Math 547: Mathematical Foundations of Statistical Learning Theory Fall 2017

Lecture 34 — November 8

Instructor: Stas Minsker *Scribe: Mose Wintner*

34.1 Exact Recovery of Sparse Signals

$Y = \mathbb{X}\lambda + v$, $\frac{1}{n}\|v\|_1 \le \varepsilon$, $\lambda \in K$.

$\|\lambda'\|_K := \inf\{t > 0 : \lambda'/t \in K\}$.

Seek to minimize $\|\lambda'\|_K$ subject to $\|\mathbb{X}\lambda' - Y\|_1 / n \le \varepsilon$.

If K is not convex, replace it by its convex hull so we get a convex optimization problem.

Example 1. Assume λ is sparse, so that the support $J(\lambda) := \{j : \lambda_j \neq 0\}$ satisfies $|J(\lambda)| = s << p$.

For instance, e_j are 1-sparse. However, its convex hull is not. Take $K = \|\lambda\|_1 B_{L_1}(0, 1)$ and consider the minimization problem. It is equivalent to minimizing $\|\lambda'\|_1$ such that $\|\mathbb{X}\lambda - Y\|_1 / n \le \varepsilon$. Let $\hat{\lambda}$ be a solution. Then we have that

$$\mathbb{E}\left\|\hat{\lambda} - \lambda\right\|_2 \le \sqrt{8\pi}\left(\frac{w(K)}{\sqrt{n}} + \varepsilon\right).$$

Question: can the error be =0? Yes, but how many measurements does this require? To figure this out we'll assume $\varepsilon = 0$ so that $Y = \mathbb{X}\lambda$.

Let E be the nullspace of $\mathbb{X}$. Then $\lambda \in K \cap \{\lambda + E\}$, i.e. the solution is unique. Consider $T_K(\lambda) = \{t(z - \lambda) : z \in K, t \ge 0\}$. $\lambda = K \cap \{\lambda + E\}$ iff $\lambda = K \cap T_K(\lambda)$. Then $0 = \{K - \lambda\} \cap T_{K-\lambda}(0)$. Letting $S(K, \lambda) = T_K(\lambda) \cap S^{p-1}$, λ is the unique solution if and only if $T_K(\lambda) \cap S^{p-1} = \{0\}$.

Theorem 1.

$$P(T_K(\lambda) \cap S^{p-1} \neq \{0\}) \le \sqrt{\frac{8\pi}{n}}(\overline{w}(S(K, \lambda) + \sqrt{\frac{2}{\pi}}),$$

where

$$\overline{w}(S(K, \lambda)) = \mathbb{E}\sup_{u \in S(K,\lambda)} \langle g, u\rangle.$$

If $n >> \overline{w}^2(S(K, \lambda))$, then the probability of exact recovery is large. In our example, $\overline{w}^2$ is about $s \log p/s$.

Proof of theorem. Apply general theorem to $T = S(K, \lambda)$, $E = \ker(X)$, $\varepsilon = 0$.