

Sufficient conditions for the oracle property in penalized linear regression

Sunghoon Kwon^a, Hyeseong Moon^a, Jaeho Chang^a, Sangin Lee^{1,b}

^aDepartment of Applied Statistics, Konkuk University;

^bDepartment of Information and Statistics, Chungnam National University

Abstract

In this paper, we introduce how to construct sufficient conditions for the oracle property in penalized linear regression model. We give formal definitions of the oracle estimator, penalized estimator, oracle penalized estimator, and the oracle property of the oracle estimator. Based on the definitions, we present a unified way of constructing optimality conditions for the oracle property and sufficient conditions for the optimality conditions that covers most of the existing penalties. In addition, we present an illustrative example and results from the numerical study.

Keywords: penalized estimator, oracle penalized estimator, oracle property

1. 서론

별점 추정법(penalized estimation)은 주어진 손실함수(loss function)에 별점함수(penalty function)를 더하여 별점 손실함수(penalized loss function)를 정의하고 이 별점 손실함수를 최소화하여 별점 추정량(penalized estimator)을 정의하는 방법이다. 예를 들어 선형 회귀모형(linear regression model)에서 잔차 제곱합(sum of squared residuals)에 LASSO 별점함수(Tibshirani, 1996)를 더하여 LASSO 별점 잔차제곱합(penalized sum of squared residuals; LASSO)을 정의하고 이를 최소화하면 LASSO 별점 추정량을 정의할 수 있다. 별점 추정법은 회귀모형의 최적 부분모형을 구성하는 연구(Tibshirani, 1996; Fan과 Li, 2001)를 시작으로 이후 일반화 선형 모형(generalized linear model)의 최적 부분모형(Fan과 Lv, 2011; Kwon과 Kim, 2012), 회귀모형의 그룹 구조(group structure) (Yuan과 Lin, 2006; Shen과 Huang, 2010; Kwon 등, 2017), 시계열 자료의 평활(trend filtering) (Tibshirani와 Taylor 2011), 다변량 자료의 네트워크 구축(network construction) (Yuan과 Lin, 2007), 주성분 및 군집 분석(Zou 등, 2006; Pan 등, 2013) 등 다양한 분야에서 활발하게 연구되고 있다.

본 논문은 선형 회귀모형에서 별점 추정량의 신의 성질(oracle property)을 소개하고 신의 성질을 가지기 위한 충분조건을 유도하는 방법에 대하여 소개한다. 독자의 이해를 돕기 위하여 신의 추정량(oracle estimator), 신의 성질, 신의 성질을 정의하는 최적조건(optimality condition), 신의 성질을 만족하기 위한 충분조건(sufficient condition)을 유도하는 과정을 일반적인 원리로 설명하였으며 다양한 예제를 사용하여 확인해 볼 수 있도록 하였다.

This paper was supported by Konkuk University Researcher Fund in 2019, and the National Research Foundation of Korea (NRF) grant (No. 2020R111A3071646).

This paper includes a part of Hyeseong Moon's master thesis.

¹ Corresponding author: Department of Information and Statistics, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Korea. E-mail: sanginlee44@gmail.com

신의 성질은 별점 추정법을 통해 신의 추정량을 구현할 수 있는가를 결정하는 성질이다. 특히 별점함수가 비볼록 함수(nonconvex function)인 경우 신의 추정량이 강한 신의 성질(strong oracle property)을 가지지 않으면 신의 추정량을 구현할 수 없다는 단점을 가진다. 따라서 신의 추정량이 강한 신의 성질을 가질 수 있도록 충분조건을 구성하고 이러한 충분조건이 점근적으로 성립하는 것을 증명하는 것이 비볼록 별점 추정법에서 매우 중요한 연구 주제 중 하나가 된다.

신의 성질에 대한 충분조건을 유도하려면 먼저 신의 성질을 정의하는 최적조건을 이해해야 한다. 특히 Lv와 Fan (2009)는 알려진 대부분의 별점함수에 적용할 수 있는 최적조건을 제시하고 있으며 본 논문에서는 이 최적조건을 차용하여 신의 성질에 대한 충분조건을 유도하였다. 이 최적조건이 별점함수 각각의 특징을 반영한 최선의 최적조건은 아니지만 (Huang 등, 2008; Kim과 Kwon, 2012; Kwon 등, 2015) 신의 성질에 대한 원리를 이해하기에 충분한 최적조건이다. 본 논문의 내용은 신의 성질을 연구한 기존의 연구들을 (Fan과 Li, 2001; Huang 등, 2008; Kim 등, 2008; Kim과 Kwon, 2012; Kwon 등, 2013, 2016) 하나의 원리로 이해할 수 있도록 구성하였으며 새로운 별점함수를 개발하거나 기존의 별점함수를 응용하는데 큰 도움이 될 것으로 판단된다.

본 논문은 다음과 같이 구성하였다. 2장에서는 별점 추정량, 신의 별점 추정량, 신의 성질을 정의하였으며 3장에서는 최적조건을 사용하여 신의 성질을 가지기 위한 충분조건을 유도하는 원리를 소개하였다. 4장에서는 간단한 예제와 가상 실험을 진행하고 그 결과를 소개한다. 5장에서는 본 논문에서 다루지 못한 몇 가지 주제에 대하여 간단히 정리하였다.

2. 별점 추정량과 신의 성질

먼저 논의에 필요한 모형을 소개한다.

$$y = X\beta^* + \varepsilon = X_{\mathcal{A}}\beta_{\mathcal{A}}^* + X_{\mathcal{A}^c}\beta_{\mathcal{A}^c}^* + \varepsilon = X_{\mathcal{A}}\beta_{\mathcal{A}}^* + \varepsilon. \quad (2.1)$$

단, $X = (X_1, \dots, X_p)$ 는 n 개의 자료와 p 개의 변수로 이루어진 $n \times p$ 크기의 디자인 행렬, $\beta^* = (\beta_1^*, \dots, \beta_p^*)^T$ 는 X 에 대응하는 회귀계수 벡터, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ 는 $E(\varepsilon_1) = 0$ 이고 $\text{var}(\varepsilon_1) = \sigma^2 < \infty$ 를 만족하는 서로 독립이고 동일한 분포를 따르는 임의오차(random error) 벡터이며 X 와 독립이다. $\mathcal{A} = \{j | \beta_j^* \neq 0\}$ 는 모형에 필요한 변수의 첨자(index) 집합이고 $X_{\mathcal{A}} = (X_j, j \in \mathcal{A})$ 와 $\beta_{\mathcal{A}}^* = (\beta_j^*, j \in \mathcal{A})$ 는 모형에 필요한 변수로 구성된 디자인 행렬과 회귀계수 벡터이다. 이하의 내용은 $p < n$ 을 만족하는 저차원 회귀모형을 다루지만 대부분의 내용은 $p > n$ 인 고차원 모형에서도 그대로 성립하며 일부는 약간의 수정을 통해 고차원 모형에 적용할 수 있다.

임의의 함수 $f(x)$ 에 대하여 \hat{x} 가 $f(x)$ 의 국소 최소해이면(local minimizer)이면 $\hat{x} = \text{arglocalmin}_x f(x)$ 로 표기하고 유일한 광역 최소해(unique global minimizer)이면 $\hat{x} = \text{argmin}_x f(x)$ 로 표기한다. 이러한 표기가 필요한 이유는 별점함수(penalty function)가 비볼록 함수(nonconvex function)인 경우 별점 추정량(penalized estimator)이 국소 최소해가 될 수도 있고 유일한 광역 최소해가 될 수도 있기 때문이다.

