第4章　誤特定された線型モデル

前章まででおいた仮定は強いものであった．すなわち, 回帰関数 $f(\boldsymbol{x})$ は線型性をもつとした．

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{\beta}.$$

この仮定がみたされない場合はどうなるであろうか?　現実には, 線型性を信じたりはしない．よい統計手法は, このモデルからの乖離に対して頑強であることが期待される．線型性の誤特定にかかわる問題である．

この章では, 次のモデルを仮定する．

$$Y_j = f(\boldsymbol{X}_j) + \epsilon_j \qquad (j = 1, 2, \ldots, n). \tag{4.1}$$

ただし, $\boldsymbol{\epsilon} := (\epsilon_1, \epsilon_2, \ldots, , \boldsymbol{\epsilon}_n)^\top$ は代行分散 $\sigma^2\,(\sigma > 0)$ の劣 Gauss 分布とし, $\boldsymbol{X}_i \in \mathbb{R}^d$ とする．任意の関数 $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ に対して, $\boldsymbol{X}_j\,(j = 1, 2, \ldots, n)$ における f の値を集めたベクトルを

$$\boldsymbol{f} = (f(\boldsymbol{X}_1), f(\boldsymbol{X}_2), \ldots, f(\boldsymbol{X}_n))^\top \in \mathbb{R}^n$$

と書くことにする．固定デザインの場合

$$\mathsf{MSE}(\widehat{\boldsymbol{f}}) = \frac{1}{n}|\boldsymbol{f} - \widehat{\boldsymbol{f}}|_{2,n}^2 = \frac{1}{n}\sum_{j=1}^{n}\bigl\{f(\boldsymbol{X}_j) - \widehat{f}(\boldsymbol{X}_j)\bigr\}^2$$

として, f の推定量 $\widehat{f}$ の精度を評価する．ただし

$$\widehat{\boldsymbol{f}} = (\widehat{f}(\boldsymbol{X}_1), \widehat{f}(\boldsymbol{X}_2), \ldots, \widehat{f}(\boldsymbol{X}_n))^\top \in \mathbb{R}^n$$

である．

真のモデルは線型性をみたしていないとき, 前章で考えた推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}$, $\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}$, $\widehat{\boldsymbol{\beta}}^{\mathcal{L}}$ の精度を議論する．これらの推定量について何らかの主張をする．

4.1　オラクル不等式

オラクルとは, 興味ある量 (ここでは, 回帰関数) の事前の知識なしでは, 構成できない量のことをいう．ここでは, データを無限個観察したときの推定量のことをいう．

最小 2 乗推定量を採用するとき, 関数 f が線型性をみたしているかどうかに関わらず, 考察を関数

$$\boldsymbol{x} \mapsto \boldsymbol{\beta}^\top \boldsymbol{x} \qquad (\boldsymbol{x} \in \mathbb{R}^d)$$

に限定していることになる. したがって, オラクル $\widehat{f}$ は f にもっとも近い線型関数である. 線型関数 $\boldsymbol{\beta}^\top \boldsymbol{x} \approx f$ で f を近似するより, もうすこし一般的なモデルを考える. 関数 $\varphi_j : \mathbb{R}^d \longrightarrow \mathbb{R}\,(j = 1, 2, \ldots, M;\ M \in \mathbb{N})$ の集まりである (これを辞書と呼ぼう.) $\{\varphi_1, \varphi_2, \ldots, \varphi_M\}$ を考える. 実際, 目標は, 辞書の中の関数の線型結合

$$\varphi_{\boldsymbol{\beta}}(\boldsymbol{x}) = \sum_{j=1}^{N} \beta_j \varphi_j(\boldsymbol{x}) \qquad \big(\beta_j \in \mathbb{R};\ \boldsymbol{x} \in \mathbb{R}^d\big)$$

を用いて, f の推定を考える.

注意 4.1. $M = d$ とし, $\varphi_j(X) = X^{(j)}\,(j = 1, 2, \ldots, d)$ とする. ただし, $X^{(j)}$ は X の第 j 要素とする. このとき, $f(\boldsymbol{x})$ を $\boldsymbol{\beta}^\top \boldsymbol{x}$ で近似することになる. 辞書を使用することで, より広い枠組みを考えることができる. □

辞書をもちいても最小 2 乗推定量や制約付き最小 2 乗推定量は変化しない.

$$\varphi_{\boldsymbol{\beta}}(\boldsymbol{x}) = \sum_{j=1}^{M} \beta_j \varphi_j(\boldsymbol{x}) \qquad (\boldsymbol{x} \in \mathbb{R}^d;\ \beta_1, \beta_2, \ldots, \beta_M \in \mathbb{R}),$$
$$\boldsymbol{\varphi}_{\boldsymbol{\beta}} = \big(\varphi_{\boldsymbol{\beta}}(\boldsymbol{X}_1), \varphi_{\boldsymbol{\beta}}(\boldsymbol{X}_2), \ldots, \varphi_{\boldsymbol{\beta}}(\boldsymbol{X}_n)\big)^\top \in \mathbb{R}^n$$

とする.

1. 最小 2 乗推定量

$$\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} \in \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^M} \frac{1}{n} \sum_{j=1}^{n} \{Y_j - \varphi_{\boldsymbol{\beta}}(\boldsymbol{X}_j)\}^2. \tag{4.2}$$

2. $K \subset \mathbb{R}^M$ に制約した最小 2 乗推定量

$$\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_K \in \arg\min_{\boldsymbol{\beta} \in K} \frac{1}{n} \sum_{j=1}^{n} \{Y_j - \varphi_{\boldsymbol{\beta}}(\boldsymbol{X}_j)\}^2.$$

3. BIC 推定量

$$\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} \in \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^M} \left\{ \frac{1}{n} \sum_{j=1}^{n} \{Y_j - \varphi_{\boldsymbol{\beta}}(\boldsymbol{X}_j)\}^2 + 2\tau |\boldsymbol{\beta}|_0 \right\}. \tag{4.3}$$

ただし, $\tau > 0$ で $|\boldsymbol{\beta}|_0$ は $\boldsymbol{\beta}$ の零でない要素の個数である.

4. Lasso 推定量

$$\widehat{\boldsymbol{\beta}}^{\mathcal{L}} \in \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^M} \left\{ \frac{1}{n}\sum_{j=1}^{n}\{Y_j - \varphi_{\boldsymbol{\beta}}(\boldsymbol{X}_j)\}^2 + \tau^2|\boldsymbol{\beta}|_1 \right\}. \tag{4.4}$$

ただし, $\tau > 0$ で $|\boldsymbol{\beta}|_1 = \sum_{j=1}^{M}|\beta_j|\,(\boldsymbol{\beta} = (\beta_1,\, \beta_2,\, \ldots,\, \beta_M)^\top)$ である.

定義 4.2. $M \in \mathbb{N}$ とし, $\mathsf{R}(\cdot)$ をリスク関数とする. $\mathcal{H} = \{\varphi_1,\, \varphi_2,\, \ldots,\, \varphi_M\}$ を $\mathbb{R}^d$ 上で定義された $\mathbb{R}$ 値関数の辞書とする. K を $\mathbb{R}^M$ の部分集合とする. リスク関数 R に関する K 上のオラクル $\varphi_{\overline{\boldsymbol{\beta}}}\,(\overline{\boldsymbol{\beta}} \in K)$ を

$$\mathsf{R}\big(\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\big) \le \mathsf{R}\big(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) \qquad (\forall \boldsymbol{\beta} \in K)$$

をみたすものと定義する. さらに, $\mathsf{R}_K = \mathsf{R}\big(\varphi_{\overline{\boldsymbol{\beta}}}\big)$ を K 上のオラクルリスクという. 推定量 $\widehat{f}$ は剰余項 ϕ の K 上のオラクル不等式をみたすとは, 次のいずれかが成立することをいう.

(1) ある定数 $C \ge 1$ が存在して

$$\mathsf{E}\big[\mathsf{R}\big(\widehat{\boldsymbol{f}}\big)\big] \le C \inf_{\boldsymbol{\beta}\in K} \mathsf{R}\big(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) + \phi_{n,\,M}(K).$$

(2) または, $\forall\delta > 0$ に対して

$$\mathsf{Pr}\left(\mathsf{R}\big(\widehat{\boldsymbol{f}}\big) \le C \inf_{\boldsymbol{\beta}\in K} \mathsf{R}\big(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) + \phi_{n,\,M}(K)\right) \ge 1 - \delta.$$

$C = 1$ のとき, 不等式は厳密という.

4.1.1 最小 2 乗推定量のオラクル不等式

モデルが誤特定された場合について, BIC 推定量と LASSO 推定量に対するスパース・オラクル不等式を証明するのが目的である. この拡張を証明する上で難しいところは, 最小 2 乗推定量を扱うときに現れる.

定理 4.3. $n > M$ とし, $\{\varphi_1,\, \varphi_2,\, \ldots,\, \varphi_M\}$ を辞書とする. 回帰モデル (4.1) で $\boldsymbol{\epsilon} \sim \mathsf{subG}_n(\sigma^2)\,(\sigma > 0)$ を考える. このとき, 最小 2 乗推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}$ は次をみたす. ある $C > 0$ が存在して, 任意の $\forall\delta > 0$ に対して,

$$\mathsf{Pr}\left(\mathsf{MSE}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\big) \le \inf_{\boldsymbol{\beta}\in\mathbb{R}^M} \mathsf{MSE}\big(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) + C\sigma^2\frac{M + \log(1/\delta)}{n}\right) \ge 1 - \delta$$

が成立する. ただし, $\widehat{\beta}^{\mathrm{LS}}$ は (4.2) で与えられ

$$\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}} = \big(\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}(\boldsymbol{X}_1),\, \varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}(\boldsymbol{X}_2),\, \ldots,\, \varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}(\boldsymbol{X}_n)\big)^\top \in \mathbb{R}^n,$$

$$\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}(\boldsymbol{x}) = \sum_{j=1}^{M}\widehat{\beta}_j^{\mathrm{LS}}\varphi_j(\boldsymbol{x}); \qquad \widehat{\boldsymbol{\beta}}^{\mathrm{LS}} = \big(\widehat{\beta}_1^{\mathrm{LS}},\, \widehat{\beta}_2^{\mathrm{LS}},\, \ldots,\, \widehat{\beta}_M^{\mathrm{LS}}\big)^\top \in \mathbb{R}^M$$

である.

Proof. 最小 2 乗推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}$ の定義より,

$$\left|\boldsymbol{Y}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right|_{2,n}^2 \le \left|\boldsymbol{Y}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}^2$$

である. ただし, $\varphi_{\overline{\boldsymbol{\beta}}}$ は $\varphi_1, \varphi_2, \ldots, \varphi_M$ の張る線型部分空間 $\mathrm{span}\{\varphi_1, \varphi_2, \ldots, \varphi_M\}$ への f の直交射影である. $\boldsymbol{Y}=\boldsymbol{f}+\boldsymbol{\epsilon}$ なので

$$\begin{aligned}
\left|\boldsymbol{f}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right|_{2,n}^2 &= \left|\boldsymbol{Y}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}-\boldsymbol{\epsilon}\right|_{2,n}^2 \\
&= \left|\boldsymbol{Y}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right|_{2,n}^2 - 2\boldsymbol{\epsilon}^\top\left(\boldsymbol{Y}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right) + \left|\boldsymbol{\epsilon}\right|_{2,n} \\
&\le \left|\boldsymbol{Y}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}^2 - 2\boldsymbol{\epsilon}^\top\left(\boldsymbol{Y}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right) + \left|\boldsymbol{\epsilon}\right|_{2,n} \\
&= \left|\boldsymbol{f}+\boldsymbol{\epsilon}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}^2 - 2\boldsymbol{\epsilon}^\top\left(\boldsymbol{Y}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right) + \left|\boldsymbol{\epsilon}\right|_{2,n} \\
&= \left|\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}^2 + 2\boldsymbol{\epsilon}^\top\left(\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right) + \left|\boldsymbol{\epsilon}\right|_{2,n}^2 - 2\boldsymbol{\epsilon}^\top\left(\boldsymbol{Y}-\right) + \left|\boldsymbol{\epsilon}\right|_{2,n} \\
&= \left|\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}^2 + 2\boldsymbol{\epsilon}^\top\left\{\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}+\boldsymbol{\epsilon}-\boldsymbol{Y}+\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right\} \\
&= \left|\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}^2 + 2\boldsymbol{\epsilon}^\top\left\{\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right\}
\end{aligned}$$

となる. よって

$$\left|\boldsymbol{f}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right|_{2,n}^2 \le \left|\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}^2 + 2\boldsymbol{\epsilon}^\top\left\{\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right\} \tag{4.5}$$

を得る. さらに, Pythagoras の定理より

$$\begin{aligned}
\left|\boldsymbol{f}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right|_{2,n}^2 &= \left|\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}+\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right|_{2,n}^2 \\
&= \left|\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n} + 2\Big(\underbrace{\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}}_{\in\mathrm{span}\{\varphi_1,\ldots,\varphi_M\}^\perp}\Big)^\top\Big(\underbrace{\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}}_{\in\mathrm{span}\{\varphi_1,\ldots,\varphi_M\}}\Big) + \left|\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right|_{2,n}^2 \\
&= \left|\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n} + \left|\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right|_{2,n}^2
\end{aligned}$$

から

$$\left|\boldsymbol{f}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right|_{2,n}^2 - \left|\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}^2 = \left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}^2 \tag{4.6}$$

