsymbol{X} = (\boldsymbol{X}_1^{\mathrm{T}}, \cdots, \boldsymbol{X}_m^{\mathrm{T}})^{\mathrm{T}}$ はランクが p の $N \times p$ 行列, $\boldsymbol{Z} = \mathrm{diag}(\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_m)$ はランクが $r = mq$ の $N \times r$ ブロック対角行列, $\boldsymbol{b} = (\boldsymbol{b}_1^{\mathrm{T}}, \cdots, \boldsymbol{b}_m^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^{\mathrm{T}}, \cdots, \boldsymbol{\varepsilon}_m^{\mathrm{T}})^{\mathrm{T}}$, そして $\boldsymbol{G}_0 = \mathrm{diag}_m(\boldsymbol{G})$ は m 個の $\boldsymbol{G}$ が対角に並んだブロック対角行列となっている.

また分散共分散行列に関するパラメータである $\sigma^2, \boldsymbol{G}_0$ は既知であるとし, 以下では (2) の表現を基に議論を進めていく.

2 LASSO に対する cAIC

本節では, 混合効果モデル, 特に条件付きモデルに対する LASSO の自由度及び LASSO に対する cAIC について導出を行う. 以下, $\boldsymbol{y}$ に関しては中心化, $\boldsymbol{X}, \boldsymbol{Z}_i$ $(i = 1, \cdots, m)$ はそれぞれ正規化の処理がなされているものとする.

LASSO 推定を考えるが, 通常の回帰モデルの時とは異なり $\boldsymbol{\beta}$ の推定だけでなく, $\boldsymbol{b}$ の予測も伴う. このことによってスパース推定の対象パラメータが $\boldsymbol{\beta}$ と $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{b}^{\mathrm{T}})^{\mathrm{T}}$ の 2 つのパターンが存在することがわかる. 本節ではこの内 $\boldsymbol{\beta}$ に関してスパース推定を行うものを考える.

この時の LASSO 推定量は Hodges and Sargent (2001) の方式に $\boldsymbol{\beta}$ の L1 罰則項を加えたもの, つまり

$$\hat{\boldsymbol{\theta}}_\lambda^{(\mathrm{c})}(\boldsymbol{y}) = (\hat{\boldsymbol{\beta}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})^{\mathrm{T}}, \hat{\boldsymbol{b}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})^{\mathrm{T}})^{\mathrm{T}} = \mathop{\mathrm{argmin}}_{\boldsymbol{\beta}, \boldsymbol{b}} \left\{ \left\| \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{pmatrix} - \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Z} \\ \boldsymbol{O} & -\sigma\boldsymbol{D} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{b} \end{pmatrix} \right\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \right\} \tag{3}$$

となる. ここで $\boldsymbol{M} = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Z} \\ \boldsymbol{O} & -\sigma\boldsymbol{D} \end{pmatrix}$ とする.

以下の条件を満たす $\lambda^{(\mathrm{c})}$ の有限部分列 $\{\lambda_i^{(\mathrm{c})}\}_{i=0,\cdots,K'}$ を考える：

$$\lambda_0^{(\mathrm{c})} > \lambda_1^{(\mathrm{c})} > \lambda_2^{(\mathrm{c})} > \cdots > \lambda_{K'}^{(\mathrm{c})} = 0 \tag{4}$$

- $\forall \lambda > \lambda_0^{(\mathrm{c})},\ \hat{\boldsymbol{\beta}}_\lambda^{(\mathrm{c})}(\boldsymbol{y}) = \boldsymbol{0}$
- 区間 $(\lambda_{k+1}^{(\mathrm{c})}, \lambda_k^{(\mathrm{c})})$ の内部において, アクティブセット $\mathcal{B}^{(\mathrm{c})}(\lambda)$ と符号関数 $\mathrm{Sgn}^{(\mathrm{c})}(\lambda)_{\mathcal{B}(\lambda)}$ が λ に関してコンスタントである. この為, 表記簡略のためにそれぞれ $\mathcal{B}_k^{(\mathrm{c})},\ \mathrm{Sgn}_k^{(\mathrm{c})}$ とする.

