DISCRETE TOMOGRAPHY MODEL FOR A DOE PROBLEM WITH LIMITED RESOURCES

Levon Aslanyan, Hasmik Sahakyan

Abstract: This work is aimed at understanding the applied value of the mathematical problem of discrete tomography. Tomography, in general, is about the reconstructing of objects by sets of observable properties. Theoretically this is a typical inverse problem of combinatorial analysis. In applied level, in addition to the well-known task of tests and testing from the electrical engineering, complementary tasks of biomedical nature are considered. An example of the second task which is about treating sets of biological linear specimens with combinations from a limited resource of drugs (antibiotics) is considered aiming at achieving as many different treatment courses as possible to take place.

Keywords: discrete tomography, design of experiments.

ITHEA Keywords: G.2.3 Applications, F2.2. Nonnumerical Algorithms and Problems

1 Introduction

Problems where the object is given and it is necessary to calculate its characteristics - are called direct/forward problems, in contrast to the inverse problems in which the object is not available, and it is necessary to recover it based on a partial information, which is often given by measurements ([Heuberger, 2014], [DemangeMonnot, 2013].

Thus, inverse problems are a class of mathematical problems that arise when it is required to obtain information on internal or hidden data through external/available measurements. Inverse problems as a rule are ill-posed.
According to the Hadamar’s definition [Hadamard, 1902], a problem should have the following three properties to be considered as “well-posed”:

1. For all admissible data a solution exists,
2. For all admissible data the solution is unique,
3. The solution depends continuously on the data.

Problems that violate any of the three properties are "ill-posed". In the inverse problems the third condition is mainly violated, often referred to as a stability condition.

Tomography is a set of inverse problems about the reconstruction of an unknown object by means of partial data coming from its projections collected by means of X-rays, and taken along given directions. Typically, physical structures have a large variety of density values, and therefore a large number of X-rays are needed to recover the density distribution. In some cases, the object has a small number of density values (or the required number of directions, along which projections are taken is very limited in order to avoid physical damaging the objects) and thus a small number of X-rays is used. Discrete tomography is a domain to deal with these cases. In recent years, discrete tomography has been of research interest because of its mathematical formulation and the variety of applications. Discrete tomography is widely used, particularly in the processing of medical images. Applications are also related to the reconstruction of crystalline structures that are accessible only through some images provided by high-resolution transmission electron microscopy, and others ([SlumpGerbrands, 1982], [PrauseOnnasch, 1996], [IrvingJerrum, 1994], [Jinschek et al, 2004], [Kisieloski et al, 1995].

Discrete tomography, in the simplest case, considers an object $T$, which is a set of cells of the $n$-dimensional integer lattice $\mathbb{Z}^n$. A projection of $T$ in any direction calculates the number of points of $T$ on the lines parallel to the projection direction. Given a set \{ $d_1, d_2, \ldots, d_l$ \} of lattice directions and projections $P_1, P_2, \ldots, P_l$ along those directions. Consider Consistency and Reconstruction problems in Discrete Tomography.

**Consistency:** Does there exist a discrete set $T \in \mathbb{Z}^n$ with given projections $P_1, P_2, \ldots, P_l$ in lattice directions $d_1, d_2, \ldots, d_l$?
Reconstruction: Construct a discrete object \( T \in \mathbb{Z}^n \) from its projections \( p_1, p_2, \cdots, p_l \).

These are NP-hard problems for \( n \geq 2 \), and \( l \geq 3 \) non-parallel projections in the integer lattice \( \mathbb{Z}^n \) ([GardGrizmPran, 1999]).

