The Degrees of Freedom of the Group Lasso

Samuel Vaiter, Charles Deledalle, Gabriel Peyré, Jalal M. Fadili, Charles Dossal

Degrees of Freedom of the Group Lasso

Samuel Vaiter, Charles Deledalle, Gabriel Peyré
CEREMADE, CNRS, Université Paris-Dauphine, France

VAITER@CEREMADE.DAUPHINE.FR

Jalal Fadili
GREYC, CNRS-ENSICAEN-Université de Caen, France

Charles Dossal
IMB, Université Bordeaux 1, France

Abstract

This paper studies the sensitivity to the observations of the block/group Lasso solution to an overdetermined linear regression model. Such a regularization is known to promote sparsity patterns structured as nonoverlapping groups of coefficients. Our main contribution provides a local parameterization of the solution with respect to the observations. As a byproduct, we give an unbiased estimate of the degrees of freedom of the group Lasso. Among other applications of such results, one can choose in a principled and objective way the regularization parameter of the Lasso through model selection criteria.

1. Introduction

This paper deals with the overdetermined linear regression model of the form $y = \mathbf{X}\beta_0 + \varepsilon$ where $y \in \mathbb{R}^Q$ is the observation/response vector, $\beta_0 \in \mathbb{R}^N$ the regression vector, $\mathbf{X}$ is the design matrix whose columns are linearly independent, and ε is an additive noise. Note that $Q > N$ and $\mathbf{X}_I^\mathsf{T}\mathbf{X}_I$ is an invertible matrix.

1.1. Group Lasso

A block segmentation $\mathcal{B}$ corresponds to a disjoint union of the set of indices i.e. $\bigcup_{b \in \mathcal{B}} = \{1, \dots, N\}$ and for each $b, b' \in \mathcal{B}, b \cap b' = \emptyset$. For $\beta \in \mathbb{R}^N$, for each $b \in \mathcal{B}$, $x_b = (\beta_i)_{i \in b}$ is a subvector of β whose entries are indexed by the block b, where $|b|$ is the cardinality of b.

We consider the Group Lasso or Block Sparse regularization introduced by (Bakin, 1999; Yuan & Lin, 2006) which reads

$$\min_{\beta \in \mathbb{R}^N} \frac{1}{2}\|y - \mathbf{X}\beta\|^2 + \lambda \sum_{b \in \mathcal{B}} \|\beta_b\|, \qquad (\mathcal{P}_\lambda(y))$$

where $\lambda > 0$ is the so-called regularization parameter and $\|\cdot\|$ is the ℓ^2-norm. Note that if each block b is of size 1, we recover the standard Lasso (Tibshirani, 1996).

1.2. Degrees of Freedom

We focus in this paper on the variations of the solution $\beta^\star(y)$ of $\mathcal{P}_\lambda(y)$ with respect to the observations y. This turns out to be a pivotal ingredient to compute the effective degrees of freedom (DOF) usually used to quantify the complexity of a statistical modeling procedure.

Let $\hat{\mu}(y) = \mathbf{X}\beta^\star(y)$ be the response or the prediction associated to the estimator $\beta^\star(y)$ of β_0, and let $\mu_0 = \mathbf{X}\beta_0$. It is worth noting that $\hat{\mu}(y)$ is always uniquely defined, although when $\beta^\star(y)$ is not as is the case of rank-deficient or underdetermined design matrix X. Note that any estimator $\hat{\mu}$ of μ_0 might be considered. We also make the assumption that ε is an additive white Gaussian noise term $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_Q)$, hence y follows the law $\mathcal{N}(\mu_0, \sigma^2 I_Q)$ and according to (Efron, 1986), the DOF is given by

$$df = \sum_{i=1}^Q \frac{\operatorname{cov}(y_i, [\hat{\mu}(y)]_i)}{\sigma^2} .$$

The well-known Stein's lemma asserts that if $\hat{\mu}$ is weakly differentiable then its divergence is an unbiased estimator of its DOF, i.e.

$$\hat{df} = \operatorname{tr}(\partial\hat{\mu}(y)) \quad \text{and} \quad \mathbb{E}_\varepsilon(\hat{df}) = df.$$

An unbiased estimator of the DOF provides an unbiased estimate for the prediction risk of $\hat{\mu}(y)$ through e.g. the Mallow's C_p (Mallows, 1973), the AIC (Akaike, 1973), the SURE (Stein, 1981) or the GCV (Golub et al., 1979). These quantities can serve as model selection criteria to assess the accuracy of a candidate model.

