

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Discrete Applied Mathematics 151 (2005) 36–54

**DISCRETE
APPLIED
MATHEMATICS**www.elsevier.com/locate/dam

An evolutionary algorithm for discrete tomography

K.J. Batenburg

*Mathematical Institute, Leiden University, Niels Bohrweg 1, 2333 CA Leiden and CWI, Kruislaan 413,
1098 SJ Amsterdam, The Netherlands*

Received 30 September 2003; received in revised form 6 May 2004; accepted 10 February 2005
Available online 5 July 2005

Abstract

One of the main problems in discrete tomography is the reconstruction of binary matrices from their projections in a small number of directions. In this paper we consider a new algorithmic approach for reconstructing binary matrices from only two projections. This problem is usually underdetermined and the number of solutions can be very large. We present an evolutionary algorithm for finding the reconstruction which maximises an evaluation function, representing the “quality” of the reconstruction, and show that the algorithm can be successfully applied to a wide range of evaluation functions. We discuss the necessity of a problem-specific representation and tailored search-operators for obtaining satisfactory results. Our new search-operators can also be used in other discrete tomography algorithms.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Discrete tomography; Evolutionary algorithm; Network flow

1. Introduction

Discrete tomography (DT) is concerned with the reconstruction of a discrete image from its projections. One of the key problems is the reconstruction of a binary (black-and-white) image from only two projections, horizontal and vertical (see Fig. 1). In 1957 Ryser [23] and Gale [11] independently derived necessary and sufficient conditions for the existence of a solution. Ryser also provided a polynomial time algorithm for finding such a solution. However, the problem is usually highly underdetermined and a large number of solutions may exist [27].

E-mail address: kbatenbu@math.leidenuniv.nl.

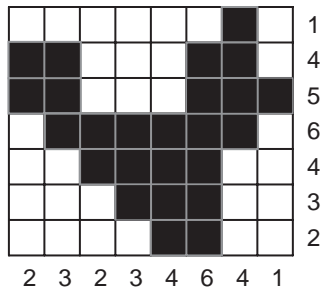


Fig. 1. A binary image with its horizontal and vertical projections.

Among the applications of discrete tomography are the reconstruction of crystal lattices from projections obtained by electron microscopy [17,25] and the reconstruction of angiographic images in medical imaging [22,26]. In such applications the projection data are the result of measurements of a physical object and we are interested in finding a reconstruction which resembles the original image as closely as possible, not just one that corresponds to the given projections. Therefore it is necessary to use all available information about the class of images to which the measured image belongs.

In this paper we assume that the given projection data are consistent (i.e., there is at least one image which has the given projections). In practice the situation is often more complicated, as noise and other errors may result in inconsistent data.

For certain classes of highly-structured images, such as hv-convex polyominoes (the black pixels in each row and column are contiguous), polynomial-time reconstruction algorithms exist (see, e.g., [4,6]). On the other hand, there are classes of images, such as the more general class of hv-convex images (which may consist of many separate polyominoes), for which the reconstruction problem is NP-hard (see [19]).

Instead of assuming specific properties of the image structure, we will focus on a more general approach. Suppose that we are able to define an evaluation function, which assigns a value to each of the solutions, reflecting how good a particular solution is in our context. An algorithm that maximises the evaluation over the set of all solutions will then yield the most desirable solution.

As an example, consider the class of hv-convex images. Suppose that the unknown original image belongs to this class. In [8] it is shown that when we define the evaluation of an image to be the number of neighbouring pairs of black pixels (either horizontal or vertical), the evaluation function is maximised by a reconstruction that has the prescribed projections if and only if the reconstruction is also hv-convex. Similarly, we can define an evaluation function for the reconstruction of hv-convex polyominoes, where we subtract a large penalty if the image consists of more than one polyomino.

Probably of greater practical relevance is the case where the measured object can be considered to be a random variable, sampling from a certain known distribution. In this case the evaluation function reflects the likelihood that the image is a random sample from this distribution.

Finally we remark that the NP-hard problem of reconstructing binary images from more than two projections, including the horizontal and vertical projections, fits in our model as

well. Define the *projection deviation* of an image to be the total difference (as a sum of absolute values) between the image's projections and the prescribed projections. Because we only deal with maximisation problems, we use the negated projection deviation as the evaluation function.

The flexibility of our model comes at a price. In general, the problem of maximising the evaluation function over the set of all solutions is NP-hard, which follows directly from the NP-hardness of the problem of reconstructing binary images from more than two projections (see [12]). In order to deal with this intractability, we resort to the use of modern approximation algorithms. Several preliminary experiments with different algorithms resulted in the choice for an evolutionary algorithm.

In Section 3 we discuss the choices that we made in designing the algorithm. Because the problem at hand has a natural binary encoding it seems attractive to apply a classical genetic algorithm (GA) (see [21]), using the bitstring representation. For several reasons however, this does not lead to good results. We will discuss the problem features that cause the application of classical GA's to be inadequate and introduce new problem-specific mutation and crossover operators. We discuss the necessity of using a hillclimb operator as a post-mutation and post-crossover operator for improving the solution quality.

Section 4 presents our experimental results. The results suggest that our algorithm can be successfully applied to a wide range of evaluation functions, making it very versatile.

2. Preliminaries

We will now introduce some notation and define the DT problem mathematically. Parts of our algorithm are based on well-known theoretical results, which we will summarise (see [15] for details). We assume that the reader is familiar with the theory of network flows and the basic principles of evolutionary algorithms.

Throughout this paper we will assume that all binary matrices are of size $m \times n$. We consider the problem of reconstructing a binary matrix from its horizontal and vertical projections.

Definition 1. Let $R = (r_1, \dots, r_m)$ and $S = (s_1, \dots, s_n)$ be nonnegative integral vectors. We denote the class of all binary matrices $A = (a_{ij})$ satisfying

$$\sum_{j=1}^n a_{ij} = r_i, \quad i = 1, \dots, m,$$

$$\sum_{i=1}^m a_{ij} = s_j, \quad j = 1, \dots, n,$$

by $\mathcal{A}(R, S)$. The vectors R and S are called the row and column projections of any matrix $A \in \mathcal{A}(R, S)$.

Because DT is strongly related to digital image processing we often refer to binary matrices as *images* and call the matrix entries *pixels* with values *black* (1) and *white* (0).