Due to the complexity of the problem, a special attention has been given to the 2-dimensional case. Subsets of \( \mathbb{Z}^2 \) can be presented as binary images or binary matrices, where the 1s determine the cells of \( T \). Various researches are devoted to the case of orthogonal projections: horizontal and vertical. In terms of binary matrices, the row sum corresponds to the horizontal projection of \( T \), and the column sum corresponds to the vertical projection. In the case with only horizontal and vertical projections the problem has polynomial complexity, but the number of solutions can be large. Any prior knowledge /constraint/ about the image to be reconstructed, can reduce the search space of possible solutions. The existence problem under different geometrical constraints /convexity, connectivity/ is investigated by various authors (R.Gardner, P.Gritzmann, A.Del Lungo, E.Barcucci, M.Nivat, R.Pinzani, G.Woeginger, and others, [GardnerGritzmann, 1999], [Barcucci et al, 1996], [Woeginger, 2001]); for some cases the NP-completeness is proved, some particular cases can be solved by polynomial algorithms, but there also exist open problems in terms of complexity.

Summarizing, discrete tomography consistency and reconstruction problems can be presented through the model of weighted binary matrices (rows’ weights correspond to the horizontal projection and columns’ weights correspond to the vertical projections). The matrix model brings its own constraints, and one of them is the requirement of non-repetitiveness of the matrix rows. This constraint naturally appears in a number of applications; one of them is the design of experiments (DOE).

In this paper we consider a DOE problem with limited resources with examples from the biomedical field.

Firstly we compare the binary matrix model of this problem with a known mathematical problem of minimum test collections (MTC) on binary tables ([YablonskiiChegis, 1955], [ChegisYablonskii, 1958], [Solov’ev, 1978], [Dmitriev
et al, 1966)]. MTC is one of the basic NP-complete problems ([Karp, 1972], [GareyJonson, 1979]). The basic reference to MTC is “unpublished papers” by M. Garey and D. Jonson (see [GareyJonson, 1979]) but much earlier S. Yablonskii and I. Chegis investigated the problem in detail. An effective machine learning application was the work [Zhuravlev et al, 1966] by Yu. Zhuravlev et al. The input of the problem is a \((0,1)\) table of a given size \(m \times n\) with different rows. It is necessary to find subsets and, if possible, minimal size subsets of columns by which the rows still remind different. Usually the interpretation of this problem is given in terms of the electrical engineering. There are observable characteristics of electrical equipment (qualitative or quantitative), corresponding to the columns of the matrix. Matrix rows correspond to individual equipment that are physically in various malfunctioning states. The problem examines at first stage such sets of observable characteristics, by which the given \(m\) states (rows) differ one from the other (forming the initial input table), and then, the goal is to minimize the sets of columns by removing from it one or another group of observations keeping the rows different.

What is the similarity and difference between the two considered problems – the minimum test sets and the planning of a large number of different experiments with limited resources. The first task is a real optimization problem about physical objects and their properties. The second task relates to virtual reality - it asks about the reality of the plan of experiments, which are many, and which are based on the framework of the available resources.

Consider another comparison of the two mentioned problems on matrices with different rows. In MTC the matrix is given so that all rows and all columns are given beforehand. We just try to find proper row fragments composed by minimal number of columns keeping all rows different. In DOE with resource limitation what is given is the column weight vector. The general idea is in organizing as much as possible different and informative experiments, with an effective use of given limited resource of the problem.

Summarizing, the use of the binary matrix model for the considered DOE problem with limited resources makes it possible to apply known approaches
and algorithms developed for the construction of binary matrices with given projections and given constraints.

As an example, a greedy algorithm, developed in [Sahakyan, 2010], [SahakyanAslanyan, 2011] is adjusted/modified and applied in the design of biomedical experiments with limited resources.

The paper is organized as follows: below in Chapter 2 the DOE problem is introduced and modeled in terms of binary matrices. Chapter 3 brings a brief description of the greedy approximation algorithm developed in [Sahakyan, 2010], [SahakyanAslanyan, 2011] for constructing binary matrices with different rows. Chapter 4 introduces some modifications of the algorithm.

2 Problem Definition

Design of experiments (DOE) is a research domain ([Fisher], [Bose, 1939], [Rao, 1996], which helps in assigning treatments to the experimental units in a way of optimizing several characteristics of experiments such as the evaluated output value dispersion, or the computational complexity, etc. Combinatorial block design ([BhattacharyaSinghi, 2013], [Beth et al, 1985], [ColbournDinitz, 2006], one of the constituents of the DOE theory, combines units into homogeneous groups to achieve the goal of the DOE.

