COMPRESS SENSING PROJECT

ON THE GAP BETWEEN RESTRICTED ISOMETRY PROPERTIES AND SPARSE RECOVERY CONDITIONS

AUTHORS

NICOLAS CLOAREC & SUZANNE SIGALLA

1 Introduction

We introduce here the framework of [6] and [3]. We place ourselves in the standard Compressed Sensing setup : we observe $y_i = \langle X_i, \hat{x} \rangle$ for $i = 1, \dots, m$ and $\hat{x} \in \mathbb{R}^n$ is the sought signal, and $m << n$. Our objective is to identify $\hat{x}$ using the measurements $(y_i)_{i=1}^m$. What allows the identification of $\hat{x}$ is the sparsity hypothesis : we assume that most of $\hat{x}$'s coordinates equal 0, i.e. $\hat{x}$'s support is of size s with $s << n$. We denote by Σ_s the set of all s sparse vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$ depending of the context and consider the measurement matrix :

$$\Gamma = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \langle X_i, \cdot \rangle f_i = \frac{1}{\sqrt{m}} (X_1, \dots, X_m)^T$$

with $(f_1, \dots, f_m)$ the canonical basis of $\mathbb{R}^m$ or $\mathbb{C}^n$ and Γ of dimension $m \times n$.

Stable and robust recovery. Through the research of conditions and class of matrices that would give us simple criteria to verify the ability of a measurement matrix to achieve uniform recovery among all vectors in Σ_s with the optimal number of measurements, we will always analyze the quality of the reconstruction process in terms of **stability and robustness** with an ℓ_p/ℓ_q**approximation guarantee** of the form

$$\left\| \hat{x} - x^{\#} \right\|_p \lesssim A \cdot \sigma_s(\hat{x})_q + B \cdot \varepsilon \tag{1}$$

where $\hat{x}$ denotes the true signal, $x^{\#}$ our approximation, $A, B > 0$ some constants that do not depend of $\hat{x}$ (but can depend of other variables such as s, p or q) and $\sigma_s(\hat{x})_q$ the smallest possible ℓ_q approximation error of $\hat{x}$ over all s-sparse vectors. This kind of guarantee is very useful as it gives us a *robustness property* thanks to the second term (with the level of noise) and *stability property* thanks to the first one, which allows us to consider signals that are not exactly s-sparse.

2 Classical tools in this framework

A natural recovery procedure to solve the problem introduced above would be the ℓ_0**-minimization**, i.e. finding the following set :

$$\operatorname{argmin}\{\|t\|_0 : \Gamma t = \Gamma \hat{x}\}$$

The ℓ_0-minimization procedure is optimal from a theoretical point of view : it allows $\hat{x}$'s identification for a minimal number of measurements ($m \geq 2s$), under a minimal hypothesis on Γ :

$\mathrm{Ker}(\Gamma) \cap \Sigma_{2s} = \{0\}$. But since solving the ℓ_0-minimization is a NP-hard problem, we need to use a more computationally-feasible-procedure. The ℓ_0-minimization is difficult to solve because the objective function is not convex.

Definition 1 (Basis Pursuit procedure) **:** The classic **Basis Pursuit procedure** constitutes a convex relaxation of ℓ_0-minimization :

$$\mathrm{argmin}\{\|t\|_1 : \Gamma t = \Gamma \hat{x}\} \tag{BP}$$

Instead of minimizing the ℓ_0 norm, we minimize its convex hull on B_∞^n, where B_∞^n designates the unit ball of $(\mathbb{R}^n, \ell_\infty^n)$: the ℓ_1 norm. The Basis Pursuit problem may be rewritten as a linear program, which means that from a computational point of view, it is solvable.
The authors of [3] extends this setting by introducing the ℓ_p**-constrained Basis Pursuit** as following

$$\min_{z \in \mathbb{C}^n} \|z\|_1 \quad \text{subject to} \quad \|\Gamma z_0 - \Gamma z\|_p \leq \varepsilon \tag{BPDN$_p$}$$

where the sparse signal now lives in $\mathbb{C}^n$ instead of $\mathbb{R}^n$, the level of noise $\Gamma z_0 - \Gamma z$ is not necessarily set to zero but smaller than a fixed level ε for the norm $\|\cdot\|_p$.

