Theoretical Computer Science 406 (2008) 15–23

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

A framework for generating some discrete sets with disjoint components by using uniform distributions[☆]

Péter Balázs [*]

University of Szeged, Department of Image Processing and Computer Graphics, Árpád tér 2., H-6720, Szeged, Hungary

ARTICLE INFO

Keywords:
Discrete tomography
Discrete set
Random generation
hv-convexity
Q-convexity

ABSTRACT

Discrete tomography deals with the reconstruction of discrete sets from few projections. Assuming that the set to be reconstructed belongs to a certain class of discrete sets with some geometrical properties is a commonly used technique to reduce the number of possibly many different solutions of the same reconstruction problem. The average performance of reconstruction algorithms are often tested on such classes by choosing elements of a given class from uniform random distributions. This paper presents a general framework for generating discrete sets with disjoint connected components using uniform distributions. Especially, the uniform random generation of hv-convex discrete sets and Q-convex discrete sets according to the size of the minimal bounding rectangle are discussed.

1. Introduction

Discrete tomography (DT) [18,19] aims to reconstruct a discrete set (a finite subset of the two-dimensional integer lattice defined up to translation) from its line integrals along several (usually horizontal, vertical, diagonal, and antidiagonal) directions. It has several applications in pattern recognition, image processing, electron microscopy, angiography, radiology, non-destructive testing, and so on. The main challenge in DT is that practical limitations every time reduce the number of available projections to at most about ten (usually far fewer) — which results in ambiguous reconstruction, i.e. the number of possible solutions of the same reconstruction task can be extremely large. This can cause the reconstructed discrete set to be quite unlike the original one. In addition, the reconstruction problem can be NP-hard, depending on the number and directions of the projections. One way of eliminating these problems is to suppose that the set to be reconstructed has some geometrical properties. In this way we can reduce the search space of the possible solutions, and we can achieve fast and rare ambiguous reconstructions.

When analysing reconstruction algorithms from the viewpoint of running time, the worst case time complexity is just one feature where the effectiveness of the algorithm can be described. In fact, when a reconstruction algorithm is applied in practice the average running time of the algorithm is also an important characteristic of its efficiency. Besides the worst case scenario, the average performance of an algorithm is also important, even when other parameters of its efficiency are tested such as accuracy and the memory requirements. Moreover, it is also important to describe under which constraining assumptions (in our case, which kind of discrete sets) the algorithm performs for the reconstruction to be fast and accurate, and what are the (perhaps less likely) situations where the algorithm attains its theoretically worst performance.

☆ This work is supported by OTKA grant T48476.

* Tel.: +36 62 546 396.
E-mail address: pbalazs@inf.u-szeged.hu.
URL: http://www.inf.u-szeged.hu/~pbalazs.

0304-3975/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2008.06.010

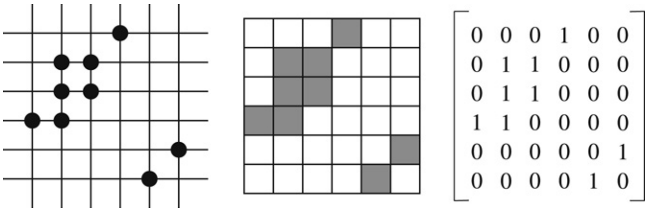

Fig. 1. A discrete set of size 6×6 represented by its elements (*left*), a binary picture (*center*), and a binary matrix (*right*).

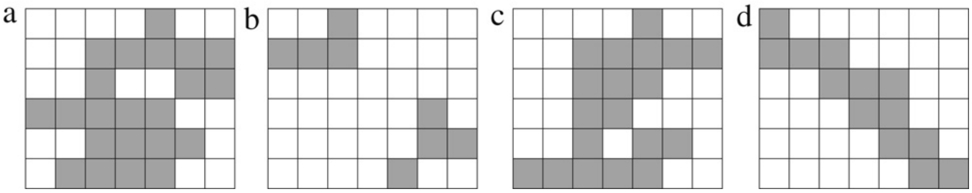

Fig. 2. (a) a polyomino, (b) an hv-convex discrete set, (c) an NE-directed polyomino, and (d) an NW-parallelogram polyomino.

To be able to analyze the average performance of a reconstruction algorithm that was developed for a given class $\mathcal{G}$ of discrete sets, one has to generate elements of the class $\mathcal{G}$ of the same size by using uniform random distributions. For some classes of discrete sets (e.g., the class of all discrete sets) this task is not so complicated, but for others it is far from being trivial. This paper addresses the topic of generating discrete sets with disjoint components by using uniform random distributions, and its structure is the following. First, the necessary definitions are introduced in Section 2. In Section 3 we describe the generation algorithm for the class of hv-convex discrete sets which can be regarded as an extension of the algorithm in [20] for generating hv-convex polyominoes. After, in Section 4 we show how the method can be extended to generate Q-convex discrete sets by using uniform random distributions. Finally, in Section 5 we discuss our results.

2. Definitions

The finite subsets of the 2D integer lattice $\mathbb{Z}^2$ are called *discrete sets*. The *size* of a discrete set is the size of its minimal bounding discrete rectangle. A discrete set F of size $m \times n$ is defined up to a translation, and it can be represented by a binary picture consisting of unitary cells or by a binary matrix $\hat{F} = (\hat{f}_{ij})_{m \times n}$ (see Fig. 1). To be consistent with the corresponding matrix representation we shall assume that the vertical axis of the 2D integer lattice is top-down directed and that the upper left corner of the minimal bounding rectangle of F is the $(1, 1)$ position. For the same reason we will refer to any element of a discrete set by its matrix position (i.e. not by the position of the element in the 2D integer lattice). Clearly, definitions given for discrete sets always have natural counterparts in matrix theory. Vice versa, a definition described in matrix theoretical form can be expressed in the language of discrete sets, too.

