SCIENCES SORBONNE UNIVERSITÉ

Conditions for ℓ^1-recovery

Claire Boyer

Summary

Minimal number of measurements for ℓ^0-min

Recall:

Theorem

Given $A \in \mathbb{C}^{m \times d}$, the following properties are equivalent:

(a) *Every s-sparse vector $x \in C^d$ is the unique s-sparse solution of $Az = Ax$, that is, if $Ax = Az$ and both x and z are s-sparse, then $x = z$.*

(b) *The null space $\ker A$ does not contain any 2s-sparse vector other than the zero vector, that is,$\ker(A) \cap \{z \in \mathbb{C}^d, \|z\|_0 \leq 2s\} = \{0\}$.*

(c) *For every $S \subset [|d|]$ with $|S| \leq 2s$, the submatrix A_S is injective as a map from $\mathbb{C}^S$ to $\mathbb{C}^m$.*

(d) *Every set of 2s columns of A is linearly independent.*

A partial proof

Zoom

(a) Every s-sparse vector $x \in \mathbb{C}^d$ is the unique s-sparse solution of $Az = Ax$, that is, if $Ax = Az$ and both x and z are s-sparse, then $x = z$.

(b) The null space $\ker A$ does not contain any $2s$-sparse vector other than the zero vector, that is, $\ker(A) \cap \{z \in \mathbb{C}^d, \|z\|_0 \leq 2s\} = \{0\}$.

(b)⇒(a). Let x and z be s- sparse with $Ax = Az$. Then $x - z$ is $2s$-sparse and $A(x - z) = 0$. If the kernel does not contain any $2s$-sparse vector different from the zero vector, then $x = z$.
(a)⇒(b). Conversely, assume that for every s-sparse vector $x \in \mathbb{C}^d$, we have $\{z \in \mathbb{C}^d, \|z\|_0 \leq 2s\} = \{x\}$. Let $v \in \ker(A)$ be $2s$-sparse. We can write $v = x - z$ for s-sparse vectors x, z with $\mathrm{supp} x \cap \mathrm{supp} z = \emptyset$. Then $Ax = Az$ and by assumption $x = z$. Since the supports of x and z are disjoint, it follows that $x = z = 0$ and $v = 0$.

About the rest of the proof

For the equivalence of (b), (c), and (d), we observe that for a $2s$-sparse vector v with $S = \mathrm{supp} v$, we have $Av = A_S v_S$, where A_S is the extracted matrix from A which columns are indexed by S.

Noting that $S = \mathrm{supp} v$ ranges through all possible subsets of $[|n|]$ of cardinality $|S| \leq 2s$ when v ranges through all possible $2s$-sparse vectors completes the proof by basic linear algebra.

Minimal number for ℓ^0 recovery

We observe, in particular, that if it is possible to reconstruct every s-sparse vector $x \in \mathbb{C}^d$ from the knowledge of its measurement vector $y = Ax \in C^m$, then (a) holds and consequently so does (d).
$\Rightarrow$ rank$(A) \geq 2s$, and
$\Rightarrow$ rank$(A) \leq m$, because the rank is at most equal to the number of rows.

Minimal number of measurements for ℓ^0 recovery:

the number of measurements needed to reconstruct every s-sparse vector always satisfies

$$m \geq 2s.$$

Summary

But ℓ^0 recovery is NP-hard

Idea: relax the ℓ_0 (pseudo) norm into a convex function because convex function can be efficiently minimized!

Theorem

For a function f, the envelope (or biconjugate)

$$f^{\star\star}(z) = \sup_y \langle y, z \rangle - f^\star(y).$$

is the pointwise supremum of all the affine functions on $\mathbb{R}^d$ majorized by f.

Proof: Left as an aside exercise.
Hint: Introduce an auxiliary variable in the definition of $f^{\star\star}$ above.

