Compressive Sensing Report 2

Group 11

Yang Liu, Dayue Zhao, Chi Wang, Mingyang Yang

***Abstract*—In this report, we continue on our learning on CS on the basis of previous work. We show newest development in RIP, and take a few numerical experiment concerning construction of measurement matrices. Then we extend our work to Robust Compressive Sensing, for this is a more popular model in practise. We give a brief introduction and show the error rate of this type of problem, following by several experiments. Using this model, we also introduce an application in wireless sensor network.**

***Index Terms*—RIP, RIC, Robust Compressive Sensing, Compressive Wireless Sensing, l^p sparse recovery, Bregman Iterative Algorithm.**

I. Introduction

FIRST we introduce the two kinds of CS problems, which are CS and RCS. For some S sparse signal **x**, assume $\mid x_i \mid \leq b$,thus $\parallel x \parallel_1 \leq b \times N = B$, also the signal can be well approximated as x_s by s largest elements. Then we can design a $M \times N$ measurement matrix Ψ to get M projections **y** of the signal. Then the recovery problem is one of l^p norm minimization problems.

1.noiseless recovery: In noiseless conditions, the problem becomes:

$\min_{x \in R^n} \parallel x \parallel_{l_p}$ where $p = 0$, subject to $\Psi x = y$

Under certain constraint on matrix Ψ(also called dictionary), the l_0 problem has a unique s-sparse solution and can be transferred into the l_1 problem:

$\min_{x \in R^n} \parallel x \parallel_{l_1}$, subject to $\Psi x = y$

Since l_1 problem has becomes a popular method for solving sparse signal recovery.

An important characteristic of a dictionary to guarantee recovery is called RIP.For a matrix Ψ the restricted isometry constantRIC δ_k,is defined as the smallest number such that

$$1 - \delta_k \leq \frac{\parallel \Psi x_T \parallel_2}{\parallel x_T \parallel_2} \leq 1 + \delta_k$$

for every vector x and every index set T with $\mid T \mid \leq k$. There are many results on δ_k and k, which we will discuss later.

2.noisy recovery: In nosiy conditions, the vector y is contaminated with noise e,thus $y = \Psi x + w$, the problem becomes:
$\min_{x \in R^n} \parallel x \parallel_{l_1}$, subject to $\parallel \Psi x - y \parallel_2 \leq \varepsilon$ We will

give a error bound later for this type of methods,and illustrate that the log-barrier recovery algorithm works well when SNR is big enough by numerical experiments.

Also, this type of problem is useful in applications since noise is common in many occasions. We will show that in sensor network, this model leads to excellent results.

The report is organized as follows.In section I, we state the basis problem and summarize main results of this report. In section II, we first discuss in detail the RIP condition and introduce recently development on this topic, after which a few matrix construction experiments are shown.In section III,we study another type of CS problem—-Robust Compressive Sensing$RCSforshort$,which involve noise in measuring process. Also, we carry out some experiment using the l1-magic package and introduce Bregman Iterative Algorithm compared with log-barrier algorithm.In section IV, we suggest an application of CS in sensor network using the noise model described in section III.Finally we get some conclusion and show what to do in later work in section V.

II. details on RIP

A. error bound

Since RIP condition has a lot of application versions, we will describe one studied by Candès and Tao:
Theorem 1: If Ψ satisfy RIP of order 3s with $\delta_{3s} < 1$, then

$$\| x^* - x \|_2 \leq C_0 s^{-1/2} \| x_s - x \|_1$$

$x_s$denotes the best S term approximation, x^* denotes solution to the l_1 problem.

Since $\| x_s - x \|_1 \leq \| x \|_1 \leq B$, we can get a bound for recovery error. We can see this result is tight following :

$$E_{s,N}(U(l_1))_{l_2} \leq 2C_0 \sqrt{\frac{log(N/s)+1}{s}}$$

If we take $k \leq C_3 s/(log(N/s)+1)$, then the two inequality will be the same.

Following the above RIP version, we will list several other conditions to guarantee recovery success.It's easy to see there exists tradeoff between the value of δ_k and the size of k.

- $\delta_{3s} < 1$
- $\delta_{2s} < \sqrt{2} - 1$
- $\delta_{2s} + \delta_{3s} < 1$
- $\delta_{3s} + \delta_{4s} < 2$

B. dictionary construction

As we mentioned in earlier report,random matrix for dictionary construction performs well currently, we will list several conclusions that certain types of matrix will follow RIP the above condition with overwhelming probability.

