ECE 8201: Low-dimensional Signal Models for High-dimensional Data Analysis

Lecture 9: Atomic norm for low-complexity signal models

Yuejie Chi
The Ohio State University

The Ohio State University

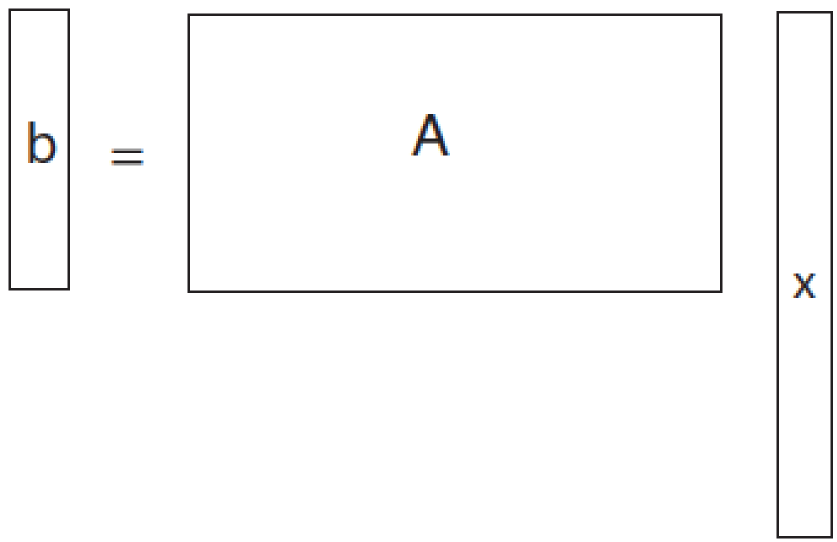
Main Reference

- Chandrasekaran, V., B. Recht, P. A. Parrilo, and A. S. Willsky. "The convex geometry of linear inverse problems." Foundations of Computational Mathematics 12, no. 6 (2012): 805-849.

Recap

- We start with an under-determined linear system:

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}$$



- Estimate $\boldsymbol{x} \in \mathbb{R}^n$ from linear measurements $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^m$, where $m \ll n$.

Recap

Hope?

- Constrain the set of interesting signals: looking for $\boldsymbol{x}$ with additional structures:
  - sparsity: $\|\boldsymbol{x}\|_0$ is small;
  - low-rankness: $\text{rank}(\text{reshape}(\boldsymbol{x}))$ is small;

- Constrain the measurement operator:
  - The measurement matrix $\boldsymbol{A}$ has to be "incoherent" to the postulated structure: for example, $\boldsymbol{A}$ cannot be a partial identity matrix for sparse $\boldsymbol{x}$;
  - More generically, this incoherence can be provided if we choose "random" measurement matrix, e.g. $\boldsymbol{A}$ composed of iid Gaussian entries provides "universal" guarantees.

- Efficient algorithms based on convex relaxations:

$$\min \|\boldsymbol{x}\|_1, \quad \text{s.t.} \quad \boldsymbol{b} = \boldsymbol{A}\boldsymbol{x},$$

  and nuclear norm minimization for rank minimization;

- advantages:
  - computationally efficient, many solvers have been developed to handle large-scale convex problems;
  - provable near-optimal performance in terms of sample complexity through the machinery of convex analysis;

- Question: can we extend this framework to other low-dimensional structures?

The Atomic Norm Approach

- The **atomic norm** is proposed by Chandrasekaran et.al. to find tightest convex relaxations of general parsimonious models.

