

On Taxicab Distance Mean Functions and their Geometric Applications: Methods, Implementations and Examples

Csaba Vincze*

Institute of Mathematics

University of Debrecen

P.O.Box 400, H-4002 Debrecen, Hungary

csvincze@science.unideb.hu

Ábris Nagy

Institute of Mathematics

University of Debrecen

Debrecen, Hungary

abris.nagy@science.unideb.hu

Abstract. A distance mean function measures the average distance of points from the elements of a given set of points (focal set) in the space. The level sets of a distance mean function are called generalized conics. In case of infinite focal points the average distance is typically given by integration over the focal set. The paper contains a survey on the applications of taxicab distance mean functions and generalized conics' theory in geometric tomography: bisection of the focal set and reconstruction problems by coordinate X-rays. The theoretical results are illustrated by implementations in Maple, methods and examples as well. ¹

Keywords: distance mean functions, generalized conics, taxicab distance, parallel x-rays

Address for correspondence: Institute of Mathematics, University of Debrecen, P.O.Box 400, H-4002 Debrecen, Hungary

*Research supported in part by the Eötvös Loránd Research Network (ELKH).

¹The paper is based on the plenary lecture presented at Meeting on Tomography and Applications (Discrete Tomography, Neuroscience and Image Reconstruction) 16th Edition, IN MEMORIAM OF CARLA PERI, 2 - 4 May 2022, Mathematics Department, Politecnico di Milano, Milano, Italy.

1. Introduction

A distance mean function measures the average distance of points from the elements of a given set of points (focal set) in the space. The level sets of a distance mean function are called generalized conics. The most important discrete examples are polyellipses (polyellipsoids) as level sets of a function measuring the arithmetic mean of distances from finitely many focal points (constant distance sum) and polynomial lemniscates as level sets of a function measuring the geometric mean of distances from finitely many focal points (constant distance product). In case of infinite focal points the average distance is typically given by integration over the focal set. Using partitions and integral sums, the level sets (generalized conics) are Hausdorff limits of polyellipsoids Section 4 [1].

1.1. Some general observations

The general form of the functions we are interested in is

$$x \mapsto f_D(x) := \frac{1}{\mu(D)} \int_D u \circ d(x, y) d_\mu y, \quad (1)$$

where d measures the distance between the points, $D \subset \mathbb{R}^n$ is a compact subset with a finite positive measure with respect to μ , $u: \mathbb{R} \rightarrow \mathbb{R}$ is a strictly monotone increasing convex function satisfying the initial condition $u(0) = 0$. In what follows we suppose that the distance function comes from a norm. The convexity of the integrand implies that the distance mean function is a convex and, consequently, a continuous function. Using the increasing slope property of convex functions, we have that

$$\liminf_{t \rightarrow \infty} \frac{u(t)}{t} > 0 \quad (2)$$

and the distance mean function inherits a growth property of the form

$$\liminf_{\|x\| \rightarrow \infty} \frac{f_D(x)}{\|x\|} > 0. \quad (3)$$

The growth property (3) implies that the sublevel sets of the form $C_D := \{x \mid f_D(x) \leq c\} \subset \mathbb{R}^n$ are bounded because the existence of a sequence $x_n \in C_D$ such that $\lim_{n \rightarrow \infty} \|x_n\| = \infty$ gives a contradiction:

$$\lim_{n \rightarrow \infty} \frac{f_D(x_n)}{\|x_n\|} \leq \lim_{n \rightarrow \infty} \frac{c}{\|x_n\|} = 0.$$

Theorem 1. [2, 3] The sublevel sets of a distance mean function are convex and compact.

Weierstrass theorem states that if all the level sets of a continuous function defined on a non-empty closed set in \mathbb{R}^n are bounded, then it has a global minimizer.

Theorem 2. [2, 3] The distance mean function has a global minimizer.

1.2. The problem of unicity

The general problem of unicity means to characterize the subsets in the space that are uniquely determined by the average distance measuring. The following theorem shows that the iteration of the averaging process determines the sublevel sets of a distance mean function under the choice $u(t) = t$ in formula (1). The result is a generalization of [2, Theorem 9].

Theorem 3. [3] Let $u(t) = t$ and let C_D be a sublevel set of f_D . If $f_{C_*} = f_{C_D}$, where C_* is a compact set such that $\mu(C_D) = \mu(C_*)$ then C_D is equal to C_* except on a set of measure zero with respect to μ .

We present some applications under a more special choice of the ingredients (1). In all the following sections we choose $u(t) = t$ and $d = d_1$, the taxicab distance. In sections 2 and 3 we let μ be the Lebesgue measure, while μ is the counting measure in section 4.

2. Bisection of bodies by coordinate hyperplanes

Suppose that distance measuring and integration are taken with respect to the taxicab distance

$$d_1(x, y) = \sum_{i=1}^n |x^i - y^i| \tag{4}$$

and the Lebesgue measure μ_n , respectively. Let K be a compact subset of measure one² in \mathbb{R}^n and consider the taxicab distance mean function

$$f_K(x) = \int_K d_1(x, y) dy = \sum_{i=1}^n \int_K |x^i - y^i| dy. \tag{5}$$

Since the derivative of the integrand at x^i is ± 1 depending on $y^i < x^i$ or $x^i < y^i$, we can conclude that the value 1 occurs as many times as many points $y \in K$ is on the left hand side of x with respect to the i -th coordinate:

$$K <_i x^i := \{y \in K \mid y^i < x^i\}.$$

In a similar way, -1 occurs as many times as many points $y \in K$ is on the right hand side of x with respect to the i -th coordinate:

$$x^i <_i K := \{y \in K \mid x^i < y^i\}.$$

Since the set

$$K =_i x^i := \{y \in K \mid y^i = x^i\} \quad (i = 1, \dots, n)$$

is of measure zero we have that

$$D_i f_K(x) = \mu_n(K \leq_i x^i) - \mu_n(x^i \leq_i K) \quad (i = 1, \dots, n). \tag{6}$$

Theorem 4. [2] The point $x \in \mathbb{R}^n$ is a minimizer of f_K if and only if each coordinate hyperplane at x divides K in two parts of equal measure.

²It is a technical condition to avoid the denominator $\mu_n(K)$.

How to bisect a set in two parts of equal measure? Formula (6) shows that

$$|D_i f_K(x) - D_i f_K(y)| = 2\mu_n (\min\{x^i, y^i\} <_i K <_i \max\{x^i, y^i\})$$

and the compactness of K implies that f_K has a Lipschitzian gradient. Therefore the gradient descent method can be used to find the minimizer bisecting the measure of the integration domain K in the sense that each coordinate hyperplane passing through the minimizer divides the set into two parts of equal measure. Let us present the gradient descent method in terms of a stochastic algorithm [4, 3]: let P_k be a sequence of K -valued independent uniformly distributed random variables and consider the recursion

$$X_{k+1} = X_k - t_{k+1}Q_{k+1}, \quad (7)$$

where $X_0 \in K$ is a (random) starting point,

$$Q_{k+1} := (\text{sgn}(X_k^1 - P_{k+1}^1), \dots, \text{sgn}(X_k^n - P_{k+1}^n)) \quad (8)$$

and the step size is a decreasing sequence of positive real numbers t_k satisfying conditions

$$\sum_{k=1}^{\infty} t_k = \infty \quad \text{and} \quad \sum_{k=1}^{\infty} t_k^2 < \infty. \quad (9)$$

Assuming that K is of measure one, we have the conditional probability

$$P(Q_{k+1} = (1, \dots, 1) | X_k) = \mu_n ((K < X_k^1) \cap \dots \cap (K < X_k^n)) \quad (10)$$

because $Q_{k+1} = (1, \dots, 1)$ means that X_k is greater than P_{k+1} with respect to the coordinatewise partial ordering $x \prec y \Leftrightarrow x^1 < y^1, \dots, x^n < y^n$ and P_{k+1} is a uniformly distributed K -valued random variable. In a similar way we have the conditional probability

$$P(Q_{k+1} = (1, -1, 1, \dots, 1) | X_k) = \mu_n ((K < X_k^1) \cap (X_k^2 < K) \cap (K < X_k^3) \cap \dots \cap (K < X_k^n)), \dots \quad (11)$$

and so on. A direct computation shows that $\mathbb{E}(Q_{k+1} | X_k) = \text{grad } f_K(X_k)$.

Remark 1. To illustrate the process let us consider the case of dimension two. The lines parallel to the coordinate axis at X_k divide the plane into four quadrants. Since we have a sequence of independent, uniformly distributed random variables, the value of P_{k+1} is most likely to fall in the quadrant containing the part of K of the highest measure. Using formula (6), it follows that the gradient of f_K at X_k is pointed in the same quadrant represented by the value of the stochastic vector Q_{k+1} . Therefore the step of the highest probability is taken into the opposite direction of the gradient in the sense that the corresponding quadrants are opposite to each other.

Definition 1. A nonempty compact set is called a body if it is the closure of its interior.

