Stability of l_1 minimisation in compressed sensing

P. Wojtaszczyk
Institut of Applied Mathematics
University of Warsaw
and
Institut of Mathematics
Polish Academy of Sciences
wojtaszczyk@mimuw.edu.pl

Abstract

We discuss known results (c.f. [16, 6]) about stability of ℓ_1 minimisation (denoted Δ_1) with respect to the measurement error and how those results depend on the measurement matrix $\mathbf{\Phi}$. Then we produce a large class of measurement matrices $\mathbf{\Phi}$ for which we can apply results from [16] so we have estimate

$$\|\Delta_1(\mathbf{\Phi}(x)+r)-x\|_2 \leq C(\|r\|_2 + k^{-1/2}\sigma_k^1(x)). \quad (1)$$

We conclude with a modification of ℓ_1 minimisation which gives (1) for most random measurement matrices considered in compressed sensing literature. We also discuss stability of instance optimality in probability.

1 General description of compressed sensing

Let us start by explaining the general setup of compressed sensing. Given $N >> n$ we look for $n \times N$ a matrix $\mathbf{\Phi}$ such that vector $y = \mathbf{\Phi}x \in \mathbb{R}^n$ preserves information about $x \in \mathbb{R}^N$. We need a decoder (generally nonlinear) $\Delta : \mathbb{R}^n \to \mathbb{R}^N$ such that $\Delta(\mathbf{\Phi}x)$ looks like x. We require a k so $\Delta(\mathbf{\Phi}x) = x$ for x any k-sparse vector. We want Δ to be numerically friendly and k big. This leads to requiring that $\mathbf{\Phi}$ has RIP(k, δ).

Definition 1.1 ([3]) *Matrix $\mathbf{\Phi}$ has RIP(k,δ), $0 < \delta < 1$ if*

$$(1-\delta)\|x\|_2 \leq \|\mathbf{\Phi}x\|_2 \leq (1+\delta)\|x\|_2$$

for every k sparse vector $x \in \mathbb{R}^N$.

The largest possible k is $\sim n/\log(N/n)$.

All matrices which perform well for this maximal range of k are random. In this paper we consider only random matrices $\mathbf{\Phi}(\omega) = (\phi_{i,j}(\omega))$ where $\phi_{i,j}$'s are independent, i.i.d. subgaussian random variables e.g. Gaussian, Bernoulli. We normalize $\mathbb{E}|\phi_{i,j}|^2 = 1/n$ so columns Φ_j of $\mathbf{\Phi}$ have typically norm one. Let us call them standard matrices.

Given $0 < \delta < 1$ standard $\mathbf{\Phi}(\omega)$ satisfies RIP(k,δ) for $k = \lfloor c_1(\delta)n/\log(N/n) \rfloor$ with probability $\geq 1 - e^{-c_2(\delta)n}$ where $c_1, c_2 > 0$. This is well known, see e.g [7, 3, 1, 11]. The important point is that RIP is not practically verifiable even for moderately large N and k.

There are two main approaches to finding Δ for the above matrices. One approach, introduced by (E. Candes, D.Donoho et. al. see e.g. [3, 7]) is ℓ_1 minimization Δ_1 i.e.

$$\Delta_1(y) = \text{Argmin}\{\|z\|_1 \ : \ \mathbf{\Phi}(z) = y\}. \quad (2)$$

Another uses greedy algorithm. We start with an algorithm AL which for $y \in \mathbb{R}^n$ and vectors $(\Phi_j)_{j=1}^N$ gives a subset $\Lambda \subset \{1, \ldots, N\}$ with $\#\Lambda \leq k$ and

$\sum_{j\in\Lambda} a_j\Phi_j$. We set $\Delta_{AL}(y) = \sum_{j=1}^{N} a_j e_j$. This approach was proposed with AL=OMP by A.Gilbert, J. Tropp et al. see e.g. [15]. Some variants of OMP were used by D.Needell, J.Tropp, R.Vershynin et al. see [13, 14, 5].

