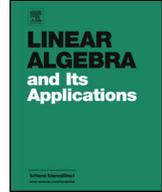




ELSEVIER

Contents lists available at [SciVerse ScienceDirect](http://SciVerse.ScienceDirect.com)

# Linear Algebra and its Applications

[www.elsevier.com/locate/laa](http://www.elsevier.com/locate/laa)


## Average case recovery analysis of tomographic compressive sensing



Stefania Petra\*, Christoph Schnörr

*Image and Pattern Analysis Group, University of Heidelberg, Speyerer Str. 6, 69115 Heidelberg, Germany*

### ARTICLE INFO

#### Article history:

Received 13 August 2012

Accepted 26 June 2013

Available online 2 August 2013

Submitted by V. Mehrmann

#### MSC:

65F22

68U10

#### Keywords:

Compressed sensing

Underdetermined systems of linear equations

Sparsity

Large deviation

Tail bound

Nonnegative least squares

Algebraic reconstruction

TomoPIV

### ABSTRACT

The reconstruction of three-dimensional sparse volume functions from few tomographic projections constitutes a challenging problem in image reconstruction and turns out to be a particular problem instance of compressive sensing. The tomographic measurement matrix encodes the incidence relation of the imaging process, and therefore is not subject to design up to small perturbations of non-zero entries. We present an average case analysis of the recovery properties and a corresponding tail bound to establish weak thresholds in excellent agreement with numerical experiments. Our results improve the state-of-the-art of tomographic imaging in experimental fluid dynamics.

© 2013 Elsevier Inc. All rights reserved.

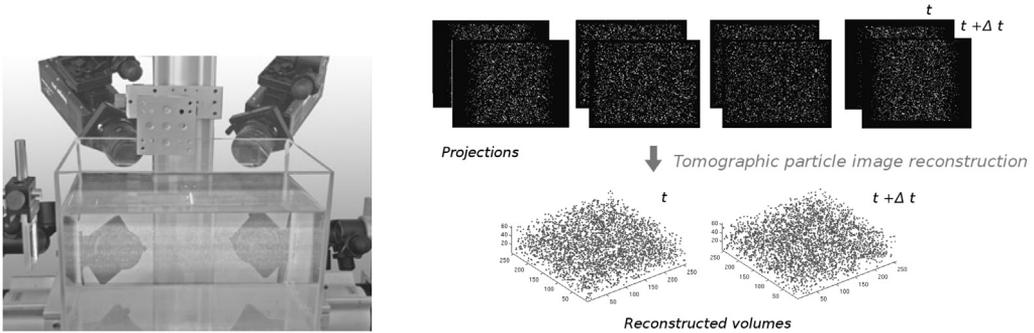
## 1. Introduction

Research on compressive sensing [8,3] focuses on properties of underdetermined linear systems

$$Ax = b, \quad A \in \mathbb{R}^{m \times n}, \quad m \ll n, \quad (1.1)$$

\* Corresponding author.

E-mail addresses: [petra@math.uni-heidelberg.de](mailto:petra@math.uni-heidelberg.de) (S. Petra), [schnoerr@math.uni-heidelberg.de](mailto:schnoerr@math.uni-heidelberg.de) (C. Schnörr).



**Fig. 1.** Compressive sensing in experimental fluid dynamics: A multi-camera setup gathers few projections from a sparse volume function. This scenario is described by a very large and highly underdetermined sparse linear system (1.1) having the additional properties (1.2). The sparsity parameter  $k$  reflects the seeding density of a given fluid with particles. Less sparse scenarios increase the spatial resolution of subsequent studies of turbulent motions, but compromise accuracy of the reconstruction. Research is concerned with working out and mathematically substantiating the best compromise.

that ensure the accurate recovery of sparse solutions  $x$  from observed measurements  $b$ . Strong assertions are based on random ensembles of measurement matrices  $A$  and measure concentration in high dimensions that allow proving good recovery properties with high probability [9,4].

A common obstacle in various application fields is the limited number of options for *designing* a measurement matrix so as to exhibit desirable mathematical properties. Accordingly, recent research has also been concerned with more restricted scenarios, spurred by their relevancy to applications (cf. Section 2.3).

Consequently, we consider a representative scenario, motivated by TomoPIV (*Tomographic Particle Image Velocimetry*) [12], an application in experimental fluid dynamics (Fig. 1). A suitable mathematical abstraction of this setup gives rise to a huge and severely underdetermined linear system (1.1) that has additional properties: a *very sparse* nonnegative measurement matrix  $A$  with *constant small support* of all column vectors, and a nonnegative sparse solution vector  $x$ :

$$A \geq 0, x \geq 0, \quad \text{supp}(A_{\bullet,j}) = \ell \ll m, \quad \forall j = 1, \dots, n. \tag{1.2}$$

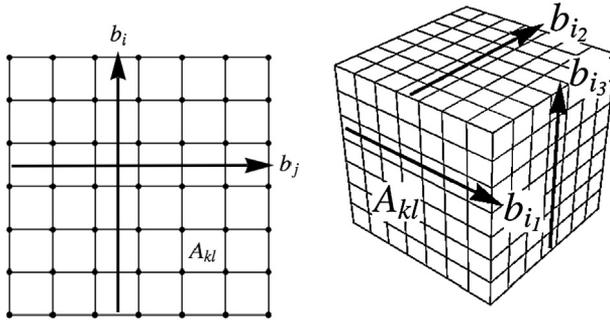
Our objective is the usual one: relating accurate recovery of  $x$  from given measurements  $b$  to the sparsity  $k = \text{supp}(x)$  of the solution  $x$  and to the dimensions  $m, n$  of the measurement matrix  $A$ . The sparsity parameter  $k$  has an immediate physical interpretation (Fig. 1). Engineers require high values of  $k$ , but are well aware that too high values lead to spurious solutions. The current practice is based on a rule of thumb leading to conservative low values of  $k$ .

In this paper, we are concerned with working out a better compromise along with a mathematical underpinning. The techniques employed are general and only specific to the class of linear systems (1.1), (1.2), rather than to a particular application domain.

We regard the measurement matrix  $A$  as *given*. Concerning the design of  $A$ , we can only resort to small random perturbations of the non-zero entries of  $A$ , thus preserving the sparse structure that encodes the underlying incidence relation of the sensor. Additionally, we exploit the fact that solution vectors  $x$  can be regarded as samples from a uniform distribution over  $k$ -sparse vectors, which represents with sufficient accuracy the underlying physical situation.

Under these assumptions, we focus on an *average case analysis* of conditions under which *unique recovery* of  $x$  can be expected with *high probability*. A corresponding tail bound implies a weak threshold effect and criterion for adequately choosing the value of the sparsity parameter  $k$ . Our results are in excellent agreement with numerical experiments and improve the state-of-the-art by a factor of three.

*Contribution and organization.* Although our approach will be developed for a specific imaging set-up, the techniques that we apply are more general. They should be applicable not only to alternative



**Fig. 2.** **Left:** 2D imaging geometry with  $d^2$  cells and  $2d$  projection rays (here:  $d = 6$ ). The incidence relation is given by the measurement matrix  $A = A_d^2$  (cf. Eq. (2.2)) which is the adjacency of a bipartite graph with constant left degree  $\ell = 2$ . **Right:** 3D imaging geometry with  $d^3$  cells and  $3d^2$  rays (here:  $d = 7$ ). The incidence relation given by the measurement matrix  $A = A_d^3$  is the adjacency of a bipartite graph with constant left degree  $\ell = 3$ .

imaging geometries but also to abstract sensors characterized by sparse expanders and nonnegativity, irrespective of any application context. This enables to exhibit mathematically stronger reconstruction performance under weaker conditions, that may be nonetheless relevant to real application scenarios, as in our case additionally demonstrated by detailed numerical experiments. In Section 2, we detail the mathematical abstraction of the imaging process and discuss directly related work. In Section 3, we examine recent results of compressive sensing based on sparse expanders. This sets the stage for an average case analysis conducted in Section 5 and corresponding weak recovery properties that are in sharp contrast to poor strong recovery properties presented in Section 4. We conclude with a discussion of quantitative results and their agreement with numerical experiments in Section 6.

