DIMACS Technical Report 98-01
January 9, 1998

Closed Sets and Generators in Ternary Hamming Spaces[1]

by

Endre Boros[2] Peter L. Hammer[2] Federica Ricca[3] Bruno Simeone[3]

[1]The authors gratefully acknowledge the partial support by the Office of Naval Research (Grants N00014-92-J-1375 and N00014-92-J-4083), the Italian Ministry of University and Research, and the Italian National Research Council.
[2]Permanent Member of DIMACS and RUTCOR, Rutgers University, Piscataway, NJ 08854, USA ({boros,hammer}@rutcor.rutgers.edu).
[3]Dipartimento di Statistica, Probabilità e Statistische Applicate, Università "La Sapienza", Roma, Italy ({federica,marsalis}@rosd.sta.uniroma1.it).

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Labs-Research, Bell Labs and Bellcore.

DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

ABSTRACT

The n-dimensional ternary Hamming space is $\mathbf{T}^n$, where $\mathbf{T} = \{0, 1, 2\}$. Three points in $\mathbf{T}^n$ form a line if they have in common excactly $n-1$ components. A subset of $\mathbf{T}^n$ is closed if, whenever it contains two points of a line, it contains also the third one. A generator is a set, whose closure is $\mathbf{T}^n$. In this paper, we investigate several properties of closed sets and generators.

Two alternative proofs of our main result, stating that the minimum cardinality of a generator is 2^n, are provided. The present study was motivated by some combinatorial questions concerning origin-destination matrices in transportation systems.

1 Introduction

In transportation modelling, Origin-Destination (O/D) matrices are frequently used to represent travel demand. The region at hand is subdivided into a fixed number of elementary zones which are at the same time origins and destinations of possible trips. The surveyed data are organized in an O/D matrix whose generic cell (i, j) contains the number of trips observed from zone i to zone j, in the given time interval (day, month, week, ...). The observed O/D matrix is used in order to estimate the global trip distribution. Often, it is interesting and useful to classify data according to some specific characteristic, e. g., the reason of the trip, the mode of transport, the hour, the age of those who move, their sex and activity. In this paper we will consider only dichotomic characteristics, such as systematic/non-systematic shifts, public/private mode of transport, rush/non-rush hour, age under/over 65, male/female, employee/professional. Although this simple case is already interesting in itself, the same problem could be analyzed in the more general case of non-dichotomic attributes.

On the basis of a selected set of attributes, each O/D matrix can be labelled by the vector of the attribute values for that matrix. For example, consider all the attributes listed before; then a matrix can be characterized by the following vector:

$$M_1 = (\text{systematic, private, rush hour, under 65, female, employee}).$$

Sometimes, it is useful to deal with matrices which are obtained by summing up other two. For example, consider matrix M_1 and another matrix M_2 defined as follows:

$$M_2 = (\text{systematic, private, rush hour, under 65, male, employee}).$$

The sum of M_1 and M_2 gives the following matrix

$$M = (\text{systematic, private, rush hour, under 65, total, employee}),$$

where the value of the fifth attribute ("total") is given by the sum of those in M_1 and M_2.

Obviously, also subtractions between matrices are allowed. If M and M_1 are available, then M_2 can be obtained by subtraction. Notice that sum and subtraction can be performed only between matrices the attribute vectors of which differ by just one element. However, matrices which differ by more than one attribute can be obtained from each other through a series of operations of the type seen above.

Moreover, it is possible to combine also matrices whose labels contain one or more attributes equal to "total", provided that they differ exactly by one attribute. Therefore, the generic element of the vector representing a specific O/D matrix refers to one attribute and it may be equal either to one of the two possible dichotomic values of the attribute, or to the corresponding total.

If n is the number of attributes, then there are 3^n different O/D matrices which can be built with the same data, when they are classified according to all the attributes. Moreover, a vector of attribute values can be replaced by a ternary n-vector whose components are 0, 1, or 2, where 0 and 1 refer to the dichotomic values of an attribute, while 2 means their sum (total). The set of all possible n-vectors of this kind describes a set $\mathbf{T}^n$ of size 3^n.

More generally, one can think of problems where every point $x \in \mathbf{T}^n$ represents a (large) data set (e.g., a large matrix A_x), and the data sets corresponding to points which differ by just one component are related in the sense that the knowledge of any two of those sets is always enough to derive the third one (e.g., $A_x + A_y = A_z$, and thus $A_x = A_z - A_y$, etc.). In such a situation it is important to know which collections of data sets are enough to be stored in order to be able to reconstruct all of them (which correspond to the complete set $\mathbf{T}^n$). Therefore, we study particular subsets of points in $\mathbf{T}^n$, which are able to *generate* all the other points in the same set, and we call such subsets *generators*.

In this paper we introduce a suitable closure operator in $\mathbf{T}^n$ and discuss the properties of the corresponding closed and open sets; in particular, the generators are those sets whose closure is the whole space $\mathbf{T}^n$. We also analyze the generation process and investigate the size and the properties of generators.

2 Definitions and main results

We shall consider *ternary Hamming spaces* $\mathbf{T}^n$, where $\mathbf{T} = \{0, 1, 2\}$, and n is a positive integer. The elements of $\mathbf{T}^n$ are called *points*, and will be represented as n-vectors of the form $x = (x_1, ..., x_n)$, where $x_i \in \mathbf{T}$ for $i = 1, ..., n$.

