East Asian Journal on Applied Mathematics
doi: 10.4208/eajam.110121.210621

Vol. **12**, No. 1, pp. 35-52
February 2022

Hard Thresholding Regularised Logistic Regression: Theory and Algorithms

Lican Kang[1], Yanyan Liu[1], Yuan Luo[1] and Chang Zhu[2,*]

[1]*School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China.*
[2]*Department of Anesthesiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China.*

Received 11 January 2021; Accepted (in revised version) 21 June 2021.

Abstract. The hard thresholding regularised logistic regression in high dimensions with larger number of features than samples is considered. The sharp oracle inequality for the global solution is established. If the target signal is detectable, it is proven that with a high probability the estimated and true supports coincide. Starting with the KKT condition, we introduce the primal and dual active sets algorithm for fitting and also consider a sequential version of this algorithm with a warm-start strategy. Simulations and a real data analysis show that SPDAS outperforms LASSO, MCP and SCAD methods in terms of computational efficiency, estimation accuracy, support recovery and classification.

AMS subject classifications: 62J12, 62J02, 62J07

Key words: Sparse logistic regression, hard thresholding regularisation, PDAS, SPDAS.

1. Introduction

Let $y \in \{0,1\}$ be the binary response variable, $\mathbf{x} \in \mathbb{R}^p$ the covariate vector and $\boldsymbol{\beta}^* \in \mathbb{R}^p$ the underlying regression coefficients vector in the logistic regression model

$$P\big(y = 1|\mathbf{x}\big) = \frac{\exp\big(\mathbf{x}^T\boldsymbol{\beta}^*\big)}{1+\exp\big(\mathbf{x}^T\boldsymbol{\beta}^*\big)},$$

cf. [10, 11]. Logistic regression is an important generalised linear model (GLM) widely used in statistics, machine learning, social and medical sciences, finance industry and so on. In this work, we focus on the variable estimation and selection in high-dimensional and sparse settings — i.e. if $n \ll p$ and $\|\boldsymbol{\beta}^*\|_0 < n$, where n is the sample size and $\|\boldsymbol{\beta}^*\|_0$ the cardinality of the set of nonzero elements in $\boldsymbol{\beta}^*$.

*Corresponding author. *Email addresses:* `kanglican@whu.edu.cn` (L. Kang), `liuyy@whu.edu.cn` (Y. Liu), `yuanluo@whu.edu.cn` (Y. Luo), `changzhu@hust.edu.cn` (C. Zhu)

To obtain the estimator of $\boldsymbol{\beta}^*$ in high-dimensional and sparse cases, many regularised methods have been proposed. In particular, works [13,17] extend the least absolute shrinkage and selection operator method (LASSO) [16] from the linear regression to GLMs. Friedman *et al.* [4] used the coordinate descent to solve the elastic net penalised GLMs [21]. In Refs. [8,19], the path following proximal gradient descent method [12] has been applied to variable estimation in GLMs with smoothly clipped absolute deviation (SCAD) and minimax concave (MC) penalties [3]. Besides, Li *et al.* [7] introduced a DC proximal Newton (DCPN) method for GLMs with sparsity promoting non-convex penalties such as MC and SCAD ones.

In this paper, we consider the hard thresholding regulariser

$$\rho_\lambda(t) = \begin{cases} \dfrac{-t^2}{2} + \lambda|t|, & \text{if} \quad |t| < \lambda, \\ \dfrac{\lambda^2}{2}, & \text{if} \quad |t| \geq \lambda, \end{cases} \tag{1.1}$$

where a non-convex and non-smooth function ρ_λ admits the hard thresholding operator (3.2), cf. [1]. The hard thresholding regularised estimator leads to the problem

$$\min_{\boldsymbol{\beta}} \mathscr{L}_n(\boldsymbol{\beta}) + \sum_{i=1}^{p} \rho_\lambda(\beta_i), \tag{1.2}$$

where

$$\mathscr{L}_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \log\Big(1 + \exp\big(\mathbf{x}_i^T \boldsymbol{\beta}\big)\Big) - \frac{\mathbf{Y}^T \mathbf{X}\boldsymbol{\beta}}{n}$$

is the negative logarithmic likelihood function and $\lambda > 0$ the tuning parameter.

The ideas of [8, 20] can be used to determine the sharp upper bound for the errors of the estimator (1.2). Nevertheless, since (1.2) is a non-convex non-smooth optimisation problem, it is difficult to develop a stable efficient computational algorithm for its solution, especially in high-dimensional and sparse settings. Inspired by [5, 15], we construct a primal and dual active set (PDAS) algorithm for solving the minimisation problem (1.2). Our approach is motivated by the KKT conditions of the hard thresholding regularised problem.

In PDAS, the active set of a relatively small size is first determined via summation of primal and dual variables generated by the previous iteration. The primal variable is then updated by solving a minimisation problem on the active set, whereas the dual variable is updated by using the gradient information. Further, in order to make PDAS more applicable, we consider a sequential version of PDAS (SPDAS), which combines PDAS with a continuation strategy on the regularisation parameter λ. Thus, by SPDAS algorithm we generate a solution path with a different regularisation parameter λ. Then we can chose one data-driven method such as the modified Bayesian information criteria [6, 18] or the voting method [5] to chose the optimal solution.

The main results of this work are as follows. Using regularity assumptions on the loss function and the covariance matrix, we establish a sharp oracle inequality of the global