ELSEVIER

Discrete Applied Mathematics 75 (1997) 39–61

DISCRETE APPLIED MATHEMATICS

# The discrete Radon transform and its approximate inversion via linear programming

Peter Fishburn[a,*], Peter Schwander[b,1], Larry Shepp[a], Robert J. Vanderbei[c]

[a] *AT&T Bell Laboratories, P.O. Box 636, 600 Mountain Avenue, Murray Hill, NJ 07974-0636, USA*

[b] *ETH, Zurich; formerly AT&T Bell Laboratories*

[c] *Princeton University, Princeton, NJ 08544, USA*

Received 20 April 1995; revised 3 June 1996

**Abstract**

Let $S$ be a *finite* subset of a lattice and let $v_S(L)$, the number of points of $S \cap L$ for each line $L$, denote the discrete Radon transform of $S$. The problem is to reconstruct $S$ from a knowledge (possibly noisy) of the restriction of $v_S$ to a subset $\mathscr{L}$ of the set of all lines in any of a few given directions through the lattice. Reconstructing a density from its line integrals is a well-understood problem, but discreteness causes many difficulties and precludes use of continuous Radon inversion algorithms. Indeed it has been shown that when the directions are *main* directions of the lattice, the case for most applications, the problem is finite but is NP-hard, so that any reconstruction algorithm will surely have to consist of exponentially many steps in the size of $S$.

We address this problem by looking instead for a *fuzzy set* $S$ with the given line sums, i.e. a function $f(z)$ with $0 \leqslant f(z) \leqslant 1$ for all points $z$ in the lattice, for which $v_f(L) \equiv v_S(L)$. The set of all such $f$ forms a convex set and those $f = \chi_S$ with each $f(z) \in \{0,1\}$ are extreme points. Finding a fuzzy set $f$ with the given line sums is a linear programming problem and so there are efficient algorithms for finding $f$ or proving that no such $f$ (and hence no $S$) exists with the given line sums.

If $S$ is an *additive set* with respect to $\mathscr{L}$, i.e. if we can write $S = \{z : \sum_{\mathscr{L} \ni L \ni z} g(L) > 0\}$ for some functional $g$ on $\mathscr{L}$, we show that there is only one fuzzy set $f$ with the given line sums. We prove here that if $S$ is *not* additive then there are *many fuzzy* sets with the given line sums, although there still may be only one actual *set*. Linear programming methods that are based on interior point methods always produce solutions that lie in the center of the convex set of all solutions. As a result, if $S$ is a set with given line sums and linear programming produces a solution that consists only of $\{0,1\}$ then this solution must be the original subset $S$, and $S$ must be a set of uniqueness. Thus interior point LP's give a polynomial and practical way to obtain the assertion of uniqueness when strong uniqueness obtains.

This problem arises in a practical situation, although it was earlier studied in the case of coordinate directions for its intrinsic interest. In the practical situation, $S$ represents a piece of a real crystal, and the line sums in any fixed direction can be measured (possibly with uncertainties) using a transmission electron microscope.

* Corresponding author. E-mail: fish@research.att.com.

[1] Formerly at AT& T Bell Laboratories, Murray Hill, NJ, USA.



*PII* S0166-218X(96)00083-2

In order to study the behavior of the linear programming reconstruction algorithm and to explore the question of whether a "typical" set $S$ with a "typical" $\mathscr{L}$ will be a case of uniqueness or near-uniqueness, we conducted simulation experiments. These indicate that the method is practical and reasonably efficient, at least on the examples we considered.

## 1. Discrete Radon transform

The interplay between the continuous and discrete is a deep and important theme in mathematics. Passage to the limit often simplifies the solution of a problem, and continuous approximate models are used to exploit this fact. On the other hand, discrete problems are often solvable by finite exhaustive search, at least in principle, whereas the continuous version may remain impossible to solve. Our problem will illustrate this theme.

Suppose $S$ is a finite subset of $\mathbb{Z}^2$ or $\mathbb{Z}^3$, the integer lattice in 2 or 3 dimensions. For any line $L$ in $\mathbb{R}^2$ or $\mathbb{R}^3$, let $v_S(L)$ be the number of points in $S \cap L$, so that $v_S(L) \in \{0, 1, \ldots\}$. Suppose $\mathscr{L}$ is a restricted set of lines, and we know $v_S(L)$ for all $L \in \mathscr{L}$. The problem is to reconstruct $S$ from its line sums $v_S(L)$ for $L \in \mathscr{L}$. In practice, there may be noise and we consider the more general problem: Given $v(L)$ on $\mathscr{L}$, we ask about existence and uniqueness of $S$ for which $v_S(L) = v(L)$ on $\mathscr{L}$. This is the problem we study in this paper and it was the problem which motivated the Mini-Symposium on Discrete Tomography which met at DIMACS, September 19, 1994. Ideas presented at this meeting due to Ron Aharoni et al. [2] and the earlier work of Gabor Herman and Richard Gordon [6] suggested to us that linear programming would play a role in the solution. We pursue their suggestion in our paper.

In the usual continuous model for tomography one attempts to reconstruct a function $f(z)$ for $z$ in $\mathbb{R}^2$ or $\mathbb{R}^3$ from a knowledge of its line integrals (rather than line sums), $\int_L f \, ds, L \in \mathscr{L}$. Appropriate [11, 8] quadratures of the Radon inversion formula are used, with Fourier transforms, Jacobians, and other concepts from calculus and continuous mathematics playing the main role. Unless $S$ is extremely large, $|S| \approx 10^{23}$, which is not the case in the practical problem motivating us, discussed below, a continuous model is not appropriate and a convolution-back-projection type approach [8] seems unlikely to work in practice. Yet there is a way to bring in some continuity via "fuzzy" sets and this is the approach we will take.

An application of discrete tomography arises in High-Resolution Transmission Electron Microscopy of crystals as follows: A parallel beam of electrons of high energy is directed at a small piece of crystal. After passage through both the crystal and a high magnification lens, the electrons form an image, either on a detector or on photographic film. The microscope resolution is sufficient that for some (main) directions individual atomic columns, each corresponding to a line sum, can be resolved. The image contrast at each atomic column depends on the number of atoms that are contained in each individual column. Although there is a complicated relationship between image contrast and number of atoms due to the physics of electron scattering and the image formation process, it has been demonstrated in [9] that this is indeed possible. The technique, named QUANTITEM, is based on vector pattern recognition. QUANTITEM deduces a signal from the image that is directly proportional to the number of atoms contained in each atomic column. For a small crystal the measured values must be approximately integral multiples of a fixed quantity. Thus, we are able to measure line sums, $v_S(L)$, along each line, $L$, corresponding to an atomic column. Of course if this direction has an irrational slope then each line $L$ will have at most one integer lattice point on it and $S$ can be reconstructed exactly from a knowledge of $v_S(L)$ for all lines $L$ in this single direction. In practice, this is impossible because along an irrational direction the atomic columns would be too closely spaced to be resolvable by any microscope. In fact, the direction must be a main direction $(0,1),(1,0),(1,1),(1,2),\ldots$ in $\mathbb{R}^2$, or $(0,0,1),(0,1,0),(1,0,0),(1,1,1),\ldots$ in $\mathbb{R}^3$. Thus for the experimenter to be able to discern the true direction, the atoms must "line up." There are other practical problems in choosing directions and taking measurements: (a) the high-energy electrons can damage the crystal by displacing the atoms (irradiation damage), and this sets a practical limit on the total exposure time and therefore also on the total number of projections; (b) the stage of the microscope is not easily tilted by more than 15° in any direction, so that e.g. in $\mathbb{R}^3, (1,1,1),(3,2,2),(2,3,2),(2,2,3)$ give a particularly nice set of four directions where the last three are "tiltable" from the first, but to develop mathematical understanding we will permit ourselves not to be bound by these engineering practicalities; (c) there may be errors in the line sum measured values, e.g., if $v_S(L) = 37.3$ after the subtraction, it is tempting to round off to 37 since $v_S(L)$ must be an integer, but noise occurs due to various sources: (1) if the atoms of the crystal are shifted away from their exact lattice sites, (2) the fact that only a finite number of electrons are used to form the image results in statistical or Poisson-like noise (think of a grainy image). This noise may mean that the true line sum is 36, 38 or even farther away from 37. Nevertheless, it is clear from the step-like nature of the measured values that the discrete Radon transform is the natural mathematical model to use in reconstructing the crystal. It is believed that a good mathematical algorithm for inverting the discrete Radon transform would offer a way to image small crystals in 3D. This might be helpful in understanding the solid-state physics of VLSI in present and future designs with ever smaller crystal sizes.

The problem of inverting the discrete Radon transform is such a natural one that it was studied [5] for its intrinsic mathematical interest for coordinate directions even before the problem became of interest in electron microscopy. It was shown in [5] that a subset of $\mathbb{Z}^2$ is uniquely reconstructible from its line sums in the $x$ and $y$ directions if and only if it is an additive set, i.e. $S = \{(x,y) \in \mathbb{Z}^2 : a(x) + b(y) > 0\}$ where $a$ and $b$ are any real-valued maps on $\mathbb{Z}$. This result is easily generalized to any set of lines $\mathscr{L}$, and we have the following *sufficient* condition on $S$ and $\mathscr{L}$ for uniqueness.