Let us denote by $[n]$ the set $\{1, 2, ..., n\}$. For $k \in [n]$ let us call a *subspace of dimension* k, or in short a *k-subspace* the set of all the points of $\mathbf{T}^n$ which have the same coordinates in some $n-k$ components. Subspaces of dimension $n-1$ are called *hyperplanes*, while subspaces of dimension 1 are called *lines*. Here, every line contains exactly 3 points. Let us denote the set of lines by $\mathcal{L}$.

For vectors $x, y \in \mathbf{T}^n$ let us define their *Hamming distance* as

$$d_H(x, y) = |\{i | x_i \neq y_i\}|. \tag{2.1}$$

For a subset $S \subseteq \mathbf{T}^n$ and a point $x \in \mathbf{T}^n$ let

$$d_H(x, S) = \min_{y \in S} d_H(x, y).$$

Finally, for a subset $S \subseteq \mathbf{T}^n$ let us introduce the notation $D_H(S) = \max_{x \in \mathbf{T}^n} d_H(x, S)$.

Clearly, two points are on a line if and only if their Hamming distance is 1. Let us call two points $x, y \in \mathbf{T}^n$ *composable*, if their Hamming distance is 1, or in other words if they belong to a line $L = \{x, y, z\} \in \mathcal{L}$. Let us define a *composition* operation $\otimes$ by associating with any pair $x, y \in L = \{x, y, z\}$ of comparable points the third point of the line containing them, i.e.

$$x \otimes y = z. \tag{2.2}$$

Given a subset $S \subseteq \mathbf{T}^n$, its *emission* is defined as the set of points

$$\partial S = \{z | z = x \otimes y \text{ for some } x, y \in S\} \tag{2.3}$$

which can be generated from S by one composition. Let us call a set $S \subseteq \mathbf{T}^n$ *independent* if $\partial S = \emptyset$, *emissive* if $S \cap \partial S = \emptyset$, and *closed* if $\partial S \subseteq S$. Let us note that independent sets are also closed. Finally, let us call a subset $S \subseteq \mathbf{T}^n$ *open*, if its complement is closed, i.e. if $S \cap \partial(\mathbf{T}^n \setminus S) = \emptyset$.

Let $S^{(0)} = S$, and let us define recursively

$$S^{(k)} = S^{(k-1)} \cup \partial S^{(k-1)} \text{ for } k \geq 1. \tag{2.4}$$

Since $\mathbf{T}^n$ is finite, for every subset $S \subseteq \mathbf{T}^n$ there exists a finite integer k such that $S^{(k')} = S^{(k)}$ for all $k' \geq k$. Let us call the set $S^{(k)}$ for this k the *closure* of S and denote it by S^*. It is easy to see that the mapping $S \mapsto S^*$ is indeed a closure operation, i.e. $S \subseteq S^*$, $S \subseteq T$ implies $S^* \subseteq T^*$, and $S^{**} = S^*$ for all subsets $S, T \subseteq \mathbf{T}^n$.

Let us call a subset $G \subseteq \mathbf{T}^n$ a *generating set*, or in short a *generator* if $G^* = \mathbf{T}^n$.

Our main result gives a sharp lower bound on the size of a generator.

Theorem 1 *For every generating set $G \subseteq \mathbf{T}^n$ we have $|G| \geq 2^n$.*

3 Basic properties of composition: the Boolean generator

For a point $x \in \mathbf{T}^n$ and for $\alpha \in \mathbf{T}$ let us denote by $\kappa_\alpha(x)$ the number of coordinates of x equal to α, i.e. if $x = (x_1, ..., x_n)$, then $\kappa_\alpha(x) = |\{i | x_i = \alpha\}|$.

Lemma 2 *Let $S \subseteq \mathbf{T}^n$, $\alpha \in \mathbf{T}$, and $L \in \mathcal{L}$. Then,*

(i) $\max_{x \in L} \kappa_\alpha(x) = 1 + \min_{x \in L} \kappa_\alpha(x)$, *where the maximum is attained exactly once on L,*

(ii) $\max_{x \in L} d_H(x, S) \leq 1 + \min_{x \in L} d_H(x, S)$, *with equality if $|S| = 1$, in which case the maximum is attained for exactly one point of L.*

Proof. Follows immediately from the definitions. □

Let us now consider a special generating set, $\mathbf{B}_n$ consisting of all "binary " points, i.e.

$$\mathbf{B}_n = \{x \in \mathbf{T}^n | \kappa_2(x) = 0\}.$$

Proposition 3 $\mathbf{B}_n^{(k)} = \{x \in \mathbf{T}^n | \kappa_2(x) \leq k\}$.

Proof. For $k = 0$ the claim is just the definition of $\mathbf{B}_n$, while for $k > 0$, the claim follows by an easy induction on k from (i) of Lemma 2. □

4 Closed, open and even sets

Closed and open sets will play an important role in our proofs.

