Near-Optimal Noisy Group Testing via Separate Decoding of Items

Jonathan Scarlett
National University of Singapore
scarlett@comp.nus.edu.sg

Volkan Cevher
LIONS, EPFL
volkan.cevher@epfl.ch

***Abstract*—In this paper, we revisit an efficient algorithm for noisy group testing in which each item is decoded separately (Malyutov and Mateev, 1980), and develop novel performance guarantees via an information-theoretic framework for general noise models. For the noiseless and symmetric noise models, we find that the asymptotic number of tests required for vanishing error probability is within a factor $\log 2 \approx 0.7$ of the information-theoretic optimum at low sparsity levels, and that when a small fraction of incorrectly-decoded items is allowed, this guarantee extends to all sublinear sparsity levels. In many scaling regimes, these are the best known theoretical guarantees for any noisy group testing algorithm.**

I. Introduction

In this paper, we consider the group testing problem [1], in which one seeks to determine a small subset S of "defective" items within a larger set of items $\{1,\dots,p\}$ based on a number of tests. In the noiseless setting, each test takes the form

$$Y = \bigvee_{i \in S} X_i, \tag{1}$$

where the test vector $X = (X_1,\dots,X_p) \in \{0,1\}^p$ indicates which items are included in the test, and Y is the resulting observation. That is, the output indicates whether at least one defective item was included in the test. One wishes to minimize the total number of tests n while still ensuring the reliable recovery of S.

We let the defective set S be uniform on the $\binom{p}{k}$ subsets of $\{1,\dots,p\}$ of cardinality k. For convenience, we will sometimes equivalently refer to a vector $\beta \in \{0,1\}^p$ whose j-th entry indicates whether or not item j is defective:

$$\beta_j = \mathbb{1}\{j \in S\}. \tag{2}$$

We consider i.i.d. Bernoulli testing, where each item is placed in a given test independently with probability $\frac{\nu}{k}$ for some constant $\nu > 0$. The vector of n observations is denoted by $\mathbf{Y} \in \{0,1\}^n$, and the corresponding measurement matrix (each row of which contains a single measurement vector $X = (X_1,\dots,X_p)$) is denoted by $\mathbf{X} \in \{0,1\}^{n\times p}$.

Generalizing (1), we consider a broad class of *noisy* group testing models. Denoting the i-th entry of $\mathbf{Y}$ by $Y^{(i)}$ and the i-th row of $\mathbf{X}$ by $X^{(i)}$, the measurement model is given by

$$(Y^{(i)}|X^{(i)}) \sim P_{Y|N(S,X^{(i)})}, \tag{3}$$

where $N(S,X^{(i)}) = \sum_{j=1}^p \mathbb{1}\{j \in S \cap X_j^{(i)} = 1\}$ denotes the number of defective items in the test. That is, we consider arbitrary noise distributions $P_{Y|N}$ for which $Y^{(i)}$ depends on $X^{(i)}$ only through $N(S,X^{(i)})$, with conditional independence among the tests $i = 1,\dots,n$. For each item $j = 1,\dots,p$, the j-th column of $\mathbf{X}$ is written as $\mathbf{X}_j \in \{0,1\}^n$.

While most of our results will be written in terms of general noise models of the form (3), we also pay particular attention to two specific models: The noiseless model in (1), and the *symmetric noise model* with parameter $\rho > 0$:

$$Y = \bigg(\bigvee_{i \in S} X_i\bigg) \oplus Z, \tag{4}$$

where $Z \sim \mathrm{Bernoulli}(\rho)$, and $\oplus$ denotes modulo-2 addition.

Given $\mathbf{X}$ and $\mathbf{Y}$, a *decoder* forms an estimate $\hat{S}$ of S, or equivalently, an estimate $\hat{\beta}$ of β. We consider two related performance measures. In the case of *exact* recovery, the error probability is given by

$$P_{\mathrm{e}} := \mathbb{P}[\hat{S} \ne S], \tag{5}$$

and is taken over the realizations of S, $\mathbf{X}$, and $\mathbf{Y}$ (the decoder is assumed to be deterministic). In addition, we consider a less stringent performance criterion in which we allow for up to $d_{\mathrm{pos}} \in \{0,\dots,p-k-1\}$ false positives and $d_{\mathrm{neg}} \in \{0,\dots,k-1\}$ false negatives, yielding an error probability of

$$P_{\mathrm{e}}(d_{\mathrm{pos}}, d_{\mathrm{neg}}) := \mathbb{P}\big[|\hat{S}\setminus S| > d_{\mathrm{pos}} \cup |S\setminus\hat{S}| > d_{\mathrm{neg}}\big]. \tag{6}$$