2.1. 별점 추정량과 변수 선택

모형 (2.1)에서 분석의 목적은 첨자 집합 \mathcal{A} 를 찾고 이에 대응하는 회귀계수 $\beta_{\mathcal{A}}$ 를 추정하는 것이다. 이러한 분석 목적을 변수 선택(variable selection) 혹은 모형 선택(model selection)이라고 하며 별점 추정법을 적용할 수 있는 가장 대표적인 주제이다. 모형 (2.1)에 대하여 별점 잔차제곱합(penalized sum of squared residuals)의 국소 최소해를 별점 추정량으로 정의한다.

$$\hat{\beta}^{\dagger} = \text{arglocalmin}_{\beta} Q_{\lambda}(\beta). \quad (2.2)$$

단, $Q_\lambda(\beta) = L(\beta) + \sum_{j=1}^p J_\lambda(|\beta_j|)$, $L(\beta) = \|y - X\beta\|^2/2n$, $J_\lambda(t)$, $t > 0$ 는 조율모수(tuning parameter)가 $\lambda > 0$ 인 벌점함수이다. 만약 벌점함수가 볼록 함수(convex function)이면 벌점 추정량은 유일한 광역 최소해가 된다.

$$\hat{\beta}^\lambda = \operatorname{argmin}_\beta Q_\lambda(\beta).$$

논의의 편의상 벌점함수가 다음의 성질을 만족한다고 가정한다.

(J1) $J_\lambda(t)$, $t > 0$ 는 증가하며 미분 가능하다.

(J2) $J_\lambda(|t|)$, $t \in \mathbb{R}$ 는 원점에서 미분 불가능하다.

Remark 1. (J1)은 벌점함수의 대략적 형태를 정의하며 대부분의 벌점함수가 만족하는 성질이다. 하지만 벌점함수가 반드시 증가할 필요는 없으며(Kwon 등, 2013) 유한개의 점에서 미분 불가능한 경우도 있다 (Choi와 Park, 2012; Shen 등, 2012). (J2)는 벌점 추정량이 성김성(sparsity)을 갖기 위한 충분조건이며 (Fan과 Li, 2001) 원점에서 미분 가능한 벌점함수는 일반적으로 성김성을 갖지 않는다 (Hoerl과 Kennard, 1970).

변수 선택을 위하여 벌점 추정법을 사용하는 이유는 벌점 추정량이 성김성을 갖기 때문이다.

$$\exists S \text{ s.t. } \hat{\beta}_j^\lambda = 0, \quad j \in S.$$

즉, 모형에 필요하지 않은 변수에 대응하는 회귀계수가 정확하게 0으로 추정되기 때문에 검정(test)과 같은 추가적인 작업을 하지 않아도 벌점 추정량으로부터 직접 \mathcal{A} 를 추정할 수 있게 된다.

$$\hat{\mathcal{A}} = S^c = \{j | \hat{\beta}_j^\lambda \neq 0\}.$$

2.2. 신의 벌점 추정량

만약 \mathcal{A} 를 알고 있다면 $X_{\mathcal{A}}$ 만 사용하여 모형을 적합하는 것이 변수 선택의 측면에서 가장 이상적이며 이러한 추정량을 신의 추정량(oracle estimator)이라고 한다. 신의 추정량은 주로 벌점 추정량의 이론적 성질을 연구하기 위해 사용되며 가장 대표적인 신의 추정량은 다음의 신의 LS 추정량(oracle least square estimator)이다 (Fan과 Li, 2001; Kim 등, 2008; Zhang, 2010).

$$\hat{\beta}^{\text{or,LS}} = \operatorname{argmin}_{\beta_j=0, j \in \mathcal{A}^c} L(\beta). \quad (2.3)$$

반면, 벌점 추정법을 사용하여 얻을 수 있는 신의 추정량을 신의 벌점 추정량(oracle penalized estimator)이라고 한다 (Um 등, 2020).

$$\hat{\beta}^{\text{or},\lambda} = \operatorname{arglocalmin}_{\beta_j=0, j \in \mathcal{A}^c} Q_\lambda(\beta). \quad (2.4)$$

신의 벌점 추정량은 제한된 모수 공간 $\mathbb{R}_{\mathcal{A}}^p = \{\beta | \beta_j = 0, j \in \mathcal{A}^c\}$ 에 포함되므로 $\hat{\beta}_j^{\text{or},\lambda} = 0, j \in \mathcal{A}^c$ 을 만족한다. 따라서 신의 벌점 추정량이 다음을 만족하는 경우 변수 선택의 측면에서 최선의 벌점 추정량이다.

$$\hat{\beta}_j^{\text{or},\lambda} \neq 0, \quad j \in \mathcal{A}. \quad (2.5)$$

Remark 2. 신의 벌점 추정량도 벌점 추정량 중 하나이므로 성김성을 가진다. 따라서 (2.5)를 만족하는 신의 벌점 추정량이 존재하는지가 벌점 추정법에서 다루어야 하는 중요한 문제 중 하나이다. 하지만 대부분의 벌점 함수에 대하여 (2.5)를 만족하는 신의 벌점 추정량이 존재하는 것으로 알려져 있으며 (Fan과 Li, 2001; Fan과 Peng, 2004; Huang 등, 2008; Kwon과 Kim, 2012; Kwon 등, 2016), 따라서 이 논문에서는 신의 벌점 추정량이 (2.5)를 만족한다고 가정한다.

Remark 3. 별점 추정량의 정의 (2.2)에서 사용된 별점함수와 신의 별점 추정량의 정의 (2.4)에서 사용된 별점함수가 반드시 같을 필요는 없으며 이 문제는 뒤에서 자세히 소개한다.

몇 가지 예제를 통해 신의 별점 추정량을 구체적으로 표현해 보면 다음과 같다.

Example 1. (신의 TLP 추정량) 별점함수가 TLP (Shen 등, 2013), $J_\lambda(t) = (\lambda/\tau) \min\{t, \tau\}$, $t > 0$, $\tau > 0$ 인 경우 신의 TLP 추정량은 다음과 같이 정의된다.

$$\hat{\beta}^{\text{or},\lambda} = \text{arglocalmin}_{\beta_j=0, j \in \mathcal{A}^c} \left\{ \frac{\|y - X\beta\|^2}{2n} + \sum_{j=1}^p \left(\frac{\lambda}{\tau} \right) \max\{\tau, |\beta_j|\} \right\}.$$

신의 TLP 추정량은 다음과 같이 매우 특별한 성질을 가진다(Shen 등, 2013).

$$\min_{j \in \mathcal{A}} |\hat{\beta}_j^{\text{or},\lambda}| > \tau \Rightarrow \hat{\beta}_{\mathcal{A}}^{\text{or},\lambda} = \hat{\beta}_{\mathcal{A}}^{\text{or,LS}}.$$

즉, 추정된 회귀계수의 절대값이 조율모수 τ 보다 큰 경우 신의 TLP 추정량이 별점함수를 사용하지 않은 신의 LS 추정량과 정확하게 일치한다는 것이다. 따라서 신의 TLP 추정량은 별점함수로 인한 축소 효과(shrinkage effect)를 가지지 않으며 불편성(unbiasedness)과 같은 신의 LS 추정량의 이론적 성질을 그대로 가지게 된다.

Example 2. (신의 ridge 추정량) 별점함수가 ridge (Hoerl과 Kennard, 1970), $J_\lambda(t) = \lambda t^2$, $t > 0$ 인 경우 신의 ridge 추정량 (Huang 등, 2016; Kwon 등, 2013)은 다음과 같이 정의된다.

$$\hat{\beta}^{\text{or},\lambda} = \text{arglocalmin}_{\beta_j=0, j \in \mathcal{A}^c} \left\{ \frac{\|y - X\beta\|^2}{2n} + \sum_{j=1}^p \lambda \beta_j^2 \right\}.$$

신의 ridge 추정량은 별점함수로 인해 축소 효과(shrinkage effect)를 가지는 편향 추정량(biased estimator)이다. 하지만 추정량의 분산이 신의 LSE 추정량보다 작기 때문에 다중공선성(multicollinearity)이 발생할 경우 신의 LS 추정량 대신 고려할 수 있다. 참고로 ridge 별점함수는 성김성을 갖지 않기 때문에 신의 ridge 추정량은 ridge 별점함수를 사용하여 구현할 수 없으며 다른 별점함수를 사용해야 구현할 수 있다는 특징을 가진다. 이 부분도 뒤에서 좀 더 자세하게 소개하도록 한다.

