Compressed sensing

Anastasiia Storozhenko

June 8, 2025

Todo list

1 Introduction

Write intro (with applications?)

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 alternative ways to solve it. One such approach is to solve the corresponding l_0-minimization problem.

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:

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

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* $\mathbf{Ax} = \mathbf{y}$.

(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 $\mathbf{Ax} = \mathbf{y}$, 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. $\square$

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* $\mathbf{Ax} = \mathbf{Az}$ *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 $\mathbf{Az} = \mathbf{0}$. On the other hand, we also have $\mathbf{A0} = \mathbf{0}$, 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 $\mathbf{Ax} = \mathbf{Az}$. Then $\mathbf{x} - \mathbf{z}$ is $2s$-sparse and $\mathbf{Ax} - \mathbf{Az} = \mathbf{A}(\mathbf{x} - \mathbf{z}) = \mathbf{0}$, 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 iff $\operatorname{Ker}\mathbf{A} = \{\mathbf{0}\}$. 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 $\mathbf{Ax} = \mathbf{A}_S\mathbf{x}_S + \mathbf{A}_{\overline{S}}\mathbf{x}_{\overline{S}} = \mathbf{0} + \mathbf{A}_{\overline{S}}\mathbf{0} = \mathbf{0}$, 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, $\operatorname{Ker}\mathbf{A}_S = \{\mathbf{0}\}$. $\square$

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 $m \geq 2s$.

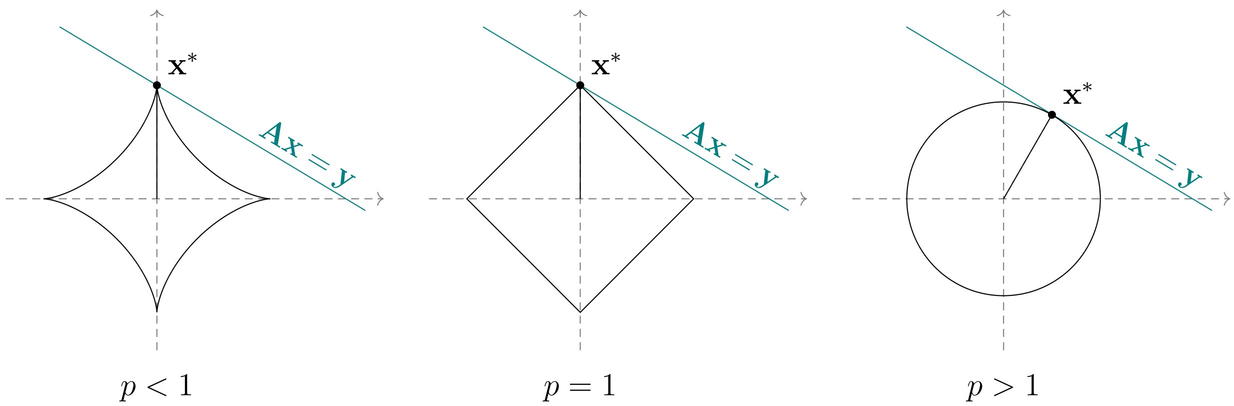

Figure 1: Unit balls

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 [..]). Another example is random matrices, where it is enough to reconstruct vector $\mathbf{x}$ with some probability of success (we will see examples of that later). add ref

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 [source] . In fact, any non-convex optimization problem is NP-hard, which means that we have to look for convex alternatives instead. And one such alternative approach is l_1-minimization. add ref

3 From l_0 to l_1

Add description to Fig1, fix spacing

In the context of this work, when we speak of $l1$-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 it is non-convex and thus 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 $\operatorname{card}(S) \le s$.

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 $\mathbf{y} = \mathbf{A}\mathbf{z}$ iff $\mathbf{A}$ satisfies the null space property relative to S.*

Proof. ($\Rightarrow$) Let $\mathbf{v} \in \operatorname{Ker} \mathbf{A}\backslash\{\mathbf{0}\}$. Then $\mathbf{A}\mathbf{v} = \mathbf{A}(\mathbf{v}_S + \mathbf{v}_{\overline{S}}) = 0$ and $\mathbf{A}\mathbf{v}_S = -\mathbf{A}\mathbf{v}_{\overline{S}}$. Vector $\mathbf{v}_S$ is supported on S, so by the hypothesis, $\mathbf{v}_S$ is the unique solution of P_1 with $\mathbf{y} = \mathbf{A}\mathbf{v}_S$. However, $-\mathbf{v}_{\overline{S}}$ is another solution of the equation $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{v}_S$, 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 $\mathbf{y} = \mathbf{A}\mathbf{z}$. Suppose that there exists a vector $\mathbf{z}^* \neq \mathbf{z}$ such that $\mathbf{z}^*$ minimizes $\|\cdot\|_1$ with $\mathbf{A}\mathbf{z}^* = \mathbf{A}\mathbf{z}$. 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 $\|\mathbf{z}\|_1 \le \|\mathbf{z} - \mathbf{z}^*_S\|_1 + \|\mathbf{z}^*_S\| < \left\|\mathbf{z}^*_{\overline{S}}\right\|_1 + \|\mathbf{z}^*_S\|_1 = \|\mathbf{z}^*\|_1$, 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 $\mathbf{y} = \mathbf{A}\mathbf{z}$ iff $\mathbf{A}$ satisfies the null 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 $\mathbf{y} = \mathbf{A}\mathbf{z}$.