ここで, $\tilde{\mathcal{B}}_k^{(\mathrm{c})} = \mathcal{B}_k^{(\mathrm{c})} \cup \{p+1, \cdots, p+r\},\ \widetilde{\mathrm{Sgn}}_k^{(\mathrm{c})} = (\mathrm{Sgn}_k^{(\mathrm{c})\mathrm{T}}, \boldsymbol{0}^\mathrm{T})^\mathrm{T}$ としておく.

以下, いくつかの補題を与える.

補題 1. $\lambda \in (\lambda_{k+1}^{(\mathrm{c})}, \lambda_k^{(\mathrm{c})})$ とする時, $\hat{\boldsymbol{\theta}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})$ は次のようになる.

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})_{\tilde{\mathcal{B}}_k^{(\mathrm{c})}} = (\hat{\boldsymbol{\beta}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})_{\mathcal{B}_k^{(\mathrm{c})}}^\mathrm{T}, \hat{\boldsymbol{b}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})^\mathrm{T})^\mathrm{T} &= \left(\boldsymbol{M}_{\tilde{\mathcal{B}}_k^{(\mathrm{c})}}^\mathrm{T} \boldsymbol{M}_{\tilde{\mathcal{B}}_k^{(\mathrm{c})}}\right)^{-1} \left(\boldsymbol{M}_{\tilde{\mathcal{B}}_k^{(\mathrm{c})}}^\mathrm{T} \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{pmatrix} - \frac{\lambda}{2} \widetilde{\mathrm{Sgn}}_k^{(\mathrm{c})}\right) \\
&= \left(\boldsymbol{M}_{\tilde{\mathcal{B}}_k^{(\mathrm{c})}}^\mathrm{T} \boldsymbol{M}_{\tilde{\mathcal{B}}_k^{(\mathrm{c})}}\right)^{-1} \left(\begin{pmatrix} \boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}}^\mathrm{T} \\ \boldsymbol{Z}^\mathrm{T} \end{pmatrix} \boldsymbol{y} - \frac{\lambda}{2} \widetilde{\mathrm{Sgn}}_k^{(\mathrm{c})}\right)
\end{aligned} \tag{5}$$

これを解くと,

$$\hat{\boldsymbol{\beta}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})_{\mathcal{B}_k^{(\mathrm{c})}} = (\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}}^\mathrm{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}})^{-1} \left(\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}}^\mathrm{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{y} - \frac{\lambda}{2} \mathrm{Sgn}_k^{(\mathrm{c})}\right)$$

$$\hat{\boldsymbol{b}}_\lambda^{(\mathrm{c})}(\boldsymbol{y}) = \frac{1}{\sigma^2} \boldsymbol{G}_0 \boldsymbol{Z}^\mathrm{T} \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}} \hat{\boldsymbol{\beta}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})_{\mathcal{B}_k^{(\mathrm{c})}})$$

となる.

補題 2. 遷移点 $\lambda_k^{(\mathrm{c})}, \lambda_{k+1}^{(\mathrm{c})}\ (\geq 0)$ に対して, $\mathcal{B}_k^{(\mathrm{c})}$ は $(\lambda_{k+1}^{(\mathrm{c})}, \lambda_k^{(\mathrm{c})})$ 内でのアクティブセットとする. i_{add} を $\lambda_k^{(\mathrm{c})}$ において $\mathcal{B}_k^{(\mathrm{c})}$ に入る添え字とし, また $\mathcal{B}_k^{(\mathrm{c})}$ における添え字を i つまり $i_{\mathrm{add}} = (\mathcal{B}_k^{(\mathrm{c})})_i$ とする. ここで $(\boldsymbol{a})_l$ はベクトル $\boldsymbol{a}$ の第 l 番目成分である. この時遷移点 $\lambda_k^{(\mathrm{c})}$ は以下のように表現できる.

$$\lambda_k^{(\mathrm{c})} = \frac{2\left((\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}}^\mathrm{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}})^{-1} \boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}}^\mathrm{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}\right)_i}{\left((\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}}^\mathrm{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}})^{-1} \mathrm{Sgn}_k^{(\mathrm{c})}\right)_i} \tag{6}$$