Consider the following medical-biological DOE problem.

Multidrug resistance problem is well known in biomedicine. According to the WHO (World Health Organization), most pathogenic species in existence today have developed resistance to one or more antimicrobials. E.g., the multidrug resistance of Escherichia coli (an essential component of the digestive tract flora of healthy humans and even most animals) has increased from 7.2% during the 1950s to 63.6% during the 2000s. One of the ways of combating multidrug resistance is the use of antibiotic combinations (cocktail) [Hickman, 2011].

Suppose that for a biomedical experiment, performed to gain a practical knowledge of antibiotic combinations, there are $n$ antibiotics of different type
given in limited quantities/portions. A cocktail combines portions of several of the given \( n \) antibiotics. Let, for the \( i \)-th antibiotic \( s_i \) portions only of the drug are given \((i = 1, \ldots, n)\).

**The problem is the following:** design/plan experiments, which means creation of given number of different cocktails so that they use the whole available drug store \( s_1, s_2, \ldots, s_n \).

It is possible to describe very large number of similar DOE scenarios. From technical point of view it is important to mention that the resources \( s_1, s_2, \ldots, s_n \) are for single use. If they describe not drugs but bacteria, then these are not cultivated.

In addition to this it is specifically important to mention that the non-repetition of experiments is a natural and valid requirement of the experimental design in these cases.

Let us also mention that each cocktail has its experimental value and that we do not consider general or economic optimization issues.

The scope of the problem solutions can be wide, but we are interested in:

- Finding/composing one of the solutions, because it clarifies the possibility of planning experiments with given quantities of antibiotics (otherwise, the composition of available samples should be changed);
- Composing a solution, which involves as much as possible different cocktails with given quantities of antibiotics.

Thus, the problem can be formulated in the following way:

**Antibiotics_Cocktail \((A_C)\):** Given \( n \) antibiotics of limited \( s_1, s_2, \ldots, s_n \) quantities, correspondingly.

(1) Decide whether it is possible to design an experiment with the given number \( m \) of different cocktails (subset, combination) so that \( s_1, s_2, \ldots, s_n \) quantities are used;

(2) Compose as much as possible different cocktails using \( s_1, s_2, \ldots, s_n \) quantities of antibiotics.
Each cocktail can be presented by a binary vector of length \( n \) such that the \( j \)-th component of the vector is 1 if and only if the \( j \)-th antibiotic is used in the cocktail. In this manner an experiment \( E \) with \( m \) different cocktails corresponds to a binary matrix \( A = \{ a_{i,j} \} \) with \( m \) rows and \( n \) columns; the number of 1s in the \( j \)-th column of \( A \) is equal to the quantity of the \( j \)-th antibiotic used in the experiment; the number of 1s in the \( i \)-th row of \( A \) equals the number of antibiotics used in the \( i \)-th cocktail.

The problem in terms of binary matrices has the following formulation.

**Existence of binary matrices with given column sum and with different rows** (**CS.D**): Given non-negative integer vector \( S = (s_1, s_2, \ldots, s_n) \) and a natural number \( m \).

1. Decide whether there is a binary matrix \( A = \{ a_{i,j} \} \) of size \( m \times n \) with all different rows and with the column sum vector \( S = (s_1, s_2, \ldots, s_n) \);  
2. Compose a binary matrix with the column sum \( S = (s_1, s_2, \ldots, s_n) \) and with maximum possible number of different rows.

Thus \( A, C \) is reduced to the combinatorial problem **CS.D**, which is known as a hard computational problem.

In the process of seeking efficient approximate solutions a greedy heuristics is investigated and an algorithm is developed in [Sahakyan, 2010], [SahakyanAslanyan, 2011] for constructing binary matrices with given column sums and with different rows. In the next sections we briefly introduce the algorithm, and also mention some peculiarities of the *Antibiotics_Cocktail* problem and the algorithm itself, which makes it expedient using the algorithm for this problem.
3 Optimization Version and Approximation Algorithm

For a given non-negative integer vector \( S = (s_1, s_2, \ldots, s_n) \) let \( U(S) \) denote the class of binary matrices of size \( m \times n \), which have the column sum vector \( S = (s_1, s_2, \ldots, s_n) \).

