ISSN 1064-2269, Journal of Communications Technology and Electronics, 2025.

Adaptive Group Testing Algorithm

K. Yu. Gorbunov*

Charkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, 127051 Russia

**e-mail: gorbunov@iitp.ru*

Received April 10, 2025; revised April 30, 2025; accepted April 30, 2025

Abstract—An algorithm is described for group testing under the conditions when, in a specified set of cardinality n, there is an unknown amount of $d > 0$ of defective elements and it is necessary to identify them with a sequence of tests. Each next test can take into account the results of the previous ones. The upper estimate for the number of tests in our algorithm improves the known value. The algorithm runtime is the minimum possible: it is of the order of the number of tests.

Keywords: group testing, adaptive testing, testing algorithm, search for defects, identification of infected individuals

DOI: 10.1134/S1064226925700238

1. INTRODUCTION

In group testing tasks, it is required to identify all defective elements (in terms of medicine, infected individuals) in specified set S of cardinality n using a minimum number of tests, i.e., inqueries about the presence of at least one defective element in a test subset. These problems differ in two criteria: whether number d of infected individuals (or its estimate) is a priori known and whether it is allowed to conduct tests sequentially (adaptive testing) or only in parallel. Parallel testing was discussed in studies [1–3], where further references were given. Sequential testing was considered in study [4], where numerous references on the history of the problem were given. We are going to improve the result of this study.

Now, let us consider the case of adaptive testing, when the number $d > 0$ is not known in advance and the tests are carried out sequentially, one after another. Under these conditions, in [4] the estimate $d\log\frac{n}{d} + (5-\log 5)d + O(\log^2 d)$ was obtained (all logarithms are hereinafter taken to base 2). The key issue of this problem is to reduce the coefficient at the second term. For this problem, we will describe an algorithm with the best upper bound $d\log\frac{n}{d} + (7-\log 21)d + O(\log^2 d)$. The general scheme of our algorithm is the same as in [4] and the proof of the estimate is somewhat simpler. In the description of the algorithm, we preserve, as far as possible, the terminology and notations from [4].

2. DESCRIPTION OF THE ALGORITHM

The algorithm uses two auxiliary procedures, DIG and 4-Split, and the numerical sequence $a_0 = 1$, $a_1 = 2$, $a_2 = 4$, $a_3 = 6$, and $a_i = 11\times 2^{i-4}$ for $i > 3$. Each procedure is fed with set X, which is known to contain at least one defective element.

DIG procedure. This procedure solves the problem of identifying one defective element in X. As long as $|X| > 1$, we choose arbitrary subset Y with cardinality $\lceil \frac{X}{2} \rceil$ (upper integer part) in X, test it, and, if the result is positive (i.e., Y contains a defective element), assume $X = Y$; otherwise, $X = X\setminus Y$. If one remains element in X (recall, $d > 0$), we identify it as defective. It can be easily proven by induction on $|X|$ that the number of tests in DIG is $\log|X|$, if $|X|$ is the power of two, and does not exceed $1+\log|X|$ otherwise. Note that, when DIG yields a negative result, we do not consider elements of the corresponding set to be identified.

4-Split procedure. It gets at the input, along with X, the parameter: natural number v, such that $|X| \le a_v$. If $|X| \le 4$, DIG is applied to X. If $|X| = 5$ or $|X| = 6$, we choose arbitrary subset X' with cardinality 2 in X and test it. If the result is positive, we assume $X = X'$; otherwise, we identify all elements from X' as nondefective and assume $X = X\setminus X'$. Then, we apply the DIG procedure to X. If $|X| \ge 7$, we randomly split X into sets Y, Z, U, and V (some of them may be empty), where $|Y| = \min\{2^{v-2}, |X|\}$, $|Z| = \min\{3\times 2^{v-4},$

$|X|-|Y|\}$, $|U| = \min\{2^{v-3}, |X|-|Y|-|Z|\}$, and $|V| = |X|-|Y|-|Z|-|U|$ (obviously, $|V| \le 2^{v-3}$). If Z is empty, we apply DIG to Y. Otherwise, we test the sets Y, Z, U, and V in this order until the first positive result, identifying elements of the sets with a negative result as nondefective (if the first three tests are negative, the fourth is not performed). Let us denote the first defective set by X'.

