University of Luxembourg

Master in Mathematics

Student project

Compressed sensing

Author:
Lucien May

Supervisor:
Prof. I. Nourdin

Winter semester 2014

1 Introduction

Let us consider an s-sparse vector $\mathbf{x_0}$ in $\mathbb{R}^n$, that is a vector in $\mathbb{R}^n$ with at most s non-zero entries. Suppose we have $m \leq n$ linear measurements of our vector $\mathbf{x_0}$, i.e. we have a vector $\mathbf{y}$ in $\mathbb{R}^m$ such that

$$\mathbf{y} = A\mathbf{x_0}, \tag{1}$$

where $A \in \mathrm{Mat}(\mathbb{R}, m, n)$. An important question in signal processing is, whether we can reconstruct the vector $\mathbf{x_0}$ out of the knowledge of $\mathbf{y}$. If $m < n$, this linear system is under-determined and we cannot hope to solve this system uniquely. We must specify additional constraints on $\mathbf{x_0}$ in order to reduce the number of interesting solutions to 1. We thus ask whether we can find an unique solution to the following optimization problem

$$\begin{array}{ll} \text{minimize} & \max\{k|1 \leq i_1 < ... < i_k \leq n : x_{i_j} \neq 0\} =: ||x||_0 \\ \text{subject to} & \mathrm{A}\mathbf{x} = A\mathbf{x_0}. \end{array}$$

The $|| \cdot ||_0$ pseudo-norm[1] controls the sparsity of the vector $\mathbf{x}$. Though it is possible under suitable conditions on A to solve this problem, it is computationally untractable if n is large [2]. To circumvent this problem, it turns out that replacing the function $|| \cdot ||_0$ by $|| \cdot ||_1$, the l_1-norm, leads to a computationally efficient algorithm. It is remarkable that the l_1-norm is adequate to control the sparsity of $\mathbf{x}$. In this project we consider that the matrix A is random, more precisely we suppose that the $n \cdot m$ entries of A are independent and identically distributed normal random variable $\mathcal{N}(0, 1/m)$. The problem we shall discuss is thus the following

$$\begin{array}{rl} \text{minimize} & ||x||_1 \\ \text{subject to} & A\mathbf{x} = A\mathbf{x_0}. \end{array} \tag{2}$$

In this work we present in detail that for $m \geq 2\beta s \log n + s$ where $\beta > 1$ is arbitrary but fixed, we can recover $\mathbf{x}$ in problem (2) with high probability. Problems of this kind fall in the domain of *compressed sensing*, which is currently a rapidly developing field. This presentation is based on the paper [1]. Then we illustrate the problem by computational experiments, which confirm the theoretical results.

[1]It is not a norm because the triangle inequality is not satisfied.

2 The sparse vector recovery problem

The aim will be to present the following result.

Theorem 1. *Let* $\mathbf{x_0}$ *be an* s*-sparse vector in* $\mathbb{R}^n$ *and* $\beta > 1$ *. For a* $m \times n$ *matrix* A *with normal* $\mathcal{N}(0, \frac{1}{m})$ *random variables as entries and* $m \geq 2\beta s \log n + s$ *the recovery of* $\mathbf{x_0}$ *in problem (2) is exact with probability at least* $1 - 2n^{-f(\beta,s)}$*, where*

$$f(\beta, s) = \left(\sqrt{\frac{\beta}{2s} + \beta - 1} - \sqrt{\frac{\beta}{2s}} \right)^2 .$$

Note that $f(\beta, s)$ is positive for all $\beta > 1$ and $s > 0$

The proof is based on special properties of the l_1 norm and on concentration of measure inequalities. Recall that the *subdifferential* $\partial f(x_0)$ of a convex function $f : \mathbb{R}^n \to \mathbb{R}$ at $x_0 \in \mathbb{R}^n$ is defined as the set

$$\partial f(x_0) = \{g \in \mathbb{R}^n | \forall x \in \mathbb{R}^n : f(x) \geq f(x_0) + \langle g, x - x_0 \rangle\},$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product on $\mathbb{R}^n$.