Proposition 4 *Let C be a closed set and let U be an open set in $\mathbf{T}^n$. Then we have*

$$|L \cap C| \neq 2 \quad \textit{and} \quad |L \cap U| \neq 1 \quad \textit{for all lines} \quad L \in \mathcal{L}.$$

Proof. Immediate from the definition of closed and open sets. □

Proposition 5 *A closed set $C \subseteq \mathbf{T}^n$ is the union of subspaces $C = H_1 \cup \cdots \cup H_s$ such that for every line $L \in \mathcal{L}$ connecting a point $x \in H_i$ with a point $y \in H_j$, $y \neq x$, one of these subspaces contains L fully, i.e. there exists an index k such that $L \subseteq H_k$.*

Proof. Let us take for the subspaces all maximal subspaces contained in C. □

Proposition 6 *Let G be a generator and U be a nonempty open set in $\mathbf{T}^n$. Then $G \cap U \neq \emptyset$.*

Proof. Let $C = \mathbf{T}^n \setminus U$. By definition, C is a closed set, i.e. $C^* = C$. If $G \subseteq C$, then $G^* \subseteq C^* = C \neq \mathbf{T}^n$, contradicting the fact that G is a generator. □

For a point $x = (x_1, ..., x_n) \in \mathbf{T}^n$ let us denote by $L_i(x)$ the line through x in the direction i, i.e. $L_i(x) = \{(x_1, ..., x_{i-1}, \alpha, x_{i+1}, ..., x_n) | \alpha \in \mathbf{T}\}$.

Proposition 7 *If $U \subseteq \mathbf{T}^n$ is a nonempty open set, then $|U| \geq 2^n$.*

Proof. Let us construct an assignment $u : 2^{[n]} \mapsto U$ in the following way. Let us choose a point $u(\emptyset) \in U$ arbitrarily, and then repeat the following steps as long as there are unassigned subsets of $[n]$. Let S be a minimal unassigned subset, and let $i \in S$ be the largest index in S. Since S is minimal, the set $S \setminus \{i\}$ is already assigned; let $x = u(S \setminus \{i\})$, and let us consider the line $L_i(x)$ through it. By Proposition 4 $|L_i(x) \cap U| \neq 1$, there exists another point $y \neq x$, $y \in L_i(x) \cap U$. Let us then set $u(S) = y$.

Clearly, the above procedure stops after 2^n steps, and assigns a point of U to every subset of $[n]$. Let us observe next that if $x = u(S)$, then $x_i = u(\emptyset)_i$ for every index $i \notin S$, while $x_i \neq u(\emptyset)_i$ for indices $i \in S$. Thus, if $A \neq B$ are two different subsets, say $i \in A \setminus B$, then the vectors $u(A)$ and $u(B)$ differ in their ith coordinate. This proves that the above assignment assigns different points to different subsets, and hence $|U| \geq |\{u(S)|S \subseteq [n]\}| = 2^n$. □

Proposition 8 *If $U \subseteq \mathbf{T}^n$ is a nonempty open set, then it contains a generator of size 2^n.*

Proof. Let us construct an assignment $u : 2^{[n]} \mapsto U$ as in the proof of Proposition 7, and set $G = \{u(S) | S \subseteq [n]\}$. We claim that G is a generator.

Let us prove this claim by induction on n. Clearly, if $n = 1$, then G contains 2 points of the line $\mathbf{T}^1$, and thus it is indeed a generator. Let us hence assume that the claim is verified for all $n' < n$, and let us consider the hyperplane $H = \{x | x_n = u(\emptyset)_n\}$. We have $|H \cap G| = 2^{n-1}$, since H contains all the points $u(S)$ for which $n \notin S$. Since H is a ternary space of dimension $n - 1$, $H \cap U$ is an open set of this subspace, thus it follows by our inductive hypothesis that $G \cap H$ generates H, i.e. $(G \cap H)^* = H$. Then the points on the lines $L_n(u(S))$ for $S \ni n$ are also generated, since $u(S)$ and $H \cap L_n(u(S))$ are two different points, already generated by G on this line. Finally, these lines intersect both of the parallel hyperplanes to H in subsets, isomorphic to $G \cap H$. Hence these subsets will generate all points of the corresponding hyperplanes, respectively, just like the set $G \cap H$ generated H. □

Proposition 9 *If U and U' are two nonempty open sets, then $U \cap U' \neq \emptyset$.*

Proof. According to Proposition 8, there exists a generating set G contained by U. Thus, according to Proposition 6, the generating set G must intersect U', and thus we have $U \cap U' \supseteq G \cap U' \neq \emptyset$. □

For a subset $S \subseteq \mathbf{T}^n$ and a point $x \in S$ let us denote by $U(x, S)$ the open set $U(x, S) = \mathbf{T}^n \setminus (S \setminus \{x\})^*$.

Let us call a set $E \subseteq \mathbf{T}^n$ an *even* subset, if every line intersects it in an even number of points, i.e. if $|E \cap L|$ is even for every $L \in \mathcal{L}$. Clearly, even sets are open, but not conversely – see e.g. the open set in Figure 1, which contains full lines, and hence is not even.