- *Gaussian measurements*: Every element of the dictionary obey $\Psi_(i,j) \sim N(0, 1/M)$,then if $s \leq CM/log(N/M)$, s will obey the RIP condition with probability $1 - O(e^{-\gamma N})$.
- *Binary measurements*:Suppose the elements obey Bernoulli distribution $P(\Psi_{i,j} = \pm 1/\sqrt{M}) = 1/2$.Then RIP is satisfied with probability $1 - O(e^{-\gamma N})$,s should also obey $s \leq CM/log(N/M)$.
- *Fourier measurements*:Suppose now that Psi is a partial Fourier matrix obtained by selecting M rows uniformly at random,and renormalizing the columns so that they are unit-normed. Then RIP is satisfied with overwhelming probability if$s \leq CM/log(N)^4$.

C. l_p recovery failure

From the RIP condition, we can see the larger δ_k is, the looser the constraint is on the measurement matrix, then it's important how much larger the RIC could we expect to guarantee l^1 recovery of any s-sparse vector. We will take k = 2s for illustration, then comes the next theorem:
Theorem 2:For any $\epsilon > 0$, there exists an integer

s and a dictionary Ψ with a RIC $\delta_{2s} \leq 1/\sqrt{2} + \epsilon$ for which l^1 fails on some s-sparse vector.
The full proof is given in[26]. This theorem shows

that near the value of 0.7, RIP may fails to do l^1 recovery.In the next figure, we can see that below the safe value of $2(3-\sqrt{2})/7 \approx 0.4531$, all l_p recovery is well performed, however above diognal line, the recovery will fail, despite the gap between the two areas. This figure may suggest there still can be some benefit in $p \ll 1$ to improve recovery. However, considering the practical aspects of l_p recovery, which use a reweighted l_1 optimization problems as the kernel. It's proved in [25], even l_p recovery $p \ll 1$ is not realizable in practise.

