

Minimization Algorithm of Model Selection Criterion for Optimizing Tuning Parameter in Lasso Estimator When Explanatory Variables are Orthogonal *

広島大学・理学研究科・数学専攻[†] 大石峰暉・柳原宏和

Mineaki Ohishi and Hirokazu Yanagihara

Department of Mathematics, Graduate School of Science

Hiroshima University

§1. 序

本論文では, 説明変数が直交する場合における線形回帰モデルを扱う. $\mathbf{y} = (y_1, \dots, y_n)'$ を n 次元目的変数ベクトル, $\mathbf{X}$ を $n \times k$ 説明変数行列とすると, 線形回帰モデルは以下のように表される.

$$\mathbf{y} = \mu \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (1)$$

ここで, μ は位置母数, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ は k 次元回帰係数ベクトル, $\mathbf{1}_n$ は全ての成分が 1 である n 次元ベクトル, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)'$ は n 次元誤差ベクトルで各成分は独立に平均 0, 分散 σ^2 を持つ分布に従うとする. 本論文では説明変数が直交する場合を扱うため, $\mathbf{X}$ には以下を仮定する.

$$\mathbf{X}'\mathbf{X} = \mathbf{D} = \text{diag}(d_1, \dots, d_k), \quad d_1 \geq \dots \geq d_k > 0.$$

さらに, $\mathbf{X}$ は中心化 ($\mathbf{X}'\mathbf{1}_n = \mathbf{0}_k$) された行列で $\text{rank}(\mathbf{X}) = k < n - 1$ とする. ここで, $\mathbf{0}_k$ は全ての成分が 0 である k 次元ベクトルである.

(1) 式の未知パラメータ μ, β の推定法の 1 つに最小二乗 (Least Square; LS) 推定法がある。LS 推定法は、以下の残差平方和 (Residual Sum of Squares; RSS) の最小化により未知パラメータ μ, β を推定する方法である。

$$(\mathbf{y} - \mu \mathbf{1}_n - \mathbf{X}\beta)'(\mathbf{y} - \mu \mathbf{1}_n - \mathbf{X}\beta). \quad (2)$$

LS 推定法によって得られた推定量である LS 推定量は、

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}, \quad (3)$$

と簡単に求めることができ、理論上良い性質をいくつか持つため、広く用いられている。しかしながら、 $\mathbf{X}$ に多重共線性がある場合、LS 推定法を用いることが好ましくない場合がある。そのような問題は、LS 推定量の代わりに、(2) 式の RSS に罰則項を加えた罰則付き残差平方和 (Penalized RSS; PRSS) を最小にする、罰則付き LS (Penalized LS; PLS) 推定量を用いることで改善される。PLS 推定量の一つに、Hoerl & Kennard (1970) によって提案された一般化リッジ回帰 (Generalized Ridge Regression; GRR) 推定量がある。GRR 推定量は、以下のような GRR 型 ℓ_2 ペナルティを RSS に加えた PRSS を最小とする β である。

$$(\mathbf{y} - \mu \mathbf{1}_n - \mathbf{X}\beta)'(\mathbf{y} - \mu \mathbf{1}_n - \mathbf{X}\beta) + \sum_{j=1}^k \theta_j \beta_j^2. \quad (4)$$

ここで、 $\theta_j \in \mathbb{R}_+ = \{\theta \in \mathbb{R} \mid \theta \geq 0\}$ ($j = 1, \dots, k$) はリッジパラメータと呼ばれるハイパーパラメータである。リッジパラメータにより GRR 推定値が変化するため、リッジパラメータの最適化問題が生じる。リッジパラメータの最適化には、モデル選択規準 (Model Selection Criterion; MSC) を最小にするリッジパラメータを最適なものとする MSC 最小化法が頻繁に用いられる。GRR での MSC は複数のリッジパラメータの多変数関数となるため、同時最小化が必要となり時間がかかると予想されるが、Ohishi *et al.* (2017) により、最小化する関数を 1 変数に落とし込むことで、リッジパラメータ最適化を高速に行うアルゴリズムが提案されている。

GRR 推定量以外の PLS 推定量として、Tibshirani (1996) により提案された Lasso 推定量がある。Lasso 推定量は、GRR 型 ℓ_2 ペナルティの代わりに Lasso 型 ℓ_1 ペナルティを (2) 式の RSS に加えた以下のような PRSS を最小にする β である。

$$\text{PRSS}_\theta(\mu, \beta) = \frac{1}{2}(\mathbf{y} - \mu \mathbf{1}_n - \mathbf{X}\beta)'(\mathbf{y} - \mu \mathbf{1}_n - \mathbf{X}\beta) + \theta \sum_{j=1}^k |\beta_j|. \quad (5)$$

ここで, $\theta \in \mathbb{R}_+$ はチューニングパラメータと呼ばれるハイパーパラメータである. 通常, (5) 式の最小化問題は陽に解くことができず, 推定値を求めるために Coordinate Descent (例えば, Friedman *et al.*, 2010), Least Angle Regression (LARS; Efron *et al.*, 2004) や Alternating Direction Method of Multipliers (ADMM; Boyd *et al.*, 2011) などの数値探索アルゴリズムを用いる必要がある. また, GRR と同様に, チューニングパラメータにより Lasso 推定値が変化するため, チューニングパラメータの最適化問題が生じる. MSC 最小化法によりチューニングパラメータを最適化する場合, MSC 最小化問題の解を陽に解くことができないため, 繰り返し計算を伴う数値探索法を用いなくてはならない.

本論文では, MSC 最小化法に基づく Lasso 推定量のチューニングパラメータ最適化法の高速化を考える. 本論文のように説明変数が直交する場合, (5) 式の最小化問題は陽に解くことができる. その解を用いることで MSC を最小にする候補点を有限個に限定することができ, それらの候補点から最小を達成する解を探すことにより最適化法の高速化が可能となる.

本論文は以下のように構成される. §2 では, チューニングパラメータ選択のための MSC を定義し, 最小化問題を解くための MSC の区分関数を定義する. §3 では, MSC を最小にする最適なチューニングパラメータを求める高速なアルゴリズムを提案する. §4 では, 最適化されたチューニングパラメータを用いたときの Lasso 推定量の性能を数値実験により確かめる. 数学的な証明は付録に記載した.

