Discrete Tomography with Distinct Rows: Relaxation

Hasmik, Sahakyan
Institute for Informatics and Automation Problems
Yerevan, Armenia
e-mail: hsahakyan@sci.am

Levon, Aslanyan
Institute for Informatics and Automation Problems
Yerevan, Armenia
e-mail: lasl@sci.am

ABSTRACT
In this paper we consider a discrete tomography problem, where a new constraint - the requirement of distinct rows is imposed. We focus on a relaxed version of the problem, where some constant number of repeated rows are allowed; and investigate the complexity of the relaxed problem, as well as obtain several properties/results.

Keywords
Discrete tomography, binary matrices, different rows, relaxation.

1. INTRODUCTION
Reconstruction algorithms that are intended to solve particular inverse problems, have many applications in areas such as the image processing, medicine, computer tomography assisted engineering and design, etc. A large number of well-known medical problems require discrete reconstruction technique ([1], [2]). For example, in angiography, the values 0 and 1 (or some other discrete values) can represent the absence or presence of a contrast agent in heart chambers. In radiation therapy planning the input data compose a three dimensional discrete matrix that is to be covered by rectangular shapes in an optimized way. This matrix is to expire by a series of beams with a special criterion on the borderline.

Reconstruction of discrete sets from given projections, - is one of the main tasks of *Discrete Tomography*. Discrete sets can be presented as binary images or matrices. The line/weight sum of a line through the image is the sum of the values or the weights of the points on this line. The projection of the image in a certain direction consists of all the weight sums of the collinear lines passing through the image in this direction. Any binary image with exactly the same projections as of the original image represents a reconstruction of that image.

Opposite to methods of Computerized Tomography which use several hundreds of projections, in Discrete Tomography a few projections are available. The main problem arising here is that there may appear different binary images with the same projections; and in case of a small number of projections the problem in this form can have a large number of solutions ([3]). For exactly two directions, the horizontal and vertical ones, in general it is possible to reconstruct an image in polynomial time ([4]). But in general, if only the horizontal and vertical projections are given, then the number of solutions can be exponentially large ([5]).

On the other hand, for any set of more than two directions, the problem of reconstructing a binary image from its projections in those directions is NP-complete.

One way to eliminate the mentioned problems is to suppose that there is some prior knowledge of the image to be reconstructed and this can reduce the search space of the possible solutions. It can be assumed that the image has some geometrical properties. Using geometrical knowledge about the discrete sets, such that convexity and connectivity, reconstruction is a well-studied area. However, the existence problems for convex matrices, as well as the existence problem for connected matrices are NP-complete ([6], [7]. In the meantime, the existence problem for horizontally and vertically convex and connected matrices can be solved in polynomial time ([8].

Another strategy here can be the search for a possibly good but not necessarily exact solution of the problem.

In this paper we impose a new and specific for the domain of discrete tomography constraint/property – assuming that the rows of the matrix to be reconstructed are distinct. The combinatorial origin of this constraint comes from the n-dimensional binary cube, and is related mainly to the problem of quantitative characterization of n-cube subsets partitioning ([9], [10]). We treat the row-difference constraint in line with the traditional constraints of connectivity or the convexity type.

Consider a binary matrix A of size $m{\times}n$. Let $S=(s_1,\cdots,s_n)$ denote the column sum vector of A, where s_j is the number of ones in the j–th column of A. Obviously, $0\le s_j\le m$ for each j, $1\le j\le n$.

For a given integer vector $S=(s_1,\cdots,s_n)$ let $U(S)$ denote the class of binary matrices of size $m{\times}n$, which have the column sum vector $S=(s_1,\cdots,s_n)$. Let $\overline{U}(S)$ denote the subclass of $U(S)$ where all matrices consist of all distinct rows.

Consider the following problem V_d:

(**V_d**)**: Existence/reconstruction** of a binary $m{\times}n$ matrix in the class $\overline{U}(S)$.

We consider also an optimization version of V_d:

For a binary matrix $A\in U(S)$, let $DR(A)$ denote the number of distinct rows of A.
(**V_d_opt**): Given an integer vector $S=(s_1,\cdots,s_n)$. Find $A_{opt}\in U(S)$ such that $DR(A_{opt})=max_{A\in U(S)}DR(A)$.

Clearly, $DR(A)\le m$, and $DR(A)=m$ if and only if $A\in\overline{U}(S)$. Thus, if $\overline{U}(S)$ is not empty then any solution of V_d_opt is a solution also for V_d .

No polynomial algorithm is known for solving the problem V_d (V_d_opt), and it is widely known that this problem is open in terms of algorithmic complexities ([11-13]).

In this paper we consider a relaxed version of the problem, where some constant number of repeated rows are allowed; and we investigate the complexity of the relaxed problem, as well as obtain several properties/results related to the topic.

