

# Extended Oracle Properties of Adaptive Lasso Estimators

Lorenzo Camponovo

SUPSI—Dipartimento Tecnologie Innovative (DTI), Viganello, Switzerland

Email: [lorenzo.camponovo@supsi.ch](mailto:lorenzo.camponovo@supsi.ch)

**How to cite this paper:** Camponovo, L. (2022) Extended Oracle Properties of Adaptive Lasso Estimators. *Open Journal of Statistics*, 12, 210-215. <https://doi.org/10.4236/ojs.2022.122015>

**Received:** March 15, 2022

**Accepted:** April 15, 2022

**Published:** April 18, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

We study the asymptotic properties of adaptive lasso estimators when some components of the parameter of interest  $\beta$  are strictly different than zero, while other components may be zero or may converge to zero with rate  $n^{-\delta}$ , with  $\delta > 0$ , where  $n$  denotes the sample size. To achieve this objective, we analyze the convergence/divergence rates of each term in the first-order conditions of adaptive lasso estimators. First, we derive conditions that allow selecting tuning parameters in order to ensure that adaptive lasso estimates of  $n^{-\delta}$ -components indeed collapse to zero. Second, in this case, we also derive asymptotic distributions of adaptive lasso estimators for nonzero components. When  $\delta > 1/2$ , we obtain the usual  $n^{1/2}$ -asymptotic normal distribution, while when  $0 < \delta \leq 1/2$ , we show  $n^\delta$ -consistency combined with (biased)  $n^{1/2-\delta}$ -asymptotic normality for nonzero components. We call these properties, *Extended Oracle Properties*. These results allow practitioners to exclude in their model the asymptotically negligible variables and make inferences on the asymptotically relevant variables.

## Keywords

Adaptive Lasso, Asymptotics, Oracle Properties

## 1. Introduction

Consider the linear regression model

$$y_t = \sum_{i=1}^p \beta_i x_{t,i} + \varepsilon_t, \quad t = 1, \dots, n, \quad (1)$$

where  $y_t$  is the response variable,  $x_t = (x_{t,1}, \dots, x_{t,p})^T$  is a vector of covariates,  $\varepsilon_t$  is the error term, and  $\beta = (\beta_1, \dots, \beta_p)$  is the unknown parameter of interest. We focus on adaptive lasso estimators defined as the solution of

$$\hat{\beta}_n = \arg \min_u \sum_{t=1}^n (y_t - u^\top x_t)^2 + \sum_{i=1}^p \lambda_{n,i}^\gamma |u_i|, \quad (2)$$

where  $\gamma > 0$ ,  $\lambda_{n,i} = 1/|\bar{\beta}_{n,i}|$ , and  $\bar{\beta}_n = (\bar{\beta}_{n,1}, \dots, \bar{\beta}_{n,p})^\top$  is  $n^\alpha$ -consistent preliminary estimator of  $\beta$ , with  $\alpha > 0$ . Adaptive lasso estimators have been introduced in [1], as refinements of the lasso approach proposed in [2]. Indeed, the adaptive lasso fulfils the oracle properties in the sense introduced in [3]. Note that in the seminal definition of adaptive lasso, the penalization term in Equation (2) is multiplied by a tuning parameter  $\lambda_n > 0$ . In our study, we simply set  $\lambda_n = 1$ , since our asymptotic results do not require that  $\lambda_n \rightarrow \infty$ .

Since the introduction in [2], the statistical properties of lasso estimators, solution of (2) with  $\lambda_{n,i} = 0$ , have been analyzed in several studies. As shown in [2], the lasso may combine estimation and variable selection. However, as pointed out in [4] and [5], among others, to achieve correct model selection lasso estimators require certain *irrepresentable conditions*. In general, the lasso selects more predictive variables than the number of true variables. To overcome this problem, [1] introduces adaptive lasso estimators defined in (2), and shows that they can achieve the oracle properties; see, e.g., [2] [6] and [7]. More precisely, adaptive lasso estimators may combine efficient parameter estimation and correct variable selection in one step. Following this intuition, recently several studies propose penalized estimators with these desirable properties in different context; see, e.g., [8] [9] [10] and [11]. We refer to [12] and [13] for a detailed discussion of lasso estimators.