A discrete set F is 4-*connected* (with an other term *polyomino*), if for any two positions $P \in F$ and $Q \in F$ of the set there exists a sequence of distinct positions $(i_0, j_0) = P, \ldots, (i_k, j_k) = Q$ such that $(i_l, j_l) \in F$ and $|i_l - i_{l+1}| + |j_l - j_{l+1}| = 1$ for each $l = 0, \ldots, k-1$ (see Fig. 2a).

A discrete set is called *hv-convex* if all the rows and columns of the set are 4-connected, i.e., the 1s of the corresponding representing matrix are consecutive in each row and column (see Fig. 2b).

A polyomino F is *northeast directed* (NE-directed for short) if there is a particular point $P \in F$ such that for each point $Q \in F$ there is a sequence $P_0 = P, \ldots, P_t = Q$ of distinct points of F such that each point P_l of the sequence is north or east of P_{l-1} for each $l = 1, \ldots, t$ (see Fig. 2c). Similar definitions can be given for SW-, SE-, and NW-directedness. An hv-convex polyomino is called *NW/NE-parallelogram polyomino* if it is both NW- and SE-directed or both NE- and SW-directed, respectively (see Fig. 2d).

The maximal 4-connected subsets of a discrete set F are called the *components of F*. For, example the discrete set in Fig. 1 has four components: $\{(1, 4)\}$, $\{(2, 2), (2, 3), (3, 2), (3, 3), (4, 1), (4, 2)\}$, $\{(5, 6)\}$, and $\{(6, 5)\}$.

For a point $P = (p_1, p_2)$ we can define the four *quadrants around P* by

$$\begin{aligned}
R_0(P) &= \{Q = (q_1, q_2) \mid q_1 \le p_1 \text{ and } q_2 \le p_2\}, \\
R_1(P) &= \{Q = (q_1, q_2) \mid q_1 \ge p_1 \text{ and } q_2 \le p_2\}, \\
R_2(P) &= \{Q = (q_1, q_2) \mid q_1 \ge p_1 \text{ and } q_2 \ge p_2\}, \\
R_3(P) &= \{Q = (q_1, q_2) \mid q_1 \le p_1 \text{ and } q_2 \ge p_2\}.
\end{aligned} \tag{1}$$

A discrete set F is *Q-convex* if $R_k(P) \cap F \neq \emptyset$ for all $k \in \{0, 1, 2, 3\}$ implies $P \in F$ (see Fig. 3).

Remark 1. Q-convexity was introduced in [7] in a more general way along an arbitrary set of directions. More precisely, the above definition corresponds to Q-convexity along the horizontal and vertical directions. Since we only consider Q-convex sets defined in the above manner, for the sake of simplicity we shall use this abbreviated form.

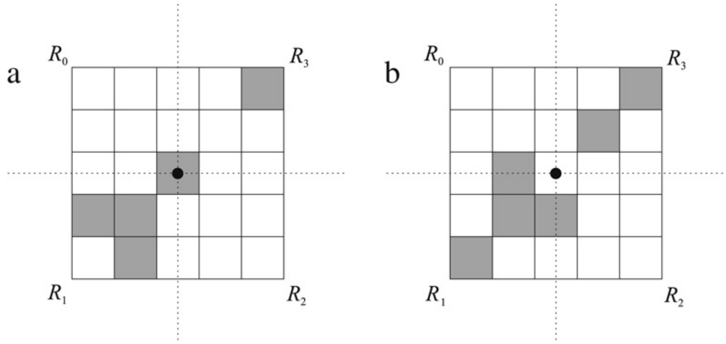

Fig. 3. (a) a Q-convex and (b) a non-Q-convex discrete set with the four quadrants around the point P (denoted by a black dot).

Our aim is to develop a framework for the uniform random generation of certain discrete sets which have disjoint components.

Definition 2. Let F be a discrete set with k components such that $I_l \times J_l = \{i_l, \ldots, i'_l\} \times \{j_l, \ldots j'_l\}$ is the minimal bounding rectangle of the l-th component of F. We say that the components of F are *disjoint* if for any $1 \le l, l' \le k$ $l \neq l'$ implies that $I_l \cap I_{l'} = \emptyset$ and $J_l \cap J_{l'} = \emptyset$.

Note that the above definition of disjointness is stronger than just saying that the minimal bounding rectangles of the components are pairwisely disjoint.

3. Random generation of hv-convex discrete sets

The class of hv-convex discrete sets is one of the most important classes in discrete tomography. Clearly, the components of an arbitrary hv-convex set are necessarily disjoint. Although the reconstruction from two projections in the class of hv-convex discrete sets is NP-hard [25] several methods can solve this problem by applying some heuristic [21], metaheuristic [6] or optimization [13] technique. And it was also shown that the reconstruction task in this class is no longer intractable if absorption in the projections is present (for certain absorption coefficients) [23]. Besides, for some subclasses of the hv-convex class (hv-convex polyominoes and hv-convex 8-connected sets) different polynomial-time reconstruction algorithms have been developed. One of them (let us call it Algorithm A) approximates the solution iteratively by a nondecreasing sequence of so-called *kernel sets* and by a nonincreasing sequence of so-called *shell sets* (see [3,4,11]). Algorithm A has a worst case time complexity of $O(mn \cdot \log mn \cdot \min\{m^2, n^2\})$. The other approach (let us call it Algorithm B) is based on an observation that the reconstruction task can be transformed into a 2SAT task that is solvable in polynomial time [12,22]. Algorithm B has a worst case time complexity of $O(mn \cdot \min\{m^2, n^2\})$. However, in [2] algorithms A and B were compared from the viewpoint of the average execution time and it was found that Algorithm A reconstructs the solution faster than Algorithm B in almost every studied case. This comparison was based on a sophisticated method of [20] by which it is possible to generate hv-convex polyominoes of given size from a uniform random distribution. The observations concerning the average execution times of the two reconstruction approaches led to the design of a third reconstruction algorithm that has the same worst case time complexity as Algorithm B but remains as fast as Algorithm A in the average case [2].