Theorem 5. [4, 3] Let $K \subset \mathbb{R}^n$ be a connected compact body. The sequence of random variables X_k converges almost surely to the unique global minimizer x^* of the function f_K .

2.1. Implementation and examples

We show an implementation of the above procedure in the Maple software for polygons in the plane. For this we first need to know how to choose a random point uniformly in a polygon. The standard way to do this is the following:

1. Triangulate the polygon with the help of non-intersecting diagonals.
2. Compute the areas of the triangles and then the area of the whole polygon.
3. Choose a random number x uniformly from the interval $[0, A]$, where A denotes the area of the polygon.
4. If A_i denotes the area of the triangle T_i in the triangulation, then find the smallest positive integer k that satisfies

$$x \leq \sum_{i=1}^k A_i.$$

5. Then choose a random point uniformly in the triangle T_k . This can be done by choosing two independent random numbers u and v with uniform distribution in the $[0, 1]$ interval. If the vertices of T_k are denoted by P_k , Q_k , and R_k , and $u + v \leq 1$, then choose the point $P_k + u(Q_k - P_k) + v(R_k - P_k)$. If $u + v > 1$, then choose the point $P_k + (1 - u)(Q_k - P_k) + (1 - v)(R_k - P_k)$ of the triangle.

Several algorithms exist for polygon triangulation. The most widely used algorithm is based on partitioning the polygon into monotone pieces first and then triangulating the monotone pieces [5, 6]. Another favorable algorithm is the ear-clipping method based on the two ears theorem [7]. Both of them are built in the computational geometry package of Maple. Using this, the following procedure in Maple produces n random points with uniform distribution in the polygon P given by listing the vertices along the boundary.

```
> randompointsinpolygon := proc (P, n)
local L, T, triangleareas, areasum, i, k, x, tf, u, v, Px, Py;
L := [];
T := ComputationalGeometry:-PolygonTriangulation(P);
triangleareas := [];
for i from 1 to nops(T) do
  geometry:-point(A, P[T[i][1]][1], P[T[i][1]][2]);
  geometry:-point(B, P[T[i][2]][1], P[T[i][2]][2]);
  geometry:-point(C, P[T[i][3]][1], P[T[i][3]][2]);
  geometry:-triangle(t, [A, B, C]);
  triangleareas := [op(triangleareas), geometry:-area(t)];
end do;
areasum := add(evalf(triangleareas[i]), i = 1..nops(triangleareas));
for j from 1 to n do
  x := RandomTools:-Generate(float(range = 0..areasum, method = uniform));
  k := 0;
  tf := true;
  while tf do
    k := k + 1;
```

```

    if x<=add(evalf(triangleareas[i]),i=1..k) then
      tf:=false;
    end if;
  end do;
  u:=RandomTools:-Generate(float(range=0..1,method=uniform));
  v:=RandomTools:-Generate(float(range=0..1,method=uniform));
  if 1<u+v then
    u:=1-u;
    v:=1-v;
  end if;
  Px:=P[T[k][1]][1]+u*(P[T[k][2]][1]-P[T[k][1]][1])
    +v*(P[T[k][3]][1]-P[T[k][1]][1]);
  Py:=P[T[k][1]][2]+u*(P[T[k][2]][2]-P[T[k][1]][2])
    +v*(P[T[k][3]][2]-P[T[k][1]][2]);
  L:=[op(L),[Px,Py]];
end do;
return(L)
end proc;

```

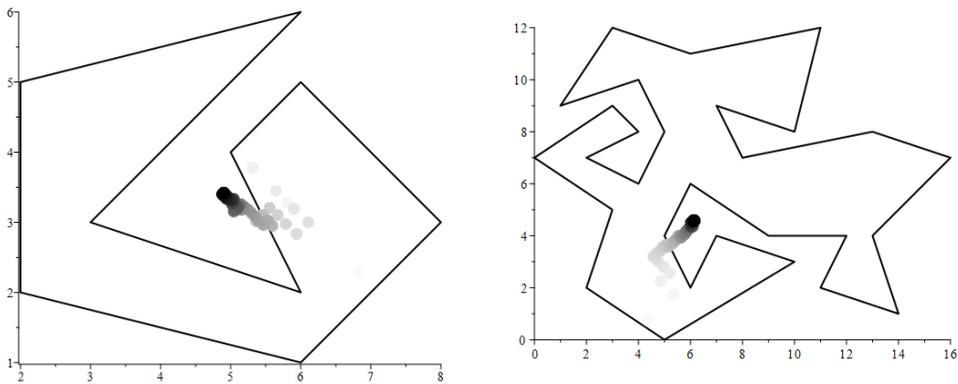


Figure 1. The sequence of points X_k generated by the above procedure for $k = 1, 2, \dots, 50$. Darker points present elements X_k with higher indices k . Notice how these sequences of points converge to the minimizer of the taxicab distance mean function (5).

Then the stochastic algorithm for finding the minimizer of the taxicab distance mean function (5) of a polygon can be implemented in Maple as follows, see Figure 1. The step size sequence is given by $t_k = 1/k$.

```

> minimizer:=proc(polygon,n)
  local P,X,k,Q;
  P:=randompointsinpolygon(polygon,n);
  X:=P[1];
  for k from 1 to nops(P)-1 do
    Q:=[signum(X[1]-P[k+1][1]),signum(X[2]-P[k+1][2])];
    X:=[X[1]-1/k*Q[1],X[2]-1/k*Q[2]];
  end do;
  return(X)
end proc;

```

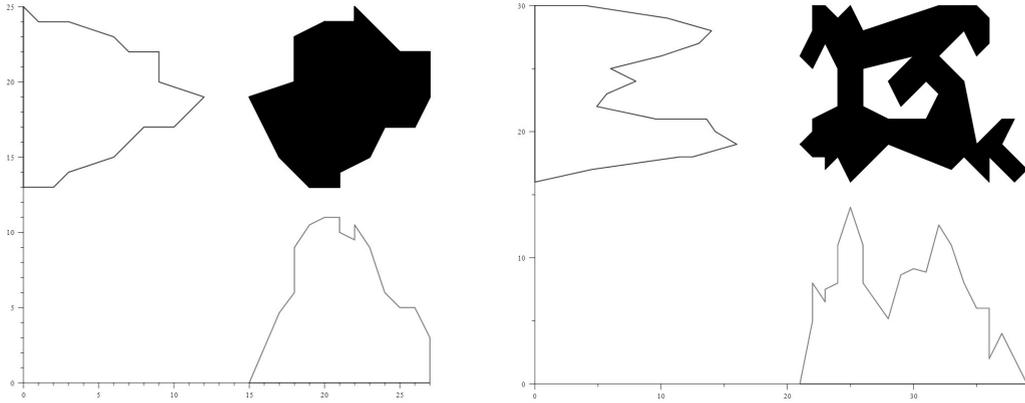


Figure 2. X-rays of compact planar bodies.

3. Applications in geometric tomography

The unweighted function (5) is strongly related to the parallel X-rays as follows: by the Cavalieri principle, formula

$$D_i f_K(x) = \mu_n(K \leq_i x^i) - \mu_n(x^i \leq_i K) \quad (i = 1, \dots, n)$$

of the first partial derivatives implies that

$$D_i D_i f_K(x) =_{\text{a.e.}} 2X_i K(x^i) \quad (i = 1, \dots, n), \tag{12}$$

where $X_i K(x^i) := \mu_{n-1}(x^i =_i K)$ is the $(n - 1)$ -dimensional Lebesgue measure of the set

$$x^i =_i K := \{y \in K \mid y^i = x^i\}. \tag{13}$$

The functions

$$X_i K(t) := \mu_{n-1}(t =_i K) \quad (t \in \mathbb{R} \text{ and } i = 1, \dots, n)$$

are called the coordinate X-rays of K , see Figure 2. In terms of the coordinate X-rays

$$f_K(x) = \int_K d_1(x, y) dy = \sum_{i=1}^n \int_K |x^i - y^i| dy = \sum_{i=1}^n \int_{-\infty}^{\infty} |x^i - t| X_i K(t) dt. \tag{14}$$

Theorem 6. [2] $f_K = f_L$ iff the coordinate X-rays of K and L coincide almost everywhere.

Since the coordinate X-rays determine both the measure and the taxicab distance mean function of the sets, we can formulate the following result as a consequence of Theorem 3.

Corollary 1. [2, 3] The sublevel sets of a taxicab distance mean function are determined by their X-rays parallel to the coordinate hyperplanes among compact sets.

Example 1. Circles are determined by their X-rays in the coordinate directions among compact sets in the plane. They are level sets of the taxicab distance mean function f_B associated to the circumscribed square: if $B := \text{conv} \{(0, 0), (1, 0), (1, 1), (0, 1)\}$, then we have that

$$f_B(x) = (x^1 - (1/2))^2 + (x^2 - (1/2))^2 + (1/2)$$

for any interior point $(x^1, x^2) \in B$. The general problem of unicity for convex bodies can be found in Gardner's basic monograph [8]: characterize those convex bodies that can be determined by two X-rays.