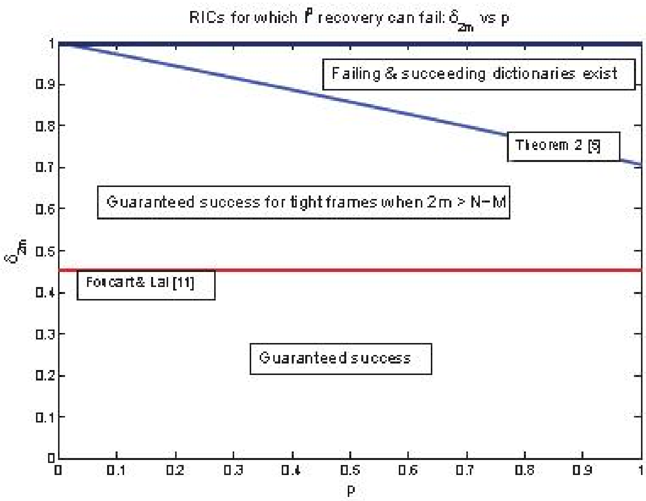

Fig. 1. A summary of known results relating RIC to l^p recovery

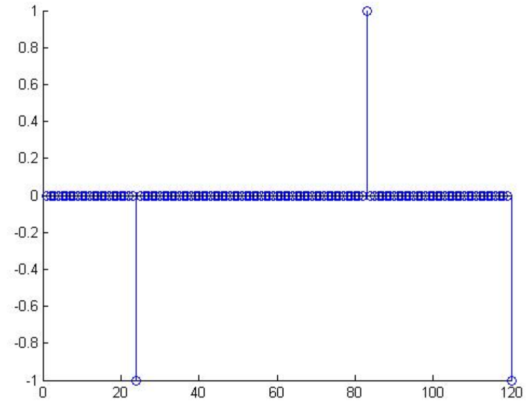

Fig. 2. The original signal

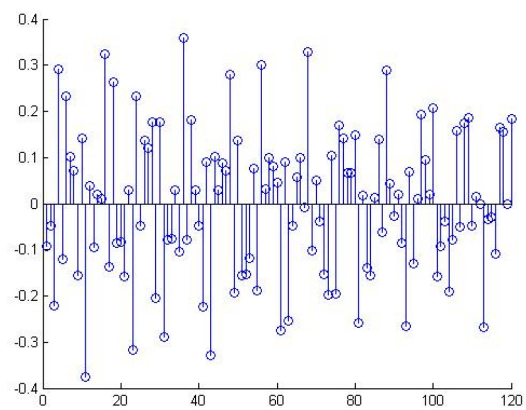

Fig. 3. The DCT spectrum for the signal

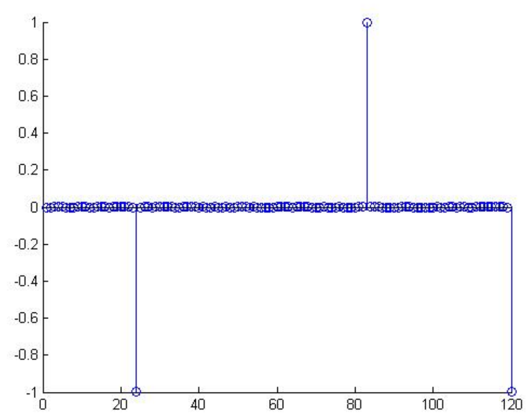

Fig. 4. Signal recovered. MSE = 2.34e-6

Using RIP to characterize matrices has been very useful in understanding when l^1 recovery can be achieved. However, to further explore the possibility to better recover sparse signal needs other more refined tools, while in Robust CS, RIP still plays important roles in signal recovery.

D. *experiment*

As we have performed experiments on matrix construction in the case of Gaussian Random Matrix and shifted PN-sequence matrix in former report, we will concentrate on the random row-selecting methods in DFT or DCT basis still using 1D signal for illustration.In this experiment we try 3 signals, one random 3-sparse, one 10-sparse square signal, and the other, one delta signal. We will see that why in general, totally random matrix is better than either fixed-pattern ones or random row-selecting ones. First, we create a real-valued discrete signal with length N=120 and have S=3 spikes, then we create measurement matrix Θ with a size of 40×120 by uniformly selecting rows of the DCT matrix. The reconstruction algorithm is primal-dual in Basic Pursuit.As we can see in figure 3, since DCT domain signal will be dense enough, the recovery will be successful for granted. This illustrate random row-selecting method is efficient for sparse signal recovery.

However, when the DCT spectrum is of some fixed patterns, this method will be unstable. Now we consider a delta signal, as we can see in figure 6, the energy will concentrate on low frequency parts. Thus we take a fixed- pattern matrix for comparision,which capture the lowest 40 rows of the DCT matrix. Since the spectrum is so special that measurement capture the biggest 40 terms in spectrum. Signal recovered is highly close to original signal.While using random row-selecting method,the signal recovered is although good, not ideal.