- "Parsimony" refers to the fact that the signal of interest $\boldsymbol{x}$ can be described by a much small number of parameters than its ambient dimension;
  - For a known room, the image of the room is determined by the location and orientation of the camera, rather than the number of pixels of the image;
  - For an K-sparse vector $\boldsymbol{x}$, it can be described by $2K$ parameters;

- This models the signal of interest $\boldsymbol{x}$ as composed of atoms in an *atomic* set:

$$\mathcal{A} = \{\boldsymbol{a}_i\}$$

which could be infinite;

- The signal $\boldsymbol{x}$ can be written as a superposition of a small number of *atoms* in an atomic set $\mathcal{A}$:

$$\boldsymbol{x} = \sum_{i=1}^{r} c_i \boldsymbol{a}_i, \quad \boldsymbol{a}_i \in \mathcal{A},\ c_i > 0.$$

  Known examples:

  – Sparse case: $\mathcal{A}$ is composed of normalized vectors of sparsity one;
  – Low-rank case: $\mathcal{A}$ is composed of normalized matrices of rank one;

- Define the atomic norm (a.k.a. the "guage of $\boldsymbol{A}$") as

$$\begin{aligned}\|x\|_{\mathcal{A}} &= \inf\left\{t > 0 : x \in t\mathrm{conv}(\mathcal{A})\right\} \\ &= \inf\left\{\sum_i c_i \middle| x = \sum_i c_i a_i,\ a_i \in \mathcal{A}, c_i > 0\right\}.\end{aligned}$$

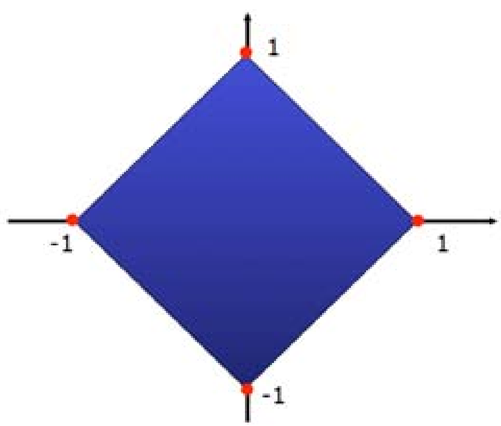
  if $\mathcal{A}$ is centrally symmetric about the origin (i.e., $\boldsymbol{a} \in \mathcal{A}$ if and only if $-a \in \mathcal{A}$) we have that $\|\cdot\|_{\mathcal{A}}$ is a norm. It is also a convex function.

Special cases of atomic norm

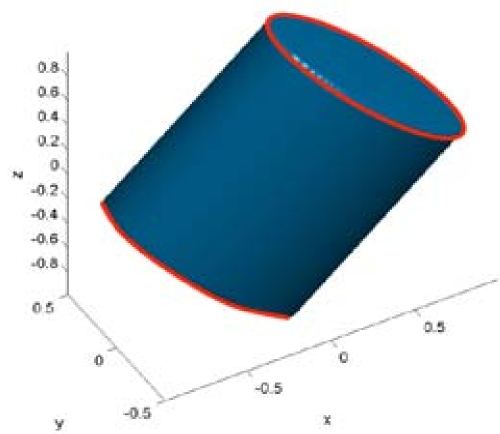
The atomic norm minimization recovers the ℓ_1 and nuclear norm:

- Sparse signals: *an atom* for *sparse signals* is a normalized vector of sparsity one, and the atomic norm is ℓ_1 norm;

- Low-rank matrices: *an atom* for *low-rank matrices* is a normalized rank-one matrix; and the atomic norm is nuclear norm;



unit ball of ℓ_1 norm



unit ball of nuclear norm

Atomic norm minimization

... for underdetermined linear systems:

- For a given atomic set $\mathcal{A}$, define the atomic norm $\|\boldsymbol{x}\|_{\mathcal{A}}$;

- Run the convex program:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_{\mathcal{A}} \quad \text{s.t.} \quad \boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}$$

- For noisy measurements:

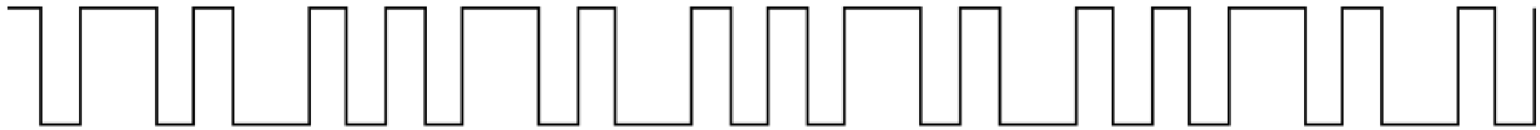
$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_{\mathcal{A}} \quad \text{s.t.} \quad \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2 \le \epsilon$$

Examples of atomic norm

- "democratic" signal representations via ℓ_∞ norm;
- "joint sparsity" or "group sparsity" via ℓ_1/ℓ_2 norm;

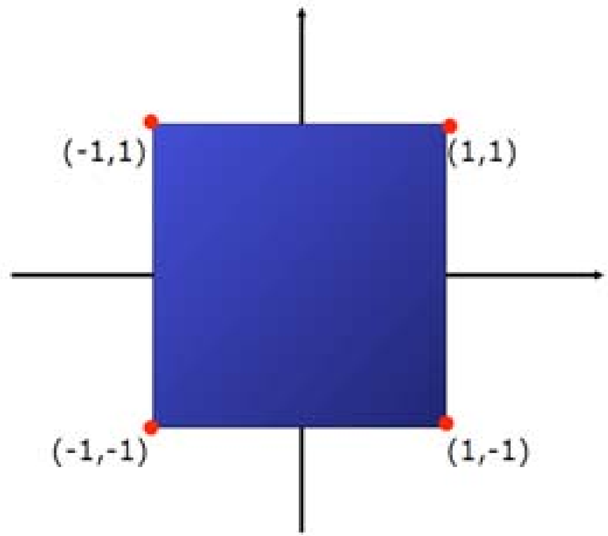
Democratic representations

- Signal representations $\boldsymbol{x}$ that have the same amplitudes for every entry;



  - motivated by integer programming $\boldsymbol{x} \in \{+1, -1\}^n$;
  - peak-to-average power ratio reduction in OFDM communication;

- The atomic set contains all sign vectors $\boldsymbol{A} = \{\{+1, -1\}^n\}$;
- The atomic norm becomes $\|\boldsymbol{x}\|_{\mathcal{A}} = \|\boldsymbol{x}\|_\infty$

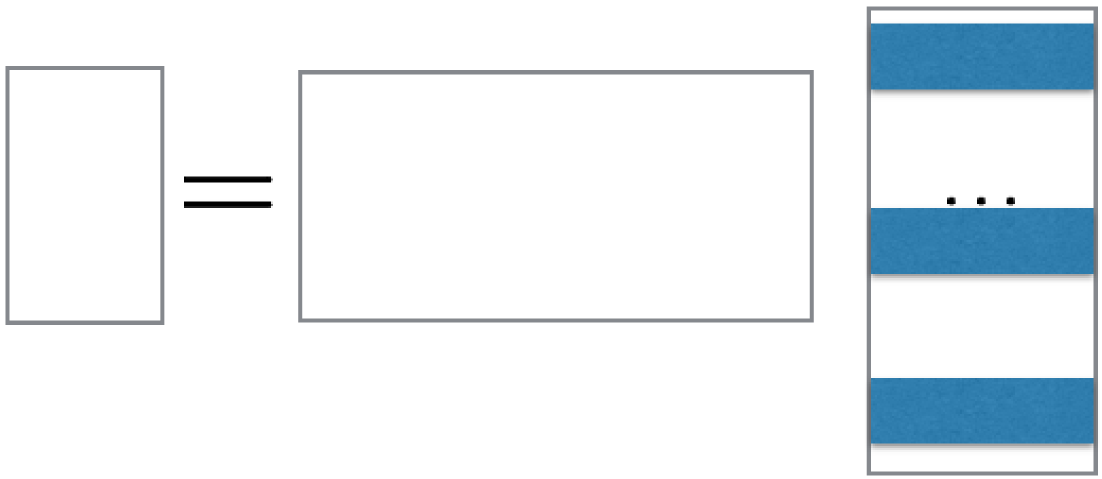


Joint sparsity

Consider sparse recovery with multiple snapshots:

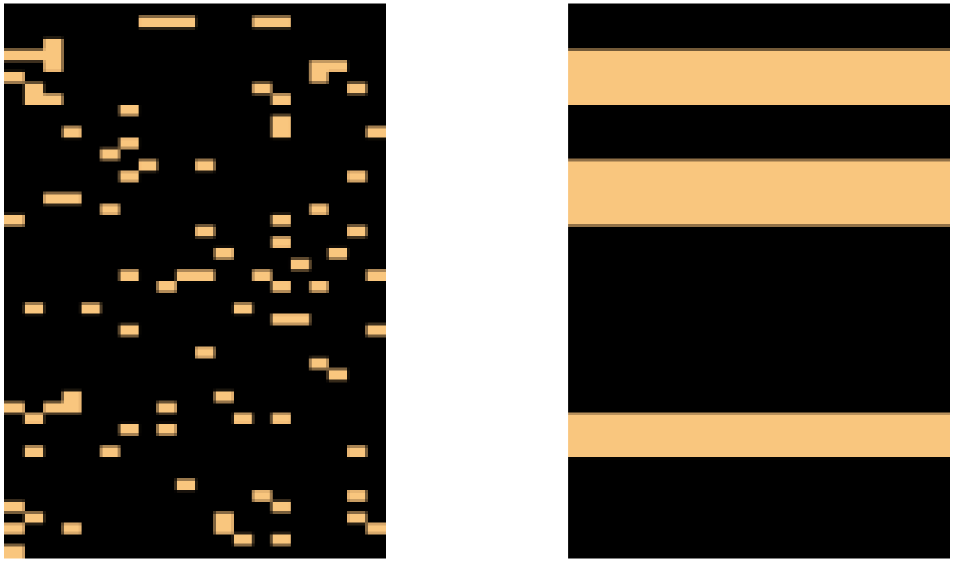
$$\boldsymbol{B} = \boldsymbol{A}\boldsymbol{X}$$

where $\boldsymbol{X} \in \mathbb{R}^{n \times T}$, where T is the number of snapshots.



- Different snapshots $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T]$ share the same support but have different coefficients;

- motivated by multi-task learning, array processing ,etc...



(a) Sparse (b) Group sparse

- The atoms in $\mathcal{A}$ can be written as rank-one matrix as

$$\mathcal{A} = \{\boldsymbol{e}_i \boldsymbol{u}_i^T | \|\boldsymbol{u}_i\|_2 = 1\}$$

where $\boldsymbol{e}_i$ is the ith standard basis vector, and $\|\boldsymbol{u}_i\|_2 = 1$.

