Stability and Instability in Discrete Tomography

Andreas Alpers[1], Peter Gritzmann[2⋆], and Lionel Thorens[3]

[1] Zentrum Mathematik, Technische Universität München
Arcisstr. 21, D-80290 München, Germany
alpers@ma.tum.de

[2] Zentrum Mathematik, Technische Universität München
Arcisstr. 21, D-80290 München, Germany
gritzman@ma.tum.de

[3] tfk GmbH, Baierbrunner Str. 39, D-81379 München, Germany
Lionel.Thorens@tfk-gmbh.de

Abstract. The paper gives strong instability results for a basic reconstruction problem of discrete tomography, an area that is particularly motivated by demands from material sciences for the reconstruction of crystalline structures from images produced by quantitative high resolution transmission electron microscopy. In particular, we show that even extremely small changes in the data may lead to entirely different solutions. We will also give some indication of how one can possibly handle the ill-posedness of the reconstruction problem in practice.

1 Introduction

The field of discrete tomography deals with the retrieval of information about discrete objects from typically noisy data. The given data describe (possibly weighted) incidences of the object with query sets. Typical query sets are lines, planes, or various kinds of windows. The field's numerous applications range from scheduling problems, [GGP00], questions of data security, [IJ94], tasks in image processing, [SG82] and many others, [Rys78], [FLRS91], to the question of the reconstruction of crystalline structures from few of their images under high resolution transmission electron microscopy, [SKB+93], [KSB+95], [GdV01]. In the present paper we focus on the question of ill-posedness of the latter *discrete inverse problem*.

The quantitative analysis of high resolution transmission electron microscopic images of a given crystal yields essentially the information how many atoms of the given object interact with sharply focused electron beams in a given viewing direction, [SKB+93], [KSB+95]; see also [GdV01]. So, in principle, we are given the information how many atoms there are on each line parallel to a given small number of directions. To be more precise, let $n \in \mathbb{N}$, $n \geq 2$, let F be a finite subset of $\mathbb{Z}^n$, let S be a line through the origin, and let $\mathcal{A}(S)$ denote the set

⋆ Supported in part by the German Federal Ministry of Education and Research Grant 03-GR7TM1.

G. Bertrand et al. (Eds.): Digital and Image Geometry, LNCS 2243, pp. 175–186, 2001.

of all lines of Euclidean n-space $\mathbb{E}^n$ that are parallel to S. Then the *(discrete) X-ray of F parallel to S* is the function

$$X_S F : \mathcal{A}(S) \to \mathbb{N}_0 = \mathbb{N} \cup \{0\}$$

defined by

$$X_S F(T) = |F \cap T| = \sum_{x \in T} \mathbb{1}_F(x),$$

for each $T \in \mathcal{A}(S)$. In the following let $\mathcal{F}^n = \{F : F \subset \mathbb{Z}^n \wedge F \text{ is finite}\}$ and $\mathcal{L}^n = \big\{\operatorname{lin}\{u\} : u \in \mathbb{Z}^n \setminus \{0\}\big\}$. The elements of $\mathcal{F}^n$ and $\mathcal{L}^n$ are called *lattice sets* and *lattice lines*, respectively. Quite typically, we have some additional a priori information available. This can be modeled by considering suitable subsets $\mathcal{G}$ of $\mathcal{F}^n$.

Given m different lattice lines $S_1, \ldots, S_m$, central questions in discrete tomography are as follows. What kind of information about a finite lattice set F can be retrieved from its X-ray images $X_{S_1}F, \ldots, X_{S_m}F$? How difficult is the reconstruction algorithmically? How sensitive is the task to data errors?

The surveys [Gri97], [GdV01] and the book [HK99] give an overview of the known results dealing with the first two questions and their relatives. (The surveys also include some further explanation for the restriction made here, give examples for relevant classes $\mathcal{G}$ and describe the imaging process in more detail.)

In the present paper we focus on the stability and instability of the reconstruction task. Of course, it is clear that small changes in the data can produce inconsistency. But this is not really a problem. We will, however, prove the following two theorems which show that a small change in the data can lead to a dramatic change in the image, already in the plane. Here the data is given in terms of functions

$$f_i : \mathcal{A}(S_i) \to \mathbb{N}_0, \qquad i = 1, \ldots, m$$

with finite support, hence the difference of two data functions with respect to the same line S is a function $f : \mathcal{A}(S) \to \mathbb{Z}$ whose size is measured in terms of its ℓ_1-norm

$$\|f\|_1 = \sum_{T \in \mathcal{A}(S)} |f(T)|.$$

Theorem 1. *Let $m \in \mathbb{N}$, $m \geq 3$, $S_1, \ldots, S_m \in \mathcal{L}^2$ and $\alpha \in \mathbb{N}$. Then there exist $F_1, F_2 \in \mathcal{F}^2$ with the following properties:*

F_1 is uniquely determined by $X_{S_1}F_1, \ldots, X_{S_m}F_1$;
F_2 is uniquely determined by $X_{S_1}F_2, \ldots, X_{S_m}F_2$;
$\sum_{i=1}^m \|X_{S_i}F_1 - X_{S_i}F_2\|_1 = 2(m-1)$;
$|F_1| = |F_2| \geq \alpha$;
$F_1 \cap F_2 = \emptyset$.