§2. チューニングパラメータ選択のためのモデル選択規準

本論文では説明変数が直交することを仮定している. そのため, 特異値分解により説明変数行列は,

$$X = P \begin{pmatrix} D^{1/2} \\ O_{n-k,k} \end{pmatrix} = P_1 D^{1/2}, \quad (6)$$

と表すことができる. ここで, P は n 次直交行列, P_1 は P の分割 $P = (P_1, P_2)$ から得られる $n \times k$ 行列で $P_1' P_1 = I_k$, $P_1' \mathbf{1}_n = \mathbf{0}_k$ を満たし, $O_{n,k}$ はすべての成分が 0 である $n \times k$ 行列である. また k 次元ベクトル $\mathbf{z}_1$ と k 次対角行列 L_θ をそれぞれ以下で定義する.

$$\mathbf{z}_1 = (z_1, \dots, z_k)' = P_1' \mathbf{y}, \quad L_\theta = \text{diag}(\ell_{\theta,1}, \dots, \ell_{\theta,k}). \quad (7)$$

ただし, L_θ の第 j 対角成分は,

$$\ell_{\theta,j} = \begin{cases} \frac{-\theta \operatorname{sign}(z_j) + \sqrt{d_j} z_j}{\sqrt{d_j^3} z_j} & (\theta < \sqrt{d_j} |z_j|) \\ 0 & (\theta \geq \sqrt{d_j} |z_j|) \end{cases}, \quad (8)$$

である. ここでは, $z_j \neq 0$ ($j = 1, \dots, k$) を仮定する. 簡単な計算から, (5) 式を最小にする μ は, (3) 式での $\hat{\mu}$ であることがわかる. そのため, Lasso 推定量は,

$$\hat{\beta}_\theta = \arg \min_{\beta \in \mathbb{R}^k} \text{PRSS}_\theta(\hat{\mu}, \beta),$$

となる. 本論文のように説明変数が直交する場合, 上記の最小化問題は陽に解くことができ, その陽解は以下の定理のように与えられる (証明は付録 A.1 参照).

定理 1 Lasso 推定量は,

$$\hat{\beta}_\theta = (\hat{\beta}_{\theta,1}, \dots, \hat{\beta}_{\theta,k})' = L_\theta X' y = L_\theta D^{1/2} z_1, \quad (9)$$

となり, その第 j 成分は,

$$\hat{\beta}_{\theta,j} = \begin{cases} \frac{-\theta \operatorname{sign}(z_j) + \sqrt{d_j} z_j}{d_j} & (\theta < \sqrt{d_j} |z_j|) \\ 0 & (\theta \geq \sqrt{d_j} |z_j|) \end{cases}, \quad (10)$$

で与えられる.

(6), (7) 式を用いると, (3) 式で与えられた β の LS 推定量は

$$\hat{\beta} = (X' X)^{-1} X' y = D^{-1/2} z_1 = \begin{pmatrix} z_1 / \sqrt{d_1} \\ \vdots \\ z_k / \sqrt{d_k} \end{pmatrix},$$

と書き換えることができ, これは $\theta = 0$ のときの Lasso 推定量に一致する. 定理 1 から y の予測値は,

$$\hat{y}_\theta = \hat{\mu} 1_n + X \hat{\beta}_\theta = (J_n + X L_\theta X') y,$$

で与えられる. ここで, J_n は $J_n = 1_n 1_n' / n$ で与えられる n 次射影行列である.

チューニングパラメータ選択のための MSC は多くの場合, 以下で定義される σ^2 の推定量と一般化自由度 (degree of freedom; df) の 2 変数関数により表すことができる.

$$\hat{\sigma}^2(\theta) = \frac{1}{n} (y - y_\theta)' (y - y_\theta) = \frac{1}{n} y' (I_n - J_n - X L_\theta X')^2 y, \quad (11)$$

$$\text{df}(\theta) = \text{tr}(\mathbf{J}_n + \mathbf{X}\mathbf{L}_\theta\mathbf{X}') = 1 + \text{tr}(\mathbf{L}_\theta\mathbf{D}). \quad (12)$$

また, $\hat{\sigma}^2(0) = \hat{\sigma}_0^2$ とする. すなわち,

$$\hat{\sigma}_0^2 = \frac{1}{n} \mathbf{y}'(\mathbf{I}_n - \mathbf{J}_n - \mathbf{X}\mathbf{D}^{-1}\mathbf{X}')\mathbf{y}. \quad (13)$$

これは正規性の下での σ^2 の最尤推定量と一致する. $k < n - 1$ のとき, ほとんどの場合で $\hat{\sigma}_0^2 \neq 0$ を満たすため, 本論文では $\hat{\sigma}_0^2 \neq 0$ を仮定する. このとき, $\hat{\sigma}^2(\theta)$, $\text{df}(\theta)$ に関して以下の補題が成り立つ (証明は付録 A.2 参照).

補題 1 $\hat{\sigma}^2(\theta)$, $\text{df}(\theta)$ は以下のように表すことができる.

$$\begin{aligned} \hat{\sigma}^2(\theta) &= \frac{1}{n} \left[n\hat{\sigma}_0^2 + \sum_{j=1}^k \left\{ \mathbf{I}(\theta < \sqrt{d_j}|z_j|) \left(\frac{\theta}{\sqrt{d_j}|z_j|} - 1 \right) + 1 \right\}^2 z_j^2 \right], \\ \text{df}(\theta) &= 1 + k - \sum_{j=1}^k \left\{ \mathbf{I}(\theta < \sqrt{d_j}|z_j|) \left(\frac{\theta}{\sqrt{d_j}|z_j|} - 1 \right) + 1 \right\}. \end{aligned}$$

ここで, $\mathbf{I}(\cdot)$ は定義関数である.

補題 1 を用いると, 以下の補題を得ることができる (証明は付録 A.3 参照).

補題 2 $\hat{\sigma}^2(\theta)$, $\text{df}(\theta)$ の定義域はそれぞれ,

$$\hat{\sigma}^2(\theta) \in [\hat{\sigma}_0^2, \hat{\sigma}_\infty^2], \quad \text{df}(\theta) \in [1, k + 1],$$

である. ここで, $\hat{\sigma}_\infty^2 = \lim_{\theta \rightarrow \infty} \hat{\sigma}^2(\theta)$ である. すなわち,

$$\hat{\sigma}_\infty^2 = \frac{1}{n} \mathbf{y}'(\mathbf{I}_n - \mathbf{J}_n)\mathbf{y}.$$

チューニングパラメータ選択のための MSC での 2 変数関数 $f(r, u)$ は, 以下の 3 条件 (Ohishi *et al.*, 2017) を満たすものを用いる.