Instance optimality originate in the work of A.Gilbert, M.Strauss and coauthors and was formally introduces in [4]. Define $\sigma_k^p(x) = \inf\{\|x - z\|_p \ : \ z \text{ is } k \text{ sparse}\}$. We would like to have

$$\|x - \Delta(\mathbf{\Phi}(x))\|_2 \leq C\sigma_k^2(x) \text{ for all } x \in \mathbb{R}^N. \tag{3}$$

Cohen Dahmen DeVore [4] showed this is impossible. So a pair random measurement matrix $\mathbf{\Phi}(\omega)$ and decoder $\Delta(\omega)$ is *instance optimal in probability* (for k with constant C) if for every $x \in \mathbb{R}^N$ there exists a set $\Omega(x)$ of probability very close to 1 such that for $\omega \in \Omega(x)$ we have $\|x - \Delta(\mathbf{\Phi}(x))\|_2 \leq C\sigma_k^2(x)$. This is possible – it was shown [4] that there exists such a decoder but it was totally impractical. In [16], [6] and [5] instance optimality in probability was shown for more practical decoders.

From practical point of view it is unrealistic to expect that we can get $\mathbf{\Phi}(x)$ exactly; generally we should expect some measurement error. So we apply our decoder to $\mathbf{\Phi}(x) + r$ for some $r \in \mathbb{R}^n$ and expect the result to be close to x. For ℓ_1 minimisation results of this type were proved first in [16] for Gaussian measurement matrix and later in [6] for standard matrices.

2 Stability of ℓ_1 minimization

Our arguments are largely geometric. On $\mathbb{R}^n$ and $\mathbb{R}^N$ we will consider the following well known norms $\|.\|_2$, $\|.\|_1$ and $\|.\|_\infty$. The unit balls in those norms will be denoted $\mathcal{B}_2$, $\mathcal{B}_1$ and $\mathcal{B}_\infty$ respectively with the dimension of the space added as superscript if needed. We will also use on $\mathbb{R}^n$ the following family of norms: $\|x\|_{J(\alpha)} = \max(\|x\|_2, \alpha\|x\|_\infty)$ for $\alpha > 1$. The unit ball in this norm will be denoted $\mathcal{B}_{J(\alpha)}$. If $\alpha = \sqrt{N}$ this norm will be denoted by $\|\cdot\|_J$ and the ball $\mathcal{B}_J$. By $\mathcal{S}$ we will denote the euclidean unit sphere.

In [16] and in [6] we introduced the following new geometric properties of the measurement matrix

Definition 2.1 *A matrix $\mathbf{\Phi}$ has LQ(μ, k) property if for every vector $y \in \mathbb{R}^n$ there exists $x \in \mathbb{R}^N$ such that $\mathbf{\Phi}x = y$ and $\|x\|_1 \leq \mu^{-1}\sqrt{k}\|y\|_2$. Equivalently $\mathbf{\Phi}(\mathcal{B}_1^N)$ (which is the convex hull of $(\pm\Phi_j)_{j=1}^N$) contains $(\mu/\sqrt{k})\mathcal{B}_2$.*

In [16] it was proved that for some $\mu > 0$ and $c > 0$ and $\delta > 0$ the Gaussian random measurement matrix satisfies LQ(μ, k) and RIP(k, δ) with $k = \lfloor cn/\log(N/n)\rfloor$ with overwhelming probability.

Definition 2.2 *A matrix $\mathbf{\Phi}$ has J(μ, k) property if for every vector $y \in \mathbb{R}^n$ there is $x \in \mathbb{R}^N$ such that $\mathbf{\Phi}x = y$ and $\|x\|_1 \leq \mu^{-1}\sqrt{k}\max(\|y\|_2, \sqrt{n/k}\|y\|_\infty)$. Equivalently $\mathbf{\Phi}(\mathcal{B}_1^N)$ (which is the convex hull of $(\pm\Phi_j)_{j=1}^N$) contains $(\mu/\sqrt{k})\mathcal{B}_{J(\sqrt{n/k})}$.*

In [9] (see also [6]) it was proved that any standard measurement matrix for some $c, \mu, \delta > 0$ has J(μ, k) and RIP(k, δ) with $k = \lfloor cn/\log(N/n)\rfloor$ with overwhelming probability.

