Math 547: Mathematical Foundations of Statistical Learning Theory

Lectures 1 — Week 13

Instructor: S. Minsker *Scribe: S. Minsker*

1.1 Exact recovery of sparse signals

Once again, assume that there is no noise ($\nu = 0$). Can we recover λ exactly (with high probability)? We will positively answer this question below.

In what follows, assume that $\mathcal{K}$ is a convex set. Let E be the null space of $\mathbf{X}$, and let $\widehat{\lambda}$ be a solution to

$$\|\lambda'\|_{\mathcal{K}} \to \min \quad \text{subject to } \mathbf{X}\lambda' = Y$$

Then $\widehat{\lambda} = \lambda \iff \mathcal{K} \cap \{\lambda + E\} = \{\lambda\}$ (in other words, the intersection of $\mathcal{K}$ and the affine subspace $\lambda + E$ contains 1 point). Consider the *descent cone* of $\mathcal{K}$ at the point λ,

$$T_{\mathcal{K}}(\lambda) = \{t(z - \lambda) : z \in K, t \geq 0\}.$$

We have the following chain of implications (see figure 1.1 for the geometric depiction):

$$\begin{aligned}
&\{\lambda\} = \mathcal{K} \cap \{\lambda + E\} \iff \\
&\{0\} = \{\mathcal{K} - \lambda\} \cap E \quad \text{if} \\
&\{0\} = T_{\mathcal{K}}(\lambda) \cap E.
\end{aligned}$$

Next, define $S(\mathcal{K}, \lambda) = T_{\mathcal{K}}(\lambda) \cap S^{p-1}$ where $S^{p-1} = \{u \in \mathbb{R}^p : \|u\|_2 = 1\}$. Then

$$T_{\mathcal{K}}(\lambda) \cap E = \{0\} \iff S(\mathcal{K}, \lambda) \cap E = \emptyset.$$

A estimate for the probability of the latter event is provided by the theorem below.

Theorem 1. The following inequality holds:

$$\Pr\left(\widehat{\lambda} \neq \lambda\right) = \Pr\left(S(\mathcal{K}, \lambda) \cap E \neq \emptyset\right) \leq 2\sqrt{\frac{8\pi}{n}}\, w(S(\mathcal{K}, \lambda)).$$

Remark 1. It follows that if $n \gg w^2(S(K, \lambda))$, then the probability of exact recovery is large. For instance, for sparse vectors, $w^2(S(K, \lambda)) \lesssim s \log p$.

Proof. Apply Theorem 1 from last week's notes, together with the fact that for any set T,

$$\mathbb{E} \sup_{z \in T} |\langle z, g\rangle| \leq 2\mathbb{E} \sup_{z \in T} \langle z, g\rangle = w(T).$$

Choosing $T = S(\mathcal{K}, \lambda),\ E = \ker(\mathbf{X}),\ \varepsilon = 0$, we get that

$$\mathbb{E} \sup_{u \in S(\mathcal{K}, \lambda) \cap E} \|u\|_2 \leq 2\sqrt{\frac{8\pi}{n}}\, w(S(\mathcal{K}, \lambda)).$$

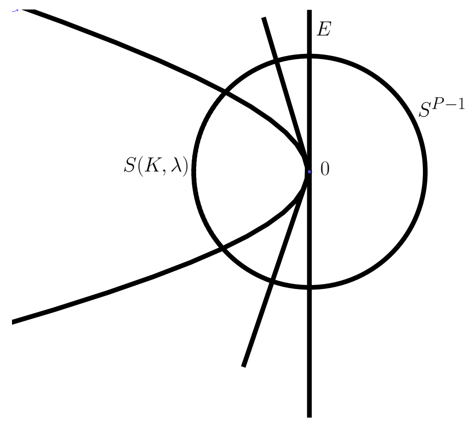

Figure 1.1. Geometry of the exact recovery condition

Moreover, it is easy to see that

$$\sup_{u\in S(\mathcal{K},\lambda)\cap E}\|u\|_2 = \begin{cases} 1, & S(\mathcal{K},\lambda)\cap E\neq\emptyset \\ 0, & S(\mathcal{K},\lambda)\cap E=\emptyset, \end{cases}$$

is the indicator function of the event $S(\mathcal{K},\lambda)\cap E\neq\emptyset$, hence

$$\mathbb{E}\sup_{u\in S(\mathcal{K},\lambda)\cap E}\|u\|_2 = \Pr\left(T_{\mathcal{K}}(\lambda)\cap S^{p-1}\neq\emptyset\right).$$

□

Note that this is not the most precise form of bound possible. Gordon's escape through a mesh theorem (that we are not going to prove) provides a sharper bound.

