Finding a Set of Healthy Individuals from a Large Population: Algorithms and Bounds

Abhay Sharma and C. R. Murthy
abhay.bits@gmail.com; cmurthy@ece.iisc.ernet.in

August 8, 2014

Group Testing Framework

- Main ingredients
  - A set of N items with a small number K of defective items
  - Group test: An enabling *tool* that tests arbitrary group of items together in a group
- Given the binary test outcomes of multiple group tests the main goal is to identify the defective set
- Different curry flavors
  - Non-adaptive vs adaptive pooling
  - Random vs Deterministic pooling
  - Noisy outcomes
- On today's menu: NNGT-R: Noisy, Non-adaptive Group Testing with Random pools

"Healthy" Subset Identification

- Sometimes identification of an healthy subset is of prime importance
  - Spectrum hole search in a cognitive radio network
    - Primary occupancy is sparse, secondary users need to find only a "small free chunk"
    - Does the secondary network need to identify all the bands with primary occupancy?
  - Data streams: Online trivia contest
  - Manufacturing: Shipping a non-defective batch on high priority
- Focus on computationally tractable algorithms to identify a healthy subset of a given size
  - Identification of defective set will do the job !!

NNGT-R Signal Model

- Noisy group testing signal model

$$\underline{y} = \bigvee_{i=1}^{N} \mathbf{D}_i \underline{x}_i \mathbb{I}_{\{i \in \mathcal{G}\}} \bigvee \underline{w}$$

  - $\mathcal{G}$ is the defective set
  - $\underline{x}_i \in \{0,1\}^M$ is the i^{th} column of $\mathbf{X}$
  - $\mathbf{X}(i,j) \sim \mathcal{B}(p)$, i.i.d., p is a design parameter, $0 < p < 1$
  - $\underline{w} \in \{0,1\}^M$ is the additive noise, $\underline{w}(i) \sim \mathcal{B}(q)$.
  - $\mathbf{D}_i \triangleq \text{diag}(\underline{d}_i)$
    - $\underline{d}_i \in \{0,1\}^M$, $\underline{d}_i(j) \sim \mathcal{B}(1-u)$ is chosen independently $\forall\, j = 1, 2, \ldots M$ and $\forall\, i = 1, 2, \ldots N$

Main goals of this work

Given the test output vector, $\underline{y}$, our goals are following:

- To find computationally tractable algorithms to identify L non-defective items, i.e., an L-sized subset belonging to $[N] \backslash \mathcal{G}$.
- Algorithm analysis
  - Finding the number of tests M for successful recovery with high probability
  - Choosing the appropriate design/algorithm parameters

Lower bounds on number of tests

	$0 \leq \alpha_0 < 1$	Small α_0, e.g., $\alpha_0 \leq 0.5$
No Noise	$O\left(\frac{K}{\log K} \log \frac{1}{(1-\alpha_0)}\right)$	$O\left(\frac{K\alpha_0}{\log K}\right)$
Dilution Noise	$O\left(\frac{K}{(1-u)\log K} \log \frac{1}{1-\alpha_0}\right)$	$O\left(\frac{K\alpha_0}{(1-u)\log K}\right)$
Additive Noise	$O\left(\frac{K}{\log \frac{1}{q}} \log \frac{1}{(1-\alpha_0)}\right)$	$O\left(\frac{K\alpha_0}{\log \frac{1}{q}}\right)$

- $\alpha_0 \triangleq \frac{L-1}{N-K}$

Ready to start . . .

- Pop quiz: Identify these parameters ?
  - N, K, L
  - q, u
  - p (or α)

Row based algorithm: **A1**

- Compute $\underline{z} = \sum_{j \in \text{supp}(\underline{y}^c)} \underline{x}_j^{(r)}$, where $\underline{x}_j^{(r)}$ is the j^{th} row of the test matrix.
- Order entries of $\underline{z}$ in descending order.
- Declare the items indexed by the top L entries as the non-defective subset.

Non-Uniform recovery with **A1**

Theorem

Let p be chosen as $\frac{\alpha}{K}$ with $\alpha = O(\frac{1}{(1-u)})$. If the number of tests are chosen as

$$M = O\left(\frac{K}{(1-q)(1-u)} \frac{\left[\log \binom{N-K}{L-1} + \log K\right]}{(N-K)-(L-1)}\right), \quad (1)$$

then for a given defective set there exist positive constants c_0, c_1, such that the algorithm **A1** *finds L non-defective items with probability exceeding $1 - \exp(-Mc_0) - \exp(-Mc_1)$.*

A Greedy algorithm: **A2**

Let $\psi_{cb} > 0$ be some normalizing constant.