さらに $\lambda_{k+1}^{(\mathrm{c})}$ において $\mathcal{B}_k^{(\mathrm{c})}$ から脱落する添え字番号が存在したとしてそれを j_{drop} とし, $j_{\mathrm{drop}} = (\mathcal{B}_k^{(\mathrm{c})})_j$ とする. この時 $\lambda_{k+1}^{(\mathrm{c})}$ は以下のように書ける.

$$\lambda_{k+1}^{(\mathrm{c})} = \frac{2\left((\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}}^\mathrm{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}})^{-1} \boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}}^\mathrm{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}\right)_j}{\left((\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}}^\mathrm{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}})^{-1} \mathrm{Sgn}_k^{(\mathrm{c})}\right)_j} \tag{7}$$

補題 3. $\forall\lambda > 0$, 零集合 $\mathcal{N}_\lambda^{(\mathrm{c})}$ が存在していて $\mathbb{R}^N$ 上超平面の有限集合であるとする. この時 $\mathcal{G}_\lambda^{(\mathrm{c})} = \mathbb{R}^N \setminus \mathcal{N}_\lambda^{(\mathrm{c})}$ とすると $\forall \boldsymbol{y} \in \mathcal{G}_\lambda^{(\mathrm{c})}$, λ は非遷移点つまり $\lambda \notin \left\{\lambda(\boldsymbol{y})_k^{(\mathrm{c})}\right\}$

補題 4. $\forall\lambda$, $\hat{\boldsymbol{\theta}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})$ は $\boldsymbol{y}$ に関して連続関数である.

補題 5. $\lambda > 0$ を固定したもとで, $\mathcal{G}_\lambda^{(\mathrm{c})}$ を補題 3において定義したものとし $\boldsymbol{y} \in \mathcal{G}_\lambda^{(\mathrm{c})}$ とする. この時アクティブセット $\mathcal{B}_\lambda^{(\mathrm{c})}$, 符号関数 $\mathrm{Sgn}_\lambda^{(\mathrm{c})}$ は $\boldsymbol{y}$ に関して局所コンスタントである.

補題 6. $\mathcal{G}_0^{(\mathrm{c})} = \mathbb{R}^N$ とする. 任意の $\lambda\ (\geq 0)$ を固定したもとで, 補題 3で定義された *(*全測度が *1* となる*)*$\mathcal{G}_\lambda^{(\mathrm{c})}$ 上において $\hat{\boldsymbol{\mu}}_\lambda^{(\mathrm{c})}(\boldsymbol{y}) = \boldsymbol{X}\hat{\boldsymbol{\beta}}_\lambda^{(\mathrm{c})}(\boldsymbol{y}) + \boldsymbol{Z}\hat{\boldsymbol{b}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})$ は一様リプシッツである. つまり十分小さな $\Delta\boldsymbol{y}$ に対して

$$\|\hat{\boldsymbol{\mu}}_\lambda^{(\mathrm{c})}(\boldsymbol{y}+\Delta\boldsymbol{y}) - \hat{\boldsymbol{\mu}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})\| \leq \|\Delta\boldsymbol{y}\| \tag{8}$$

が成立し, さらに次のことが言える.

$$\nabla\cdot\hat{\boldsymbol{\mu}}_\lambda^{(\mathrm{c})}(\boldsymbol{y}) = \mathrm{tr}\left\{\begin{pmatrix} \boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}} & \boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}}^{\mathrm{T}}\boldsymbol{Z} \\ \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}} & \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z} + \sigma^2\boldsymbol{G}_0^{-1}\end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}} & \boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}}^{\mathrm{T}}\boldsymbol{Z} \\ \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{c})}} & \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\end{pmatrix}\right\} \tag{9}$$

以上, 補題 1から補題 6によって次の定理が導かれる.