Theorem 2 (Gordon's "Escape Through a Mesh" theorem)**.** Let S be a fixed subset of S^{p-1}, and let E be a subspace of $\mathbb{R}^p$ of dimension $p-n$ chosen uniformly at random. Assume that

$$\bar{w}(S) := \mathbb{E}\sup_{u\in S}\langle g,u\rangle < \sqrt{n}.$$

Then

$$\Pr\left(S\cap E\neq\emptyset\right)\leq 2.5\exp\left(-\left(n/\sqrt{n+1}-\bar{w}(S)\right)^2/18\right).$$

Finally, we will bbtain a bound on $w\left(S_{\mathcal{K}}(\lambda)\right)$ where λ is s-sparse and $\mathcal{K}=\|\lambda\|_1 B_{\|\cdot\|_1}$. Let $J=J(\lambda)$ be the support of λ defined as

$$J=\{j\in\{1,\ldots,p\}:\ \lambda_j\neq 0\},$$

and let u_J be the vector u restricted to the set of coordinates with indices in the set J. Note that the cardinality of J is equal to s.

Lemma 1. The following statements hold: for any $\lambda' \in \mathcal{K}$,

(a) $\|(\lambda' - \lambda)_{J^c}\|_1 \le \|(\lambda' - \lambda)_J\|_1$;

(b) $\|\lambda' - \lambda\|_1 \le 2\sqrt{s}\|\lambda' - \lambda\|_2$.

Proof.
(a) Note that

(1) $\|\lambda'\|_1 \le \|\lambda\|_1$ by definition of $\mathcal{K}$;

(2) $\|\lambda'\|_1 = \|\lambda + (\lambda' - \lambda)\|_1 = \|\lambda_J + (\lambda' - \lambda)_J\|_1 + \|\underbrace{\lambda_{J^c}}_{=0} + (\lambda' - \lambda)_{J^c}\|_1$.

Applying the triangle inequality, we get $\|\lambda'\|_1 \ge \|\lambda_J\|_1 - \|(\lambda - \lambda)_J\|_1 + \|(\lambda' - \lambda)_{J^c}\|_1$. Combined with (1), it implies part (a).
(b) To prove the second inequality, note that

$$\begin{aligned}\|\lambda' - \lambda\|_1 &= \|(\lambda' - \lambda)_J\|_1 + \|(\lambda' - \lambda)_{J^c}\|_1 \\ &\le 2\|(\lambda' - \lambda)_J\|_1 \ \le 2\sqrt{s}\|(\lambda' - \lambda)_J\|_2 \le 2\sqrt{s}\|\lambda' - \lambda\|_2\end{aligned}$$

□

Let $u := u(\lambda') = \frac{\lambda' - \lambda}{\|\lambda' - \lambda\|_2}$, and note that

1. $u \in S^{p-1}$ and
2. $u \in T_{\mathcal{K}}(\lambda)$.

Hence, $u \in S_{\mathcal{K}}(\lambda)$. It follows from part (b) of Lemma 1 that

$$u \in T_s = \{v \in S^{p-1} : \|u\|_1 \le 2\sqrt{s}\},$$

meaning that $S_{\mathcal{K}}(\lambda) \subseteq T_s$. Hence, it is enough to estimate $w(T_s)$. Note that

$$\begin{aligned}w(T_s) &= 2\mathbb{E}\sup_{u \in T_s}\langle g, u\rangle = 4\sqrt{s}\,\mathbb{E}\sup_{u \in B_{\|\cdot\|}(1)}\langle g, u\rangle \\ &= 4\sqrt{s}\,\mathbb{E}\max_{j=1,\dots,p}|\langle g, e_j\rangle|,\end{aligned}$$

where we used the fact that $B_{\|\cdot\|}(1) = \mathrm{co}\,\{\pm e_j,\ j = 1, \dots, p\}$ where $e_1, \dots, e_p$ is the canonical basis, and the invariance of the Gaussian mean width with respect to taking the convex hull. Finally, $4\sqrt{s}\,\mathbb{E}\max_{j=1,\dots,p}|\langle g, e_j\rangle| \le 4\sqrt{2}\sqrt{s\log(2p)}$, hence we deduce that

$$w(S_{\mathcal{K}}(\lambda)) \le 4\sqrt{2}\sqrt{s\log(2p)},$$

which implies that recovery is exact with high probability whenever $n \gg s\log(2p)$.

Remark 2. The latter condition can be improved to $n \gg s\log(p/s)$ via a more delicate argument. This argument will be outlined in your last (optional) homework assignment.