**Proposition 1.** *If $\mathscr{L}$ is any set of lines $L$ and if $S$ is an additive set in the sense that there exists $g(L) \in \mathbb{R}$ for each $L \in \mathscr{L}$ for which $S$ is represented as*

$$S = \{z \in \mathbb{Z}^n : \sum_{\mathscr{L} \ni L \ni z} g(L) > 0\} , \tag{1.1}$$

*then $S$ is unique. That is, if $T \subset \mathbb{Z}^n$ and $v_T(L) = v_S(L)$ for all $L \in \mathscr{L}$, then $T = S$.*

This holds for $n = 2, 3, \ldots$ The proof is simple but we delay it because the same proof will show that $S$ is unique even among fuzzy sets that we discuss first.

Irwing and Jerrum [7] have proven that the problem of reconstructing $S$ from $v_S(L)$ or $v(L)$ for $L \in \mathscr{L}$ is in general NP-hard, so that it is expected on the basis of NP-completeness theory that no algorithm that runs in polynomial time in $|S|$ can be found. The problem becomes amenable if continuity is introduced into it by relaxing the deterministic membership requirement of a lattice point in the set $S$. We do this by looking for a fuzzy set $f$, which is a function $f(z)$ for $z \in \mathbb{Z}^n$ with $0 \leqslant f(z) \leqslant 1$ at each $z$, that has the right line sums:

$$v_f(L) = \sum_{z \in L} f(z) = v(L) \quad \text{for all } L \in \mathscr{L} . \tag{1.2}$$

If $f(z) \in \{0,1\}$ for every $z \in \mathbb{Z}^n$ then $f$ is the characteristic function $\chi_S$ where $S = \{z \in \mathbb{Z}^n : f(z) = 1\}$. Hence a fuzzy set $f$ can be a set. Of course, every atom in a crystal either is or is not there so that fuzzy sets do not correspond to real crystals unless they are zero-one. However, fuzzy sets are mathematically convenient.

It is obvious that the family of all fuzzy sets $f$ with the given $v(L)$ line sums forms a compact and convex set $C$ that we refer to later as $F_{S,D}$. Any point $f \in C$ which is actually a set is of course an extreme point of $C$ but the converse is not necessarily true. Indeed there may be no sets in $C$ even though $C$ is non-empty. An example of this situation is as follows: $\mathscr{L}$ consists of all lines in $\mathbb{R}^2$ with direction parallel to the $x$-axis, $y$-axis, or at 135° except for 3 lines: the vertical line through (1,0), the horizontal line through (0,1) and the 135° line through (0,0). $v(L) = 0$ except for 3 lines: the vertical line through (0,0) and (0,1), the horizontal line through (0,0) and (1,0), and the 135° line through (1,0) and (0,1). For these 3 lines $v(L) = 1$: see Fig. 1.

It is clear from the figure that $f(z) = \frac{1}{2}$ at (0,0), (1,0), and (0,1), and $f(z) = 0$ otherwise, is a fuzzy set which belongs to $C$ but that no *set* $S$ can satisfy $v_S = v$ in this case. Thus, $C$ is non-empty but has no points (fuzzy sets) which are actually sets. We will see later that it is very common even in typical cases with real crystals and error-free measurements that extreme points of $C$ are only fuzzy sets and not sets; in this case the $f(z)$ will be rational numbers.

There are several reasons for introducing fuzzy sets when the object is to find actual sets with the given line sums. For one, the problem of deciding whether $C$ is empty or not is a feasibility problem of linear programming because each $0 \leqslant f(z) \leqslant 1$ is a linear constraint and each $v_f(L) = v(L)$ is a linear equation in the unknowns $f(z)$. Efficient algorithms which run in polynomial time in $|S|$ exist for linear programming

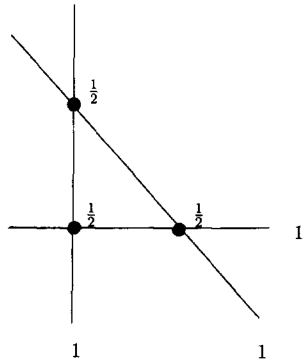


Fig. 1.

problems. Further, if $C$ is found to be empty then there are no sets $S$ with $v_S = v$. Thus, the problem of deciding whether there is a set with given line sums is no longer NP-complete or NP-hard. Of course if $v$ is obtained by noise-free measurements from an actual crystal, then $C$ is certainly not empty and contains at least one set $S$. If $S$ is additive for $\mathscr{L}$ in the sense of Proposition 1 then, as we now prove, there is no fuzzy set $f$ with the given line sums other than $S$. Indeed suppose $f$ is a fuzzy set and $S$ is an additive set for $\mathscr{L}$ with the same line sums:

$$v_f(L) = v_S(L), \quad L \in \mathscr{L} . \tag{1.3}$$

We show that $f = \chi_S$. Since $S$ is additive, (1.1) holds for some functional $g$ on $\mathscr{L}$, and we may write from (1.3)

$$\begin{aligned} 0 = \sum_{L\in\mathscr{L}} g(L)(v_S(L) - v_f(L)) &= \sum_{L\in\mathscr{L}}\sum_{z\in L} g(L)(\chi_S(z) - f(z)) \\ &= \sum_{z\in\mathbb{Z}^n}(\chi_S(z) - f(z))\sum_{L\ni z} g(L) . \end{aligned} \tag{1.4}$$

But if we sum first on $L \in \mathscr{L}$ holding $z \in \mathbb{Z}^n$ fixed, then for $z \in S$, $\sum_{L\ni z} g(L) > 0$ and $\chi_S(z) - f(z) \geqslant 0$ since $\chi_S(z) = 1$ when $z \in S$ and $0 \leqslant f(z) \leqslant 1$, so the terms in the sum on $z \in S$ on the right of (1.4) are nonnegative. But if $z \notin S$, then $\sum_{L\ni z} g(L) < 0$ (note that it is always possible if $S$ is additive to define $g(L)$ so that $\sum_{L\ni z} g(L) \neq 0$ for all lines $L \in \mathscr{L}$, as can be seen by adding $\varepsilon$ to $g$), and $\chi_S(z) - f(z) \leqslant 0$ since $\chi_S(z) = 0$ when $z \notin S$ and $0 \leqslant f(z) \leqslant 1$. Thus every term in the sum on $z$ on the right of (1.4) is nonnegative and, since the total sum is zero, every term must be zero and $\sum_{L\ni z} g(L) \neq 0$, we must have $f \equiv \chi_S$. This proves Proposition 1 even for fuzzy sets. Proposition 1 generalizes an earlier version [5] proved for special sets $\mathscr{L}$.

The converse of Proposition 1 is false (as was observed in [5]) in the sense that there are examples of $v$'s for which there exist exactly one *set* $S$ with $v_S(L) = v(L)$ on $\mathscr{L}$, but $S$ is not additive for $\mathscr{L}$. Indeed, simulations indicate that this is probably typical, so additivity is probably not a good indicator that $\mathscr{L}$ is rich enough to determine

$S$ uniquely. But the linear program itself can be used to give insight into the uniqueness question, as we shall see. The following proposition is stated more formally in Section 2.

**Proposition 2.** *If $S$ is a set and $v_S(L) = v(L)$ on a set $\mathscr{L}$ of all lines parallel to some line in a finite set, and if $S$ is not additive, then $S$ is also not unique among fuzzy sets (although there may be no other* set *with the same line sums as $S$).*

Since $C$ is a convex set, once there is a nondegenerate fuzzy set with the same line sums as $S$, then $C$ is not a singleton and so there is a continuum of fuzzy sets $\lambda\chi_S + (1-\lambda)f$, $0 \leqslant \lambda \leqslant 1$, all with the same line sums. We give the proof of a more precise formulation of Proposition 2 in the next section: see Theorem 2.

It is possible to give examples of a pair of sets $S$ and $T$ with the same line sums in any $n$ described directions. Techniques for doing this are described in the following section. We are grateful to Ron Graham for first pointing this out and for the additional observation that the *existence* of such pairs also follows from a simple pigeon-hole counting argument, illustrated in the special case of dimension 2. We show here that there exists a pair of sets $S, T$ each contained in $\{1,\ldots,k\} \times \{1,\ldots,k\}$ if $k > c\ n \log n$ which have the same line sums for any fixed set of *main* directions, where a main direction means that the direction cosines are rational. Indeed, Graham observes that each line sum of a set $S \subset \{1,\ldots,k\}^2$ is an integer in $\{0,\ldots,k\}$ and so the set of possible vectors $\{v(L) : L \in \mathscr{L}\}$ of line sums has no more than $(k+1)^{cn^2}$ elements. But the number of subsets of $\{1,\ldots,k\}^2$ is $2^{k^2}$, which is larger than $(k+1)^{cn^2}$ if $k > c\ n \log n$, and the result follows.

Let us return now to the general problem. We are given $v(L)$ on $\mathscr{L}$, where $\mathscr{L}$ is typically the set of all lines in a few fixed directions. We solve the linear program and find, say, that it is feasible, i.e. $C$ is not empty and so there exist fuzzy set solutions.