For the case of three directions Theorem 1 shows that even if the X-ray data coincide on all but four single lines the corresponding solutions may still be completely disjoint. The next result shows that this kind of instability persists even if we weaken our measure of the 'distance' from F_1 to F_2 by replacing $|F_1 \triangle F_2|$ by the affinely invariant difference

$$\delta_{\text{aff}}(F_1, F_2) := \min\{|F_1 \triangle h(F_2)|\colon h \text{ is an affine transformation}\}.$$

For practical purposes the special case of translations is particularly interesting. It refers to the situation that we regard an image the same no matter where we look at it. As a generalization of Theorem 1 we show the following instability even for the much weaker affinely invariant difference.

Theorem 2. *Let $m \in \mathbb{N}$ with $m \geq 3$ let $S_1, \ldots, S_m \in \mathcal{L}^2$ be m different lines, and let $\alpha \in \mathbb{N}$. Then there exist $F_1, F_2 \in \mathcal{F}^2$ with the following properties:*

F_1 is uniquely determined by $X_{S_1}F_1, \ldots, X_{S_m}F_1$;
F_2 is uniquely determined by $X_{S_1}F_2, \ldots, X_{S_m}F_2$;
$\sum_{i=1}^{m} \|X_{S_i}F_1 - X_{S_i}F_2\|_1 = 2(m-1)$;
$|F_1| = |F_2| \geq \alpha$;
$\delta_{\mathit{aff}}(F_1, F_2) \geq |F_1|$.

Our two main theorems will be proved in Section 2. Section 3 will contain a rather weak stability result and some remarks towards a possible regularization of the ill-posed discrete inverse problem with a view towards our prime application in semi-conductor industry.

2 Proof of Theorems 1 and 2

We will now prove our main theorems. In [KH98], [KH99] a construction is given showing that whenever $m \geq 3$ and $S_1, \ldots, S_m \in \mathcal{L}^2$ there exist arbitrarily large irreducible switching components. Our construction can be regarded as a modification of a generalization of this result. In the following we are not aiming at smallest possible sets F_1, F_2 but try to give the most transparent construction that covers both, Theorem 1 and Theorem 2 simultaneously.

Throughout this section, let $m \geq 3$ and $S_1, \ldots, S_m \in \mathcal{L}^2$ be m different lattice lines. Let $v_1, \ldots, v_m \in \mathbb{Z}^2$ such that $S_i = \text{lin}\{v_i\}$ for $i = 1, \ldots, m$. We may assume without loss of generality that $v_1 = (1, 0)^T$ and $v_2 = (0, 1)^T$; see [GGP99].

2.1 Some Technical Lemmas on Polynomials

Our construction given in Subsection 2.2 is based on two functionals. Rather than giving it in all possible generality we will use two specific polynomials p and q here, defined by

$$p(t) = t^5 + t^4 + t^3 \qquad \text{and} \qquad q(t) = t^6 + t^5 + t^4.$$

In addition, we use a parameter $\omega \in \mathbb{N}$ that will be fixed later. The functionals are then of the form

$$f_\omega(x) = \omega p(x) \quad \text{and} \quad g_\omega(x) = q(x/\omega).$$

Let us begin with a simple remark.

Remark 3. *For any* $\omega \in \mathbb{N}$,

(a) $f_\omega(\mathbb{N}), g_\omega(\omega \cdot \mathbb{N}) \subset \mathbb{N}$;
(b) f_ω *and* g_ω *are strictly increasing on* $[0, \infty[$ *whence invertible;*
(c) $f_\omega(t) > g_\omega^{-1}(t)$ *for every* $t \in]1, \infty[$.

Next we prove auxilliary results on certain functional equations involving f_ω and g_ω.

Lemma 1. *Let* $\alpha, \beta, \gamma, \delta, \sigma, \tau \in \mathbb{R}$ *such that* $\alpha\delta - \beta\gamma \neq 0$.

(a) Then the equation

$$\alpha t + \beta f_\omega(t) + \sigma = g_\omega\big(\gamma t + \delta f_\omega(t) + \tau\big)$$

holds for at most 30 *different values of* t.

