Initiation to Research Report
Compressed Sensing

Anastasiia Storozhenko
supervised by Guillaume Garrigos

June 20, 2025

1 Introduction

The field of *comressed sensing* (often also called compressive sensing) treats the problem of solving an underdetermined linear system under the assumption of sparsity. It emerged from signal processing, where such problems often occur; many signals are sparse at least in some basis, e.g. Fourier or wavelet. Tools developed in compressed sensing allow to take a smaller amount of linear measurements (linear combinations of signal components) than would normally be required to restore the signal. These tools can also be used to *compress* any sparse data for efficient storage or transmission with guarantees of successful restoration. Another field of application is medical imaging; with the help of compressed sensing it is possible to significally shorten the time needed to take the image.

The list of applications goes on, but our focus will be on the mathematical side of things. During this project I worked on three main sources: the book "A mathematical Introduction to Compressive Sensing" by S. Foucart and H. Rauhut [5], the paper "The Convex Geometry of Linear Inverse Problems" by V. Chandrasekaran et al. [4], and the paper "Living on the Edge: Phase Transitions in Convex Programs with Random Data" by D. Amelunxen et al. [1]. We will first cover some general results from the book in sections 2 and 3 and then switch to the study of the two papers in section 4. We will look for connections between those papers and try to reproduce some results, while filling any gaps we find.

2 Studying the l_0-minimization

We want to recover an s-sparse vector $\mathbf{x} \in \mathbb{R}^N$ knowing a vector of m measurements $\mathbf{y} \in \mathbb{R}^m$ and a measurement matrix $\mathbf{A} \in M_{m\times N}(\mathbb{R})$ with $m < N$, such that $\mathbf{Ax} = \mathbf{y}$. The system is underdetermined, so we have to look for more creative ways to solve it. One such way is to solve the corresponding l_0-minimization problem. We start by giving basic definitions as seen in [5].

Definition 2.1. The *support* of a vector $\mathbf{x} \in \mathbb{R}^N$ is the set of indices of its nonzero entries:

$$\operatorname{supp}(\mathbf{x}) = \{j \in [\![1, N]\!] \colon x_j \neq 0\}$$

Definition 2.2. We define $\|\mathbf{x}\|_0$ as the cardinality of $\operatorname{supp}(\mathbf{x})$. We say that the vector $\mathbf{x}$ is *s-sparse* if $\|\mathbf{x}\|_0 \leq s$.

Note that $\|\cdot\|_0$ is not an actual norm, nor is it a semi-norm. Now we can formalize the problem in the following form:

$$\text{minimize } \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{y}. \tag{P_0}$$

Even though it is easy to prove, there is no reason for the problem P_0 to be automatically equivalent to what we started this. To highlight this, we write it down as a small proposition below.

Proposition 2.3. *Let $\mathbf{A} \in M_{m\times N}(\mathbb{R})$, $\mathbf{x} \in \mathbb{R}^N$ s-sparse and $\mathbf{y} \in \mathbb{R}^m$. The following two statements are equivalent:*

(i) The vector $\mathbf{x}$ is the unique solution of the compressed sensing problem, i.e, it's the unique s-sparse vector such that $\mathbf{A}\mathbf{x} = \mathbf{y}$.

(ii) The vector $\mathbf{x}$ is the unique solution of P_0.

Proof. $(i) \Rightarrow (ii)$ If $\mathbf{x}$ is the only s-sparse vector that satisfies $\mathbf{A}\mathbf{x} = \mathbf{y}$, then there exists no such vector $\mathbf{z}$, that $\|\mathbf{z}\|_0 \leq \|\mathbf{x}\|_0 \leq s$, which makes $\mathbf{x}$ the unique minimizer of P_0.

$(ii) \Rightarrow (i)$ Immediate. □

One of the biggest questions in this field (and the biggest in this report) is how many measurements m we need for a successful recovery of the unknown vector. In the following theorem we denote by $\mathbf{A}_S$ the matrix consisting of the columns of $\mathbf{A}$ indexed by S and by $\mathbf{x}_S$ – the vector consisting of the entries of $\mathbf{x}$ indexed by S.

Theorem 2.4. *Let $\mathbf{A} \in M_{m\times N}(\mathbb{R})$ and $\mathbf{x}, \mathbf{z} \in \mathbb{R}^N$. The following statements are equivalent:*

(i) If $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{z}$ and both $\mathbf{x}$ and $\mathbf{z}$ are s-sparse, then $\mathbf{x} = \mathbf{z}$.

(ii) $\operatorname{Ker}\mathbf{A} \cap \{\mathbf{z} \in \mathbb{R}^N \colon \|\mathbf{z}\|_0 \leq 2s\} = \{\mathbf{0}\}$, i.e., $\mathbf{0}$ is the only $2s$-sparse vector in the $\operatorname{Ker}\mathbf{A}$.

(iii) For every $S \subset [\![1, N]\!]$ with $\operatorname{card}(S) \leq 2s$, the submatrix $\mathbf{A}_S$ is injective as a map from $\mathbb{R}^{\operatorname{card}(S)}$ to $\mathbb{R}^m$.

(iv) Every set of $2s$ columns of $\mathbf{A}$ is linearly independent, i.e., $\operatorname{rang}\mathbf{A} \geq 2s$.

Proof. $(i) \Rightarrow (ii)$ Let $\mathbf{z} \in \mathbb{R}^N$ be $2s$-sparse and satisfy $\mathbf{A}\mathbf{z} = \mathbf{0}$. On the other hand, we also have $\mathbf{A}\mathbf{0} = \mathbf{0}$, so by the hypothesis it has to be that $\mathbf{z} = \mathbf{0}$.

$(ii) \Rightarrow (i)$ Now let $\mathbf{x}$ and $\mathbf{z}$ be two s-sparse vectors such that $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{z}$. Then $\mathbf{x} - \mathbf{z}$ is $2s$-sparse and $\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{z} = \mathbf{A}(\mathbf{x} - \mathbf{z}) = \mathbf{0}$, which implies that $\mathbf{x} - \mathbf{z} \in \operatorname{Ker}\mathbf{A}$. By hypothesis, we conclude that $x = z$.

$(ii) \Rightarrow (iii)$ We recall that linear map $\mathbf{A}$ is injective if and only if $\operatorname{Ker}\mathbf{A} = \{\mathbf{0}\}$. Let $\in \mathbb{R}^N$ be a $2s$-sparse vector, such that $\mathbf{x}_S \in \operatorname{Ker}\mathbf{A}_S$, where $S = \operatorname{supp}(\mathbf{x})$. Then $\mathbf{A}\mathbf{x} = \mathbf{A}_S\mathbf{x}_S + \mathbf{A}_{\overline{S}}\mathbf{x}_{\overline{S}} = \mathbf{0} + \mathbf{A}_{\overline{S}}\mathbf{0} = \mathbf{0}$, where $\overline{S} = [\![1, N]\!] \setminus S$. Then by hypothesis, $\mathbf{x} = 0$ and $\mathbf{x}_S = 0$, and thus $\mathbf{A}_S$ is injective.

