Sub-linear time Compressed Sensing and Group Testing via sparse-graph codes

Avinash Vem, Nagaraj T. Janakiraman, Krishna R. Narayanan
ECE Dept.,Texas A&M University

Compressed Sensing(CS)

Recover a sparse signal $\mathbf{x}$ from $\mathbf{y}$: $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$

- $\mathbf{x}$ -$N \times 1$ sparse signal
- $\mathbf{A}$ -$M \times N$ measurement matrix
- $\mathbf{w}$ -additive Gaussian noise
- $\mathbf{y}$ -$M \times 1$ measurement vector
- $|\{i : x_i \neq 0\}| = K$. $K << N$

Metric of interest:

- Prob. of failure of support recovery $\mathbb{P}_F := \Pr(\mathrm{supp}(\hat{\mathbf{x}}) \neq \mathrm{supp}(\mathbf{x}))$

Group Testing(GT)

Recover a sparse signal $\mathbf{x}$ from $\mathbf{y}$: $\mathbf{y} = \mathbf{A} \odot \mathbf{x}$

- $\mathbf{x}, \mathbf{y}, \mathbf{A}$ are binary vectors/matrix respectively
- $\odot$: Matrix multiplication with **"binary** OR" instead of addition
- $\mathbf{y}$- $M \times 1$ measurement vector

Metric of interest:

- Prob. of missing defective item: $\mathbb{P}_m := \Pr(\hat{x}_i = 0, x_i = 1)$

Known Bounds for CS

- In 2007, Wainwright gave information theoretic limits for compressed sensing: support recovery
- For sub-linear sparsity, $K = o(N)$, $M = \Theta\left(K \log(\frac{N}{K})\right)$ is shown to be necessary and sufficient.
- In the linear sparse regime, $K = \alpha N$, it was shown that $M = \Theta(N)$ measurements are sufficient. [1]

Main Result (CS)

Sub-linear sparsity: For a given SNR, our scheme has

- Sample complexity of $M = c_1\ K \log(\frac{c_2 N}{K})$
- Decoding complexity of $O\left(K \log(\frac{N}{K})\right)$
- $\mathbb{P}_F \to 0$ asymptotically in K and N

Linear sparsity: Our scheme has

- Sample complexity of $M = c_3 K \log K$
- Decoding complexity of $O\left(K \log(\frac{N}{K})\right)$
- $\mathbb{P}_F \to 0$ asymptotically in K and N

Bounds for Group Testing

- We assume all the $\binom{N}{K}$ K-sparse sets are equi-probable
- At least $\log_2\binom{N}{K}$ tests are necessary
- For large K and N, $\log_2\binom{N}{K} \approx K \log_2(\frac{N}{K})$

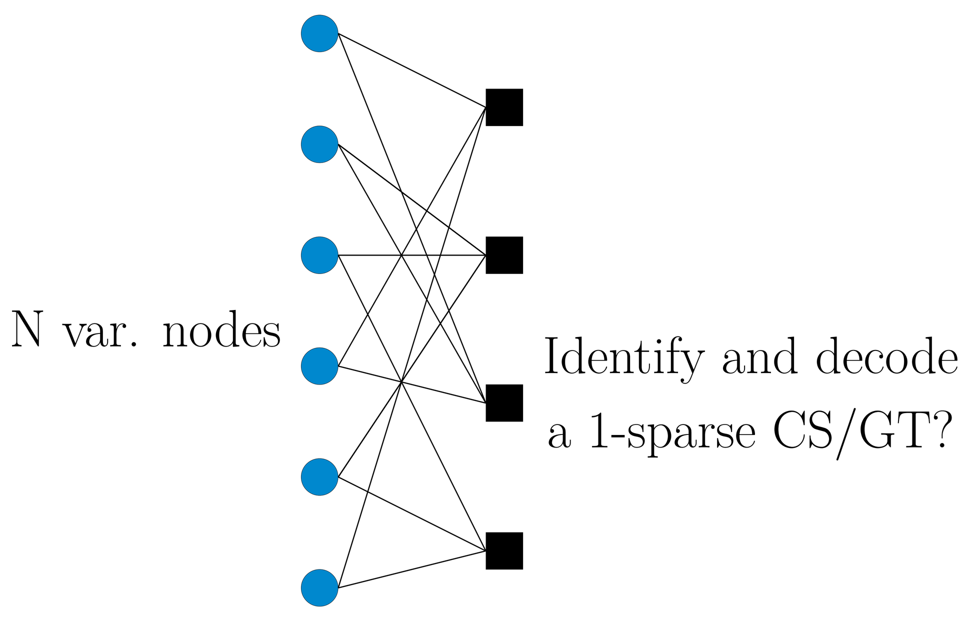
Main Result (GT)