*Notation.*  $|X|$  denotes the cardinality of a finite set  $X$  and  $[n] = \{1, 2, \dots, n\}$  for  $n \in \mathbb{N}$ . We will denote by  $\|x\|_0 = |\{i: x_i \neq 0\}|$  and  $\mathbb{R}_k^n = \{x \in \mathbb{R}^n: \|x\|_0 \leq k\}$  the set of  $k$ -sparse vectors. The corresponding sets of nonnegative vectors are denoted by  $\mathbb{R}_+^n$  and  $\mathbb{R}_{k,+}^n$ , respectively. The support of a vector  $x \in \mathbb{R}^n$ ,  $\text{supp}(x) \subseteq [n]$ , is the set of indices of non-vanishing components of  $x$ . With  $I^+(x) = \{i: x_i > 0\}$ ,  $I^0(x) = \{i: x_i = 0\}$  and  $I^-(x) = \{i: x_i < 0\}$ , we have  $\text{supp}(x) = I^+(x) \cup I^-(x)$  and  $\|x\|_0 = |\text{supp}(x)|$ .

For a finite set  $S$ , the set  $\mathcal{N}(S)$  denotes the union of all neighbors of elements of  $S$  where the corresponding relation (graph) will be clear from the context.

$\mathbb{1} = (1, \dots, 1)^T$  denotes the one-vector of appropriate dimension.

$A_{\bullet,i}$  denotes the  $i$ -th column vector of a matrix  $A$ . For given index sets  $I, J$ , matrix  $A_{IJ}$  denotes the submatrix of  $A$  with rows and columns indexed by  $I$  and  $J$ , respectively.  $I^c, J^c$  denote the respective complement sets. Similarly,  $b_I$  denotes a subvector of  $b$ .

$\mathbb{E}[\cdot]$  denotes the expectation operation applied to a random variable and  $\text{Pr}(B)$  the probability to observe an event  $B$ .

## 2. Preliminaries

### 2.1. Imaging set-up and representation

We refer to Fig. 2 for an illustration of the mathematical abstraction of the scenario depicted by Fig. 1. In order to handle in parallel the 2D and 3D cases, we will use the variable

$$D \in \{2, 3\}. \tag{2.1}$$

We measure the **problem size** in terms of  $d \in \mathbb{N}$  and consider  $n := d^D$  **cells** in a square ( $D = 2$ ) or cube ( $D = 3$ ) and  $m := Dd^{D-1}$  **rays**, compare Fig. 2, left and right. It will be useful to denote the set of cells by  $C = [n]$  and the set of rays by  $R = [m]$ . The incidence relation between cells and rays is given by an  $m \times n$  **measurement matrix**  $A_d^D$

$$(A_d^D)_{ij} = \begin{cases} 1, & \text{if } i\text{-th ray intersects } j\text{-th cell,} \\ 0, & \text{otherwise,} \end{cases} \tag{2.2}$$

for all  $i \in [m], j \in [n]$ . Thus, cells and rays correspond to columns and rows of  $A_d^D$ .

The incidence relation encoded by  $A_d^D$  gives rise to the equivalent representation in terms of a **bipartite graph**  $G = (C, R; E)$  with left and right vertices  $C$  and  $R$ , and edges  $cr \in E$  iff  $(A_d^D)_{rc} = 1$ . Fig. 2 illustrates that  $G$  has **constant left degree**  $\ell = D$ . It will be convenient to use a separate symbol  $\ell$ .

For a fixed vertex  $i$ , any adjacent vertex  $j \sim i$  is called **neighbor** of  $i$ . For any nonnegative measurement matrix  $A$  and the corresponding graph, the set

$$\mathcal{N}(S) = \{i \in [m]: i \sim j, j \in S\} = \{i \in [m]: A_{ij} > 0, j \in S\}$$

contains all neighbors of  $S$ . The same notation applies to neighbors of subsets  $S \subset [m]$  of right nodes.

With slight abuse, we call the matrix  $A_d^D$  that encodes the adjacency  $r \sim c$  of vertices  $r \in R$  and  $c \in C$  **adjacency matrix** of the induced bipartite graph  $G$ , deviating from the usual definition of the adjacency matrix of a graph that encodes the adjacency of *all* nodes  $v_i \sim v_j, V = C \cup R$ . Moreover, in this sense, we will call any nonnegative matrix adjacency matrix, based on its non-zero entries.

Let  $A$  be the nonnegative adjacency matrix of a bipartite graph with constant left degree  $\ell$ . The **perturbed matrix**  $\tilde{A}$  is computed by uniformly perturbing the non-zero entries  $A_{ij} > 0$  to obtain  $\tilde{A}_{ij} \in [A_{ij} - \varepsilon, A_{ij} + \varepsilon]$ , and by normalizing subsequently all column vectors of  $\tilde{A}$ . In practice, such perturbation can be implemented by discretizing the image by radial basis functions and choose their locations on an irregular grid, see [16].

The following class of graphs plays a key role in the present context and in the field of compressed sensing in general.

**Definition 2.1.** A  $(\nu, \delta)$ -**unbalanced expander** is a bipartite simple graph  $G = (L, R; E)$  with constant left degree  $\ell$  such that for any  $X \subset L$  with  $|X| \leq \nu$ , the set of neighbors  $\mathcal{N}(X) \subset R$  of  $X$  has at least size  $|\mathcal{N}(X)| \geq \delta \ell |X|$ .

### 2.2. Deviation bound

We will apply the following inequalities for bounding the deviation of a random variable from its expected value based on martingales, that is on sequences of random variables  $(X_i)$  defined on a finite probability space  $(\Omega, \mathcal{F}, \mu)$  satisfying

$$\mathbb{E}[X_{i+1} | \mathcal{F}_i] = X_i, \quad \text{for all } i \geq 1, \tag{2.3}$$

where  $\mathcal{F}_i$  denotes an increasing sequence of  $\sigma$ -fields in  $\mathcal{F}$  with  $X_i$  being  $\mathcal{F}_i$ -measurable.

This setting applies to random variables associated to measurements that are statistically *dependent* due to the intersection of projection rays (cf. Fig. 2).