- For each $i = 1, \ldots, N$, compute

$$\mathcal{T}(i) = \psi_{cb}(\underline{x}_i^T \underline{y}) - \underline{x}_i^T \underline{y}^c, \quad (2)$$

  where $\underline{x}_i$ is the i^{th} column of $\mathbf{X}$.
- Sort $\mathcal{T}(i)$ in descending order.
- Output the last L entries as the healthy subset.

Non-Uniform recovery with **A2**

Theorem

Let $\Gamma \triangleq (1-q)\,(1-(1-u)p)^K$ *and let* $\gamma_0 \triangleq (u/(1-(1-u)p))$. *Let* p *be chosen as* α/K *with* $\alpha = O(\frac{1}{(1-u)})$ *sufficiently small.* *Set* $\psi_{cb} = \sqrt{\frac{\Gamma}{1-\gamma_0\Gamma}}$. *If the number of tests are chosen as*

$$M = O\left(\frac{K}{(1-u)(1-q)}\frac{\left[\log\binom{N-K}{L-1} + \log K\right]}{(N-K)-(L-1)}\right), \qquad (3)$$

then for a given defective set there exists a positive constant c_0 *such that the algorithm* **A2** *finds* L *non-defective items with probability exceeding* $1-\exp(-Mc_0)$.

LP relaxation based algorithms

- Setup and solve a linear program (LP0, LP1 or LP2). Let $\underline{\hat{z}}$ be the solution.
- Sort $\underline{\hat{z}}$ and choose the items indexed by the largest L entries as $\hat{S}_L$.

- **A3**: LP0
- **A3++**: LP1
- **A4**: LP2

LP0

- Let $Y_z \triangleq \{\underline{y} = 0\}$, $M_z = |Y_z|$, $Y_p \triangleq \{\underline{y} = 1\}$, $M_p = |Y_p|$

$$\begin{aligned} \underset{\underline{z},\underline{\eta}}{\text{minimize}} \quad & \underline{1}_{M_z}^T \underline{\eta} \\ \textit{subject to} \quad & \mathbf{X}(Y_z,:)(\underline{1}_N - \underline{z}) - \underline{\eta} = \underline{0}_{M_z}, \qquad \textbf{(LP0)} \qquad (4) \\ & \underline{0}_N \preccurlyeq \underline{z} \preccurlyeq \underline{1}_N, \quad \underline{\eta} \succcurlyeq \underline{0}_{M_z}, \\ & \underline{1}_N^T \underline{z} \leq L. \end{aligned}$$

LP1

$$
\begin{aligned}
\underset{\underline{z},\underline{\eta}_z}{\text{minimize}} \quad & \underline{1}_{M_z}^T \underline{\eta}_z \\
\text{subject to} \quad & \mathbf{X}(Y_z,:)(\underline{1}_N - \underline{z}) - \underline{\eta}_z = \underline{0}_{M_z} \\
& \mathbf{X}(Y_p,:)(\underline{1}_N - \underline{z}) \succcurlyeq (1-\epsilon_0)\underline{1}_{M_p} \\
& \underline{0}_N \preccurlyeq \underline{z} \preccurlyeq \underline{1}_N, \quad \underline{\eta}_z \succcurlyeq \underline{0}_{M_z} \\
& \underline{1}_N^T \underline{z} \leq L
\end{aligned}
\qquad \textbf{(LP1)}
$$

Non-Uniform recovery with **A3** and **A3++**

Theorem

Let p be chosen as $\frac{\alpha}{K}$ *with* $\alpha = O\left(\frac{1}{1-u}\right)$. *If the number of tests are chosen as (1), then for a given defective set there exist positive constants* c_0, c_1, *such that the algorithm* **A3** *(and* **A3++***) finds L non-defective items with probability exceeding* $1 - \exp(-Mc_0) - \exp(-Mc_1)$.

$$\begin{aligned}
\underset{\underline{z}}{\text{minimize}} \quad & \underline{1}_{M_z}^T \mathbf{X}(Y_z,:)(\underline{1}_N - \underline{z}) - \psi_{lp}\left[\underline{1}_{M_p}^T \mathbf{X}(Y_p,:)(\underline{1}_N - \underline{z})\right] \\
\text{subject to} \quad & \underline{0}_N \preccurlyeq \underline{z} \preccurlyeq \underline{1}_N \qquad \textbf{(LP2)} \\
& \underline{1}_N^T \underline{z} \leq L
\end{aligned}$$