Example of biconjugate

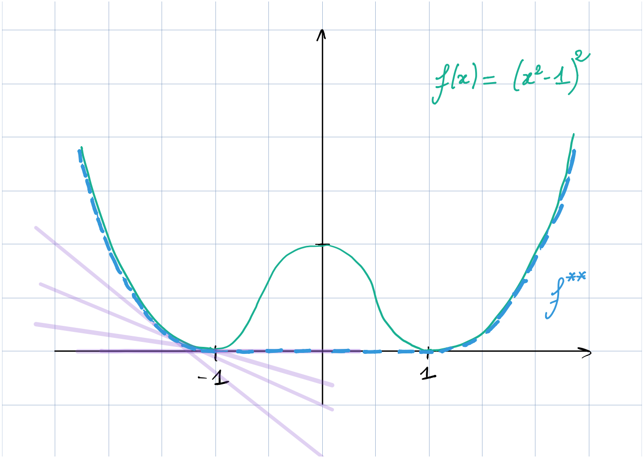

Figure: The biconjugate $f^{\star\star}$ is the close-convexification of f, as the pointwise supremum of all affine functions minorizing f.

The biconjugate of the ℓ^0-norm

Theorem

The ℓ^1-norm is the convex envelope of the counting ℓ^0-norm on the ℓ^∞ unit-ball $B_\infty^d = [-1,1]^d$.

Exercise: prove it.

Recall that the conjugate function is defined as follows for $y \in \mathbb{R}^d$

$$f^\star(y) = \sup_{x \in [-1,1]^d} \langle x, y \rangle - f(x).$$

Assume some ordering σ such that $y_{\sigma(i)}^2 \geq y_{\sigma(i-1)}^2$ and introduce the auxiliary variable r, then

$$f^\star(y) = \max_r \sup_{x \in [-1,1]^d} \left\langle x, y_{\sigma(1:r)} \right\rangle - f(x).$$

The biconjugate of the ℓ^0-norm

Once r is fixed, the supremum is achieved when the inner product is maximized and given the constraint $x \in [-1,1]^d$ that is $x_i = \mathrm{sign}(y_{\sigma(i)})$. Hence,

$$f^\star(y) = \max_r \sum_{i=1}^{r} (|y_{\sigma(i)}| - 1) = \sum_{i=1}^{d} (|y_i| - 1)_+.$$

The biconjugate can be then computed, for $z \in [-1,1]^d$,

$$f^{\star\star}(z) = \sup_{y \in \mathbb{R}^d} \langle z, y \rangle - \sum_{i=1}^{d} (|y_i| - 1)_+.$$

The last supremum is separable in coordinates, so one can focus only on one coordinate and consider cases when the positive part is zero and non-negative. Elementary reasoning leads to

$$f^{\star\star}(z) = \sum_{i=1}^{d} |z_i|, \qquad \text{for} \quad z \in B_\infty^d.$$

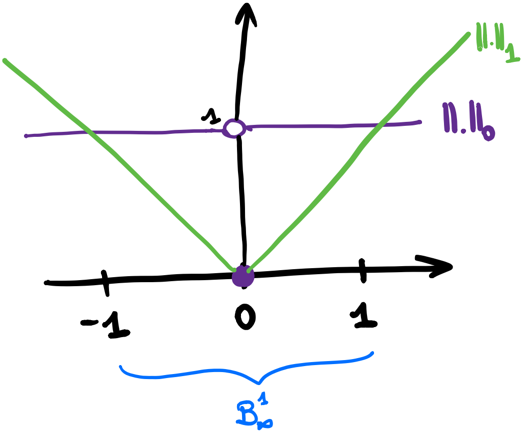

Figure: On B_∞, the ℓ^1 norm is the supremum of affine functions minorizing the ℓ^0-norm.