From this point on we assume that the row and column projections are *consistent*, meaning that $\mathcal{A}(R, S)$ is nonempty. In particular, this implies that $\sum_{i=1}^m r_i = \sum_{j=1}^n s_j$. Necessary and sufficient conditions for the nonemptiness of $\mathcal{A}(R, S)$ are given in [18].

One of the basic problems in DT is the *reconstruction problem*:

Problem 2. Let R and S be given integral vectors. Construct a binary matrix $A \in \mathcal{A}(R, S)$.

Because the number of possible solutions of the reconstruction problem can be very large it is necessary to impose additional properties on the solution being sought. In this paper we consider the following problem:

Problem 3 (MAXEVAL). Let R and S be given integral vectors and let $f : \mathcal{A}(R, S) \rightarrow \mathbb{Z}$ be a given evaluation function. Find a binary matrix $A \in \mathcal{A}(R, S)$ such that $f(A)$ is maximal.

Although in Problem 3 the domain of f is restricted to the class $\mathcal{A}(R, S)$, f is usually defined on the entire set $\{0, 1\}^{m \times n}$.

The notion of *switching components* plays an important role in the characterisation of the class $\mathcal{A}(R, S)$.

Definition 4. Let $A \in \mathcal{A}(R, S)$. A switching component of A is a 2×2 submatrix of the form

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Switching components have the property that if we interchange the 0's and 1's, the projections do not change. We call such an interchange operation an *elementary switching operation*. An important theorem of Ryser [24] describes how the class $\mathcal{A}(R, S)$ is characterised by a single element of this class and the set of switching components:

Theorem 5. Let $A \in \mathcal{A}(R, S)$. There exists $B \in \mathcal{A}(R, S)$, $B \neq A$, if and only if A contains a switching component. Moreover, if such a matrix B exists, then A can be transformed into B by a sequence of elementary switching operations.

We remark that every matrix B that is the result of applying a sequence of elementary switching operations to A is also in $\mathcal{A}(R, S)$. The fact that we can transform a matrix in $\mathcal{A}(R, S)$ into any other matrix having the same projections by means of elementary switching operations, makes the use of elementary switching operations very suitable for local search procedures. In our evolutionary algorithm, we make extensive use of these operations.

An important operation in our algorithm is the computation of matrices in $\mathcal{A}(R, S)$, given R and S . We use a network flow approach for computing these matrices, which was first introduced by Gale [11]. First, we construct a directed graph N . The set V of nodes consists of a source node S , a sink node T , one layer V_1, \dots, V_m of nodes that correspond to the image rows (*row nodes*) and one layer W_1, \dots, W_n of nodes that correspond to the image

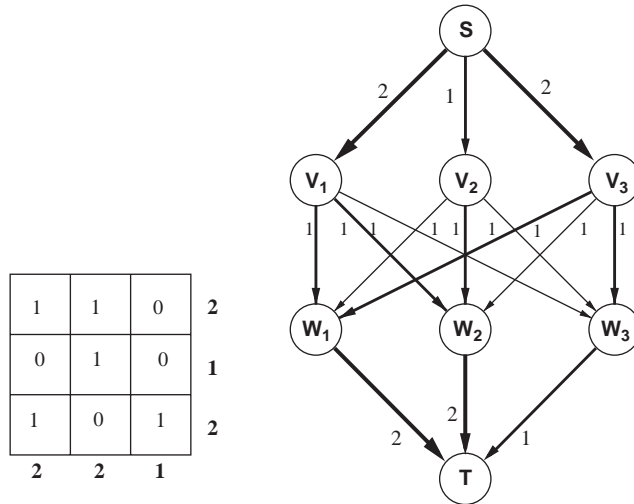


Fig. 2. A 3×3 image instance and its corresponding network flow.

columns (*column nodes*). The set A of arcs consists of arcs (S, V_i) for $i = 1, \dots, m$, arcs (V_i, W_j) for $i = 1, \dots, m, j = 1, \dots, n$, and arcs (W_j, T) for $j = 1, \dots, n$. We remark that the two layers of row nodes and column nodes form a complete bipartite graph.

We also define a *capacity function* $C : A \rightarrow \mathbb{N}_0$ which assigns an integral capacity to every arc. All arcs between a row node and a column node are assigned capacity 1. We will denote these arcs as *pixel-arcs*, because each of them corresponds to a single, unique pixel of the image (determined by the row and column). The prescribed projections determine the capacities of the arcs from the source node and the arcs to the sink node. Every arc (S, V_i) has capacity r_i and every arc (W_j, T) has capacity s_j . Fig. 2 shows the network N for an example 3×3 image.

It is well-known that a maximum flow from S to T can be found in polynomial time. From the theory of linear programming we know that there is a maximum flow which is integral. Suppose that this maximum flow fully saturates all outgoing arcs from the source (and all incoming arcs towards the sink). Every pixel-arc carries a flow of either 0 or 1. Assign a 1 to pixel (i, j) if arc (V_i, W_j) carries a flow of 1 and assign 0 otherwise. It is easy to verify that the resulting image has the prescribed projections in both directions. Conversely, if the horizontal and vertical projections are consistent, i.e., there is an image that satisfies them, the network must have a maximal flow that saturates all outgoing arcs from the source. We see that the problem of finding a maximum integral flow in N is equivalent to the reconstruction problem.

We can use algorithms for solving the max flow problem to compute solutions of the DT reconstruction problem. In our case we want to impose additional requirements on the resulting image. In particular, given a binary image $M \in \{0, 1\}^{m \times n}$, we want to find the binary image $A \in \mathcal{A}(R, S)$ that differs from M in as few pixels as possible. We will refer to M as the *model image*.

Problem 6 (BESTMATCH). Let R and S be given integral vectors and let $M = (m_{ij})$ be a given binary matrix. Find $A = (a_{ij}) \in \mathcal{A}(R, S)$ such that

$$\sum_{i=1}^m \sum_{j=1}^n |m_{ij} - a_{ij}|$$

is minimal.

Problem 6 can be solved efficiently using an extension of the network flow model, incorporating a cost function. We assign a cost $c_{ij} = -m_{ij}$ to every arc (V_i, W_j) . The remaining arcs are all assigned a cost of 0. The process of finding a maximum flow of minimal cost now comes down to finding an integral flow that saturates the maximum number of pixel-arcs for which the corresponding model image pixel has value 1. In other words, if A is the image that corresponds to the maximal flow of minimal cost, then A has as many 1's in common with M as possible.