$(iii) \Rightarrow (ii)$ Let $\mathbf{x}$ be $2s$-sparse, $S = \operatorname{supp}(\mathbf{x})$ and $\mathbf{A}_S$ an injective map. Suppose that $\mathbf{x} \in \operatorname{Ker}\mathbf{A}$. Then by extension, $\mathbf{x}_S \in \operatorname{Ker}\mathbf{A}_S$ and $\mathbf{x}_S = \mathbf{0}$. Thus, $\mathbf{x} = 0$.

For the last two implications we have to assume that $2s \leq m$.

$(iii) \Rightarrow (iv)$ Let $S \subset [\![1, N]\!]$, $\operatorname{card}(S) = 2s$. Then $\operatorname{rang}(A_S) = 2s - \dim(\operatorname{Ker}\mathbf{A}_S)) = 2s - 0 = 2s$.

$(iv) \Rightarrow (iii)$ Let $S \subset [\![1, N]\!]$, $\operatorname{card}(S) \leq 2s$. Then $\operatorname{rang}(A_S) = \operatorname{card}(S)$ and $\dim(\operatorname{Ker}\mathbf{A}_S)) = \operatorname{card}(S) - \operatorname{rang}(A_S) = 0$. Thus, $\operatorname{Ker}\mathbf{A}_S = \{\mathbf{0}\}$. □

The importance of this theorem is that it gives us the necessary condition for a successful recovery of all s-sparse vectors $\mathbf{x}$ from P_0: the number of measurements m has to be at least $2s$. Indeed, if it is possible to reconstruct the vector by solving l_0-problem, then statement (i) holds, and then according to the theorem, $\operatorname{rang} \mathbf{A} \geq 2s$. On the other hand, the rank of a matrix cannot be greater than its smallest dimension, which in this case is m. This gives us the necessary condition $m \geq 2s$.

Remark 2.5. However, in some cases we can successfully reconstruct the vector with fewer measurements m. For example, if it is possible to construct a matrix $\mathbf{A}$ that depends on the specific vector $\mathbf{x}$ we want to recover, then the necessary condition becomes $m = s + 1$ instead (see Theorem 2.16 in [5]). Alternatively, if we are ready to sacrifice the absolute certainty of the potential recovery, then we can turn to random matrices, where it's possible to get recovery guarantees in probabilistic form (we will discuss this in more details later).

Unfortunately, as good as it seems in theory, the l_0-minimization is not effective in practice: problem P_0 happens to be NP-hard (see, for example, section 2.3 in [5]). The same can be said for any l_p-minimization problem with $p < 1$, which means that we have to look for convex alternatives instead, for example, l_1-minimization.

3 From the l_0 to the l_1 minimization

In the context of this work, when we speak of l_1-minimization, we mean the following convex optimization problem (also known as basis pursuit):

$$\text{minimize } \|\mathbf{x}\|_1 \quad \text{subject to } \quad \mathbf{A}\mathbf{x} = \mathbf{y}. \tag{P_1}$$

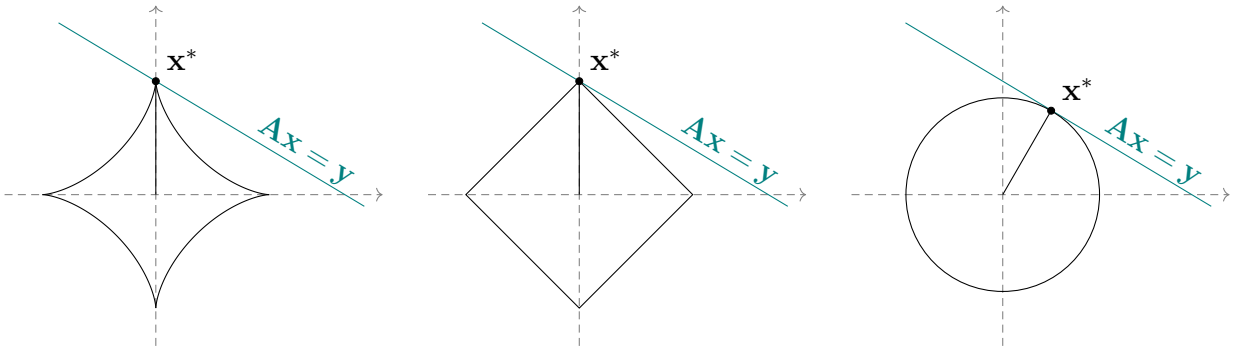
Going from $\|\cdot\|_0$ to $\|\cdot\|_p$ is quite intuitive, as $\|\mathbf{x}\|_p^p \xrightarrow[p\to 0]{} \|\mathbf{x}\|_0$, but why do we specifically want to work with l_1-minimization? There are three cases here: $p < 1$, $p = 1$ and $p > 1$. For $p < 1$, the biggest problem is that the corresponding problem is NP-hard, which makes it not very useful in practice (however, it is still used to build some theory around compressive sensing). With $p > 1$ the problem becomes convex, however, a bigger issue arises: in most cases the solution of the corresponding minimization problem won't be sparse. It is easy to see why from a simple visualisation in Fig. 1 in dimension 2, where we show unit balls under different norms. We see that it would only work if the line $\mathbf{A}\mathbf{x} = \mathbf{y}$ was parallel to one of the axes, in which case the solution would be indeed sparse. In any other case the solution would be a non-sparse vector that is closer to the origin under this norm. Which leaves us with $p = 1$; for this value of p we have neither of those problems. Moreover, it is a very well studied problem in convex optimization and many effective algorithms exist for solving it.

Now a new question arises: under which conditions does the minimizer of P_1 solve P_0? To answer it we introduce the notion of null space property.

Definition 3.1. A matrix $\mathbf{A} \in M_{m\times N}(\mathbb{R})$ is said to satisfy the *null space property* relative to a set $S \subset [\![1, N]\!]$ if

$$\|\mathbf{v}_S\|_1 < \|\mathbf{v}_{\overline{S}}\|_1, \quad \forall \mathbf{v} \in \operatorname{Ker} \mathbf{A} \backslash \{\mathbf{0}\}.$$

It is said to satisfy the *null space property of order* s if it satisfies the null space property relative to any $S \subset [\![1, N]\!]$ with $\operatorname{card}(S) \leq s$.



$p < 1$ $p = 1$ $p > 1$

Figure 1: Unit balls under l_p norm for different values of p. Here $\mathbf{x}^*$ denotes the solution of the corresponding minimization problem under constraints $\mathbf{Ax} = \mathbf{y}$.

