Discrete tomography with two directions

Dalen, B.E. van

Discrete tomography with two directions

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 20 september 2011
klokke 15:00 uur

door

Birgit Ellen van Dalen

geboren te 's-Gravenhage
in 1984

Samenstelling van de promotiecommissie

Promotoren

prof.dr. R. Tijdeman
prof.dr. K.J. Batenburg (Centrum Wiskunde & Informatica, Universiteit Antwerpen)

Overige leden

prof.dr. S.J. Edixhoven
prof.dr. H.W. Lenstra, Jr.
prof.dr. A. Schrijver (Centrum Wiskunde & Informatica)
prof.dr. P. Stevenhagen

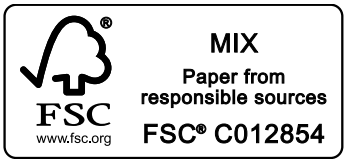
Discrete tomography with two directions

Birgit van Dalen

ISBN/EAN 9789461081803

bevandalen@gmail.com

Typeset using LaTeX
Printed by Gildeprint Drukkerijen, Enschede
Cover design by Ad van den Broek



Contents

CHAPTER 1

Introduction

In this chapter we introduce the topic of discrete tomography and explain the basic concepts. We then describe the part of discrete tomography that this thesis is focused on. We discuss the problems that are considered as well as the main results of the thesis.

1.1 Discrete tomography

Let F be a finite subset of $\mathbb{Z}^2$. If a point of $\mathbb{Z}^2$ is an element of F, we say that the point has value one, or that there is a one in this point. If on the other hand a point of $\mathbb{Z}^2$ is not an element of F, we say that the point has value zero, or that there is a zero in this point. In this way we can view the set F as a function that attaches a value from $\{0, 1\}$ to every point in $\mathbb{Z}^2$, where only finitely many points have value one. We also call this a *binary image*. Rather than considering the whole of $\mathbb{Z}^2$, we usually restrict the image to a rectangle containing all points with value one.

For integers a and b we can consider a line in the direction (a, b), that is, all points $(x, y) \in \mathbb{Z}^2$ satisfying $ay - bx = h$ for a certain integer h. We can count the number of elements of F on this line; this is called the *line sum* of F along this line. We can take all lines in the direction (a, b) that pass through integer points by varying h over $\mathbb{Z}$. The infinite sequence of line sums we find in this way we call the *projection* of the binary image in the direction (a, b). Instead of considering all possible lines in

the direction (a, b), we usually consider a finite set of consecutive lines that contains all lines that pass through points of F. Then the projection becomes a finite sequence of line sums containing all the nonzero line sums.

Given a binary image, the projection in any lattice direction is of course determined. If on the other hand the image is unknown, but the projections in several directions are given, it is not so clear whether the image is determined by these projections, or even whether there exists an image satisfying these projections. The problem of reconstructing binary images from given projections in several lattice directions is what *discrete tomography* is concerned with. An image satisfying given projections is called a *reconstruction*. There may be more than one reconstruction corresponding to given projections, or none at all. If there is exactly one reconstruction, then we say that the projections *uniquely determine* the image.

The term *discrete tomography* is also used for a wider scope of reconstruction problems, such as reconstructing a binary image on $\mathbb{R}^2$ rather than $\mathbb{Z}^2$. Then the domain of the function is no longer discrete, but the possible values of the function form a discrete set, which is why this is still called discrete tomography. And even if we restrict ourselves to functions on lattices, there are still some variations possible. For example, one may consider a function on $\mathbb{Z}^2$ that has a (small) discrete set of values, rather than just $\{0, 1\}$. It is also possible to do discrete tomography in more dimensions, using $\mathbb{Z}^k$ rather than $\mathbb{Z}^2$, or on a hexagonal grid rather than a square grid. A complete overview of discrete tomography is given in [14].

1.2 Applications

The most direct application of discrete tomography is the reconstruction of nanocrystals at atomic resolution. In such a crystal, the atoms usually lie on a regular grid, and only a few types of atoms occur. By electron microscopy, two-dimensional projection images are acquired from various angles by tilting the sample. Recently, new algorithms have been developed that allow a fast and accurate reconstruction from a small number of projection images [7, 17].

There are also some applications in medical imaging [15, 25]. However, much more widely used in medical imaging (among other fields) is the technique of *continuous* or *computerised tomography* [13]. Here images can have values in a continuous set rather than a discrete set, and the object that is being reconstructed does not have a lattice structure, but a continuous structure. For the reconstruction of such images projections in very many directions are needed. The most well-known application of this type of tomography is the CT-scan, where CT stands for "computerised tomography".

Further applications of discrete tomography are for example in nuclear science [19, 20] and materials science [27].

1.3 Two directions

The first discrete tomography problems arose in the literature in 1957, when Ryser published a paper on reconstructing binary images from their projections in the horizontal and vertical directions [24]. He was the first to describe an algorithm to do this, and he gave sufficient and necessary conditions on the projections for a reconstruction to exist.

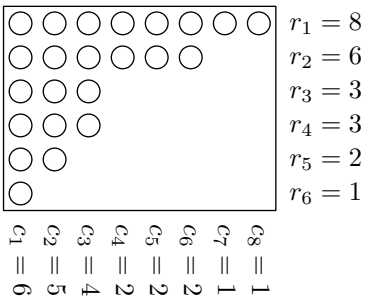


Figure 1.1: A uniquely determined set. The row and column sums are indicated.

Let $(r_1, r_2, \ldots, r_m)$ be the sequence of row sums (the horizontal projection) and let $(c_1, c_2, \ldots, c_n)$ be the sequence of column sums (the vertical projection). We must have $\sum_{i=1}^{m} r_i = \sum_{j=1}^{n} c_j$, since both sums are equal to the number of elements of the binary image. As long as we are only interested in the number of possible reconstructions (and not in special properties of those reconstructions) we can without loss of generality order the rows and columns such that $r_1 \geq r_2 \geq \ldots \geq r_m$ and $c_1 \geq c_2 \geq \ldots \geq c_n$. For $i = 1, 2, \ldots, m$ define $b_i = \#\{j : c_j \geq i\}$. Ryser proved that there exists a set F with those row and column sums if and only if

$$\sum_{i=1}^{k} b_i \geq \sum_{i=1}^{k} r_i \qquad \text{for } k = 1, 2, \ldots, m.$$

He also showed that the reconstruction is unique if and only if

$$\sum_{i=1}^{k} b_i = \sum_{i=1}^{k} r_i \qquad \text{for } k = 1, 2, \ldots, m,$$

or, equivalently,

$$b_i = r_i \qquad \text{for } i = 1, 2, \ldots, m.$$

Such a uniquely determined image has a particular shape [26]. After all, $r_1 = b_1 = \#\{j : c_j \geq 1\}$ means that for every column j with $c_j \geq 1$ there must be an element of F in $(1, j)$. And then $r_2 = b_2 = \#\{j : c_j \geq 2\}$ implies that for every column j with $c_j \geq 2$ there must be an element of F in $(2, j)$, since any column j with $c_j = 1$ contains only one element of F, which is $(1, j)$. By continuing this argument, we find that $(i, j) \in F$ if and only if $c_j \geq i$. This means that

- in row i the elements of F are precisely the points $(i, 1)$, $(i, 2)$, $\ldots$, (i, r_i);
- in column j the elements of F are precisely the points $(1, j)$, $(2, j)$, $\ldots$, (c_j, j).

See Figure 1.1 for an example of a uniquely determined set.

Unfortunately, in discrete tomography with three or more directions such nice properties do not exist. The problem of deciding whether an image is uniquely determined, given projections in three or more directions, is NP-hard. The same holds for the problem of reconstructing an image from its projections in three or more directions [11].

The research in this thesis concerns only discrete tomography in two directions, the horizontal and vertical directions. In the remainder of this chapter we will therefore always use discrete tomography with only horizontal and vertical line sums, unless explicitly mentioned otherwise.

1.4 Stability

Suppose line sums that uniquely determine an image are given. If we slightly tweak those line sums, say by adding 1 to a few row sums and subtracting 1 from exactly as many other row sums, then the resulting line sums may no longer uniquely determine an image. A question that naturally arises from this is: do the reconstructions of the new line sums still look a lot like the original, uniquely determined image, or is it possible that an image satisfying the new line sums is completely different from the original image? This concerns what we call *stability*: the more the reconstructions from the new line sums have in common with the original image, the more *stable* the original image is.

In the case of three or more directions Alpers et al. showed that there can exist two images, both uniquely determined by their line sums, that are disjoint but have almost the same line sums [1, 3]. So in the case of three or more directions, even uniquely determined images are highly unstable. However, this does not hold for discrete tomography with two directions.

Consider given column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$, and define $\mathcal{B} = (b_1, b_2, \ldots, b_m)$ as $b_i = \#\{j : c_j \geq i\}$ for $1 \leq i \leq m$. We have seen in the previous section that the row sums $\mathcal{B}$ and column sums $\mathcal{C}$ uniquely determine an image F_1. Now suppose we have slightly different row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$, such that there exists at least one binary image F_2 with row sums $\mathcal{R}$ and column sums $\mathcal{C}$. Let $N = \sum_{j=1}^{n} c_j$. Furthermore define

$$\alpha = \frac{1}{2} \sum_{i=1}^{m} |r_i - b_i|.$$

Note that α is an integer, since 2α is congruent to

$$\sum_{i=1}^{m} (r_i + b_i) = \sum_{i=1}^{m} r_i + \sum_{i=1}^{m} b_i = 2N \equiv 0 \mod 2.$$

The parameter α measures the difference in the row sums of F_1 and F_2. The stability question now translates into: can it happen that the symmetric difference $F_1 \triangle F_2$ is large (compared to N, the number of elements of F_1), while α is small?

Alpers et al. [1, 2] proved two results related to this question. They showed that if $F_1 \cap F_2 = \emptyset$, then

$$N \leq \alpha^2.$$

So if F_1 and F_2 are disjoint, then α must be large compared to N. On the other hand, they considered the case $\alpha = 1$ and showed that

$$|F_1 \triangle F_2| \leq \sqrt{8N + 1} - 1.$$

In Chapter 2 of this thesis we consider the stability problem for general α. We generalise the above bound to

$$|F_1 \triangle F_2| \leq \alpha\sqrt{8N + 1} - \alpha.$$

We also prove a different bound. Write $p = |F_1 \cap F_2|$, then

$$|F_1 \triangle F_2| \leq 2\alpha + 2(\alpha + p) \log(\alpha + p).$$

By using this bound with $p = 0$, we can derive that if F_1 and F_2 are disjoint, then

$$N \leq \alpha(1 + \log \alpha),$$

which improves the bound of Alpers et al. for disjoint F_1 and F_2.

1.5 Difference between reconstructions

Another interesting question, related to stability, is how much two reconstructions from the same projections can possibly differ. We already know that there exist

images that are uniquely determined. On the other hand, it is not so hard to find images that are disjoint, but have the same line sums. See Figure 1.2(a) for the smallest example and Figure 1.2(b) for a more complex example. But perhaps it is possible to define a collection of "almost uniquely determined images" of which any two reconstructions always must have large intersection?

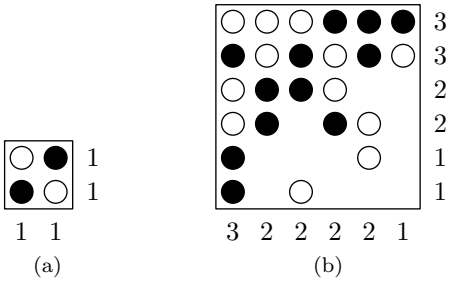


Figure 1.2: Each picture shows two disjoint sets with the same line sums. One set consists of the white points, the other set consists of the black points.

In Chapter 3 we consider this question. First we define a parameter that indicates in some sense how close an image is to being uniquely determined. For this we use the parameter α that we introduced before. As we have seen in the previous section, α measures the distance between a given set F_2 and a given uniquely determined set F_1. For a fixed F_2 we can characterise the sets F_1 that yield the smallest α, and the α corresponding to such a set F_1 is the one we will use.

We study the difference between two sets with the same line sums and small α, and we prove that this difference is bounded from above, using the results from Chapter 2. We also indicate a subset of points that must contain a sizeable part of any reconstruction. On the other hand, we show that α must be large if there exist two disjoint reconstructions. And finally, we generalise everything to reconstructions from different sets of row and column sums.

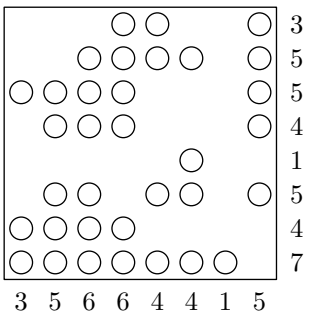
In Chapter 4 we consider the complementary problem: given line sums, find two reconstructions that are as different as possible. Again the parameter α plays an important role, and we show constructively that if $\alpha \geq 1$ (that is, if the projections do not uniquely determine the image) there exist two reconstructions that have a symmetric difference of at least $2\alpha + 2$.

1.6 Boundary length

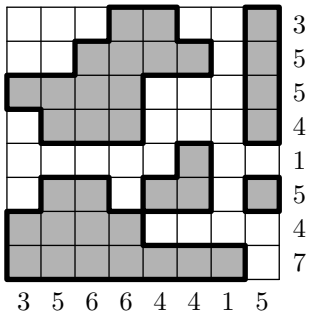
Rather than viewing a binary image as consisting of points in $\mathbb{Z}^2$ that each have value zero or one, we can also view a binary image as consisting of pixels (cells of 1 by 1) that each are white or black. See also Figure 1.3. Now there is a natural way to define the *boundary* of the image: it consists precisely of all the line segments that separate black cells from white cells. Equivalently, the boundary is the set of pairs of points (i, j) and (i', j') in $\mathbb{Z}^2$ such that

- the points are adjacent, that is: $i = i'$ and $|j - j'| = 1$, or $|i - i'| = 1$ and $j = j'$;
- $(i, j) \in F$ and $(i', j') \notin F$.

The *length of the boundary* is the number of pairs of points in this set.



(a) The image is represented by the white points.



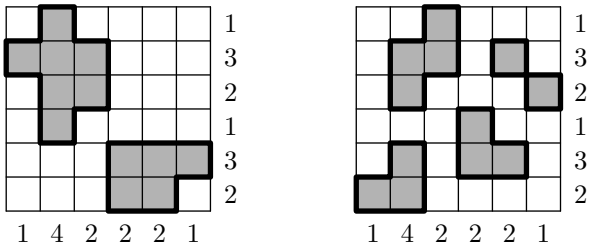
(b) The image is represented by the grey cells. The length of the boundary of this image is 62.

Figure 1.3: The same binary image represented in two different ways. The numbers indicate the row and column sums.

Recall from Section 1.3 the special shape of a uniquely determined set with monotone row and column sums. In every row and columns all the points with value one (or the black cells) are connected, so each row and each column with a nonzero line sum contributes 2 to the length of the boundary. So if there are m nonzero row sums and n nonzero column sums, then the total length of the boundary is $2m + 2n$. This is obviously the smallest possible length of the boundary of any set with the same number of nonzero row sums and nonzero column sums.

This minimum is not only attained for uniquely determined sets with monotone line sums. There are also other sets that have this property. In general a set with m

nonzero row sums, n nonzero column sums and a boundary of length $2m + 2n$ is called *hv-convex*. See Figure 1.4 for an example of an hv-convex set and another set (not hv-convex) that have the same line sums (so this hv-convex set is not uniquely determined). Deciding whether there exists an hv-convex reconstruction for given row and columns sums, is NP-complete [28] and hence it is also NP-complete to decide whether there exists a reconstruction with boundary length equal to $2m+2n$.



(a) This image is hv-convex. The length of the boundary is 24.

(b) This image is not hv-convex. The length of the boundary is 34.

Figure 1.4: Two binary images with the same line sums.

However, that does not mean that it is *always* hard to decide from the line sums whether the boundary can have length $2m + 2n$ or not. There exist arguments that can be used in part of the cases to prove easily that a boundary of length $2m + 2n$ is impossible. Suppose for example that we have 10 columns with nonzero sum, and that the first three row sums are (in that order) 10, 2 and 10. Then all columns have black cell rows 1 and 3, while only two columns have a black cell in row 2. Hence it is certain that there are at most two columns in which the black cells are connected. The other eight columns must contribute at least 4 each to the length of the boundary, so the length of the boundary must be at least $2m + 2 \cdot 2 + 8 \cdot 4$.

In Chapter 5 we generalise this principle to find a new lower bound on the length of the boundary, depending not only on m and n but on all row and column sums. In many cases our bound gives a better result than the straightforward lower bound $2m + 2n$.

In Chapter 6 we consider the complementary problem: given line sums, can you construct an image that satisfies these line sums and has relatively small boundary? Here we restrict ourselves to the case that the line sums are monotone. In this chapter α makes another appearance. Above we had already seen that when a set is uniquely determined by its line sums (that is equivalent with $\alpha = 0$) the length of the boundary is equal to $2m + 2n$. One of the main results of this chapter is a generalisation of this: when for the row and column sums we have $n = r_1 \geq r_2 \geq \ldots \geq r_m$ and $m = c_1 \geq c_2 \geq \ldots \geq c_n$, and the line sums are consistent, then there exists a reconstruction for which the length of the boundary is at most $2m+2n+4\alpha$.

1.7 Shape of binary images

In Chapter 7 we study the connection between the length of the boundary, the number of black cells, and the general shape of a binary image. Intuitively, it seems clear that when the number of black cells is large, but the boundary is small, the black cells must form some solid, roundish object. In this chapter, we will make this more precise.

Suppose we are given the length of the boundary and the number of black cells of an unknown binary image. We study the following question: what is the minimal size of the largest connected component in this image? Here we use *4-adjacency* [21] to define *connected*; that is, two cells are *adjacent* if they share an edge (and not just a vertex).

We can define the distance of a black cell to the boundary as follows: a black cell has distance 0 to the boundary if it is adjacent to a white cell, and it has distance $k+1$ to the boundary if k is the minimal distance to the boundary of the cells it is adjacent to. This distance function is also called the city block distance [23]. This leads to the second question we are interested in: what is the largest distance to the boundary that must occur in the image? A different way to phrase this: what is the minimal size of the largest ball of black cells that is contained in the image? We derive results about this question both in the case that the connected components are all simply connected (that is, they do not have any holes [21]) and in the general case.

Note that this chapter is only about properties of binary images, and discrete tomography plays no role here.

1.8 Overview

In **Chapter 2** we prove new stability results for the reconstruction of binary images from their horizontal and vertical projections. We consider an image that is uniquely determined by its projections and possible reconstructions from slightly different projections. We show that for a given difference in the projections, the reconstruction can only be disjoint from the original image if the size of the image is not too large. We also prove an upper bound on the size of the image given the error in the projections and the size of the intersection between the image and the reconstruction.

In **Chapter 3** we consider different reconstructions from the same horizontal and vertical projections. We present a condition that the projections must necessarily satisfy when there exist two disjoint reconstructions from those projections. More

generally, we derive an upper bound on the symmetric difference of two reconstructions from the same projections. We also consider two reconstructions from two different sets of projections and prove an upper bound on the symmetric difference in this case.

In **Chapter 4** we prove constructively that if there exists more than one reconstruction from given horizontal and vertical projections, then there exist two reconstructions that have a symmetric difference of at least $2\alpha + 2$. Here α is a parameter depending on the line sums and indicating how close (in some sense) the image is to being uniquely determined.

In **Chapter 5** we study the following question: for given horizontal and vertical projections, what is the smallest length of the boundary that a reconstruction from those projections can have? We prove a new lower bound that, in contrast to simple bounds that have been derived previously, combines the information of both row and column sums.

In **Chapter 6** we construct from given monotone row and column sums an image satisfying those line sums that has a small boundary. We prove several bounds on the length of this boundary, and we give a few examples for which we show that no smaller boundary is possible than the one of our construction.

In **Chapter 7** we consider an unknown binary image, of which the length of the boundary and the area of the image are given. We derive from this some properties about the general shape of the image. First, we prove sharp lower bounds on the size of the largest connected component. Second, we derive some results about the size of the largest ball containing only ones, both in the case that the connected components of the image are all simply connected and in the general case.

Each of the chapters can be read independently of the others. When results from earlier chapters are used, these are explicitly referred to. The notation will be defined separately for each chapter. Although the notation is fairly consistent throughout the thesis, there sometimes are subtle changes from one chapter to another.

CHAPTER 2

Stability results for uniquely determined sets

This chapter (with minor modifications) has been published as: Birgit van Dalen, "Stability results for uniquely determined sets from two directions in discrete tomography", Discrete Mathematics 309 (2009) 3905-3916.

2.1 Introduction

An interesting problem in discrete tomography is the stability of reconstructions. This concerns the following question: for a given binary image that is uniquely determined, can there exist a second image that is very different from the first one, but has almost the same line sums? For three or more directions, the answer is yes: there even exist two disjoint, arbitrarily large, uniquely determined images of which the line sums differ only very slightly [1, 3].

Here we focus on the same question, but with only two directions. Alpers et al. [1, 2] showed that in this case a total error of at most 2 in the projections can only cause a small difference in the reconstruction. They also obtained a lower bound on the error if the reconstruction is disjoint from the original image.

In this chapter we improve this bound, and we resolve the open problem of stability with a projection error greater than 2.

2.2 Notation and statement of the problems

Let F_1 and F_2 be two finite subsets of $\mathbb{Z}^2$ with characteristic functions χ_1 and χ_2. (That is, $\chi_h(x,y) = 1$ if and only if $(x,y) \in F_h$, $h \in \{1,2\}$.) For $i \in \mathbb{Z}$, we define *row* i as the set $\{(k,l) \in \mathbb{Z}^2 : k = i\}$. We call i the index of the row. For $j \in \mathbb{Z}$, we define *column* j as the set $\{(k,l) \in \mathbb{Z}^2 : l = j\}$. We call j the index of the column. Note that we follow matrix notation: we indicate a point (i,j) by first its row index i and then its column index j. Also, we use row numbers that increase when going downwards and column numbers that increase when going to the right.

The *row sum* $r_i^{(h)}$ is the number of elements of the set F_h in row i, that is $r_i^{(h)} = \sum_{j\in\mathbb{Z}} \chi_h(i,j)$. The *column sum* $c_j^{(h)}$ of F_h is the number of elements of F_h in column j, that is $c_j^{(h)} = \sum_{i\in\mathbb{Z}} \chi_h(i,j)$. We refer to both row and column sums as the *line sums* of F_h.

Throughout this chapter, we assume that F_1 is uniquely determined by its row and column sums. Such sets were studied by, among others, Ryser [24] and Wang [26]. Let a be the number of rows and b the number of columns that contain elements of F_1. We renumber the rows and columns such that we have

$$r_1^{(1)} \geq r_2^{(1)} \geq \ldots \geq r_a^{(1)} > 0,$$

$$c_1^{(1)} \geq c_2^{(1)} \geq \ldots \geq c_b^{(1)} > 0,$$

and such that all elements of F_2 are contained in rows and columns with positive indices. By [26, Theorem 2.3] we have the following property of F_1 (see Figure 2.1):

- in row i the elements of F_1 are precisely the points $(i,1)$, $(i,2)$, $\ldots$, $(i, r_i^{(1)})$,
- in column j the elements of F_1 are precisely the points $(1,j)$, $(2,j)$, $\ldots$, $(c_j^{(1)}, j)$.

We will refer to this property as the *triangular shape* of F_1.

Everywhere except in Section 2.6 we assume that $|F_1| = |F_2|$. Note that we do not assume F_2 to be uniquely determined.

As F_1 and F_2 are different and F_1 is uniquely determined by its line sums, F_2 cannot have exactly the same line sums as F_1. Define the *difference* or *error in the line sums* as

$$\sum_{j\geq 1} |c_j^{(1)} - c_j^{(2)}| + \sum_{i\geq 1} |r_i^{(1)} - r_i^{(2)}|.$$

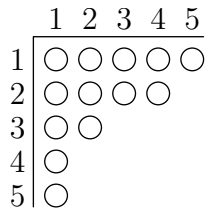


Figure 2.1: A uniquely determined set with the assumed row and column ordering.

As in general $|t-s| \equiv t+s \bmod 2$, the above expression is congruent to

$$\sum_{j\geq 1}\left(c_j^{(1)}+c_j^{(2)}\right)+\sum_{i\geq 1}\left(r_i^{(1)}+r_i^{(2)}\right) \equiv 2|F_1|+2|F_2| \equiv 0 \mod 2,$$

hence the error in the line sums is always even. We will denote it by 2α, where α is a positive integer.

For notational convenience, we will often write p for $|F_1 \cap F_2|$.

We consider two problems concerning stability.

Problem 2.1. *Suppose $F_1 \cap F_2 = \emptyset$. How large can $|F_1|$ be in terms of α?*

Alpers et al. [2, Theorem 29] proved that $|F_1| \leq \alpha^2$. They also showed that there is no constant c such that $|F_1| \leq c\alpha$ for all F_1 and F_2. In Section 2.4 we will prove the new bound $|F_1| \leq \alpha(1+\log\alpha)$ and show that this bound is asymptotically sharp.

Problem 2.2. *How small can $|F_1 \cap F_2|$ be in terms of $|F_1|$ and α, or, equivalently, how large can $|F_1|$ be in terms of $|F_1 \cap F_2|$ and α?*

Alpers ([1, Theorem 5.1.18]) showed in the case $\alpha = 1$ that

$$|F_1 \cap F_2| \geq |F_1| + \tfrac{1}{2} - \sqrt{2|F_1| + \tfrac{1}{4}}.$$

This bound is sharp: if $|F_1| = \frac{1}{2}n(n+1)$ for some positive integer n, then there exists an example for which equality holds. A similar result is stated in [2, Theorem 19].

While [1, 2] only deal with the case $\alpha = 1$, we will give stability results for general α. In Section 2.5 we will give two different upper bounds for $|F_1|$. The bounds have different asymptotic behaviour. Writing p for $|F_1 \cap F_2|$, the second bound (Theorem 2.8) reduces to

$$|F_1| \leq p + 1 + \sqrt{2p+1}$$

in case $\alpha = 1$, which is equivalent to

$$p \geq |F_1| - \sqrt{2|F_1|}.$$

Hence the second new bound can be viewed as a generalisation of Alpers' bound. The first new bound (Corollary 2.5) is different and better in the case that α is very large.

In Section 2.6 we will generalise the results to the case $|F_1| \neq |F_2|$.

2.3 Staircases

Alpers introduced the notion of a staircase to characterise $F_1 \triangle F_2$ in the case $\alpha = 1$. We will use a slightly different definition and then show that for general α the symmetric difference $F_1 \triangle F_2$ consists of α staircases.

Definition 2.1. *A set of points $(p_1, p_2, \ldots, p_n)$ in $\mathbb{Z}^2$ is called a* staircase *if the following two conditions are satisfied:*

- *for each i with $1 \leq i \leq n-1$ one of the points p_i and p_{i+1} is an element of $F_1 \backslash F_2$ and the other is an element of $F_2 \backslash F_1$;*
- *either for all i the points p_{2i} and p_{2i+1} are in the same column and the points p_{2i+1} and p_{2i+2} are in the same row, or for all i the points p_{2i} and p_{2i+1} are in the same row and the points p_{2i+1} and p_{2i+2} are in the same column.*

This definition is different from [1, 2] in the following way. Firstly, the number of points does not need to be even. Secondly, the points p_1 and p_n can both be either in $F_1 \backslash F_2$ or in $F_2 \backslash F_1$. So this definition is slightly more general than the one used in [1, 2] for the case $\alpha = 1$.

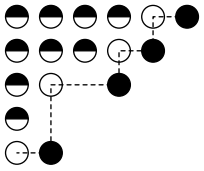


Figure 2.2: A staircase. The set F_1 consists of the white and the black-and-white points, while F_2 consists of the black and the black-and-white points. The staircase is indicated by the dashed line segments.

Consider a point $p_i \in F_1 \backslash F_2$ of a staircase $(p_1, p_2, \ldots, p_n)$. Assume p_{i-1} is in the same column as p_i and p_{i+1} is in the same row as p_i. Because of the triangular shape

of F_1, the row index of p_{i-1} must be larger than the row index of p_i, and the column index of p_{i+1} must be larger than the column index of p_i. Therefore, the staircase looks like a real-world staircase (see Figure 2.2). From now on, we assume for all staircases that p_1 is the point with the largest row index and the smallest column index, while p_n is the point with the smallest row index and the largest column index. We say that the staircase *begins* with p_1 and *ends* with p_n.

Lemma 2.1. *Let F_1 and F_2 be finite subsets of $\mathbb{Z}^2$ such that*

- *F_1 is uniquely determined by its row and column sums, and*
- *$|F_1| = |F_2|$.*

Let α be defined as in Section 2.2. Then the set $F_1 \triangle F_2$ is the disjoint union of α staircases.

Proof. We will construct the staircases one by one and delete them from $F_1 \triangle F_2$. For a subset A of $F_1 \triangle F_2$, define

$$\begin{aligned} \rho_i(A) &= |\{j \in \mathbb{Z} : (i,j) \in A \cap F_1\}| - |\{j \in \mathbb{Z} : (i,j) \in A \cap F_2\}|, \quad i \in \mathbb{Z}, \\ \sigma_j(A) &= |\{i \in \mathbb{Z} : (i,j) \in A \cap F_1\}| - |\{i \in \mathbb{Z} : (i,j) \in A \cap F_2\}|, \quad j \in \mathbb{Z}, \\ \tau(A) &= \sum_i |\rho_i(A)| + \sum_j |\sigma_j(A)|. \end{aligned}$$

We have $2\alpha = \tau(F_1 \triangle F_2)$.

Assume that the rows and columns are ordered as in Section 2.2. Because of the triangular shape of F_1, for any point $(i,j) \in F_1 \backslash F_2$ and any point $(k,l) \in F_2 \backslash F_1$ we then have $k > i$ or $l > j$.

Suppose we have deleted some staircases and are now left with a non-empty subset A of $F_1 \triangle F_2$. Let $(p_1, p_2, \ldots, p_n)$ be a staircase of maximal length that is contained in A. Let (x_1, y_1) and (x_n, y_n) be the coordinates of the points p_1 and p_n respectively. Each of those two points can be either in $A \cap F_1$ or in $A \cap F_2$, so there are four different cases. (If $n = 1$, so p_1 and p_n are the same point, then there are only two cases.) We consider two cases; the other two are similar.

First suppose $p_1 \in A \cap F_1$ and $p_n \in A \cap F_2$. If (x, y_1) is a point of $A \cap F_2$ in the same column as p_1, then $x > x_1$, so we can extend the staircase by adding this point. That contradicts the maximal length of the staircase. So there are no points of $A \cap F_2$ in column y_1. Therefore $\sigma_{y_1}(A) > 0$.

Similarly, since $p_n \in A \cap F_2$, there are no points of $A \cap F_1$ in the same column as p_n. Therefore $\sigma_{y_n}(A) < 0$.

All rows and all columns that contain points of the staircase, except columns y_1 and y_n, contain exactly two points of the staircase, one in $A \cap F_1$ and one in $A \cap F_2$. Let $A' = A \backslash \{p_1, p_2, \ldots, p_n\}$. Then $\rho_i(A') = \rho_i(A)$ for all i, and $\sigma_j(A') = \sigma_j(A)$ for all $j \neq y_1, y_n$. Furthermore, $\sigma_{y_1}(A') = \sigma_{y_1}(A) - 1$ and $\sigma_{y_n}(A') = \sigma_{y_n}(A) + 1$. Since $\sigma_{y_1}(A) > 0$ and $\sigma_{y_n}(A) < 0$, this gives $\tau(A') = \tau(A) - 2$.

Now consider the case $p_1 \in A \cap F_1$ and $p_n \in A \cap F_1$. As above, we have $\sigma_{y_1}(A) > 0$. Suppose (x_n, y) is a point of $A \cap F_2$ in the same row as p_n. Then $y > y_n$, so we can extend the staircase by adding this point. That contradicts the maximal length of the staircase. So there are no points of $A \cap F_2$ in row x_n. Therefore $\rho_{x_n}(A) > 0$.

All rows and all columns that contain points of the staircase, except column y_1 and row x_n, contain exactly two points of the staircase, one in $A \cap F_1$ and one in $A \cap F_2$. Let $A' = A \backslash \{p_1, p_2, \ldots, p_n\}$. Then $\rho_i(A') = \rho_i(A)$ for all $i \neq x_n$, and $\sigma_j(A') = \sigma_j(A)$ for all $j \neq y_1$. Furthermore, $\sigma_{y_1}(A') = \sigma_{y_1}(A) - 1$ and $\rho_{x_n}(A') = \rho_{x_n}(A) - 1$. Since $\sigma_{y_1}(A) > 0$ and $\rho_{x_n}(A) > 0$, this gives $\tau(A') = \tau(A) - 2$.

We can continue deleting staircases in this way until all points of $F_1 \triangle F_2$ have been deleted. Since $\tau(A) \geq 0$ for all subsets $A \subset F_1 \triangle F_2$, this must happen after deleting exactly α staircases. □

Remark 2.1. *Some remarks about the above lemma and its proof.*

(i) The α staircases from the previous lemma have 2α endpoints in total (where we count the same point twice in case of a staircase consisting of one point). Each endpoint contributes a difference of 1 to the line sums in one row or column. Since all these differences must add up to 2α, they cannot cancel each other.

(ii) A staircase consisting of more than one point can be split into two or more staircases. So it may be possible to write $F_1 \triangle F_2$ as the disjoint union of more than α staircases. However, in that case some of the contributions of the endpoints of staircases to the difference in the line sums cancel each other. On the other hand, it is impossible to decompose $F_1 \triangle F_2$ into fewer than α staircases.

(iii) The endpoints of a staircase can be in $F_1 \backslash F_2$ or $F_2 \backslash F_1$. For a staircase T of which the two endpoints are in different sets, we have $|T \cap F_1| = |T \cap F_2|$. For a staircase T of which the two endpoints are in the same set, we have $|T \cap F_1| = 1 + |T \cap F_2|$ or $|T \cap F_2| = 1 + |T \cap F_1|$. Since $|F_1 \backslash F_2| = |F_2 \backslash F_1|$, the number of staircases with two endpoints in $F_1 \backslash F_2$ must be equal to the number of staircases with two endpoints in $F_2 \backslash F_1$. This implies that of the 2α endpoints, exactly α are in the set $F_1 \backslash F_2$ and α are in the set $F_2 \backslash F_1$.

Consider a decomposition of $F_1 \triangle F_2$ as in the proof of Lemma 2.1. We will now show that for our purposes we may assume that all these staircases begin with a

point $p_1 \in F_1 \backslash F_2$ and end with a point $p_n \in F_2 \backslash F_1$.

Suppose there is a staircase beginning with a point $(x, y) \in F_2 \backslash F_1$. Then there also exists a staircase ending with a point $(x', y') \in F_1 \backslash F_2$: otherwise more than half of the 2α endpoints would be in $F_2 \backslash F_1$, which is a contradiction to Remark 2.1(iii). Because of Remark 2.1(i) we must have $r_x^{(1)} < r_x^{(2)}$ and $r_{x'}^{(1)} > r_{x'}^{(2)}$.

Let y'' be such that $(x', y'') \notin F_1 \cup F_2$. Delete the point (x, y) from F_2 and add the point (x', y'') to F_2. Then $r_x^{(2)}$ decreases by 1 and $r_{x'}^{(2)}$ increases by 1, so the difference in the row sums decreases by 2. Meanwhile, the difference in the column sums increases by at most 2. So α does not increase, while F_1, $|F_2|$ and $|F_1 \triangle F_2|$ do not change. So the new situation is just as good or better than the old one. The staircase that began with (x, y) in the old situation now begins with a point of $F_1 \backslash F_2$. The point that we added becomes the new endpoint of the staircase that previously ended with (x', y').

Therefore, in our investigations we may assume that all staircases begin with a point of $F_1 \backslash F_2$ and end with a point of $F_2 \backslash F_1$. This is an important assumption that we will use in the proofs throughout the chapter. An immediate consequence of it is that $r_i^{(1)} = r_i^{(2)}$ for all i. The only difference between corresponding line sums occurs in the columns.

2.4 A new bound for the disjoint case

Using the concept of staircases, we can prove a new bound for Problem 2.1.

Theorem 2.2. *Let F_1 and F_2 be finite subsets of $\mathbb{Z}^2$ such that*

- *F_1 is uniquely determined by its row and column sums,*
- *$|F_1| = |F_2|$, and*
- *$F_1 \cap F_2 = \emptyset$.*

Let α be defined as in Section 2.2. Then

$$|F_1| \leq \sum_{i=1}^{\alpha} \left\lfloor \frac{\alpha}{i} \right\rfloor .$$

Proof. Assume that the rows and columns are ordered as in Section 2.2. Let a be the number of rows and b the number of columns that contain elements of F_1. Let

$(k,l) \in F_1$. Then all the points in the rectangle $\{(i,j) : 1 \le i \le k, 1 \le j \le l\}$ are elements of F_1. Since F_1 and F_2 are disjoint, none of the points in this rectangle is an element of F_2, and all the points belong to $F_1 \triangle F_2$. So all of the kl points must belong to different staircases, which implies $\alpha \ge kl$. For all i with $1 \le i \le a$ we have $(i, r_i^{(1)}) \in F_1$, hence $r_i^{(1)} \le \frac{\alpha}{i}$. Since $r_i^{(1)}$ must be an integer, we have

$$|F_1| = \sum_{i=1}^{a} r_i^{(1)} \le \sum_{i=1}^{a} \left\lfloor \frac{\alpha}{i} \right\rfloor .$$

Since $(a,1) \in F_1$, we have $a \le \alpha$, so

$$|F_1| \le \sum_{i=1}^{\alpha} \left\lfloor \frac{\alpha}{i} \right\rfloor .$$

□

Corollary 2.3. *Let F_1, F_2 and α be defined as in Theorem 2.2. Then*

$$|F_1| \le \alpha(1 + \log \alpha).$$

Proof. We have

$$|F_1| \le \sum_{i=1}^{\alpha} \left\lfloor \frac{\alpha}{i} \right\rfloor \le \alpha \sum_{i=1}^{\alpha} \frac{1}{i} \le \alpha \left(1 + \int_1^{\alpha} \frac{1}{x} dx \right) = \alpha \left(1 + \log \alpha \right).$$

□

The following example shows that the upper bound cannot even be improved by a factor $\frac{1}{2 \log 2} \approx 0.72$.

Example 2.1. (taken from [1]) Let $m \ge 1$ be an integer. We construct sets F_1 and F_2 as follows (see also Figure 2.3).

- Row 1:
  - $(1,j) \in F_1$ for $1 \le j \le 2^m$,
  - $(1,j) \in F_2$ for $2^m + 1 \le j \le 2^{m+1}$.
- Let $0 \le l \le m-1$. Row i, where $2^l + 1 \le i \le 2^{l+1}$:
  - $(i,j) \in F_1$ for $1 \le j \le 2^{m-l-1}$,
  - $(i,j) \in F_2$ for $2^{m-l-1} + 1 \le j \le 2^{m-l}$.

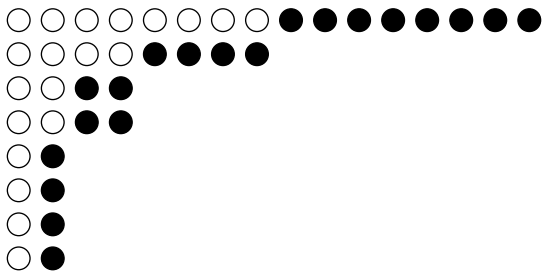


Figure 2.3: The construction from Example 2.1 with $m = 3$.

The construction is almost completely symmetrical: if $(i, j) \in F_1$, then $(j, i) \in F_1$; and if $(i, j) \in F_2$ with $i > 1$, then $(j, i) \in F_2$. Since it is clear from the construction that each row contains exactly as many points of F_1 as points of F_2, we conclude that each column j with $2 \leq j \leq 2^m$ contains exactly as many points of F_1 as points of F_2 as well. The only difference in the line sums occurs in the first column (which has 2^m points of F_1 and none of F_2) and in columns $2^m + 1$ up to 2^{m+1} (each of which contains one point of F_2 and none of F_1). So we have

$$\alpha = 2^m.$$

Furthermore,

$$|F_1| = 2^m + \sum_{l=0}^{m-1} 2^l 2^{m-l-1} = 2^m + m2^{m-1}.$$

Hence for this family of examples it holds that

$$|F_1| = \alpha + \frac{1}{2}\alpha \log_2 \alpha,$$

which is very close to the bound we proved in Corollary 2.3.

2.5 Two bounds for general α

In case F_1 and F_2 are not disjoint, we can use an approach very similar to Section 2.4 in order to derive a bound for Problem 2.2.