In order to have a well-posed problem, we are interested in the following property :

Definition 2 (Exact reconstruction property or order s) **:** $\Gamma \in \mathbb{R}^{m \times n}$ satisfies the **exact reconstruction property** of order s if for every $\hat{x} \in \Sigma_s$:

$$\mathrm{argmin}\{\|t\|_1 : \Gamma t = \Gamma \hat{x}\} = \{\hat{x}\} \tag{ER(s)}$$

To reconstruct a s-sparse vector $\hat{x}$ through the Basis-Pursuit using m measurements $\Gamma \hat{x}$, we must have $m \geq c_0 s \log(en/s)$. Conversely, it is possible to show that there are (random) matrices Γ which satisfy ER(s) with $m \sim s \log(en/s)$. This means that in comparison with the optimality condition we had for the ℓ_0-minimization, "$m \geq 2s$", we only lose a logarithmic factor, whereas the computational gain is huge.

We are now interested in the matrices which satisfies ER(s) with an optimal number of measurements. We introduce the Null Space Property of order s :

Definition 3 (Null Space Property of order s) **:** $\Gamma \in \mathbb{R}^{m \times n}$ satisfies the **Null Space Property of order** s if for every $J \subset \{1, \ldots, n\}$ such that $|J| = s$ and for all $v \in \mathrm{Ker}(\Gamma) - \{0\}$:

$$\|v_J\|_1 < \|v_{J^C}\|_1 \tag{NSP(s)}$$

The Null Space Property means that the vectors in $\mathrm{Ker}(\Gamma)$ must be the opposite of sparse, i.e. they must be "well-spread" : for such a vector v, less than s coordinates of v cannot concentrate more than half of v's mass. We have the following result : Γ satisfies RE(s) if and only if Γ satisfies NSP(s). In other words, for a matrix Γ, it is equivalent to check if Γ satisfies the exact construction property of order s and to check it Γ satisfies the null space property of order s. As for the RIP condition, the definition of $NSP(s)$ is extended in [3] for any norm $\|\cdot\|_q$ with $q \geq 1$. Γ satisfies the ℓ_q**-robust null space property of order** s with constants $0 < \rho < 1$ and $\tau > 0$ *with respect to a norm* $\|\cdot\|$ if for every $J \subset \{1, \ldots, n\}$ such that $|J| = s$ and for all $v \in \mathbb{C}^n$:

$$\|v_J\|_q \leq \frac{\rho}{s^{1-1/q}} \cdot \|v_{J^C}\|_1 + \tau \cdot \|\Gamma v\| \tag{ℓ_q NSP(s)}$$

If we set $q = 1$ and $v \in \mathrm{Ker}(\Gamma) - \{0\}$ the condition above is equivalent to

$$\|v_J\|_1 \leq \rho \cdot \|v_{J^C}\|_1 \tag{2}$$

which is a bit stronger than NSP(p) as it ensures that the supremum of $\|v_J\|_1/\|v_{J^C}\|_1$ for $v \in \mathrm{Ker}(\Gamma) - \{0\}$ such that $\|v_{J^C}\|_1 \neq 0$ is strictly less than 1 which is not guaranteed for the NSP(s).

We now focus on our initial problem again : if Γ that satisfies RE(s), the Basis Pursuit procedure allows to reconstruct $\hat{x}$ with an optimal number of measurements. It is equivalent to prove the exact reconstruction property of order s and the null space property of order s. However, it is difficult to check if a matrix satisfies the null space property of order s. Consequently, we may introduce sufficient (but not necessary) conditions for exact reconstruction ; in particular, the Restricted Isometry Property of order s is quite popular :

Definition 4 (Restricted Isometry Property of order s) **:** $\Gamma \in \mathbb{R}^{m\times n}$ satisfies the **Restricted Isometry Property** of order s with constant $\delta \in]0,1[$ if for every $t \in \Sigma_s$:

$$(1-\delta)\cdot\|t\|_2 \leq \|\Gamma t\|_2 \leq (1+\delta)\cdot\|t\|_2 \tag{RIP(s)}$$