2.3. 약한 신의 성질과 강한 신의 성질

별점 추정법에서 가장 중요한 문제는 별점 추정법을 사용하여 신의 별점 추정량을 구할 수 있는지의 여부이다. 이 문제에 대한 연구 결과는 크게 약한 신의 성질(weak oracle property)과 강한 신의 성질(strong oracle property)로 구분되어 연구되고 있으며 각각의 정의는 다음과 같다. 만약 신의 별점 추정량이 다음을 만족하면 약한 신의 성질을 가진다고 말한다.

$$\hat{\beta}^{\text{or},\lambda} = \text{arglocalmin}_\beta Q_\lambda(\beta). \quad (2.6)$$

또한 신의 별점 추정량이 다음의 성질을 만족하면 강한 신의 성질을 가진다고 말한다.

$$\hat{\beta}^{\text{or},\lambda} = \text{argmin}_\beta Q_\lambda(\beta). \quad (2.7)$$

신의 별점 추정량이 약한 신의 성질을 가지면 $Q_\lambda(\beta)$ 의 국소 최소해 중 하나가 되며 (Fan과 Li, 2001; Fan과 Peng, 2004) 강한 신의 성질을 가지면 $Q_\lambda(\beta)$ 의 유일한 광역 최소해가 된다(Kim과 Kwon, 2012; Zhang, 2010). 따라서 식(2.6)과 (2.7)은 각각 신의 별점 추정량의 존재성(existence)과 유일성(uniqueness)을 의미하며 신의

성질의 성립 여부에 따라 별점 추정법을 통해 신의 별점 추정량을 찾을 수 있는지의 여부가 결정된다. 만약 강한 신의 성질이 성립하면 별점함수가 비볼록 함수인 경우에도 알고리즘이나 초기값의 종류에 상관없이 항상 신의 별점 추정량을 구현할 수 있으므로 별점 추정법에서 가장 중요한 성질은 강한 신의 성질임을 알 수 있다.

Remark 4. 신의 성질은 $\hat{\beta}^\lambda = \hat{\beta}^{\text{or},\lambda}$ 의 관계를 만족하는 조율모수 $\lambda = \lambda^{\text{or}}$ 이 존재한다는 것을 의미하지만 λ^{or} 을 어떻게 찾을 수 있는지에 대하여 알려 주는 것은 아니다. 조율모수의 선택은 별점 추정법에서 다루는 또 다른 중요한 문제이며 이 논문에서는 다루지 않는다. 다만 교차 타당성(cross validation)과 같은 예측 중심의 조율모수 선택 기법을 사용할 경우 λ^{or} 을 찾을 수 없음이 알려져 있고 (Wang 등, 2007), Bayesian Information Criterion (BIC)와 유사한 측도를 사용하면 찾을 수 있다는 것이 알려져 있다 (Wang 등, 2007, 2009).

3. 신의 성질을 갖기 위한 충분조건

이 장에서는 신의 별점 추정량이 앞에서 소개한 두 가지 신의 성질을 가지기 위한 충분조건을 소개한다. 특히 별점 추정량을 정의할 때 사용되는 별점함수와 신의 별점 추정량을 정의할 때 사용되는 별점함수가 서로 같은 경우뿐만 아니라 서로 다른 경우도 고려하여 신의 성질에 대한 이해를 돕고자 한다.

3.1. 신의 별점 추정량의 최적조건

신의 LS 추정량은 다음의 최적조건(optimality condition)을 만족한다.

$$\hat{\beta}^{\text{or},\text{LS}} = \operatorname{argmin}_{\beta_j=0, j \in \mathcal{A}^c} L(\beta) \Leftrightarrow \frac{\partial L(\hat{\beta}^{\text{or},\text{LS}})}{\partial \beta_j} = 0, \quad j \in \mathcal{A}. \quad (3.1)$$

이러한 성질은 미분 가능한 볼록 함수의 광역 최소해가 경사도(gradient) 혹은 미분(derivative)을 0으로 만드는 원리를 표현한 것이다. 이와 마찬가지로 신의 별점 추정량이 신의 성질을 만족하기 위한 최적조건을 찾을 수 있다 (Lv와 Fan, 2009).

$$\hat{\beta}^{\text{or},\lambda} = \operatorname{arglocalmin}_{\beta} Q_{\lambda}(\beta) \Leftrightarrow \begin{cases} \frac{\partial L(\hat{\beta}^{\text{or},\lambda})}{\partial \beta_j} = -\operatorname{sign}(\hat{\beta}_j^{\text{or},\lambda}) \nabla J_{\lambda}(|\hat{\beta}_j^{\text{or},\lambda}|), & j \in \mathcal{A}, & (3.2a) \\ \left| \frac{\partial L(\hat{\beta}^{\text{or},\lambda})}{\partial \beta_j} \right| \leq \lim_{t \rightarrow 0^+} \nabla J_{\lambda}(t), & j \in \mathcal{A}^c, & (3.2b) \\ \rho_{\mathcal{A}} > \max_{j \in \mathcal{A}} \bar{\nabla}^2 J_{\lambda}(|\hat{\beta}_j^{\text{or},\lambda}|). & & (3.2c) \end{cases}$$

단, $\rho_{\mathcal{A}} = \lambda_{\min}(X_{\mathcal{A}}^T X_{\mathcal{A}}/n)$ 는 $X_{\mathcal{A}}^T X_{\mathcal{A}}/n$ 의 최소 고유값(minimum eigenvalue), $\nabla J_{\lambda}(t) = dJ_{\lambda}(t)/dt$, $t > 0$ 이고 $\bar{\nabla}^2 J_{\lambda}(t)$, $t > 0$ 는 $\nabla J_{\lambda}(t)$ 의 t 에서의 최대 오목도(maximum concavity)이다 (Zhang, 2010).

$$\bar{\nabla}^2 J_{\lambda}(t) = \lim_{\varepsilon \rightarrow 0^+} \sup_{t-\varepsilon < t_1 < t_2 < t+\varepsilon} \frac{\nabla J_{\lambda}(t_1) - \nabla J_{\lambda}(t_2)}{t_2 - t_1}, \quad t > 0.$$

참고로 별점함수가 두 번 미분 가능하면 $\bar{\nabla}^2 J_{\lambda}(t) = -\nabla^2 J_{\lambda}(t) = -d^2 J_{\lambda}(t)/dt^2$, $t > 0$ 을 만족한다. 또한 대부분의 경우 (3.2a)와 (3.2b)는 필요조건이고 (3.2c)를 추가하면 충분조건이 된다.

Remark 5. 위 최적조건은 $Q_{\lambda}(\beta)$ 의 $\hat{\beta}^{\text{or},\lambda}$ 에서의 부경사 집합(subdifferential)이 0을 포함하여 $\hat{\beta}^{\text{or},\lambda}$ 가 국소 최소해가 된다는 것을 의미한다.

• $\hat{\beta}_j^{\text{or},\lambda} \neq 0$ 인 경우 $J_{\lambda}(|\beta_j|)$ 는 $\hat{\beta}_j^{\text{or},\lambda}$ 에서 미분 가능한 함수이며 (3.2a)를 만족하면 $Q_{\lambda}(\beta)$ 의 β_j 방향 편미분(partial derivative)이 $\hat{\beta}^{\text{or},\lambda}$ 에서 0이 된다.