The next step is to find a set solution if one exists. That is, we seek a feasible solution all of whose values are zero or one. This is a special case of an integer programming problem, and these problems are generally NP-hard. However, certain subclasses of the family of integer programming problems can be solved in polynomial time (such as network flow problems with integer data), and algorithms that work fairly well in practice for the general case are known. The most common of these is the so-called *branch-and-bound* algorithm.

Since our problem is just a feasibility problem, simpler heuristics suggest themselves. For example, we can introduce an arbitrary linear function and maximize it to obtain an extreme point. One can hope that this extreme point solution will correspond to a set. But, as we shall see in Section 7, these extreme point solutions may fail to be zero–one valued. So, one can next look at the fuzzy set given by the feasibility program and the set $T = \{z : f(z) > \lambda\}$ where $0 < \lambda < 1$, perhaps $\lambda = \frac{1}{2}$ or $\frac{3}{4}$, and set up a new linear program which maximizes $\sum_{z\in T} f(z)$ over $C$. The solution to the new problem is likely to be even closer to integrality. If it is not integral, one can continue in the same way until eventually an integer solution is found. Our experience with the above practical uses of linear programming is described in Section 7, which is devoted to simulation experiments.

## 2. Definitions and theorems

This section presents our main definitions and analytical results. We begin with the dimensionality $n \geqslant 2$ of our basic space, the integer lattice $\mathbb{Z}^n \subseteq \mathbb{R}^n$, and a finite family $D = \{d^1, d^2, \ldots, d^m\} \subseteq \mathbb{Z}^n \backslash \{\mathbf{0}\}$ of $m \geqslant 2$ directional vectors with integer components. We refer to $d^j = (d_1^j, \ldots, d_n^j)$ as a *direction* and assume that the $m$ directions are distinct in the sense that if $i \neq j$ and $\lambda \in \mathbb{R}$ then $d^i \neq \lambda d^j$.

Let $\mathscr{L}_D$ denote the set of all lines $L$ in $\mathbb{R}^n$ that are parallel to $d^j$ for some $d^j \in D$ and contain at least one point (hence an infinite number of points) in $\mathbb{Z}^n$. For every $x \in \mathbb{Z}^n$ and $L \in \mathscr{L}_D$ let

$$v_x(L) = \begin{cases} 1 & \text{if } x \in L, \\ 0 & \text{otherwise} . \end{cases}$$

Hence each $x \in \mathbb{Z}^n$ has $v_x(L) = 1$ for exactly $m$ lines in $\mathscr{L}_D$. For every nonempty finite $S \subseteq \mathbb{Z}^n$ we define a mapping $v_S : \mathscr{L}_D \to \{0, 1, 2, \ldots\}$ by

$$v_S(L) = |\{x \in S : x \in L\}| = \sum_{x \in S} v_x(L) ,$$

so that $v_S$ counts the number of points in $S$ that lie on each line in $\mathscr{L}_D$. It is natural to refer to $v_S(L)$ as a line sum to distinguish our discrete approach from discussions of "continuous" sets of uniqueness based on line integrals.

Let $S$ be a nonempty finite subset of $\mathbb{Z}^n$, and let $D$ be a family of $m$ directions in $\mathbb{Z}^n \backslash \{\mathbf{0}\}$. We say that $S$ *is a set of uniqueness with respect to* $D$, or that $S$ *is* $D$*-unique*, if there exists no $T \subseteq \mathbb{Z}^n$ such that $T \neq S$ and $v_T = v_S$. Consequently, if $v : \mathscr{L}_D \to \{0, 1, 2, \ldots\}$ happens to be the line sum function for some set of uniqueness with respect to $D$, then precisely one $S \subseteq \mathbb{Z}^n$ has $v_S = v$. And if $v$ is not the line sum function for some $D$-unique set, then $\{S \subseteq \mathbb{Z}^n : v_S = v\}$ either is empty or has at least two members.

We consider a notion of additive set alongside the notion of a set of uniqueness. With $S$ and $D$ as above, we say that $S$ *is additive with respect to* $D$, or that $S$ *is* $D$*-additive*, if there is a mapping $g : \mathscr{L}_D \to \mathbb{R}$ such that, for all $x \in \mathbb{Z}^n$,

$$x \in S \Leftrightarrow \sum_{L \in \mathscr{L}_D} v_x(L) g(L) > 0 .$$

Because the sum's positivity is preserved when a small positive constant is subtracted from $g$, we will assume that, when $S$ is $D$-additive with $g$ as specified, it is true also that

$$\sum_{L \in \mathscr{L}_D} v_x(L) g(L) < 0 \quad \text{for every } x \in \mathbb{Z}^n \backslash S .$$

Additivity is more restrictive than uniqueness in the sense that, for some $D$ with $m \geqslant 3$, every $D$-additive set is also $D$-unique, but not every $D$-unique set is $D$-additive. We refer to the latter type of $S$ as a *nonadditive set of uniqueness*. Our comparison between uniqueness and additivity will be enhanced by considering functions that map $\mathbb{Z}^n$ into $\{0,1\}$ or $[0,1]$. Let

$$E_{S,D} = \{f : \mathbb{Z}^n \to \{0,1\} : \sum_{x \in L \cap \mathbb{Z}^n} f(x) = v_S(L) \quad \text{for every } L \in \mathscr{L}_D\},$$

$$F_{S,D} = \{f : \mathbb{Z}^n \to [0,1] : \sum_{x \in L \cap \mathbb{Z}^n} f(x) = v_S(L) \quad \text{for every } L \in \mathscr{L}_D\}.$$

Functions in $E_{S,D}$ are extreme members of the set $F_{S,D}$ of fuzzy sets, and each such function is the characteristic function $\chi_T$ of some $T \subseteq \mathbb{Z}^n$. Our definition of uniqueness implies directly that $S$ is $D$-unique if and only if $E_{S,D} = \{\chi_S\}$. The more inclusive set $F_{S,D}$, which allows $f$ values strictly between 0 and 1 and was denoted by $C$ in the preceding section, is obviously closed under convex combinations. We will see below that if $S$ is $D$-additive then $F_{S,D} = E_{S,D}$, whereas if $S$ is a nonadditive set of uniqueness then $E_{S,D} \subset F_{S,D}$.

The final definitions needed for our main theorems involve balances of line sums between members of $S$ and lattice points not in $S$. A *K-bad D-configuration* for $S (K \geqslant 2)$ is a pair of lists consisting of $K$ distinct points $x^1, \ldots, x^K$ in $S$ and $K$ distinct points $y^1, \ldots, y^K$ in $\mathbb{Z}^n \backslash S$ such that

$$\sum_{k=1}^{K} v_{x^k}(L) = \sum_{k=1}^{K} v_{y^k}(L) \quad \text{for every } L \in \mathscr{L}_D.$$

We refer to a 2-bad $D$-configuration as a *bad rectangle*. Because each point in $\mathbb{Z}^n$ lies on $m$ lines in $\mathscr{L}_D$, bad rectangles can exist only when $m = 2$.

A *weakly K-bad D-configuration* for $S$ is the same as a $K$-bad $D$-configuration except that points in the lists $x^1, \ldots, x^K$ and $y^1, \ldots, y^K$ need not be distinct. We will also express this by noting explicitly the multiplicity of each distinct point in the lists. Let $(\gamma_1 z^1, \ldots, \gamma_I z^I)$ denote a list of $\Sigma\gamma_i$ points in $\mathbb{Z}^n$ in which $z^i$ appears exactly $\gamma_i$ times ($i = 1, \ldots, I$), the $z^i$ are distinct, and each $\gamma_i$ is a positive integer. Then a weakly $K$-bad $D$-configuration consists of $(\alpha_1 x^1, \ldots, \alpha_I x^I)$ from $S$ and $(\beta_1 y^1, \ldots, \beta_J y^J)$ from $\mathbb{Z}^n \backslash S$ such that $\Sigma\alpha_i = \Sigma\beta_j = K$ and

$$\sum_{i=1}^{I} \alpha_i v_{x^i}(L) = \sum_{j=1}^{J} \beta_j v_{y^j}(L) \quad \text{for every } L \in \mathscr{L}_D .$$

We say that $S$ has *no bad D-configuration (no weakly bad D-configuration)* if there is no $K$-bad $D$-configuration (no weakly $K$-bad $D$-configuration) for every $K \geqslant 2$.

The following theorems apply to all $n \geqslant 2$, all nonempty finite $S \subseteq \mathbb{Z}^n$, and all direction families $D \subseteq \mathbb{Z}^n \backslash \{\mathbf{0}\}$ with $|D| = m \geqslant 2$, unless restricted otherwise in context. The first two theorems characterize uniqueness and additivity, respectively.

**Theorem 1.** *The following are mutually equivalent:*
1. *$S$ is $D$-unique;*
2. *$S$ has no bad $D$-configuration;*
3. $E_{S,D} = \{\chi_S\}$.

**Theorem 2.** *The following are mutually equivalent:*
1. *$S$ is $D$-additive;*
2. *$S$ has no weakly bad $D$-configuration;*
3. $F_{S,D} = \{\chi_S\}$.