1.3. Previous Works

In the special case of standard Lasso with a linearly independent design, (Zou et al., 2007) show that the number of nonzero coefficients is an unbiased estimate for the degrees of freedom. This work is extended in (Dossal et al.) to an arbitrary design matrix. The DOF of the analysis sparse regularization (a.k.a. generalized Lasso in statistics) is studied in (Tibshirani & Taylor, 2012; Vaiter et al., 2012). A formula of an estimate of the DOF for the group Lasso when the design is orthogonal within each group is conjectured in (Yuan & Lin, 2006). (Kato, 2009) studies the DOF of a generalization of the Lasso where now the regression coefficients are constrained to a closed convex set. He provides an unbiased estimate of the DOF for the constrained version of the group Lasso under the same orthogonality assumption on $\mathbf{X}$ as (Yuan & Lin, 2006). An estimate of the DOF for the group Lasso is also given in (Solo & Ulfarsson, 2010) by an heuristic proof in the full column rank case, but its unbiasedness is not proved.

1.4. Contributions

This paper proves a general result (Theorem 1) on the variations of the solutions to $\mathcal{P}_\lambda(y)$ with respect to the observation/response vector y. With such a result on hand, Theorem 2 provides a provably unbiased estimate of the DOF. These contributions are detailed in Sections 2.1 and 2.2 below. The proofs are deferred to Section 3 awaiting inspection by the interested reader.

1.5. Notations

We start by some notations used in the sequel. We extend the notion of support, commonly used in sparsity by defining the $\mathcal{B}$-support $\operatorname{supp}_{\mathcal{B}}(\beta)$ of $\beta \in \mathbb{R}^N$ as

$$\operatorname{supp}_{\mathcal{B}}(\beta) = \{b \in \mathcal{B} \setminus \|\beta_b\| \neq 0\}.$$

The size of $\operatorname{supp}_{\mathcal{B}}(\beta)$ is defined as $|\operatorname{supp}_{\mathcal{B}}(\beta)| = \sum_{b \in \mathcal{B}} |b|$. We denote by $\mathbf{X}_I$, where I is a $\mathcal{B}$-support, the matrix formed by the columns $\mathbf{X}_i$ where i is an element of $b \in I$. We introduce the following block-diagonal operator

$$\delta_\beta : v \in \mathbb{R}^{|I|} \mapsto (v_b/\|\beta_b\|)_{b \in \mathcal{B}} \in \mathbb{R}^{|I|}.$$

and

$$P_\beta : v \in \mathbb{R}^{|I|} \mapsto (P_{\beta_b^\perp}(v_b))_{b \in \mathcal{B}} \in \mathbb{R}^{|I|}$$

where $P_{\beta_b^\perp}$ is the projector orthogonal to x_b. For any operator A, we denote A^{T} its adjoint.

2. Main results

Note that as the $\mathbf{X}$ is assumed full column rank, $\mathcal{P}_\lambda(y)$ has exactly one global minimizer $\beta^\star(y)$. Hence, we define the single-valued mapping $y \mapsto \beta^\star(y)$.

2.1. Local Parameterization

Let I be the $\mathcal{B}$-support of some vector β. For any block $b \notin I$, we define

$$\mathcal{H}_{I,b} = \left\{ y \in \mathbb{R}^Q \setminus \exists \beta : \forall c \in I, (\|\mathbf{X}_b^{\mathrm{T}} r\|, \mathbf{X}_c^{\mathrm{T}} r) = (\lambda, \lambda \frac{\beta_c}{\|\beta_c\|}) \right\}.$$

where $r = y - \mathbf{X}_I \beta$.

Definition 1. *The transition space $\mathcal{H}$ is defined as*

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

where $\mathcal{I}$ is the set of sub-sets of $\{0, \ldots, N-1\}$ obtained as unions of blocks.