For a matrix \( A \) of \( U(S) \) let \( DP(A) \) denote the number of disjoint pairs of rows of \( A \); Obviously, \( 0 \leq DP(A) \leq C_m^2 \), and \( DP(A) = C_m^2 \) if and only if the rows of \( A \) are different.

Now we consider the following optimization version of \( CS_D \) (1) with the objective function \( DP \).

\((CS_{D^{opt}})\): find \( A_{opt} \in U(S) \) such that \( DP(A_{opt}) = \max_{A \in U(S)} DP(A) \).

It is clear that any solution of \( (CS_{D^{opt}}) \) is also a solution for \( CS_D \) for the cases when \( U(S) \) contains also matrices with different rows.

Now we bring a brief description of the algorithm \( G \) introduced in [SahakyanAslanyan, 2011], [Sahakyan, 2010].

Algorithm \( G \) for the problem \( CS_{D^{opt}} \)

**Input:** non-negative integer vector \( S = (s_1, \ldots, s_n) \) and natural number \( m \).

The algorithm \( G \) constructs a binary matrix \( A_{G,n} \) of size \( m \times n \) in the column-by-column manner, putting \( s_i \) 1s in the \( i \)-th column \( (i = 1, \ldots, n) \) having the goal to maximize the increase of the objective function (the number of disjoint pairs of rows) in each step.

**Step 1.** Construction of the first column: put 1s in the first \( s_1 \) rows and put 0s in the remaining \( m - s_1 \) rows. In the first column we get an interval of 1s of length \( s_1 \) (i.e. \( s_1 \) consecutive 1s) and an interval of 0s of length \( m - s_1 \) (\( m - s_1 \) consecutive 0s). Then, the number of different pairs of rows after the construction of the first column will be equal to \( s_1 \cdot (m - s_1) \). We notice that any other distribution of \( s_1 \) 1s and \( m - s_1 \) 0s in the first column would produce the same number of different pairs of rows.
In each next step the algorithm splits every interval of the previous step (column) into two parts, and puts 1s in one of them, and 0s in the other, such that the summary number of 1s equals the corresponding component of the column sum vector \( S = (s_1, \cdots, s_n) \).

Suppose that the first \((k - 1)\) columns of the matrix are constructed, and let the \((k - 1)\)-th column consists of \(p\) intervals of non-zero lengths, denoted by:
\[
d_{k-1,1}^C, d_{k-1,2}^C, \ldots, d_{k-1,p}^C.
\]

**Step \(k\).** Construction of the \(k\)-th column: for \(i = 1, \cdots, p\), split the \(d_{k-1,i}^C\)-length interval into two parts, -denoting them by \(d_{k-1,i,0}^C\) and \(d_{k-1,i,1}^C\) - and put 0s and 1s respectively, such that:
\[
\sum_{i=1}^{p} d_{k-1,i,0}^C = m - s_k, \quad \sum_{i=1}^{p} d_{k-1,i,1}^C = s_k.
\]

Then the increase of the objective function will be equal to: and
\[
\sum_{i=1}^{p} d_{k-1,i,1}^C \cdot d_{k-1,i,0}^C.
\]

All rows of the matrix will be different if and only if the last column of the matrix consists of only one-length intervals.

The detailed description of the algorithm \(G\) and the proof of its local optimality (the maximal increase of the objective function is achieved in each step) is given in [Sahakyan, 2010]; and its performance guarantee is estimated in [Sahakyan, 2017]. On the other hand, experimental results given in [SahakyanAslanyan, 2011] (algorithm run for all binary matrices with \(n \leq 6\) columns and with \(m, m = 1, 2, \cdots, 2^n\) rows) show that the constructed matrices in the last column can have at most 2-length intervals; moreover in most cases there is only one 2-length interval in the last column (and for small values of \(m\) (\(m \leq 11\), all rows are different), which means that constructed matrices in most cases contain at most 1 pair of coinciding rows.
4 Modified Algorithm for the Problem “Antibiotics Cocktail”