がわかる. (4.5) と (4.6) を合わせると

$$\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}^2 \le 2\boldsymbol{\epsilon}^\top\left(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right) = 2\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}\cdot\frac{\boldsymbol{\epsilon}^\top\left(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right)}{\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}}$$

いま, $\boldsymbol{\Psi}\in\mathbb{R}^{n\times M}$ を

$$\begin{bmatrix} \varphi_1(\boldsymbol{X}_1) \\ \varphi_1(\boldsymbol{X}_2) \\ \vdots \\ \varphi_1(\boldsymbol{X}_n) \end{bmatrix}, \begin{bmatrix} \varphi_2(\boldsymbol{X}_1) \\ \varphi_2(\boldsymbol{X}_2) \\ \vdots \\ \varphi_2(\boldsymbol{X}_n) \end{bmatrix}, \cdots, \begin{bmatrix} \varphi_M(\boldsymbol{X}_1) \\ \varphi_M(\boldsymbol{X}_2) \\ \vdots \\ \varphi_M(\boldsymbol{X}_n) \end{bmatrix}$$

の張る線型部分空間の正規直交基底とする．すると, $\boldsymbol{\nu} \in \mathbb{R}^M$ が存在して,

$$\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}} - \boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}} = \Phi\boldsymbol{\nu}$$

と書ける．よって

$$\frac{\boldsymbol{\epsilon}^\top\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}} - \boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\big)}{\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}} - \boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\big|_{2,\,n}} = \frac{\boldsymbol{\epsilon}^\top\Phi\boldsymbol{\nu}}{|\Phi\boldsymbol{\nu}|_{2,\,n}} = \frac{\boldsymbol{\epsilon}^\top\Phi\boldsymbol{\nu}}{|\boldsymbol{\nu}|_{2,\,M}} = \tilde{\boldsymbol{\epsilon}}^\top\frac{\boldsymbol{\nu}}{|\boldsymbol{\nu}|_{2,\,M}} \le \sup_{\boldsymbol{\mu}\in B2}\tilde{\boldsymbol{\epsilon}}^\top\boldsymbol{\mu}.$$

ただし, $B_2 := \big\{\boldsymbol{\mu}\in\mathbb{R}^M;\ |\boldsymbol{\mu}|_{2,\,M} = 1\big\}$ は $\mathbb{R}^M$ の単位球で, $\tilde{\boldsymbol{\epsilon}} = \Phi^\top\boldsymbol{\epsilon}$ である．したがって

$$\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}} - \boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\big|_2^2 \le 4\sup_{\boldsymbol{\mu}\in B_2}\big(\tilde{\boldsymbol{\epsilon}}^\top\boldsymbol{\mu}\big)^2. \tag{4.7}$$

次に, Φ の列は正規直交基底なので, $\boldsymbol{v}\in\mathbb{S}^{M-1}$ に対して

$$|\boldsymbol{\Phi v}|_{2,\,n}^2 = \boldsymbol{v}^\top\underbrace{\boldsymbol{\Phi}^\top\boldsymbol{\Phi}}_{=\boldsymbol{I}_M}\boldsymbol{v} = \boldsymbol{v}^\top\boldsymbol{v} = 1$$

となることがわかる．よって, $\boldsymbol{\epsilon} \sim \mathsf{subG}_n(\sigma^2)$ から, $\forall s\in\mathbb{R}$ に対して,

$$\mathsf{E}\big[\exp\{s\tilde{\boldsymbol{\epsilon}}\boldsymbol{v}\}\big] = \mathsf{E}\big[\exp\{s\big(\Phi\boldsymbol{v}\big)^\top\boldsymbol{\epsilon}\}\big] \le \exp\left\{\frac{s^2}{2}\sigma^2\right\}$$

がわかる．したがって, $\tilde{\boldsymbol{\epsilon}} \sim \mathsf{subG}_M(\sigma^2)$ となる．期待値の上限を得るために, 補題 2.10 から

$$\widetilde{\epsilon}_j \sim \mathsf{subG}(\sigma^2) \Rightarrow \mathsf{Pr}\big(\widetilde{\epsilon}_j > t\big) \le \exp\left(-\frac{t^2}{2\sigma^2}\right)$$

がわかる．さらに, 2.11 から

$$\mathsf{Pr}\big(\widetilde{\epsilon}_j > t\big) \le \exp\left(-\frac{t^2}{2\sigma^2}\right) \Rightarrow \mathsf{E}\big[\widetilde{\epsilon}_j^2\big] \le 2\sigma^2\times 2\times\Gamma(1) = 4\sigma^2$$

がわかる．よって

$$4\mathsf{E}\left[\sup_{\boldsymbol{\mu}\in B_2}\big(\tilde{\boldsymbol{\epsilon}}^\top\boldsymbol{\mu}\big)^2\right] = 4\sum_{j=1}^M\mathsf{E}\big[\tilde{\epsilon}_j^2\big] \le 16\sigma^2 M$$

を得る．ただし, $\tilde{\boldsymbol{\epsilon}} = (\tilde{\epsilon}_1, \tilde{\epsilon}_2, \ldots, \tilde{\epsilon}_M)$ である．さらに, $\widetilde{\boldsymbol{\epsilon}} \sim \mathsf{subG}_M(\sigma^2)$ に注意して, (2.5) の議論を繰り返すと

$$\mathsf{Pr}\left(\sup_{\boldsymbol{\mu}\in B_2}\big(\widetilde{\epsilon}^\top\boldsymbol{\mu}\big)^2 > t^2\right) = \mathsf{Pr}\left(\sup_{\boldsymbol{\mu}\in B_2}\widetilde{\epsilon}^\top\boldsymbol{\mu} > t\right) \le 6^M\exp\left(-\frac{t^2}{8\sigma^2}\right)$$

となるので

$$1 - 6^M\exp\left(-\frac{t^2}{8\sigma^2}\right) \le \mathsf{Pr}\left(\sup_{\boldsymbol{\mu}\in B_2}\big(\widetilde{\epsilon}^\top\boldsymbol{\mu}\big)^2 \le t^2\right)$$

となる．すると $t>0$ を以下の不等式を満たすように定めると

$$6^M \exp\left(-\frac{t^2}{8\sigma^2}\right) \le \delta \Leftrightarrow t^2 \ge 8\sigma^2 M \log 6 + 8\sigma^2 \log\left(\frac{1}{\delta}\right)$$

となる．$\sqrt{8\log 6} < 4$ なので

$$t := 4\sigma\sqrt{M} + 2\sigma\sqrt{2\log\left(\frac{1}{\delta}\right)}$$

とおけば

$$t^2 \ge 8\sigma^2 M \log 6 + 8\sigma^2 \log\left(\frac{1}{\delta}\right)$$

となる．したがって

$$\begin{aligned}
\mathsf{Pr}&\left(\sup_{\boldsymbol{\mu}\in B_2}\left(\boldsymbol{\mu}^\top\widetilde{\boldsymbol{\epsilon}}\right)^2 > 8\sigma^2 M\log 6 + 8\sigma^2\log\left(\frac{1}{\delta}\right)\right)\\
&\le \mathsf{Pr}\left(\sup_{\boldsymbol{\mu}\in B_2}\left(\boldsymbol{\mu}^\top\widetilde{\boldsymbol{\epsilon}}\right)^2 > t^2\right)\\
&\le 6^M\exp\left(-\frac{t^2}{8\sigma^2}\right)\\
&\le \exp\left(-\frac{t^2}{8\sigma^2} + M\log 6\right)\\
&\le \delta
\end{aligned}$$

から

$$\mathsf{Pr}\left(\sup_{\boldsymbol{\mu}\in B_2}\left(\tilde{\boldsymbol{\epsilon}}^\top\boldsymbol{\mu}\right)^2 \le 8\log(6)\sigma^2 M + 8\sigma^2\log\left(\frac{1}{\delta}\right)\right) \ge 1-\delta$$

がわかる．ここで

$$\sup_{\boldsymbol{\mu}\in B_2}\left(\tilde{\boldsymbol{\epsilon}}^\top\boldsymbol{\mu}\right)^2 \ge \frac{1}{4}\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}} - \boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}^2 = \frac{1}{4}\left\{\left|\boldsymbol{f} - \boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right|_{2,n}^2 - \left|\boldsymbol{f} - \boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}^n\right\}$$

を思い出すと

$$\begin{aligned}
\mathsf{Pr}&\left(\frac{1}{4}\left\{\left|\boldsymbol{f} - \boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right|_{2,n}^2 - \left|\boldsymbol{f} - \boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}^2\right\} \le 8\log(6)\sigma^2 M + 8\sigma^2\log\left(\frac{1}{\delta}\right)\right)\\
&\ge \mathsf{Pr}\left(\sup_{\boldsymbol{\mu}\in B_2}\left(\tilde{\boldsymbol{\epsilon}}^\top\boldsymbol{\mu}\right)^2 \le 8\log(6)\sigma^2 M + 8\sigma^2\log\left(\frac{1}{\delta}\right)\right)\\
&\ge 1-\delta
\end{aligned}$$

がわかる．上式を整理すると

$$\begin{aligned}
\mathsf{Pr}&\left(\frac{1}{n}\left|\boldsymbol{f} - \boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right|_{2,n}^2 \le \frac{1}{n}\left|\boldsymbol{f} - \boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right|_{2,n}^2 + \frac{32\sigma^2\left\{M\log 6 + \log\left(\delta^{-1}\right)\right\}}{n}\right)\\
&\ge 1-\delta
\end{aligned}$$

を得る．よって, 定理は証明された． □

4.1.2 BIC 推定量のスパース・オラクル不等式

定理 4.4. 回帰モデル (4.1) と $\boldsymbol{\epsilon} \sim \mathsf{subG}_n(\sigma^2)\,(\sigma > 0)$ を仮定する．このとき, 調整パラメータ

$$\tau^2 = 32\log(12)\frac{\sigma^2}{n} + 32\frac{\sigma^2\log(Me)}{n} \tag{4.8}$$

をもつ BIC 推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}$ は次をみたす．$\forall\delta > 0$ とする．ある定数 $C > 0$ が存在して,

$$\mathsf{Pr}\Bigg(\mathsf{MSE}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}\big) \le \inf_{\boldsymbol{\beta}\in\mathbb{R}^M}\left\{3\mathsf{MSE}\big(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) + \frac{C\sigma^2}{n}|\boldsymbol{\beta}|_{0,\,M}\log(eM)\right\} + \frac{C\sigma^2}{n}\log\left(\frac{1}{\delta}\right)\Bigg) \ge 1-\delta$$

が成立する．ただし, $\widehat{\beta}^{\mathrm{BIC}}$ は (4.4) で与えられ

$$\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} = \big(\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}(\boldsymbol{X}_1),\, \varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}(\boldsymbol{X}_2),\, \ldots,\, \varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}(\boldsymbol{X}_n)\big)^\top \in \mathbb{R}^n,$$
$$\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}(\boldsymbol{x}) = \sum_{j=1}^M \widehat{\beta}_j^{\mathrm{BIC}}\varphi_j(\boldsymbol{x}); \qquad \widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} = \big(\widehat{\beta}_1^{\mathrm{BIC}},\, \widehat{\beta}_2^{\mathrm{BIC}},\, \ldots,\, \widehat{\beta}_M^{\mathrm{BIC}}\big)^\top \in \mathbb{R}^M$$

である．

Proof. BIC 推定量の定義より

$$\frac{1}{n}\big|\boldsymbol{Y} - \boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}\big|_{2,\,n}^2 + \tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\big|_{0,\,M} \le \frac{1}{n}\big|\boldsymbol{Y} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n}^2 + \tau^2\big|\boldsymbol{\beta}\big|_{0,\,M}$$

が成立することに注意する．これより

$$\big|\boldsymbol{f} - \boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}\big|_{2,\,n}^2 + n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\big|_{0,\,M} \le \big|\boldsymbol{f} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n}^2 + 2\boldsymbol{\epsilon}^\top\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) + n\tau^2\big|\boldsymbol{\beta}\big|_{0,\,M} \tag{4.9}$$

を得る．$\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} = \boldsymbol{\beta}$ のとき, 主張は自明なので, $\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} \ne \boldsymbol{\beta}$ と仮定する．このとき, $\alpha > 0$ に対して

$$\begin{aligned}
&2\boldsymbol{\epsilon}^\top\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) && (4.10)\\
&\quad= 2\boldsymbol{\epsilon}^\top\left(\frac{\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}}{\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n}}\right)\big|\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n} \\
&\quad\le \frac{2}{\alpha}\left[\boldsymbol{\epsilon}^\top\left(\frac{\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}}{\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n}}\right)\right]^2 + \frac{\alpha}{2}\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n}^2 && (4.11)
\end{aligned}$$

が成立する．これは Young の不等式

$$2ab \le \frac{2}{\alpha}a^2 + \frac{\alpha}{2}b^2 \qquad (\forall a,\, b \in \mathbb{R};\, \alpha > 0)$$

よりわかる．次に, $(a+b)^2 \le 2a^2 + 2b^2\,(a,\, b \in \mathbb{R})$ から

$$\frac{1}{4}\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n}^2 \le \frac{1}{2}\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{f}\big|_{2,\,n}^2 + \frac{1}{2}\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,\,n}^2 \tag{4.12}$$