For any matrix $B = (b_{ij}) \in \{0, 1\}^{m \times n}$, denote the number of 1's in B by n_B . n_M is defined analogously. Suppose that $B \in \mathcal{A}(R, S)$. Clearly, we have $n_B = \sum_{i=1}^m r_i = \sum_{j=1}^n s_j$. Put

$$v_B = |\{(i, j) : b_{ij} = 1 \wedge m_{ij} = 1\}|.$$

Then

$$|\{(i, j) : b_{ij} \neq m_{ij}\}| = (n_B - v_B) + (n_M - v_B) = n_B + n_M - 2v_B,$$

so

$$|\{(i, j) : b_{ij} = m_{ij}\}| = mn - (n_B + n_M - 2v_B),$$

which shows that maximising the number of common 1's between B and M is equivalent to solving Problem 6, because mn , n_B and n_M are constant for $B \in \mathcal{A}(R, S)$. Numerous standard algorithms are available for solving the min cost max flow problem in polynomial time. We can use any of these algorithms for solving Problem 6, see, e.g., [2].

3. Algorithmic approach

3.1. Overview of the approach

In this section we will describe a heuristic algorithm for solving Problem 3 (MAXEVAL). Because finding a reconstruction that maximises the evaluation function is an NP-hard problem, we have to resort to approximation algorithms. In our approach we make extensive use of the fact that Problem 6 (BESTMATCH) can be solved in polynomial time. In preliminary experiments we implemented and tested several approaches, based on simulated annealing [1], tabu search [13] and evolutionary algorithms [3]. On the basis of these preliminary experiments, we believe that simulated annealing and tabu search, which only use local search operators, are not well suited for this task. The main reason for this is that the DT problem usually has a great number of local optima and moving between different optima may require a large number of “uphill” steps.

Evolutionary algorithms can handle this problem in two ways. Firstly, the algorithm uses a diverse population of candidate solutions, instead of a single solution. Secondly, the crossover operator is capable of performing large steps in the search space.

Because the problem at hand allows for a natural bitstring representation of candidate solutions it fits directly into the framework of classical genetic algorithms on bitstrings. Simply use $\{0, 1\}^{mn}$ as the candidate solution space. We now have two optimisation criteria, the evaluation function and the deviation from the prescribed projections.

We found this approach to be inadequate for several reasons. In the first place, the crossover operator usually results in candidate solutions that deviate greatly from the prescribed projections. This is particularly common when the two parent solutions have many differences. This behaviour makes the crossover operator hardly effective at all, reducing the algorithm to a simple hillclimb procedure that only uses mutation to improve upon the solutions. In addition, the most common crossover operators, single-point crossover and uniform crossover, do not take into account the two-dimensional spatial structure of the candidate solutions. Intuitively, building blocks for this problem are made up of regions of pixels, not necessarily horizontally or vertically aligned. A problem-specific operator could benefit from this structure.

We have designed a problem-specific evolutionary algorithm to overcome the disadvantages of the classical GA. Instead of optimising over all bitstrings in $\{0, 1\}^{mn}$, the algorithm optimises exclusively over $\mathcal{A}(R, S)$. At the end of each generation all candidate solutions have the prescribed horizontal and vertical projections. This requires a new crossover operator, that is not only capable of mixing features from both parents, but also of ensuring that the produced children adhere to the prescribed projections. Similar requirements apply to the mutation operator.

Our algorithm is a *memetic algorithm* (see [7]): after every crossover or mutation operation a stochastic hillclimb is performed until the solution has reached a local optimum. In this way, individuals always represent local optima in the search space. We chose this approach, because although the proposed crossover operator generates children that have the prescribed projections, the children are often of inferior quality with respect to the evaluation function (when compared to the parents). In order to fully exploit the explorative power of the crossover operator, the hillclimb procedure increases the solution quality while still remaining quite close to the solution that resulted from the crossover operation. Another reason for choosing this approach is that our crossover and mutation operators are computationally expensive operations. Memetic algorithms typically require only a small number of generations (while performing a lot of work for each generation).

Fig. 3 summarises our algorithm. The parameters λ and μ are the population size and the number of children that are created in each generation, respectively. In the next sections we will describe the various operators in more detail.

3.2. A new crossover operator

The crossover operator is one of the main parts of our algorithm. The input of the operator consists of two *parent images*. The output is a *child image*, which has certain features from both parents. Because all images in the population are members of $\mathcal{A}(R, S)$, the resulting

```

generate initial population  $P_0$  of size  $\lambda$ , consisting of matrices in  $\mathcal{A}(R, S)$ ;
perform a hillclimb operation on each matrix in  $P_0$ ;
 $t := 0$ ;
while (stop criterium is not met) do
begin
 $P'_t = \emptyset$ ;
for  $i := 1$  to  $\mu$  do
begin
generate a child matrix  $C$ , by crossover or mutation;
perform a hillclimb operation on  $C$ ;
 $P'_t := P'_t \cup \{C\}$ ;
end;
select new population  $P_{t+1}$  from  $P_t \cup P'_t$ ;
 $t := t + 1$ ;
end
output the best individual found;

```

Fig. 3. Outline of the evolutionary algorithm.

```

while (there are border cells) do
begin
randomly select a border cell  $(x, y)$ ;
mark all unassigned neighbouring cells of  $(x, y)$  as border cells;
assign these cells to the same parent as  $(x, y)$ ;
remove the border status of  $(x, y)$ ;
end

```

Fig. 4. Main loop of the procedure that generates the crossover mask.

image should have the prescribed projections. In order to enforce this constraint, we use the network flow formulation of the DT problem.

First, a *crossover mask* $X = (x_{ij}) \in \{0, 1\}^{m \times n}$ is computed, which determines for each pixel from which parent image it inherits. In the following description, *assigning cell* (i, j) *to the first parent* means setting $x_{ij} = 0$ and *assigning cell* (i, j) *to the second parent* means setting $x_{ij} = 1$. By *neighbouring cells* of a cell (i, j) we mean the four horizontal and vertical neighbours. As discussed in Section 3.1 we want the crossover operator to take the spacial structure of our image into account. Therefore the crossover mask should assign *areas* of the image to each of the parents. We use the following procedure for realising this. First, one random cell is selected in each of the four quadrants (where the origin is considered to be in the center of the image). Two (randomly chosen) of those cells are assigned to the first parent, the other two are assigned to the second parent. The four cells are marked as *border cells*. Next, the algorithm enters a loop, shown in Fig. 4. Because the first four border cells are each in a different quadrant, this procedure results in a crossover mask that roughly assigns the same number of cells to both parents.