(b) The equation

$$f_\omega\big(\alpha t + \beta f_\omega(t) + \sigma\big) = \gamma t + \delta f_\omega(t) + \tau$$

holds for all $t \in \mathbb{R}$ *if and only if* $\alpha = \delta = 1$ *and* $\beta = \gamma = \sigma = \tau = 0$. *Otherwise it holds for at most* 25 *different values of* t.

(c) The equation

$$\alpha g_\omega(t) + \beta t + \sigma = g_\omega\big(\gamma g_\omega(t) + \delta t + \tau\big)$$

holds for all $t \in \mathbb{R}$ *if and only if* $\alpha = \delta = 1$ *and* $\beta = \gamma = \sigma = \tau = 0$. *Otherwise it holds for at most* 36 *different values of* t.

Proof. To prove assertion (a) just note that $\deg(g_\omega) = 6 > \deg(f_\omega) = 5 \geq 2$ implies that the equation cannot hold as an identity. Hence there are at most $\deg(f_\omega) \cdot \deg(g_\omega)$ many solutions.

Similarly, if (b) holds identically, $\beta = 0$, and we have the equation

$$f_\omega\big(\alpha t + \sigma\big) = \gamma t + \delta f_\omega(t) + \tau.$$

Taking the k th derivative on both sides for $k = 2, 3, 4, 5$ yields the condition

$$\alpha^k f_\omega^{(k)}\big(\alpha t + \sigma\big) = \delta f_\omega^{(k)}(t).$$

The equation for $k = 5$ yields $\alpha^5 = \delta$, which implies $\delta = 1$ if $\alpha = 1$. Suppose that $\alpha \neq 1$. By choosing $t_0 = \sigma/(1-\alpha) \neq 1/5$ we obtain $\alpha^5 = \alpha^4 = \delta$. For $t_0 = 1/5$ it follows that $f_\omega^{(4)}(t_0) = 0$, but $f_\omega^{(2)}(t_0) \neq 0$. Thus, $\alpha^5 = \alpha^2 = \delta$. In any case, this implies $\alpha = \delta = 1$. Taking the fourth derivative shows now that $\sigma = 0$. Plugging $t = 0$ into the original equation yields $\tau = 0$, and finally $\gamma = 0$.

Of course, (c) follows similarly. □

2.2 The Basic Construction

Now let $k \in \mathbb{N}$ such that $2^{m-2} \cdot k > \alpha$, where α is the positive integer of Theorem 1 or Theorem 2. For the proof of Theorem 2 we may assume that $\alpha \geq 2^{2(m+2)}$. We begin by constructing two lattice sets U_m and V_m with $|U_m| = |V_m| = 2^{m-2} \cdot k > \alpha$ that are *tomographically equivalent* with respect to $S_1, \dots, S_m$ i.e., they have the same X-rays parallel to $S_1, \dots, S_m$.

Let $\omega \geq 2$, set

$$\lambda_1 = g_\omega\big(f_\omega(2)\big),$$

and define

$$\lambda_{i+1} = g_\omega\big(f_\omega(\lambda_i)\big) \qquad \text{for } i = 2, \dots, k-1.$$

Remark 4. *(a) The sequence $(\lambda_i)_{i=1,\dots,k}$ is independent of the specific choice of ω;*
(b) $\lambda_1, \dots, \lambda_k \in \mathbb{N}$
(c) $\lambda_1 < \cdots < \lambda_k$.

Now let

$$B_{k-1} = \left\{ \begin{pmatrix} \lambda_i \\ f_\omega(\lambda_i) \end{pmatrix} \;:\; i = 1, \dots, k-1 \right\}, \qquad B_k = B_{k-1} \cup \left\{ \begin{pmatrix} \lambda_k \\ f_\omega(\lambda_k) \end{pmatrix} \right\}$$

and

$$C_{k-1} = \left\{ \begin{pmatrix} \lambda_{i+1} \\ f_\omega(\lambda_i) \end{pmatrix} \;:\; i = 1, \dots, k-1 \right\}, \qquad C_k = C_{k-1} \cup \left\{ \begin{pmatrix} \lambda_1 \\ f_\omega(\lambda_k) \end{pmatrix} \right\}.$$

Observe that the points of C_{k-1} are all of the form $(t, g_\omega^{-1}(t))^T$.

The two sets B_k and C_k are tomographically equivalent with respect to S_1 and S_2. The rest of the construction will depend on suitably chosen positive integers $\theta_3, \dots, \theta_m \in \mathbb{N}$ which will be fixed later. For each additional line in $\mathcal{L}^2$, the sizes of the sets are doubled. For $j = 3, \dots, m$ we define

$$\begin{aligned} U_2 &= B_k & V_2 &= C_k \\ U_j &= U_{j-1} \cup (V_{j-1} + \theta_j v_j) \qquad & V_j &= V_{j-1} \cup (U_{j-1} + \theta_j v_j). \end{aligned}$$

Clearly, we have the following properties.