It is clear from the triangle inequality that the l_1 norm is a convex function on $\mathbb{R}^n$. The subdifferentials of $|| \cdot ||_1$ have a special structure. For $x \in \mathbb{R}^n$ having at least one zero entry we have that there exists a subspace $T \subseteq \mathbb{R}^n$ and $e \in T$ such that

$$\partial ||x||_1 = \{z \in \mathbb{R}^n : P_T(z) = \mathbf{e} \text{ and } ||P_{T^\perp}||_\infty \leq 1\},$$

where P_V designates the projection operator on V. We say in this case that the norm $|| \cdot ||_1$ is *decomposable* at x. Moreover we have that for $w \in T^\perp$

$$||w||_1 = \sup_{v \in T^\perp, ||v||_\infty \leq 1} \langle v, w \rangle. \tag{3}$$

Let us illustrate this in the case of $\mathbb{R}^2$. For $x = (x_1, x_2) \in \mathbb{R}$ we easily find the following subdifferentials.[2]

$\text{sign}(x_1)$	$\text{sign}(x_2)$	$\partial \|\|(x_1, x_2)\|\|_1$	T	e
0	+1	$\{(x, 1) \in \mathbb{R}^2 \| \|x\| \leq 1\}$	$\mathbb{R}(0, 1)$	$(0, 1)$
0	-1	$\{(x, -1) \in \mathbb{R}^2 \| \|x\| \leq 1\}$	$\mathbb{R}(0, -1)$	$(0, -1)$
+1	0	$\{(1, x) \in \mathbb{R}^2 \| \|x\| \leq 1\}$	$\mathbb{R}(1, 0)$	$(1, 0)$
-1	0	$\{(-1, x) \in \mathbb{R}^2 \| \|x\| \leq 1\}$	$\mathbb{R}(-1, 0)$	$(-1, 0)$
0	0	$\{(x, y) \in \mathbb{R}^2 \| \|x\| \leq 1, \|y\| \leq 1\}$	$\{(0, 0)\}$	$(0, 0)$

[2] sign(x)=0 means that $x = 0$.

Formula (3) is easily seen to be valid. Similar types of computations can easily be done also in higher dimensions. If $x_0 \in \mathbb{R}^n$ is an s-sparse vector, one sees that dim(T)= s, and $\mathbf{e}$=sign($\mathbf{x_0}$) where sign($\cdot$) return the signs of the components. This shows that

$$||\mathbf{e}||_2 = \sqrt{s}. \tag{4}$$

The following proposition gives a condition under which a unique solution to (2) exists.

Proposition 1. *Let $\mathbf{x_0} \in \mathbb{R}$ be a point at which the l_1-norm is decomposable and T, $\mathbf{e}$ the corresponding subspace, respectively vector as previously described. Let A be a random matrix as in theorem 2.1. Suppose that A is one-to-one on the subspace T and that there exists $\mathbf{y} \in imA^* \subseteq \mathbb{R}^n$ such that*

1. $P_T(\mathbf{y}) = \mathbf{e}$

2. $||P_{T^\perp}(\mathbf{y})||_\infty < 1$.

Then $\mathbf{x_0}$ is the unique solution to the linear inverse problem (2).

Proof. One has to check that if $\mathbf{h} \in \mathbb{R}^n$ is such that $A\mathbf{h} = 0$, then $||\mathbf{x_0} + \mathbf{h}||_1 \geq ||\mathbf{x_0}||_1$. Using formula (3) we see that there exists $\mathbf{v} \in T^\perp$ such that $||\mathbf{v}||_\infty \leq 1$ and $\langle \mathbf{v}, P_{T^\perp}(\mathbf{h})\rangle = ||P_{T^\perp}(\mathbf{h})||_1$. Note that $P_T(\mathbf{e}+\mathbf{v}) = \mathbf{e}$ and $||P_{T^\perp}(\mathbf{e}+\mathbf{v})||_\infty = ||P_{T^\perp}(\mathbf{v})||_\infty \leq 1$, so that $\mathbf{e}+\mathbf{v} \in \partial||\mathbf{x_0}||_1$. Thus

$$\begin{aligned}
||\mathbf{x_0}+\mathbf{h}||_1 &\geq ||\mathbf{x_0}||_1 + \langle \mathbf{e}+\mathbf{v}, \mathbf{h}\rangle \\
&= ||\mathbf{x_0}||_1 + \langle \mathbf{e}+\mathbf{v}-\mathbf{y}, \mathbf{h}\rangle \\
&= ||\mathbf{x_0}||_1 + \langle \mathbf{v}-P_{T^\perp}(\mathbf{y}), \mathbf{h}\rangle \\
&= ||\mathbf{x_0}||_1 + \langle \mathbf{v}-P_{T^\perp}(\mathbf{y}), P_{T^\perp}(\mathbf{h})\rangle \\
&\geq ||\mathbf{x_0}||_1 + (1-||P_{T^\perp}(\mathbf{y})||_\infty)||P_{T^\perp}(\mathbf{h})||_1.
\end{aligned}$$

The first equality comes from the fact that $\mathbf{y} = A^*\mathbf{w}$, so that $\langle \mathbf{y}, \mathbf{h}\rangle = \langle A^*\mathbf{w}, \mathbf{h}\rangle = \langle \mathbf{w}, A\mathbf{h}\rangle = 0$. The second equality simply uses that $P_T(\mathbf{y}) = \mathbf{e}$. The third equality comes from the fact that $\mathbf{v} - P_{T^\perp}(\mathbf{y}) \in T^\perp$. The last inequality comes from the fact that $\langle P_{T^\perp}(\mathbf{y}), P_{T^\perp}(\mathbf{h})\rangle \leq ||P_{T^\perp}(\mathbf{h})||_1||P_{T^\perp}(\mathbf{y})||_\infty$, where we have used (3).