Using those concepts we proved

Theorem 2.3 ([16]) *If $\mathbf{\Phi}$ satisfies RIP(k, δ) and LQ(μ, k) then there exists C such that for every $x \in \mathbb{R}^N$ and $r \in \mathbb{R}^n$*

$$\|\Delta_1(\mathbf{\Phi}(x) + r) - x\|_2 \leq C\big(\|r\|_2 + \frac{\sigma_k^1(x)}{\sqrt{k}}\big) \tag{4}$$

and

$$\|\Delta_1(\mathbf{\Phi}(x)+r)-x\|_2 \leq C\big(\|r\|_2+\sigma_k^2(x)+\|\mathbf{\Phi}(x|S^c)\|_2\big) \tag{5}$$

where S is a k element set for which $\sigma_k^2(x) = \|x|S^c\|_2$.

Theorem 2.4 ([6]) *If $\mathbf{\Phi}$ satisfies RIP(k, δ) and J(μ, k) then there exists C such that for every $x \in \mathbb{R}^N$ and $r \in \mathbb{R}^n$*

$$\|\Delta(\mathbf{\Phi}(x) + r) - x\|_2 \leq C\big(\|r\|_J + \frac{\sigma_k^1(x)}{\sqrt{k}}\big) \tag{6}$$

and

$$\|\Delta(\mathbf{\Phi}(x)+r)-x\|_2 \leq C\big(\|r\|_J+\sigma_k^2(x)+\|\mathbf{\Phi}(x|S^c)\|_J\big) \tag{7}$$

where S is a k element set for which $\sigma_k^2(x) = \|x|S^c\|_2$.

This in particular implies that Δ_1 is instance optimal in probability for all standard matrices. But it also shows stability of this decoder with respect to the measurement error r.

Let us note that (4) and (6) build upon and improve results from [2].

Actually Theorem 2.4 was not stated in [6]. We were interested in instance optimality and did not want to get into abstract formulations. However all the arguments needed to the proof of Theorem 2.4 are in [6] (see also [16]). If we compare Theorems 2.3 and 2.4 we see that when $r = 0$ they have the same conclusions. The only difference is in how the measurement error influences the estimate. Since $\|x\|_2 \le \|x\|_J$ the estimates in Theorem 2.3 are better. Later on we will comment on this in more detail. Unfortunately LQ in the maximal range of k was shown in [16] only for Gaussian standard measurement matrix and for matrix whose columns are independently drawn from $\mathcal{S}$ (but this is quite close to Gaussian). LQ clearly fails for Bernoulli random matrix. This is the main drawback of Theorem 2.3.

3 Random matrices satisfying LQ

Now we want to produce a large family of measurement matrices which satisfy LQ(μ, k) for $k = \lfloor cn\log(N/n)\rfloor$. First we need a unitary matrix in $\mathbb{R}^n$. In his fundamental paper [8] B.S. Kashin formulated the following

Theorem 3.1 *There exists a constant $c > 0$ such that with overwhelming probability a unitary matrix U satisfies*

$$c\|x\|_2 \le \frac{1}{2\sqrt{n}}(\|U(x)\|_1 + \|x\|_1) \le \|x\|_2 \qquad (8)$$

for all $x \in \mathbb{R}^n$.

