

Group Testing using Left and Right regular sparse graphs

Avinash Vem, Krishna R. Narayanan
Department of Electrical and Computer Engineering
Texas A&M University
{vemavinash@gmail.com, krn@tamu.edu}

Abstract—To be added.

I. INTRODUCTION

The problem of Group Testing refers to testing a large population for sick/defective individual people when the fraction of sick people is known to be small. This problem was first introduced to the literature of statistics by Dorfman [1] during World War II for testing the soldiers for syphilis without having to test each soldier individually. Since then many schemes and algorithms were designed for this problem.

In [2] Lee, Pedarsani, Ramchandran applied the sparse-graph codes and a simple peeling decoder, which are popular tools in the error control coding community, to the non-adaptive group testing problem. [3], [4], [2]

II. PROBLEM STATEMENT

Formally, we define the group testing problem as follows. Let N be the total number of items, K be the number of defective items and m be the number of tests for identifying the defective items. For now we consider only the noiseless group testing problem i.e., the result of each test is exactly equal to the boolean OR of all the items participating in the test.

Formally, let the support vector $\mathbf{x} \in \{0, 1\}^N$ denote the list of items with non-zero indices indicating the defective items. A non-adaptive testing scheme of m measurements can be denoted by a matrix $\mathbf{A} \in \{0, 1\}^{m \times N}$ where each row \mathbf{a}_i corresponds to a test. The non-zero indices in row \mathbf{a}_i correspond to the items that participate in i^{th} test. With these notations the output corresponding to \mathbf{A} can be expressed in matrix form as:

$$\mathbf{y} = \mathbf{A} \odot \mathbf{x}$$

where \odot is the usual matrix multiplication in which the arithmetic multiplications are replaced by the boolean AND operation and the arithmetic additions are replaced by the boolean OR operation.

III. TESTING SCHEME

In [2] Lee, Pedarsani and Ramchandran introduced a framework based on sparse graph codes for non-adaptive group testing problem. We will briefly review the SAFFRON

testing scheme, decoder and their main results. The SAFFRON testing scheme consists of two stages: the first stage is based on a left-regular sparse graph code and the second stage is a signature matrix at each bin derived from the first stage. For the first stage consider a bipartite graph with N variable nodes and M_1 bin nodes. Each variable node is connected to l bin nodes chosen uniformly at random. All the variable nodes (historically depicted on the left side of the graph) have a degree l , hence the left-regular, whereas the bin nodes on the right have a degree ranging from $[1 : n]$.

Let $\mathbf{T} \in \{0, 1\}^{M_1 \times N}$ be the adjacency matrix corresponding to the bipartite graph. At universal signature matrix $\mathbf{U} \in \{0, 1\}^{h \times N}$ is defined corresponding to h tests at each bin. Thus the total number of tests is $M = M_1 \times h$. More formally, the overall testing matrix $\mathbf{A} := [\mathbf{A}_1^T, \dots, \mathbf{A}_{M_1}^T]^T$ where $\mathbf{A}_i = \mathbf{U} \circ \text{diag}(\mathbf{t}_i)$.

Theorem 1. SAFFRON framework introduced in [2] recovers atleast a $(1 - \epsilon)$ fraction of the defective items for arbitrarily-small ϵ with high probability. The number of tests is $m = 6C(\epsilon)K \log_2 N$ and the computational complexity is $O(K \log N)$.

Note that the computational complexity is order optimal for both the noiseless and noisy settings as mentioned in [?]. Regarding the optimality of the number of tests for the noiseless setting where both K and N scale satisfying $K = o(N)$, it was shown [3] that the number of tests need to be atleast as large as $CK \log(\frac{N}{K})$ for some constant C such that the probability of error approaches zero. As far as we are aware this is the tightest lower bound. In the same work it is shown that $CK \log N$ is the sufficient number of tests. In our work we show that in fact $C(\epsilon)K \log(\frac{N}{K})$ tests is sufficient to recover $(1 - \epsilon)$ fraction of the defective items with high probability. More survey needs to be done regarding the lower and upper bounds for the number of tests in noiseless and noisy settings especially under different performance evaluation criteria. For e.g., in [3] the upper bound(achievable) on the minimal number of tests $O(K \log N)$ is when the performance metric considered is the average probability of error that the decoded support set is not exactly equal to the original support set. But for the framework where ϵ -fraction of the defective items are allowed to be missed, only the lower

bound on the number of tests required is given.

IV. PROPOSED SCHEME

REFERENCES

- [1] R. Dorfman, "The detection of defective members of large populations," *The Annals of Mathematical Statistics*, vol. 14, no. 4, pp. 436–440, 1943.
- [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] G. K. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1880–1901, 2012.
- [4] A. Mazumdar, "Nonadaptive group testing with random set of defectives via constant-weight codes," *arXiv preprint arXiv:1503.03597*, 2015.