**Theorem 2.1 (Azuma's inequality).** (See [1,6].) Let  $(X_i)_{i=0,1,2,\dots}$  be a sequence of random variables satisfying (2.3) such that for each  $i$ ,

$$|X_i - X_{i-1}| \leq c_i. \tag{2.4}$$

Then, for all  $j \geq 0$  and any  $\tau > 0$ ,

$$\Pr(|X_j - X_0| \geq \tau) \leq 2 \exp\left(-\frac{\tau^2}{2 \sum_{i=1}^j c_i^2}\right). \tag{2.5}$$

The second inequality stated below will be applied to the following setting. Let  $\mathbf{X} = (X_1, \dots, X_m)$  be family of – possibly dependent – random variables taking values in  $\mathcal{X} = \{0, 1\}^m$ . Suppose  $x_i \in \{0, 1\}$  for each  $i = 1, \dots, r - 1$ , and let  $B$  denote the event

$$B: X_i = x_i \quad \text{for } i = 1, \dots, r - 1. \tag{2.6}$$

For a given function  $f(\mathbf{X})$  define

$$g(x) = \mathbb{E}[f(\mathbf{X})|B, X_r = x] - \mathbb{E}[f(\mathbf{X})|B] \tag{2.7}$$

and the related quantities

$$\text{dev}(x_1, \dots, x_{r-1}) = \max\{|g(0)|, |g(1)|\}, \tag{2.8a}$$

$$\text{ran}(x_1, \dots, x_{r-1}) = \max_{x,y \in \{0,1\}} |g(x) - g(y)|, \tag{2.8b}$$

and with  $\mathbf{x} = (x_1, \dots, x_m)$  and index sets  $\mathcal{I}_r \subseteq [m]$ ,  $|\mathcal{I}_r| = r \leq m$ ,

$$\begin{aligned} \max_{\text{dev}} &= \max_{\mathbf{x} \in \mathcal{X}, \mathcal{I}_r} \text{dev}(x_1, \dots, x_{r-1}), \\ R^2(\mathbf{x}) &= \sum_{r=1}^m (\text{ran}(x_1, \dots, x_{r-1}))^2, \\ \hat{r}^2 &= \max_{\mathbf{x} \in \mathcal{X}} R^2(\mathbf{x}). \end{aligned} \tag{2.9}$$

**Theorem 2.2.** (See [15, Thm. 3.9].) Let  $f$  denote a bounded real-valued function defined on  $\mathcal{X}$ . Let  $\mu$  denote the mean of  $f(\mathbf{X})$ . Suppose that, for any given values taken by  $X_1, \dots, X_{r-1}$ , the random variable  $X_r$  takes two values, and the smaller corresponding probability is at most  $p \leq \frac{1}{2}$ . Then for any  $\tau \geq 0$ ,

$$\Pr(|f(\mathbf{X}) - \mu| \geq \tau) \leq 2 \exp\left(-\frac{\tau^2}{2p\hat{r}^2(1 + \frac{\tau \max_{\text{dev}}}{3p\hat{r}^2})}\right). \tag{2.10}$$

### 2.3. Related work

Although it was shown [3] that random measurement matrices are optimal for Compressive Sensing, in the sense that they require a minimal number of samples to recover efficiently a  $k$ -sparse vector, recent trends [2,22] tend to replace random dense matrices by adjacency matrices of “high quality” expander graphs. Explicit constructions of such expanders exist, but are quite involved. However, random  $m \times n$  binary matrices with nonreplicative columns, having  $\lfloor \ell n \rfloor$  ones, perform numerically extremely well, even if  $\ell$  is small, as shown in [2]. Ref. [13] shows that perturbing the elements of adjacency matrices of expander graphs with low expansion improves performance. These findings complement our prior work in [16], where we observed that by slightly perturbing the entries of a tomographic projection matrix its reconstruction performance can be improved significantly.

We wish to inspect bounds on the required sparsity that guarantee exact reconstruction of *most* sparse signals and corresponding critical parameter values similar to weak thresholds in [10,11]. The authors have computed sharp reconstruction thresholds for Gaussian measurements, such that for given a signal length  $n$  and numbers of measurements  $m$ , the maximal sparsity value  $k$  which guarantees perfect reconstruction can be determined precisely.

For a matrix  $A \in \mathbb{R}^{m \times n}$ , Donoho and Tanner define the undersampling ratio  $\delta = \frac{m}{n} \in (0, 1)$  and the sparsity as a fraction of  $m$ ,  $k = \rho m$ , for  $\rho \in (0, 1)$ . The so called *strong phase transition*  $\rho_S(\delta)$  indicates the necessary undersampling ratio  $\delta$  to recover *all*  $k$ -sparse solutions, while the *weak phase transition*  $\rho_W(\delta)$  indicates when  $x^*$  with  $\|x^*\|_0 \leq \rho_W(\delta) \cdot m$  can be recovered with overwhelming probability by linear programming.

Relevant for TomoPIV is the setting as  $\delta \rightarrow 0$  and  $n \rightarrow \infty$ , that is severe undersampling, since the number of measurements is of order  $O(10^4)$  and discretization of the volume can be made accordingly fine. For Gaussian ensembles a strong asymptotic threshold  $\rho_S(\delta) \approx (2e \log(1/\delta))^{-1}$  and weak asymptotic threshold  $\rho_W(\delta) \approx (2 \log(1/\delta))^{-1}$  holds, see e.g. [10]. In this highly undersampled regime, the asymptotic thresholds are the same for nonnegative and unsigned signals. Exact sparse recovery of

nonnegative vectors has been also studied in a series of recent papers [13,20], while [17,18] additionally assumes that all non-zero elements are equal to each other. As expected, additional information, improves the recoverable sparsity thresholds.

2.3.1. Strong recovery

The maximal sparsity  $k$  depending on  $m$  and  $n$ , such that all sparse signals are unique and coincide with the unique positive solution of  $Ax = b$ , is investigated in [10,11] from the perspective of convex geometry by studying the face lattice of the convex polytope  $\text{conv}\{A_{\bullet,1}, \dots, A_{\bullet,n}, 0\}$ . It is related to the nullspace property for nonnegative signals in what follows.

**Theorem 2.3.** (See [10,13,20,16].) Let  $A \in \mathbb{R}^{m \times n}$  be an arbitrary matrix. Then the following statements are equivalent:

- (a) Every  $k$ -sparse nonnegative vector  $x^*$  is the unique positive solution of  $Ax = Ax^*$ .
- (b) The convex polytope defined as the convex hull of the columns in  $A$  and the zero vector, i.e.  $\text{conv}\{A_{\bullet,1}, \dots, A_{\bullet,n}, 0\}$  is outwardly  $k$ -neighborly.
- (c) Every non-zero null space vector has at least  $k + 1$  negative (and positive) entries.

2.3.2. Weak recovery

Thm. 2 in [10] shows the equivalence between  $(k, \epsilon)$ -weakly (outwardly) neighborliness and weak recovery, i.e. uniqueness of all except a fraction  $\epsilon$  of  $k$ -sparse nonnegative vectors. Weak neighborliness is the same thing as saying that  $A\Delta_0^{n-1}$  has at least  $(1 - \epsilon)$ -times as many  $(k - 1)$ -faces as the simplex  $\Delta_0^{n-1}$ . A different form of weak recovery is to determine the probability that a random  $k$ -sparse nonnegative vector is unique by probabilistic nullspace analysis. These concepts are related to each other in the next theorem for an arbitrary sparse vector with exactly  $k$  nonnegative entries.

**Theorem 2.4.** Let  $A \in \mathbb{R}^{m \times n}$  be an arbitrary matrix. Then the following statements are equivalent:

- (a) The  $k$ -sparse nonnegative vector  $x^*$  supported on  $S$ ,  $|S| = k$ , is the unique positive solution of  $Ax = Ax^*$ .
- (b) Every non-zero null space vector cannot have all its negative components in  $S$ .
- (c)  $A_S \mathbb{R}_+^k$  is a  $k$ -face of  $A\mathbb{R}_+^n$ , i.e. there exists a hyperplane separating the cone generated by the linearly independent columns  $\{A_{\bullet,j}\}_{j \in S}$  from the cone generated by the columns of the off-support  $\{A_{\bullet,j}\}_{j \in S^c}$ .

**Proof.** Statement (a) holds if and only if there is no  $v \neq 0$  such that  $Av = 0$  and  $v_{S^c} \geq 0$ , compare for e.g. [14, Thm. 1]. Thus (a)  $\Leftrightarrow$  (b). By [11, Lem. 5.1], (a)  $\Leftrightarrow$  (c) holds as well.  $\square$

If in addition all  $k$  non-zero entries are equal to each other, then a stronger characterization holds.