Proposition 10 *Even subsets are minimally open.*

Proof. Let E be a nonempty even subset, and let us assume indirectly that $U \subset E$ is an open proper subset of it. We claim first that in this case $E \setminus U$ is also an open subset. For this, let $L \in \mathcal{L}$ be a line for which $L \cap (E \setminus U) \neq \emptyset$. If this intersection were a singleton, then $|U \cap L| = 1$ would be implied by the fact that E is even, and thus U could not be an open set. Hence both U and $E \setminus U$ are nonempty open sets, contradicting Proposition 9. □

One might think that the reverse of the above statement is true. This in fact would substantially simplify many of our proofs. There are however, minimal open sets which are not even – see again the open set in Figure 1, which is in fact minimal, and does not contain (and itself is not) an even set.

Proposition 11 *The symmetric difference of even sets is even again.*

Proof. Let E and E' be even sets. Then, for every line $L \in \mathcal{L}$ we have $L \cap (E \triangle E') = (L \cap E) \triangle (L \cap E')$. Since on the right hand side of this equation both sets are even, the left hand side has an even cardinality, too. □

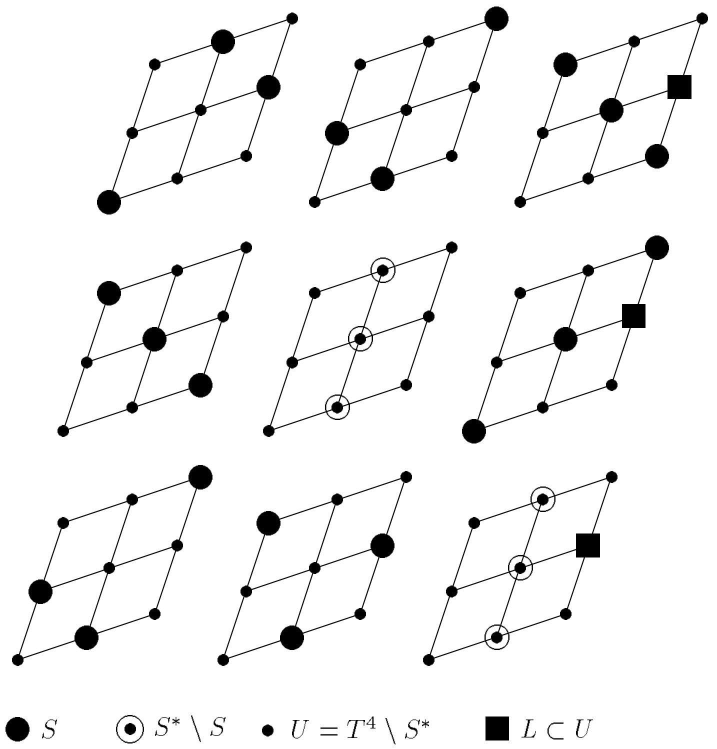

Figure 1: An example for a closed set S^* in $\mathbf{T}^4$, generated by the set S, the complement of which $U = \mathbf{T}^4 \setminus S^*$ is an open set not containing nonempty even subsets. The open set U contains several lines in full, one of them is indicated by black squares on the picture.

For a subset $S \subseteq \mathbf{T}^n$ and $x \in S$ let us introduce the notation $U(x,S) = S^* \setminus (S \setminus \{x\})^*$, i.e. $U(x,S)$ denotes the set of points which are generated by S, only if x is not deleted from S.

Proposition 12 *For the Boolean generator* $\mathbf{B}^n$ *and for any* $g \in \mathbf{B}_n$ *the set* $U(g,\mathbf{B}_n)$ *is even.*

Proof. It is easy to see that for $g \in \mathbf{B}_n$ we have $U(g,\mathbf{B}_n) = \{x | x_i \in \{g_i, 2\} \ \text{ for } \ i = 1,...,n\}$, and for such sets the statement is straightforward to verify. $\square$

Before proving Theorem 1, we need a combinatorial lemma, which may be interesting on its own.

Lemma 13 *Let* S, H_i, $i = 1,...,m$ *be subsets*[1] *of a finite set* X, *such that*

$$S \cap (\triangle_{i \in I} H_i) \neq \emptyset \tag{4.5}$$

holds for all nonempty subsets $I \subseteq [m]$. Then,

$$\left| S \cap \left(\bigcup_{i \in I} H_i \right) \right| \geq |I| \ \ \textit{for all} \ \ I \subseteq [m]. \tag{4.6}$$

Proof. Let us prove (4.6) by induction on the size of I. Clearly, for $|I| = 0$ the inequality (4.6) holds true. Let us assume now that it holds for all $k' < k$, and let us consider a subset I of size $|I| = k$. By our induction hypothesis, $|S \cap (\cup_{i \in I} H_i)| \geq k-1$. Let us denote $k-1$ of these elements by x_1,...,x_{k-1}, i.e. $S \cap (\cup_{i \in I} H_i) \supseteq \{x_1, ..., x_{k-1}\}$, and let us associate binary vectors χ^i, $i = 1, ..., k-1$ with these points by defining

$$\chi^i_j = \begin{cases} 1 & \text{if } \ x_i \in H_j, j \in I \\ 0 & \text{otherwise.} \end{cases}$$