定理 1. $\forall\lambda$, $\hat{\boldsymbol{\mu}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})$ は $\boldsymbol{y}$ に関して一様リプシッツ関数. $\boldsymbol{M}$ がフルランク $(\mathrm{rank}(\boldsymbol{M}) = p+r)$ であるならば $\hat{\boldsymbol{\mu}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})$ の自由度は

$$df(\lambda) = \mathrm{E}[\nabla\cdot\hat{\boldsymbol{\mu}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})] \tag{10}$$

となる.

この時, $\mathrm{BC} = \mathrm{E}_{f(\boldsymbol{y}|\boldsymbol{u})}[(\boldsymbol{y}-\boldsymbol{\mu})^{\mathrm{T}}\hat{\boldsymbol{\mu}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})]/\sigma^2 = \nabla\cdot\hat{\boldsymbol{\mu}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})$ となり, したがって, LASSO に対する cAIC は,

$$\mathrm{cAIC}^{\mathrm{LASSO}} = -2\log g^{(\mathrm{c})}(\boldsymbol{y} \mid \hat{\boldsymbol{\beta}}_\lambda^{(\mathrm{c})}(\boldsymbol{y}), \hat{\boldsymbol{b}}_\lambda^{(\mathrm{c})}(\boldsymbol{y})) + 2\rho^*$$

ただし,

$$\rho^* = \mathrm{tr}\left\{\begin{pmatrix} \boldsymbol{X}_{\mathcal{B}_\lambda^{(\mathrm{c})}}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{B}_\lambda^{(\mathrm{c})}} & \boldsymbol{X}_{\mathcal{B}_\lambda^{(\mathrm{c})}}^{\mathrm{T}}\boldsymbol{Z} \\ \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{B}_\lambda^{(\mathrm{c})}} & \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z} + \sigma^2\boldsymbol{G}_0^{-1}\end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{X}_{\mathcal{B}_\lambda^{(\mathrm{c})}}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{B}_\lambda^{(\mathrm{c})}} & \boldsymbol{X}_{\mathcal{B}_\lambda^{(\mathrm{c})}}^{\mathrm{T}}\boldsymbol{Z} \\ \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{B}_\lambda^{(\mathrm{c})}} & \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\end{pmatrix}\right\}$$

となる.

この ρ^* は次のように計算できる.

$$\rho^* = |\mathcal{B}_\lambda^{(\mathrm{c})}| + \sum_{i=1}^{r_0^*}\frac{\lambda_i^*}{1+\lambda_i^*} \tag{11}$$

ここで, $\lambda_1^*, \cdots, \lambda_{r_0^*}^*$ は
$(\sigma^{-2}\boldsymbol{G}_0)^{1/2}\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{X}_{\mathcal{B}_\lambda^{(\mathrm{c})}}(\boldsymbol{X}_{\mathcal{B}_\lambda^{(\mathrm{c})}}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{B}_\lambda^{(\mathrm{c})}})^{-1}\boldsymbol{X}_{\mathcal{B}_\lambda^{(\mathrm{c})}}^{\mathrm{T}})\boldsymbol{Z}(\sigma^{-2}\boldsymbol{G}_0)^{1/2}$ の非ゼロ固有値である.

3　LASSOに対するmAIC

本節では, 混合効果モデル, 特に周辺モデルに対するLASSOの自由度及びLASSOに対するmAICについて導出を行う. 前節と同様, $\boldsymbol{y}$ に関しては中心化, $\boldsymbol{X}, \boldsymbol{Z}_i$ $(i = 1, \cdots, m)$ はそれぞれ正規化の処理がなされているものとする.

混合効果モデルの周辺尤度に対し, LASSO推定を考えると,

$$\hat{\boldsymbol{\beta}}_{\lambda}^{(\mathrm{m})}(\boldsymbol{y}) = \operatorname*{argmin}_{\boldsymbol{\beta}} \left\{ \frac{1}{\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_1 \right\} \tag{12}$$

となる.