The biconjugate of the ℓ^0-norm

The ℓ^1-norm can also appear natural as a convex function able to capture sparsity:

- The k-sparse vectors set can be constructed by a combination of k elements of the so-called atomic set $\mathcal{A} = \{\pm e_i\}_{1 \leq i \leq n}$.
- The convex hull of $\mathcal{A}$ is given by the unit ball of the ℓ^1-norm.
- One can then construct an atomic norm associated to conv$\mathcal{A}$, as the gauge associated to conv$\mathcal{A}$ which leads to the ℓ^1-norm.

Relax and conquer

Definition

Given a matrix $A \in \mathbb{R}^{m\times d}$ and a measurement vector $y \in \mathbb{R}^m$ the basis pursuit procedure returns

$$\hat{x} \in \operatorname*{arg\,min}_{\substack{z\in\mathbb{R}^d \\ y=Az}} \|z\|_1 \tag{BP}$$

when there exists a solution to the equation $Az = y$ and $\emptyset$ otherwise.

Recall that

Basis Pursuit $\equiv$ ℓ^1-minimization with equality constraints

Relax and conquer

Definition

Let $A : \mathbb{R}^d \to \mathbb{R}^m$ such that $m \leq n$. We say that A satisfies the exact recovery property of order s (ER(s)), if for any s-sparse vector$x$, one has

$$\arg\min_{\substack{z \in \mathbb{R}^d \\ y = Ax = Az}} \|z\|_1 = \{x\}$$

In this lecture, we will introduce conditions on the sensing matrix A to ensure exact recovery via (BP) for the whole set of sparse vectors

$$\Sigma_s := \{z \in \mathbb{R}^d : \|z\|_0 \leq s\}.$$

Why ℓ^1-min captures sparsity?

Define the descent cone of the norm $\|\cdot\|$ at a point x as

$$\mathcal{D}_{\|\cdot\|_1}(x) = \left\{ d \in \mathbb{R}^d : \exists c > 0, \|x + cd\|_1 \leq \|x\|_1 \right\}$$

This convex cone[1] is the set of non-ascent directions of $\|\cdot\|_1$ at x.

A first condition

A (feasible) point x is the unique solution to (BP) if and only if the null space of A misses the cone descent at x, i.e.

$$\ker(A) \cap \mathcal{D}_{\|\cdot\|_1}(x) = \{0\}. \tag{1}$$

[1] A cone is a set closed under positive linear combinations

Why ℓ^1-min captures sparsity?

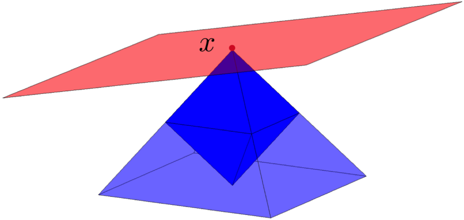

Figure: Ball associated with the ℓ^1 norm with the affine feasible set for (BP). When the feasible set is tangent to the ball, the solution to (BP) is exact.

The descent cone to the ℓ^1 norm is "narrow" at sparse vectors and, therefore, even though the null space is of small codimension m, it is likely that if m is large enough, it will miss the descent cone.

Relax and conquer

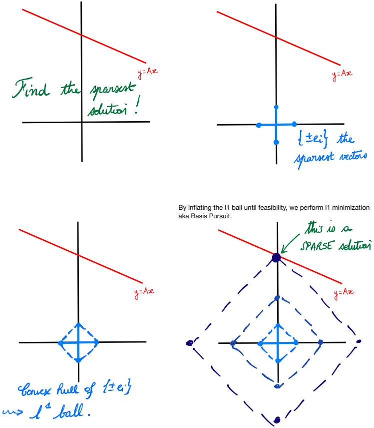

Figure: ℓ^1-minimization leads to sparse recovery: the ℓ^1 ball being spiky, it will hit the feasible set on one of its corner,i.e. for a sparse vector.

Summary

Price to pay for relaxation?

- Since we relax the ℓ^0 cost function into the ℓ^1 one, going from an NP-hard problem to a linear program, one should expect to pay this price in some way.
- We have seen that the ℓ^0 minimization problem can solve CS problem using only at least m measurements. of the order of $2s$.