Based on these findings, in the last few years the class of hv-convex discrete sets has become one of the statistical indicators of newly developed exact or heuristical reconstruction algorithms which characterises the effectiveness of a given technique. This means that the performance of the newly developed techniques are often tested on this class. Unhappily, researchers had to acknowledge the fact that no method was known for generating hv-convex sets of a given size using uniform random distributions, and hence no exact comparison of the techniques was possible. In actual fact, the efficiency of these techniques was usually tested on only 5–10 elements (or just one!) of the class of hv-convex discrete sets that were chosen ad hoc. Obviously, this does not say too much about the average performance of the reconstruction method on that class.

In this section we present an algorithm for generating hv-convex discrete sets of a moderate size from uniform random distributions which can (partially) eliminate the problem mentioned above. The generating algorithms proposed in this paper are partially based on the decomposition technique deeply studied in [16]. In that work, the authors presented a systematic approach to the random generation of labelled combinatorial structures according to one parameter. Later, in [15] the method was extended to be able to generate objects according to several parameters simultaneously, too. This result will form the basis of our generating framework.

3.1. Canonical hv-convex discrete sets

We first will consider a special class of hv-convex discrete sets. The following new concept will help us to introduce this class.

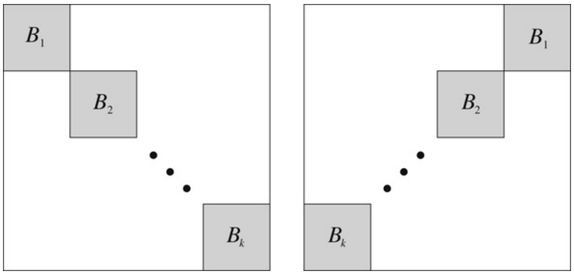

Fig. 4. The relative position of the minimal bounding rectangles of the components $B_1, \ldots, B_k$ of a canonical (left) and an anticanonical (right) discrete set.

Definition 3. Let F be an arbitrary discrete set with $k \geq 2$ disjoint components such that the minimal bounding rectangle of the l-th component of F is $\{i_l, \ldots, i'_l\} \times \{j_l, \ldots, j'_l\}$ $(l = 1, \ldots, k)$ and $i_l < i_{l+1}$ for each $l = 1, \ldots, k-1$. F is called *canonical* if $j_l < j_{l+1}$ for each $l = 1, \ldots, k-1$. F is called *anticanonical* if $j_l > j_{l+1}$ for each $l = 1, \ldots, k-1$.

The notion of F being canonical corresponds, in matrix language, the fact that the corresponding matrix is the direct sum of smaller matrices $B_1, \ldots, B_k$. That is, the discrete set is canonical (anticanonical) if – omitting empty rows and columns – the minimal bounding rectangles of the components are connected to each other with their bottom right hand and upper left hand (bottom left hand and upper right hand) corners (see Fig. 4).

The perimeter (i.e, the length of the boundary) of an hv-convex polyomino of size $m \times n$ is trivially $2(m+n)$. The first result concerning the uniform random generation of hv-convex discrete sets was given in [14], where the authors proved that the number P_n of hv-convex polyominoes with a perimeter value of $2n+8$ is

$$P_n = (2n+11)4^n - 4(2n+1)\binom{2n}{n}. \tag{2}$$

Later, based on the above result in [17] it was shown that the number $P_{m+1,n+1}$ of hv-convex polyominoes of size $(m+1) \times (n+1)$ is

$$P_{m+1,n+1} = \frac{m+n+mn}{m+n}\binom{2m+2n}{2m} - \frac{2mn}{m+n}\binom{m+n}{m}^2. \tag{3}$$

Let us first concentrate on the $\mathcal{S}'$ class of canonical hv-convex discrete sets with nonempty rows and columns. Denoting the number of discrete sets of $\mathcal{S}'$ with size $m \times n$ by $S'_{m,n}$ we obtain $S'_{1,j} = P_{1,j} = 1$ $(j = 1, \ldots, n)$, $S'_{i,1} = P_{i,1} = 1$ $(i = 1, \ldots, m)$, and the following

Theorem 4. *For every* $m, n > 1$

$$S'_{m,n} = P_{m,n} + \sum_{k<m,\ l<n} P_{k,l} \cdot S'_{m-k,n-l}. \tag{4}$$

Proof. It is not hard to see that a discrete set $F \in \mathcal{S}'$ of size $m \times n$ is either an arbitrary hv-convex polyomino (i.e. it has just one component) or it contains an hv-convex polyomino of size $k \times l$ (where $k < m$ and $l < n$) as a subset in the upper left hand corner and the remaining part of F is an arbitrary discrete set of size $(m-k) \times (n-l)$ which also belongs to the $\mathcal{S}'$ class (see Fig. 4 again). This observation can be concisely expressed by the recursive formula (4). $\square$

We now can describe the algorithm for generating hv-convex discrete sets of $\mathcal{S}'$ of a given size by using a uniform random distribution. A basic idea behind this algorithm is to use the method described in [20] for generating the components which are hv-convex polyominoes.