The taxicab distance mean function f_K accumulates the coordinate X-ray information. Instead of the X-rays we can investigate a convex function independently of the convexity of the integration domain. Techniques and results based on f_K are typically working in higher dimensional spaces as well.

3.1. Reconstruction of planar sets by their coordinate X-rays

[2, 9, 10] In what follows we restrict ourselves to the coordinate plane \mathbb{R}^2 . Let $K \subset \mathbb{R}^2$ be a compact subset. The coordinate X-rays of K enable us to construct an axis-parallel bounding box containing K . Since f_K is also given by the coordinate X-rays, the reconstruction is based on the best approximation of f_K by the distance mean functions of a special class of sets. They are constituted by unions of subrectangles of the bounding box under a given resolution: $f_{L_n} \rightarrow f_K$. Taking K^* as the limit set of a convergent subsequence in L_n , we have to provide the continuity of the mapping $L \mapsto f_L$ for a convergent reconstruction process. The continuity implies that the taxicab distance mean functions of K^* and K coincide. So do their coordinate X-rays (almost everywhere). In general X-rays can have deviant behavior under the Hausdorff convergence of the sets [11]. The taxicab distance mean functions are more regular objects in some sense. This makes them to be a natural starting point of the reconstruction.

Let K be a compact subset in the plane. The outer parallel body K_ε is the union of closed Euclidean disks centered at the points of K with radius $\varepsilon > 0$. The Hausdorff distance between the compact subsets K and L is given by the formula

$$\delta(K, L) := \inf\{\varepsilon > 0 \mid K \subset L_\varepsilon \text{ and } L \subset K_\varepsilon\}.$$

Definition 2. [2] The Hausdorff convergence $L_n \rightarrow K$ is called regular if and only if

$$\lim_{n \rightarrow \infty} \mu_2(L_n) = \mu_2(K).$$

It is X-regular if and only if $\lim_{n \rightarrow \infty} \mu_2(I_n) = \mu_2(K)$, where $I_n := \bigcap_{i=n}^{\infty} L_i$.

It can be easily seen that under the hypothesis of the Hausdorff convergence, the regularity is equivalent to the convergence in the symmetric difference metric (or Lebesgue metric). In general they are not equivalent metrics as the following theorem shows.

Theorem 7. [12] The sequence L_n converges in Hausdorff distance to K if and only if

$$\lim_{n \rightarrow \infty} \mu_2((L_n)_\varepsilon \triangle K_\varepsilon) = 0 \text{ for each } \varepsilon > 0,$$

where K_ε is the parallel body of K with radius ε .

Theorem 8. [2, 10] If $L_n \rightarrow K$ with respect to the Hausdorff metric then

$$\limsup_{n \rightarrow \infty} f_{L_n}(x) \leq f_K(x).$$

If the Hausdorff convergence $L_n \rightarrow K$ is regular then $\lim_{n \rightarrow \infty} f_{L_n}(x) = f_K(x)$ and the convergence $f_{L_n} \rightarrow f_K$ is uniform over any compact subset in \mathbb{R}^2 . If the Hausdorff convergence $L_n \rightarrow K$ is X-regular then it is regular and the coordinate X-rays converge to the coordinate X-rays of the limit set almost everywhere:

$$\lim_{n \rightarrow \infty} X_1 L_n(t) =_{\text{a.e.}} X_1 K(t), \quad \lim_{n \rightarrow \infty} X_2 L_n(t) =_{\text{a.e.}} X_2 K(t).$$

We have the following examples:

- (i) If each L_n is obtained from a compact set L via finitely many Steiner symmetrizations and Euclidean isometries then the Hausdorff convergence $L_n \rightarrow K$ is regular [12, Lemma 3.2].
- (ii) Any outer Hausdorff approximation $K \subset L_n \rightarrow K$ is X-regular [2, Lemma 1, Remark 2].
- (iii) Let $f: K \rightarrow D_1 \subset \mathbb{C}$ be a homeomorphism, where D_1 denotes the unit disk of the plane centered at the origin. If the mapping f is differentiable (in complex sense) at each inner point of K then, by Mergelyan's theorem, f can be approximated uniformly on K by polynomials: $P_n \rightarrow f$. Therefore we have an approximation of K by polynomial lemniscate domains of the form $|P_n(z)| \leq c_n$ in the sense that the maximal connected components $L_n^* \subset K$ of the lemniscate domains tends to K with respect to the Hausdorff metric. If K has a boundary of measure zero then the Hausdorff convergence is X-regular [2, Section 5].
- (iv) If L_n is a sequence of compact connected hv-convex sets tending to the limit K with respect to the Hausdorff metric, then the convergence is regular [10, Section 3].
- (v) The Hausdorff convergence of compact convex subsets L_n to K with non-empty interior is X-regular [3, Section 4.1].

In the sense of the last example, the Hausdorff convergence in the class of compact convex sets (with nonempty interior) implies the X-regularity and, by Theorem 8, the reconstruction can be based on direct comparisons of X-rays; see Gardner and Kiderlen [13] (four directions, compact convex planar bodies). Indeed, if the sequence L_n is constructed by the approximation of the X-rays of K , then the X-regularity implies that the X-rays of L_n tend to the X-rays of the accumulation points which also equal to the X-rays of K (almost everywhere). Example (iv) shows that the Hausdorff convergence in the class of compact connected hv-convex sets implies the regularity and the reconstruction can be based on direct comparisons of the taxicab distance mean functions.

Theorem 9. [10] Let \mathcal{M}_B^{hv} denote the set of non-empty compact connected hv-convex sets contained in the axis parallel bounding box $B \subset \mathbb{R}^2$ and let $K \in \mathcal{M}_B^{hv}$. For any $\varepsilon > 0$ there exists $\sigma > 0$ such that whenever

$$\int_B |f_L(x) - f_K(x)| dx < \sigma$$

holds for $L \in \mathcal{M}_B^{hv}$, then there exists K^* , satisfying $\delta(L, K^*) < \varepsilon$ and $f_K = f_{K^*}$. Therefore K and K^* have the same coordinate X-rays almost everywhere.

3.2. An algorithm for the reconstruction

[9] Let $n \in \mathbb{N}$ be a natural number and suppose that the coordinate X-rays X_1K, X_2K of a non-empty compact connected hv-convex planar body $K \subset \mathbb{R}^2$ are given. The Cartesian product of the supports of the coordinate X-rays gives a box

$$B = \text{supp}(X_1K) \times \text{supp}(X_2K) = [a, b] \times [c, d] \quad (15)$$

containing K . The function f_K associated to K is defined by the formula

$$f_K(x) = \int_{-\infty}^{\infty} |x_1 - t| X_1K(t) dt + \int_{-\infty}^{\infty} |x_2 - t| X_2K(t) dt. \quad (16)$$

Let

$$t_i^1 = a + i \frac{b-a}{n} \text{ and } t_j^2 = d - j \frac{d-c}{n} \quad (i, j = 0, \dots, n)$$

be equally spaced points. The control grid $G_K^n := \{y_{ij} \in B_K \mid i, j = 1, \dots, n\}$ consists of the centers of the subrectangles

$$B_{ij}^n = [t_{i-1}^1, t_i^1] \times [t_j^2, t_{j-1}^2], \quad (17)$$

where $i, j = 1, \dots, n$. The feasible set \mathcal{H}_n contains an element L if and only if it is a compact connected hv-convex set which can be written as union of elements of the collection (17) and

$$f_L(y_{ij}) \geq f_K(y_{ij}) \text{ for any } i, j = 1, \dots, n. \quad (18)$$

For the output we choose $L_n \in \mathcal{H}_n$ that minimizes

$$\sum_{i,j=1}^n \frac{f_{L_n}(y_{ij}) - f_K(y_{ij})}{n^2}, \quad (19)$$

see Figures 3, 4 and 5. The procedure can be formulated in terms of a linear 0 - 1 programming because any element L in the feasible set can be represented as a 0 - 1 interval matrix by the variables x_{kl} and $\bar{x}_{kl} = 1 - x_{kl}$, where $x_{kl} = 1$ if $B_{kl}^n \subset L$ and $x_{kl} = 0$ otherwise ($k, l = 1, \dots, n$). The linearization of the constraints is based on [14, chapters 11 and 12]. The applications of the greedy or the antigreedy algorithmic paradigms are also possible [9, sections 7 and 8]. They are based on deleting the subrectangle which causes the extremal (the greatest or the least) average descent of f_{L_n} at the control points. In general the antigreedy version increases the number of the possible outputs for making some voting processes more effective. The algorithm is adapted to finitely many and/or noisy measurements of the coordinate X-rays as well [11].