Theorem 2.4. *Let F_1 and F_2 be finite subsets of $\mathbb{Z}^2$ such that*

- *F_1 is uniquely determined by its row and column sums, and*
- *$|F_1| = |F_2|$.*

Let α be defined as in Section 2.2, and let $p = |F_1 \cap F_2|$. Then

$$|F_1| \leq \sum_{i=1}^{\alpha+p} \left\lfloor \frac{\alpha+p}{i} \right\rfloor .$$

Proof. Assume that the rows and columns are ordered as in Section 2.2. Let $(k,l) \in F_1$. Then all the points in the rectangle $\{(i,j) : 1 \leq i \leq k, 1 \leq j \leq l\}$ are elements of F_1. At most p of the points in this rectangle are elements of F_2, so at least $kl - p$ points belong to $F_1 \triangle F_2$. None of the points in the rectangle is an element of $F_2 \backslash F_1$, so all of the $kl - p$ points of $F_1 \triangle F_2$ in the rectangle must belong to different staircases, which implies $\alpha + p \geq kl$. For all i with $1 \leq i \leq a$ we have $(i, r_i^{(1)}) \in F_1$, hence $r_i^{(1)} \leq \frac{\alpha+p}{i}$. Since $r_i^{(1)}$ must be an integer, we have

$$|F_1| = \sum_{i=1}^{a} r_i^{(1)} \leq \sum_{i=1}^{a} \left\lfloor \frac{\alpha+p}{i} \right\rfloor .$$

Since $(a,1) \in F_1$, we have $a \leq \alpha + p$, so

$$|F_1| \leq \sum_{i=1}^{\alpha+p} \left\lfloor \frac{\alpha+p}{i} \right\rfloor .$$

□

Corollary 2.5. *Let F_1, F_2, α and p be defined as in Theorem 2.4. Then*

$$|F_1| \leq (\alpha+p)(1+\log(\alpha+p)).$$

Proof. Analogous to the proof of Corollary 2.3. □

The following example shows that the upper bound cannot even be improved by a factor $\frac{1}{2\log 2} \approx 0.72$, provided that $\alpha > \frac{p+1}{2\log 2 - 1} \log(p+1)$.

Example 2.2. Let k and m be integers satisfying $k \geq 2$ and $m \geq 2k-2$. We construct sets F_1 and F_2 as follows (see also Figures 2.4 and 2.5).

- Row 1:
  - $(1, j) \in F_1 \cap F_2$ for $1 \leq j \leq 2^{k-1}$,
  - $(1, j) \in F_1$ for $2^{k-1} + 1 \leq j \leq 2^m - 2^{k-1} + 1$,
  - $(1, j) \in F_2$ for $2^m - 2^{k-1} + 2 \leq j \leq 2^{m+1} - 2^k - 2^{k-1} + 2$.
- Let $0 \leq l \leq k - 2$. Row i, where $2^l + 1 \leq i \leq 2^{l+1}$:
  - $(i, 1) \in F_1 \cap F_2$,
  - $(i, j) \in F_1$ for $2 \leq j \leq 2^{m-l-1} - 2^{k-l-2} + 1$,
  - $(i, j) \in F_2$ for $2^{m-l-1} - 2^{k-l-2} + 2 \leq j \leq 2^{m-l} - 2^{k-l-1} + 1$.
- Let $k - 1 \leq l \leq m - k$. Row i, where $2^l + 1 \leq i \leq 2^{l+1}$:
  - $(i, j) \in F_1$ for $1 \leq j \leq 2^{m-l-1}$,
  - $(i, j) \in F_2$ for $2^{m-l-1} + 1 \leq j \leq 2^{m-l}$.
- Let $m - k + 1 \leq l \leq m - 1$. Row i, where $2^l - 2^{l-m+k-1} + 2 \leq i \leq 2^{l+1} - 2^{l-m+k} + 1$:
  - $(i, j) \in F_1$ for $1 \leq j \leq 2^{m-l-1}$,
  - $(i, j) \in F_2$ for $2^{m-l-1} + 1 \leq j \leq 2^{m-l}$.

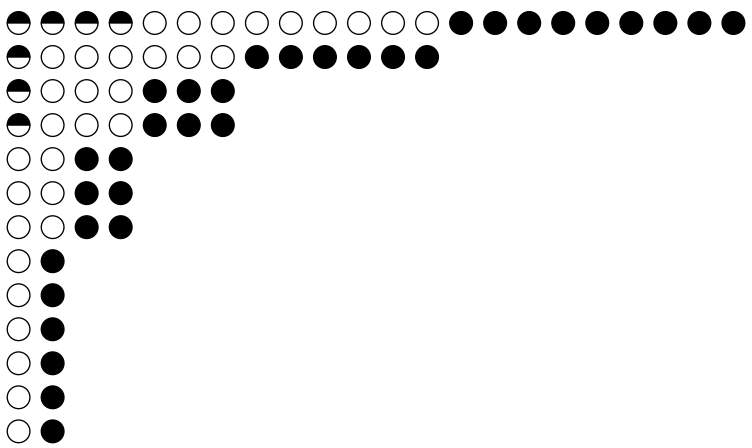


Figure 2.4: The construction from Example 2.2 with $k = 3$ and $m = 4$.

The construction is almost symmetrical: if $(i, j) \in F_1$, then $(j, i) \in F_1$; if $(i, j) \in F_1 \cap F_2$, then $(j, i) \in F_1 \cap F_2$; and if $(i, j) \in F_2$ with $i > 1$, then $(j, i) \in F_2$. Since it is clear from the construction that each row contains exactly as many points of F_1 as points of F_2, we conclude that each column j with $2 \leq j \leq 2^m - 2^{k-1} + 1$ contains exactly as many points of F_1 as points of F_2 as well. The only difference in

the line sums occurs in the first column (which has $2^m - 2^{k-1} + 1$ points of F_1 and only 2^{k-1} of F_2) and in columns $2^m - 2^{k-1} + 2$ up to $2^{m+1} - 2^k - 2^{k-1} + 2$ (each of which contains one point of F_2 and none of F_1). So we have

$$\begin{aligned} \alpha &= \frac{1}{2}\left((2^m - 2^{k-1} + 1) - 2^{k-1} + (2^{m+1} - 2^k - 2^{k-1} + 2) - (2^m - 2^{k-1} + 1)\right) \\ &= 2^m - 2^k + 1. \end{aligned}$$

It is easy to see that

$$p = |F_1 \cap F_2| = 2^k - 1.$$

Now we count the number of elements of F_1.

- Row 1 contains $2^m - 2^{k-1} + 1$ elements of F_1.
- Let $0 \leq l \leq k-2$. Rows $2^l + 1$ up to 2^{l+1} together contain $2^l(2^{m-l-1} - 2^{k-l-2} + 1) = 2^{m-1} - 2^{k-2} + 2^l$ elements of F_1.
- Let $k-1 \leq l \leq m-k$. Rows $2^l + 1$ up to 2^{l+1} together contain $2^l \cdot 2^{m-l-1} = 2^{m-1}$ elements of F_1.
- Let $m - k + 1 \leq l \leq m - 1$. Rows $2^l - 2^{l-m+k-1} + 2$ up to $2^{l+1} - 2^{l-m+k} + 1$ together contain $(2^l - 2^{l-m+k-1})(2^{m-l-1}) = 2^{m-1} - 2^{k-2}$ elements of F_1.

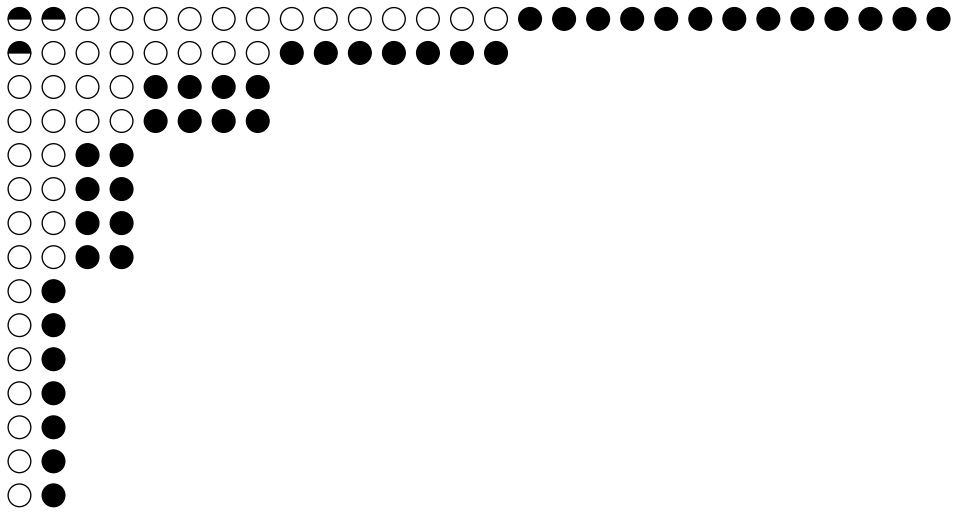


Figure 2.5: The construction from Example 2.2 with $k = 2$ and $m = 4$.

Hence the number of elements of F_1 is

$$\begin{aligned} |F_1| &= 2^m - 2^{k-1} + 1 + (k-1)(2^{m-1} - 2^{k-2}) + \sum_{l=0}^{k-2} 2^l \\ &\quad + (m-2k+2)2^{m-1} + (k-1)(2^{m-1} - 2^{k-2}) \\ &= 2^m + m2^{m-1} + 2^{k-1} - k2^{k-1}. \end{aligned}$$

For this family of examples we now have

$$|F_1| = \alpha + p + \frac{\alpha + p}{2} \log_2(\alpha + p) + \frac{p+1}{2} - \frac{p+1}{2} \log_2(p+1).$$

We will now prove another bound, which is better if $p = |F_1 \cap F_2|$ is large compared to α. Let u be an integer such that $2u = |F_1 \triangle F_2|$. We will first derive an upper bound on u in terms of a, b and α. Then we will derive a lower bound on $|F_1|$ in terms of a, b and α. By combining these two, we find an upper bound on u in terms of α and p.

Lemma 2.6. *Let F_1 and F_2 be finite subsets of $\mathbb{Z}^2$ such that*

- *F_1 is uniquely determined by its row and column sums, and*
- *$|F_1| = |F_2|$.*

Let α, a and b be defined as in Section 2.2. Define u as $2u = |F_1 \triangle F_2|$. Then we have

$$u^2 \leq \frac{\alpha}{4}(a+b)(a+b+\alpha-1).$$

Proof. Decompose $F_1 \triangle F_2$ into α staircases as in Lemma 2.1, and let $\mathcal{T}$ be the set consisting of these staircases. Let $T \in \mathcal{T}$ be a staircase and $i \leq a+1$ a positive integer. Consider the elements of $T \cap F_2$ in rows i, $i+1$, ..., a. If such elements exist, then let $w_i(T)$ be the largest column index that occurs among these elements. If there are no elements of $T \cap F_2$ in those rows, then let $w_i(T)$ be equal to the smallest column index of an element of $T \cap F_1$ (no longer restricted to rows i, ..., a). We have $w_i(T) \geq 1$. Define $W_i = \sum_{T \in \mathcal{T}} w_i(T)$.

Let d_i be the number of elements of $F_1 \backslash F_2$ in row i. Let $y_1 < \ldots < y_{d_i}$ be the column indices of the elements of $F_1 \backslash F_2$ in row i, and let $y'_1 < \ldots < y'_{d_i}$ be the column indices of the elements of $F_2 \backslash F_1$ in row i. Let $\mathcal{T}_i \subset \mathcal{T}$ be the set of staircases with elements in row i. The elements in $F_2 \backslash F_1$ of these staircases are in columns y'_1, y'_2, ..., y'_{d_i}, hence the set $\{w_i(T) : T \in \mathcal{T}_i\}$ is equal to the set $\{y'_1, y'_2, \ldots, y'_{d_i}\}$. The elements in $F_1 \backslash F_2$ are in columns y_1, y_2, ..., y_d and are either the first element of

a staircase or correspond to an element of $F_2 \backslash F_1$ in the same column but in a row with index at least $i+1$. In either case, for a staircase $T \in \mathcal{T}_i$ we have $w_{i+1}(T) = y_j$ for some j. Hence the set $\{w_{i+1}(T) : T \in \mathcal{T}_i\}$ is equal to the set $\{y_1, y_2, \ldots, y_{d_i}\}$. We have

$$\sum_{T \in \mathcal{T}_i} w_{i+1}(T) = \sum_{j=1}^{d_i} y_j \leq \sum_{j=1}^{d_i} (y_{d_i} - j + 1) = d_i y_{d_i} - \frac{1}{2}(d_i - 1)d_i,$$

and

$$\sum_{T \in \mathcal{T}_i} w_i(T) = \sum_{j=1}^{d_i} y'_j \geq \sum_{j=1}^{d_i} (y_{d_i} + j) = d_i y_{d_i} + \frac{1}{2}(d_i + 1)d_i.$$

Hence

$$\begin{aligned} W_i &= W_{i+1} + \sum_{T \in \mathcal{T}_i} (w_i(T) - w_{i+1}(T)) \\ &\geq W_{i+1} + \frac{1}{2}(d_i + 1)d_i + \frac{1}{2}(d_i - 1)d_i \\ &= W_{i+1} + d_i^2. \end{aligned}$$

Since $W_{a+1} \geq \alpha$, we find

$$W_1 \geq \alpha + d_1^2 + \cdots + d_a^2.$$

We may assume that if (x, y) is the endpoint of a staircase, then (x, y') is an element of $F_1 \cup F_2$ for $1 \leq y' < y$ (i.e. there are no gaps between the endpoints and other elements of $F_1 \cup F_2$ on the same row). After all, by moving the endpoint of a staircase to another empty position on the same row, the error in the columns can only become smaller (if the new position of the endpoint happens to be in the same column as the first point of another staircase, in which case the two staircases fuse together to one) but not larger, and u, a and b do not change.

So on the other hand, as W_1 is the sum of the column indices of the endpoints of the staircases, we have

$$W_1 \leq (b+1) + (b+2) + \cdots + (b+\alpha) = \alpha b + \frac{1}{2}\alpha(\alpha+1).$$

We conclude

$$\alpha + \sum_{i=1}^{a} d_i^2 \leq \alpha b + \frac{1}{2}\alpha(\alpha+1).$$

Note that $\sum_{i=1}^{a} d_i = u$. By the Cauchy-Schwarz inequality, we have

$$\left(\sum_{i=1}^{a} d_i^2\right)\left(\sum_{i=1}^{a} 1\right) \geq \left(\sum_{i=1}^{a} d_i\right)^2 = u^2,$$

so

$$\sum_{i=1}^{a} d_i^2 \geq \frac{u^2}{a}.$$

From this it follows that

$$\alpha b + \frac{1}{2}\alpha(\alpha+1) \geq \alpha + \frac{u^2}{a},$$

or, equivalently,

$$u^2 \leq \alpha ab + \frac{1}{2}\alpha(\alpha-1)a.$$

By symmetry we also have

$$u^2 \leq \alpha ab + \frac{1}{2}\alpha(\alpha-1)b.$$

Hence

$$u^2 \leq \alpha ab + \frac{1}{4}\alpha(\alpha-1)(a+b).$$

Using that $\sqrt{ab} \leq \frac{a+b}{2}$, we find

$$u^2 \leq \alpha\left(\frac{(a+b)^2}{4} + \frac{(\alpha-1)(a+b)}{4}\right) = \frac{\alpha}{4}(a+b)(a+b+\alpha-1).$$

□

Lemma 2.7. *Let F_1 and F_2 be finite subsets of $\mathbb{Z}^2$ such that*

- *F_1 is uniquely determined by its row and column sums, and*
- *$|F_1| = |F_2|$.*

Let α, a and b be defined as in Section 2.2. Then we have

$$|F_1| \geq \frac{(a+b)^2}{4(\alpha+1)}.$$

Proof. Without loss of generality, we may assume that all rows and columns that contain elements of F_1 also contain at least one point $F_1 \triangle F_2$: if a row or column does not contain any points of $F_1 \triangle F_2$, we may delete it. By doing so, $F_1 \triangle F_2$ does not change, while $|F_1|$ becomes smaller, so the situation becomes better.

First consider the case $r_{i+1}^{(1)} < r_i^{(1)} - \alpha$ for some i. We will show that this is impossible. If a column does not contain an element of $F_2 \backslash F_1$, then by the assumption above it contains an element of $F_1 \backslash F_2$, which must then be the first point of a staircase.

Consider all points of $F_2 \backslash F_1$ and all first points of staircases in columns $r_{i+1}+1$, $r_{i+1}+2$, …, r_i. Since these are more than α columns, at least two of those points must belong to the same staircase. On the other hand, if $(x, y) \in F_1 \backslash F_2$ is the first point of a staircase with $r_{i+1} < y \leq r_i$, then we have $x \leq i$, so the second point (x', y') in the staircase, which is in $F_2 \backslash F_1$, must satisfy $x' \leq i$ and therefore $y' > r_i$. So the second point cannot also be in one of the columns $r_{i+1}+1$, $r_{i+1}+2$, …, r_i. If two points of $F_2 \backslash F_1$ in columns $r_{i+1}+1$, $r_{i+1}+2$, …, r_i belong to the same staircase, then they must be connected by a point of $F_1 \backslash F_2$ in the same columns. However, by a similar argument this forces the next point to be outside the mentioned columns, while we assumed that it was in those columns. We conclude that it is impossible for row sums of two consecutive rows to differ by more than α.

By the same argument, column sums of two consecutive columns cannot differ by more than α. Hence we have $r_{i+1}^{(1)} \geq r_i^{(1)} - \alpha$ for all i, and $c_{j+1}^{(1)} \geq c_j^{(1)} - \alpha$ for all j.

We now have $r_2^{(1)} \geq b - \alpha$, $r_3^{(1)} \geq b - 2\alpha$, and so on. Also, $c_2^{(1)} \geq a - \alpha$, $c_3^{(1)} \geq a - 2\alpha$, and so on. Using this, we can derive a lower bound on $|F_1|$ for fixed a and b. Consider Figure 2.6. The points of F_1 are indicated by black dots. The number of points is equal to the grey area in the picture, which consists of all 1×1-squares with a point of F_1 in the upper left corner. We can estimate this area from below by drawing a line with slope α through the point $(a+1, 1)$ and a line with slope $\frac{1}{\alpha}$ through the point $(b+1, 1)$; the area closed in by these two lines and the two axes is less than or equal to the number of points of F_1.

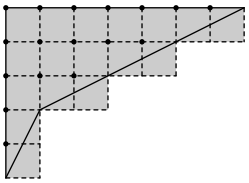


Figure 2.6: The number of points of F_1 (indicated by small black dots) is equal to the grey area.

For $\alpha = 1$ those lines do not have a point of intersection. Under the assumption we made at the beginning of this proof, we must in this case have $a = b$ and the number of points of F_1 is equal to

$$\frac{a(a+1)}{2} \geq \frac{a^2}{\alpha+1} = \frac{(a+b)^2}{4(\alpha+1)},$$

so in this case we are done.

In order to compute the area for $\alpha \geq 2$ we switch to the usual coordinates in $\mathbb{R}^2$, see Figure 2.7. The equation of the first line is $y = \alpha x - a$, and the equation of the

second line is $y = \frac{1}{\alpha}x - \frac{1}{\alpha}b$. We find that the point of intersection is given by

$$(x, y) = \left(\frac{a\alpha - b}{\alpha^2 - 1}, \frac{-b\alpha + a}{\alpha^2 - 1} \right).$$

The area of the grey part of Figure 2.7 is equal to

$$\frac{1}{2}a \cdot \frac{a\alpha - b}{\alpha^2 - 1} + \frac{1}{2}b \cdot \frac{b\alpha - a}{\alpha^2 - 1} = \frac{a^2\alpha + b^2\alpha - 2ab}{2(\alpha^2 - 1)}.$$

We now have

$$|F_1| \geq \frac{\alpha(a^2 + b^2) - 2ab}{2(\alpha^2 - 1)} \geq \frac{\alpha\frac{(a+b)^2}{2} - \frac{(a+b)^2}{2}}{2(\alpha^2 - 1)} = \frac{(a+b)^2}{4(\alpha + 1)}.$$

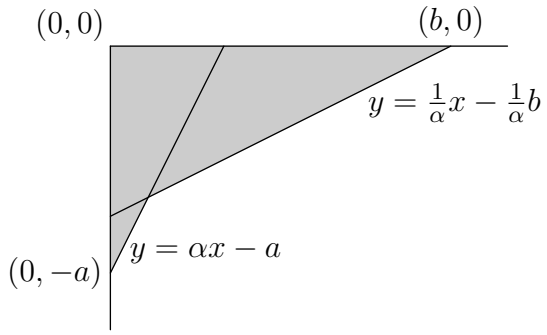


Figure 2.7: Computing the area bounded by the two lines and the two axes.

□

Theorem 2.8. *Let F_1 and F_2 be finite subsets of $\mathbb{Z}^2$ such that*

- *F_1 is uniquely determined by its row and column sums, and*
- *$|F_1| = |F_2|$.*

Let α be defined as in Section 2.2, and let $p = |F_1 \cap F_2|$. Write $\beta = \sqrt{\alpha}(\alpha+1)$. Then

$$|F_1| \leq p + \sqrt{\frac{\alpha}{4}\left(\beta + \sqrt{\beta(\alpha - 1) + 4(\alpha + 1)p + \beta^2} + \frac{\alpha - 1}{2}\right)^2 - \frac{(\alpha - 1)^2\alpha}{16}}.$$

Proof. Write $s = a + b$ for convenience of notation. From Lemma 2.6 we derive

$$u \leq \frac{\sqrt{\alpha}}{2}\left(s + \frac{\alpha - 1}{2}\right).$$

We substitute $|F_1| = u + p$ in Lemma 2.7 and use the above bound for u:

$$\frac{\sqrt{\alpha}}{2}\left(s + \frac{\alpha - 1}{2}\right) + p \geq |F_1| \geq \frac{s^2}{4(\alpha + 1)}.$$

Solving for s, we find

$$\begin{aligned} s &\leq \sqrt{\alpha}(\alpha + 1) + \sqrt{\sqrt{\alpha}(\alpha^2 - 1) + 4(\alpha + 1)p + \alpha(\alpha + 1)^2} \\ &= \beta + \sqrt{\beta(\alpha + 1) + 4(\alpha + 1)p + \beta^2} \end{aligned}$$

Finally we substitute this in Lemma 2.6:

$$u \leq \sqrt{\frac{\alpha}{4}\left(\beta + \sqrt{\beta(\alpha - 1) + 4(\alpha + 1)p + \beta^2} + \frac{\alpha - 1}{2}\right)^2 - \frac{(\alpha - 1)^2\alpha}{16}}.$$

This, together with $|F_1| = u + p$, yields the claimed result. □

Remark 2.2. *By a straightforward generalisation of [2, Proposition 13 and Lemma 16], we find a bound very similar to the one in Theorem 2.8:*

$$|F_1| \leq p + (\alpha + 1)(\alpha - \frac{1}{2}) + (\alpha + 1)\sqrt{2p + \frac{(2\alpha - 1)^2}{4}}.$$

Theorem 2.8 says that $|F_1|$ is asymptotically bounded by $p + \alpha\sqrt{p} + \alpha^2$. The next example shows that $|F_1|$ can be asymptotically as large as $p + 2\sqrt{\alpha p} + \alpha$.

Example 2.3. Let N be a positive integer. We construct F_1 and F_2 with total difference in the line sums equal to 2α as follows (see also Figure 2.8). Let $(i, j) \in F_1 \cap F_2$ for $1 \leq i \leq N$, $1 \leq j \leq N$. Furthermore, for $1 \leq i \leq N$:

- Let $(i, j), (j, i) \in F_1 \cap F_2$ for $N + 1 \leq j \leq N + (N - i)\alpha$.
- Let $(i, j), (j, i) \in F_1$ for $N + (N - i)\alpha + 1 \leq j \leq N + (N - i + 1)\alpha$.
- Let $(i, j), (j, i) \in F_2$ for $N + (N - i + 1)\alpha + 1 \leq j \leq N + (N - i + 2)\alpha$.

Finally, for $1 \leq t \leq \alpha$, let $(i, j) \in F_2$ with $i = N + t$ and $j = N + \alpha + 1 - t$.

The only differences in the line sums occur in the first column (a difference of α) and in columns $N + N\alpha + 1$ up to $N + N\alpha + \alpha$ (a difference of 1 in each column). We have

$$p = N^2 + 2 \cdot \frac{1}{2}N(N - 1)\alpha = N^2 + N^2\alpha - N\alpha,$$

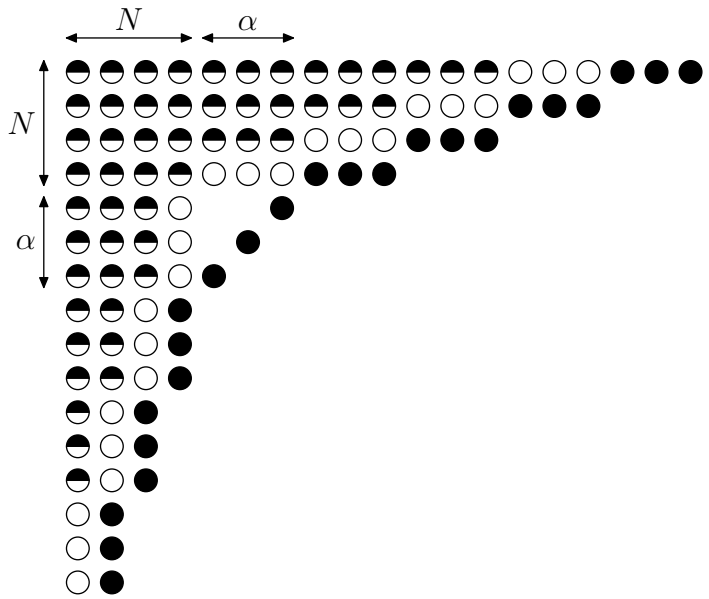


Figure 2.8: The construction from Example 2.3 with $N = 4$ and $\alpha = 3$.

and

$$|F_1| = N^2 + 2 \cdot \frac{1}{2} N(N+1)\alpha = N^2 + N^2\alpha + N\alpha.$$

From the first equality we derive

$$N = \frac{\alpha}{2(\alpha+1)} + \sqrt{\frac{p}{\alpha+1} + \frac{\alpha^2}{4(\alpha+1)^2}}.$$

Hence

$$|F_1| = p + 2N\alpha = p + \frac{\alpha^2}{\alpha+1} + \sqrt{\frac{4\alpha^2 p}{\alpha+1} + \frac{\alpha^4}{(\alpha+1)^2}}.$$

2.6 Generalisation to unequal sizes

Until now, we have assumed that $|F_1| = |F_2|$. However, we can easily generalise all the results to the case $|F_1| \neq |F_2|$.

Suppose $|F_1| > |F_2|$. Then there must be a row i with $r_i^{(1)} > r_i^{(2)}$. Let $j > b$ be such that $(i, j) \notin F_2$ and define $F_3 = F_2 \cup \{(i, j)\}$. We have $r_i^{(3)} = r_i^{(2)} + 1$, so the error

in row i has decreased by one, while the error in column j has increased by one. In this way, we can keep adding points until F_2 together with the extra points is just as large as F_1, while the total difference in the line sums is still 2α. Note that $p = |F_1 \cap F_2|$ and $|F_1|$ have not changed during this process, so the results from Theorem 2.8 and Corollary 2.5 are still valid in exactly the same form.

Suppose on the other hand that $|F_1| < |F_2|$. Then there must be a row with $r_i^{(1)} < r_i^{(2)}$. Let j be such that $(i,j) \in F_2 \backslash F_1$ and define $F_3 = F_2 \backslash \{(i,j)\}$. The error in row i has now decreased by one, while the error in column j has at most increased by one, so the total error in the line sums has not increased. We can keep deleting points of F_2 until there are exactly $|F_1|$ points left, while the total difference in the line sums is at most 2α.

By using $|F_1 \triangle F_2| = 2(|F_1| - p)$, we can state the results from Theorem 2.8 and Corollary 2.5 in a more symmetric way, not depending on the size of F_1.

Theorem 2.9. *Let F_1 and F_2 be finite subsets of $\mathbb{Z}^2$ such that F_1 is uniquely determined by its row and column sums. Let α be defined as in Section 2.2, and let $p = |F_1 \cap F_2|$. Write $\beta = \sqrt{\alpha}(\alpha + 1)$. Then*

1. $|F_1 \triangle F_2| \leq 2\alpha + 2(\alpha + p)\log(\alpha + p)$.
2. $|F_1 \triangle F_2| \leq \sqrt{\alpha\left(\beta + \sqrt{\beta(\alpha-1) + 4(\alpha+1)p + \beta^2} + \frac{\alpha-1}{2}\right)^2 - \frac{(\alpha-1)^2\alpha}{4}}$.

CHAPTER 3

Upper bounds for the difference between reconstructions

This chapter (with minor modifications) has been published as: Birgit van Dalen, "On the difference between solutions of discrete tomography problems", Journal of Combinatorics and Number Theory 1 (2009) 15-29.

3.1 Introduction

When a binary image is not uniquely determined by its projections, the reconstruction may not be equal to the original image. In such a situation, it is interesting to know whether the reconstruction is a good approximation of the original image. In other words, we would like to find bounds on how much two images with the same projections can differ, and to have conditions under which the two images can be completely disjoint.

There exists a very simple such bound. If the image is contained in an $m \times n$-rectangle and a certain row sum is equal to $a \geq n/2$, then the difference in that row can be at most $2a - n$. If on the other hand a row sum is equal to $b < n/2$, then the difference in the row can be at most $2b$. Summing over all m rows gives an upper bound on the size of the symmetric difference of two different reconstructions. While this bound may be quite good in some special cases, it is not very good in general.

In this chapter we will use a different approach, based on the work in Chapter 2. There the concept of staircases, introduced by Alpers [1], was used to compare an arbitrary image to a uniquely determined image. Here we generalise this method in order to compare two arbitrary binary images. We use a uniquely determined image that is as close as possible to the original image. We characterise such images in Theorem 3.4. We then consider two reconstructions from the same horizontal and vertical projections and prove bounds on the intersection and symmetric difference of the two reconstructions in Theorems 3.5 and 3.7. As a consequence of these results, we find a condition on the projections that must hold when the reconstruction and the original image are disjoint.

In Theorem 3.6 we show that we can construct a uniquely determined image that is guaranteed to have a large intersection with the original image. To complement this result, we state conditions under which no individual point must necessarily belong to the original image (these conditions are a direct consequence of a theorem by Anstee [4]). Finally, we will consider two reconstructions from two different sets of horizontal and vertical projections and prove an upper bound for the difference between the two reconstructions.

3.2 Notation

Let F be a finite subset of $\mathbb{Z}^2$ with characteristic function χ. (That is, $\chi(k,l) = 1$ if $(k,l) \in F$ and $\chi(k,l) = 0$ otherwise.) For $i \in \mathbb{Z}$, we define *row* i as the set $\{(k,l) \in \mathbb{Z}^2 : k = i\}$. We call i the index of the row. For $j \in \mathbb{Z}$, we define *column* j as the set $\{(k,l) \in \mathbb{Z}^2 : l = j\}$. We call j the index of the column. Note that we follow matrix notation: we indicate a point (i,j) by first its row index i and then its column index j. Also, we use row numbers that increase when going downwards and column numbers that increase when going to the right.

The *row sum* r_i is the number of elements of F in row i, that is $r_i = \sum_{j\in\mathbb{Z}} \chi(i,j)$. The *column sum* c_j of F is the number of elements of F in column j, that is $c_j = \sum_{i\in\mathbb{Z}} \chi(i,j)$. We refer to both row and column sums as the *line sums* of F. We will usually only consider finite sequences $(r_1, r_2, \ldots, r_m)$ and $(c_1, c_2, \ldots, c_n)$ of row and column sums that contain all the nonzero line sums.

We call F *uniquely determined by its line sums* or simply *uniquely determined* if the following property holds: if F' is a subset of $\mathbb{Z}^2$ with exactly the same row and column sums as F, then $F' = F$. Suppose F is uniquely determined and has row sums r_1, r_2, …, r_m. For each j with $1 \le j \le \max_i r_i$ we can count the number $\#\{l : r_l \ge j\}$ of row sums that are at least j. These numbers are exactly the nonzero column sums of F (in some order). This is an immediate consequence of Ryser's theorem ([24], see also [14, Theorem 1.7]).

Suppose we have two finite subsets F_1 and F_2 of $\mathbb{Z}^2$. For $h = 1, 2$ we denote the row and column sums of F_h by $r_i^{(h)}$, $i \in \mathbb{Z}$, and $c_j^{(h)}$, $j \in \mathbb{Z}$, respectively. Define

$$\alpha(F_1, F_2) = \frac{1}{2}\left(\sum_{j\in\mathbb{Z}} |c_j^{(1)} - c_j^{(2)}| + \sum_{i\in\mathbb{Z}} |r_i^{(1)} - r_i^{(2)}|\right).$$

Note that $\alpha(F_1, F_2)$ is an integer, since $2\alpha(F_1, F_2)$ is congruent to

$$\sum_{j\in\mathbb{Z}} \left(c_j^{(1)} + c_j^{(2)}\right) + \sum_{i\in\mathbb{Z}} \left(r_i^{(1)} + r_i^{(2)}\right) = 2|F_1| + 2|F_2| \equiv 0 \mod 2.$$

We will sometimes refer to $\sum_{j\in\mathbb{Z}} |c_j^{(1)} - c_j^{(2)}|$ as the *difference in the column sums* and to $\sum_{i\in\mathbb{Z}} |r_i^{(1)} - r_i^{(2)}|$ as the *difference in the row sums.*

In order to describe the symmetric difference between two sets F_1 and F_2, we use the notion of a staircase, first introduced by Alpers [1].

Definition 3.1. *A set of points $(p_1, p_2, \ldots, p_n)$ in $\mathbb{Z}^2$ is called a* staircase *if the following two conditions are satisfied:*

- *for each i with $1 \le i \le n-1$ one of the points p_i and p_{i+1} is an element of $F_1 \backslash F_2$ and the other is an element of $F_2 \backslash F_1$;*
- *either for all i the points p_{2i} and p_{2i+1} are in the same column and the points p_{2i+1} and p_{2i+2} are in the same row, or for all i the points p_{2i} and p_{2i+1} are in the same row and the points p_{2i+1} and p_{2i+2} are in the same column.*

3.3 Some lemmas

We prove some lemmas that we will use later for our main results.

Lemma 3.1. *Let $a_1 \ge a_2 \ge \ldots \ge a_n$ be non-negative integers. Let $m \ge \max_j a_j$. For $1 \le i \le m$, define $b_i = \#\{j : a_j \ge i\}$. Then for $1 \le j \le n$ we have $a_j = \#\{i : b_i \ge j\}$.*

Proof. We have $b_1 \ge b_2 \ge \ldots \ge b_m$. Hence for $1 \le j \le n$ we have

$$\#\{i : b_i \ge j\} = \max\{i : b_i \ge j\} = \max\{i : \max\{l : a_l \ge i\} \ge j\}.$$

For a fixed i we have

$$\max\{l : a_l \ge i\} \ge j \quad \Longleftrightarrow \quad a_j \ge i,$$

hence

$$\max\{i : \max\{l : a_l \geq i\} \geq j\} = \max\{i : a_j \geq i\} = a_j.$$

This completes the proof. □

Lemma 3.2. *Let F be a uniquely determined finite subset of $\mathbb{Z}^2$ with row sums r_i, $i \in \mathbb{Z}$, and column sums c_j, $j \in \mathbb{Z}$, respectively. If for integers i_1, i_2 and j_0 we have $(i_1, j_0) \in F$ and $(i_2, j_0) \notin F$, then $r_{i_1} > r_{i_2}$.*

Proof. As F is uniquely determined, we have the following characterisation of its elements [14, p. 17]: a point (x, y) is an element of F if and only if $r_x \geq \#\{l : c_l \geq c_y\}$. So if $(i_1, j_0) \in F$ and $(i_2, j_0) \notin F$, we have $r_{i_1} \geq \#\{l : c_l \geq c_{j_0}\} > r_{i_2}$. □

Let F_1 and F_2 be finite subsets of $\mathbb{Z}^2$, such that F_1 is uniquely determined and $|F_1| = |F_2|$. Denote the row sums of F_1 by r_i, $i \in \mathbb{Z}$. Let $\alpha = \alpha(F_1, F_2)$. The symmetric difference $F_1 \triangle F_2$ is the disjoint union of α staircases (see Lemma 2.1). Consider such a staircase with points $(x_1, y_1), (x_2, y_2), \ldots, (x_t, y_t) \in F_1 \backslash F_2$ and $(x_2, y_1), (x_3, y_2) \ldots, (x_t, y_{t-1}) \in F_2 \backslash F_1$. (The staircase may contain another point of $F_2 \backslash F_1$ in row x_1 and another one in column y_t, but this is irrelevant here.) By Lemma 3.2 we have

$$r_{x_1} > r_{x_2} > \ldots > r_{x_t}.$$

Hence the rows x_1, x_2, …, x_t of F_1 have pairwise different line sums.

Lemma 3.3. *We have*

$$|F_1 \triangle F_2| \leq \alpha\sqrt{8|F_1| + 1} - \alpha.$$

Proof. Let n be the largest positive integer such that $|F_1| \geq n(n+1)/2$. Suppose F_1 has at least $n + 1$ distinct positive row sums. Then

$$|F_1| \geq 1 + 2 + \cdots + n + (n + 1) = \frac{1}{2}(n + 1)(n + 2),$$

which contradicts the maximality of n. So F_1 has at most n distinct positive row sums. Any staircase of $F_1 \triangle F_2$ therefore contains elements of $F_1 \backslash F_2$ in at most n different rows. So the total number of elements of $F_1 \backslash F_2$ cannot exceed αn. Hence $|F_1 \triangle F_2| \leq 2\alpha n$. On the other hand, we have $2|F_1| \geq n^2 + n = (n + 1/2)^2 - 1/4$, thus $n \leq \sqrt{2|F_1| + 1/4} - 1/2$. We conclude

$$|F_1 \triangle F_2| \leq \alpha\sqrt{8|F_1| + 1} - \alpha.$$

□

Remark 3.1. *We will also use the slightly weaker estimate*

$$|F_1 \triangle F_2| \leq 2\alpha\sqrt{2|F_1|}.$$

3.4 Uniquely determined neighbours

Consider a set F_2 that is not uniquely determined by its line sums. We are interested in how close – in some sense – this set is to being uniquely determined. We define the distance between F_2 and a uniquely determined set F_1 as $\alpha(F_2, F_1)$. The smallest possible value of $\alpha(F_2, F_1)$ then indicates how close F_2 is to being uniquely determined. It turns out that we can characterise in a very simple way the sets F_1 for which $\alpha(F_2, F_1)$ is minimal.

Theorem 3.4. *Let F_2 be a finite subset of $\mathbb{Z}^2$ with nonzero row sums $r_1 \geq r_2 \geq \ldots \geq r_m$ and nonzero column sums $c_1 \geq c_2 \geq \ldots \geq c_n$. Put $a_j = \#\{i : r_i \geq j\}$, $1 \leq j \leq n$, and $b_i = \#\{j : c_j \geq i\}$, $1 \leq i \leq m$. Define $\alpha_0 = \min\{\alpha(F_2, F) : F$ is a uniquely determined set$\}$. Let F_1 be a uniquely determined set with row sums $u_1 \geq u_2 \geq \ldots$, and column sums $v_1 \geq v_2 \geq \ldots$. Then the following conditions are equivalent:*

(i) $\alpha(F_2, F_1) = \alpha_0$,

(ii) for all $j \geq 1$ we have $\begin{cases} \min(a_j, c_j) \leq v_j \leq \max(a_j, c_j) & \text{if } 1 \leq j \leq n, \\ v_j = 0 & \text{otherwise,} \end{cases}$

(iii) for all $i \geq 1$ we have $\begin{cases} \min(b_i, r_i) \leq u_i \leq \max(b_i, r_i) & \text{if } 1 \leq i \leq m, \\ u_i = 0 & \text{otherwise.} \end{cases}$

Proof. We will prove the equivalence of (i) and (ii). By symmetry the equivalence of (i) and (iii) then follows. During the proof, we will use several times the fact that $u_i = \#\{j : v_j \geq i\}$, $i \geq 1$, as F_1 is uniquely determined (see Section 3.2).

$(i) \Rightarrow (ii)$. Suppose F_1 does not satisfy (ii). Then either $v_j \neq 0$ for some $j > n$, or $v_j < \min(a_j, c_j)$ for some j with $1 \leq j \leq n$, or $v_j > \max(a_j, c_j)$ for some j with $1 \leq j \leq n$. In each of those three cases we will prove that there exists a uniquely determined set F_1' such that $\alpha(F_2, F_1') < \alpha(F_2, F_1)$, which implies that F_1 does not satisfy (i).

Case 1: there is an $l > n$ such that $v_l \neq 0$. As for all $j \in \{1, 2, \ldots, n\}$ we have $v_j \geq v_l$, we must have $u_{v_l} = \#\{j : v_j \geq v_l\} \geq n + 1$. Now consider the set F_1' with the same row and column sums as F_1, except that the column sum with index l is exactly 1 smaller and the row sum with index v_l is exactly 1 smaller. Note that this set is uniquely determined. Since either $v_l > m$ (so r_{v_l} does not exist) or $r_{v_l} \leq n$, the difference in the row sums of F_1' and F_2 is 1 less than the difference in the row sums of F_1 and F_2. The same holds for the differences in the column sums. So $\alpha(F_2, F_1') < \alpha(F_2, F_1)$.

Case 2: there is a $k \in \{1, 2, \ldots, n\}$ such that $v_k < \min(a_k, c_k)$. Assume that k is the smallest positive integer with this property. Define F_1' such that its row sums u_i' and column sums v_j' are as follows:

$$u_i' = \begin{cases} u_i + 1 & \text{if } i = v_k + 1, \\ u_i & \text{otherwise,} \end{cases}$$

$$v_j' = \begin{cases} v_k + 1 & \text{if } j = k, \\ v_j & \text{otherwise.} \end{cases}$$

If $k = 1$, then the column sums of F_1' are obviously non-increasing. If $k \geq 2$, then

$$v_{k-1} \geq \min(a_{k-1}, c_{k-1}) \geq \min(a_k, c_k) > v_k,$$

so $v_{k-1}' = v_{k-1} \geq v_k + 1 = v_k'$, hence the column sums are non-increasing in this case as well. For the row sums we have $u_i' = \#\{j : v_j' \geq i\}$, which shows that the row sums are non-increasing and that F_1' is uniquely determined.
Clearly, the difference in the column sums has decreased by 1 when changing from F_1 to F_1'. The difference in the row sums has changed by $|u_{v_k+1} + 1 - r_{v_k+1}| - |u_{v_k+1} - r_{v_k+1}|$. We have $u_{v_k+1} = \#\{j : v_j \geq v_k + 1\} < k$. By Lemma 3.1 we have $r_{v_k+1} = \#\{j : a_j \geq v_k + 1\}$ and therefore $r_{v_k+1} \geq k$, using $a_k \geq \min(a_k, c_k) \geq v_k + 1$. Hence $u_{v_k+1} < r_{v_k+1}$ and therefore the difference in the row sums has decreased by 1. So $\alpha(F_2, F_1') < \alpha(F_2, F_1)$.

