Generalized theory of Compressed Sensing (CS)

Vanilla CS: $y = A x + z$ ← noise

known (y, A)

unknown, sparse signal (x)

General CS: $y = Ax + z$

x has "some structure" i.e. $x \in K \subseteq \mathbb{R}^n$, where K encodes structure.

Ex)
- x has small total variation
- x is sparse in a basis
- x is a low-rank matrix/tensor

Recovery of x depends on conditioning of A restricted to K, or $K \cap S^{n-1}$.

For random A, we can use the tools from class to determine this restricted conditioning.

Thm (Master theorem) Let $A \in \mathbb{R}^{m \times n}$ have independent entries satisfying:
$\mathbb{E} A_{ij} = 0$, $\mathbb{E} A_{ij}^2 = 1$, $\|A_{ij}\|_{\psi_2} \leq 10$. ← Can be any const.
Let $0 \in K \subseteq \mathbb{R}^n$. Then

$$\mathbb{E} \sup_{x \in K} \left| \|Ax\| - \sqrt{m}\|x\|_2 \right| \leq C\, w(K)$$

← Gaussian mean width

Def (Gaussian mean width) Let $K \subseteq \mathbb{R}^n$, $g \sim N(0, I_n)$ then we call

$$w(K) := \mathbb{E} \sup_{x \in K} \langle x, g \rangle$$

the Gaussian mean width.

Remarks:

(1) $w(K)$ is an established measure of size/complexity of a set K. Closely related to Gaussian complexity.

(2) When K is a cone, i.e. $K = \lambda K$ for $\lambda > 0$,

$w^2(K \cap B_2^n)$ gives a robust measure of $\dim(K)$.

Ex) $K = \mathbb{R}^n$. $w(K \cap B_2^n) = \mathbb{E} \sup_{x \in S^{n-1}} \langle x, g \rangle = \mathbb{E}\|g\|_2$

$$\leq \sqrt{\mathbb{E}\|g\|_2^2} = \sqrt{n}$$

And, using Gaussian concentration, $|\mathbb{E}\|g\| - \sqrt{n}| \leq C$.

$\Rightarrow$ $w^2(K \cap B_2^n) \approx n$ (as expected).

- Combined w/ theorem, this instantly recovers: $\mathbb{E}\sigma_1(A) \leq \sqrt{m} + C\sqrt{n}$, $\mathbb{E}\sigma_{min}(A) \geq \sqrt{m} - C\sqrt{n}$.

Ex) Similarly, if K is a d-dimensional subspace, then

$$w^2(K \cap B_2^n) = d + O(1)$$

Ex). $K \subseteq S^{n-1}$ & $|K| < \infty$

$w(K) =$ Expected max of $|K|$ $N(0,1)$ r.v.s

$\leq \sqrt{2 \log |K|}$

- Immediately implies JL lemma.

Ex) $K = \{x \in \mathbb{R}^n : \|x\|_0 \leq s\} = \{$"s-sparse vectors"$\}$

$\boxed{\text{HW}}$ Show that: $cs\log\left(\frac{cn}{s}\right) \leq w^2(K \cap B_2^n) \leq C\, s \log\left(\frac{Cn}{s}\right)$
For $n \geq Cs$

- Implies restricted isometry property.

Ex) $K = \{ X \in \mathbb{R}^{n\times n} : \text{rank}(X) \le r \}$

$$w^2(K \cap B_F^{n\times n}) \le C\, n \cdot r \qquad \text{(matches manifold dimension)}$$

Proof

$$w(K) = \mathbb{E} \sup_{\substack{X \in K \\ \|X\|_F \le 1}} \langle X, G \rangle \qquad G \sim N(0, I_{n\times n})$$

Note, for any $X \in K \cap S^{n\times n-1}$, w/ svd $X = \sum_{i=1}^{r} \sigma_i u_i v_i^T$ we have

$$\begin{aligned}
\langle X, G \rangle &= \sum_{i=1}^{r} \sigma_i \, u_i^T G v_i \\
&\le \sum_{i=1}^{r} \sigma_i \|G\| \\
&= \|G\| \, \|\vec{\sigma}\|_1 , \qquad \vec{\sigma} := (\sigma_1, \dots, \sigma_r) \\
&\le \|G\| \cdot \sqrt{r} \|\vec{\sigma}\|_2 \qquad \text{(by Cauchy-schwartz)} \\
&= \|G\| \sqrt{r} \qquad (\text{since } \|\vec{\sigma}\|_2 = \|X\|_F \le 1)
\end{aligned}$$

$$\Rightarrow w(K \cap B_F^{n\times n}) \le \sqrt{r}\, \mathbb{E}\|G\|$$
$$\le C \sqrt{nr}. \qquad \text{(By many theorems we have proven)}$$

QED

We will now prove the thm above using:

Lemma (sub-Gaussian increments)

> Let A be as in the theorem. Then the random process $X_s := \big| \|As\|_2 - \sqrt{m}\|s\|_2 \big|$, $s \in \mathbb{R}^n$, has sub-Gaussian increments:
>
> $$\|X_s - X_t\|_{\psi_2} \le C \|s - t\|_2 .$$

$\boxed{\boxed{\text{HW}}\ \text{Prove lemma.}}$

Proof (of thm)

By lemma, $\underbrace{\|X_s - X_t\|_{\psi_2}}_{C} \le \|Y_s - Y_t\|_{\psi_2}$

where $Y_s := \langle g, s \rangle$, $g \sim N(0, I_n)$.

Thus, by the Gaussian-sub-Gaussian comparison* (a.k.a. Majorizing measures thm) from Lec. 12,

$$\mathbb{E} \sup_{t \in T} X_t \le C \,\mathbb{E} \sup_{t \in T} Y_t = C\, w(T)$$

QED !

(*) Technical note: X_t is not a centered random process as required in our statement of the Gaussian-sub-Gauss. comparison. This can be remedied by subtracting $\mathbb{E} X_t$, and noting that $\mathbb{E} X_t \le C\|t\|_2 \le C\, w(K)$ since $0 \in K$. Also, the sub-Gaussian norm between increments is at most doubled by centering.