Non-Uniform recovery with **A4**

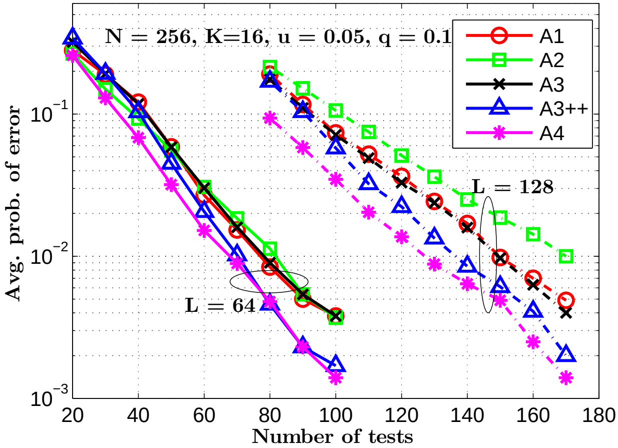
Theorem

Let p be chosen as $\frac{\alpha}{K}$ *with* $\alpha = O(\frac{1}{(1-u)})$ *sufficiently small and let* $\psi_{lp} = O(\frac{1}{Mp})$ *sufficiently small. If the number of tests are chosen as (3) then for a given defective set there exist positive constants* c_0, c_1, c_2 *such that the algorithm* **A4** *finds at least L non-defective items with probability exceeding* $1 - \exp(-Mc_0) - \exp(-Mc_1) - \exp(-Mc_2)$.

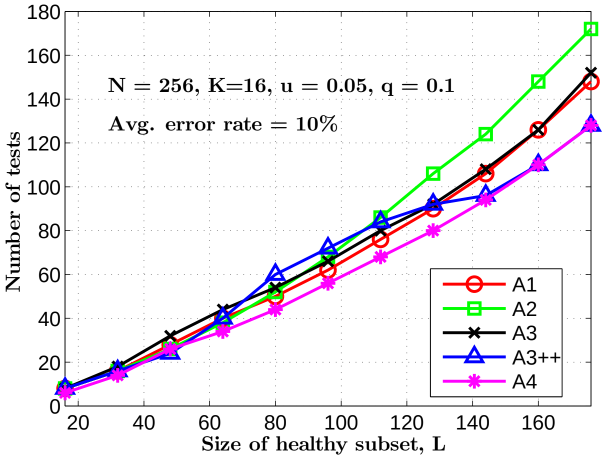
Some observations

- Comparisons with lower bounds
  - Within $\log(K)$ factor of lower bounds
  - Optimal with respect to impact of u and q
- Penalties due to imperfect knowledge of K and u
  - For K: Let $\hat{K} = \Delta_k K$ and let $p = O(1/\hat{K})$
    - A factor of $f_M(\Delta_k) \triangleq \Delta_k \exp\left(-(1-u)(\frac{1}{\Delta_k} - 1)\right)$ increase
    - $f_M(\Delta_k)$ is asymmetric in Δ_k
  - A factor of $1/(1-u)$ increase for not using information about u in choosing p

