Journal of Complexity 32 (2016) 20–39

Contents lists available at ScienceDirect

Journal of Complexity

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

Stability of the elastic net estimator[☆]

Yi Shen[a,b,c,*], Bin Han[b], Elena Braverman[c]

[a] Department of Mathematics, Zhejiang Sci–Tech University, Hangzhou, 310028, China

[b] Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

[c] Department of Mathematics and Statistics, University of Calgary, 2500 University Drive N.W., Calgary, Alberta T2N 1N4, Canada

ARTICLE INFO

Article history:
Received 30 August 2014
Accepted 16 July 2015
Available online 30 July 2015

Keywords:
Compressed sensing
Elastic net
Stable recovery
Correlated random matrix
Grouping effect

ABSTRACT

The elastic net is a regularized least squares regression method that has been widely used in learning and variable selection. The elastic net regularization linearly combines an l_1 penalty term (like the lasso) and an l_2 penalty term (like ridge regression). The l_1 penalty term enforces sparsity of the elastic net estimator, whereas the l_2 penalty term ensures democracy among groups of correlated variables. Compressed sensing is currently an extensively studied technique for efficiently reconstructing a sparse vector from much fewer samples/observations. In this paper we study the elastic net in the setting of sparse vector recovery. For recovering sparse vectors from few observations by employing the elastic net regression, we prove in this paper that the elastic net estimator is stable provided that the underlying measurement/design matrix satisfies the commonly required restricted isometry property or the sparse approximation property. It is well known that many independent random measurement matrices satisfy the restricted isometry property while random measurement matrices generated by highly correlated Gaussian random variables satisfy the sparse approximation property. As a byproduct, we establish a uniform bound for the grouping effect of the elastic net. Some numerical

☆ Research was supported in part by Natural Sciences and Engineering Research Council of Canada (NSERC Canada) under grants 05865 and 261351, and Pacific Institute for the Mathematical Sciences (PIMS) CRG grant. Research of Yi Shen was also supported in part by a PIMS postdoctoral fellowship, the Zhejiang Provincial Natural Science Foundation of China under Grant No. LY15A010020 and the Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province under grants No. OBDMA201505.

* Corresponding author at: Department of Mathematics, Zhejiang Sci–Tech University, Hangzhou, 310028, China.
E-mail addresses: yshen@zstu.edu.cn (Y. Shen), bhan@ualberta.ca (B. Han), maelena@math.ucalgary.ca (E. Braverman).
URL: http://www.ualberta.ca/~bhan (B. Han).

http://dx.doi.org/10.1016/j.jco.2015.07.002
0885-064X/

experiments are provided to illustrate our theoretical results on stability and grouping effect of the elastic net estimator.

1. Introduction

The standard model of linear regression can be stated as follows:

$$\boldsymbol{y} = X\boldsymbol{\beta}^* + \boldsymbol{z}, \tag{1.1}$$

where $\boldsymbol{y} \in \mathbb{R}^n$ is an observed signal/vector, $X \in \mathbb{R}^{n\times p}$ is a given measurement/design matrix, $\boldsymbol{\beta}^* \in \mathbb{R}^p$ is an unknown true signal/vector to be recovered, and $\boldsymbol{z}$ is the white Gaussian noise term which is a vector of independent normal/Gaussian random variables.

In the setting of compressed sensing, one is interested in the linear regression model (1.1) for the particular case that n is much smaller than p; that is, the dimension p of the ambient space of the unknown signal $\boldsymbol{\beta}^*$ is large but the number n of the observations $\boldsymbol{y}$ is extremely small. If the unknown signal $\boldsymbol{\beta}^*$ is sparse, the central task of compressed sensing is to recover in a stable way the unknown sparse true signal $\boldsymbol{\beta}^*$ from its few noisy observations/measurements $\boldsymbol{y}$ (see e.g. [8,15]).