The RIP(s) condition implies that Γ must behave like an isometry on the subset Σ_s. We have the following result : if Γ satisfies RIP($2s$) then Γ satisfies RE(s). Other forms of RIP(s) conditions exist and take the following forms :

$$c\cdot\|t\|_q \leq \|\Gamma t\|_p \leq C\cdot\|t\|_q \tag{RIP$_{p,q}$(s)}$$

where c and C are two positive real constants. Let's remark that if we do not impose some constraint on the form of c and C like in the classic RIP(s) condition where we had $(c+C)/2 = 1$, then the RIP$_{p,q}$(s) condition does not depend of the norm $\|\cdot\|_q$. In fact, by the equivalence of norms in finite-dimensional spaces, if Γ satisfies the RIP$_{p,q}$(s) condition then for any $q' \geq 1$, it exists α and β positive real constants such that

$$\forall t \in \mathbb{R}^n, \qquad \alpha\cdot\|t\|_{q'} \leq \|t\|_q \leq \beta\cdot\|t\|_{q'} \tag{3}$$

and by introducing new constants $c' = \alpha\cdot c$ and $C' = \beta\cdot C$, Γ also satisfies the RIP$_{p,q'}$(s) condition. In fact, it is common to have c and C such that

$$\begin{cases} c = \mu_{p,q}\,(1-\delta_s)^{1/q} \\ C = \mu_{p,q}\,(1+\delta_s)^{1/q} \end{cases} \tag{4}$$

with $\delta_s \in (\,0\,,\,1\,)$ and $\mu_{p,q} > 0$.

3 The problem stated in [6] and [3]

We have introduced all necessary elements to the comprehension of problem. The main idea of the authors of [6] and [3] is to show that $RIP_{p,q}$ as introduced above may not be the right tool to consider when it comes to analyze the performance of sparse recovery methods as

1. it enforces the number of rows m for the measurement matrix to be much larger than the optimal number of rows for uniform recovery of all s-sparse vectors,
2. only a small class of available measurement matrices verify this criteria.

We expose here these phenomena through known results on two very different class of matrices : standard gaussian matrices and adjacency matrices of random left d-regular bipartite graphs with n left vertices and m right vertices. What is very interesting with the second example, introduced in [6] and detailed in [2], is the fact that as opposed to gaussian matrices

1. we now consider $\Gamma_{i,j}$ in $\{0,1\}$ instead of having $\Gamma_{i,j}$ continuous on $\mathbb{R}$, *i.e.* Γ is *binary*.

2. we now deal with *sparse* matrices. To see this, let's considered Γ a standard gaussian matrix and $\varepsilon > 0$ a given threshold. By setting the coefficients in Γ to zero for those below zero, the average number of non-null coefficients in such new designed matrix denoted $\tilde{\Gamma}$ is:

$$\frac{1}{n \cdot m} \left\| \tilde{\Gamma} \right\|_F^2 = \frac{1}{n \cdot m} \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbb{1}_{|\Gamma_{i,j}| > \varepsilon} \tag{5}$$

where $\| \cdot \|_F$ denotes the Frobenius norm on $\mathbb{R}^{m \times n}$. As the coefficients $\Gamma_{i,j}$ are independent and identically distributed, by the law of large number, this fraction converges almost surely to

$$\mathbb{E}\left[\mathbb{1}_{|\Gamma_{i,j}| > \varepsilon} \right] = \mathbb{P}\left(\left| \Gamma_{i,j} \right| > \varepsilon \right) = 1 - 2 \int_0^{\varepsilon} \underbrace{\frac{1}{\sqrt{2\pi}} \mathbf{e}^{\frac{-x^2}{2}}}_{:=\varphi(x)} \mathrm{d}x. \tag{6}$$

Derivating φ gives $\varphi'(x) = \frac{-x}{\sqrt{2\pi}} \mathbf{e}^{\frac{-x^2}{2}}$, which yields to

$$\forall x \in [0\,;\varepsilon], \qquad |\varphi(x) - \varphi(0)| \leq \frac{\varepsilon}{\sqrt{2\pi}} \qquad \text{so} \qquad \int_0^{\varepsilon} \varphi(x)\,\mathrm{d}x \leq \frac{\varepsilon}{\sqrt{2\pi}} (1+\varepsilon). \tag{7}$$

For instance, setting $\varepsilon = 0.1$ gives a probability greater than 91% for entries in $\tilde{\Gamma}$ not being null. For Γ being the random adjacency matrix mentioned above, the number of non-null coefficients is straightforward and equals to $(d \cdot n)/(m \cdot n) = d/m$. The degree d here denotes the level of sparsity in the measurement matrix which is assumed to be much smaller than m.