Remark 5. *For each ω the integers $\theta_3, \dots, \theta_m$ can be chosen suitably large, so that $U_m \cap V_m = \emptyset$, $|V_m| = |U_m| = 2^{m-2} \cdot k$, and U_m and V_m are tomographically equivalent with respect to $S_1, \dots, S_m$.*

The required sets F_1 and F_2 are now constructed by removing a point of U_m and V_m each. In fact, we choose points $z_0 \in U_m$ and $z_1 \in V_m$ that lie on the same line parallel to S_1 and set

$$F_1 = U_m \setminus \{z_0\}, \qquad F_2 = V_m \setminus \{z_1\}.$$

Note that the construction shows that

$$F_1 \cap F_2 = \emptyset \quad \text{and} \quad |F_1| = |F_2| = 2^{m-2} \cdot k - 1.$$

Clearly,

$$\sum_{i=1}^{m} ||X_{S_i} F_1 - X_{S_i} F_2||_1 = 2(m-1).$$

Our theorems will follow if we can show that the integers ω and $\theta_3, \ldots, \theta_m$ can be chosen in such a way that no other finite lattice set is tomographically equivalent to U_m and V_m with respect to $S_1, \ldots, S_m$, and U_m and V_m are such that even affine transformations cannot place them 'too closely' on top of each other.

In fact, from the former property which will be derived in Lemma 3 it follows that F_1 is uniquely determined by its X-rays parallel to $S_1, \ldots, S_m$. Indeed, if $F \in \mathcal{F}^2$ is tomographically equivalent to F_1, then $z_0 \notin F$, and $F \cup \{z_0\}$ is tomographically equivalent to U_m. Since $F \cup \{z_0\} \neq V_m$ we must have $F \cup \{z_0\} = U_m$ hence $F_1 = F$. Analogously, F_2 is uniquely determined by $X_{S_1} F_2, \ldots, X_{S_m} F_2$.

2.3 Properties of U_m and V_m

Now we derive some crucial properties of U_m and V_m for appropriate choices of the parameters. First note that every set that is tomographically equivalent to U_j or, what is the same, to V_j is contained in the *grid* of U_j

$$G_j = \mathbb{Z}^2 \cap \bigcap_{i=1}^{j} \bigcup_{x \in U_j} (x + S_i),$$

for $j = 2, \ldots, m$. Of course,

$$G_2 = \{(\lambda_i, f_\omega(\lambda_j))^T \, : \, i, j = 1, \ldots, k\}.$$

Lemma 2. *The integer ω can be chosen in such a way that there is no line parallel to $S_3, \ldots, S_m$ which intersects G_2 in more than one point.*

Proof. Suppose there are indices $i_1, i_2, j_1, j_2 \in \{1, \ldots, k\}$ with $i_1 \neq i_2$, $j_1 \neq j_2$, and $v \in \{v_3, \ldots, v_m\}$, and there exists a real μ such that

$$\begin{pmatrix} \lambda_{i_1} \\ f_\omega(\lambda_{j_1}) \end{pmatrix} - \begin{pmatrix} \lambda_{i_2} \\ f_\omega(\lambda_{j_2}) \end{pmatrix} = \mu v.$$

By Remark 4 (a) this equation is equivalent to

$$\begin{pmatrix} \lambda_{i_1} - \lambda_{i_2} & 0 & -\nu_1 \\ 0 & p(\lambda_{j_1}) - p(\lambda_{j_2}) & -\nu_2 \end{pmatrix} \begin{pmatrix} 1 \\ \omega \\ \mu \end{pmatrix} = 0,$$

where $v = (\nu_1, \nu_2)^T$. The rank of the 2×3 matrix is 2, hence its nullspace is of dimension 1. On the other hand, for any two different numbers ω_1, ω_2 and arbitrary μ_1, μ_2, the vectors $(1, \omega_1, \mu_1)^T$ and $(1, \omega_2, \mu_2)^T$ are always linearly independent, whence two such vectors cannot be in the nullspace simultaneously. Taking all possible quadruples of indices $i_1, i_2, j_1, j_2 \in \{1, \ldots, k\}$ leads then to the union of $k^2(k-1)^2$ one-dimensional subspaces that we have to avoid. Hence among any set of $k^2(k-1)^2+1$ different values of ω there must be one for which none of the vectors $(1, \omega, \mu)^T$ lie in any of the nullspaces. □

Note that the result of Lemma 2 is invariant with respect to translations of G_2. In the following we assume that ω is fixed according to Lemma 2. The next lemma implies Theorem 1 already.