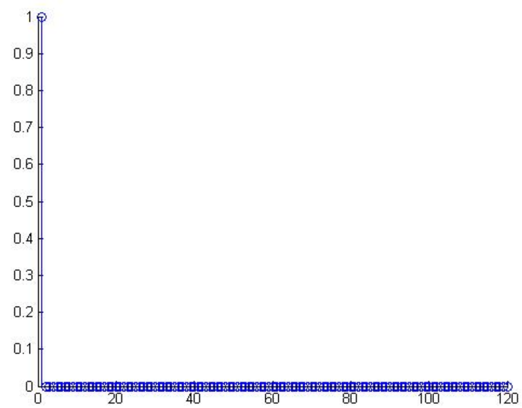

Fig. 5. The original signal

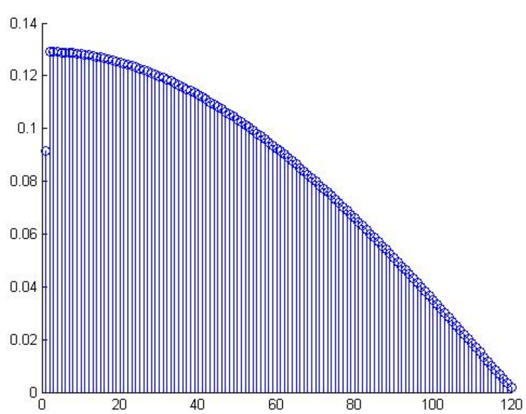

Fig. 6. The DCT spectrum for the signal

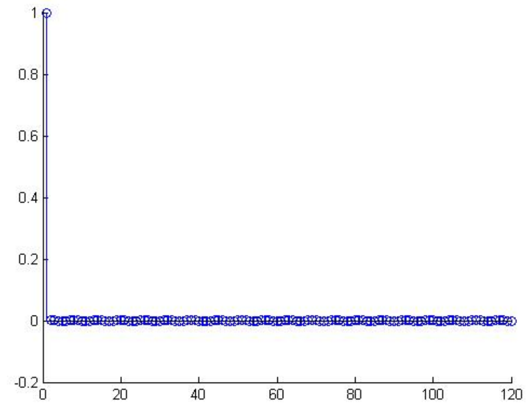

Fig. 7. Signal recovered using low freuency pattern . MSE = 1.6326e-6

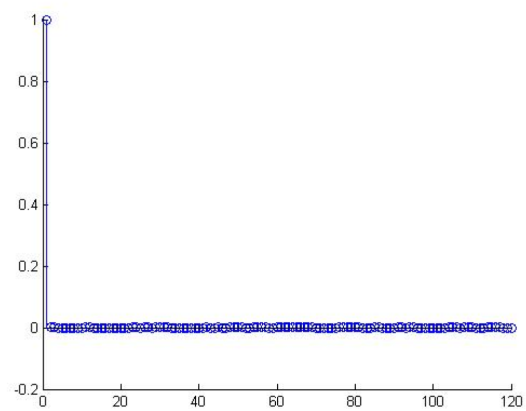

Fig. 8. Signal recovered using random row-selecting . MSE = 4.4739e-6

Then we consider a 10-sparse square signal, this time the spectrum is distributed and fixed pattern method totally failed. The random row-selecting method will be better,but worse than random matrix methods.

