Open Problem: Restricted Eigenvalue Condition for Heavy Tailed Designs

Arindam Banerjee  
University of Minnesota Twin Cities, USA  
BANERJEE@CS.UMN.EDU

Sheng Chen  
University of Minnesota Twin Cities, USA  
SHENGC@CS.UMN.EDU

Vidyashankar Sivakumar  
University of Minnesota Twin Cities, USA  
SIVAKUMA@CS.UMN.EDU

Abstract

The restricted eigenvalue (RE) condition characterizes the sample complexity of accurate recovery in the context of high-dimensional estimators such as Lasso and Dantzig selector (Bickel et al., 2009). Recent work has shown that random design matrices drawn from any thin-tailed (sub-Gaussian) distributions satisfy the RE condition with high probability, when the number of samples scale as the square of the Gaussian width of the restricted set (Banerjee et al., 2014; Tropp, 2015). We pose the equivalent question for heavy-tailed distributions: Given a random design matrix drawn from a heavy-tailed distribution satisfying the small-ball property (Mendelson, 2015), does the design matrix satisfy the RE condition with the same order of sample complexity as sub-Gaussian distributions? An answer to the question will guide the design of high-dimensional estimators for heavy tailed problems.

1. Introduction

Progress has been made over the last decade in characterizing conditions under which high-dimensional structured estimation, i.e., methods such as Lasso (Tibshirani, 1996), Dantzig Selector (Candes and Tao, 2007), etc., yields accurate results. Typically, one assumes a linear random measurement model: \( y_i = \langle x_i, \theta^* \rangle + \epsilon_i, i = 1, \ldots, n \), where \( \theta^* \) is assumed to be structured (Negahban et al., 2012; Chandrasekaran et al., 2012). While the true structure in \( \theta^* \) may be complex, e.g., needing a combinatorial or non-convex specification, one considers the tightest convex relaxation of the structure, in terms of the gauge function of the convex hull, yielding a norm \( \mathcal{R}(\cdot) \) (Chandrasekaran et al., 2012; Banerjee et al., 2014).

The literature has considered various forms of estimators for such problems. Two common forms are (i) Lasso-type estimators, which consider norm regularized regression (Tibshirani, 1996; Negahban et al., 2012; Banerjee et al., 2014), and (ii) Dantzig-type estimators, which consider a constrained form of the problem (Candes and Tao, 2007; Chatterjee et al., 2014). Let \( y \in \mathbb{R}^n \) denote the vector of observations and \( X = [x_1 \cdots x_n]^T \in \mathbb{R}^{n \times p} \) denote the random design matrix. Then, lasso-type estimators take the form \( \hat{\theta}^{\text{lasso}}_n = \arg\min_{\theta \in \mathbb{R}^p} \frac{1}{2n} \| y - X\theta \|^2 + \lambda_n \mathcal{R}(\theta) \), where \( \mathcal{R}(\cdot) \) is the norm encoding the structure, and \( \lambda_n > 0 \) is a regularization constant. Instead of squared loss, one can use general convex losses corresponding to generalized linear models (Negahban et al., 2012; Banerjee et al., 2014). The Dantzig-type estimators take the form \( \hat{\theta}^{\text{dantzig}}_n = \arg\min_{\theta \in \mathbb{R}^p} \mathcal{R}(\theta) \) s.t. \( \mathcal{R}^*(X^T(y - X\theta)) \leq \varphi_n \), where \( \mathcal{R}^*(\cdot) \) is the dual norm of \( \mathcal{R}(\cdot) \), and \( \varphi_n > 0 \) is a suitable constant for the constraint (Candes and Tao, 2007; Chatterjee et al., 2014).
There are other related estimators which have been considered in the literature, which have similar requirements for recovery (Chandrasekaran et al., 2012; Tropp, 2015).

A key requirement for accurate recovery relies on a certain restricted eigenvalue (RE) condition associated with the design matrix $X$ (Bickel et al., 2009). Let $A \subseteq S^{p-1}$ be a spherical cap, i.e., a subset of the unit sphere, which characterizes a set of directions. For the estimators discussed above, $A$ gets determined by the norm $R(\cdot)$, e.g., for Dantzig-type estimators, we have $A = \text{cone}\{\Delta \in \mathbb{R}^p : R(\theta^* + \Delta) \leq R(\theta^*)\} \cap S^{p-1}$. Given a random design matrix $X$, the RE condition boils down to showing that the following uniform lower bound holds with high probability:

$$\inf_{u \in A} \|Xu\|_2 \geq c_1 n - c_2 w^2(A)$$  \hspace{1cm} (2)

holds with high probability? An affirmative answer will imply that the RE condition holds for heavy tailed distributions for arbitrary spherical caps $A$.