If $X' = Y$, $X' = U$, or $X' = V$, we apply DIG to X'. If $X' = Z$, we choose arbitrary subset Z' with the cardinality $\min\{|Z|, 2^{v-4}\}$ in Z and test it. If the result is positive, we assume $X = Z'$; otherwise, we identify all elements from Z' as nondefective and, if $Z \backslash Z'$ is non-empty, assume $X = Z \backslash Z'$. Then, we apply the DIG procedure to$X$. Note that the aim of the 4-Split procedure is either to reduce the search area by more than half or to obtain nondefective identified elements.

Basic algorithm. At each iteration, current set S is formed to which certain natural rank k is assigned. At the first iteration, we assume S to be equal to initial S and its rank to be equal to the minimum k value, such that $n \le a_k$. The iterations are performed until S is non-empty.

Let us describe the content of the iteration. We are given set S and its rank k. Let us choose, if possible, arbitrary subset S' of cardinality a_k in S or, otherwise, put it $S' = S$. We test S'. If the result is negative, we identify all elements from S' as nondefective, remove them from S, increase k by 1, and pass to the next iteration. If the result is positive, we do the following. If $k > 0$, we apply the 4-Split procedure with the parameter $v = k$ to S', remove the identified defective and nondefective elements from S, reduce k by 1, and pass to the next iteration. If $k = 0$, then S' contains exactly one element. We identify it as defective, remove it from S, retain $k = 0$, and pass to the next iteration.

According to the S' test result, the iteration is hereinafter called *positive* or *negative.*

3. ESTIMATION OF THE NUMBER OF TESTS IN THE ALGORITHM

Theorem. *The number of tests in the described algorithm does not exceed*

$$d\log\frac{n}{d} + (7 - \log 21)d + \frac{\log^2 d}{2} + 2\log d + 3.$$

Proof.

Lemma 1. *The number of tests in the 4-Split procedure with parameter* v *is no more than* v. *If, as a result of the 4-Split procedure, all elements from* X *are identified, the number of tests in it does not exceed* $2 + \log|X|$. *The number of tests in the last iteration of the algorithm is no more than* $3 + \log|X|$. *The ranks of all iterations do not exceed the rank* k *of the first iteration, which is no more than* $\log n + 5 - \log 11$.

Proof. Let $N(m)$ be the number of tests in 4-Split on set X of cardinality m. Obviously, $N(1) = 0$ ($v \ge 0$), $N(2) = 1$ ($v \ge 1$), $N(3) = N(4) = 2$ ($v \ge 2$), $N(5) \le 3$ ($v \ge 3$), and $N(6) \le 3$ ($v \ge 3$) and, for $m \le 6$, the first statement is satisfied. If $m \ge 7$, let us consider the cases.

Case 1. Subset Y is defective. Since $|Y| \le 2^{v-2}$, $N(m) \le 1 + (v-2) = v - 1$ is valid.

Case 2. Subset Y is nondefective and subset Z' is defective. Since $|Z'| \le 2^{v-4}$, $N(m) \le 3 + (v-4) = v - 1$ is valid.

Case 3. Subsets Y and Z' are nondefective and subset $Z \backslash Z'$ is defective. Since $|Z \backslash Z'| \le |Z| - 2^{v-4} \le 3 \times 2^{v-4} - 2^{v-4} = 2^{v-3}$, $N(m) \le 3 + (v-3) = v$ is valid.

Case 4. Subsets Y and Z are nondefective and U is defective. Since $|U| \le 2^{v-3}$, $N(m) \le 3 + (v-3) = v$ is valid.

Case 5. Subsets Y, Z, and U are nondefective and subset V is defective. Since $|V| \le 2^{v-3}$ and V is not tested, $N(m) \le 3 + (v-3) = v$ is valid.

The second statement of the lemma at $|X| = 1$ is obvious. In the case $|X| \ge 2$, note that the last non-empty set of Y, Z', $Z \backslash Z'$, U, and V is single-element; therefore, the number of tests in the procedure is no more than 3. The third statement follows from the second one. The first part of the last statement of the lemma follows from the fact that the iteration of rank k acts on the set S' containing all unidentified elements from S and therefore is positive or is the last. The inequality $k \le \log n + 5 - \log 11$ is checked directly at $n \le 11$ and, at other n, follows from the fact $n \ge a_{k-1}$. □

Let us designate iterations of the algorithm by I_1, $I_2, \ldots, I_q$. As in [4], we divide the set of iterations into three classes: C_1, C_2, and C_3. Let r_i be the rank of iteration I_i, t_i be the number of tests on it, and n_i be the number of elements (defective and nondefective) identified on it. The class C_1 contains positive iterations of rank 0 and iteration I_q. The class C_2 contains iteration I_1 and, for each natural number r, $r_q < r < r_1$, the first iteration of rank r in the sequence $I_1, I_2, \ldots, I_q$. The class C_3 contains all the rest iterations.