A. Separate Decoding of Items

In this paper, we study a decoding method introduced in an early work of Malyutov and Mateev [5] (see also [6], [7]), which we refer to as *separate decoding of items*. Specifically, we adopt this terminology to mean any decoding scheme in which $\hat{\beta}_j$ is only a function of $\mathbf{X}_j$ and $\mathbf{Y}$, i.e.,

$$\hat{\beta}_j = \phi_j(\mathbf{X}_j, \mathbf{Y}), \quad j = 1,\dots,p \tag{7}$$

for some functions $\{\phi_j\}_{j=1}^p$. All of our results will choose ϕ_j not depending on j; more specifically, following [5], each decoder is of the following form for some $\gamma > 0$:[1]

$$\phi_j(\mathbf{X}_j, \mathbf{Y}) = \mathbb{1}\bigg\{\sum_{i=1}^n \log \frac{P_{Y|X_j,\beta_j}(Y^{(i)}|X_j^{(i)},1)}{P_Y(Y^{(i)})} > \gamma\bigg\}, \tag{8}$$

[1] Here and subsequently, the function $\log(\cdot)$ has base e, and the corresponding information quantities are in units of nats.

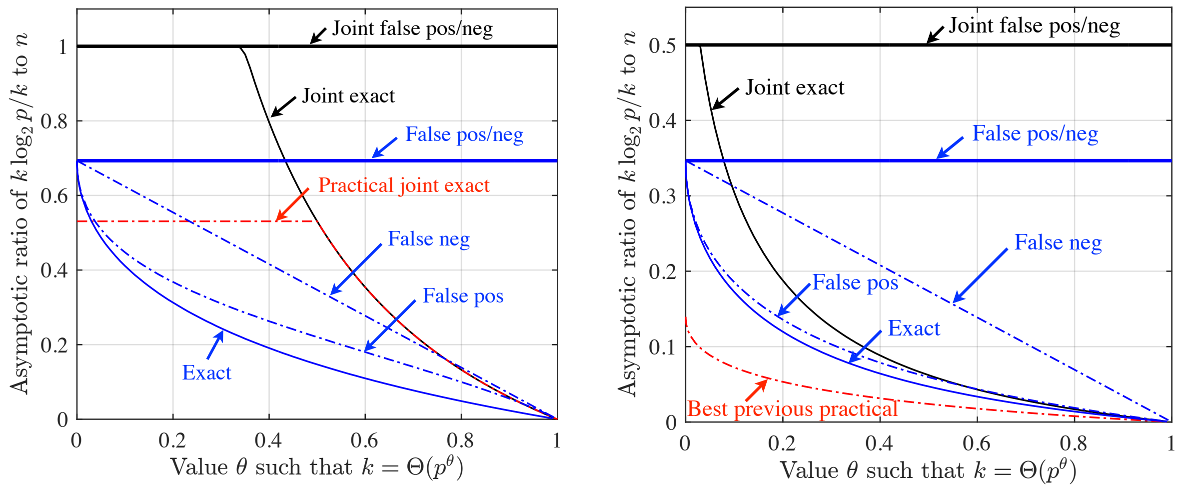

Figure 1: Asymptotic thresholds on the number of tests required for vanishing error probability in (Left) the noiseless setting, and (Right) the symmetric noise setting with $\rho = 0.11$. The blue curves correspond to separate decoding of items, with exact recovery (Exact), $\Theta(k)$ false positives only (False pos), $\Theta(k)$ false negatives only (False neg), or both $\Theta(k)$ false positives and $\Theta(k)$ false negatives (False pos/neg). The "practical joint exact" curve is DD [2] or LP [3] in the noiseless case, and the "best previous practical" curve is NCOMP [4] in the noisy case.

where P_Y is the unconditional distribution of a given observation, and $P_{Y|X_j,\beta_j}(\cdot|\cdot,1)$ is the conditional distribution given $\beta_j = 1$ and the value of X_j. This can be interpreted as the Neyman-Pearson test for binary hypothesis testing with hypotheses $H_0 : \beta_j = 0$ and $H_1 : \beta_j = 1$.