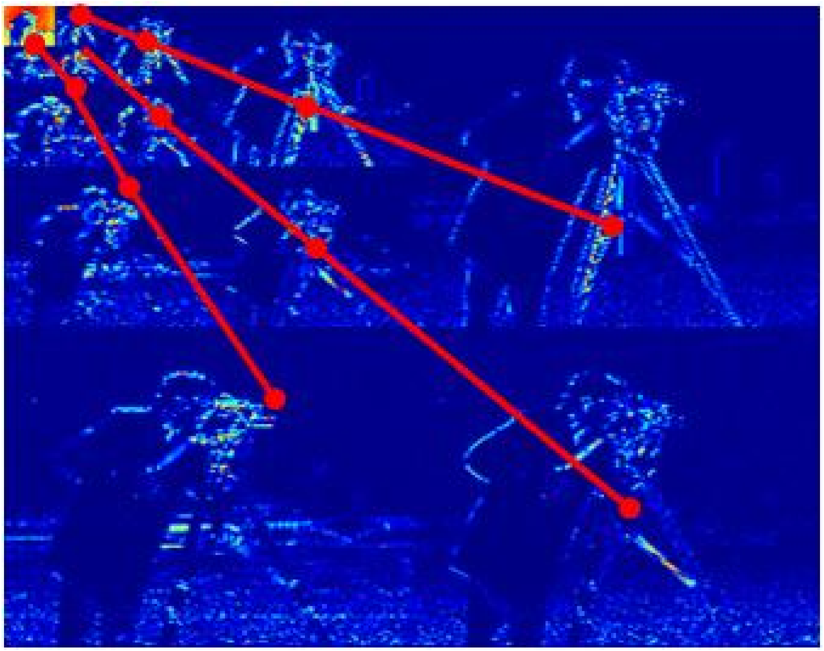
- The atomic norm becomes

$$\begin{aligned}
\|\boldsymbol{X}\|_{\mathcal{A}} &= \inf\{t > 0 : \boldsymbol{X} \in t\mathrm{conv}(\mathcal{A})\} \\
&= \inf\left\{\sum_i c_i \middle| \boldsymbol{X} = \sum_i c_i \boldsymbol{e}_i \boldsymbol{u}_i^T,\ \|\boldsymbol{u}_i\|_2 = 1, c_i > 0\right\} \\
&= \sum_{i=1}^{n} \left(\sum_{j=1}^{T} |x_{ij}|^2\right)^{1/2} \quad := \|\boldsymbol{X}\|_{1,2}
\end{aligned}$$

which is the ℓ_1/ℓ_2 norm of $\boldsymbol{X}$.

Group sparsity

The joint sparsity for multiple snapshots problem is related to the so-called group sparsity/structured sparsity.



We can group the coefficients into overlapping/non-overlapping groups, such that we motivate sparsity between groups, but not within the groups.

Performance with Gaussian measurements

If $\boldsymbol{A}$ is composed of i.i.d. Gaussian entries, how many measurements do we need to guarantee success recovery?

$$\hat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{x}} \|\boldsymbol{x}\|_{\mathcal{A}} \quad \text{s.t.} \quad \boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}$$

Surprisingly, this can be answered with a single notion called "Gaussian width"

- A convex set is a cone if it is closed under positive linear combinations.

- The tangent cone at $\boldsymbol{x}$ with respect to the scaled unit ball $\|\boldsymbol{x}\|_{\mathcal{A}}\text{conv}(\mathcal{A})$ is

$$T_{\mathcal{A}}(\boldsymbol{x}) = \text{cone}\{\boldsymbol{z} - \boldsymbol{x} : \|\boldsymbol{z}\|_{\mathcal{A}} \leq \|\boldsymbol{x}\|_{\mathcal{A}}\}$$

which is the set of descent directions of the atomic norm at the point $\boldsymbol{x}$,

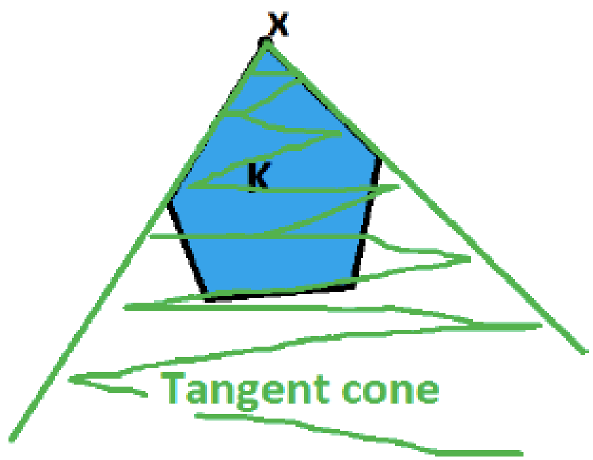


FIGURE 1. The tangent cone

- **Proposition:** We have that $\hat{\boldsymbol{x}} = \boldsymbol{x}^\star$ is the unique optimal solution if and only if $null(\boldsymbol{A}) \cap T_{\mathcal{A}}(\boldsymbol{x}^\star) = \{0\}$.

