

# Improved Bounds for Sub-linear Time Compressed Sensing for <sup>1</sup>

## Support Recovery using Sparse-Graph Codes

Krishna R. Narayanan

Dept. of Elect. & Comp. Engg.,

Texas A&M University,

College Station, TX, U.S.A

### I. INTRODUCTION

In [1] (and in the expanded version in [2]), Li, Pawar and Ramchandran have considered the problem of recovering the support of a  $K$ -sparse,  $N$ -dimensional signal from  $M$  linear measurements in the presence of noise. Based on sparse-graph codes and a peeling decoder, they have proposed an elegant design of the measurement matrix and a recovery algorithm that is nearly optimal in terms of measurement complexity and computational complexity. In particular, they have proposed two designs - the first design requires  $M = O(K \log N)$  measurements and a near-linear ( $O(N \log N)$ ) decoding complexity, whereas the second design requires  $M = O(K \log^{1.3} N)$  measurements with a sub-linear decoding complexity.

Specifically, the following theorems are given by Li, Pawar and Ramchandran in [2]

*Theorem 1 ([2] Noiseless recovery):* In the absence of noise, given any  $K$ -sparse signal  $x$  with  $x[k] \in \mathcal{X}$  for  $k \in \text{supp}(x)$ , their noiseless recovery schemes achieve a vanishing failure probability  $\mathbb{P}_F(O(\frac{1}{K})) \rightarrow 0$  asymptotically in  $K$  and  $N$  with

	Measurement cost $M$	Computational Complexity $T$
Fourier noisy	$2(1 + \epsilon)K$	$O(K)$
Binary noisy	$(1 + \epsilon)K(\log_2 N + 1)$	$O(K \log N)$

*Theorem 2 ([2] Sub-linear Time Noisy Recovery):* . In the presence of i.i.d. Gaussian noise with zero mean and variance  $\sigma^2$ , given any  $K$ -sparse signal  $x$  with  $x[k] \in \mathcal{X}$  for  $k \in \text{supp}(x)$ , our noiseless recovery schemes achieve a vanishing failure probability  $\mathbb{P}_F \rightarrow 0$  asymptotically in  $K$  and  $N$  with

	Measurement cost $M$	Computational Complexity $T$
Fourier noisy	$O(K \log^{1.3} N)$	$O(K \log^{1.3} N)$
Binary noisy	$O(K \log N)$	$O(K \log N)$

*Theorem 3 ([2] Near-linear Time Noisy Recovery):* . In the presence of i.i.d. Gaussian noise with zero mean and variance  $\sigma^2$ , given any  $K$ -sparse signal  $x$  with  $x[k] \in \mathcal{X}$  for  $k \in \text{supp}(x)$ , the RandomNoisy scheme achieves a vanishing failure probability  $\mathbb{P}_F \rightarrow 0$  asymptotically in  $K$  and  $N$  with a measurement complexity of  $M = O(K \log N)$  and computational complexity of  $T = O(N \log N)$ .

## II. MAIN RESULTS

We believe that Theorem 2 and Theorem 3 can be sharpened to the following new theorems

*Theorem 4 (Sub-linear Time Noisy Recovery):* . In the presence of i.i.d. Gaussian noise with zero mean and variance  $\sigma^2$ , given any  $K$ -sparse signal  $x$  with  $x[k] \in \mathcal{X}$  for  $k \in \text{supp}(x)$ , our noisy recovery schemes achieve a vanishing failure probability  $\mathbb{P}_F \rightarrow 0$  asymptotically in  $K$  and  $N$  with

	Measurement cost $M$	Computational Complexity $T$
Fourier noisy	$O\left(K \log^{1.3} \frac{N}{K}\right)$	$O\left(K \log^{1.3} \frac{N}{K}\right)$
Binary noisy	$O\left(K \log \frac{N}{K}\right)$	$O\left(K \log \frac{N}{K}\right)$

*Theorem 5 (Near-linear Time Noisy Recovery):* . In the presence of i.i.d. Gaussian noise with zero mean and variance  $\sigma^2$ , given any  $K$ -sparse signal  $x$  with  $x[k] \in \mathcal{X}$  for  $k \in \text{supp}(x)$ , the RandomNoisy scheme achieves a vanishing failure probability  $\mathbb{P}_F \rightarrow 0$  asymptotically in  $K$  and  $N$  with a measurement complexity of  $M = O\left(K \log \frac{N}{K}\right)$  and computational complexity of  $T = O\left(N \log \frac{N}{K}\right)$ .