Since $||P_{T^\perp}(\mathbf{y})||_\infty < 1$, we get that $||\mathbf{x_0}+\mathbf{h}||_1 > ||\mathbf{x_0}||_1$ unless that $P_{T^\perp}(\mathbf{h}) = 0$. In the latter case we then have that $\mathbf{h} \in T$. Since $A\mathbf{h} = 0$ and A is injective on T, we conclude that $\mathbf{h} = 0$. This shows that $\mathbf{x_0}$ is a minimizer of (2).
Now observe that the previous arguments show that the minimizer $\mathbf{x_0}$ is unique. □

If the entries of A are independent and identically distributed centered gaussian random variables with variance $1/m$, then it is clear that provided $\dim T =: d_T \leq m$, A restricted to T is one-to-one.
In order to find the vector $\mathbf{y}$ in proposition 1 we need a priori to solve the equation $P_T(\mathbf{y}) = \mathbf{e}$, which is of course not uniquely solvable. The plan is to find the least square solution of the equation

$$P_T(A^*\mathbf{q}) = \mathbf{e}, \tag{5}$$

and then show that for $\mathbf{y} := A^*\mathbf{q}$ we have $||P_{T^\perp}(\mathbf{y})||_\infty < 1$ with high probability. This means that with high probability the problem (2) has a unique solution.

Note that the map $P_T \circ A^*$ is a linear map from $\mathbb{R}^m \to \mathbb{R}^n$, where $m \leq n$. Let A_T and $A_{T^\perp}$ denote the restriction of A to T and $T^\perp$,respectively. We then have the following proposition.

Proposition 2. *If A_T is one-to-one, the least-square solution of (5) is given by*

$$\mathbf{q} = A_T(A_T^*A_T)^{-1}\mathbf{e}. \tag{6}$$

Moreover we have that

$$P_{T^\perp}(\mathbf{y}) = A_{T^\perp}^*\mathbf{q}. \tag{7}$$

Proof. Note that by assumption $A_T^*A_T$ is invertible, so that the MoorePenrose pseudoinverse of A_T^* is given by $A_T(A_T^*A_T)^{-1}$. The first formula and uniqueness then follows from the relation of the MoorePenrose pseudoinverse and the least square method, see for instance [4] . □

The following proposition will be crucial for the sequel.

Proposition 3. *Let A, $A_{T^\perp}^*$ and q as before. Then, for any T, $A_{T^\perp}^*$ and $\mathbf{q}$ are independent.*

Proof. By Cochran's theorem (see for example [5] theorem 5.17), if $X \sim \mathcal{N}(0, I_d)$ and E_1, E_2 are two orthogonal spaces in $\mathbb{R}^d$, then the projections $\Pi_{E_1}X, \Pi_{E_2}X$ are independent. Now note that $\mathbf{q}$ depends of A_T, and the claim then follows. □

In order to see that the second condition in proposition 1 is verified with high probability for the y previously constructed, we need to know the distribution of $P_{T^\perp}(\mathbf{y})$. It is given by the following proposition.

Proposition 4. *Let A and $\mathbf{q}$ be as in proposition 2. Then, conditioned on $\mathbf{q}$, $P_{T^\perp}(\mathbf{y})$ is distributed as $\iota_{T^\perp}\mathbf{g}$,where $\iota_{T^\perp}$ is an isometry from $\mathbb{R}^{n-d_T}$ onto $T^\perp$ and $\mathbf{g} \sim \mathcal{N}(0, \frac{||\mathbf{q}||_2^2}{m}I)$.*

Proof. Note that we have $P_{T^\perp}(\mathbf{y}) = A^*_{T^\perp}(\mathbf{q})$. Then the claim follows immediately from the fact that if X_1, X_2 are two independent normal random variables with distribution $\mathcal{N}(0,1)$, then $aX_1 + bX_2$, $a, b \in \mathbb{R}$, has distribution $\mathcal{N}(0, a^2 + b^2)$. □

In the following we will need to bound probabilities of the following type.