Despite this result being fundamental for the theoretical study of compressed sensing, it is not easy to verify if a matrix satisfies this property in practice. Turns out, in the case of random matrices it is not needed: in fact, we can obtain satisfying recovery guaranties in a probabilistic form. That is exactly what we are going to focus on for the rest of this report.

4 Minimal number of measurements

Prop 3.10 from Convex Geometry...

Living on the Edge (definitions, theorem II, prop 4.5)

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. However, another field of application is data compression; in that setting we are free to choose the matrix $\mathbf{A}$ however we like. Considering the vast number of results for normally distributed matrices, they become the obvious candidates for study.

Before proceeding to recovery criteria for random matrices, we first have to recall some definitions from convex geometry.

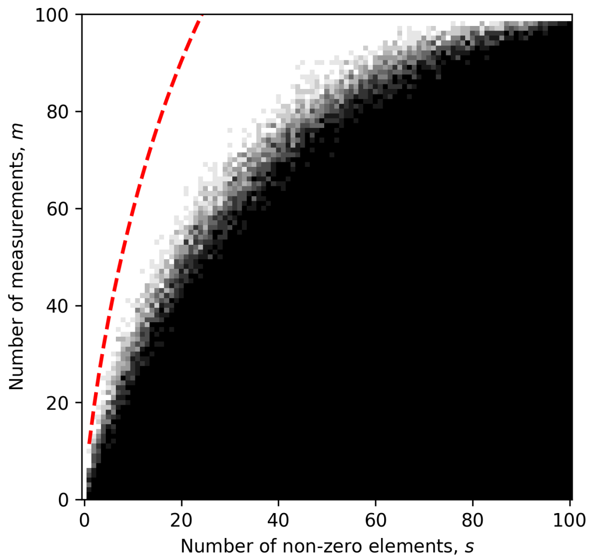

Figure 2: Log

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 $T_C(\mathbf{x}) = \mathrm{cl}\{\alpha \mathbf{x} : \mathbf{x} \in C, \alpha > 0\}$. 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

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

Remark 4.4. Tangent cone to level sets

Proposition 4.5. *The vector* $\mathbf{x}$ *is the unique solution of* P_1 *iff* $\mathcal{D}(\|\cdot\|_1, \mathbf{x}) \cap \mathrm{Ker}\, \mathbf{A} = \{\mathbf{0}\}$.

Proof. □

Here write about "Convex geometry of..."?

