Journal of Combinatorial Theory, Series A 126 (2014) 70–91

Contents lists available at ScienceDirect

Journal of Combinatorial Theory, Series A

www.elsevier.com/locate/jcta

Strict group testing and the set basis problem ☆

Peter Damaschke [a,*], Azam Sheikh Muhammad [a], Gábor Wiener [b]

[a] *Department of Computer Science and Engineering, Chalmers University, 41296 Göteborg, Sweden*
[b] *Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Budapest 1521, Hungary*

ARTICLE INFO

Article history:
Received 15 October 2013
Available online 8 May 2014

Keywords:
Group testing
Hypergraph
Set basis
Graph coloring
d-Disjunct matrix

ABSTRACT

Group testing is the problem to identify up to d defectives out of n elements, by testing subsets for the presence of defectives. Let $t(n, d, s)$ be the optimal number of tests needed by an s-stage strategy in the strict group testing model where the searcher must also verify that at most d defectives are present. We start building a combinatorial theory of strict group testing. We compute many exact $t(n, d, s)$ values, thereby extending known results for $s = 1$ to multistage strategies. These are interesting since asymptotically nearly optimal group testing is possible already in $s = 2$ stages. Besides other combinatorial tools we generalize d-disjunct matrices to any candidate hypergraphs, and we reveal connections to the set basis problem and communication complexity. As a proof of concept we apply our tools to determine almost all test numbers for $n \leq 10$ and some further $t(n, 2, 2)$ values. We also show $t(n, 2, 2) \leq 2.44 \log_2 n + o(\log_2 n)$.

☆ This is an extended version of a conference paper of the first two authors that appeared in the Proceedings of the *19th International Computing and Combinatorics Conference COCOON 2013*, Hangzhou (China), *Lecture Notes in Computer Science*, Springer, vol. 7936, pp. 446–457.

* Fax: +46 31 772 3663.
E-mail addresses: ptr@chalmers.se (P. Damaschke), azams@chalmers.se (A.S. Muhammad), wiener@cs.bme.hu (G. Wiener).

http://dx.doi.org/10.1016/j.jcta.2014.04.005
0097-3165/

1. Introduction

In the *group testing* problem, a set of n elements is given, each being either *defective* (*positive*) or *non-defective* (*negative*). Let P denote the unknown set of positive elements. A *group test* takes any subset Q of elements, called a *pool*. The test (or pool) is positive if $Q \cap P \neq \emptyset$, and negative otherwise. In the latter case all elements in Q are recognized as negative. The goal of a *searcher* is to identify P using a minimum number of tests. A group testing strategy may be organized in s *stages*, where the tests applied in a stage may depend on the outcomes of all tests in previous stages, and all tests within a stage are executed in parallel. In *adaptive* strategies s is not limited, hence tests can be done sequentially. Case $s = 1$ is called *nonadaptive*. Small s is desired in applications where the tests take much time. The term *pooling design* refers to a set of pools, especially within one stage. A pooling design can be written as a binary *matrix* whose rows and columns are the pools and elements, respectively. A matrix entry is 1 if the element is in the pool, and 0 else.

We consider the following scenario. The searcher expects $|P| \leq d$ for some previously known bound d, and $|P| > d$ is unlikely but not impossible. She wants to identify P if $|P| \leq d$, and just report "$|P| > d$" otherwise. This setting is called *strict group testing*, in contrast to *hypergeometric group testing* where $|P| \leq d$ is "promised". It was argued in, e.g., [1] that strict group testing is preferable. It does not rely on the assumption $|P| \leq d$, and the searcher is sure about not having missed any defective.

For complexity results and various applications of group testing we refer to [8,9,3,24,5,20,30,34] as entry points to further studies. For the test number in $s = 1$ stage, a lower bound $\frac{d^2}{2\log_2[e(d+1)/2]}\log_2 n + o(\log_2 n)$ was given in [13] and later refined in [16,17]. As opposed to that, $O(d \log n)$ tests are sufficient already if $s = 2$, as shown by the random coding upper bound in [14], followed by several improved constructions [15,10,6,18,17]. However, even asymptotically optimal strategies do not necessarily entail optimal strategies for specific input sizes n. Furthermore, pool sizes increase with n, whereas in some applications large pools may be infeasible. Still we can split an instance into many small instances and solve them independently, each with optimal efficiency. To mention a practical example, screening millions of blood donations for infectious diseases is performed at some labs in instances ("minipools") of, e.g., 16 samples [25], and group testing is proposed [35] to reduce the waiting times and costs. We also refer to [12] for biological applications of 2-stage strategies, with tests in the last stage being individual (whereas we will drop this restriction).

We define $t(n, d, s)$ to be the optimal worst-case number of tests needed for strict group testing for n elements, up to d defectives, and at most s stages. Some *monotonicity* relations hold trivially: If $n \leq n'$, $d \leq d'$, and $s \geq s'$ then $t(n, d, s) \leq t(n', d', s')$. If $t(n, d, s) = t(n, d, n)$, we write $t(n, d, s+)$ to indicate that more stages would not lower the test number.

To our best knowledge, the strict group testing model and the construction of optimal strategies for specific problem sizes in the multistage case are under-researched so far.