Malyutov and Mateev [5] showed that when $k = O(1)$ and the decoder (8) is used with suitably-chosen γ, one can achieve exact recovery with vanishing error probability provided that

$$n \geq \frac{\log p}{I_1}(1 + o(1)), \tag{9}$$

where the *single-item mutual information* I_1 is defined as follows, with implicit conditioning on item 1 being defective:

$$I_1 := I(X_1; Y). \tag{10}$$

As noted in [7], in the noiseless setting with $\nu = \ln 2$ we have $I_1 = \frac{(\log 2)^2}{k}(1 + o(1))$ as $k \to \infty$, and in this case, (9) matches the optimal information-theoretic threshold [8], [9] up to a factor of $\log 2 \approx 0.7$. Characterizations of the mutual information I_1 for other noise models were given in [10].

In this paper, we move beyond the highly sparse regime $k = O(1)$, and give theoretical guarantees for separate decoding of items for sublinear scalings of the form $k = \Theta(p^\theta)$, where $\theta \in (0, 1)$. As with joint decoding [8], [11], this regime comes with significant challenges, with additional requirements of n arising from *concentration inequalities* and often dominating (9). In addition, we show that far fewer tests may be needed under the partial recovery criterion in (6).

B. Other Related Work and Our Results

Figure 1 plots the asymptotic number of tests for achieving $P_e \to 0$ or $P_e(d_{\text{pos}}, d_{\text{neg}}) \to 0$ under Bernoulli testing with $\nu = \log 2$, including existing bounds and our novel contributions. In this figure, the number of allowed false positives and/or false negatives (if any) is always assumed to be $\Theta(k)$, with an arbitrarily small implied constant.

The information-theoretic limits of group testing for $k = O(1)$ have long been well-understood in the Russian literature [5], [12], and have recently become increasingly well-understood for $k = \Theta(p^\theta)$ [8], [9], [13], [14]. The information-theoretic joint decoding results in Figure 1 come from the recent works [8], [9], [11]. Note that the flat line at the top is not only sufficient for partial recovery, but also necessary, i.e., it is the exact information-theoretic threshold.

When it comes to *practical* algorithms in the regime $k = \Theta(p^\theta)$, near-optimal theoretical guarantees are known in the noiseless setting [2], [15], but the constant factors in the noisy setting are far from the information-theoretic limits [4], [16]–[19]. To our knowledge, the best known existing bounds are those of the Noisy Combinatorial Orthogonal Matching Pursuit (NCOMP) algorithm [4].

We make the following observations regarding Figure 1:

- In the noiseless case, our asymptotic bounds are within a factor $\log 2$ of the optimal threshold for joint decoding as $\theta \to 0$, are reasonable for all $\theta \in (0, 1)$ with improvements when false positives or false negatives are allowed, and are within a factor $\log 2$ of the optimal joint decoding threshold *for all* θ when both are allowed. Moreover, with exact recovery and $\theta \in (0, 0.0398)$, we strictly improve on the best known bound for any efficient algorithm under Bernoulli testing.
- For the symmetric noise model, the general behavior is similar, but we significantly outperform the best known previous bound (NCOMP [4]) for all $\theta \in (0, 1)$. Once again, when both false positives and false negatives are allowed, we are within a factor $\log 2$ of the optimal threshold for joint decoding.

II. Achievability Results with Exact Recovery

In this section, we develop the theoretical results for exact recovery leading to the asymptotic bounds for the noiseless and noisy settings in shown Figure 1. To do this, we first establish non-asymptotic bounds on the error probability, then present the tools for performing an asymptotic analysis, and finally give the details of the applications to specific models.