以下の条件を満たす $\lambda^{(\mathrm{m})}$ の有限部分列 $\{\lambda_i^{(\mathrm{m})}\}_{i=0,\cdots,K''}$ を考える

$$\lambda_0^{(\mathrm{m})} > \lambda_1^{(\mathrm{m})} > \lambda_2^{(\mathrm{m})} > \cdots > \lambda_{K''}^{(\mathrm{m})} = 0 \tag{13}$$

- $\forall\lambda > \lambda_0^{(\mathrm{m})},\ \hat{\boldsymbol{\beta}}_{\lambda}^{(\mathrm{m})}(\boldsymbol{y}) = \boldsymbol{0}$
- 区間 $(\lambda_{k+1}^{(\mathrm{m})}, \lambda_k^{(\mathrm{m})})$ の内部において, アクティブセット $\mathcal{B}^{(\mathrm{m})}(\lambda)$ と符号関数 $\mathrm{Sgn}^{(\mathrm{m})}(\lambda)_{\mathcal{B}^{(\mathrm{m})}(\lambda)}$ が λ に関してコンスタントである. この為, 表記簡略のためにそれぞれ $\mathcal{B}_k^{(\mathrm{m})},\ \mathrm{Sgn}_k^{(\mathrm{m})}$ とする.

以下, いくつかの補題を与える.

補題 7. $\lambda \in (\lambda_{k+1}^{(\mathrm{m})}, \lambda_k^{(\mathrm{m})})$ とする時, $\hat{\boldsymbol{\beta}}_{\lambda}^{(\mathrm{m})}(\boldsymbol{y})$ は次のようになる.

$$\hat{\boldsymbol{\beta}}_{\lambda}^{(\mathrm{m})}(\boldsymbol{y})_{\mathcal{B}_k^{(\mathrm{m})}} = \left(\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{m})}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{m})}}\right)^{-1}\left(\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{m})}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{y} - \frac{\lambda\cdot\sigma^2}{2}\,\mathrm{Sgn}_k^{(\mathrm{m})}\right) \tag{14}$$

補題 8. 遷移点 $\lambda_k^{(\mathrm{m})}, \lambda_{k+1}^{(\mathrm{m})}$ (≥ 0) に対して, $\mathcal{B}_k^{(\mathrm{m})}$ は $(\lambda_{k+1}^{(\mathrm{m})}, \lambda_k^{(\mathrm{m})})$ 内でのアクティブセットとする. i_{add} を $\lambda_k^{(\mathrm{m})}$ において $\mathcal{B}_k^{(\mathrm{m})}$ に入る添え字とし, また $\mathcal{B}_k^{(\mathrm{m})}$ における添え字を i^* つまり $i_{\mathrm{add}} = (\mathcal{B}_k^{(\mathrm{m})})_{i^*}$ とする. この時, 遷移点 $\lambda_k^{(\mathrm{m})}$ は以下のように表現できる.

$$\lambda_k^{(\mathrm{m})} = \frac{2\left((\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{m})}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{m})}})^{-1}\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{m})}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\right)_{i^*}}{\sigma^2\left((\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{m})}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{m})}})^{-1}\mathrm{Sgn}_k^{(\mathrm{m})}\right)_{i^*}} \tag{15}$$

さらに $\lambda_{k+1}^{(\mathrm{m})}$ において $\mathcal{B}_k^{(\mathrm{m})}$ から脱落する添え字番号が存在したとしてそれを j_{drop} とし, $j_{\mathrm{drop}} = (\mathcal{B}_k^{(\mathrm{m})})_{j^*}$ とする. この時 $\lambda_{k+1}^{(\mathrm{m})}$ は以下のように書ける.

$$\lambda_{k+1}^{(\mathrm{m})} = \frac{2\left((\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{m})}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{m})}})^{-1}\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{m})}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\right)_{j^*}}{\sigma^2\left((\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{m})}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_{\mathcal{B}_k^{(\mathrm{m})}})^{-1}\mathrm{Sgn}_k^{(\mathrm{m})}\right)_{j^*}} \tag{16}$$

補題 9. $\forall\lambda > 0$, 零集合 $\mathcal{N}_\lambda^{(\mathrm{m})}$ が存在していて $\mathbb{R}^N$ 上超平面の有限集合であるとする. この時 $\mathcal{G}_\lambda^{(\mathrm{m})} = \mathbb{R}^N \setminus \mathcal{N}_\lambda^{(\mathrm{m})}$ とすると $\forall \boldsymbol{y} \in \mathcal{G}_\lambda^{(\mathrm{m})}$, λ は非遷移点つまり $\lambda \notin \left\{ \lambda(\boldsymbol{y})_k^{(\mathrm{m})} \right\}$

補題 10. $\forall\lambda$, $\hat{\boldsymbol{\beta}}_\lambda^{(\mathrm{m})}(\boldsymbol{y})$ は $\boldsymbol{y}$ に関して連続関数である.