To provide some necessary background for compressed sensing and the elastic net, we first recall some definitions. Let $\boldsymbol{\beta} = (\beta_{[1]}, \dots, \beta_{[p]})^T$ denote a vector in $\mathbb{R}^p$. For $0 < q < \infty$, the l_q norm of $\boldsymbol{\beta}$ is $\|\boldsymbol{\beta}\|_q = \left(|\beta_{[1]}|^q + \cdots + |\beta_{[p]}|^q\right)^{1/q}$, and the l_∞ norm of $\boldsymbol{\beta}$ is $\|\boldsymbol{\beta}\|_\infty = \max\{|\beta_{[1]}|, \dots, |\beta_{[p]}|\}$. By $\mathrm{supp}(\boldsymbol{\beta})$ we denote the support of $\boldsymbol{\beta}$, i.e., the set $\{1 \le j \le p : \beta_{[j]} \ne 0\}$. Moreover, $\#\,\mathrm{supp}(\boldsymbol{\beta})$ is defined to be the cardinality of the set $\mathrm{supp}(\boldsymbol{\beta})$, that is, the number of nonzero entries in the vector $\boldsymbol{\beta}$. We further define the l_0 "norm" of $\boldsymbol{\beta}$ to be $\|\boldsymbol{\beta}\|_0 = \#\,\mathrm{supp}(\boldsymbol{\beta})$.

For a nonnegative integer s, we say that a vector $\boldsymbol{\beta} \in \mathbb{R}^p$ is *s-sparse* if the number of all its nonzero entries is no more than s, that is, $\|\boldsymbol{\beta}\|_0 \le s$. For a positive integer s, we say that a measurement matrix $X \in \mathbb{R}^{n\times p}$ has the *restricted isometry property* (RIP) with RIP constant $0 < \delta_s < 1$ if

$$(1 - \delta_s)\|\boldsymbol{\beta}\|_2^2 \le \|X\boldsymbol{\beta}\|_2^2 \le (1 + \delta_s)\|\boldsymbol{\beta}\|_2^2, \quad \text{for all } \boldsymbol{\beta} \in \Sigma_s, \tag{1.2}$$

where the set Σ_s of all s-sparse vectors in $\mathbb{R}^p$ is defined to be

$$\Sigma_s := \{\boldsymbol{\gamma} \in \mathbb{R}^p : \|\boldsymbol{\gamma}\|_0 \le s\}. \tag{1.3}$$

The restricted isometry property in (1.2) with a small RIP constant δ_s implies that every group of arbitrary s column vectors from the measurement matrix X must be nearly orthogonal to each other. It is well known [8] that with overwhelming probability, a measurement matrix X generated by many known independent random variables has the restricted isometry property with a small RIP constant. If the measurement matrix X has the restricted isometry property with RIP constants satisfying $\delta_s + \delta_{2s} + \delta_{3s} < 1$, [8, Theorem 1.4] shows that one can always recover an s-sparse signal $\boldsymbol{\beta}^*$ from few measures in $\boldsymbol{y}$ in (1.1). Even for the case $n = O(s\ln(p/s))$ so that the number n of measurements in $\boldsymbol{y}$ is much less than the number p of all entries in the unknown true signal $\boldsymbol{\beta}^*$, a random measurement matrix can satisfy the restricted isometry property with RIP constants satisfying $\delta_s + \delta_{2s} + \delta_{3s} < 1$ with overwhelming probability. Compressed sensing employs the l_1 regularization technique which has been adopted in the basis pursuit [15] and in the lasso [37]. The lasso is an l_1 penalized least squares regression method that can fit the observation data well while seeking a sparse solution simultaneously. However, it is known that the lasso may not be an ideal method if a group of columns of a measurement matrix is highly correlated, for example, in microarray data analysis [40]. To avoid such a limitation of lasso, the elastic net has been proposed in Zou and Hastie [40] by linearly combining an l_1 penalty term (like lasso) and an l_2 penalty term (like ridge regression). Through experiments, the elastic net shows the "grouping effect" which could include automatically all the highly correlated variables in a group. The theoretical properties of the grouping effect of the elastic net have been studied in [40] and further improved in [39].