An immediate corollary is that $S$ is a nonadditive set of uniqueness if and only if $F_{S,D}$ has an infinity of solutions but only one 0–1 solution, namely $\chi_S$.

**Theorem 3.** *There exist nonadditive sets of uniqueness. In particular, when $n = 2$ and $m = 3$, there is a nonadditive $D$-unique $S \subseteq \mathbb{Z}^2$ with $|S| \leqslant 11$.*

The explicit construction of nonadditive sets of uniqueness has proved rather difficult even though it might be true that nonadditive uniqueness is the rule rather than the exception. However, nonadditive sets of uniqueness cannot occur when $D$ has only two directions.

**Theorem 4.** *Suppose $m = 2$. Then the following are mutually equivalent:*
1. *$S$ is $D$-unique;*
2. *$S$ is $D$-additive;*
3. $F_{S,D} = \{\chi_S\}$;
4. *$S$ has no weakly bad $D$-configuration;*
5. *$S$ has no bad $D$-configuration;*
6. *$S$ has no bad rectangle.*

Our final three theorems illustrate aspects of planar uniqueness with three or four directions. We note first that, unlike Theorem 4(6) for $m = 2$, there is no upper bound on $K$, apart from $|S|$, in deciding whether $S$ has a $K$-bad $D$-configuration when $m = 3$. In the language of logic [10], this says that the planar theory of uniqueness for three directions is not axiomatizable by a universal sentence.

**Theorem 5.** *Suppose $n = 2$ and $m = 3$. Then for every $K \geqslant 3$ there is an $S \subseteq \mathbb{Z}^2$ with $|S| = K$ such that $S$ is not $D$-unique but every nonempty proper subset of $S$ is $D$-unique.*

Theorem 1 implies that such an $S$ has a $K$-bad $D$-configuration but no $J$-bad $D$-configuration for $J < K$. A similar result obtains for even $K \geqslant 4$ at $m = 4$ when the four directions are the most natural ones.

**Theorem 6.** *Suppose* $n = 2$ *and* $D^* = \{(1,0),(0,1),(1,1),(-1,1)\}$. *For every even* $K \geqslant 4$ *there is an* $S \subseteq \mathbb{Z}^2$ *with* $|S| = K$ *such that* $S$ *is not* $D^*$*-unique but every nonempty proper subset of* $S$ *is* $D^*$*-unique. For every odd* $K \geqslant 7$ *there is an* $S \subseteq \mathbb{Z}^2$ *with* $|S| = K$ *that has a* $K$*-bad* $D^*$*-configuration.*

Theorem 6 leaves two matters unresolved for the given $D^*$. The first is whether any odd $K \geqslant 7$ has an $S$ with $|S| = K$ that is not $D^*$-unique while every nonempty proper subset is $D^*$-unique. We leave this as an open problem.

The second matter concerns $K = 5$. An example in [3, pp. 234–235] shows that, when $|S| = 5$, most families of four directions admit an $S$ which has a 5-bad configuration. Because their analysis is not confined to $\mathbb{Z}^2$, we generalize to their context.

**Theorem 7.** *Suppose* $n = 2$ *and* $D = \{(1,0),(0,1),(1,r),(-1,s)\}$ *with* $r$ *and* $s$ *any positive numbers. Then there exist disjoint* $A, B \subseteq \mathbb{R}^2$ *with* $|A| = |B| = 5$*, such that every line in a* $D$ *direction contains the same number of* $A$ *points as* $B$ *points if, and only if,* $r \neq s$.

The conclusion for $r = s = 1$ says that, in sharp distinction to the conclusion of Theorem 6 for odd $K \geqslant 7$, there is no five-point $S$ that has a 5-bad $D^*$-configuration. Theorem 7 corrects an oversight in [3, p. 234] which alleged a 5-bad configuration for every family of four directions.

We prove Theorems 1, 2 and 4 in the next section, Theorem 3 in Section 4, Theorems 5 and 6 in Section 5, and Theorem 7 in Section 6.

## 3. Equivalence proofs: Theorems 1, 2 and 4

**Proof of Theorem 1 for Uniqueness.** Statements (1) and (3) are equivalent by the definitions.

(1) $\Rightarrow$ (2): If $S$ has a bad $D$-configuration, we contradict uniqueness by replacing in $S$ the points in the $S$-list of the bad configuration by those in its $\mathbb{Z}^n \backslash S$ list.

(2) $\Rightarrow$ (1): If $S$ is not $D$-unique because $T \neq S$ and $v_T = v_S$, then $v_{T\backslash(S\cap T)} = v_{S\backslash(S\cap T)}$ and the points in $S\backslash(S \cap T)$ and in $T\backslash(S \cap T)$ form two lists that comprise a bad $D$-configuration.

**Proof of Theorem 2 for Additivity.** If $S$ has a weakly bad $D$-configuration, let it consist of $(\alpha_1 x^1, \ldots, \alpha_I x^I)$ from $S$ and $(\beta_1 y^1, \ldots, \beta_J y^J)$ from $\mathbb{Z}^n \backslash S$ such that $\Sigma\alpha_i = \Sigma\beta_j$ and

$$\sum_{i=1}^{I} \alpha_i v_{x^i}(L) = \sum_{j=1}^{J} \beta_j v_{y^j}(L) \quad \text{for every } L \in \mathscr{L}_D \ .$$

(1) $\Rightarrow$ (2): Suppose (2) fails so that $S$ has a weakly bad $D$-configuration. Let $g$ be any mapping from $\mathscr{L}_D$ into $\mathbb{R}$. We multiply both sides of the preceding equation by

$g(L)$, then sum over $\mathscr{L}_D$ and reverse orders of summation to conclude that

$$\sum_{i=1}^{I} \alpha_i \sum_{\{L \in \mathscr{L}_D : x^i \in L\}} g(L) = \sum_{j=1}^{J} \beta_j \sum_{\{L \in \mathscr{L}_D : y^j \in L\}} g(L) .$$

If $\sum_{\{L:x^i \in L\}} g(L) > 0$ for all $x^i$, then $\sum_{\{L:y^j \in L\}} g(L) > 0$ for at least one $y^j$. But $y^j \notin S$, so $S$ cannot be $D$-additive, and (1) fails.

(2) $\Rightarrow$ (1): This implication is an application of the linear separation theorem (a.k.a. Farkas's lemma, Motzkin's lemma, fundamental theorem of linear programming, and the theorem of the alternative, among others) with rational coefficient vectors. The rational version of the separation theorem is stated, for example, in [4, Ch. 5]. Its application in our present context is very similar to its use in the proof of Theorem 2 in [5]. Briefly, either there exists $g : \mathscr{L}_D \to \mathbb{R}$ which shows that $S$ is $D$-additive, or (exclusive form) there exist integer multiplicities of points in $S$ and in $\mathbb{Z}^n \backslash S$ that exhibit the line sum balance that defines a weakly bad $D$-configuration. We omit details. It may be noted that the application also establishes (1) $\Rightarrow$ (2) by the exclusive *or*, in a manner tantamount to the demonstration in the preceding paragraph.

(1) $\Rightarrow$ (3): See our proof of Proposition 1 around (1.3) and (1.4) in Section 1.

(3) $\Rightarrow$ (2): We show that not (2) $\Rightarrow$ not (3). Assume that $S$ has a weakly bad $D$-configuration. For small positive $\lambda$ let

$$b_j = \lambda \beta_j \text{ and } a_i = 1 - \lambda \alpha_i \quad \text{for } i = 1, \ldots, I \text{ and } j = 1, \ldots, J ,$$

so that $0 < b_j < 1$ and $0 < a_i < 1$ for all $i$ and $j$. It then follows immediately from $\Sigma_i \alpha_i v_{x^i}(L) = \Sigma_j \beta_j v_{y^j}(L)$ that, for all $L \in \mathscr{L}_D$,

$$\sum_{i=1}^{I} a_i v_{x^i}(L) + \sum_{j=1}^{J} b_j v_{y^j}(L) = \sum_{i=1}^{I} v_{x^i}(L) .$$

Define $f : \mathbb{Z}^n \to [0,1]$ by

$$\begin{aligned} f(x^i) &= a_i, \quad i = 1, \ldots, I, \\ f(y^j) &= b_j, \quad j = 1, \ldots, J, \\ f(z) &= \chi_S(z) \quad \text{for all } z \in \mathbb{Z}^n \backslash (\{x^i\} \cup \{y^j\}) . \end{aligned}$$

Then $\Sigma_{x \in L \cap \mathbb{Z}^n} f(x) = v_S(L) = \Sigma_{x \in L \cap \mathbb{Z}^n} \chi_S(x)$ for all $L \in \mathscr{L}_D$, so $f \in F_{S,D}$. Because $f \neq \chi_S$, (3) fails.

We have (2) $\Rightarrow$ (1) $\Rightarrow$ (3) $\Rightarrow$ (2), and the proof of Theorem 2 is complete.