This algorithm works as follows. First, in Step 1, it calculates $S'_{m,n} = P_{m,n} + P_{1,1} \cdot S'_{m-1,n-1} + P_{1,2} \cdot S'_{m-1,n-2} + P_{2,1} \cdot S'_{m-2,n-1} + \cdots + P_{m-1,n-1} \cdot S'_{1,1}$. By choosing a number r with a uniform random distribution in the interval $[1, S'_{m,n}]$ (Step 2), one must determine which one of the following intervals r falls into.

$$\begin{aligned}
I_0 \quad &= [1, P_{m,n}]\,, \\
I_{1,1} \quad &= [P_{m,n}+1, P_{m,n} + P_{1,1} \cdot S'_{m-1,n-1}]\,, \\
I_{1,2} \quad &= [P_{m,n} + P_{1,1} \cdot S'_{m-1,n-1} + 1, P_{m,n} + P_{1,1} \cdot S'_{m-1,n-1} + P_{1,2} \cdot S'_{m-1,n-2}]\,, \\
&\ \vdots \\
I_{m-1,n-1} \quad &= [S'_{m,n} - P_{m-1,n-1} \cdot S'_{1,1} + 1, S'_{m,n}].
\end{aligned} \tag{5}$$

If $r \in I_0$ then the whole discrete set to be generated will consist of just one component, i.e. it will be an hv-convex polyomino. If $r \in I_{1,1}$ then the upper left hand component of the set to be generated will have a size 1×1. Similarly, if $r \in I_{1,2}$ then the upper left hand component will have a size 1×2. And so on. Finally, if $r \in I_{m-1,n-1}$ then the upper left hand component will have a size $(m-1) \times (n-1)$. The intervals of (5) give a (uniquely determined) partitioning of the interval $[1, S'_{m,n}]$. The size of each partition I_{ij} $(1 \leq i \leq m-1,\ 1 \leq j \leq n-1)$ is equal to the number of sets of $\mathcal{S}'$ which have a component

Algorithm GENHV-S' for generating hv-convex discrete sets of $\mathcal{S}'$ by using uniform random distributions

Input: The integers m and n.
Output: An hv-convex discrete set of $\mathcal{S}'$ with size $m \times n$.

Step 1 let Q be an empty queue and calculate $S'_{m,n}$;
Step 2 choose a number $r \in [1, S'_{m,n}]$ from a uniform random distribution;
Step 3 **if** ($r > P_{m,n}$)
{ $r = r - P_{m,n}$;
for $k = 1$ **to** $m - 1$
for $l = 1$ **to** $n - 1$
if ($r > P_{k,l} \cdot S'_{m-k,n-l}$) $r = r - P_{k,l} \cdot S'_{m-k,n-l}$;
else { put (k, l) into Q;
goto Step 2 with parameters $m - k$ and $n - l$; }
}
else put (m, n) into Q;
Step 4 **foreach** (k, l) in Q **do** generate an hv-convex polyomino of size $k \times l$ by using a uniform random distribution;

of size $i \times j$ in the upper left corner. Hence the size $k \times l$ of the upper left hand component of the set to be generated can be determined (in this case we put the pair (k, l) into the queue Q) and this method can be repeated for the remaining set as well (Step 3). Next, we only have to generate the components themselves from uniform random distributions by knowing the sizes of their bounding rectangles (stored in Q) — which is possible with the algorithm given in [20] (Step 4).

From the above points it follows that Algorithm GENHV-S' generates sets of $\mathcal{S}'$ of size $m \times n$ from a uniform random distribution. Turning to the analysis of the algorithm we can state the following. Step 1 is for the preprocessing. By using the initial values $S'_{1,j} = P_{1,j}$ $(j = 1, \ldots, n)$ and $S'_{i,1} = P_{i,1}$ $(i = 1, \ldots, m)$, and the Eqs. (3) and (4), $S'_{m,n}$ can be calculated via a dynamic programming approach. During the calculation of $S'_{m,n}$ every $S'_{k,l}$ $(1 \le k \le m,\ 1 \le j \le n)$ will be computed and stored. Thus, Step 1 can be performed in $O(m^2n^2)$ time with $O(mn)$ memory requirements (here and in the following we consider that an arithmetical operation takes a constant time). In steps 2 and 3 the size of the components of the set to be generated are determined. Since the number of components can be at most $\min\{m, n\}$, steps 2 and 3 takes $O(mn \cdot \min\{m, n\})$ time. Finally, in Step 4 we generate the components from uniform random distributions by knowing the size of their bounding rectangles. The complexity of this step depends on the generation algorithm applied. In [20] two methods are proposed for this task. The first one computes a random hv-convex polyomino of size $k \times l$ in $O((k + l)^3)$ time with $O((k + l)^5)$ preprocessing time and $O((k + l)^3)$ memory requirements. The other one is a probabilistic generator which has $O(k + l)$ time and memory requirements, and succeeds with asymptotic probability 0.5.

Our method can also be extended to arbitrary canonical hv-convex discrete sets which may have empty rows or/and columns. This class will be denoted by $\mathcal{S}$. Naturally, $\mathcal{S}' \subset \mathcal{S}$. Then a discrete set $F \in \mathcal{S}$ of size $m \times n$ is either a polyomino or it contains a polyomino of size $k \times l$ (where $k < m$ and $l < n$) as a subset in the upper left hand corner, and the remaining part of F is a discrete set of $\mathcal{S}$ of size $(m - k) \times (n - l)$ or smaller as there may be some empty rows or/and columns between the two parts. Denoting the number of discrete sets of $\mathcal{S}$ with size $m \times n$ by $S_{m,n}$, we get a formula similar to Eq. (4). That is,

$$S_{m,n} = P_{m,n} + \sum_{k<m,\ l<n} P_{k,l} \cdot \left(\sum_{i \le m-k,\ j \le n-l} S_{i,j} \right). \tag{6}$$

As before, using (3) and the initial values $S_{1,j} = P_{1,j} = 1$ $(j = 1, \ldots, n)$ and $S_{i,1} = P_{i,1} = 1$ $(i = 1, \ldots, m)$, (6) can be evaluated via a dynamic programming approach in $O(m^2n^2)$ time with $O(mn)$ memory requirements. Next, an algorithm similar to Algorithm GENHV-S' can be supplied to generate hv-convex discrete sets of given size from the $\mathcal{S}$ class based on a uniform random distribution (let us call it Algorithm GENHV-S).

