The degrees of freedom of the group Lasso for a general design
Samuel Vaiter, Gabriel Peyré, Jalal M. Fadili, Charles-Alban Deledalle,
Charles Dossal

To cite this version:
Samuel Vaiter, Gabriel Peyré, Jalal M. Fadili, Charles-Alban Deledalle, Charles Dossal. The degrees of freedom of the group Lasso for a general design. SPARS’13, Jul 2013, Lausanne, Switzerland. 1 page. hal-00926929

HAL Id: hal-00926929
https://hal.archives-ouvertes.fr/hal-00926929
Submitted on 10 Jan 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract—In this paper, we are concerned with regression problems where covariates can be grouped in nonoverlapping blocks, from which a few are active. In such a situation, the group Lasso is an attractive method for variable selection since it promotes sparsity of the groups. We study the sensitivity of any group Lasso solution to the observations and provide its precise local parameterization. When the noise is Gaussian, this allows us to derive an unbiased estimator of the degrees of freedom of the group Lasso. This result holds true for any fixed design, no matter whether it is under- or over-determined. Our results specialize to those of [1], [2] for blocks of size one, i.e. $\ell^1$ norm. These results allow objective choice of the regularization parameter through e.g. the SURE.

I. GROUP LASSO AND DEGREES OF FREEDOM

Consider the linear regression problem $Y = X_\beta_0 + \varepsilon$, where $Y$ is the real $n$-dimensional response vector, $\beta_0 \in \mathbb{R}^p$ is the unknown vector of regression coefficients to be estimated, $X \in \mathbb{R}^{n \times p}$ is the design matrix whose columns are the $p$ covariate vectors, and $\varepsilon$ is the error term. In this paper, we do not make any specific assumption on $\varepsilon$ with respect to $p$.

Let $B$ be a disjoint union of the set of indices i.e. $\bigcup_{b \in B} = \{1, \ldots, p\}$ such that $b, b' \in B, b \cap b' = 0$. For $\beta \in \mathbb{R}^p$, for each $b \in B$, $\hat{\beta}(\beta)_b$ is a subvector of $\beta$ whose entries are indexed by the block $b$, and $|b|$ is the cardinality of $b$. The group Lasso amounts to solving

$$\hat{\beta}(Y) = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{2} |y - X\beta|^2 + \lambda \sum_{b \in B} |\beta_b|,$$

from an observation $y \in \mathbb{R}^n$ of the regression model, where $\lambda > 0$ is the regularization parameter and $|\cdot|$ is the $\ell^1$-norm.

Let $y \mapsto \hat{\mu}(y) = X\hat{\beta}(Y)$ be the response or the prediction associated to $\hat{\beta}(Y)$, and let $\mu_0 = X\beta_0$. We recall that $\hat{\mu}(y)$ is always uniquely defined (by strict convexity of the fidelity term), although $\hat{\beta}(Y)$ may not as is the case when $X$ is rank-deficient or underdetermined design matrix. Suppose that $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$. Following [3], the DOF is given by $df = \sum_{i=1}^n \frac{\cos(Y_i, \hat{\mu}(Y))}{\sqrt{|\hat{\mu}(Y)|}}$. The well-known Stein’s lemma asserts that, if $y \mapsto \hat{\mu}(y)$ is a weakly differentiable function with an essentially bounded gradient, then an unbiased estimator of $df$ is $d(Y) = \text{div}(\hat{\mu}(Y)) = \text{tr}((\partial \hat{\mu}(Y))^T)$ and $E_\varepsilon(d(Y)) = df$, where $\partial \hat{\mu}(\cdot)$ is the Jacobian of $\hat{\mu}(\cdot)$.

In the sequel, we define the $B$-support $\text{supp}_B(\beta)$ of $\beta \in \mathbb{R}^p$ as $\text{supp}_B(\beta) = \{b \in B \mid |\beta_b| \neq 0\}$. The size of $\text{supp}_B(\beta)$ is $|\text{supp}_B(\beta)| = \sum_{b \in B} |b|$. The set of all $B$-supports is denoted $\mathcal{I}_B$ and $\mathcal{I}$, where $I$ is a $B$-support, is the matrix formed by the columns $X_i$ where $i$ is an element of $b \in I$. We also introduce the following block-diagonal operators

$$\delta_b : v \in \mathbb{R}^{|I|} \mapsto (v_{b}/|\beta_b|)_{b \in I} \in \mathbb{R}^{|I|}$$

and

$$P_{\beta} : v \in \mathbb{R}^{|I|} \mapsto (\text{Proj}_{b \supset \beta}(v_{b}))_{b \in I} \in \mathbb{R}^{|I|},$$

where $\text{Proj}_{b \supset \beta} = \text{Id} - \beta_b \delta_b$ is the orthogonal projector on $\beta_b$.

II. MAIN CONTRIBUTIONS

The first difficulty we need to overcome when $X$ is not full column rank is that $y \mapsto \hat{\beta}(Y)$ is set-valued. Toward this goal, we are led to impose a sufficient assumption on $X$ with respect to the block structure.

Assumption (A($\beta$)): Given a vector $\beta \in \mathbb{R}^p$ of $B$-support $I$, we assume that the finite subset of vectors $\{X_b\beta_b \mid b \in I\}$ is linearly independent.

It is important to notice that (A($\beta$)) is weaker than imposing that $X_I$ is full column rank, which is standard when analyzing the Lasso. The two assumptions coincide for the Lasso, i.e. $|b| = 1, \forall b \in I$.

Definition 1: Let $\lambda > 0$. The transition space $\mathcal{H}$ is defined as

$$\mathcal{H} = \bigcup_{I \in \mathcal{I}} \mathcal{H}_{I,b},$$

where we have denoted

$$\pi : \mathbb{R}^n \to \mathbb{R}^{I^1 \times I^2 \times I^3} \mapsto \mathbb{R}^n, \quad I^1 = \bigcup_{b \in I} \{\beta|b\} \setminus \{0\}$$

the canonical projection on $\mathbb{R}^n$ (with respect to the first component), $\text{bd } C$ is the boundary of the set $C$, and

$$A_{I,b} = \{ (y, \beta, u) \in \mathbb{R}^n \times \mathbb{R}^{I^1} \times \mathbb{R}^{I^2} \mid \|X_b^T(y - X_I\beta_I)\| = \lambda, X_I^T(X_I\beta_I - y) + \lambda \delta_{I,b} = 0, \forall g \in I, v_{g} = \beta_g / |\beta_g| \}.$$