A. Additional Notation

We define some further notation in addition to that in the introduction. Our analysis will apply for any given choice of the defective set S, due to the symmetry of the observation model (3) and the i.i.d. test matrix $\mathbf{X}$. Hence, throughout this section we will focus on the specific set $S = \{1, \dots, k\}$. In particular, we assume that item 1 is defective, and we define $P_{Y|X_1}$ accordingly:

$$P_{Y|X_1}(y|x_1) = P_{Y|X_1,\beta_1}(y|x_1, 1). \tag{11}$$

Hence, the summation in (8) can be written as

$$\imath_1^n(\mathbf{X}_j, \mathbf{Y}) := \sum_{i=1}^{n} \imath_1(X_j^{(i)}, Y^{(i)}), \tag{12}$$

where

$$\imath_1(x_1, y) := \log \frac{P_{Y|X_1}(y|x_1)}{P_Y(y)}. \tag{13}$$

Following the terminology of the channel coding literature [20]–[22], we refer to this quantity as the *information density*. Denoting the distribution of a single entry of $\mathbf{X}$ by $P_X \sim \mathrm{Bernoulli}\big(\frac{\nu}{k}\big)$, we find that the average of (13) with respect to $(X_1, Y) \sim P_X \times P_{Y|X_1}$ is the mutual information I_1 in (10). With the above distributions in place, we define $P_X^n(\mathbf{x}_1) = \prod_{i=1}^n P_X(x_1^{(i)})$, $P_Y^n(\mathbf{y}) = \prod_{i=1}^n P_Y(y^{(i)})$, and $P_{Y|X_1}^n(\mathbf{y}|\mathbf{x}_1) = \prod_{i=1}^n P_{Y|X_1}(y^{(i)}|x_1^{(i)})$.

When we specialize our results to the noiseless and symmetric noise models, we will choose

$$\nu = \nu_{\mathrm{symm}} := \left\{ \text{unique value such that } \left(1 - \frac{\nu}{k}\right)^k = \frac{1}{2} \right\} \tag{14}$$

$$= (\log 2)(1 + o(1)). \tag{15}$$

For $k \to \infty$ (as we consider), there is essentially no difference between setting $\nu = \nu_{\mathrm{symm}}$ or $\nu = \log 2$, but we found the former to be slightly more convenient mathematically.

B. Initial Non-Asymptotic Bound

The following theorem provides an initial non-asymptotic upper bound on the error probability for general models. The result is proved using simple thresholding techniques that appeared in early studies of channel coding [23], [24], and have also been applied previously in the context of group testing [5], [8], [11].

Theorem 1. (Non-asymptotic, exact recovery) *For a general group testing model with with* $\mathrm{Bernoulli}\big(\frac{\nu}{k}\big)$ *testing and separate decoding of items according to* (8)*, we have*

$$P_{\mathrm{e}} \le k\mathbb{P}[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \le \gamma] + (p-k)e^{-\gamma}, \tag{16}$$

where $(\mathbf{X}_1, \mathbf{Y}) \sim P_X^n(\mathbf{x}_1) P_{Y|X_1}^n(\mathbf{y}|\mathbf{x}_1)$*, and* γ *is as in* (8)*.*

Proof. For the exact recovery criterion, correct decoding requires the k defective items to pass the threshold test, and the $p-k$ non-defective items to fail the threshold test. Hence, by the union bound, we have

$$P_{\mathrm{e}} \le k\mathbb{P}[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \le \gamma] + (p-k)\mathbb{P}[\imath_1^n(\overline{\mathbf{X}}_1, \mathbf{Y}) > \gamma], \tag{17}$$

where $(\mathbf{X}_1, \mathbf{Y}, \overline{\mathbf{X}}_1) \sim P_X^n(\bar{\mathbf{x}}_1) P_{Y|X_1}^n(\mathbf{y}|\bar{\mathbf{x}}_1) P_X^n(\bar{\mathbf{x}}_1)$, i.e., $\overline{\mathbf{X}}_1$ is an independent copy of $\mathbf{X}_1$ (recall that the columns of $\mathbf{X}$ are i.i.d.). We bound the second term by writing

$$\begin{aligned}
&\mathbb{P}[\imath_1^n(\overline{\mathbf{X}}_1, \mathbf{Y}) > \gamma] \\
&\quad = \sum_{\bar{\mathbf{x}}_1, \mathbf{y}} P_X^n(\bar{\mathbf{x}}_1) P_Y^n(\mathbf{y}) \mathbb{1}\left\{ \log \frac{P_{Y|X_1}^n(\mathbf{y}|\bar{\mathbf{x}}_1)}{P_Y^n(\mathbf{y})} > \gamma \right\} && (18) \\
&\quad \le \sum_{\bar{\mathbf{x}}_1, \mathbf{y}} P_X^n(\bar{\mathbf{x}}_1) P_{Y|X}^n(\mathbf{y}|\bar{\mathbf{x}}_1) e^{-\gamma} && (19) \\
&\quad = e^{-\gamma}, && (20)
\end{aligned}$$

where (18) follows from the preceding joint distribution and the definition of $\imath_1^n$, and (20) bounds $P_Y^n(\mathbf{y})$ according to the event in the indicator function, and then bounds the indicator function by one. Combining (17) and (20) completes the proof. □