**Theorem 2.5.** (See [14, Prop. 2].) Let  $A \in \mathbb{R}^{m \times n}$  be an arbitrary matrix. Then the following statements are equivalent:

- (a) The  $k$ -sparse binary vector  $x^* \in \{0, 1\}^n$  supported on  $S$ ,  $|S| = k$ , is the unique solution of  $Ax = Ax^*$  with  $x \in [0, 1]^n$ .
- (b) Every non-zero null space vector cannot have all its negative components in  $S$  and the positive ones in  $S^c$ .
- (c) There exists a vector  $r$  such that  $\text{Diag}(z^*)A^\top r > 0$ , with  $z^* := \mathbb{1} - 2x^*$ .
- (d)  $0 \in \mathbb{R}^m$  is not contained in the convex hull of the columns of  $A\text{Diag}(z^*)$ , i.e.  $0 \notin \text{conv}\{z_1^*A_{\bullet,1}, \dots, z_n^*A_{\bullet,n}, 0\}$ , with  $z^* := \mathbb{1} - 2x^*$ .

**Proof.** Uniqueness of  $x^*$  in  $\{x: Ax = Ax^*, x \in [0, 1]^n\}$  holds, for e.g. by [14, Thm. 1], if there is no  $v \neq 0$  such that  $Av = 0$ ,  $v_{S^c} \geq 0$  and  $v_S \leq 0$ , which shows equivalence to (b). With  $D := \text{Diag}(\mathbb{1} - 2x^*)$  and  $DD = I$ , (b) can be rewritten as follows: there is no  $v \neq 0$  such that  $ADDv = 0$ ,  $Dv \geq 0$ ,  $Dv \neq 0$ . With  $u := Dv$ , the above condition becomes:

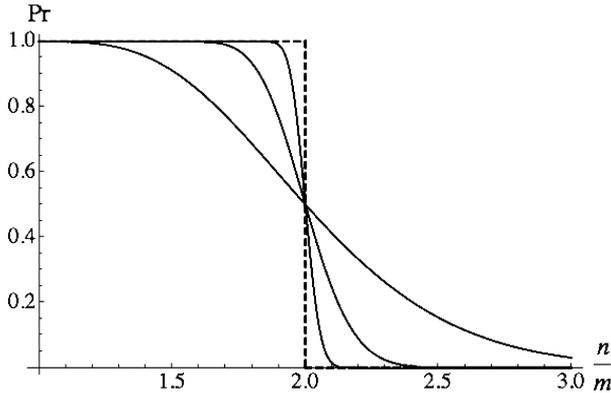


Fig. 3. The probability  $\Pr(n, m)$  given by (2.12) that  $n$  points in general position in  $\mathbb{R}^m$  can be linearly separated [21]. This holds with probability  $\Pr(n, m) = 1$  for  $n/m \leq 1$ , and with  $\Pr(n, m) \rightarrow 1$  if  $m \rightarrow \infty$  and  $1 \leq n/m < 2$ .

$$ADu = 0, \quad u \geq 0, \quad u \neq 0, \quad \text{has no solution,}$$

which by Gordon’s theorem of alternative gives the equivalent certificate (c):

$$\exists r \quad \text{such that} \quad DA^T r > 0. \tag{2.11}$$

In other words, a small  $k$ -subset of the columns of  $A$ , are “flipped” by multiplication with  $-1$ , and these modified columns together with all remaining ones can be separated from the origin, which shows equivalence to (d), i.e.  $0$  is not contained in the convex hull of these points.  $\square$

Note that statement (d) is related to the necessary condition for uniqueness in [20, Thm. 1]. We further comment on Theorem 2.5 (c) from a probabilistic viewpoint. Condition (c) says that all points defined by the columns of  $A \text{Diag}(\mathbb{1} - 2x^*)$  are located in a single half space defined by a hyperplane through the origin with normal  $r$ . Conditions under which this is likely to hold were studied by Wendel [21]. This problem is also directly related to the basic pattern recognition problem concerning the linear classification<sup>1</sup> of any dichotomy of a finite point set [5].

Assuming  $n$  points in  $\mathbb{R}^m$  to be in general position, that is any subset of  $m$  vectors is linearly independent, and that the distribution from which the given point set is regarded as an i.i.d. sample set is symmetric with respect to the origin, then condition (2.11) holds with probability

$$\Pr(n, m) = \frac{1}{2^{n-1}} \sum_{i=0}^{m-1} \binom{n-1}{i}. \tag{2.12}$$

As Fig. 3 illustrates,  $\Pr(n, m) = 1$  if  $n/m \leq 1$ , due to the well-known fact that any dichotomy of  $m + 1$  points in  $\mathbb{R}^m$  can be separated by a hyperplane [19,7]. For increasing dimension  $m \rightarrow \infty$ , this also holds almost surely if  $n/m < 2$ , which can be easily deduced by applying a binomial tail bound. Accordingly, assuming that the measurement matrix  $A$  conforms to the assumptions, the authors of [14] conclude that an existing binary solution to (1.1) is unique with probability (2.12) for underdetermined systems with ratio  $m/n > 1/2$ .

We adopt this viewpoint in Section 5.4 and develop a criterion for unique recovery with high probability using the given measurement matrix (2.2), based on a probabilistic average case analysis of condition (5.39) (Section 5.4). This criterion currently characterizes best the design of tomographic scenarios (Fig. 2), with recovery performance guaranteed with high probability. We conclude this section by mentioning that exact nonasymptotic recovery results for a  $k$ -sparse nonnegative vector

<sup>1</sup> In this context, “linear” means affine decision functions.

are obtained in [11, Thm. 1.10] by exploiting Wendel’s theorem. Donoho and Tanner show that the probability of uniqueness of a  $k$ -sparse nonnegative vector equals  $\Pr(n - m, n - k)$ , provided  $A$  satisfies certain conditions which do not hold in our considered application.

### 3. Expanders, perturbation, and recovery via reduced systems

Both Sections 3.1 and 3.2 collect recent results of recovery properties based on expanders associated with sparse measurement matrices, possibly after a random perturbation of the non-zero matrix entries. In Section 3.3 we show that the support of a sparse nonnegative vector can always be recovered by solving an appropriate reduced system. Section 3.4 applies these results to our specific setting in a form suitable for a probabilistic analysis of recovery performance presented in Section 5.

#### 3.1. Expanders and recovery

The following theorem is a slight variation of Theorem 4 in [20] tailored to our specific setting.

**Theorem 3.1.** *Let  $A$  be the adjacency matrix of a  $(\nu, \delta)$ -unbalanced expander and  $1 \geq \delta > \frac{\sqrt{5}-1}{2}$ . Then for any  $k$ -sparse nonnegative vector  $x^*$  with  $k \leq \frac{\nu}{(1+\delta)}$ , the solution set  $\{x: Ax = Ax^*, x \geq 0\}$  is a singleton.*

**Proof.** We will show that every non-zero null space vector has at least  $\frac{\nu}{(1+\delta)} + 1$  negative and positive entries. Then Theorem 2.3 will provide the desired assertion.

Suppose without loss of generality that there is a vector  $v \in \ker(A) \setminus \{0\}$  with

$$s := |I^-(v)| \leq \frac{\nu}{(1 + \delta)}. \tag{3.1}$$

Then

$$\ell |I^-(v)| \geq |\mathcal{N}(I^-(v))| \geq \delta \ell s, \tag{3.2}$$

where the second inequality follows by assumption due to the expansion property.

Denoting by  $S$  the support of  $v$ ,  $S = I^-(v) \cup I^+(v)$ , we have

$$\mathcal{N}(I^-(v)) = \mathcal{N}(I^+(v)) = \mathcal{N}(S), \tag{3.3}$$

since otherwise  $Av \neq 0$  because  $A$  is nonnegative.

