Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2020-0250

Vol. **30**, No. 1, pp. 190-209
July 2021

Truncated L_1 Regularized Linear Regression: Theory and Algorithm

Mingwei Dai[1], Shuyang Dai[2,3], Junjun Huang[2], Lican Kang[2] and Xiliang Lu[2,3,*]

[1] *Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu, P.R. China.*
[2] *School of Mathematics and Statistics, Wuhan University, Wuhan, P.R. China.*
[3] *Hubei Key Laboratory of Computational Science, Wuhan University, Wuhan, P.R. China.*

Received 9 December 2020; Accepted (in revised version) 26 February 2021

Abstract. Truncated L_1 regularization proposed by Fan in [5], is an approximation to the L_0 regularization in high-dimensional sparse models. In this work, we prove the non-asymptotic error bound for the global optimal solution to the truncated L_1 regularized linear regression problem and study the support recovery property. Moreover, a primal dual active set algorithm (PDAS) for variable estimation and selection is proposed. Coupled with continuation by a warm-start strategy leads to a primal dual active set with continuation algorithm (PDASC). Data-driven parameter selection rules such as cross validation, BIC or voting method can be applied to select a proper regularization parameter. The application of the proposed method is demonstrated by applying it to simulation data and a breast cancer gene expression data set (bcTCGA).

AMS subject classifications: 62J99

Key words: High-dimensional linear regression, sparsity, truncated L_1 regularization, primal dual active set algorithm.

1 Introduction

In this paper, we consider the high-dimensional sparse linear regression model

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

where $\mathbf{y}\in\mathbb{R}^n$ is the response vector, $\mathbf{X}=[\mathbf{x}_1,\cdots,\mathbf{x}_p]\in\mathbb{R}^{n\times p}$ is the covariance matrix, $\boldsymbol{\beta}^*\in\mathbb{R}^p$ is the underlying regression coefficients vector, $\boldsymbol{\epsilon}=(\epsilon_1,\cdots,\epsilon_n)^T\in\mathbb{R}^n$ is the random noise.

*Corresponding author. *Email addresses:* `daimw@swufe.edu.cn` (M. Dai), `shuyang_dai@whu.edu.cn` (S. Dai), `hjj_wd@whu.edu.cn` (J. Huang), `kanglican@whu.edu.cn` (L. Kang), `xllv.math@whu.edu.cn` (X. Lu)

Without loss generality, we assume that $\mathbf{X}$ is normalized such that each column of $\mathbf{X}$ is $\sqrt{n}$-length. We focus on the case $n \ll p$ and $\|\boldsymbol{\beta}^*\|_0 < n$ for the high dimensional and sparsity assumptions for (1.1), where $\|\boldsymbol{\beta}^*\|_0$ denotes the cardinality of nonzero element of $\boldsymbol{\beta}^*$.

There are various convex and non-convex regularization methods for variable estimation and selection of model (1.1). The popular convex regularization methods include the least absolute shrinkage and selection operator method (Lasso) [19], the adaptive Lasso [28] and Elastic net [29]. Thanks to the convexity of these regularizers, people have designed a lot of efficient numerical algorithms to solve above models, see e.g. [4, 21]. The convex model also has its drawback: it produces biased estimates for large coefficients [14] and lacks oracle property [6]. Some useful nonconvex regularization methods are proposed to circumvent this drawback, such as the bridge penalty method [9, 10], the truncated L_1 regularization [5], the smoothly clipped absolute deviation (SCAD) penalty [7], the Dantzig selector [3], the minimax concave penalty (MCP) [23], the capped-L_1 penalty [26], etc.

The above mentioned nonconvex regularizers can be viewed as an approximation of original L_0 penalty ($\|\cdot\|_0$). Among these regularization methods, the truncated L_1 regularization has an attractive property: its thresholding operator is exactly same as the thresholding for L_0 regularizer. In this work, we will consider the truncated L_1 regularization for variable estimation and selection, i.e., we want to solve

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2n}\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|^2+\sum_{i=1}^{p}\rho_\lambda(\beta_i), \tag{1.2}$$

where $\lambda>0$ is the regularization parameter and $\rho_\lambda(\cdot)$ is defined by

$$\rho_\lambda(t)=\begin{cases} \lambda|t|, & \text{if } |t|<\lambda, \\ \frac{\lambda^2}{2}, & \text{if } |t|\geq\lambda. \end{cases} \tag{1.3}$$

We will prove that if the covariance matrix $\mathbf{X}$ satisfies a certain incoherence condition, then one can obtain the nonasymptotic error bound for the global optimal solution to (1.2). And the support recovery property is also studied. Due to the non-convex and non-smooth structure of the truncated L_1 regularization, (1.2) is a non-convex and non-smooth optimization problem. Then it is very difficult to design a stable and efficient numerical algorithm.

Inspired by [8, 12, 15, 17], we will propose a primal dual active set algorithm (PDAS) to compute the optimal solution to (1.2). PDAS can be viewed as a generalized Newton method, which involves two steps for each iteration. The active set is first determined using the summation of the primal and dual variables. Then the primal variable is updated by solving an optimization problem on the active set with small size, and the dual variable is updated based on a closed-form expression. Combining PDAS with a continuation strategy on the regularization parameter λ can make the whole algorithm more