Theorem 3.2. *Let $\mathbf{A} \in M_{m\times N}(\mathbb{R})$. A vector $\mathbf{z} \in \mathbb{R}^N$ supported on a set S is the unique solution of P_1 with $\mathbf{y} = \mathbf{Az}$ if and only if $\mathbf{A}$ satisfies the null space property relative to S.*

Proof. ($\Rightarrow$) Let $\mathbf{v} \in \operatorname{Ker}\mathbf{A}\backslash\{\mathbf{0}\}$. Then $\mathbf{Av} = \mathbf{A}(\mathbf{v}_S + \mathbf{v}_{\overline{S}}) = 0$ and $\mathbf{Av}_S = -\mathbf{Av}_{\overline{S}}$. Vector $\mathbf{v}_S$ is supported on S, so by the hypothesis, $\mathbf{v}_S$ is the unique solution of P_1 with $\mathbf{y} = \mathbf{Av}_S$. However, $-\mathbf{v}_{\overline{S}}$ is another solution of the equation $\mathbf{Ax} = \mathbf{Av}_S$, so it has to be that $\|\mathbf{v}_{\overline{S}}\|_1 > \|\mathbf{v}_S\|_1$.

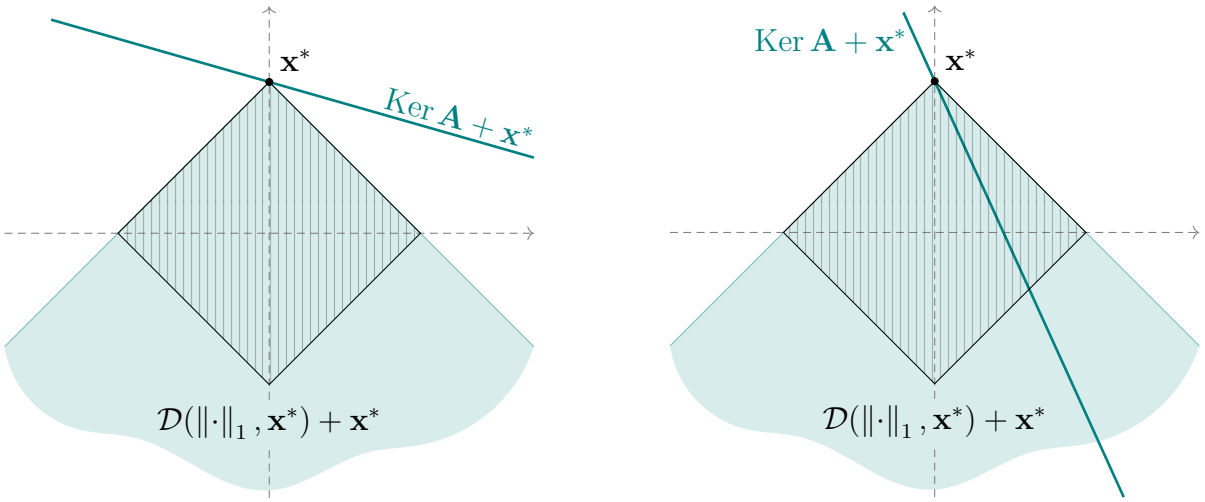
($\Leftarrow$) Now suppose that $\forall \mathbf{v} \in \operatorname{Ker}\mathbf{A}\backslash\{\mathbf{0}\}$, $\|\mathbf{v}_S\|_1 < \|\mathbf{v}_{\overline{S}}\|_1$. Let $\mathbf{z} \in \mathbb{R}^N$ be supported on S . We want to show that it is the unique solution of P_1 with $\mathbf{y} = \mathbf{Az}$. Suppose that there exists a vector $\mathbf{z}^* \neq \mathbf{z}$ such that $\mathbf{z}^*$ minimizes $\|\cdot\|_1$ with $\mathbf{Az}^* = \mathbf{Az}$. Then $\mathbf{A}(\mathbf{z}-\mathbf{z}^*) = 0$ and $\mathbf{z} - \mathbf{z}^* \in \operatorname{Ker}\mathbf{A}\backslash\{\mathbf{0}\}$. According to the null space property, $\|(\mathbf{z} - \mathbf{z}^*)_S\|_1 < \|(\mathbf{z} - \mathbf{z}^*)_{\overline{S}}\|_1$. Combining this with the fact that $\mathbf{z}_S = \mathbf{z}$ and $\mathbf{z}_{\overline{S}} = \mathbf{0}$, we get $\|\mathbf{z} - \mathbf{z}^*_S\|_1 < \left\|\mathbf{z}^*_{\overline{S}}\right\|$. Then we obtain $\|\mathbf{z}\|_1 \leq \|\mathbf{z} - \mathbf{z}^*_S\|_1 + \|\mathbf{z}^*_S\| < \left\|\mathbf{z}^*_{\overline{S}}\right\|_1 + \|\mathbf{z}^*_S\|_1 = \|\mathbf{z}^*\|_1$, which contradicts the assumption. □

Theorem 3.3. *Let $\mathbf{A} \in M_{m\times N}(\mathbb{R})$. An s-sparse vector $\mathbf{z} \in \mathbb{R}^N$ is the unique solution of P_1 with $\mathbf{y} = \mathbf{Az}$ if and only if $\mathbf{A}$ satisfies the null space property of order s.*

Proof. Immediate from Theorem 3.2. □

Remark 3.4. Note that this result gives us conditions on when the solution $\mathbf{z}$ of P_1 is also the solution of P_0 with $\mathbf{y} = \mathbf{Az}$.

Despite this result being fundamental in the theoretical study of compressed sensing, it is not easy to verify if a matrix satisfies this property in practice. New tools were developed over the years to study conditions for the solution of the problem P_1 to also be the solution of P_0. Probably the most important one is the *restricted isometry property* (see, for example, [2]). We won't be discussing this topic in details, just state the fact we will be interested in the most: the corresponding condition for equivalence between P_1 and P_0 is satisfied by certain types of random matrices with high probability. In the following section we will focus on two specific results for matrices with independent normally distributed entries, that give us the conditions for finding the unique solution of P_1.



(a) (b)

Figure 2: Colored area shows the descent cone to the l_1 norm in $\mathbf{x}^*$; dashed area is the set of vectors for which the norm is decreasing compared to $\mathbf{x}^*$.

4 Minimal number of measurements with random matrices

From now on we will consider only random matrices with $\mathcal{N}(0,1)$-distributed entries. It might seem unnatural at first if we think of compressive sensing only in the context of natural phenomena, where we have no or little control over how vector $\mathbf{y}$ is obtained. However, we could also look, for example, at data compression, where we are free to choose the matrix $\mathbf{A}$ however we like. Considering the vast number of results for normally distributed matrices and normal distribution in general, they become convenient candidates for the task.

Before proceeding to the recovery conditions for random matrices, we first have to recall some definitions from convex geometry, relying mainly on [7] as source material.

Definition 4.1. A convex set $\mathcal{C} \subset \mathbb{R}^N$ is called a *cone* if it is closed under non-negative scalar multiplications, i.e. $\forall \alpha \geq 0, \mathbf{x} \in \mathcal{C} : \alpha\mathbf{x} \in \mathcal{C}$.

The cone $\mathcal{C}^* = \{\mathbf{x} \in \mathbb{R}^N : \langle \mathbf{x}, \mathbf{y} \rangle \leq 0,\ \forall \mathbf{y} \in \mathcal{C}\}$ is called the *polar* of $\mathcal{C}$.