- $\hat{\beta}_j^{\text{or},\lambda} = 0$ 인 경우 $J_\lambda(|\beta_j|)$ 는 $\hat{\beta}_j^{\text{or},\lambda}$ 에서 미분 불가능한 함수이지만 (3.2b)를 만족하면 $L(\beta)$ 의 β_j 방향 편미분이 $\hat{\beta}_j^{\text{or},\lambda}$ 에서 $J_\lambda(|\beta_j|)$ 의 부경사가 되고 $Q_\lambda(\beta)$ 의 β_j 방향 부경사 집합이 0을 포함하게 된다.
- 조건 (3.2c)를 만족하는 경우 $Q_\lambda(\beta)$ 는 제한된 모수공간 $\mathbb{R}_{\mathcal{A}}^p = \{\beta | \beta_j = 0, j \in \mathcal{A}\}$ 의 한점 $\hat{\beta}_j^{\text{or},\lambda}$ 의 주위에서 국소적으로 엄밀한 볼록 함수(locally strictly convex function)이며 $\hat{\beta}_j^{\text{or},\lambda}$ 는 $\mathbb{R}_{\mathcal{A}}^p$ 에서 $Q_\lambda(\beta)$ 의 국소 최소해 중 하나(one of local minimizers)이다. 따라서 $\hat{\beta}_j^{\text{or},\lambda}$ 는 약한 신의 성질을 만족하게 된다 (Fan과 Li, 2001; Huang 등, 2008).
- 조건 (3.2c) 보다 강한 조건 (3.3)를 만족하는 경 $Q_\lambda(\beta)$ 는 전체 모수 공간 \mathbb{R}^p 에서 엄밀한 볼록 함수이며 $\hat{\beta}_j^{\text{or},\lambda}$ 는 \mathbb{R}^p 에서 유일한 광역 최소해가 된다 (Kim과 Kwon, 2012; Zhang, 2010).

$$\rho > \sup_{t>0} \bar{\nabla}^2 J_\lambda(t). \quad (3.3)$$

단, $\rho = \lambda_{\min}(X^T X/n)$ 이다. 따라서 $\hat{\beta}_j^{\text{or},\lambda}$ 는 강한 신의 성질을 만족한다(Kim 과 Kwon, 2012; Zhang, 2010)

3.2. 신의 성질을 갖기 위한 충분조건: 벌점함수가 동일한 경우

이 장에서는 신의 벌점 추정량과 동일한 벌점함수를 사용하여 벌점 추정량을 정의하고 앞에서 소개한 최적조건 (3.2a), (3.2b), (3.2c), (3.3)에 대한 충분조건을 구성하는 원리를 몇 가지 예제를 통해 소개한다. 먼저 신의 벌점 추정량은 다음의 최적조건을 만족한다.

$$\hat{\beta}_j^{\text{or},\lambda} = \text{arglocalmin}_{\beta_j=0, j \in \mathcal{A}^c} \left\{ L(\beta) + \sum_{j=1}^p J_\lambda(|\beta_j|) \right\} \Leftrightarrow \frac{\partial L(\hat{\beta}_j^{\text{or},\lambda})}{\partial \beta_j} = -\text{sign}(\hat{\beta}_j^{\text{or},\lambda}) \nabla J_\lambda(|\hat{\beta}_j^{\text{or},\lambda}|), \quad j \in \mathcal{A}. \quad (3.4)$$

따라서 조건 (3.4)에 의하여 조건 (3.2a)가 당연히 성립하고 $\partial L(\hat{\beta}_j^{\text{or},\lambda})/\partial \beta_j = -X_j^T(y - X_{\mathcal{A}}\hat{\beta}_{\mathcal{A}}^{\text{or},\lambda})/n$ 이므로 신의 성질을 만족하기 위한 최적조건은 다음과 같이 간략하게 요약할 수 있다.

$$\Delta_{\mathcal{A}^c} \leq \kappa, \quad \rho_{\mathcal{A}} > \phi_{\mathcal{A}}, \quad \rho > \phi. \quad (3.5)$$

단, $\Delta_{\mathcal{A}^c} = \max_{j \in \mathcal{A}^c} |X_j^T(y - X_{\mathcal{A}}\hat{\beta}_{\mathcal{A}}^{\text{or},\lambda})/n|$, $\kappa = \lim_{t \rightarrow 0^+} \nabla J_\lambda(t)$, $\phi_{\mathcal{A}} = \max_{j \in \mathcal{A}} \bar{\nabla}^2 J_\lambda(|\hat{\beta}_j^{\text{or},\lambda}|)$, $\phi = \sup_{t>0} \bar{\nabla}^2 J_\lambda(t)$ 이다. 또한 (3.5)에서 처음 두 조건을 만족하면 약한 신의 성질을 가지고 마지막 조건도 만족하면 강한 신의 성질을 가진다. 이제 몇 가지 벌점함수를 통해 신의 성질에 대한 충분조건을 구성하는 원리를 구체적으로 기술한다. $\alpha_* = \min_{j \in \mathcal{A}} |\hat{\beta}_j^{\text{or},\lambda}|$ 라 하자.

Example 3. (신의 LASSO 추정량) 벌점함수가 LASSO(Tibshirani, 1996), $J_\lambda(t) = \lambda t, t > 0$ 인 경우 벌점함수는 다음을 만족한다.

$$\nabla J_\lambda(t) = \lambda, \quad \bar{\nabla}^2 J_\lambda(t) = -\nabla^2 J_\lambda(t) = 0, \quad t > 0.$$

따라서 $\kappa = \lambda$, $\phi_{\mathcal{A}} = \phi = 0$ 이므로 최적조건 (3.5)는 다음과 동치이다.

$$\Delta_{\mathcal{A}^c} \leq \lambda, \quad \rho_{\mathcal{A}} > 0, \quad \rho > 0. \quad (3.6)$$

Remark 6. 참고로 최적조건 (3.4)와 (3.6)이 동시에 성립하려면 디자인 행렬이 Strong Irrepresentable(SI) condition과 같은 매우 특별한 성질을 만족해야 한다는 것이 알려져 있다 (Zhao와 Yu, 2006). 따라서 일반적인 조건하에서는 LASSO 벌점함수를 사용하여 신의 LASSO 추정량을 구할 수는 없으며(Zou, 2006) LASSO 벌점함수에서 조율모수에 가중치를 주거나(Zou, 2006), LASSO가 아닌 다른 벌점함수를 사용해야 한다 (Kwon 등, 2015).

Example 4. (신의 MCP 추정량) 별점함수가 MCP (Zhang, 2010), $J_\lambda(t) = \lambda \int_0^t (1 - x/(a\lambda))_+ dx$, $t > 0$, $a > 1$ 인 경우 별점함수는 다음을 만족한다.

$$\nabla J_\lambda(t) = \lambda \left(1 - \frac{t}{a\lambda}\right)_+, \quad \bar{\nabla}^2 J_\lambda(t) = \frac{1}{a} I(t \leq a\lambda), \quad t > 0.$$

따라서 $\kappa = \lambda$, $\phi_{\mathcal{A}} = (1/a)I(\alpha_* < a\lambda)$, $\phi = 1/a$ 이므로 최적조건 (3.5)는 다음과 동치이다.

$$\Delta_{\mathcal{A}^c} \leq \lambda, \quad \rho_{\mathcal{A}} > \frac{1}{a} I(\alpha_* < a\lambda), \quad \rho > \frac{1}{a}. \quad (3.7)$$

Remark 7. (신의 LS 추정량) 만약 신의 MCP 추정량이 $\alpha_* > a\lambda$ 를 만족하면 $\nabla J_\lambda(\hat{\beta}_j^{\text{or},\lambda}) = 0$, $j \in \mathcal{A}$ 이므로 (3.2a)는 (3.1)과 동치이다. 따라서 신의 MCP 추정량은 신의 LS 추정량과 동일하다. 또한 $\phi_{\mathcal{A}} = 0$ 이 되어 최적조건 (3.5)에 대한 충분조건은 다음과 같다.

$$\alpha_* > a\lambda, \quad \Delta_{\mathcal{A}^c} \leq \lambda, \quad \rho_{\mathcal{A}} > 0, \quad \rho > \frac{1}{a}. \quad (3.8)$$

위 충분조건은 추정량의 크기에 제한을 두므로 강한 신의 성질에 대하여 최적조건 (3.7)보다 조금 더 강한 조건이 된다. 이러한 성질은 MCP뿐만 아니라 TLP, SCAD (Fan과 Li, 2001)와 같이 꼬리 부분이 평평(flat)하고 최대 오목도가 유한(bounded)인 별점함수에 대하여 동일하게 성립하는 성질이다 (Fan과 Li, 2001; Kim 등, 2008; Kim 과 Kwon, 2012; Shen 등, 2013).