In this paper, we extend the analysis of the statistical properties of lasso estimators by studying the asymptotic properties of adaptive lasso estimators when some components of the unknown parameter  $\beta$  are strictly different than zero, while other components may be zero or may converge to zero with rate  $n^{-\delta}$ , with  $\delta > 0$ . To achieve this objective, we analyze the convergence/divergence rates of each term in the first-order conditions of adaptive lasso estimators. We derive conditions that allow selecting the tuning parameter  $\gamma$  as a function of the  $n^\alpha$ -consistent estimator  $\bar{\beta}_n$ , in order to ensure that adaptive lasso estimates of  $n^{-\delta}$ -coefficients of the unknown parameter  $\beta$  indeed collapse to zero. Furthermore, in this case, we also derive the asymptotic properties of adaptive lasso estimators for nonzero components. These results allow practitioners to exclude in their model the asymptotically negligible variables and make inferences on the asymptotically relevant variables.

## 2. Asymptotic Properties

Before presenting the main results, we formally introduce the following assumptions.

### Assumption 1

- 1) The parameter  $\beta = (\beta_A^\top, \beta_{A^c}^\top)^\top$ , where  $\beta_A \neq 0$ , and  $\beta_{A^c} = cn^{-\delta}$ , with  $c \in \mathbb{R}$  and  $\delta > 0$ .
- 2) The preliminary estimator  $\bar{\beta}_n$  satisfies  $n^\alpha (\bar{\beta}_n - \beta) = O_p(1)$ , for some

$\alpha > 0$ .

Assumption 1 1) implies that some components of the unknown parameter  $\beta$  are strictly different than zero, while other components may be zero or may converge to zero with rate  $\delta > 0$ . Assumption 1 2) simply implies that  $\bar{\beta}_n$  is  $n^\alpha$ -consistent estimator of  $\beta$ .

Furthermore, consider also following high-level assumptions.

**Assumption 2**

- 1)  $n^{-1} \sum_{t=1}^n x_t x_t^T$  converges in probability to a positive definite matrix  $D$ , as  $n \rightarrow \infty$ .
- 2)  $n^{-1/2} \sum_{t=1}^n x_t \varepsilon_t$  converges weakly to normal with mean 0 and variance  $\Omega$ , as  $n \rightarrow \infty$ .

Assumption 2 provides a set of high-level conditions that are typically satisfied in linear regression models (1) when  $p$  is fixed and  $n$  is large. Currently, we are not interested in analyzing high-dimensional models. We believe that first, it is important to understand the asymptotic properties of adaptive lasso estimators of  $n^{-\delta}$ -coefficients in standard settings with standard (high-level) assumptions. The extension to high-dimensional linear regression models is left for future research.

To simplify the exposition of our results, we consider the linear regression model (1) with  $p = 2$ ,  $\beta_1 = c_1 \neq 0$ ,  $\beta_2 = c_2 n^{-\delta}$ ,  $c_2 \neq 0$ , and  $\delta > 0$ . We introduce the function

$$f(u_1, u_2) = \sum_{t=1}^n (y_t - u_1 x_{t,1} - u_2 x_{t,2})^2 + |\bar{\beta}_{n,1}|^{-\gamma} \cdot |u_1| + |\bar{\beta}_{n,2}|^{-\gamma} \cdot |u_2|, \tag{3}$$

where  $\gamma > 0$ . For  $i = 1, 2$ , let  $f_i(u_1, u_2) = n^{-1/2} \frac{\partial f}{\partial u_i}(u_1, u_2)$ . Then, we have,