(A1) $f(r, u)$ は $(0, \hat{\sigma}_\infty^2] \times [1, n)$ 上の連続関数.

(A2) $(r, u) \in (0, \hat{\sigma}_\infty^2] \times [1, n)$ 上で $f(r, u) > 0$.

(A3) $f(r, u)$ は $(0, \hat{\sigma}_\infty^2] \times [1, n)$ 上で 1 階偏微分可能で, $(0, \hat{\sigma}_\infty^2] \times [1, n)$ 上で,

$$\frac{\partial}{\partial r} f(r, u) > 0, \quad \frac{\partial}{\partial u} f(r, u) > 0.$$

この $f(u, v)$ を用いて、チューニングパラメータ選択のための MSC を以下のように定義する.

$$\text{MSC}(\theta) = f(\hat{\sigma}^2(\theta), \text{df}(\theta)). \quad (14)$$

一般によく知られる規準である, Generalized C_p (GC_p ; Atkinson, 1980), Extended Generalized Cross Validation (EGCV; Ohishi *et al.*, 2017), Generalized Information Criterion (GIC; Nishii, 1984) を表現するには以下の $f(r, u)$ を用いればよい.

$$f(r, u) = \begin{cases} nr/s_0^2 + \alpha u & (GC_p) \\ r/(1 - u/n)^\alpha & (\text{EGCV} : u \neq n, \alpha \geq 2) \\ r \exp(\alpha u/n) & (\text{GIC}) \end{cases} \quad (15)$$

ここで, α はモデルの複雑さに対する罰則の強さを表す正の値であり, s_0^2 は σ^2 の不偏推定量, すなわち,

$$s_0^2 = \frac{n}{n-k-1} \hat{\sigma}_0^2, \quad (16)$$

である. $k < n-1$, $\hat{\sigma}_0^2 \neq 0$ を満たすことから $s_0^2 \neq 0$ である. これらの MSC は α を決めることで他のよく知られた規準と一致する. 例えば GC_p において, $\alpha = 2$ のときは C_p (Maows, 1973, 1995), $\alpha = 1 + 2/(n-k-1)$ のときは MC_p (Fujikoshi & Satoh, 1997; Yanagihara *et al.*, 2009) に一致する. EGCV は Ohishi *et al.* (2017) によって提案された規準で, モデルの複雑さに対する罰則を変えることができるように Generalized Cross Validation (GCV; Craven & Wahba, 1979) を $\alpha \geq 2$ で拡張したものであり, $\alpha = 2$ のとき通常の GCV に一致する. 本論文では, GIC は正規性の下での通常の GIC を n で割って指数変換したものをを用いており, $\alpha = 2$ で AIC (Akaike, 1973), $\alpha = 2 \log \log n$ で HQC (Hannan & Quinn, 1978), $\alpha = \log n$ で BIC (Schwarz, 1978) に一致する.

本論文での目標は, 以上のように定義された MSC を最小とするような θ を求める高速アルゴリズムを提案することにある. この目標を達成するために, Ohishi *et al.* (2017) のアイデアを用いて, 定義域を分割することにより MSC を区分関数として個別最小化を考える. t_j ($j = 1, \dots, k$) を $\sqrt{d_1}|z_1|, \dots, \sqrt{d_k}|z_k|$ の第 j 順序統計量, すなわち,

$$t_j = \begin{cases} \min\{\sqrt{d_1}|z_1|, \dots, \sqrt{d_k}|z_k|\} & (j = 1) \\ \min\{\{\sqrt{d_1}|z_1|, \dots, \sqrt{d_k}|z_k|\} \setminus \{t_1, \dots, t_{k-1}\}\} & (j = 2, \dots, k) \end{cases}, \quad (17)$$

とし, これらを用いて区間 R_j ($j = 0, \dots, k$) を,

$$R_j = \begin{cases} (0, t_1] & (j = 0) \\ (t_j, t_{j+1}] & (j = 1, \dots, k-1) \\ (t_k, \infty] & (j = k) \end{cases}, \quad (18)$$

で定義する. また, 添え字 τ_j ($j = 1, \dots, k$) を,

$$\tau_j \in \{1, \dots, k\} \text{ s.t. } t_j = \sqrt{d_{\tau_j}} |z_{\tau_j}|,$$

で定義し, $q_j = d_{\tau_j}$ とする. $\theta \in R_a$ のとき, 補題 1 の $\hat{\sigma}^2(\theta)$ と $\text{df}(\theta)$ はそれぞれ以下のよう置き換えることができる.

$$\hat{\sigma}^2(\theta) = \hat{\sigma}_a^2(\theta) = \hat{\sigma}_0^2 + \frac{1}{n}(c_{1,a} + c_{2,a}\theta^2), \quad (19)$$

$$\text{df}(\theta) = \text{df}_a(\theta) = 1 + k - a - c_{3,a}\theta. \quad (20)$$

ここで, $c_{1,a}$, $c_{2,a}$, $c_{3,a}$ はそれぞれ以下のように与えられる正の定数である.

$$c_{1,a} = \sum_{j=1}^a \frac{t_j^2}{q_j}, \quad c_{2,a} = \sum_{j=a+1}^k \frac{1}{q_j}, \quad c_{3,a} = \sum_{j=a+1}^k \frac{1}{t_j}. \quad (21)$$

$\hat{\sigma}_a^2(\theta)$, $\text{df}_a(\theta)$ を用いることにより, (14) 式の MSC は区分関数として,

$$\text{MSC}(\theta) = \phi_a(\theta) = f(\hat{\sigma}_a^2(\theta), \text{df}_a(\theta)) \quad (\theta \in R_a), \quad (22)$$

で与えられる. この $\phi_a(\theta)$ の $\theta \in R_a$ における最小化問題を解き, その解を用いて MSC 最小化問題を解く高速アルゴリズムを提案する.