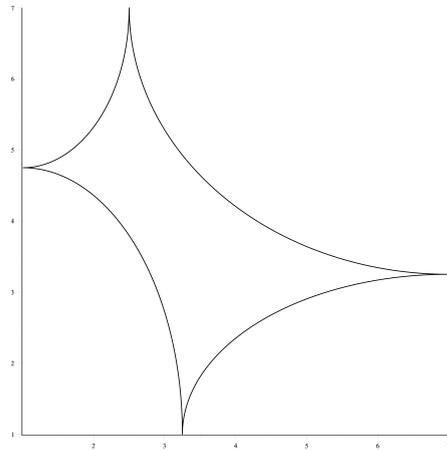


Figure 3. The set we are looking for.

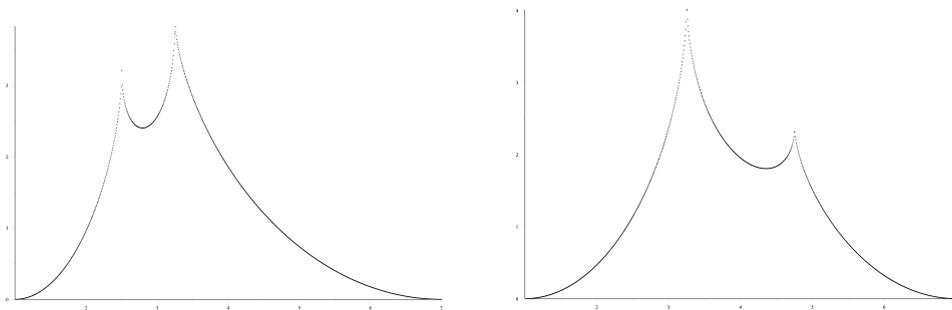


Figure 4. The coordinate X-rays.

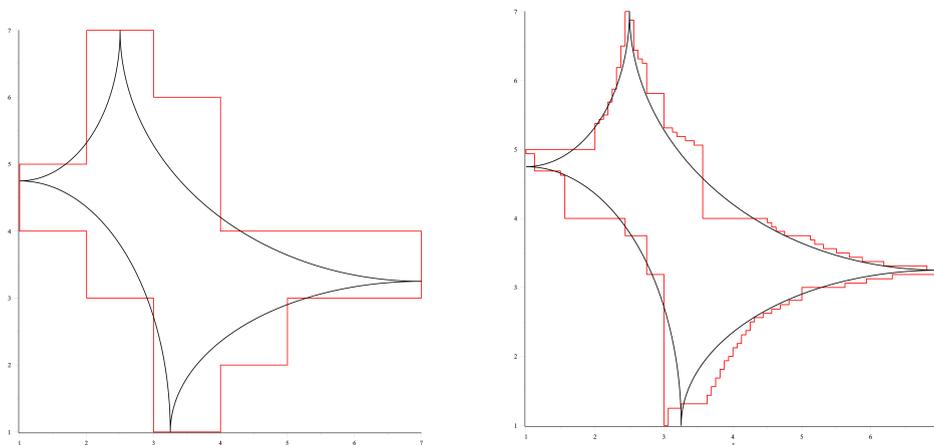


Figure 5. The optimal solution under low resolution (left) and a greedy version under high resolution (right).

4. Reconstruction by the least average values

To illustrate the process let us consider the discrete version [15] of the presented tomographic tools. It is a special case of the general theory with counting measure in the integral formulas. Let $F = \{x_i \in \mathbb{R}^n \mid i = 1, \dots, m\}$ be a finite set of different points in the coordinate space and consider the taxicab distance sum function

$$f(x) := \sum_{i=1}^m d_1(x, x_i) = \sum_{i=1}^m \sum_{j=1}^n |x^j - x_i^j|. \quad (20)$$

Introducing the one-sided partial derivatives

$$D_j^+ f(x) := \lim_{\varepsilon \rightarrow 0^+} \frac{f(x^1, \dots, x^j + \varepsilon, \dots, x^n) - f(x)}{\varepsilon},$$

$$D_j^- f(x) := \lim_{\varepsilon \rightarrow 0^-} \frac{f(x^1, \dots, x^j + \varepsilon, \dots, x^n) - f(x)}{\varepsilon}$$

we have the following collection of formulas:

$$D_j^+ f(x) = |F \leq_j x^j| - |F >_j x^j|, \quad D_j^- f(x) = |F <_j x^j| - |F \geq_j x^j|,$$

where

$$F >_j t := \{x_i \in F \mid x_i^j > t\}, \quad F =_j t := \{x_i \in F \mid x_i^j = t\}, \quad F <_j t := \{x_i \in F \mid x_i^j < t\},$$

$$F \geq_j t := \{x_i \in F \mid x_i^j \geq t\}, \quad F \leq_j t := \{x_i \in F \mid x_i^j \leq t\},$$

$$\frac{D_j^+ f(x) - D_j^- f(x)}{2} = |F =_j x^j| \quad (j = 1, \dots, n).$$

The cardinality $|F =_j x^j|$ is the number of the points in the intersection of F with the hyperplane $x + H_j$, where $H_j := \{x \in \mathbb{R}^n \mid x^j = 0\}$. The $(n - 1)$ -dimensional X-ray function parallel to the coordinate hyperplane H_j is defined as

$$X_j: \mathbb{R} \rightarrow \mathbb{R}, \quad X_j(t) := |F =_j t| \quad (j = 1, \dots, n). \quad (21)$$

X-rays take the zero value except at finitely many $t \in \mathbb{R}$. In terms of X-rays

$$f(x) = \sum_{i=1}^m d_1(x, x_i) = \sum_{i=1}^m \sum_{j=1}^n |x^j - x_i^j| = \sum_{j=1}^n \sum_{t \in \mathbb{R}} X_j(t) |x^j - t| \quad (x \in \mathbb{R}^n). \quad (22)$$

Therefore the taxicab distance sum function accumulates the coordinate X-ray information.

4.1. The least average value principle for the reconstruction of planar lattice sets by coordinate X-rays

Let G be the intersection of the integer lattice $\mathbb{Z} \times \mathbb{Z}$ and the rectangular picture region $[1, n] \times [1, m]$, where $m, n \in \mathbb{Z}$ are integers. Then we can write

$$G = \{1, 2, \dots, n\} \times \{1, 2, \dots, m\} \subset \mathbb{R}^2.$$

For each $j \in \{1, 2, \dots, n\}$ let $l_{1,j}$ denote the vertical line intersecting the horizontal axis at the point $(j, 0)$, and for each $i \in \{1, 2, \dots, m\}$ let $l_{2,i}$ denote the horizontal line intersecting the vertical axis at the point $(0, m - i + 1)$. The problem is to reconstruct the unknown lattice set $F \subset G$, if the numbers

$$p_{1,j} = |l_{1,j} \cap F|, \quad j \in \{1, 2, \dots, n\},$$

and

$$p_{2,i} = |l_{2,i} \cap F|, \quad i \in \{1, 2, \dots, m\}$$

are given. These are the numbers of elements of F contained by each horizontal and vertical lattice line respectively, i.e. $p_{1,j} = X_1(j)$ and $p_{2,i} = X_2(m - i + 1)$. The characteristic function of F can be presented by the binary matrix $A = (a_{ij})$ of size $m \times n$, where

$$a_{ij} = \begin{cases} 1, & \text{if } (j, m - i + 1) \in F, \\ 0, & \text{otherwise,} \end{cases}$$

hence the above problem is equivalent to the following.

Problem 1. Given two integral vectors $R = (r_1, r_2, \dots, r_m)$ and $S = (s_1, s_2, \dots, s_n)$, find a binary matrix $A = (a_{ij})$ of size $m \times n$ such that

$$r_i = \sum_{j=1}^n a_{ij}, \quad i \in \{1, 2, \dots, m\},$$

and

$$s_j = \sum_{i=1}^m a_{ij}, \quad j \in \{1, 2, \dots, n\}.$$

Certainly the above problem may have a solution only if $0 \leq r_i \leq n$, $0 \leq s_j \leq m$ for all $i \in \{1, 2, \dots, m\}$, $j \in \{1, 2, \dots, n\}$, and

$$\sum_{i=1}^m r_i = \sum_{j=1}^n s_j.$$

If this property holds, then the integral vectors R and S are called compatible. In the rest of the paper we assume that the integral vectors R and S are compatible in Problem 1. This problem was first solved independently by Ryser [16, 17] and Gale [18]. Ryser's approach is based on the construction of a maximal matrix³ and shifting certain ones to the right within rows to attain the correct column

³It is a matrix (a_{ij}) such that $a_{ij} = 1$ whenever $j \leq r_i$, otherwise $a_{ij} = 0$.

sums. Gale's approach is based on network flows. This method was later improved by Anstee [19] and Batenburg [20]. The advantage of the network flow method is that it can be easily applied for any pair of lattice directions, and even under the possible restriction, that we accept only those solutions, where a given set of entries are equal to zero and another given set of entries are equal to one. Before discussing the details of the network flow approach we introduce the basic concepts of flows in a network.