となることに注意する．(4.9) に (4.11) を代入し，さらに (4.12) を用いると

$$\begin{aligned}
\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{f}\big|_{2,\,n}^2 \;\le\;& \big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,\,n}^2 + n\tau^2\big|\boldsymbol{\beta}\big|_{0,\,M} - n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\big|_{0,\,M} \\
&+2\boldsymbol{\epsilon}^\top\frac{\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)}{\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n}} \\
\le\;& \big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2}^2 + n\tau^2\big|\boldsymbol{\beta}\big|_{0,\,M} - n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\big|_{0,\,M} \\
&+4\left[\boldsymbol{\epsilon}^\top\frac{\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)}{\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n}}\right]^2 + \frac{1}{4}\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n}^2 \\
\le\;& \frac{3}{2}\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,\,n}^2 + n\tau^2\big|\boldsymbol{\beta}\big|_{0,\,M} - n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\big|_{0,\,M} \\
&+4\left[\boldsymbol{\epsilon}^\top\frac{\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)}{\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n}}\right]^2 + \frac{1}{2}\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{f}\big|_{2,\,n}^2 \\
\le\;& \frac{3}{2}\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,\,n}^2 + 2n\tau^2\big|\boldsymbol{\beta}\big|_{0} - n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \boldsymbol{\beta}\big|_{0,\,M} \\
&+4\left[\boldsymbol{\epsilon}^\top\frac{\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)}{\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n}}\right]^2 + \frac{1}{2}\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{f}\big|_{2,\,n}^2
\end{aligned}$$

を得る．これを整理すれば

$$\begin{aligned}
\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{f}\big|_{2,\,n}^2 \;\le\;& 3\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,\,n}^2 + 4n\tau^2\big|\boldsymbol{\beta}\big|_{0,\,M} - 2n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\big|_{0,\,M} \\
&+8\big[\boldsymbol{\epsilon}^\top\mathcal{U}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)\big]^2
\end{aligned} \tag{4.13}$$

を得る．ただし, $\mathcal{U}(\boldsymbol{z}) = \boldsymbol{z}/|\boldsymbol{z}|_{2,\,n}$ である．$S \subset \{1,\, 2,\, \ldots,\, M\}$ に対して

$$\boldsymbol{\Phi}_{r_{S,\,*}} = [\boldsymbol{\phi}_1,\, \boldsymbol{\phi}_2,\, \ldots,\, \boldsymbol{\phi}_{r_{S,\,*}}]$$

を集合 $\{\boldsymbol{\varphi}_j(\boldsymbol{x});\, j \in S \cup \mathrm{supp}(\boldsymbol{\beta})\}$ の張る線型部分空間の正規直交基底とする．ただし

$$\boldsymbol{\varphi}_j(\boldsymbol{x}) = \big(\varphi_j(\boldsymbol{X}_1),\, \varphi_j(\boldsymbol{X}_2),\, \ldots,\, \varphi_j(\boldsymbol{X}_n)\big)^\top; \qquad (j = 1,\, 2,\, \ldots,\, M)$$

で, $r_{S,\,*} \le \#(S) + \big|\boldsymbol{\beta}\big|_{0,\,M}$ である．このとき

$$\begin{aligned}
&4\big[\boldsymbol{\epsilon}^\top\mathcal{U}\big(\boldsymbol{\phi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\phi}_{\boldsymbol{\beta}}\big)\big]^2 - n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\big|_{0,\,M} \\
&\quad = \max_{1 \le k \le M}\left\{\max_{\#(S)=k}\left\{\sup_{\mathrm{supp}(\tilde{\boldsymbol{\beta}})=S}\left\{4\big[\boldsymbol{\epsilon}^\top\mathcal{U}\big(\boldsymbol{\varphi}_{\tilde{\boldsymbol{\beta}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)\big]^4 - n\tau^2 k\right\}\right\}\right\}
\end{aligned}$$

となる. よって, $\forall t > 0$ に対して

$$\mathsf{Pr}\left(\max_{\#(S)=k}\left\{\max_{\#(S)=k}\left\{4\sup_{\mathrm{supp}(\tilde{\boldsymbol{\beta}})=S}\left[\boldsymbol{\epsilon}^\top\mathcal{U}\left(\boldsymbol{\varphi}_{\tilde{\boldsymbol{\beta}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right)\right]^4 - n\tau^2 k\right\}\right\}\right\} > \frac{t}{2}\right)$$
$$\leq \sum_{k=1}^{M}\sum_{\#(S)=k}\mathsf{Pr}\left(\sup_{\boldsymbol{u}\in\mathbb{B}_{r_{S,*}}}\left[\boldsymbol{\epsilon}^\top\Phi_{r_{S,*}}\boldsymbol{u}\right]^2 > \frac{t}{8}+\frac{1}{4}n\tau^2 k\right)$$

がわかる. 次に, 定理 2.30 の ϵ 網論法を用いれば

$$\mathsf{Pr}\left(\sup_{\boldsymbol{u}\in\mathbb{B}_{r_{S,*}}}\left[\boldsymbol{\epsilon}^\top\Phi_{r_{S,*}}\boldsymbol{u}\right]^2 > \frac{t}{8}+\frac{1}{4}n\tau^2 k\right)$$
$$\leq 2\cdot 6^{r_{S,*}}\exp\left(-\frac{\frac{t}{8}+\frac{1}{4}n\tau^2 k}{8\sigma^2}\right)$$
$$\leq 2\exp\left(-\frac{t}{64\sigma^2}-\frac{n\tau^2 k}{32\sigma^2}+(k+|\boldsymbol{\beta}|_0)\log 6\right)$$
$$= 2\exp\left(-\frac{t}{64\sigma^2}-k\log 6-k\log(eM)+(k+|\boldsymbol{\beta}|_0)\log 6\right)$$
$$= \exp\left(\log 2-\frac{t}{64\sigma^2}-k\log(eM)+|\boldsymbol{\beta}|_{0,M}\log 6\right)$$
$$\leq \exp\left(-\frac{t}{64\sigma^2}-k\log(eM)+|\boldsymbol{\beta}|_{0,M}\log(12)\right)$$

がわかる. さらに, (4.13) より

$$\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{f}\right|_{2,n}^2 > 3\left|\boldsymbol{\varphi}_{\boldsymbol{\beta}}-\boldsymbol{f}\right|_{2,n}^2+2n\tau^2|\boldsymbol{\beta}|_{0,M}+t$$

ならば

$$4\left[\boldsymbol{\epsilon}^\top\left(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right)\right]^2-n\tau^2\left|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\right|_{0,M}\geq\frac{t}{2}$$

である. したがって

$$\mathsf{Pr}\left(\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{f}\right|_{2,n}^2 > 2n\tau^2\left|\boldsymbol{\beta}\right|_{0,M}+t+3\left|\boldsymbol{\varphi}_{\boldsymbol{\beta}}-\boldsymbol{f}\right|_{2,n}^2\right)$$
$$\leq \mathsf{Pr}\left(4\left[\boldsymbol{\epsilon}^\top\mathcal{U}\left(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right)\right]^2-n\tau^2\left|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\right|_{0,M}\geq\frac{t}{2}\right)$$
$$\leq \sum_{k=1}^{M}\sum_{\#(S)=k}\exp\left(-\frac{t}{64\sigma^2}-k\log(eM)+|\boldsymbol{\beta}|_{0,M}\log(12)\right)$$
$$\leq \sum_{k=1}^{M}\left(\frac{eM}{k}\right)^k\exp\left(-\frac{t}{64\sigma^2}-k\log(eM)+|\boldsymbol{\beta}|_{0,M}\log(12)\right)$$
$$(\because \text{補題 } 3.10 \text{ より})$$
$$\leq \exp\left(-\frac{t}{64\sigma^2}+|\boldsymbol{\beta}|_{0,M}\log(12)\right)$$

を得る．ここで

$$t := 64\sigma^2|\boldsymbol{\beta}|_{0,M}\log(12) + 64\sigma^2\log\left(\frac{1}{\delta}\right)$$

と取れば

$$\begin{aligned}
&\mathsf{Pr}\Bigg(\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{f}\right|_{2,n}^2 > 2n\tau^2|\boldsymbol{\beta}|_{0,M} + 64\sigma^2|\boldsymbol{\beta}|_{0,M}\log(12) + 64\sigma^2\log\left(\frac{1}{\delta}\right) \\
&\qquad\qquad + 3\left|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\right|_{2,n}^2\Bigg) \\
&\qquad \le \mathsf{Pr}\Bigg(\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{f}\right|_{2,n}^2 > 64\sigma^2\left|\boldsymbol{\beta}\right|_{0,M}\left\{\log(6) + \log(Me)\right\} \\
&\qquad\qquad + 64\sigma^2\log\left(\frac{1}{\delta}\right) + 3\left|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\right|_{2,n}^2\Bigg) \\
&\qquad \le \delta
\end{aligned}$$

を得る．よって

$$\begin{aligned}
1-\delta \quad &\le \mathsf{Pr}\Bigg(\tfrac{1}{n}\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\right|_{2,n}^2 \le 3\mathsf{MSE}\left(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right) + |\boldsymbol{\beta}|_{0,M}\frac{C\sigma^2}{n}\log(eM) \\
&\qquad\qquad + \frac{C\sigma^2}{n}\log\left(\frac{1}{\delta}\right)\Bigg)
\end{aligned}$$

より

$$\begin{aligned}
\mathsf{Pr}\Bigg(\mathsf{MSE}\left(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}\right) \le \inf_{\boldsymbol{\beta}\in\mathbb{R}^M}&\left\{3\mathsf{MSE}\left(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right) + |\boldsymbol{\beta}|_{0,M}\frac{C\sigma^2}{n}\log(eM)\right\} \\
&+ \frac{C\sigma^2}{n}\log\left(\frac{1}{\delta}\right)\Bigg) \ge 1-\delta
\end{aligned}$$

を得る． □

注意 4.5. BIC 推定量は関数空間の近似誤差と母数 $\boldsymbol{\beta}$ の複雑さのトレード・オフを一番うまく行っている．このことをスパース・オラクル不等式ということもある．いま, $\overline{\boldsymbol{\beta}} \in \mathbb{R}^M$ を

$$\mathsf{MSE}\left(\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right) = \inf_{\boldsymbol{\beta}\in\mathbb{R}^M}\mathsf{MSE}\left(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right)$$

をみたすものとする．このとき, $\forall\delta > 0$ に対して

$$\begin{aligned}
&\mathsf{Pr}\Bigg(\mathsf{MSE}\left(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}\right) \le 3\mathsf{MSE}\left(\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right) + \frac{C\sigma^2}{n}\left\{|\overline{\boldsymbol{\beta}}|_0\log(eM) + \log\left(\frac{1}{\delta}\right)\right\}\Bigg) \\
&\qquad \ge 1-\delta
\end{aligned}$$

が成り立つ．真のモデルが線型性をもてば,

$$\mathsf{MSE}\left(\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\right) = 0$$

である． □

4.1.3 LASSO に対するスパースオラクル不等式

LASSO に対するオラクル不等式を証明するために, incoherence 条件のようなデザイン行列に対する仮定が必要になってくる. $n \times M$ のデザイン行列 $\mathbf{\Psi}$ の $(i,\, j)$ 要素と第 j 列を以下のように書く. すなわち

$$\begin{aligned}\mathbf{\Psi}_{ij} &= \varphi_j(\boldsymbol{X}_i) \qquad (i = 1,\, 2,\, \dots,\, n;\, j = 1,\, 2,\, \dots,\, M)\\ \boldsymbol{\varphi}_j &= (\varphi_j(\boldsymbol{X}_1),\, \varphi_j(\boldsymbol{X}_2),\, \dots,\, \varphi_j(\boldsymbol{X}_n))^\top\end{aligned}$$

である.

定理 4.6. 一般の回帰モデル (4.1) を仮定し, $\boldsymbol{\epsilon} \sim \mathsf{subG}_n(\sigma^2)\,(\sigma > 0)$ とする. さらに, ある正の整数 k があって, デザイン行列 $\mathbf{\Psi}$ は, $\mathsf{INC}(k)$ 条件をみたすとする. このとき, 調整パラメータ τ は, $\delta > 0$ としたとき,

$$2\tau := 8\sigma\sqrt{\frac{2\log(2M)}{n}} + 8\sigma\sqrt{\frac{2\log(1/\delta)}{n}}$$

で与えらる LASSO 推定量 $\widehat{\boldsymbol{\beta}}^{\mathcal{L}}$ は次をみたす. $\forall\delta > 0$ とある定数 (a generic constant) があって

$$\mathsf{Pr}\left(\mathsf{MSE}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}\big) \le \inf_{\boldsymbol{\beta}\in\mathbb{R}^M;\, |\boldsymbol{\beta}|_0\le k}\left\{\mathsf{MSE}\big(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) + \frac{C\sigma^2}{n}|\boldsymbol{\beta}|_{0,\,M}\log\left(\frac{2M}{\delta}\right)\right\}\right) \ge 1-\delta$$

が成り立つ. ただし, $\widehat{\beta}^{\mathcal{L}}$ は (??) で与えられ

$$\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} = \big(\varphi_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}(\boldsymbol{X}_1),\, \varphi_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}(\boldsymbol{X}_2),\, \dots,\, \varphi_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}(\boldsymbol{X}_n)\big)^\top \in \mathbb{R}^n,$$

$$\varphi_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}(\boldsymbol{x}) = \sum_{j=1}^{M}\widehat{\beta}_j^{\mathcal{L}}\varphi_j(\boldsymbol{x}); \qquad \widehat{\boldsymbol{\beta}}^{\mathcal{L}} = \big(\widehat{\beta}_1^{\mathcal{L}},\, \widehat{\beta}_2^{\mathcal{L}},\, \dots,\, \widehat{\beta}_M^{\mathcal{L}}\big)^\top \in \mathbb{R}^M$$

である.