Sub-linear sparsity: Let $K \in o(N^{\frac{p}{p+1}})$, for some $p \in \mathbb{Z}$

- Number of tests $M = 2(p+1)c(\epsilon)\ K \log_2(\frac{c_1 N}{K})$
- Decoding complexity of $O\left(K \log(\frac{N}{K})\right)$
- $\mathbb{P}_m \leq \epsilon$ asymptotically in K and N
- e.g. for $\epsilon = 10^{-5}$, $c(\epsilon) = 9.63$

Idea: Divide-and-Conquer

- Original problem is K-sparse CS/GT
- Divide N nodes into non-disjoint bins
- Can you solve for 1-sparse CS/GT at a bin?



K-sparse CS/GT

Divide: (ℓ, r) Bipartite Graph

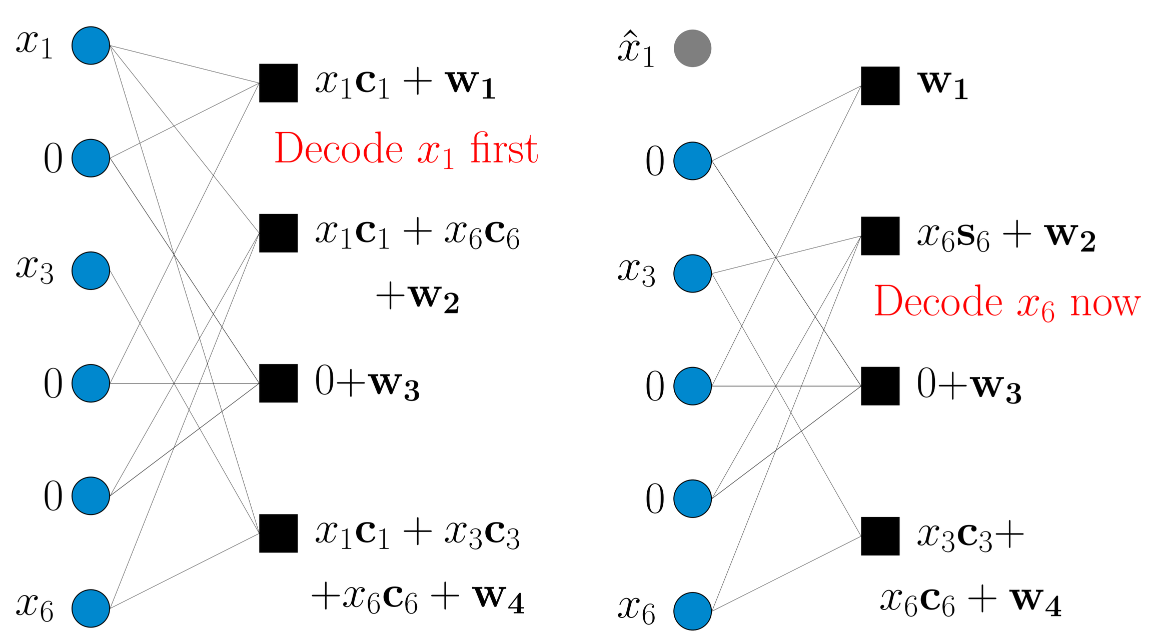
- N Variable(left) nodes: x_i. Each node has (left) degree: ℓ
- Bin(right) nodes: Choose $M_1 = cK$ bins (sub problems)
- Each bin has (right) degree r. Gives $r = \frac{N\ell}{cK}$
- Connections between $N\ell$ edges on each side are random