From the crossover mask and both parent images $P = (p_{ij})$ and $Q = (q_{ij})$, a *model image* $M = (m_{ij})$ is computed, as follows:

$$m_{ij} = \begin{cases} p_{ij} & \text{if } x_{ij} = 0, \\ q_{ij} & \text{if } x_{ij} = 1. \end{cases}$$

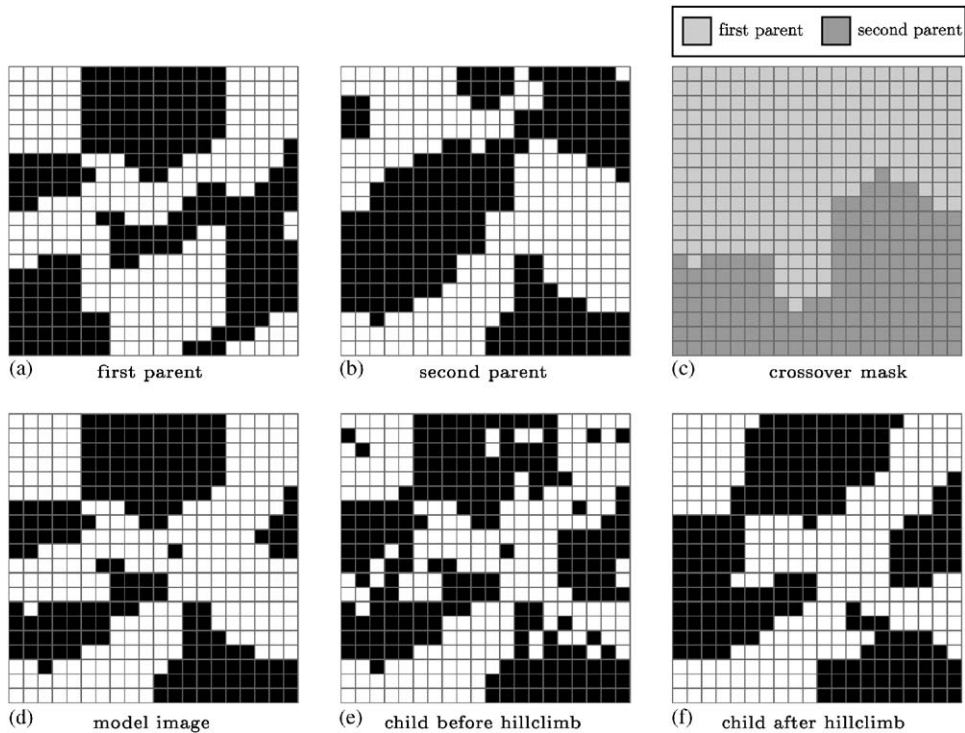


Fig. 5. The crossover operator combines two parent images into a child image.

Subsequently, we construct the child image C by solving Problem 6, using M as the model image. This will result in a child image that is in $\mathcal{A}(R, S)$, resembles the first parent in a certain part and resembles the other parent in the rest of the image. Fig. 5 shows two parent images (having the same projections), a crossover mask, the corresponding model image and the resulting child image. In this example, we use the number of neighbouring pairs of black pixels as the evaluation function. Although the child image resembles both parents in their corresponding parts, it is clear that the child image is far from a local optimum with respect to the evaluation function. To ensure that the child image has sufficient quality, we apply a local hillclimb operator after the crossover operation, which we describe in the next section.

3.3. The hillclimb operator

The hillclimb operator applies a sequence of small modifications, all of which increase the evaluation function, until a local optimum is reached. Because we want the resulting image to be in $\mathcal{A}(R, S)$, we use elementary switching operations as local steps. The outline of the procedure is shown in Fig. 6. It is of great importance that the switching component in each step is chosen randomly among all those yielding an increase of the evaluation

```

while (a switching operation that can improve the evaluation exists) do
begin
  randomly select such a switching operation;
  apply the switching operation;
end

```

Fig. 6. Outline of the hillclimb procedure.

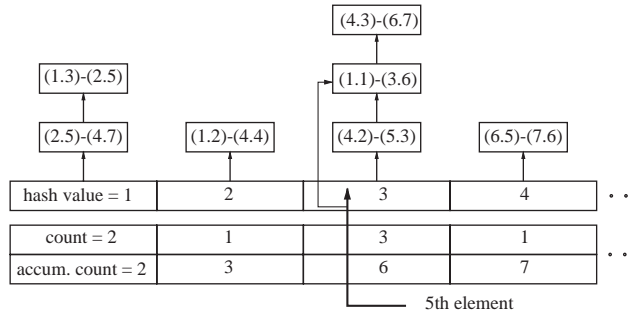


Fig. 7. The hash table of switching components is extended to allow binary search for an indexed element.

function. We have performed experiments with implementations for which this was not the case: the search for switching components always started in the topleft corner and proceeded from left to right, top to bottom. However, this introduces a bias in the hillclimb process, resulting in “skew images”, clearly showing the consequences of the biased operator.

Although the hillclimb operation can be easily described, its implementation offers several computational challenges. Finding all *applicable switching components* (switching components that are present in the image) is an $O((mn)^2)$ operation, making it computationally expensive. Although the *number* of applicable switching components is also $O((mn)^2)$, it is usually much smaller. We store all applicable switching components in a hash table. Switching components are stored as two integer-pairs: (top–left, bottom–right). We can use the hash table to iterate over all applicable switching components. The search efficiency of hash tables is required to update the datastructure after a switching operation has been applied by first removing all switching components containing any of the four pixels involved and subsequently adding all new switching components that contain any of those four pixels. If the hash table is large enough we can perform this update operation in $O(mn)$ time.

