Journal of Multivariate Analysis 100 (2009) 1338–1352

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On the degrees of freedom in shrinkage estimation

Kengo Kato

Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

ARTICLE INFO

Article history:
Received 16 March 2008
Available online 13 December 2008

AMS subject classifications:
62J05
62J07

Keywords:
AIC
Degrees of freedom
Fused Lasso
Group Lasso
Lasso
Mallows' C_p
Second fundamental form
Shrinkage estimation
Stein's lemma
Tubal coordinates

ABSTRACT

We study the degrees of freedom in shrinkage estimation of regression coefficients. Generalizing the idea of the Lasso, we consider the problem of estimating the coefficients by minimizing the sum of squares with the constraint that the coefficients belong to a closed convex set. Based on a differential geometric approach, we derive an unbiased estimator of the degrees of freedom for this estimation method, under a smoothness assumption on the boundary of the closed convex set. The result presented in this paper is applicable to estimation with a wide class of constraints. As an application, we obtain a C_p type criterion and AIC for selecting tuning parameters.

1. Introduction

In recent years, there has been considerable attention on shrinkage methods in estimating the coefficients of a linear model. Compared with the ordinary least squares (OLS), shrinkage methods often improve the prediction accuracy. In addition, if the constraint set towards which the estimator is shrunk has edges or corners, some coefficients can be set exactly equal to zero.

To be precise, suppose $y = (y_1, \ldots, y_n)'$ is the response vector and $x_j = (x_{1j}, \ldots, x_{nj})'$, $j = 1, \ldots, p$ are the covariates. Let $X = [x_1 \cdots x_p]$ be the design matrix. We consider a linear model

$$y = X\beta + \epsilon, \tag{1.1}$$

where $\beta = (\beta_1, \ldots, \beta_p)'$ is the coefficient vector and $\epsilon \sim N_n(0, \sigma^2 I_n)$. By centering the covariates if required, we assume that the intercept term is not included in the above linear model.

A canonical example of shrinkage methods is the Lasso [1]. Let $\|\cdot\|_2$ be the ordinary Euclidean norm: $\|z\|_2 = (z'z)^{\frac{1}{2}}$ for a vector z with appropriate dimension. The Lasso estimate is defined by an optimal solution of the following minimization problem:

$$\min_{\beta} \ \|y - X\beta\|_2^2 \quad \text{s.t. } \|\beta\|_1 \le t, \tag{1.2}$$

or equivalently

$$\min_{\beta} \ \left\{ \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\}, \tag{1.3}$$

E-mail address: ee077006@mail.ecc.u-tokyo.ac.jp.

0047-259X/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2008.12.002

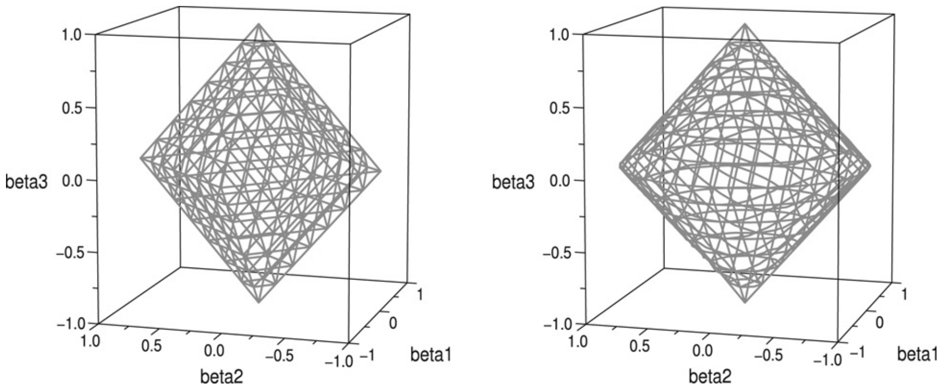

Fig. 1. The constraint sets of the Lasso (left) and the group Lasso (right).

where $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$, t and λ are non-negative tuning parameters. The Lasso shrinks the coefficients towards zero as t decreases or λ increases. An important feature of the Lasso is that, depending on the tuning parameter t or λ, some coefficients are set exactly equal to zero. It should be noted that although (1.2) and (1.3) are equivalent as *minimization problems*, the optimal solutions of these two problems are different as *estimators* since the correspondence between t and λ generally depends on the data.