$$\begin{aligned} \mathbb{P}[||P_{T^\perp}(\mathbf{y})||_\infty \geq 1] &= \mathbb{P}[||P_{T^\perp}(\mathbf{y})||_\infty \geq 1| \ ||\mathbf{q}||_2 \leq \tau] \cdot \mathbb{P}[||\mathbf{q}||_2 \leq \tau] \\ &\quad + \mathbb{P}[||P_{T^\perp}(\mathbf{y})||_\infty \geq 1| \ ||\mathbf{q}||_2 \geq \tau] \cdot \mathbb{P}[||\mathbf{q}||_2 \geq \tau] \\ &\leq \mathbb{P}[||P_{T^\perp}(\mathbf{y})||_\infty \geq 1| \ ||\mathbf{q}||_2 \leq \tau] + \mathbb{P}[||\mathbf{q}||_2 \geq \tau] \end{aligned} \tag{8}$$

Here τ is some positive parameter to be specified later. Each of the probabilities in (8) will now be bounded. To bound the second probability on the right hand side of (8) we will use the following proposition.

Proposition 5. *We have that*

$$||\mathbf{q}||_2^2 = \left\langle \mathbf{e}, (A_T^* A_T)^{-1} \mathbf{e} \right\rangle,$$

and $||\mathbf{q}||_2^2$ *is distributed as* $||\mathbf{e}||_2^2 m B_{11}$, *where* B_{11} *is the first entry in the first column of an inverse Wishart matrix with* m *degrees of freedom and covariance* $\mathbf{I}_{d^T}$.

Proof. We first recall the distribution of a Wishart matrix. Suppose X is an $n \times p$ matrix, each row of which is drawn independently from a p-variate random normal vector with distribution $\mathcal{N}(0, V)$. Then the *Wishart distribution* with n degrees of freedom and covariance V is the probability distribution of the $p \times p$ random matrix $X^T X$.
Note that (6) implies by the definition of the adjoint that $||\mathbf{q}||_2^2 = \langle \mathbf{e}, (A_T^* A_T)^{-1} \mathbf{e} \rangle$.
By what has been said before we have that $(A_T^* A_T)^{-1}$ is a $d_T \times d_T$ inverse Wishart matrix with m degrees of freedom and covariance $\frac{1}{m}\mathbf{I}_{d^T}$.
Now let P denote an orthogonal matrix of size $n \times n$ such that $P\mathbf{e} = ||\mathbf{e}||_2 e_1$, where $e_1 = (1, 0, ..., 0)$. We have then that the law of $||\mathbf{q}||_2^2$ is equal in law to

$$\begin{aligned} \left\langle \mathbf{e}, (A_T^* A_T)^{-1} \mathbf{e} \right\rangle &= \left\langle \mathbf{e}, (P^* A_T^* A_T P)^{-1} \mathbf{e} \right\rangle \\ &= \left\langle \mathbf{e}, P^* (A_T^* A_T)^{-1} P^{-1} \mathbf{e} \right\rangle \\ &= \left\langle P\mathbf{e}, P^* (A_T^* A_T)^{-1} P\mathbf{e} \right\rangle \\ &= ||e||_2^2 \left\langle e_1, (A_T^* A_T)^{-1} e_1 \right\rangle \\ &= ||e||_2^2 m B_{11}, \end{aligned}$$

where we have used in the first equality that the Gaussian distribution is isotropic. □

One will now show that B_{11} is distributed as an inverse chi-squared random variable with $m - d_T + 1$ degrees of freedom .

Proposition 6. *Let $m \geq n$. The first entry in the first column of an inverse Wishart matrix with m degrees of freedom and covariance $\mathbf{I}_n$ is distributed as an inverse chi-squared random variable with $m - n + 1$ degrees of freedom.*

Proof. We first show that the squared distance of a vector $v \in \mathbb{R}^m$ to a subspace $L \subseteq \mathbb{R}^m$ spanned by the linearly independent vectors $v_1, ..., v_n$ is given by

$$d^2(v, L) = \frac{G(v_1, \ldots, v_n, v)}{G(v_1, \ldots, v_n),}$$

where

$$G(x_1, \ldots, x_k) = \det \begin{pmatrix} \langle x_1, x_1 \rangle & \ldots & \langle x_1, x_k \rangle \\ \ldots & \ldots & \ldots \\ \langle x_k, x_1 \rangle & \ldots & \langle x_k, x_k \rangle \end{pmatrix}$$

is a Gram determinant. Note that $(x_1, ..., x_k)$ is a linear dependent family if and only if $G(x_1, ..., x_k) = 0$. Indeed, if $(x_1, ..., x_k)$ is a linear dependent family, then there is $(\lambda_1, ..., \lambda_k) \neq 0$ such that $\sum_{i=1}^{k} \lambda_i x_i = 0$. Then $\sum_{i=1}^{k} \lambda_i L_i = 0$, where $L_1, ..., L_k$ are the lines of Gram matrix of $x_1, ..., x_k$. Conversely if $G(x_1, ..., x_k) = 0$, then there is $(\lambda_1, ..., \lambda_k) \neq 0$ such that $\sum_{i=1}^{k} \lambda_i L_i = 0$, where $L_1, ..., L_k$ are the lines of the Gram matrix of $x_1, ..., x_k$ as previously. Then we can see that $\sum_{i=1}^{k} \lambda_i x_i = 0$ is orthogonal to all x_i, and is thus 0. Linear dependence follows.