A second operation that our datastructure must support is efficient selection of a randomly chosen element. Each hash table entry points to a linked list of switching components. In a separate array, we store the number of switching components that reside in each list. In another array, we compute the accumulated number of switching components, up to and including that entry. Using this information, we can efficiently find the i th element in the hash table, given any index i : first perform a binary search to find the right table entry and then perform a (tiny) linear search to find the actual switching component. Fig. 7 illustrates the approach.

```

initialize the hash table, storing all  $s$  applicable switching components;
repeat
  initialise the accumulated count array;
  generate a random permutation  $\pi$  of  $1, \dots, s$ ;
  for  $i := 1$  to  $s$  do
    begin
      find switching component  $\pi_i$  in the hash table;
      if (applying  $\pi_i$  increases the evaluation) then
        begin
          apply  $\pi_i$ ;
          update the hash table;
          break out of the for-loop and restart in the outer loop;
        end;
      end;
    end;
until (no improving switching component has been found);

```

Fig. 8. The hillclimb procedure.

Fig. 8 shows a more precise formulation of the hillclimb operator. In every iteration of the outer loop the evaluation of the image is increased. If the evaluation function is bounded from above and from below (as is the case for the evaluation functions we consider in more detail later), these bounds yield a bound on the number of iterations of the outer loop. Note that the evaluation function of Problem 3 is discrete.

As the evaluation of the solution increases, the number of iterations in the inner loop until an improvement is found will generally increase. Because the efficiency of the inner loop contributes greatly to the efficiency of the hillclimb procedure we want to make the inner loop as efficient as possible. To improve efficiency, we can use the fact that for many evaluation functions the evaluation function after application of a switching component does not have to be computed from scratch, but can be derived from the previous evaluation (*delta updating*). It is essential to exploit this feature, because the inner loop is executed so often. Despite the performance optimisations that we proposed, the hillclimb procedure is still the bottle-neck for the runtime of our algorithm.

3.4. The mutation operator

Besides the crossover operator, our algorithm also uses a mutation operator. This operator “distorts” a region of the image, while still adhering to the prescribed projections. The mutation operator is composed of several algorithmic steps, very similar to the crossover operator. Instead of the crossover mask, a *mutation mask* $X = (x_{ij}) \in \{0, 1\}^{m \times n}$ is computed. First, a random number $k \in [k_{\min}, k_{\max}]$ and a random cell (i, j) are selected. We assign $x_{ij} = 1$ and mark (i, j) as border cell. Subsequently, a similar procedure as in Fig. 4 is executed. In this case, however, the loop is only executed k times. All cells that are still unassigned after the loop terminates are assigned 0 in the mutation mask.

From the mutation mask X and a parent image P , a model image $M = (m_{ij})$ is computed by assigning $m_{ij} = p_{ij}$ if $x_{ij} = 0$ and assigning m_{ij} a random value (0 or 1) if $x_{ij} = 1$. In other words, the mutation mask determines which part of the parent image will be distorted in the model image.

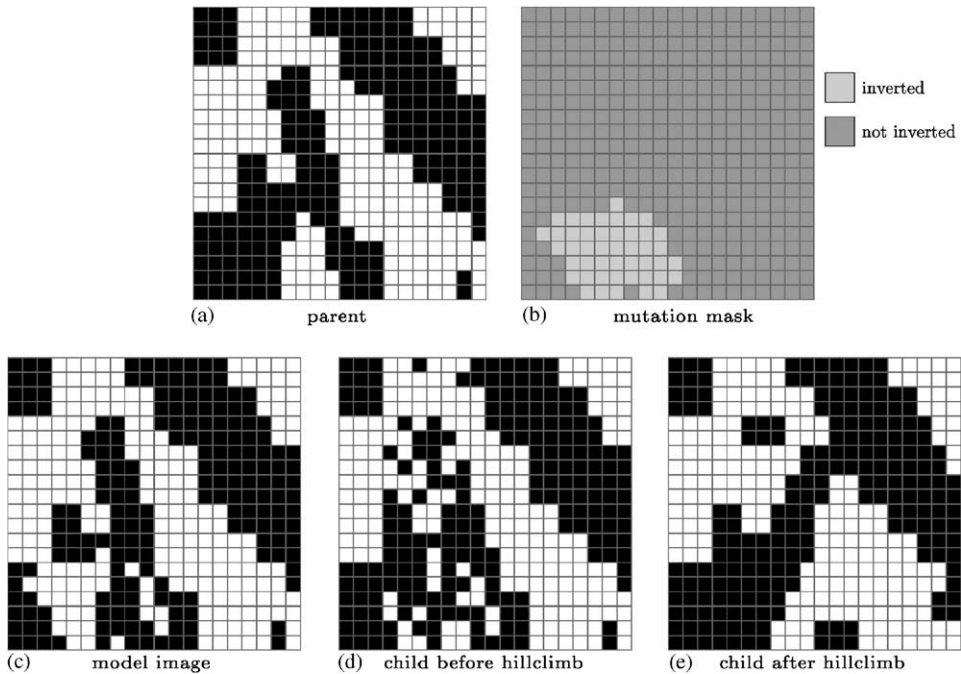


Fig. 9. The mutation operator transforms a parent image into a child image.

After the model image has been computed, the same steps are performed as for the crossover operator: Problem 6 is solved using the weight mask as M . Subsequently, the hillclimb operator is applied, yielding the child individual. Fig. 9 shows the different steps of the mutation operator. As the figure shows, there may be differences between the child and the parent individual outside the distorted region, as a result of the projection constraints.

3.5. Algorithm details

In this section we address several details of our evolutionary algorithm that have not been discussed in the preceding sections.

The initial population is generated by repeatedly solving Problem 6, each time using a new random binary matrix as the weight mask M . Although this procedure does not generate all matrices in $\mathcal{A}(R, S)$ with equal probability, it leads to sufficient diversity in the initial population.

The algorithm is designed to work with a variety of evaluation functions. We use tournament selection, because it is independent of the actual evaluation values (it only depends on their ordering) and because the selective pressure can be adjusted easily, by changing the tournament size.

Before applying the crossover operator, the two parents are randomly selected among the current population. The same applies to parent selection for the mutation operator. The

probabilities of crossover vs. mutation are adjusted dynamically after each generation. For every generated child image, the algorithm keeps track of the operator that generated it. It also counts the number of children generated by crossover (a_c) and by mutation (a_m). After the new population has been selected, the algorithm counts the number of individuals in the new population that were created by crossover (b_c) and by mutation (b_m). From these numbers the average *crossover yield* $y_c = b_c / (a_c + 1)$ and *mutation yield* $y_m = b_m / (a_m + 1)$ are computed. (We add 1 to the denominator to avoid division by zero). The probability that a child will be generated through crossover in the next generation is set to $y_c / (y_c + y_m)$. In this way, the algorithm creates more children by crossover in the next generation when the crossover operation in the current generation creates children of high quality. An explicit lower- and upperbound for y_c are used: $0.1 < y_c < 0.9$. If the computed value of y_c is outside these bounds, it is adjusted.