Lemma 2. *The number of iterations from C_3 is even and they can be divided into pairs such that the ranks of the iterations from each pair differ by exactly 1 (we will denote these ranks as v and $v-1$), the iteration of rank v is positive, the iteration of rank $v-1$ is negative, and the subset S' corresponding to it has the cardinality a_{v-1} exactly.*

Since Lemma 2 coincides with Lemma 3.5 from [4], its proof is not given. □

In Lemma 2, we fix the partition into pairs called *zigzag pairs*. We will denote a zigzag pair as (I_i, I_j), where the rank of iteration I_i is v and the rank of iteration I_j is $v-1$. The corresponding number v is called the *rank of a pair*.

Lemma 3. For each zigzag pair $P = (I_i, I_j)$, the inequality

$$t_i + t_j \le \log(n_i + n_j) + 7 - \log 21$$

is valid.

Proof. Let $v \ge 1$ be the rank of pair P. At $v = 1$, we have $t_i + t_j = 3$ and $n_i + n_j = 2$, from which the statement of the lemma follows. At $v = 2$, we have $t_i + t_j = 4$ and $n_i + n_j = 3$ and the statement of the lemma. At $v = 3$, we have $t_i + t_j = 4$ and $n_i + n_j = 5$ or $t_i + t_j = 5$ and $n_i + n_j = 7$ and the statement of the lemma. At $v \ge 4$, the procedure 4-Split is applied in iteration I_i and splits the corresponding set S_i' into four parts Y, Z, U, and V. Let us consider the cases.

Case 1. Subset Y is defective. Since $|Y| \le 2^{v-2}$, $t_i + t_j \le 3 + (v-2) = v+1$ is valid. Since $\left|S_j'\right| = a_{v-1} = 11 \times 2^{v-5}$, we have $n_i + n_j \ge 11 \times 2^{v-5}$, from which we have $v + 1 \le \log(n_i + n_j) + 6 - \log 11$ and the statement of the lemma.

Case 2. Subset Y is nondefective and subset Z' is defective. Since $|Z'| \le 2^{v-4}$, $t_i + t_j \le 5 + (v-4) = v+1$ is valid. Then, $n_i + n_j \ge 2^{v-2} + 11 \times 2^{v-5} = 19 \times 2^{v-5}$, from which we have $v + 1 \le \log(n_i + n_j) + 6 - \log 19$ and the statement of the lemma.

Case 3. Subsets Y and Z' are nondefective and subset $Z \backslash Z'$ is defective. Since $|Z \backslash Z'| \le |Z| - 2^{v-4} \le 3 \times 2^{v-4} - 2^{v-4} = 2^{v-3}$, we have $t_i + t_j \le 5 + (v-3) = v+2$. Then, $n_i + n_j \ge 2^{v-2} + 2^{v-4} + 11 \times 2^{v-5} = 21 \times 2^{v-5}$, from which we have $v + 2 \le \log(n_i + n_j) + 7 - \log 21$ and the statement of the lemma.

Case 4. Subsets Y and Z are nondefective and U is defective. Since $|U| \le 2^{v-3}$, we have $t_i + t_j \le 5 + (v-3) = v+2$. Then, $n_i + n_j \ge 2^{v-2} + 3 \times 2^{v-4} + 11 \times 2^{v-5} = 25 \times 2^{v-5}$ is valid, from which we have $v + 2 \le \log(n_i + n_j) + 7 - \log 25$ and the statement of the lemma.

Case 5. Subsets Y, Z, and U are nondefective and subset V is defective. Since $|V| \le 2^{v-3}$ and V are no tested, we have $t_i + t_j \le 5 + (v-3) = v+2$. Then, $n_i + n_j \ge 2^{v-2} + 3 \times 2^{v-4} + 2^{v-3} + 11 \times 2^{v-5} = 25 \times 2^{v-5}$ is valid, from which we have $v + 2 \le \log(n_i + n_j) + 7 - \log 29$ and the statement of the lemma. □

Let T_i $(i = 1,2,3)$ be the number of tests executed across all iterations of C_i.