3.1 Adjacency matrices

We give here an interesting result of [1] showing the dependency between the requirements in terms of rows of Γ to satisfy the $RIP_{p,q}$ condition with $q = p$.

Definition 5 (Sparsity of a binary matrix) **:** We say that an $m \times n$ matrix Γ is a random binary matrix with sparsity $d \in [m]$, if Γ is generated by assigning $d^{-1/p}$ to d random entries per column (selected uniformly at random without replacement), and assigning 0 to the remaining entries.

The connection with graph bipartite graph is straightforward. For a given random graph $G = (U, V, E)$ with $|U| = n$, $|V| = m$ and $E \subseteq U \times V$ such that all vertices u_i from U have the same degree $d := \mathrm{Card}\left\{ v_j \in V : (u_i, v_j) \in E \right\}$, if we denote by $\Gamma = \left(\mathbb{1}_{(u_i, v_j) \in E} \right)_{j,i} \in \{0,1\}^{m \times n}$ its associated adjacency matrix, then Γ is a random binary matrix with sparsity d.

Theorem. *If $\varepsilon \in (0, 1/2)$, $p \geq 2$ and Γ is a random binary matrix with sparsity such that*

$$\begin{cases} m = p^{\alpha_p} \cdot \frac{s^p}{\varepsilon^2} \cdot \log^{p-1} n \\ d = p^{\alpha_p} \cdot \frac{s^{p-1}}{\varepsilon} \cdot \log^{p-1} n \leq m \end{cases} \tag{8}$$

where $\alpha_p = O(p)$, then Γ satisfies the $RIP_{p,p}$ with constants $\mu_{p,p} = 1$ and $\delta_s = \varepsilon$ in (8).

Let's make some remarks:

1. The case $p = 2$ which is allowed by hypothesis give $m \sim s^2 \log(n)$ which is optimal on the class of binary matrices as $m \geq s^2 \log(en/s)$ is necessary in order to satisfy the $\mathrm{RIP}_{2,2}$ condition ([1]).

2. For $p > 2$, one may need substantially more than $s\log(en/s)$ measurements for uniform recovery of all s-sparse vectors. Is this due to a real phenomena that occurs considering a different norm than ℓ_2 to measure the level of noise, or is only due to an ill $\mathrm{RIP}_{p,q}$ criteria ? Authors of [3] answer to this question for a special class of random (continuous) matrices.

3.2 Gaussian matrices

For Γ is standard Gaussian, we know from [4] that if m satisfies

$$m \gtrsim \left(\delta^{-2} s\log(en/(s\delta)) + \delta^{-2}\log\left(\eta^{-1}\right)\right)^{p/2} + (p-1)2^{p-1} \tag{9}$$

where p still denotes the ℓ_p norm used in the Basis Pursuit then the $\mathrm{RIP}_{p,2}$ condition is satisfied for the measurement matrix $\mu_p^{-1}\Gamma$ with probability greater than $1-\eta$ where $\mu_p = \mathbb{E}[\,\|G\|\,]_p$ and G is a standard m-dimensional Gaussian random vector, with constant $\mu_{p,2} = 1$ in (8). Moreover, if Γ satisfies the latter condition for sparsity levels $s, 2s, 3s$ then we have an ℓ_2/ℓ_1 guarantee for the BPDN_p procedure with $A = s^{-1/2}$ and $B = \mu_p^{-1}$ in (1).

If we assume that $s\log(en/(s\delta)) >> \log\left(\eta^{-1}\right)$ and neglecting the last term $(p-1)2^{p-1}$ which only depends of fixed $p \geq 1$, the condition on rows simplify to $m \gtrsim (s\log(en/(s\delta)))^{p/2}$ which is sub-optimal.

For $p \geq 2$, the authors of [3] shows that Γ can achieve an optimal reconstruction with only

$$m \gtrsim s\log(en/(s\delta)) + \log\left(\eta^{-1}\right) \tag{10}$$

and an ℓ_2/ℓ_1 guarantee for the BPDN_p procedure with $A = s^{-1/2}$ and $B = m^{-1/p}$ in (1) where μ_p has been replaced by the number of rows. Condition (9) is clearly stronger than condition (10) when $p > 2$.