Lemma 3. *The integers $\theta_3, \ldots, \theta_m$ can be chosen large enough such that no other finite lattice set is tomographically equivalent to U_m and V_m with respect to $S_1, \ldots, S_m$.*

Proof. Since G_m is the grid of U_m and V_m we clearly have $U_m \cup V_m \subset G_m$. Setting $G^{(2)} = G_2$, it is also clear that we can choose $\theta_3, \ldots, \theta_m$ large enough such that

$$G_i \subset G^{(i)} = G^{(i-1)} \cup (\theta_i v_i + G^{(i-1)})$$

for $i = 3, \ldots, m$. For a formal proof see [GGP99, Proof of Lemma 3.1].

The assertion will now follow if we make sure that $G_m = U_m \cup V_m$ since the only two subsets of $U_m \cup V_m$ that are tomographically equivalent to U_m or V_m are precisely these two sets. We will show that for each point $g \in G^{(m)} \setminus (U_m \cup V_m)$ the line $g + S_3$ does not intersect $U_m \cup V_m$. So, let $g \in G^{(m)} \setminus (U_m \cup V_m)$ and suppose $g + S_3$ intersects $U_m \cup V_m$. Of course, $U_m \cup V_m$ consists of disjoint translates of $U_2 \cup V_2$, with translation vectors of the form $\sum_{i=3}^{m} \delta_i \theta_i v_i$ with $\delta_1, \ldots, \delta_m \in \{0, 1\}$. Let t_1 be such a translation vector with $g \in t_1 + G_2$. By our assumption there is another such translation vector t_2 with $(g + S_3) \cap (t_2 + G_2) \neq \emptyset$. By Lemma 2 no line parallel to S_3 intersects $t_1 + G_2$ and $t_2 + G_2$ in more than one point each. Hence these points must be g and $(t_2 - t_1) + g$. But $g \notin t_1 + (U_2 \cup V_2)$ hence $(t_2 - t_1) + g \notin t_2 + (U_2 \cup V_2)$. This contradiction concludes the proof of the assertion. □

2.4 The Affinely Invariant Difference

In order to prove Theorem 2 we need to study the power of affine transformations on F_2 to reduce the symmetric difference to F_1. The next result is a corollary to Lemma 1.

Lemma 4. *Let $A \in \mathbb{R}^{2 \times 2}$ be nonsingular, let $a \in \mathbb{R}^2$, and let h be the affine transformation defined by $h(x) = Ax + a$. Then*

(a) $|B_k \cap h(C_k)| \leq 31$;
(b) $|B_k \cap h(B_k)| \geq 26 \Rightarrow h \equiv id$;
(c) $|C_k \cap h(C_k)| \geq 39 \Rightarrow h \equiv id$.

Proof. Since the one point in $C_k \setminus C_{k-1}$ is of slightly different form than the points of $B_k \cup C_{k-1}$ we always count it as 'matched'. Then the assertion follows easily from Lemma 1. □

Lemma 4 shows already that the 'overlay' power of affine transformations is rather limited. In fact, suppose that A is not the unit matrix i.e., h is not just a translation. If we use again that $U_m \cup V_m$ consists of disjoint translates of $U_2 \cup V_2$ with translation vectors of the form $\sum_{i=3}^{m} \delta_i \theta_i v_i$ with $\delta_1, \ldots, \delta_m \in \{0,1\}$, and of course that $U_2 = B_k$ and $V_2 = C_k$ then Lemma 4 implies that

$$\begin{aligned}
&|h(U_m \cup V_m) \cap (U_m \cup V_m)| \\
&\quad \leq 2^{m-2} \cdot 2^{m-2} \cdot \Big(|h(U_2) \cap U_2| + |h(U_2) \cap V_2| + |h(V_2) \cap U_2| + |h(V_2) \cap V_2| \Big) \\
&\quad \leq 125 \cdot 2^{2m-4} \leq 125 \cdot \alpha \cdot 2^{-8} \leq \tfrac{1}{2}\alpha \leq \tfrac{1}{2}|F_1|.
\end{aligned}$$

So it remains to study the effect of translations. To do this appropriately we require one more property of the numbers $\theta_3, \ldots, \theta_m$. It is most easily stated with the help of a geometric functional, the *breadth* $b_v(S)$ of a planar point set S in a given direction $v \in \mathbb{R}^2$, $\|v\|_2 = 1$; it is defined by

$$b_v(S) = \max_{x \in S} v^T x - \min_{x \in S} v^T x.$$

We choose the integers $\theta_3, \ldots, \theta_m$ so that

$$\|\theta_i v_i\|_2 > b_{\hat{v}_i}(U_{i-1} \cup V_{i-1}) \qquad \text{for } i = 3, \ldots, m,$$

where $\hat{v}_i = v_i / \|v_i\|_2$. Of course, this choice does not interfere with the previous requirements.