In the experiments, which are related to the design of antibiotic combinations the general idea is in organizing as much as possible different and informative experiments (*), with an effective use (**) of the given limited resource of the problem. Combining (*) and (**) we define 2 simple strategies. The first is direct maximization of the number of experiments; and the second, focused on effective use of the problem resources, can be formulated as high column weights, i.e. every antibiotic is used at least in certain part of antibiotic combinations. In “Antibiotics Cocktail” problem we will assume that \( s_i \geq \frac{m}{2}, i = 1, \ldots, n \), that is every antibiotic is used at least in the half of antibiotic combinations.

Now we apply algorithm \( G \) to \textit{Antibiotics Cocktail (A\_C)} problem:

(1) Decide whether it is possible to design an experiment with given number of different cocktails such that \( s_1, s_2, \ldots, s_n \) quantities of antibiotics are used.

It is worth mentioning that with the supposition \( s_i \geq \frac{m}{2} \), the algorithm \( G \) can organize the splitting of intervals (in each column of the constructed matrix \( A_{G,n} \)) in such a way that the resulting matrix contains a row consisting of all 1s. It follows that in the case when \( A_{G,n} \) contains coinciding rows, at least two of them will be consisting of all 1s. Thus, \( A_{G,n} \) will contain either:

a) at most two coinciding rows (coinciding combinations of antibiotics, where each of them contains all types of antibiotics); or

b) more than two coinciding rows/combinations.

In the Case a) if all \( m \) rows of the matrix \( A_{G,n} \) are different, then \( A_{G,n} \) is the required solution. Otherwise, \( A_{G,n} \) contains one pair of coinciding rows consisting of all 1s (this can be either due to the non-optimality of the algorithm \( G \), or non-possibility to design \( m \) different experiments with \( s_1, s_2, \ldots, s_n \))
quantities. We remove one of the coinciding rows, and output the matrix with $m - 1$ rows and with the adjusted column sum vector $(s_1 - 1, \cdots, s_n - 1)$ (one amount of every antibiotic will remain unused).

In the Case b) we may assume that it is not possible to design $m$ different experiments with $s_1, s_2, \cdots, s_n$ quantities; however the constructed matrix $A_{G,n}$ provides a guaranteed number of pairs of different combinations of antibiotics (related to the optimal number).

Before considering part 2 of the problem, we formulate and prove the following lemma.

**Lemma**

Let $A$ be a binary matrix of size $m \times n$ with all different rows and with the column sum vector $S = (s_1, s_2, \cdots, s_n)$, where $s_i \geq m/2$ for $i = 1, \cdots, n$, and $s_j > m/2$ for some $j$, $1 \leq j \leq n$. Then, there exists a binary matrix of size $(m + 1) \times n$ with all different rows and with the same column sum vector $S = (s_1, s_2, \cdots, s_n)$.

**Proof.**

$s_j > m/2$ implies that $A$ contains a row (let it be the $i$-th row) with 1 in the $j$-th position such that $A$ does not contain the row differing from the $i$-th only by the $j$-th position, i.e. $(a_{i,1}, \cdots, a_{i,j-1}, 1, a_{i,j+1}, \cdots, a_{i,n}) \in A$, and $(a_{i,1}, \cdots, a_{i,j-1}, 0, a_{i,j+1}, \cdots, a_{i,n}) \notin A$. We append $(a_{i,1}, \cdots, a_{i,j-1}, 0, a_{i,j+1}, \cdots, a_{i,n})$ to the matrix $A$ (it will not cause row repetitions). The resulting matrix will have $m + 1$ different rows and a column sum vector $(s_1', s_2', \cdots, s_{j-1}', s_j, s_{j+1}', \cdots, s_n')$, where $s_k' \geq s_k$ for $k = 1, \cdots, n, k \neq j$.

Now we will modify $A$ into a binary matrix $A'$ of size $(m + 1) \times n$ with all different rows and with the column sum vector $S = (s_1, s_2, \cdots, s_n)$.