$$f_1(u_1, u_2) = -2 \left( n^{-1/2} \sum_{t=1}^n x_{t,1} \varepsilon_t + n^{1/2} (\beta_1 - u_1) n^{-1} \sum_{t=1}^n x_{t,1}^2 + n^{1/2} (\beta_2 - u_2) n^{-1} \sum_{t=1}^n x_{t,1} x_{t,2} \right) + n^{-1/2} |\bar{\beta}_{n,1}|^{-\gamma} \text{sign}(u_1), \tag{4}$$

$$f_2(u_1, u_2) = -2 \left( n^{-1/2} \sum_{t=1}^n x_{t,2} \varepsilon_t + n^{1/2} (\beta_1 - u_1) n^{-1} \sum_{t=1}^n x_{t,1} x_{t,2} + n^{1/2} (\beta_2 - u_2) n^{-1} \sum_{t=1}^n x_{t,2}^2 \right) + n^{-1/2} |\bar{\beta}_{n,2}|^{-\gamma} \text{sign}(u_2). \tag{5}$$

Note that since  $\beta_1 \neq 0$ , then  $n^{-1/2} |\bar{\beta}_{n,1}|^{-\gamma} \text{sign}(u_1) = o_p(1)$ . On the other hand, since  $\beta_2 = c_2 n^{-\delta}$ , then when  $u_2 \neq 0$ ,  $n^{-1/2} |\bar{\beta}_{n,2}|^{-\gamma} \text{sign}(u_2) = O_p(n^{\gamma \cdot \min(\delta, \alpha) - 1/2})$ . It turns out that for large  $n$ , when

$$\gamma \cdot \min(\delta, \alpha) > 1, \tag{6}$$

then, the dominant term in Equation (5) is  $n^{-1/2} |\bar{\beta}_{n,2}|^{-\gamma} \text{sign}(u_2)$ . Furthermore, the sign of  $u_2$  fully determines the sign of  $f_2(u_1, u_2)$ . Indeed,  $f_2(u_1, u_2) < 0$ , when  $u_2 < 0$ , while  $f_2(u_1, u_2) > 0$ , when  $u_2 > 0$ . Therefore,  $\hat{\beta}_{n,2} = 0$ , where  $\hat{\beta}_{n,2}$  is the adaptive lasso estimator of  $\beta_2$ .

In the next step, we derive the asymptotic properties of the adaptive lasso estimator  $\hat{\beta}_{n,1}$  of  $\beta_1$ , when  $\hat{\beta}_{n,2} = 0$ . We consider first the case  $\delta > 1/2$ . Then, in Equation (4), besides  $n^{-1/2} \left| \hat{\beta}_{n,1} \right|^{-\gamma} \text{sign}(u_1) = o_p(1)$ , also  $n^{1/2} (\beta_2 - \hat{\beta}_{n,2}) n^{-1} \sum_{t=1}^n x_{t,1} x_{t,2} = o_p(1)$ . Therefore, for large  $n$ ,  $f_1(u_1, u_2) = 0$ , when,

$$n^{1/2} (u_1 - \beta_1) = \left( n^{-1} \sum_{t=1}^n x_{t,1}^2 \right)^{-1} n^{-1/2} \sum_{t=1}^n x_{t,1} \varepsilon_t + o_p(1).$$

It turns out that by Assumption 2, for the adaptive lasso estimator  $\hat{\beta}_{n,1}$  we obtain the usual  $n^{1/2}$ -asymptotic normal distribution.