Example 5. (신의 bridge 추정량) 별점함수가 bridge (Huang 등, 2008), $J_\lambda(t) = \lambda t^{1/2}$, $t > 0$ 인 경우 별점함수는 다음을 만족한다.

$$\nabla J_\lambda(t) = \frac{\lambda}{2} t^{-\frac{1}{2}}, \quad \bar{\nabla}^2 J_\lambda(t) = -\nabla^2 J_\lambda(t) = \frac{\lambda}{4} t^{-\frac{3}{2}}, \quad t > 0.$$

따라서 $\kappa = \infty$, $\phi_{\mathcal{A}} = (\lambda/4)\alpha_*^{-3/2}$, $\phi = \infty$ 이므로 최적조건 (3.5)는 다음과 동치이다.

$$\Delta_{\mathcal{A}} < \infty, \quad \rho_{\mathcal{A}} > \frac{\lambda}{4} \alpha_*^{-\frac{3}{2}}, \quad \rho > \infty.$$

따라서 약한 신의 성질을 가지기 위한 충분조건은 성립할 수 있지만 세 번째 조건이 절대 성립하지 않으므로 강한 신의 성질을 가질 수 없다 (Kim 등, 2016). 이러한 성질은 log(Zou와 Li, 2008), h-likelihood (Lee와 Oh, 2014; Kwon 등, 2016), ML(Lee와 Kim, 2019) 별점함수와 같이 원점에서 미분 값이 유계가 아닌(unbounded) 경우 동일하게 성립한다.

Remark 8. 참고로 bridge 별점함수를 원점에서 미분 값이 유계가 되도록 선형 근사(linear approximation) 하면 신의 bridge 별점 추정량이 강한 신의 성질을 가지도록 할 수 있다 (Um 등, 2020). 이 부분은 다음 장에서 자세히 소개하도록 한다.

3.3. 신의 성질을 가지기 위한 충분조건: 별점함수가 다른 경우

앞에서 소개한 신의 ridge 추정량, 신의 LASSO 추정량, 신의 bridge 추정량은 동일한 별점함수를 사용한 경우 강한 신의 성질을 만족할 수 없다. 이 장에서는 신의 별점 추정량과 별점 추정량에 사용된 별점함수를 서로 다르게 정의하여 이러한 신의 별점 추정량이 신의 성질을 가질 수 있도록 개선하는 원리에 대하여 소개한다.

신의 별점 추정량이 별점함수 $J_\lambda^{\text{or}}(t)$, $t > 0$ 를 사용하여 정의되면 신의 별점 추정량은 다음의 최적조건을 만족한다.

$$\hat{\beta}^{\text{or},\lambda} = \text{arglocalmin}_{\beta_j=0, j \in \mathcal{A}^c} \left\{ L(\beta) + \sum_{j=1}^p J_\lambda^{\text{or}}(|\beta_j|) \right\} \Leftrightarrow \frac{\partial L(\hat{\beta}^{\text{or},\lambda})}{\partial \beta_j} = -\text{sign}(\hat{\beta}_j^{\text{or},\lambda}) \nabla J_\lambda^{\text{or}}(|\hat{\beta}_j^{\text{or},\lambda}|), \quad j \in \mathcal{A}. \quad (3.9)$$

만약 별점 추정량에 사용되는 별점함수가 $J_\gamma(t)$, $t > 0$, $\gamma > 0$ 이면 신의 별점 추정량이 신의 성질을 만족하기 위한 최적조건은 다음과 같다.

$$\hat{\beta}^{\text{or},\lambda} = \text{arglocalmin}_\beta Q_\gamma(\beta) \Leftrightarrow \begin{cases} \frac{\partial L(\hat{\beta}^{\text{or},\lambda})}{\partial \beta_j} = -\text{sign}(\hat{\beta}_j^{\text{or},\lambda}) \nabla J_\gamma(|\hat{\beta}_j^{\text{or},\lambda}|), & j \in \mathcal{A}, & (3.10a) \\ \left| \frac{\partial L(\hat{\beta}^{\text{or},\lambda})}{\partial \beta_j} \right| \leq \lim_{t \rightarrow 0^+} \nabla J_\gamma(t), & j \in \mathcal{A}^c, & (3.10b) \\ \rho_{\mathcal{A}} > \max_{j \in \mathcal{A}} \bar{\nabla}^2 J_\gamma(|\hat{\beta}_j^{\text{or},\lambda}|). & & (3.10c) \end{cases}$$

만약 $\nabla J_\gamma(t) \neq \nabla J_\lambda^{\text{or}}(t)$ 이면 일반적으로 (3.9)와 (3.10a)는 서로 다른 조건이며 위 충분조건은 더이상 간단하게 표현 할 수 없다. 하지만 $J_\gamma(t)$ 가 다음 성질을 만족한다면

$$\nabla J_\gamma(t) = \nabla J_\lambda^{\text{or}}(t), \quad t > \alpha_*, \quad (3.11)$$

식(3.9)와 (3.10a)는 동일한 조건이 된다. 이제 이러한 원리가 적용된 몇 가지 예제를 구체적으로 소개한다.

Example 6. (신의 ridge 추정량) 별점함수가 ridge, $J_\lambda^{\text{or}}(t) = (\lambda/2)t^2$, $t > 0$ 인 경우 별점함수는 다음을 만족한다.

$$\nabla J_\lambda^{\text{or}}(t) = \lambda t, \quad \bar{\nabla}^2 J_\lambda^{\text{or}}(t) = \lambda, \quad t > 0.$$

이 경우 다음의 sparse ridge (Kwon 등, 2013) 별점함수를 사용하면 신의 성질을 가지기 위한 충분조건을 찾을 수 있다.

$$\nabla J_\gamma(t) = \nabla J_\gamma^{\text{MCP}}(t)I(t \leq c) + \lambda t I(t > c), \quad t > 0, \quad \gamma > 0.$$

단, $J_\gamma^{\text{MCP}}(t)$ 는 MCP 별점함수이고 $c = a\gamma/(1 + a\lambda)$ 이다. 만약 $\alpha_* > c$ 이면 $\nabla J_\gamma(t) = \nabla J_\lambda^{\text{or}}(t)$, $t > \alpha_*$ 이므로 식 (3.11)을 만족한다. 또한 $\bar{\nabla}^2 J_\gamma(t) = (1/a)I(t \leq c) - \lambda I(t > c)$, $t > 0$ 이므로 $\kappa = \gamma$, $\phi_{\mathcal{A}} = -\lambda$, $\phi = 1/a$ 가 되어 다음과 같이 신의 성질을 만족하기 위한 충분조건을 찾을 수 있다.

$$\alpha_* > c, \quad \Delta_{\mathcal{A}^c} \leq \gamma, \quad \rho_{\mathcal{A}} > -\lambda, \quad \rho > \frac{1}{a}.$$

이와 같이 적당한 상수 $c > 0$ 에 대하여 $\nabla J_\gamma(t) = \nabla J_\gamma^{\text{MCP}}(t)$, $t \leq c$ 가 되도록 근사하고 나머지 구간에서 $\nabla J_\gamma(t) = \nabla J_\lambda^{\text{or}}(t)$, $t > c$ 를 만족하도록 $J_\gamma(t)$ 를 정의하면 항상 신의 성질을 가지기 위한 충분조건을 찾을 수 있다.

Remark 9. 다음의 MNET (Huang 등, 2016) 별점함수도

$$\nabla J_\gamma(t) = \nabla J_\gamma^{\text{MCP}}(t) + \lambda t, \quad t > 0, \quad \gamma > 0,$$

동일한 원리를 사용하여 신의 ridge 별점 추정량이 신의 성질을 가지기 위한 충분조건을 찾을 수 있다.