2. RELAXATION/RESULTS

Let A be a binary matrix such that some rows of A allowed to appear more than one time. Let $i_1, \cdots, i_k$ denote the rows (first appearances) that have repetitions in the matrix, and let $t_1, \cdots, t_k$ denote the repetition group sizes, respectively. Then, $C = t_1 + \cdots + t_k - k$ is the number of rows repeated (appearing as repetitions) in A.

Possible values of the number of groups of repeated rows (denoted by k) in an m row matrix may vary between: $1 \le k \le m/2$.

And the possible values of the group sizes when the number of groups is k vary between: $2 \le t_r \le m - 2(k-1)$, where t_r denotes the size of the r-th group.

For a given integer vector $S = (s_1, \cdots, s_n)$ and a constant number $C > 0$ let $\overline{U}^C(S)$ denote the subclass of $U(S)$, where at most C number of repeated rows are allowed in matrices of $\overline{U}^C(S)$.

We consider the existence/reconstruction problem for the class $\overline{U}^C(S)$:

(***V_C_rep***): **Existence/reconstruction** of binary $m\times n$ matrices in the class $\overline{U}^C(S)$ given the column weight vector $(s_1, \cdots, s_n)$ and the repetition limitation C.

Let $D_m^C(n)$ denote the set of all integer vectors $S = (s_1, \cdots, s_n)$ for which $\overline{U}^C(S)$, the class of $m\times n$ binary matrices with column sum $S = (s_1, \cdots, s_n)$ and with at most C number of repeated rows, - is not empty.

$D_m^0(n)$ (or simply $D_m(n)$) denotes the case of distinct rows.

Obviously, $D_m(n) \subseteq D_m^1(n) \subseteq \cdots \subseteq D_m^C(n)$.
Let us note that the complete characterization of the set $D_m(n)$ is known [11-13]. Now we obtain results/properties for $D_m^C(n)$ (some of them have their analogues for $D_m(n)$ given in [11-13]).

Lemma 1. If $(s_1, \cdots, s_n) \in D_m^C(n)$ then $(s_1, \cdots, m - s_j, \cdots, s_n) \in D_m^C(n)$, for every j, $1 \le j \le n$.

Proof.

Let $(s_1, \cdots, s_n) \in D_m^C(n)$, and A be a binary $m\times n$ matrix with column sum vector $(s_1, \cdots, s_n)$, and with at most C number of repeated rows. If to invert elements of the j-th column (interchange ones and zeros), the resulting matrix will have the column sum $(s_1, \cdots, m - s_j, \cdots, s_n)$ and again will contain the same number of repeated rows. ■

It follows from Lemma 1 that it is to restrict the attention to the "upper" subclass $\widehat{D}_m^C(n)$ of vectors of $D_m^C(n)$, where $s_j \ge m/2$ for all $1 \le j \le n$.

Lemma 2. If $(s_1, \cdots, s_n) \in D_m^C(n)$ and $s_j > m/2$ for some index j then $(s_1, \cdots, s_j - 1, \cdots, s_n) \in D_m^C(n)$.

Proof.

Let $(s_1, \cdots, s_n) \in D_m^C(n)$ and A be a binary $m\times n$ matrix with column sum vector $(s_1, \cdots, s_n)$, and with at most C number of repeated rows (that is, $A \in \overline{U}^C(S)$). Consider the index j with $s_j > m/2$.

Consider cases:

a) All groups of repeated rows contain 0 in the j-th position, and thus all rows with 1 in the j-th position are distinct. $s_j > m/2$ implies that the number of ones in the j-th column of A is greater than the number of zeros, and hence within the distinct rows of A there exist a row (say, the i-th one) with 1 in the j-th position, such that A does not contain the row which differs from the i-th row only in the j-th position. Replacing this 1 in the j-th position of the i-th row with 0, we will not cause new repeating rows.

b) Some group of repeated rows contain 1 in the j-th position. Then, replacing 1 in the j-th position of some row of the group with 0 will decrease the number of repeating rows in this group, at the same time this will cause at most one new row repetition. ■

Definition. $(s_1, \cdots, s_n)$ is an *upper boundary element* of $D_m^C(n)$ if no $(q_1, \cdots, q_n)$ with $(q_1, \cdots, q_n) > (s_1, \cdots, s_n)$ belong to $D_m^C(n)$, i.e., there is no matrix with at most C repeated rows and with the column sum vector - $(s_1, \cdots, s_j + 1, \cdots, s_n)$ (for an arbitrary position j).

Lemma 3. Let $\hat{S}$ be an upper boundary element of $D_m^C(n)$. Then for every matrix A of $U(\hat{S})$:

a) $r = 1$ (number of groups of repeated rows), and the repeated rows consist of all ones, and
b) $t_1 - 1 = C$, i.e., there are exactly C repeated rows in the group mentioned in point a).