He provided only a sketch of the proof, and stated only the existence of a matrix satisfying (8) but his proof easily gives that it holds with overwhelming probability on U. By duality we get that if a unitary matrix U satisfies (8) then $\operatorname{conv}\big(U^*(\frac{1}{\sqrt{n}}B_\infty) \cup \frac{1}{\sqrt{n}}B_\infty\big) \supset cB_2$. To see this note that if this is not the case then there exists x with $\|x\|_2 = 1$ such that $|\langle x, U^*z\rangle| < c\sqrt{n}$ and $|\langle x, z\rangle| < c\sqrt{n}$ for all $z \in B_\infty$. Thus we obtain $\|x\|_1 < c\sqrt{n}$ and $\|Ux\|_1 < c\sqrt{n}$ what contradicts (8). Since $B_J \supset \frac{1}{\sqrt{n}}B_\infty$ we get

Corollary 3.2 *With overwhelming probability on U we have* $\operatorname{conv} U(B_J) \cup B_J \supset cB_2$.

We fix one such unitary matrix. Now we define a random matrix $\mathbf{\Phi}(\omega)$ as follows. We take two symmetric, subgaussian random variables η and τ such that $\mathbb{E}|\eta|^2 = \mathbb{E}|\tau|^2 = 1$. Let $\eta_{j,i}$ and $\tau_{j,i}$ denote independent copies of η (resp. τ) and also $\eta_{j,i}$'s are independent from $\tau_{i,j}$'s. Our matrix $\mathbf{\Phi}$ has columns ϕ_j where $\phi_j = n^{-1/2}(\eta_{j,1}, \dots, \eta_{j,n})$ for $j < N/2$ and $\phi_j = n^{-1/2}U(\tau_{j,1}, \dots, \tau_{j,n})$ for $j \ge N/2$. From [6] we know that with overwhelming probability $\operatorname{conv}(\phi_j)_{j<N/2} \supset c\mathcal{B}_J$ and $\operatorname{conv}(\phi_j)_{j\ge N/2} \supset cU(\mathcal{B}_J)$ so $\operatorname{conv}(\phi_j)_{j=1}^N \supset c'\mathcal{B}_2$. So we need to show that the matrix $\mathbf{\Phi}$ satisfies RIP. Actually we will show the appropriate concentration inequality. Let us denote $\mathbf{\Phi}_1 = (\phi_j)_{j=1}^{N/2}$ and $\mathbf{\Phi}_2 = (\phi_j)_{j\ge N/2}$. We know (see e.g. [6]) that each of those matrices satisfies the concentration estimate: for each $\epsilon > 0$ there exists a constant $c(\epsilon) > 0$ such that for each x

$$\mathbb{P}(|\|\mathbf{\Phi}_s(x)\|^2 - \|x\|^2| > \epsilon\|x\|^2) \le 2e^{-nc(\epsilon)}. \qquad (9)$$

Thus (see [1]) matrices $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ satisfy RIP(k, δ). Now given $x \in \mathbb{R}^N$ we write $x = x_1 + x_2$ where x_1 equals x on coordinates $< N/2$ and x_2 equals x on coordinates $\ge N/2$. Note that

$$\begin{aligned} & \|\mathbf{\Phi}(x)\|^2 = \|\mathbf{\Phi}_1(x_1) + \mathbf{\Phi}_2(x_2)\|^2 \qquad (10) \\ = \; & \|\mathbf{\Phi}_1(x_1)\|^2 + \|\mathbf{\Phi}_2(x_2)\|^2 + 2\langle \mathbf{\Phi}(x_1), \mathbf{\Phi}_2(x_2)\rangle. \end{aligned}$$

Since $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are independent for a fixed $b = \mathbf{\Phi}_2(x_2)$ we have

$$\begin{aligned} & \mathbb{P}(|\langle \mathbf{\Phi}_1(x_1), b\rangle| > \tfrac{\epsilon}{3}\|b\| \cdot \|x_1\|) \\ = \; & \mathbb{P}\Big(\big|\sum_i^n \sum_j^{N/2} x_j b_i \eta_{i,j}\big| > \tfrac{\epsilon}{3}\|b\| \cdot \|x_1\|\Big). \end{aligned}$$

Since $\eta_{i,j}$ are independent subgaussian variables and $\sum_i^n \sum_j^{N/2} |x_j|^2 |b_i|^2 = \|b\|^2 \|x_1\|^2$ we can continue as

$$\le 2e^{-nc(\epsilon/3)} \qquad (11)$$

From (9) for $s = 2$ we see that $\|b\| \sim \|x_2\|$ with big probability so from (11), (11) and (9) we get

$$\mathbb{P}\Big(|\|\mathbf{\Phi}(x)\|^2 - \|x\|^2| > \epsilon\|x\|^2\Big) \leq 6e^{-nc(\epsilon/3)}.$$

So we proved a concentration inequality and thus we infer (see [1]) that $\mathbf{\Phi}$ satisfies RIP(k,δ) for $\delta > 0$ with $k = \lfloor c(\delta)n/\log(n/N)\rfloor$.