$\rightsquigarrow$ What about the ℓ^1-minimization problem?

Proposition

Let $A : \mathbb{R}^d \to \mathbb{R}^m$ be a matrix satisfying $ER(2s)$, then

$$m \geq \frac{1}{\log 3} \left\lfloor \frac{s}{2} \right\rfloor \log \left(\left\lfloor \frac{n}{8es} \right\rfloor \right).$$

Two main ingredients for the proof

Lemma

Let $s \leq n/2$. There exists a family of sets $\mathcal{S}$ from $[|n|]$ such that

1. $\forall S \in \mathcal{S}$, $|S| = s$;
2. $\forall S_1, S_2 \in \mathcal{S}$, $S_1 \neq S_2 \Longrightarrow |S_1 \cap S_2| \leq \lfloor s/2 \rfloor$;
3. $\log(|\mathcal{S}|) \geq \left\lfloor \frac{s}{2} \right\rfloor \log \left\lfloor \frac{n}{8es} \right\rfloor$.

Proof: Admitted.

Lemma

Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$. Let B be its associated unit ball. Let $0 < \epsilon \leq 1$ and $\lambda \subset B$ such that for all $x, y \in \Lambda$, $\|x - y\| \geq \epsilon$. Necessarily,

$$|\Lambda| \leq \left(1 + \frac{2}{\epsilon}\right)^d.$$

Proof: To be done.

Conclusion on the min nb of meas. for ℓ^1 min.

Minimal number of measurements for ℓ^1 recovery (BP)

The number of measurements needed to reconstruct every s-sparse vector always satisfies

$$m \gtrsim s \log(en/s)$$

up to a universal constant.

- It means that we are paying an extra log factor compared to the ℓ^0-minimization procedure.
- A log factor is (almost) nothing.
- We gain more on a computational side than we lose on the theoretical number of measurements.

Summary

Going further than a necessary condition

- We have already seen the necessary condition

$$\ker(A) \cap \Sigma_{2s} = \{0\}$$

  to find a decoder to reconstruct any s-sparse vector x from the linear measurements $y = Ax$.
- Then we allow only 0 to be the only s-sparse element of $\ker(A)$.
- This condition ensures that (injectivity on the set of s-sparse vectors)

$$Ax \neq Ax' \qquad \text{for } x \neq x' \in \Sigma_s$$

In this section, we also want to ensure that not only their intersection is $\{0\}$ but that $\ker(A)$ is far enough from Σ_{2s}. This property is called the null-space property.

Null Space Property

Definition

A matrix $A \in \mathbb{K}^{m \times d}$ is said to satisfy the null space property relative to a set $S \subset [|n|]$ if

$$\|v_S\|_1 < \|v_{S^c}\|_1,$$

for all $v \in \ker(A) \setminus \{0\}$. It is said to satisfy the null space property of order s if it satisfies the null space property relative to any set $S \subset [|n|]$ with $|S| \leq s$.

- The NSP quantifies the notion that vectors in the null space of A should not be too concentrated on a small subset of indices. ex: if a vector h is exactly s-sparse, then there exists a set of indices Λ such that $\|h_{\Lambda^c}\|_1 = 0$ and hence NSP implies that $h_\Lambda = 0$ as well.

Remarks on NSP

- If a matrix A satisfies the NSP then the only s-sparse vector in $\ker(A)$ is $h = 0$.
- It is important to observe that, for a given $v \in \ker(A) \setminus \{0\}$, the condition $\|v_S\|_1 \leq \|v_{S^c}\|_1$ holds for any set $S \subset [|n|]$ with $|S| \leq s$ as soon as it holds for an index set of s largest (in modulus) entries of v.

Reformulation of NSP

Remark (Two convenient versions of NSP)

1. *Adding $\|v_S\|_1$ to both sides, NSP relative to S reads*

$$2\|v_S\|_1 < \|v\|_1$$

for all $v \in \ker(A) \setminus \{0\}$.