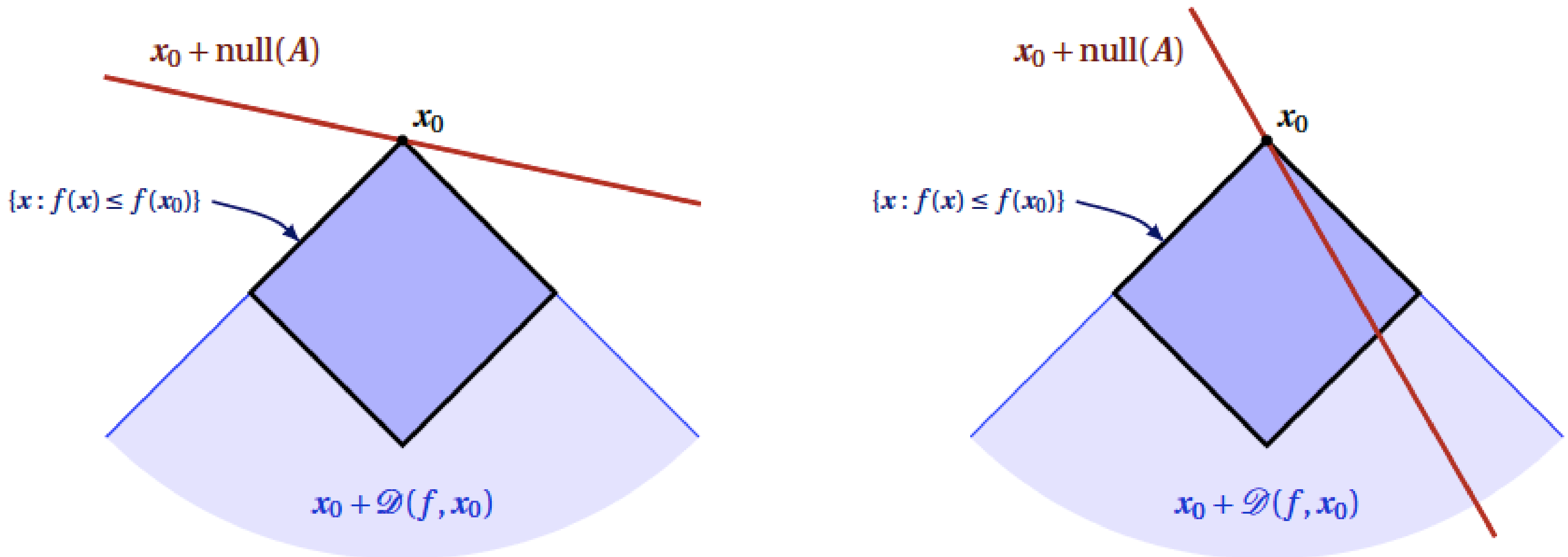


FIGURE 2.3: **The optimality condition for a regularized inverse problem.** The condition for the regularized linear inverse problem (2.4) to succeed requires that the descent cone $\mathscr{D}(f, x_0)$ and the null space null(A) do not share a ray. **[left]** The regularized linear inverse problem succeeds. **[right]** The regularized linear inverse problem fails.