Finally, we consider the case  $0 < \delta \leq 1/2$ . Note that in this case, when  $\hat{\beta}_{n,2} = 0$ , then  $n^{1/2} (\beta_2 - \hat{\beta}_{n,2}) = n^{1/2-\delta} c_2$ . It turns out that in Equation (4), for large  $n$ ,  $f_1(u_1, u_2) = 0$ , when

$$n^{-1/2} \sum_{t=1}^n x_{t,1} \varepsilon_t + n^{1/2} (\beta_1 - u_1) n^{-1} \sum_{t=1}^n x_{t,1}^2 + n^{1/2-\delta} c_2 n^{-1} \sum_{t=1}^n x_{t,1} x_{t,2} = o_p(1). \quad (7)$$

Note that we can rearrange the left term of Equation (7) as follows,

$$n^{-1/2} \sum_{t=1}^n x_{t,1} \varepsilon_t + n^{-1} \sum_{t=1}^n x_{t,1}^2 n^{1/2-\delta} \left( n^\delta (\beta_1 - u_1) + c_2 \left( n^{-1} \sum_{t=1}^n x_{t,1}^2 \right)^{-1} n^{-1} \sum_{t=1}^n x_{t,1} x_{t,2} \right). \quad (8)$$

Consequently, we have,

$$\begin{aligned} & n^{1/2-\delta} \left( n^\delta (u_1 - \beta_1) - c_2 \left( n^{-1} \sum_{t=1}^n x_{t,1}^2 \right)^{-1} n^{-1} \sum_{t=1}^n x_{t,1} x_{t,2} \right) \\ &= \left( n^{-1} \sum_{t=1}^n x_{t,1}^2 \right)^{-1} n^{-1/2} \sum_{t=1}^n x_{t,1} \varepsilon_t + o_p(1). \end{aligned} \quad (9)$$

Under Assumption 2, Equation (9) establishes  $n^\delta$ -consistency combined with (biased)  $n^{1/2-\delta}$ -asymptotic normality for the adaptive lasso estimator  $\hat{\beta}_{n,1}$  of  $\beta_1$ .

Obviously, this approach can be easily applied also to the general case  $p > 2$ , under Assumption 1. These results suggest the introduction of following *Extended Oracle Properties*.

#### Extended Oracle Properties

Consider the linear regression model (1) with Assumption 1. Then, a statistical procedure fulfils the Extended Oracle Properties when:

- 1) Identifies the (asymptotically) nonzero components  $\beta_A$ .
- 2) For  $\delta > 1/2$ , ensures the optimal  $n^{1/2}$ -asymptotic normality for the (asymptotically) nonzero components  $\beta_A$ .
- 3) For  $0 < \delta \leq 1/2$ , ensures  $n^\delta$ -consistency combined with (biased)  $n^{1/2-\delta}$ -asymptotic normality for the (asymptotically) nonzero components  $\beta_A$ .

From a practical point of view, in Equation (6) we provide conditions that allow to select the tuning exponent  $\gamma$  as a function of the preliminary  $n^\alpha$ -consistent estimator, in order to ensure that adaptive lasso estimates of  $n^{-\delta}$ -components indeed collapse to zero. For instance, for  $\delta = \alpha = 1/2$ , we simply have  $\gamma > 2$ . These conditions are quite intuitive. Indeed, when  $\delta$  decreases (large neigh-

borhoods of zero), then  $\gamma$  increases (stronger penalization). These results may be useful for practitioner that would like to include in their model only the (asymptotically) relevant variables. From a theoretical point of view to the best of our knowledge, we have never seen  $n^\delta$ -consistency combined with (biased)  $n^{1/2-\delta}$ -asymptotic normality. Since  $\delta$  is unknown, inference based on these results seems quite challenging. Also modified residual [14] and pairs [15] bootstrap seem not appropriate. Probably, multiplicative bootstrap procedures recently introduced in penalized regression models may help also in this case.