### A. Main Idea

The main difference between [2] and our approach is that we replace the left  $d$ -regular ensemble of graphs  $\mathcal{G}_{\text{reg}}^N(R, d)$  in [2] by  $(d, \frac{dN}{R})$  left and right regular ensemble with fixed degree. This allows us to use a bin detection matrix with  $P' = O\left(\log_2 \left(\frac{N}{K}\right)\right)$  rows instead of using a bin detection matrix with  $P = O(\log_2 N)$  rows.

*Definition 6 (Left and right regular graph ensemble):* Let  $\mathcal{G}_{\text{reg,reg}}^N(R, d, \frac{dN}{R})$  denote the ensemble of graphs with  $N$  left nodes and  $R$  right nodes, where each left node  $k \in [N]$  is connected to  $d$  right nodes and each right node  $j \in [R]$  is connected to  $\frac{dN}{R}$  left nodes.

The number of observations  $R$  will be chosen to be  $R = \eta K = O(K)$ . A matrix  $\mathbf{H}$  is chosen at random from this ensemble and used as the coding matrix. A bin detection matrix  $\mathbf{S}$  is chosen similar to that in [2] however now the bin detection matrix has only  $P' = \log_2 \left(\frac{N}{R}\right)$  rows. Since  $R = O(K)$ , this implies that  $P' = O(N/K)$ . The probability of error for the bin detection algorithm can be analyzed exactly as in [2] with the new  $P'$  and shown to have exactly the same behavior in  $K$ . However, there will be a penalty to pay in terms of the minimum SNR required and this is discussed next.

### B. Minimum SNR required

We will take the RandomNoisy case as an example and let us consider the case when  $K = O(N^\delta)$  for some  $0 < \delta < 1$ . In the derivation of the probability of error in [2] in Proposition 3 and Proposition 4 in Appendix D, it can be seen that the probability of error includes terms of the form

$$e^{-\frac{P}{4} \frac{\gamma^2}{1+4\gamma}} + 2e^{-c_6 P \left(1 - \frac{\gamma \sigma^2}{A_{\min}^2}\right)}$$

Right after Proposition 5, it is mentioned that by choosing  $P = O(\log N)$ , the probability of error can be made to decay as  $O\left(\frac{1}{N^3}\right) < O\left(\frac{1}{K^3}\right)$ . First, it should be noted that when  $P = O(\log N)$ , in order for the probability of error to decay as  $O\left(\frac{1}{K^3}\right)$ , the following two conditions must be satisfied on  $\gamma$

$$\frac{\gamma^2}{4(1+4\gamma)} > 2\delta, \quad \text{and} \quad c_6 \left(1 - \frac{\gamma \sigma^2}{A_{\min}^2}\right) > 2\delta$$

Hence, for the overall probability of error to decay as  $O(1/K^3)$ ,  $A_{\min}^2/\sigma^2$  should be large enough such that  $0 < 2\delta < \gamma < A_{\min}^2/\sigma^2$ . Hence, the theorems appear to be valid only for  $A_{\min}^2/\sigma^2 > SNR_{\text{Th}}$ .

### C. SNR versus Measurements

It should be noted that if the objective is to get the probability of error to decay only as  $O(1/K^3)$ , it is not required for the probability of error to decay as  $O(1/N^3)$  and hence, we could directly choose  $P = O(N/K) = N^{1-\delta}$  or in fact,  $P = O(N^\beta)$  for any  $\beta$  and that there will be a corresponding  $SNR_{\text{Th}}$ . Indeed, this appears to be reasonable since the assumption that the non-zero values of the signals are from a finite set and hence, when the  $\sigma^2 \rightarrow 0$ , the number of samples required should smoothly go to  $O(K)$ . Hence, it appears that the order of measurements required should be a function of SNR.