C. Asymptotic Analysis

In order to apply Theorem 1, we need to characterize the probability appearing in the first term. The idea is to exploit the fact that $\imath_1^n(\mathbf{X}_1, \mathbf{Y})$ is an i.i.d. sum, and hence concentrates around its mean. While the following corollary is essentially a simple rewriting of Theorem 1, it makes the application of such concentration bounds more transparent Here and subsequently, asymptotic notation such as $\to, o(\cdot), O(\cdot)$ is with respect to $p \to \infty$, and we assume that $k \to \infty$ with $k = o(p)$.

Theorem 2. (Asymptotic bound, exact recovery) *Under the setup of Theorem 1, suppose that the information density satisfies a concentration inequality of the following form:*

$$\mathbb{P}[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \le nI_1(1-\delta_2)] \le \psi_n(\delta_2) \tag{21}$$

for some function $\psi_n(\delta_2)$*. Moreover, suppose that the following conditions hold for some* $\delta_1 \to 0$ *and* $\delta_2 > 0$*:*

$$n \ge \frac{\log\big(\frac{1}{\delta_1}(p-k)\big)}{I_1(1-\delta_2)} \tag{22}$$

$$k \cdot \psi_n(\delta_2) \to 0. \tag{23}$$

Then $P_{\mathrm{e}} \to 0$ *under the decoder in* (8) *with* $\gamma = \log \frac{p-k}{\delta_1}$*.*

Proof. Setting $\gamma = \log \frac{p-k}{\delta_1}$ in Theorem 1, we obtain

$$P_{\mathrm{e}} \le k\mathbb{P}\left[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \le \log \frac{p-k}{\delta_1}\right] + \delta_1, \tag{24}$$

By the condition in (22), the probability in (24) is upper bounded by $\mathbb{P}[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \le nI_1(1-\delta_2)]$, which in turn is upper bounded by $\psi_n(\delta_2)$ by (21). We therefore have from (24) that $P_{\mathrm{e}} \le k\psi_n(\delta_2) + \delta_1$, and hence the theorem follows from the assumption $\delta_1 \to 0$ along with (23). □

D. Concentration Bounds

In order to apply Theorem 2 to specific models, we need to characterize the concentration of $\imath_1^n(\mathbf{X}_1, \mathbf{Y})$ and attain an explicit expression for $\psi_n(\delta_2)$ therein. The following lemma brings us one step closer to attaining explicit expressions, giving a general concentration result based on Bernstein's inequality [25, Ch. 2].

Lemma 1. (Concentration via Bernstein's inequality) *Defining*

$$c_{\mathrm{mean}} := k\mathbb{E}[\imath(X_1, Y)] = kI_1 \tag{25}$$

$$c_{\mathrm{var}} := k\mathrm{Var}[\imath(X_1, Y)] \tag{26}$$

$$c_{\mathrm{max}} := \max_{x_1, y} \big|\imath(x_1, y)\big|, \tag{27}$$

we have for any $\delta_2 > 0$ *that*

$$\mathbb{P}\Big[\big|\imath_1^n(\mathbf{X}_1, \mathbf{Y}) - nI_1\big| \le n\delta_2\Big] \le 2\exp\left(\frac{-\frac{1}{2}\cdot\frac{n}{k}\cdot c_{\mathrm{mean}}^2\delta_2^2}{c_{\mathrm{var}} + \frac{1}{3}c_{\mathrm{mean}}c_{\mathrm{max}}\delta_2}\right) \tag{28}$$

We will use Lemma 1 to establish the results shown for the symmetric noise model in Figure 1 (Right). While we could also use Lemma 1 for the noiseless model, it turns out that we can in fact do better via the following.