§3. モデル選択規準最小化アルゴリズム

ここでは, MSC 最小化法に基づくチューニングパラメータ最適化問題の解決法について述べる. 特に本論文では GC_p , EGCV, GIC の最小化を考える. (15) 式と (22) 式から, チューニングパラメータ選択のための MSC の区分関数は具体的に以下のように与えられる.

$$\phi_a(\theta) = \begin{cases} \frac{1}{s_0^2}(n\hat{\sigma}_0^2 + c_{1,a} + c_{2,a}\theta^2) + \alpha(1 + k - a - c_{3,a}\theta) & (GC_p) \\ \frac{\hat{\sigma}_0^2 + c_{1,a}/n + c_{2,a}\theta^2/n}{(b + a/n + c_{3,a}\theta/n)^\alpha} & (\text{EGCV}) \\ \left\{ \hat{\sigma}_0^2 + \frac{1}{n}(c_{1,a} + c_{2,a}\theta^2) \right\} \exp \left\{ \frac{1}{n}\alpha(1 + k - a - c_{3,a}\theta) \right\} & (\text{GIC}) \end{cases} \quad (23)$$

ここで, $b = 1 - (k+1)/n$ であり, $\hat{\sigma}_0^2$, s_0^2 , R_a はそれぞれ (13) 式, (16) 式, (18) 式, $c_{1,a}$, $c_{2,a}$, $c_{3,a}$ は (21) 式で与えられる. また, $\phi_a(\theta)$ に基づく区分関数 $\psi_a(\theta)$ ($\theta \in R_a$) を以下

で定義する.

$$\psi_a(\theta) = \begin{cases} \frac{2}{s_0^2} c_{2,a} \theta - \alpha c_{3,a} & (GC_p) \\ -(\alpha - 2) c_{2,a} c_{3,a} \theta^2 + 2c_{2,a}(a + nb)\theta - \alpha c_{3,a}(c_{1,a} + n\hat{\sigma}_0^2) & (\text{EGCV}) \\ -\alpha c_{2,a} c_{3,a} \theta^2 + 2nc_{2,a}\theta - \alpha c_{3,a}(c_{1,a} + n\hat{\sigma}_0^2) & (\text{GIC}) \end{cases} \quad (24)$$

このとき, MSC 最小化問題の解は以下の定理で与えられる (証明は付録 A.4 参照).

定理 2 MSC 最小化に基づく最適なチューニングパラメータを $\hat{\theta}$ とする. すなわち,

$$\hat{\theta} = \arg \min_{\theta \in R_a} \text{MSC}(\theta).$$

このとき, GC_p , EGCV , GIC の最小化に基づく $\hat{\theta}$ は以下で定義される集合 $\mathcal{S}$ に含まれる.

$$\mathcal{S} = \left\{ \bigcup_{a \in \mathcal{A}} \{\xi_a\} \right\} \cup \left\{ \bigcup_{j \in \mathcal{J}} \{t_j\} \right\}. \quad (25)$$

すなわち,

$$\hat{\theta} = \arg \min_{\theta \in \mathcal{S}} \text{MSC}(\theta).$$

ここで, $\mathcal{A}$, $\mathcal{J}$ はそれぞれ,

$$\mathcal{A} = \{a \in \{0, \dots, k-1\} \mid \xi_a \in R_a\}, \quad (26)$$

$$\mathcal{J} = \{j \in \{1, \dots, k\} \mid \psi_{j-1}(t_j) \leq 0\} \cap \{j \in \{1, \dots, k-1\} \mid \psi_j(t_j) \geq 0\}, \quad (27)$$

で定義される添え字集合, t_j は (17) 式で与えられ, ξ_a は,

$$\xi_a = \begin{cases} \frac{\alpha s_0^2 c_{3,a}}{2c_{2,a}} & (GC_p) \\ \frac{c_{3,a}(c_{1,a} + n\hat{\sigma}_0^2)}{c_{2,a}(a + nb)} & (\text{EGCV} : \alpha = 2) \\ \frac{c_{2,a}(a + nb) - \sqrt{\mathcal{D}_{1,a}}}{(\alpha - 2)c_{2,a}c_{3,a}} & (\text{EGCV} : \alpha > 2 \wedge \mathcal{D}_{1,a} > 0) \\ \frac{nc_{2,a} - \sqrt{\mathcal{D}_{2,a}}}{\alpha c_{2,a}c_{3,a}} & (\text{GIC} : \mathcal{D}_{2,a} > 0) \end{cases}. \quad (28)$$

で与えられる. ただし, $\mathcal{D}_{1,a}$, $\mathcal{D}_{2,a}$ はそれぞれ EGCV ($\alpha > 2$), GIC の場合の $\psi_a(\theta) = 0$ の判別式でありそれぞれ,

$$\mathcal{D}_{1,a} = c_{2,a}^2(a + nb)^2 - \alpha(\alpha - 2)c_{2,a}c_{3,a}^2(c_{1,a} + n\hat{\sigma}_0^2),$$

$$\mathcal{D}_{2,a} = n^2 c_{2,a}^2 - \alpha^2 c_{2,a} c_{3,a}^2 (c_{1,a} + n \hat{\sigma}_0^2),$$

で与えられる.

$\mathcal{S}$ は元がすべて陽な形で求められる有限な離散集合であり, $\mathcal{A}$ と $\mathcal{J}$ の定義から $\#(\mathcal{S}) \leq 2k$ であることがわかる. 元々最適な θ は非負の実数全体での連続点の探索によって求めるものであったが, 定理 2 の結果から有限離散探索で十分であることが導けたため, チューニングパラメータの最適化を高速で行うことができる. 実際に MSC 最小化法を用いてチューニングパラメータの最適化をするときは, 以下のアルゴリズムを用いて計算すればよい.

• MSC 最小化に基づくチューニングパラメータ最適化アルゴリズム:

- (1) $\xi_a, \mathcal{A}, \mathcal{J}$ を用いて $\mathcal{S}$ の元を求める.
- (2) $\mathcal{S}$ の各元に対して $\text{MSC}(\theta)$ を計算し値を比較する.
- (3) (2) で $\text{MSC}(\theta)$ を最小とした値を最適なチューニングパラメータ $\hat{\theta}$ とする.