Now let t be a translation vector. Suppose first that

$$v_m^T t \geq b_{\hat{v}_m}(U_{m-1} \cup V_{m-1}).$$

This means on the one hand that t has the capability of moving a point of $U_{m-1} \cup V_{m-1}$ into $\theta_m v_m + \big(U_{m-1} \cup V_{m-1}\big)$. But on the other hand

$$\big(U_{m-1} \cup V_{m-1}\big) \cap \Big(t + \big(U_{m-1} \cup V_{m-1}\big)\Big) = \emptyset.$$

Hence

$$\Big|\big(U_m \cup V_m\big) \bigtriangleup \Big(t + \big(U_m \cup V_m\big)\Big)\Big| \geq |F_1|,$$

which is the assertion. So we can assume that

$$\begin{aligned}
&\big(U_m \cup V_m\big) \bigtriangleup \Big(t + \big(U_m \cup V_m\big)\Big) \\
&\quad = \bigg(\big(U_{m-1} \cup V_{m-1}\big) \cap \Big(t + \big(U_{m-1} \cup V_{m-1}\big)\Big) \bigg) \cup \\
&\qquad \cup \bigg(\Big(\theta_m v_m + \big(U_{m-1} \cup V_{m-1}\big)\Big) \cap \Big(t + \theta_m v_m + \big(U_{m-1} \cup V_{m-1}\big)\Big) \bigg).
\end{aligned}$$

Hence by an induction argument we can reduce the assertion to $B_k \cup C_k$, and it suffices to show that $|B_k \cap (t + C_k)| \leq \frac{1}{2}|B_k|$. This follows from Lemma 4, which concludes the proof of Theorem 2.

3 Dealing with Instability in Practice

The approximation algorithms given in [GdVW00] show that the 'combinatorial optimization part' of the reconstruction problem of discrete tomography can be handled very efficiently in spite of its computational complexity. In particular, [GdVW00] considered the problem Best-Inner-Fit$(S_1, \ldots, S_m)$ [BIF] that accepts as input data functions $f_1, \ldots, f_m$, and its task is to find a set $F \in \mathcal{F}^n$ of maximal cardinality such that

$$X_{S_i} F(T) \leq f_i(T) \qquad \text{for all } T \in \mathcal{A}(S) \text{ and } i = 1, \ldots, m.$$

The results show that in spite of the underlying $\mathbb{NP}$-hardness some simple heuristics already yield good a priori approximation guarantees and behave surprisingly well in practice. In fact a dynamic greedy/postprocessing strategy actually produced solutions with small *absolute error*; for instances with 250000 variables and point density of 50% an average of +only 1.07, 23.28, 64.58 atoms were missing for 3, 4, 5 directions, respectively.

As we have seen the reconstruction task of discrete tomography is intrinsically instable. There is however the following (rather weak) stability result.

Lemma 5. *Let $F_1 \in \mathcal{F}^n$ and let $f_1, \ldots, f_m$ be given (generally noisy) data functions. Then there exists a finite lattice set F_2 with $X_{S_i} F_2(T) \leq f_i(T)$ for all $T \in \mathcal{A}(S)$ and $i = 1, \ldots, m$ such that*

$$|F_1 \triangle F_2| \leq \sum_{i=1}^{m} \|f_i - X_{S_i} F_1\|_1.$$

Proof. F_2 is constructed from F_1 by deleting at most $\sum_{i=1}^{m} \|f_i - X_{S_i} F_1\|_1$ points in order to satisfy the constraints $X_{S_i} F_2(T) \leq f_i(T)$ for all $T \in \mathcal{A}(S)$ and $i = 1, \ldots, m$. □

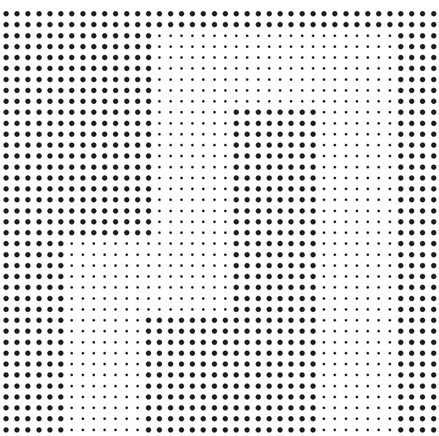
Lemma 5 shows in particular that under data errors the cardinality of a solution behaves in a stable way. Also it shows that even with noisy data there is always a solution of [BIF] that is 'close' to the original object in terms of the symmetric difference. (Of course, it is not even clear how to recognize a 'good' solution if the original image is not known, let alone how to find one.) In view of Theorem 1 this means that it is generally not a good strategy in practice to try to satisfy the measured constraints as well as possible. In fact, Theorem 1 shows that the 'perfect solution' for the noisy data may be disjoint from the original image while by Lemma 5 a 'nonperfect solution' may be quite close.