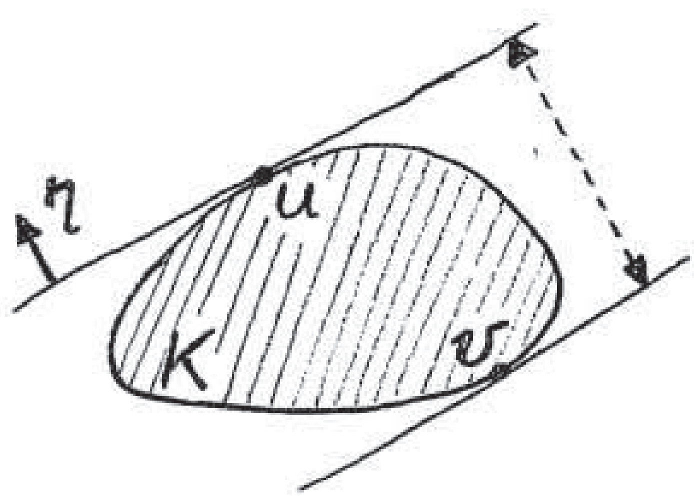
Gaussian width

- Definition: The Gaussian width of a set $S \subset \mathbb{R}^n$ is defined as:

$$w(S) := \mathbb{E}_{\boldsymbol{g}}\left[\sup_{\boldsymbol{z}\in S} \boldsymbol{g}^T\boldsymbol{z}\right], \quad \boldsymbol{g} \sim \mathcal{N}(0, I)$$

- Related to the "mean width":

$$w(S) = \frac{\lambda_p}{2}\int_{\mathbb{S}^{p-1}}\left(\max_{\mathbf{z}\in S}\mathbf{u}^T\mathbf{z} - \min_{\mathbf{z}\in S}\mathbf{u}^T\mathbf{z}\right) d\mathbf{u} = \frac{\lambda_p}{2}\, b(S)$$



- Gordon's escape through the mesh theorem: Let Ω be a closed subset of $\mathbb{S}^{n-1}$. Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ be composed of i.i.d. standard Gaussian entries. Then

$$\mathbb{E}\left[\min_{\boldsymbol{z}\in\Omega} \|\boldsymbol{A}\boldsymbol{z}\|_2\right] \geq \lambda_m - w(\Omega)$$

  where $\lambda_m = \mathbb{E}[\|\boldsymbol{g}\|_2] \leq \sqrt{m}$.

- Immediately, we have the number of measurements we need is $m \gtrsim w(\Omega)^2 + 1$: to guarantee

$$null(\boldsymbol{A}) \cap T_{\mathcal{A}}(\boldsymbol{x}^\star) = \{0\}$$

  by setting $\Omega = T_{\mathcal{A}}(\boldsymbol{x}^\star) \cap \mathbb{S}^{n-1}$.

Measurement bound with Gaussian width

Corollary 3.3. *Let $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}^n$ be a random map with i.i.d. zero-mean Gaussian entries having variance $1/n$. Further let $\Omega = T_{\mathcal{A}}(\mathbf{x}^\star) \cap \mathbb{S}^{p-1}$ denote the spherical part of the tangent cone $T_{\mathcal{A}}(\mathbf{x}^\star)$.*

1. *Suppose that we have measurements $\mathbf{y} = \Phi\mathbf{x}^\star$ and solve the convex program (5). Then $\mathbf{x}^\star$ is the unique optimum of (5) with probability at least $1 - \exp\left(-\frac{1}{2}\left[\lambda_n - w(\Omega)\right]^2\right)$ provided*

$$n \geq w(\Omega)^2 + 1\,.$$

Underlying model	Convex heuristic	# Gaussian measurements
s-sparse vector in $\mathbb{R}^p$	ℓ_1 norm	$2s\log(p/s) + 5s/4$
$m \times m$ rank-r matrix	nuclear norm	$3r(2m - r)$
sign-vector $\{-1, +1\}^p$	ℓ_∞ norm	$p/2$
$m \times m$ permutation matrix	norm induced by Birkhoff polytope	$9\,m\log(m)$
$m \times m$ orthogonal matrix	spectral norm	$(3m^2 - m)/4$

Table 1: A summary of the recovery bounds obtained using Gaussian width arguments.