From  $\ell |I^+(v)| \geq |\mathcal{N}(I^+(v))|$ , (3.3) and (3.2), we obtain

$$|I^+(v)| \geq \delta s.$$

Thus,

$$|S| = |I^-(v)| + |I^+(v)| \geq (1 + \delta)s.$$

Let  $\tilde{S} \subseteq S$  such that  $|\tilde{S}| = \lceil (\delta + 1)s \rceil$ . Since  $\nu \geq \lceil (\delta + 1)s \rceil$  due to (3.1) and  $\nu \in \mathbb{N}$ ,  $|\tilde{S}| \leq \nu$  holds. Now

$$|\mathcal{N}(\tilde{S})| \geq \delta \ell |\tilde{S}| \geq \delta \ell (\delta + 1)s > s \ell \tag{3.4}$$

provided  $\delta(1 + \delta) > 1 \Leftrightarrow \delta > (\sqrt{5} - 1)/2$ . Summarizing, we get  $s \ell < |\mathcal{N}(\tilde{S})| \leq |\mathcal{N}(S)| = |\mathcal{N}(I^-(v))| \leq s \ell$ , hence a contradiction.  $\square$

The assertion of Theorem 3.1 solely relies on the expansion property of the measurement matrix  $A$ . Theorem 3.5 below will be based on it and in turn the results of Section 5.2.

3.2. Perturbed expanders and recovery

We describe next an alternative route based on the **complete (Kruskal) rank**  $r_0 = r_0(A)$  of a measurement matrix  $A$ . This is the maximal integer  $r_0$  such that every subset of  $r_0$  columns of  $A$  is linearly independent.

While this number is combinatorially difficult to compute in practice, both the number and the corresponding recovery performance can be enhanced by relating it to a particular expansion property of the bipartite graph associated to a *perturbed* measurement matrix  $\tilde{A}$ . The latter can be easily computed in practice while preserving its sparsity, i.e. the constant left degree  $\ell$ .

**Theorem 3.2.** (See [16, Thm. 6.2], [13, Thm. 4.1].) *Let  $A$  be a nonnegative matrix with  $\ell$  non-zero entries in each column and complete rank  $r_0 = r_0(A)$ . Then  $|I^-(v)| \geq r_0/\ell$  for all nullspace vectors  $v \in \ker(A)$ .*

**Remark 3.1.** In view of Theorem 2.3 (c), Theorem 3.2 says that all  $k$ -sparse nonnegative vectors  $x$  can be uniquely recovered if  $k \leq \lceil r_0/\ell - 1 \rceil$ .

The following lemma asserts that by a perturbation of the measurement matrix the complete rank, and hence the recovery property, may be enhanced provided all subsets of columns, up to a related cardinality, entail an expansion that is *less* however than the one required by Theorem 3.1.

**Lemma 3.3.** (See [13, Lem. 4.2].) *Let  $A$  be a nonnegative matrix with  $\ell$  non-zero entries in each column. Suppose that for any submatrix formed by  $\tilde{r}_0$  columns of  $A$  it holds that  $|\mathcal{N}(X)| \geq |X|$ , for each subset  $X \subset C$  of columns of cardinality  $|X| \leq \tilde{r}_0$ . Then there exists a perturbed matrix  $\tilde{A}$  that has the same structure as  $A$  such that its complete rank satisfies  $r_0(\tilde{A}) \geq \tilde{r}_0$ .*

3.3. Reduced systems

Due to the nonnegativity and sparsity of both  $A$  and  $x$  the data vector  $b$  will be sparse and non-negative as well. As a consequence zero measurements and redundant columns in  $A$  can be removed from the original system. In this section we formalize this procedure and check equivalence to the unreduced system.

Recall from Section 2.1 that we regard a given measurement matrix  $A$  also as adjacency matrix of a bipartite graph  $G = (C, R; E)$ .

**Definition 3.1.** The **reduced system** corresponding to a given nonnegative vector  $b$ ,

$$A_{red}x = b_{red}, \quad A_{red} \in \mathbb{R}_+^{m_{red} \times n_{red}}, \tag{3.5}$$

results from  $A, b$  by choosing the subsets of rows and columns

$$R_b := \text{supp}(b), \quad C_b := C \setminus \mathcal{N}(R_b^c) = \mathcal{N}(R_b^c)^c \tag{3.6}$$

with

$$m_{red} := |R_b|, \quad n_{red} := |C_b|. \tag{3.7}$$

We further define

$$S^+ := \{x: Ax = b, x \geq 0\} \tag{3.8}$$

and

$$S_{red}^+ := \{x: A_{R_b C_b}x = b_{R_b}, x \geq 0\}. \tag{3.9}$$

The following proposition asserts that solving the reduced system (3.5) will always recover the support of the solution to the original system  $Ax = b$ .

**Proposition 3.4.** Let  $A \in \mathbb{R}^{m \times n}$  and  $b \in \mathbb{R}^m$  have nonnegative entries only, and let  $S^+$  and  $S_{red}^+$  be defined by (3.8) and (3.9), respectively. Then

$$S^+ = \{x \in \mathbb{R}^n: x_{(C_b)^c} = 0 \text{ and } x_{C_b} \in S_{red}^+\}. \tag{3.10}$$

**Proof.** Let  $S := \{x \in \mathbb{R}^n: x_{(C_b)^c} = 0 \text{ and } x_{C_b} \in S_{red}^+\}$ . We first show  $S \subseteq S^+$ . Let  $x \in S$ . From this  $x \geq 0$  follows directly. Thus, we just have to show  $\sum_{j=1}^n a_{ij}x_j = b_i, \forall i \in [n]$ . Indeed, for

$$i \in R_b: \sum_{j=1}^n a_{ij}x_j = \sum_{j \in C_b} \underbrace{a_{ij}x_j}_{=b_i} + \sum_{j \in (C_b)^c} a_{ij} \underbrace{x_j}_{=0} = b_i,$$

whereas for

$$i \in (R_b)^c: \sum_{j=1}^n a_{ij}x_j = \sum_{j \in C_b} \underbrace{a_{ij}}_{=0} x_j + \sum_{j \in (C_b)^c} a_{ij} \underbrace{x_j}_{=0} = 0 = b_i.$$

Now let  $x \in S^+$  and consider any  $i \in (R_b)^c$ . By definition  $(C_b)^c = \mathcal{N}(R_b^c)$  and thus  $a_{ij} > 0$  for every  $i \in (R_b)^c$  and  $j \in \mathcal{N}(R_b^c)$ . Then

$$0 = b_i = \sum_{j=1}^n a_{ij}x_j = \sum_{j \in C_b} \underbrace{a_{ij}}_{=0} x_j + \sum_{j \in (C_b)^c} \underbrace{a_{ij}}_{>0} x_j \tag{3.11}$$

holds. Since  $x \geq 0$ , we obtain from (3.11) that  $x_j = 0, \forall j \in (C_b)^c$ . To show that  $A_{R_b C_b} x_{C_b} = b_{R_b}$ , consider

$$i \in R_b: \sum_{j \in C_b} a_{ij}x_j = \sum_{j \in C_b} a_{ij}x_j + \sum_{j \in (C_b)^c} a_{ij} \underbrace{x_j}_{=0} = \sum_{j=1}^n a_{ij}x_j = b_i.$$

Hence,  $x_{(C_b)^c} = 0$  and  $x_{C_b} \in S_{red}^+$ . Thus  $x \in S$ .  $\square$

### 3.4. Recovery via reduced systems

In view of the previous section, the reconstruction of a random  $k$ -sparse vector  $x$  supported on  $X$  will be based on a reduced linear system restricted to the rows  $\mathcal{N}(X)$  and the columns  $\mathcal{N}(\mathcal{N}(X)^c)^c$ . Dimensions of reduced systems will be the same for most random sets  $X = \text{supp}(x)$ , with  $|X| = k$ , contained in  $C$ . Consequently, in view of a probabilistic average case analysis conducted in Section 5, it suffices to measure the expansion with respect to these sets.