Definition 4.2. Let $C \subset \mathbb{R}^N$ be a convex set and $\mathbf{x} \in \mathbb{R}^N$. We call a *tangent cone* of C the cone $T_C(\mathbf{x}) = \mathrm{cl}\{\alpha\mathbf{x} : \mathbf{x} \in C, \alpha \geq 0\}$. We call a *normal cone* of C the cone $N_C(\mathbf{x}) = T_C^*(\mathbf{x})$.

Definition 4.3. The *descent cone* $\mathcal{D}(f, \mathbf{x})$ of a proper convex function $f : \mathbb{R}^N \to \overline{\mathbb{R}}$ at $\mathbf{x} \in \mathbb{R}^N$ is defined as

$$\mathcal{D}(f, \mathbf{x}) = \bigcup_{\tau > 0} \left\{ \mathbf{z} \in \mathbb{R}^N : f(\mathbf{x} + \tau\mathbf{z}) \leq f(\mathbf{x}) \right\}.$$

Remark 4.4. Descent cones are closely related to sublevel sets; in fact, $T_C(\mathbf{x}) = \mathrm{cl}\,\mathcal{D}(f, \mathbf{x})$ for $C = \{\mathbf{z} : f(\mathbf{z}) \leq f(\mathbf{x})\}$. In [4], authors formulate the following result with tangent cones to sublevel sets instead. However, we will use descent cones as is done in [1].

Proposition 4.5. *The vector $\mathbf{x} \in \mathbb{R}^N$ is the unique solution of P_1 with $\mathbf{y} = \mathbf{A}\mathbf{x}$ if and only if $\mathcal{D}(\|\cdot\|_1, \mathbf{x}) \cap \mathrm{Ker}\,\mathbf{A} = \{\mathbf{0}\}$.*

Proof. ($\Leftarrow$) Let $\mathbf{x} \in \mathbb{R}^N$ and $\mathcal{D}(\|\cdot\|_1, \mathbf{x}) \cap \operatorname{Ker} \mathbf{A} = \{\mathbf{0}\}$. Additionally, we assume that vector $\mathbf{x}$ satisfies linear constraints $\mathbf{Ax} = \mathbf{y}$. Suppose that there exists $\mathbf{z} \in \mathbb{R}^N$, $\mathbf{z} \neq \mathbf{x}$, such that $\mathbf{Az} = \mathbf{y}$ and $\|z\|_1 \leq \|x\|_1$. Then $\mathbf{A}(\mathbf{z}-\mathbf{x}) = 0$ and $z - x \in \operatorname{Ker} \mathbf{A}$. By definition, $\mathcal{D}(f, \mathbf{x}) = \bigcup_{\tau>0} \left\{\mathbf{z} \in \mathbb{R}^N : \|\mathbf{x} + \tau\mathbf{z}\|_1 \leq \|\mathbf{x}\|\right\}$. So, $\mathbf{z} - \mathbf{x} \in \mathcal{D}(\|\cdot\|_1, \mathbf{x})$ (with $\tau = 1$ we have $\|\mathbf{x} + (\mathbf{z} - \mathbf{x})\|_1 \leq \|\mathbf{x}\|_1$). But then $\mathcal{D}(\|\cdot\|_1, \mathbf{x}) \cap \operatorname{Ker} \mathbf{A} \neq \{\mathbf{0}\}$, which contradicts the hypothesis. Thus, $\mathbf{x}$ is the unique minimizer of P_1.

($\Rightarrow$) Let $\mathbf{x} \in \mathbb{R}^N$ be the unique solution of P_1, i.e, $\mathbf{x}$ satisfies $\mathbf{Ax} = \mathbf{y}$ and for any $\mathbf{x}' \neq \mathbf{x}$ such that $\mathbf{Ax}' = \mathbf{y}$, $\|\mathbf{x}'\|_1 > \|\mathbf{x}\|_1$. Let $\mathbf{z} \in \mathcal{D}(\|\cdot\|_1, \mathbf{x})$ and $\mathbf{z} \neq \mathbf{0}$. Then there exists $\tau > 0$ such that $\|\mathbf{x} + \tau\mathbf{z}\|_1 \leq \|\mathbf{x}\|_1$. If we suppose that $\mathbf{z} \in \operatorname{Ker} \mathbf{A}$, then $\mathbf{Az} = \mathbf{0}$ and $\mathbf{A}(\mathbf{x} + \tau\mathbf{z}) = \mathbf{y}$. As $\mathbf{x}$ is the unique minimizer, it has to be that $\|\mathbf{x} + \tau\mathbf{z}\|_1 > \|\mathbf{x}\|_1$, which leads to a contradiction. Thus, $\mathcal{D}(\|\cdot\|_1, \mathbf{x}) \cap \operatorname{Ker} \mathbf{A} = \{\mathbf{0}\}$. □

The visual interpretation of this result can be seen in Fig 2. If kernel intersects the descent cone in more than just $\mathbf{0}$, then there exists $\mathbf{z} \in \operatorname{Ker} \mathbf{A}$ such that $\|\mathbf{z} + \mathbf{x}^*\|_1 \leq \|\mathbf{x}^*\|_1$. Also, $\mathbf{A}(\mathbf{z} + \mathbf{x}^*) = \mathbf{y}$ and thus $\mathbf{x}^*$ is not the unique solution of P_1.

Now that we have the foundation laid down, let us move onto the main part of this report. Most of the time during the project was spent on working with two papers: "The Convex Geometry of Linear Inverse Problems" by V. Chandrasekaran et al. (2012) and "Living on the edge: Phase transitions in convex programs with random data" by D. Amelunxen et al. (2013). Those papers take Proposition 4.5 and ask a question: what is the probability of the descent cone and the kernel to intersect (other than in $\mathbf{0}$)? To answer that question they propose two different ways to "measure" how big the cone is compared to the ambient dimension. The second paper goes further than that and also studies the phase transition that happens as we increase m. We will discuss the minimal theory behind those two papers and look at some numerical experiments in the following sections.

4.1 The Convex Geometry of Linear Inverse Problems

Authors of this paper didn't limit their study to l_1-minimization; they also study adjacent sparsity-related problems by introducing *atomic norms*. However, we are not interested in generality right now, so we will formulate all the results in restriction to l_1 norm. Conveniently enough, it also makes connections with the next paper more apparent, as we can utilise descent cones here as well. The only thing left to define is Gaussian width.

