



Robust Alternatives to Least Squares for Sparse Estimation with Applications

Vahid Nassiri

Supervisors: Philippe Cara and Ignace Loris

Abstract

In this work, sparse robust estimation of the parameters of a linear regression model is considered. Consider y , the response variable that we study using a set of k so-called explanatory variables $X = (x_1, x_2, \dots, x_k) \in \mathbb{R}^{n \times k}$ which are linearly related to the variable y :

$$y = X\beta + \epsilon,$$

where β is a vector containing k parameters, and ϵ is the error of the model. A sparse estimation of β can be obtained using a penalized model. Such a penalized model consists of three main aspects: the loss function T , the penalty function g and the solution method to obtain the estimated parameters $\hat{\beta}$ by solving the optimization problem:

$$\hat{\beta} = \operatorname{argmin}_{\beta} T(y - X\beta) + \lambda g(\beta),$$

($\lambda \geq 0$ is the penalty parameter). We are going to study each of these three aspects throughout this thesis.

For the loss function part, we will introduce some robust alternatives to the least squares loss function. These (possibly asymmetric) loss functions are convex and differentiable. The first loss function is based on a novel class of probability distributions: the so-called connected double truncated gamma distribution. We show that using this class as the error distribution of a linear model leads to a generalized quantile regression model that combines desirable properties of both least-squares and quantile regression methods: robustness to outliers and differentiable loss function. Furthermore, the possibility of using a skew-symmetric class of distributions as error distribution in a linear model is investigated. Selecting appropriate kernels in this class, would provide us with a robust alternative to least-squares which is convex and differentiable.

For the penalty part, we consider mixed norm penalties of types $g(\beta) = \|\beta\|_{1,\infty}$ and $\|\beta\|_{1,2}$, which allow the researcher to impose structured sparsity on groups of variables instead of the individual ones. Three main applications of such penalties will also be considered. We will use it to include a nominal variable (with more than 2 levels) in a penalized quantile regression model. It will also be used for simultaneous variable selection in a multivariate quantile regression model. In addition, a novel application of such structured sparsity will be introduced in the field of impact force identification.

For the solution method, we will consider two classes of efficient algorithms: iterative algorithms and non-iterative algorithms. By efficiency we mean speed of convergence (in the iterative case). But for non-iterative algorithms, efficiency means the ability to give the solution $\hat{\beta}(\lambda)$ for a range of possible values of λ at once. A novel algorithm will be derived to solve a structured sparse quantile regression with a $\ell_{1,\infty}$ -norm penalty. The algorithm gives the piece-wise linear solution path for $\hat{\beta}(\lambda)$ in a finite number of steps.