Proof. LASSO 推定量 $\widehat{\boldsymbol{\beta}}^{\mathcal{L}}$ の定義より

$$\frac{1}{n}\big|\boldsymbol{Y} - \boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}\big|_{2,\,n}^2 \le \frac{1}{n}\big|\boldsymbol{Y} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n}^2 + 2\tau|\boldsymbol{\beta}|_{1,\,M} - 2\tau\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big|_{1,\,M} \tag{4.14}$$

が成立する. 2 次式の項を展開する.

$$\begin{aligned}\big|\boldsymbol{Y} - \boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}\big|_{2,\,n}^2 &= |\boldsymbol{Y} - \boldsymbol{f}|_{2,\,n}^2 + \big|\boldsymbol{f} - \boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}\big|_{2,\,n}^2 - 2\boldsymbol{\epsilon}^\top\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{f}\big)\\ \big|\boldsymbol{Y} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n}^2, &= |\boldsymbol{Y} - \boldsymbol{f}|_{2,\,n}^2 + \big|\boldsymbol{f} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,\,n}^2 - 2\boldsymbol{\epsilon}^\top\big(\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big)\end{aligned}$$

となる. これらの式を (4.14) に代入して, 両辺に $\tau\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}\big|_{1,\,M}$ を加えて, n 倍すれば

$$\begin{aligned}&\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{f}\big|_{2,\,n}^2 - \big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,\,n}^2 + n\tau\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}\big|_{1,\,M}\\ &\qquad \le 2\boldsymbol{\epsilon}^\top\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) + n\tau\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}\big|_{1,\,M} + 2n\tau|\boldsymbol{\beta}|_{1,\,M} - 2n\tau\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big|_{1,\,M}\end{aligned} \tag{4.15}$$

を得る. 任意の $k \geq 1$ に対する $\mathsf{INC}(k)$ 条件より

$$|\boldsymbol{\varphi}_j|_{2,n} \leq \sqrt{2n} \leq 2\sqrt{n} \qquad (j = 1, 2, \ldots, M)$$

が成立することに注意する. このことより,

$$\boldsymbol{\varphi}_j^\top \boldsymbol{\epsilon} \sim \mathsf{subG}(4n\sigma^2) \tag{4.16}$$

がわかる. 次に, $\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} = \boldsymbol{\Psi}^\top \widehat{\boldsymbol{\beta}}^{\mathcal{L}}$ に注意して, Hölder の不等式を用いると

$$\left|\boldsymbol{\epsilon}^\top \big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)\right| \leq \left|\boldsymbol{\Psi}^\top \boldsymbol{\epsilon}\right|_\infty \cdot \left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}\right|_{1,M} \tag{4.17}$$

を得る. (4.16) と (4.17) を踏まえて, 補題 2.16 を用いると

$$\mathsf{Pr}\left(\max_{1\leq j\leq M} \left|\boldsymbol{\varphi}_j^\top \boldsymbol{\epsilon}\right| > t\right) \leq 2M \exp\left(-\frac{t^2}{8n\sigma^2}\right)$$

を得る. よって

$$\begin{aligned}
&\mathsf{Pr}\left(2\boldsymbol{\epsilon}^\top \big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) > \frac{n\tau}{2}\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}\right|_{1,M}\right) \\
&\quad \leq \mathsf{Pr}\left(2\left|\boldsymbol{\epsilon}^\top \big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)\right| > \frac{n\tau}{2}\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}\right|_{1,M}\right) \\
&\quad \leq \mathsf{Pr}\left(\left|\Psi^\top \boldsymbol{\epsilon}\right|_\infty > \frac{n\tau}{4}\right) \\
&\quad = \mathsf{Pr}\left(\max_{1\leq j\leq M} \left|\boldsymbol{\varphi}_j^\top \boldsymbol{\epsilon}\right| > \frac{n\tau}{4}\right) \\
&\quad \leq 2M \exp\left(-\frac{n\tau^2}{64\sigma^2}\right)
\end{aligned}$$

を得る. したがって, $S = \mathrm{supp}(\boldsymbol{\beta})$ とおけば

$$2\boldsymbol{\epsilon}^\top \big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) \leq \frac{n\tau}{2}\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}\right|_{1,M}$$

のとき, 以下のような (4.15) の右辺の上限を得る.

$$\begin{aligned}
&2\boldsymbol{\epsilon}^\top \big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) + n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}\right|_{1,M} + +2n\tau|\boldsymbol{\beta}|_{1,M} - 2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\right|_{1,M} \\
&\quad \leq 2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}\right|_1 + 2n\tau|\boldsymbol{\beta}|_{1,M} - 2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\right|_{1,M} \\
&\quad \leq 2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S - \boldsymbol{\beta}_S\right|_{1,M} + 2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_{S^C}\right|_{1,M} + 2n\tau\left|\boldsymbol{\beta}_S\right|_{1,M} + 2n\tau|\boldsymbol{\beta}_S|_{1,M} \\
&\qquad -2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S\right|_{1,M} - 2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_{S^C}\right|_{1,M} \\
&\quad \leq 2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S - \boldsymbol{\beta}\right|_{1,M} + 2n\tau\left|\boldsymbol{\beta}_S\right|_1 + 2n\tau|\boldsymbol{\beta}_S|_{1,M} - 2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S\right|_{1,M} \\
&\quad \leq 4n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S - \boldsymbol{\beta}_S\right|_{1,M} \\
&\quad = 4n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S - \boldsymbol{\beta}_S\right|_{1,M}.
\end{aligned}$$

したがって

$$
\begin{aligned}
&\mathsf{Pr}\Big(\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}-\boldsymbol{f}\big|_{2,n}^2-\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}}-\boldsymbol{f}\big|_{2,n}^2+n\tau\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}-\boldsymbol{\beta}_S\big|_{1,M}\le 4n\tau\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S-\boldsymbol{\beta}\big|_{1,M}\Big)\\
&\qquad\ge \mathsf{Pr}\Big(2\big|\boldsymbol{\epsilon}^\top(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}})\big|\le\frac{n\tau}{2}\Big)\\
&\qquad\ge \mathsf{Pr}\Big(\max_{1\le j\le M}\big|\boldsymbol{\varphi}_j^\top\boldsymbol{\epsilon}\big|\le\frac{n\tau}{4}\Big)\\
&\qquad\ge 1-2M\exp\Big(-\frac{n\tau^2}{64\sigma^2}\Big) \qquad (4.18)
\end{aligned}
$$

となる．ここで

$$
\begin{aligned}
\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_{S^C}-\boldsymbol{\beta}_{S^C}\big|_{1,M}&\ge 3\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S-\boldsymbol{\beta}_S\big|_{1,M}\\
\Leftrightarrow\ 0&\le 3\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S-\boldsymbol{\beta}_S\big|_{1,M}-\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_{S^C}-\boldsymbol{\beta}_{S^C}\big|_{1,M}\\
&=4\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S-\boldsymbol{\beta}_S\big|_{1,M}-\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S-\boldsymbol{\beta}_S\big|_{1,M}-\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_{S^C}-\boldsymbol{\beta}_{S^C}\big|_{1,M}\\
&\le 4\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S-\boldsymbol{\beta}_S\big|_{1,M}-\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}-\boldsymbol{\beta}\big|_{1,M}
\end{aligned}
$$

となることから

$$
\begin{aligned}
\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_{S^C}-\boldsymbol{\beta}_{S^C}\big|_{1,M}&\ge 3\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S-\boldsymbol{\beta}_S\big|_{1,M}\\
&\Leftrightarrow 4\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S-\boldsymbol{\beta}_S\big|_{1,M}-\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}-\boldsymbol{\beta}\big|_{1,M} \qquad (4.19)
\end{aligned}
$$

がわかる．したがって

$$
\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}-\boldsymbol{f}\big|_{2,n}^2-\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}}-\boldsymbol{f}\big|_{2,n}^2+n\tau\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}-\boldsymbol{\beta}\big|_{1,M}\le 4n\tau\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S-\boldsymbol{\beta}_S\big|_{1,M} \qquad (4.20)
$$

が起きたとき (4.19) から

$$
\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_{S^C}-\boldsymbol{\beta}_{S^C}\big|_{1,M}\ge 3\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S-\boldsymbol{\beta}_S\big|_{1,M}\ \Rightarrow\ \mathsf{MSE}(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}})\le\mathsf{MSE}(\boldsymbol{\varphi}_{\boldsymbol{\beta}})
$$

が成立する．これは, (4.20) が起きれば

$$
\mathsf{MSE}(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}})\le\mathsf{MSE}(\boldsymbol{\varphi}_{\boldsymbol{\beta}}) \qquad (4.21)
$$

または

$$
\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_{S^C}-\boldsymbol{\beta}_{S^C}\big|_{1,M}< 3\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S-\boldsymbol{\beta}_S\big|_{1,M} \qquad (4.22)
$$

のいずれかが起きることがわかる．よって, (4.20) と (4.19) が起これば

$$
\mathsf{MSE}(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}})\le\inf_{\boldsymbol{\beta}\in\mathbb{R}^M;\,|\boldsymbol{\beta}|_0\le k}\Big\{\mathsf{MSE}(\boldsymbol{\varphi}_{\boldsymbol{\beta}})+\frac{C\sigma^2}{n}|\boldsymbol{\beta}|_{0,M}\log\Big(\frac{2M}{\delta}\Big)\Big\} \qquad (4.23)
$$

が明らかに成り立つので, (4.22) が起こったときに, (4.23) が成り立つことを示せばよい．そのために, $\tilde{\boldsymbol{\beta}}:=\widehat{\boldsymbol{\beta}}^{\mathcal{L}}-\boldsymbol{\beta}$ に対して, 凸条件 (3.22)

をみたすことを用いる. Cauchy-Schwarz の不等式と補題 3.24 ($X\boldsymbol{\beta}$ を $\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}} = \Psi^\top\big(\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}\big)$ と読み替える) より

$$\begin{aligned}4n\tau\big|\widehat{\boldsymbol{\beta}}_S^{\mathcal{L}} - \boldsymbol{\beta}_S\big|_{1,M} &\le 4n\tau\sqrt{\#(S)}\big|\widehat{\boldsymbol{\beta}}_S^{\mathcal{L}} - \boldsymbol{\beta}_S\big|_{2,M} \\ &\le 4n\tau\sqrt{\#(S)}\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}\big|_{2,M} \\ &\le 4n\tau\sqrt{\#(S)}\sqrt{\frac{2}{n}}\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,n} \\ &= 4\tau\sqrt{2n\#(S)}\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,n}\end{aligned}$$

を得る. ここで, Young の不等式

$$2ab \le \frac{2}{\alpha}a^2 + \frac{\alpha}{2}b^2 \qquad (a,\, b \in \mathbb{R},\, \alpha > 0)$$

を用いると

$$\begin{aligned}4n\tau\big|\widehat{\boldsymbol{\beta}}_S^{\mathcal{L}} - \boldsymbol{\beta}\big|_{1,M} &\le \frac{16\tau^2 n|\boldsymbol{\beta}|_{0,M}}{\alpha} + \frac{\alpha}{2}\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,n}^2 \\ &\le \frac{16\tau^2 n|\boldsymbol{\beta}|_{0,M}}{\alpha} + \alpha\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{f}\big|_{2,n}^2 + \alpha\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,n}^2\end{aligned}$$

を得る. この式と (4.15), (4.20) を合わせると

$$\begin{aligned}\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{f}\big|_{2,n}^2 + \big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,n}^2 &\le \big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{f}\big|_{2,n}^2 + \big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,n}^2 + n\tau\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}\big|_{1,M} \\ &\le 4n\tau\big|\widehat{\boldsymbol{\beta}}_S^{\mathcal{L}} - \boldsymbol{\beta}_S\big|_{1,M} \\ &\le \frac{16\tau^2 n|\boldsymbol{\beta}|_0}{\alpha} + \alpha\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{f}\big|_{2,n}^2 + \alpha\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,n}^2\end{aligned}$$

を得る. この式を整理して, 辺々を n で割ると

$$(1-\alpha)\mathsf{MSE}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}\big) \le (1+\alpha)\mathsf{MSE}\big(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) + \frac{16\tau^2|\boldsymbol{\beta}|_{0,M}}{\alpha}$$

を得る. この式の両辺を $1-\alpha$ で割り, $\alpha = 1/2$ とすれば

$$\mathsf{MSE}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}\big) \le 3\mathsf{MSE}\big(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) + 32\tau^2|\boldsymbol{\beta}|_{0,M}$$

を得る. 以上の議論を踏まえると

$$\begin{aligned}&1 - 2M\exp\left(-\frac{n\tau^2}{64\sigma^2}\right) \\ &\quad < \mathsf{Pr}\left(\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}} - \boldsymbol{f}\big|_{2,n}^2 - \big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,n}^2 + n\tau\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}\big|_{1,M} \le 4n\tau\big|\widehat{\boldsymbol{\beta}}_S^{\mathcal{L}} - \boldsymbol{\beta}_S\big|_{1,M}\right) \\ &\quad \le \mathsf{Pr}\left(\mathsf{MSE}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}\big) \le 3\mathsf{MSE}\big(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) + 32\tau^2|\boldsymbol{\beta}|_{0,M}\right) \\ &\quad = \mathsf{Pr}\left(\mathsf{MSE}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathcal{L}}}\big) \le 3\mathsf{MSE}\big(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) + 128\frac{\sigma^2}{n}|\boldsymbol{\beta}|_{0,M}\log\left(\frac{2M}{\delta}\right)\right)\end{aligned}$$