§4. 数値実験

ここでは, シミュレーションにより, §3 で提案した MSC 最小化アルゴリズムによって得られた最適な推定量の性能を比較する. GC_p , EGCV, GIC の α には $\alpha = 2, 2 \log \log n, \log n$ を用いる. シミュレーションに用いる疑似データは $N_n(\mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n)$ から発生させる. ここで, $\mathbf{X}, \boldsymbol{\beta}$ は以下のように決定する. $\mathbf{Z}_0$ を $n \times k$ 行列として各成分を独立に $U(-1, 1)$ から発生させる. また $\boldsymbol{\Phi}(\rho)$ ($|\rho| < 1$) は第 (i, j) 成分が $\rho^{|i-j|}$ となる k 次対称行列とする. これらを用いて $\mathbf{Z}$ を $\mathbf{Z} = (\mathbf{I}_n - \mathbf{J}_n)\mathbf{Z}_0\boldsymbol{\Phi}(\rho)^{1/2}$ で定義する. さらに $\mathbf{P}_1, \mathbf{D}, \mathbf{Q}$ をそれぞれ $\mathbf{Z}$ の特異値分解により得られる $n \times k, k \times k, k \times k$ 行列とする. すなわち, $\mathbf{Z} = \mathbf{P}_1\mathbf{D}\mathbf{Q}'$. これらを用いて, $\mathbf{X} = \mathbf{P}_1\mathbf{D}^{1/2}, \boldsymbol{\beta} = \mathbf{D}^{-1/2}\mathbf{P}_1'\boldsymbol{\eta}$ とする. ここで, $\boldsymbol{\eta}$ は第 j 成分が,

$$\sqrt{\frac{12n(n-1)}{4n^2+6n-1}} \left\{ (-1)^{j-1} \left(1 - \frac{j-1}{n} \right) - \frac{1}{2n} \right\},$$

で与えられる n 次元ベクトルである. このシミュレーションモデルは Yanagihara (2013) のものを参考にして構成した.

$\hat{\theta}$ を MSC 最小化法によって最適化されたチューニングパラメータとし, $\hat{\mathbf{y}}$ を $\boldsymbol{\beta}$ の LS 推定量に基づく $\mathbf{y}$ の予測値, $\hat{\mathbf{y}}_{\hat{\theta}}$ を $\hat{\theta}$ から得られる $\boldsymbol{\beta}$ の最適な Lasso 推定量に基づく $\mathbf{y}$ の

予測値とする。すなわち、

$$\hat{\mathbf{y}} = (\mathbf{J}_n + \mathbf{X}\mathbf{D}^{-1}\mathbf{X}')\mathbf{y}, \quad \hat{\mathbf{y}}_{\hat{\theta}} = (\mathbf{J}_n + \mathbf{X}\mathbf{L}_{\hat{\theta}}\mathbf{X}')\mathbf{y},$$

である。今回のシミュレーションでは以下のような平均二乗誤差 (Mean Square Error; MSE) の比を用いてそれぞれの MSC 最小化法の性能を評価する。

$$\frac{\mathbb{E}[(\hat{\mathbf{y}}_{\hat{\theta}} - \mathbf{X}\boldsymbol{\beta})'(\hat{\mathbf{y}}_{\hat{\theta}} - \mathbf{X}\boldsymbol{\beta})]}{\mathbb{E}[(\hat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})'(\hat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})]} \times 100 (\%).$$

今回の設定では $\mathbb{E}[(\hat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})'(\hat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})] = k + 1$ である。上記の MSE の比は 10,000 回の繰り返しによるモンテカルロシミュレーションにより計算され、値が小さいほど最適な Lasso 推定量に基づく予測値の MSE が LS 推定量に基づく予測値の MSE よりも小さい、すなわち Lasso 推定量が LS 推定量をより改善していることを表していることになる。逆に値が 100 に近いほど Lasso 推定量が LS 推定量を改善できていない、すなわち Lasso を使う意味がないことを表す。

表 1: それぞれの規準により最適化された Lasso 推定での MSE

n	k	$\alpha = 2$			$\alpha = 2 \log \log n$			$\alpha = \log n$		
		GC_p	EGCV	GIC	GC_p	EGCV	GIC	GC_p	EGCV	GIC
50	25	99.26	80.69	86.65	99.79	89.55	86.88	103.61	110.86	100.50
	40	86.91	79.96	88.95	87.95	84.49	86.67	91.08	94.12	91.87
200	100	84.05	73.51	78.63	83.34	74.84	76.99	90.68	92.72	93.48
	160	90.55	82.66	90.46	92.28	99.68	89.61	98.61	100.61	99.89
500	250	74.59	67.65	72.45	73.35	72.91	69.45	81.99	82.56	82.58
	400	72.33	67.51	81.68	73.15	80.95	69.90	82.32	82.55	82.55

表 1 は $\rho = 0.99$, $k = 0.5n$, $0.8n$ に対する $n = 50, 200, 500$ での結果を示している。表 1 から、ほとんどの場合で n が大きくなるにつれて性能が良くなっている。また α は大きいほうが良いわけではなく、それぞれの規準や状況に合わせた適切な α の選択が必要であるといえる。

参考文献

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd. International Symposium on Information Theory*, (B. N. Petrov & F. Csáki eds.), 267–281, Akadémiai Kiadó, Budapest.

- [2] Atkinson, A. C. (1980). A note on the generalized information criterion for choice of a model. *Biometrika*, **67**, 413–418.
- [3] Boyd, S., Parikh, N., Chu, E., Peleato, B. & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, **3**, 1–122.
- [4] Craven, P. & Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.*, **31**, 377–403.
- [5] Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R. (2004). Least angle regression. *Ann. Statist.*, **32**, 407–499.
- [6] Friedman, J. H., Hastie, T. & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.*, **33**, 1–22.
- [7] Fujikoshi, Y. & Satoh, K. (1997). Modified AIC and C_p in multivariate linear regression. *Biometrika*, **84**, 707–716.
- [8] Hannan, E. J. & Quinn, B. G. (1979). The determination of the order of an autoregression. *J. Roy. Stat. Soc. Ser. B*, **41**, 190–195.
- [9] Hoerl, A. E. & Kennard, R. W. (1970) Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55–67.
- [10] Mallows, C. L. (1973). Some comments on C_p . *Technometrics*, **15**, 661–675.
- [11] Mallows, C. L. (1995). Some comments on C_p . *Technometrics*, **37**, 362–372.
- [12] Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *Ann. Statist.*, **12**, 758–765.
- [13] Ohishi, M., Yanagihara, H. & Fujikoshi, Y. (2017). A fast algorithm for optimizing ridge parameters in a generalized ridge regression by minimizing an extended GCV criterion. *TR No. 17-07, Hiroshima Statistical Research Group*, Hiroshima University.
- [14] Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.
- [15] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. Ser. B*, **58**, 267–288.
- [16] Yanagihara, H. (2013). Explicit solution to the minimization problem of generalized cross-validation criterion for selecting ridge parameters in generalized ridge regression. *TR No. 13-07, Hiroshima Statistical Research Group*, Hiroshima University.

versity.

- [17] Yanagihara, Y., Nagai, I. & Satoh, K. (2009). A bias-corrected C_p criterion for optimizing ridge parameters in multivariate generalized ridge regression. *Japanese J. Appl. Statist.*, **38**, 151-172 (in Japanese).