It is well recognized that the choice of the tuning parameter is a crucial part of shrinkage methods. As indicated by Efron [2], the degrees of freedom plays an important role in selecting the tuning parameter. The degrees of freedom reflects the model complexity controlled by the shrinkage and corresponds to the penalty term of model selection criteria such as Mallows' C_p [3] and Akaike's information criterion (AIC, [4]). Recently, Zou et al. [5] show that, under the penalization formulation (1.3), the number of non-zero coefficients is an unbiased estimator of the degrees of the freedom of the Lasso.

In this paper, we study the degrees of freedom for the estimation method formulated as

$$\min_{\beta} \ \|y - X\beta\|_2^2 \quad \text{s.t. } \beta \in K, \tag{1.4}$$

where $K \subset \mathbb{R}^p$ is a closed convex set. The minimization problem (1.4) is a generalization of the Lasso and its variants. The problem of selecting the tuning parameter is viewed as the problem of selecting the constraint set among a given collection of closed convex sets. The main goal of this paper is to derive an unbiased estimator of the degrees of freedom for this estimation method under both $p \le n$ and $p > n$ settings.

Suppose first that $\operatorname{rank} X = p$ (hence $p \le n$) and denote the OLS estimator by $\widehat{\beta}^0$. Define the inner product $\langle \cdot, \cdot \rangle$ in $\mathbb{R}^p$ by

$$\langle u, v \rangle = u'Vv, \quad \text{for } u, v \in \mathbb{R}^p \tag{1.5}$$

where $V = X'X$ and let $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$. Note that the inner product $\langle \cdot, \cdot \rangle$ here is not same as the ordinary Euclidean inner product. Then, the minimization problem (1.4) can be rewritten as

$$\min_{\beta} \ \|\beta - \widehat{\beta}^0\| \quad \text{s.t. } \beta \in K. \tag{1.6}$$

It is seen that the optimal solution $\widehat{\beta}_K$ of the problem (1.6) is given by the projection of $\widehat{\beta}^0$ onto K. Since K is closed and convex, $\widehat{\beta}_K$ is uniquely defined.

Here we present illustrative examples of the constraint sets of the Lasso and the group Lasso. The left of Fig. 1 corresponds to the Lasso constraint $|\beta_1| + |\beta_2| + |\beta_3| \le 1$. The right one corresponds to the group Lasso constraint $(\beta_1^2 + \beta_2^2)^{\frac{1}{2}} + |\beta_3| \le 1$.

To derive an unbiased estimator of the degrees of freedom, we utilize Stein's lemma [6]. Stein's lemma yields that an unbiased estimator of the degrees of freedom is given by the divergence of $\widehat{\mu}_K$ with respect to y, which coincides with the divergence of $\widehat{\beta}_K$ with respect to $\widehat{\beta}^0$. However, in general, the estimator $\widehat{\beta}_K$ cannot be expressed in an explicit form. Thus it is often impossible to directly calculate the divergence.

To overcome this difficulty, we use an idea of the "tubal coordinates" [7]. From an approach similar to that of Kuriki and Takemura [8], we derive the divergence of the projection onto K in terms of geometric quantities under a regularity condition on the boundary ∂K of K. Hence we obtain an unbiased estimator of the degrees of freedom of $\widehat{\mu}_K$, which is computable in principle.

Next, we consider a more challenging situation: $p > n$ case. When $p > n$, the OLS estimator is not uniquely defined. Furthermore, in general, shrinkage estimators such as the Lasso estimator are not uniquely defined as functions of y. However, the fitted model $\widehat{\mu}_K = X\widehat{\beta}_K$ is in fact always uniquely determined for a compact convex K where $\widehat{\beta}_K$ is any optimal solution of (1.4), since $\widehat{\mu}_K$ is the projection of y onto the closed convex set $\widetilde{K} = \{X\beta \mid \beta \in K\}$. Hence, it remains to