Now let $v = u + k$, where $u \in L$ and $k \in L^{\perp}$. By using the multi-linearity of the determinant, we find that

$$\begin{aligned} G(v_1, \ldots, v_n, v) &= G(v_1, \ldots, v_n, u) + G(v_1, \ldots, v_n, k) \\ &= G(v_1, \ldots, v_n, u) + \det \begin{pmatrix} \langle v_1, v_1 \rangle & \ldots & \langle v_1, v_n \rangle & 0 \\ \ldots & \ldots & \ldots & \ldots \\ \langle v_n, v_1 \rangle & \ldots & \langle v_n, v_n \rangle & 0 \\ * & * & * & ||k||^2 \end{pmatrix} \\ &= G(v_1, \ldots, v_n, u) + G(v_1, \ldots, v_n) ||k||^2 \end{aligned}$$

Now $G(v_1, \ldots, v_n, u) = 0$, because $(v_1, \ldots, v_n, u)$ is a linear dependent family. So we get that

$$||u||^2 = d^2(v, L) = \frac{G(v_1, \ldots, v_n, v)}{G(v_1, \ldots, v_n)}.$$

Let Φ be an $m \times n$ matrix with independent and identically distributed standard normal random variables. Denote the columns of Φ by C_1 to C_n. Since $m \geq n$ it is clear that the family $(C_1, ..., C_n)$ is linearly independent. Then it is easy to see that $(\Phi^*\Phi)_{ij} = \langle C_i, C_j \rangle$. Let us denote the elements of the inverse Wishart matrix $(\Phi^*\Phi)^{-1} =: B$ by B_{ij} $(i, j = 1, ..., n)$. Using the basic formulas for computing the inverse of matrices we find that

$$\begin{aligned} B_{11} &= \frac{\det \begin{pmatrix} \langle C_2, C_2 \rangle & \dots & \langle C_2, C_n \rangle \\ \dots & \dots & \dots \\ \langle C_p, C_2 \rangle & \dots & \langle C_p, C_n \rangle \end{pmatrix}}{\det \begin{pmatrix} \langle C_1, C_1 \rangle & \dots & \langle C_1, C_n \rangle \\ \dots & \dots & \dots \\ \langle C_n, C_1 \rangle & \dots & \langle C_n, C_n \rangle \end{pmatrix}} \\ &= \frac{G(C_2, ..., C_n)}{G(C_1, ..., C_n)} \end{aligned}$$

By what we have seen previously we see that

$$B_{11} = \frac{1}{d(C_1, L)^2},$$

where L is the vector space spanned by $C_2, ..., C_n$. Now

$$d(C_1, L)^2 = ||\Pi_{L^\perp}(C_1)||^2,$$

where $\Pi_{L^\perp}(C_1)$ denotes the orthogonal projection of C_1 to $L^\perp$, a subspace of $\mathbb{R}^n$ of dimension $m - n - 1$ (because L is of dimension $n - 1$). By Cochran's theorem (see for example theorem 5.17 in [5]) we have that

$$||\Pi_{L^\perp}(C_1)||^2 \sim \chi^2(m - n + 1).$$

□

Then we can use the following concentration result.

Proposition 7. *Let $(Y_1, ..., Y_D)$ be independent and identically distributed standard normal random variables and let $\alpha_1, ..., \alpha_D$ be nonnegative. We set $||\alpha||_\infty = \sup_{i=1,...,D} |a_i|$, $||\alpha||_2^2 = \alpha_1^2 + ... + \alpha_D^2$. and*

$$Z = \sum_{i=1}^{D} \alpha_i (Y_i^2 - 1).$$

Then the following inequalities hold for all positive x:

$$\mathbb{P}[Z \leq -2||\alpha||_2 \sqrt{x}] \leq e^{-x}. \tag{9}$$

For a proof we refer to ([3]), lemma 1, section 4. Now let U be a random variable with a chi-squared distribution with D degrees of freedom. Then it follows readily from (9) that for any positive x

$$\mathbb{P}[D - U \geq 2\sqrt{Dx}] \leq e^{-x}. \tag{10}$$

In our case for $D = m - d_T + 1$ and $\tau = \sqrt{\frac{m}{m-d_T+1-t}}||\mathbf{e}||_2$[3] in 8 we can deduce from (10) the following large deviation result for $||\mathbf{q}||_2$. For each $t > 0$,

$$\mathbb{P}\left[||\mathbf{q}||_2 \geq \sqrt{\frac{m}{m - d_T + 1 - t}}||\mathbf{e}||_2\right] \leq \mathbb{P}[z \leq m - d_T + 1 - t] \\ \leq \exp\left(-\frac{t^2}{4(m - d_T + 1)}\right), \tag{11}$$

where z is a chi-squared random variable with $m - d_T + 1$ degrees of freedom.