Taking Proposition 3.4 into account, the following theorem tailors Theorem 3.1 to our specific setting.

**Theorem 3.5.** Let  $X$  be a random subset  $X \subset C$  of left nodes, with  $R_b = \mathcal{N}(X)$  and  $C_b := \mathcal{N}(R_b^c)^c$ . If any  $Y \subset C_b$  satisfies

$$|\mathcal{N}(Y)| \geq \delta \ell |Y| \quad \text{with} \quad \delta = \frac{\sqrt{5} - 1}{2}, \tag{3.12}$$

then the solution set  $\{x: A_{red}x = A_{red}x^*, x \geq 0\}$  is a singleton for any  $\frac{|C_b|}{(1+\delta)}$ -sparse nonnegative vector  $x^*$ .

Now if  $X \subset C_b$  fulfills  $|X| \leq \frac{|C_b|}{(1+\delta)}$  then we obtain recovery of  $x$  supported on  $X$ .

The next theorem follows from Proposition 3.4 and Lemma 3.3 and simply states that recovery is guaranteed via full rank overdetermined perturbed reduced systems.

**Theorem 3.6.** Let  $X$  be a random subset  $X \subset C$  of left nodes, with  $R_b = \mathcal{N}(X)$  and  $C_b := \mathcal{N}(R_b^c)^c$ . If any  $Y \subset C_b$  satisfies

$$|\mathcal{N}(Y)| \geq |Y|, \tag{3.13}$$

then there exists a perturbation  $\tilde{A}_{red}$  of  $A_{red}$  such that for any  $|C_b|$ -sparse nonnegative vector  $x^*$ , the solution set  $\{x: \tilde{A}_{red}x = \tilde{A}_{red}x^*, x \geq 0\}$  is a singleton.

Since  $X \subset C_b$  this result also implies recovery of  $x$  supported on  $X$ . The consequences of Theorems 3.5 and 3.6 are investigated in Section 5 by working out critical values of the sparsity parameter  $k$  for which the respective conditions are satisfied with high probability.

**4. Strong equivalence**

In [16] we tested the properties of the discrete tomography matrix in focus against various conditions, like the null space property, the restricted isometry property, etc., and predicted an extremely poor worst case performance of such a measurement system. In the 3D case we showed that the strong threshold on sparsity, that is the maximal sparsity level  $k_0$  for which recovery of all  $k$ -sparse (positive) vectors,  $k \leq k_0$ , is guaranteed, is a constant, not depending on the undersampling ratio  $d$ .

**4.1. Unperturbed systems**

Given an indexing of cells and rays, we can rewrite the projection matrix  $A_d^D \in \mathbb{R}^{Dd^{D-1} \times d^D}$  from (2.2) in closed form as

$$A_d^D := \begin{cases} \begin{pmatrix} I_d \otimes \mathbb{1}_d^T \\ \mathbb{1}_d^T \otimes I_d \end{pmatrix}, & \text{if } D = 2, \\ \begin{pmatrix} \mathbb{1}_d^T \otimes I_d \otimes I_d \\ I_d \otimes \mathbb{1}_d^T \otimes I_d \\ I_d \otimes I_d \otimes \mathbb{1}_d^T \end{pmatrix}, & \text{if } D = 3. \end{cases} \tag{4.1}$$

Since for these matrices a sparse nullspace basis can be computed, we can derive the maximal sparsity via the nullspace property, as shown next.

**Proposition 4.1.** (See [16, Prop. 2.2, Prop. 3.2].) Let  $D \in \{2, 3\}$ ,  $d \in \mathbb{N}$ ,  $d \geq 3$  and  $A_d^D$  from (4.1). Define  $B_d^D \in \mathbb{R}^{d^D \times (d-1)^D}$  as

$$B_d^D := \begin{cases} \begin{pmatrix} -\mathbb{1}_{d-1}^T \\ I_{d-1} \end{pmatrix} \otimes \begin{pmatrix} -\mathbb{1}_{d-1}^T \\ I_{d-1} \end{pmatrix}, & \text{if } D = 2, \\ \begin{pmatrix} -\mathbb{1}_{d-1}^T \\ I_{d-1} \end{pmatrix} \otimes \begin{pmatrix} -\mathbb{1}_{d-1}^T \\ I_{d-1} \end{pmatrix} \otimes \begin{pmatrix} -\mathbb{1}_{d-1}^T \\ I_{d-1} \end{pmatrix}, & \text{if } D = 3. \end{cases} \tag{4.2}$$

Then the following statements hold:

- (a)  $A_d^D B_d^D = 0$ .
- (b) Every column in  $B_d^D$  has exactly  $2^D$  non-zero ( $2^{D-1}$  positive,  $2^{D-1}$  negative) elements.
- (c)  $B_d^D$  is a full rank matrix and  $\text{rank}(B_d^D) = (d-1)^D$ .
- (d)  $\ker(A_d^D) = \text{span}\{B_d^D\}$ , i.e. the columns of  $B_d^D$  provide a basis for the null space of  $A_d^D$ .
- (e)  $\text{rank}(A_d^D) = d^D - (d-1)^D$ .
- (f)  $\sum_{i=1}^n v_i = 0$  holds for all  $v \in \ker(A_d^D)$ .

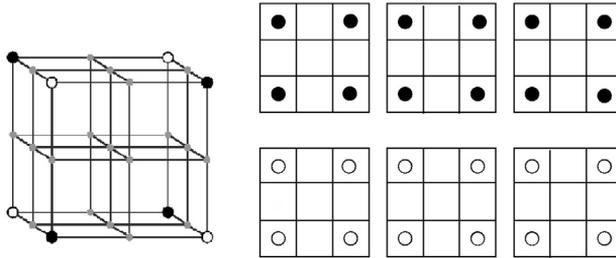


Fig. 4. Two different *nonunique* 4-sparse “particle” distributions in a  $3 \times 3 \times 3$  volume. Both configurations (represented by black and white dots) yield identical projections in all three directions. Such nonunique configurations correspond to positive or negative entries in an 8-sparse nullspace vector of  $A_d^3$ , compare to Proposition 4.1.

(g) The Kruskal rank of  $A_d^D$  is  $2^D - 1$ , i.e.

$$\min_{\substack{v \in \ker(A_d^D) \\ v \neq 0}} \|v\|_0 = 2^D.$$

(h) Every non-zero nullspace vector has at least  $2^{D-1}$  negative entries, i.e.

$$\min_{\substack{v \in \ker(A_d^D) \\ v \neq 0}} |I^-(v)| = 2^{D-1}.$$

Thus, (g) and (h) imply

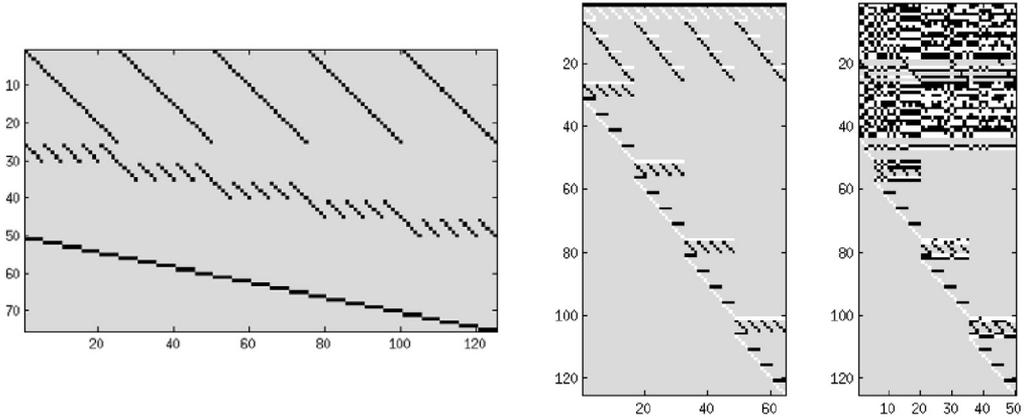
**Corollary 4.2.** For all  $d \in \mathbb{N}$ ,  $d \geq 3$ , every  $(2^{D-1} - 1)$ -sparse vector  $x^*$  is the unique sparsest solution of  $A_d^D x = A_d^D x^*$ . Moreover, for every  $(2^{D-1} - 1)$ -sparse positive vector  $x^*$   $\{x: A_d^D x = A_d^D x^*\}$  is a singleton.