Lemma 2. (Concentration for noiseless model) *Under the noiseless model with* $\nu = \nu_{\mathrm{symm}}$ (cf., (14))*, we have for any* $\delta_2 \in (0,1)$ *that*

$$\mathbb{P}[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \le nI_1(1-\delta_2)] \le \exp\bigg(-\frac{n(\log 2)^2}{k} \times \Big((1-\delta_2)\log(1-\delta_2) + \delta_2\Big)(1+o(1))\bigg) \tag{29}$$

as $p \to \infty$ *and* $k \to \infty$ *simultaneously.*

The proofs of the preceding lemmas can be found in the full version [26], and are based on Bernstein's inequality (Lemma 1) and the multiplicative Chernoff bound (Lemma 2).

E. Applications to Specific Models

Noiseless model: For the noiseless group testing model (*cf.*, (1)), we immediately obtain the following from Theorem 2 and Lemma 2.

Corollary 1. (Noiseless, exact recovery) *For the noiseless group testing problem with* $\nu = \nu_{\mathrm{symm}}$ (cf., (14)) *and* $k = \Theta(p^\theta)$ *for some* $\theta \in (0,1)$*, we can achieve* $P_{\mathrm{e}} \to 0$ *with separate decoding of items provided that*

$$n \ge \min_{\delta_2 > 0} \max\left\{\frac{k\log p}{(\log 2)^2(1-\delta_2)}, \frac{k\log k}{(\log 2)^2\big((1-\delta_2)\log(1-\delta_2)+\delta_2\big)}\right\}(1+\eta) \tag{30}$$

for some $\eta > 0$.

Proof. It is known that $I_1 = \frac{(\log 2)^2}{k}(1+o(1))$ [10], and hence the first term in (30) follows from (22) with $\delta_1 \to 0$ sufficiently slowly. Moreover, by equating $\psi_n(\delta_2)$ with the right-hand side of (29) and performing simple rearranging, we find that the second term in (30) follows from (23). □

Symmetric noise model: For the symmetric noisy model (*cf.*, (1)), we make use of Lemma 1, with the constants c_{mean}, c_{var} and c_{max} therein characterized in the following. Here H_2 is the binary entropy function in nats.

Lemma 3. (Bernstein parameters for symmetric noise) *Under the symmetric noise model with a fixed parameter* $\rho \in \big(0, \frac{1}{2}\big)$ *(not depending on* p*) and* $\nu = \nu_{\mathrm{symm}}$ (cf., (14))*, we have*

$$k\mathbb{E}[\imath(X_1, Y)] = (\log 2)\big(\log 2 - H_2(\rho)\big)(1+o(1)) \tag{31}$$

$$k\mathrm{Var}[\imath(X_1, Y)] \le (\log 2)\bigg((1-\rho)\log^2\big(2(1-\rho)\big) + \rho\log^2(2\rho)\bigg)(1+o(1)) \tag{32}$$

$$\max_{x_1, y}\big|\imath(x_1, y)\big| = \log\frac{1}{2\rho} \tag{33}$$

as $p \to \infty$ *and* $k \to \infty$ *simultaneously.*

The proof is based on directly analyzing the information density, and can be found in [26]. From this lemma, we immediately obtain the following.

Corollary 2. (Symmetric noise, exact recovery) *For noisy group testing with* $\rho \in \big(0, \frac{1}{2}\big)$ *(not depending on* p*),* $\nu = \nu_{\mathrm{symm}}$*, and* $k = \Theta(p^\theta)$ *for some* $\theta \in (0,1)$*, we can achieve* $P_{\mathrm{e}} \to 0$ *with separate decoding of items provided that*

$$n \ge \min_{\delta_2 > 0}\max\left\{\frac{k\log p}{(\log 2)(\log 2 - H_2(\rho))(1-\delta_2)}, \frac{(k\log k)\cdot\big(c_{\mathrm{var}} + \frac{1}{3}c_{\mathrm{mean}}c_{\mathrm{max}}\delta_2\big)}{\frac{1}{2}\cdot c_{\mathrm{mean}}^2\delta_2^2}\right\}(1+\eta) \tag{34}$$

for some $\eta > 0$*, where* c_{mean}*,* c_{var}*, and* c_{max} *are respectively given by the right-hand sides of* (31)–(33).