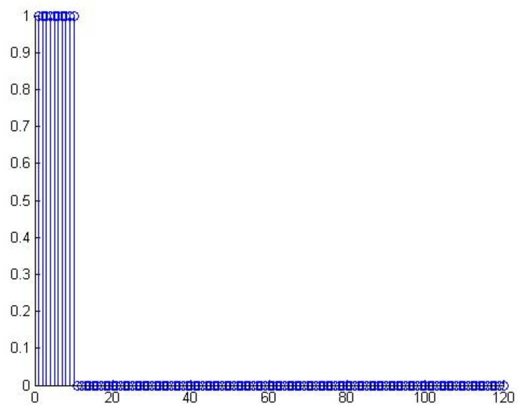

Fig. 9. The original signal

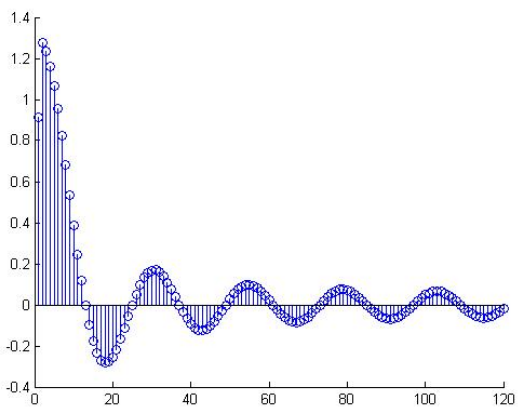

Fig. 10. The DCT spectrum for the signal

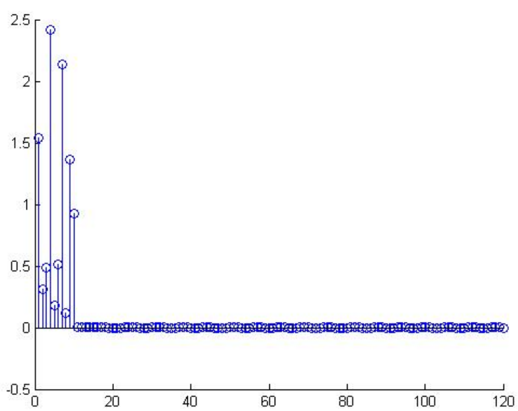

Fig. 11. Signal recovered using low freuency pattern . MSE = 2.4798

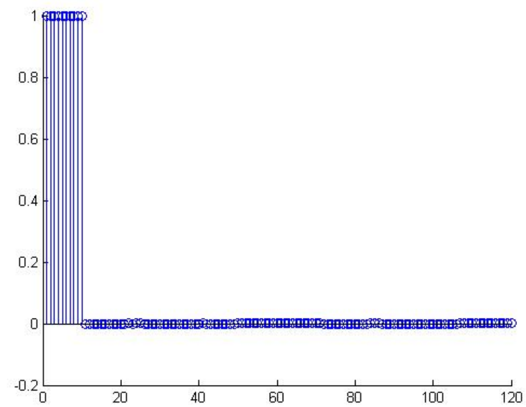

Fig. 12. Signal recovered using random row-selecting . MSE = 2.9065e-4

Then we can conclude that in general cases, pure random matrices are better than random row-selecting ones, even better than fixed pattern ones.

III. Robust Compressive Sensing

A. error bound

When the measurement is corrupted with noise, We observe

$$y = \Psi x + w$$

As mentioned in section I, we can put the problem as follows:
$\min_{x \in R^n} \parallel x \parallel_{l_1}$, subject to $\parallel \Psi x - y \parallel_2 \leq \varepsilon$

Candè[4] prove one error bound for this problem:
Theorem 1:Assume that $\delta_2 s < \sqrt{2} - 1$ and $\parallel w \parallel_2 \leq \epsilon$. Then the solution x^* obeys:

$$\parallel x^* - x \parallel_2 \leq C_0 s^{-1/2} \parallel x_s - x \parallel_1 + C_1 \epsilon$$

This result is to bound the error using $\parallel x_s - x \parallel_1$, if we choose $\parallel x_s - x \parallel_2 /N$ as the bound, then comes the following results. First, we difine the α compressible signal if:

$$\frac{\parallel x^* - x_s \parallel_2}{N} = O(s^{-2\alpha})$$

If x is α compressible T such that $x = T\theta$, then the optimization problem becomes:

$$\theta^* = argmin_{\theta\in\Theta}\{\| y - \Psi T\theta \|_2{}^2 + \frac{2log(2)log(N)}{\epsilon} \| \theta \|_0$$

If we define the distortion as $D = E(\frac{\|x^* - x\|_2}{N})$ then:

$$D = O((\frac{M}{log(N)})^{-2\alpha/(-2\alpha+1)})$$

If the signal is truly sparse that α is big enough (only has m non-zero elements on T basis),then

$$D = O((\frac{M}{mlog(N)})^{-1})$$

We will illustrate an application in sensor network later using this type of model.

B. experiment

In this part, we try some experiment using the l1magic package. First, we create a real-valued discrete signal with length N=120 and have S=10 spikes, then we create measurement matrix iid gaussian matrix Θ with a size of 40×120 . The reconstruction algorithm is log-barrier in Basic Pursuit. Then we corrupt the measurement by adding noise whose magnetitude $< \sigma$,We use different SNR levels for experiments.