According to Section 3.4, the mutation operation randomly selects an integer k in the interval $[k_{\min}, k_{\max}]$. For all experiments in Section 4 we used $k_{\min} = mn/64$ and $k_{\max} = 5mn/64$.

Because the algorithm can be used with many different evaluation functions, it is difficult to design a universal termination condition. For some specific problems, such as the reconstruction from three projections, an upper bound on the evaluation function value is known in advance and the algorithm can terminate as soon as an individual having that evaluation is found. For such problems, having solutions for which optimality can easily be checked, separate termination procedures can be implemented in our system. In general, however, the algorithm terminates when no improvement on the best solution so far has been found in the last 20 generations.

As indicated in Fig. 3 the selection procedure selects individuals for the new population from both the old population and the group of new children. Individuals are allowed to be selected more than once for the new population. We performed experiments to find suitable values for the population size λ and the number of children μ that is created in each generation. We found that our algorithm performs well when $\lambda > \mu$, and with a large population. For all the experiments in Section 4 we used $\lambda = 1000$, $\mu = 500$ and a tournament size of 3. We remark that although we have chosen suitable values for the algorithm parameters, other values may work well too.

4. Computational results

We have implemented our algorithm in C⁺⁺, using the gcc compiler. The implementation is fully object oriented. By using inheritance and virtual methods, different evaluation functions can be used without modifying the algorithm code. The evolutionary algorithm uses a base class *Individual* to represent a member of the population (a single image). For each evaluation function we implemented a derived class, which computes the evaluation function and performs delta updating computations. In this way new evaluation functions can be added with very little extra coding effort and without modifying existing code.

For the implementation of the network flow algorithm we used the *RelaxIV solver* [5], combined with the *MCFClass* library [10]. The source code of our algorithm (except the MCF library) is available from the author.

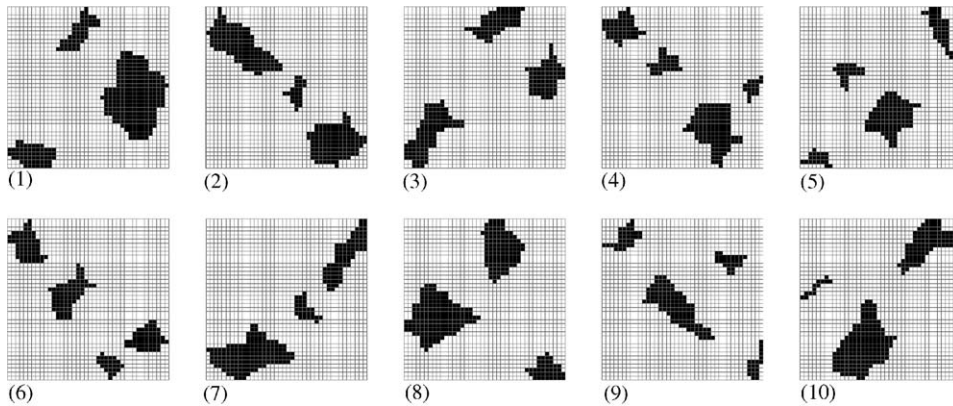


Fig. 10. hv-convex test images.

Table 1
Reconstruction results for the set of hv-convex images

Image	1	2	3	4	5	6	7	8	9	10
#Perfect	9	5	8	8	9	6	6	6	6	9
Average time (min)	25.6	16.6	15.5	8.8	8.5	10.8	17.7	17.7	8.8	15.7
Average #generations	24.6	17.3	18.5	14.8	15.8	18.5	30.2	15.0	21.0	23.6

All our experiments were run on a Pentium IV 2800 MHz PC with 512 Mb of memory. We performed experiments with several evaluation functions. The evaluation functions are very different from each other, in order to demonstrate the versatility of the problem model and the algorithm. The experimental results are intended to demonstrate the feasibility and flexibility of our approach, not to provide extensive results on each of the individual problems that we consider.

4.1. Reconstruction of hv-convex images

The first class of test images consists of *hv-convex* images. Using methods from [16] we generated 10 hv-convex images of size 40×40 . Each image consists of three or four hv-convex objects. There are no empty horizontal or vertical lines between the objects. The test images are shown in Fig. 10. For this image class, the evaluation function is the total number of neighbouring black pixels (horizontal and vertical). We performed ten test runs for each of the images. Table 1 shows the results. We call a reconstruction *perfect* if it is the same as the original image. The table shows the number of runs that achieved a perfect reconstruction, the average runtime of the algorithm (in min) and the average number of generations after which the best reconstruction was found. When the algorithm did not find a perfect reconstruction, the reconstruction was always very different from the original image. Therefore, we do not report solution quality for those cases.

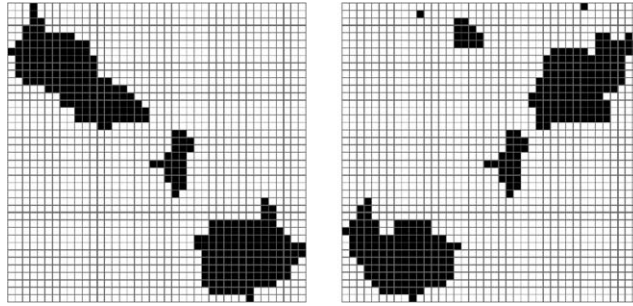


Fig. 11. Two locally optimal images having the same projections, yet being very dissimilar.

In the cases that the algorithm did not find the optimal solution, it converged to a local optimum. Fig. 11 demonstrates the difficulties involved in this optimisation problem. It shows the second test image, having 155 neighbouring black pixels, and another locally optimal image having the same projections and 175 neighbouring black pixels. The images are very dissimilar, except for the object in the center. This problem is due to the presence of large blocks of switching components.

4.2. Reconstruction from three projections

The second group of test images are *phantom* images from [9]. These images have been used as test images in several publications [14,20]. In this case we reconstruct the images from three projections: horizontal, vertical and diagonal (top left to bottom right). The evaluation function to be maximised is the negated total deviation of the image's diagonal projection from the prescribed projection (as a sum of absolute values). The three test images and their reconstructions are shown in Fig. 12. We performed a single test run for each of the images. The first two images were reconstructed perfectly (the reconstruction was equal to the original image) within 8 s. For the third image, however, the algorithm did not find a reconstruction with the given diagonal projection. The resulting reconstruction has a total difference of 4 from the prescribed diagonal projection, and took 30 s to compute.