Definition 4.6. The *Gaussian width* of a set $S \subset \mathbb{R}^N$ is defined as

$$w(S) = \mathbb{E}\left[\sup_{\mathbf{z} \in S} \mathbf{g}^T \mathbf{z}\right],$$

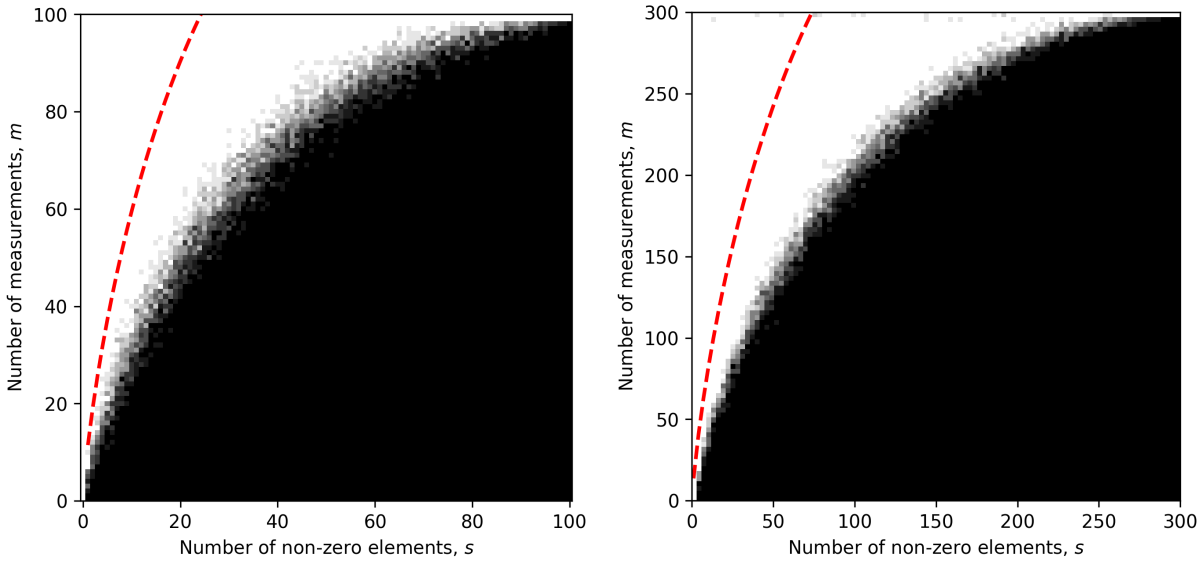
where $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is a vector of i.i.d. random variables with standard normal distribution.

The following two results are the most important ones for us; they tell us when we can recover vector $\mathbf{x}^*$ from P_1.

Theorem 4.7. *Let $\mathbf{A} \in M_{m \times N}$ be a random matrix with i.i.d components with $\mathcal{N}(0, 1)$ distribution, $\mathbf{x}^* \in \mathbb{R}^N$ and let $\Omega = \mathcal{D}(\|\cdot\|_1, \mathbf{x}^*) \cap \mathbb{S}^{N-1}$. Suppose that $\mathbf{y} = \mathbf{Ax}^*$. Then*

$$m \geq w(\Omega)^2 + 1 \implies \mathbb{P}\{\mathbf{x}^* \text{ is the unique solution of } P_1\} \geq 1 - \exp\left(-\frac{1}{2}(\lambda_m - w(\Omega))^2\right),$$

where λ_m is the expected length of an m-dimensional gaussian vector.



(a) (b)

Figure 3: Results of numerical experiments in dimensions 100 and 300. The red line represents the minimal m required by Corollary 4.9.

Proposition 4.8. *Let $\mathbf{x}^* \in \mathbb{R}^N$ be an s-sparse vector. We have the following inequality:*

$$w(\mathcal{D}(\|\cdot\|_1, \mathbf{x}^*) \cap \mathbb{S}^{N-1})^2 \leq 2s \log \frac{N}{s} + \frac{5}{4}s.$$

If we combine this result with Theorem 4.7, we get the following condition for successful recovery:

Corollary 4.9. *Let $\mathbf{x}^* \in \mathbb{R}^N$ be an s-sparse vector with $\mathbf{y} = \mathbf{A}\mathbf{x}^*$. Then $2s \log \frac{N}{s} + \frac{5}{4}s + 1$ measurements suffice to recover $\mathbf{x}^*$ from P_1 with high probability, i.e.,*

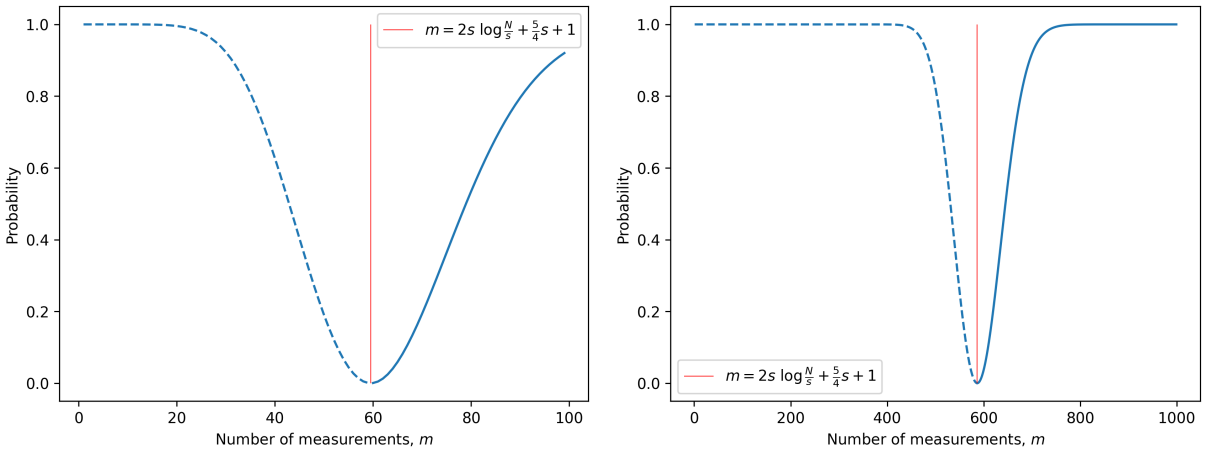
$$m \geq 2s \log \frac{N}{s} + \frac{5}{4}s + 1 \implies$$
$$\mathbb{P}\{\mathbf{x}^* \text{ is the unique solution of } P_1\} \geq 1 - \exp\left(-\frac{1}{2}\left[\lambda_m - w(\Omega)\right]^2\right). \quad (4.1)$$

In Fig. 3 we can see this estimate applied to dimensions 100 and 300.

Remark 4.10. In the paper, authors formulate this result without specifying the probability and just referring to it as being "high". And it is, no doubt, high; however, we can try studying that in more detail. At this point, the expression for probability contains $w(\Omega)$, which is not great for numerical experiments. The next lemma gives a probability estimate that depends explicitly on m, s and N.

Lemma 4.11. *Let $\mathbf{x}^* \in \mathbb{R}^N$ be an s-sparse vector with $\mathbf{y} = \mathbf{A}\mathbf{x}^*$. Then for $m \geq 2s \log \frac{N}{s} + \frac{5}{4}s + 1$ we have*

$$\mathbb{P}\{\mathbf{x}^* \text{ is the unique solution of } P_1\} \geq 1 - \exp\left(-\frac{1}{2}\left[\frac{m}{\sqrt{m+1}} - \sqrt{2s \log \frac{N}{s} + \frac{5}{4}s}\right]^2\right).$$



(a) (b)

Figure 4: Plots for the probability estimate as seen in (**??**). In (a) $N = 100$, $s = 10$, while in (b) $N = 1000$, $s = 100$.