補題 11. $\lambda > 0$ を固定したもとで, $\mathcal{G}_\lambda^{(\mathrm{m})}$ を補題 9において定義したものとし $\boldsymbol{y} \in \mathcal{G}_\lambda^{(\mathrm{m})}$ とする. この時アクティブセット $\mathcal{B}_\lambda^{(\mathrm{m})}$, 符号関数 $\mathrm{Sgn}_\lambda^{(\mathrm{m})}$ は $\boldsymbol{y}$ に関して局所コンスタントである.

補題 12. $\boldsymbol{D}^{*\mathrm{T}}\boldsymbol{D}^* = \boldsymbol{\Sigma}^{-1}$ なる $\boldsymbol{D}^*$ に対し, $\boldsymbol{y}^* = \boldsymbol{D}^*\boldsymbol{y}$, $\boldsymbol{X}^* = \boldsymbol{D}^*\boldsymbol{X}$ つまり $\boldsymbol{y}^* \sim N(\boldsymbol{X}^*\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$ とする.

この時, $\boldsymbol{y}^*, \boldsymbol{X}^*, \{\lambda^{(\mathrm{m})}\}, \mathcal{B}_\lambda^{(\mathrm{m})}$ に関して補題 7から補題 11が成立し, 補題 9で定義された集合に対応するものを $\mathcal{G}_\lambda^*$ $(\mathcal{G}_0^* = \mathbb{R}^N)$ とすると, 任意の λ (≥ 0) を固定したもとで $\boldsymbol{y}^* \in \mathcal{G}_\lambda^*$ 上において $\hat{\boldsymbol{\mu}}_\lambda^*(\boldsymbol{y}^*) = \boldsymbol{X}^*\hat{\boldsymbol{\beta}}_\lambda^*(\boldsymbol{y}^*)$ は一様リプシッツである. つまり十分小さな $\Delta\boldsymbol{y}^*$ に対して

$$\|\hat{\boldsymbol{\mu}}_\lambda^*(\boldsymbol{y}^* + \Delta\boldsymbol{y}^*) - \hat{\boldsymbol{\mu}}_\lambda^*(\boldsymbol{y}^*)\| \leq \|\Delta\boldsymbol{y}^*\| \tag{17}$$

が成立し, さらに次のことが言える.

$$\nabla \cdot \hat{\boldsymbol{\mu}}_\lambda^*(\boldsymbol{y}^*) = |\mathcal{B}_\lambda^{(\mathrm{m})}| \tag{18}$$

以上, 補題 7から補題 12によって次の定理が導かれる.

定理 2. $\forall\lambda$, $\hat{\boldsymbol{\mu}}_\lambda^*(\boldsymbol{y}^*)$ は $\boldsymbol{y}^*$ に関して一様リプシッツ関数. $\boldsymbol{X}$ がフルランク $(\mathrm{rank}(\boldsymbol{X}) = p)$ であるならば $\hat{\boldsymbol{\mu}}_\lambda^*(\boldsymbol{y}^*)$ の自由度は

$$df(\lambda) = \mathrm{E}|\mathcal{B}_\lambda^{(\mathrm{m})}| \tag{19}$$

となる.

したがって, LASSO に対する mAIC は,

$$\mathrm{mAIC}^{\mathrm{LASSO}} = -2\log g^{(\mathrm{m})}(\boldsymbol{y} \mid \hat{\boldsymbol{\beta}}_\lambda^{(\mathrm{m})}(\boldsymbol{y})) + 2|\mathcal{B}_\lambda^{(\mathrm{m})}|$$

となる.