Compressed sensing with convex optimization

Note Title 2015-10-31

Recall

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

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

Markov ≠

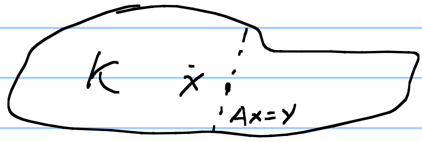
$$\Rightarrow P\Big(\sup_{x\in K} \big| \|Ax\|_2 - \sqrt{m}\|x\|_2 \big| \le C' w(K)\Big) \ge 99\%$$

Can be any const.

Application: Signal recovery in structured linear model:

$$y = Ax, \quad x \in K \subseteq \mathbb{R}^n$$

sub-Gaussian as in "master thm"



Recover, or estimate x, by solving program:

$$\text{Find } x' \in K \text{ s.t. } y = Ax' \qquad (1)$$

Let $\hat{x}$ be the solution.

Accuracy of $\hat{x}$?

Note: $\hat{x}, x \in K$ $\quad A\hat{x} = Ax = y$.

Let $h = \hat{x} - x$. Then $h \in K - K$, $h \in \operatorname{null}(A)$

$\|h\|_2$ is controlled via the low M^* estimate, which follows from the above thm.

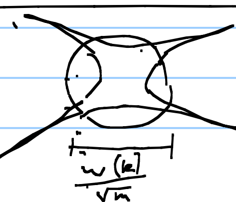
Thm (Low M^* estimate)

Under the conditions of theorem above

$$\operatorname{diam}(\operatorname{null}(A) \cap K) \le \frac{C\,w(K)}{\sqrt{m}}, \qquad \text{w.p.} > 99\%$$

Euclidean diameter

random subspace w/ codimension m.



Thus, w.p. > 99%,

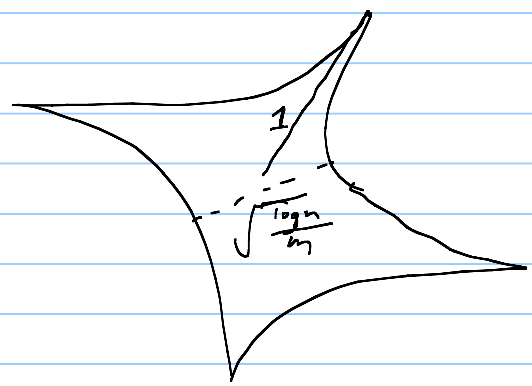
$$\boxed{\|h\|_2 \le C \frac{w(K-K)}{\sqrt{m}}}$$

Ex) $K = B_1$

$$w(K-K) = w(2B_1) = \mathbb{E} \sup_{x \in 2B_1} \langle x, g\rangle = \mathbb{E}\, 2\|g\|_\infty \le C\sqrt{\log n}$$

$$\Rightarrow \|h\|_2 \le C\sqrt{\frac{\log n}{m}}$$

$2B_1 =$



Signal processing interpretation:

As soon as $m \gg \sqrt{\log n}$, sol$\underline{n}$ becomes very accurate.

Geometric interpretation: Intersection of B_1 w/ a random hyperplane of codimension $m \gg \sqrt{\log n}$ has much smaller Euclidean diameter than that of B_1.

However, the approach above can be considerably improved.

Q: What if K is unbounded $\Rightarrow w(K) = \infty$?
Q: What if the program (1) is computationally intractible?

Both problems occur when $K_s = \{x \in \mathbb{R}^n : \|x\|_0 \le s\}$.

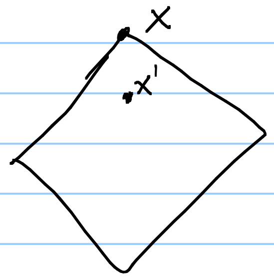
Solution: convexify and examine local properties of feasible set.

Motivating example:

$y = Ax$, $\quad x \in K_s$ $\quad$ Not convex. Replace w/ B_1^n.

Let $\hat{x}$ be a soln to the convex program:

Find x' s.t. $Ax = Ax'$, $\|x'\|_1 \le \|x\|_1$

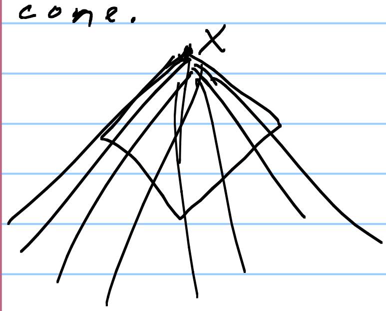


Set $h := \hat{x} - x$. $\quad$ Note:

(1) $h \in N(A)$

(2) $\|x+h\|_1 \le \|x\|_1$

Locally, the second constraint looks like a cone.



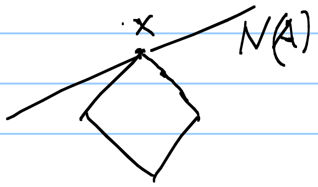
Def (Tangent cone)

$$\boxed{D(K,x) := \{\tau(v-x) : v \in K, \tau \ge 0\}}$$

Observe:

a) $h \in D(K,x)$

b) If $D(K,x) \cap N(A) = \{0\}$ then $h = 0$.



i.e. $N(A)$ "escapes" the tangent cone.

The prob. of this good event is captured by the following theorem:

Thm (Gordon's "escape through the mesh" thm) —

Let A be sub-Gaussian as in the "master theorem". Let D be a cone. Let

$$\epsilon = C \frac{w(D \cap B_2^n)}{\sqrt{m}}.$$

Then w.p. $\geq 99\%$, $\forall x \in D$,

A) $(1-\epsilon)\|x\|_2 \leq \frac{\|Ax\|_2}{\sqrt{m}} \leq (1+\epsilon)\|x\|_2$.

In particular, if $m \geq C w(D \cap B_2^n)$, then

B) $n(A) \cap D = \{0\}$ w.p. $\geq 99\%$.

A) "A is well conditioned when restricted to D if $m \geq C w(D \cap B_2^n)^2$."

Proof B) follows from the lower bound of A).

Exercize: Prove A) based on "master thm".

Return to example w/ $x \in K_s$, $h \in D(\|x\|_1 B_1^n, x)$.

One can show that

$$w\left(D(\|x\|_1 B_1^n, x) \cap B_2^n\right)^2 \leq C s \log\left(\frac{cn}{s}\right).$$

Thus, "escape through mesh" thm implies that $\hat{x} = x$ w.p. $\geq 99\%$ if $m \geq C s \log\left(\frac{cn}{s}\right)$.

Astonished researcher (2004): "x is recovered exactly! By a convex program! w/ hardly more measurements than would be needed for ℓ_0 minimization!"

General analysis using tangent cone:

Let $\hat{x}$ be a soln to

convex set containing x.

Find x' s.t. $Ax' = y$, $x' \in K$ (2)

$\boxed{\text{Corollary: If } m \geq C\, w(D(K,x) \cap B_2^n)^2, \text{ then w.p.} \geq 99\%, \ \hat{x} = x.}$

Proof: Same steps as in example above.

Remarks

- (2) is often computationally tractable.
- K is not necessarily the signal structure, rather it is a convex surrogate. Given a certain signal structure, determining what K to use is a question of interest.
- In some cases $w(K,x)^2$ can be tiny. This is what allows $m \ll n$, i.e., "dimension reduction".