Case 3: there is a $k \in \{1, 2, \ldots, n\}$ such that $v_k > \max(a_k, c_k)$. This is analogous to Case 2.

$(ii) \Rightarrow (i)$. Suppose F_1 satisfies (ii). Consider the uniquely determined set with column sums $\min(a_j, c_j)$, $1 \leq j \leq n$, and non-increasing row sums. Then we can build F_1 starting from this set by adding new points one by one. Starting in the column with index 1, we add points to each column until there are v_j points in column j. The points added in column j are in rows $\min(a_j, c_j) + 1, \ldots, v_j$ in that order. In this way, in every step the constructed set has non-increasing row and column sums and is uniquely determined. We will prove that the value of α does not change in each step, which implies that the value of α of the set we started with is equal to $\alpha(F_2, F_1)$. That proves that all sets F_1 satisfying (ii) have the same value $\alpha(F_2, F_1)$. This must then be the minimal value α_0, since we proved in the first part that the minimal value occurs among the sets F_1 satisfying (ii).

Now assume that F_1 satisfies (ii) and let k be such that $v_k < \max(a_k, c_k)$ and if $k \geq 2$, then $v_k < v_{k-1}$. It suffices to prove that if we add the point $(k, v_k + 1)$ to F_1, then the value of α does not change. (Whenever we add a point in the procedure described above, the conditions $v_k < \max(a_k, c_k)$ and $v_k < v_{k-1}$ hold.) So define F_1' as the uniquely determined set with row sums u_i' and column sums v_j' satisfying

$$u_i' = \begin{cases} u_i + 1 & \text{if } i = v_k + 1, \\ u_i & \text{otherwise,} \end{cases}$$

$$v'_j = \begin{cases} v_k + 1 & \text{if } j = k, \\ v_j & \text{otherwise.} \end{cases}$$

We will prove that $\alpha(F_2, F'_1) = \alpha(F_2, F_1)$. We distinguish between two cases.

Case 1: $a_k \leq v_k < c_k$. By changing from F_1 to F'_1 the difference in the column sums has decreased by 1. We have $u_{v_k+1} = \#\{j : v_j \geq v_k + 1\} = k - 1$, as either $k = 1$ or $v_{k-1} \geq v_k + 1$. Also, by Lemma 3.1 we have $r_{v_k+1} = \#\{j : a_j \geq v_k + 1\} \leq k - 1$, since $a_k < v_k + 1$. So $u_{v_k+1} \geq r_{v_k+1}$, which shows that the difference in the row sums has increased by 1. Hence $\alpha(F_2, F'_1) = \alpha(F_2, F_1)$.

Case 2: $c_k \leq v_k < a_k$. By changing from F_1 to F'_1 the difference in the column sums has increased by 1. We have $u_{v_k+1} = k-1$ as in Case 1. Also, by Lemma 3.1 we have $r_{v_k+1} = \#\{j : a_j \geq v_k + 1\} \geq k$, since $a_k \geq v_k + 1$. So $u_{v_k+1} < r_{v_k+1}$, which shows that the difference in the row sums has decreased by 1. Hence $\alpha(F_2, F'_1) = \alpha(F_2, F_1)$.

This completes the proof of the theorem. □

Remark 3.2. *We can always permute the rows and columns such that the row and column sums of F_2 are non-increasing, so this condition in the above theorem is not a restriction. However, the monotony of the line sums of F_1 is a slight restriction. There may be a uniquely determined set F_1 satisfying $\alpha(F_2, F_1) = \alpha_0$ while its row and column sums are not non-increasing. However, reordering the row and column sums so that they are non-increasing never increases the differences with the row and column sums of F_2. So define in that case a set F'_1 with the same row and column sums as F_1, except that the line sums of F'_1 are ordered non-increasingly. Then $\alpha(F_2, F'_1) = \alpha(F_2, F_1) = \alpha_0$, so F'_1 satisfies the conditions of the theorem and (i) and therefore satisfies (ii) and (iii).*

Let F_2 be a set with row sums $r_1, r_2, \ldots, r_m$ and column sums $c_1, c_2, \ldots, c_n$, not necessarily non-increasing. Let σ be a permutation of $\{1, 2, \ldots, n\}$ such that $c_{\sigma(1)} \geq c_{\sigma(2)} \geq \ldots \geq c_{\sigma(n)}$. Consider the uniquely determined set F_1 with row sums $u_1 = r_1, u_2 = r_2, \ldots, u_m = r_m$ and column sums $v_1, v_2, \ldots, v_n$ such that $v_{\sigma(1)} \geq v_{\sigma(2)} \geq \ldots \geq v_{\sigma(n)}$. According to Theorem 3.4 we have $\alpha(F_2, F_1) = \alpha_0$, where $\alpha_0 = \min\{\alpha(F_2, F) : F \text{ is a uniquely determined set}\}$. Such a set F_1 we call a *uniquely determined neighbour of F_2*. Note that F_2 may have more than one uniquely determined neighbour, as there may be more possibilities for σ if some of the column sums of F_2 are equal. Also note that if F_3 is another set with row sums $r_1, r_2, \ldots, r_m$ and column sums $c_1, c_2, \ldots, c_n$, then F_1 is a uniquely determined neighbour of F_3 if and only if it is a uniquely determined neighbour of F_2.

It is easy to compute the line sums of a uniquely determined neighbour of F_2 and hence it is easy to find α_0.

Example 3.1. Consider the set F_2 with row sums $(r_1, \ldots, r_6) = (5, 5, 3, 2, 2, 1)$ and column sums $(c_1, \ldots, c_6) = (3, 1, 5, 4, 2, 3)$. To find a uniquely determined neighbour

of F_2 and to compute α_0, we first sort the column sums such that they are non-increasing: $(5, 4, 3, 3, 2, 1)$. Note that we can use two permutations to achieve this: either the first column ends up as the third column, while the sixth column ends up as the fourth column, or the other way around.

Now we compute the column sums of the uniquely determined set having the same row sums as F_2 and having non-increasing column sums. The column sums are the numbers $\#\{l : r_l \geq j\}$ for $j = 1, \ldots, 6$, which gives $(6, 5, 3, 2, 2, 0)$. Comparing this to the ordered column sums $(5, 4, 3, 3, 2, 1)$ of F_2, we see that the total difference in the column sums is 4, which means that $\alpha_0 = 2$.

As there are two possible permutations, there exist two different uniquely determined neighbours of F_2. The first one has column sums $(3, 0, 6, 5, 2, 2)$, while the second one has column sums $(2, 0, 6, 5, 2, 3)$.

3.5 Sets with equal line sums

Consider a set F_2 that is not uniquely determined by its line sums. When attempting to reconstruct F_2 from its line sums, one may end up with a different set F_3 that has the same line sums as F_2. It is interesting to know whether F_3 is a good approximation of F_2 or not. In some cases, F_3 may be disjoint from F_2, but in other cases, F_2 and F_3 must have a large intersection. We shall derive an upper bound on $F_2 \triangle F_3$ that depends on the size of F_2 and on how close F_2 is to being uniquely determined, in the sense of the previous section. Both parameters can easily be computed from the line sums of F_2.

Theorem 3.5. *Let F_2 and F_3 be finite subsets of $\mathbb{Z}^2$ with the same line sums. Let F_1 be a uniquely determined neighbour of F_2 and F_3. Put $\alpha = \alpha(F_2, F_1)$. Then*

$$|F_2 \triangle F_3| \leq 2\alpha\sqrt{8|F_2| + 1} - 2\alpha.$$

Proof. By Lemma 3.3 we have $\alpha\sqrt{8|F_2| + 1} - \alpha$ as an upper bound for both $|F_1 \triangle F_2|$ and $|F_1 \triangle F_3|$. Hence

$$|F_2 \triangle F_3| \leq |F_1 \triangle F_2| + |F_1 \triangle F_3| \leq 2\alpha\sqrt{8|F_2| + 1} - 2\alpha.$$

□

While we may not be able to reconstruct the set F_2, as it is not uniquely determined, we can reconstruct a uniquely determined neighbour F_1 of F_2. When F_2 is quite close to being uniquely determined, it must have a large intersection with F_1. Hence we know that at least a certain fraction of the points of F_1 must belong to F_2. The next theorem gives a bound for this fraction.

Theorem 3.6. *Let F_2 be a subset of $\mathbb{Z}^2$. Let F_1 be a uniquely determined neighbour of F_2. Put $\alpha = \alpha(F_2, F_1)$. Then*

$$\frac{|F_2 \cap F_1|}{|F_2|} \geq 1 - \frac{\sqrt{2}\alpha}{\sqrt{|F_2|}}.$$

Proof. By Remark 3.1 we have $|F_1 \triangle F_2| \leq 2\alpha\sqrt{2|F_2|}$. Hence

$$|F_1 \cap F_2| = |F_2| - \frac{1}{2}|F_1 \triangle F_2| \geq |F_2| - \alpha\sqrt{2|F_2|}.$$

Dividing by $|F_2|$ yields the theorem. □

Similarly, we can find a lower bound on the part of F_2 that must belong to any other reconstruction F_3.

Theorem 3.7. *Let F_2 and F_3 be finite subsets of $\mathbb{Z}^2$ with the same line sums. Let F_1 be a uniquely determined neighbour of F_2 and F_3. Put $\alpha = \alpha(F_2, F_1)$. Then*

$$\frac{|F_2 \cap F_3|}{|F_2|} \geq 1 - \frac{2\sqrt{2}\alpha}{\sqrt{|F_2|}}.$$

Proof. By Remark 3.1 we have $|F_1 \triangle F_2| \leq 2\alpha\sqrt{2|F_2|}$ and $|F_1 \triangle F_3| \leq 2\alpha\sqrt{2|F_2|}$. Hence

$$|F_2 \triangle F_3| \leq 4\alpha\sqrt{2|F_2|}.$$

So

$$|F_2 \cap F_3| = |F_2| - \frac{1}{2}|F_2 \triangle F_3| \geq |F_2| - 2\alpha\sqrt{2|F_2|}.$$

Dividing by $|F_2|$ yields the theorem. □

Corollary 3.8. *If F_2 and F_3 are disjoint sets with the same line sums, then*

$$|F_2| \leq 8\alpha^2.$$

Proof. If F_2 and F_3 are disjoint sets, then $|F_2 \cap F_3| = 0$, so by Theorem 3.7

$$0 \geq 1 - \frac{2\sqrt{2}\alpha}{\sqrt{|F_2|}},$$

which we can rewrite as $|F_2| \leq 8\alpha^2$. □

Theorem 3.6 shows that for given row and column sums that a set F_2 must satisfy, we can find a set of points F_1 such that any possible set F_2 must contain a subset of F_1 of a certain size. However, it may happen that none of the individual points of F_1 must necessarily belong to such a set F_2. It is possible to determine from the line sums the intersection of all possible sets F_2, see e.g. [4, Theorem 3.4]. The following statement is a particular case of that theorem.

Theorem 3.9. *Let F_2 be a subset of $\mathbb{Z}^2$ with column sums $c_1^{(2)} \geq c_2^{(2)} \geq \ldots \geq c_n^{(2)}$. Let F_1 be a uniquely determined neighbour of F_2 with column sums $c_1^{(1)} \geq c_2^{(1)} \geq \ldots \geq c_n^{(1)}$. Suppose*

$$\sum_{j=1}^{l} c_j^{(1)} > \sum_{j=1}^{l} c_j^{(2)} \quad \text{for } l = 1, 2, \ldots, n-1.$$

Then for all $(i,j) \in F_2$ there exists a set F_3 with the same row and column sums as F_2 such that $(i,j) \notin F_3$.

We illustrate the theorems in this section by the following example.

Example 3.2. Let m and n be positive integers. Let row sums $r_1, r_2, \ldots, r_n$ be given by $r_i = (n-i+1)m$ for $1 \leq i \leq n$. Let column sums $c_1, c_2, \ldots, c_{(n+1)m}$ be given by

- $c_j = n-1$ for $1 \leq j \leq m$,
- $c_{lm+j} = n-l$ for $1 \leq l \leq n-1$, $1 \leq j \leq m$.
- $c_j = 1$ for $nm+1 \leq j \leq (n+1)m$.

The uniquely determined set F_1 with row sums $r_1, r_2, \ldots, r_n$ has column sums $c'_1, c'_2, \ldots, c'_{(n+1)m}$ given by $c'_{lm+j} = n-l$ for $0 \leq l \leq n$, $1 \leq j \leq m$. For any set F_2 with row sums $r_1, r_2, \ldots r_n$ and column sum $c_1, c_2, \ldots, c_{(n+1)m}$ we have $\alpha = \alpha(F_1, F_2) = m$: the row sums of F_1 and F_2 are the same, while the column sums of the first m and last m columns differ by exactly 1.

Construct sets F_2 and F_3 as follows. In row i, $1 \leq i \leq n$, the elements of F_2 are in columns $1, 2, \ldots, (n-i)m$ and in columns $(n-i+1)m+1, (n-i+1)m+2, \ldots, (n-i+2)m$. In row i, $1 \leq i \leq n-1$, the elements of F_3 are in columns $1, 2, \ldots, (n-i+1)m$. In row n the elements of F_3 are in columns $nm+1, nm+2, \ldots, (n+1)m$. The sets F_2 and F_3 both have row sums $r_1, r_2, \ldots r_n$ and column sum $c_1, c_2, \ldots, c_{(n+1)m}$. We have $|F_2| = |F_3| = |F_1| = mn(n+1)/2$.

Theorem 3.5 states that

$$|F_2 \triangle F_3| \leq 2m\sqrt{4mn(n+1)+1} - 2m,$$

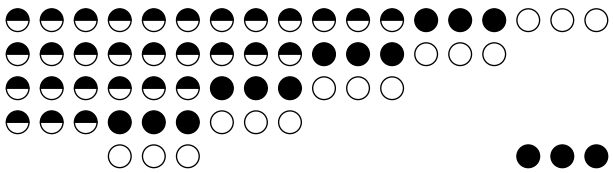


Figure 3.1: Example 3.2 with $n = 5$ and $m = 3$. The set F_2 consists of the white and black-and-white points, while F_3 consists of the black and black-and-white points.

while it actually holds that $|F_2 \triangle F_3| = 2mn$.

Theorem 3.6 states that

$$\frac{|F_1 \cap F_2|}{|F_2|} \geq 1 - \frac{\sqrt{2}m}{\sqrt{\frac{1}{2}mn(n+1)}} \geq 1 - \frac{2\sqrt{m}}{n},$$

while it actually holds that

$$\frac{|F_1 \cap F_2|}{|F_2|} = \frac{\frac{1}{2}mn(n-1)}{\frac{1}{2}mn(n+1)} = \frac{n-1}{n+1} = 1 - \frac{2}{n+1}.$$

Finally note that F_2 meets the conditions of Theorem 3.9, so none of the points of F_2 is contained in every set that has the same line sums as F_2.

3.6 Sets with different line sums

First consider two uniquely determined finite subsets F_1 and F_1' of $\mathbb{Z}^2$. Let the row sums of F_1 be denoted by $r_1, r_2, \ldots, r_m$ and let the row sums of F_1' be denoted by $r_1', r_2', \ldots, r_m'$. Without loss of generality, we may assume that $r_1 \geq r_2 \geq \ldots \geq r_m$.

Define $\alpha_1 = \alpha(F_1, F_1')$. According to Lemma 2.1, the symmetric difference $F_1 \triangle F_1'$ of the two sets can be decomposed into α_1 staircases. (In the aforementioned lemma the assumption is made that both sets considered have equal size; however, this is not used in the proof. Therefore, the statement holds for sets of any size, which we use here.) Let T be one of those staircases, of which the elements are contained in the rows $i_1 < i_2 < \ldots < i_k$. Let $(i_t, j) \in F_1 \backslash F_1'$ and $(i_{t+1}, j) \in F_1' \backslash F_1$ be elements of T. By Lemma 3.2 we have $r_{i_t} > r_{i_{t+1}}$ and $r'_{i_t} < r'_{i_{t+1}}$. Row i_1 must contain an element of $F_1 \backslash F_1'$ of T, and row i_k must contain an element of $F_1' \backslash F_1$ of T. Hence we can apply this for $t = 1, 2, \ldots, k-1$, and we find

$$r_{i_1} > r_{i_2} > \ldots > r_{i_k},$$

$$r'_{i_1} < r'_{i_2} < \ldots < r'_{i_k}.$$

Assume without loss of generality that there is at least one value of t for which $r'_{i_t} - r_{i_t} \geq 0$. (Otherwise, reverse the roles of r'_i and r_i in what follows.) Let

$$u = \min\{r'_{i_t} - r_{i_t} : r'_{i_t} - r_{i_t} \geq 0\}$$

and let s be such that $r'_{i_s} - r_{i_s} = u$. We distinguish two cases: $u = 0$ and $u \geq 1$.

Case 1: suppose $u = 0$. For $t \geq s$ we have $r_{i_t} \leq r_{i_s} - (t-s)$ and $r'_{i_t} \geq r'_{i_s} + (t-s)$, hence

$$r'_{i_t} - r_{i_t} \geq r'_{i_s} - r_{i_s} + 2(t-s) = 2(t-s) \geq 0,$$

so

$$|r'_{i_t} - r_{i_t}| \geq 2(t-s).$$

For $t < s$ we have $r_{i_t} \geq r_{i_s} + (s-t)$ and $r'_{i_t} \leq r'_{i_s} - (s-t)$, hence

$$r'_{i_t} - r_{i_t} \leq r'_{i_s} - r_{i_s} - 2(s-t) = -2(s-t) < 0,$$

so

$$|r'_{i_t} - r_{i_t}| \geq 2(s-t).$$

Now we have

$$\begin{aligned}
\sum_{t=1}^{k} |r'_{i_t} - r_{i_t}| &\geq \sum_{t=1}^{s-1} 2(s-t) + \sum_{t=s}^{k} 2(t-s) \\
&= 2s^2 + (-2k-2)s + (k^2+k) \\
&\geq 2\left(\frac{k+1}{2}\right)^2 + (-2k-2)\frac{k+1}{2} + (k^2+k) \\
&= \tfrac{1}{2}k^2 - \tfrac{1}{2}.
\end{aligned}$$

Case 2: suppose $u \geq 1$. Similarly to the first case, we have for $t \geq s$:

$$|r'_{i_t} - r_{i_t}| \geq 2(t-s) + 1.$$

If $s = 1$, there are no $t < s$ to consider. Assume $s \geq 2$. Then $r'_{i_{s-1}} - r_{i_{s-1}} < r'_{i_s} - r_{i_s} = u$, so by the minimality of u we must have $r'_{i_{s-1}} - r_{i_{s-1}} \leq -1$. Similarly to above, we have

$$|r'_{i_t} - r_{i_t}| \geq 2(s-t) - 1.$$

Hence

$$\begin{aligned}
\sum_{t=1}^{k} |r'_{i_t} - r_{i_t}| &\geq \sum_{t=1}^{s-1} (2(s-t)-1) + \sum_{t=s}^{k} (2(t-s)+1) \\
&= 2s^2 + (-2k-4)s + (k^2+2k+2) \\
&\geq 2\left(\frac{k+2}{2}\right)^2 + (-2k-4)\frac{k+2}{2} + (k^2+2k+2) \\
&= \tfrac{1}{2}k^2.
\end{aligned}$$

In both cases we have $\sum_{t=1}^{k} |r'_{i_t} - r_{i_t}| \geq \frac{1}{2}k^2 - \frac{1}{2}$, and since the sum must be an integer, we have

$$\sum_{t=1}^{k} |r'_{i_t} - r_{i_t}| \geq \lfloor \tfrac{1}{2}k^2 \rfloor.$$

Hence the difference between the row sums of F_1 and F'_1 is at least $\lfloor k^2/2 \rfloor$. Similarly, if T is a staircase that contains elements in k columns, the difference between the column sums of F_1 and F'_1 is at least $\lfloor k^2/2 \rfloor$.

Theorem 3.10. *Let F_1 and F'_1 be uniquely determined finite subsets of $\mathbb{Z}^2$. Put $\alpha_1 = \alpha(F_1, F'_1)$. Then*

$$|F_1 \triangle F'_1| \leq 2\alpha_1\sqrt{2\alpha_1 + 1} - \alpha_1.$$

Proof. Consider all staircases in $F_1 \triangle F'_1$, and let T be one with the maximal number of elements. We distinguish two cases.

- Suppose T has $2k + 1$ elements for some $k \geq 0$. Then exactly $k + 1$ rows and $k + 1$ columns contain elements of T. By the argument above, we have

  $$2\alpha_1 \geq \lfloor \tfrac{1}{2}(k+1)^2 \rfloor + \lfloor \tfrac{1}{2}(k+1)^2 \rfloor \geq (k+1)^2 - 1 = k^2 + 2k.$$

  This implies $k + 1 \leq \sqrt{2\alpha_1 + 1}$ and therefore $2k + 1 \leq 2\sqrt{2\alpha_1 + 1} - 1$.

- Suppose T has $2k$ elements for some $k \geq 1$. Then either k rows and $k + 1$ columns or $k + 1$ rows and k columns contain elements of T. By the argument above, we have

  $$2\alpha_1 \geq \lfloor \tfrac{1}{2}(k+1)^2 \rfloor + \lfloor \tfrac{1}{2}k^2 \rfloor = \tfrac{1}{2}(k+1)^2 + \tfrac{1}{2}k^2 - \tfrac{1}{2} = k^2 + k.$$

  This implies $k + 1/2 \leq \sqrt{2\alpha_1 + 1/4}$ and therefore $2k \leq 2\sqrt{2\alpha_1 + 1/4} - 1$.

All α_1 staircases of $F_1 \triangle F'_1$ have at most as many elements as T, so in both cases we have

$$|F_1 \triangle F'_1| \leq 2\alpha_1\sqrt{2\alpha_1 + 1} - \alpha_1.$$

□

Remark 3.3. *It is remarkable that the bound in Theorem 3.10 does not depend on the sizes of F_1 and F'_1. Such a dependency cannot be avoided if one of the two sets is not uniquely determined, as in Lemma 3.3. To show this, notice that in Example 3.2 for fixed $\alpha = m$ the symmetric difference $|F_1 \triangle F_2|$ becomes arbitrarily large when n tends to infinity. Theorem 3.10 shows that this cannot happen if both sets are uniquely determined.*

Example 3.3. Let $n > 1$ be an integer. Define $r_i = n - i$ for $1 \leq i \leq n$ and $r'_n = n$. Let F_1 be the uniquely determined set with row and column sums $r_1, r_2, \ldots, r_n$. Let F'_1 be the uniquely determined set with row and column sums $r_1, r_2, \ldots, r_{n-1}, r'_n$. We have $\alpha_1 = \alpha(F_1, F'_1) = n$. Consider row i, where $1 \leq i \leq n-1$. The elements of F_1 in this row are in columns $1, 2, \ldots, n-i$, while the elements of F'_1 in this row are in columns $1, 2, \ldots, n-i-1$ and n. In row n there are n elements of F'_1 and none of F_1.

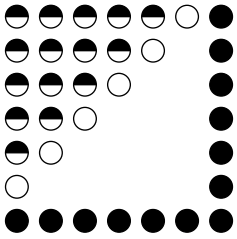


Figure 3.2: Example 3.3 with $n = 7$. The set F_1 consists of the white and black-and-white points, while F'_1 consists of the black and black-and-white points.

Hence

$$|F_1 \triangle F'_1| = 2(n-1) + n = 3n - 2,$$

while Theorem 3.10 states that

$$|F_1 \triangle F'_1| \leq 2n\sqrt{2n+1} - n.$$

Finally we derive a bound on the symmetric difference of two sets F_2 and F_3 with arbitrary line sums.

Theorem 3.11. *Let F_2 and F_3 be finite subsets of $\mathbb{Z}^2$. Let F_1 be a uniquely determined neighbour of F_2, and let F'_1 be a uniquely determined neighbour of F_3. Put $\alpha_2 = \alpha(F_1, F_2)$, $\alpha_3 = \alpha(F'_1, F_3)$ and $\alpha_1 = \alpha(F_1, F'_1)$. Then*

$$|F_2 \triangle F_3| \leq \alpha_2\sqrt{8|F_2|+1} - \alpha_2 + \alpha_3\sqrt{8|F_3|+1} - \alpha_3 + 2\alpha_1\sqrt{2\alpha_1+1} - \alpha_1.$$

Proof. This is an immediate result of Lemma 3.3 and Theorem 3.10. □

Example 3.4. Let n be a positive integer. We construct sets F_2 and F_3 as follows.

- In row i, where $1 \leq i \leq n$, the elements of F_2 are in columns $1, 2, \ldots, 2(n-i)$ as well as columns $2(n-i)+2$ and $2(n-i)+3$.
- In row $n+1$, there is a single element of F_2 in column 1.

- In row i, where $1 \leq i \leq n$, the elements of F_3 are in columns $1, 2, \ldots, 2(n-i)+1$ as well as column $2(n-i)+4$.
- In row $n+1$ there are no elements of F_3.

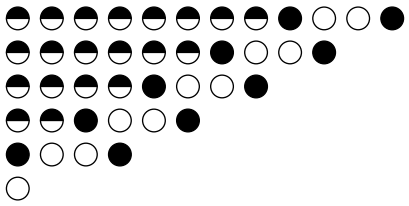


Figure 3.3: Example 3.4 with $n = 5$. The set F_2 consists of the white and black-and-white points, while F_3 consists of the black and black-and-white points.

The row sums of F_2 are given by

$$r_i^{(2)} = \begin{cases} 2(n-i+1) & \text{if } 1 \leq i \leq n, \\ 1 & \text{if } i = n+1. \end{cases}$$

The column sums of F_2 are given by

$$c_j^{(2)} = \begin{cases} n - \lfloor \frac{j-1}{2} \rfloor & \text{if } 1 \leq j \leq 2n, \\ 1 & \text{if } j = 2n+1, \\ 0 & \text{if } j = 2n+2. \end{cases}$$

The row sums of F_3 are given by

$$r_i^{(3)} = 2(n-i+1), \quad 1 \leq i \leq n+1.$$

The column sums of F_3 are given by

$$c_j^{(3)} = \begin{cases} n & \text{if } j = 1, \\ n-1 & \text{if } j = 2, \\ n - \lfloor \frac{j-1}{2} \rfloor & \text{if } 3 \leq j \leq 2n, \\ 0 & \text{if } j = 2n+1, \\ 1 & \text{if } j = 2n+2. \end{cases}$$

Let F_1 be the uniquely determined set with the same row sums as F_2 and non-increasing column sums. Let F_1' be the uniquely determined set with the same row sums as F_3 and non-increasing column sums. We have

$$\alpha_2 = \alpha(F_2, F_1) = 1, \quad \alpha_3 = \alpha(F_3, F_1') = 1, \quad \alpha_1 = \alpha(F_1, F_1') = 1.$$

Furthermore, $|F_2| = n(n+1)+1$ and $|F_3| = n(n+1)$.

Theorem 3.11 states that

$$|F_2 \triangle F_3| \leq \sqrt{8n(n+1)+9} + \sqrt{8n(n+1)+1} + 2\sqrt{3} - 3 \approx 4\sqrt{2}n,$$

while actually

$$|F_2 \triangle F_3| = 4n + 1.$$

3.7 Concluding remarks

We have proved an upper bound on the difference between two images with the same row and column sums, as well as on the difference between two images with different row and column sums. The bounds heavily depend on the parameter α, which indicates how close an image is to being uniquely determined. If a set of given line sums "almost uniquely determines" the image (i.e. α is very small) it may still happen that no points belong to all possible images with those line sums. However, using the results from this chapter we can find a set of points of which a subset of certain size is guaranteed to belong to the image.

There is still a gap between the examples we have found and the bounds we have proved. It appears that all bounds can be improved by a factor $\sqrt{\alpha}$. For this it would suffice to improve both Lemma 3.3 and Theorem 3.10 by a factor $\sqrt{\alpha}$, but so far we did not manage to improve either of those.

The results of this chapter can be applied to projections in more than two directions as well: simply pick two directions and forget about the others. One would expect this to give bad results, but that is actually not always the case. It is possible to construct examples with projections in more than two directions where the bound using only two of the directions is still only a factor $\sqrt{\alpha}$ off. However, in many cases it should be (somehow) possible to use the projections in all directions to get better results.

CHAPTER 4

A lower bound on the largest possible difference

This chapter (with minor modifications) has been published as: Birgit van Dalen "On the difference between solutions of discrete tomography problems II", Pure Mathematics and Applications 20 (2009) 103-112.

4.1 Introduction

In Chapter 3 we studied the possible difference between two binary images with the same line sums. We introduced a parameter α that indicates how close given line sums are to line sums that uniquely determine an image. We proved upper bounds on the size of the symmetric difference between two solutions of the same projections, depending on α.

In this chapter we consider the complementary problem: find the best lower bound for the symmetric difference between two solutions that you can at least achieve given a set of projections. For each set of projections that has at least two solutions, we construct two solutions that have a symmetric difference of at least $2\alpha + 2$. We also show that this bound is sharp.

4.2 Definitions and notation

Let F be a finite subset of $\mathbb{Z}^2$ with characteristic function χ. (That is, $\chi(k,l) = 1$ if $(k,l) \in F$ and $\chi(k,l) = 0$ otherwise.) For $i \in \mathbb{Z}$, we define *row* i as the set $\{(k,l) \in \mathbb{Z}^2 : k = i\}$. We call i the index of the row. For $j \in \mathbb{Z}$, we define *column* j as the set $\{(k,l) \in \mathbb{Z}^2 : l = j\}$. We call j the index of the column. Note that we follow matrix notation: we indicate a point (i,j) by first its row index i and then its column index j. Also, we use row numbers that increase when going downwards and column numbers that increase when going to the right.

The *row sum* r_i is the number of elements of F in row i, that is $r_i = \sum_{j\in\mathbb{Z}} \chi(i,j)$. The *column sum* c_j of F is the number of elements of F in column j, that is $c_j = \sum_{i\in\mathbb{Z}} \chi(i,j)$. We refer to both row and column sums as the *line sums* of F. We will usually only consider finite sequences $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and $\mathcal{C} = (c_1, c_2, \ldots, c_n)$ of row and column sums that contain all the nonzero line sums. We may assume without loss of generality that $r_1 \geq r_2 \geq \ldots \geq r_m$ and $c_1 \geq c_2 \geq \ldots \geq c_n$.

Given sequences of integers $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and $\mathcal{C} = (c_1, c_2, \ldots, c_n)$, we say that $(\mathcal{R}, \mathcal{C})$ is *consistent* if there exists a set F with row sums $\mathcal{R}$ and column sums $\mathcal{C}$. We say that the line sums $(\mathcal{R}, \mathcal{C})$ uniquely determine such a set F if the following property holds: if F' is another subset of $\mathbb{Z}^2$ with line sums $(\mathcal{R}, \mathcal{C})$, then $F' = F$. In this case we call F *uniquely determined*.

We will now define a *uniquely determined neighbour* of a set F. This is a uniquely determined set that is in some sense the closest to F. See also Section 3.4.

Definition 4.1. *Suppose F has row sums $r_1 \geq r_2 \geq \ldots \geq r_m$ and column sums $c_1 \geq c_2 \geq \ldots \geq c_n$. For $1 \leq j \leq n$, let $v_j = \#\{l : r_l \geq j\}$. Then the row sums r_1, r_2, …, r_m and column sums v_1, v_2, …, v_n uniquely determine a set F_1, which we will call the* uniquely determined neighbour of F.

Note that if F' is another set with row sums r_1, r_2, …, r_m and column sums c_1, c_2, …, c_n, then F_1 is a uniquely determined neighbour of F' if and only if it is a uniquely determined neighbour of F. Hence F_1 only depends on the row and column sums and not on the choice of the set F. We will therefore also speak about the *uniquely determined neighbour corresponding to the line sums* $(\mathcal{R}, \mathcal{C})$, without mentioning the set F.

Suppose line sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and $\mathcal{C} = (c_1, c_2, \ldots, c_n)$ are given, where $r_1 \geq r_2 \geq \ldots \geq r_m$ and $c_1 \geq c_2 \geq \ldots \geq c_n$. Let the uniquely determined neighbour corresponding to $(\mathcal{R}, \mathcal{C})$ have column sums $v_1 \geq v_2 \geq \ldots \geq v_n$. Then we define

$$\alpha(\mathcal{R}, \mathcal{C}) = \frac{1}{2} \sum_{j=1}^{n} |c_j - v_j|.$$

Note that $\alpha(\mathcal{R},\mathcal{C})$ is an integer, since $2\alpha(\mathcal{R},\mathcal{C})$ is congruent to

$$\sum_{j=1}^{n}(c_j+v_j)=\sum_{j=1}^{n}c_j+\sum_{j=1}^{n}v_j=2\sum_{j=1}^{n}c_j\equiv 0 \mod 2.$$

Consider a set F with line sums $(\mathcal{R},\mathcal{C})$ and its uniquely determined neighbour F_1. Let $\alpha=\alpha(\mathcal{R},\mathcal{C})$. It was proved in Lemma 2.1 that the symmetric difference $F\triangle F_1$ consists of α *staircases*. In this chapter we will only use staircases of length 2, which we will define below. For the general definition of a staircase, see Chapter 2.

Definition 4.2. *A* staircase of length 2 *in $F\triangle F_1$ is a pair of points (p_1,p_2) in $\mathbb{Z}^2$ such that*

- *p_1 and p_2 are in the same row,*
- *p_1 is an element of $F\backslash F_1$,*
- *p_2 is an element of $F_1\backslash F$.*

4.3 Main result

Suppose row sums $\mathcal{R}=(r_1,r_2,\ldots,r_m)$ and column sums $\mathcal{C}=(c_1,c_2,\ldots,c_n)$ are given, where $r_1\geq r_2\geq\ldots\geq r_m$ and $c_1\geq c_2\geq\ldots\geq c_n$. Assume that the line sums are consistent but do not uniquely determine a set F (hence at least two different sets with these line sums exist). Let $\alpha=\alpha(\mathcal{R},\mathcal{C})$.

In Chapter 3 it was shown that for all F_2 and F_3 satisfying these line sums, we have

$$|F_2\triangle F_3|\leq 4\alpha\sqrt{2|F_2|}.$$

One may wonder how close we can get to achieving this bound. Our theorem shows that we can construct two sets that have a symmetric difference of size at least $2\alpha+2$.

Theorem 4.1. *Let be given row sums $\mathcal{R}=(r_1,r_2,\ldots,r_m)$ and column sums $\mathcal{C}=(c_1,c_2,\ldots,c_n)$, where $r_1\geq r_2\geq\ldots\geq r_m$ and $c_1\geq c_2\geq\ldots\geq c_n$. Assume that the line sums are consistent but do not uniquely determine a set F. Let $\alpha=\alpha(\mathcal{R},\mathcal{C})$. Then there exist sets F_2 and F_3 with these line sums such that*

$$|F_2\triangle F_3|\geq 2\alpha+2.$$

This bound is sharp: for each $\alpha\geq 1$ there are line sums $(\mathcal{R},\mathcal{C})$ with $\alpha=\alpha(\mathcal{R},\mathcal{C})$ such that for any F_2 and F_3 satisfying these line sums we have $|F_2\triangle F_3|\leq 2\alpha+2$.

4.4 Proof

In this entire section, the row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$ with $r_1 \geq r_2 \geq \ldots \geq r_m$ and $c_1 \geq c_2 \geq \ldots \geq c_n$ are fixed. Furthermore, F_1 is the uniquely determined neighbour corresponding to $(\mathcal{R}, \mathcal{C})$, and $\alpha = \alpha(\mathcal{R}, \mathcal{C})$. We denote the column sums of F_1 by $v_1 \geq v_2 \geq \ldots \geq v_n$.

The proof is constructive. We will construct F_2 and F_3 such that they have the desired property. We will do this by changing a set F step by step. Only the final result of the construction will be called F_2 (or F_3); the intermediate sets will always be called F or F'. In Section 4.5 the construction is illustrated by an example.

Let the columns j for which $v_j > c_j$ have indices $j_1 \leq j_2 \leq \ldots \leq j_\alpha$, where each such j occurs $v_j - c_j$ times. Similarly, let the columns i for which $v_i < c_i$ have indices $i_1 \leq i_2 \leq \ldots \leq i_\alpha$, where each such i occurs $c_i - v_i$ times. Define a *column pair* as a pair (i_t, j_t). The consistency of the given line sums assures that $i_t > j_t$ for all t. For convenience, define $i_0 = j_0 = 0$ and $i_{\alpha+1} = j_{\alpha+1} = n + 1$.

We will construct both F_2 and F_3 by starting from $F = F_1$ and then for each t moving an element of F from column j_t to column i_t in the same row. After we have done that for $t = 1, 2, \ldots, \alpha$, the row sums of F have not changed, while the columns of F have changed from v_1, v_2, …, v_n to c_1, c_2, …, c_n. The symmetric difference $F_1 \triangle F$ then consists of α staircases of length 2. Each staircase is confined to a single row and corresponds to a column pair (i_t, j_t). We will show that we have a certain freedom in choosing the staircases.

Suppose we have moved an element for each of the column pairs (i_1, j_1), (i_2, j_2), …, (i_{t-1}, j_{t-1}), where $t \geq 1$. The resulting set is called F and has column sums c'_1, c'_2, …, c'_n. Now we want to move an element from column j_t to column i_t. For this we need a row l such that the point $(l, j_t) \in F$ and $(l, i_t) \notin F$. We have $c'_{j_t} > c_{j_t} \geq c_{i_t} > c'_{i_t}$, so $c'_{j_t} \geq c'_{i_t} + 2$. Hence there must be at least two rows that contain an element of F in column j_t but not in column i_t. This proves the existence of such a row l, and in fact at least two choices for l are possible. Now we move the element (l, j_t) to (l, i_t). The row sums of F do not change, while the column sum of column j_t decreases by one and the column sum of column i_t increases by one.

We construct both F_2 and F_3 using the construction above. First we construct F_2, making arbitrary choices for the rows in which we move elements. Then we will construct F_3. For this we let the choices in the construction depend on F_2, in a way we will describe below.

Let P_1, P_2, …, P_r be the *distinct* column pairs, where P_h has multiplicity k_h: the column pair P_1 is equal to each of the pairs (i_1, j_1), …, (i_{k_1}, j_{k_1}), the column pair

P_2 is equal to each of the pairs (i_{k_1+1}, j_{k_1+1}), ..., $(i_{k_1+k_2}, j_{k_1+k_2})$, and so on. We have $k_1 + k_2 + \cdots + k_r = \alpha$. For two consecutive column pairs (i_t, j_t) and (i_{t+1}, j_{t+1}) that are not equal we have $i_{t+1} > i_t$, $j_{t+1} \geq j_t$ or $i_{t+1} \geq i_t$, $j_{t+1} > j_t$, so the second pair contains a column that did not occur in any of the previous pairs. This means that in P_1, ..., P_r at least $r + 1$ different columns are involved. For each P_h, we denote one of the columns in P_h as the *final* column of P_h in the following way.

- If one of the columns in P_h also occurs in P_{h+1}, then the other does not occur in P_{h+1}, ..., P_r. We call the latter the final column of the pair.
- If both columns in P_h do not occur in P_{h+1}, ..., P_r, and one of the columns occurs in P_{h-1}, then the other does not occur in P_1, ..., P_{h-1}. We call the former the final column of the pair.
- If both columns in P_h do not occur in P_1, ..., P_{h-1} nor in P_{h+1}, ..., P_r, then we arbitrarily pick one of the columns in P_h and call it the final column of the pair.

By definition, we have the following properties: the final column of P_h does not occur in P_{h+1}, ..., P_r, and if the other column does not occur in P_{h+1}, ..., P_r either, then the latter column only occurs in P_h.

Our goal is to construct F_3 in such a way that, for all h, in the final column of P_h the symmetric difference between F_2 and F_3 is at least $2k_h$, while in any other column that occurs in one of the column pairs the symmetric difference between F_2 and F_3 is at least 2. (There is at least one such a column, since there are exactly r final columns, while at least $r + 1$ columns are involved in the column pairs.) If we can achieve that, then we have

$$|F_2 \triangle F_3| \geq 2k_1 + 2k_2 + \ldots + 2k_r + 2 = 2\alpha + 2.$$

To achieve this, we choose the rows in which elements are moved for all equal column pairs at once. First we choose the rows for all pairs equal to P_1, then for all pairs equal to P_2, and so on.

Let t be the index of the last column pair in a sequence of k equal column pairs

$$(i_{t-k+1}, j_{t-k+1}) = (i_{t-k+2}, j_{t-k+2}) = \ldots = (i_t, j_t),$$

where $(i_{t-k}, j_{t-k}) \neq (i_{t-k+1}, j_{t-k+1})$ and $(i_t, j_t) \neq (i_{t+1}, j_{t+1})$. Suppose we have moved elements already for the column pairs (i_1, j_1), ..., (i_{t-k}, j_{t-k}). Call the resulting set F, with column sums c'_1, ..., c'_n. Assume that i_t is the final column of (i_t, j_t) (the case where j_t is the final column, is analogous). So we have $i_t \neq i_{t+1}$. Also, we have one of the following two properties:

(A) $j_t = j_{t+1}$,

(B) $j_t \neq j_{t+1}$, and $j_{t-k} \neq j_{t-k+1}$.