Based on this, indeed, it appears that when using RandomNoisy measurement matrix, it is not even required to switch to the  $\mathcal{G}_{\text{reg-reg}}$  ensemble and we could have simply chosen  $P = O(N^\beta)$ . However, if we want to choose the Random DFT ensemble (see Page 22 [2]), then we have to switch to the  $\mathcal{G}_{\text{reg-reg}}$ . It appears convenient to switch to this ensemble for all results

## III. PROOFS

We have to prove that we can use the  $\mathcal{G}_{\text{reg,reg}}^N(R, d, \frac{dN}{R})$  for optimal peeling. This indeed appears to be straightforward. In the next section we consider a  $\mathcal{G}(R, d, \frac{dN}{R})$  ensemble and show that this ensemble will provide a probability of failure of at most  $O(\frac{1}{K})$ . Although it appears simple to use a spatially-coupled code ensemble and use existing results to show this there are two main obstacles to this:

- In traditional LDPC codes and peeling decoder over binary erasure channel, the input to the decoder is the channel output corresponding to  $N$  bit nodes and the check nodes on the right are mere parity checks whose sum modulo 2 is zero. Whereas in this problem the values corresponding to the  $N$  bit nodes on the left need to be evaluated by the decoder given the values corresponding to the  $R$  check nodes (the real sum of the bit nodes connected) on the right node are non-zero and are given as input to the decoder.
- In traditional LDPC case a constant fraction  $\epsilon$  of these  $N$  bit nodes are erased by the channel and usually the emphasis is on analyzing the performance of peeling decoder asymptotically in  $N$  or  $R$  when rate  $= 1 - \frac{R}{N}$  is fixed. But in our case the fraction of the nodes erased  $= 1 - \frac{K}{N}$ , where  $K$  is usually of the form  $K = N^\delta$ , tend to one and the rate of the code  $= 1 - \frac{R}{N} = 1 - \frac{\eta N^\delta}{N}$  tend to one asymptotically in  $N$ .

Consider a left and right regular LDPC code  $\mathcal{G}_{\text{LDPC}}(N, l, r)$  where  $N$  is the number of variable nodes on the left and  $l, r$  are the regular left and right degrees respectively. Let  $P_{\text{BEC}}^{(i)}(\mathbf{y})$  be the degree distribution of the number of check nodes after iteration  $i$  of peeling decoder given  $\mathbf{y}$  is the channel output. And similarly  $\mathcal{G}_{\text{reg-reg}}(N, l, r)$  be the graph corresponding to the parity check matrix in compressed sensing problem and  $P_{\text{CS}}^{(i)}(\mathbf{z})$  be the degree distribution of the check nodes the right after iteration  $i$  of the oracle-based peeling decoder, given  $\mathbf{z}$  is the CS equivalent of syndrome corresponding to  $\mathbf{x}$  i.e.,  $\mathbf{z} = \mathbf{H}\mathbf{x}$ . Note that  $\mathbf{y}$  is a vector of dimension  $N$  whereas  $\mathbf{z}$  is of dimension  $\frac{Nl}{r}$ .

Note that in the peeling decoder we consider, we remove one degree-1 check node and the bit node connected to it from the graph in each iteration. In the following theorems we will show that we can indeed use a left and right regular ensemble and show similar results.

*Lemma 7 (Equivalence to LDPC-BEC):* Whenever  $\mathbf{y}$  and  $\mathbf{z}$  satisfy

$$\mathbf{z} = \mathbf{H}\mathbf{x} \text{ such that } |\text{supp}(\mathbf{x})| = |\{i : y_i = \mathcal{E}\}| \quad \text{where } \mathcal{E} \text{ denotes erasure,}$$

then  $P_{\text{BEC}}^{(i)}(\mathbf{y}) = P_{\text{CS}}^{(i)}(\mathbf{z}) \quad \forall i$ .

The above lemma gives us the equivalence of peeling decoder of LDPC codes on BEC and oracle based peeling decoder of compressed sensing problem. Thus by considering an BEC of erasure probability  $\frac{K}{N}$  we can equivalently consider peeling decoder of LDPC codes on BEC channel and use a lot of the existing results.