We now need to bound the first probability on the right hand side of (8). Using the same reasoning as in proposition 4, it follows directly from (7) that conditioned on $\mathbf{q}$, the components of $P_{T^\perp}(\mathbf{y})$ in $T^\perp$ are independent and identically distributed with distribution $\mathcal{N}(0, ||\mathbf{q}||_2^2/m)$. Thus for any $\tau > 0$, we have

$$\mathbb{P}[||P_{T^\perp}(\mathbf{y})||_\infty \geq 1| \ ||\mathbf{q}||_2 \leq \tau] \leq (n - s)\mathbb{P}[|v| \geq \sqrt{m}/\tau] \\ \leq n \exp\left(-\frac{m}{2\tau^2}\right), \tag{12}$$

where in the first inequality we have used the basic union bound, the fact that $\dim T = n - s$ and $v \sim \mathcal{N}(0, 1)$. In the second inequality we have used the fact that $\mathbb{P}[|v| \geq t] \leq e^{-t^2/2}$ for positive t.

Using the same τ as above with (4), i.e. $\tau = \sqrt{\frac{ms}{m-s+1-t}}$, we get that

$$\mathbb{P}[||P_{T^\perp}(\mathbf{y})||_\infty \geq 1| \ ||\mathbf{q}||_2 \leq \tau] \leq n \exp\left(-\frac{m - s + 1 - t}{2s}\right), \tag{13}$$

Let $D = m - s + 1$. As a consequence of (8), (11) and (13) we have thus proven until now that

$$\mathbb{P}[||P_{T^\perp}(\mathbf{y})||_\infty < 1] = 1 - \mathbb{P}[||P_{T^\perp}(\mathbf{y})||_\infty \geq 1] \\ \leq 1 - \exp\left(-\frac{t^2}{4D}\right) - n \exp\left(-\frac{D - t}{2s}\right). \tag{14}$$

[3]This is natural in view of what follows.

We now want to choose t on the right hand side of (14) such that both exponential terms are equal. For this set $t = 2\beta \log(n)\left(\sqrt{1+\frac{2s(\beta-1)}{\beta}}-1\right)$, and choose $m = 2\beta s \log(n) + s - 1$ with $\beta > 1$.

Observe that

$$\begin{aligned} t &= 2\beta \log(n)\left(\sqrt{1+\frac{2s(\beta-1)}{\beta}}-1\right) \\ &= 2\log(n)\sqrt{2s\beta}\sqrt{f(\beta,s)}, \end{aligned}$$

where $f(\beta,s) = \left(\sqrt{\frac{\beta}{2s}+\beta-1}-\sqrt{\frac{\beta}{2s}}\right)^2$ as in theorem 2.1. We thus get that

$$\frac{t^2}{4(m-s+1)} = f(\beta,s)\log(n),$$

and using $m = 2\beta s \log(n) + s - 1$ we find that the first exponential term on the right hand side of 14 equals to

$$e^{-\frac{t^2}{4(m-s+1)}} = n^{-f(\beta,s)}.$$

Now observe that

$$\begin{aligned} \sqrt{f(\beta,s)} &= \sqrt{\frac{\beta}{2s}+\beta-1}-\sqrt{\frac{\beta}{2s}} \\ \Rightarrow \left(\sqrt{f(\beta,s)}+\sqrt{\frac{\beta}{2s}}\right)^2 &= \frac{\beta}{2s}+\beta-1 \\ \Leftrightarrow f(\beta,s)+2\sqrt{f(\beta,s)}\sqrt{\frac{\beta}{2s}}+\frac{\beta}{2s} &= \frac{\beta}{2s}+\beta-1 \\ \Leftrightarrow \sqrt{\frac{2\beta}{s}f(\beta,s)} &= \beta-1-f(\beta,s) \end{aligned}$$

From this we get that

$$t = 2s\log(n)(\beta-1-f(\beta,s)).$$

Using once more that $m = 2\beta s \log(n) + s - 1$ we get that

$$\begin{aligned} m-s+1-t &= 2s(1+f(\beta,s))\log(n) \\ \Rightarrow -\frac{m-s+1-t}{2s} &= \log(n^{-1-f(\beta,s)}) \end{aligned}$$

so that

$$ne^{-\frac{m-s+1-t}{2s}} = n^{-f(\beta,s)}.$$

Taking the previous computations all together, we find that

$$\mathbb{P}[||P_{T^\perp}(\mathbf{y})||_\infty < 1] \leq 1 - 2n^{-f(\beta,s)}.$$

We get thus theorem 2.1.