Conclusion. If $\mathrm{RIP}_{p,q}$ condition seems to differentiate the case $p = 2$ from $p > 2$ in terms of number of rows required, we can see from the above result that this behavior does not reflect the real optimality of standard gaussian measurement matrices for any chosen $p \geq 2$ for the ℓ_p-constrained Basis Pursuit procedure.

4 Extension to a larger class of matrices

The Gaussian behavior for the measurement matrix seen in the previous section is, in fact, a strong hypothesis, and random matrices with weaker hypothesis may still verify the reconstruction property from above.

Theorem. *If $X_1, \ldots, X_m$ are*

1. *independent ;*
2. *sub-isotropic, i.e. for every $t \in \mathbb{C}^n$, $\mathbb{E}\left[\,\langle X, t\rangle\,\right] \geq \|t\|_2^2$;*
3. *L-subgaussian, i.e. for every $t \in \mathbb{C}^n$ and $p \geq 1$, $\|\langle X, t\rangle\|_{L_p} \leq L\sqrt{p}\|\langle X, t\rangle\|_{L_2}$;*

the measurement matrix Γ with rows $X_1, \ldots, X_m$ and $m \gtrsim s^{2-2/q}\log(en/s) + \log\left(\eta^{-1}\right)$ satisfies with a probability exceeding $1-\eta$ the following ℓ_2/ℓ_1 guarantee

$$\left\|\hat{x} - x^{\#}\right\|_2 \lesssim s^{-1/2}\sigma_s(\hat{x})_1 + m^{-1/p}\varepsilon \tag{11}$$

for the reconstruction via ℓ_p-constrained basis pursuit.

A random vector X is told to have a L-subgaussian behaviour if it concentrates around its mean at a "$\exp(-t^2)$" speed. If we suppose X symmetric, it means that there are some constants c_0, c_1 so that for every $t \geq c_0$, $\mathbb{P}(|X| \geq t) \leq 2\exp -c_1 t^2$. To prove that Γ, with the latter hypothesis, satisfies RE(s) we prove that Γ satisfies RIP($2s$), using the rapid decay of the functionals $\langle X, t\rangle$. The question which naturally occurs is : if the decay of the functionals $\langle X, t\rangle$ is slower, may we still prove that Γ satisfies RIP with the same optimal number of measurements ? Here, we consider sub-exponential variables, i.e. variables which concentrates around their mean at a "$\exp(-t)$" speed :

Definition 6 (Sub-exponential variable) **:** a variable X is told **L-sub-exponential** if for every $t \in \mathbb{R}^n$ and every $p \geq 2$:

$$\|\langle X, t\rangle\|_{L_p} \leq Lp\|\langle X, t\rangle\|_{L_2}$$

For isotropic sub-exponential variables, the RIP condition is more difficult to prove and in addition, it may be satisfied with high probability only when $m > c_2(L)s\log^2(en/s)$, bound which cannot be improved : the behaviour of RIP is sub-optimal. However, [5] proved that for such variables, exact reconstruction can still be achieved with the optimal number of measurements. The question which naturally results is : may much weaker hypothesis on the measurement matrice still permit to achieve exact recovery, when RIP is not satisfied ?

In this perspective, we introduce the small-ball condition.

4.1 The small-ball condition

Definition 7 (Small-ball condition) **:** a random vector X satisfies the small-ball condition in Σ_s with constants $u, \beta > 0$ if for every $t \in \Sigma_s$:

$$P(|\langle X, t\rangle| > u\|t\|_2) \geq \beta$$

The small-ball condition is not a concentration result. We know that if a random variable has enough well-behaved moments, then its empirical mean will concentrate around its true mean. But this moment condition is restrictive, whereas only a very weak moment assumption guarantees a small-ball condition. For heavy-tailed variables whose moments we cannot hope to control, nor apply concentration results, we may still prove that such variables satisfy a small-ball condition. Then, the small-ball condition is rather a minimal hypothesis.

In order to prove that a variable satisfies a small-ball condition, one often shows that it satisfies a Paley-Zygmund inequality :

Definition 8 (Paley-Zygmund inequality) **:** any nonnegative random vector X with a finite variance satisfies the Paley-Zygmund inequality i.e., for every $\theta \in (0, 1)$:

$$\mathbb{P}\left(X \geq \theta \cdot \mathbb{E}[X]\right) \geq (1-\theta)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}$$

We note however that the small-ball condition does not even require that X admits a covariance matrix.