As this is the last time column i_t occurs, we need to choose the rows in such a way that by moving the elements of F the symmetric difference between F and F_2 in this column becomes at least $2k$. Also, in case (B) we want the symmetric difference in column j_t to be at least 2.

Since we need to move k elements out of column j_t into column i_t, we have $c'_{j_t} \geq c_{j_t} + k \geq c_{i_t} + k \geq c'_{i_t} + 2k$, so there are at least $2k$ rows l such that $(l, j_t) \in F$ and $(l, i_t) \notin F$. Let R be the set of those $2k$ rows. (If there are more than $2k$ possible rows, then pick $2k$ of them.) We distinguish between two cases.

Case 1. Suppose there are k different rows l in R such that $(l, i_t) \notin F_2$. Then we move elements from column j_t to column i_t in each of those k rows. Call the resulting set F'. We have $(l, i_t) \in F' \backslash F_2$ for k different values of l. The number of elements of F' in column i_t must be equal to the number of elements of F_2 in column i_t, so there are also k different values of l for which $(l, i_t) \in F_2 \backslash F'$. Hence the symmetric difference between F' and F_2 in this column is at least $2k$.

In case (A) we are now done, as column j_t will be handled in a later column pair. Suppose we are in case (B). The column j_t only occurs in the column pairs $(i_{t-k+1}, j_{t-k+1}), \ldots, (i_t, j_t)$, which are all equal. If for a row l we have $(l, i_t) \notin F_2$, then in the construction of F_2 this row was not used for a staircase corresponding to the column pair (i_t, j_t) (or one of the equal ones), so we must have $(l, j_t) \in F_2$. Hence after moving elements we have k different values of l for which $(l, j_t) \in F_2 \backslash F'$. So in column j_t the symmetric difference between F' and F_2 is at least $2k \geq 2$.

Case 2. Suppose there are at least $k+1$ different rows l in R such that $(l, i_t) \in F_2$. Let R' be a set of $k+1$ of those rows. Pick one of the rows in R' and call it l_0. Let R'' consist of l_0 and the $k-1$ other rows in $R \backslash R'$ (for which it may or may not hold that $(l, i_t) \in F_2$). Move elements from column j_t to column i_t in each of the k rows in R''. Call the resulting set F'. Then for all k rows l in $R \backslash R''$ we have $(l, i_t) \in F_2 \backslash F'$. Similarly to above, we find that the symmetric difference between F' and F_2 in column i_t is at least $2k$.

Again, in case (A) we are done. Suppose we are in case (B). As column j_t only occurs in the column pairs $(i_{t-k+1}, j_{t-k+1}), \ldots, (i_t, j_t)$, which are all equal, for at most k rows l in R we have $(l, j_t) \notin F_2$. This means that we can choose l_0 above in such a way that $(l_0, j_t) \in F_2$. After moving the elements, we then have $(l_0, j_t) \in F_2 \backslash F'$. So the symmetric difference between F' and F_2 in column j_t is at least 2.

At least one of Case 1 and Case 2 above must hold, since there are $2k$ rows in R. Therefore we have finished the construction of F_2 and F_3 such that $F_2 \triangle F_3 \geq 2\alpha + 2$.

We will now prove the second part of Theorem 4.1 by giving a family of examples for which the bound of $2\alpha+2$ is sharp. Let $s \geq 1$ be an integer. Take $m = n = s+1$ and let all row and column sums be equal to 1. These line sums are consistent. The uniquely determined neighbour F_1 has column sums $v_1 = s+1$, $v_2 = v_3 = \ldots = v_{s+1} = 0$, so $\alpha = s$.

Suppose F_2 and F_3 satisfy the given row and column sums. We have $|F_2| = |F_3| = s+1$, hence

$$|F_2 \triangle F_3| \leq |F_2| + |F_3| = 2(s+1) = 2\alpha + 2.$$

This completes the proof of Theorem 4.1. □

Remark 4.1. *There do not seem to be very many examples for which the bound of $2\alpha+2$ is sharp. In particular, they all seem to have $m = n = \alpha+1$. However, even in more general cases, when α is much larger than n, the bound is not very far off. Take for example $m = n$ and let all line sums be equal to k, where $k \leq \frac{1}{2}n$. The uniquely determined neighbour has k column sums equal to n and $n-k$ column sums equal to 0, so $\alpha = k(n-k)$. As $n-k \geq \frac{1}{2}n$, we have $\alpha \geq \frac{1}{2}kn$. Suppose F_2 and F_3 satisfy the given row and column sums, then $|F_2| = |F_3| = kn$, hence*

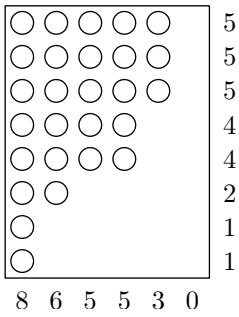
$$|F_2 \triangle F_3| \leq |F_2| + |F_3| = 2kn \leq 4\alpha.$$

4.5 Example

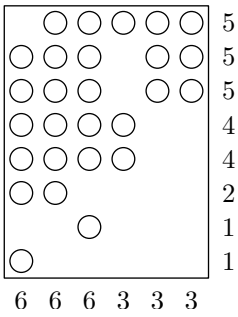
We illustrate the construction in the proof by an example. Let be given row sums $(5,5,5,4,4,2,1,1)$ and column sums $(6,6,6,3,3,3)$. The uniquely determined neighbour F_1 has the same row sums, but column sums $(8,6,5,5,3,0)$ (see Figure 4.1(a)). From this we derive that $\alpha = 4$ and that the four column pairs are $(3,1)$, $(6,1)$, $(6,4)$ and $(6,4)$.

To construct F_2, we move one element from column 1 to column 3, one element from column 1 to column 6, and two elements from column 4 to column 6. We choose the rows to move elements in arbitrarily from the available rows. If we choose rows 7, 1, 2 and 3 respectively, we arrive at the set F_2 shown in Figure 4.1(b).

Now we construct the set F_3 step-by-step, following the proof of the theorem. We start with F_1, shown again in Figure 4.2(a). For the first column pair, we need to move an element from column 1 to column 3. The available rows are 6, 7 and 8. We need only two of them, so let us take $R = \{7,8\}$. Column 3 is the final column in this column pair, so in this column we need to make sure that we achieve a symmetric difference of at least 2 with F_2. We have $(8,3) \notin F_2$, so we are in case 1 and we pick row 8 for our staircase. Hence we delete the element $(8,1)$ and add the element $(8,3)$. The new situation is shown in Figure 4.2(b).



(a) The set F_1 with its row and column sums.



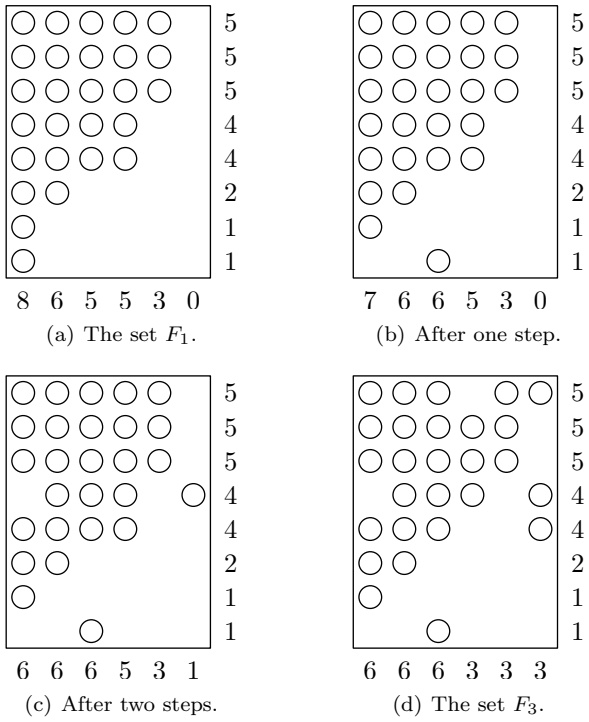
(b) The set F_2 with its row and column sums.

Figure 4.1

The next column pair is $(6, 1)$. Now column 1 is the final column of the pair, and all rows except row 8 are available. We are again in case 1 and pick row 4. Figure 4.2(c) shows the new situation, after deleting $(4, 1)$ and adding $(4, 6)$.

Finally, we need to move two elements at once for the column pair $(6, 4)$, which occurs twice. Column 6 is the final column, so we need to achieve a symmetric difference of at least 4 with F_2 in this column. We also need a symmetric difference of at least 2 in column 4 (case (B)). We have $R = \{1, 2, 3, 5\}$. As $(1, 8)$, $(2, 8)$ and $(3, 8)$ are all elements of F_2, we are in case 2. We have $R' = \{1, 2, 3\}$ and we need to find an $l_0 \in R'$ such that $(l_0, 4) \in F_2$. The only possible choice is $l_0 = 1$. We find $R'' = \{1, 5\}$, so we delete $(1, 4)$ and $(5, 4)$, and we add $(1, 6)$ and $(5, 6)$. This completes the construction of F_3. The resulting set is shown in Figure 4.2(d).

The construction guarantees that the symmetric difference between F_2 and F_3 is at least $2\alpha + 2 = 10$, but we have in fact constructed two sets with symmetric difference 14.



(a) The set F_1.

(b) After one step.

(c) After two steps.

(d) The set F_3.

Figure 4.2: The construction of the set F_3.

CHAPTER 5

Minimal boundary length of a reconstruction

This chapter (with minor modifications) will be published in SIAM Journal on Discrete Mathematics. A preprint is available as Birgit van Dalen, "Boundary length of reconstructions in discrete tomography", arXiv:1006.4449 [math.CO] (2010) 25 pp.

5.1 Introduction

If there are multiple images corresponding to one set of line sums, it is interesting to reconstruct an image with a special property. In order to find reconstructions that look rather like a real object, two special properties in particular are often imposed on the reconstructions. The first is *connectivity* of the points with value one in the picture [6, 8, 28]. The second is *hv-convexity*: if in each row and each column, the points with value one form one connected block, the image is called *hv-convex*. The reconstruction of hv-convex images, either connected or not necessarily connected, has been studied extensively [5, 6, 8, 9, 28].

Another relevant concept in this context is the *boundary* of a binary image. The boundary can be defined as the set of pairs consisting of two adjacent points, one with value 0 and one with value 1. Here we use 4-adjacency: that is, a point is adjacent to its two vertical and to its two horizontal neighbours [21]. The number of such pairs of adjacent points with two different values is called the *length of the boundary* or sometimes the *perimeter length* [12].

In this chapter we will consider given line sums that may correspond to more than one binary image. Since the boundary of real objects is often small compared to the area, it makes sense to look for reconstructions of which the length of the boundary is as small as possible. In particular, if there exists an hv-convex reconstruction, then the length of the boundary of that image is the smallest possible. In that sense, the length of the boundary is a more general concept than hv-convexity.

The question we are interested in in this chapter is: given line sums, what is the smallest length of the boundary that a reconstruction fitting those line sums can have? We can give two straightforward lower bounds on the length of the boundary, given the row and column sums. Both are equivalent to bounds given by Dahl and Flatberg in [9, Section 2].

The first is that every column with a nonzero sum contributes at least 2 to the length of the horizontal boundary, while every row with nonzero sum contributes at least 2 to the length of the vertical boundary. So if there are m nonzero row sums and n nonzero column sums, then the total length of the boundary is at least $2n + 2m$.

For the second bound we use that if the row sums of two consecutive rows are different, then the length of the horizontal boundary between those rows is at least the absolute difference between those row sums. A similar result holds for the column sums and the vertical boundary. So if an image has row sums $r_1, r_2, \ldots, r_m$ and column sums $c_1, c_2, \ldots, c_n$, then the length of the boundary is at least

$$r_1 + \sum_{i=1}^{m-1} |r_i - r_{i+1}| + r_m + c_1 + \sum_{j=1}^{n-1} |c_j - c_{j+1}| + c_n.$$

Despite being simple, these bounds are sharp in many cases. For example, the first bound is sharp if and only if there exists a hv-convex image that satisfies the line sums. On the other hand it is clear that much information is disregarded in these bounds. The first bound does not use the actual value of the nonzero line sums at all, while the second bound only uses the column sums to estimate the length of the vertical boundary and only the row sums to estimate the length of the horizontal boundary.

In this chapter we prove a new lower bound on the length of the boundary that combines the row and column sums. After introducing some notation in Section 5.2, we prove this bound in Section 5.3. Some examples and a corollary are in Section 5.4. Finally, in Section 5.5 we derive an extension of the bound that gives better results in certain cases.

5.2 Definitions and notation

Let F be a finite subset of $\mathbb{Z}^2$ with characteristic function χ. (That is, $\chi(k,l) = 1$ if $(k,l) \in F$ and $\chi(k,l) = 0$ otherwise.) For $i \in \mathbb{Z}$, we define *row* i as the set $\{(k,l) \in \mathbb{Z}^2 : k = i\}$. We call i the index of the row. For $j \in \mathbb{Z}$, we define *column* j as the set $\{(k,l) \in \mathbb{Z}^2 : l = j\}$. We call j the index of the column. Note that we follow matrix notation: we indicate a point (i,j) by first its row index i and then its column index j. Also, we use row numbers that increase when going downwards and column numbers that increase when going to the right.

The *row sum* r_i is the number of elements of F in row i, that is $r_i = \sum_{j\in\mathbb{Z}} \chi(i,j)$. The *column sum* c_j of F is the number of elements of F in column j, that is $c_j = \sum_{i\in\mathbb{Z}} \chi(i,j)$. We refer to both row and column sums as the *line sums* of F. We will usually only consider finite sequences $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and $\mathcal{C} = (c_1, c_2, \ldots, c_n)$ of row and column sums that contain all the nonzero line sums.

Given sequences of integers $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and $\mathcal{C} = (c_1, c_2, \ldots, c_n)$ with $0 \le r_i \le n$, $0 \le c_j \le m$, we say that $(\mathcal{R}, \mathcal{C})$ is *consistent* if there exists a set F with row sums $\mathcal{R}$ and column sums $\mathcal{C}$. Define $b_i = \#\{j : c_j \ge i\}$ for $i = 1, 2, \ldots, m$. Note that by definition we have $\sum_{i=1}^{m} b_i = \sum_{j=1}^{n} c_j$. Ryser's theorem [24] states that if $r_1 \ge r_2 \ge \ldots \ge r_m$, then the line sums $(\mathcal{R}, \mathcal{C})$ are consistent if and only if $\sum_{j=1}^{n} c_j = \sum_{i=1}^{m} r_i$ and for each $k = 1, 2, \ldots, m$ we have $\sum_{i=1}^{k} b_i \ge \sum_{i=1}^{k} r_i$. From this we can conclude a similar result for the case of not necessarily non-increasing row sums: if the line sums $(\mathcal{R}, \mathcal{C})$ are consistent, then $\sum_{j=1}^{n} c_j = \sum_{i=1}^{m} r_i$ and for each $k = 1, 2, \ldots, m$ we have

$$\sum_{i=1}^{k} b_i \ge \sum_{i=1}^{k} r_i. \tag{5.1}$$

The converse clearly does not hold.

We can view the set F as a picture consisting of cells with zeroes and ones. Rather than $(i,j) \in F$, we might say that (i,j) has value 1 or that there is a one at (i,j). Similarly, for $(i,j) \notin F$ we sometimes say that (i,j) has value zero or that there is a zero at (i,j).

We define the *boundary* of F as the set consisting of all pairs of points $\big((i,j),(i',j')\big)$ such that

- $i = i'$ and $|j - j'| = 1$, or $|i - i'| = 1$ and $j = j'$, and
- $(i,j) \in F$ and $(i',j') \notin F$.

One element of this set we call *one piece of the boundary*. We can partition the

boundary into two subsets, one containing the pairs of points with $i = i'$ and the other containing the pairs of points with $j = j'$. The former set we call the *vertical boundary* and the latter set we call the *horizontal boundary*. We define the *length of the (horizontal, vertical) boundary* as the number of elements in the (horizontal, vertical) boundary.

5.3 The main theorem

Theorem 5.1. *Let be given row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$, where $r_1 = n$, $r_m = 0$. Let L_h be the total length of the horizontal boundary of an image with line sums $(\mathcal{R}, \mathcal{C})$. Define $b_i = \#\{j : c_j \geq i\}$ and $d_i = b_i - r_i$ for $i = 1, 2, \ldots, m$. For any integer $t \geq 0$ and any subset $\{i_1, i_2, \ldots, i_{2t+1}\} \subset \{1, 2, \ldots, m\}$ with $i_1 < i_2 < \ldots < i_{2t+1}$ we have*

$$L_h \geq 2n + d_{i_1} - d_{i_2} + d_{i_3} - \cdots - d_{i_{2t}} + 2d_{i_{2t+1}}, \tag{5.2}$$

$$L_h \geq 2n - d_{i_{2t+1}} + d_{i_{2t}} - d_{i_{2t-1}} + \cdots + d_{i_2} - 2d_{i_1}. \tag{5.3}$$

Proof. First we prove (5.2) by induction on n. In the initial case $n = 0$ we have $d_i = b_i = r_i = 0$ for all i, hence we have to prove that $L_h \geq 0$, which is obviously true.

Now let $n \geq 1$ and consider a binary image F with line sums $(\mathcal{R}, \mathcal{C})$. Let $I \subset \{1, 2, \ldots, m\}$ be the set of indices i such that cell (i, n) has value 1. Note that $\#I = c_n$. Let F' be the binary image we obtain by deleting column n from F. Let $(r'_1, r'_2, \ldots, r'_m)$ be the row sums of F'. The column sums of F' are $(c_1, c_2, \ldots, c_{n-1})$, and define $b'_i = \#\{j \leq n-1 : c_j \geq i\}$ and $d'_i = b'_i - r'_i$ for $i = 1, 2, \ldots, m$. We have

$$r'_i = \begin{cases} r_i & \text{if } i \notin I, \\ r_i - 1 & \text{if } i \in I, \end{cases}$$

$$b'_i = \begin{cases} b_i - 1 & \text{if } i \leq c_n, \\ b_i & \text{if } i > c_n, \end{cases}$$

and therefore

$$d'_i = \begin{cases} d_i - 1 & \text{if } i \notin I \text{ and } i \leq c_n, \\ d_i & \text{if } i \notin I \text{ and } i > c_n, \text{ or } i \in I \text{ and } i \leq c_n, \\ d_i + 1 & \text{if } i \in I \text{ and } i > c_n. \end{cases}$$

As induction hypothesis we assume that (5.2) is true for the smaller image F'. So for the total length L'_h of the horizontal boundary of F' we have

$$L'_h \geq 2(n-1) + d'_{i_1} - d'_{i_2} + d'_{i_3} - \cdots - d'_{i_{2t}} + 2d'_{i_{2t+1}}.$$

Let $2B$ be equal to the horizontal boundary in column n of F. Then $L_h = L'_h + 2B$. We want to prove (5.2), hence it suffices to prove

$$2B-2 \geq (d_{i_1}-d'_{i_1})-(d_{i_2}-d'_{i_2})+(d_{i_3}-d'_{i_3})-\cdots-(d_{i_{2t}}-d'_{i_{2t}})+2(d_{i_{2t+1}}-d'_{i_{2t+1}}). \quad (5.4)$$

Write the right-hand side as

$$\sum_{s=1}^{t}\Big((d_{i_{2s-1}} - d'_{i_{2s-1}}) - (d_{i_{2s}} - d'_{i_{2s}})\Big) + 2(d_{i_{2t+1}} - d'_{i_{2t+1}}).$$

Note that

$$d_i - d'_i = \begin{cases} 1 & \text{if } i \notin I \text{ and } i \leq c_n, \\ 0 & \text{if } i \notin I \text{ and } i > c_n, \text{ or } i \in I \text{ and } i \leq c_n, \\ -1 & \text{if } i \in I \text{ and } i > c_n. \end{cases}$$

The only possible values of $(d_{i_{2s-1}} - d'_{i_{2s-1}}) - (d_{i_{2s}} - d'_{i_{2s}})$ are therefore -1, 0, 1 and 2. If we have $i_{2s-1}, i_{2s} \leq c_n$ or $i_{2s-1}, i_{2s} > c_n$, then the value 2 is not possible and

$$(d_{i_{2s-1}} - d'_{i_{2s-1}}) - (d_{i_{2s}} - d'_{i_{2s}}) = 1 \qquad \Leftrightarrow \qquad i_{2s-1} \notin I \text{ and } i_{2s} \in I.$$

Furthermore note that of the $2B$ pieces of horizontal boundary in column n, one is above row 1 (as $r_1 = n$, so $1 \in I$) and exactly $B-1$ are between a pair of cells with row indices i and $i+1$, such that $i \notin I$ and $i+1 \in I$. We now distinguish between four cases.

Case 1. Suppose $i_{2t+1} \leq c_n$ and $i_{2t+1} \notin I$. Then $2(d_{i_{2t+1}} - d'_{i_{2t+1}}) = 2$. In the first c_n cells of column n, there is at least one cell (the one with row index i_{2t+1}) that has value 0, hence $B \geq 2$ and there is a cell with row index greater than i_{2t+1} with value 1. This means that there are at most $B-2$ pairs (i_{2s-1}, i_{2s}) such that $i_{2s-1} \notin I$ and $i_{2s} \in I$. Also, $i_{2s-1}, i_{2s} \leq c_n$ for all s. So

$$\sum_{s=1}^{t}\Big((d_{i_{2s-1}} - d'_{i_{2s-1}}) - (d_{i_{2s}} - d'_{i_{2s}})\Big)+2(d_{i_{2t+1}}-d'_{i_{2t+1}}) \leq (B-2)+2 = B \leq 2B-2.$$

Case 2. Suppose $i_{2t+1} \leq c_n$ and $i_{2t+1} \in I$. Then $2(d_{i_{2t+1}} - d'_{i_{2t+1}}) = 0$. Now there are at most $B-1$ pairs (i_{2s-1}, i_{2s}) such that $i_{2s-1} \notin I$ and $i_{2s} \in I$. Also, $i_{2s-1}, i_{2s} \leq c_n$ for all s. So

$$\sum_{s=1}^{t}\Big((d_{i_{2s-1}} - d'_{i_{2s-1}}) - (d_{i_{2s}} - d'_{i_{2s}})\Big) + 2(d_{i_{2t+1}} - d'_{i_{2t+1}}) \leq B-1 \leq 2B-2.$$

Case 3. Suppose $i_{2t+1} > c_n$ and $B \geq 2$. Then $2(d_{i_{2t+1}} - d'_{i_{2t+1}}) \leq 0$. Again there are at most $B-1$ pairs (i_{2s-1}, i_{2s}) such that $i_{2s-1} \notin I$ and $i_{2s} \in I$. If there does not

exist an u such that $i_{2u-1} \leq c_n$ and $i_{2u} > c_n$, then we are done, as in the previous case. If there does exist such an u, then

$$(d_{i_{2u-1}} - d'_{i_{2u-1}}) - (d_{i_{2u}} - d'_{i_{2u}}) = 2 \qquad \Leftrightarrow \qquad i_{2u-1} \notin I \text{ and } i_{2u} \in I.$$

If $(d_{i_{2u-1}} - d'_{i_{2u-1}}) - (d_{i_{2u}} - d'_{i_{2u}}) = 2$, then on the right-hand side of (5.4) we have a 2 and at most $B-2$ times a 1. If not, then we have no 2 and at most B times a 1. In both cases we find

$$\sum_{s=1}^{t} \Big((d_{i_{2s-1}} - d'_{i_{2s-1}}) - (d_{i_{2s}} - d'_{i_{2s}}) \Big) + 2(d_{i_{2t+1}} - d'_{i_{2t+1}}) \leq B \leq 2B - 2.$$

Case 4. Suppose $B = 1$. Then $i \in I \Leftrightarrow i \leq c_n$, hence $d'_i = d_i$ for all i. Therefore

$$\sum_{s=1}^{t} \Big((d_{i_{2s-1}} - d'_{i_{2s-1}}) - (d_{i_{2s}} - d'_{i_{2s}}) \Big) + 2(d_{i_{2t+1}} - d'_{i_{2t+1}}) = 0 = 2B - 2.$$

In all possible cases we have now proved inequality (5.4), which finishes the proof of (5.2).

Now we prove (5.3). Let F be a binary $m \times n$ image with row sums $\mathcal{R}$ and column sums $\mathcal{C}$. Define $\bar{F}$ as the binary $m \times n$ image that has zeroes where F has ones and ones where F has zeroes. Let $(\bar{r}_1, \ldots, \bar{r}_m)$ be the row sums of $\bar{F}$ and $(\bar{c}_1, \ldots, \bar{c}_n)$ the column sums. Define $\bar{b}_i = \#\{j : \bar{c}_j \geq i\}$ and $\bar{d}_i = \bar{b}_i - \bar{r}_{m+1-i}$ for $i = 1, 2, \ldots, m$. As $\bar{r}_i = n - r_i$ and $\bar{c}_j = m - c_j$ for all i and j, we have

$$\bar{b}_i = \#\{j : m - c_j \geq i\} = \#\{j : c_j \leq m - i\} = n - \#\{j : c_j \geq m + 1 - i\} = n - b_{m+1-i}.$$

Hence

$$\bar{d}_i = \bar{b}_i - \bar{r}_{m+1-i} = n - b_{m+1-i} - n + r_{m+1-i} = -d_{m+1-i}.$$

As $\bar{r}_1 = 0$ and $\bar{r}_m = n$, we may apply (5.2) to the row sums $(\bar{r}_m, \bar{r}_{m-1}, \ldots, \bar{r}_1)$. We write the subset of the row indices we use as $(m+1-i_{2t+1}, m+1-i_{2t}, \ldots, m+1-i_1)$ with $i_1 < i_2 < \ldots < i_{2t+1}$. We find that for the total length $\bar{L}_h$ of the horizontal boundary of $\bar{F}$ holds:

$$\begin{aligned} \bar{L}_h &\geq 2n + \bar{d}_{m+1-i_{2t+1}} - \bar{d}_{m+1-i_{2t}} + \bar{d}_{m+1-i_{2t-1}} - \cdots - \bar{d}_{m+1-i_2} + 2\bar{d}_{m+1-i_1} \\ &= 2n - d_{i_{2t+1}} + d_{i_{2t}} - d_{i_{2t-1}} + \cdots + d_{i_2} - 2d_{i_1}. \end{aligned}$$

In each column of $\bar{F}$, the number of horizontal pieces of boundary is equal to the number of pairs of neighbouring cells such that one cell has value 1 and the other has value 0, plus one for the boundary below row m. In each column of F, the number of horizontal pieces of boundary is equal to the number of pairs of neighbouring cells such that one cell has value 1 and the other has value 0, plus one for the boundary

above row 1. As in each column the number of pairs of neighbouring cells such that one cell has value 1 and the other has value 0, is the same in F and in $\bar{F}$, we have $\bar{L}_h = L_h$. Hence

$$L_h \geq 2n - d_{i_{2t+1}} + d_{i_{2t}} - d_{i_{2t-1}} + \cdots + d_{i_2} - 2d_{i_1}.$$

□

5.4 Some examples and a corollary

To illustrate Theorem 5.1, we apply it to two small examples.

Example 5.1. Let $m = n = 10$ and let row sums $(10, 7, 7, 5, 4, 3, 5, 6, 1, 0)$ and column sums $(8, 8, 8, 8, 6, 3, 2, 2, 2, 1)$ be given. We compute b_i and d_i, $i = 1, 2, \ldots, 10$ as shown below.

i	1	2	3	4	5	6	7	8	9	10
b_i	10	9	6	5	5	5	4	4	0	0
r_i	10	7	7	5	4	3	5	6	1	0
d_i	0	+2	−1	0	+1	+2	−1	−2	−1	0

We take $t = 1$, $i_1 = 2$, $i_2 = 3$ and $i_3 = 6$. Now (5.2) tells us that

$$L_h \geq 20 + 2 - (-1) + 2 \cdot 2 = 27.$$

Alternatively, we take $t = 2$, $i_1 = 2$, $i_2 = 3$, $i_3 = 6$, $i_4 = 8$ and $i_5 = 10$. Now (5.2) tells us that

$$L_h \geq 20 + 2 - (-1) + 2 - (-2) + 2 \cdot 0 = 27.$$

As L_h must be even, we conclude $L_h \geq 28$. This bound is sharp: in Figure 5.1(a) a binary image F with the given row and column sums is shown, for which $L_h = 28$.

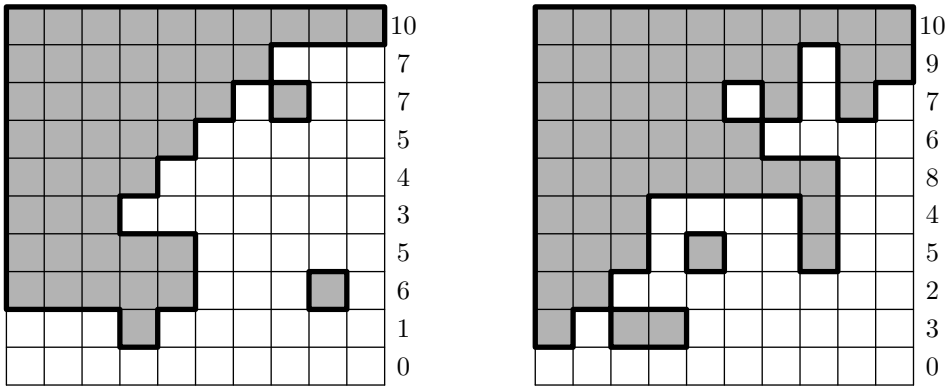
Example 5.2. Let $m = n = 10$ and let row sums $(10, 9, 7, 6, 8, 4, 5, 2, 3, 0)$ and column sums $(9, 8, 8, 6, 6, 4, 4, 4, 3, 2)$ be given. We compute b_i and d_i, $i = 1, 2, \ldots, 10$ as shown below.

i	1	2	3	4	5	6	7	8	9	10
b_i	10	10	9	8	5	5	3	3	1	0
r_i	10	9	7	6	8	4	5	2	3	0
d_i	0	+1	+2	+2	−3	+1	−2	+1	−2	0

We take $t = 2$, $i_1 = 5$, $i_2 = 6$, $i_3 = 7$, $i_4 = 8$ and $i_5 = 9$. Now (5.3) tells us that

$$L_h \geq 20 - (-2) + 1 - (-2) + 1 - 2 \cdot (-3) = 32.$$

This bound is sharp: in Figure 5.1(b) a binary image F with the given row and column sums is shown, for which $L_h = 32$.



(a) The length of the horizontal boundary of this image is 28.

(b) The length of the horizontal boundary of this image is 32.

Figure 5.1: The binary images from Examples 5.1 and 5.2. The grey cells have value 1, the other cells value 0. The numbers indicate the row and column sums.

In the Introduction we mentioned two simple bounds of the length of the boundary. We recall them here, just for the horizontal boundary. The first one uses that in every column, there are at least two pieces of boundary, so if there are n columns with nonzero sums, then

$$L_h \geq 2n. \tag{5.5}$$

The other bound computes the sum of the absolute differences between consecutive row sums, which yields

$$L_h \geq r_1 + \sum_{i=1}^{m-1} |r_i - r_{i+1}| + r_m. \tag{5.6}$$

In order to compare the bounds in Theorem 5.1 to these two simple bounds, we construct two families of examples.

Example 5.3. Let the number of columns n be even. Let $m = n + 2$. Define line sums

$$\mathcal{C} = (n, n, n-2, n-2, \ldots, 4, 4, 2, 2), \quad \mathcal{R} = (n, n-1, n-1, n-3, n-3, \ldots, 3, 3, 1, 1, 0).$$

We calculate

$$(b_1, b_2, \ldots, b_m) = (n, n, n-2, n-2, \ldots, 2, 2, 0, 0),$$

$$(d_1, d_2, \ldots, d_m) = (0, +1, -1, +1, -1, \ldots, +1, -1, +1, -1, 0).$$

Now (5.2) tells us that

$$L_h \geq 2n + \frac{n}{2} \cdot (1 - -1) + 2 \cdot 0 = 3n.$$

On the other hand, (5.5) says $L_h \geq 2n$, while (5.6) gives

$$L_h \geq n + 1 + \frac{n-2}{2} \cdot 2 + 1 = 2n.$$

So Theorem 5.1 gives a much better bound in this family of examples. In fact, it is sharp: there exists a binary image with the length of the boundary equal to $3n$. Such an image is easy to construct; see for an example Figure 5.2(a).

Example 5.4. Let $m = n + 2$. Define line sums

$$\mathcal{C} = (2, 2, 2, \ldots, 2, 2, 2), \quad \mathcal{R} = (n, 1, 1, 1, \ldots, 1, 1, 1, 0).$$

We calculate

$$(b_1, b_2, \ldots, b_m) = (n, n, 0, 0, 0, \ldots, 0, 0, 0),$$

$$(d_1, d_2, \ldots, d_m) = (0, +(n-1), -1, -1, -1, \ldots, -1, -1, -1, 0).$$

Now (5.2) tells us that

$$L_h \geq 2n + 2 \cdot (n-1) = 4n - 2.$$

On the other hand, (5.5) says $L_h \geq 2n$, while (5.6) gives

$$L_h \geq n + (n-1) + 1 = 2n.$$

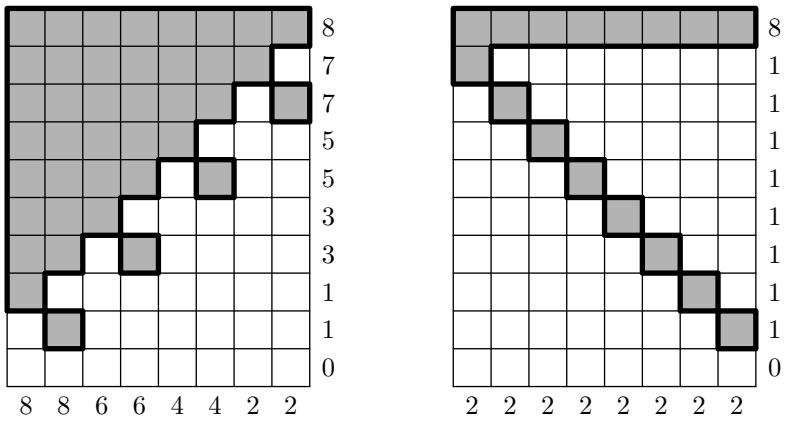
So again Theorem 5.1 gives a much better bound. In fact, it is sharp: there exists a binary image with the length of the boundary equal to $4n - 2$. Such an image is easy to construct; see for an example Figure 5.2(b).

We can easily generalise the result from Theorem 5.1 to the case where the conditions $r_1 = n$ and $r_m = 0$ are not satisfied.

Corollary 5.2. *Let be given row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$. Let L_h be the total length of the horizontal boundary of an image with line sums $(\mathcal{R}, \mathcal{C})$. Define $b_i = \#\{j : c_j \geq i\}$ and $d_i = b_i - r_i$ for $i = 1, 2, \ldots, m$. Also set $d_0 = d_{m+1} = 0$. For any integer $t \geq 0$ and any subset $\{i_1, i_2, \ldots, i_{2t+1}\} \subset \{0, 1, 2, \ldots, m, m+1\}$ with $i_1 < i_2 < \ldots < i_{2t+1}$ we have*

$$L_h \geq 2r_1 + d_{i_1} - d_{i_2} + d_{i_3} - \cdots - d_{i_{2t}} + 2d_{i_{2t+1}}, \tag{5.7}$$

$$L_h \geq 2r_1 - d_{i_{2t+1}} + d_{i_{2t}} - d_{i_{2t-1}} + \cdots + d_{i_2} - 2d_{i_1}. \tag{5.8}$$



(a) The length of the horizontal boundary of this image is $24 = 3n$.

(b) The length of the horizontal boundary of this image is $30 = 4n - 2$.

Figure 5.2: Binary images from Examples 5.3 and 5.4, with $n = 8$. The grey cells have value 1, the other cells value 0. The numbers indicate the row and column sums.

Proof. Let F be a binary image with line sums $(\mathcal{R}, \mathcal{C})$ and a horizontal boundary of total length L_h. Construct F' by adding a row above row 1 with row sum n and a row below row m with row sum 0. Let L'_h be the length of the horizontal boundary of F'. We have $L'_h = L_h + 2(n - r_1)$. The column sums of F' are $c'_j = c_j + 1$, $j = 1, 2, \ldots, n$. The row sums are $r'_1 = n$, $r'_i = r_{i-1}$ for $i = 2, 3, \ldots, m + 1$ and $r'_{m+2} = 0$. Let $b'_i = \#\{j : c'_j \geq i\}$ and $d'_i = b'_i - r'_i$ for $i = 1, 2, \ldots, m$. Then for all $i = 2, 3, \ldots, m + 1$ we have

$$b'_i = \#\{j : c_j + 1 \geq i\} = \#\{j : c_j \geq i - 1\} = b_{i-1},$$

so $d'_i = b_{i-1} - r_{i-1} = d_{i-1}$. Also, $d'_1 = d_0 = 0$ and $d'_{m+2} = d_{m+1} = 0$. We apply Theorem 5.1 to F' with the set of indices $\{i_1 + 1, i_2 + 1, \ldots, i_{2t+1} + 1\}$ and we find

$$\begin{aligned} L'_h &\geq 2n + d'_{i_1+1} - d'_{i_2+1} + d'_{i_3+1} - \cdots - d'_{i_{2t}+1} + 2d'_{i_{2t+1}+1} \\ &= 2n + d_{i_1} - d_{i_2} + d_{i_3} - \cdots - d_{i_{2t}} + 2d_{i_{2t+1}}, \\ L'_h &\geq 2n - d'_{i_{2t+1}+1} + d'_{i_{2t}+1} - d'_{i_{2t-1}+1} + \cdots + d'_{i_2+1} - 2d'_{i_1+1} \\ &= 2n - d_{i_{2t+1}} + d_{i_{2t}} - d_{i_{2t-1}} + \cdots + d_{i_2} - 2d_{i_1}, \end{aligned}$$

and therefore

$$\begin{aligned} L_h &\geq 2r_1 + d_{i_1} - d_{i_2} + d_{i_3} - \cdots - d_{i_{2t}} + 2d_{i_{2t+1}}, \\ L_h &\geq 2r_1 - d_{i_{2t+1}} + d_{i_{2t}} - d_{i_{2t-1}} + \cdots + d_{i_2} - 2d_{i_1}. \end{aligned}$$

□

5.5 An extension

Theorem 5.3. *Let be given row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$, where $r_1 = n$, $r_m = 0$. Suppose there exists an image F with line sums $(\mathcal{R}, \mathcal{C})$ and let $L_h(F)$ be the total length of the horizontal boundary of this image. Define $b_i = \#\{j : c_j \geq i\}$ and $d_i = b_i - r_i$ for $i = 1, 2, \ldots, m$. Let k be an integer with $2 \leq k \leq m-1$ such that $d_k < 0$ and $d_{k+1} \geq 0$. Let $\sigma = \sum_{i=1}^{k} d_i$. For any integers $t, s \geq 0$ and any sets $\{i_1, i_2, \ldots, i_{2t+1}\} \subset \{1, 2, \ldots, k-1, k, m\}$ with $i_1 < i_2 < \ldots < i_{2t+1}$ and $\{\tilde{i}_1, \tilde{i}_2, \ldots, \tilde{i}_{2s+1}\} \subset \{1, k+1, k+2, \ldots, m-1, m\}$ with $\tilde{i}_1 < \tilde{i}_2 < \ldots < \tilde{i}_{2s+1}$ we have*

$$\begin{aligned} L_h(F) \geq 2n + d_{i_1} - d_{i_2} + d_{i_3} - \cdots - d_{i_{2t}} + 2d_{i_{2t+1}} \\ + d_{\tilde{i}_1} - d_{\tilde{i}_2} + d_{\tilde{i}_3} - \cdots - d_{\tilde{i}_{2s}} + 2d_{\tilde{i}_{2s+1}} - \sigma. \end{aligned} \tag{5.9}$$

Proof. We will prove the theorem by induction on σ. Note that by (5.1) we have $\sigma \geq 0$, since the line sums are consistent.

As we are only considering the horizontal boundary, we may for convenience assume that $c_1 \geq c_2 \geq \ldots \geq c_n$.

Suppose $\sigma = 0$. Then

$$\sum_{i=1}^{k} r_i = \sum_{i=1}^{k} b_i = \sum_{i=1}^{k} \#\{j : c_j \geq i\} = \sum_{j | c_j \leq k} c_j + \sum_{j | c_j > k} k.$$

So in any column j with $c_j > k$ we must have $(i, j) \in F$ for $1 \leq i \leq k$, and in any column j with $c_j \leq k$ we must have $(i, j) \notin F$ for $k+1 \leq i \leq m$. This means that we can split the image F into four smaller images, one of which contains only ones and one of which contains only zeroes. The other two parts we call F_1 and F_2 (see Figure 5.3). In order to have images with the first row filled with ones and the last row filled with zeroes, we glue row m to F_1 and row 1 to F_2. More precisely, let F_1 consist of rows $1, 2, \ldots, k-1, k$ and m of F and the columns j with $c_j \leq k$; let F_2 consist of rows 1 and $k+1, k+2, \ldots, m-1, m$ of F and the columns j with $c_j > k$.