3.2. Arbitrary hv-convex discrete sets

The method employed in Section 3.1 provides a useful tool for reconstructing some special hv-convex discrete sets by using uniform random distributions. It needs, though, some rethinking to adapt algorithms GENHV-S' and GENHV-S to the whole class of hv-convex discrete sets. This class will be denoted by $\mathcal{HV}$. We will use the following proposition which holds for every discrete set with disjoint components.

Proposition 5. *Let F be an arbitrary discrete set with $k \ge 2$ disjoint components. Then there is a uniquely determined canonical discrete set F' with disjoint components, and a uniquely determined permutation π of order k, such that all of the following ones hold:*

- *F and F' have the same components*

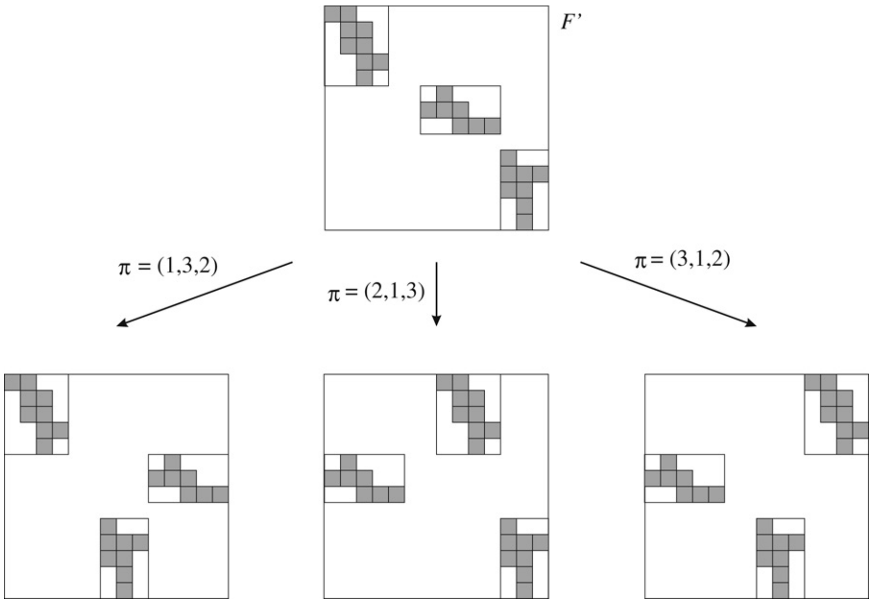

Fig. 5. An example which shows the connection between elements of the $\mathcal{HV}$ and $\mathcal{S}$ classes.

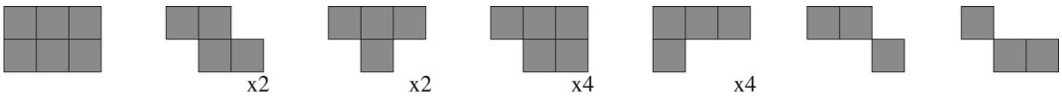

Fig. 6. All the sets of $\mathcal{S}'$ of size 2×3. The numbers tell us that there are other solutions which can be obtained by mirroring or/and rotating the given set.

- *if the minimal bounding rectangle of the l-th component of F′ is* $\{i_l, \ldots, i'_l\} \times \{j_l, \ldots, j'_l\}$ *then the minimal bounding rectangle of the l-th component of F is* $\{i_l, \ldots, i'_l\} \times \{j''_l, \ldots, j'''_l\}$ *(with* $1 \le j''_l, j'''_l \le n$*)*
- *for every* $1 \le t, s \le k$ $\; j''_t < j''_s$ *if and only if* $\pi^{-1}(t) < \pi^{-1}(s)$
- *for each* $1 \le l \le k-1$ $\; j_{l+1} - j'_l = j''_{l+1} - j'''_l$.

As the components of an arbitrary hv-convex set are necessarily disjoint Proposition 5 holds for hv-convex sets as well. Fig. 5 shows an example of this connection between sets of $\mathcal{S}$ and $\mathcal{HV}$. Of course, if the hv-convex set does not have empty rows and columns then the corresponding set F' also belongs to the $\mathcal{S}'$ class.

The reader may think that it is now straightforward to generate sets of $\mathcal{HV}$ from a uniform random distribution by doing the following. One simply generates a set of $\mathcal{S}$ from a uniform random distribution, and then one gets the final hv-convex set by permuting the column sets of the components with a randomly chosen permutation. Unfortunately, this method does not generate the sets from a uniform random distribution. Let us see why this is so in an illustrative example.

Example 6. Suppose we want to generate a discrete set of $\mathcal{S}'$ with size 2×3 from a uniform random distribution. Then,

$$S'_{2,3} = P_{2,3} + P_{1,2} \cdot S'_{1,1} + P_{1,1} \cdot S'_{1,2} = 13 + 1 \cdot 1 + 1 \cdot 1 = 15. \tag{7}$$

Fig. 6 shows the 15 possible discrete sets of $\mathcal{S}'$ with size 2×3. That is, the probability that the generated set will be a polyomino is $13/15$, while the probability that the result will be the last or second last set of Fig. 6 is $1/15$ in both cases. There are two other hv-convex sets of size 2×3 that we can get by mirroring the last two sets of Fig. 6 horizontally (or equivalently, by permuting the column sets of the components of these sets according to the permutation $\pi = (2, 1)$). Thus there exist 17 hv-convex discrete sets of size 2×3 that do not contain empty rows and columns. Following the naive generation method described above, the probability that the generated set will be the last set of Fig. 6, say, is $\frac{1}{15} \cdot \frac{1}{2}$ (since either we keep this set in its original form or we apply the permutation $\pi = (2, 1)$ on the column sets of its components). But the true probability for a uniform random generation must be $1/17$ – which shows that the naive method does not work well.