3 Numerical simulations

We use Matlab to show how a phase transition in the convex optimization problem (2) takes place. We proceed as described in appendix A of [2].

1. Construct a vector $x_0 \in \mathbb{R}^d$ with s nonzero entries. The locations of the nonzero entries are random, and the nonzero entries are equal to $+1$ or -1 with equal probability.
2. Draw a random standard normal matrix $A \in \mathbb{R}^{m \times d}$.
3. Solve 2 to find $\hat{x}$ using the *cvx* package.
4. Declare success if $||\hat{x} - x_0||_2 \leq 10^{-5}$.

This procedure is repeated for a given m, s and d times, where d will be specified.

3.1 Experimental results

The following diagrams indicate the empirical probability of success to solve the convex optimization problem 2. The color bar indicates the probability of success. s is the number of non-zero elements in the vector $x_0 \in \mathbb{R}^d$ to be recovered and m is the number of random linear measurements done to recover x_0. Note the sharp transition between certain failure (blue) and certain correct recovery (red).

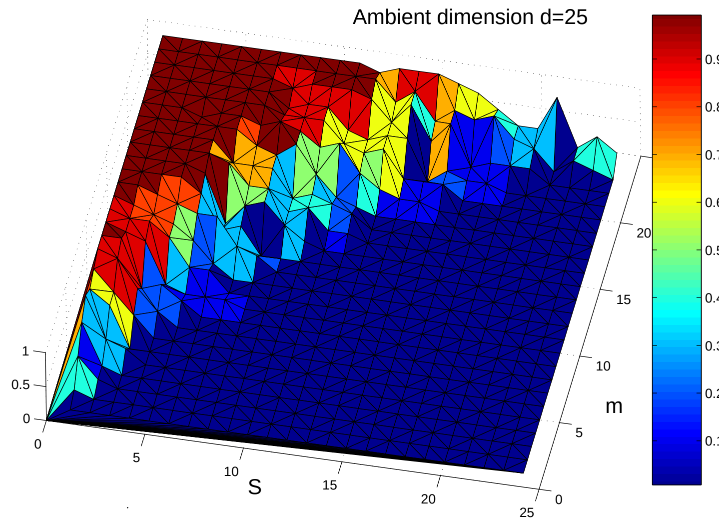


Figure 1: The dimension of the ambient space is $d = 25$. For each point the convex optimization problem has been repeated 10 times and the arithmetic mean is then used as an estimate for the probability of success.

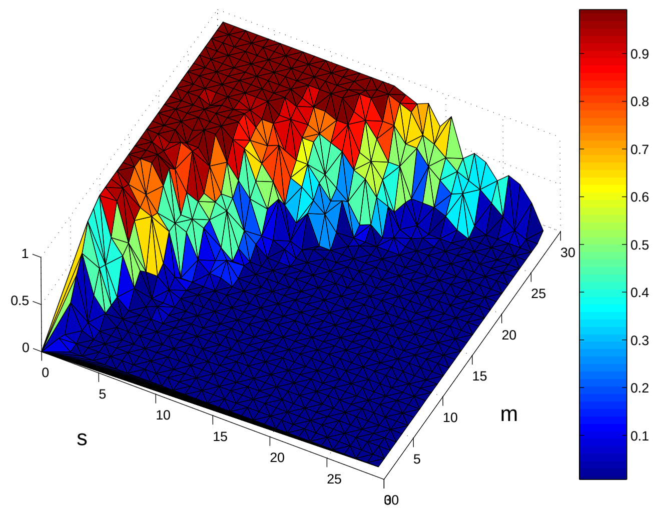
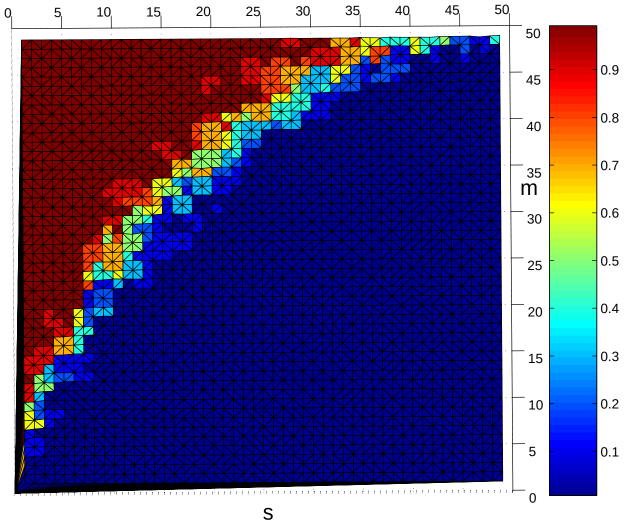


Figure 2: The dimension of the ambient space is $d = 30$. For each point the convex optimization problem has been repeated 20 times and the arithmetic mean is then used as an estimate for the probability of success.