2. *Choosing S as an index set of s largest (in modulus) entries of v and this time by adding $\|v_{S^c}\|_1$ to both sides of the inequality, NSP relative to S reads*

$$\|v\|_1 < 2\sigma_{s,1}(v)$$

for all $v \in \ker(A) \setminus \{0\}$, where $\sigma_{s,1}(v)$ is the ℓ^1 error of the best s-term approximation of v defined by $\sigma_{s,1}(v) := \min_{z:\|z\|_0 \leq s} \|x - z\|_1$.

How strong is NSP?

Question : Is NSP is a necessary and sufficient condition for recovery success via (BP)? Is it for ℓ^0-minimization ?

Recall: a necessary and sufficient condition for ℓ^0 recovery

For $A \in \mathbb{K}^{m \times d}$, the following assertions are equivalent:

1. *Every s-sparse vector $x \in \mathbb{C}^d$ is the unique s-sparse solution of $Az = Ax$, that is, if $Ax = Az$ and both x and z are s-sparse, then $x = z$.*
2. *The null space $\ker A$ does not contain any $2s$-sparse vector other than the zero vector, that is, $\ker(A) \cap \Sigma_{2s} = \{0\}$.*

Proposition

Let $A \in \mathbb{K}^{m \times d}$. If A satisfies the NSP of order s then

$$\ker(A) \cap \Sigma_{2s} = \{0\}.$$

- Though, NSP is stronger than $\ker(A) \cap \Sigma_{2s} = \{0\}$.

Proof: NSP $\Rightarrow \ell^0$-recovery

Let $v \in \ker(A) \cap \Sigma_{2s}$ such that $v \neq 0$. If A satisfies NSP(s), it means that for (any vector) $v \in \ker(A) \setminus \{0\}$, one has

$$v_1' + v_2' + \ldots + v_s' < \frac{\|v\|_1}{2},$$

where $(v_i')_i$ is the sorting permutation in decreasing order of $(|v_j|)_j$, i.e. $v_1' \geq v_2' \geq \ldots \geq v_d' \geq 0$. Since $v \in \Sigma_{2s}$, one can also write

$$v_1' + v_2' + \ldots + v_s' \geq v_{s+1}' + \ldots + v_d' = \|v\|_1 - (v_1' + v_2' + \ldots + v_s'),$$

$$v_1' + v_2' + \ldots + v_s' \geq \frac{\|v\|_1}{2}.$$

Therefore, by contradiction, $v = 0$.

NSP and approximately sparse vectors

Remark

Nearly s-sparse vectors are also prohibited to be in $\ker(A)$ *under NSP*(s)*. For instance, if A satisfies NSP*(s)*, then*

$$v = (1, \ldots 1, \frac{s}{n-s}, \ldots, \frac{s}{n-s}) \notin \ker(A),$$

since $\|v_S\|_1 = \|v_{S^c}\|_1$ *with* $S = \{1, \ldots, s\}$*.*

NSP $\iff$ uniform ℓ^1-recovery

Theorem

Let $A \in \mathbb{K}^{m \times d}$. The following assertions are equivalent:

1. *A satisfies $ER(s)$.*
2. *A satisfies $NSP(s)$.*

NSP is a Necessary and Sufficient Condition for ℓ^1-recovery!
Proof: let's do it (blackboard time).

