JMLR: Workshop and Conference Proceedings vol 40:1–4, 2015

Open Problem: Restricted Eigenvalue Condition for Heavy Tailed Designs

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

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

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

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 θ^* is assumed to be structured (Negahban et al., 2012; Chandrasekaran et al., 2012). While the true structure in θ^* 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 $R(\cdot)$ (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 $\hat{\theta}_n^{\text{lasso}} = \text{argmin}_{\theta \in \mathbb{R}^p} \frac{1}{2n} \|y - X\theta^*\|^2 + \lambda_n R(\theta)$, where $R(\cdot)$ 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 $\hat{\theta}_n^{\text{dantzig}} = \text{argmin}_{\theta \in \mathbb{R}^p} R(\theta)$ s.t. $R^*(X^T(y - X\theta)) \leq \varphi_n$, where $R^*(\cdot)$ is the dual norm of $R(\cdot)$, 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: $\inf_{u \in A} \|Xu\|^2 \geq c_1 n - c_2 w^2(A)$ for suitable constants $c_1, c_2 > 0$, where $w(A) = E_g[\sup_{u \in A}\langle g, u\rangle]$, where $g \sim N(0, \mathbb{I}_{p \times p})$ is a p-dimensional isotropic normal vector. The quantity $w(A)$ is referred to as the *Gaussian width* of the spherical cap A (Chandrasekaran et al., 2012; Banerjee et al., 2014). While earlier analysis had illustrated the RE condition to hold only for special design matrices and special sets A, recent results have shown the RE condition is satisfied by all sub-Gaussian design matrices and all spherical caps A (Banerjee et al., 2014; Tropp, 2015). In the sequel, constants are denoted by c, c_1, c_2, etc, whose value may change from line to line.

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

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

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

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

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 γ for the sum of indicator functions $\inf_{u \in A} \sum_{i=1}^{n} \mathbb{I}_{|\langle x_i, u\rangle| > \frac{\alpha}{2}}$ implying $\inf_{u \in A} \|Xu\|_2^2 \geq \frac{\alpha^2}{4}\gamma$, and bounds γ 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 $u \in A$, $\|Xu\|_2^2 = \sum_{i=1}^n \langle x_i, u\rangle^2$ 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:

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

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, $T_\xi = \{\mathbb{I}_{|\langle x,u\rangle|>\xi} : u \in A\}$. Using empirical process theory, they prove that with high probability,

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

where $\mathbf{VC}(\cdot)$ denote VC dimension. When A is the unit sphere S^{p-1} it can be shown that $\mathbf{VC}(T_\xi) = O(p) = O(w^2(S^{p-1}))$, 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

A. Banerjee, S. Chen, F. Fazayeli, and V. Sivakumar. Estimation with norm regularization. In *Advances in Neural Information Processing Systems (NIPS)*, 2014.

P. J. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009.

E. Candes and T Tao. The Dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6):2313–2351, 2007.

V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6):805–849, 2012.

S. Chatterjee, S. Chen, and A. Banerjee. Generalized dantzig selector: Application to the k-support norm. In *Advances in Neural Information Processing Systems (NIPS)*, 2014.

V. Koltchinskii and S. Mendelson. Bounding the smallest singular value of a random matrix without concentration. *arXiv:1312.3580*, 2013.

G. Lecué and S. Mendelson. Sparse recovery under weak moment assumptions. *arXiv:1401.2188*, 2014.

S. Mendelson. Learning without concentration. *Journal of the ACM*, 2015.

S. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu. A unified framework for the analysis of regularized M-estimators. *Statistical Science*, 27(4):538–557, 2012.

R. I. Oliveira. The lower tail of random quadratic forms, with applications to ordinary least squares and restricted eigenvalue properties. *arXiv:1312.2903*, 2013.

M. Talagrand. *The Generic Chaining*. Springer, 2005.

R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58(1):267–288, 1996.

J. A. Tropp. Convex recovery of a structured signal from independent random linear measurements. In *Sampling Theory, a Renaissance*. 2015.