Figure 3: The dimension of the ambient space is $d = 50$. For each point the convex optimization problem has been repeated 10 times and the arithmetic mean is then used as an estimate for the probability of success.



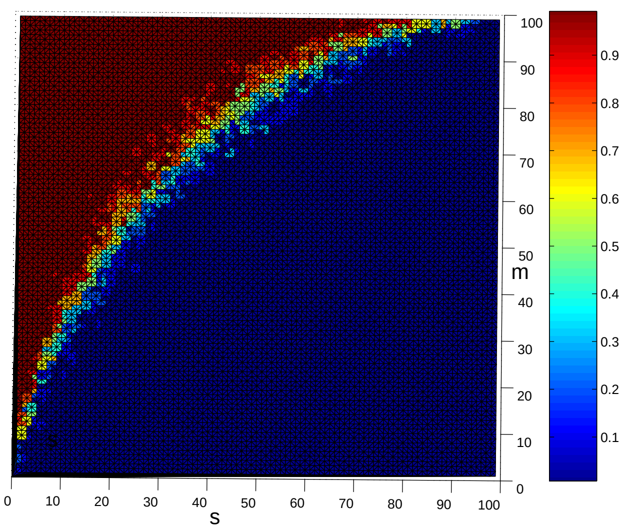


Figure 4: The dimension of the ambient space is $d = 100$. For each point the convex optimization problem has been repeated 10 times and the arithmetic mean is then used as an estimate for the probability of success.

3.2 Matlab code

```
function [y]=RandomVector(d,s)
%d is the size of the random vector
%s the number of non-zero elements.

a=zeros(d,1);
l=1;
    while l<s+1,

    x = floor(1 + (d)*rand(1,1));
%This a random element between 1 and d.
    if a(x)==0,
        a(x)=2*binornd(1,1/2)-1;
%This a -1/1 random element
        l=l+1;
    end,
end
y=a;
```

```

function [Y]=RandomSystem(m,n)
% Cr e une matrice m fois n avec de v.a. nomales standard.
Y=randn(m, n);
```

```
function w=successes(N,m,d,s)
%N is the number of experiments to do, f.ex. 50
%The size of the gaussian matrix is m times d. s is the
    sparsity
%of the vector to be recovered.
%Successes estimated the probability of success given the
    data.

l=0;
%Counting the number successes.
t=[];
y=[];
r=[]
z=[];
```

```
for i=1:N,
      t=RandomVector(d,s);
      y=MinimizationProblem(RandomSystem(m,d),t);
      r=y-t;
      z=norm(r);
if z<10^(-5),
%This threshold can be modified
      l=l+1;
end
end
w=l/N;
```

```
function s=MinimizationProblem(A,y)
%Solves the convex linear optimization problem
%using the cvx package
b=A*y;
n=size(y,1);
cvx_begin
      variable x(n);
      minimize( norm(x,1) );
      A*x == b;
cvx_end
s=x;
```

```
function PlottingFunction(N,d)
%N is the number of experiments to do, f.ex. 50
%d is the dimension of the ambient space
%This plots the probabilities of success as a function
%of m(number of measurements) and s(sparsity)

A=[];
 %This matrix will contain de data

for s=1:(d-1),
      for m=1:(d-1),
          A(max((s-1)*d-1,0)+m)=successes(N,m,d,s);
      end
end


S=[];
```

```
M=[];
for i=1:(d-1),
    for j=1:(d-1),
        S(max((i-1)*d-1,0)+j)=i;
    end
end

for i=1:(d-1),
    for j=1:(d-1),
        M(max((i-1)*d-1,0)+j)=j;
    end
end

tri = delaunay(S,M);
trisurf(tri,S,M,A);
```

References

[1] Candès, E.; Recht, B.: *Simple Bound for Recovering Low-complexity Models.* arXiv:1106.1474v2.

[2] Amelunxen, D.; Lotz, N.; Mccoy, M.; A. Tropp, Joel: *Living on the edge: phase transitions in convex programs with random data.* arXiv:1303.6672v2

[3] Laurent, B.; Massart, P.: *Adaptive estimation of quadratic functional by model selection.* Ann. Statist.,28(5):1302-1338,2000.

[4] Roman, S.: *Advanced linear algebra*, Third edition, Springer.

[5] Rivoirard, V; Stoltz, V.: *Statistique mathématique en action*, 2e édition, Vuibert