The main strength of our algorithm is its flexibility. All test images are completely 4-connected and have no “holes” in them. When we assume this as *a priori information*, we can incorporate it in the evaluation function.

For an image A , put

$$f(A) = -c \times d(A) + b(A),$$

where $d(A)$ denotes the total difference between the diagonal projection of A and the prescribed projection, $b(A)$ denotes the total number of pairs of neighbouring black pixels in A and c is a very large constant. Maximising $f(A)$ will result in the reconstruction that satisfies the prescribed diagonal projection which has as many neighbouring black pixels as possible. Using this new evaluation function, the algorithm reconstructed the original images

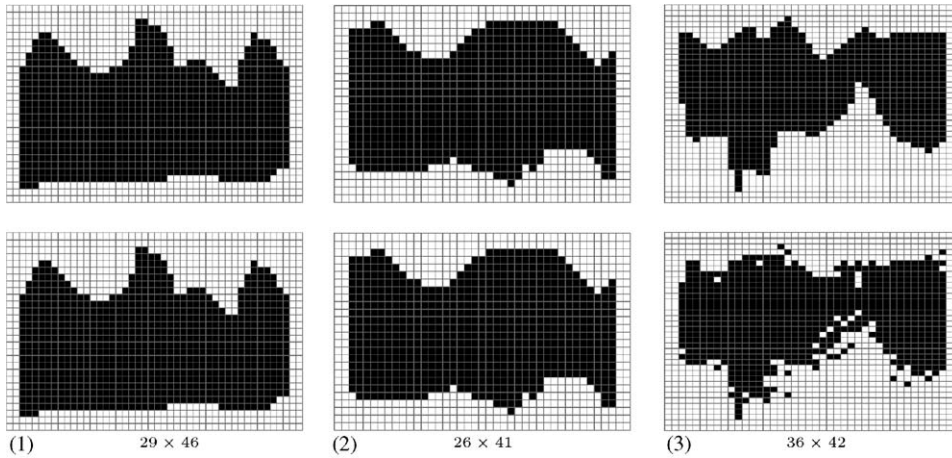


Fig. 12. Three phantom images and their reconstructions.

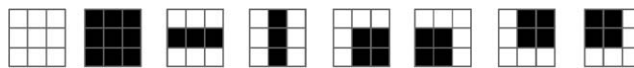


Fig. 13. The eight patterns having a nonzero local energy.

perfectly within 1 min, even the third image. This experiment shows that reconstructions from three projections can be very accurate if appropriate a priori information is used and that such information can easily be incorporated in our algorithm.

4.3. Reconstruction using Gibbs priors

The third evaluation function involves a probability distribution, known as a *Gibbs* distribution. A method for using Gibbs priors in the reconstruction process was presented in [20]. We use similar definitions. We assume that the original image is a random sample of a certain known probability distribution. Let Π be such a probability distribution, on the set $\{0, 1\}^{m \times n}$, which determines for each matrix the probability that it is sampled. The probability $\Pi(A)$ of a matrix A is given by

$$\Pi(A) = \frac{1}{Z} e^{\beta \sum_{i=1}^m \sum_{j=1}^n I_{ij}(A)}.$$

Here, Z is a normalising factor, β is a parameter defining the “peakedness” of the distribution and $I_{ij}(A)$ is the *local energy function* of cell (i, j) . The local energy function of a cell is determined by the value of the cell and each of its eight neighbours. Border pixels, which have less than eight neighbours, are treated as if the lacking neighbours are white. Fig. 13 shows the eight patterns that have nonzero local energy: *homogeneous white and black*, horizontal and vertical *edges*, and four *corners*. The corresponding local energy values are

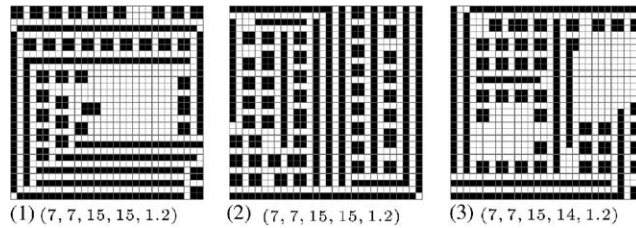


Fig. 14. Three samples from Gibbs distributions.

Table 2
Reconstruction results for the Gibbs samples

Image	1	2	3
#Perfect	10	5	10
Average time (min)	16.4	19.5	14.4
Average #generations	17.3	36.8	20.1

k_w, k_b, k_e and k_c respectively. We denote a Gibbs distribution by the 5-tuple $(k_w, k_b, k_e, k_c, \beta)$, which we call the *parameters* of the distribution. As we know that the original image is a random sample from the given Gibbs distribution, we want to find a reconstruction that has maximal likelihood. We remark that this reconstruction is not guaranteed to be equal to the original image.

Maximising $\Pi(A)$ is equivalent to maximising $E(A) = \sum_{i=1}^m \sum_{j=1}^n I_{ij}(A)$, since $\Pi(A)$ is a monotonously increasing function of $E(A)$. Computing $E(A)$ involves only the known five parameters of the distribution and the image A , we do not need to compute Z .

We implemented a *Metropolis algorithm*, described in [20], for generating random samples from Gibbs distributions. Three such samples of size 30×30 and their parameters are shown in Fig. 14. The reason that we selected three images, instead of showing results for a large number of images, is that many generated images are not tomographically challenging. For example, many contain several lines that only consist of black or white pixels. The fact that all three images consist of lines and blocks of four pixels is due to the selected patterns in Fig. 13. Different patterns lead to other types of images. As our goal here is to illustrate the feasibility of our approach, we only show results for this set of patterns. We performed ten test runs for each of the three images. Table 2 shows the reconstruction results. For the third image, the algorithm actually found a reconstruction that is slightly different from the original image, but has higher likelihood. We still marked this reconstruction as a perfect reconstruction. When the algorithm did not find a perfect reconstruction (image 2), the reconstruction was always very different from the original image. Therefore, we do not report solution quality for those cases. The results show clearly that perfect reconstruction from only two projections and a given Gibbs distribution is possible in nontrivial cases.