[6] prove that a small-ball condition and a weak moment assumption suffices to achieve exact recovery with the optimal number of measurements. Thus, the small-ball condition combined with a weak moment assumption constitute much weaker requirements that the RIP condition. We now state that result :

4. EXTENSION TO A LARGER CLASS OF MATRICES

Theorem. *(A) There are absolute constants c_0, c_1, c_2 and, for every $\alpha \geq 1/2$, there is a constant $c_3(\alpha)$ for which the following holds.*
Let $X = (x_i)_{i=1}^n$ a random vector on $\mathbb{R}^n$ (with potentially dependent coordinates). We suppose that :

1. *there are $\kappa_1, \kappa_2, \omega > 1$ that satisfy for every $1 \leq j \leq n$: $\|x_j\|_{L_2} = 1$; and for every $4 \leq p \leq 2\kappa_2 \log(\omega n)$: $\|x_i\|_{L_p} \leq \kappa_1 p^\alpha$.*

2. *X satisfies the small ball condition in Σ_s with constants u and β.*

If :

$$m \geq c_0 \max\left(s \log\left(\frac{en}{s}\right), (c_3(\alpha)\kappa_1^2)^2 (\kappa_2 \log(\omega n))^{\max(4\alpha-1,1)}\right)$$

and $X_1, \ldots, X_m$ are independent copies of X, then, with probability at least :

$$1 - 2\exp(-c_1\beta^2 m) - 1/\omega^{\kappa_2} n^{\kappa_2 - 1}$$

Γ the measurements matrix defined as above satisfies the exact reconstruction property in Σ_{s_1} with $s_1 = c_2 u^2 \beta s$.

This theorem implies that if x is a centered random variable with variance 1, such that $\|x\|_{L_p} \leq c\sqrt{p}$ for $1 \leq p \leq 2\log(n)$, if X's coordinates are independent copies of x then the corresponding measurement matrix Γ with $m \leq c_1 s \log(en/s)$ permits to recover any s-sparse vector with high probability.

As we previously saw (and since it is the point of the whole article), we cannot use a RIP-based-argument to prove Theorem A. We state a second theorem which allows Theorem A's proof :

Theorem. *(B) Let $\Gamma : \mathbb{R}^n \to \mathbb{R}^m$ and $(e_1, \ldots, e_n)$ the canonical basis of $\mathbb{R}^n$. We suppose that :*

1. *for every $x \in \Sigma_s$, $\|\Gamma x\|_2 \geq c_0 \|x\|_2$*

2. *for every $j \in \{1, \ldots, n\}$, $\|\Gamma e_j\|_2 \leq c_1$*

With $s_1 = \left\lfloor c_0^2 (s-1)/(4c_1^2) \right\rfloor - 1$, Γ satisfies the exact reconstruction property in Σ_{s_1}.

Theorem B's conditions are weaker to the RIP conditions. We indeed need to check the right-side of

$$(1-\delta)\|t\|_2 \leq \|\Gamma t\|_2 \leq (1+\delta)\|t\|_2$$

for *only* 1-sparse vector and not *every* s-sparse vector : here lies the fundamental gap between RIP and the exact reconstruction property. Indeed, if the lower bound in the latter inequality is satisfied for rather general ensembles, the upper bound amounts to require that X's coordinates exhibit a subgaussian behaviour of enough moments, which is precisely the hypothesis we wish to release here. Consequently, if a small-ball condition suffices to provide the lower bound, the upper bound constitutes the most demanding part of the RIP property.

4.2 Exact Reconstruction still has a cost, i.e. a moment condition – the price of convex relaxation

However, checking the lower bound of RIP(s) still has a cost, which is the first condition of Theorem A, a moment condition. Without that hypothesis, we cannot achieve Exact Reconstruction.

Definition 9 (Generated matrix) **:** a random matrix Γ is generated by the random variable x if $\Gamma = \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\langle X_i, \cdot\rangle f_i$ and $X_1, \dots, X_m$ are independent copies of $X = (x_1, \dots, x_n)^T$ whose coordinates are independent copies of x.

We have the following result, which is denoted Theorem C' in [6].