We prove the following sensitivity theorem

Theorem 1. *Let $y \notin \mathcal{H}$, and $I = \operatorname{supp}_{\mathcal{B}}(\beta^\star(y))$ the $\mathcal{B}$-support of $\beta^\star(y)$. There exists a neighborhood $\mathcal{O}$ of y such that*

1. *the $\mathcal{B}$-support of $\beta^\star(y)$ is constant on $\mathcal{O}$, i.e.*

$$\forall y \in \mathcal{O}, \quad \operatorname{supp}_{\mathcal{B}}(\beta^\star(\bar{y})) = I,$$

2. *the mapping $\beta^\star$ is $\mathcal{C}^1$ on $\mathcal{O}$ and its differential is such that*

$$[\partial \beta^\star(\bar{y})]_{I^c} = 0 \quad and \quad [\partial \beta^\star(\bar{y})]_I = d(y), \quad (1)$$

where

$$d(y) = \left(\mathbf{X}_I^{\mathrm{T}} \mathbf{X}_I + \lambda \delta_{\beta^\star(y)} \circ P_{\beta^\star(y)}\right)^{-1} \mathbf{X}_I^{\mathrm{T}}$$

and

$$I^c = \{b \in \mathcal{B} \setminus b \notin I\}.$$

2.2. Degrees of Freedom

We consider the estimator $\hat{\mu}(y) = \mathbf{X}\beta^\star(y)$.

Theorem 2. *Let $\lambda > 0$. The mapping $y \mapsto \hat{\mu}(y)$ is of class $\mathcal{C}^1$ on $\mathbb{R}^Q \setminus \mathcal{H}$ and,*

$$\operatorname{div}(\hat{\mu}(y)) = \operatorname{tr}(\mathbf{X}_I d(y)), \quad (2)$$

where $\beta^\star(y)$ is the solution of $\mathcal{P}_\lambda(y)$ and $I = \operatorname{supp}_{\mathcal{B}}(\beta^\star(y))$. Moreover, the set $\mathcal{H}$ has zero Lebesgue measure, thus if $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_Q)$, equation (2) is an unbiased estimate of the DOF of the group Lasso.

We specify this result for the Block Soft Thresholding

Corollary 1. *If $\mathbf{X} = \mathrm{Id}$, one has*

$$\hat{df} = |J| - \lambda \sum_{b \subset J} \frac{|b| - 1}{\|y_b\|}$$

where $J = \bigcup \{b \in \mathcal{B} \setminus \|y_b\| > \lambda\}$.

3. Proofs

This section details the proofs of our results. We introduce the following normalization operator

$$\mathcal{N}(\beta_I) = v \quad \text{where} \quad \forall b \in I, v_b = \frac{\beta_b}{\|\beta_b\|}.$$

We use the following lemma in our proofs which is a straightforward consequence of the first order necessary and sufficient condition of a minimizer of the group Lasso problem $\mathcal{P}_\lambda(y)$.

Lemma 1. *A vector $\beta \in \mathbb{R}^N$ is the solution of $\mathcal{P}_\lambda(y)$ if, and only if, these two conditions holds*

1. *On the $\mathcal{B}$-support $I = \operatorname{supp}_\mathcal{B}(\beta)$,*

$$\mathbf{X}_I^{\mathrm{T}}(y - \mathbf{X}_I \beta_I) = \lambda \mathcal{N}(\beta_I).$$

2. *For all $b \in \mathcal{B}$ such that $b \notin I$, one has*

$$\|\mathbf{X}_b^{\mathrm{T}}(y - \mathbf{X}_I \beta_I)\| \leqslant \lambda.$$

A proof of this lemma can be found in (Bach, 2008).

3.1. Proof of Theorem 1

We will use this following lemma.

Lemma 2. *Let $\beta \in \mathbb{R}^N$ and $\lambda > 0$. Then $\mathbf{X}_I^{\mathrm{T}}\mathbf{X}_I + \lambda\delta_\beta \circ P_\beta$ is invertible.*

Proof. We prove that $\mathbf{X}_I^{\mathrm{T}}\mathbf{X}_I + \lambda\delta_\beta \circ P_\beta$ is symmetric positive definite. Remark that $\mathbf{X}_I^{\mathrm{T}}\mathbf{X}_I$ is a positive definite matrix. Moreover, $\delta_\beta \circ P_\beta$ is symmetric positive semi-definite since both δ_β and P_β are SDP and commute. We conclude using the fact that the sum of a symmetric positive definite matrix and a symmetric positive semi-definite matrix is symmetric positive definite. $\square$

Let $y \notin \mathcal{H}$. We define $I = \operatorname{supp}_\mathcal{B}(\beta^\star(y))$ the $\mathcal{B}$-support of the solution $\beta^\star(y)$ of $\mathcal{P}_\lambda(y)$. We define the following mapping

$$\Gamma(\alpha_I, y) = \mathbf{X}_I^{\mathrm{T}}(\mathbf{X}_I \alpha_I - y) + \lambda \mathcal{N}(\alpha_I).$$

Observe item 1 of Lemma 1 is equivalent to $\Gamma([\beta^\star(y)]_I, y) = 0$.