4 Norm $\|\cdot\|_J$

In Theorem 2.4 the norm $\|\cdot\|_J$ appears as a measure which estimates the influence of the measurement error on the accuracy of the recovery. Now we want to make some remarks on this norm. First we observe that

$$\|x\|_J \leq \sqrt{\log N}\|x\|_2 \tag{12}$$

so Theorem 2.4 implies for any standard measurement matrix the error estimate

$$\|\Delta(\mathbf{\Phi}(x)+r)-x\|_2 \leq C\big(\sqrt{\log N}\|r\|_2 + \frac{\sigma_k^1(x)}{\sqrt{k}}\big) \tag{13}$$

and also shows that for each $x \in \mathbb{R}^N$ with big probability on the draw of $\mathbf{\Phi}$ we have

$$\|\Delta(\mathbf{\Phi}(x)+r)-x\|_2 \leq C\big(\sqrt{\log N}\|r\|_2 + \sigma_k^2(x)\big). \tag{14}$$

To see (14) we must use that with the big probability on the draw of $\mathbf{\Phi}$ we have $\|\mathbf{\Phi}(x|S^c)\|_J \leq \sigma_k^2(x)$. But this was already proved in [6] Lemma 5.5. The point is that for a random vector $z = \sum_j z_j\eta_j$ where η_j's are independent symmetric random variables with the same subgaussian distribution it is very unlikely that $\|z\|_2 < \|z\|_J$.

But we can make the same observation about the error. If the error is as above e.g. random Gaussian it is exponentially unlikely that $\|z\|_2 < \|z\|_J$. The same is true if our error is uniformly drawn from $\mathcal{S}$, σ is the normalized surface measure on $\mathcal{S}$. The following is well known (see e.g. [12, p.5])

Lemma 4.1 *For* $0 \leq \alpha \leq 1$

$$\sigma\left(\{x \in \mathcal{S} \;:\; x_1 \geq \alpha\}\right) \leq \sqrt{\tfrac{\pi}{2}}e^{-\frac{\alpha^2 n}{2}}. \tag{15}$$

Note that $\{x \in \mathcal{S} \;:\; x_1 \geq \alpha\} \subset \{x \in \mathcal{S} \;:\; x_1 \geq \sin\alpha\}$. Let us denote $I_n = \int_{-\pi/2}^{\pi/2}\cos^n\theta\,d\theta$. From classical formulas for σ we get

$$\sigma\left(\{x \in \mathcal{S} \;:\; x_1 \geq \sin\alpha\}\right) = I_n^{-1}\int_\alpha^{\pi/2}\cos^n\theta\,d\theta.$$

Since $\cos\theta \leq e^{-\theta^2/2}$ for $0 \leq \theta \leq \pi/2$ we get

$$\begin{aligned}\int_\alpha^{\pi/2}\cos^n\theta\,d\theta &= \tfrac{1}{\sqrt{n}}\int_{\alpha\sqrt{n}}^{\sqrt{n}\pi/2}\cos^n(\tfrac{t}{\sqrt{n}})\,dt \\ &\leq \frac{1}{\sqrt{n}}\int_{\alpha\sqrt{n}}^{\sqrt{n}\pi/2}e^{-t^2/2}\,dt\end{aligned}$$

and substituting $u = t - \alpha\sqrt{n}$ we can continue

$$\leq \tfrac{1}{\sqrt{n}}e^{-\alpha^2 n/2}\int_0^\infty e^{-u^2/2}\,du = \sqrt{\tfrac{\pi}{2n}}e^{-\alpha^2 n/2}.$$

To estimate I_n from below we integrate by parts to get $I_n = \frac{n-2}{n-1}I_{n-2} \geq \sqrt{\frac{n-3}{n-1}}I_{n-2}$. From this we get $\sqrt{n}I_n \geq 1$.