**Proof of Theorem 4 for two directions.** Our definitions and Theorems 1 and 2 give (4) $\Rightarrow$ (3) $\Rightarrow$ (2) $\Rightarrow$ (1) $\Rightarrow$ (5) $\Rightarrow$ (6). It remains to prove that (6) $\Rightarrow$ (4). We do this by showing that (6) $\Rightarrow$ (5) and (5) $\Rightarrow$ (4), i.e., that if $S$ has a weakly bad $D$-configuration then it has a bad $D$-configuration [not (4) $\Rightarrow$ not (5)] and that if $S$ has a bad $D$-configuration then it has a bad rectangle [not (5) $\Rightarrow$ not (6)]. The proof is similar to the proof of Lemma 1 in [5], but since that proof pertains to $D = \{(1,0),(0,1)\}$ in two dimensions we note its generalization to $D = \{d^1, d^2\}$ in $n$ dimensions.

(5) $\Rightarrow$ (4): For $i = 1,2$ and $x, y \in \mathbb{Z}^n$ let $x \sim_i y$ mean that $x$ and $y$ are contained in a line parallel to the line through $\mathbf{0}$ and $d^i$. Suppose that $S$ has a weakly bad $D$-configuration composed of lists $x^1,\ldots,x^K$ from $S$ and $y^1,\ldots,y^K$ from $\mathbb{Z}^n \backslash S$, multiplicities allowed. Beginning with $a_1 = x^1$, form an alternating sequence

$$a_1 \sim_1 b_1 \sim_2 a_2 \sim_1 b_2 \sim_2 a_3 \sim_1 b_3 \sim_2 a_4 \sim_1 b_4 \cdots$$

in which each $a_i$ is an $x^k$ and each $b_i$ is a $y^k$. Our weakly bad hypothesis ensures that the sequence continues until we come to an $x^k$ or $y^k$ encountered previously, at which point we stop. The part of the sequence between the stopping point and its earlier identical twin yields a bad $D$-configuration with lists of distinct $x^k$ and distinct $y^k$. The line sum balance required for a bad $\{d^1, d^2\}$-configuration is guaranteed by the alternating character of the sequence.

(6) $\Rightarrow$ (5): Suppose $S$ has a $K$-bad $D$-configuration, $K \geqslant 3$, with distinct-point lists $x^1,\ldots,x^K$ from $S$ and $y^1,\ldots,y^K$ from $\mathbb{Z}^n \backslash S$. If two $x^k$ and two $y^k$ form a bad rectangle, we are done. Otherwise, we can assume that

$$x^1 \sim_1 y^1 \sim_2 x^2 \sim_1 y^2 \sim_2 x^3 \sim_1 y^3 .$$

Let $t \in \mathbb{R}^n$ complete the parallelogram whose other three corners are $x^2, y^2$ and $x^3$, so that $x^2 \sim_2 t \sim_1 x^3$ and

$$y^1 \sim_2 t \sim_1 y^3 .$$

The lattice structure implies that $t$ is in $\mathbb{Z}^n$. If $t \notin S$ then $\{x^2, x^3, y^2, t\}$ forms a bad rectangle. If $t \in S$, we replace $x^2$ and $x^3$ by $t$ in the $S$ list and delete $y^2$ from the non-$S$ list to produce a $(K-1)$-bad $D$-configuration. We then repeat the process on the smaller bad configuration, and continue until a bad rectangle appears.

## 4. A nonadditive set of uniqueness: Theorem 3

We present an 11-point example of an $S$ in $\mathbb{Z}^2$ that is a nonadditive set of uniqueness with respect to $D = \{(1,0),(0,1),(1,1)\}$. We then show that, for any other family of three directions, an affine transformation maps our example into an 11-point nonadditive set of uniqueness with respect to that family.

Fig. 2 denotes our 11 points in $S$ by $a, b, \ldots, k$. Their specific positions in the grid are

$$\begin{array}{llll} a = (10,11), & d = (11,0), & g = (0,32), & j = (38,24), \\ b = (7,9), & e = (24,19), & h = (48,70), & k = (49,8) . \\ c = (9,7), & f = (16,35), & i = (35,25), & \end{array}$$

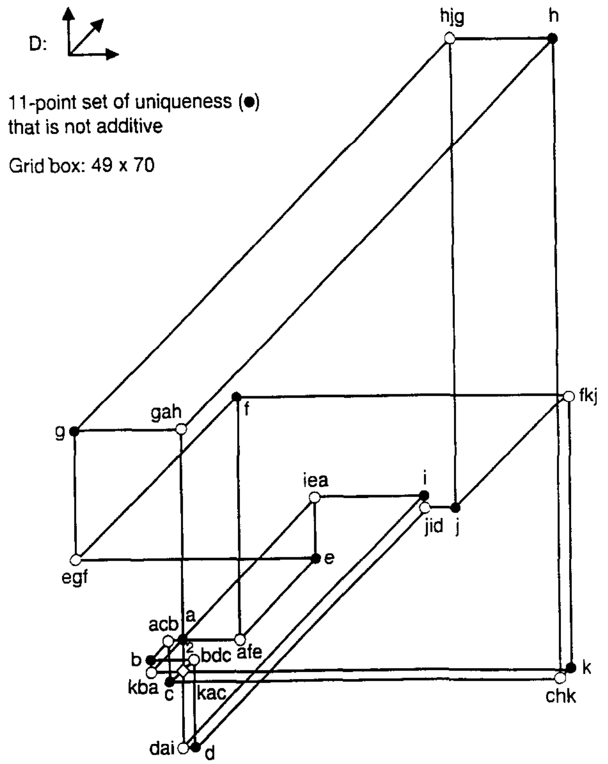


Fig. 2. A nonadditive set of uniqueness.

Let *rst* denote a non-$S$ grid point that lies on the horizontal line through $r \in S$, the vertical line through $s \in S$, and the 45° line through $t \in S$. There are 13 *rst* points. They appear in Fig. 2 as the diamond *kac* and 12 open circles, namely

| | | | |
|---|---|---|---|
| *acb* | *gah* | *bdc* | *chk* |
| *afe* | *kba* | *egf* | *fkj* |
| *dai* | *iea* | *jid* | *hjg* . |

These 12 form the non-$S$ list of a weakly 12-bad $D$-configuration whose $S$ list is $a, a, b, c, d, \ldots, k$. By Theorem 2, $S$ is not $D$-additive.

We show that $S$ has no bad $D$-configuration, so it is $D$-unique. Note that if *rst* is part of the non-$S$ list of a bad configuration then this list must have triples with $r$ second, $r$ third, $s$ first, $s$ third, $t$ first, and $t$ second. We show that this leads to a contradiction.

Suppose $S$ has a bad $D$-configuration with $S$ list $A$ and non-$S$ list $B$. Suppose *kac* is in $B$. Then $c$ is in $A$, so *acb* is in $B$; then $b$ is in $A$, so *bdc* is in $B$. But then $c$ appears twice in the third position of $B$ members, i.e., in *kac* and *bdc*, so $c$ would have multiplicity two or more in the $A$ list, a contradiction. We conclude that *kac* is not in $B$.

Suppose $h$ appears in some triple of the non-$S$ list $B$. Then all letters in $\{a,b,\dots,k\}$ appear in $B$. In particular, since the only $rst$ triples with $h$ are $hjg, chk$, and $gah$, each of $j,g,c,k$, and $a$ appears in $B$; since the only triples with $j$ in first or third position are $jid$ and $fkj$, each of $i,d$ and $f$ appears in $B$; then $i$'s presence requires $e$ (by $iea$), and $d$'s presence requires $b$ (by $bdc$). Moreover, if any letter other than $h$ appears in $B$, then $h$ also appears in $B$ by similar reasoning. It follows that all 12 of the $rst$ triples other than $kac$ are in $B$. This is a contradiction because $A$ has at most 11 terms.

Now consider another family $D' = \{(\alpha_1,\alpha_2),(\beta_1,\beta_2),(\gamma_1,\gamma_2)\}$ of three distinct directions in $\mathbb{Z}^2\backslash\{\mathbf{0}\}$. Because the directions are different, we have $\alpha_1\beta_2 \neq \alpha_2\beta_1$, $\alpha_1\gamma_2 \neq \alpha_2\gamma_1$ and $\beta_1\gamma_2 \neq \beta_2\gamma_1$. We base our affine transformation for $D'$ on the nonsingular matrix

$$A = \begin{bmatrix} \alpha_1(\beta_1\gamma_2-\beta_2\gamma_1) & \gamma_1(\alpha_1\beta_2-\alpha_2\beta_1) \\ \alpha_2(\beta_1\gamma_2-\beta_2\gamma_1) & \gamma_2(\alpha_1\beta_2-\alpha_2\beta_1) \end{bmatrix},$$

which maps

$$\begin{aligned} &(1,0) \text{ into } c_1(\alpha_1,\alpha_2), && c_1 = \beta_1\gamma_2-\beta_2\gamma_1,\\ &(0,1) \text{ into } c_2(\gamma_1,\gamma_2), && c_2 = \alpha_1\beta_2-\alpha_2\beta_1,\\ &(1,1) \text{ into } c_3(\beta_1,\beta_2), && c_3 = \alpha_1\gamma_2-\alpha_2\gamma_1 . \end{aligned}$$

The transformation by $A$ is one-to-one from $\mathbb{R}^2$ into $\mathbb{R}^2$, preserves the parallel property, and maps grid points into grid points. Hence $A$ maps our example for $D$ into an example of a nonadditive set of uniqueness with respect to $D'$.