Let us consider these vectors as elements of $GF(2)^k$, the vector space of dimension k over the finite field of mod 2 computations, and let W denote the subspace spanned by χ^1, ..., χ^{k-1}. Then we have $\dim W \leq k-1$, since $k-1$ vectors cannot span a subspace of higher dimension. Thus, $\dim W^\perp > 0$ is implied for the orthogonal subspace $W^\perp$ by the equality $\dim W + \dim W^\perp = k$. Hence, there exists a non-zero binary vector $\chi \in W^\perp$ for which the scalar products $< \chi, \chi^i >= \sum_{j=1}^k \chi_j \chi^i_j \equiv 0 (\text{mod } 2)$ for all $i = 1, ..., k-1$. In other words, the number of indices j for which $\chi_j = 1$ and $x_i \in H_j$ is even for every $i = 1, ..., k-1$. Hence, by denoting $J = \{j | \chi_j = 1\}$, we have that the subfamily $\{H_j | j \in J\}$ involves an even number of those sets which contain x_i for every $i = 1, ..., k-1$, and thus

$$(\triangle_{j \in J} H_j) \cap \{x_1, ..., x_{k-1}\} = \emptyset.$$

Since by (4.5) $S \cap (\triangle_{j \in J} H_j) \neq \emptyset$, it follows that $\{x_1, ..., x_{k-1}\} \neq S \cap (\bigcup_{i \in I} H_i)$, hence proving (4.6). □

We are now ready to give our first proof for Theorem 1.

First proof of Theorem 1. Let G be an arbitrary generator, and let us consider the Boolean generator $\mathbf{B}_n$. We shall show that $|G \setminus \mathbf{B}_n| \geq |\mathbf{B}_n \setminus G|$.

Let us define a bipartite graph, the vertices of which are the sets $V_1 = G \setminus \mathbf{B}_n$ and $V_2 = \mathbf{B}_n \setminus G$, and let us connect a vertex $x \in V_1$ to a vertex $y \in V_2$ if and only if $x \in U(y, \mathbf{B}_n)$. Our claim, i.e. $|V_1| \geq |V_2|$, will follow by Hall's theorem (see [2]). To see this, we shall show that for any subset $A \subseteq V_2$ the set of neighbors of A (in V_1) is at least as large as A, i.e.

$$\left| V_1 \cap \left(\bigcup_{a \in A} U(a, \mathbf{B}_n) \right) \right| \geq |A|. \tag{4.7}$$

Let us note first that for any subset $A \subseteq \mathbf{B}_n$, the set $D(A) = \triangle_{a \in A} U(a, \mathbf{B}_n)$ is a nonempty even set. It is nonempty, since $D(A) \cap \mathbf{B}_n = A$, and it is even by Proposition 11. Proposition 6 thus implies that $G \cap D(A) \neq \emptyset$. Hence, the conditions of Lemma 13 are fulfilled by the sets $S = V_1$ and $H_a = U(a, \mathbf{B}_n)$ for $a \in V_2$, implying the inequality (4.7) for all nonempty $A \subseteq V_2$. Thus, the theorem follows by Hall's theorem. $\square$

We believe that for a minimal generator in fact there is equality in the above theorem.

Conjecture 1 *Every minimal generator of $\mathbf{T}^n$ is of size 2^n.*

5 Signature sets

We shall investigate further properties of even sets.

Let us observe that if G is the Boolean generator $\mathbf{B}^n$, and we assign a set $\sigma(x)$ to every point $x \in \mathbf{T}^n$ by defining $\sigma(x) = \{g \in G | x \in U(g, G)\}$, then this mapping satisfies the following properties:

(i) For every $g \in G$, we have $\sigma(g) = \{g\}$.

(ii) For every line $L = \{x, y, z\} \in \mathcal{L}$, we have $\sigma(x) \triangle \sigma(y) = \sigma(z)$.

It is easy to verify that property (i) follows for every minimal generator, while (ii) is true for the Boolean generator, because the sets $U(g, \mathbf{B}_n)$ are even.

Let us call, in general, a mapping $\sigma : \mathbf{T}^n \mapsto 2^G$ a *signature* of the generator G if it satisfies both conditions (i) and (ii).

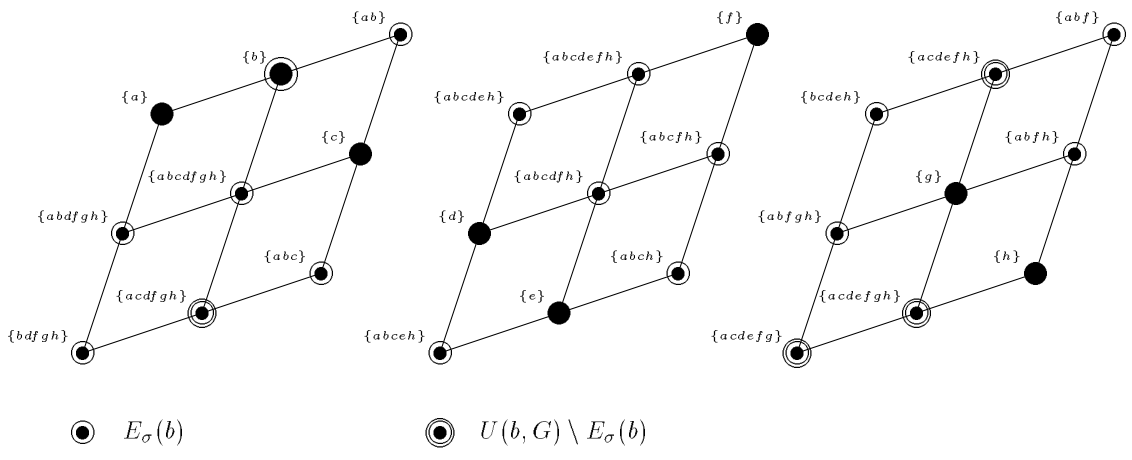

Figure 2: A generator $G \subseteq \mathbf{T}^3$ with a signature σ, where the sets $\sigma(x)$ are indicated at the vertices. For $b \in G$ the sets $E_\sigma(b)$ and $U(b, G)$ are indicated on the picture.