This bound is tight, since we can construct two  $2^{D-1}$ -sparse solutions  $x^1$  and  $x^2$  such that  $A_d^D x^1 = A_d^D x^2$ , compare Fig. 4 for the 3D case. However, when  $D = 3$ , not every 8-column combination, or more, in  $A_d^3$  is linearly dependent. In fact, only a limited number of  $k$ -column combinations can be dependent without violating  $\text{rank}(A_d^3) = 3d^2 - 3d + 1$ . It turns out that this number is tiny for smaller  $k$  when compared to  $\binom{n}{k}$ . As  $k$  increases the probability of linear dependency of  $k$  arbitrary columns in  $A_d^3$  also grows and equals 1 only when  $k > \text{rank}(A_d^3)$ . Likewise, not every 4-sparse binary vector is nonunique. Due to the simple geometry of the problem it is not difficult to count the “bad” 4-sparse configurations in 3D. Since they are always located in 4 out of 8 corners of a cuboid in the  $d^3$  cube, compare Fig. 4 left, and there are only two possibilities to choose them, the probability that a 4-sparse binary vector is unique, equals

$$1 - \frac{2 \binom{d}{2}^3}{\binom{d^3}{4}} = 1 - \frac{6(d-1)^2}{(d^2+d+1)(d^3-2)(d^3-3)} = 1 - \mathcal{O}(d^{-6}) \xrightarrow{d \rightarrow \infty} 1.$$

#### 4.2. Perturbed systems

The weak performance of  $A_d^D$  rests upon its small Kruskal rank. In order to increase the maximal number  $k$  of columns such that all  $k$  (or less) column combinations are linearly independent we perturb the non-zero entries of the original matrix  $A_d^D$ . Fig. 5, right, indicates that if we could estimate the Kruskal rank  $\tilde{r}_0$  of the perturbed system we could apply Theorem 3.2 and obtain a lower bound on the sparsity yielding strong recovery for all  $\lceil \tilde{r}_0 / \ell - 1 \rceil$ -sparse vectors. However, determining  $\tilde{r}_0$  for the perturbed matrix seems impossible. We believe however that it increases with  $d$ , in contrast to the



**Fig. 5.** **Left:** The 3D projection matrix  $A_{5,3}^3$  and, **middle,** a sparse basis which spans its nullspace. **Right:** If we allow a small perturbation of the non-zero entries of  $A_d^D$ , all corresponding nullspace vectors of the perturbed matrix will be less sparse and lie in a  $d^{D-1}(d-D)$ -dimensional subspace, compared to  $(d-1)^D$  in the unperturbed case.

constant  $2^D - 1$  in the case of unperturbed systems. Luckily, it turns out in Section 5.2 that the weak recovery threshold for unperturbed systems will give a *lower bound* on the strong recovery threshold for perturbed matrices, since reduced systems will be strictly overdetermined and guaranteed of full rank.

**5. Weak recovery**

In this section, we consider the recovery properties of the 3D setup depicted in Fig. 2 and establish conditions for weak recovery, that is conditions for unique recovery that holds *on average with high probability*. We clearly point out that our conditions do *not* guarantee unique recovery in *each* concrete problem instance.

**Remark 5.1.** In what follows, the phrase **with high probability** refers to values of the sparsity parameter  $k$  for which random supports  $|\text{supp}(b)|$  concentrate around the crucial expected value  $N_R$ , according to Proposition 5.2, and for which the number of corresponding non-redundant cells  $|C_b|$  concentrate around the expected value  $N_C$ , according to Proposition 5.8, thus yielding a desired threshold effect.

We first inspect in Section 5.1 the effect of sparsity on the expected dimensions of a reduced system of linear equations. Moreover, a corresponding tail bound implies a threshold effect for these dimensions. Subsequently, we establish the conditions for weak recovery based on Theorems 3.5 and 3.6, and on the *expected* quantities involved in the corresponding conditions.

In particular, we establish such uniqueness conditions for reduced underdetermined systems of dimension  $m/n > (\sqrt{5} - 1)/2 \approx 0.618$ . Our results are in excellent agreement with numerical experiments discussed in Section 6.

5.1. *Expected dimensions of reduced systems*

We compute the expected values of the reduced system dimension (3.7).

5.1.1. *Expected number of non-zero measurements*

We consider the uniform random assignment of  $k$  particles to the  $n = |C|$  cells  $c \in C$ . A single cell may be occupied by more than a single particle. This corresponds to the physical situation that real particles are very small relative to the discretization depicted by Fig. 2. The imaging optics enlarges

the appearance of particles, and the action of physical projection rays is adequately represented by linear superposition.

This scenario gives rise to a random vector  $x \in \mathbb{R}_{k,+}^n$  with support  $|\text{supp}(x)| \leq k$ . It generates a vector

$$b = A_d^D x \in \mathbb{R}_+^m \tag{5.1}$$

of measurements. We are interested in the expected size of the support of  $b$ ,

$$N_R := \mathbb{E}[|\text{supp}(b)|], \quad N_R^0 := m - N_R, \tag{5.2}$$

that equals the number of projection rays  $r \in R$  with non-vanishing measurements  $b_r \neq 0$ . We denote the event  $b_r = 0$  by the binary random variable<sup>2</sup>  $X_r = 1$ , i.e.  $X_r = 0$  corresponds to the event  $b_r > 0$  that at least a single particle meets ray  $r$ .

The probability that a single  $c$  is met by ray  $r$  is

$$q_d := \frac{d}{|C|} = \frac{d}{n} = \frac{1}{d^{D-1}}. \tag{5.3}$$

For  $k$  particles, the probability that  $0 \leq i \leq k$  particles meet projection ray  $r$  is

$$\Pr(b_r = i) = \binom{k}{i} q_d^i p_d^{k-i}, \quad p_d := 1 - q_d. \tag{5.4}$$

Consequently, we have

$$\Pr[X_r = 1] = \mathbb{E}[X_r] = p_d^k, \tag{5.5a}$$

$$\Pr[X_r = 0] = \sum_{i=1}^k \binom{k}{i} q_d^i p_d^{k-i} = 1 - p_d^k. \tag{5.5b}$$

**Lemma 5.1.** *The expected number of non-zero measurements defined by (5.2) is*

$$N_R = N_R(k) = |R|(1 - p_d^k) = Dd^{D-1} \left( 1 - \left( 1 - \frac{1}{d^{D-1}} \right)^k \right), \tag{5.6a}$$

