The degrees of freedom of the Lasso in underdetermined linear regression models

Maher Kachour, Jalal M. Fadili, Christophe Chesneau, Charles Dossal, Gabriel Peyré

The degrees of freedom of the Lasso in underdetermined linear regression models

Maher Kachour and Jalal Fadili, GREYC, ENSICAEN
Christophe Chesneau, LMNO Université de Caen
Charles Dossal, IMB Université Bordeaux 1
Gabriel Peyré, Ceremade Université Paris-Dauphine

***Abstract*—In this paper, we investigate the degrees of freedom (df) of penalized ℓ_1 minimization (also known as the Lasso) for an underdetermined linear regression model. We show that under a suitable condition on the design matrix, the number of nonzero coefficients of the Lasso solution is an unbiased estimate for the degrees of freedom. An effective estimator of the number of degrees of freedom may have several applications including an objectively guided choice of the regularization parameter in the Lasso through the SURE or GCV frameworks.**

***Index Terms*—Lasso, degrees of freedom, SURE.**

I. Introduction

Consider the following linear regression model

$$y = Ax^0 + \varepsilon, \tag{1}$$

where $y \in \mathbb{R}^n$ is the response vector, $A = (a_1, \cdots, a_p) \in \mathbb{R}^{n\times p}$ is a deterministic design matrix with $n < p$, $x^0 \in \mathbb{R}^p$ is the unknown regression vector, and $\varepsilon \in \mathbb{R}^n$ is the noise vector whose entries are i.i.d. $\mathcal{N}(0, \sigma^2)$. The goal is to solve (1) when the solution is assumed to be sparse. Towards this goal, a now popular estimator is the Lasso [4]. The Lasso estimate amounts to solving the following convex problem

$$\min_{x\in\mathbb{R}^p} \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1, \tag{2}$$

where $\lambda > 0$ is the regularization or tuning parameter. In the last years, there has been a huge amount of work where efforts have focused on investigating the theoretical guarantees of the Lasso as a sparse recovery procedure from noisy measurements in the underdetermined case $n < p$.

Degrees of freedom *df* is a familiar phrase in statistics. In overdetermined linear regression *df* is the number of estimated predictors. Degrees of freedom is often used to quantify the model complexity of a statistical modeling procedure (e.g. it corresponds to the penalty term of model selection criteria such as AIC and BIC). However, generally speaking, there is no exact correspondence between the degrees of freedom *df* and the number of parameters in the model. On the other hand, the degrees of freedom plays an important role for an objective selection of the tuning parameter.

Let us denote by $\hat{x}$ any estimator of x^0 which depends on y and let $\hat{y} = A\hat{x}$. Since $y \sim \mathcal{N}(Ax^0, \sigma^2)$, according to [2], the degrees of freedom of $\hat{y}$ is

$$df(\hat{y}) = \sum_{i=1}^{n} \frac{\mathrm{cov}(\hat{y}_i, y_i)}{\sigma^2}. \tag{3}$$

If $\hat{y}$ is almost differentiable, Stein's lemma [3] yields the following unbiased estimator of *df*

$$\hat{df}(\hat{y}) = \mathrm{div}\,\hat{y} = \sum_{i=1}^{n} \frac{\partial \hat{y}_i}{\partial y_i}. \tag{4}$$

Contributions Let $\hat{\mu}_\lambda = \hat{\mu}_\lambda(y) = A\hat{x}_\lambda(y)$ be the Lasso response vector, where $\hat{x}_\lambda(y)$ is a solution of the Lasso problem (2). In the overdetermined case, i.e. $n > p$, $\hat{x}_\lambda(y)$ is unique, and the authors in [5] showed that for any given λ the number of non-zero coefficients of $\hat{x}_\lambda$ is an unbiased estimator of the degrees of the freedom of the Lasso. Though their proof contains a gap. The contribution of this paper is to extend their result to the underdetermined case where the Lasso solution is not unique. To ensure the uniqueness of the solution, we introduce the condition (UC) on the design matrix.

II. Main results

Let $z \in \mathbb{R}^p$, $S \subseteq \{1, 2, \cdots, p\}$ and $|S|$ its cardinality. We denote by A_S the submatrix $A_S = [\cdots, a_j, \cdots]_{j\in S}$, where a_j is the jth column of A and the pseudo-inverse $(A_S^t A_S)^{-1} A_S^t$ of A_S is denoted A_S^+. Let z_j be the jth component of z. Similarly, we define $z_S = (\cdots, z_j, \cdots)_{j\in S}$ for z. Let $\mathrm{supp}(z) = \{j : z_j \neq 0\}$ be the support or the active set of z.

Definition 1 (Condition (UC) [1]): A matrix A satisfies condition (UC) if, for all subsets $I \subset \{1, \cdots, p\}$ with $|I| \leq n$, such that $(a_i)_{i\in I}$ are linearly independent, for all indices $j \notin I$ and all vectors $V \in \{-1, 1\}^{|I|}$,

$$|\langle a_j, (A_I^+)^t V\rangle| \neq 1. \tag{5}$$

Theorem 1: Suppose that A satisfies condition (UC). For any $y \in \mathbb{R}^n$, there exists a finite set of values λ, denoted by $\{\lambda_m\}$, for which we have

$$\max_{j\notin I} |\langle a_j, y - A\hat{x}_\lambda(y)\rangle| = \lambda, \tag{6}$$

where $I = \mathrm{supp}(\hat{x}(\lambda))$ and $\hat{x}_\lambda(y)$ is the solution of the Lasso. Furthermore, if $\lambda \in \left]0, \|A^t y\|_\infty\right[\setminus \{\lambda_m\}$, then

$$\max_{j\notin I} |\langle a_j, y - A\hat{x}_\lambda(y)\rangle| < \lambda. \tag{7}$$

Theorem 2: Suppose that A satisfies condition (UC). For any $y \in \mathbb{R}^n$, and all values of λ for which (7) is satisfied, we have

- The Lasso response $\hat{\mu}_\lambda(y) = A\hat{x}_\lambda(y)$ is a uniformly Lipschitz function of y;
- The support and vector sign of the Lasso solution are locally constant with respect to y, and consequently

$$\mathrm{div}\ \hat{\mu}_\lambda(y) = |\,\mathrm{supp}(\hat{x}_\lambda(y))|. \tag{8}$$

That is, using Stein's lemma [3] and the divergence formula (8), the number of non-zero coefficients of $\hat{x}_\lambda$ is an unbiased estimator of the degrees of the freedom of the Lasso.

References

[1] Dossal, C (2007). A necessary and sufficient condition for exact recovery by l1 minimization. Technical report, HAL-00164738:1.

[2] Efron, B. (1981). How biased is the apparent error rate of a prediction rule. J. Amer. Statist. Assoc. vol. 81 pp. 461?470.

[3] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. Ann. Statist. 9 1135-1151.

[4] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. J. Roy. Statist. Soc. Ser. B 58(1) 267-288.

[5] Zou, H., Hastie, T. and Tibshirani, R. (2007). On the "degrees of freedom" of the Lasso. Ann. Statist. Vol. 35, No. 5. 2173-2192.