Now if $\|z\|_J > \|z\|_2$ then for some coordinate s we have $|x_s| > \frac{1}{\sqrt{\log N}}$. The probability that this happens is at most

$$2n\sigma\big\{x \in \mathcal{S} \,:\, x_1 > \frac{1}{\sqrt{\log N}}\big\} \leq 2n\sqrt{\frac{\pi}{2}}e^{-\frac{n}{2\log N}}.$$

5 New algorithm

Now we want to suggest a modified ℓ_1 minimization algorithm which will have stability properties in ℓ_2 norm. The main point in our improvement is that we do not change the measurement matrix as it may be prescribed by other considerations. So suppose we have a fixed measurement matrix $\mathbf{\Phi} : \mathbb{R}^N \to \mathbb{R}^n$ satisfying RIP(k,δ) for $k = \lfloor cn/\log n/N\rfloor$ and some $\delta > 0$. Given a signal $x \in \mathbb{R}^N$ we observe $y = \mathbf{\Phi}(x)$ or $y = \mathbf{\Phi}(x) + r$. Our algorithm works as follows:

1. We fix at random $n \times N$ Gaussian matrix $\mathbf{\Psi} = \Big(\frac{1}{\sqrt{n}}\gamma_{i,j}\Big)_{i=1,j=1}^{n\quad N}$.

2. We consider $n \times 2N$ matrix $\Gamma = (\mathbf{\Phi}, \mathbf{\Psi})$ and solve the ℓ_1 minimization problem

$$\bar{z} = \text{Argmin}\, z\{z \in \mathbb{R}^{2N} \;:\; \Gamma(z) = y\}$$

3. We define the decoder as $\Delta(y) = \bar{z}|\{1, \ldots, N\}$

If the matrix $\mathbf{\Phi}$ is not fixed but is a random standard measurement matrix we proceed exactly the same but we must make sure that $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are independent.

The main observation is that with big probability on the draw of $\mathbf{\Psi}$ (or with big probability on the joint draw of $\mathbf{\Phi}$ and $\mathbf{\Psi}$) the matrix Γ satisfies RIP and LQ for some constants and $k = \lfloor cn/\log N/n \rfloor$. Since $\|x - \Delta(y)\| \leq \|x - \bar{z}\|$ for every norm we are considering we get:

1. $\|x - \Delta(y)\|_2 \leq C(\|r\|_2 + \frac{\sigma_k^1(x)}{\sqrt{k}})$

2. If $\mathbf{\Phi}$ is random then for every $x \in \mathbb{R}^N$ with big probability we get

$$\|x - \Delta(x)\|_2 \leq C(\|r\|_2 + \sigma_k^2(x)).$$

This is an immediate consequence of Theorem 2.3 once we show that the matrix Γ with big probability satisfies RIP and LQ for some constants and $k = \lfloor cn/\log N/n \rfloor$. But Γ has LQ because $\mathbf{\Psi}$ has. We will show that Γ has RIP$(k, \frac{4}{5}\delta)$. The argument for this is standard so we will provide only a sketch.

• We take k columns from Γ. If all are columns of $\mathbf{\Phi}$ or $\mathbf{\Psi}$ then we are done, so assume we have $k > s > 0$ such columns $(\phi_j)_{j \in A}$ from matrix $\mathbf{\Phi}$ and $k - s$ columns $(\psi_j)_{j \in B}$ from $\mathbf{\Psi}$. There are $\sum_{s=1}^{k-1} \binom{N}{s}\binom{N}{k-s} < N^{k-1}$ such possibilities.