Proof. In order to prove this result, we will want to approximate λ_m and $w(\Omega)$ with some inequalities. For $w(\Omega)$ we can utilise the result of the Proposition 4.9, and as for λ_m, it can be shown via integration that $\frac{m}{\sqrt{m+1}} \leq \lambda_m \leq \sqrt{m}$. Then we get

$$\lambda_m - w(\Omega) \geq \frac{m}{\sqrt{m+1}} - \sqrt{2s \log \frac{N}{s} + \frac{5}{4}s}.$$

We also notice that the function $f(x) = 1 - \exp\left(-\frac{1}{2}x^2\right)$ is monotonically increasing for $x \geq 0$. If the right side of the inequality above is non-negative, then we can insert it in the inequality for the probability and get the desired result. Let us denote for convenience $a = 2s \log \frac{N}{s} + \frac{5}{4}s$. According to corollary 4.9, "success" requires $m \geq a + 1$. As we are only interested in this case, we can conclude that

$$\frac{m^2}{m+1} - a \geq \frac{(a+1)^2}{a+2} - a = \frac{1}{a+2} > 0,$$

which means that

$$\lambda_m - w(\Omega) \geq \frac{m}{\sqrt{m+1}} - \sqrt{2s \log \frac{N}{s} + \frac{5}{4}s} > 0.$$

Inserting this inequality into (4.1) completes the proof. □

This is a more practical result, suitable for numerical study. The curves for probability from the last inequality are depicted in Fig. 4. Note, that the only relevant information on those plots lies to the right of the red line, as in the derivation of this result we assumed that $m > 2s \log \frac{N}{s} + \frac{5}{4}s + 1$.

Remark 4.12. In the figure we can see that "high probability" doesn't necessarily happen immediately after m crosses the threshold. This seems dissapointing, but we will excuse it as being the worst case due to the nature of estimates we did.

4.2 Living on the Edge

Another way to "measure" a cone for our purpose is proposed in [1]. Authors of the paper look at the result of the Proposition 4.5 in the context of conic integral geometry and connect it with another problem: "What is the probability of a randomly rotated convex cone to intersect a fixed convex cone?". It was known before that this question is answered by the kinematic formula for cones, which relies on the notion of conic intrinsic volumes. We will not dive into details of intrinsic volumes, but will instead look at one simple example: polyhedral cones. It is much easier to interpret and describes the intuition behind this concept.

As usual, we first list some definitions from convex geometry as given by [7]. We will denote by $[\mathbf{x}, \mathbf{y}]$ the line segment connecting points $\mathbf{x}$ and $\mathbf{y}$, i.e., the set $\{t\mathbf{x} + (1-t)\mathbf{y}, t \in [0,1]\}$.

Definition 4.13. The *relative interior* $\operatorname{ri} C$ of a convex set $C \subset \mathbb{R}^N$ is defined as

$$\operatorname{ri} C = \{\mathbf{x} \in \operatorname{aff} C : \exists \varepsilon > 0, B(\mathbf{x}, \varepsilon) \cap \operatorname{aff} C \subset C\},$$

where $\operatorname{aff} C$ is the affine hull of the set C, i.e., the smallest affine space that contains C, and $B(\mathbf{x}, \varepsilon)$ is the ball of radius ε centered in $\mathbf{x}$.

Definition 4.14. Let $C' \subset C \subset \mathbb{R}^N$ be convex sets. If for every $\mathbf{x}, \mathbf{y} \in C$, such that $\operatorname{ri}[\mathbf{x}, \mathbf{y}] \cap C' \neq \emptyset$, both x and y are in C', then C' is called a *face* of C.

Definition 4.15. Let $\mathcal{C} \subset \mathbb{R}^N$ be a polyhedral cone. For $k \in [\![1..N]\!]$, the kth *conic intrinsic volume* is defined as

$$\nu_k(\mathcal{C}) = \mathbb{P}\left\{\operatorname{proj}_{\mathcal{C}} \mathbf{g} \text{ lies in the relative interior of a } k\text{-dimensional face of } \mathbb{C}\right\},$$

where $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is a vector of independent random variables with standard normal distribution.

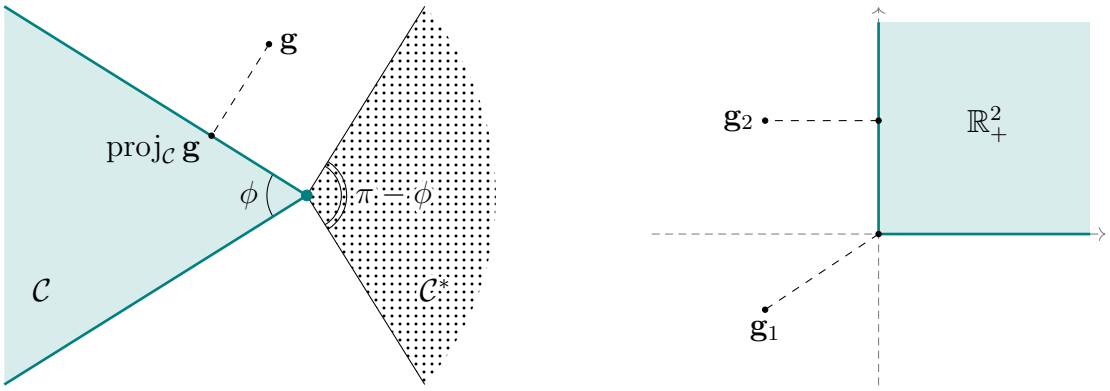
For an arbitrary cone, we first approximate it with polyhedral cones and then define the kth intrinsic volume as the limit of the sequence formed by kth intrinsic volumes of the approximating sequence.

Definition 4.16. Let $\mathcal{C} \subset \mathbb{R}^N$ be a closed convex cone. The statistical dimension $\delta(\mathcal{C})$ of the cone $\mathcal{C}$ is then defined as

$$\delta(\mathcal{C}) = \sum_{k=0}^{N} k\nu_k(\mathcal{C}).$$

Example 4.17. Let $\mathcal{C} \subset \mathbb{R}^2$ be a convex cone as shown in Fig. 5a. It has one zero-dimensional face ($\{\mathbf{0}\}$, the origin point), two one-dimensional faces (the two boundary rays) and one two-dimensional face (the cone itself). For the projection of vector $\mathbf{g}$ to be in the relative interior of a zero-dimensional face (which is also $\{\mathbf{0}\}$), the vector $\mathbf{g}$ has to be inside of the polar of $\mathcal{C}$, i.e., $\mathbf{g} \in \mathcal{C}^*$. The standard normal distribution is rotationally symmetric, so $\mathbb{P}\{\mathbf{g} \in \mathcal{C}^*\} = \frac{\pi - \phi}{2\pi}$. Similarly, for $\operatorname{proj}_C \mathbf{g}$ to lie inside $\mathcal{C}$, $\mathbf{g}$ itself must be inside of $\mathcal{C}$, which happens with probability $\frac{\phi}{2\pi}$. Finally, for one-dimensional faces we have the probability $\frac{1}{2}$. The statistical dimension can be easily calculated from the definition as

$$\delta(\mathcal{C}) = \frac{1}{2} + 2\frac{\phi}{2\pi} = \frac{\pi + 2\phi}{2\pi}.$$