Lemma 4. Let x $(0 \le x \le d)$ be the number of elements identified as defective on the interations from $C_1 \cup C_3$. If $x = 0$, then $|T_1 \cup T_3| = 1$. Otherwise, $|T_1 \cup T_3| \le 1 + (7 - \log 21)x + x \log \dfrac{n}{x}$.

Proof. If $x = 0$, the ranks of the iterations decrease monotonically from r_1 to r_q and all iterations, except for the negative iteration I_q, lie in C_2, from which the first statement follows. To prove the second statement, let us consider the cases.

Case 1. Iteration I_q is negative. Denoting the zigzag pair by (I_i, I_j), according to Lemma 3, we obtain

$$|T_1 \cup T_3| \le |C_1| + \sum_{(I_i, I_j)} (\log(n_i + n_j) + 7 - \log 21)$$
$$= |C_1| + \frac{|C_3|(7 - \log 21)}{2} + \sum_{(I_i, I_j)} \log(n_i + n_j).$$

Note that $x = |C_1| - 1 + \dfrac{|C_3|}{2}$. Let us add the zero term to the resulting expression $(|C_1| - 1)\log 1$, so that the number of logarithmic terms be equal to x and the sum of the arguments of the logarithms be no more than n. Using the convexity of the logarithmic function and the inequality $|C_1| \ge 1$, we obtain

$$|T_1 \cup T_3| \le |C_1| + \frac{|C_3|(7 - \log 21)}{2} + x \frac{\sum_{(I_i, I_j)} \log(n_i + n_j) + (|C_1| - 1)\log 1}{x}$$
$$\le |C_1| + \frac{|C_3|(7 - \log 21)}{2} + x \log \frac{n}{x} \le 1 + (7 - \log 21)x + x \log \frac{n}{x}.$$

Case 2. Iteration I_q is positive. According to Lemmas 1 and 3, we have

$$|T_1 \cup T_3| \le |C_1| + \sum_{(I_i,I_j)} (\log(n_i + n_j) + 7 - \log 21)$$
$$\le (|C_1| - 1) + 3 + n_q + \frac{|C_3|(7 - \log 21)}{2} + \sum_{(I_i,I_j)} \log(n_i + n_j)$$
$$= 2 + |C_1| + \frac{|C_3|(7 - \log 21)}{2} + \sum_{(I_i,I_j)} \log(n_i + n_j).$$

Note that $x = |C_1| + \frac{|C_3|}{2}$. Let us add the zero term to the resulting expression $(|C_1| - 1)\log 1$. As above, obtain

$$|T_1 \cup T_3| \le 2 + |C_1| + \frac{|C_3|(7 - \log 21)}{2} + x\frac{\sum_{(I_i,I_j)} \log(n_i + n_j) + (|C_1| - 1)\log 1 + \log n_q}{x}$$
$$\le 2 + |C_1| + \frac{|C_3|(7 - \log 21)}{2} + x\log\frac{n}{x} \le 1 + (7 - \log 21)x + x\log\frac{n}{x}. \square$$

Lemma 5. Let x ($0 \le x \le d$) be the number of elements identified as defective on the iterations from $C_1 \cup C_3$ and k be the rank of iteration I_1. Then, we have $|T_2| \le \frac{((2k+3) - (d-x))(d-x)}{2}$.

Since Lemma 5 coincides with Lemma 3.10 from [4], its proof is not given. □

Lemma 6. At $1 \le x \le d$, the inequality $x\log x \ge d\log d - (\log d + \log e)(d - x)$ is valid.

Proof. Let us consider the function $f(y) = (d - y)\log(d - y)$ at $0 \le y \le d - 1$ and differentiate it twice: $f'(y) = -\log(d - y) - \log e$, $f''(y) = \frac{\log e}{d - y}$. Since the second derivative is positive everywhere, the approximation of the $f(y)$ function tangent at the point $y = 0$ is no more than the $f(y)$ value. Hence, $d\log d - x\log x \le d\log d - (d\log d - (\log d + \log e)(d - x))$, from which the statement of the lemma follows. □

Let us first check the theorem for $x = 0$. According to Lemmas 1, 4, and 5, we obtain

$$|T_1 \cup T_2 \cup T_3| \le 1 + \frac{((2k+3) - d)d}{2}$$
$$\le \frac{(2(\log n + 5 - \log 11) + 3 - d)d}{2}$$
$$= 1 + d\log n + (6.5 - \log 11)d - 0.5d^2.$$