5. Conclusions

We have designed a new algorithm for finding a binary image that satisfies prescribed horizontal and vertical projections and has an optimal or near-optimal evaluation function value. Our experimental results demonstrate that the algorithm is effective for several different evaluation functions, corresponding to diverse reconstruction problems.

As demonstrated by the results in Section 4.2 a main feature of our approach is its ability to incorporate various forms of a priori information in the evaluation function.

The bottleneck of the algorithm is the hillclimb operator. Because our hillclimb procedure involves keeping track of all applicable switching components, our algorithm is limited to images of size 50×50 or less. Perhaps making the hillclimb operation less exhaustive could result in improved time complexity of the operation.

In our experiments we often observed that really finding the global optimum of the evaluation function was necessary to find an image that resembled the unknown original image, which was used to generate the projection data. Near-optimal solutions were usually very different. In order to increase the stability of the search procedure, it is necessary to incorporate as much information as possible in the evaluation function, in order to provide proper gradients throughout the search space.

The various reconstruction results show that when sufficient information is incorporated, our algorithm is capable of finding accurate reconstructions.

Acknowledgements

The author would like to thank Walter Kusters, Hendrik Jan Hoogetboom, Herman te Riele and Willem Jan Palenstijn for their valuable comments and suggestions.

References

- [1] E. Aarts, J. Korst, *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*, Wiley, Englewood Cliffs, NJ, 1996.
- [2] R.K. Ahuja, T.L. Magnanti, J.B. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Prentice-Hall, New York, 1993.
- [3] Th. Bäck, D.B. Fogel, T. Michalewicz (Eds.), *Evolutionary Computation 1*, Institute of Physics Publishing, Bristol, Philadelphia, 2000.
- [4] E. Barucci, A. Del Lungo, M. Nivat, R. Pinzani, Reconstructing convex polyominoes from horizontal and vertical projections, *Theoret. Comput. Sci.* 155 (1996) 321–347.
- [5] D.P. Bertsekas, P. Tseng, Relax-IV: a faster version of the relax code for solving minimum cost flow problems, Technical Report LIDS-P-2276, M.I.T., 1994.
- [6] 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.
- [7] D. Corne, M. Dorigo, F. Glover, *New Ideas in Optimisation*, McGraw-Hill Education, 1999.
- [8] G. Dahl, T. Flatberg, Optimization and reconstruction of hv-convex $(0, 1)$ -matrices, in: A. Del Lungo, V. Di Gesù, A. Kuba (Eds.), *Electronic Notes in Discrete Mathematics*, vol. 12, Elsevier, Amsterdam, 2003.
- [9] P. Fishburn, P. Schwander, L. Shepp, R. Vanderbei, The discrete Radon transform and its approximate inversion via linear programming, *Discrete Appl. Math.* 75 (1997) 39–61.
- [10] A. Frangioni, C. Gentile, The MCFClass Project. <http://www.di.unipi.it/di/groups/optimize/Software/MCF.html>, 2003.

- [11] D. Gale, A theorem on flows in networks, *Pacific J. Math.* 7 (1957) 1073–1082.
- [12] R.J. Gardner, P. Gritzmann, D. Prangenberg, On the computational complexity of reconstructing lattice sets from their X-rays, *Discrete Math.* 202 (1999) 45–71.
- [13] F. Glover, M. Laguna, *Tabu Search*, Kluwer Academic Publishers, Dordrecht, 1998.
- [14] L. Hajdu, R. Tijdeman, An algorithm for discrete tomography, *Linear Algebra Appl.* 339 (2001) 147–169.
- [15] G.T. Herman, A. Kuba (Eds.), *Discrete Tomography: Foundations, Algorithms and Applications*, Birkhäuser, Boston, 1999.
- [16] W. Hochstättler, M. Loebl, C. Moll, Generating convex polyominoes at random, *Discrete Math.* 153 (1996) 165–176.
- [17] C. Kisielowski, P. Schwander, F. Baumann, M. Seibt, Y. Kim, A. Ourmazd, An approach to quantitative high-resolution transmission electron microscopy of crystalline materials, *Ultramicroscopy* 58 (1995) 131–155.
- [18] A. Kuba, G.T. Herman, Discrete tomography: a historical overview, in: G.T. Herman, A. Kuba (Eds.), *Discrete Tomography: Foundations, Algorithms and Applications*, Birkhäuser, Boston, 1999, pp. 3–34, (Chapter 1).
- [19] A. Del Lungo, M. Nivat, Reconstruction of connected sets from two projections, in: G.T. Herman, A. Kuba (Eds.), *Discrete Tomography: Foundations, Algorithms and Applications*, Birkhäuser, Boston, 1999, pp. 163–188, (Chapter 7).
- [20] S. Matej, A. Vardi, G.T. Herman, E. Vardi, Binary tomography using gibbs priors, in: G.T. Herman, A. Kuba (Eds.), *Discrete Tomography: Foundations, Algorithms and Applications*, Birkhäuser, Boston, 1999, pp. 191–212, (Chapter 8).
- [21] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, 3rd Revision ed., Springer, Berlin, 1996.
- [22] D.G.W. Onnasch, G.P.M. Prause, Heart chamber reconstruction from biplane angiography, in: G.T. Herman, A. Kuba (Eds.), *Discrete Tomography: Foundations, Algorithms and Applications*, Birkhäuser, Boston, 1999, pp. 385–401, (Chapter 17).
- [23] H.J. Ryser, Combinatorial properties of matrices of zeros and ones, *Canad. J. Math.* 9 (1957) 371–377.
- [24] H.J. Ryser, *Combinatorial Mathematics*. The Carus Mathematical Monographs no. 14, American Mathematical Society, Providence, RI, 1963 (Chapter 6).
- [25] P. Schwander, C. Kisielowski, F. Baumann, Y. Kim, A. Ourmazd, Mapping projected potential, interfacial roughness, and composition in general crystalline solids by quantitative transmission electron microscopy, *Phys. Rev. Lett.* 71 (1993) 4150–4153.
- [26] C.H. Slump, J.J. Gerbrands, A network flow approach to reconstruction of the left ventricle from two projections, *Comput. Graphics Image Process.* 18 (1982) 18–36.
- [27] B. Wang, F. Zhang, On the precise number of $(0, 1)$ -matrices in $\mathcal{A}(R, S)$, *Discrete Math.* 187 (1998) 211–220.