(a) (b)

Figure 5: Examples of cones used for computation of statistical dimension in examples 4.17 and 4.18

Example 4.18. The non-negative orthant $\mathbb{R}^N_+$ forms a polyhedral cone as an intersection of two half-spaces. It has one zero-dimensional face $\{\mathbf{0}\}$, N one-dimensional faces of the form $\{0\}^k \times \mathbb{R}_+ \times \{0\}^{N-k-1}$ for $k \in [\![0..N-1]\!]$, $\frac{N(N-1)}{2}$ two-dimensional faces of the form $\{0\}^k \times \mathbb{R}_+ \times \{0\}^j \times \mathbb{R}_+ \times \{0\}^{N-k-j-2}$ for $(k,j) \in [\![0..N-2]\!]^2$ with $k+j = N-2$ and so on. In other words, k-dimensional faces only contain vectors with exactly k non-zero components. Then $\text{proj}_{\mathbb{R}^N_+} \mathbf{g}$ lies in the relative interior of a k-dimensional face of $\mathbb{R}^N_+$ if and only if $\mathbf{g}$ has exactly k positive components, so

$$\nu_k(\mathbb{R}^N_+) = \mathbb{P}\{\mathbf{g} \text{ has exactly } k \text{ positive components}\}.$$

There are $\binom{N}{k}$ ways to "choose" which components will be positive. The probability of each component to be positive is $1/2$; the probability of it to be negative is also $1/2$. As they are independent from each other, we get the result

$$\nu_k(\mathbb{R}^N_+) = \frac{1}{2^N}\binom{N}{k}.$$

The statistical dimension is then

$$\delta(\mathbb{R}^N_+) = \frac{1}{2^N}\sum_{k=0}^{N} k\binom{N}{k} = \frac{N}{2^N}\sum_{k=0}^{N-1}\binom{N-1}{k} = \frac{N}{2}.$$

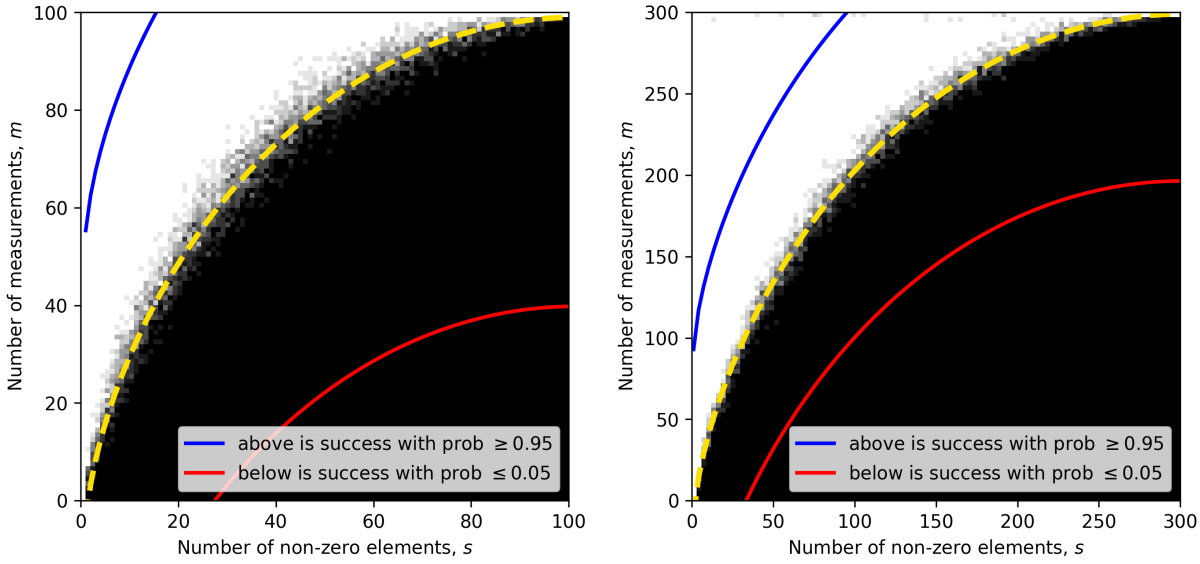
For example, $\delta(\mathbb{R}_+) = 1/2$ and $\delta(\mathbb{R}^2_+) = 1$.

The topic of statistical dimension and conic intrinsic volumes is discussed in more detail in [6]. Now we can look at the main result of this paper that predicts the location of the phase transition, as well as gives minimal number of measurements corresponding to a desired level of success.

Theorem 4.19. *Let $\mathbf{x}^* \in \mathbb{R}^N$ be a fixed vector and $p \in (0,1)$. Suppose $\mathbf{A} \in M_{m\times N}(\mathbb{R})$ is a matrix with independent $\mathcal{N}(0,1)$-distributed entries and $\mathbf{y} = \mathbf{A}\mathbf{x}^*$. Then*

$$m \le \delta(\mathcal{D}(\|\cdot\|_1, \mathbf{x}^*)) - a_\eta\sqrt{N} \implies \mathbb{P}\{\mathbf{x}^* \text{ is the unique solution of } P_1\} \le \eta$$
$$m \ge \delta(\mathcal{D}(\|\cdot\|_1, \mathbf{x}^*)) + a_\eta\sqrt{N} \implies \mathbb{P}\{\mathbf{x}^* \text{ is the unique solution of } P_1\} \ge 1-\eta,$$

where $a_\eta = \sqrt{8\log(4/\eta)}$.



(a) (b)

Figure 6: Results of numerical experiments in dimensions 100 and 300. Yellow line corresponds to the estimated statistical dimension of the descent cone.

Fig. 6 shows the estimates from this theorem and the statistical dimension plotted over the results of numerical experiments. We can see that the curve of statistical dimension sits exactly in the middle of the transition region; we also observe both the bounds and the transition region becoming relatively thinner in higher dimension.

Remark 4.20. In the paper, authors only show on plots the curve corresponding to the statistical dimension and omit the actual estimates for m, which are plotted in blue and red here. And understandably so — those predictions look terrible in dimension 100; at first I thought I had made a mistake in the code, as they look even worse than the ones given by the previous paper ([4]), that was published a year earlier. However, we also see that situation gets better in bigger dimension; moreover, from the statement of the theorem itself we can predict that the relative width of the region between these two lines will approach 0 as $N \to \infty$. We can observe that behavior in Fig. 7, where we also plot the curve for m from 4.11. To try and make this comparison fair, we fix the same probability (0.95) for both results, based on the probability estimate in 4.11. Though we have to keep in mind that this estimate is less optimistic than the original statement of 4.9, so this comparison might still be a bit unfair.