付録

A.1 定理 1 の証明

(5) 式の PRSS は (6) 式で与えられた直交行列 $\mathbf{P}$ を用いて以下のように変形できる.

$$\begin{aligned} \text{PRSS}_\theta(\hat{\mu}, \beta) &= \frac{1}{2}(\mathbf{y} - \hat{\mu}\mathbf{1}_n - \mathbf{X}\beta)'(\mathbf{y} - \hat{\mu}\mathbf{1}_n - \mathbf{X}\beta) + \theta \sum_{j=1}^k |\beta_j| \\ &= \frac{1}{2}\{(\mathbf{I}_n - \mathbf{J}_n)\mathbf{y} - \mathbf{X}\beta\}'\mathbf{P}\mathbf{P}'\{(\mathbf{I}_n - \mathbf{J}_n)\mathbf{y} - \mathbf{X}\beta\} + \theta \sum_{j=1}^k |\beta_j| \\ &= \frac{1}{2}\left\{z - \begin{pmatrix} \mathbf{D}^{1/2} \\ \mathbf{O}_{n-k,k} \end{pmatrix} \beta\right\}'\left\{z - \begin{pmatrix} \mathbf{D}^{1/2} \\ \mathbf{O}_{n-k,k} \end{pmatrix} \beta\right\} + \theta \sum_{j=1}^k |\beta_j|. \end{aligned}$$

ここで, z は,

$$z = (z_1, \dots, z_n)' = \mathbf{P}'(\mathbf{I}_n - \mathbf{J}_n)\mathbf{y} = \begin{pmatrix} \mathbf{P}_1'\mathbf{y} \\ \mathbf{P}_2'(\mathbf{I}_n - \mathbf{J}_n)\mathbf{y} \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix},$$

で与えられる n 次元ベクトルである. よって, 上記の PRSS は,

$$\text{PRSS}_\theta(\hat{\mu}, \beta) = \frac{1}{2} \sum_{j=1}^k \left\{ (z_j - \sqrt{d_j}\beta_j)^2 + 2\theta|\beta_j| \right\} + \sum_{j=k+1}^n z_j^2,$$

となり, PRSS の最小化には第 1 項の j ごとの最小化, すなわち,

$$g_\theta(\beta_j | d_j) = (z_j - \sqrt{d_j}\beta_j)^2 + 2\theta|\beta_j| \quad (j = 1, \dots, k),$$

の最小化を考えればよい. ここで, $g_\theta(\beta_j | d_j)$ は,

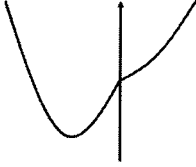
$$g_\theta(\beta_j | d_j) = \begin{cases} d_j\beta_j^2 - 2(z_j\sqrt{d_j} - \theta)\beta_j + z_j^2 & (\beta_j \geq 0) \\ d_j\beta_j^2 - 2(z_j\sqrt{d_j} + \theta)\beta_j + z_j^2 & (\beta_j < 0) \end{cases},$$

と変形でき, $d_j > 0$ であることに注意すると $g_\theta(\beta_j | d_j)$ は下に凸な 2 次関数であることがわかる. 頂点の β_j 軸座標はそれぞれ $(z_j\sqrt{d_j} - \theta)/d_j$, $(z_j\sqrt{d_j} + \theta)/d_j$ であり, それぞ

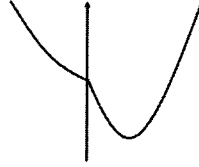
れの位置に注意することで図1のような3通りの場合分けができる．よって (10) 式が得られる．また, $\mathbf{X} = \mathbf{P}_1 \mathbf{D}^{1/2}$, $\mathbf{z}_1 = \mathbf{P}'_1 \mathbf{y}$ であることから,

$$\hat{\beta}_\theta = \begin{pmatrix} \hat{\beta}_{\theta,1} \\ \vdots \\ \hat{\beta}_{\theta,k} \end{pmatrix} = \begin{pmatrix} \ell_{\theta,1} \sqrt{d_1} z_1 \\ \vdots \\ \ell_{\theta,k} \sqrt{d_k} z_k \end{pmatrix} = \mathbf{L}_\theta \mathbf{D}^{1/2} \mathbf{z}_1 = \mathbf{L}_\theta \mathbf{D}^{1/2} \mathbf{P}'_1 \mathbf{y} = \mathbf{L}_\theta \mathbf{X}' \mathbf{y},$$

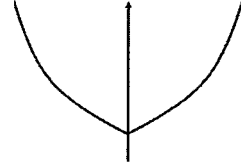
が成り立ち, (9) 式が得られる．



(a) $\theta < -\sqrt{d_j} z_j$ の場合



(b) $\theta < \sqrt{d_j} z_j$ の場合



(c) $\theta \geq \sqrt{d_j} |z_j|$ の場合

図1: $g_\theta(\beta_j | d_j)$ の形状

A.2 補題 1 の証明

まず $\hat{\sigma}^2(\theta)$ についての結果を示す. (11) 式の $\hat{\sigma}^2(\theta)$ は以下のように変形できる.

$$\begin{aligned} \hat{\sigma}^2(\theta) &= \frac{1}{n} \mathbf{y}' (\mathbf{I}_n - \mathbf{J}_n - \mathbf{X} \mathbf{L}_\theta \mathbf{X}')^2 \mathbf{y} \\ &= \frac{1}{n} \mathbf{y}' (\mathbf{I}_n - \mathbf{J}_n) (\mathbf{I}_n - \mathbf{X} \mathbf{L}_\theta \mathbf{X}')^2 (\mathbf{I}_n - \mathbf{J}_n) \mathbf{y} \\ &= \frac{1}{n} \mathbf{y}' (\mathbf{I}_n - \mathbf{J}_n) \left\{ \mathbf{P} \mathbf{P}' - \mathbf{P} \begin{pmatrix} \mathbf{I}_k \\ \mathbf{O}_{n-k,k} \end{pmatrix} \mathbf{D}^{1/2} \mathbf{L}_\theta \mathbf{D}^{1/2} (\mathbf{I}_k, \mathbf{O}_{k,n-k}) \mathbf{P}' \right\}^2 \\ &\quad \times (\mathbf{I}_n - \mathbf{J}_n) \mathbf{y} \\ &= \frac{1}{n} \mathbf{z}' \left\{ \mathbf{I}_n - \begin{pmatrix} \mathbf{D}^{1/2} \mathbf{L}_\theta \mathbf{D}^{1/2} & \mathbf{O}_{k,n-k} \\ \mathbf{O}_{n-k,k} & \mathbf{O}_{n-k,n-k} \end{pmatrix} \right\}^2 \mathbf{z} \\ &= \frac{1}{n} \mathbf{z}' \begin{pmatrix} \mathbf{I}_k - \mathbf{D}^{1/2} \mathbf{L}_\theta \mathbf{D}^{1/2} & \mathbf{O}_{k,n-k} \\ \mathbf{O}_{n-k,k} & \mathbf{I}_{n-k,n-k} \end{pmatrix}^2 \mathbf{z} \\ &= \frac{1}{n} (\mathbf{z}'_1, \mathbf{z}'_2) \begin{pmatrix} (\mathbf{I}_k - \mathbf{D}^{1/2} \mathbf{L}_\theta \mathbf{D}^{1/2})^2 & \mathbf{O}_{k,n-k} \\ \mathbf{O}_{n-k,k} & \mathbf{I}_{n-k,n-k} \end{pmatrix} \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix} \\ &= \frac{1}{n} \mathbf{z}'_1 (\mathbf{I}_k - \mathbf{D}^{1/2} \mathbf{L}_\theta \mathbf{D}^{1/2})^2 \mathbf{z}_1 + \mathbf{z}'_2 \mathbf{z}_2. \end{aligned}$$