The columns of F with sum at most k are exactly the columns with indices greater than b_{k+1}. Define $h = b_{k+1}$. Let $r_1^{(1)}, r_2^{(1)}, \ldots, r_k^{(1)}, r_m^{(1)}$ be the row sums of F_1, and let $r_1^{(2)}, r_{k+1}^{(2)}, \ldots, r_{m-1}^{(2)}, r_m^{(2)}$ be the row sums of F_2. We have

$$r_i^{(1)} = r_i - h, \quad \text{for } 1 \leq i \leq k, \text{ and } \quad r_m^{(1)} = r_m,$$

$$r_i^{(2)} = r_i \quad \text{for } k+1 \leq i \leq m, \text{ and } \quad r_1^{(2)} = h = r_1 - (n - h).$$

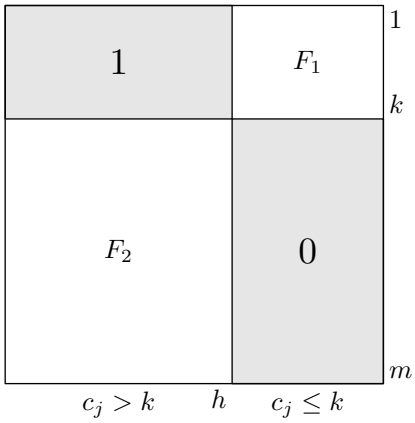


Figure 5.3: Splitting the image F into four smaller images.

Let $c_{h+1}^{(1)}, c_{h+2}^{(1)}, \ldots, c_{n-1}^{(1)}, c_n^{(1)}$ be the column sums of F_1, and let $c_1^{(2)}, c_2^{(2)}, \ldots, c_{h-1}^{(2)}, c_h^{(2)}$ be the column sums of F_2. We have

$$c_j^{(1)} = c_j, \quad \text{and} \quad c_j^{(2)} = c_j - (k-1) \quad \text{for all } j.$$

Define

$$\begin{aligned}
b_1^{(1)} &= \#\{j \ge h+1 : c_j^{(1)} \ge 1\}, & b_1^{(2)} &= \#\{j \le h : c_j^{(2)} \ge 1\},\\
b_2^{(1)} &= \#\{j \ge h+1 : c_j^{(1)} \ge 2\}, & b_{k+1}^{(2)} &= \#\{j \le h : c_j^{(2)} \ge 2\},\\
&\vdots & &\vdots\\
b_k^{(1)} &= \#\{j \ge h+1 : c_j^{(1)} \ge k\}, & b_{m-1}^{(2)} &= \#\{j \le h : c_j^{(2)} \ge m-k\},\\
b_m^{(1)} &= \#\{j \ge h+1 : c_j^{(1)} \ge k+1\}, & b_m^{(2)} &= \#\{j \le h : c_j^{(2)} \ge m-k+1\}.
\end{aligned}$$

For $1 \le i \le k$ we have

$$b_i^{(1)} = \#\{j \ge h+1 : c_j^{(1)} \ge i\} = \#\{j \le n : c_j \ge i\} - \#\{j \le h : c_j \ge i\} = b_i - h.$$

Also, $b_m^{(1)} = 0 = b_m$. For $k+1 \le i \le m$ we have

$$\begin{aligned}
b_i^{(2)} &= \#\{j \le h : c_j^{(2)} \ge i-k+1\} = \#\{j \le h : c_j \ge i\}\\
&= \#\{j \le n : c_j \ge i\} - \#\{j \ge h+1 : c_j \ge i\} = b_i - 0 = b_i.
\end{aligned}$$

Also, $b_1^{(2)} = h = b_1 - (n-h)$. Now define $d_i^{(1)} = b_i^{(1)} - r_i^{(1)}$ for $i \in \{1, 2, \ldots, k-1, k, m\}$ and $d_i^{(2)} = b_i^{(2)} - r_i^{(2)}$ for $i \in \{1, k+1, k+2, \ldots, m-1, m\}$. We find

$$d_i^{(1)} = b_i - h - (r_i - h) = d_i, \quad \text{for } 1 \le i \le k,$$

$$d_m^{(1)} = b_m - r_m = d_m,$$
$$d_i^{(2)} = b_i - r_i = d_i \quad \text{for } k+1 \le i \le m$$
$$d_1^{(2)} = b_1 - (n-h) - (r_1 - (n-h)) = d_1.$$

All in all we conclude $d_i^{(1)} = d_i$ and $d_i^{(2)} = d_i$ for all i.

The total length of the horizontal boundary of F in the columns j with $c_j \le k$ is exactly the same as the total length $L_h(F_1)$ of the horizontal boundary of F_1. The total length of the horizontal boundary of F in the columns j with $c_j > k$ is exactly the same as the total length $L_h(F_2)$ of the horizontal boundary of F_2. So $L_h(F) = L_h(F_1) + L_h(F_2)$. Note that F_1 has $n - b_{k+1}$ columns and F_2 has b_{k+1} columns. By Theorem 5.1 applied to F_1 we know that for any integer $t \ge 0$ and any set $\{i_1, i_2, \ldots, i_{2t+1}\} \subset \{1, 2, \ldots, k-1, k, m\}$ with $i_1 < i_2 < \ldots < i_{2t+1}$ we have

$$L_h(F_1) \ge 2(n - b_{k+1}) + d_{i_1} - d_{i_2} + d_{i_3} - \cdots - d_{i_{2t}} + 2d_{i_{2t+1}}.$$

By the same theorem applied to F_2 we know that for any integer $t \ge 0$ and any set $\{\tilde{i}_1, \tilde{i}_2, \ldots, \tilde{i}_{2s+1}\} \subset \{1, k+1, k+2, \ldots, m-1, m\}$ with $\tilde{i}_1 < \tilde{i}_2 < \ldots < \tilde{i}_{2s+1}$ we have

$$L_h(F_2) \ge 2b_{k+1} + d_{\tilde{i}_1} - d_{\tilde{i}_2} + d_{\tilde{i}_3} - \cdots - d_{\tilde{i}_{2s}} + 2d_{\tilde{i}_{2s+1}}.$$

Adding these two results yields (5.9).

Now let $\sigma \ge 1$ and suppose that we have already proven the theorem for any image with $\sum_{i=1}^{k} d_i < \sigma$. Let

$$A_1 = \max\{d_{i_1} - d_{i_2} + d_{i_3} - \cdots - d_{i_{2t}} + 2d_{i_{2t+1}}\},$$
$$A_2 = \max\{d_{\tilde{i}_1} - d_{\tilde{i}_2} + d_{\tilde{i}_3} - \cdots - d_{\tilde{i}_{2s}} + 2d_{\tilde{i}_{2s+1}}\},$$

where the first maximum is taken over all integers $t \ge 0$ and sets $\{i_1, i_2, \ldots, i_{2t+1}\} \subset \{1, 2, \ldots, k-1, k, m\}$ with $i_1 < i_2 < \ldots < i_{2t+1}$, and the second maximum over all integers $s \ge 0$ and sets $\{\tilde{i}_1, \tilde{i}_2, \ldots, \tilde{i}_{2s+1}\} \subset \{1, k+1, k+2, \ldots, m-1, m\}$ with $\tilde{i}_1 < \tilde{i}_2 < \ldots < \tilde{i}_{2s+1}$. Furthermore, fix $i_1, i_2, \ldots, i_{2t+1}$ and $\tilde{i}_1, \tilde{i}_2, \ldots, \tilde{i}_{2s+1}$ such that these maxima are attained.

Since $d_k < 0$ by definition of k, and since $d_m = 0$, we have

$$d_{i_1} - d_{i_2} + d_{i_3} - \cdots - d_{i_{2t}} + 2d_k < d_{i_1} - d_{i_2} + d_{i_3} - \cdots - d_{i_{2t}} + 2d_m.$$

If $i_{2t+1} = k$, this would contradict the maximality of A_1, so we conclude

$$i_{2t+1} \ne k. \tag{5.10}$$

We also know $d_{k+1} \ge 0$ by definition of k, and $d_1 = 0$. So if $s \ge 1$, then

$$d_1 - d_{k+1} + d_{\tilde{i}_3} - \cdots - d_{\tilde{i}_{2s}} + 2d_{\tilde{i}_{2s+1}} \le d_{\tilde{i}_3} - \cdots - d_{\tilde{i}_{2s}} + 2d_{\tilde{i}_{2s+1}}.$$

This means that if $s \geq 1$, we may assume without loss of generality that $(\tilde{i}_1, \tilde{i}_2) \neq (1, k+1)$. Also,

$$d_1 - d_{\tilde{i}_2} + d_{\tilde{i}_3} - \cdots - d_{\tilde{i}_{2s}} + 2d_{\tilde{i}_{2s+1}} \leq d_{k+1} - d_{\tilde{i}_2} + d_{\tilde{i}_3} - \cdots - d_{\tilde{i}_{2s}} + 2d_{\tilde{i}_{2s+1}}.$$

This means that if $s \geq 1$ and $\tilde{i}_2 > k+1$, we may assume that $\tilde{i}_1 \neq 1$. Finally, $2d_1 \leq 2d_{k+1}$, so if $s = 1$ we may also assume that $\tilde{i}_1 \neq 1$.

All in all we may assume in all cases that

$$\tilde{i}_1 \neq 1. \tag{5.11}$$

It suffices to prove

$$L_h(F) \geq 2n + A_1 + A_2 - \sigma. \tag{5.12}$$

Let j with $1 \leq j \leq n$ be such that $\#\big(\{(1,j),(2,j),\ldots,(k,j)\} \cap F\big) < \min(c_j, k)$, i.e. in column j there is at least one one in rows $k+1, k+2, \ldots, m$ and at least one zero in rows $1, 2, \ldots, k$. Such a column exists, because

$$\sum_{i=1}^{k} r_i < \sum_{i=1}^{k} b_i = \sum_{i=1}^{k} \#\{j : c_j \geq i\} = \sum_{j|c_j \leq k} c_j + \sum_{j|c_j > k} k.$$

We will now consider various cases.

Case 1. Suppose that there exist integers $l \geq 2$, $h \geq k+1$ and $u \geq 0$ such that $l + u \leq k$, $h + u \leq m - 1$ and

- $(l-1, j) \in F$, and
- $(l, j), (l+1, j), \ldots, (l+u, j) \notin F$, and
- $(h, j), (h+1, j), \ldots, (h+u, j) \in F$, and
- $(h+u+1, j) \notin F$, and
- $(l+u+1, j) \in F$ or $(h-1, j) \notin F$.

We define a new image F' by moving the ones at $(h, j), (h+1, j), \ldots, (h+u, j)$ to $(l, j), (l+1, j), \ldots, (l+u, j)$; that is,

$$F' = F \cup \{(l,j),(l+1,j),\ldots,(l+u,j)\} \backslash \{(h,j),(h+1,j),\ldots,(h+u,j)\}.$$

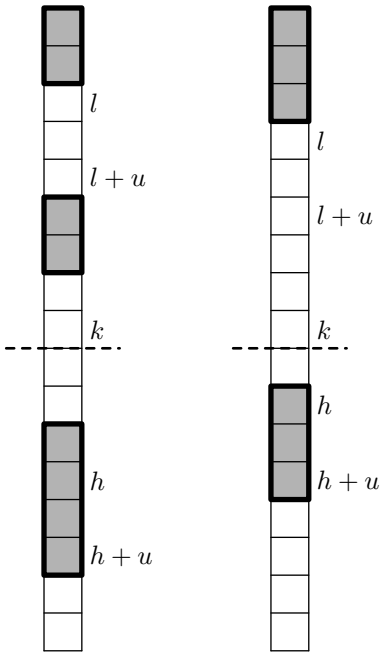


Figure 5.4: Two possibilities for column j in Case 1. The grey cells have value 1, the other cells value 0.

The column sums of F' are identical to the column sums of F. The row sums r'_i of F' are given by

$$r'_i = \begin{cases} r_i + 1 & \text{if } l \le i \le l + u, \\ r_i - 1 & \text{if } h \le i \le h + u, \\ r_i & \text{else.} \end{cases}$$

Define $d'_i = b_i - r'_i$ and $\sigma' = \sum_{i=1}^{k} d'_i = \sigma - (u+1)$. By the induction hypothesis, we have for the total length $L_h(F')$ of the horizontal boundary of F'

$$L_h(F') \ge 2n + A'_1 + A'_2 - \sigma',$$

where

$$A'_1 = d'_{i_1} - d'_{i_2} + d'_{i_3} - \cdots - d'_{i_{2t}} + 2d'_{i_{2t+1}},$$
$$A'_2 = d'_{\tilde{i}_1} - d'_{\tilde{i}_2} + d'_{\tilde{i}_3} - \cdots - d'_{\tilde{i}_{2s}} + 2d'_{\tilde{i}_{2s+1}}.$$

By moving the $u+1$ ones in column j, the piece of horizontal boundary between row $l-1$ and row l has vanished, just like the piece of horizontal boundary between row $h+u$ and $h+u+1$. If $(l+u+1, j) \in F$, the piece of horizontal boundary between row $l+u$ and row $l+u+1$ has also vanished, but there may be a new piece of

horizontal boundary between row $h-1$ and h. On the other hand, if $(h-1,j) \notin F$, the piece of horizontal boundary between row $h-1$ and row h has vanished, but there may be a new piece of horizontal boundary between row $l+u$ and $l+u+1$. At least one of both is the case. All in all, we have $L_h(F') \leq L_h(F) - 2$.

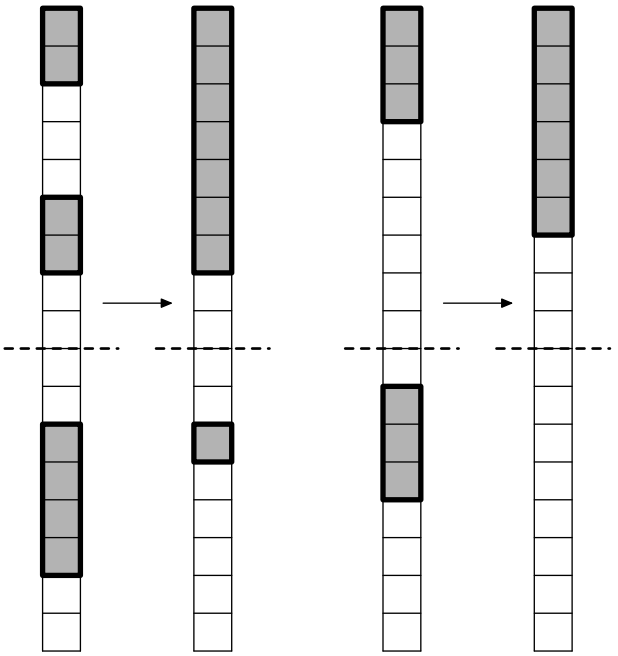


Figure 5.5: Moving ones in Case 1, in both possible configurations. The grey cells have value 1, the other cells value 0.

Furthermore, some of the d'_i involved in A'_1 or A'_2 may be different from the corresponding d_i. Since $\{i_1, i_2, \ldots, i_{2t+1}\} \subset \{1, 2, \ldots, k-1, k, m\}$, we have $d'_i = d_i$ or $d'_i = d_i - 1$ for $i \in \{i_1, i_2, \ldots, i_{2t+1}\}$. The values of i for which $d'_i = d_i - 1$, are all consecutive. Since the coefficients for d_i in A_1 are alternatingly positive and negative, and there is only one positive coefficient that is $+2$ rather than $+1$, we have

$$A'_1 = d'_{i_1} - d'_{i_2} + d'_{i_3} - \cdots - d'_{i_{2t}} + 2d'_{i_{2t+1}} \geq d_{i_1} - d_{i_2} + d_{i_3} - \cdots - d_{i_{2t}} + 2d_{i_{2t+1}} - 2 = A_1 - 2.$$

Since $\{\tilde{i}_1, \tilde{i}_2, \ldots, \tilde{i}_{2s+1}\} \subset \{1, k+1, k+2, \ldots, m-1, m\}$, we have $d'_i = d_i$ or $d'_i = d_i + 1$ for $i \in \{\tilde{i}_1, \tilde{i}_2, \ldots, \tilde{i}_{2s+1}\}$. By a similar argument as above and by the fact that all negative coefficients in A_2 are equal to -1, we have

$$A'_2 \geq A_2 - 1.$$

Finally, we have $\sigma' = \sigma - (u+1) \leq \sigma - 1$. We conclude

$$\begin{aligned} L_h(F) &\geq L_h(F') + 2 \\ &\geq 2n + A_1' + A_2' - \sigma' + 2 \\ &\geq 2n + (A_1 - 2) + (A_2 - 1) - (\sigma - 1) + 2 \\ &= 2n + A_1 + A_2 - \sigma. \end{aligned}$$

This proves (5.12) in Case 1.

Case 2. Suppose that the conditions of Case 1 do not hold and furthermore that $(k, j) \in F$ and $(k+1, j) \in F$. Then there exist integers $l \geq 2$, $h \leq k$ and $u \geq 0$ such that $h \geq l+1$, $k+1 \leq h+u \leq m-1$ and

- $(l-1, j) \in F$, and
- $(l, j), (l+1, j), \ldots, (h-1, j) \notin F$, and
- $(h, j), (h+1, j), \ldots, (h+u, j) \in F$, and
- $(h+u+1, j) \notin F$.

As Case 1 does not apply, we cannot change all zeroes in $(l, j), (l+1, j), \ldots, (h-1, j)$ into ones by moving ones from $(k+1, j), (k+2, j), \ldots, (h+u, j)$. This implies that $h - l > (h+u) - k \geq 1$, so $l < h-1$. We will now distinguish between several cases.

Case 2a. Suppose that there does not exist an integer r with $0 \leq r \leq t$ such that $l = i_{2r+1}$. We define a new image F' by moving the one at $(h+u, j)$ to (l, j); that is,

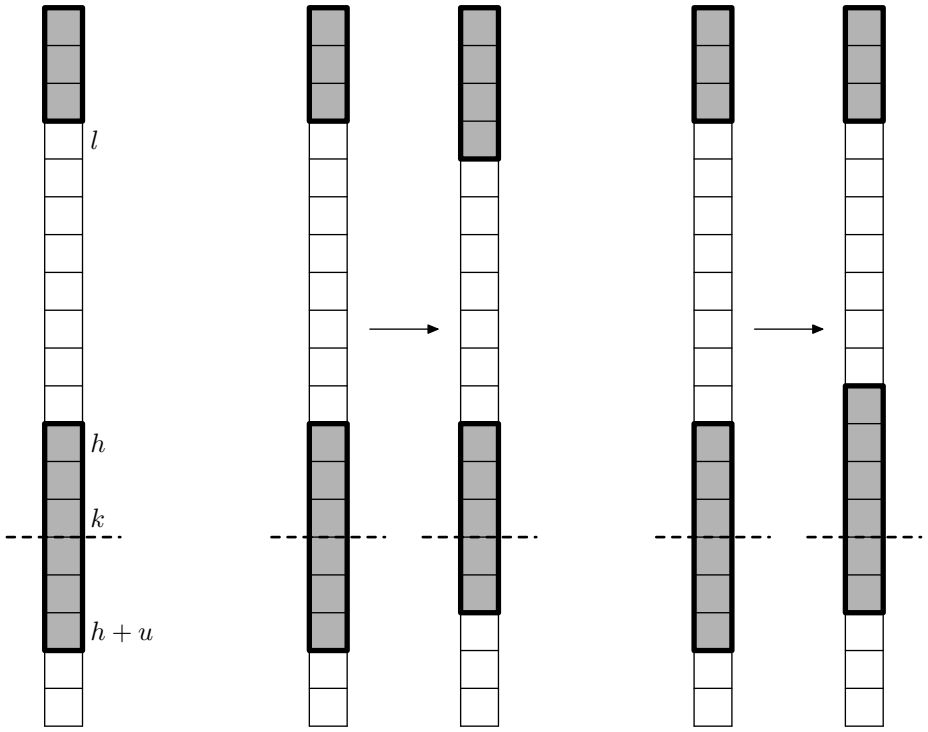
$$F' = F \cup \{(l, j)\} \backslash \{(h+u, j)\}.$$

We define r_i', d_i', σ', A_1', A_2' and $L_h(F')$ similarly as in Case 1. As in Case 1 we have $A_2' \geq A_2 - 1$. However, of the d_i with $i \in \{1, 2, \ldots, k-1, k, m\}$ only one has changed (namely $d_l' = d_l - 1$), and we know that d_l does not have a positive coefficient in A_1. So $A_1' \geq A_1$. Furthermore, $L_h(F') = L_h(F)$ and $\sigma' = \sigma - 1$. By applying the induction hypothesis to F', we find

$$\begin{aligned} L_h(F) &= L_h(F') \\ &\geq 2n + A_1' + A_2' - \sigma' \\ &\geq 2n + A_1 + (A_2 - 1) - (\sigma - 1) \\ &= 2n + A_1 + A_2 - \sigma. \end{aligned}$$

This proves (5.12) in Case 2a.

Case 2b. Suppose that there does not exist an integer r with $0 \leq r \leq t$ such that $h-1 = i_{2r+1}$. We define a new image F' by moving the one at $(h+u, j)$ to $(h-1, j)$; the rest of the proof is the same as in Case 2a.



(a) An example of column j in Case 2.

(b) Moving the ones in Case 2a.

(c) Moving the ones in Case 2b.

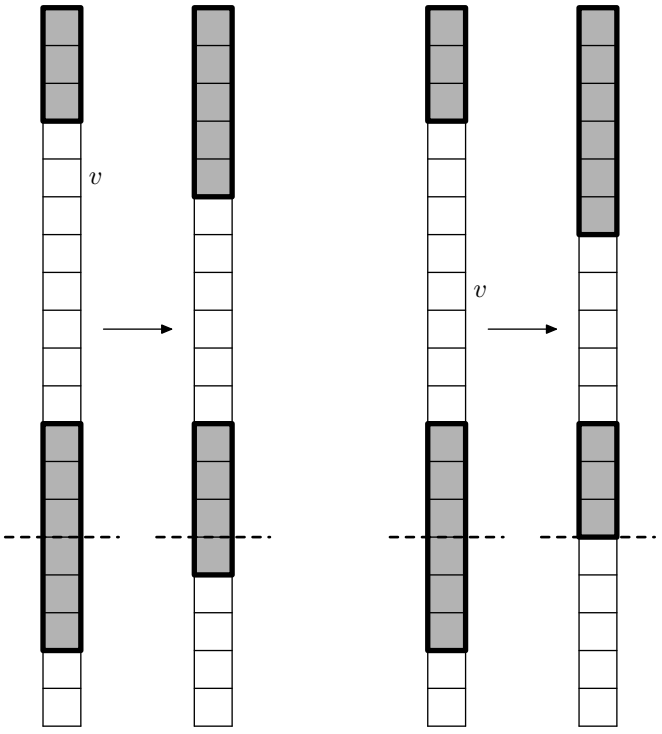
Figure 5.6: Illustrations for Case 2 of the proof. The grey cells have value 1, the other cells value 0.

Case 2c. Suppose neither Case 2a nor Case 2b applies. Then there are integers r_1 and r_2 with $0 \leq r_1 < r_2 \leq t$ such that $l = i_{2r_1+1}$ and $h-1 = i_{2r_2+1}$. Note that $r_1 < t$, so d_l has coefficient $+1$ in A_1. Now let $v = i_{2r_1+2} < h-1$. Again, we distinguish between two cases.

Case 2c1. Suppose that $k+1 \leq h+u-v+l$. Then we define a new image F' by moving the ones at $(h+u-v+l, j)$, $(h+u-v+l+1, j)$, $\ldots$, $(h+u, j)$ to (l, j), $(l+1, j)$, $\ldots$, (v, j); that is,

$$F' = F \cup \{(l,j), (l+1,j), \ldots, (v,j)\} \setminus \{(h+u-v+l, j), (h+u-v+l+1, j), \ldots, (h+u, j)\}.$$

We define r_i', d_i', σ', A_1', A_2' and $L_h(F')$ similarly as in Case 1. As in Case 2a we have $A_2' \geq A_2 - 1$ and $L_h(F') = L_h(F)$. Also, $\sigma' \leq \sigma - 1$. Furthermore, of the d_i with $i \in \{1, 2, \ldots, k-1, k, m\}$ exactly two have changed: $d_l' = d_l - 1$ and $d_v' = d_v - 1$. As



(a) Moving the ones in Case 2c1.

(b) Moving the ones in Case 2c2.

Figure 5.7: More illustrations for Case 2 of the proof. The grey cells have value 1, the other cells value 0.

d_l has coefficient $+1$ in A_1 and d_v has coefficient -1 in A_1, we have $A_1' = A_1$. By applying the induction hypothesis to F', we find

$$\begin{aligned} L_h(F) &= L_h(F') \\ &\geq 2n + A_1' + A_2' - \sigma' \\ &\geq 2n + A_1 + (A_2 - 1) - (\sigma - 1) \\ &= 2n + A_1 + A_2 - \sigma. \end{aligned}$$

This proves (5.12) in Case 2c1.

Case 2c2. Suppose that $k + 1 > h + u - v + l$. Then we define a new image F' by moving the ones at $(k + 1, j)$, $(k + 2, j)$, $\ldots$, $(h + u, j)$ to (l, j), $(l + 1, j)$, $\ldots$, $(l + h + u - k - 1, j)$; that is,

$$F' = F \cup \{(l,j),(l+1,j),\ldots,(l+h+u-k-1,j)\} \setminus \{(k+1,j),(k+2,j),\ldots,(h+u,j)\}.$$

We define r'_i, d'_i, σ', A'_1, A'_2 and $L_h(F')$ similarly as in Case 1. As in Case 2c1 we have $L_h(F') = L_h(F)$ and $\sigma' \leq \sigma - 1$. Since $l + h + u - k - 1 < v$, of the d_i with $i \in \{1, 2, \ldots, k-1, k, m\}$ exactly one has changed: $d'_l = d_l - 1$. As d_l has coefficient $+1$ in A_1, we have $A'_1 = A_1 - 1$.

Now we consider A'_2. Some of the d_i with $i \in \{\tilde{i}_1, \tilde{i}_2, \ldots, \tilde{i}_{2s+1}\}$ may have increased by 1. If $\tilde{i}_1 > h + u$, none of the row indices $k+1$, $k+2$, $\ldots$, $h+u$ occurs in $\{\tilde{i}_1, \tilde{i}_2, \ldots, \tilde{i}_{2s+1}\}$, and we have $A'_2 = A_2$. If not, then $k+1 \leq \tilde{i}_1 \leq h+u$ (using (5.11)). The values of i for which $d'_i = d_i + 1$, are all consecutive. Since the coefficients for d_i in A_1 are alternatingly positive and negative, and since $\tilde{i}_1$ (which has a positive coefficient in A_1) is included in $\{k+1, k+2, \ldots, h+u\}$, we have $A'_2 \geq A_2$.

By applying the induction hypothesis to F', we find

$$\begin{aligned} L_h(F) &= L_h(F') \\ &\geq 2n + A'_1 + A'_2 - \sigma' \\ &\geq 2n + (A_1 - 1) + A_2 - (\sigma - 1) \\ &= 2n + A_1 + A_2 - \sigma. \end{aligned}$$

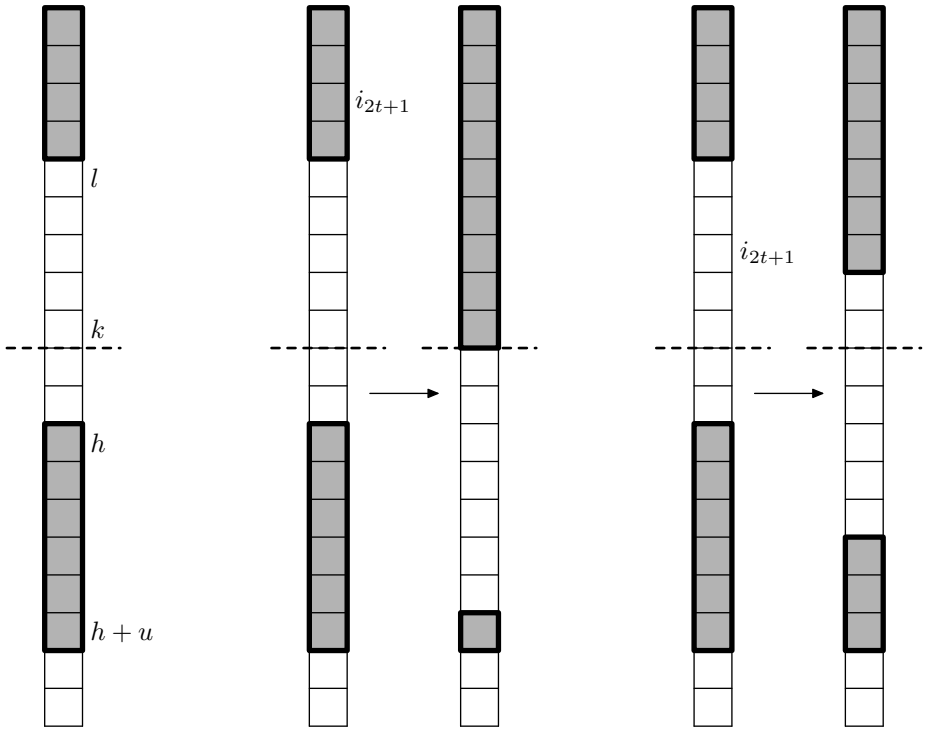
This proves (5.12) in Case 2c2, which completes the proof of Case 2.

Case 3. Suppose that the conditions of Case 1 and Case 2 do not hold. By definition of j we know that in column j there is at least one one in rows $k+1$, $k+2$, $\ldots$, m. As Case 2 does not apply, we have $(k, j) \notin F$ or $(k+1, j) \notin F$. If $(k, j) \in F$ (so $(k+1, j) \notin F$) we can apply Case 1: let l be the smallest integer such that $(l, j) \notin F$, let h' be the greatest integer such that $(h', j) \in F$, and let u be maximal such that $(i, j) \notin F$ for $l \leq i \leq l+u$ and $(i, j) \in F$ for $h' - u \leq i \leq h'$. Define $h = h' - u$. Since $(k, j) \in F$ and $(k+1, j) \notin F$, we have $l + u < k$ and $h > k + 1$, so all conditions of Case 1 are satisfied.

Hence we have $(k, j) \notin F$. Now there exist integers $h \geq k + 1$ and $u \geq 0$ such that $h + u \leq m - 1$ and

- $(h-1, j) \notin F$, and
- $(i, j) \in F$ for $h \leq i \leq h + u$, and
- $(h + u + 1, j) \notin F$.

Furthermore, let $l \leq k$ be such that $(l-1, j) \in F$ and $(l, j) \notin F$. Since Case 1 does not apply, there does not exist an integer u' such that $l + u' \leq k$, $(i, j) \notin F$ for $l \leq i \leq l + u'$ and $(l + u' + 1, j) \in F$. This means that $(i, j) \notin F$ for all i with $l \leq i \leq k+1$. Also, we could still apply Case 1 if there are at least as many zeroes in (l, j), $(l+1, j)$, $\ldots$ (k, j) as there are ones in (h, j), $(h+1, j)$, $\ldots$, $(h+u, j)$. Hence we must have $u + 1 > k - l + 1$.



(a) An example of column j in Case 3.

(b) Moving the ones in Case 3a.

(c) Moving the ones in Case 3b.

Figure 5.8: Illustrations for Case 3 of the proof. The grey cells have value 1, the other cells value 0.

We will distinguish between various cases.

Case 3a. Suppose that either $i_{2t+1} < l$ or $i_{2t+1} = m$. This means that none of the d_i with $l \leq i \leq k$ has coefficient $+2$ in A_1. Since $u + 1 > k - l + 1$, we have $h + k - l < h + u$, so there are ones at (h, j), $(h + 1, j)$, ..., $(h + k - l, j)$. We define a new image F' by moving those ones to (l, j), $(l + 1, j)$, ..., (k, j); that is

$$F' = F \cup \{(l, j), (l + 1, j), \ldots, (k, j)\} \backslash \{(h, j), (h + 1, j), \ldots, (h + k - l, j)\}.$$

We define r'_i, d'_i, σ', A'_1, A'_2 and $L_h(F')$ similarly as in Case 1. As in Case 1 we have $A'_2 \geq A_2 - 1$. Furthermore, $L_h(F') = L_h(F)$.

Suppose $l = k$. Then only one d_i with $i \in \{1, 2, \ldots, k-1, k, m\}$ has changed, namely $d'_k = d_k - 1$. We know that d_k does not have a positive coefficient in A_1, since

$k \neq i_{2t+1}$ (see (5.10)) and $i_{2t-1} \leq k-1$. So $A_1' \geq A_1$. Also, $\sigma' = \sigma - 1$, so by applying the induction hypothesis to F', we find

$$\begin{aligned} L_h(F) &= L_h(F') \\ &\geq 2n + A_1' + A_2' - \sigma' \\ &\geq 2n + A_1 + (A_2 - 1) - (\sigma - 1) \\ &= 2n + A_1 + A_2 - \sigma. \end{aligned}$$

Now suppose that $l < k$. Then we have $\sigma' \leq \sigma - 2$. Furthermore, none of the d_i with $l \leq i \leq k$ has coefficient $+2$ in A_1, so $A_1' \geq A_1 - 1$. By applying the induction hypothesis to F', we find

$$\begin{aligned} L_h(F) &= L_h(F') \\ &\geq 2n + A_1' + A_2' - \sigma' \\ &\geq 2n + (A_1 - 1) + (A_2 - 1) - (\sigma - 2) \\ &= 2n + A_1 + A_2 - \sigma. \end{aligned}$$

This proves (5.12) in Case 3a.

Case 3b. Suppose that $i_{2t+1} \geq l$, $i_{2t+1} \neq m$ and $i_{2t+1} \neq k-1$. Using (5.10), we then have $l \leq i_{2t+1} \leq k-2$. Since $u + 1 > k - l + 1$, we find that $u \geq k - l + 1 \geq (l+2) - l + 1 \geq 3$. We define a new image F' by moving the ones at (h, j), $(h+1, j)$ and $(h+2, j)$ to (l, j), $(l+1, j)$ and $(l+2, j)$; that is,

$$F' = F \cup \{(l, j), (l+1, j), (l+2, j)\} \backslash \{(h, j), (h+1, j), (h+2, j)\}.$$

We define r_i', d_i', σ', A_1', A_2' and $L_h(F')$ similarly as in Case 1. As in Case 1, we have $A_1' \geq A_1 - 2$ and $A_2' \geq A_2 - 1$. Furthermore, $L_h(F') = L_h(F)$ and $\sigma' = \sigma - 3$. By applying the induction hypothesis to F', we find

$$\begin{aligned} L_h(F) &= L_h(F') \\ &\geq 2n + A_1' + A_2' - \sigma' \\ &\geq 2n + (A_1 - 2) + (A_2 - 1) - (\sigma - 3) \\ &= 2n + A_1 + A_2 - \sigma. \end{aligned}$$

This proves (5.12) in Case 3b.

Case 3c. Suppose that neither Case 3a nor Case 3b applies. Then we have $i_{2t+1} = k-1$. Using (5.11), this means that $\tilde{i}_1 \geq k+1 > k-1 = i_{2t+1}$. We now apply Theorem 5.1 to the image F and the row indices $\{i_1, i_2, \ldots, i_{2t}, k-1, k, \tilde{i}_1, \tilde{i}_2, \ldots, \tilde{i}_{2s+1}\}$:

$$\begin{aligned} L_h(F) &\geq 2n + d_{i_1} - d_{i_2} + \cdots - d_{i_{2t}} + d_{k-1} - d_k + d_{\tilde{i}_1} - d_{\tilde{i}_2} + \cdots - d_{\tilde{i}_{2s}} + 2d_{\tilde{i}_{2s+1}} \\ &= 2n + A_1 - d_{k-1} - d_k + A_2. \end{aligned}$$

By Ryser's theorem [24] we have $\sum_{i=1}^{k-2} d_i \geq 0$, since the line sums are consistent, so

$$\sigma = \sum_{i=1}^{k} d_i = \sum_{i=1}^{k-2} d_i + d_{k-1} + d_k \geq d_{k-1} + d_k.$$

Hence

$$L_h(F) \geq 2n + A_1 - d_{k-1} - d_k + A_2 \geq 2n + A_1 + A_2 - \sigma,$$

which proves (5.12) in Case 3c.

This finishes the proof of the theorem. □

Example 5.5. Let $m = n = 12$ and let row sums $(12, 8, 9, 8, 8, 5, 5, 2, 3, 2, 1, 0)$ and column sums $(10, 8, 8, 8, 6, 6, 6, 3, 2, 2, 2, 2)$ be given. We compute b_i and d_i, $i = 1, 2, \ldots, 12$ as shown below.

i	1	2	3	4	5	6	7	8	9	10	11	12
b_i	12	12	8	7	7	7	4	4	1	1	0	0
r_i	12	8	9	8	8	5	5	2	3	2	1	0
d_i	0	+4	−1	−1	−1	+2	−1	+2	−2	−1	−1	0

Here (5.2) yields at most

$$L_h \geq 24 + 4 - (-1) + 2 - (-1) + 2 \cdot 2 = 36,$$

and (5.3) yields at most

$$L_h \geq 24 - (-2) + 2 - (-1) + 2 - (-1) + 4 - 2 \cdot 0 = 36.$$

However, we can apply Theorem 5.3 with $k = 5$ (note that $d_5 < 0$ and $d_6 \geq 0$). We have $\sigma = 1$. If we take $t = 0$, $s = 0$, $i_1 = 2$, $\tilde{i}_1 = 6$, $\tilde{i}_2 = 7$ and $\tilde{i}_3 = 8$, then we find

$$L_h \geq 24 + 2 \cdot 4 + 2 - (-1) + 2 \cdot 2 - 1 = 38.$$

So in this example, Theorem 5.3 gives a better bound than Theorem 5.1. In fact, the bound of Theorem 5.3 is sharp in this example: in Figure 5.9 a binary image F with the given row and column sums is shown, for which $L_h = 38$.

Corollary 5.4. *Let be given row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$. Suppose there exists an image F with line sums $(\mathcal{R}, \mathcal{C})$ and let $L_h(F)$ be the total length of the horizontal boundary of this image. Define $b_i = \#\{j : c_j \geq i\}$ and $d_i = b_i - r_i$ for $i = 1, 2, \ldots, m$. Also set $d_0 = d_{m+1} = 0$. Let k be an integer with $1 \leq k \leq m$ such that $d_k < 0$ and $d_{k+1} \geq 0$. Let $\sigma = \sum_{i=1}^{k} d_i$. For any integers $t, s \geq 0$ and any sets $\{i_1, i_2, \ldots, i_{2t+1}\} \subset \{0, 1, \ldots, k-1, k, m+1\}$ with $i_1 < i_2 < \ldots < i_{2t+1}$ and $\{\tilde{i}_1, \tilde{i}_2, \ldots, \tilde{i}_{2s+1}\} \subset \{0, k+1, k+2, \ldots, m, m+1\}$ with $\tilde{i}_1 < \tilde{i}_2 < \ldots < \tilde{i}_{2s+1}$ we have*

$$\begin{aligned} L_h(F) \geq 2r_1 &+ d_{i_1} - d_{i_2} + d_{i_3} - \cdots - d_{i_{2t}} + 2d_{i_{2t+1}} \\ &+ d_{\tilde{i}_1} - d_{\tilde{i}_2} + d_{\tilde{i}_3} - \cdots - d_{\tilde{i}_{2s}} + 2d_{\tilde{i}_{2s+1}} - \sigma. \end{aligned}$$

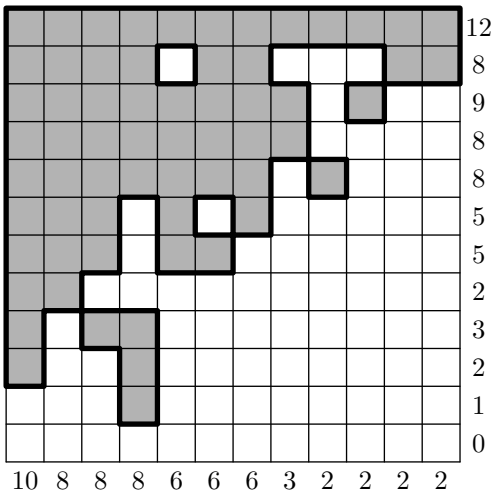


Figure 5.9: The binary image from Examples 5.5. The grey cells have value 1, the other cells value 0. The numbers indicate the row and column sums. The length of the horizontal boundary of this image is 38.

Proof. Completely analogous to the proof of Corollary 5.2. □

CHAPTER 6

Reconstructions with small boundary

This chapter (with minor modifications) is available as a preprint as: Birgit van Dalen, "Discrete tomography reconstructions with small boundary", arXiv:1011.5351 [math.CO] (2010) 18 pp.

6.1 Introduction

In Chapter 5 we proved a lower bound on the length of the boundary for any reconstruction of an image with given line sums. In this chapter we complement this result by giving a reconstruction that has a relatively small boundary in the case that both the row and the column sums are monotone.

After introducing some notation in Section 6.2, we describe the construction of a solution to the discrete tomography problem in Section 6.3. In Section 6.4 we prove upper bounds on the length of the boundary of this constructed solution. We show by examples that these bounds are sharp in Section 6.5, and finally in Section 6.6 we generalise the results slightly.

6.2 Definitions and notation

Let F be a finite subset of $\mathbb{Z}^2$ with characteristic function χ. (That is, $\chi(k,l) = 1$ if $(k,l) \in F$ and $\chi(k,l) = 0$ otherwise.) For $i \in \mathbb{Z}$, we define *row* i as the set $\{(k,l) \in \mathbb{Z}^2 : k = i\}$. We call i the index of the row. For $j \in \mathbb{Z}$, we define *column* j as the set $\{(k,l) \in \mathbb{Z}^2 : l = j\}$. We call j the index of the column. Note that we follow matrix notation: we indicate a point (i,j) by first its row index i and then its column index j. Also, we use row numbers that increase when going downwards and column numbers that increase when going to the right.