Given a generator G and a signature $\sigma : \mathbf{T}^n \mapsto 2^G$ of it, let us define $E_\sigma(g) = \{x \in \mathbf{T}^n | g \in \sigma(x)\}$ for every $g \in G$.

Proposition 14 *Let us assume that G is a generator and σ is a signature of it. Then the sets $E_\sigma(g)$ are even and $E_\sigma(g) \subseteq U(g, G)$ for all $g \in G$.*

Proof. Let $g \in G$, $L = \{x, y, z\} \in \mathcal{L}$, and let us assume that $x \in E_\sigma(g)$. This means that $g \in \sigma(x)$ by the definition of the set $E_\sigma(g)$, and hence by property (ii) it follows that $g \in \sigma(y)\triangle\sigma(z)$. Thus, exactly one of the points, y or z belongs to $E_\sigma(g)$, implying $|L \cap E_\sigma(g)| = 2$. Since this holds for an arbitrary line intersecting $E_\sigma(g)$, this set is indeed even.

Property (i) implies that $G \cap E_\sigma(g) = \{g\}$, and thus the set $G \setminus \{g\}$ is a subset of the complement $\mathbf{T}^n \setminus E_\sigma(g)$, which is a closed set, and hence $(G \setminus \{g\})^* \cap E_\sigma(g) = \emptyset$, i.e. $U(g, G) \supseteq E_\sigma(g)$. □

Proposition 15 *If a generator has a signature, then it is a minimal generator.*

Proof. Let us assume that the generator G has a signature σ, and that $(G \setminus \{g\})^* = \mathbf{T}^n$ holds for some $g \in G$. Then, property (i) and Proposition 14 imply that $g \in E_\sigma(g) \subseteq U(g, G) = \mathbf{T}^n \setminus (G \setminus \{g\})^* = \emptyset$, a contradiction, proving the claim. □

As we observed at the beginning of this section, the Boolean generator has a signature, and we believe that in fact every minimal generator does have one.

Conjecture 2 *Every minimal generator has a signature.*

Proposition 16 *Every minimal generator has at most one signature.*

Proof. Let us assume that both σ and σ' are signatures of a minimal generator G. Let us then define $\sigma''(x) = \sigma(x)\triangle\sigma'(x)$ for every $x \in \mathbf{T}^n$, and let $D = \{x|\sigma''(x) \neq \emptyset\}$.

Then, we have $\sigma''(g) = \emptyset$ for all $g \in G$, since both σ and σ' satisfy condition (i), and thus the set D is disjoint from G. Furthermore, σ'' also satisfies property (ii), and thus the set D is even. Thus, $D = \emptyset$ follows by Proposition 6 and hence $\sigma = \sigma'$. □

Let us observe next that, although we defined the signature for generator sets, non-generators may also have a signature mapping.

Proposition 17 *Let us assume that the set G has a signature, i.e. there exists a mapping $\sigma : \mathbf{T}^n \mapsto 2^G$ satisfying conditions (i) and (ii). Let $g \in G$, and $g' \in E_\sigma(g)$, $g' \neq g$. Then the set $G' = (G \setminus \{g\}) \cup \{g'\}$ does also have a signature.*

Proof. Let us define a mapping σ' by

$$\sigma'(x) = \begin{cases} \sigma(x) & \text{if } x \notin E_\sigma(g), \\ \sigma(x)\triangle(\sigma(g') \cup \{g'\}) & \text{if } x \in E_\sigma(g). \end{cases}$$

Let us show first that $\sigma'(x) \subseteq G'$ for every $x \in \mathbf{T}^n$. By the definition of σ' it is clear that $\sigma'(x) \subseteq G \cup \{g'\}$, thus it is enough to show that $g \notin \sigma'(x)$ for all $x \in \mathbf{T}^n$. For $x \notin E_\sigma(g)$ it is obvious, and for $x \in E_\sigma(g)$ it follows from the fact that g belongs to both $\sigma(x)$ and $(\sigma(g') \cup \{g'\})$.

Let us verify next that σ' satisfies condition (i). For $x \in G \setminus \{g\}$ we have $\sigma'(x) = \sigma(x) = \{x\}$, since σ is a signature, and for $x = g'$ we have $\sigma'(g') = \sigma(g') \triangle (\sigma(g') \cup \{g'\}) = \{g'\}$.