In any case, it is important to utilize any additional physical knowledge and experience available in order to actually produce solutions that are close to the original physical objects. For the task of quality control in semi-conductor industry we can, e.g., utilize the fact that we know the 'perfect' object in advance. Suppose we are given a few images under high resolution transmission electron microscopy of parts of a silicon wafer that carries an etched pattern of some electrical circuit. Our instability results seem to suggest that there is not much hope for a reasonably good reconstruction of the object from these measurements since all physical measurements are noisy. On the other hand, we do not need to reconstruct the object 'from scratch'. We know what the object would look like if the production process had been exact and the measurements were correct. Hence it seems reasonable to measure the distance from the theoretical perfect chip. This means we are trying to reconstruct a finite lattice set approximately satisfying the (noisy) X-ray measurements so that its distance from the perfect template is as small as possible. If there does not exist a good enough approximation, we know that the chip is faulty. But if there exists a good enough approximation, all we know is that the data from high resolution transmission electron microscopy is not sufficient to rule out the possibility that the chip is correct. Let us look at the following simple example of two images F_1 and F_2 depicted in Figure 1.

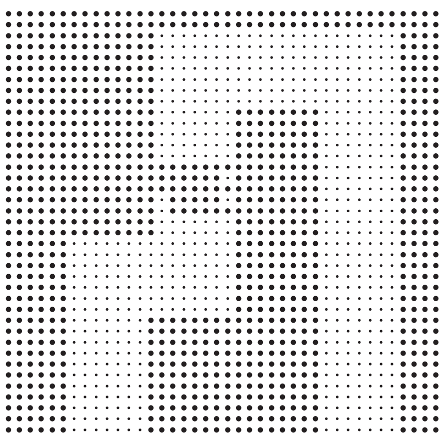
The corresponding grid G has the size 40×40; whence the reconstruction problem involves 1600 0-1-variables. The sets F_1 and F_2 are depicted as fat black points ($\bullet$); the empty places are represented as small dots ($\cdot$). Template F_1 is regarded as proper, while template F_2 is faulty since the conductor path is blocked. Now we consider X-rays parallel to the lines $S_1 = \operatorname{lin}\{(1,0)^T\}$, $S_2 = \operatorname{lin}\{(0,1)^T\}$ and $S_3 = \operatorname{lin}\{(1,1)^T\}$. Note that F_1 and F_2 are uniquely determined by their X-rays parallel to these lines.

However, as shown in Figure 2 there are finite lattice sets F_1' and F_2' with

$$\begin{array}{rl} & -1 \leq X_{S_i}F_1(T) - X_{S_i}F_1'(T) \leq 1 \quad \text{for each} \quad T \in \mathcal{A}(S_i),\ i = 1,2,3, \\ \text{and} & -1 \leq X_{S_i}F_2(T) - X_{S_i}F_2'(T) \leq 1 \quad \text{for each} \quad T \in \mathcal{A}(S_i),\ i = 1,2,3 \end{array}$$

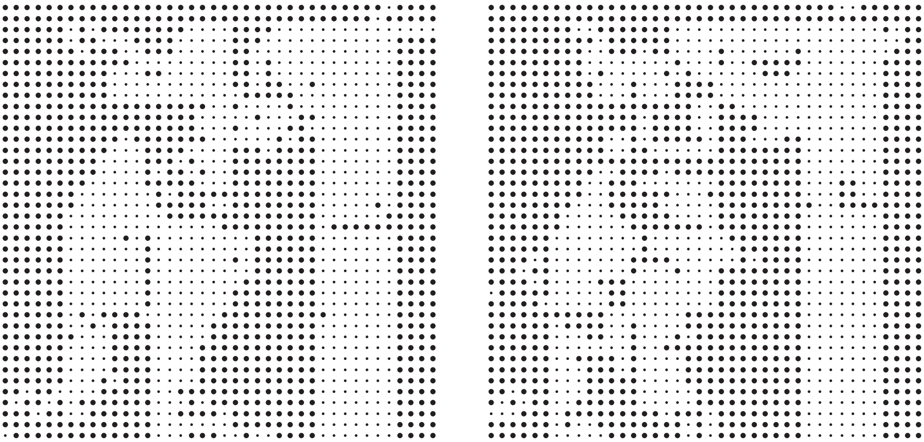


(a) phantom F_1



(b) phantom F_2

Fig. 1. Two lattice sets in a 40×40 grid



(a) F_1' (b) F_2'