## 5. Special nonunique sets in the plane: Theorems 5 and 6

**Proof of Theorem 5 for three directions.** Fix $(n,m) = (2,3)$ and $D = \{(1,0),(0,1),(1,1)\}$. We show for every $K \geqslant 3$ that there is a $K$-point $S \subseteq \mathbb{Z}^2$ that has a $K$-bad configuration but no $J$-bad configuration for $J < K$. The affine transformation at the end of the preceding section gives the same result for every family $D'$ of three directions.

We define $x^k$ and $y^k$ for $k = 1,\dots,K$, $K \geqslant 3$, by

$$\begin{aligned} &x^1 = (1,K-1), && y^1 = (2,K), &&\\ &x^k = (k+1,K-k+2), && y^k = (k-1,K-k) && \text{for } k \text{ even}, k<K,\\ &x^k = (k-1,K-k), && y^k = (k+1,K-k+2) && \text{for } k \text{ odd}, 3 \leqslant k < K \end{aligned}$$

and

$$\begin{aligned} &x^K = (K,2), && y^K = (K-1,1), && \text{if } K \text{ is even},\\ &x^K = (K-1,1), && y^K = (K,2) && \text{if } K \text{ is odd}. \end{aligned}$$

Fig. 3 illustrates this for $K = 6,7$.

Set $S = \{x^1,\dots,x^K\}$ and let an alleged bad configuration have $S$ list $A$ and non-$S$ list $B$. All $B$ points must be in $[K]^2 = \{1,\dots,K\}^2$.

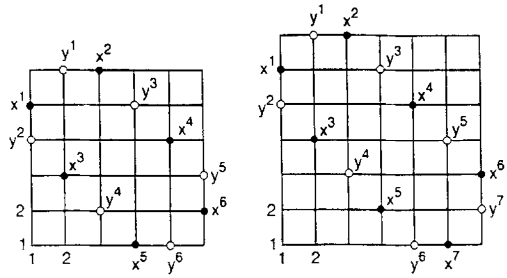


Fig. 3. Bad configurations for $D = \{(1,0),(0,1),(1,1)\}$.

Suppose $x^1$ is in $A$. Then $y^1$ is in $B$ because $y^1$ is the only non-$S$ point in $[K]^2$ on the 45° line through $x^1$, and $x^2$ is in $A$ because it is the only $S$ point on the horizontal line through $y^1$. The possible positions for a $B$ match for $x^2$ on its 45° line are $(2, K-1)$ and $(1, K-2) = y^2$, but the first of these is infeasible because $y^1$ is on the same vertical line and there is only one $S$ point on that line, namely $x^3$, which must be in $A$. The only feasible point for $B$ on the 45° line through $x^3$ is $y^3 = (4, K-1)$ because rows one and three (top down) and column one already have points ($y^1$ and $y^2$) in $B$. Next, $x^4$ is in $A$ because it is the only $S$ point in $y^2$'s row, and $y^4$ is in $B$ because it is the only point in $[K]^2$ on the 45° line through $x^4$ that does not already have a point for $B$ in the same row or column. Continuation leads us to conclude that the bad configuration has every $x^k$ in $A$ and every $y^k$ in $B$. It is evident that the resulting lists comprise a $K$-bad configuration.

Suppose $x^1$ is not in $A$. Then there is no $B$ point in the row, column, or 45° line through $x^1$, and it follows that $x^2$ is not in $A$. Then there is no $B$ point in the row, column, or 45° line through $x^2$, and it follows (consider the 45° line through $x^3$) that $x^3$ is not in $A$. Continuation implies that $A$ is empty, a contradiction.

We conclude that $S$ has no $J$-bad configuration for $J < K$.

**Proof of Theorem 6 for four directions.** Fix $n = 2$ and $D^* = \{(1,0),(0,1),(1,1),(-1,1)\}$. The construction of the preceding proof shows that, for every *even* $K \geqslant 4$, there is a $K$-point $S \subseteq \mathbb{Z}^2$ that has a $K$-bad $D^*$-configuration but no $J$-bad $D^*$-configuration for $J < K$. When $K$ is odd, the preceding construction (see $K = 7$ in Fig. 3) does not produce a bad $D^*$-configuration because of imbalance in direction $(-1,1)$, i.e., on lines of $-45°$ slope.

We defer consideration of $K = 5$ to the next section.

The conclusion of Theorem 6 for odd $K \geqslant 7$ is obtained by splicing a mirror image of our preceding $|S| = 4$ array onto an $|S| = K-3$ array by placing the left-most $y^k$ of the $|S| = 4$ array on top of the right-most $x^k$ of the $|S| = K-3$ array and then deleting those two. Fig. 4 shows the result for $K = 7, 9$. This corrects the $-45°$ imbalance in our preceding odd-$K$ array and yields a $K$-bad $D^*$-configuration. However, $S$ in

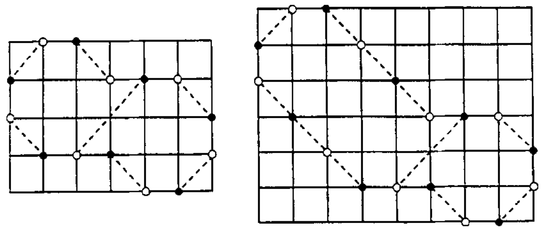


Fig. 4. Bad configurations for $D^* = \{(1,0),(0,1),(1,1),(-1,1)\}$.

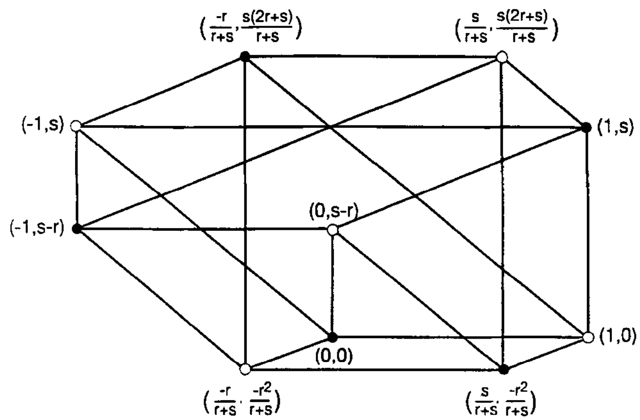


Fig. 5. A 5-bad configuration.

these cases also has a 4-bad $D^*$-configuration, so at least one proper subset is not $D^*$-unique.

## 6. Proof of Theorem 7 for 5-bad planar configurations

Let $n = 2$ and $D = \{(1,0),(0,1),(1,r),(-1,s)\}$ with $r,s > 0$. Suppose $r \neq s$. Fig. 5, patterned on Fig. 2 in [3], displays a 5-bad $D$-configuration with $r < s$. For Theorem 7, we can let $A$ be the solid points and $B$ be the open points, or vice versa. A 5-bad configuration obtains whenever $r \neq s$, but if $r = s$ then the interior points coincide, and their deletion leaves a 4-bad configuration.

We now prove that a 5-bad $D$-configuration is impossible when $r = s$. Assume that $r = s$, let $d^1 = (1,0), d^2 = (0,1), d^3 = (1,r)$, and $d^4 = (-1,r)$, and let $xd^jy$ mean that $x$ and $y$ lie on the same line in direction $d^j$. We will say that geometric conclusions implied by $r = s$ follow "by the geometry". For example, if $x$ and $y$ lie on a rectangle's top edge, $u$ and $v$ lie on the rectangle's right edge, $z$ and $w$ lie on its bottom edge, and $xd^4vd^3z$, $yd^4ud^3w$, and $xd^2w$ (same vertical line), then $yd^2z$ by the geometry.

Suppose, contrary to Theorem 7, that disjoint

$$A = \{a_1, a_2, a_3, a_4, x\} \quad \text{and} \quad B = \{b_1, b_2, b_3, b_4, y\}$$

have equal numbers of points on every line in a $D$ direction. Let $R$ be the smallest rectangle that includes $A \cup B$. The following are assumed with no loss of generality.

(i) $\{a_1, b_1\}, \{a_2, b_2\}, \{a_3, b_3\}$, and $\{a_4, b_4\}$ lie in the left, top, right, and bottom edge of $R$, respectively.

(ii) The bad-configuration matches in $B$ for $x$ in directions $d^3$ and $d^4$ lie below $x$. Then for balance, the bad-configuration matches in $A$ for $y$ in directions $d^3$ and $d^4$ lie above $y$. Horizontal balance then implies that $x$ and $y$ lie in the interior of $R$.

(iii) $a_2$ is to the left of $b_2$ on the top edge of $R$.

Three exclusive possibilities for $x$ versus $y$ are

(I) $xd^3y$: $x$ above $y$ on a line sloping upward to the right;
(II) $xd^4y$: $x$ above $y$ on a line sloping upward to the left;
(III) neither $xd^3y$ nor $xd^4y$.

We consider each of these in turn.

(I) $xd^3y$ implies $xd^2b_4$ or $xd^2b_2$.