The *row sum* r_i is the number of elements of F in row i, that is $r_i = \sum_{j\in\mathbb{Z}} \chi(i,j)$. The *column sum* c_j of F is the number of elements of F in column j, that is $c_j = \sum_{i\in\mathbb{Z}} \chi(i,j)$. We refer to both row and column sums as the *line sums* of F. We will usually only consider finite sequences $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and $\mathcal{C} = (c_1, c_2, \ldots, c_n)$ of row and column sums that contain all the nonzero line sums. In this chapter we will always assume that the line sums are monotone, that is $r_1 \geq r_2 \geq \ldots \geq r_m$ and $c_1 \geq c_2 \geq \ldots \geq c_n$.

Given sequences of integers $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and $\mathcal{C} = (c_1, c_2, \ldots, c_n)$ with $0 \leq r_i \leq n$, $0 \leq c_j \leq m$, we say that $(\mathcal{R}, \mathcal{C})$ is *consistent* if there exists a set F with row sums $\mathcal{R}$ and column sums $\mathcal{C}$. Define $b_i = \#\{j : c_j \geq i\}$ for $i = 1, 2, \ldots, m$. Note that by definition we have $\sum_{i=1}^{m} b_i = \sum_{j=1}^{n} c_j$. Ryser's theorem [24] states that if $r_1 \geq r_2 \geq \ldots \geq r_m$, the line sums $(\mathcal{R}, \mathcal{C})$ are consistent if and only if $\sum_{j=1}^{n} c_j = \sum_{i=1}^{m} r_i$ and for each $k = 1, 2, \ldots, m$ we have $\sum_{i=1}^{k} b_i \geq \sum_{i=1}^{k} r_i$.

We say that the line sums $(\mathcal{R}, \mathcal{C})$ uniquely determine such a set F if the following property holds: if F' is another subset of $\mathbb{Z}^2$ with line sums $(\mathcal{R}, \mathcal{C})$, then $F' = F$. In this case we call F *uniquely determined*.

We will now define a *uniquely determined neighbour corresponding to line sums* $(\mathcal{R}, \mathcal{C})$. This is a uniquely determined set that is in some sense the closest to any set with those line sums. See also Section 3.4.

Definition 6.1. *Let be given row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$, where $n = r_1 \geq r_2 \geq \ldots \geq r_m$ and $m = c_1 \geq c_2 \geq \ldots \geq c_n$. Let $b_i = \#\{j : c_j \geq i\}$ for $i = 1, 2, \ldots, m$. Then the column sums c_1, c_2, …, c_n and row sums b_1, b_2, …, b_m uniquely determine a set F_1, which we will call the* uniquely determined neighbour corresponding to line sums $(\mathcal{R}, \mathcal{C})$.

Suppose line sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and $\mathcal{C} = (c_1, c_2, \ldots, c_n)$ are given, where $r_1 \geq r_2 \geq \ldots \geq r_m$ and $c_1 \geq c_2 \geq \ldots \geq c_n$. Let the uniquely determined neighbour

corresponding to $(\mathcal{R}, \mathcal{C})$ have row sums $b_1 \geq b_2 \geq \ldots \geq b_n$. Then we define

$$\alpha(\mathcal{R}, \mathcal{C}) = \frac{1}{2} \sum_{i=1}^{m} |r_i - b_i|.$$

Note that $\alpha(\mathcal{R}, \mathcal{C})$ is an integer, since $2\alpha(\mathcal{R}, \mathcal{C})$ is congruent to

$$\sum_{i=1}^{m} (r_i + b_i) = \sum_{i=1}^{m} r_i + \sum_{i=1}^{m} b_i = 2 \sum_{i=1}^{m} r_i \equiv 0 \mod 2.$$

If we write $d_i = b_i - r_i$ for all i, then because $\sum_{i=1}^{m} r_i = \sum_{i=1}^{m} b_i$, we have

$$\alpha = \sum_{d_i > 0} d_i = - \sum_{d_i < 0} d_i.$$

We can view the set F as a picture consisting of cells with zeroes and ones. Rather than $(i, j) \in F$, we might say that (i, j) has value 1 or that there is a one at (i, j). Similarly, for $(i, j) \notin F$ we sometimes say that (i, j) has value zero or that there is a zero at (i, j).

We define the *boundary* of F as the set consisting of all pairs of points $\big((i, j), (i', j')\big)$ such that

- $i = i'$ and $|j - j'| = 1$, or $|i - i'| = 1$ and $j = j'$, and
- $(i, j) \in F$ and $(i', j') \notin F$.

One element of this set we call *one piece of the boundary*. We can partition the boundary into two subsets, one containing the pairs of points with $i = i'$ and the other containing the pairs of points with $j = j'$. The former set we call the *vertical boundary* and the latter set we call the *horizontal boundary*. We define the *length of the (horizontal, vertical) boundary* as the number of elements in the (horizontal, vertical) boundary. For a given set F we denote the length of the horizontal boundary by $L_h(F)$ and the length of the vertical boundary by $L_v(F)$.

6.3 The construction

In this section we will construct a set F_2 satisfying given monotone row and column sums that are consistent. First we will describe one step of this construction.

Let row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$ be given, such that $n = r_1 \geq r_2 \geq \ldots \geq r_m$ and $m = c_1 \geq c_2 \geq \ldots \geq c_n$. Assume that

those line sums are consistent. For $i = 1, 2, \ldots, m$ define $b_i = \#\{j : c_j \geq i\}$ and $d_i = b_i - r_i$. For convenience we define $r_{m+1} = b_{m+1} = d_{m+1} = 0$. We have $n = b_1 \geq b_2 \geq \ldots \geq b_m > b_{m+1}$.

Let F_1 be the uniquely determined neighbour corresponding to the line sums $(\mathcal{R}, \mathcal{C})$. Then F_1 has row sums $(b_1, b_2, \ldots, b_m)$ and column sums $(c_1, c_2, \ldots, c_n)$. Moreover, in every column j the elements of F_1 are exactly in the first c_j rows.

If $r_i = b_i$ for all i, then F_1 already satisfies the line sums $(\mathcal{R}, \mathcal{C})$, and there is nothing to be done. Now assume that not for all i we have $r_i = b_i$. Then there is at least one i with $d_i > 0$ and one i with $d_i < 0$. Also, because of the consistency of the line sums the smallest i with $d_i \neq 0$ satisfies $d_i > 0$.

Let i_1 be minimal such that $d_{i_1} > 0$ and let i_2 be minimal such that $d_{i_2} > 0$ and $d_{i_2+1} \leq 0$. Let $R^+ = \{i_1, i_1+1, \ldots, i_2\}$. Similarly, let i_3 be minimal such that $d_{i_3} < 0$ and let i_4 be minimal such that $d_{i_4} < 0$ and $d_{i_4+1} \geq 0$. Such i_4 exists, since $d_{m+1} = 0$. Let $R^- = \{i_3, i_3 + 1, \ldots, i_4\}$. Now $d_i > 0$ for all $i \in R^+$ and $d_i < 0$ for all $i \in R^-$.

If $|R^+| \leq |R^-|$, we execute an **A-step**, while if $|R^+| > |R^-|$, we execute a **B-step**. We will now describe these two different steps.

A-step. Let j be maximal such that $c_j \in R^+$. Such a j exists, because as $b_{i_2+1} \leq r_{i_2+1} \leq r_{i_2} < b_{i_2}$, there exists a column with sum i_2. Define $s = c_j - i_1 + 1$; this is the number of rows i with $i_1 \leq i \leq c_j$. We will be moving the ones in the s cells (i_1, j), $\ldots$, (c_j, j) to other cells. To determine to which cells those ones are moved, consider $i_3, i_3 + 1, \ldots, i_3 + s - 1$. Since $i_4 - i_3 + 1 = |R^-| \geq |R^+| \geq s$, we have $i_3 + s - 1 \leq i_4$, so $\{i_3, i_3 + 1, \ldots, i_3 + s - 1\} \subset R^-$. If $r_{i_3+s-1} > r_{i_3+s}$, then let $I = \{i_3, i_3 + 1, \ldots, i_3 + s - 1\}$.

Now suppose $r_{i_3+s-1} = r_{i_3+s}$. Let t_1 be minimal such that $i_3 \leq t_1 \leq i_3 + s - 1$ and $r_{t_1} = r_{i_3+s-1}$. Let t_2 be such that $t_2 \geq i_3 + s$ and $r_{i_3+s-1} = r_{t_2} > r_{t_2+1}$. Since we have $d_{i_4+1} \geq 0$, we have $r_{i_4+1} \leq b_{i_4+1} \leq b_{i_4} < r_{i_4}$, hence $t_2 \leq i_4$. Let $t_3 = t_2 + t_1 - i_3 - s + 1$. As $t_2 \geq i_3 + s$, we have $t_3 \geq t_1 + 1$, and as $t_1 \leq i_3 + s - 1$, we have $t_3 \leq t_2$. Now define $I = \{i_3, i_3 + 1, \ldots, t_1 - 1\} \cup \{t_3, t_3 + 1, \ldots, t_2\}$. We have $|I| = (t_1 - i_3) + (-t_1 + i_3 + s) = s$.

In both cases we have now defined a set $I \subset R^-$ with $|I| = s = c_j - i_1 + 1$ and satisfying the following property: if $i \in I$ and $i + 1 \notin I$, then $r_i > r_{i+1}$.

Now we move the ones from the rows i with $i_1 \leq i \leq c_j$ to the rows $i \in I$. This column will later be one of the columns of F_2. We delete the column and change the line sums accordingly: define for $i = 1, 2, \ldots, m$ the new row sums r'_i, which is equal to r_i if there was no one in this row in column j, and equal to $r_i - 1$ if there was a

one in this row in column j. We have

$$r'_i = \begin{cases} r_i - 1 & \text{for } i < i_1, \\ r_i & \text{for } i_1 \leq i \leq c_j, \\ r_i - 1 & \text{for } i \in I, \\ r_i & \text{for } i > c_j \text{ and } i \notin I. \end{cases}$$

Also let b'_i be the number of columns not equal to j with column sum at least i. We have

$$b'_i = \begin{cases} b_i - 1 & \text{for } i \leq c_j, \\ b_i & \text{for } i > c_j. \end{cases}$$

Note that the set F'_1, defined as F_1 without column j, has row sums $b'_1, b'_2, \ldots, b'_m$.

We now want to show that the new row sums are non-increasing and that they are consistent, together with the column sums without column j, that is, that $\sum_{i=1}^{k} b'_i \geq \sum_{i=1}^{k} r'_i$ for $k = 1, 2, \ldots, m$.

Suppose for some i we have $r'_i < r'_{i+1}$. Then we must have $r'_i = r_i - 1$ and $r'_{i+1} = r_{i+1}$, since $r_i \geq r_{i+1}$. So either $i = i_1 - 1$ or $i \in I$ and $i + 1 \notin I$. In the latter case we know $r_i > r_{i+1}$, hence $r'_i \geq r'_{i+1}$. If on the other hand $i = i_1 - 1$, we have $d_i = 0$ and $d_{i+1} > 0$, so $r_i = b_i \geq b_{i+1} > r_{i+1}$, hence $r'_i \geq r'_{i+1}$. We conclude that it can never happen that $r'_i < r'_{i+1}$. So $n - 1 = r'_1 \geq r'_2 \geq \ldots \geq r'_m$.

Now we prove consistency. For $i < i_1$ we have $d_i = 0$, hence

$$b'_i - r'_i = (b_i - 1) - (r_i - 1) = d_i = 0.$$

For $i_1 \leq i \leq c_j$ we have $d_i > 0$, hence

$$b'_i - r'_i = (b_i - 1) - r_i = d_i - 1 \geq 0.$$

For $c_j + 1 \leq i \leq i_3 - 1$ we have $d_i \geq 0$, hence

$$b'_i - r'_i = b_i - r_i = d_i \geq 0.$$

So for $k \leq i_3 - 1$ we clearly have

$$\sum_{i=1}^{k} (b'_i - r'_i) \geq 0.$$

On the other hand, for $k \geq i_4$ we have $\sum_{i=1}^{k} (b_i - r_i) \geq 0$ because of the consistency of the original line sums, hence

$$\sum_{i=1}^{k} (b'_i - r'_i) = \left(\sum_{i=1}^{k} b_i - c_j \right) - \left(\sum_{i=1}^{k} r_i - c_j \right) = \sum_{i=1}^{k} (b_i - r_i) \geq 0.$$

For $i_3 \le i \le i_4$ we have $d_i < 0$, so

$$b'_i - r'_i = b_i - r_i = d_i < 0 \qquad \text{if } i \notin I,$$
$$b'_i - r'_i = b_i - (r_i - 1) = d_i + 1 \le 0 \qquad \text{if } i \in I.$$

Hence for $i_3 \le k \le i_4 - 1$ we have

$$\sum_{i=1}^{k}(b'_i - r'_i) = \sum_{i=1}^{i_4}(b'_i - r'_i) - \sum_{i=k+1}^{i_4}(b'_i - r'_i) \ge 0.$$

This proves the consistency.

B-step. Let j be minimal such that $c_j + 1 \in R^-$. Such a j exists, because as $b_{i_3-1} \ge r_{i_3-1} \ge r_{i_3} > b_{i_3}$, there exists a column with sum $i_3 - 1$. Similarly to the A-step, we find a set $I \subset R^+$ such that $|I| = i_4 - c_j$ with the following property: if $i \notin I$ and $i + 1 \in I$, then $r_i > r_{i+1}$.

Now we move the ones from the rows i with $i \in I$ to the rows i with $c_j + 1 \le i \le i_4$. This column will later be one of the columns of F_2. We delete the column and change the line sums accordingly. Analogously to above we prove that the new line sums are non-increasing and consistent, and that the set F'_1 that we have left, is the uniquely determined neighbour corresponding to these new line sums.

The procedure described above, which changes line sums $(\mathcal{R}, \mathcal{C})$ and their uniquely determined neighbour F_1 to new line sums $(\mathcal{R}', \mathcal{C}')$ and their uniquely determined neighbour F'_1, we denote by φ. Since the new line sums satisfy all the necessary properties, we can apply φ also to $(\mathcal{R}', \mathcal{C}')$ and F'_1. We can repeat this until we arrive at a situation where the uniquely determined neighbour already satisfies the line sums. One by one we can then put the deleted columns back in the right position (first the column that was last deleted, then the one that was deleted before that, and so on, to make sure that the resulting set F_2 has its columns in the right order). Every time we put back a column, the line sums change back to what they were before that instance of φ was applied. When all the columns are back in place, the line sums are therefore equal to $(\mathcal{R}, \mathcal{C})$ and the resulting set satisfies these line sums. This proves the following theorem.

Theorem 6.1. *Let be given row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$, where $n = r_1 \ge r_2 \ge \ldots \ge r_m$ and $m = c_1 \ge c_2 \ge \ldots \ge c_n$. Assume that the line sums are consistent. Let F_1 be the uniquely determined neighbour corresponding to the line sums $(\mathcal{R}, \mathcal{C})$. If we start with F_1 and repeatedly apply φ until this is no longer possible, and then put all the deleted columns back in the right position, then the result is a set F_2 that satisfies the line sums $(\mathcal{R}, \mathcal{C})$.*

Now we show an example of this construction. Let $m = 12$, $n = 11$ and define line sums

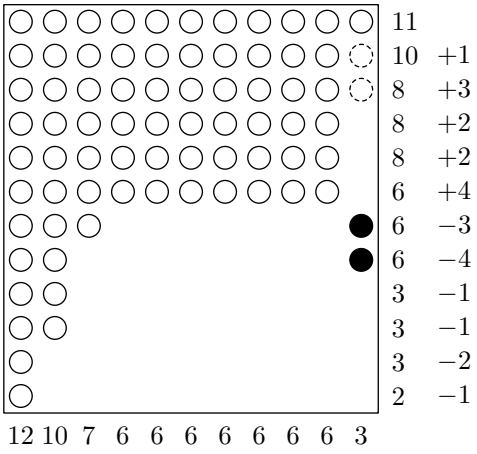
$$\mathcal{R} = (11, 10, 8, 8, 8, 6, 6, 6, 3, 3, 3, 2), \qquad \mathcal{C} = (12, 10, 7, 6, 6, 6, 6, 6, 6, 6, 3).$$

We have

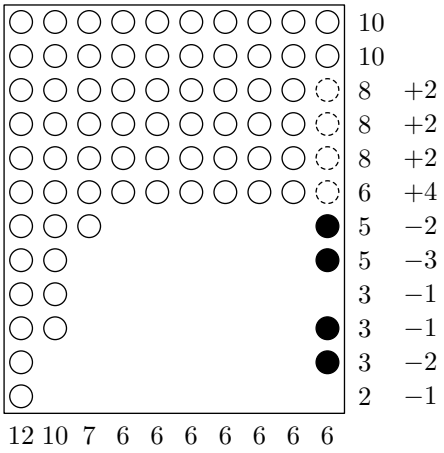
$$(b_1, \ldots, b_{12}) = (11, 11, 11, 10, 10, 10, 3, 2, 2, 2, 1, 1),$$

$$(d_1, \ldots, d_{12}) = (0, +1, +3, +2, +2, +4, -3, -4, -1, -1, -2, -1).$$

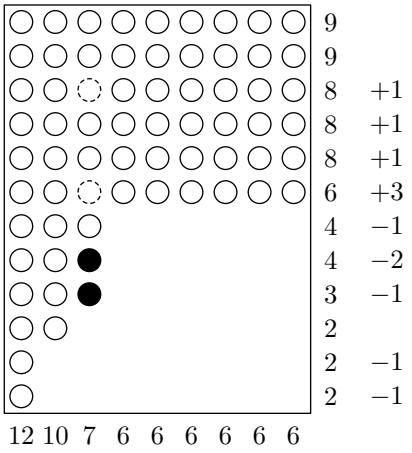
We will now do the construction step by step, illustrated by Figures 6.1 and 6.2. The r_i and d_i in every step are indicated in the figure. We start with the uniquely determined neighbour F_1, that is, the set with column sums $\mathcal{C}$ and row sums $(b_1, \ldots, b_{12})$.



(a) Step 1.



(b) Step 2.



(c) Step 3.

Figure 6.1: The first steps of the construction of the set F_2. The ones are indicated by white circles. The dashed circles are ones that are deleted in that step, while the black circles are ones that are newly added in that step. The numbers directly next to each figure are the row sums, while the numbers next to that are the d_i.

Step 1. We have $R^+ = \{2,3,4,5,6\}$, $R^- = \{7,8,9,10,11,12\}$. Since $|R^+| \leq |R^-|$, we execute an A-step. The rightmost column j with $c_j \in R^+$ is column 11, with sum 3. We delete the ones in $(2,11)$ and $(3,11)$. We find $I = \{7,8\}$, since $r_8 > r_9$. So we add ones in $(7,11)$ and $(8,11)$. We then delete column 11.

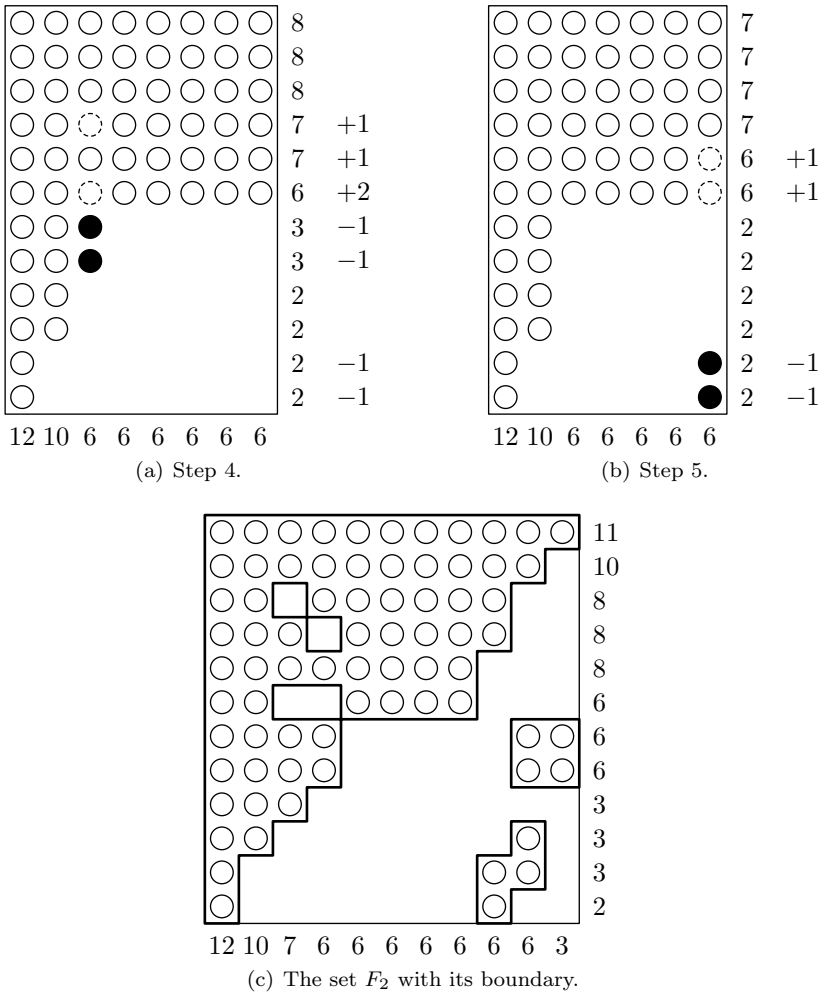
Step 2. We have $R^+ = \{3,4,5,6\}$ and $R^- = \{7,8,9,10,11,12\}$. Since $|R^+| \leq |R^-|$, we execute an A-step. The rightmost column j with $c_j \in R^+$ is column 10, with sum 6. We delete the ones in this column in rows 3, 4, 5 and 6. Since $r_{10} = r_{11}$, we cannot use $I = \{7,8,9,10\}$. Instead we take $I = \{7,8,10,11\}$. This works since $r_8 > r_9$ and $r_{11} > r_{12}$. So we add ones in column 10 in rows 7, 8, 10 and 11. We then delete column 10.

Step 3. In row 10, the new row sum is 2, while the new b_{10} is also 2. So the new d_{10} is 0. This means that while R^+ is still equal to $\{3,4,5,6\}$, we now have $R^- = \{7,8,9\}$. Hence $|R^+| > |R^-|$ and therefore we execute a B-step. The leftmost column j with $c_j + 1 \in R^-$ is column 3 with sum 7. So we add ones in $(8,3)$ and $(9,3)$. As $r_5 = r_4 = r_3$, we cannot take $I = \{6,5\}$, but we have to take $I = \{6,3\}$. Hence we delete ones in $(3,3)$ and $(6,3)$. We then delete column 3.

Step 4. We have $R^+ = \{4,5,6\}$ and $R^- = \{7,8\}$. As $|R^+| > |R^-|$, we execute a B-step. The leftmost column j with $c_j + 1 \in R^-$ is column 3 (which was originally column 4) with sum 6. We add ones in $(7,3)$ and $(8,3)$. As $r_5 = r_4$, we take $I = \{6,4\}$, so we delete ones from $(4,3)$ and $(6,3)$. We then delete column 3.

Step 5. We have $R^+ = \{5,6\}$ and $R^- = \{11,12\}$. As $|R^+| \leq |R^-|$, we execute an A-step. The rightmost column j with $c_j \in R^+$ is column 7 (which was originally column 9) with sum 6. We deletes ones from $(5,7)$ and $(6,7)$, and we add ones in $(11,7)$ and $(12,7)$. We then delete column 7.

Now all d_i have become 0, so we are done. We put back the deleted columns in their original places and find the set F_2 that satisfies the original line sums, see Figure 6.2(c).



(a) Step 4.

(b) Step 5.

(c) The set F_2 with its boundary.

Figure 6.2: The last steps of the construction of the set F_2. The ones are indicated by white circles. The dashed circles are ones that are deleted in that step, while the black circles are ones that are newly added in that step. The numbers directly next to each figure are the row sums, while the numbers next to that are the d_i.

6.4 Boundary length of the constructed solution

In this section we prove upper bounds on the length of the boundary of the set that results from the construction described in the previous section.

Theorem 6.2. *Let be given row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$, where $n = r_1 \geq r_2 \geq \ldots \geq r_m$ and $m = c_1 \geq c_2 \geq \ldots \geq c_n$. Assume that the line sums are consistent. Let $\alpha = \alpha(\mathcal{R}, \mathcal{C})$. For the set F_2 constructed in Theorem 6.1 we have*

$$L_h(F_2) \leq 2n + 2\alpha, \qquad L_v(F_2) \leq 2m + 2\alpha.$$

Proof. Let F_1 be the uniquely determined neighbour corresponding to the line sums $(\mathcal{R}, \mathcal{C})$. Starting with F_1, we apply φ repeatedly, moving ones in several columns accordingly and deleting those columns. After that, to analyse what happens to the boundary, we start again with F_1 and repeat the entire procedure, moving exactly the same ones, but this time keeping the columns that were supposed to be deleted.

The length of the horizontal boundary of F_1 is equal to $2n$, since there are n columns that each contain one connected set of ones. The length of the vertical boundary of F_1 is $2m$. Note that the ones that are moved when applying ϕ are always deleted from a row i with $d_i > 0$ and added to a row i with $d_i < 0$. In fact for each row i with $d_i > 0$ ones are deleted exactly d_i times during the construction, and for each row i with $d_i < 0$ ones are added exactly $-d_i$ times. Therefore the total number of ones that are moved is equal to α. We now want to show that when in one application of φ exactly s ones are moved, both the horizontal and vertical boundary do not increase with more than $2s$. From this the theorem follows.

We will only consider what happens at an A-step; the other case is analogous. So suppose we execute an A-step and move s ones, while either the horizontal or vertical boundary increases by more than $2s$. First consider the horizontal boundary. Since the ones in the rows i with $i_1 \leq i \leq c_j$ are removed, and there never was a one in $(c_j + 1, j)$, this does not yield any additional boundary. Adding the ones in the rows i with $i \in I$ may yield additional boundary, but only 2 for each one that is added, so at most $2s$ in total.

So we may assume that the vertical boundary has increased by more than $2s$. Adding the ones leads to additional vertical boundary of at most $2s$, so deleting the ones must also have led to additional boundary. This means that there was a one in (i, j), which is now deleted, while there are still ones in $(i, j-1)$ and $(i, j+1)$. As $d_i > 0$, those ones cannot have been added during earlier steps in the construction, so they must have been there from the beginning. This means in particular that $c_{j+1} \geq i \geq i_1$, while also $c_{j+1} \leq c_j \leq i_2$, so $c_{j+1} \in R^+$. But j was chosen maximally

such that $c_j \in R^+$, so apparently column $j+1$ was in the original construction deleted in an earlier application of φ.

Suppose this earlier application has been an A-step. Since rows l with $d_l \leq 0$ at some point in the construction can never have $d_l > 0$ at a later point in the construction, we know that all rows l with $i_1 \leq l \leq c_{j+1}$ were contained in R^+ in this earlier application of φ. In particular should the one in $(i, j+1)$ have been moved during this step. So this is impossible.

Now suppose that the earlier application has been a B-step. Then column $j+1$ can only have been chosen to execute this step in if $d_{c_{j+1}+1} < 0$. Since $c_{j+1} \leq c_j$ and $d_{c_j} > 0$ (now, and therefore also earlier), we then must have $c_j = c_{j+1}$. Hence $d_{c_j+1} < 0$, which means that to execute this B-step column j, rather than column $j+1$, should have been chosen. So this case is impossible as well.

We conclude that the vertical boundary has increased by at most $2s$ as well, and this completes the proof of the theorem. □

In light of this theorem it is interesting to note that α cannot become arbitrarily large while n and m are fixed. In fact, we have the following result.

Proposition 6.3. *Let be given row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$, where $n = r_1 \geq r_2 \geq \ldots \geq r_m$ and $m = c_1 \geq c_2 \geq \ldots \geq c_n$. Assume that the line sums are consistent. Let $\alpha = \alpha(\mathcal{R}, \mathcal{C})$. Then*

$$\alpha \leq \frac{(m-1)(n-1)}{4}.$$

Proof. For $i = 1, 2, \ldots, m$, let $b_i = \#\{j : c_j \geq i\}$ and $d_i = b_i - r_i$. Let a be the number of rows (indices i) with $d_i > 0$ and b the number of rows with $d_i < 0$. We assume $\alpha > 0$, so $a, b > 0$. Define $d^+ = \max\{d_i : d_i > 0\}$ and $d^- = \max\{-d_i : d_i < 0\}$. We have $b_1 = n = r_1$, so $d_1 = 0$, hence $a + b \leq m - 1$.

Now we prove that $d^+ + d^- \leq n - 1$. Let k and l be such that $b_k - r_k = d^+$ and $r_l - b_l = d^-$. First suppose $k < l$. Then since $r_1 \geq r_2 \geq \ldots \geq r_m$ and $b_1 \geq b_2 \geq \ldots \geq b_m$ we have $b_1 \geq b_k = b_k - r_k + r_k = d^+ + r_k$ and $-b_m \geq -b_l = r_l - b_l - r_l = d^- - r_l$, hence

$$d^+ + d^- \leq (b_1 - r_k) + (-b_m + r_l) \leq b_1 - b_m \leq n - 1.$$

If on the other hand $k > l$, then $r_1 \geq r_l = r_l - b_l + b_l = d^- + b_l$ and $-r_m \geq -r_k = b_k - r_k - b_k = d^+ - b_k$, and hence

$$d^+ + d^- \leq (-r_m + b_k) + (r_1 - b_l) \leq r_1 - r_m \leq n - 1.$$

Now note that we have

$$\alpha = \sum_{d_i>0} d_i = \sum_{d_i<0} (-d_i),$$

so

$$\alpha^2 = \left(\sum_{d_i>0} d_i\right)\left(\sum_{d_i<0} (-d_i)\right) \leq \big(a \cdot d^+\big)\big(b \cdot d^-\big) = \big(a \cdot b\big)\big(d^+ \cdot d^-\big)$$

$$\leq \left(\frac{a+b}{2}\right)^2 \left(\frac{d^+ + d^-}{2}\right)^2 \leq \left(\frac{m-1}{2}\right)^2 \left(\frac{n-1}{2}\right)^2.$$

Therefore

$$\alpha \leq \frac{(m-1)(n-1)}{4}.$$

□

In case of large α, the construction of Theorem 6.1 actually gives a much smaller horizontal boundary than the bound in Theorem 6.2, as the following theorem shows.

Theorem 6.4. *Let be given row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$, where $n \geq 2$, $n = r_1 \geq r_2 \geq \ldots \geq r_m$ and $m = c_1 \geq c_2 \geq \ldots \geq c_n$. Assume that the line sums are consistent. For the set F_2 constructed in Theorem 6.1 we have*

$$L_h(F_2) \leq 4n - 4.$$

Proof. We will prove this by induction on n. Let $\alpha = \alpha(\mathcal{R}, \mathcal{C})$. If $\alpha > 0$, then there are l_1 and l_2 such that $2 \leq l_1 < l_2$ and $d_{l_1} > 0$ and $d_{l_2} < 0$. Then

$$b_1 \geq b_{l_1} \geq r_{l_1} + 1 \geq r_{l_2} + 1 \geq b_{l_2} + 2 \geq 1 + 2 = 3.$$

Hence $n \geq 3$. So when $n = 2$, we have $\alpha = 0$ and the construction gives $F_2 = F_1$, with $L_h = 2n = 4n - 2n = 4n - 4$.

Now let $k \geq 3$ and suppose that we have proved the theorem in case $n < k$. Let $n = k$. Let F_1 be the uniquely determined neighbour corresponding to the line sums $(\mathcal{R}, \mathcal{C})$. We apply φ to F_1 once. Assume without loss of generality that an A-step is executed in column j.

First suppose that I consists of consecutive numbers. Then after moving the ones in column j, the length of the horizontal boundary in this column is equal to 4. When we delete this column, we are left with $k-1$ columns, so we can apply the induction hypothesis, which yields that the total length of the horizontal boundary at the end of the construction will be

$$L_h \leq 4(k-1) - 4 + 4 = 4k - 4.$$

Now suppose that I does not consist of consecutive numbers. Then we know that I is of the form $I = \{i_3, i_3+1, \ldots, t_1-1\} \cup \{t_3, t_3+1, \ldots, t_2\}$. So after moving the ones, the length of the horizontal boundary in column j is equal to 6. Also, we know in particular that the one in (c_j, j) was deleted and a one was added in (i_3, j).

The new parameters, after moving the ones and deleting column j, we denote by r'_i, b'_i and d'_i. The construction will in later steps execute an A-step in at most d'_{i_3-1} columns with sum i_3-1 and a B-step in at most $-d'_{i_3}$ columns with sum i_3-1. On the other hand, we currently have $b'_{i_3-1} - b'_{i_3}$ columns with sum i_3-1.

We know that $r_{i_3-1} \geq r_{i_3}$, and $r'_{i_3} = r_{i_3} - 1$. Both in the case $c_j = i_3 - 1$ and in the case $c_j < i_3 - 1$, we have $r'_{i_3-1} = r_{i_3-1}$, so

$$(b'_{i_3-1} - b'_{i_3}) - (d'_{i_3-1} - d'_{i_3}) = r'_{i_3-1} - r'_{i_3} = r_{i_3-1} - r_{i_3} + 1 \geq 1.$$

This means that there is at least one column with sum $i_3 - 1$ in which none of the later steps of the construction will be executed. This column will at the end of the construction therefore still have a horizontal boundary of length 2. If we delete this column entirely and then do the construction, exactly the same steps will be carried out. After all, the deleted column would never have been chosen to execute a step in anyway; also, deleting the column does not influence the choice of the set I in each step of the construction, as the only difference between the row sums of two consecutive rows that is changed, is between rows $i_3 - 1$ and i_3, but as $d_{i_3-1} \geq 0$ and $d_{i_3} < 0$, these rows will never both be in R^+ or both be in R^-.

By applying the induction hypothesis to the new situation with $n = k-2$, we find that the total length of the horizontal boundary at the end of the construction will be

$$L_h \leq 4(k-2) - 4 + 6 + 2 = 4k - 4.$$

This completes the induction step. □

Unfortunately, we cannot prove a similar result for the vertical boundary. In fact, we can find examples for which our construction gives a vertical boundary with a length as large as $\frac{4}{9}m^2 + \frac{4}{9}m + \frac{10}{9}$, see Example 6.5. However, we believe that there always exists a solution with a small boundary length, both horizontal and vertical.

Conjecture 6.5. *Let be given row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$, where $n = r_1 \geq r_2 \geq \ldots \geq r_m$ and $m = c_1 \geq c_2 \geq \ldots \geq c_n$. Assume that the line sums are consistent. There exists a set F_3 with line sums $(\mathcal{R}, \mathcal{C})$ for which*

$$L_h(F_3) \leq 4n - 4, \qquad L_v(F_3) \leq 4m - 4.$$

6.5 Examples

We give two families of examples for which we can prove that the construction of Theorem 6.1 gives the smallest possible length of the boundary.

Example 6.1. Let the number of columns n be odd and let $m = n$. Define line sums

$$\mathcal{C} = \mathcal{R} = (n, n-1, n-1, n-3, n-3, \ldots, 4, 4, 2, 2).$$

We calculate

$$(b_1, b_2, \ldots, b_n) = (n, n, n-2, n-2, \ldots, 3, 3, 1),$$
$$(d_1, d_2, \ldots, d_n) = (0, +1, -1, +1, -1, \ldots, +1, -1, +1, -1).$$

So $\alpha = \alpha(\mathcal{R}, \mathcal{C}) = \frac{n-1}{2}$. Theorem 6.2 tells us that the set F_2 constructed with Theorem 6.1 satisfies

$$L_h(F_2) \leq 2n + 2\alpha = 3n - 1, \qquad L_v(F_2) \leq 2m + 2\alpha = 3n - 1.$$

On the other hand, by Corollary 5.2 we know that for any set F with these line sums, we have

$$L_h(F) \geq 2n + \frac{n-1}{2} \cdot (1 - (-1)) + 2 \cdot 0 = 3n - 1,$$

and by symmetry also $L_v(F) \geq 3n-1$. This shows that F_2 has the smallest boundary among all sets F with these line sums. See for the constructed set F_2 in the case that $n = 9$ Figure 6.3(a). (This example is in fact a slightly modified version of Example 5.3.)

Example 6.2. Let $m = n \geq 2$. Define line sums

$$\mathcal{C} = \mathcal{R} = (n, 2, 2, 2, \ldots, 2, 2).$$

We calculate

$$(b_1, b_2, \ldots, b_n) = (n, n, 1, 1, \ldots, 1, 1),$$
$$(d_1, d_2, \ldots, d_n) = (0, n-2, -1, -1, \ldots, -1).$$

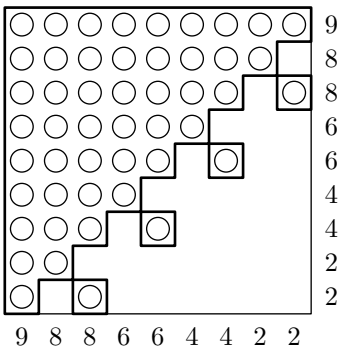
So $\alpha = \alpha(\mathcal{R}, \mathcal{C}) = n - 2$. Theorem 6.2 tells us that the set F_2 constructed with Theorem 6.1 satisfies

$$L_h(F_2) \leq 2n + 2\alpha = 4n - 4, \qquad L_v(F_2) \leq 2m + 2\alpha = 4n - 4.$$

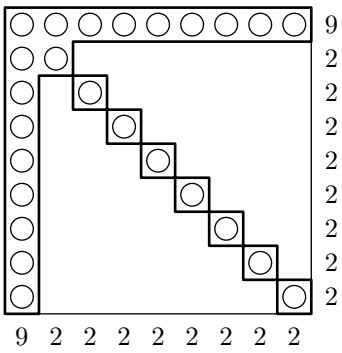
On the other hand, by Corollary 5.2 we know that for any set F with these line sums, we have

$$L_h(F) \geq 2n + 2(n-2) = 4n - 4,$$

and by symmetry also $L_v(F) \geq 4n-4$. This shows that F_2 has the smallest boundary among all sets F with these line sums. See for the constructed set F_2 in the case that $n = 9$ Figure 6.3(b). (This example is in fact a slightly modified version of Example 5.4.)



(a) Example 6.1 with $n = 9$. The horizontal and vertical boundary both have length 26.



(b) Example 6.2 with $n = 9$. The horizontal and vertical boundary both have length 32.

Figure 6.3: The constructed sets F_2 for two of the examples.

We can generalise Example 6.2 to larger α, in which case the bound of Theorem 6.2 is no longer sharp. However, in this case we can use Theorem 6.4 to prove that the horizontal boundary is the smallest possible, as shown below.

Example 6.3. Let k be a positive integer and let $m = kn - k + 1$. Define line sums

$$\mathcal{C} = (kn - k + 1, k + 1, k + 1, \ldots, k + 1, k + 1), \qquad \mathcal{R} = (n, 2, 2, \ldots, 2).$$

We calculate

$$(b_1, b_2, \ldots, b_m) = (\underbrace{n, n, \ldots, n}_{k+1 \text{ times}}, \underbrace{1, 1, \ldots, 1}_{kn-2k \text{ times}}),$$

$$(d_1, d_2, \ldots, d_m) = (0, \underbrace{n-2, n-2, \ldots, n-2}_{k \text{ times}}, \underbrace{-1, -1, \ldots, -1}_{kn-2k \text{ times}}).$$

Theorem 6.4 tells us that the set F_2 constructed with Theorem 6.1 satisfies

$$L_h(F_2) \leq 4n - 4.$$

On the other hand, by Corollary 5.2 we know that for any set F with these line sums, we have

$$L_h(F) \geq 2n + 2(n - 2) = 4n - 4.$$

This shows that F_2 has the smallest horizontal boundary among all sets F with these line sums.

The next example shows that the upper bound on α given in Proposition 6.3 can be achieved.

Example 6.4. Let k be a positive integer and let $m = n = 2k+1$. Define line sums

$$\mathcal{C} = \mathcal{R} = (2k+1, k+1, k+1, \ldots, k+1).$$

We calculate

$$(b_1, b_2, \ldots, b_m) = (\underbrace{2k+1, 2k+1, \ldots, 2k+1}_{k+1 \text{ times}}, \underbrace{1, 1, \ldots, 1}_{k \text{ times}}),$$

$$(d_1, d_2, \ldots, d_m) = (0, \underbrace{k, k, \ldots, k}_{k \text{ times}}, \underbrace{-k, -k, \ldots, -k}_{k \text{ times}}).$$

Hence

$$\alpha = \alpha(\mathcal{R}, \mathcal{C}) = k^2 = \frac{(m-1)(n-1)}{4}.$$

Finally we show by an example that the vertical boundary of the set F_2 constructed in Theorem 6.1 can become quite large, so it is not possible to prove a similar result as Theorem 6.4 for the vertical boundary.

Example 6.5. Let k be a positive integer and let $m = 3k+1$, $n = 3k$. Define line sums

$$\mathcal{C} = (3k+1, k+1, k+1, \ldots, k+1), \qquad \mathcal{R} = (3k, \underbrace{k+1, k+1, \ldots, k+1}_{2k \text{ times}}, \underbrace{k, k, \ldots, k}_{k \text{ times}}).$$

We calculate

$$(b_1, b_2, \ldots, b_m) = (\underbrace{3k, 3k, \ldots, 3k}_{k+1 \text{ times}}, \underbrace{1, 1, \ldots, 1}_{2k \text{ times}}),$$

$$(d_1, d_2, \ldots, d_m) =$$

$$(0, \underbrace{2k-1, 2k-1, \ldots, 2k-1}_{k \text{ times}}, \underbrace{-k, -k, \ldots, -k}_{k \text{ times}}, \underbrace{-(k-1), -(k-1), \ldots, -(k-1)}_{k \text{ times}}).$$

Hence $\alpha = \alpha(\mathcal{R}, \mathcal{C}) = 2k^2 - k$.