*Lemma 8:* The evolution of the left and right degree distribution as the peeling decoder progresses can be given by

$$\begin{aligned} \tilde{L}_d(y) &= \epsilon y^i, \\ \tilde{R}_0(y) &= 1 - \sum_{j \geq 1} \tilde{R}_j(y), \\ \tilde{R}_1(y) &= R'(1) \epsilon y^{d-1} [y - 1 + (1 - \epsilon y^{d-1})^{r-1}] \\ \tilde{R}_j(y) &= \sum_{j \geq 2} R_j \binom{j}{i} (\epsilon y^{d-1})^i (1 - \epsilon y^{d-1})^{j-1}, \quad i \geq 2 \end{aligned}$$

where  $\epsilon = \frac{K}{N}$  and  $r = \frac{Nd}{\eta K}$ .

*Proof:* In fact these are exact results for an LDPC code on BEC channel for a given  $N, R$  and  $\epsilon$ . We just restated the result. ■

*Definition 9 (BP Threshold):* We define the BP threshold,  $\eta^{\text{BP}}$  to be the value of  $\eta$  for which there is a unique non-zero solution for the equation:

$$\begin{aligned} y &= \lim_{N \rightarrow \infty} 1 - \left( 1 - \frac{y^{d-1}}{N^{1-\delta}} \right)^{\frac{dN^{1-\delta}}{\eta}} \\ &= 1 - e^{-\frac{dy^{d-1}}{\eta}} \end{aligned}$$

*Lemma 10:* If  $\eta > \eta^{\text{BP}}$  then with probability at least  $1 - O\left(N^{1/6} e^{-\frac{\sqrt{N(d-1)}}{(dr)^3}}\right)$  the peeling decoder of a specific instance progresses until the number of residual nodes in the graph has reached size  $\gamma N$  where  $\gamma$  is an arbitrary positive constant.

*Theorem 11:* Consider the ensemble  $\mathcal{G}_{\text{reg-reg}}(N, l, \frac{Nl}{\eta K})$ , the oracle based peeling decoder peels off all the edges in the pruned graph in  $O(K)$  iterations with probability at least  $1 - O(1/K)$ .

*Proof:* Here is the brief outline of the proof. We first show that expressions for the evolution of degree distributions from LDPC-BEC are valid here as shown in Lemma. 8. Then Lemma. 10 shows us that the peeling decoder fails to peel off till the residual graph has  $\gamma N$  variable nodes remaining with an exponentially low probability. Then similar to [1] we show that the left and right regular graphs are good expander graphs with probability  $O(\frac{1}{K})$  and hence the remaining  $\gamma N$  nodes can be peeled off. ■

#### IV. EXTENSIONS

For the group testing problem addressed in [3] also, it appears we can replace the ensemble used by Lee, Pedarsani and Ramchandran by the regular-regular ensemble and obtain a sharper result, i.e.,  $O(K \log(N/K))$  measurements for the case of  $K = N^\delta$ , where  $\delta < 2/3$ .

When the right regular ensemble is considered, then the probability of error at each measurement node will be  $O\left(\frac{K^2}{N^2}\right)$  and there are  $K$  stages in the peeling process. Thus, for the overall error probability to decay with  $K$ , we require that

$$\frac{K^3}{N^2} < 1 \Rightarrow \frac{N^{3\delta}}{N^2} < 1 \Rightarrow \delta < \frac{2}{3}.$$

The main idea in the proof is to look at the reduced subgraph induced only by the defective items and when  $N/K \rightarrow \infty$ , this graph will be a graph from the left regular, right Poisson ensemble considered in [3].

#### REFERENCES

- [1] X. Li, S. Pawar, and K. Ramchandran, "Sub-linear time compressed sensing using sparse-graph codes," in *Proc. IEEE Symposium on Information Theory*, 2015, pp. 1645–1649.
- [2] —, "Sub-linear time compressed sensing using sparse-graph codes," [http://www.eecs.berkeley.edu/~xiaoli/TR\\_CS\\_sublinear.pdf](http://www.eecs.berkeley.edu/~xiaoli/TR_CS_sublinear.pdf).
- [3] K. Lee, R. Pedarsani, and K. Ramchandran, "Saffron: A fast, efficient, and robust framework for group testing based on sparse-graph codes," *arXiv preprint arXiv:1508.04485*, 2015.