Finally, let us prove that σ' satisfies property (ii). Let us consider a line $L = \{x, y, z\} \in \mathcal{L}$. If $L \cap E_\sigma(g) = \emptyset$, then $\sigma'(a) = \sigma(a)$ for all $a \in L$. If L intersects $E_\sigma(g)$, then since this set is even by Proposition 14, this intersection contains two points, say e.g. $\{x, y\} = L \cap E_\sigma(g)$. In this case we have $\sigma'(x) = \sigma(x) \triangle (\sigma(g') \cup \{g'\})$, $\sigma'(y) = \sigma(y) \triangle (\sigma(g') \cup \{g'\})$, and $\sigma'(z) = \sigma(z)$. In both cases, property (ii) for σ' follows straight from the fact that σ is a signature. $\square$

Proposition 18 *Every generator G of $\mathbf{T}^n$ has a subset $G' \subseteq G$ of size 2^n, which has a signature.*

Proof. Let us consider all the subsets of size 2^n, which have a signature. This is a nonempty family, since it includes the Boolean generator. Let us then choose a set G' from this family, for which $|G' \setminus G|$ is minimal. We claim that in fact $G' \subseteq G$.

To this end, let us assume indirectly that there is an element $a \in G' \setminus G$. Let us denote by σ' the signature of G'. The set $E_{\sigma'}(a)$ is an even set by Proposition 14, and hence it intersects G by Proposition 6. Let $b \in G \cap E_{\sigma'}(a)$. Proposition 17 then can be applied, and we obtain that the set $G'' = (G' \setminus \{a\}) \cup \{b\}$ is again a subset of size 2^n, having a signature. For this set we have $|G \setminus G''| < |G \setminus G'|$, since we substituted the element $a \notin G$ with $b \in G$, a contradiction with the selection of G'. This contradiction proves that $G' \subseteq G$. $\square$

Second proof of Theorem 1. Immediate consequence of Proposition 18. $\square$

Proposition 19 *Conjectures 1 and 2 are equivalent.*

Proof.

1 $\Longrightarrow$ 2 Let G be a minimal generator. Then by Proposition 18 it has a subset $G' \subseteq G$ of size 2^n, which has a signature. Conjecture 1 implies that $|G| = 2^n$, and hence $G = G'$ and thus G has a signature.

2 $\Longrightarrow$ 1 Let G be a minimal generator and let σ be its signature, which exists by Conjecture 2. Let us then repeat the first proof for Theorem 1, but interchanging the roles of G and $\mathbf{B}_n$, and using the even sets $E_\sigma(g)$ for $g \in G$ instead of $U(g, \mathbf{B}_n)$ for $g \in \mathbf{B}_n$. We shall obtain $|G \setminus \mathbf{B}_n| \leq |\mathbf{B}_n \setminus G|$, which thus implies $|G| = |\mathbf{B}_n| = 2^n$.

$\square$

6 Counterexamples and open problems

6.1 Generators and matroids

Some properties of generators might lead one to suspect the existence of an underlying matroidal structure. In fact, there are several possible ways of associating hypergraphs to a Hamming space, each of which could be a matroid (for definitions see e.g. [4].)

Let us define $\mathcal{M}'_n$ as the family of sets $I \subseteq \mathbf{T}^n$ contained in a minimal generator $G \supseteq I$. Let $\mathcal{M}''_n$ be the family of subsets $I \subseteq \mathbf{T}^n$ for which $(\mathbf{T}^n \setminus I)* = \mathbf{T}^n$, i.e. the complement of which contains a generator. It is easy to see that these two hypergraphs are dual to each other, i.e. the maximal elements of $\mathcal{M}'$ are the minimal generators, while the maximal elements of $\mathcal{M}''$ are the complements of minimal generators.

One can easily verify that $\mathcal{M}'_2$ and $\mathcal{M}''_2$ are both matroidal, and the same is true for $\mathcal{M}'_3$ and $\mathcal{M}''_3$. This however cannot be generalized further. On can find examples of consisting two minimal generators $G_1, G_2 \subset \mathbf{T}^n$, and a point $g \in G_1 \setminus G_2$ such that $G' = (G_1 \setminus \{g\}) \cup \{g'\}$ is not a generator for all $g' \in G_2 \setminus G_1$, implying that neither of $\mathcal{M}'_n$ and $\mathcal{M}''_n$ are matroids. The smallest such example we found is in $\mathbf{T}^5$, and there seems to be many such pairs of minimal generators in $\mathbf{T}^6$.

One may also think that the closure operation $S \mapsto S^*$ is in fact induced by a matroid, i.e. that the hypergraph, $\mathcal{M}'''_n$ consisting of those sets $I \subseteq \mathbf{T}^n$ for which $x \notin (I \setminus \{x\})^*$ holds for all $x \in I$, forms a matroid. This can also be seen as a false hope by many different ways.

It is easy to see that minimal generators G as well as subsets $S \subseteq \mathbf{T}^n$ with $\partial S = \emptyset$, i.e. independent sets, belong to $\mathcal{M}'''_n$. Thus, among the maximal elements of $\mathcal{M}'''_n$ we can find e.g. $\mathbf{B}_n$ of size 2^n, and the maximum independent sets of $\mathbf{T}^n$, which are known from coding theory (see e.g. [1]) to be of size 3^{n-1} (an easy example being $I = \{(x_1, ..., x_n) | \sum_{i=1}^n x_i = 0 \bmod 3\}$.) Hence the maximal elements of $\mathcal{M}'''_n$ are not all of the same size.