Proof: NSP $\Longleftrightarrow$ uniform ℓ^1-recovery

[$\Rightarrow$] If A satisfies $\mathrm{ER}(s)$ then for all $x \in \Sigma_s$, x is the unique element of $x + \ker(A)$ with a minimal ℓ^1-norm, i.e. $\forall v \in \ker(A) \setminus \{0\}$, $\|x\|_1 < \|x+v\|_1$. Let $J := \mathrm{supp}(x)$. One has $\|x+v\|_1 = \|x+v_J\|_1 + \|v_{J^c}\|_1$. This being true for any $x \in \Sigma_s$, one can choose x such that

$$\|x + v_J\|_1 = \|x\|_1 - \|v_J\|_1$$

(take (x_i)'s large enough and with opposite signs of the $(v_i)'s$). Then for all $J \subset [|n|]$ such that $|J| = s$, for all $v \in \ker(A) \setminus \{0\}$, there exists $x \in \Sigma_s$, such that

$$\|x\|_1 < \|x+v\|_1 = \|x\|_1 - \|v_j\|_1 + \|v_{J^c}\|_1,$$

and then $\|v_J\| < \|v_{J^c}\|_1$. This is true for any $v \in \ker(A) \setminus \{0\}$, then A satisfies the $\mathrm{NSP}(s)$.

[$\Leftarrow$]. Suppose that A satisfies the $\mathrm{NSP}(s)$. Let $x \in \Sigma_s$, note $J = \mathrm{supp}(x)$. Let $v \in \ker(A) \setminus \{0\}$,

$$\|x+v\|_1 = \|x+v_J\|_1 + \|v_{J^c}\|_1 \geq \|x\|_1 - \|v_J\|_1 + \|v_{J^c}\|_1 > \|x\|_1.$$

Then A satisfies $\mathrm{ER}(s)$.

What's next?

- NSP is thus a necessary and sufficient condition for a matrix A to ensure uniform recovery via (BP).
- We have to construct sensing matrices satisfying NSP(s) with a minimal number of rows (i.e. a minimal number of measurements) to solve CS problem using (BP).

- However, verifying such a condition is impossible in practice.
- Use instead two stronger/more restrictive properties that imply NSP(s).
- These two conditions are then sufficient - whereas NSP is necessary and sufficient - for (BP) success.

Summary

Towards a stronger condition

- Another condition is linked to asymptotic theory in Banach spaces.
- It relies on euclidean sections of the unit ℓ^1 ball B_1^d in dimension d.
- It says that the null space of A is going to intersect the unit ℓ^1 ball in a very peculiar way.

Some notation

- L^p-ball in dimension n: $B_p^d = \{x \in \mathbb{K}^d, \quad \|x\|_p \leq 1\}$.
- One has $\rightsquigarrow$ Draw it!

$$\frac{1}{\sqrt{d}} B_2^d \subset B_1^d \subset B_2^d$$

- $\forall x \in \mathbb{K}^d$,

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2$$

  (think of these norms as gauges associated to the convex sets that are the balls)
- The bounds are reached respectively for $x = e_1$ and $x = \mathbb{1}_d$.
- In particular

$$\operatorname{diam}(B_1^d, \ell^2) := \sup_{x \in \mathbb{K}^d, x \in B_1^d} \|x\|_2 = 1,$$

  reached for $x = e_1$.

The ℓ^2 diameter is maximal on the canonical axes of $\mathbb{K}^d$, which are the sparsest vectors in $\mathbb{K}^d$ since only one coordinate is non-zero.

Getting intuition

- Under NSP(s), ker(A) does not contain the sparse vectors set Σ_{2s}. Then ker(A) is far enough from the canonical axes and from all the $2s$-dimensional spaces spanned by canonical basis vectors.
- Intuitively, ker(A) will be directed only diagonal directions in $\mathbb{K}^d$.
- For these vectors said to be well-spread, the ratio of their ℓ^2 norm over their ℓ^1-norm is much smaller than 1.
- Intuition: NSP(s) will be verified if the ℓ_2 norm of vectors in $B_1^d \cap \ker(A)$ is much smaller than 1.