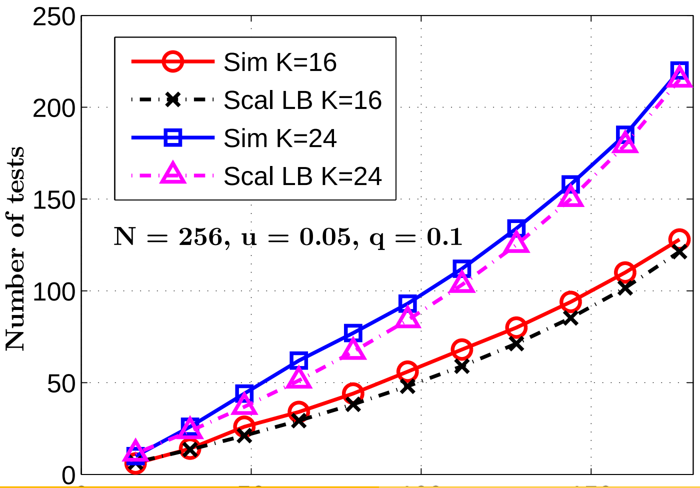
Average Probability of Error Vs M



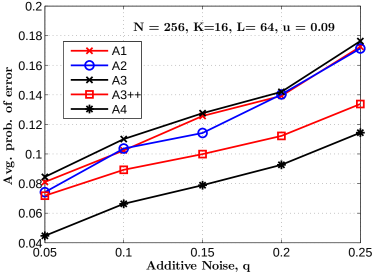
M Vs L


N = 256, K=16, u = 0.05, q = 0.1
Avg. error rate = 10%
Number of tests
Size of healthy subset, L
A1
A2
A3
A3++
A4
0
20
40
60
80
100
120
140
160
180
20
40
60
80
100
120
140
160

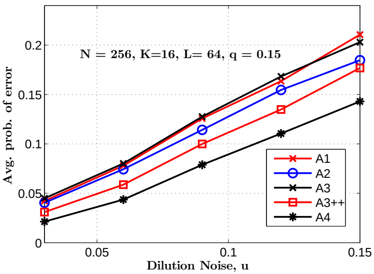
A4 Vs Scaled Lower Bounds



Performance variation with noise



(a)



(b)

Robustness to uncertainity in the knowledge of K

$K_t = 16, N = 256, L = 128, q = 0.1, u = 0.05$			
	$\Delta_K = 0.75$	$\Delta_K = 1.5$	$\Delta_K = 2.0$
A1	1.13	1.06	1.20
A2	1.55	1.0	1.16
A3	1.08	1.04	1.17
A3++	1.04	1.0	1.17
A4	1.1	1.02	1.17

Proof for **A3**

- Equivalent formulation

$$\begin{aligned} \underset{\underline{z}}{\text{minimize}} \quad & \underline{1}_{M_z}^T \mathbf{X}_0 \underline{z} \\ \text{subject to} \quad & \underline{0}_N \preccurlyeq \underline{z} \preccurlyeq \underline{1}_N \qquad \textbf{(LP0a)} \\ & \underline{1}_N^T \underline{z} \geq (N - L) \end{aligned}$$

- KKT conditions

$$\underline{1}_{M_z}^T \mathbf{X}_0 - \underline{\lambda}_1 + \underline{\lambda}_2 - \nu \underline{1}_N = \underline{0}_N \quad (5)$$

$$\underline{\lambda}_1 \circ \underline{z} = \underline{0}_N; \ \underline{\lambda}_2 \circ (\underline{z} - \underline{1}_N) = \underline{0}_N; \ \nu(\underline{1}_N^T \underline{z} - (N - L)) = 0; \quad (6)$$

$$\underline{0}_N \preccurlyeq \underline{z} \preccurlyeq \underline{1}_N; \ \underline{1}_N^T \underline{z} \geq (N - L); \ \underline{\lambda}_1 \succcurlyeq \underline{0}_N; \ \underline{\lambda}_2 \succcurlyeq \underline{0}_N; \ \nu \geq 0; \quad (7)$$

Proof for **A3** (Cotd.)

- Let $(\underline{z}, \underline{\lambda}_1, \underline{\lambda}_2, \nu)$ be the primal, dual optimal points
- Claim: If $\underline{\lambda}_2(i) > 0,\ \forall\, i \in S_d$ then $\hat{S}_L \cap S_d = \{\emptyset\}$
- Thus, $\mathcal{E} \implies \{\underline{\lambda}_2(i) = 0\} \implies \underline{1}_{M_z}^T \mathbf{X}_0(:, i) = \underline{\lambda}_1(i) + \nu \geq \nu$
- Define, $\theta_0 \triangleq \max_{\{i : \underline{\lambda}_1(i) = 0\}} \underline{1}_{M_z}^T \mathbf{X}_0(:, i)$
- Claim: $\nu \geq \theta$
- Claim: $\exists$ at most L items for which $\underline{\lambda}_1(i) > 0$
- Thus, for a given i, $\exists$ at least $(N - K) - (L - 1)$ non-defective items that have $\lambda_1(i) = 0$
- $\mathcal{E} \subseteq \cup_{i \in S_d} \cup_{S_z \in \mathcal{S}_z} \left\{ \underline{1}_{M_z}^T \mathbf{X}_0(:, i) \ \geq \ \underline{1}_{M_z}^T \mathbf{X}_0(:, j), \forall j \in S_z \right\}$
- This is exactly the same error event as was analysed for **A1** !

Thank You