Example 7. (신의 LASSO 추정량) 별점함수가 LASSO, $J_\lambda^{\text{or}}(t) = \lambda t$, $t > 0$ 인 경우 별점함수는 다음을 만족한다.

$$\nabla J_\lambda^{\text{or}}(t) = \lambda, \quad \bar{\nabla}^2 J_\lambda^{\text{or}}(t) = 0, \quad t > 0.$$

이 경우 다음의 clipped LASSO (Kwon 등, 2015) 벌점함수를 사용하면 디자인 행렬이 SI와 같은 특별한 조건을 만족하지 않아도 강한 신의 성질을 가지기 위한 충분조건을 찾을 수 있다.

$$\nabla J_\gamma(t) = \nabla J_\gamma^{\text{MCP}}(t)I(t \leq c) + \lambda I(t > c), \quad t > 0, \quad \gamma > \lambda.$$

단, $c = a(\gamma - \lambda)$ 이다. 만약 $\alpha_* > c$ 이면 $\nabla J_\gamma(t) = \nabla J_\lambda^{\text{or}}(t) = \lambda, t > \alpha_*$ 이므로 (3.11)을 만족한다. 또한 $\bar{\nabla}^2 J_\gamma(t) = (1/a)I(t \leq c), t > 0$ 이므로 $\kappa = \gamma, \phi_{\mathcal{A}} = 0, \phi = 1/a$ 가 되어 다음과 같이

$$\alpha_* > c, \quad \Delta_{\mathcal{A}^c} \leq \gamma, \quad \rho_{\mathcal{A}} > 0, \quad \rho > \frac{1}{a},$$

신의 성질을 만족하기 위한 충분조건을 찾을 수 있다.

Example 8. (신의 bridge 추정량) 벌점함수가 bridge (Huang 등, 2008), $J_\lambda(t) = \lambda t^{1/2}, t > 0$ 인 경우 벌점함수는 다음을 만족한다.

$$\nabla J_\lambda^{\text{or}}(t) = \frac{\lambda}{2}t^{-\frac{1}{2}}, \quad \nabla^2 J_\lambda^{\text{or}}(t) = -\frac{\lambda}{4}t^{-\frac{3}{2}}, \quad t > 0.$$

만약 동일한 벌점함수를 사용하면 $\phi = \infty$ 가 되어 강한 신의 성질을 만족할 수 없다. 이 경우 다음과 같이 $J_\lambda^{\text{or}}(t)$ 를 원점 근처에서 선형으로 근사하면 (Um 등, 2020) 강한 신의 성질을 가지기 위한 충분조건을 찾을 수 있다.

$$\nabla J_\gamma(t) = \nabla J_\lambda^{\text{or}}(\gamma)I(t \leq \gamma) + \nabla J_\lambda^{\text{or}}(t)I(t > \gamma), \quad t > 0, \quad \gamma > 0.$$

만약 $\alpha_* > \gamma$ 이면 $\nabla J_\gamma(t) = \nabla J_\lambda^{\text{or}}(t) = (\lambda/2)t^{-1/2}, t > \alpha_*$ 이므로 (3.11)을 만족한다. 또한 $\bar{\nabla}^2 J_\gamma(t) = (\lambda/4)t^{-3/2}I(t > \alpha_*), t > 0$ 이므로 $\kappa = \nabla J_\lambda^{\text{or}}(\gamma), \phi_{\mathcal{A}} = (\lambda/4)\alpha_*^{-3/2}, \phi = (\lambda/4)\gamma^{-3/2}$ 이므로 다음과 같이 신의 성질을 만족하기 위한 충분조건을 찾을 수 있다.

$$\alpha_* > \gamma, \quad \Delta_{\mathcal{A}^c} \leq \nabla J_\lambda^{\text{or}}(\gamma), \quad \rho_{\mathcal{A}} > \frac{\lambda}{4}\alpha_*^{-\frac{3}{2}}, \quad \rho > \frac{\lambda}{4}\gamma^{-\frac{3}{2}}.$$

참고로 위 조건은 약한 신의 성질에 대하여는 동일한 벌점함수를 사용하는 경우보다 조금 더 강한 조건이 된다.

4. 예제와 가상실험

이 장에서는 단순 선형 회귀모형을 사용하여 벌점 추정량의 성질을 간단히 살펴보고 앞에서 소개한 최적조건이 유한 표본에서 성립하는지 확인하기 위하여 가상실험을 진행한다.

4.1. 예제: 단순 선형 회귀모형에서 벌점 추정량의 성질

절편이 없는 단순 선형 회귀모형 $y = x\beta + \varepsilon$ 에 대하여 적당한 표본이 주어지고 잔차제곱합이 다음과 같다고 하자.

$$L(\beta) = \frac{\|y - x\beta\|^2}{2n} = \rho(\beta - 3)^2 + 1, \quad \rho > 0.$$

MCP 벌점함수에 대하여 벌점 잔차제곱합은 다음과 같이 조각난 이차 함수(piecewise quadratic function)가 된다.

$$Q_\gamma(\beta) = \begin{cases} \rho(\beta - 3)^2 - \frac{1}{2a}(|\beta| - a\lambda)^2 + a\lambda^2/2 + 1, & |\beta| \leq a\lambda, \\ \rho(\beta - 3)^2 + \frac{a\lambda^2}{2} + 1, & |\beta| > a\lambda. \end{cases}$$

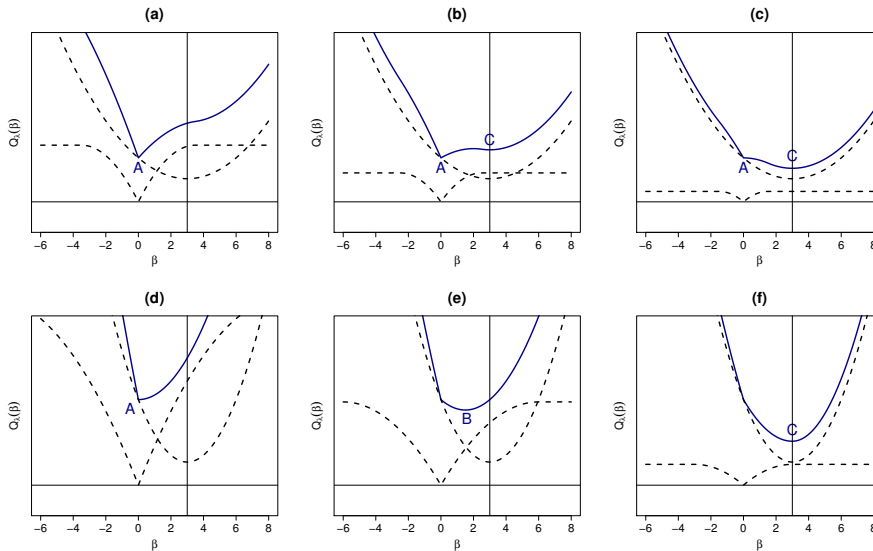


Figure 1: Examples of penalized estimators for MCP penalty: local (top) and unique global (bottom).

A, B, C indicate that the minimizer is zero, not zero but shrunken and the same as LS estimator, respectively.

Figure 1은 위 벌점 잔차제곱합에 대하여 ρ, a, λ 의 값에 따라 벌점 추정량 $\hat{\beta}^\lambda$ 가 어떤 값을 가지는지 설명하는 그림이다. Figure 1에서 검은 색 점선은 각각 잔차제곱합과 벌점함수이며 파란색 실선은 벌점 잔차제곱합을 나타낸다. 또한 A, B, C는 각각 $\hat{\beta}^\lambda = 0, 0 < \hat{\beta}^\lambda < 3, \hat{\beta}^\lambda = 3$ 인 경우를 나타낸다. 그림 (a), (b), (c)에서 $\nu = 1/10, a = 2.5, \lambda = 7/5, 5/5, 3/5$ 이고 (d), (e), (f)에서 $\nu = 3/10, a = 5, \lambda = 9/5, 6/5, 3/5$ 이다. Figure 1을 통해 다음과 같은 사실을 확인 할 수 있다.