Our proof is done in three steps. We first (**1.**) prove there exists a mapping $\bar{y} \mapsto \beta(\bar{y})$ such that for every element of a neighborhood of y one has $\Gamma([\beta(\bar{y})]_I, \bar{y}) = 0$ and $[\beta(\bar{y})]_{I^c} = 0$. Then, we prove (**2.**) that $\beta(\bar{y}) = \beta^\star(\bar{y})$ is the solution of $\mathcal{P}_\lambda(\bar{y})$ in a neighborhood of y. Finally, we obtain (**3.**) equation (1) from the implicit function theorem.

1. The derivative of Γ with respect to the first variable reads on $(\mathbb{R}^{|I|} \setminus U) \times \mathbb{R}^Q$

$$\partial_1 \Gamma(\beta_I, y) = \mathbf{X}_I^{\mathrm{T}}\mathbf{X}_I + \lambda\delta_{\beta_I} \circ P_{\beta_I}.$$

where $U = \{\alpha \in \mathbb{R}^{|I|} \setminus \exists b \in I : \alpha_b = 0\}$. The mapping $\partial_1 \Gamma$ is invertible according to Lemma 2. Hence, using the implicit function theorem, there exists a neighborhood $\tilde{\mathcal{O}}$ of y such that we can define a mapping $\beta_I : \mathcal{O} \to \mathbb{R}^{|I|}$ of class $\mathcal{C}^1$ over $\tilde{\mathcal{O}}$ that satisfies for $\bar{y} \in \tilde{\mathcal{O}}$

$$\Gamma(\beta_I(\bar{y}), \bar{y}) = 0 \quad \text{and} \quad \beta_I(y) = [\beta^\star(y)]_I.$$

We then extend β_I on I^c as $\beta_{I^c}(\bar{y}) = 0$, which defines a mapping $\beta(\bar{y}) : \tilde{\mathcal{O}} \to \mathbb{R}^N$.

2. Writting the first-order conditions on $\beta^\star(y)$ on the blocks not included in the $\mathcal{B}$-support, one has

$$\forall b \notin I, \quad \|\mathbf{X}_b^{\mathrm{T}}(y - \mathbf{X}_I [\beta^\star(y)]_I)\| \leqslant \lambda.$$

Suppose there exists $b \notin I$ such that $\|\mathbf{X}_b^{\mathrm{T}}(\mathbf{X}_I [\beta^\star(y)]_I - y)\| = \lambda$. Then $y \in \mathcal{H}_{I,b}$ since

$$\|\mathbf{X}_b^{\mathrm{T}} r\| = \lambda \quad \text{and} \quad \mathbf{X}_c^{\mathrm{T}} r = \lambda \frac{[\beta^\star(y)]_c}{\|[\beta^\star(y)]_c\|},$$

for $r = y - \mathbf{X}_I [\beta^\star(y)]_I$, which is a contradiction with $y \notin \mathcal{H}$. Hence,

$$\forall b \notin I, \quad \|\mathbf{X}_b^{\mathrm{T}}(\mathbf{X}_I [\beta^\star(y)]_I - y)\| < \lambda.$$

By continuity of $\bar{y} \mapsto \beta_I(\bar{y})$ and since $\beta_I(y) = [\beta^\star(y)]_I$, we can find a neighborhood $\mathcal{O}$ included in $\tilde{\mathcal{O}}$ such that for every $\bar{y} \in \mathcal{O}$, one has

$$\forall b \notin I, \quad \|\mathbf{X}_b^{\mathrm{T}}(\mathbf{X}_I \beta_I(\bar{y}) - \bar{y})\| \leqslant \lambda.$$

Moreover, by definition of the mapping β_I, one has

$$\mathbf{X}_I^{\mathrm{T}}(y - \mathbf{X}_I \beta_I(\bar{y})) = \lambda \mathcal{N}(\beta_I(\bar{y})) \text{ and } \operatorname{supp}_\mathcal{B}(\beta_I(\bar{y})) = I.$$

According to Lemma 1, the vector $\beta(\bar{y})$ is solution of $\mathcal{P}_\lambda(\bar{y})$. Since $\mathcal{P}_\lambda(\bar{y})$ admits a unique solution, $\beta^\star(\bar{y}) = \beta(\bar{y})$ for every $\bar{y} \in \mathcal{O}$.