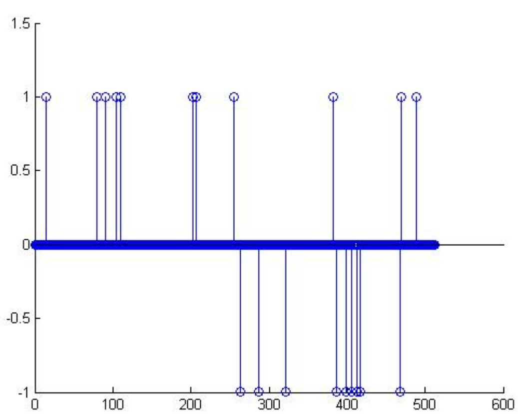

Fig. 13. The original signal,l1=20

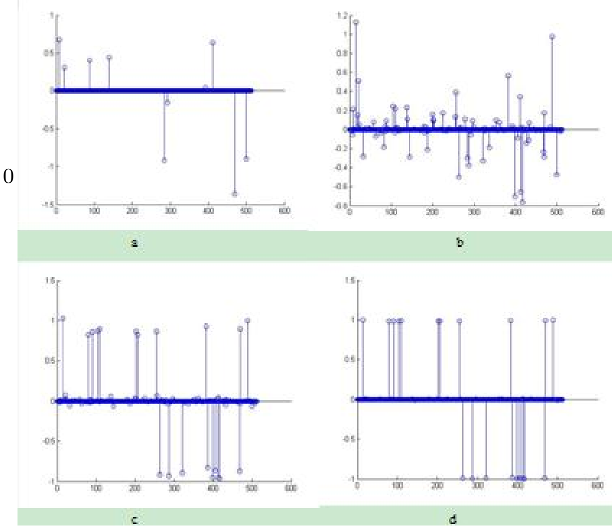

Fig. 14. Signal recovered under different conditions: (a)SNR=1,l1=5.859,MSE=4.9857 (b)SNR=0.1,l1=13.587, MSE=3.4428 (c)SNR=0.001,l1=19.940,MSE=0.0546 (d)SNR=0.0001,l1=19.995,MSE=0.0056

As we can see as the SNR increases, the recovery accuracy increases. However when the measurement is noiseless, this algorithm will fail. An alternative will be the so-called Bregman Iterative Algorithm which can be used as a method for noiseless as well as noisy conditions. We will focus on this algorithm in later work.

IV. Compressive Wireless Sensing

In this section, we first introduce the basic system structure of the sensor network, then we introduce how CWS changes the system equality.

Consider a wireless sensor network with n nodes where each node takes a noisy sample of the form:

$$x_j = x^* + w_j, j = 1, 2...n$$

the errors w_j are iid gaussian variables such that $w_j \sim N(0, \delta_w{}^2)$ and the signal satisfy $x_j^* \leq B$. In the next figure, we show the basic communication structure of SNW.

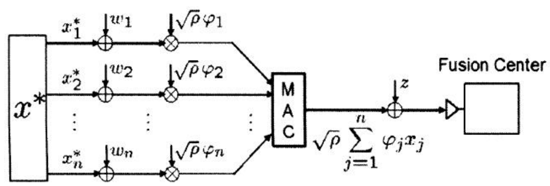

Fig. 15. The original signal,l1=20

Each node multiplies its measurement x_j with $(\sqrt{\rho}\psi_j)$ to obtain $m_j = \sqrt{\rho}\psi_j x_j$,where ρ is a scaling factor used to satisfy sensor's transmission power constraints P, and the nodes coherent transmit their m_j over network-to-FC channel.
The signal is then transformed into AWGN MAC channel and received signal at the FC is given by

$$r = \sqrt{\rho}\Psi_T(x^* + w) + z = \sqrt{\rho}(v + \tilde{w}) + z$$

where $z \sim N(0, {\delta_z}^2)$ is the channel noise and $\tilde{w} \sim N(0, {\delta_w}^2)$.If we get k projections of the signal, then the problem become one RCS problem.

By carefully selecting the parameters, we can get the following two results concerning the Power-Distortion-Latency Trade-offs:

- If there is enough prior about the signal structure and measurement basis.We have to get $k = n^{\frac{1}{2\alpha+1}}$ projections. We can get:

$$D \sim O(n^{\frac{-2\alpha}{2\alpha+1}})$$

$$L \sim O(n^{\frac{1}{2\alpha+1}})$$

$$P_{tot} \sim O(n^{\frac{1}{2\alpha+1}})$$

  where $P_{tot} = \sum_{i=1}^{k} P_{v_i}$,and $P_{v_i} = \sum_{i=1}^{n} P_{i,j} \sim O(1)$
  Then we can get the trade-off:

$$D \sim P_{tot}^{-2\alpha} \sim L^{-2\alpha}$$

- If no prior of the signal is known, random measurement matrices are needed, and we can get:

$$D = O((\frac{k}{log(n)})^{-2\alpha/(-2\alpha+1)})$$

  ignoring the log(n) term, we can get the trade-off:

$$D \sim P_{tot}^{-2\alpha/(2\alpha+1)} \sim L^{-2\alpha/(2\alpha+1)}$$

If the signal is truly sparse, then:

$$D \sim P_{tot}^{-1} \sim L^{-1}$$

V. Conclusions

In this report, we discuss extensively on the topic of RIP including 1.different versions of RIP 2.matrix size constraint for several types of dictionary 3.known results on l_p recovery failure. We also introduce a new type of CS problem we've learned – RCS. Then we give an application in wireless sensor network. We shall focus on algorithm analysis, especially for BIA algorithm in later work and study into details about different noise sources in RCS problems.Also we may try to learn other evaluation methods for matrix other than RIP.