を得る. 最後に

$$\begin{aligned}\frac{n\tau^2}{64\sigma^2} &= \frac{n}{256\sigma^2}\left\{8\sigma\sqrt{\frac{2\log(2M)}{n}} + 8\sigma\sqrt{\frac{2\log(1/\delta)}{n}}\right\}^2 \\ &\leq \frac{n}{256\sigma^2}\left\{2\times 64\sigma^2\frac{2\log(2M)}{n} + 2\times 64\sigma^2\frac{2\log(1/\delta)}{n}\right\} \\ &= \log(2M) + \log\left(\frac{1}{\delta}\right)\end{aligned}$$

に注意すれば

$$\begin{aligned}1 - 2M\exp\left(-\frac{n\tau^2}{64\sigma^2}\right) &= 1 - \exp\left(\log(2M) - \frac{n\tau^2}{64\sigma^2}\right) \\ &\geq 1 - \exp\left(-\log\left(\frac{1}{\delta}\right)\right) \\ &\geq 1 - \delta\end{aligned}$$

がわかり, 定理は証明された. □

4.1.4 Maurey の論法

MSE$(\varphi_{\boldsymbol{\beta}})$ が小さくなるようなスパース性を持つベクトル $\boldsymbol{\beta}$ はあるのだろうか? LASSO 推定量の収束のスピードは遅ければ, 定理 4.6 の主張は役に立たない. 現実では, スパースベクトル $\boldsymbol{\beta}$ の存在など確証を持てない. むしろ, Zip の法則から期待できることは, $\boldsymbol{\beta}$ の要素の絶対値を取ったものを降順に並べたときに, 多項式のオーダーで減衰していくことである. そのような $\boldsymbol{\beta}$ に対しても, 定理 3.20 にある遅い収束レートの結果を誤特定の設定へ拡張できる. incoherence 条件が成立しているとき, LASSO 推定量は BIC 推定量よりも好ましい. なぜならば, LASSO 推定量は遅く収束レートであるが, BIC 推定量は早い収束レートである.

多項式のスピードで減衰するベクトル $\boldsymbol{\beta}$ は, 次 n 意味で, スパースベクトルで近似できる. 任意のベクトル $\boldsymbol{\beta} \in \mathbb{R}^M$ で, $|\boldsymbol{\beta}|_1 \leq 1$ なるものに対して, あるスパースベクトル $\boldsymbol{\beta}' \in \mathbb{R}^M$ が存在して, MSE$(\boldsymbol{\varphi}_{\boldsymbol{\beta}'})$ は MSE$(\boldsymbol{\varphi}_{\boldsymbol{\beta}})$ よりあまり大きくならないようにできる. 次の命題は, スパース性と MSE のトレード・オフを表現する主張である. これは, Mauery 論法と言われるが, Pisier (1981) によって知られるようになったが, 慣例に従い Mauery 論法ということにする.

命題 4.7. 集合 $\{\varphi_1, \varphi_2, \ldots, \varphi_M\}$ を規準化された辞書とする.

$$\max_{1\leq j\leq M} |\boldsymbol{\varphi}_j|_{2,n} \leq D\sqrt{n} \qquad (j = 1, 2, \ldots, M).$$

ただし, D は generic 定数である. このとき, 任意の正の整数 $1 \le \forall k \le M$ と任意の正の数 R に対して

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^M;\,|\boldsymbol{\beta}|_{0,\,M}\le k} \mathsf{MSE}\big(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) \le \min_{\boldsymbol{\beta}\in\mathbb{R}^M;\,|\boldsymbol{\beta}|_{1,\,M}\le R} \mathsf{MSE}\big(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big) + \frac{D^2R^2}{k}$$

が成り立つ.

Proof. $\overline{\boldsymbol{\beta}} \in \mathbb{R}^M$ を

$$\overline{\boldsymbol{\beta}} \in \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^M;\,|\boldsymbol{\beta}|_1\le R} \big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,\,n}^2$$

で定める. 一般性を失わず

$$|\overline{\beta}_1| \ge |\overline{\beta}_2| \ge \cdots \ge |\overline{\beta}_M|; \qquad \overline{\boldsymbol{\beta}} = (\overline{\beta}_1,\, \overline{\beta}_2,\, \ldots,\, \overline{\beta}_M)^\top$$

としてよい.

いま, $\boldsymbol{U} \in \mathbb{R}^n$ を $\{\boldsymbol{0},\, \pm R\boldsymbol{\varphi}_1,\, \pm R\boldsymbol{\varphi}_2,\, \ldots,\, \pm R\boldsymbol{\varphi}_M\}$ に値を取る確率ベクトルとする. さらに, その確率を

$$\begin{aligned}
\mathsf{Pr}\big(\boldsymbol{U} = R\,\mathrm{sign}\,(\overline{\beta}_j)\boldsymbol{\varphi}_j\big) &= \frac{|\overline{\beta}_j|}{R} \qquad (j = 1,\, 2,\, \ldots,\, M),\\
\mathsf{Pr}\big(\boldsymbol{U} = -R\,\mathrm{sign}\,(\overline{\beta}_j)\boldsymbol{\varphi}_j\big) &= 0,\\
\mathsf{Pr}(\boldsymbol{U} = \boldsymbol{0}_n) &= 1 - \frac{1}{R}\sum_{j=1}^M |\overline{\beta}_j|
\end{aligned}$$

で与える. このとき

$$\begin{aligned}
\mathsf{E}[\boldsymbol{U}] &= \sum_{j=1}^M R\,\mathrm{sign}\,(\overline{\beta}_j)\boldsymbol{\varphi}_j \frac{|\overline{\beta}_j|}{R} = \sum_{j=1}^M \overline{\beta}_j \boldsymbol{\varphi}_j = \boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}},\\
|\boldsymbol{U}|_{2,\,n} &\le \sqrt{\max_{1\le j\le M} R^2 |\boldsymbol{\varphi}_j|_{2,\,n}^2} \le RD\sqrt{n}
\end{aligned}$$

となる. いま, $\boldsymbol{U}_1,\, \boldsymbol{U}_2,\, \ldots,\, \boldsymbol{U}_k$ を $\boldsymbol{U}$ の独立複製とし

$$\overline{\boldsymbol{U}} := \frac{1}{k}\sum_{j=1}^k \boldsymbol{U}_j$$

と定めると, あるランダムベクトル $\widetilde{\boldsymbol{\beta}} \in \mathbb{R}^M$ で $\big|\widetilde{\boldsymbol{\beta}}\big|_{0,\,M} \le k$ なるものがあって

$$\overline{\boldsymbol{U}} = \boldsymbol{\varphi}_{\widetilde{\boldsymbol{\beta}}}$$

と書ける. なぜならば, $\overline{\boldsymbol{U}}$ は $R\mathrm{sign}(\overline{\beta}_j)\boldsymbol{\varphi}_j$ に値を取る k 個の $\boldsymbol{U}_i$ に基づく標本平均だからである. さらに

$$\mathsf{E}\big[\big(\boldsymbol{f} - \boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\big)^\top\big(\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}} - \overline{\boldsymbol{U}}\big)\big] = \big(\boldsymbol{f} - \boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\big)^\top\big(\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}} - \mathsf{E}[\overline{\boldsymbol{U}}]\big)\big] = 0$$

となるので

$$\begin{aligned}
\mathsf{E}\big[\big|\boldsymbol{f}-\overline{\boldsymbol{U}}\big|_{2,n}^2\big] &= \mathsf{E}\big[\big|\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}+\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}-\overline{\boldsymbol{U}}\big|_2^2\big] \\
&= \mathsf{E}\big[\big|\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\big|_2^2+\big|\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}-\overline{\boldsymbol{U}}\big|_2^2\big] \\
&= \big|\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\big|_{2,n}^2+\mathsf{E}\big[\big|\overline{\boldsymbol{U}}-\mathsf{E}[\overline{\boldsymbol{U}}]\big|_{2,n}^2\big] \\
&= \big|\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\big|_{2,n}^2+\frac{\mathsf{E}[\{\boldsymbol{U}-\mathsf{E}[\boldsymbol{U}]\}^2]}{k} \\
&\le \big|\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\big|_{2,n}^2+\frac{(DR\sqrt{n})^2}{k}
\end{aligned}$$

を得る．よって

$$\mathsf{E}\big[\big|\boldsymbol{f}-\overline{\boldsymbol{U}}\big|_{2,n}^2\big]=\mathsf{E}\big[\big|\boldsymbol{f}-\boldsymbol{\varphi}_{\tilde{\beta}}\big|_{2,n}^2\big]\ge\min_{\boldsymbol{\beta}\in\mathbb{R}^M;\,|\boldsymbol{\beta}|_{0,M}\le k}\big|\boldsymbol{f}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,n}^2$$

に注意すれば

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^M;\,|\boldsymbol{\beta}|_{0,M}\le k}\frac{1}{n}\big|\boldsymbol{f}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,n}^2\le\frac{1}{n}\big|\boldsymbol{f}-\boldsymbol{\varphi}_{\overline{\boldsymbol{\beta}}}\big|_{2,n}^2+\frac{D^2R^2}{k}$$

を得る．これから

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^M;\,|\boldsymbol{\beta}|_{0,M}\le k}\frac{1}{n}\big|\boldsymbol{f}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,n}^2\le\min_{\boldsymbol{\beta}\in\mathbb{R}^M;\,|\boldsymbol{\beta}|_{0,M}\le k}\frac{1}{n}\big|\boldsymbol{f}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big|_{2,n}^2+\frac{D^2R^2}{k}$$

がわかる． □

系 4.8. 定理 4.6 と同じ仮定をする．辞書 $\{\varphi_1,\,\varphi_2,\,\ldots,\,\varphi_M\}$ は規準化され,

$$\begin{aligned}
\max_{1\le j\le M}|\boldsymbol{\varphi}_j|_{2,n} &= \max_{1\le j\le M}\sqrt{\sum_{i=1}^n\varphi_j(\boldsymbol{X}_i)^2}\le\sqrt{n}, \\
\boldsymbol{\varphi}_j &= \big(\varphi_j(\boldsymbol{X}_1),\,\varphi_j(\boldsymbol{X}_2),\,\ldots,\,\varphi_j(\boldsymbol{X}_n)\big)^\top \qquad (j=1,\,2,\,\ldots,\,M)
\end{aligned}$$

をみたすとする．このとき, ある定数 C が存在して, 任意の $\delta>0$ に対して, BIC 推定量は次をみたす.

$$\begin{aligned}
\mathsf{Pr}\Bigg(\mathsf{MSE}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}\big)\le\inf_{\boldsymbol{\beta}\in\mathbb{R}^M}\Bigg\{&2\mathsf{MSE}\big(\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)+\frac{C\sigma^2\log(1/\delta)}{n} \\
&+C\left[\frac{\sigma^2|\boldsymbol{\beta}|_{0,M}\log(eM)}{n}\wedge\sigma|\boldsymbol{\beta}|_{0,M}\sqrt{\frac{\log(eM)}{n}}\right]\Bigg\}\Bigg) \\
\ge 1-\delta&
\end{aligned}$$

が成り立つ.

Proof. $\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}$ の定義より

$$\left|\boldsymbol{Y}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}\right|_{2,n}^2+n\tau^2\left|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\right|_{0,M}\le\left|\boldsymbol{Y}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right|_{2,n}^2+n\tau^2|\boldsymbol{\beta}|_{0,M}$$

を得る．これより

$$\begin{aligned}\left|\boldsymbol{f}-\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}\right|_{2,n}^2+n\tau^2\left|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\right|_{0,N} &\le \left|\boldsymbol{f}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right|_{2,n}^2+n\tau^2|\boldsymbol{\beta}|_{0,M}\\ &\quad+2\boldsymbol{\epsilon}^\top\left(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right)\end{aligned}\tag{4.24}$$

を得る．$\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}=\boldsymbol{\beta}$ のときは, 系の主張は自明なので, $\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\ne\boldsymbol{\beta}$ とする．このとき, Young の不等式

$$2ab\le\frac{2}{\alpha}a^2+\frac{\alpha}{2}b^2\qquad(a,\,b\in\mathbb{R},\,\alpha>0)$$

において, $\alpha=1/3$ とおけば

$$\begin{aligned}2\boldsymbol{\epsilon}^\top\left(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right) &= 2\boldsymbol{\epsilon}^\top\left(\frac{\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}}{\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right|_{2,n}}\right)\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right|_{2,n}\\ &\le 6\left[\boldsymbol{\epsilon}^\top\left(\frac{\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}}{\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right|_{2,n}}\right)\right]^2+\frac{1}{6}\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right|_{2,n}^2\end{aligned}$$

を得る．三角不等式より

$$\frac{1}{6}\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right|_{2,n}^2\le\frac{1}{3}\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{f}\right|_{2,n}^2+\frac{1}{3}\left|\boldsymbol{\varphi}_{\boldsymbol{\beta}}-\boldsymbol{f}\right|_{2,n}^2\tag{4.25}$$