Conquer: 1-sparse CS

- At each bin, use code words of error control code $\mathcal{C}$:
$$\mathbf{y}_i = x_{i1}\mathbf{c}_1 + x_{i2}\mathbf{c}_2 + \ldots x_{ir}\mathbf{c}_r + \mathbf{w}_i$$
- If it is 1-sparse, only one $x_{ik} \neq 0$, (call it singleton):
$$\mathbf{y}_i = x_{ik}\mathbf{c}_k + \mathbf{w}_i$$
- A channel coding problem if $\mathrm{sign}(x_{ik})$ is known
- From channel coding, $\dim(\mathbf{c}_j) > \frac{\log r}{C_{\mathrm{Sh}}} \approx \Theta(\log \frac{N}{K})$
- After decoding index k, need to decode the value of x_{ik}
- Let $x \in \mathcal{X} := \{\pm A : A \in \mathcal{A}\}$ be a discrete and finite set
- We decode $\hat{x}_{ik} \in \mathcal{X}$ to be the value that best fits $\mathbf{y}_i\mathbf{c}_k^\dagger / ||\mathbf{c}_k||^2$

Reconstruction via Peeling

- Assume we can conquer the 1-sparse CS/GT at a bin
- If a singleton bin is found, decode the index and the value
- Peel off the decoded variable node value from other bins



- Continue peeling iteratively until no new singletons are found
- $\exists$ threshold c_* such that for $M_1 > c_* K$ bins, peeling decoder recovers all nodes w.h.p asymptotically

Conquer: 1 and 2-sparse GT

- Let $\mathbf{b}_k$ be the binary expansion of k; $\bar{\mathbf{b}}_k$ it's complement
- At each bin, test result vector would be:
$$\mathbf{y}_i = x_{i1}\begin{bmatrix}\mathbf{b}_1\\ \bar{\mathbf{b}}_1\end{bmatrix} \vee x_{i2}\begin{bmatrix}\mathbf{b}_2\\ \bar{\mathbf{b}}_2\end{bmatrix} \vee \ldots \vee x_{ir}\begin{bmatrix}\mathbf{b}_r\\ \bar{\mathbf{b}}_r\end{bmatrix}$$
- If it is 1-sparse i.e. only one $x_{ij} = 1$, trivial to decode j:
$$\mathbf{y}_i = \begin{bmatrix}\mathbf{b}_k\\ \bar{\mathbf{b}}_k\end{bmatrix}$$
- Non-linear OR operation poses problem in peeling; can't remove a decoded variable node from connected bins
- If a bin is 2-sparse i.e. $x_{ij}, x_{ik} = 1$ and an index j is known: refer to as resolvable double-ton
$$\mathbf{y}_i = \begin{bmatrix}\mathbf{b}_j\\ \bar{\mathbf{b}}_j\end{bmatrix} \vee \begin{bmatrix}\mathbf{b}_k\\ \bar{\mathbf{b}}_k\end{bmatrix}$$
- bins with more than 2 non-zero variables (multi-ton) are unusable due to *OR*
- For $M_1 > c(\epsilon)K$ bins, just using singletons and double-tons, peeling like iterative decoder recovers $1 - \epsilon$ fraction of nodes w.h.p asymptotically

Conlusions

Compressed Sensing: We propose a scheme that has

- order optimal sample complexity of $O(K \log(\frac{N}{K}))$
- sub-linear optimal decoding complexity: $O(K \log(\frac{N}{K}))$

Group testing: We propose a scheme that achieves

- order optimal testing complexity: $O(K \log(\frac{N}{K}))$
- sub-linear optimal decoding complexity: $O(K \log(\frac{N}{K}))$

References

[1] M. Wainwright, "Information-Theoretic Limits on Sparsity Recovery in the High-Dimensional and Noisy Setting." *IEEE Trans. Inform Theory*, vol. 55, no. 12, pp. 5728-5741, 2009.

[2] 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.

[3] X. Li, S. Pawar, and K. Ramchandran, "Sub-linear time compressed sensing using sparse-graph codes", in *Proc. Int. Symp. Inform. Theory*, pp. 1645-1649, 2015.