The construction executes $2k-1$ times an A-step, in each of the columns $3k$, $3k-1$, $\ldots$, $k+2$. In the first step (and every odd-numbered step after that) we have $I = \{k+2, k+3, \ldots, 2k+1\}$. At the beginning of the second step, however, the row sums in rows $k+2$, $k+3$, $\ldots$, $3k+1$ are all equal, so we have $I = \{2k+2, 2k+3, \ldots, 3k+1\}$. The same holds for every other even-numbered step. This means that at the end of the construction, the vertical boundary in each of the rows $k+2$, $k+3$, $\ldots$, $2k+1$ will be equal to $2(k+1)$, while the vertical boundary in each of the rows $2k+2$, $2k+3$, $\ldots$, $3k+1$ will be equal to $2k$. Adding the boundary of 2 in each of the rows 1, 2, $\ldots$, $k+1$, we find

$$L_v(F_2) = (k+1) \cdot 2 + k \cdot 2(k+1) + k \cdot 2k = 4k^2 + 4k + 2.$$

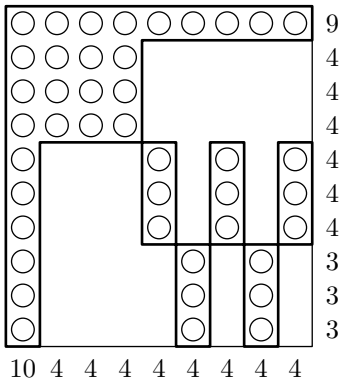


Figure 6.4: The constructed set F_2 from Example 6.5 with $k = 3$. The vertical boundary has length 50.

This is not linear in $m = 3k + 1$. It is in fact equal to $\frac{4}{9}m^2 + \frac{4}{9}m + \frac{10}{9}$. For the constructed set F_2 in the case that $k = 3$ see Figure 6.4.

It is clear that in fact there exists a set F with the same line sums, but with a much smaller vertical boundary, which supports Conjecture 6.5.

6.6 Generalising the results for arbitrary c_1 and r_1

In all results of the previous sections, we used the condition that $c_1 = m$ and $r_1 = n$. This is purely for convenience; it is not a necessary condition. We can easily generalise the results for the case that these conditions do not necessarily hold.

Consider a given set F with row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$, where $r_1 \geq r_2 \geq \ldots \geq r_m$ and $c_1 \geq c_2 \geq \ldots \geq c_n$, but not necessarily $c_1 = m$ and $r_1 = n$. Let F' be a set that is equal to F, except that we add a full row with index 0 and a full column with index 0, i.e.

$$F' = F \cup \{(0, j) : 0 \leq j \leq n\} \cup \{(i, 0) : 1 \leq i \leq m\}.$$

The row sums of F' are

$$\mathcal{R}' = (r'_0, r'_1, r'_2, \ldots, r'_m) = (n, r_1 + 1, r_2 + 1, \ldots, r_m + 1).$$

and the column sums of F' are

$$\mathcal{C}' = (c'_0, c'_1, c'_2, \ldots, c'_n) = (m, c_1 + 1, c_2 + 1, \ldots, c_n + 1).$$

It is easy to see that $\alpha(\mathcal{R}', \mathcal{C}') = \alpha(\mathcal{R}, \mathcal{C})$. Now consider the length of the horizontal boundary. For every j with $(1, j) \in F$, the horizontal boundary in column j of F' is equal to the horizontal boundary of column j in F. For every j with $(1, j) \notin F$, however, the horizontal boundary in column j of F' is 2 larger than the horizontal boundary in column j of F. (This also holds for column 0, where the horizontal boundary of F had length 0 and the horizontal boundary of F' has length 2.) Hence

$$L_h(F') = L_h(F) + 2(n + 1 - r_1).$$

Analogously, we have

$$L_v(F') = L_v(F) = 2(m + 1 - c_1).$$

By applying Theorems 6.2 and 6.4 as well as Proposition 6.3 to F' (with $n + 1$ columns and $m + 1$ rows), we acquire the following results.

Proposition 6.6. *Let be given row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$, where $r_1 \geq r_2 \geq \ldots \geq r_m$ and $c_1 \geq c_2 \geq \ldots \geq c_n$. Assume that the line sums are consistent. Let $\alpha = \alpha(\mathcal{R}, \mathcal{C})$. Then*

$$\alpha \leq \frac{mn}{4}.$$

Theorem 6.7. *Let be given row sums $\mathcal{R} = (r_1, r_2, \ldots, r_m)$ and column sums $\mathcal{C} = (c_1, c_2, \ldots, c_n)$, where $r_1 \geq r_2 \geq \ldots \geq r_m$ and $c_1 \geq c_2 \geq \ldots \geq c_n$. Assume that the line sums are consistent. Let $\alpha = \alpha(\mathcal{R}, \mathcal{C})$. Then there exists a set F_2 satisfying these line sums such that*

$$L_h(F_2) \leq \min(\ 2r_1 + 2\alpha,\ 2r_1 + 2n - 2\ )$$

and

$$L_v(F_2) \leq 2c_1 + 2\alpha.$$

CHAPTER 7

Boundary and shape of binary images

This chapter (with minor modifications) has been published as: Birgit van Dalen, "The boundary and shape of binary images", Discrete Mathematics 310 (2010) 2910-2918.

7.1 Introduction

In this chapter we will consider an unknown binary image, of which the length of the boundary and the area of the picture are given. These two values together contain some information about the general shape of the picture. We will study two properties of the shape in particular. First, using 4-adjacency, we can define the connected components of the picture [21]. We will prove sharp lower bounds for the size of the largest connected component.

The second question that we are interested in is: what is the size of the largest ball containing only ones? Or equivalently, considering for each cell the city block distance to the boundary [23], what is the maximal distance that occurs? We will derive some results related to this question, both in the case that the connected components are all simply connected (that is, they do not have any holes [21]) and in the general case.

After introducing some notation in Section 7.2, we will tackle the first question in

Section 7.3 and the second question in Section 7.4.

7.2 Definitions and notation

Let a *cell* in $\mathbb{R}^2$ be a square of side length 1 of which the vertices have integer coordinates. A *binary image* is a rectangle in $\mathbb{R}^2$ consisting of a number of cells, such that each cell inside the rectangle has been assigned a value 0 or 1. We will often refer to *a one* or *a zero* of a binary image, meaning a cell that has been assigned that value. When exactly N of the cells of a binary image have been assigned the value 1, we say that the image *consists of N ones*.

We will only consider 4-adjacency [21], and hence we will simply call two cells *neighbours* if they have a common edge. Two cells c and c' with value 1 in a binary image are called *connected* if there is a path $c = c_1, c_2, \ldots, c_n = c'$ of cells with value 1 such that c_i and c_{i+1} are neighbours for $1 \leq i \leq n-1$. Being connected is an equivalence relation and the equivalence classes are called the *connected components* of the image.

A connected component is said to contain a *hole* if there is a zero or a group of zeroes that is completely surrounded by ones of the connected component.

The *boundary* of a binary image consists of edges of cells. An edge belongs to the boundary if

- it is the common edge of two neighbouring cells, one of which has value 1 and one of which has value 0, or
- it belongs to exactly one cell within the rectangle (i.e. it is part of the outer edge of the rectangle) and that cell has value 1.

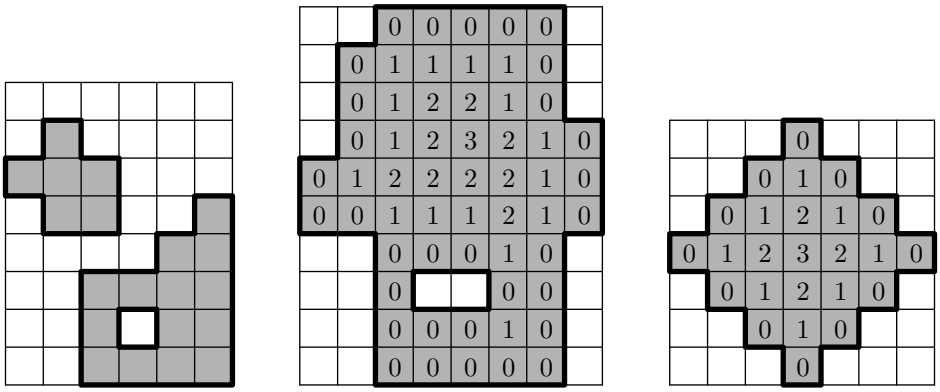
We define the *length of the boundary* as the number of edges that belong to the boundary. Some binary images with their boundaries are shown in Figure 7.1.

For each cell c with value 1 in a binary image, we define the *distance to the boundary* $d(c)$ recursively. A cell of which one of the edges belongs to the boundary has distance 0 to the boundary. For any other cell c with value 1, we set

$$d(c) = 1 + \min\{\ d(c') \mid c' \text{ and } c \text{ are neighbours }\}.$$

See Figure 7.1(b) for an example. In the literature this specific distance function is often referred to as city block distance [23].

For any integer $i \geq 1$ we define the *i-boundary* similarly to the boundary. An edge belongs to the i-boundary if it is a common edge of two cells with value 1, one of



(a) The length of the boundary of this image is 34.

(b) In each cell with value 1 the distance to the boundary is indicated.

(c) A ball with radius 3.

Figure 7.1: Three binary images. The grey cells have value 1.

which has distance $i-1$ to the boundary and the other of which has distance i to the boundary. The i-boundary separates the cells c with value 1 and $d(c) \geq i$ from the cells c with value 0 or $d(c) \leq i-1$.

We say that a binary image contains a *ball with radius* k if there is a cell with value 1 that has distance at least k to the boundary. In that case the connected component containing this cell must contain at least $2k^2+2k+1$ cells. See also Figure 7.1(c).

7.3 Largest connected component

Let F be a binary image consisting of m^2 ones. If the ones are arranged into one square with side length m, then the boundary of F has length $4m$. This is the smallest possible boundary for this number of ones (see also Lemma 7.2). If the length of the boundary is greater than $4m$, then the image may contain more than one connected component. We can, however, still prove a good lower bound on the size of the largest connected component. We will do this in two cases: when the boundary has length $4m$ plus some constant, and when the boundary has length $4m$ times some constant. In the second case we will also generalise to an image consisting of N ones, where N does not need to be a square.

First we prove two lemmas.

Lemma 7.1. *Let $r \geq 2$ and $0 \leq A < B$ be integers and let S be an integer satisfying $rA \leq S \leq rB$. The minimal value of*

$$f(k_1, k_2, \ldots, k_r) = \sqrt{k_1} + \sqrt{k_2} + \ldots + \sqrt{k_r}$$

where $k_1, k_2, \ldots, k_r$ are integers in the interval $[A, B]$ for which $k_1 + k_2 + \cdots + k_r = S$, is attained at some r-tuple $(k_1, k_2, \ldots, k_r)$ for which $k_i \notin \{A, B\}$ holds for at most one value of i.

Proof. We argue by contradiction. Suppose the minimal value of f is attained at some r-tuple $(k_1, k_2, \ldots, k_r)$ for which we have $k_1, k_2 \notin \{A, B\}$. Let $S' = k_1 + k_2$. Consider all possible values of $g(x) = \sqrt{x} + \sqrt{S' - x}$, where x is an integer in the interval $[A, B]$ such that $S' - x \in [A, B]$ as well. Our assumption implies that the minimal value of g is attained when $x = k_1$ and also when $x = k_2$. We now distinguish between two cases.

First suppose $k_1 + k_2 \leq A + B$. When we take $x = A$, we have $S' - x = k_1 + k_2 - A \leq B$ and $S' - x \geq A$, so $S' - x \in [A, B]$. Hence by our assumption $g(A) \geq g(k_1) = g(k_2)$. On the other hand, the continuous function $g(x) = \sqrt{x} + \sqrt{S' - x}$ on the interval $[0, S'] \subset \mathbb{R}$ is monotonically increasing on $[0, S'/2]$ and monotonically decreasing on $[S'/2, S']$. At least one of k_1, k_2 must be in $[0, S'/2]$ and $A < k_1, k_2$, so we must have $g(A) < g(k_1) = g(k_2)$, which yields a contradiction.

Now suppose $k_1 + k_2 > A + B$. When we take $x = B$, we have $S' - x = k_1 + k_2 - B > A$ and $S' - x \leq B$, so $S' - x \in [A, B]$. Similarly to above, this leads to a contradiction. □

Note that one could also prove Lemma 7.1 by using classical results from convex geometry.

Lemma 7.2. *Let k be a positive integer. A binary image consisting of k ones has a boundary of length at least $4\sqrt{k}$.*

Proof. First suppose that there is just one connected component. Let the smallest rectangle containing the component have side lengths a and b. The boundary of the rectangle has length equal to or smaller than the boundary of the original image, so the boundary of the image has length at least $2a + 2b$. On the other hand, we have $k \leq ab$, since all k ones are contained in the rectangle. As $\frac{a+b}{2} \geq \sqrt{ab} \geq \sqrt{k}$, the boundary has length at least $4\sqrt{k}$.

Now suppose that there are r connected components consisting of $k_1, k_2, \ldots, k_r$ ones respectively. Then the boundary of the image has length at least $4\sqrt{k_1} + 4\sqrt{k_2} + \cdots + 4\sqrt{k_r}$. So it suffices to prove

$$\sqrt{k_1} + \sqrt{k_2} + \cdots + \sqrt{k_r} \geq \sqrt{k_1 + k_2 + \cdots + k_r},$$

which can easily be done by squaring both sides. □

Note that similar results as Lemma 7.2 are in [22], although there a slightly different definition for the length of the boundary is used.

We will now prove our first theorem, concerning an image with boundary only an additive constant larger than the minimal length.

Theorem 7.3. *Let m and c be positive integers. Suppose a binary image F consists of m^2 ones and has a boundary of length $4m+4c$. If m is sufficiently large compared to c, then the largest connected component of F consists of at least $m^2 - c^2$ ones.*

Proof. Suppose to the contrary that the largest connected component of F consists of $t \leq m^2-c^2-1$ ones. We distinguish between two cases. First assume that $t \geq c^2+1$. By Lemma 7.2 the boundary has length at least $4\sqrt{t} + 4\sqrt{m^2 - t}$, while it is given to be equal to $4m + 4c$. So we have

$$\sqrt{t} + \sqrt{m^2 - t} \leq m + c.$$

By Lemma 7.1 the smallest possible value of $\sqrt{t} + \sqrt{m^2 - t}$ is attained when $t = m^2 - c^2 - 1$ (and when $t = c^2 + 1$). So we must have

$$\sqrt{m^2 - c^2 - 1} + \sqrt{c^2 + 1} \leq m + c.$$

Subtracting $\sqrt{c^2 + 1}$ from both sides and squaring gives

$$m^2 - c^2 - 1 \leq m^2 + 2mc + 2c^2 + 1 - 2(m + c)\sqrt{c^2 + 1}.$$

This is equivalent to

$$m \leq \frac{3c^2 + 2 - 2c\sqrt{c^2 + 1}}{2\sqrt{c^2 + 1} - 2c}.$$

Hence for sufficiently large m, this case is impossible.

Now consider the case that $t \leq c^2$. Suppose we have r connected components. Then $r \geq \frac{m^2}{t} \geq \frac{m^2}{c^2}$. The boundary of each connected component has length at least 4, so the total length of the boundary is at least $4r \geq 4\frac{m^2}{c^2}$. Therefore, we must have

$$\frac{m^2}{c^2} \leq m + c.$$

For sufficiently large m, this is also impossible. We conclude that the largest connected component must consist of at least $m^2 - c^2$ ones. □

The bound given in this theorem is sharp: suppose the ones in the image are grouped in two connected components, an $(m-c)\times(m+c)$ rectangle and a $c\times c$ square. The boundary of the rectangle then has length $4m$, while the boundary of the square has length $4c$, so in total the boundary of F has length $4m+4c$.

The next theorem concerns a binary image consisting of m^2 ones and having a boundary of length a constant times $4m$.

Theorem 7.4. *Let m and c be positive integers such that m is divisible by c and $m \geq c(c+1)$. Suppose a binary image F consists of m^2 ones and has a boundary of length $4mc$. Then the largest connected component of F consists of at least $\frac{m^2}{c^2}$ ones.*

Proof. Let n be an integer such that $m = nc$. Then F contains c^2n^2 ones and the boundary of F has length $4c^2n$. We want to prove that the largest connected component of F consists of at least n^2 ones. Suppose to the contrary that the largest connected component of F consists of $t \leq n^2-1$ ones. Let r be the number of connected components, and let k_i be the number of ones in the i-th component, $1 \leq i \leq r$. Then by Lemma 7.2 the boundary of F is at least equal to

$$4\left(\sqrt{k_1}+\sqrt{k_2}+\cdots+\sqrt{k_r}\right). \tag{7.1}$$

We will try to determine the minimal value of this and show that it is greater than $4c^2n$.

The integers $k_1, \ldots, k_r$ are all in the interval $[1,t]$ and at least one of them is equal to t. For our purposes we may as well assume that $k_i \in [1, n^2-1]$: by doing so we may find a minimal value that is even smaller than the actual minimal value, but if we can still prove that it is greater than $4c^2n$, we are done anyway.

The integers $k_1, \ldots, k_r$ furthermore satisfy $k_1+k_2+\cdots+k_r = c^2n^2$. Also, since $c^2\cdot(n^2-1) < c^2n^2$, we know that $r \geq c^2+1$.

By Lemma 7.1 the minimal value is attained at some r-tuple $(k_1,\ldots,k_r)$ of which at least $r-1$ elements are equal to 1 or n^2-1. Up to order, there is only one such r-tuple satisfying $k_1+\cdots+k_r = c^2n^2$. After all, suppose there are two such r-tuples, $(k_1 \leq k_2 \leq \ldots \leq k_r)$ and $(k'_1 \leq k'_2 \leq \ldots \leq k'_r)$. Let i be such that $k_i = 1$, $k_{i+1} > 1$ and let j be such that $k'_j = 1$, $k'_{j+1} > 1$. If $i = j$, then the two r-tuples must be equal, as the sum of the elements is equal. So assume that $i \neq j$, say, $i > j$. Then $k_{i+2} = \ldots = k_r = n^2-1$ and $k'_{j+2} = \ldots = k'_r = n^2-1$. Since the two sums of the r-tuples must be equal, we must have $k_{i+1} - k'_{j+1} = (i-j)(n^2-2)$. Since $k'_{j+1} \geq 2$ and $k_{i+1} \leq n^2-1$, the left-hand side can be at most n^2-3, while the right-hand side is at least n^2-2, which is a contradiction.

The unique r-tuple (ordered non-decreasingly) that satisfies the requirements is given

by

$$k_1 = \ldots = k_{r-v-1} = 1, \quad k_{r-v} = (c^2-v)n^2+2v+1-r, \quad k_{r-v+1} = \ldots = k_r = n^2-1,$$

where v is the unique positive integer such that

$$(c^2 - v - 1)n^2 + 2v + 3 \leq r \leq (c^2 - v)n^2 + 2v.$$

Note that the choice of v ensures that $1 \leq k_{r-v} \leq n^2 - 1$. This r-tuple must give the minimal value of (7.1) under the conditions that $k_i \in [1, n^2 - 1]$ and $k_1 + \cdots + k_r = c^2n^2$. Therefore it now suffices to prove that

$$(r - v - 1) + \sqrt{(c^2 - v)n^2 + 2v + 1 - r} + v\sqrt{n^2 - 1} > c^2n. \tag{7.2}$$

From $m \geq c(c + 1)$ we have $n \geq c + 1$. This implies $n^2 > c^2 + 1$, and from that we derive $v \leq c^2$: if $v \geq c^2+1$, then $\sum_i k_i \geq (c^2+1)(n^2-1) = c^2n^2+n^2-c^2-1 > c^2n^2$, which contradicts $\sum_i k_i = c^2n^2$. We now distinguish between two cases: $v \leq c^2 - 1$ and $v = c^2$.

First suppose $v \leq c^2 - 1$. Consider the function $f(x) = x + \sqrt{S - x}$ on the interval $[A, S - 1]$. Its derivative is $f'(x) = 1 - \frac{1}{2\sqrt{S-x}}$, which is positive for $x \leq S - 1$, so the function is strictly increasing on the interval. Hence for all $x \in [A, S - 1]$ we have $f(x) \geq f(A)$. If we apply this for $A = (c^2 - v - 1)n^2 + 2v + 3$, $S = (c^2 - v)n^2 + 2v + 1$ and $x = r$, we find that

$$(r - v - 1) + \sqrt{(c^2 - v)n^2 + 2v + 1 - r} \geq (c^2 - v - 1)n^2 + v + 2 + \sqrt{n^2 - 2}.$$

As $n \geq c + 1 \geq 2$, we have $n^2 - 2 \geq (n - 1)^2$, hence the left-hand side of (7.2) is at least

$$(c^2 - v - 1)n^2 + v + 2 + \sqrt{(n-1)^2} + v\sqrt{(n-1)^2}$$

As $c^2 - v - 1 \geq 0$ and $n^2 \geq n$, this is at least

$$(c^2 - v - 1)n + v + 2 + (v + 1)(n - 1) = c^2n + 1 > c^2n,$$

which proves that (7.2) holds in this case.

Now suppose $v = c^2$. Then $r \leq 2c^2$. Recall that we also have $r \geq c^2 + 1$. We have to prove

$$r - c^2 - 1 + \sqrt{2c^2 + 1 - r} + c^2\sqrt{n^2 - 1} > c^2n.$$

We again apply $f(x) \geq f(A)$ with $f(x)$ as above, now with $A = c^2 + 1$, $S = 2c^2 + 1$ and $x = r$. We find

$$r - c^2 - 1 + \sqrt{2c^2 + 1 - r} \geq (c^2 + 1) - c^2 - 1 + \sqrt{2c^2 + 1 - (c^2 + 1)} = c.$$

Hence it suffices to prove

$$c + c^2\sqrt{n^2 - 1} > c^2n.$$

This is equivalent to

$$c^4(n^2-1) > (c^2n-c)^2,$$

which we can rewrite as

$$n > \tfrac{1}{2}(c+\tfrac{1}{c}).$$

This follows from $n \geq c+1$, hence (7.2) holds in this case as well. This completes the proof of the theorem. □

The bound given in this theorem is sharp: suppose the ones in the image are grouped in c^2 squares of side length $\frac{m}{c}$, containing $\frac{m^2}{c^2}$ ones each. Then the boundary of each square has length $4\frac{m}{c}$, so in total the boundary of F has length $4mc$.

The condition that m, c and $\frac{m}{c}$ be integers does not seem to be very essential in the above theorem or proof. In fact, in a similar way (though slightly more technical) we can prove a more general result in which this condition is omitted.

Theorem 7.5. *Let N be a positive integer and $c > 1$ a real number. Suppose a binary image F consists of N ones and has a boundary of length at most $4c\sqrt{N}$. If N is sufficiently large compared to c, then the largest connected component of F consists of more than $\frac{N}{c^2} - 1$ ones.*

Proof. Let $q = \frac{\sqrt{N}}{c} \in \mathbb{R}$. Then F contains c^2q^2 ones and the boundary has length at most $4c^2q$. Let $1 \leq \varepsilon < 2$ be such that $q^2 - \varepsilon$ is an integer, and suppose there are $t \leq q^2 - \varepsilon$ ones in the largest connected component of F. We will derive a contradiction, from which the theorem then follows. Let r be the number of connected components, and let k_i be the number of ones in the i-th connected component, $1 \leq i \leq r$.

Similarly to the proof of Theorem 7.4 it suffices to prove that (for sufficiently large q compared to c) the minimal value of

$$\sqrt{k_1} + \sqrt{k_2} + \cdots + \sqrt{k_r},$$

where $k_1, \ldots, k_r$ are integers in the interval $[1, q^2 - \varepsilon]$ satisfying $k_1 + k_2 + \cdots + k_r = c^2q^2$, is greater than c^2q. Also similarly to the proof of Theorem 7.4, that minimal value is attained when

$$\begin{aligned} k_1 &= \ldots = k_{r-v-1} = 1, \\ k_{r-v} &= (c^2 - v)q^2 + (\varepsilon+1)v + 1 - r, \\ k_{r-v+1} &= \ldots = k_r = q^2 - \varepsilon, \end{aligned}$$

where v is the unique positive integer such that

$$(c^2 - v - 1)q^2 + (\varepsilon+1)v + \varepsilon + 2 \leq r \leq (c^2 - v)q^2 + (\varepsilon+1)v.$$

It suffices to prove that

$$(r - v - 1) + \sqrt{(c^2 - v)q^2 + (\varepsilon + 1)v + 1 - r} + v\sqrt{q^2 - \varepsilon} > c^2 q. \qquad (7.3)$$

Let $c^2 + \delta$ be the smallest integer strictly greater than c^2. Then we can choose q large enough such that $\delta q^2 > 2(c^2 + \delta)$, which is equivalent to $(c^2 + \delta)(q^2 - 2) > c^2 q^2$. As $\varepsilon < 2$, we then also have $(c^2 + \delta)(q^2 - \varepsilon) > c^2 q^2$. As $c^2 q^2 \geq v(q^2 - \varepsilon)$, we find $v \leq c^2 + \delta - 1 \leq c^2$. We now distinguish between three cases: the case $v \leq c^2 - 1$, the case $c^2 - 1 < v < c^2$ and the case $v = c^2$. (Note that depending on whether c^2 is an integer, only one of the two latter cases may occur.)

First suppose $v \leq c^2 - 1$. We have $r \geq (c^2 - v - 1)q^2 + (\varepsilon + 1)v + \varepsilon + 2$ and therefore (similarly to the proof of Theorem 7.4)

$$(r - v - 1) + \sqrt{(c^2 - v)q^2 + (\varepsilon + 1)v + 1 - r} \geq (c^2 - v - 1)q^2 + \varepsilon v + \varepsilon + 1 + \sqrt{q^2 - \varepsilon - 1}.$$

Furthermore, assuming $q \geq 2$ we have $\sqrt{q^2 - \varepsilon} > q - \varepsilon$ and $\sqrt{q^2 - \varepsilon - 1} \geq q - \varepsilon - 1$, hence the left-hand side of (7.3) is strictly greater than

$$(c^2 - v - 1)q^2 + \varepsilon v + \varepsilon + 1 + (q - \varepsilon - 1) + v(q - \varepsilon) = (c^2 - v - 1)q^2 + (v + 1)q.$$

As $c^2 - v - 1 \geq 0$ and $q^2 \geq q$, this is at least

$$(c^2 - v - 1)q + (v + 1)q = c^2 q,$$

which proves (7.3) in this case.

Now suppose $c^2 - 1 < v < c^2$. The largest connected component of F contains fewer than q^2 ones, and F contains $c^2 q^2$ ones; hence the number of connected components is greater than c^2. This implies

$$\begin{aligned}(r - v - 1) + \sqrt{(c^2 - v)q^2 + (\varepsilon + 1)v + 1 - r} \\ \geq \; c^2 - v - 1 + \sqrt{(c^2 - v)q^2 + (\varepsilon + 1)v + 1 - c^2}.\end{aligned}$$

We have $(\varepsilon + 1)v - c^2 + 1 > 0$, hence

$$\sqrt{(c^2 - v)q^2 + (\varepsilon + 1)v + 1 - c^2} > \sqrt{(c^2 - v)q^2} = q\sqrt{c^2 - v}.$$

Also, $c^2 - v - 1 > 0$ and (as above) $\sqrt{q^2 - \varepsilon} > q - \varepsilon$. Therefore it suffices to prove

$$q\sqrt{c^2 - v} + v(q - \varepsilon) \geq c^2 q,$$

which is equivalent to

$$(\sqrt{c^2 - v} - (c^2 - v))q \geq \varepsilon v.$$

As $\varepsilon \leq 2$, it also suffices to prove

$$(\sqrt{c^2 - v} - (c^2 - v))q \geq 2v.$$

Since $0 < c^2 - v < 1$, we have $(\sqrt{c^2 - v} - (c^2 - v)) > 0$. Now note that for a given c, there is at most one possible value for v satisfying $c^2 - 1 < v < c^2$, as v is an integer. This value does not depend on q. Therefore we can choose q large enough such that it satisfies

$$(\sqrt{c^2 - v} - (c^2 - v))q \geq 2v.$$

Hence (7.3) holds for sufficiently large q.

Finally suppose $v = c^2$. In this case (7.3) transforms into

$$(r - c^2 - 1) + \sqrt{(\varepsilon + 1)c^2 + 1 - r} + c^2\sqrt{q^2 - \varepsilon} > c^2 q.$$

As above, we have $r \geq c^2$, hence

$$(r - c^2 - 1) + \sqrt{(\varepsilon + 1)c^2 + 1 - r} \geq (c^2 - c^2 - 1) + \sqrt{(\varepsilon + 1)c^2 + 1 - c^2} = -1 + \sqrt{\varepsilon c^2 + 1}.$$

As $\varepsilon \geq 1$, we have $\sqrt{\varepsilon c^2 + 1} > c$. Also, $\varepsilon \leq 2$. Therefore it suffices to prove

$$-1 + c + c^2\sqrt{q^2 - 2} > c^2 q.$$

After some rewriting, this is equivalent to

$$q(2c^3 - 2c^2) \geq 2c^4 + c^2 - 2c + 1.$$

Since $2c^3 - 2c^2 > 0$, this is true for sufficiently large q. Hence also in this case (7.3) holds for sufficiently large q. This completes the proof of the theorem. □

7.4 Balls of ones in the image

In the previous section we proved bounds on the size of the largest connected component of an image. However, we are also interested in the shapes of such components. It seems likely that if the boundary is small compared to the number of ones, then there needs to be a large ball-shaped cluster of ones somewhere in the image. In this section we will prove lower bounds on the radius of such a ball.

First we prove some lemmas about the length of the i-boundary of an image.

Lemma 7.6. *In a binary image, the length of the* 1*-boundary is at most three times the length of the boundary.*

Proof. We can split the boundary into a number of simple, closed paths. (If there is more than one way to do this, we just pick one.) Let $\mathcal{P}$ be one of those paths, and denote its length by L_0. Let S be the set of cells that have value 1 and have an edge in common with $\mathcal{P}$. Either the cells in S are all on the outside of the path, or they

are all on the inside of the path. (Note that we are using a discrete analog of the Jordan Curve Theorem [18].) Let L_1 be the number of edges of cells in S that are part of the 1-boundary. (These edges do not necessarily form a simple, closed path.) We will prove a bound on L_1 in terms of L_0.

Consider all the pairs of edges of $\mathcal{P}$ having a vertex in common. There are three possible configurations, as shown in Figure 7.2. We call a pair of edges that form a straight line segment a *straight connection*. The other two types we call *corners*. A corner is of type I if both edges belong to the same cell with value 0; it is of type II if both edges belong to the same cell with value 1.

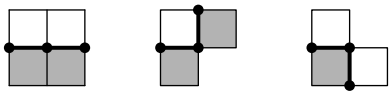


Figure 7.2: From left to right: a straight connection, a corner of type I and a corner of type II. Such corners may also be called *reentrant* and *salient* respectively [10].

We distinguish between three cases.

Case 1. The path $\mathcal{P}$ consists of only four edges, and the cell enclosed by $\mathcal{P}$ has value 1. In this case $L_0 = 4$ and $L_1 = 0$.

Case 2. The path $\mathcal{P}$ consists of more than four edges, and the cells in S are on the inside of $\mathcal{P}$. Let a be the number of straight connections and let b be the number of corners of type I. Then the number of corners of type II must be $b + 4$. We have $L_0 = a + 2b + 4$. Each edge of $\mathcal{P}$ is the edge of a cell in S, and each cell in S has at least one edge in $\mathcal{P}$. In a corner of type II, we count the same cell in S twice, so the number of cells in S is $a + 2b + 4 - (b + 4) = a + b$. Now we calculate an upper bound for L_1. Each cell in S has four edges, of which in total $a + 2b + 4$ belong to $\mathcal{P}$. Also, the two cells in S next to a straight connection share an edge that does not belong to either the boundary or the 1-boundary. Hence

$$L_1 \leq 4(a + b) - (a + 2b + 4) - 2a = a + 2b - 4 = L_0 - 8.$$

Case 3. The cells in S are on the outside of $\mathcal{P}$. Let a be the number of straight connections and let b be the number of corners of type I. Then $b \geq 4$ and there are $b - 4$ corners of type II. Similarly to above, we find $L_0 = a + 2b - 4$, the number of cells in S is $a + b$ and

$$L_1 \leq 4(a + b) - (a + 2b - 4) - 2a = a + 2b + 4 = L_0 + 8.$$

Since $L_0 \geq 4$, we have $L_1 \leq 3L_0$. This inequality obviously also holds in Cases 1 and 2.

Let l_0 be the length of the boundary and let l_1 be the length of the 1-boundary of this image. Then l_0 is the sum of the lengths L_0 of all the paths $\mathcal{P}$, while l_1 is at most the sum of the lengths L_1 (we have counted each edge of the 1-boundary at least once). We conclude $l_1 \leq 3l_0$. □

Lemma 7.7. *Let $i \geq 1$ be an integer. In a binary image, the length of the $(i+1)$-boundary is at most $\frac{2i+3}{2i+1}$ times the length of the i-boundary.*

Proof. Recall that the i-boundary consists of the edges between cells with distance $i-1$ to the boundary and cells with distance i to the boundary. Just like the boundary, we can split the i-boundary into a number of simple, closed paths. Let $\mathcal{P}$ be one of those paths, and denote its length by L_i. Let S be the set of cells that have distance i to the boundary and have an edge in common with $\mathcal{P}$. Either the cells in S are all on the outside of the path, or they are all on the inside of the path. Let L_{i+1} be the number of edges of cells in S that are part of the $(i+1)$-boundary. (These edges do not necessarily form a simple, closed path.) Analogously to the proof of Lemma 7.6 we can prove a bound on L_{i+1} in terms of L_i:

- In Case 1, $L_i = 4$ and $L_{i+1} = 0$.
- In Case 2, $L_{i+1} \leq L_i - 8$.
- In Case 3, $L_{i+1} \leq L_i + 8$.

In Case 3, where in Lemma 7.6 we had $L_0 \geq 4$, we now have $L_i \geq 8i + 4$. We will prove this here. Somewhere within $\mathcal{P}$ there must be a cell c with value 0. A horizontal line drawn through c must cross $\mathcal{P}$ somewhere to the left of c and somewhere to the right of c. Between those two edges of $\mathcal{P}$ there must be at least $2i+1$ cells: c and two cells at distance j for each j with $0 \leq j \leq i-1$. Similarly, there are at least $2i+1$ cells stacked in the vertical direction between two pieces of $\mathcal{P}$. Hence $L_i \geq 4(2i+1)$.

Since we have $L_{i+1} \leq L_i + 8$, we may conclude in Case 3 that

$$\frac{L_{i+1}}{L_i} \leq 1 + \frac{8}{L_i} \leq 1 + \frac{8}{8i+4} = \frac{2i+3}{2i+1},$$

and hence $L_{i+1} \leq \frac{2i+3}{2i+1} \cdot L_i$. Obviously this inequality holds in Cases 1 and 2 as well.

Let l_i be the length of the i-boundary and let l_{i+1} be the length of the $(i+1)$-boundary of this image. As in the proof of Lemma 7.6 we conclude $l_{i+1} \leq \frac{2i+3}{2i+1} l_i$. □

Lemma 7.8. *Let $i \geq 0$ be an integer. In a binary image, the number of cells at distance i from the boundary is at most $2i+1$ times the length of the boundary.*

Proof. For $i \geq 0$, let A_i be the number of cells at distance i from the boundary. For $i \geq 1$, let l_i be the length of the i-boundary. Let l_0 be the length of the boundary. Each cell at distance i from the boundary, $i \geq 1$, has at least one neighbour at distance $i-1$ from the boundary, hence the number of cells at distance i from the boundary is at most equal to the length of the i-boundary. Similarly, the number of cells at distance 0 from the boundary is at most l_0. Furthermore, for $i \geq 1$ we have by Lemmas 7.6 and 7.7 that

$$l_i \leq \frac{2i+1}{2i-1} \cdot l_{i-1} \leq \frac{2i+1}{2i-1} \cdot \frac{2i-1}{2i-3} \cdot l_{i-2} \leq \ldots \leq \frac{2i+1}{2i-1} \cdot \frac{2i-1}{2i-3} \cdot \cdots \cdot \frac{3}{1} \cdot l_0 = (2i+1)l_0.$$

For $i = 0$ it trivially holds that $l_i \leq (2i+1)l_0$. Hence for $i \geq 0$ we have

$$A_i \leq (2i+1)l_0.$$

□

We now use these lemmas to prove our next theorem.

Theorem 7.9. *Let N and l be positive integers. Suppose a binary image F consists of N ones and has a boundary of length l. Then the image contains a ball of radius $\left\lceil \sqrt{\frac{N}{l}} - 1 \right\rceil$.*

Proof. For $i \geq 0$, let A_i be the number of cells with value 1 at distance i from the boundary. Let k be a positive integer. Recall that F contains a ball with radius k if there is a cell with value 1 that has distance at least k to the boundary. Using Lemma 7.8 we can find an upper bound for the number of cells with value 1 and distance to the boundary at most $k-1$:

$$A_0 + A_1 + A_2 + \cdots + A_{k-1} \leq (1 + 3 + \cdots + 2k-1)l = k^2 l.$$

Hence if $N > k^2 l$, then F contains a ball with radius k.

Now let $k = \left\lceil \sqrt{\frac{N}{l}} - 1 \right\rceil$ and assume that it is a positive integer (if it is not, then the theorem is trivial). Then $k < \sqrt{\frac{N}{l}}$, hence $N > k^2 l$. Therefore F contains a ball with radius $\left\lceil \sqrt{\frac{N}{l}} - 1 \right\rceil$. □

Remark 7.1. *Suppose as in Theorem 7.5 that the boundary of F has length $4c\sqrt{N}$ for some $c \in \mathbb{R}$. Then Theorem 7.9 says that F contains a ball of radius $\left\lceil \sqrt{\frac{\sqrt{N}}{4c}} - 1 \right\rceil$. This ball contains approximately $\frac{\sqrt{N}}{2c}$ ones. On the other hand, Theorem 7.5 tells us that there exists a connected component with more than $\frac{N}{c^2} - 1$ ones. This is roughly four times the square of the size of the ball, but this component does not need to be ball-shaped.*

If the binary image contains no holes, then we can prove a much stronger result, by sharpening the lemmas in this section.

Theorem 7.10. *Let N and l be positive integers. Suppose a binary image F consists of N ones and has a boundary of length l. Furthermore assume that none of the connected components of F contains any holes. Then the image contains a ball of radius $\lfloor \frac{N}{l} \rfloor$.*

Proof. For $i \geq 0$, let A_i be the number of cells with value 1 at distance i from the boundary. Case 3 in the proofs of Lemmas 7.6 and 7.7 does not occur if the connected components of F do not contain any holes. This means that in Lemma 7.6 we can conclude that the length of the 1-boundary is strictly smaller than the length of the boundary, and in Lemma 7.7 that the length of the $(i+1)$-boundary is strictly smaller than the length of the i-boundary. Hence we have for all $i \geq 0$

$$A_i < A_{i-1} < \ldots < A_0 < l.$$

Let k be a positive integer. Then the number of cells with value 1 and distance to the boundary at most $k-1$ is

$$A_0 + A_1 + A_2 + \cdots + A_{k-1} < kl.$$

Hence if $N \geq kl$, then F contains a ball of radius k. This is obviously the case for $k = \lfloor \frac{N}{l} \rfloor$. □

We will show by two examples that the bounds from the previous two theorems are nearly sharp.

Example 7.1. Let u and c be positive integers. Consider a square of ones of side length $cu^2 + u - 1$. Denote the cells in the square by coordinates (i, j), where $1 \leq i, j \leq cu^2 + u - 1$. For all i and j that are divisible by u, we change the value of cell (i, j) from 1 to 0. Let F be the resulting binary image (see also Figure 7.3(a)). The number of ones of F is

$$N = (cu^2 + u - 1)^2 - (cu)^2 = c^2u^4 + 2cu^3 + (-c^2 - 2c + 1)u^2 - 2u + 1.$$

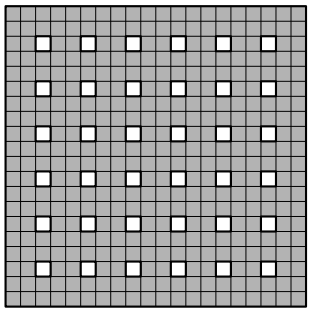
The length of the boundary is

$$l = 4(cu^2 + u - 1) + 4c^2u^2 = 4(c^2 + c)u^2 + 4u - 4.$$

If u is very large, we have $N \approx c^2u^4$ and $l \approx 4(c^2 + 2)u^2$. So according to Theorem 7.9, F should contain a ball of radius approximately

$$\sqrt{\frac{N}{l}} \sim \sqrt{\frac{c^2u^4}{4(c^2 + c)u^2}} = \frac{1}{2} \cdot \sqrt{\frac{c^2}{c^2 + c}} \cdot u, \qquad u \to \infty.$$

If u is odd, F in fact contains a ball of radius $u - 2$. If u is even, then F contains a ball of radius $u - 1$. See also Figures 7.3(b) and 7.3(c).



(a) The binary image F from the example, where $u = 3$ and $c = 2$.

	0	1	1	0	
0	1	2	2	1	0
1	2	3	3	2	1
1	2	3	3	2	1
0	1	2	2	1	0
	0	1	1	0	

(b) When u is odd, the radius of the largest ball that fits in the image is $u - 2$.

	0	1	2	1	0	
0	1	2	3	2	1	0
1	2	3	4	3	2	1
2	3	4	5	4	3	2
1	2	3	4	3	2	1
0	1	2	3	2	1	0
	0	1	2	1	0	

(c) When u is even, the radius of the largest ball that fits in the image is $u - 1$.

Figure 7.3: Some illustrations for Example 7.1.