References

[1] Richard G. Baraniuk, *Compressive Sensing.*

[2] Emmanuel Candes Justin Romberg and Terence Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. on Information Theory, 52(2) pp. 489 - 509, February 2006.

[3] David L. Donoho, *Compressive Sensing*, IEEE Trans. on Information Theory, 52(4), pp. 1289 - 1306, April 2006.

[4] Emamnuel J. Candes, *Compressive Sampling.*

[5] Emmanuel J. Cands and Michael B. Wakin, *An introduction to Compressive Sampling.*

[6] David Donoho and Yaakov Tsaig, *Extensions of compressed sensing*, Signal Processing, 86(3), pp. 533-548, March 2006.

[7] Richard Baraniuk and Philippe Steeghs, *Compressive Radar Imaging*, Department of Electrical and Computer Engineering Rice University, E. P. Wigner Institute Bucharest, Romania.

[8] Justin Romberg, *Imaging via compressive sampling*, IEEE Signal Processing Magazine, 25(2), pp. 14 - 20, March 2008.

[9] Piotr Indyk, *Explicit constructions for compressed sensing of sparse signals*, Symp. on Discrete Algorithms, 2008.

[10] P. Wojtaszczyk, *Stability and instance optimality for Gaussian measurements in compressed sensing*, Preprint, 2008.

[11] Florian Sebert, Leslie Ying, and Yi Ming Zou, *Toeplitz block matrices in compressed sensing*, Preprint, 2008.

[12] Sina Jafarpour, Weiyu Xu, Babak Hassibi, and Robert Calderbank, *Efficient compressed sensing using high-quality expander graphs*, Preprint, 2008.

[13] Emmanuel Cands, *The restricted isometry property and its implications for compressed sensing*, Compte Rendus de l'Academie des Sciences, Paris, Series I, 346, pp. 589-592, 2008.

[14] Shamgar Gurevich and Ronny Hadani, *Incoherent dictionaries and the statistical restricted isometry property*, Preprint, 2008.

[15] Jarvis Haupt, Waheed U. Bajwa, Gil Raz, and Robert Nowak, *Toeplitz compressed sensing matrices with applications to sparse channel estimation*, Preprint, 2008.

[16] Gabriel Peyr, *Best basis compressed sensing*, Preprint, 2006.

[17] Lawrence Carin, Dehong Liu, and Ya Xue, *In Situ Compressive Sensing*, Inverse Problems, 24(1), Feb. 2008.

[18] Petros Boufounos and Richard G. Baraniuk, *1-Bit compressive sensing*, Conf. on Info. Sciences and Systems (CISS), Princeton, New Jersey, March 2008.
[19] Lawrence Carin, Dehong Liu, and Bin Guo, *In situ compressive sensing for multi-static scattering: Imaging and the restricted isometry property*, Preprint, 2008.
[20] Giuseppe Valenzise, Giorgio Prandi, Mario Tagliasacchi, and Augusto Sarti, *Identification of sparse audio tampering using distributed source coding and compressive sensing techniques*, Preprint, 2008.
[21] Waheed Bajwa, Jarvis Haupt, Akbar Sayeed, and Robert Nowak, *Compressive Wireless Sensing*, Preprint, 2008.
[22] Jarvis Haupt and Robert Nowak, *Signal Reconstruction From Noisy Random Projections*, IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 52, NO. 9, SEPTEMBER 2006.
[23] Wotao Yin,Stanley Osher,Donald Goldfarb,and Jerome Darbon, *Bregman Iterative Algorithms for l_1-Minimization with Applications to Compressed Sensing*, SIAM J. IMAGING SCIENCES Vol. 1, No. 1, pp. 143C168.
[24] Richard Baraniuk ,Mark Davenport ,Ronald DeVore ,MichaelWakin, *A Simple Proof of the Restricted Isometry Property for Random Matrices*, Constr Approx (2008) 28: 253C263DOI 10.1007/s00365-007-9003-x
[25] Mike Davies,Remi Gribonval, *The Restricted Isometry Property and l^p sparse recovery failure*, Author manuscript,2009
[26] Michael Evan Davies, and Remi Gribonval, *Restricted Isometry Constants Where l^p Sparse Recovery Can Fail for* $0 < p \leq 1$, IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 55, NO. 5, MAY 2009