More directly, we can see that the closure operation does not satisfy the closure axioms of a matroidal system (see e.g. [4]), see e.g. the example in Figure 3.

6.2 Switching and alike

Let G be a generator and $L = \{x, y, z\}$ be a line in $\mathbf{T}^n$. If $G \cap L = \{x, y\}$, then obviously, the set $G' = (G \setminus \{x\}) \cup \{z\}$ is a generator again. Let us call the operation by which we obtained G' from G a *switch* of the elements x and z of the line L. Clearly, a switch operation is reversible, hence we can call two generators G and G' *switch equivalent*, and denote it by $G \equiv G'$ if there is a sequence of switch operation which transforms G to G'.

Clearly, $\equiv$ is an equivalence relation over the set of generators, and we believe that

Conjecture 3 *Any two minimal generators of* $\mathbf{T}^n$ *are switch equivalent.*

Clearly, this conjecture would imply readily Conjecture 1.

Let us observe next that if $(G \cap L) \ni x$ for a line $L = \{x, y, z\}$ and $y \in (G \setminus \{x\})^*$, then again the set $G' = (G \setminus \{x\}) \cup \{z\}$ is a generator. Let us call such a transformation a *shift*

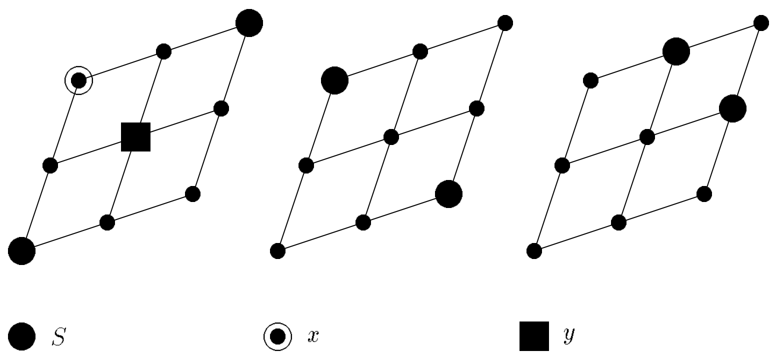

Figure 3: An example for a set $S \subset \mathbf{T}^3$ and points $x, y \in \mathbf{T}^3$ such that $y \in (S \cup \{x\})^*$ and $x \notin (S \cup \{y\})^*$.

of element x along the line L. Shift operations are also reversible, and hence we can define two generators G and G' *shift equivalent* if G' can be obtained by a series of shifts from G, and let us denote this relation by $G \approx G'$. Shift equivalence is again an equivalence relation over the set of generators, and it is also clear that every switch is a shift, too, but not vice versa, see e.g. Figure 4.

Let us observe that although the shift operation $c \mapsto v$ in this example is not a switch, we can obtain it by a series of switches, namely, switch first a to u, then switch c to v and finally switch u back to a. We believe that in fact this always can be done, i.e.

Conjecture 4 *Every shift can be realized by a series of switches, i.e. the relations $\equiv$ and $\approx$ are the same.*

Notice that Conjecture 4 implies the following

Conjecture 5 *Any two minimal generators of $\mathbf{T}^n$ are shift-equivalent.*

Let G be a generator, $g \in G$ and $g' \notin G$. If the set $G' = (G \setminus \{g\}) \cup \{g'\}$ is a generator again, then let us say that it was obtained by the *exchange* of elements g and g'. Clearly, switches and shifts are special type of exchange operations.

Since the exchange operation is also reversible, we can say that generators G and G' are *exchange-equivalent*, or in notation $G \sim G'$, if G' can be obtained from G by a series of exchange operations.

Conjecture 6 *If two minimal generators G and G' are exchange-equivalent, then either of them can be obtained from the other by a sequence of switches, i.e. the relations $\equiv$, $\approx$ and $\sim$ are identical.*

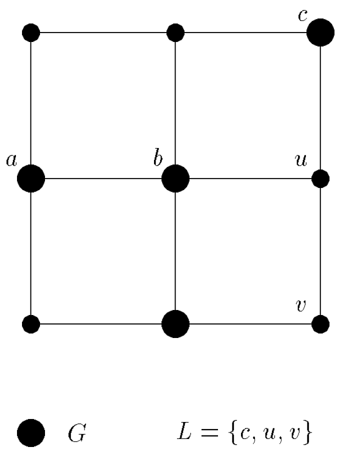

Figure 4: An example for a generating set $G \subset \mathbf{T}^3$ and a line $L = \{c, u, v\}$ in $\mathbf{T}^2$ for which the element $c \in G \cap L$ can be shifted to v, but cannot switched to any other point.

References

[1] Cameron, P.J., *Combinatorics*, (Cambridge University Press, Cambridge, 1994)

[2] Hall, P., On representatives of subsets, *J. London Math. Soc.* **10** (1935), pp. 26-30.

[3] Mac Williams, F.J. and Sloane, N.A.J., *The theory of error-correcting codes*, (North-Holland, Amsterdam, 1977)

[4] Welsh, D.J.A., *Matroid Theory*, (London Academic Press, 1976)