3. Using the implicit function theorem, one obtains the derivative of $[\beta(y)]_I$ as

$$[\partial\beta^\star(\bar{y})]_I = -\left(\partial_1 \Gamma([\beta^\star(y)]_I, y)\right)^{-1} \circ \left(\partial_2 \Gamma([\beta^\star(y)]_I, y)\right)$$

where $\partial_2 \Gamma([\beta^\star(y)]_I, y) = -\mathbf{X}_I^{\mathrm{T}}$, which leads us to (1).

3.2. Proof of Theorem 2

We define for each $b \in \mathcal{B}$

$$\mathcal{H}_{I,b}^\uparrow = \Big\{ (r, \beta) \in \mathbb{R}^Q \times \mathbb{R}^{|I|} \setminus$$

$$\|\mathbf{X}_b^{\mathrm{T}} r\| = 1 \quad \text{and} \quad \forall g \in I, \quad \mathbf{X}_g^{\mathrm{T}} r = \frac{\beta_g}{\|\beta_g\|} \Big\}.$$

We prove (**1.**) that for each $b \in I$, $\mathcal{H}_{I,b}^{\uparrow}$ is a manifold of dimension $Q - 1$. Then (**2.**) we prove that $\mathcal{H}$ if of measure zero with respect to the Lebesgue measure on $\mathbb{R}^Q$. Finally (**3.**), we prove that $\hat{df}$ is an unbiased estimate of the DOF.

1. Note that $\mathcal{H}_{I,b}^{\uparrow} = \psi^{-1}(\{0\})$ where

$$\psi(r, \beta) = \left(\|\mathbf{X}_b^{\mathrm{T}} r\|^2 - 1, \mathbf{X}_I^{\mathrm{T}} r - \mathcal{N}(\beta) \right).$$

Remark that the adjoint of the differential of ψ

$$(\partial \psi)^{\mathrm{T}}(r, \beta) = \left(\begin{array}{c|c} 2\mathbf{X}_b \mathbf{X}_b^{\mathrm{T}} r & \mathbf{X}_I \\ \hline 0 & \lambda \delta_\beta \circ P_\beta \end{array} \right)$$

has full rank. Indeed, consider the matrix $A = (2\mathbf{X}_b \mathbf{X}_b^{\mathrm{T}} r | \mathbf{X}_I)$. Let $\alpha = (s, u)^{\mathrm{T}} \in \mathbb{R} \times \mathbb{R}^{|I|}$ such that $A\alpha = 0$. Then $(2s\mathbf{X}_b r, u)^{\mathrm{T}} \in \mathrm{Ker}(\mathbf{X}_b | \mathbf{X}_I)$. Since $(\mathbf{X}_b | \mathbf{X}_I)$ has full rank, we conclude that $\alpha = 0$. As a consequence, $\partial \psi(r, \beta)$ is non-degenerated. Finally, $\mathcal{H}_{I,b}^{\uparrow}$ is a manifold of dimension $Q - 1$.

2. We prove that $\mathcal{H}_{I,b}$ is of Hausdorff dimension less or equal to $Q - 1$. Consider the following mapping

$$\varphi : \left\{ \begin{array}{ccc} \mathbb{R}^Q \times \mathbb{R}^{|I|} & \to & \mathbb{R}^Q \times \mathbb{R}^{|I|} \\ (r, \beta) & \mapsto & (r + \mathbf{X}_I \beta, \mathbf{X}_I^{\mathrm{T}} r) \end{array} \right. .$$

The mapping φ is a $\mathcal{C}^1$-diffeomorphism between $\mathbb{R}^Q \times \mathbb{R}^{|I|}$ and itself. Thus, $A = \varphi(\mathcal{H}_{I,b}^{\uparrow})$ is a manifold of dimension $Q - 1$. We now introduce the projection

$$\pi : \left\{ \begin{array}{ccc} A & \to & \mathbb{R}^Q \\ (y, \alpha) & \mapsto & y \end{array} \right. .$$

Observe that $\mathcal{H}_{I,b} = \pi(A)$. According to Hausdorff measure properties (Rogers, 1998), since π is 1-Lipschitz, the Hausdorff dimension of $\pi(A)$ is less or equal to the Hausdorff dimension of A which is the dimension of A as a manifold, namely $Q - 1$. Hence, the measure of $\mathcal{H}_{I,b}$ w.r.t the Lebesgue measure of $\mathbb{R}^Q$ is zero.