Let E and V be two finite sets, and $\varphi: E \rightarrow V \times V$ a function. Then the triple (E, V, φ) is called a directed graph, where E is the set of edges, and V is the set of vertices. If $\varphi(e) = (v_i, v_j)$ for some edge $e \in E$ and ordered pair of vertices $(v_i, v_j) \in V \times V$, then we say v_i is connected to v_j by the edge e . The vertex v_i is called the initial vertex, and v_j is called the terminal vertex of the edge e . The directed path connecting the vertex v_0 to the vertex v_k in a directed graph (E, V, φ) is a sequence

$$(v_0, e_1, v_1, e_2, v_2, e_3, v_3, \dots, v_{k-1}, e_k, v_k),$$

where v_0, v_1, \dots, v_k are pairwise different vertices, and e_1, e_2, \dots, e_k are pairwise different edges, such that $\varphi(e_i) = (v_{i-1}, v_i)$ for all $i = 1, 2, \dots, k$. The undirected path connecting the vertex v_0 to the vertex v_k in a directed graph (E, V, φ) is similar to the directed path except, that now any of $\varphi(e_i) = (v_{i-1}, v_i)$ or $\varphi(e_i) = (v_i, v_{i-1})$ is possible for all $i = 1, 2, \dots, k$. An edge e_i of an undirected path is called forward edge if $\varphi(e_i) = (v_{i-1}, v_i)$, and it's called backward edge if $\varphi(e_i) = (v_i, v_{i-1})$. We say that the length of a directed or undirected path is k if it consists of k edges. There are different efficient methods to find the shortest directed/undirected path connecting a vertex to another, for example with the help of breadth-first search.

A network is a directed graph (E, V, φ) together with a non-negative capacity function $U: E \rightarrow \mathbb{R}$ and two special vertices s and t , such that there's no edge with terminal vertex s and there's no edge with initial vertex t . Then the vertex s is called source, while the vertex t is called sink. The capacity of any edge $e \in E$ is denoted by $U(e)$. A flow on the network (E, V, φ, U, s, t) is a function $Y: E \rightarrow \mathbb{R}$ which satisfies the following two conditions:

- $0 \leq Y(e) \leq U(e)$ for any edge e ,
- for any vertex $v \in V$, except the source and the sink, it's true that the sum of the flow values on all the edges with terminal vertex v is equal to the sum of the flow values on all the edges with initial vertex v .

The later is called the flow conservation property. The value of the flow on any edge $e \in E$ is denoted by $Y(e)$. We say that the edge e is saturated if $Y(e) = U(e)$. The size of a flow Y is the sum of the flow values on all the edges with the source s being the initial vertex. The flow conservation property ensures that the size of the flow also equals to the sum of the flow values on all the edges with the sink t being the terminal vertex.

Now we discuss how a network is constructed for Problem 1, and how a flow of maximal size on the network helps to determine a solution of Problem 1. Let's define the network (E, V, φ, U, s, t) in the following way:

- a vertex v_i is assigned to each horizontal line $l_{2,i}$,

- a vertex w_j is assigned to each vertical line $l_{1,j}$,
- each vertex v_i is connected to every vertex w_j ,
- the source s is connected to every vertex v_i ,
- every vertex w_j is connected to the sink t .

Hence the vertex set is

$$V = (s, t, v_1, v_2, \dots, v_m, w_1, w_2, \dots, w_n),$$

the edge set is

$$E = \{e_{ij} \mid i \in \{1, 2, \dots, m\}, j \in \{1, 2, \dots, n\}\} \cup \{e_{si} \mid i \in \{1, 2, \dots, m\}\} \cup \{e_{jt} \mid j \in \{1, 2, \dots, n\}\},$$

where

$$\varphi(e_{ij}) = (v_i, w_j), \quad \varphi(e_{si}) = (s, v_i), \quad \varphi(e_{jt}) = (w_j, t).$$

Then the capacity function is defined as

- $U(e_{ij}) = 1$, for all $i \in \{1, 2, \dots, m\}$ and $j \in \{1, 2, \dots, n\}$,
- $U(e_{si}) = r_i$, for all $i \in \{1, 2, \dots, m\}$,
- $U(e_{jt}) = s_j$, for all $j \in \{1, 2, \dots, n\}$.

If Y is an integer flow (i.e. flow with integer values), then the flow values $Y(e_{ij})$ are all equal to 0 or 1, since the capacities of the edges e_{ij} are all equal to 1. Thus, having an integer flow Y , we can construct the binary matrix $A = (a_{ij})$ of size $m \times n$ as

$$a_{ij} = Y(e_{ij}), \quad \text{for all } i \in \{1, 2, \dots, m\}, j \in \{1, 2, \dots, n\}.$$

Actually, there's a one-to-one correspondence between integer flows on the network and binary matrices of size $m \times n$. The flow conservation property shows that the binary matrix A corresponding to the integer flow Y has row sums equal to $Y(e_{si})$ and column sums equal to $Y(e_{jt})$, where $Y(e_{si}) \leq r_i$ and $Y(e_{jt}) \leq s_j$. Furthermore it's easy to see that no flow can have size larger than

$$\sum_{i=1}^m U(e_{si}) = \sum_{j=1}^n U(e_{jt}),$$

and whenever the size equals to the above number, then all the edges e_{si} and e_{jt} are saturated. Now our task is to find a flow Y^* of maximal size on the network. All the capacities of the network are integer numbers, and it can be proved that the maximal flow on such network also has integer values. Hence we have the following theorem.

Theorem 10. Problem 1 has a solution if and only if the size of the maximal flow Y^* on the network (E, V, φ, U, s, t) is equal to

$$\sum_{i=1}^m r_i = \sum_{j=1}^n s_j.$$

Then the binary matrix $A = (a_{ij})$ of size $m \times n$ with

$$a_{ij} = Y(e_{ij}), \quad \text{for all } i \in \{1, 2, \dots, m\}, j \in \{1, 2, \dots, n\}$$

is a solution of Problem 1.

The construction of the maximal flow has two stages. In the first stage we construct an initial flow, and in the second stage we increase the size of the flow by changing the flow values along so called flow-augmenting paths. A flow-augmenting path for the flow Y on the network (E, V, φ, U, s, t) is an undirected path connecting the source s to the sink t , such that the forward edges are all unsaturated and the values of the flow Y on the backward edges are strictly positive. If all the capacities are integer numbers and Y is an integer flow, then it's easy to see that increasing the flow values by 1 along the forward edges and decreasing the flow values by 1 along the backward edges results a new flow \hat{Y} with larger size. It's possible to prove that a flow on a network is maximal if and only if there exists no flow-augmenting path. Hence starting with an initial flow, such as the zero flow, we can find a maximal flow by searching for flow-augmenting paths and increasing the size of the flow as long as such path exists. The process is accelerated much, if we always choose a shortest flow-augmenting path, which can be efficiently found with help of breadth-first search.

The construction of the network associated to Problem 1 implies that any flow-augmenting path consists of at least 3 edges. This shows that if we choose the zero flow as initial flow at the first stage, then we first try to increase the flow values along flow-augmenting paths of 3 edges by 1 as long as possible without exceeding the capacities. This means for the matrix A that we first let A be the zero matrix, and try to switch entries of A to 1 as long as possible without exceeding the corresponding row and column sums. How further (i.e. longer) flow-augmenting paths can be found and what they imply on the matrix will be discussed in section 4.3. Now we mention that it's also an important question which ordering of the entries of A is considered when we try to switch them to 1. It turns out that putting different preferences on the entries has a large effect on the number of further flow-augmenting paths required later to attain the maximal flow.

The preference on the entries of A can be based directly on the row sums or column sums, but since the distance mean function has a nice connection to the coordinate X-rays, we can choose preferences upon the values of discrete version of the distance mean function, which is called taxicab distance sum function defined by formula (22). The value of the taxicab distance sum function corresponding to the unknown set F at any point $x = (x^1, x^2) \in \mathbb{R}^2$ can be computed as

$$\begin{aligned} f(x) &= \sum_{j=1}^n X_1(j) |x^1 - j| + \sum_{i=1}^m X_2(i) |x^2 - i| \\ &= \sum_{j=1}^n s_j \cdot |x^1 - j| + \sum_{i=1}^m r_{m-i+1} \cdot |x^2 - i| \\ &= \sum_{j=1}^n s_j \cdot |x^1 - j| + \sum_{i=1}^m r_i \cdot |x^2 - (m - i + 1)|. \end{aligned} \tag{23}$$

The least average value principle means that points of the set G , or equivalently entries of the matrix A , with lowest taxicab distance sum values are preferred first. This implies the following steps:

1. Create the preference list L of entries of A by sorting the entries into increasing order with respect to taxicab distance sum values at the corresponding points of G .
2. Try to switch the entry of A that corresponds to the first element of the preference list. This is possible unless the prescribed row sum or column sum corresponding to that element is zero. Then we say that the first element of the preference list is visited.
3. After having a preference list with some visited elements, look for the first unvisited element of the list and try to switch the corresponding element of A . This is possible if none of the prescribed row sums and column sums are exceeded by the switch. We repeat this as long as the preference list has unvisited elements.