5 Details of numerical experiments

All of the plots shown here were done in Python. At first, I tried to write my own implementation of the Chambolle-Pock algorithm (see [3]), but it wasn't fast enough for large numbers of tests, so I switched to `cvxpy` library. For plots in figures 3 and 6 I generated for each pair (s, m) 10 random normal matrices and 10 s-sparse vectors $\mathbf{x}$ with non-zero components being either 1 or -1. Then corresponding vectors of measurements $\mathbf{y}$ were computed and the obtained problem was passed into the solver. The result of the procedure was considered to be a success if the absolute error between $\mathbf{x}$ and the obtained

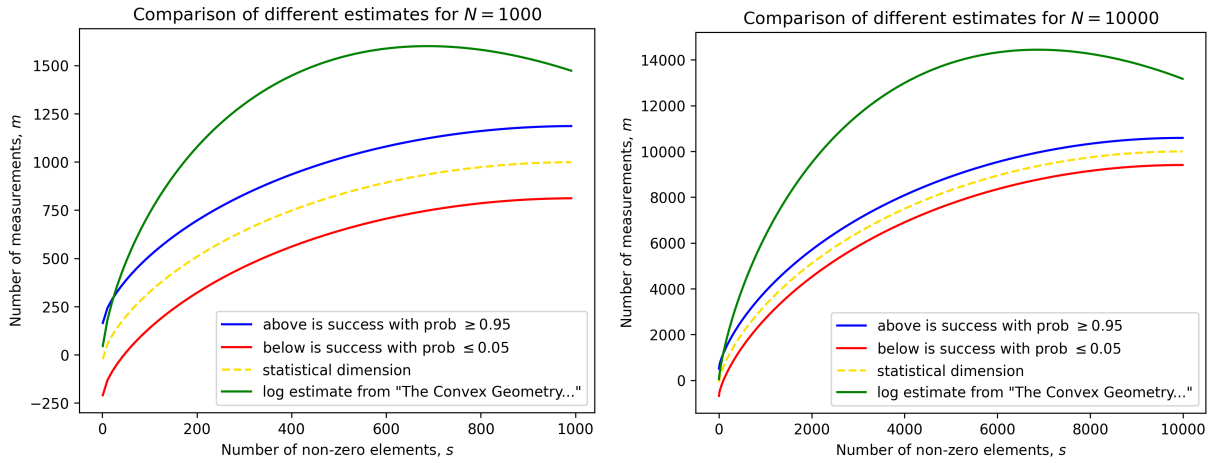


Figure 7: Comparison of different estimates in higher dimensions.

solution was less then 10^{-4}. Then the frequency of successful tests was calculated and plotted as a corresponding shade of gray, with white representing 100% success.

For the study of transition phase, the first major question was how to compute the statistical dimension of $\|\cdot\|_1$. Thankfully, the paper [1] provides us with the result below, that gives us tight bounds for the statistical dimension.

Proposition 5.1. *Let $\mathbf{x} \in \mathbb{R}^N$ be an s-sparse vector. Then the statistical dimension of the descent cone of the l_1 norm satisfies the inequality*

$$\psi\left(\frac{s}{N}\right) - \frac{2}{\sqrt{sN}} \leq \frac{\delta(\mathcal{D}(\|\cdot\|_1, \mathbf{x}))}{N} \leq \psi\left(\frac{s}{N}\right), \tag{5.1}$$

where $\psi : [0,1] \to [0,1]$ is defined as

$$\psi(\rho) := \inf_{\tau \geq 0} \left[\rho(1+\tau^2) + (1-\rho)\sqrt{\frac{2}{\pi}} \int_\tau^\infty (u-\tau)^2 e^{-u^2/2} \mathrm{d}u\right]. \tag{5.2}$$

The infimum in 5.2 is achieved for the unique value of τ that solves the equation

$$\sqrt{\frac{2}{\pi}} \int_\tau^\infty \left(\frac{u}{\tau} - 1\right) e^{-u^2/2} \mathrm{d}u = \frac{\rho}{1-\rho}. \tag{5.3}$$

Integrals in 5.2 and 5.3 can be simplified with the use of the error function (erf) to obtain more suitable for computation quantities:

$$\sqrt{\frac{2}{\pi}} \tau^{-1} e^{-\tau^2/2} + \operatorname{erf}\left(\frac{\tau}{\sqrt{2}}\right) - \frac{1}{1-\rho} = 0, \tag{5.4}$$

$$\psi(\rho) := \inf_{\tau \geq 0} \left[\rho(1+\tau^2) + (1-\rho)\left(\sqrt{\frac{2}{\pi}} \tau e^{-\tau^2/2} + (1+\tau^2)\left(\operatorname{erf}\left(\frac{\tau}{\sqrt{2}}\right) - 1\right)\right)\right]. \tag{5.5}$$

Then these results were used to plot the curves corresponding to the statistical dimension and bounds for the number of measurements from Theorem 4.19.

Another moment worth expanding on is how the green curve in Fig. 7 was obtained. In the inequality from 4.11 probability depends on m, so to make this comparison fair,

we would have to choose a specific threshold for the probability and then compute the corresponding m, which is exactly what was done.

Let p be the desired minimal probability of the successful recovery. Then we have an equation in terms of m:

$$1 - \exp\left(-\frac{1}{2}\left[\frac{m}{\sqrt{m+1}} - \sqrt{2s\log\frac{N}{s} + \frac{5}{4}s}\right]^2\right) = p.$$

With some simple manipulations we get the solution of this equation:

$$m = \frac{L^2 + L\sqrt{L^2+4}}{2},$$

where $L = \sqrt{\log\frac{1}{(1-p)^2}} + \sqrt{2s\log\frac{N}{s} + \frac{5}{4}s}$.

The code is available on GitHub as a Jupyter Notebook.

References

[1] D. Amelunxen, M. Lotz, M. McCoy, and J. Tropp. "Living on the edge: Phase transitions in convex programs with random data". In: *Information and Inference* 3 (Mar. 2013). DOI: 10.1093/imaiai/iau005.

[2] E. Candès, J. Romberg, and T. Tao. "Stable Signal Recovery from Incomplete and Inaccurate Measurements". In: *Communications on Pure and Applied Mathematics* 59 (Aug. 2006). DOI: 10.1002/cpa.20124.

[3] A. Chambolle and T. Pock. "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging". In: *Journal of Mathematical Imaging and Vision* 40 (May 2011). DOI: 10.1007/s10851-010-0251-1.

[4] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. "The Convex Geometry of Linear Inverse Problems". In: *Foundations of Computational Mathematics* 12.6 (Oct. 2012), pp. 805–849. DOI: 10.1007/s10208-012-9135-7.

[5] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Applied and Numerical Harmonic Analysis. Springer New York, Jan. 2013.

[6] M. B. McCoy and J. A. Tropp. "From Steiner Formulas for Cones to Concentration of Intrinsic Volumes". In: *Discrete Computational Geometry* 51 (Apr. 2014), pp. 926–963. DOI: 10.1007/s00454-014-9595-4.

[7] R. Rockafellar. *Convex Analysis*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 1997.