Suppose $xd^2b_4$. Then $xd^4b_3, b_2d^2a_4$ and $a_2d^2y$. Because three $B$ points are below $x$ (i.e., $b_4, b_3$ and $y$), we need $a_1$ and $a_3$ below $x$ for $d^1$ matches with $b_3$ and $y$, respectively. Then $xd^1b_1$. We then require $b_1d^3a_2, a_1d^3b_2, a_4d^4b_1$ and $yd^4a_1$. See Fig. 6(Ia) for a slightly warped picture. The last four $d^j$ relationships in conjunction with $a_2d^2y$ and $b_2d^2a_4$ imply $yd^1a_4$ by the geometry. But then $y$ is on the lower edge of $R$, a contradiction.

Suppose $xd^2b_2$. Then $b_4d^2a_2, yd^2a_4, xd^4b_3$, $a_2d^4y$ and, because three $B$ points lie below $x$, $a_1d^1b_3, yd^1a_3$ and $b_1d^1x$. Balance also requires $b_1d^3a_2, a_1d^3b_2, a_1d^4b_4$ and $b_1d^4a_4$. An inaccurate picture appears in Fig. 6(Ib). The last four $d^j$ along with $a_2d^2b_4$ imply $b_2d^2a_4$ by the geometry. This forces $yd^2x$, a contradiction.

(II) $xd^4y$ implies $xd^2b_4$ or $xd^2b_2$.

Suppose $xd^2b_4$. Then $xd^3b_1$ and $xd^1b_3$. Because $b_1, y$ and $b_4$ are below $x$, we have $a_3$ and $a_1$ below $x$. Then $b_3d^4a_2$. But this forces $a_3$ below $b_3$ and, since $b_2$ is to the right of $a_2$, there is no $d^4$ match for $b_2$: see Fig. 6(IIa).

Suppose $xd^2b_2$. Then $a_2d^2b_4$, $a_2d^3b_1$, $a_1d^3b_2$, $xd^3b_4$, $yd^2a_4$, $a_1d^4b_4$, and $b_1d^4a_4$: see Fig. 6(IIb). By the geometry, $b_2d^2a_4$. This forces $yd^2x$, a contradiction.

(III) Neither $xd^3y$ nor $xd^4y$ implies $a_2d^4yd^3a_3$ or $a_1d^4yd^3a_3$. ($a_1d^4yd^3a_2$ forces $a_1d^3b_2$, placing $b_2$ left of $a_2$.)

Suppose $a_2d^4yd^3a_3$. Then $a_2d^2b_4$, $a_3d^1b_1$, $b_1d^3a_2$ ($x$ is to the right of $a_2$ for its vertical match), and $a_1$ is below $b_1$ for $a_1$'s $d^3$ match. Also, $b_2d^4a_3$, $b_4d^3xd^4b_3$, $a_1d^4b_4$, $a_1d^3b_2$, $b_1d^4a_4$ and, by the geometry, $a_4d^2b_2$. In addition, $a_4d^3b_3$ and, by the geometry, $a_1d^1b_3$ : see Figure 6 (IIIa). This requires both $xd^1y$ and $xd^2y$, a contradiction.

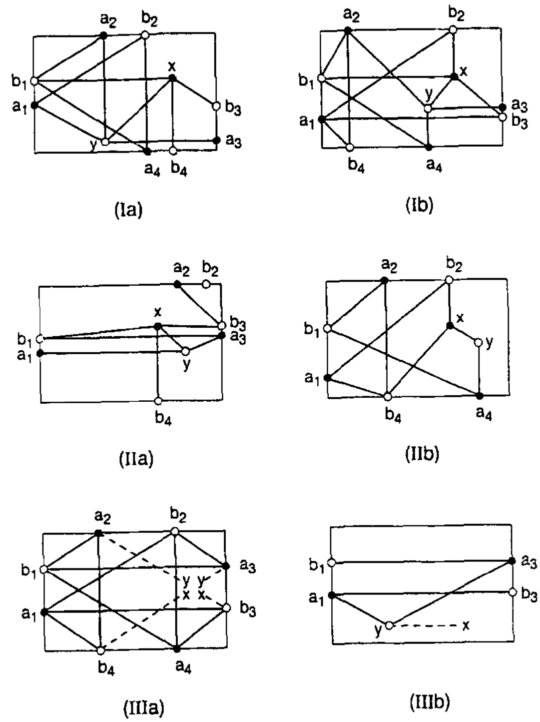


Fig. 6.

Suppose $a_1 d^4 y d^3 a_3$. Then $a_1 d^1 b_3$ and $a_3 d^1 b_1$. These require $x d^1 y$: see Fig. 6(IIIb). Then $b_4$ is the only $B$ point below $x$, contradicting (ii).

Because (I)–(III) yield contradictions, we conclude that there do not exist disjoint five-point sets $A$ and $B$ that have equal numbers of points on every line in a $D$ direction.

## 7. Simulation experiments

In order to test the efficacy and efficiency of the linear programming approach and to help formulate a *fully definite algorithm* based on linear programming for reconstructing a set (or sets) $T$ with $v_T = v_S$ from the given line sums $v_S$, we consider a few "typical" sets in two dimensions which we felt to be representative of real crystals. We refer to these typical sets as *phantoms*. They are pictured in the left half of Figs 7–9, where each "dot" represents a lattice point which is not in the phantom $S$, and each "one" represents a point in the phantom.

For each phantom, we first constructed the set of nonzero line sums $v_S(L)$, $L \in \mathscr{L}$ using directions $(1,0),(0,1),(1,1)$. We then used these line sums as data in the

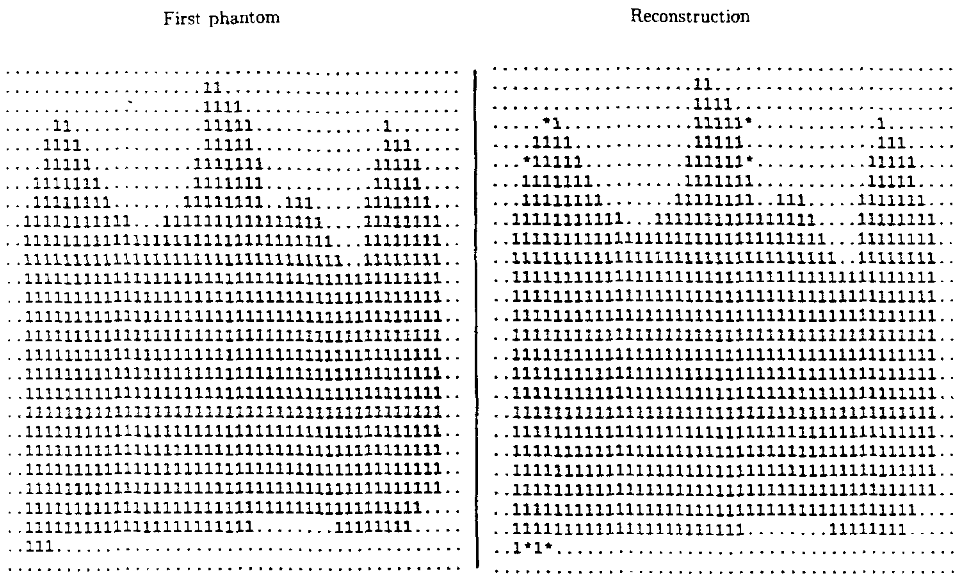


Fig. 7. First phantom and reconstruction.

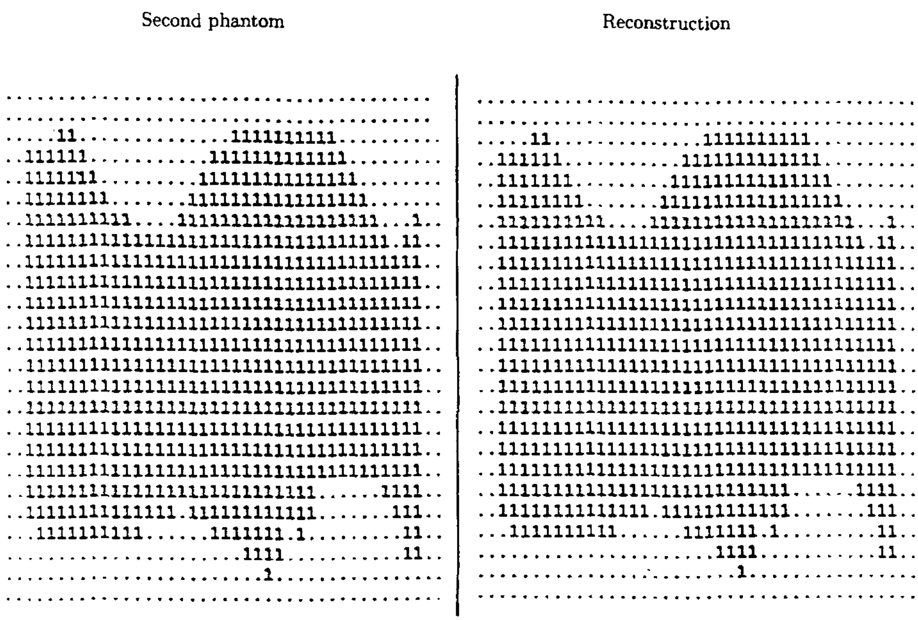


Fig. 8. Second phantom and reconstruction.