Two accelerating methods can be applied here. The first one is that once we attain a row where the row sum of A equals to the prescribed row sum, then we set all the unvisited entries of A in that row to be visited without switching them to 1. These entries of A remain equal to zero until all elements of the preference list are visited. A similar step can be applied for columns as well. The second accelerating technique is that once we attain a row where the difference of the prescribed row sum and the actual row sum of A equals to the number of unvisited entries in that row, then we switch these unvisited entries of A to 1 in those columns, where the prescribed column sum is not exceeded by the switch. Then we also set all the unvisited entries of A in that row to be visited. A similar step can be applied for columns as well. Certainly applying these accelerating methods has a computational cost, but in change it reduces the number of further necessary switches along flow-augmenting paths.

4.2. Example.

Now we present an example of Problem 1, and starting with the zero matrix A , we show how entries of A are replaced by ones based on the preference list determined by the taxicab distance sum values. We also use the two accelerating method described above. Let $m = n = 5$ and hence $G = \{1, 2, 3, 4, 5\} \times \{1, 2, 3, 4, 5\}$. Furthermore, let the row sums and column sums be

$$(r_1, r_2, r_3, r_4, r_5) = (3, 1, 4, 4, 2) \quad \text{and} \quad (s_1, s_2, s_3, s_4, s_5) = (4, 3, 1, 4, 2)$$

respectively. The matrix A initially equals to the zero matrix. The matrix

$$M := \begin{bmatrix} 54 & 48 & 48 & 50 & 60 \\ 46 & 40 & 40 & 42 & 52 \\ 40 & 34 & 34 & 36 & 46 \\ 42 & 36 & 36 & 38 & 48 \\ 52 & 46 & 46 & 48 & 58 \end{bmatrix}$$

shows the values of the taxicab distance sum function f at the points of the set G , where

$$m_{kl} = f(l, m - k + 1) = \sum_{j=1}^n s_j \cdot |l - j| + \sum_{i=1}^m r_i \cdot |k - i|$$

for all $k, l \in \{1, 2, 3, 4, 5\}$. The lowest taxicab distance sum function value is 34. The first element of the preference list L is the entry with the lowest taxicab distance sum function value. There are two such entries: $(3, 2)$ and $(3, 3)$. We choose, for example, the former one and switch a_{32} to 1 in the matrix A . Then we say that the entry $(3, 2)$ is visited, and the next unvisited element of the preference list is $(3, 3)$. We switch a_{33} to 1.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We see that the 3rd column sum of A equals to $s_3 = 1$, thus we say that all entries in the 3rd column are visited, but we don't change them. Then we see that the difference of r_4 and the 4th row sum of A equals to the number of unvisited entries in the 4th row, thus we switch all the unvisited entries to 1 in the 4th row and say that they are visited.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & \mathbf{1} & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & 0 & \mathbf{0} & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & \mathbf{1} & \mathbf{1} & 0 & 0 \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} \\ 0 & 0 & \mathbf{0} & 0 & 0 \end{bmatrix}$$

The next two unvisited elements of the preference list (i.e. unvisited elements with lowest taxicab distance sum function value) are $(3, 4)$ and then $(2, 2)$ where we switch the matrix to 1, and we call these entries visited.

$$\begin{bmatrix} 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & \mathbf{1} & \mathbf{1} & 0 & 0 \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} \\ 0 & 0 & \mathbf{0} & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & \mathbf{1} & \mathbf{1} & \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} \\ 0 & 0 & \mathbf{0} & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & \mathbf{1} & \mathbf{0} & 0 & 0 \\ 0 & \mathbf{1} & \mathbf{1} & \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} \\ 0 & 0 & \mathbf{0} & 0 & 0 \end{bmatrix}$$

We see that the 2nd row sum of A equals to $r_2 = 1$ and the 2nd column sum of A equals to $s_2 = 3$, thus we say that all entries in the 2nd row and 2nd column are visited, but we don't change them.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Now the difference of r_1 and the 1st row sum of A equals to the number of unvisited entries in the 1st row, thus we switch all the unvisited entries to 1 in the 1st row and say that they are visited.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

At this point we see that the 5th column sum of A equals to the column sum s_5 , while in the 1st, and 4th columns the differences of the prescribed column sums s_1, s_4 and the actual column sums of A equal to the number of unvisited entries in the corresponding columns. Thus, we first say that all entries in the 5th column are visited, but we don't change them. Then we switch the unvisited entries a_{31} and a_{51} of the 1st column to 1 and the unvisited entry a_{54} of the 4th column to 1.

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Finally all elements of the preference list are visited. The final matrix A has row sums $(3, 1, 4, 4, 2) = (r_1, r_2, r_3, r_4, r_5)$ and has column sums $(4, 3, 1, 4, 2) = (s_1, s_2, s_3, s_4, s_5)$, hence no further augmentation is required. In other examples, especially for larger matrices, it's possible that further augmentation is required as some of prescribed row sums and column sums are not attained yet. Anyway, the least average value principle transforms the coordinate X-rays into some geometric information by the values of the taxicab distance sum function. They are working as probability-like quantities whenever the subsequent step of the algorithm is not determined by the X-rays.

4.3. Switching chains

Here we discuss how to find a flow-augmenting path for a flow Y in the network (E, V, φ, U, s, t) constructed for Problem 1, if Y is not maximal, but there exists no flow augmenting path of length 3. Then there exists at least one unsaturated edge e_{si} (connecting the source s to the vertex v_i) and at least one unsaturated edge e_{jt} (connecting the vertex w_j to the sink t). However the edge e_{ij} must be saturated for any such pair of unsaturated edges e_{si} and e_{jt} , because there's no flow augmenting path of length 3. This means that the i -th row sum is less than r_i and the j -th column sum is less than s_j , while $a_{ij} = 1$ in the binary matrix A corresponding to Y . Finding a (shortest) flow-augmenting path containing the edges e_{si} and e_{jt} is equivalent to finding a (shortest) switching chain, i.e. sequence of pair of indexes $(i_0, j_0), (i_1, j_1), \dots, (i_l, j_l)$ that satisfies the following conditions:

- l is an odd number,
- $i_0 = i$ and $j_0 = j$,
- $1 \leq i_k \leq m$ and $1 \leq j_k \leq n$ for all $k \in \{0, 1, \dots, l\}$,
- $i_k = i_{k+1}$ and $j_k \neq j_{k+1}$ for all even numbers $k \in \{0, 1, \dots, l-1\}$,
- $j_k = j_{k+1}$ and $i_k \neq i_{k+1}$ for all odd numbers $k \in \{1, 2, \dots, l-1\}$,
- $j_l = j_0$ and $i_l \neq i_0$,
- $a_{i_k j_k} = 1$ for all even numbers $k \in \{1, 2, \dots, l\}$,
- $a_{i_k j_k} = 0$ for all odd numbers $k \in \{1, 2, \dots, l\}$,

see Figure 6. We can find a shortest flow-augmenting path, and hence a shortest switching chain with the above properties by assigning labels to the entries of A based on a breadth-first search in the associated network. This means that we first assign the label 0 to the entry a_{ij} . If we assume that the highest label assigned to any element of A is the even number k , then we look for unlabeled entries equal to 0 in rows containing entries with label k . If we find such an entry, then we assign the label $k + 1$ to it. If we assume that the highest label assigned to any element of A is the odd number k , then we look for unlabeled entries equal to 1 in columns containing entries with label k . If we find such an entry, then we assign the label $k + 1$ to it. We repeat these steps as long as possible or the column of a_{ij} contains an entry with the highest non-zero label. Otherwise there's no switching chain that

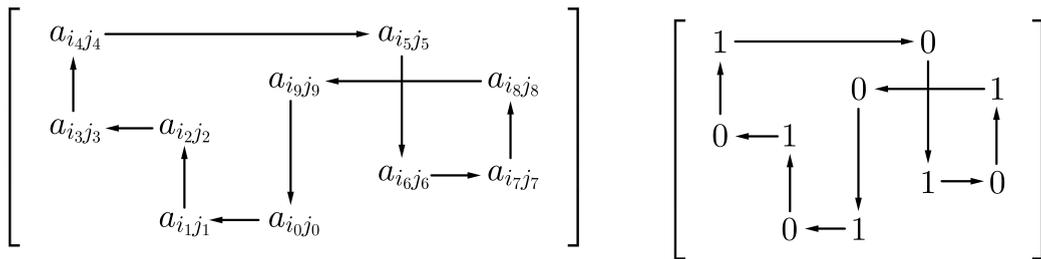


Figure 6. A switching chain of 10 elements.

satisfies the above conditions. Let $a_{i_l j_l} = 0$ be an entry with the highest label l in the column of a_{ij} . Then there must be at least one entry, $a_{i_{l-1} j_{l-1}} = 1$, with label $l - 1$ in the row of $a_{i_l j_l}$. The column of $a_{i_{l-1} j_{l-1}}$ must contain at least one entry, $a_{i_{l-2} j_{l-2}} = 0$, with label $l - 2$. This can be continued until we find an entry $a_{i_1 j_1} = 0$ with label 1 in the column of $a_{i_2 j_2}$ and in the row of a_{ij} . Thus $(i, j) = (i_0, j_0), (i_1, j_1), \dots, (i_l, j_l)$ is a shortest switching chain that satisfies the above conditions.