We need to make some small changes in Algorithm GENHV-S' to achieve our goal. In fact, what we have to do is to have proper weights of the discrete sets of $\mathcal{S}'$ that can represent several hv-convex sets. The recursive formulas used in the generation will change in the following way. Let $HV'^{(t)}_{m,n}$ denote the number of arbitrary hv-convex discrete sets with nonempty rows and columns that have minimal bounding rectangles of size $m \times n$ and exactly t components. Then $HV'^{(t)}_{i,j} = 0$ if $i < t$ or $j < t$, and $HV'^{(1)}_{i,j} = P_{i,j}$ for each $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Finally, the following recursion holds

Theorem 7. *For every $t > 1$ and $m, n \geq 1$*

$$HV'^{(t)}_{m,n} = \sum_{k<m,\ l<n} P_{k,l} \cdot HV'^{(t-1)}_{m-k,n-l} \cdot t. \tag{8}$$

Proof. By Proposition 5, sets of $\mathcal{S}'$ with k components represent $k!$ different hv-convex discrete sets. In Eq. (8) the factor t decreases by one at each step as we go deeper and deeper in the recursion. This will yield a factor of $t!$ for a set that consists of t components — which is exactly the weight we need. Then formula (8) follows from similar observations as in formula (4). □

With a simple calculation we find that the total number $HV'_{m,n}$ of arbitrary hv-convex discrete sets of size $m \times n$ with nonempty rows and columns is

$$HV'_{m,n} = \sum_{t=1}^{\min\{m,n\}} HV'^{(t)}_{m,n}. \tag{9}$$

With this formula we can compute the right probabilities. For the problem presented in Example 6 it results

$$\begin{aligned} HV'_{2,3} &= HV'^{(1)}_{2,3} + HV'^{(2)}_{2,3} = P_{2,3} + P_{1,2} \cdot HV'^{(1)}_{1,1} \cdot 2 + P_{1,1} \cdot HV'^{(1)}_{1,2} \cdot 2 \\ &= 13 + 1 \cdot 1 \cdot 2 + 1 \cdot 1 \cdot 2 = 17. \end{aligned} \tag{10}$$

Now, analogously to Algorithm GENHV-S', we get the generation method, called Algorithm GENHV', for the whole class of hv-convex discrete sets with nonempty rows and columns. We must also take into account what Proposition 5 says, namely that the components need to be permuted by a randomly chosen permutation.

Algorithm GENHV' for generating hv-convex discrete sets with nonempty rows and columns by using a uniform random distribution

---

Input: The integers m and n.
Output: An hv-convex discrete set of size $m \times n$.

Step 1 let Q be an empty queue and calculate $HV'_{m,n}$;
Step 2 choose a number $r \in [1, HV'_{m,n}]$ from a uniform random distribution;
Step 3 **if** ($r \in [1, HV'^{(1)}_{m,n}]$) $t = 1$;
    else let t be the integer for which $r \in [HV'^{(t-1)}_{m,n} + 1, HV'^{(t)}_{m,n}]$;
Step 4 choose a number $r \in [1, HV'^{(t)}_{m,n}]$ from a uniform random distribution;
Step 5 **if** ($t \neq 1$)
    { **for** $k = 1$ **to** $m - 1$
        for $l = 1$ **to** $n - 1$
            if ($r > P_{k,l} \cdot HV'^{(t-1)}_{m-k,n-l} \cdot t$) $r = r - P_{k,l} \cdot HV'^{(t-1)}_{m-k,n-l} \cdot t$;
            else { put (k, l) into Q;
                goto Step 4 with parameters $m - k$, $n - l$, and $t - 1$; }
    }
    else put (m, n) into Q;
Step 6 **foreach** (k, l) in Q **do** generate an hv-convex polyomino of size $k \times l$
    by using a uniform random distribution;
Step 7 choose a permutation π of order $|Q|$ from a uniform random distribution
    and permute the column sets of the components according to π;

Unfortunately, the formula calculated in Step 1 of Algorithm GENHV' is more complex than the formula used in Algorithm GENHV-S'. In fact, the preprocessing step of Algorithm GENHV' takes $O(m^2n^2 \cdot \min\{m, n\})$ time and $O(mn \cdot \min\{m, n\})$ memory requirements. Similarly as in Algorithm GENHV-S', the determination of the number and the sizes of the components (steps 2–5) takes $O(mn \cdot \min\{m, n\})$ time. Regarding the complexity of Step 6 of Algorithm GENHV' we can make the same observations as for Step 4 of Algorithm GENHV-S'. Finally, in Step 7 a random permutation of order $|Q|$ will be generated, where for the number of elements in Q $1 \leq |Q| \leq \min\{m, n\}$ holds. Thus, the last step can be performed in $O(\min\{m, n\})$ time.

We should also add that, by using jointly (6) and (8), we can develop Algorithm GENHV in a straightforward way for generating arbitrary hv-convex sets – perhaps with empty rows or/and columns – from a uniform random distribution. After, the total number of hv-convex discrete sets can be calculated by a formula similar to (9).

4. Random generation of Q-convex discrete sets

Beside the class of hv-convex sets an other important class in discrete tomography is the class of Q-convex discrete sets that was introduced in [7]. Up to now, this class is one of the biggest known classes where the reconstruction from two projections can be performed in polynomial time. Inspired by [20], in [8] a probabilistic method was described that can generate elements of this class from uniform random distributions as well, by which important statistics from the Q-convex sets could be collected [9,10]. The authors of [8] also described an asymptotic formula for the number of Q-convex discrete sets.