It is required to show that the last expression does not exceed $d\log n - d\log d + (7 - \log 21)d + 3$, i.e., $0.5d^2 - d\log d + (0.5 - \log 21 + \log 11)d + 2 \ge 0$. Since $0.5 - \log 21 + \log 11 \ge -0.5$ is valid, it is sufficient to demonstrate that $0.5d^2 - d\log d - 0.5d + 2 \ge 0$. Let us check this by examining the function $g(d) = 0.5d^2 - d\log d - 0.5d$ for a minimum at natural d. We have $g'(d) = d - \log d - \log e - 0.5$, $g''(d) = 1 - \frac{\log e}{d}$; therefore, $g(d)$ increases at $d \ge 4$. Thus, it is sufficient to calculate the minimum from $g(1)$, $g(2)$, $g(3)$, and $g(4)$, which, as can be easily seen, is $g(4) = -2$.

Let $x \ge 1$. According to Lemmas 1, 4, and 5, we have

$$|T_1 \cup T_2 \cup T_3| \le 1 + (7 - \log 21)x + x\log n - x\log x + \frac{(2(\log n + 5 - \log 11) + 3 - (d - x))(d - x)}{2}$$

$$= -0.5(d - x)^2 + (\log 21 - \log 11 - 0.5)(d - x) + (\log n + 7 - \log 21)d + 1 - x\log x$$

$$\le -0.5(d - x)^2 + 0.5(d - x) + (\log n + 7 - \log 21)d + 1 - x\log x.$$

According to Lemma 6, the last expression does not exceed

$$-0.5(d - x)^2 + (\log d + \log e + 0.5)(d - x) + (\log n + 7 - \log 21)d - d\log d + 1.$$

The sum of the first two terms is the quadratic function $g(y) = -0.5y^2 + (\log d + \log e + 0.5)y$ of $y = d - x$. Let us estimate its value at the maximum point: $g'(y) = -y + \log d + \log e + 0.5 = 0$, $y_{\max} = \log d + \log e + 0.5$,

$$g(y_{\max}) = -0.5(\log d + \log e + 0.5)^2 + (\log d + \log e + 0.5)^2 = 0.5(\log d + \log e + 0.5)^2$$

$$= 0.5\log^2 d + (\log e + 0.5)\log d + 0.5\log^2 e + 0.5\log e + \frac{1}{8} \le 0.5\log^2 d + 2\log d + 2.$$

We obtain $|T_1 \cup T_2 \cup T_3| \le d\log\frac{n}{d} + (7 - \log 21)d + g(y_{\max}) + 1$, which completes the proof of the theorem. □

4. CONCLUSIONS

We decreased the constant c in the estimate $d\log\frac{n}{d} + cd + O(\log^2 d)$ of the number of tests in the adaptive group testing algorithm to $c = 7 - \log 21$. At the same time, the algorithm time is minimum possible: of the order of the number of tests. The question of further decreasing this constant, possibly by increasing the algorithm runtime, is of interest. It would also be interesting to obtain a lower bound on the number of tests.

FUNDING

This study was supported by the State assignment of the Ministry of Science and Higher Education of the Russian Federation for the Kharkevich Institute for Information Transmission Problems of the Russian Academy of Sciences.

CONFLICT OF INTEREST

The author of this work declares that he has no conflicts of interest.

REFERENCES

1. A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth, and P. Loick, “Optimal group testing,” Combinatorics, Probab. Comput. **30**, 811–848 (2021). https://doi.org/10.1017/s096354832100002x
2. N. H. Bshouty, “Improved lower bound for estimating the number of defective items,” J. Comb. Optim. **49**, 33 (2025). https://doi.org/10.1007/s10878-025-01264-9
3. A. Coja-Oghlan, M. Hahn-Klimroth, L. Hintze, D. Kaaser, L. Krieg, M. Rolvien, and O. Scheftelowitsch, “Noisy group testing via spatial coupling,” Combinatorics, Probab. Comput. **34**, 210–258 (2025). https://doi.org/10.1017/s0963548324000336
4. J. Wu, Yo. Cheng, and D.-Z. Du, “An improved zig zag approach for competitive group testing,” Discrete Optim. **43**, 100687 (2022). https://doi.org/10.1016/j.disopt.2022.100687

Translated by E. Bondareva