Other noise models: While we specifically applied Lemma 1 to the symmetric noise model, it can also be applied more generally, yielding an analogous result for any model in which the quantities c_{mean}, c_{var}, and c_{max} in (25)–(27) behave as $\Theta(1)$. In particular, for any such model and any fixed $\nu > 0$, in the limit as $\theta \to 0$, it suffices to have

$$n \ge \frac{\log p}{I_1}(1+\eta) = \frac{k\log p}{c_{\mathrm{mean}}}(1+\eta), \tag{35}$$

for arbitrarily small $\eta > 0$. In contrast, for θ strictly greater than zero, the conditions on n resulting from Bernstein's inequality may dominate (35), similarly to Corollary 2.

III. Outline of Extensions to Partial Recovery

Due to space constraints, we provide only an outline of the extension of the preceding analysis to partial recovery, where false positives and/or false negatives are allowed. The full details can be found in [26].

The main tool we need is the following, whose proof is in fact implicit in our analysis for the exact recovery criterion.

Lemma 4. (Auxiliary result for partial recovery) *For any group testing model of the form* (3)*, under the decoder in* (8) *with threshold $\gamma > 0$, we have the following:*

(i) For any $j \notin S$, the probability of passing the threshold test is upper bounded by $e^{-\gamma}$.

(ii) Suppose that the information density satisfies a concentration inequality of the form (21) *for some function $\psi_n(\delta_2)$, and that the number of tests satisfies $n \geq \frac{\gamma}{I_1(1-\delta_2)}$. Then for any $j \in S$, the probability of failing the threshold test is upper bounded by $\psi_n(\delta_2)$.*

Letting $N_{\rm pos}$ and $N_{\rm neg}$ denote the number of false positives and false negatives, we note that the analysis of exact recovery shows that $\mathbb{E}[N_{\rm pos}]$ and $\mathbb{E}[N_{\rm neg}]$ behave as $o(1)$, from which Markov's inequality implies that the probability of any false positives or negatives vanishes. Instead, when $\Theta(k)$ false positives are allowed, we simply show that $\mathbb{E}[N_{\rm pos}] = o(k)$, and use Markov's inequality to conclude that the probability of having $\Theta(k)$ false positives tends to zero. The same argument is used when $\Theta(k)$ false negatives are allowed.

Using the the first part of Lemma 4, we find that obtaining $\mathbb{E}[N_{\rm pos}] = o(k)$ instead of the stricter $\mathbb{E}[N_{\rm pos}] = o(1)$ amounts to replacing $\log \frac{p-k}{\delta_1}$ by $\log \frac{p-k}{k\delta_1}$ in (22). Moreover, using the second part of Lemma 4, we find that obtaining $\mathbb{E}[N_{\rm neg}] = o(k)$ instead of the stricter $\mathbb{E}[N_{\rm neg}] = o(1)$ amounts to replacing the requirement $k \cdot \psi_n(\delta_2) \to 0$ by $\psi_n(\delta_2) \to 0$ in (23).

Consequently, with false positives we can replace $k \log p$ by $k \log \frac{p}{k}$ in the first terms of (30) and (34), and with false negatives, we can remove the second terms therein. Generally, when *both* false positives and false negatives are allowed, we have the following simple corollary.

Corollary 3. (General noise models, partial recovery) *For any group testing model such that the quantities $c_{\rm mean}$, $c_{\rm var}$, and $c_{\rm max}$ in* (25)–(27) *behave as $\Theta(1)$, we can achieve $P_{\rm e}(d_{\rm pos}, d_{\rm neg}) \to 0$ with separate decoding provided that $d_{\rm pos} = \Theta(k)$, $d_{\rm neg} = \Theta(k)$, and*

$$n \geq \frac{\log \frac{p}{k}}{I_1}(1+\eta) = \frac{k \log \frac{p}{k}}{c_{\rm mean}}(1+\eta) \tag{36}$$

for some $\eta > 0$.

Hence, while we only obtained the threshold on the right-hand side of (36) in the limit $\theta \to 0$ under exact recovery (see (35)), when we allow a small fraction of false positives and false negatives, this extends to *all* sublinear sparsity levels. We again refer to the examples in Figure 1, where we are within a factor $\log 2$ of the optimal information-theoretic threshold.