Proof.

a) Assume that the assertion is not true, and there is a matrix A in $U(\hat{S})$ that contains a group of repeated rows containing 0 in some position. Replacing 0 with 1 in some row of the group will move out the row from that group, decreasing the group size by 1. At the same time this may increase by 1 the size of some other group of repetitions (if such group exists). We get a matrix with the column sum vector greater than $\hat{S}$, and with at most C repeated rows. This contradicts the definition of upper boundary elements.

b) Suppose that the assertion is not true, and consider some matrix A in $U(\hat{S})$ that has less than C repeated rows. Recall that there is only one group of repetitions, and the repeated rows consist of all ones. Then, this group size is less than C. Consider an arbitrary row out of that group, and replace all 0s in this row with 1. This will lead to a matrix with the column sum vector greater than $\hat{S}$, and with one additional repeated row so that the total row repetition is restricted again by the same C, which contradicts the definition of upper boundary elements in this class. ■

The converse assertion, which is easy to check, is given by the following Lemma 4:

Lemma 4. If a binary $m\times n$ matrix A contains $C+1$ rows consisting of all ones, and the remaining rows are all distinct, then $\hat{S}$, the column sum vector of A is an upper boundary element of $D_m^C(n)$.

The following Lemma 5 gives the relation between the upper boundary elements of $D_m(n)$ and $D_{m+C}^{C-1}(n)$.

Lemma 5. If $(s_1,\cdots,s_n)$ is an upper boundary element of $D_m(n)$, then $(s_1+C,\cdots,s_n+C)$ is an upper boundary element of $D_{m+C}^C(n)$; and vice versa: for each upper boundary element $(q_1,\cdots,q_n)$ of $D_{m+C}^C(n)$, $(q_1-C,\cdots,q_n-C)$ is an upper boundary element of $D_m(n)$.

Proof.

Suppose that $(s_1,\cdots,s_n)$ is an upper boundary element of $D_m(n)$, and A is an $m\times n$ binary matrix with the column sum $(s_1,\cdots,s_n)$ and with distinct rows. Compose new matrix A' by appending C new rows, consisting of all ones, into A. Since A has already such a row (it follows from the result that the upper boundary elements of $D_m(n)$ correspond to the "one" values sets of corresponding monotone Boolean functions [11]), then A' will have $C+1$ rows consisting of all ones. Then, $(s_1+C,\cdots,s_n+C)$, the column sum vector of A' belongs to $D_{m+C}^C(n)$, and according to Lemma 4, it is an upper boundary element of $D_{m+C}^C(n)$.

Now suppose that $(q_1,\cdots,q_n)$ is an upper boundary element of $D_{m+C}^C(n)$, and A is an $m\times n$ binary matrix with the column sum $(q_1,\cdots,q_n)$. According to Lemma 3 A contains exactly C repeated rows consisting of all ones. It follows that $q_i\geq C$ for all i. Removing C number of repeated rows will lead to a matrix with all distinct rows and with column sum $(q_1-C,\cdots,q_n-C)$. ■

Theorem 1. Given an integer vector $S=(s_1,\cdots,s_n)$ with $m/2\leq s_i\leq m$. Then, $(V\_d)\propto(V\_C\_rep)$, i.e., V_d is polynomial reducible to V_C_rep.

Proof.

Let I_1 denote an instance of the problem V_d: given $S=(s_1,\cdots,s_n)$, and m. We construct the following instance I_2 of V_C_rep: $S'=(s_1+C,\cdots,s_n+C)$ and $m'=m+C$. Now we prove that I_1 is a positive instance for V_d if and only if I_2 is a positive instance of V_C_rep.

Suppose that I_1 is a positive instance for V_d, and A is a $m\times n$ binary matrix with the column sum $(s_1,\cdots,s_n)$ and with distinct rows. Then the matrix A', constructed by appending C rows consisted of all ones, to A, - will be a solution of instance I_2 for V_C_rep.

Now suppose that I_2 is a positive instance of V_C_rep, and A' is a $(m+C)\times n$ binary matrix A' with the column sum $(s_1+C,\cdots,s_n+C)$ and with at most C repeated rows. Assume that there are k groups of repeated rows in A', and $t_1,\cdots,t_k$ are the corresponding group sizes: $C'=(t_1-1)+\cdots+(t_k-1)$, and $C'\leq C$. Now, we leave only one row in each group of repeated rows, and remove the C' repeated rows from A' (in case if $C'<C$, we remove from A' also $C-C'$ arbitrary rows). The obtained $m\times n$ matrix will consist of all distinct rows, and let $S''=(s''_1,\cdots,s''_n)$ denote its column sum vector.