上式の $\mathbf{I}_k - \mathbf{D}^{1/2} \mathbf{L}_\theta \mathbf{D}^{1/2}$ は対角行列であり, その第 j 対角成分は以下のように表すことができる.

$$\begin{aligned} 1 - \ell_{\theta,j} d_j &= \begin{cases} 1 - \frac{-\theta \operatorname{sign}(z_j) + \sqrt{d_j} z_j}{\sqrt{d_j} z_j} & (\theta < \sqrt{d_j} |z_j|) \\ 1 & (\theta \geq \sqrt{d_j} |z_j|) \end{cases} \\ &= \begin{cases} \frac{\theta}{\sqrt{d_j} |z_j|} & (\theta < \sqrt{d_j} |z_j|) \\ 1 & (\theta \geq \sqrt{d_j} |z_j|) \end{cases} \\ &= \mathbf{I}(\theta < \sqrt{d_j} |z_j|) \left(\frac{\theta}{\sqrt{d_j} |z_j|} - 1 \right) + 1. \end{aligned}$$

また, $\mathbf{z}'_2 \mathbf{z}_2$ は以下のように表すことができる.

$$\begin{aligned} \mathbf{z}'_2 \mathbf{z}_2 &= \mathbf{y}' (\mathbf{I}_n - \mathbf{J}_n) \mathbf{P}_2 \mathbf{P}'_2 (\mathbf{I}_n - \mathbf{J}_n) \mathbf{y} \\ &= \mathbf{y}' (\mathbf{I}_n - \mathbf{J}_n) (\mathbf{I}_n - \mathbf{P}_1 \mathbf{P}'_1) (\mathbf{I}_n - \mathbf{J}_n) \mathbf{y} \\ &= \mathbf{y}' \{ (\mathbf{I}_n - \mathbf{J}_n)^2 - (\mathbf{I}_n - \mathbf{J}_n) \mathbf{X} \mathbf{D}^{-1} \mathbf{X}' (\mathbf{I}_n - \mathbf{J}_n) \} \mathbf{y} \\ &= \mathbf{y}' \{ \mathbf{I}_n - \mathbf{J}_n - \mathbf{X} \mathbf{D}^{-1} \mathbf{X}' \} \mathbf{y} \\ &= n \hat{\sigma}_0^2. \end{aligned}$$

よって, $\hat{\sigma}^2(\theta)$ は補題 1 のように与えられる.

次に, $\operatorname{df}(\theta)$ についての結果を示す. (12) 式の $\mathbf{L}_\theta \mathbf{D}$ の第 j 対角成分は以下のように表すことができる.

$$\begin{aligned} \ell_{\theta,j} d_j &= \begin{cases} \frac{-\theta \operatorname{sign}(z_j) + \sqrt{d_j} z_j}{\sqrt{d_j} z_j} & (\theta < \sqrt{d_j} |z_j|) \\ 0 & (\theta \geq \sqrt{d_j} |z_j|) \end{cases} \\ &= \begin{cases} 1 - \frac{\theta}{\sqrt{d_j} |z_j|} & (\theta < \sqrt{d_j} |z_j|) \\ 1 - 1 & (\theta \geq \sqrt{d_j} |z_j|) \end{cases} \\ &= 1 - \left\{ \mathbf{I}(\theta < \sqrt{d_j} |z_j|) \left(\frac{\theta}{\sqrt{d_j} |z_j|} - 1 \right) + 1 \right\}. \end{aligned}$$

よって, $\operatorname{df}(\theta)$ は補題 1 のように与えられる. 以上から補題 1 が示された.

A.3 補題 2 の証明

$\theta \in R_a$ とする. (19), (20) 式から, $\hat{\sigma}_a^2(\theta)$ は θ に関して単調増加, $\text{df}_a(\theta)$ は θ に関して単調減少であることがわかる. また, $\hat{\sigma}_a^2(\theta)$, $\text{df}_a(\theta)$ に関してそれぞれ以下が成り立つ.

$$\begin{aligned}
 \hat{\sigma}_a^2(t_{a+1}) &= \hat{\sigma}_0^2 + \frac{1}{n}(c_{1,a} + c_{2,a}t_{a+1}^2) \\
 &= \hat{\sigma}_0^2 + \frac{1}{n} \left\{ \left(c_{1,a+1} - \frac{t_{a+1}^2}{q_{a+1}} \right) + \left(c_{2,a+1} + \frac{1}{q_{a+1}} \right) t_{a+1}^2 \right\} \\
 &= \hat{\sigma}_0^2 + \frac{1}{n}(c_{1,a+1} + c_{2,a+1}t_{a+1}^2) \\
 &= \hat{\sigma}_{a+1}^2(t_{a+1}), \\
 \text{df}_a(t_{a+1}) &= 1 + k - a - c_{3,a}t_{a+1} \\
 &= 1 + k - \{(a+1) - 1\} - \left(c_{3,a+1} + \frac{1}{t_{a+1}} \right) t_{a+1} \\
 &= 1 + k - (a+1) - c_{3,a+1}t_{a+1} \\
 &= \text{df}_{a+1}(t_{a+1}).
 \end{aligned}$$