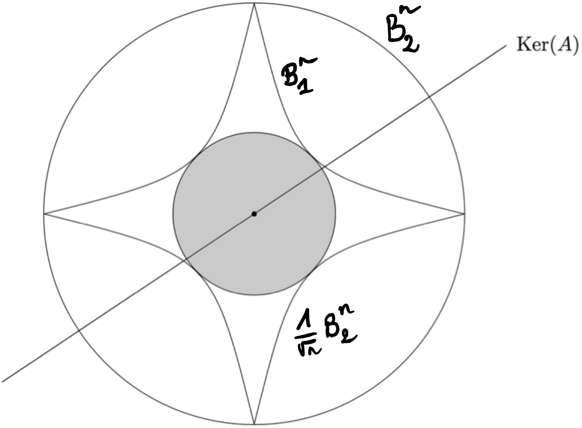

Figure: In high dimension, the ℓ^1 ball is very spiky.

Gelfand property

"Intuition: NSP(s) will be verified if the ℓ_2 norm of vectors in $B_1^d \cap \ker(A)$ is much smaller than 1."

Definition

We say that A verifies the Gelfand property of order s if

$$\mathrm{diam}(B_1^d \cap \ker(A), \ell_2) < \frac{1}{2\sqrt{s}}$$

We will write that A satisfies Gelfand(s).

Gelfand property

- The dimension of the null space of A is at least of $n - m$.
- Usually $m \ll n$, then $\ker(A)$ is a very big linear space, of dimension almost n.
- We have seen that

$$\operatorname{diam}(B_1^d, \ell^2) := \sup_{x \in \mathbb{K}^d, x \in B_1^d} \|x\|_2 = 1,$$

and the Gelfand condition requires that

$$\operatorname{diam}(B_1^d \cap \ker(A), \ell_2) < \frac{1}{2\sqrt{s}}$$

that is to say that the restriction of B_1^d has a much smaller diameter than 1, even if $\ker(A)$ is of large dimension.

A sufficient condition: Gelfand $\Rightarrow$ NSP

However, this restrictive condition implies NSP.

Theorem

Let $A \in \mathbb{R}^{m \times d}$. If A satisfies Gelfand(s), then A satisfies NSP(s).

Proof: Let us suppose that A satisfies Gelfand(s). Then for any $v \in \ker(A) \setminus \{0\}$,

$$\|v\|_2 < \|v\|_1 / 2\sqrt{s}.$$

Let $v \in \ker(A) \setminus \{0\}$, let $J \subset [|n|]$ such that $|J| = s$, then

$$\|v_J\|_1 \leq \sqrt{s}\|v_J\|_2 \leq \sqrt{s}\|v\|_2 < \sqrt{s}\frac{1}{2\sqrt{s}}\|v\|_1 = \frac{\|v\|_1}{2}.$$

where we used Cauchy-Schwarz in the first inequality.

Summary

Another condition: the RIP

Now let us introduce a famous (while non-trivial to verify) condition to ensure (BP) success.

Definition

Let $A \in \mathbb{K}^{m \times d}$. We say that A satisfies the restricted isometry property (RIP) of order s if for all $x \in \Sigma_s$,

$$\frac{1}{2}\|x\|_2^2 \leq \|Ax\|_2^2 \leq \frac{3}{2}\|x\|_2^2.$$

We say that A satisfies RIP(s).

RIP condition is the most popular in CS. If A satisfies RIP(s) then it acts like an isometry on the set of s-sparse vectors. Of course, A cannot be an isometry on $\mathbb{K}^d$, because it is a rectangular matrix, it is a compression matrix: $\ker(A)$ is not reduced to $\{0\}$ but of high dimension. RIP(s) only asks for A to be an isometry on a subset of $\mathbb{K}^d$, Σ_s.

Comments on the RIP

Remark (A general version of RIP)

More generally, one can associate a constant δ_s defined as the smaller positive number δ such that for all $x \in \Sigma_s$

$$(1 - \delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta)\|x\|_2^2.$$

We say that A satisfies the restricted isometry property if δ_s is small for reasonably large s.
We keep the constants 1/2 and 3/2 for simplicity.