を得る．これらからの式を (4.24) に代入すれば

$$\begin{aligned}\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right|_{2,n}^2 &\le \left|\boldsymbol{\varphi}_{\boldsymbol{\beta}}-\boldsymbol{f}\right|_{2,n}^2+n\tau^2\left|\boldsymbol{\beta}\right|_{0,M}-n\tau^2\left|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\right|_{0,M}\\ &\quad+2\boldsymbol{\epsilon}^\top\left(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right)\\ &\le \left|\boldsymbol{\varphi}_{\boldsymbol{\beta}}-\boldsymbol{f}\right|_{2,n}^2+n\tau^2\left|\boldsymbol{\beta}\right|_{0,M}-n\tau^2\left|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\right|_{0,M}\\ &\quad+6\left[\boldsymbol{\epsilon}^\top\left(\frac{\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}}{\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right|_{2,n}}\right)\right]^2+\frac{1}{6}\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right|_{2,n}^2\\ &\quad+2\boldsymbol{\epsilon}^\top\left(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right)\\ &= \left|\boldsymbol{\varphi}_{\boldsymbol{\beta}}-\boldsymbol{f}\right|_{2,n}^2+n\tau^2\left|\boldsymbol{\beta}\right|_{0,M}-n\tau^2\left|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\right|_{0,M}\\ &\quad+6\left[\boldsymbol{\epsilon}^\top\mathcal{U}\left(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right)\right]^2+\frac{1}{6}\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right|_{2,n}^2\\ &= \left|\boldsymbol{\varphi}_{\boldsymbol{\beta}}-\boldsymbol{f}\right|_{2,n}^2+n\tau^2\left|\boldsymbol{\beta}\right|_{0,M}-n\tau^2\left|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\right|_{0,M}\\ &\quad+6\left[\boldsymbol{\epsilon}^\top\mathcal{U}\left(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\right)\right]^2+\frac{1}{3}\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{f}\right|_{2,n}^2+\frac{1}{3}\left|\boldsymbol{\varphi}_{\boldsymbol{\beta}}-\boldsymbol{f}\right|_{2,n}^2\end{aligned}$$

を得る．これを整理すると

$$\begin{aligned}\frac{2}{3}\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{f}\big|_{2,n}^2 &\le \frac{4}{3}\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}}-\boldsymbol{f}\big|_{2,n}^2+n\tau^2\big|\boldsymbol{\beta}\big|_{0,M}-n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\big|_{0,M}\\ &\quad+6\big[\boldsymbol{\epsilon}^\top\mathcal{U}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)\big]^2 \qquad (4.26)\\ &\Longleftrightarrow\\ \big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{f}\big|_{2,n}^2 &\le 2\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}}-\boldsymbol{f}\big|_{2,n}^2-3n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}-\boldsymbol{\beta}\big|_{0,M}n\tau^2\big|\boldsymbol{\beta}\big|_{0,M}-n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\big|_{0,M}\\ &\quad+18\big[\boldsymbol{\epsilon}^\top\mathcal{U}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)\big]^2+6n\tau^2\big|\boldsymbol{\beta}\big|_{0,M} \qquad (4.27)\end{aligned}$$

を得る．$S\subset\{1,2,\ldots,M\}$ に対して

$$\boldsymbol{\Psi}_{S,*}=[\boldsymbol{\phi}_1,\,\boldsymbol{\phi}_2,\,\ldots,\,\boldsymbol{\phi}_{r_{S,*}}]$$

を集合

$$\{\boldsymbol{\varphi}_j;\, j\in S\cup\mathrm{supp}(\boldsymbol{\beta})\}$$

が張る線型部分空間の基底とする．ただし

$$\boldsymbol{\varphi}_j=(\boldsymbol{\varphi}_j(\boldsymbol{X}_1),\,\boldsymbol{\varphi}_j(\boldsymbol{X}_2),\,\ldots,\,\boldsymbol{\varphi}_j(\boldsymbol{X}_n))^\top$$

であり，

$$r_{S,*}\le\#(S)+|\boldsymbol{\beta}|_{0,M}$$

である．このとき

$$\begin{aligned}&18\big[\boldsymbol{\epsilon}^\top\mathcal{U}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)\big]^2-3n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}-\boldsymbol{\beta}\big|_{0,M}\\ &=\max_{1\le k\le M}\left\{\max_{\#(S)=k}\left\{\sup_{\mathrm{supp}(\tilde{\boldsymbol{\beta}})=S}\boldsymbol{\epsilon}^\top\mathcal{U}\big(\boldsymbol{\varphi}_{\tilde{\boldsymbol{\beta}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)\big]^2-3n\tau^2k\right\}\right\}\end{aligned}$$

となる．これより，$\forall t>0$ に対して，

$$\begin{aligned}&\mathsf{Pr}\left(\max_{1\le k\le M}\left\{\max_{\#(S)=k}\left\{\sup_{\mathrm{supp}(\tilde{\boldsymbol{\beta}})=S}\big[\boldsymbol{\epsilon}^\top\mathcal{U}\big(\boldsymbol{\varphi}_{\tilde{\boldsymbol{\beta}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)\big]^2-3n\tau^2k\right\}\right\}>t\right)\\ &\le\sum_{k=1}^m\sum_{\#(S)=k}\mathsf{Pr}\left(\sup_{\mathrm{supp}(\tilde{\boldsymbol{\beta}})=S}\left\{\big[\boldsymbol{\epsilon}^\top\mathcal{U}\big(\boldsymbol{\varphi}_{\tilde{\boldsymbol{\beta}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)\big]^2\right\}>\frac{t}{18}+\frac{1}{6}n\tau^2k\right)\end{aligned}$$

命題 2.30 の ϵ 網論法を用いる．

$$\begin{aligned}&\mathsf{Pr}\left(\sup_{\boldsymbol{u}\in\mathbb{B}_{r_{S,*}}}\big[\boldsymbol{\epsilon}^\top\Psi_{r_{S,*}}\boldsymbol{u}\big]^2>\frac{t}{18}+\frac{n\tau^2k}{6}\right)\\ &\le 2\cdot 6^{r_{S,*}}\exp\left(-\frac{\frac{t}{18}+\frac{n\tau^2k}{6}}{8\sigma^2}\right)\\ &\le 2\exp\left(-\frac{t}{8\cdot18\sigma^2}-\frac{n\tau^2k}{48\sigma^2}+|k+|\boldsymbol{\beta}|_{0,M})\log 6\right)\\ &\le 2\exp\left(-\frac{t}{8\cdot18\sigma^2}-k\log(6)-2k\log(eM)+\big(k+|\boldsymbol{\beta}|_{0,M}\big)\log(6)\right)\\ &\le 2\exp\left(-\frac{t}{8\cdot18\sigma^2}-2k\log(eM)+|\boldsymbol{\beta}|_{0,M}\log(12)\right)\end{aligned}$$

を得る．(4.25) より

$$6n\tau^2|\boldsymbol{\beta}|_{0,M} + t + 2\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,n}^2 \le \big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{f}\big|_{2,n}^2$$

が起これば

$$\begin{aligned} 6n\tau^2|\boldsymbol{\beta}|_{0,M} + t + 2\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,n}^2 \ \le\ & 2\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,n}^2 - 3n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \boldsymbol{\beta}\big|_{0,M} \\ & + 18\big[\boldsymbol{\epsilon}^\top \mathcal{U}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)\big]^2 + 6n\tau^2|\boldsymbol{\beta}|_{0,M} \end{aligned}$$

となる．よって

$$18\big[\boldsymbol{\epsilon}^\top \mathcal{U}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)\big]^2 - 3n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \boldsymbol{\beta}\big|_{0,M} \ge t$$

が起こる．よって

$$\begin{aligned} & \mathsf{Pr}\Big(\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{f}\big|_2^2\Big) > 2\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_2^2 + 6n\tau^2|\boldsymbol{\beta}|_{0,M} + t\Big) \\ & \le \mathsf{Pr}\Big(18\big[\boldsymbol{\epsilon}^\top \mathcal{U}\big(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}}\big)\big]^2 - n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \boldsymbol{\beta}\big|_{0,M} > t\Big) \\ & \le \sum_{k=1}^{M} \sum_{\#(S)=k} \mathsf{Pr}\Big(-\frac{t}{8\cdot 18\sigma^2} - 2k\log(eM) + |\boldsymbol{\beta}|_{0,M}\log(12)\Big) \\ & \le \sum_{k=1}^{M} \Big(\frac{eM}{k}\Big)^k \exp\Big(-\frac{t}{8\cdot 18\sigma^2} - 2k\log(eM) + |\boldsymbol{\beta}|_{0,M}\Big) \\ & \qquad (\text{補題 3.10 より}) \\ & = \sum_{k=1}^{M} \frac{1}{k^k} \exp\Big(-\frac{t}{18\cdot\ 8\sigma^2} - k\log(eM) + |\boldsymbol{\beta}|_{0,M}\log(12)\Big) \\ & \le \frac{M}{(keM)^k} \exp\Big(-\frac{t}{8\cdot\ 18\sigma^2} + |\boldsymbol{\beta}|_{0,M}\log(12)\Big) \\ & \le \exp\Big(-\frac{t}{8\cdot\ 18\sigma^2} + |\boldsymbol{\beta}|_{0,M}\log(12)\Big) \end{aligned}$$

を得る．ここで

$$t := 8\cdot\ 18\sigma^2|\boldsymbol{\beta}|_{0,M}\log(12) + 8\cdot\ 18\sigma^2\log\Big(\frac{1}{\delta}\Big)$$

とおけば

$$\begin{aligned} \mathsf{Pr}\Big(\big|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}} - \boldsymbol{f}\big|_{2,n}^2 \ >\ & 2\big|\boldsymbol{\varphi}_{\boldsymbol{\beta}} - \boldsymbol{f}\big|_{2,n}^2 + 6n\tau^2|\boldsymbol{\beta}|_{0,M} + 8\cdot\ 18\sigma^2|\boldsymbol{\beta}|_{0,M}\log(12) \\ & + 8\cdot\ 18\sigma^2\log\Big(\frac{1}{\delta}\Big)\Big) \le \delta \end{aligned}$$

となる. よって

$$\tau^2 = \frac{8 \cdot 18\sigma^2 \log(eM)}{n}$$

なので

$$\mathsf{Pr}\Bigg(\mathsf{MSE}(\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}}) \le 2 \inf_{\boldsymbol{\beta}\in\mathbb{R}^M} \mathsf{MSE}(\boldsymbol{\varphi}_{\boldsymbol{\beta}}) + C\frac{\sigma^2|\boldsymbol{\beta}|_{0,\,M}\log(eM)}{n} + C\frac{\sigma^2\log(1/\delta)}{n}\Bigg) \\ \ge 1-\delta$$

を得る. □

4.2 ノンパラメトリック回帰

記号

$$\begin{aligned} \mathcal{H} &= \{\varphi_1,\,\varphi_2,\,\ldots,\,\varphi_M\}: \text{選択された辞書}, \\ f &: [0,\,1] \longrightarrow \mathbb{R}, \\ \varphi_j &: [0,\,1] \longrightarrow \mathbb{R} \qquad (j=1,\,2,\,\ldots,\,M). \end{aligned}$$

4.2.1 フーリエ分解

辞書を $L_2([0,\,1])$ の正規直交系とする. すなわち,

$$\int_0^1 \varphi_j^x(x)\,\mathrm{d}x = 1; \qquad \int_0^1 \varphi_j(x)\varphi_k(x)\,\mathrm{d}x \qquad (1 \le j \ne k \le M).$$

ただし, $M=\infty$ の場合も扱う. その場合には, $1 \le j \ne k < \infty$ と考える.

$\mathcal{H}$ は正規直交系で, 係数 $\beta_j \in \mathbb{R}\,(j=1,\,2,\,\ldots)$ を

$$\beta_j = \int_0^1 f(x)\varphi_j(x)\,\mathrm{d}x$$

で定義する. これを f のフーリエ係数という.

回帰関数 f は

$$f = \sum_{j=1}^{\infty} \beta_j\varphi_j$$

なる分解を持つとする. ただし, 等号は

$$\lim_{M\to\infty}\int_0^1 \Big|f(x) - \sum_{j=1}^{M}\beta_j\varphi_j(x)\Big|^2\,\mathrm{d}x = 0$$

の意味である.

例 4.9. (三角関数基底) これは $L_2([0, 1])$ の正規直交基底となる.

$$
\begin{aligned}
\varphi_1(x) &= 1, \\
\varphi_{2k} &= \sqrt{2}\cos(2\pi kx), \\
\varphi_{2k+1} &= \sqrt{2}\sin(2\pi kx) \qquad (k = 1, 2, \ldots).
\end{aligned}
$$

□

例 4.10. (ウェーブレット) 関数 $\psi : \mathbb{R} \longrightarrow \mathbb{R}$ を十分滑らかでコンパクトな台を持つものとする. この ψ をマザーウェーブレットという. 関数系を

$$
\psi_{jk}(x) = 2^{j/2}\psi\bigl(2^j x - k\bigr) \qquad (j, k \in \mathbb{Z})
$$

で定義する. 適切な ψ に対して, 辞書 $\{\psi_{jk};\, j, k \in \mathbb{Z}\}$ は $L_2([0, 1])$ の正規直交系となり, 時には基底となる. 基底となる場合には, $\forall g \in L_2([0, 1])$ に対して

$$
g = \sum_{j=-\infty}^{j=\infty} \sum_{k=-\infty}^{k=\infty} \beta_{jk}\psi_{jk}; \qquad \beta_{jk} = \int_0^1 g(x)\psi_{jk}(x)\,\mathrm{d}x
$$

とすればよい. 係数 β_{jk} を g のウェーブレット係数という.