The main advantage of the methods employed in Section 3 is that they can be readily extended to any class of discrete sets with disjoint components (even when the components are not hv-convex) if the components can be generated themselves from a uniform random distribution – knowing their bounding rectangles – and if it is possible to enumerate them. For example, we can generalize our method to enumerate Q-convex discrete sets with a given size, and to give an other uniform generator for the class of Q-convex discrete sets as well. Again, we will study just discrete sets with nonempty rows and columns. The reader can directly extend the results for the class of arbitrary Q-convex sets. Let us introduce the notation $D^{NW}_{m,n}, D^{NE}_{m,n}, D^{SW}_{m,n}, D^{SE}_{m,n}, L^{NW}_{m,n}, L^{NE}_{m,n}$, and $Q_{m,n}$ for the number of NW-directed, NE-directed, SW-directed, SE-directed hv-convex polyominoes, NW-parallelogram, NE-parallelogram polyominoes, and Q-convex discrete sets (with no empty rows and columns) of size $m \times n$, respectively. Moreover, let $T^{NW}_{m,n}$ denote the number of canonical discrete sets which components are all NW-parallelogram polyominoes and which do not contain empty rows and columns. With this notation

Theorem 8. *For each* $m, n > 1$

$$T^{NW}_{m,n} = L^{NW}_{m,n} + \sum_{k<m,\ l<n} L^{NW}_{k,l} \cdot T^{NW}_{m-k,n-l} \tag{11}$$

and

$$Q_{m,n} = P_{m,n} + 2 \sum_{i<m,\ j<n} D^{NW}_{i,j} \cdot \left(\sum_{k\le m-i,\ l\le n-j} D^{NW}_{k,l} \cdot T^{NW}_{m-i-k,n-j-l} \right). \tag{12}$$

Proof. Eq. (11) follows analogously to (4).

To prove Eq. (12) we recall the following observations from [1]. A Q-convex set is either a hv-convex polyomino or it consists of several hv-convex components. Let F be a non-4-connected Q-convex set having components $F_1, \dots, F_k$ such that $\{i_l, \dots, i'_l\} \times \{j_l, \dots, j'_l\}$ is the minimal bounding rectangle of the l-th ($l = 1, \dots, k$) component of F. Without loss of generality we can assume that $1 = i_1 \le i'_1 < i_2 \le i'_2 < \cdots \le i'_k = m$. Then, either $1 = j_1 \le j'_1 < j_2 \le j'_2 < \cdots \le j'_k = n$, or $n = j_1 \ge j'_1 > j_2 \ge j'_2 > \cdots \ge j'_k = 1$. Consequently, a Q-convex set is always canonical or anticanonical. In the former case $F_1, \dots, F_{k-1}$ are NW-directed and $F_2, \dots, F_k$ are SE-directed (that is, $F_2, \dots, F_{k-1}$ are NW-parallelogram polyominoes). In the latter case $F_1, \dots, F_{k-1}$ are NE-directed and $F_2, \dots, F_k$ are SW-directed (that is, $F_2, \dots, F_{k-1}$ are NE-parallelogram polyominoes). In particular, we also get that there are Q-convex sets which have just two components and which do not contain any parallelogram polyomino between these two components.

Due to symmetry $D^{NW}_{m,n} = D^{NE}_{m,n} = D^{SW}_{m,n} = D^{SE}_{m,n}, L^{NW}_{m,n} = L^{NE}_{m,n}$, and the number of canonical and anticanonical Q-convex sets which are not polyominoes is the same. Therefore, it is sufficient to calculate the number of canonical Q-convex discrete sets and multiply the result by 2. The structure of such a set is the following. It contains an NW-directed polyomino of size $i \times j$ in the upper-left corner (where $i < m$ and $j < n$), an SE-directed polyomino of size $k \times l$ in the bottom right corner (where $k \le m - i$ and $l \le n - j$), and the remaining part is a canonical discrete set of size $(m - i - k) \times (n - j - l)$ whose components are all NW-parallelogram polyominoes. Thus, we get the formula (12). □

Now, setting $D^{NW}_{1,j} = D^{NW}_{i,1} = L^{NW}_{0,j} = L^{NW}_{1,j} = L^{NW}_{i,0} = L^{NW}_{i,1} = T^{NW}_{1,j} = T^{NW}_{i,1} = 1$, and $T_{0,j} = T_{i,0} = 0$ for each $i = 1, \dots, m$ and $j = 1, \dots, n$ our method can be adapted to generate Q-convex discrete sets (with nonempty rows and columns) by using uniform random distributions. This occurs if $D^{NW}_{m,n}$ and $L^{NW}_{m,n}$ can be calculated for any arbitrary m and n, and if we can generate NW-directed hv-convex polyominoes and NW-parallelogram polyominoes with minimal bounding rectangles of a fixed size $m \times n$ by using uniform random distributions. For the number of hv-convex NW-directed (parallelogram) polyominoes of size $m \times n$ we obtain the necessary formulas from [5]. Namely,

$$D^{NW}_{m,n} = \binom{m+n-2}{m-1}\binom{m+n-2}{n-1} \tag{13}$$

and

$$L^{NW}_{m,n} = \frac{1}{m+n-1}\binom{m+n-1}{m-1}\binom{m+n-1}{n-1}. \tag{14}$$

The complexity of the modified algorithm is the same as of GENHV-S', except the step where the components are generated from uniform random distributions. Based on the bijective proofs of [5], an hv-convex NW-directed polyomino of size $m \times n$ can be generated from a uniform distribution in $O(m + n)$ time, while a NW-parellelogram polyomino (as a special hv-convex NW-directed polyomino) can be generated by using the method of rejection.

5. Conclusions and discussion

In this paper we developed several variants of a technique for generating hv-convex discrete sets by using uniform random distributions. The main advantage of our method is that it can be applied to any class of discrete sets that have disjoint components if the components themselves can be generated from uniform random distributions by knowing their bounding rectangles, and if it is possible to enumerate them. To show this, we presented the idea of a uniform random generator for the class of Q-convex discrete sets as well. In addition, similar techniques can be developed that generate discrete sets according to other parameters such as area and perimeter.

Another way of generalization is to drop the uniform random distributions, so generating some sets with bigger probabilities than others. By manipulating the weight factor t in the recursive formula (8) used in the generation algorithm, we can assign bigger weights to some sets we prefer and smaller ones to the sets that we want to occur less often in the generation process. For example, we can select generated sets that have more components and not ones that have fewer components. Actually, $w(t) = t$ is just a special weight function that yields a uniform random distribution.