To our knowledge, these partial recovery guarantees are the best known for any practical group testing algorithm for all $\theta \in (0, 1)$, in both the noiseless and symmetric noise settings.

Acknowledgment

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 725594 – time-data), and by an NUS startup grant.

References

[1] R. Dorfman, "The detection of defective members of large populations," *Ann. Math. Stats.*, vol. 14, no. 4, pp. 436–440, 1943.

[2] M. Aldridge, L. Baldassini, and O. Johnson, "Group testing algorithms: Bounds and simulations," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3671–3687, June 2014.

[3] D. Malioutov and M. Malyutov, "Boolean compressed sensing: LP relaxation for group testing," in *IEEE Int. Conf. Acoust. Sp. Sig. Proc. (ICASSP)*, March 2012, pp. 3305–3308.

[4] C. L. Chan, P. H. Che, S. Jaggi, and V. Saligrama, "Non-adaptive probabilistic group testing with noisy measurements: Near-optimal bounds with efficient algorithms," in *Allerton Conf. Comm., Ctrl., Comp.*, Sep. 2011, pp. 1832–1839.

[5] M. B. Malyutov and P. S. Mateev, "Screening designs for non-symmetric response function," *Mat. Zametki*, vol. 29, pp. 109–127, 1980.

[6] M. Malyutov and H. Sadaka, "Maximization of ESI. Jaynes principle for testing significant inputs of linear model," *Rand. Opt. Stoch. Eq.*, vol. 6, no. 4, pp. 339–358, 1998.

[7] M. B. Malyutov, "Search for sparse active inputs: A review," in *Inf. Theory, Comb. and Search Theory*, 2013, pp. 609–647.

[8] J. Scarlett and V. Cevher, "Phase transitions in group testing," in *Proc. ACM-SIAM Symp. Disc. Alg. (SODA)*, 2016.

[9] M. Aldridge, "The capacity of Bernoulli nonadaptive group testing," 2015, http://arxiv.org/abs/1511.05201.

[10] T. Laarhoven, "Asymptotics of fingerprinting and group testing: Tight bounds from channel capacities," *IEEE Trans. Inf. Forens. Sec.*, vol. 10, no. 9, pp. 1967–1980, 2015.

[11] J. Scarlett and V. Cevher, "Limits on support recovery with probabilistic models: An information-theoretic framework," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 593–620, 2017.

[12] M. Malyutov, "The separating property of random matrices," *Math. Notes Acad. Sci. USSR*, vol. 23, no. 1, pp. 84–91, 1978.

[13] G. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1880–1901, March 2012.

[14] M. Aldridge, L. Baldassini, and K. Gunderson, "Almost separable matrices," *J. Comb. Opt.*, pp. 1–22, 2015.

[15] M. Aldridge, "On the optimality of some group testing algorithms," in *IEEE Int. Symp. Inf. Theory*, 2017.

[16] C. L. Chan, S. Jaggi, V. Saligrama, and S. Agnihotri, "Non-adaptive group testing: Explicit bounds and novel algorithms," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 3019–3035, May 2014.

[17] O. Johnson, M. Aldridge, and J. Scarlett, "Performance of group testing algorithms with constant tests-per-item," 2016, submitted to *IEEE Trans. Sig. Proc.*

[18] S. Cai, M. Jahangoshahi, M. Bakshi, and S. Jaggi, "Efficient algorithms for noisy group testing," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2113–2136, Apr. 2017.

[19] K. Lee, R. Pedarsani, and K. Ramchandran, "SAFFRON: A fast, efficient, and robust framework for group testing based on sparse-graph codes," 2015, https://arxiv.org/abs/1508.04485.

[20] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer, 2003.

[21] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.

[22] Y. Polyanskiy, V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[23] A. Feinstein, "A new basic theorem of information theory," *IRE Prof. Group. on Inf. Theory*, vol. 4, no. 4, pp. 2–22, Sept. 1954.

[24] C. E. Shannon, "Certain results in coding theory for noisy channels," *Information and Control*, vol. 1, no. 1, pp. 6–25, 1957.

[25] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.

[26] J. Scarlett and V. Cevher, "Near-optimal noisy group testing via separate decoding of items," https://arxiv.org/abs/1710.08704.