Remark

In all the previous conditions (ER, NSP and Gelfand), only the null-space of A is involved. In particular, these conditions are invariant to renormalizations, since $\ker(\lambda A) = \ker(A)$ for all $\lambda \neq 0$.
Though, in RIP definition, the matrix A is directly involved, then one should be careful to its normalization.

How strong is the NSP?

The RIP condition is very strong, as it can be seen from the following theorem.

Theorem

If $A \in \mathbb{K}^{m \times d}$ satisfies RIP(s) then A satisfies Gelfand$(s/65)$.

Proof: Let's do it. (Blackboard time)

Remark

We did not try to optimize the constants!

Summary

Exercise ((☕) NSP when positivity constraints)

Given $A \in \mathbb{R}^{m \times d}$, *prove that every nonnegative s-sparse vector* $x \in \mathbb{R}^d$ *is the unique solution of*

$$\min_{z \in \mathbb{R}^d} \|z\|_1 \quad \text{such that} \quad Az = Ax \quad \text{and} \quad z \geq 0,$$

if and only if

$$v_{S^c} \geq 0 \Rightarrow \sum_{j=1}^{d} v_j > 0,$$

for all $v \in \ker(A) \setminus \{0\}$ *and all* $S \subset [|n|]$ *with* $|S| \leq s$.

Exercise ((☕) Minimal number of measurements and RIP)

Let $A \in \mathbb{R}^{m \times d}$ be a matrix verifying the RIP of order 2s with constant 1/2, i.e. such that for all $x \in \Sigma_{2s}$, $1/2\|x\|_2 \leq \|Ax\|_2 \leq 3/2\|x\|_2$. Then, show that there exist universal constants $c_0, c_1 > 0$ such that

$$m \geq c_0 s \log(c_1 n/s).$$

Hint: we already showed that one can construct a 1-separated set of $\Sigma_s \cap B_2^d$ for the ℓ_d^2 metric with cardinality at least $(c_0 n/s)^{s/c_1}$.
Hint 2: use this to construct a separated set in B_2^m.
Hint 3: use the volumic argument.

Exercise ((☕) RIP⇒NSP)

Show that RIP implies the NSP.
More explicitly, let $A \in \mathbb{R}^{m \times d}$ *satisfy the RIP of order* $2s$ *with constant* $\delta_{2s} < 1/3$, *i.e.*

$$(1 - \delta_{2s})\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_{2s})\|x\|_2^2,$$

holds for all $2s$*-sparse vectors* x.
show that A *satisfies the NSP of order* s, *i.e. for any* S *such that* $|S| \leq s$, *and any* $v \in \ker(A) \setminus \{0\}$, *it holds*

$$2\|v_S\|_1 < \|v\|_1.$$

RIP $\Rightarrow$ NSP

Hint 1: first show that if x and y are s-sparse with disjoint supports,

$$\langle Ax, y\rangle \leq \delta_{2s}\|x\|_2\|y\|_2$$

Hint 2: for $v \in \ker(A) \setminus \{0\}$, let T_0 the set of indexes corresponding to the s-largest entries of v. Let $T_0^c = T_1 \cup T_2 \cup \ldots$ be a partition of T_0^c such that T_1 contains the s-largest entries of $v_{T_0^c}$, T_2 contains the s-largest entries of $v_{T_0^c \setminus T_1}$ and so on...

Hint 3: Show the null space property for T_0 (and the NSP for all T will follow) using RIP + Hint 1.

Take-home message

To put in a nutshell

We showed that

$$\mathrm{RIP}(65s) \Longrightarrow \mathrm{Gelfand}(s) \Longrightarrow NSP(s) \Longleftrightarrow ER(s)$$

(uniform ℓ^1-recovery)

Next step: construct sensing matrices satisfying RIP.

Remark

- *The NSP(s) is equivalent to ER(s), then this is a tight condition for ER(s).*
- *However it is unpractical because it is quasi not verifiable.*
- *The RIP condition, even if it is more restrictive, gives a quantitative criterion, which is much easier to deal with.*