Example 7.2. Let F consist of a rectangle of ones, with side lengths a and ta, where $t \geq 1$. Then the number of ones is equal to ta^2, while the length of the boundary is equal to $2(t + 1)a$. So according to Theorem 7.10, F should contain a ball of radius $\lfloor \frac{ta^2}{2(t+1)a} \rfloor = \lfloor \frac{t}{t+1} \frac{a}{2} \rfloor$. The actual radius of the largest ball contained in F is equal to $\lfloor \frac{a-1}{2} \rfloor$.

Bibliography

[1] A. Alpers, *Instability and stability in discrete tomography*, Ph.D. thesis, Technische Universität München, ISBN 3-8322-2355-X, Shaker Verlag, Aachen (2003).

[2] A. Alpers, S. Brunetti, Stability results for the reconstruction of binary pictures from two projections, *Image and Vision Computing* 25 (2007) 1599-1608.

[3] A. Alpers, P. Gritzmann, L. Thorens, Stability and instability in discrete tomography, *Lectures Notes in Computer Science 2243: Digital and Image Geometry* (2001) 175-186.

[4] R.P. Anstee, The network flows approach for matrices with given row and column sums, *Discrete Mathematics* 44 (1983) 125-138.

[5] E. Balogh, A. Kuba, C. Dévényi, A. Del Lungo, Comparison of algorithms for reconstructing hv-convex discrete sets, *Linear Algebra and its Applications* 339 (2001) 23-35.

[6] E. Barcucci, A. Del Lungo, M. Nivat, R. Pinzani, Reconstructing convex polyominoes from horizontal and vertical projections, *Theoretical Computer Science* 155 (1996) 321-347.

[7] K.J. Batenburg, S. Bals, J. Sijbers, C. Kübel, P.A. Midgley, J.C. Hernandez, U. Kaiser, E.R. Encina, E.A. Coronado, G. Van Tendeloo, 3D imaging of nanomaterials by discrete tomography, *Ultramicroscopy* 109 (2009) 730-740.

[8] M. Chrobak, C. Dürr, Reconstructing hv-convex polyominoes from orthogonal projections, *Information Processing Letters* 69 (1999) 283-289.

[9] G. Dahl, T. Flatberg, Optimization and reconstruction of hv-convex (0,1)-matrices, *Discrete Applied Mathematics* 151 (2005) 93-105.

[10] A. Daurat, M. Nivat, Salient and reentrant points of discrete sets, *Discrete Applied Mathematics* 151 (2005) 106-121.

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

[12] S.B. Gray, Local properties of binary images in two dimensions, *IEEE Transactions on Computers* 20 (1971) 551-561.

[13] G.T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, Springer (2009).

[14] G.T. Herman, A. Kuba, editors, *Discrete Tomography: Foundations, Algorithms and Applications*, Birkhäuser, Boston (1999).

[15] G.T. Herman, A. Kuba, Discrete tomography in medical imaging, *Proceedings of the IEEE* 91 (2003) 1612-1626.

[16] G.T. Herman, A. Kuba, editors, *Advances in Discrete Tomography and Its Applications*, Birkhäuser, Boston (2007).

[17] J.R. Jinschek, K.J. Batenburg, H.A. Calderon, R. Kilaas, V. Radmilovic and C. Kisielowski, 3-D reconstruction of the atomic positions in a simulated gold nanocrystal based on discrete tomography, *Ultramicroscopy* 108 (2008) 589-604.

[18] R. Kopperman, J.L. Pfaltz, Jordan surfaces in discrete topologies IWCIS, *10th International Workshop on Combinatorial Image Analysis (IWCIA)* (2004).

[19] A. Kuba, L. Rodek, Z. Kiss, L. Ruskó, A. Nagy, M. Balaskó, Discrete tomography in neutron radiography, *Nuclear Instruments and Methods in Physics Research, Section A* 542 (2005) 376-382.

[20] J.C. Palacios, L.C. Longoria, J. Santos, R.T. Perry, A PC-based discrete tomography imaging software system for assaying radioactive waste containers, *Nuclear Instruments and Methods in Physics Research, Section A* 508 (2003) 500-511.

[21] A. Rosenfeld, Connectivity in digital pictures, *Journal of the Association for Computing Machinery* 17 (1970) 146-160.

[22] A. Rosenfeld, Compact Figures in Digital Pictures, *IEEE Transactions on System, Man and Cybernetics* 4 (1974) 221-223.

[23] A. Rosenfeld, J.L. Pfaltz, Distance functions on digital pictures, *Pattern Recognition* 1 (1968) 33-61.

[24] H.J. Ryser, Combinatorial properties of matrices of zeros and ones, *Canadian Journal of Mathematics* 9 (1957) 371-377.

[25] H. Slump, J.J. Gerbrands, A network flow approach to reconstruction of the left ventricle from two projections, *Computer Graphics and Image Processing* 18 (1982) 18-36.

[26] Y.R. Wang, Characterization of binary patterns and their projections, *IEEE Trans. Comput.* 24 (1975) 1032-1035.

[27] Linbing Wang, Jin-Young Park, Yanrong Fu, Representation of real particles for DEM simulation using X-ray tomography, *Construction and Building Materials* 21 (2007) 338-346.

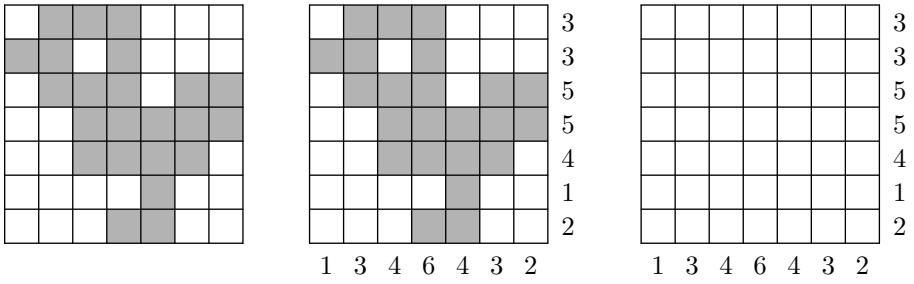
[28] G.J. Woeginger, The reconstruction of polyominoes from their orthogonal projections, *Information Processing Letters* 77 (2001) 225-229.

Samenvatting

Deze samenvatting is voor iedereen die graag wil weten waar mijn proefschrift over gaat, maar de wiskundige notatie in de andere hoofdstukken wat te veel van het goede vindt. Ga er even voor zitten en laat je meenemen in de wondere wereld van de zwarte en witte vakjes, waarin ik de afgelopen jaren met veel plezier rondgedoold heb.

1 Binaire plaatjes en Japanse puzzels

Dit proefschrift gaat over *binaire plaatjes*. Dat zijn plaatjes die je op een velletje ruitjespapier kunt tekenen door sommige vakjes zwart te kleuren en andere vakjes open te laten. Zie figuur 1(a) voor een eenvoudig voorbeeld van zo'n plaatje.



(a) Een binair plaatje van een vogeltje.

(b) Het vogeltje met rij- en kolomsommen.

(c) De DT-puzzel van het vogeltje.

Figuur 1

Als we zo'n binair plaatje hebben, kunnen we in elke rij (horizontaal) en elke kolom (verticaal) tellen hoeveel vakjes er zwart gemaakt zijn. We noemen het aantal zwarte

vakjes in een rij de *rijsom* van die rij en het aantal zwarte vakjes in een kolom de *kolomsom* van die kolom. We hebben deze rij- en kolomsommen aangegeven in figuur 1(b).

Je kunt nu van dit plaatje een puzzel maken door de zwarte vakjes weer uit te gummen, maar de rij- en kolomsommen te laten staan. De puzzel wordt dan: vind het plaatje terug aan de hand van de rij- en kolomsommen. Zie figuur 1(c). Dit soort puzzels wordt bestudeerd in de *discrete tomografie*.

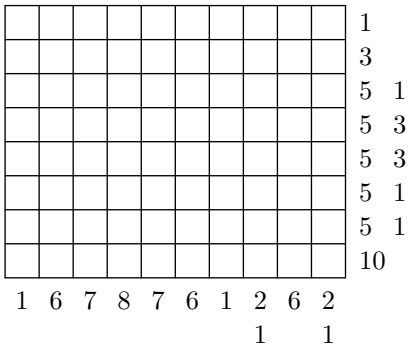
Discrete tomografie gaat in het algemeen over het reconstrueren van plaatjes waarvan alleen maar in een aantal richtingen (bijvoorbeeld horizontaal en verticaal zoals hierboven, maar andere richtingen kunnen ook) bekend is hoeveel vakjes van elke kleur er zijn. Tomografie wordt bijvoorbeeld toegepast bij het maken van een CT-scan in het ziekenhuis. (De T in "CT-scan" staat dan ook voor tomografie.) Daar wordt met behulp van een röntgenfoto per richting bepaald hoeveel weefsel zich in die richting bevindt. Door dit in veel verschillende richtingen te doen, kan vervolgens berekend worden hoe de patiënt er van binnen uitziet, zonder hem open te hoeven snijden.

Dit proefschrift gaat niet over de toepassingen, maar bestudeert de theoretische eigenschappen van puzzels zoals die in figuur 1(c). Die puzzels, die we verder afgekort *DT-puzzels* noemen, zijn namelijk al heel leuk op zich, zoals hopelijk duidelijk zal worden in de rest van deze samenvatting.

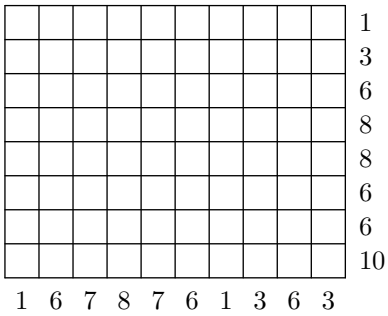
DT-puzzels lijken erg op een ander soort puzzels: Japanse puzzels, ook wel nonogrammen genoemd. Bij een Japanse puzzel vertellen de getallen buiten het veld je niet alleen hoeveel zwarte vakjes er in een rij of kolom staan, maar ook hoeveel er daarvan aaneengesloten zijn. In figuur 2(a) zie je een Japanse puzzel. De getallen 5 en 1 in de derde rij betekenen dat ergens in die rij 5 aaneengesloten zwarte vakjes zitten en ergens rechts daarvan nog 1 los zwart vakje.

We kunnen deze Japanse puzzel veranderen in een DT-puzzel door bij elke rij en elke kolom steeds de getallen bij elkaar op te tellen. Dan krijgen we immers het totaal aantal zwarte vakjes in die rij of kolom. Zie figuur 2(b). Deze nieuwe puzzel (de DT-puzzel) is lastiger dan de oorspronkelijke Japanse puzzel. In de DT-puzzel weet je namelijk niet zeker of de 3 zwarte vakjes die in de tweede rij moeten komen, allemaal aan elkaar zitten, of allemaal los, of 2 aan elkaar en 1 los. En net zo goed weet je in de derde rij niet dat de 6 vakjes verdeeld zijn als 5 en 1, wat je bij de Japanse puzzel nog wel wist.

Ondanks dat hij lastiger is, is de DT-puzzel van figuur 2(b) nog prima op te lossen. Probeer het maar!



(a) Een Japanse puzzel.



(b) De DT-puzzel behorende bij de Japanse puzzel uit (a).

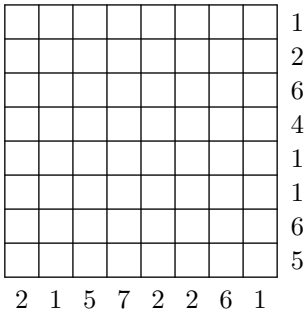
Figuur 2

2 Onoplosbare puzzels

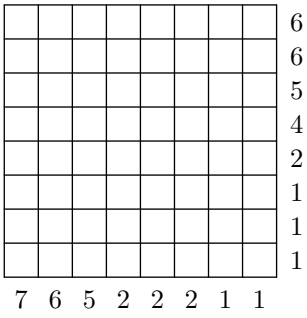
Als je een rechthoekig ruitjesveld neemt en bij elke rij en elke kolom een getal neerzet, heb je nog niet meteen een goede DT-puzzel. Het kan gebeuren dat er helemaal geen oplossing bestaat voor je puzzel. Dat kan diverse redenen hebben. Als je bijvoorbeeld een rijsom 10 hebt, terwijl er maar 8 beschikbare vakjes in je rij zijn (omdat je rechthoek maar 8 kolommen breed is), dan kan dat natuurlijk niet: je kunt nooit van 8 vakjes er 10 zwart kleuren. Iets anders om op te letten is dat als je alle rijsommen optelt, er hetzelfde uitkomt als wanneer je alle kolomsommen optelt. Beide getallen staan immers voor het totaal aantal zwarte vakjes in je rechthoek.

Zelfs als je ervoor zorgt dat de rijsommen nooit groter zijn dan het aantal kolommen, dat de kolomsommen nooit groter zijn dan het aantal rijen en dat de rijsommen en de kolomsommen dezelfde som hebben, kan het nog steeds gebeuren dat er geen oplossing voor je puzzel is. Dit is niet altijd direct duidelijk. Bekijk bijvoorbeeld figuur 3(a). Op het eerste gezicht lijkt dit een prima puzzel. Maar we kunnen laten zien dat er hier geen oplossing is.

Dit wordt duidelijker als we de rijsommen en kolomsommen even op grootte sorteren. Dit kunnen we overigens gewoon doen zonder de puzzel essentieel te veranderen. Als het plaatje echt iets voorstelt, zoals het vogeltje van figuur 1(a), dan wordt het natuurlijk een rommeltje als je een paar rijen met elkaar verwisselt. Maar als je alleen maar wilt weten of er al dan niet een oplossing is, dan maakt het helemaal niet uit als je een paar rijen met elkaar verwisselt. Als de puzzel in figuur 3(b) een oplossing heeft, dan heeft de oorspronkelijke puzzel, figuur 3(a), ook een oplossing, die we kunnen vinden door in de oplossing van figuur 3(b) de rijen en kolommen weer



(a) Heeft deze puzzel een oplossing?



(b) Dezelfde puzzel met de rij- en kolomsommen gesorteerd.

Figuur 3

terug te wisselen naar de oorspronkelijke positie. Als figuur 3(b) juist geen oplossing heeft, dan heeft de figuur 3(a) natuurlijk ook geen oplossing.

We bekijken dus nu de puzzel in figuur 3(b), waar de rijen en kolommen op grootte gesorteerd zijn. Voor zo'n puzzel is er een methode uitgevonden om te bepalen of er een oplossing is.

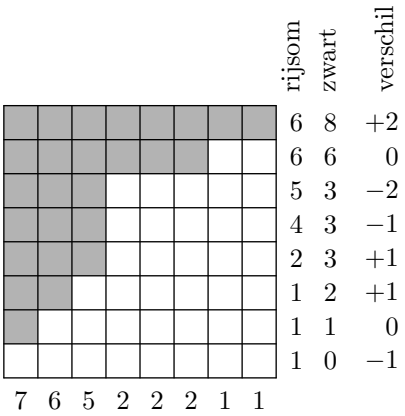
- Allereerst vergeten we de rijsommen en kleuren we in elke kolom precies het aantal vakjes zwart dat de kolomsom aangeeft, en wel van boven naar beneden. Als de kolomsom 5 is, kleuren we dus de 5 bovenste vakjes in de kolom zwart. Zie figuur 4.

- Vervolgens tellen we in elke rij hoeveel zwarte vakjes daar gekleurd zijn en schrijven dit aantal rechts naast de rijsom die al gegeven was.

- We berekenen daarna per rij hoeveel vakjes er te veel of te weinig gekleurd zijn. Als er te veel vakjes gekleurd zijn, schrijven we dit verschil met een + op en als er te weinig vakjes gekleurd zijn, schrijven we dit verschil met een − op.

- Tel van boven naar beneden deze verschillen op. In ons voorbeeld beginnen we dus met +2, dan tellen we er 0 bij op, dus blijven we op +2, dan tellen we er −2 bij op, dus komen we op 0, enzovoorts. Het is belangrijk om van boven naar beneden één voor één de getallen op te tellen, want juist deze tussenresultaten hebben we nodig.

- Er geldt nu: als het tussenresultaat altijd minstens 0 blijft, dus 0 of iets positiefs, dan is er een oplossing. Wordt het tussenresultaat ergens negatief, dan is er geen oplossing.



Figuur 4: In elke kolom zijn de vakjes van boven naar beneden zwart gekleurd, precies het aantal aangegeven door de kolomsom. De getallen naast elke rij geven van links naar rechts aan: de gewenste rijsom, het aantal vakjes dat zwart gekleurd is, en het verschil tussen die twee.

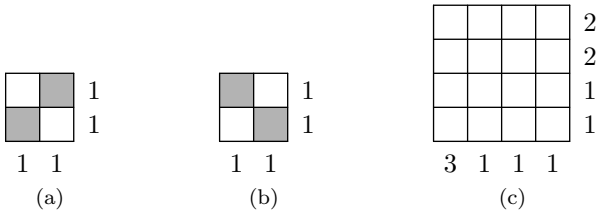
Waarom werkt deze methode? Dat heeft te maken met het feit dat we begonnen zijn om alle vakjes van bovenaf zwart te kleuren. Het verschil dat we berekenen in de eerste rij is het aantal vakjes dat we te veel gekleurd hebben in die rij. Als dit positief is, is dat niet erg: dan kunnen we gewoon weer wat vakjes uitgummen en verder naar beneden neerzetten. Maar als het negatief is, is het wel erg, want dan zouden we te weinig vakjes gekleurd hebben in de bovenste rij, terwijl we juist alle vakjes zoveel mogelijk van bovenaf gekleurd hadden. Zelfs met zoveel mogelijk gekleurde vakjes bovenin zijn er dan niet genoeg gekleurde vakjes in de eerste rij, dus kan er geen oplossing bestaan.

Als we vervolgens de verschillen in de eerste en tweede rij optellen, vinden we het totaal aantal vakjes dat in de eerste twee rijen te veel zwart gekleurd is. Ook hier geldt: als dit negatief is, moeten er dus meer vakjes bovenin gekleurd worden, maar dat kan niet, want we hadden juist al zoveel mogelijk vakjes bovenin gekleurd. Hetzelfde verhaal geldt voor de eerste drie rijen, de eerste vier rijen, enzovoorts.

In het voorbeeld van figuur 4 wordt het tussenresultaat na vier rijen negatief. Dat betekent concreet het volgende. We hebben alle vakjes zoveel mogelijk van bovenaf zwart gekleurd. Hierdoor zijn er in de eerste vier rijen samen 20 vakjes zwart gekleurd. Dat is dus het grootste aantal vakjes dat we ooit zwart zouden kunnen kleuren in de eerste vier rijen, zolang we ons aan de gegeven kolomsommen houden. Maar als je de rijsommen van de eerste vier rijen optelt, dan blijkt dat we daar 21 vakjes zwart moeten kleuren. Dat is dus onmogelijk.

3 Saaie puzzels

We weten nu hoe we bij het maken van DT-puzzels de grootste frustratiebron van menig puzzelaar, namelijk een puzzel die geen oplossing heeft, kunnen voorkomen. Maar daarmee ben je er nog niet, want het kan onverhoopt ook nog gebeuren dat er meer dan één oplossing is. Dat zou wel eens de op één na grootste frustratie van een puzzelaar kunnen zijn, want dan is de puzzel niet uniek op te lossen.

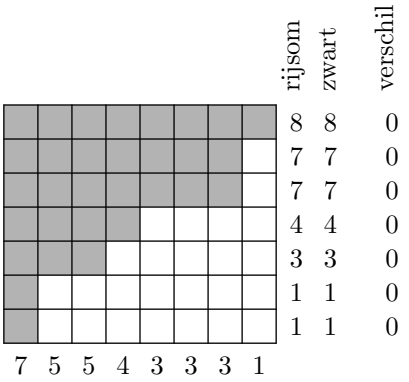


Figuur 5: DT-puzzels kunnen soms meerdere oplossingen hebben.

Er bestaan al heel kleine DT-puzzels die geen unieke oplossing hebben. Figuur 5(a) en figuur 5(b) laten twee oplossingen zien van dezelfde DT-puzzel die slechts 2×2 groot is. Als we hem iets groter maken, kunnen er nog veel meer verschillende oplossingen zijn. Zo blijkt de puzzel in figuur 5(c) maar liefst 18 verschillende oplossingen te hebben.

Laten we nog eens beter kijken naar de methode van hiervoor om te bepalen of er al dan niet een oplossing bestaat van een gegeven DT-puzzel. Stel dat het toevallig zo uitkomt dat de verschillen die je opschrijft in de derde stap van de methode, stuk voor stuk gelijk zijn aan 0. Zie voor een voorbeeld figuur 6. Dat betekent dat de aantallen zwarte vakjes per rij die je in de tweede stap opgeschreven hebt, allemaal gelijk zijn aan de oorspronkelijke rijsommen. En dat betekent weer dat de kleuring die we in de eerste stap gemaakt hebben, meteen een goede oplossing is.

Zouden er nog meer oplossingen kunnen zijn van zo'n puzzel? Laten we dit bekijken aan de hand van het voorbeeld. In de eerste rij hebben we 8 zwarte vakjes nodig. Er zijn maar 8 vakjes in die rij, dus die moeten allemaal gekleurd zijn. In deze rij zit dus geen speelruimte meer. In de volgende rij hoeven we slechts 7 van de 8 vakjes zwart te kleuren. Toch zit ook hier geen speelruimte, want de meest rechterkolom is al vol door dat ene zwarte vakje in de eerste rij. Er blijven dus nog maar 7 kolommen over om vakjes in zwart te kleuren. De tweede en derde rij liggen dus ook helemaal vast. Maar nadat we die gekleurd hebben, blijken weer drie kolommen al klaar te zijn. We kunnen nu alleen nog maar de linker vier kolommen gebruiken, dus ook bij het kleuren van de vierde rij (met som 4) hebben we weer niets te kiezen.



Figuur 6: Bij deze DT-puzzel levert de methode allemaal verschillen van 0 op.

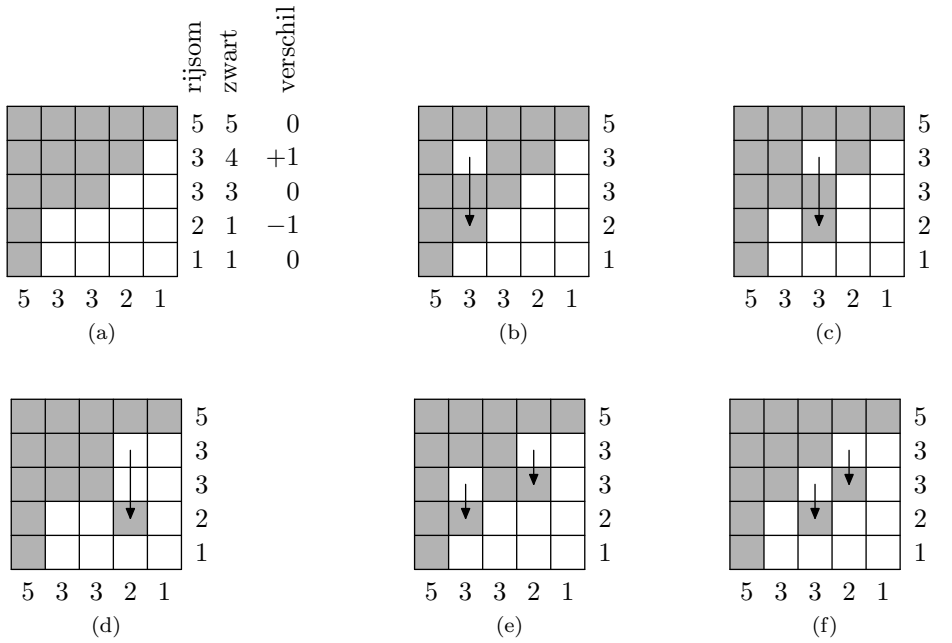
Dit is geen toeval. Er is maar één manier om zoveel mogelijk vakjes bovenin te kleuren terwijl je je aan de kolomsommen houdt. Als die manier meteen een oplossing oplevert (dus als de gekleurde vakjes ook kloppen met de rijsommen) dan moet dat de enige oplossing zijn. Kortom, als je in de derde stap van de methode alleen maar nullen krijgt, heb je een puzzel gevonden met een unieke oplossing.

We weten nu hoe we een puzzel kunnen maken met een unieke oplossing. Tegelijkertijd zijn we er ook achter gekomen dat dit heel saaie puzzels worden. Je kunt namelijk altijd eerst de rij met de grootste rijsom helemaal inkleuren, dan de (eventuele) kolommen afstrepen die hierdoor al klaar zijn (omdat de kolomsom 1 was), vervolgens de rij met de grootste overgebleven rijsom inkleuren, enzovoorts.

4 Puzzels met meerdere oplossingen

Omdat de puzzels met een unieke oplossing saai zijn, kijken we vanaf nu naar puzzels met meerdere oplossingen. Voor een puzzelaar zijn deze puzzels misschien minder leuk, maar voor een wiskundige is er heel wat aan te beleven.

We gaan weer terug naar de methode die we hiervoor gebruikt hebben. We hebben gezien dat er een unieke oplossing is als er in de derde stap van de methode alleen maar nullen tevoorschijn komen. We hebben ook gezien dat er geen oplossingen zijn als we bij het optellen van de verschillen (van boven naar beneden) een keer een negatief tussenresultaat krijgen. Maar hoe zit het als het tussenresultaat niet negatief wordt, maar er ook niet altijd 0 staat? Dan blijken er meerdere oplossingen te zijn.



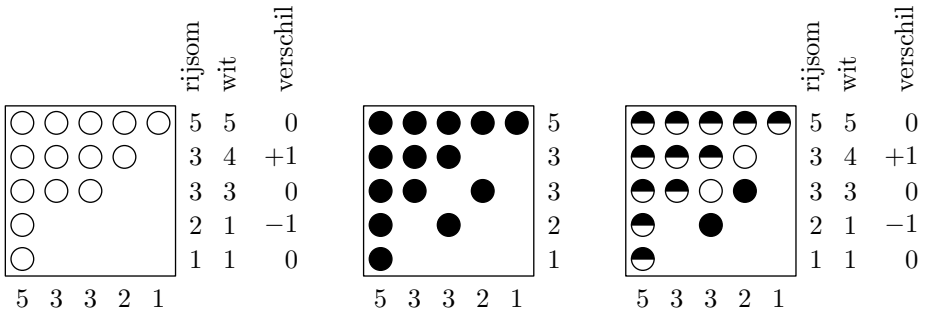
Figuur 7: Met behulp van de methode vinden we vijf verschillende oplossingen van deze DT-puzzel.

Kijk bijvoorbeeld naar de puzzel in figuur 7(a). We vinden hier door de methode toe te passen één $+1$ en één -1 en verder nullen. Dat betekent dat de tweede rij een zwart vakje te veel heeft en de vierde rij een zwart vakje te weinig. We kunnen nu een oplossing van de puzzel vinden door een zwart vakje van de tweede rij naar de vierde rij te verplaatsen. Dat vakje moet wel binnen één kolom verhuizen, want de kolomsommen waren al goed, dus daar mogen we niet meer aanzitten. Er zijn drie geschikte kolommen waarin we een zwart vakje uit de tweede rij kunnen verhuizen naar de vierde rij, dus dat geeft al drie verschillende oplossingen; zie figuur 7(b), 7(c) en 7(d).

Er is nog een andere mogelijkheid. We kunnen ook eerst een zwart vakje van de tweede naar de derde rij verhuizen. Dan heeft de tweede rij vervolgens precies genoeg zwarte vakjes, maar de derde rij eentje te veel. Dus moeten we nog een ander vakje van de derde naar de vierde rij verhuizen. Zo vinden we nog twee oplossingen: figuur 7(e) en 7(f).

Laten we nog eens beter kijken naar hoe we van de situatie met zoveel mogelijk zwarte vakjes bovenin, figuur 7(a), naar een echte oplossing van de puzzel komen, bijvoorbeeld de oplossing in figuur 7(f). Hiervoor willen we deze twee situaties in één

plaatje weergeven. Dat doen we als volgt. Allereerst kunnen we figuur 7(a) weergeven met witte bolletjes in plaats van de zwarte vakjes, zie figuur 8(a). Vervolgens geven we oplossing van de puzzel uit figuur 7(f) weer met zwarte bolletjes, zie figuur 8(b). Deze twee kunnen we nu tegelijk weergeven zoals in figuur 8(c): daar waar zowel een zwart als een wit bolletje staat, tekenen we een zwart-wit bolletje.



(a) Figuur 7(a) weergegeven met witte bolletjes.

(b) Figuur 7(f) weergegeven met zwarte bolletjes.

(c) Beide figuren in één.

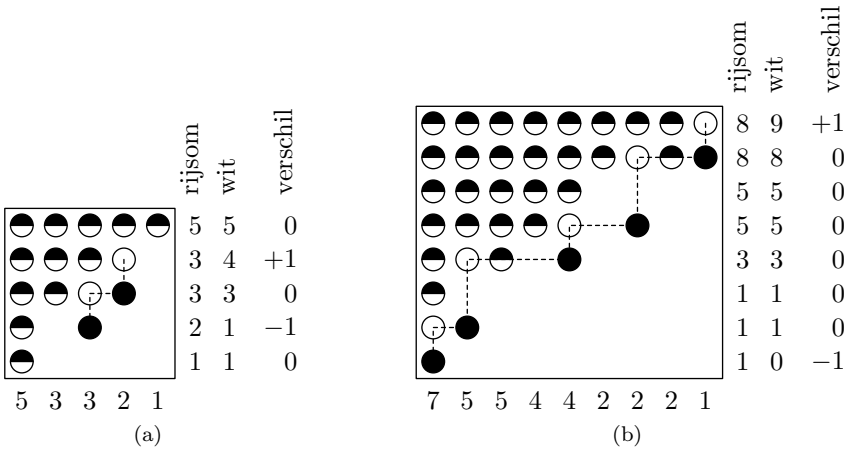
Figuur 8

De zwart-witte bolletjes zijn nu de vakjes die niet verhuisd zijn bij het maken van een oplossing van deze puzzel. De zwarte en de witte bolletjes geven juist het spoor aan van de vakjes die wel verhuisd zijn: op een wit bolletje was eerst wel een gekleurd vakje, maar nu niet meer, en bij een zwart bolletje is het andersom. Dit spoor heeft de vorm van een trappetje, zie ook figuur 9(a). Deze trapvorm wordt nog duidelijker als we een groter voorbeeld nemen: in figuur 9(b) zie je een grotere puzzel, waar de witte en zwart-witte bolletjes samen een oplossing van de puzzel vormen. Ook hier is het spoor van de verhuisde vakjes (de zwarte en witte bolletjes) trapvormig.

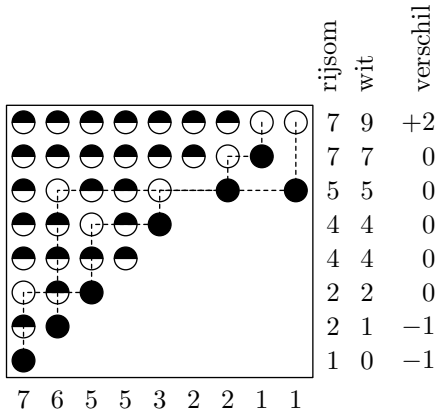
Deze twee puzzels hebben gemeen dat we in de methode van hiervoor alleen één keer een $+1$ en en één keer een -1 tegenkomen. Je kunt laten zien dat in dat geval de zwarte en witte bolletjes altijd een trap vormen.

Maar wat als dat niet zo is? Stel dat je bijvoorbeeld twee keer $+1$ hebt en twee keer een -1. Of één keer $+2$ en twee keer -1. In dat geval vormen de zwarte en witte bolletjes samen niet één, maar twee trappen. Zie figuur 10 voor een voorbeeld.

In het algemeen blijkt het aantal trappen precies gelijk te zijn aan de som van de positieve verschillen die je uitrekent in de methode. Met deze kennis kun je allerlei leuke dingen doen, zoals bepalen hoeveel gekleurde vakjes twee verschillende oplossingen van dezelfde puzzel altijd gemeen moeten hebben. Hoofdstukken 2 tot en met 4 van dit proefschrift gaan hierover.



Figuur 9: De witte en zwarte bolletjes (zonder de zwart-witte) vormen samen een trap.



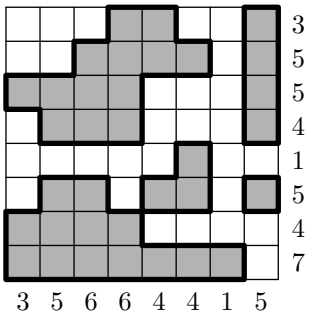
Figuur 10: Hier zijn er twee trappen, die door elkaar heen lopen.

5 Rand

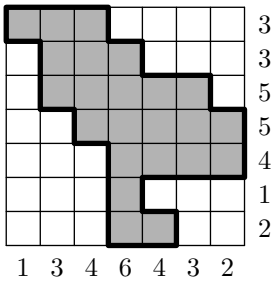
Als een DT-puzzel meerdere oplossingen heeft, dan biedt dat gelegenheid om nieuwe uitdagingen toe te voegen, bijvoorbeeld: vind de oplossing met de kleinste rand. Als rand tellen we hier alle grenzen van de zwarte vakjes, dus ook die aan de buitenrand van het veld. Zo heeft het plaatje van figuur 11(a) een rand van 62.

We gaan even terug naar het allereerste binaire plaatje dat we bekeken hebben:

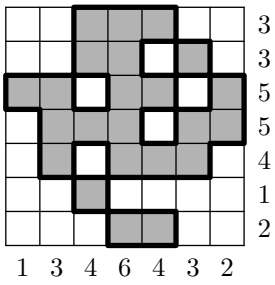
het vogeltje uit figuur 1(a). De DT-puzzel behorende bij dit plaatje heeft heel veel oplossingen. De oplossing met de kleinst mogelijke rand zie je in figuur 11(b) en een oplossing met juist een heel grote rand staat in figuur 11(c).



(a) Dit binaire plaatje heeft een rand van lengte 62.



(b) De DT-puzzel van het vogeltje: een oplossing met rand 30.



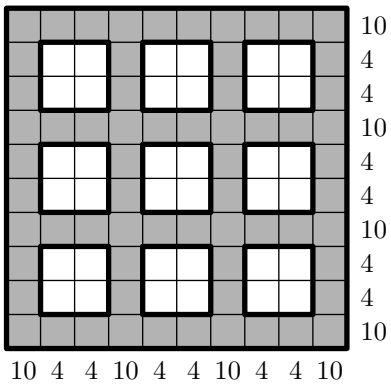
(c) De DT-puzzel van het vogeltje: een oplossing met rand 52.

Figuur 11

Bepalen wat de kleinste rand is die een oplossing van een gegeven DT-puzzel kan hebben, is heel lastig. Maar we kunnen er wel een paar dingen over zeggen. Zo weet je dat in elke kolom toch minstens één zwart vakje moet staan (aangenomen dat er geen kolomsommen gelijk aan 0 zijn). Dus als je van boven naar beneden door zo'n kolom wandelt, kom je zeker twee keer een stukje rand tegen, namelijk aan de bovenkant van het eerste zwarte vakje dat je tegenkomt en aan de onderkant van het laatste zwarte vakje dat je tegenkomt. Dus elke kolom draagt minstens 2 bij aan de totale lengte van de rand. Dat geldt ook voor elke rij. We zien dus dat elke oplossing van de DT-puzzel in figuur 11(b), die 7 rijen en 7 kolommen heeft, een rand van minstens $7 \times 2 + 7 \times 2 = 28$ heeft. Een uitdaging voor de lezer: laat zien dat de minimale rand van deze puzzel 30 is, zodat de oplossing van figuur 11(b) echt degene met de kleinste rand is.

Als we dezelfde tactiek toepassen op de puzzel in figuur 12, komen we uit op een minimale rand van $10 \times 2 + 10 \times 2 = 40$. De werkelijke rand van de oplossing in deze figuur is echter 112. Dat is een enorm verschil, maar toch lijkt het er niet op dat deze puzzel een andere oplossing met veel minder rand heeft. In feite kunnen we laten zien dat de rand altijd minstens 112 is. Dat gaat als volgt. Vergelijk de eerste rij met de tweede rij. In de eerste rij moeten 10 zwarte vakjes komen, in de tweede rij 4. Dat betekent dat er altijd minstens 6 zwarte vakjes in de eerste rij zijn die boven een wit vakje in de tweede rij zitten (want er kunnen er maar 4 van de 10 boven een zwart vakje zitten). Dus bij de overgang van de eerste naar de tweede rij hebben we minstens 6 randstukjes.

Bij de overgang van de tweede naar de derde rij zit er misschien helemaal geen rand,



Figuur 12: Dit binaire plaatje heeft een rand van 112.

want het zou kunnen dat die 4 zwarte vakjes in de tweede rij precies boven de 4 zwarte vakjes in de eerste rij zitten. Bij de overgang van de derde naar de vierde rij weten we wel weer zeker dat er randstukjes tevoorschijn komen, want die 10 zwarte vakjes uit de vierde rij kunnen aansluiten op hoogstens 4 zwarte vakjes in de derde rij, dus zijn er hier minstens 6 randstukjes te vinden.

Zo vinden we bij de overgangen tussen de rijen al $6+0+6+6+0+6+6+0+6 = 36$ randstukjes. En dan moeten we ook nog de randstukjes aan de bovenrand en de onderrand van het veld meetellen: in de eerste rij zitten 10 zwarte vakjes, dus dat geeft 10 randstukjes aan de bovenrand, en zo ook zijn er 10 randstukjes aan de onderrand. Nu zitten we al op 56. Als we het hele verhaal nog een keer overdoen voor de kolommen (in plaats van de rijen) vinden we ook daar nog 56 randstukjes. Bij elkaar opgeteld hebben we nu laten zien dat elke oplossing een rand van minstens 112 heeft.

Deze nieuwe techniek om de minimale lengte van de rand te bepalen, is dus in dit voorbeeld veel beter dan de vorige. Maar als we deze nieuwe techniek toepassen op figuur 11(b), dan komen we juist slechter uit dan eerst, namelijk op een rand van minstens 24, terwijl we eerder al 28 hadden. We hebben dus twee technieken, waarvan de ene de ene keer beter is en de andere de andere keer.

Hoofdstukken 5 tot en met 7 van dit proefschrift gaan over dit soort technieken en over andere interessante dingen die je kunt zeggen over de rand van binaire plaatjes. De methode van het eerste deel (om te bepalen of een oplossing uniek is of hoeveel trappen je nodig hebt om hem te maken) komt hier ook weer terug: hoe minder trappen je nodig hebt om een oplossing te maken, hoe kleiner je de rand van een oplossing kunt krijgen.

Curriculum Vitae

Birgit van Dalen werd geboren op 14 november 1984 in Den Haag. Ze doorliep van 1996 tot 2002 het gymnasium aan de Vlaardingse Openbare Scholengemeenschap. Daarna ging ze studeren aan de Universiteit Leiden. Ze behaalde het eerste jaar twee propedeuses, wiskunde en sterrenkunde, en ontving bovendien de Stieltjes propedeuseprijs wiskunde. In 2005 behaalde ze haar Bachelordiploma wiskunde en in 2007 haar Masterdiploma wiskunde, beide cum laude.

Aansluitend begon Birgit met haar promotieonderzoek in de discrete tomografie aan de Universiteit Leiden. Dit werd begeleid door haar promotoren prof.dr. Rob Tijdeman en prof.dr. Joost Batenburg. Het onderzoek leverde in vier jaar een aantal gepubliceerde artikelen op en resulteerde in dit proefschrift.

Op 12-jarige leeftijd nam Birgit voor het eerst deel aan een zomerkamp van Stichting Vierkant voor Wiskunde. Dit beviel haar zo goed dat de jaren erna de vakantieplannen van het gezin hierop afgestemd werden. Na vijf keer deelnemer te zijn geweest, werd Birgit in 2002 begeleider van de Vierkant zomerkampen. Ze heeft sindsdien elke zomer tot en met 2010 één of twee kampen begeleid. Vanaf 2005 had ze samen met een collega-kampleider de algehele leiding over zo'n kamp. Bovendien was ze van 2007 tot en met 2010 eindverantwoordelijk voor de algehele organisatie van alle Vierkant zomerkampen.

Ondertussen is Birgit ook actief bij de Nederlandse Wiskunde Olympiade. Na als scholier drie jaar te hebben deelgenomen aan het het trainingsprogramma voor de International Mathematical Olympiad (IMO) en in 2002 een eervolle vermelding behaald te hebben bij de IMO in Schotland, werd ze eind 2004 zelf trainer van de Nederlandse kandidaten voor het IMO-team. In de jaren daarna is ze steeds meer betrokken geraakt bij de organisatie van dit trainingsprogramma en van de Nederlandse Wiskunde Olympiade in het algemeen. In deze periode is onder andere het trainingsprogramma uitgebreid, de finaletraining opgezet, de regionale tweede ronde ingevoerd en de Benelux Mathematical Olympiad ontstaan. Van 2007 t/m 2010 was ze vice-teamleider van het Nederlandse team bij de IMO. Sinds 2011 is ze

ook bestuurslid van de Stichting Nederlandse Wiskunde Olympiade.

In 2011 vond voor het eerst in de geschiedenis de IMO in Nederland plaats. Birgit maakte vanaf de zomer van 2008 deel uit van het organisatiecomité dat dit negendaagse evenement voorbereidde. Er namen aan deze IMO zo'n 900 buitenlandse gasten deel en er werkten meer dan 300 vrijwilligers mee. Eerder al werd als voorproefje de Benelux Mathematical Olympiad 2010 in Nederland georganiseerd. Birgit was co-voorzitter van het organisatiecomité van dit evenement, waar (voor de gelegenheid) zes landen aan deelnamen.