よって, $\hat{\sigma}^2(\theta)$ は θ に関する連続な単調増加関数, $\text{df}(\theta)$ は θ に関する連続な単調減少関数である. さらに, 補題 1 から,

$$\begin{aligned}
 \hat{\sigma}^2(0) &= \hat{\sigma}_0^2, \quad \text{df}(0) = k + 1, \\
 \lim_{\theta \rightarrow \infty} \hat{\sigma}^2(\theta) &= \lim_{\theta \rightarrow \infty} \hat{\sigma}_k^2(\theta) \\
 &= \frac{1}{n}(n\hat{\sigma}_0^2 + \mathbf{z}'_1 \mathbf{z}_1) \\
 &= \frac{1}{n}(n\hat{\sigma}_0^2 + \mathbf{y}' \mathbf{P}_1 \mathbf{P}'_1 \mathbf{y}) \\
 &= \frac{1}{n} \{ \mathbf{y}' (\mathbf{I}_n - \mathbf{J}_n - \mathbf{X} \mathbf{D}^{-1} \mathbf{X}') \mathbf{y} + \mathbf{y}' \mathbf{X} \mathbf{D}^{-1} \mathbf{X}' \mathbf{y} \} \\
 &= \frac{1}{n} \mathbf{y}' (\mathbf{I}_n - \mathbf{J}_n) \mathbf{y}, \\
 \lim_{\theta \rightarrow \infty} \text{df}(\theta) &= \lim_{\theta \rightarrow \infty} \text{df}_k(\theta) = 1,
 \end{aligned}$$

となる. したがって, 補題 2 が示された.

A.4 定理 2 の証明

(23) 式の $\phi_a(\theta)$ を最小にするような θ を求めるため, $\phi_a(\theta)$ の導関数を考え, 極小値を探索する. $\phi_a(\theta)$ の導関数を以下のように表す.

$$\frac{\partial}{\partial \theta} \phi_a(\theta) = \lambda_a(\theta) \psi_a(\theta).$$

ここで, $\lambda_a(\theta)$ は,

$$\lambda_a(\theta) = \begin{cases} 1 & (GC_p) \\ \frac{1}{n(b + a/n + c_{3,a}\theta/n)^{\alpha+1}} & (EGCV : \alpha \geq 2) \\ \frac{1}{n^2} \exp \left\{ -\frac{\alpha}{n} (c_{3,a}\theta - 1 - k + a) \right\} & (GIC) \end{cases} \quad (\theta \in R_a),$$

で与えられ, $\psi_a(\theta)$ は (24) 式で与えられる. $\lambda_a(\theta)$ は常に正の値をとるため, $\phi_a(\theta)$ の導関数の符号は $\lambda_a(\theta)$ の値に依存しない. したがって極小値の探索には $\psi_a(\theta)$ を考えれば十分である. $\phi_a(\theta)$ が極小値となる点は $\psi_a(\theta)$ の符号が負から正に変わる点であるため, これを満たす点を区間ごと, また全区間を繋げた場合について探索すればよい. 以下 $\psi_a(\theta)$ の種類で場合分けして考える.

まず GC_p と $\alpha = 2$ の場合の EGCV について考える. この場合いずれも (24) 式から $\psi_a(\theta)$ は右肩上がりの 1 次関数であることがわかる. したがって区間ごとにみた場合, $\phi_a(\theta)$ が極小値となるのは $\psi_a(\theta)$ がその区間で θ 軸との交点を持つ場合である. すなわち ξ_a を方程式 $\psi_a(\theta) = 0$ の解とすると, $\xi_a \in R_a$ を満たすときの $\phi_a(\xi_a)$ が極小値となる. ここで ξ_a は (28) 式のように与えられる. また, 全区間で見た場合, 注目すべき点は各区間の境界点と両端点である. 各境界点に関して $\psi_a(\theta)$ は各区間では繋がっていない. すなわち, $\psi_a(t_{a+1}) \neq \psi_{a+1}(t_{a+1})$ ($a = 0, \dots, k-2$) である. したがって $\psi_a(t_{a+1}) \leq 0$ かつ $\psi_{a+1}(t_{a+1}) \geq 0$ を満たすとき, $\phi_a(t_{a+1})$ は極小値となる. 左端点である 0 に関しては,

$$\psi_0(0) = \begin{cases} -\alpha c_{3,0} & (GC_p) \\ -n\alpha c_{3,0} \hat{\sigma}_0^2 & (GCV : \alpha = 2) \end{cases} < 0,$$

が成り立つので $\phi_0(0)$ は極小値となり得ない. 右端点である t_k に関しては, $\phi_{k-1}(t_k)$ は極小値にはならないが, もし $\psi_{k-1}(t_k) \leq 0$ を満たすのであれば, $\theta \in R_{k-1}$ で $\phi_{k-1}(\theta)$ は単調減少であるため $\phi_{k-1}(t_k)$ は最小値になり得る.

次に GIC と $\alpha > 2$ の場合の EGCV について考える. この場合いずれも (24) 式から $\psi_a(\theta)$ は上に凸な 2 次関数であることがわかる. したがって区間ごとにみた場合, $\phi_a(\theta)$

が極小値となるのは $\psi_a(\theta)$ が θ 軸と 2 点で交わり、交点の小さい方がその区間に含まれる場合である。すなわち方程式 $\psi_a(\theta) = 0$ の実数解が 2 つ存在し、その小さい方を ξ_a とするとき、 $\xi_a \in R_a$ を満たすときの $\phi_a(\xi_a)$ が極小値となる。ここで ξ_a は (28) 式のように与えられる。また、全区間で見えた場合、 $\psi_a(t_{a+1}) \neq \psi_{a+1}(t_{a+1})$ ($a = 0, \dots, k-2$) であるため上記の場合と同様に $\psi_a(t_{a+1}) \leq 0$ かつ $\psi_{a+1}(t_{a+1}) \geq 0$ を満たすときの $\phi_a(t_{a+1})$ が極小値となる。さらに左端点である 0 に関しては、

$$\psi_0(0) = -n\alpha c_{3,0}\hat{\sigma}_0^2 < 0,$$

が成り立つので $\phi_0(0)$ は極小値となり得ない。右端点である t_k に関しては、上記の場合のように $\psi_{k-1}(t_k) \leq 0$ を満たすのであれば $\phi_{k-1}(t_k)$ が最小値になり得る。

以上から (26) 式、(27) 式の添え字集合 $\mathcal{A}$, $\mathcal{J}$ が導かれ、求める最小値の候補点集合が (25) 式の S で与えられることがわかる。したがって定理 2 が示された。