Theorem (C'). *There are absolute constants c_0, c_1, c_2 and κ for which the following holds.*
If $n \geq c_0$ and $2 < p < c_1 \log(n)$, there exists a centered random variable x with variance 1, for which $\|x\|_{L_q} \leq \kappa\sqrt{q}$ for $2 < q \leq p$, such that :
if $m \leq c_2\sqrt{p}(n/\log(n))^{1/p}$ and Γ is the $m \times n$ matrix generated by x then with probability at least $1/2$, Γ does not satisfy the exact reconstruction property of order 1.

Such a variable x as stated in Theorem C' cannot satisfy the first hypothesis of Theorem A. Theorem C' implies that Basis Pursuit may not return the true vector when the coordinates of X do not have enough moments ; Exact Reconstruction then requires a polynomial number of measurements in n. This may be seen as the price of convex relaxation : indeed, [6] prove that under an even weaker assumption that the small-ball condition, the $\ell_0$minimisation procedure still achieves Exact Recovery with the optimal number of measurements. First, we introduce this even-weaker property :

Definition 10 (Weak small ball-condition) **:** X satisfies a weak small-ball condition in Σ_s with constant β if for every $t \in \Sigma_s$:

$$P(|\langle X, t\rangle| > 0) \geq \beta$$

We now state the following (striking) result :

Theorem (D). *For every $0 < \beta < 1$, there are constants c_0 and c_1, that depends only on β, for which the following holds.*
Let X be a random vector that satisfies the weak small-ball condition in Σ_s with a constant β.
Let $X_1, \dots, X_m$ be m independent copies of X and Γ the corresponding measurement matrix.
If $m \geq c_0 s\log(en/s)$ then with probability at least $1 - 2\exp(-c_1 m)$, for every $\hat{x} \in \Sigma_{\lfloor s/2\rfloor}$, ℓ_0-minimisation has a unique solution which is $\hat{x}$ itself.

We see what convex relaxation costs relatively to the number of measurements required for exact reconstruction. $m \geq c_0 s\log(en/s)$ must be compared with $m \geq c_2\sqrt{p}(n/\log(n))^{1/p}$: if X satisfies the hypothesis of Theorem C' for $p = 4$, then it also satisfies the hypothesis of Theorem D : the ℓ_0-minimisation permits then to reconstruct s-sparse vectors with only $c_0 s\log(en/s)$ measurements, whereas Basis Pursuit demands more than $(n/\log(n))^{1/4}$ measurements to reconstruct 1-sparse vectors.

4.3 Conclusion

In this article, we have seen that some measurement ensembles whose behaviour was not gaussian nor subgaussian but "heavy-tailed" did not satisfy the RIP property, but still achieves Exact Recovery under much weaker assumptions : the small-ball condition and a compulsory moment condition, which is the price we have to pay for the convex relaxation of the ℓ_0-minimisation procedure.

Resources

[1] Zeyuan Allen-Zhu, Rati Gelashvili, and Ilya Razenshteyn. "Restricted Isometry Property for General p-Norms". In: (July 8, 2014). arXiv: http://arxiv.org/abs/1407.2178v3 [cs.DS].

[2] R. Berinde et al. "Combining geometry and combinatorics: A unified approach to sparse signal recovery". In: (Apr. 29, 2008). arXiv: http://arxiv.org/abs/0804.4666v1 [cs.DM].

[3] Sjoerd Dirksen, Guillaume Lecué, and Holger Rauhut. "On the gap between RIP-properties and sparse recovery conditions". In: (Apr. 20, 2015). arXiv: http://arxiv.org/abs/1504.05073v1 [cs.IT].

[4] Laurent Jacques, David K. Hammond, and M. Jalal Fadili. "Dequantizing Compressed Sensing: When Oversampling and Non-Gaussian Constraints Combine". In: (Feb. 13, 2009). arXiv: http://arxiv.org/abs/0902.2367v4 [math.OC].

[5] Vladimir Koltchinskii. *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: Ecole d'Eté de Probabilités de Saint-Flour XXXVIII-2008*. Vol. 2033. Springer Science & Business Media, 2011.

[6] Guillaume Lecué and Shahar Mendelson. "Sparse recovery under weak moment assumptions". In: (Jan. 9, 2014). arXiv: http://arxiv.org/abs/1401.2188v5 [math.ST].