### 3. Conclusion

In this paper, we extend the analysis of the statistical properties of lasso estimators by studying the asymptotic properties of adaptive lasso estimators when some components of the unknown parameter  $\beta$  are strictly different than zero, while other components may be zero or may converge to zero with rate  $n^{-\delta}$ , with  $\delta > 0$ . To achieve this objective, we analyze the convergence/divergence rates of each term in the first-order conditions of adaptive lasso estimators. First, we derive conditions that allow selecting tuning parameters of adaptive lasso estimators in order to ensure that adaptive lasso estimates of  $n^{-\delta}$ -components of the unknown parameter of interest  $\beta$  indeed collapse to zero when  $\delta > 0$ . Second, in this case, we also derive asymptotic distributions of adaptive lasso estimators for strictly nonzero components. When  $\delta > 1/2$ , we obtain the usual  $n^{1/2}$ -asymptotic normal distribution, while when  $0 < \delta \leq 1/2$ , we show  $n^\delta$ -consistency combined with (biased)  $n^{1/2-\delta}$ -asymptotic normality for nonzero components. Several Monte Carlo simulations about 1) selection of the tuning exponent  $\gamma$ , and 2)  $n^\delta$ -consistency combined with (biased)  $n^{1/2-\delta}$ -asymptotic normality confirm the theoretical results and are available from the author. These results allow practitioners to exclude in their model the asymptotically negligible variables and make inferences on the asymptotically relevant variables.

### Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

### References

- [1] Zou, H. (2006) The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, **101**, 1418-1429.  
<https://doi.org/10.1198/016214506000000735>
- [2] Tibshirani, R.J. (1996) Regression Analysis and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, **21**, 267-288.  
<https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- [3] 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.  
<https://doi.org/10.1198/016214501753382273>
- [4] Zhao, P. and Yu, B. (2006) On Model Selection Consistency of Lasso. *Journal of*

- Machine Learning Research*, **7**, 2541-2563.
- [5] Meinshausen, N. and Yu, B. (2009) Lasso-Type Recovery of Sparse Representation for High-Dimensional Data. *Annals of Statistics*, **37**, 246-270. <https://doi.org/10.1214/07-AOS582>
- [6] Fan, J. and Peng, H. (2004) On Nonconcave Penalized Likelihood with Diverging Number of Parameters. *Annals of Statistics*, **32**, 928-961. <https://doi.org/10.1214/009053604000000256>
- [7] Zou, H. and Zhang, H. (2009) On the Adaptive Elastic Net with a Diverging Number of Parameters. *Annals of Statistics*, **37**, 1733-1751. <https://doi.org/10.1214/08-AOS625>
- [8] Caner, M. (2009) Lasso-Type GMM Estimator. *Econometric Theory*, **25**, 270-290. <https://doi.org/10.1017/S0266466608090099>
- [9] Caner, M. and Zang, H.H. (2014) Adaptive Elastic Net for Generalized Methods of Moment. *Journal of Business and Economic Statistics*, **32**, 30-47. <https://doi.org/10.1080/07350015.2013.836104>
- [10] Audrino, F. and Camponovo, L. (2018) Oracle Properties, Bias Correction, and Inference of the Adaptive Lasso for Time Series Extremum Estimators. *Journal of Time Series Analysis*, **39**, 111-128. <https://doi.org/10.1111/jtsa.12270>
- [11] Camponovo, L. (2020) Bootstrap Inference for Penalized GMM Estimators with Oracle Properties. *Econometric Reviews*, **39**, 362-372. <https://doi.org/10.1080/07474938.2019.1630076>
- [12] Bühlmann, P. and van de Geer, S. (2011) *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Series in Statistics, Springer, Heidelberg. <https://doi.org/10.1007/978-3-642-20192-9>
- [13] Zhang, C.H. and Zhang, T. (2012) A General Theory of Concave Regularization for High-Dimensional Sparse Estimation Problems. *Statistical Science*, **27**, 576-593. <https://doi.org/10.1214/12-STS399>
- [14] Chatterjee, A. and Lahiri, S.N. (2011) Bootstrapping Lasso Estimators. *Journal of the American Statistical Association*, **106**, 608-625. <https://doi.org/10.1198/jasa.2011.tm10159>
- [15] Camponovo, L. (2015) On the Validity of the Pairs Bootstrap for Lasso Estimators. *Biometrika*, **102**, 981-987. <https://doi.org/10.1093/biomet/asv039>