• Let $X = \text{span}(\phi_j)_{j \in A}$ and $Z = \text{span}(\psi_j)_{j \in B}$. In order to show RIP it suffices to show that $|\langle x, z\rangle| \leq \frac{\delta}{10}\|x\|\|z\|$ for $x \in X$ and $z \in Z$.

• It is sufficient to show this for x and z from η nets in the unit balls (with appropriate η. Since (see e.g. [10, Ch. 15 Prop. 1.3.]) the unit ball in d dimensional space has an η net of cardinality not exceeding $(6/\eta)^d$ applying this to X and Z we see that we must to consider $(6/\eta)^k$ pairs x, z.

• A standard estimate shows that $\mathbb{P}(|\langle x, z\rangle| \geq \lambda\|x\|\|z\|) \leq 4e^{-\lambda^2 n/2}$. This is true both when $\mathbf{\Phi}$ is fixed and when $\mathbf{\Phi}$ is random.

• We put those estimates together (see e.g. [1]) and infer that we can find k of the desired magnitude.

References

[1] R.Baraniuk, M.Davenport, R.DeVore, M.Wakin *A simple proof of the restricted isometry property for random matrices* Constructive Approximation (2008) 28(3), pp. 253-263

[2] E.Candes, J.Romberg, T.Tao *Stable signal recovery from incomplete and inaccurate measurements* Comm. Pure Appl. Math. 59 (2006), no. 8, 1207–1223.

[3] E.Candes, T.Tao *Decoding by linear programming* IEEE Trans. Inf. Theory, 51(2005) no. 12, 4203–4215

[4] A.Cohen, W.Dahmen, R.DeVore *Compressed sensing and best k term approximation* J. Amer. Math. Soc. 22 (2009), no. 1, 211–231

[5] A.Cohen, W.Dahmen, R.DeVore, *Instance optimal Decoding by Thresholding in Compressed Sensing* preprint `http://www.dsp.ece.rice.edu/cs/`

[6] R.DeVore, G.Petrova, P.Wojtaszczyk *Instance-optimality in Probability with an ℓ_1-Minimization Decoder*`http://www.dsp.ece.rice.edu/cs/`

[7] D.Donoho *Compressed Sensing* IEEE Trans. Inf. Theory, 52(2006) no. 4, 1289–1306.

[8] B.S. Kashin, *The widths of certain finite-dimensional sets and classes of smooth functions.* (in Russian) Izv. Akad. Nauk SSSR Ser. Mat. 41 (1977), no. 2, 334–351,

[9] A.Litvak, A.Pajor, M.Rudelson, N.Tomczak-Jaegermann *Smallest singular value of random matrices and geometry of random polytopes* Adv. Math. 195(2005) no. 2, 491–523

[10] G.G.Lorentz, M. v.Golitschek, Y.Makovoz, **Constructive Approximation; Advanced Problems**, Springer Verlag 1996

[11] S.Mendelson, A.Pajor, N.Tomczak-Jaegermann , *Uniform uncertainty principle for Bernoulli and subgaussian ensembles*, Constr. Approx. 28 (2008), no. 3, 277–289.

[12] V.D. Milman,G. Schechtman, **Asymptotic Theory of finite Dimensional Normed Spaces**, Lecture Notes in Math. 1200, Springer Verlag, Berlin 1986

[13] D.Needell, R.Vershynin, *Signal Recovery from Inaccurate and Incomplete Measurements via Orthogonal Matching Pursuit*, preprint `http://www.dsp.ece.rice.edu/cs/`

[14] D.Needell, J. Tropp, *CoSaMP: Iterative signal recovery from incomplete and inaccurate samples*, preprint `http://www.dsp.ece.rice.edu/cs/`

[15] J.A.Tropp, A.C. Gilbert *Signal recovery from random measurements via Orthogonal Matching Pursuit*, IEEE Trans. Info. Theory, 53(2007), 1289-1306

[16] P. Wojtaszczyk *Stability and instance optimality for Gaussian measurements in compressed sensing* `http://www.dsp.ece.rice.edu/cs/`