Definition 4.6. Atomic norm

Definition 4.7. Gaussian width

Proposition 4.8. *Corollary 3.3(1)*

Proposition 4.9. *Prop 3.10*

Then about Living on the Edge

Definition 4.10. Intrinsic volume

Definition 4.11. Statistical dimension

Theorem 4.12. *Theorem II*

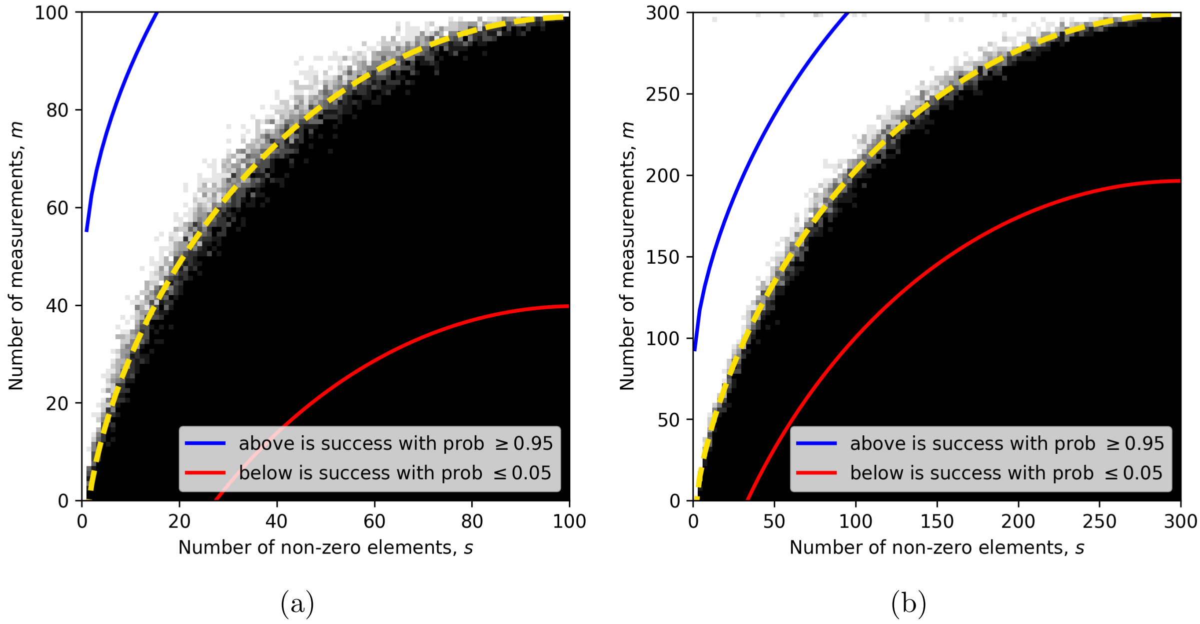

Figure 3: Living on the Edge

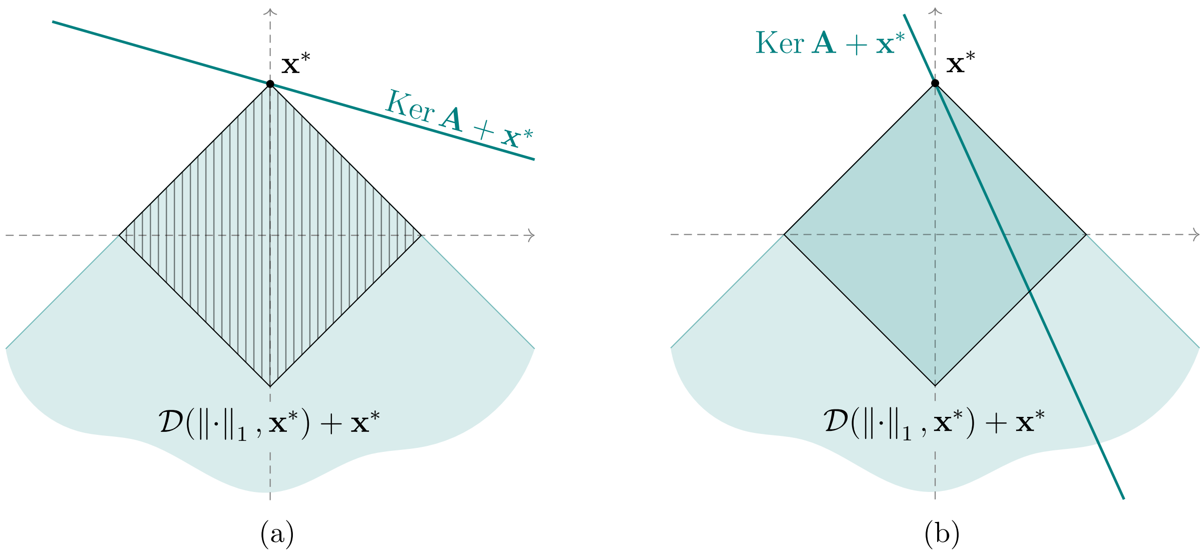

Figure 4: descent cone

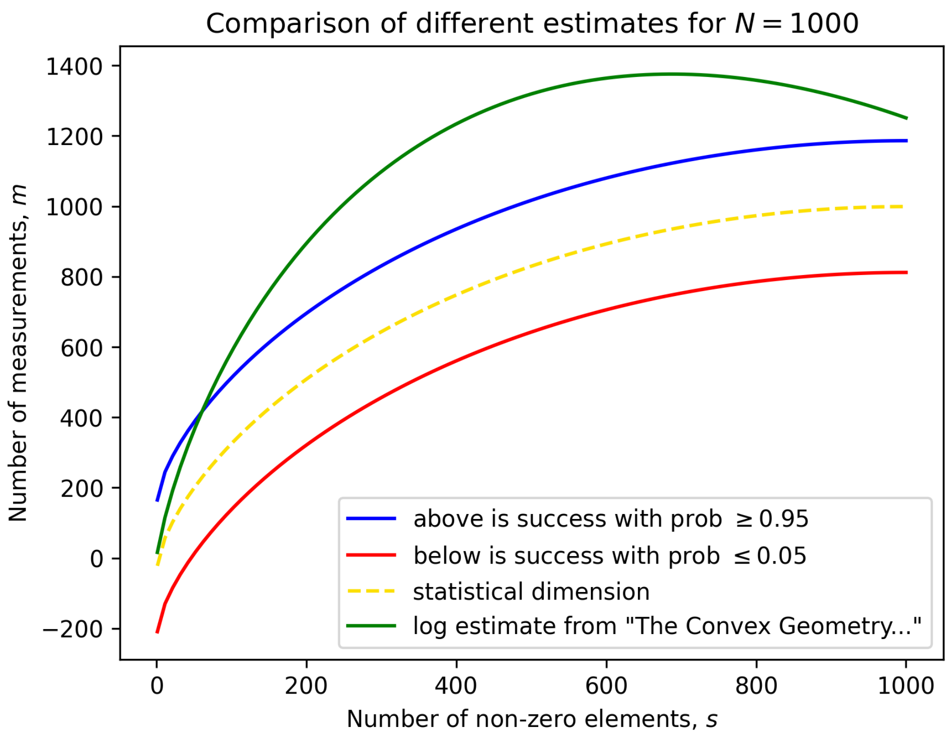

Figure 5: Comparison

5 Details of umerical experiments

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:

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

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

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] 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.

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

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