最も簡単な例は Haar ウェーブレットである.

$$
\psi(x) = \begin{cases} 1 & (0 < x < 1/2), \\ -1 & (1/2 \le x \le 1), \\ 0 & (\text{その他}). \end{cases}
$$

□

4.2.2 Sobolev 族と楕円体

滑らかな関数族を考える. 滑らかさは微分できる回数で表現される. $f^{(k)}$ を f の k 回導関数とする.

定義 4.11. $p \in \{1, 2, \ldots\}$ と $L > 0$ を固定する. 関数族である Sobolev クラス $W(p, L)$ を次で定義する.

$$
\begin{aligned}
W(p, L) \;:=\; \Bigl\{ & f : [0, , 1] \longrightarrow \mathbb{R};\, f \in L_2([0, 1]), f^{(p-1)} \text{ は絶対連続で}, \\
& \int_0^1 \{f^{(p)}(x)\}^2\,\mathrm{d}x \le L^2, \\
& f^{(j)}(0) = f^{(j)}(1) = 0\,(j = 0, 1, \ldots, p-1) \Bigr\}.
\end{aligned}
$$

任意の $f \in W(p,\, L)$ は

$$f = \beta_1\varphi_1 + \sum_{k=1}^{\infty}\{\beta_{2k}\varphi_{2k} + \beta_{2k+1}\varphi_{2k+1}$$

と展開される. ただし, 等号の意味は

$$\lim_{M\to\infty}\int_0^1 \Bigl|f(x) - \sum_{j=1}^{M}\beta_j\varphi_j(x)\Bigr|^2 \,\mathrm{d}x = 0$$

である. $\boldsymbol{\beta} = \{\beta_j\}_{j\geq 1}$ は

$$\ell_1(\mathbb{N}) = \Bigl\{\boldsymbol{\beta} = (\beta_1,\, \beta_2,\, \ldots) : \sum_{j=1}^{\infty}|\beta_j| < \infty\Bigr\}$$

に属する.

$\forall p > 0$ に対して, $a_j\ (j = 1,\, 2,\, \ldots)$ を

$$a_j = \begin{cases} j^p & (j \text{ は偶数}), \\ (j-1)^p & (j \text{ は奇数}), \end{cases} \tag{4.28}$$

で定める.

命題 4.12. $p \geq 1$ と L を固定する. $\{\varphi_j\}_{j\geq 1}$ を $L_2([0,\, 1])$ の三角関数基底とする. $\{a_j\}_{j\geq 1}$ を (4.28) で定義する. 関数 $f \in W(p,\, L)$ は

$$f = \sum_{j=1}^{\infty}\beta_j\varphi_j$$

と表現される. ただし, フーリエ係数 $\{\beta_j\}_{j\geq 1}$ は次で定義される $\ell_2(\mathbb{N})$ の Sobolev 楕円体に属する.

$$\begin{aligned} \Theta(p,\, L) &:= \Bigl\{\boldsymbol{\beta} \in \ell_2(\mathbb{N});\, \sum_{j=1}^{\infty} a_j\beta_j^2 \leq Q\Bigr\}, \\ Q &= \frac{L^2}{\pi^{2p}}. \end{aligned}$$

Proof. $j = 1,\, 2,\, \ldots,\, p$ の j 回導関数 $f^{(j)}$ のフーリエ係数を $\{s_k(j)\}_{k\geq 1}$ で定義する.

$$\begin{aligned} s_1(j) &= \int_0^1 f^{(j)}(t)\,\mathrm{d}t = f^{(j-1)}(1) - f^{(j-1)}(0) = 0, \\ s_{2k}(j) &= \sqrt{2}\int_0^1 f^{(j)}(t)\cos(2\pi kt)\,\mathrm{d}t, \\ s_{2k+1}(j) &= \sqrt{2}\int_0^1 f^{(j)}(t)\sin(2\pi kt)\,\mathrm{d}t. \end{aligned}$$

f のフーリエ係数は $\beta_k = s_k(0)$ で与えられる. 部分積分の公式より

$$\begin{aligned} s_{2k}(p) &= \sqrt{2} f^{(p-1)}(t)\cos(2\pi kt)\Big|_0^1 + (2\pi k)\sqrt{2}\int_0^1 f^{(p-1)}(t)\sin(2\pi kt)\,\mathrm{d}t \\ &= \sqrt{2}\big\{f^{(p-1)}(1) - f^{(p-1)}(0)\big\} + (2\pi k)\sqrt{2}\int_0^1 f^{(p-1)}(t)\sin(2\pi kt)\,\mathrm{d}t \\ &= (2pik)s_{2k+1}(p-1) \end{aligned}$$

となる. さらに

$$\begin{aligned} s_{2k+1}(p) &= = \sqrt{2} f^{(p-1)}(t)(2\pi kt)\Big|_0^1 - (2\pi k)\sqrt{2}\int_0^1 f^{(p-1)}(t)\cos(2\pi kt)\,\mathrm{d}t \\ &= -(2\pi k)s_{2k}(p-1) \end{aligned}$$

である. 特に, 帰納法より, $\forall k \geq 1$ に対して

$$\{s_{2k}(p)\}^2 + \{s_{2k+1}(p)\}^2 = (2\pi k)^{2p}\{\beta_{2k}^2 + \beta_{2k+1}^2\}$$

となる. 次に, (4.28) の a_j の定義より

$$\begin{aligned} \sum_{k=1}^{\infty}(2\pi k)^{2p}\{\beta_{2k}^2 + \beta_{2k+1}^2\} &= \pi^{2p}\Big\{\sum_{k=1}^{\infty} a_{2k}^2\beta_{2k}^2 + \sum_{k=1}^{\infty} a_{2k+1}^2\beta_{2k+1}^2\Big\} \\ &= \pi^{2p}\sum_{j=1}^{\infty} a_j^2\beta_j^2 \end{aligned}$$

である. よって, パーセルの等式より

$$\int_0^1 \{f^{(p)}(t)\}^2\,\mathrm{d}t = \sum_{k=1}^{\infty}\Big\{\{s_{2k}(p)\}^2 + \{s_{2k+1}(p)\}^2\Big\} = \pi^{2p}\sum_{j=1}^{\infty} a_j^2\beta_j^2$$

がわかる. $f \in W(p,\,L)$ なので

$$\int_0^1 \{f^{(p)}(t)\}^2\,\mathrm{d}t \leq L^2$$

となる. よって, $\boldsymbol{\beta} \in \Theta(p,\,L^2/\pi^{2p})$ である. □

注意 4.13. 命題 4.12 の逆も成立することが知られている. この本では, この事実を使わないので, 証明しない.

命題 4.14. Sobolev 楕円体は次の性質を持つ.
(1) $\forall Q > 0$ に対して

$$0 < p' < p \Longrightarrow \Theta(p,\,Q) \subset \Theta(p',\,Q)$$

である.
(2) $\forall Q > 0$ に対して

$$p > \frac{1}{2} \Longrightarrow f \text{ は連続}$$

である.

Proof. 証明は省略. $\square$

補題 4.15. $n \in \mathbb{N}\,(n \geq 2)$ とする. $\{X_1, X_2, \ldots, X_n\}$ を正則デザインとする. すなわち

$$X_j = \frac{j-1}{n} \qquad (j = 1, 2, \ldots, n)$$

とする. このとき, $\forall M \leq n-1$ に対して, $n \times M$ のデザイン行列

$$\Psi = (\varphi_j(X_i))_{1\leq i\leq n;\, 1\leq j\leq M}$$

は ORT 条件をみたす.

Proof. $j, j' \in \{1, 2, \ldots, M-1\}$ を固定する.

$$\begin{aligned} \boldsymbol{\varphi}_j &= (\varphi_j(X_1), \varphi_j(X_2), \ldots, \varphi_j(X_n))^\top, \\ \boldsymbol{\varphi}_{j'} &= (\varphi_{j'}(X_1), \varphi_{j'}(X_2), \ldots, \varphi_{j'}(X_n))^\top \end{aligned}$$

の内積

$$\boldsymbol{\varphi}_j^\top \boldsymbol{\varphi}_{j'} = \sum_{j=1}^{n} \varphi_j(X_i)\varphi_{j'}(X_i)$$

を考える. $k_j = \lfloor j/2 \rfloor$ とおく. すなわち, $j/2$ の整数部分である. ベクトル $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{a}', \boldsymbol{b}' \in \mathbb{R}^n$ を以下のように定義する.

$$\begin{aligned} \exp\left(\frac{\sqrt{-1}2\pi k_j s}{n}\right) &= a_{s+1} + \sqrt{-1}b_{s+1}, \\ \exp\left(\frac{\sqrt{-1}2\pi k_{j'} s}{n}\right) &= a'_{s+1} + \sqrt{-1}b'_{s+1} \qquad (s = 0, 1, \ldots, n-1). \end{aligned}$$

このとき

$$\frac{1}{2}\boldsymbol{\varphi}_j^\top \boldsymbol{\varphi}_{j'} \in \{\boldsymbol{a}^\top \boldsymbol{a}', \boldsymbol{b}^{\top\prime}, \boldsymbol{b}^\top \boldsymbol{b}', \boldsymbol{a}^\top \boldsymbol{b}'\}.$$

一方, $k_j \neq k_{j'}$ のとき, $\sigma \in \{-1, +1\}$ に対して

$$\sum_{s=0}^{n-1} \exp\left(\frac{\sqrt{-1}2\pi k_j s}{n}\right) \exp\left(\frac{\sqrt{-1}2\pi k_{j'} s}{n}\right) = \sum_{s=0}^{n-1} \exp\left(\frac{\sqrt{-1}2\pi (k_j + \sigma k_{j'}) s}{n}\right)$$

となる. 他方

$$\begin{aligned}&\sum_{s=0}^{n-1}\exp\left(\frac{\sqrt{-1}2\pi k_j s}{n}\right)\exp\left(\frac{\sqrt{-1}2\pi k_{j'} s}{n}\right)\\&\quad=(\boldsymbol{a}+\sqrt{-1}\boldsymbol{b})^\top(\boldsymbol{a}'+\sqrt{-1}\boldsymbol{b}')\\&\quad=\boldsymbol{a}^\top\boldsymbol{a}'-\sigma\boldsymbol{b}^\top\boldsymbol{b}'+\sqrt{-1}\{\boldsymbol{b}^\top\boldsymbol{a}'+\sigma\boldsymbol{a}^\top\boldsymbol{b}'\}\end{aligned}$$

である. よって, $k_j \neq k_{j'}$ のとき

$$\boldsymbol{a}^\top\boldsymbol{a}'=\pm\boldsymbol{b}^\top\boldsymbol{b}';\quad \boldsymbol{b}^\top\boldsymbol{a}'=\pm\boldsymbol{a}^\top\boldsymbol{b}'$$

となる. 次に, $k_j = k_{j'}$ のときを考える. このとき

$$\sum_{s=0}^{n-1}\exp\left(\frac{\sqrt{-1}\,2\pi k_j s}{n}\right)\exp\left(\frac{\sqrt{-1}\,2\pi k_j s}{n}\right)=\begin{cases}0 & (\sigma=1),\\ n & (\sigma=-1)\end{cases}$$

となることがわかる. □

4.2.3 積分 2 乗誤差

$\boldsymbol{\beta}\in\Theta(p,Q)$ のとき,

$$a_j^2\beta_j \to 0\,(j\to 0)$$

なので,

$$|\beta_j| = o(j^{-p}).$$

したがって, β_j は多項式のスピードで　0 に減衰するので, 固定した M に対して, f の打ち切りフーリエ列

$$\sum_{j=1}^{M}\beta_j\varphi_j =: \varphi_{\boldsymbol{\beta}}^M$$

で f を近似するのは妥当であろう. 打ち切り $M\to\infty$ のとき, 誤差が 0 に収束するからである. 誤差の収束のレートに興味がある. Sobolev 族に f が属していると仮定すると, 誤差は打ち切り M と滑らかさの程度 p で制御されることがわかる.

補題 4.16. 任意の正の整数 M と $f\in\Theta(p,Q)\,(p>1/2)$ に対して, 以下が成立する.

$$\|\varphi_{\boldsymbol{\beta}}^M - f\|_{L_2([0,1])}=\sum_{j>M}|\beta_j|^2\le QM^{-2p}. \tag{4.29}$$

特に, $M = n - 1$ のとき

$$\left|\varphi_{\boldsymbol{\beta}}^{M} - \boldsymbol{f}\right|_2 \leq 2n \left(\sum_{j>M} |\beta_j|^2\right)^2 \precsim Qn^{2-2p} \tag{4.30}$$

となる. ただし

$$\begin{aligned} \varphi_{\boldsymbol{\beta}}^{M} &= (\varphi_{\boldsymbol{\beta}}^{M}(X_1),\, \varphi_{\boldsymbol{\beta}}^{M}(X_2),\, \ldots, \varphi\varphi_{\boldsymbol{\beta}}^{M}(X_n))^\top, \\ \boldsymbol{f} &= (f(X_1),\, f(X_2),\, \ldots,\, f(X_n))^\top \end{aligned}$$

である.