Suppose that $C'=C$, $k=1$, and the repeated rows consist of all ones. In this case, $S''=S$. Hence, I_1 is a positive instance for V_d.

Otherwise, $S''>S$, and in the meantime in all components where $s''_i>s_i$, we have also, $s''_i>m/2$. Then, there exists (and we can construct) a matrix with all distinct rows and the column sum $S=(s_1,\cdots,s_n)$ [11,13]. Hence, I_1 is a positive instance for V_d. ■

It follows from Theorem 1 that: if there is a polynomial time algorithm that solves the problem V_C_rep, then also there is a polynomial time algorithm that solves V_d.

Let A_d and A_{C_rep} denote solution matrices for $V\_d^{\ opt}$ and V_C_rep, respectively. Then:

$\mathrm{m}-\mathrm{C}\leq\mathrm{DP}(\mathrm{A_d})\leq\mathrm{m}$,

$DP(A_{C\_rep})\geq m-C$, and hence:

$DP(A_d)-DP(A_{C\_rep})\leq C$.

In this manner, any algorithm that gives a solution to V_C_rep, will serve as absolute approximation algorithm for $V\_d^{\ opt}$. And thus, the following theorem holds:

Theorem 2.

If there exists a polynomial time absolute C-approximation algorithm for $V\_d^{\ opt}$ then there exists also a polynomial time exact algorithm for V_d.

4. CONCLUSION

A discrete tomography problem is considered in this paper with a new constraint: the requirement of distinct rows in the matrix to be reconstructed. This is a specific constraint for the domain, whilst the combinatorial origin comes from the n-dimensional binary cube, and mainly is related to the problem of quantitative characterization of n-cube subsets partitioning.

This problem is known through a number of alternative representations being an open problem in terms of the algorithmic complexities.

In the process of seeking good approximate solutions for the problem we considered a relaxed version of the problem, where some constant number of repeated rows are allowed. We investigated the complexity of the relaxed version, and concluded that the relaxed version is not easier than the original problem.

5. ACKNOWLEDGEMENT

The research is partially supported by the Grant 15T-1B417 of SCS of MES RA, "Application Driven Inverse Type Constraint Satisfaction Problems". It was also supported in the frame of the Grant grant 15RF-083 of SCS of MES RA, and RFBR 15-51-05059 Арм_а, Russian Foundation.

REFERENCES

[1] G. Prause and D. Onnasch, Binary reconstruction of the heart chambers from biplane angiographic image sequence, *IEEE Transactions Medical Imaging*, 15, pp. 532-559, 1996.

[2] C. Slump, J. Gerbrands, A network flow approach to reconstruction of the left ventricle from two projections, *Comput. Gr. Im. Proc.*, 18, pp.18-36, 1982.

[3] Gardner R. J., Gritzmann P., Prangenberg D., On the computational complexity of reconstructing lattice sets from their X-rays, *Discrete Mathematics,* 202, pp. 45-71, (1999).

[4] H. Ryser, Combinatorial properties of matrices of zeros and ones, *Canad. J. Math. 9*, pp.371–377, 1957.

[5] A. Del Lungo, Polyominoes deffined by two vectors, *Theoretical Computer Science*, 127, pp.187-198, 1994.

[6] E. Barcucci, A. Del Lungo, M. Nivat, and R. Pinzani, Reconstructing convex polyominoes from horizontal and vertical projections, *Theoret. Comput. Sci.,* 155, pp.321-347, 1996.

[7] G. Woeginger, The reconstruction of polyominoes from their orthogonal projections, *Inform. Process. Lett.*, 77, pp. 225-229, 2001.

[8] Ch. Durr, M. Chrobak, Reconstructing hv-convex polyominoes from orthogonal projections, *Information Processing Letters*, 69, pp. 283-291, 1999.

[9] L. Aslanyan, The discrete isoperimetry problem and related extremal problems of discrete spaces, *Problemy. Kibernetiki*. 36, pp. 85–127, 1979 (in Russian).

[10] L. Aslanyan, H. Danoyan, On the optimality of the Hash-Coding type nearest neighbour search algorithm, *CSIT 2013 - 9th International Conference on Computer Science and Information Technologies, Revised Selected Papers*, 2013, 7 pages.

[11] H. Sahakyan, Numerical characterization of n-cube subset partitioning, *Discrete Applied Mathematics*, 157, 9, pp. 2191-2197, 2009.

[12] H. Sahakyan, Essential points of n-cube subsets partitioning characterization, *Discrete Applied Mathematics*, 163, 2, pp. 205-213, 2014.

[13] H. Sahakyan, (0,1)-matrices with different rows, *CSIT 2013 Revised Selected Papers*, IEEE conference proceedings, DOI: 10.1109/CSITechnol.2013.6710342.