- 벌점함수를 사용하지 않은 경우 최소 제곱 추정량은 $\hat{\beta}^{LS} = 3$ 이다.
- (a), (b), (c)의 경우 $\rho = 1/10 < 1/a = 1/2.5$ 이므로 $Q_\lambda(\beta)$ 는 비볼록 함수이고 여러 개의 국소 최소해를 가질 수 있다.
 - (a)는 λ 의 값이 가장 큰 경우이다. 이 경우 원점 근방에서 $J_\lambda(\beta)$ 의 변화가 $L(\beta)$ 보다 더 크기 때문에 $Q_\lambda(\beta)$ 는 A에서 최소값을 가지며 $\hat{\beta}^\lambda = 0$ 이 된다.
 - (b)는 λ 의 값이 적당한 경우이다. $Q_\lambda(\beta)$ 는 A와 C 두 곳에서 국소 최소해를 가지게 된다. 특히 C에서 최소값이 발생하는 경우 $J_\lambda(\beta)$ 가 국소적으로 상수 함수이므로 $\hat{\beta}^\lambda = \hat{\beta}^{LS} = 3$ 이 되고 $\hat{\beta}^\lambda$ 와 $\hat{\beta}^{LS}$ 는 약한 신의 성질을 갖는다.
 - (c)는 λ 의 값이 가장 작은 경우이다. (b)의 경우와 동일하지만 C에서 약한 신의 성질을 가진다는 것을 좀 더 분명하게 확인할 수 있다.
- (d), (e), (f)의 경우 $\rho = 3/10 > 1/a = 1/5$ 이므로 $Q_\lambda(\beta)$ 는 엄밀한 볼록 함수이고 유일한 광역 최소해를 가진다.
 - (d)는 λ 의 값이 가장 큰 경우이며 A에서 $\hat{\beta}^\lambda = 0$ 이 된다.
 - (e)는 λ 의 값이 적당한 경우이며 B에서 $0 < \hat{\beta}^\lambda < \hat{\beta}^{LS} = 3$ 이 된다. 즉 $\hat{\beta}^\lambda$ 는 축소 효과와 강한 신의 성질을 가진다.
 - (f)는 λ 의 값이 가장 작은 경우이며 C에서 $\hat{\beta}^\lambda = \hat{\beta}^{LS} = 3$ 이 된다. 따라서 $\hat{\beta}^\lambda$ 뿐만 아니라 $\hat{\beta}^{LS}$ 도 강한 신의 성질을 가진다.

4.2. 가상실험: 강한 신의 성질에 대한 충분조건

앞의 예제에서 소개한 신의 추정량이 유한 표본에서 강한 신의 성질을 가지는지 확인하기 위하여 다음 모형을 사용하여 가상실험을 진행하였다.

$$X_j \sim \text{i.i.d. } N(0, 1), \quad \varepsilon \sim N(0, 1), \quad \beta_j^* = \alpha \left\{ \frac{1 + (j-1)}{(q-1)} \right\} I(j \leq q), \quad j \leq p.$$

위 모형에서 $\alpha \in \{0.6, 1.2\}$, $n \in \{50, 100, \dots, 600\}$, $p \in \{20, 40\}$, $q \in \{5, 10\}$ 으로 두고 가능한 모든 조합에 대하여 각각 100회씩 신의 추정량이 강한 신의 성질을 만족하는지 확인하였다. 실험에 사용된 신의 추정량과 벌점함수는 다음과 같다.

- 신의 LS 추정량과 $a = 4$ 인 MCP 벌점함수 (Zhang, 2010).
- $\lambda = 0.1$ 인 신의 LASSO 추정량과 $\lambda = 0.1$, $a = 4$ 인 clipped LASSO 벌점함수 (Kwon 등, 2015).
- $\lambda = 0.1$ 인 신의 ridge 추정량과 $\lambda = 0.1$, $a = 4$ 인 sparse ridge 벌점함수 (Kwon 등, 2013).
- $\lambda = 0.1$ 인 신의 bridge 추정량과 $\gamma = 0.5$ 인 modified bridge 벌점함수 (Um 등, 2020).

만약 신의 추정량의 회귀계수 중 하나라도 0이 되는 경우가 발생하면, 모두 0이 되지 않을 때까지 λ 의 값을 계속하여 반으로 줄여 사용하였다.

실험의 결과는 Table 1에 정리하였으며 잘 알려진 것처럼 n 과 a 가 클 수록, p 와 q 가 작을 수록 신의 추정량이 강한 신의 성질을 더 잘 만족하는 것을 확인할 수 있다. 참고로 이 실험에서는 신의 추정량을 고정해 두고 신의 성질이 성립하는지 확인한 것이며 신의 성질을 만족하는 조율모수의 존재성과 같이 좀 더 일반적인 실험을 진행하는 방법에 대하여는 Kim과 Kwon(2012)을 참고하기 바란다.

5. 정리의 글

많은 연구자들이 벌점 추정법에 대한 연구를 어려워하는 이유는 기술적인 난이도 보다는 전체적인 연구의 흐름을 파악하기 어렵기 때문이라고 판단된다. 만약 본 논문에서 소개한 내용을 바탕으로 벌점 추정법의 연구 과정을 간략하게 작성해 본다면 다음과 같다. 먼저 신의 벌점 추정량에 사용할 벌점함수를 결정하고 신의 벌점 추정량의 최적조건을 유도한다. 그 다음 벌점 추정량에 사용할 벌점함수를 결정하고 벌점 추정량의 최적조건을 유도한다. 마지막으로 신의 벌점 추정량이 신의 성질을 가지는 충분조건을 구성한다. 이러한 과정이 최적의 답은 아니겠지만 독자들에게 어느 정도 도움이 될 것으로 기대한다.

신의 벌점 추정량의 존재성과 추정오차, 최적 조율모수의 선택법, 추정량 구현을 위한 알고리즘, 고차원 모형에서의 신의 성질 등 본 논문에서 다루지 못한 내용 중 벌점 추정법에서 반드시 이해해야 할 주제들이 많이 남아 있으며 기회가 된다면 이러한 주제들에 대하여 소개하는 논문을 작성하고자 한다.

References

- Choi H and Park C (2012). Approximate penalization path for smoothly clipped absolute deviation, *Journal of Statistical Computation and Simulation*, **82**, 643–652.
- Fan J and Li R (2001). Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, **96**, 1348–1360.
- Fan J and Lv J (2011). *Nonconcave Penalized Likelihood with NP-Dimensionality*, IEEE Transactions on information theory, **57**, 5467–5484.
- Fan J and Peng H (2004). Nonconcave penalized likelihood with a diverging number of parameters, *The Annals of Statistics*, **32**, 928–961.