Suppose that $s_k' > s_k$ for some $k$ (in fact, $s_k' = s_k + 1$), then $s_k' > m/2$. It follows that there exists a row in $A$ (let it be the $t$-th row) with 1 in the $j$-th position:

$(a_{t,1}, \cdots, a_{t,k-1}, 1, a_{t,k+1}, \cdots, a_{t,n}) \in A$, 

such that \((a_{t,1}, \ldots, a_{t,k-1}, 0, a_{t,k+1}, \ldots, a_{t,n}) \notin A\). We replace 
\((a_{t,1}, \ldots, a_{t,k-1}, 1, a_{t,k+1}, \ldots, a_{t,n})\) with \((a_{t,1}, \ldots, a_{t,k-1}, 0, a_{t,k+1}, \ldots, a_{t,n})\); this will decrease \(s_k'\) by 1 and will not cause row repetitions.

By the same reasoning we make relevant row replacements for all \(s_k' > s_k\). □

\textbf{Antibiotics Cocktail (A,C) problem (part 2):}

(2) Suppose that it is possible to design \(m\) different cocktails with \(s_1, s_2, \ldots, s_n\) quantities of antibiotics; compose maximum possible number of different cocktails using the same quantities of antibiotics.

Firstly, we apply algorithm \(G\) and get as a result a binary matrix \(A\) (possibly with a small number repeated rows, which are further removed, and the column sum vector and the number of rows are adjusted) of size \(m \times n\) with all different rows and with the column sum vector \(S = (s_1, s_2, \ldots, s_n)\).

Now we introduce an algorithm \(M\) that increases the number of rows keeping the same column sum vector \(S\).

\textbf{Algorithm M.}

**Input:** matrix \(A\) of size \(m \times n\) with all different rows and with the column sum vector \(S = (s_1, s_2, \ldots, s_n)\);

\[ A' := A ; m' := m ; S' := S ; \]

\[ \text{While ( \(A'\) satisfies the Lemma conditions (all } s'_i \geq m'/2 \text{ and } s'_j > m'/2 \text{ for some } j ) )} \]

\{

find a row (according to the Lemma), and append it to \(A'\); \(m' := m' + 1\);

calculate new column sum vector \(S' = (s'_1, s'_2, \ldots, s_n)\) of \(A'\);

\[ \text{While ( } s'_k > s_k \text{ for some } k )} \]

\}
find and make appropriate row replacements in $A'$ (according to the proof of the Lemma),
calculate new column sum vector $S' = (s'_1, s'_2, \ldots, s'_n)$ of $A'$;
\}
\}

Output: the matrix $A'$ of size $m' \times n$ with all different rows and with the column sum vector $S = (s_1, s_2, \ldots, s_n)$, where $m' > m$.

Conclusion

A DOE problem with limited resources from the biomedical field is modeled by binary matrices as a discrete tomography problem with the constraint of non-repetitive rows. The problem is hard computationally, and in the process of seeking efficient approximate solutions, a known greedy algorithm developed for the construction of binary matrices with given projections and with all different rows, is adjusted/modified and applied to solve the problem.

Acknowledgements

This work is partially supported by the grants № 18RF-144, and № 18T-1B407 of the Science Committee of the Ministry of Education and Science of Armenia.

Bibliography


[Hickman, 2011] Hickman R., Could antibiotic cocktails be the answer? An in vitro study into antibiotic combinations with enhanced activity against Escherichia coli, Biology Education Centre and Department of Cell and Molecular Biology, Uppsala University, 2011.


Authors' Information

**Levon Aslanyan** – Institute for Informatics and Automation Problems of the National Academy of Sciences of Armenia, head of department; 1 P.Sevak str., Yerevan 0014, Armenia; e-mail: lasl@sci.am

Major Fields of Scientific Research: Discrete analysis – algorithms and optimization, pattern recognition theory, information technologies.

**Hasmik Sahakyan** – Institute for Informatics and Automation Problems of the National Academy of Sciences of Armenia, scientific secretary; 1 P.Sevak str., Yerevan 0014, Armenia; e-mail: hsahakyan@sci.am

Major Fields of Scientific Research: Combinatorics, Discrete Tomography, Data Mining.