It's easy to see, that by interchanging zeros and ones in a switching chain doesn't change the row sums and column sums of the matrix A . Hence, if the i -th row sum of A is less than the prescribed row sum r_i and the j -th column sum of A is less than the prescribed column sum s_j , but there exists a switching chain containing the entry $a_{ij} = 1$, then interchanging the zeros and ones in the switching chain makes $a_{ij} = 0$, and we can switch this to $a_{ij} = 1$ to increase the i -th row sum and j -th column sum of A by 1. Note that the switch leaves other row sums and column sums unchanged. It's a well-known result in the theory of network flows, that a flow is maximal if and only if there exists no flow-augmenting path. This ensures that, if the tomographic problem has a solution, then a switching chain exists. Therefore, starting with any initial matrix we can find a solution with the help of finitely many switches. The existence of the switching chain can be proved directly with the help of Mirsky's theorem on integer matrices [21] as the last section shows.

5. The existence of the switching chain

Let $R = (r_1, r_2, \dots, r_m)$ and $S = (s_1, s_2, \dots, s_n)$ be a pair of compatible integral vectors and let $\mathfrak{A}(R, S)$ denote the set of all 0 - 1 matrices of size $m \times n$ with row sums equal to R and column sums equal to S . We would like show that if both $\mathfrak{A}(R, S)$ and $\mathfrak{A}(R + \delta, S + \varepsilon)$ are non-empty for some integral vectors $\delta = (\delta_1, \delta_2, \dots, \delta_m)$ and $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ such that $R + \delta$ and $S + \varepsilon$ are compatible too, then having a binary matrix $A \in \mathfrak{A}(R, S)$ it's either possible to switch a zero entry of A to 1 without exceeding the row sums in $R + \delta$ and column sums in $S + \varepsilon$, or there exist at least one switching chain in A . The proof is based on the following theorem.

Theorem 11. Mirsky [21] Let $0 \leq r'_i \leq r''_i, 0 \leq s'_j \leq s''_j$ and $c_{ij} \geq 0$ be integers ($i = 1, 2, \dots, m$), ($j = 1, 2, \dots, n$). Then there exists an integral matrix $A = (a_{ij})$ of size $m \times n$ such that

$$\begin{aligned} r'_i &\leq \sum_{j=1}^n a_{ij} \leq r''_i && (i = 1, 2, \dots, m) \\ s'_j &\leq \sum_{i=1}^m a_{ij} \leq s''_j && (j = 1, 2, \dots, n) \\ 0 &\leq a_{ij} \leq c_{ij} && (1 \leq i \leq m, 1 \leq j \leq n) \end{aligned}$$

if and only if, for all $I \subset \{1, 2, \dots, m\}, J \subset \{1, 2, \dots, n\}$,

$$\sum_{i \in I} \sum_{j \in J} c_{ij} \geq \max \left\{ \sum_{i \in I} r'_i - \sum_{j \notin J} s''_j, \sum_{j \in J} s'_j - \sum_{i \notin I} r''_i \right\}.$$

Corollary 2. The set $\mathfrak{A}(R, S)$ is non-empty with compatible integral vectors $R = (r_1, r_2, \dots, r_m)$ and $S = (s_1, s_2, \dots, s_n)$ if and only if for all $I \subset \{1, 2, \dots, m\}$, $J \subset \{1, 2, \dots, n\}$,

$$|I| \cdot |J| \geq \sum_{i \in I} r_i - \sum_{j \notin J} s_j = \sum_{j \in J} s_j - \sum_{i \notin I} r_i. \tag{24}$$

Proof:

Let's choose $r'_i = r''_i = r_i$, $s'_i = s''_i = s_i$, and $c_{ij} = 1$ for all $i \in \{1, 2, \dots, m\}$ and $j \in \{1, 2, \dots, n\}$. Then, by Mirsky's theorem, the set $\mathfrak{A}(R, S)$ is non-empty if and only if for all $I \subset \{1, 2, \dots, m\}$, and $J \subset \{1, 2, \dots, n\}$,

$$|I| \cdot |J| \geq \max \left\{ \sum_{i \in I} r_i - \sum_{j \notin J} s_j, \sum_{j \in J} s_j - \sum_{i \notin I} r_i \right\}.$$

On the other hand

$$\begin{aligned} \sum_{i \in I} r_i - \sum_{j \notin J} s_j &= \sum_{i \in I} r_i + \sum_{i \notin I} r_i - \sum_{i \notin I} r_i - \sum_{j \notin J} s_j = \\ &= \sum_{i=1}^m r_i - \sum_{i \notin I} r_i - \sum_{j \notin J} s_j = \sum_{j=1}^n s_j - \sum_{i \notin I} r_i - \sum_{j \notin J} s_j = \\ &= \sum_{j \in J} s_j + \sum_{j \notin J} s_j - \sum_{i \notin I} r_i - \sum_{j \notin J} s_j = \sum_{j \in J} s_j - \sum_{i \notin I} r_i \end{aligned}$$

and we are done. □

Given the compatible integral vectors $R = (r_1, r_2, \dots, r_m)$ and $S = (s_1, s_2, \dots, s_n)$ let $\delta = (\delta_1, \delta_2, \dots, \delta_m)$ and $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ be a pair of integral vectors with $\delta_i \in [0, n - r_i]$ for all $i \in \{1, 2, \dots, m\}$ and $\varepsilon_j \in [0, m - s_j]$ for all $j \in \{1, 2, \dots, n\}$. Then we define the sets $I_0 = \{i \mid \delta_i > 0\}$ and $J_0 = \{j \mid s_j > 0\}$. Let's assume that I_0 and J_0 are nonempty.

Theorem 12. If the sets $\mathfrak{A}(R, S)$ and $\mathfrak{A}(R + \delta, S + \varepsilon)$ are non-empty, then there are indices $i' \in I_0$ and $j' \in J_0$ and there is a matrix $A = (a_{ij})$ in $\mathfrak{A}(R, S)$, such that $a_{i'j'} = 0$.

Proof:

Let's choose arbitrary elements $i_0 \in I_0$ and $j_0 \in J_0$. By Mirsky's theorem and Corollary 2, there exists a matrix $A \in \mathfrak{A}(R, S)$ with $a_{i_0j_0} = 0$ if and only if for all subsets $I \subset \{1, 2, \dots, m\}$, $J \subset \{1, 2, \dots, n\}$,

$$\sum_{i \in I} \sum_{j \in J} c_{ij} \geq \sum_{i \in I} r_i - \sum_{j \notin J} s_j = \sum_{j \in J} s_j - \sum_{i \notin I} r_i,$$

where

$$c_{ij} = \begin{cases} 0 & \text{if } i = i_0 \text{ and } j = j_0 \\ 1 & \text{otherwise} \end{cases}$$

This means that, if we assume that there's no matrix $A \in \mathfrak{A}(R, S)$ with $a_{i_0 j_0} = 0$, then there are subsets $I \subset \{1, 2, \dots, m\}$, $J \subset \{1, 2, \dots, n\}$, such that

$$\sum_{i \in I} \sum_{j \in J} c_{ij} < \sum_{i \in I} r_i - \sum_{j \notin J} s_j = \sum_{j \in J} s_j - \sum_{i \notin I} r_i. \quad (25)$$

It's not possible that $i_0 \notin I$ or $j_0 \notin J$, since otherwise inequality (25) means

$$|I| \cdot |J| < \sum_{i \in I} r_i - \sum_{j \notin J} s_j = \sum_{j \in J} s_j - \sum_{i \notin I} r_i,$$

which implies, that $\mathfrak{A}(R, S)$ is empty. Thus inequality (25) has the following form

$$|I| \cdot |J| - 1 < \sum_{i \in I} r_i - \sum_{j \notin J} s_j = \sum_{j \in J} s_j - \sum_{i \notin I} r_i.$$