Proof. 任意の $\boldsymbol{\beta} \in \Theta(p,\, Q)$ に対して, $p > 1/2$ のとき

$$\begin{aligned} \sum_{j=2}^{\infty} |\beta_j| &= \sum_{j=2}^{\infty} a_j |\beta_j| \frac{1}{a_j} \leq \sqrt{\sum_{j=2}^{\infty} a_j^2 \beta_j^2} \sqrt{\sum_{j=2}^{\infty} \frac{1}{a_j^2}} \qquad \text{(Cauchy-Schwarz の不等式)} \\ &\leq \sqrt{Q \sum_{j=1}^{\infty} \frac{1}{j^{2p}}} < \infty \end{aligned}$$

である. $\{\varphi_j \ {}_{j\geq 1}$ は $L_2([0,\, 1])$ の正規直交系なので

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^M} \|\varphi_{\boldsymbol{\theta}} - \boldsymbol{f}\|_{L_2} = \|\varphi_{\boldsymbol{\beta}} - \boldsymbol{f}\|_{L_2}$$

である. さらに, $\boldsymbol{\beta} \in \Theta(p,\, Q)$ のとき

$$\sum_{j>M} |\beta_j|^2 = \sum_{j>M} a_j^2 |\beta_j|^2 \frac{1}{a_j^2} \leq \frac{1}{a_{M+1}^2} Q \leq \frac{Q}{M^{2p}}$$

である. 2 番目の主張を証明するために

$$\left|\varphi_{\boldsymbol{\beta}} - \boldsymbol{f}\right|_2 = \big| \sum_{j \geq n2} \beta_j \varphi_j \big|_2 \leq 2\sqrt{2n} \sum_{j \geq n} |\beta_j$$

に注意する. 最後の不等式は, いかなるデザイン $X_1,\, X_2,\, \ldots,\, X_n$ に対しても

$$|\varphi_j|_2 \leq \sqrt{2n} \qquad (j \geq 1)$$

であることを用いた. $\boldsymbol{\beta} \in \Theta(p,\, Q)$ のとき

$$\begin{aligned} \sum_{j \geq n} |\beta_j| \sum_{j \geq n} a_j |\beta_j| \frac{1}{a_j} &\leq \sqrt{\sum_{j \geq n} a_j^2 \beta_j^2} \sqrt{\sum_{j \geq n} \frac{1}{a_j^2}} \precsim \sqrt{Q} n^{1/2-p}. \\ &\left(\because \sum_{j \geq n}^{\infty} \frac{1}{j^{2p}} \approx \int_n^{\infty} x^{-2p}\, \mathrm{d}x = [x^{-2p+1}]_n^{\infty} = n^{-2p+1} \right) \end{aligned}$$

である. 一方, $f \in \Theta(p, Q)$ に対して

$$\beta_j^* = \int_0^1 f(x)\varphi_j(x)\,\mathrm{d}x \qquad (j = 1, 2, \ldots)$$

であった. いま

$$\varphi_{\boldsymbol{\beta}^*} := \varphi_{\boldsymbol{\beta}^*}^M = \sum_{j=2}^M \beta_j^* \varphi_j$$

とする. $\varphi_{\boldsymbol{\beta}^*}^M$ を推定するために, 最小 2 乗推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}$ を用いる.

$$\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} \in \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^M} \sum_{j=1}^n \{Y_j - \varphi_{\boldsymbol{\beta}}(X_j)\}^2.$$

定理 4.3 より, MSE に対するオラクル不等式は, $\forall \delta > 0$ に対して,

$$\mathsf{Pr}\left(\left|\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}^M - \boldsymbol{f}\right|_{2,n}^2 \le \inf_{\boldsymbol{\beta} \in \mathbb{R}^M} \left|\varphi_{\boldsymbol{\beta}} - \boldsymbol{f}\right|_{2,n}^2 + C\sigma^2 \log\left(\frac{1}{\delta}\right) \right) \ge 1 - \delta$$

であった. ただし, $C > 0$ は定数である. これより

$$\begin{aligned}
\left|\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}^M - \varphi_{\boldsymbol{\beta}}^M\right|_{2,n}^2 &= \left|\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}^M - \boldsymbol{f}\right|_{2,n}^2 + \left|\varphi_{\boldsymbol{\beta}^*}^M - \boldsymbol{f}\right|_{2,n}^2 - 2\big(\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}^M - \boldsymbol{f}\big)^\top \big(\varphi_{\boldsymbol{\beta}^*}^M - \boldsymbol{f}\big) \\
&\le 2\left|\varphi_{\boldsymbol{\beta}^*}^M - \boldsymbol{f}\right|_{2,n}^2 - -2\big(\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}^M - \boldsymbol{f}\big)^\top \big(\varphi_{\boldsymbol{\beta}^*}^M - \boldsymbol{f}\big) + C\sigma^2 \log\left(\frac{1}{\delta}\right) \\
&= 2\big(\varphi_{\boldsymbol{\beta}^*}^M - \varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}^M\big)^\top \big(\varphi_{\boldsymbol{\beta}^*}^M - \boldsymbol{f}\big) + C\sigma^2 \log\left(\frac{1}{\delta}\right) \\
&= 2\big(\varphi_{\boldsymbol{\beta}^*}^M - \varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}^M\big)^\top \left(\sum_{j>M} \beta_j^* \varphi_j\right) + C\sigma^2 \log\left(\frac{1}{\delta}\right) \\
&= 2\big(\varphi_{\boldsymbol{\beta}^*}^M - \varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\big)^\top \left(\sum_{j>M} \beta_j^* \varphi_j\right) + C\sigma^2 \log\left(\frac{1}{\delta}\right)
\end{aligned}$$

となる. 最後の等号は補題 4.15 よりわかる. すなわち, $j = j'$ のとき

$$\varphi_j^\top \varphi_{j'} = 0.$$

(4.29) と Young の不等式より

$$2ab \le \alpha a^2 + \frac{b^2}{\alpha} \qquad (a, b \in \mathbb{R}, \alpha > 0)$$

より

$$\begin{aligned}
2\big(\varphi_{\boldsymbol{\beta}^*}^M - \varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\big)^\top \left(\sum_{j>M} \beta_j^* \varphi_j\right) &\le \alpha\left|\varphi_{\boldsymbol{\beta}^*}^M - \varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right|_2^2 + \frac{C}{\alpha} \sum_{j \ge 1} (\beta_j^*)^2 \varphi_j^2 \\
&= \alpha\left|\varphi_{\boldsymbol{\beta}^*}^M - \varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}\right|_2^2 + \frac{C}{\alpha} Q n^{2-2p}.
\end{aligned}$$

が成り立つ. 最後の等号は

$$\sum_{j\geq n}\beta_j^* \approx \int_n^\infty x^{-2p}\,\mathrm{d}x \approx n^{-2p+1}, |\varphi_j^2 \ \leq 2n, \boldsymbol{\beta}^* \in \Theta(p, Q)$$

よりわかる. 以上を合わせると

$$\begin{aligned}\left|\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}} - \varphi_{\boldsymbol{\beta}^*}\right|_2^2 &\leq 2\left(\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}} - \varphi_{\boldsymbol{\beta}^*}\right)^\top\left(\sum_{j\geq n}\beta_j^*\varphi_j\right) + C\sigma^2\log\left(\frac{1}{\delta}\right) \\ &\leq \alpha\left|\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}} - \varphi_{\boldsymbol{\beta}^*}\right|_2^2 + \frac{C}{\alpha}Qn^{n-2p} + C\sigma^2\log\left(\frac{1}{\delta}\right)\end{aligned}$$

なので, $0 < \alpha < 1$ に対して

$$(1-\alpha)\left|\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}} - \varphi_{\boldsymbol{\beta}^*}\right|_2^2 \leq \frac{C}{\alpha}Qn^{2-2p} + C\sigma^2 M\log\left(\frac{1}{\delta}\right)$$

である. よって

$$\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}^*}\right|_{2,n}^2 \leq \frac{C}{\alpha(1-\alpha)}Qn^{2-2p} + \frac{\sigma^2 M}{1-\alpha}\log\left(\frac{1}{\delta}\right) \tag{4.31}$$

がわかる. ORT 条件をみたすので

$$\begin{aligned}\left|\boldsymbol{\varphi}^M_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}} - \boldsymbol{\varphi}_{\boldsymbol{\beta}^*}\right|_{2,n}^2 &= \int_0^1\Big\{\sum_{j=2}^M\widehat{\beta}_j^{\mathrm{LS}}\varphi_j(x) - \sum_{j=2}^M\beta_j^*\varphi_j(x)\Big\}^2\,\mathrm{d}x \\ &= \sum_{j=1}^M(\widehat{\beta}_j^{\mathrm{LS}} - \beta_j^*)^2\end{aligned}$$

である. 一方, ORT 条件をみたすので

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n\varphi_j(X_i)\varphi_j(X_i) &= 1, \\ \frac{1}{n}\sum_{i=1}^n\varphi_j(X_i)\varphi_{j'}(X_i) &= 0 \qquad (j\neq j')\end{aligned}$$

である. したがて

$$\begin{aligned}\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}} - \boldsymbol{\varphi}^M_{\boldsymbol{\beta}^*}\right|_{2,n}^2 &= \left|\sum_{j=1}^M(\widehat{\beta}_j^{\mathrm{LS}} - \beta_j^*)\boldsymbol{\varphi}_j\right|_{2,n}^2 = \sum_{i=1}^n\left(\sum_{j=1}^M(\widehat{\beta}_j^{\mathrm{LS}} - \beta_j^*)\varphi_j(X_i)\right)^2 \\ &= \sum_{j=1}^M\sum_{j'=1}^M\big(\widehat{\beta}_j^{\mathrm{LS}} - \beta_j^*\big)\big(\widehat{\beta}_{j'}^{\mathrm{LS}} - \beta_{j'}^*\big)\sum_{i=1}^n\varphi_j(X_i)\varphi_{j'}(X_i) \\ &= \sum_{j=1}^M\sum_{j'=1}^M\big(\widehat{\beta}_j^{\mathrm{LS}} - \beta_j^*\big)\big(\widehat{\beta}_{j'}^{\mathrm{LS}} - \beta_{j'}^*\big)\delta_{j,j'} = n\sum_{j=1}^M\big(\widehat{\beta}_j^{\mathrm{LS}} - \beta_j^*\big)^2\end{aligned}$$

となる．よって

$$\left|\boldsymbol{\varphi}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}-\boldsymbol{\varphi}_{\boldsymbol{\beta}^*}\right|_{2,n}^2=n\|\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}-\varphi_{\boldsymbol{\beta}^*}^M\|_{L_2}^2$$

がわかる． □

命題 4.17. $p\geq(1+\sqrt{5})/4$, $Q>0$ とする．回帰モデル (4.1) を仮定し，$\epsilon\sim\mathsf{subG}_n(\sigma^2)\,(0<\sigma\leq 1)$ とする．さらに，$M=\lceil n^{1/(2p+1)}\rceil$ とし，n は十分大きく $M\leq n-1$ とする．最小 2 乗推定量

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} &\in \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^M}\frac{1}{n}\sum_{j=1}^n\big\{Y_j-\varphi_{\boldsymbol{\beta}}^M(X_i)\big\}^2,\\
\varphi_{\boldsymbol{\beta}}^M(x) &= \sum_{k=1}^M\beta_k\varphi_k(x),\\
\varphi_k(x) &= \begin{cases}\sqrt{2}\cos(\pi kx) & (j\ \text{は偶数}),\\ \sqrt{2}\sin(\pi(k-1)x) & (j\ \text{は奇数}).\end{cases}
\end{aligned}$$

で定義する．このとき，最小 2 乗推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}$ は，$\forall\delta>0$ に対して

$$\mathsf{Pr}\left(\left\|\varphi_{\widehat{\boldsymbol{\beta}}}^M-f\right\|_{L_2([0,1])}\precsim n^{-2p/(2p+1)}\big(1+\sigma^2\log(1/\delta)\big)\right)\geq 1-\delta$$

をみたす．

Proof. たとえば，$\alpha=1/2$ とし，(4.31) と補題 4.15 を用いると，$M\leq n-1$ のとき，

$$\left\|\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}^M-\varphi_{\boldsymbol{\beta}^*}^M\right\|_{L_2([0,1])}\precsim n^{1-2p}+\sigma^2\frac{M\log(1/\delta)}{n}.$$

補題 4.16 と $\sigma^2\leq 1$ を用いると

$$\begin{aligned}
\left\|\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}^M-f\right\|_{L_2([0,1])} &\leq 2\left\|\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}^M-\varphi_{\boldsymbol{\beta}^*}^M\right\|_{L_2([0,1])}+2\left\|\varphi_{\boldsymbol{\beta}^*}-f\right\|_{L_2([0,1])}\\
&\precsim M^{-2p}+n^{1-2p}+\sigma^2\frac{M\log(1/\delta)}{n}
\end{aligned}$$

となる．ここで，$M=\lceil n^{1/(2p+1)}\rceil\leq n-1$ になるように n を十分大きく取れば

$$\left\|\varphi_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}}^M-f\right\|_{L_2([0,1])}\precsim n^{-2p/(2p+1)}+n^{1-2p}\sigma^2\log\left(\frac{1}{\delta}\right)$$

となる．最後に，$p\geq(1+\sqrt{5})/4$ であることに注意すれと

$$\begin{aligned}
n^{1-2p}<n^{-2p/(2p+1)} &\iff 1-2p<-\frac{2p}{2p+1}\iff(1-2p)(2p+1)<-p\\
&\iff 4p^2-2p-1>0\iff p>\frac{1+\sqrt{5}}{4},\ p<\frac{1-\sqrt{5}}{4}
\end{aligned}$$

となる. だから

$$\left\|\varphi^{M}_{\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}} - f\right\|_{L_2([0,1])} \lesssim n^{-2p/(2p+1)}\left(1 + \sigma^2 \log\left(\frac{1}{\delta}\right)\right)$$

を得る. □

4.3 章末注釈と参考文献

この章は [3] を参考にした.

4.4 演習問題

演習問題 4.1.

演習問題 4.1 の解答 □