We implemented algorithms GENHV-S', GENHV-S, GENHV', and GENHV in C++ using the long integer functions of library NTL-5.4 [24] on a PC with Intel Pentium 4 processor of 3.2 GHz and 1 GB RAM under Debian GNU/Linux 3.1 with Kernel 2.6.17.13. We found that 1000 sets of $\mathcal{S}$ or $\mathcal{S}'$ up to a size of 200×200 can be generated in about a day. But – because of its huge computational complexity (which is mainly due to the preprocessing step) – our generation method is usually applicable for discrete sets of $\mathcal{HV}$ of moderate size only. For example, for the generation of 1000 sets of the $\mathcal{HV}$ class of size 80×80 the generation algorithm takes a few days to run. This seems to be the only disadvantage of the method presented. The on-line generation of sets of large size with this method is not feasible. We think that this is not really such a big problem as benchmark sets used in testing reconstruction algorithms have to be generated once and can then be used over and over again. Nevertheless, it is an important and interesting open question whether more sophisticated and more efficient generation techniques for the classes of hv-convex and Q-convex discrete sets can be developed to overcome current difficulties.

The presented methods allow us to examine some important properties of hv-convex or Q-convex discrete sets. Statistics, for an example, about the expected number of components of a discrete set, can be especially useful in the reconstruction task. It tells us something about the discrete set to be reconstructed before we attempt to reconstruct it. Thus, such statistics opens the way to designing reconstruction algorithms in the future that exploit information known beforehand about the expected number of components. We believe that such algorithms could be more effective in practice than the previously developed ones which do not make use of such a priori knowledge.

References

[1] P. Balázs, A decomposition technique for reconstructing discrete sets from four projections, Image and Vision Computing 25 (10) (2007) 1609–1619.
[2] E. Balogh, A. Kuba, Cs. Dévényi, A. Del Lungo, Comparison of algorithms for reconstructing hv-convex discrete sets, Linear Algebra Appl. 339 (2001) 23–35.
[3] E. Barcucci, A. Del Lungo, M. Nivat, R. Pinzani, Reconstructing convex polyominoes from horizontal and vertical projections, Theoret. Comput. Sci. 155 (1996) 321–347.
[4] E. Barcucci, A. Del Lungo, M. Nivat, R. Pinzani, Medians of polyominoes: A property for the reconstruction, Int. J. Imaging Syst. Technol. 9 (1998) 69–77.
[5] E. Barcucci, A. Frosini, S. Rinaldi, On directed-convex polyominoes in a rectangle, Discrete Math. 298 (2005) 62–78.
[6] K.J. Batenburg, An evolutionary algorithm for discrete tomography, Discrete Appl. Math. 151 (2005) 36–54.
[7] S. Brunetti, A. Daurat, An algorithm reconstructing convex lattice sets, Theoret. Comput. Sci. 304 (2003) 35–57.
[8] S. Brunetti, A. Daurat, Random generation of Q-convex sets, Theoret. Comput. Sci. 347 (2005) 393–414.
[9] S. Brunetti, A. Daurat, Stability in discrete tomography: Some positive results, Discrete Appl. Math. 147 (2–3) (2005) 207–226.
[10] S. Brunetti, A. Daurat, A. Kuba, Fast filling operations used in the reconstruction of convex lattice sets, Lecture Notes in Comput. Sci. 4245 (2006) 98–109.
[11] S. Brunetti, A. Del Lungo, F. Del Ristoro, A. Kuba, M. Nivat, Reconstruction of 4- and 8-connected convex discrete sets from row and column projections, Linear Algebra Appl. 339 (2001) 37–57.
[12] M. Chrobak, Ch. Dürr, Reconstructing hv-convex polyominoes from orthogonal projections, Inform. Process. Lett. 69 (6) (1999) 283–289.
[13] G. Dahl, T. Flatberg, Optimization and reconstruction of hv-convex (0, 1)-matrices, Discrete Appl. Math. 151 (2005) 93–105.
[14] M.P. Delest, G. Viennot, Algebraic languages and polyominoes enumeration, Theoret. Comput. Sci. 34 (1984) 169–206.
[15] I. Dutour, J.M. Fédou, Object grammars and random generation, Disc. Math. and Theoret. Comput. Sci. 2 (1998) 49–63.
[16] P. Flajolet, P. Zimmermann, B. Van Cutsem, A calculus for the random generation of labelled combinatorial structures, Theoret. Comput. Sci. 132 (1994) 1–35.
[17] I. Gessel, On the number of convex polyominoes, Ann. Sci. Math. Québec 24 (2000) 63–66.
[18] G.T. Herman, A. Kuba (Eds.), Discrete Tomography: Foundations, Algorithms and Applications, Birkhäuser, Boston, 1999.
[19] G.T. Herman, A. Kuba (Eds.), Advances in Discrete Tomography and its Applications, Birkhäuser, Boston, 2007.
[20] W. Hochstättler, M. Loebl, C. Moll, Generating convex polyominoes at random, Discrete Math. 153 (1996) 165–176.
[21] A. Kuba, The reconstruction of two-directionally connected binary patterns from their two orthogonal projections, Comput. Vis. Graph. Image Process 27 (1984) 249–265.
[22] A. Kuba, Reconstruction in different classes of 2D discrete sets, Lecture Notes in Comput. Sci. 1568 (1999) 153–163.
[23] A. Kuba, A. Nagy, E. Balogh, Reconstruction of hv-convex binary matrices from their absorbed projections, Discrete Appl. Math. 139 (2004) 137–148.
[24] V. Shoup, NTL: A library for doing number theory. http://www.shoup.net/ntl.
[25] G.W. Woeginger, The reconstruction of polyominoes from their orthogonal projections, Inform. Process. Lett. 77 (2001) 225–229.