2. Open Problem: Restricted Eigenvalue Condition for Heavy Tails

Let $X \in \mathbb{R}^{n \times p}$ be a random design matrix with heavy tailed entries. For simplicity, we assume that each row $X_i$ is independent. The heavy tail of each row $Z \equiv X_i$ is characterized by the so-called ‘small-ball’ property (Mendelson, 2015; Tropp, 2015): for some set $E \subseteq \mathbb{R}^p$, there exists some $\alpha, \beta > 0$ such that

$$\inf_{v \in E} P(|\langle Z, v \rangle| \geq \alpha \|v\|_2) \geq \beta. \hspace{1cm} (1)$$

The small-ball property characterizes the tail probability along directions $\frac{v}{\|v\|_2}$ without making any restrictive assumptions regarding the existence of moments of a certain order.

**Open Problem:** Let $X \in \mathbb{R}^{n \times p}$ be a random matrix with independent rows, where each row sampled identically from a heavy-tailed distribution satisfying the small-ball property in (1). Given any spherical cap $A \subseteq S^{p-1}$, can we show that the following uniform lower bound

$$\inf_{u \in A} \|Xu\|_2^2 = \inf_{u \in A} \sum_{i=1}^n \langle x_i, u \rangle^2 \geq c_1 n - c_2 w^2(A)$$  \hspace{1cm} (2)

holds with high probability? An affirmative answer will imply that the RE condition holds for heavy tailed distributions for arbitrary spherical caps $A$.

3. Related Work: Existing Approaches and Results

In recent work, Banerjee et al. (2014), Tropp (2015) proved the RE condition for sub-Gaussian $X$. Banerjee et al. (2014) illustrate exponential concentration for a single $u \in A$, and obtains the uniform bound based on generic chaining (Talagrand, 2005). Tropp (2015) builds on the arguments in Mendelson (2015), considers a lower bound $\gamma$ for the sum of indicator functions $\inf_{u \in A} \sum_{i=1}^n I_{\{\langle x_i, u \rangle > \gamma/2\}}$, implying $\inf_{u \in A} \|Xu\|_2^2 \geq \frac{\gamma^2}{2}$, and bounds $\gamma$ using standard tools from empirical processes and generic chaining. However, both arguments yield a dependency on the Gaussian width only for sub-Gaussian designs.
In addition to the results for sub-Gaussian $X$ and general $A$, the RE condition for some special $A$ has been established when $X$ is heavy tailed. Oliveira (2013) makes the observation that for a given $u \in A$, $\|Xu\|_2^2 = \sum_{i=1}^n (x_i, u)^2$ has sub-Gaussian lower tails under weak moment assumptions on $X$. In order to turn this into a uniform bound for all $u \in A$, tools like generic chaining or covering arguments require upper bounds on $\|Xu\|_2^2$, thus fail for heavy tailed $X$. But in the special case when $A$ is the unit sphere $S^{p-1}$, PAC Bayesian based arguments show a result similar to the following with high probability:

$$\inf_{u \in A} \|Xu\|_2^2 \geq c_1 n - c_3 p. \quad (3)$$

Since $w^2(S^{p-1}) = O(p)$, the RE condition holds for the special case of $A = S^{p-1}$. The result in Koltchinskii and Mendelson (2013) relies on the small-ball property of $A$ and the VC dimension of the class of functions, $T_\xi = \{\mathbb{1}_{|\langle x, u \rangle| > \xi} : u \in A\}$. Using empirical process theory, they prove that with high probability,

$$\inf_{u \in A} \|Xu\|_2^2 \geq c_1 n - c_3 \text{VC}(T_\xi), \quad (4)$$

where $\text{VC}(\cdot)$ denote VC dimension. When $A$ is the unit sphere $S^{p-1}$ it can be shown that $\text{VC}(T_\xi) = O(p) = O(w^2(S^{p-1}))$, and the result coincides with Oliveira (2013). Based on this VC dimension argument, Lecué and Mendelson (2014) show that the RE condition is also true when $A$ is the set of all unit $s$-sparse vectors. While these results illustrate that RE condition for heavy tails is true for certain special cases of $A$, the result for general $A \subseteq S^{p-1}$ under the small-ball property remains open.

Acknowledgements: We thank Joel Tropp for helpful discussions. The research was supported by NSF grants IIS-1447566, IIS-1422557, CCF-1451986, CNS-1314560, IIS-0953274, IIS-1029711, and by NASA grant NNX12AQ39A.

References