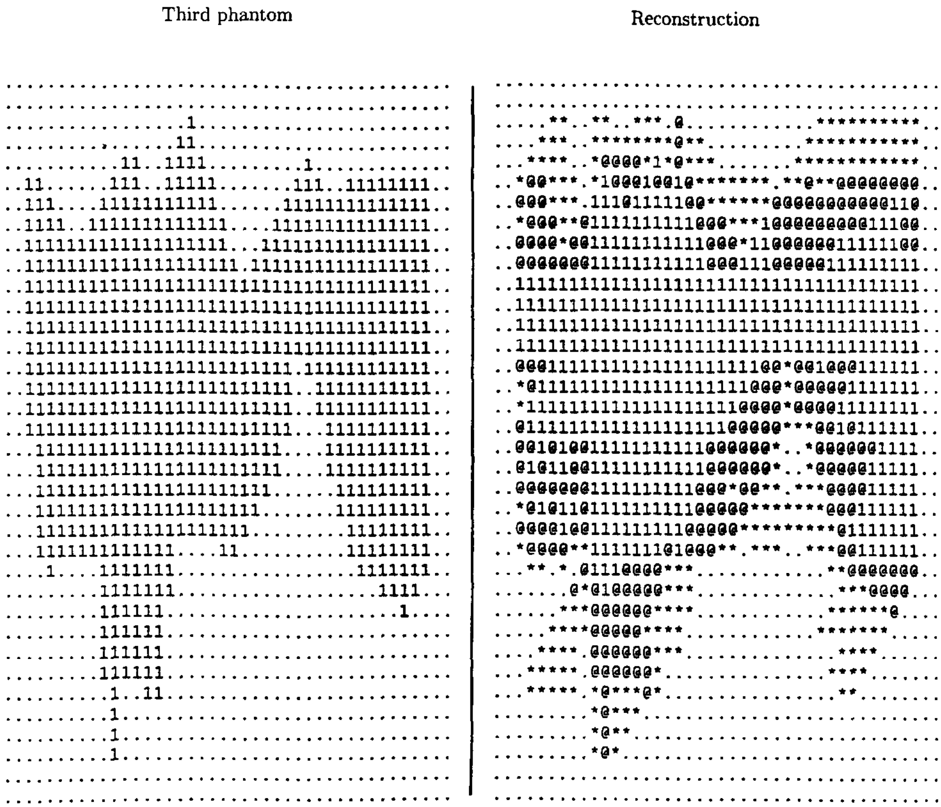


Fig. 9. Third phantom and reconstruction.

following linear program:

$$\begin{aligned} &\text{minimize} \quad 0 \\ &\text{subject to} \quad \sum_{z \in L} f(z) = v_S(L), \quad L \in \mathscr{L}, \qquad (7.1) \\ &\qquad\qquad 0 \leqslant f(z) \leqslant 1, \quad z \in I, \end{aligned}$$

where $I$ denotes a subset of $\mathbb{Z}^2$ which is assumed known to contain the phantom (we used $I = \{-30, -29, \ldots, 29, 30\}^2$).

To solve the linear program, we used an interior point solver called LOQO: see [12]. The reconstructions are shown on the right-hand sides in Fig. 7–9. In the reconstructions, each "dot" represents a lattice point having value 0, a "one" represents a lattice point having value 1, and each "asterisk" represents a lattice point having a value strictly between 0 and 1.

It is well known that interior point methods produce solutions that lie in the center of the face of optimality [1]. This means that any variable which can be away from

```
...........................................
.................1.........................
.....*.............................1.@.....
...........1111.11...........1.............
...@1@....1111111111.........*1*11111*11...
...111....1111111111111....@.1*111111111...
...*111..11111111111.1..1111@@111111111...
...1111.111111111111111.1.11111111111111...
...1111111111111111111.11111111111111111...
...1111111111111111111111111111111111111...
...1111111111111111111111111111111111111...
...1111111111111111111111111111111111111...
...1111111111111111111111111111111111111...
...1111111111111111111111111.11111111111...
....1111111111111111111@@@1111@111111111...
....11111111111111111111111111.111111111...
...11111111111111111111111.**@1111111111...
...111111111111111111111@1*...@1111*1111...
...11111111111111111111111*..*@.11111111...
...111111111111111111@11.....*.111111111...
....1111111111111111*.1@.....1.111111111...
....11111.1111111111111..*....@11111111...
.....1111.1111111111.1..........11111111...
.........11111111................1111111...
.........11111.1*.1................@111....
.......*..1111@1......................1....
..........111*11@..........................
.......@..11111*...........................
....1.....11*1*1................*..........
...........1@@.@...........................
..............1............................
...........1...............................
...........1...............................
...........................................
```

Fig. 10. Another reconstruction for third phantom.

its bounds at optimality will be. In our setting, the variables are the $f(z), z \in I$ and the bounds are 0 and 1. Hence, any lattice point at which $f(z)$ can lie strictly between 0 and 1 will. The fact that most values in the reconstructions are zeros and ones gives us very strong information. Namely, they are lattice points at which every possible reconstruction must take the given extreme value. Note that this tells us that the second phantom, on the left of Fig. 8, has only one reconstruction; i.e., it is fuzzy unique. The first phantom, on the left of Fig. 7, is almost unique. All but six of the lattice points are forced to be at one of the two bounds. A little reflection reveals that these six points represent a 3-bad configuration and so there are precisely two sets having the given line sums: the fuzzy set reconstruction shown in Fig. 7 is simply a convex combination of these two extreme solutions.

The third phantom turns out to be the least unique. In Fig. 9, we have broken those lattice points whose linear program solution lies strictly between zero and one into two cases: an "asterisk" is used to represent values between 0 and 0.5 whereas an "at-sign" is used for values between 0.5 and 1. This example appears to have more than one $K$-bad configuration. To try to get a better understanding of the set of feasible reconstructions, we replaced the zero objective function in the linear program with an objective function in which each coefficient was chosen independently from a Normal mean 0, variance 1 distribution. Fig. 10 shows the reconstruction for this objective function. In this case, the optimal solution will with probability one be an extreme point of the convex set of feasible solutions. It is interesting that this extreme point solution is still a fuzzy solution: it does not represent the indicator of a set. It is also interesting that all of the noninteger values are multiples of $\frac{1}{5}$.

## Acknowledgements

The idea to use linear programming as the basic algorithm for inverting the discrete Radon transform occurred to us only after hearing a presentation at the DIMACS minisymposium on discrete tomography, 19 September, 1994, at Rutgers University. The presenter was Attila Kuba and he was discussing joint work with Ron Aharoni and Gabor Herman. We are grateful to them for presenting their work at the meeting and for their suggestion that linear programming would be very useful in this context, as indeed it is.

We are indebted to Abbas Ourmazd for his continuing interest and many stimulating discussions on applying discrete tomography to lattice images of crystals.

We thank Neal Young for helping us with our NP-completeness questions and guiding us to [7].

We thank Peter Winkler for his amazingly quick (during a talk at the conference on discrete tomography at DIMACS) observation that led to Fig. 4, and Ron Graham for the observation credited in Section 1.

## References

[1] I. Adler and R.D.C. Monteiro, Limiting behavior of the affine scaling continuous trajectories for linear programming, Math. Programming 50 (1991) 29–51.

[2] R. Aharoni, G. Herman and A. Kuba, Binary vectors partially determined by linear equation systems, Discrete Math.

[3] G. Bianchi and M. Longinetti, Reconstructing plane sets from projections, Discrete Comput. Geom. 5 (1990) 223–242.

[4] P.C. Fishburn, Mathematics of Decision Theory, (Mouton, Paris, 1972).

[5] P.C. Fishburn, J.C. Lagarias, J.A. Reeds, and L.A. Shepp, Sets uniquely determined by projections on axes II. Discrete case, Discrete Math. 91 (1991) 149–159.

[6] R. Gordon and G.T. Herman, Reconstruction of pictures from their projections, Commun. ACM 14 (1971) 759–768.

[7] R.W. Irwing and M.R. Jerrum, Three-dimensional statistical data security problems, SIAM J. Comput. 23 (1994) 170–184.

[8] L.A. Shepp and J.B. Kruskal, Computerized tomography: the new medical X-ray technology, Amer. Math. Monthly 85 (1978) 420–439.

[9] P. Schwander, C. Kisielowski, M. Seibt, F.H. Baumann, Y. Kim and A. Ourmazd, Mapping projected potential, interfacial roughness, and composition in general crystalline solids by quantitative transmission electron microscopy, Phys. Rev. Lett. 71 (1993) 4150–4153.

[10] D. Scott and P. Suppes, Foundational aspects of theories of measurement, J. Symbolic Logic 23 (1958) 113–128.

[11] L.A. Shepp and B.F. Logan, The Fourier reconstruction of a head section, IEEE Trans. Nucl. Sci. NS-21 (1974) 21–43.

[12] R.J. Vanderbei, An interior point code for quadratic programming, Princeton University, SOR 94-15 (1994).