This inequality is equivalent to the equation

$$|I| \cdot |J| = \sum_{i \in I} r_i - \sum_{j \notin J} s_j = \sum_{j \in J} s_j - \sum_{i \notin I} r_i, \quad (26)$$

because $\mathfrak{A}(R, S)$ is non-empty and inequality (24) holds. It's not possible that $I_0 \subset I$, since otherwise $\sum_{j \in J} \varepsilon_j > 0$ and $\sum_{i \notin I} \delta_i = 0$, and hence equation (26) leads to

$$|I| \cdot |J| = \sum_{j \in J} s_j - \sum_{i \notin I} r_i < \sum_{j \in J} s_j + \varepsilon_j - \sum_{i \notin I} r_i = \sum_{j \in J} s_j + \varepsilon_j - \sum_{i \notin I} r_i + \delta_i,$$

which implies, that $\mathfrak{A}(R + \delta, S + \varepsilon)$ is empty. Similarly, it's not possible that $J_0 \subset J$, since otherwise $\sum_{i \in I} \delta_i > 0$ and $\sum_{j \notin J} \varepsilon_j = 0$, and hence equation (26) leads to

$$|I| \cdot |J| = \sum_{i \in I} r_i - \sum_{j \notin J} s_j < \sum_{i \in I} r_i + \delta_i - \sum_{j \notin J} s_j = \sum_{i \in I} r_i + \delta_i - \sum_{j \in J} s_j + \varepsilon_j,$$

which implies, that $\mathfrak{A}(R + \delta, S + \varepsilon)$ is empty. Thus there are elements $i_1 \in I_0 \setminus I$ and $j_1 \in J_0 \setminus J$. On the other hand if $A = (a_{ij})$ is any matrix $A \in \mathfrak{A}(R, S)$, then

$$\begin{aligned} \sum_{i \notin I} \sum_{j \notin J} a_{ij} &= \sum_{i \notin I} r_i - \sum_{i \notin I} \sum_{j \in J} a_{ij} = \sum_{i \notin I} r_i - \sum_{j \in J} s_j + \sum_{i \in I} \sum_{j \in J} a_{ij} = \\ &= \sum_{i \notin I} r_i - \sum_{j \in J} s_j + |I| \cdot |J| - |I| \cdot |J| + \sum_{i \in I} \sum_{j \in J} a_{ij}. \end{aligned}$$

Hence

$$\sum_{i \notin I} \sum_{j \notin J} a_{ij} + \left(|I| \cdot |J| - \sum_{i \in I} \sum_{j \in J} a_{ij} \right) = |I| \cdot |J| - \left(\sum_{j \in J} s_j - \sum_{i \notin I} r_i \right).$$

This gives, by equality (26), that

$$\sum_{i \notin I} \sum_{j \notin J} a_{ij} + \left(|I| \cdot |J| - \sum_{i \in I} \sum_{j \in J} a_{ij} \right) = 0$$

and here

$$|I| \cdot |J| - \sum_{i \in I} \sum_{j \in J} a_{ij} \geq 0,$$

thus

$$\sum_{i \notin I} \sum_{j \notin J} a_{ij} = 0.$$

This is possible only if $a_{ij} = 0$ for all $i \notin I$ and $j \notin J$, including $i_1 \in I_0 \setminus I$ and $j_1 \in J_0 \setminus J$, which gives $a_{i_1 j_1} = 0$. Thus finally we can conclude the following.

If $i_0 \in I_0$ and $j_0 \in J_0$ and there's no matrix $A \in \mathfrak{A}(R, S)$ with $a_{i_0 j_0} = 0$, then there are indices $i_1 \in I_0$ and $j_1 \in J_0$, such that $a_{i_1 j_1} = 0$ for any matrix $A \in \mathfrak{A}(R, S)$. Hence $(i', j') = (i_0, j_0)$ or $(i', j') = (i_1, j_1)$ makes the statement true. \square

Consider now two compatible integral vectors R and S . Assume that none of the row sums of a binary matrix $A = (a_{ij})$ are larger than the corresponding row sums in R , and none of the column sums of A are larger than the corresponding column sums in S . Let I_0 denote the set of all those indexes i , where the i -th row sum of A is strictly less than the prescribed row sum r_i , and let J_0 denote the set of all those indexes j , where the j -th column sum of A is strictly less than the prescribed column sum s_j . If $a_{ij} = 1$ for all pair of indexes $(i, j) \in I_0 \times J_0$, then we can't switch any of the zero entries of A to 1 without exceeding the prescribed row and column sums. By Theorem 12, there must be another binary matrix $\tilde{A} = (\tilde{a}_{ij})$ with the same row sums and column sums as A , and with $\tilde{a}_{i' j'} = 0$ for some $i' \in I_0$ and $j' \in J_0$ provided that the set $\mathfrak{A}(R, S)$ is nonempty. Ryser showed in [16] that if two matrices have the same row and column sums, then they can be transformed into each other with the help of finitely many switches in so-called switching components, i.e. switching chains of 4 elements. The entries of A and \tilde{A} are different in the intersection of i' -th row and j' -th column, hence (i', j') must be contained in one of the switching components that transform A to \tilde{A} . Merging all these switching components results in a switching chain containing (i', j') . Thus Theorem 12 ensures the existence of the switching chain.

References

- [1] Vincze C, Kovács Z, Csorvássy Z. On the generalization of Erdős-Vincze's theorem about the approximation of closed convex plane curves by polyellipses. *Annales Mathematicae et Informaticae*, 2018. **49**:181–197. doi:10.33039/ami.2018.11.002.
- [2] Vincze C, Nagy A. On the theory of generalized conics with applications in geometric tomography. *J. of Approx. Theory*, 2012. **164**:371–390. doi:10.1016/j.jat.2011.11.004.

- [3] Vincze C, Nagy A. On the average taxicab distance function and its applications. *Acta Appl. Math.*, 2019. **161**:201–220. doi:10.1007/s10440-018-0210-1.
- [4] Barczy M, Nagy A, Noszály C, Vincze C. A Robbins-Monro type algorithm for computing global minimizer of generalized conic functions. *Optimization*, 2015. **64**(9):1999–2020. doi:10.1080/02331934.2014.919499.
- [5] Garey MR, Johnson DS, Preparata FP, Tarjan RE. Triangulating a simple polygon. *Information Processing Letters*, 1978. **7**(4):175–179. doi:10.1016/0020-0190(78)90062-5.
- [6] Lee D, Preparata FP. Location of a point in a planar subdivision and its applications. *SIAM Journal on Computing*, 1977. **6**(3):594–606. doi:10.1137/0206043.
- [7] Meisters GH. Polygons have ears. *The American Mathematical Monthly*, 1975. **82**(6):648–651. doi:10.2307/2319703.
- [8] Gardner RJ. Geometric Tomography. Cambridge University Press, New York, 2006.
- [9] Vincze C, Nagy A. Reconstruction of hv-convex sets by their coordinate X-ray functions. *Journal of Mathematical Imaging and Vision*, 2014. **49**(3):569 – 582. doi:10.1007/s10851-013-0487-7.
- [10] Vincze C, Nagy A. Generalized conic functions of hv-convex planar sets: continuity properties and X-rays. *Aequationes Mathematicae*, 2015. **89**:1015 – 1030. doi:10.1007/s00010-014-0322-2.
- [11] Vincze C, Nagy A. An algorithm for the reconstruction of hv-convex planar bodies by finitely many and noisy measurements of their coordinate X-rays. *Fundamenta Informaticae*, 2015. **141**(2-3):169 – 189. doi:10.3233/FI-2015-1270.
- [12] Bianchi G, Burchard A, Gronchi P, Volcic A. Convergence in Shape of Steiner Symmetrization. *Indiana University Math. Journal*, 2012. **61**(4):1695–1709. doi:10.1512/iumj.2012.61.5087.
- [13] Gardner RJ, Kiderlen M. A solution to Hammer’s X-ray reconstruction problem. *Advances in Mathematics*, 2007. **214**(1):323–343. doi:10.1016/j.aim.2007.02.005.
- [14] Li D, Sun X. Nonlinear Integer Programming. Springer, New York, 2006.
- [15] Vincze C. On the taxicab distance sum function and its applications in discrete tomography. *Periodica Mathematica Hungarica*, 2019. **79**:177 – 190. doi:10.1007/s10998-018-00278-7.
- [16] Ryser HJ. Combinatorial properties of matrices of zeros and ones. *Canadian Journal of Mathematics*, 1957. **9**:371–377. doi:10.4153/CJM-1957-044-3.
- [17] Ryser HJ. Matrices of zeros and ones. *Bulletin of the American Mathematical Society*, 1960. **66**(6):442 – 464. doi:10.1090/S0002-9904-1960-10494-6.
- [18] Gale D. A theorem on flows in networks. *Pacific Journal of Mathematics*, 1957. **7**(2):1973–1982. doi:10.2140/pjm.1957.7.1073.
- [19] Anstee RP. The network flow approach for matrices with given row and column sums. *Discrete Mathematics*, 1983. **44**(2):125–138. doi:10.1016/0012-365X(83)90053-5.
- [20] Batenburg KJ. A network flow algorithm for reconstructing binary images from discrete X-rays. *Journal of Mathematical Imaging and Vision*, 2007. **27**:175–191. doi:10.1007/s10851-006-9798-2.
- [21] Mirsky L. Combinatorial theorems and integral matrices. *Journal of Combinatorial Theory*, 1968. **5**(1):30–44. doi:10.1016/S0021-9800(68)80026-2.