3. According to Theorem 1, $y \mapsto \beta^\star(y)$ is $\mathcal{C}^1$ on $\mathbb{R}^Q \setminus \mathcal{H}$. Composing by $\mathbf{X}$ gives that $\hat{\mu}$ is differentiable almost everywhere, hence weakly differentiable. Moreover, taking the divergence of the differential (1), one obtains (2). This formula is verified almost everywhere, outside the set $\mathcal{H}$. Stein's Lemma (Stein, 1981) gives the unbiased propertiy of our estimator $\hat{df}$ of the DOF.

3.3. Proof of Corollary 1

When $\mathbf{X} = \mathrm{Id}$, the solution of $\mathcal{P}_\lambda(y)$ is a block soft thresholding

$$[\beta^\star(y)]_b = \begin{cases} 0 & \text{if } \|y_b\| \leqslant \lambda \\ (1 - \frac{\lambda}{\|y_b\|})y_b & \text{otherwise} \end{cases} . \qquad (3)$$

For every $b \in J$, we differentiate equation (3)

$$[\partial \beta^\star(y)]_b : \alpha \in \mathbb{R}^{|b|} \mapsto \alpha - \frac{\lambda}{\|y_b\|} P_{y_b^{\perp}}(\alpha).$$

Since $P_{y_b^{\perp}}(\alpha)$ is a projector on space of dimension $|b| - 1$, one has $\mathrm{tr}(P_{y_b^{\perp}}) = |b| - 1$.

References

Akaike, H. Information theory and an extension of the maximum likelihood principle. In *Second international symposium on information theory*, volume 1, pp. 267–281. Springer Verlag, 1973.

Bach, F.R. Consistency of the group lasso and multiple kernel learning. *The Journal of Machine Learning Research*, 9:1179–1225, 2008.

Bakin, S. Adaptive regression and model selection in data mining problems, 1999. Thesis (Ph.D.)–Australian National University, 1999.

Dossal, C., Kachour, M., Fadili, J., Peyré, G., and Chesneau, C. The degrees of freedom of the Lasso for general design matrix. Technical report. URL http://hal.archives-ouvertes.fr/hal-00638417.

Efron, B. How biased is the apparent error rate of a prediction rule? *Journal of the American Statistical Association*, 81(394):461–470, 1986.

Golub, G.H., Heath, M., and Wahba, G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, pp. 215–223, 1979.

Kato, K. On the degrees of freedom in shrinkage estimation. *Journal of Multivariate Analysis*, 100(7):1338–1352, 2009.

Mallows, C. L. Some comments on cp. *Technometrics*, 15 (4):661–675, 1973.

Rogers, C.A. *Hausdorff measures*. Cambridge Univ Pr, 1998.

Solo, V. and Ulfarsson, M. Threshold selection for group sparsity. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 3754–3757. IEEE, 2010.

Stein, C.M. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9(6):1135–1151, 1981.

Tibshirani, R. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B. Methodological*, 58(1):267–288, 1996.

Tibshirani, R. J. and Taylor, J. Degrees of freedom in lasso problems. Technical report, arXiv:1111.0653, 2012.

Vaiter, S., Deledalle, C., Peyré, G., Fadili, J., and Dossal, C. Local behavior of sparse analysis regularization: Applications to risk estimation. Technical report, Preprint Hal-00687751, 2012. URL http://hal.archives-ouvertes.fr/hal-00687751/.

Yuan, M. and Lin, Y. Model selection and estimation in regression with grouped variables. *J. of The Roy. Stat. Soc. B*, 68(1):49–67, 2006.

Zou, H., Hastie, T., and Tibshirani, R. On the "degrees of freedom" of the lasso. *The Annals of Statistics*, 35(5):2173–2192, 2007.