Table 1: Number of cases when the strong oracle property holds

α	p	q	oracle	n										
				100	150	200	250	300	350	400	450	500	550	600
0.6	20	5	LS	0	1	3	2	7	7	9	22	32	32	32
			LASSO	1	8	18	25	38	56	67	66	77	84	84
			ridge	3	8	18	19	32	41	58	58	64	81	79
		bridge	27	41	33	32	35	39	46	50	39	47	34	
		10	LS	1	0	7	6	10	17	15	27	27	37	47
			LASSO	4	8	25	35	42	65	66	81	82	85	92
	ridge		1	4	20	31	36	52	55	71	70	79	83	
	bridge	8	12	19	17	26	24	25	29	26	31	32		
	40	5	LS	0	0	0	1	0	0	1	2	0	10	12
			LASSO	0	0	2	7	21	22	40	42	56	64	76
			ridge	0	1	2	4	17	23	32	32	43	48	58
		bridge	0	7	40	34	46	36	38	44	43	39	43	
10		LS	0	0	0	0	1	1	4	3	7	6	14	
		LASSO	0	0	1	7	16	22	36	53	61	72	88	
	ridge	0	0	0	5	14	11	21	41	50	64	75		
bridge	0	3	15	22	31	19	28	26	27	31	39			
0.9	20	5	LS	2	15	20	26	51	61	71	65	73	80	91
			LASSO	7	39	53	65	89	91	94	97	99	99	100
			ridge	16	49	61	71	92	94	97	100	98	100	100
		bridge	48	97	97	99	99	100	100	100	100	100	100	
		10	LS	8	16	37	39	48	67	81	75	87	93	93
			LASSO	26	44	69	78	81	91	98	97	99	100	100
	ridge		30	48	75	80	85	94	97	98	99	100	100	
	bridge	63	93	97	98	96	97	100	100	100	100	100		
	40	5	LS	0	0	8	18	29	23	50	54	61	68	76
			LASSO	0	5	37	62	70	79	92	93	92	96	97
			ridge	0	11	44	71	80	80	93	96	92	97	97
		bridge	0	13	80	99	100	99	100	100	100	100	100	
10		LS	0	0	4	15	26	28	41	51	73	64	85	
		LASSO	0	5	26	52	66	80	88	93	93	95	99	
	ridge	0	11	40	60	74	87	93	93	94	96	99		
bridge	0	27	91	95	99	100	100	100	100	100	100			
1.2	20	5	LS	14	33	49	83	86	90	95	97	98	99	100
			LASSO	33	71	82	96	97	98	100	100	100	100	100
			ridge	55	85	91	97	99	99	100	100	100	100	100
		bridge	58	99	100	100	100	100	100	100	100	100	100	
		10	LS	20	47	66	87	82	94	96	98	98	100	99
			LASSO	49	80	89	96	98	98	100	100	100	100	100
	ridge		63	87	97	98	100	100	100	100	100	100	100	
	bridge	67	100	100	100	100	100	100	100	100	100	100		
	40	5	LS	0	13	31	64	63	85	89	95	92	98	100
			LASSO	0	37	71	93	94	99	99	100	100	100	100
			ridge	0	52	83	95	96	100	99	100	100	100	100
		bridge	0	26	84	98	100	100	100	100	100	100	100	
10		LS	0	7	37	64	67	90	96	92	96	100	100	
		LASSO	0	25	70	90	94	98	98	99	100	100	100	
	ridge	0	36	81	98	98	99	100	100	100	100	100		
bridge	0	20	78	100	100	100	100	100	100	100	100			

Hoerl AE and Kennard RW (1970). Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, **12**, 55–67.

Huang J, Brehehy P, Lee S, Ma S, and Zhang CH (2016). The mnet method for variable selection, *Statistica*

- Sinica*, 903–923.
- Huang J, Horowitz JL, and Ma S (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models, *The Annals of Statistics*, **36**, 587–613.
- Kim Y, Choi H, and Oh HS (2008). Smoothly clipped absolute deviation on high dimensions, *Journal of the American Statistical Association*, **103**, 1665–1673.
- Kim Y, Jeon JJ, and Han S (2016). A necessary condition for the strong oracle property, *Scandinavian Journal of Statistics*, **43**, 610–624.
- Kim Y and Kwon S (2012). Global optimality of nonconvex penalized estimators, *Biometrika*, **99**, 315–325.
- Kwon S, Ahn J, Jang W, Lee S, and Kim Y (2017). A doubly sparse approach for group variable selection, *Annals of the Institute of Statistical Mathematics*, **69**, 997–1025.
- Kwon S and Kim Y (2012). Large sample properties of the scad-penalized maximum likelihood estimation on high dimensions, *Statistica Sinica*, 629–653.
- Kwon S, Kim Y, and Choi H (2013). Sparse bridge estimation with a diverging number of parameters, *Statistics and Its Interface*, **6**, 231–242.
- Kwon S, Lee S, and Kim Y (2015). Moderately clipped lasso, *Computational Statistics & Data Analysis*, **92**, 53–67.
- Kwon S, Oh S, and Lee Y (2016). The use of random-effect models for high-dimensional variable selection problems, *Computational Statistics & Data Analysis*, **103**, 401–412.
- Lee S and Kim S (2019). Marginalized lasso in sparse regression, *Journal of the Korean Statistical Society*, **48**, 396–411.
- Lee Y and Oh HS (2014). A new sparse variable selection via random-effect model, *Journal of Multivariate Analysis*, **125**.
- Lv J and Fan Y (2009). A unified approach to model selection and sparse recovery using regularized least squares, *The Annals of Statistics*, **37**, 3498–3528.
- Pan W, Shen X, and Liu B (2013). Cluster analysis: unsupervised learning via supervised learning with a non-convex penalty, *The Journal of Machine Learning Research*, **14**, 1865–1889.
- Shen X and Huang HC (2010). Grouping pursuit through a regularization solution surface, *Journal of the American Statistical Association*, **105**, 727–739.
- Shen X, Pan W, and Zhu Y (2012). Likelihood-based selection and sharp parameter estimation, *Journal of the American Statistical Association*, **107**, 223–232.
- Shen X, Pan W, Zhu Y, and Zhou H (2013). On constrained and regularized high-dimensional regression, *Annals of the Institute of Statistical Mathematics*, **65**, 807–832.
- Tibshirani R (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society : Series B (Methodological)*, **58**, 267–288.
- Tibshirani RJ and Taylor J (2011). The solution path of the generalized lasso, *The Annals of Statistics*, **39**, 1335–1371.
- Um S, Kim D, Lee S, and Kwon S (2020). On the strong oracle property of concave penalized estimators with infinite penalty derivative at the origin, *Journal of the Korean Statistical Society*, **49**, 439–456.
- Wang H, Li B, and Leng C (2009). Shrinkage tuning parameter selection with a diverging number of parameters, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**, 671–683.
- Wang H, Li R, and Tsai CL (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method, *Biometrika*, **94**, 553–568.

- Yuan M and Lin Y (2006). Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**, 49–67.
- Yuan M and Lin Y (2007). Model selection and estimation in the gaussian graphical model, *Biometrika*, **94**, 19–35.
- Zhang CH (2010). Nearly unbiased variable selection under minimax concave penalty, *The Annals of Statistics*, **38**, 894–942.
- Zhao P and Yu B (2006). On model selection consistency of lasso, *The Journal of Machine Learning Research*, **7**, 2541–2563.
- Zou H (2006). The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*, **101**, 1418–1429.
- Zou H, Hastie T, and Tibshirani R (2006). Sparse principal component analysis, *Journal of Computational and Graphical Statistics*, **15**, 265–286.
- Zou H and Li R (2008). One-step sparse estimates in nonconcave penalized likelihood models, *Annals of Statistics*, **36**, 1509.

Received March 16, 2021; Revised March 29, 2021; Accepted March 30, 2021

선형 회귀모형에서 별점 추정량의 신의 성질에 대한 충분조건

권성훈^a, 장재호^a, 문혜성^a, 이상인^{1,b}

^a건국대학교 응용통계학과, ^b충남대학교 정보통계학과

요 약

본 논문은 선형 회귀모형에서 별점 추정량의 신의 성질에 대한 충분조건을 구성하는 방법을 소개하였다. 신의 추정량, 별점 추정량, 신의 별점 추정량, 신의 성질을 명확히 정의하였으며 이를 바탕으로 신의 성질에 대한 최적조건과 최적조건에 대한 충분조건을 구성하는 방법을 대부분의 별점함수에 적용 가능하도록 하나의 통합된 원리로 소개하였다. 추가로 신의 성질에 대한 이해를 돕기 위해 간단한 예제와 함께 가상실험 결과를 첨부하였다.

주요용어: 별점 추정량, 신의 별점 추정량, 신의 성질

본 논문은 2019년도 건국대학교 우수연구인력 양성사업과 한국연구재단 지원에 의한 논문임 (No. 2020R111A3071646).

본 논문은 문혜성의 석사 학위논문의 내용을 일부 차용하고 있음.

¹교신저자: 대전광역시 유성구 대학로99, 충남대학교 정보통계학과. E-mail: sanginlee44@gmail.com