Fig. 2. Reconstructions of F_1 and F_2, respectively, based on data with error at most one on each line

that do not share the same 'conductor path structure'. Also, F_1' and F_2' look similar and it does not seem justified to distinguish a 'proper chip' from a 'faulty' one based on this reconstruction. But after all, the goal is only to see whether there is a chance to detect faulty chips even in the presence of considerable data errors. And in fact, if the error on each X-ray line is reasonably small such information can be derived. Let $\epsilon \in \mathbb{N}$, and suppose that for $j = 1, 2$ the data functions f_1^j, f_2^j, f_3^j are measured with

$$-\epsilon \leq X_{S_i} F_j(T) - f_i^j \leq \epsilon \qquad \text{for each } T \in \mathcal{A}(S_i),\ i = 1, 2, 3,\ j = 1, 2.$$

We pursue an approach motivated by Lemma 5, choose some $\delta \in \mathbb{N}$ and determine two sets F^1 and F^2 that minimize the symmetric difference to F_1 under the constraints that

$$-\delta \leq X_{S_i} F^j(T) - f_i^j \leq \delta \qquad \text{for each } T \in \mathcal{A}(S_i),\ i = 1, 2, 3,\ j = 1, 2.$$

Clearly, if $\delta \geq \epsilon$, the set F^1 coincides with the template F_1. Further, if

$$\delta + \epsilon < \max_{i=1,2,3} \|X_{S_i} F_1 - X_{S_i} F_2\|_\infty,$$

then $F^2 \neq F_1$. In our specific example this is for instance the case when $\delta = \epsilon = 3$. The question is whether this trivial observation which is only based on the 'local' information of what happens on single X-ray lines can be substantially extended when all the given information is taken into account and if the error is regarded as random variable with some given underlying distribution. Also, what can be said for real-world data? These questions will require further theoretical investigations and an extended experimental study for the above and other conceivable regularization techniques.

References

[FLRS91] 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(2):149–159, 1991. 175

[GdV01] P. Gritzmann and S. de Vries. Reconstructing crystalline structures from few images under high resolution transmission electron microscopy. In *Mathematics: Key Technology for the Future, (ed. by W. Jäger)*. Springer, 2001. in print. 175, 176

[GdVW00] P. Gritzmann, S. de Vries, and M. Wiegelmann. Approximating binary images from discrete X-rays. *SIAM J. Opt.*, 11:522–546, 2000. 183

[GGP99] R. J. Gardner, P. Gritzmann, and D. Prangenberg. On the computational complexity of reconstructing lattice sets from their X-rays. *Discrete Math.*, 202:45–71, 1999. 177, 181

[GGP00] R. J. Gardner, P. Gritzmann, and D. Prangenberg. On the computational complexity of determining polyatomic structures by X-rays. *Th. Comput. Sci.*, 233:91–106, 2000. 175

[Gri97] P. Gritzmann. On the reconstruction of finite lattice sets from their X-rays. In *Discrete Geometry for Computer Imagery, (ed. by E. Ahronovitz and C. Fiorio)*, pages 19–32. Springer, Berlin, 1997. 176

[HK99] G. T. Herman and A. Kuba (eds.). *Discrete Tomography: Foundations, Algorithms, and Applications*. Birkhäuser, Boston, 1999. 176

[IJ94] R. W. Irving and M. R. Jerrum. Three-dimensional statistical data security problems. *SIAM J. Comput.*, 23:170–184, 1994. 175

[KH98] T. Y. Kong and G. T. Herman. On which grids can tomographic equivalence of binary pictures be characterized in terms of elementary switching operations. *Int. J. Imaging Syst. Technol.*, 9:118–125, 1998. 177

[KH99] T. Y. Kong and G. T. Herman. Tomographic equivalence and switching operations. In *Discrete tomography: foundations, algorithms, and applications (ed. by G. Herman and A. Kuba)*, pages 59–84. Birkhäuser, Boston, 1999. 177

[KSB+95] C. Kisielowski, P. Schwander, F. H. Baumann, M. Seibt, Y. Kim, and A. Ourmazd. An approach to quantitative high-resolution transmission electron microscopy of crystalline materials. *Ultramicroscopy*, 58:131–155, 1995. 175

[Rys78] H. J. Ryser. *Combinatorial Mathematics*, chapter 6, Matrices of zeros and ones. Mathematical Association of America and Quinn & Boden, Rahway, NJ, 1963, pp. 61–78. 175

[SG82] C. H. Slump and J. J. Gerbrands. A network flow approach to reconstruction of the left ventricle from two projections. *Comput. Graphics Image Process.*, 18:18–36, 1982. 175

[SKB+93] P. Schwander, C. Kisielowski, 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(25):4150–4153, 1993. 175