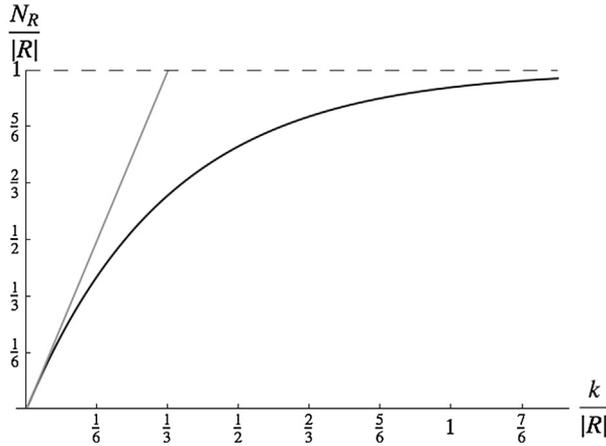
$$N_R^0 = N_R^0(k) = |R| - N_R = |R|p_d^k = Dd^{D-1} \left( 1 - \frac{1}{d^{D-1}} \right)^k. \tag{5.6b}$$

**Proof.** Due to the linearity of expectation, summing over all rays gives

$$N_R = \mathbb{E} \left[ \sum_{r \in R} (1 - X_r) \right] = |R|(1 - p_d^k). \quad \square$$

**Remark 5.2.** Note that  $N_R$  specifies the expected value of  $m_{red}$  in (3.7) induced by random  $k$ -sparse vectors  $x \in \mathbb{R}_{k,+}^n$ . See Fig. 6 for an illustration.

<sup>2</sup> We economize notation here by re-using the symbol  $X$ , a random indicator vector indexed by rays (right nodes)  $r \in R$ . Due to the context, there should be no danger of confusion with  $X = \text{supp}(x)$  denoting random subsets of left nodes used in other sections.



**Fig. 6.** The expected number  $N_R$  of non-zero measurements (5.5). For highly sparse scenarios (small  $k$ ), the expected support (5.2) of the measurement vector  $|\text{supp}(b)| \approx 3k$ . For large values of  $k$ , this rate decreases due to the multiple incidence of cells and projection rays.

*Bounding the deviation of  $N_R^0$ .* We are interested in how sharply the random number  $X = \sum_{r \in R} X_r$  of zero measurements peaks around its expected value  $N_R^0 = \mathbb{E}[X]$  given by (5.6). We derive next a corresponding tail bound by regarding a sequence of  $k$  randomly located cells and by bounding the difference of subsequent conditional expected values of the random variable  $X$ . Theorem 2.1 then provides a bound for the deviation  $|X - \mathbb{E}[X]|$ .

Let the set of rays  $R$  represent the elementary events corresponding to the observations  $X_r = 1$  or  $X_r = 0$  for each ray  $r \in R$ , i.e. ray  $r$  corresponds to a zero measurement or not.

Let  $\mathcal{F}_i \subset 2^R$ ,  $i = 0, 1, 2, \dots$ , denote the  $\sigma$ -field generated by the collection of subsets of  $R$  that correspond to all possible events after having observed  $i$  randomly selected cells. We set  $\mathcal{F}_0 = \{\emptyset, R\}$ . Because observing cell  $i + 1$  just further partitions the current state based on the previously observed  $i$  cells by possibly removing some ray (or rays) from the set of zero measurements, we have a nested sequence (filtration)  $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_k$  of the set  $2^R$  of all subsets of  $R$ .

Based on this, for a fixed value of the sparsity parameter  $k$ , we define the sequence of random variables

$$Y_i = \mathbb{E}[X | \mathcal{F}_i], \quad i = 0, 1, \dots, k, \tag{5.7}$$

where  $Y_i$ ,  $i = 0, 1, \dots, k - 1$ , are the random variables specifying the expected number of zero measurements after having observed  $k$  randomly selected cells, conditioned on the subset of events  $\mathcal{F}_i$  determined by the observation of  $i$  randomly selected cells. Consequently,  $Y_0 = \mathbb{E}[X] = N_R^0$  due to the absence of any information, and  $Y_k = X$  is just the observed number of zero measurements. The sequence  $(Y_i)_{i=0, \dots, k}$  is a martingale by construction satisfying  $\mathbb{E}[Y_{i+1} | \mathcal{F}_i] = Y_i$ , that is condition (2.3).

**Proposition 5.2.** Let  $N_R^0 = \mathbb{E}[X]$  be the expected number of zero measurements for a given sparsity parameter  $k$ , given by (5.6). Then, for any  $\tau > 0$ ,

$$\begin{aligned} \Pr(|X - N_R^0| \geq \tau) &\leq 2 \exp\left(-\frac{1 - p_d^2}{(1 - p_d^{2k})} \frac{\tau^2}{2D^2}\right) \\ &\nearrow 2 \exp\left(-\frac{\tau^2}{2D^2k}\right) \quad \text{if } d \rightarrow \infty. \end{aligned} \tag{5.8}$$

This result shows that for large problem sizes  $d$  occurring in applications, concentration of observations of  $N_R^0$  primarily depends on the sparsity parameter  $k$ . As a consequence, the bound enables suitable choices of  $k = k(d)$  of the sparsity parameter depending on the problem size.

For example, typical values

$$k = \begin{cases} 0.05d & \text{in 2D,} \\ 0.05d^2 & \text{in 3D,} \end{cases} \tag{5.9}$$

chosen by engineers<sup>3</sup> in applications according to a rule of thumb, result in

$$\Pr(|X - N_R^0| \geq \tau) \leq \begin{cases} 2 \exp(-\frac{5}{2d} \tau^2) & \text{in 2D,} \\ 2 \exp(-\frac{10}{9d^2} \tau^2) & \text{in 3D.} \end{cases} \tag{5.10}$$

For the 3D case (5.9), the probability to observe deviations from  $N_R^0$  larger than 1% drops below 0.01 for problem sizes  $d \geq 77$ , which is common in practice.

Thus, the bound (5.8) is strong enough to indicate not only that (5.9) is a particular sensible choice, but also leads to more proper choices of  $k$  for applications, which still give highly concentrated values of observations of  $N_R^0$ . This is the essential prerequisite for threshold effects of unique recovery from sparse measurements.

**Proof of Proposition 5.2.** Let  $R_{i-1}^0 \subset R$  denote the subset of rays with zero measurements after the random selection of  $i - 1 < k$  cells. For the remaining  $k - (i - 1)$  trials, the probability that not any cell incident with some ray  $r \in R_{i-1}^0$  will be selected, is

$$p_d^{k-(i-1)} = \mathbb{E}[X_r | \mathcal{F}_{i-1}], \tag{5.11}$$

with  $p_d$  given by (5.4). Consequently, by linearity, the expectation  $Y_{i-1}$  of zero measurements given  $|R_{i-1}^0|$  zero measurements after the selection of  $i - 1$  cells, is

$$Y_{i-1} = \mathbb{E}[X | \mathcal{F}_{i-1}] = \sum_{r \in R_{i-1}^0} p_d^{k-(i-1)}. \tag{5.12}$$

Now suppose we observe the random selection of the  $i$ -th cell. We distinguish two possible cases.

- (1) Cell  $i$  is not incident with any ray  $r \in R_{i-1}^0$ . Then the number of zero measurements remains the same, and

$$Y_i = \sum_{r \in R_{i-1}^0} p_d^{k-i}. \tag{5.13}$$

Furthermore,

$$\begin{aligned} Y_i - Y_{i-1} &= \sum_{r \in R_{i-1}^0} (p_d^{k-i} - p_d^{k-(i-1)}) = |R_{i-1}^0| p_d^{k-i} (1 - p_d) \\ &\leq (|R| - 1) p_d^{k-i} q_d. \end{aligned} \tag{5.14}$$

- (2) Cell  $i$  is incident with  $1, \dots, D$  rays contained in  $R_{i-1}^0$ . Let  $R_i^0$  denote the set  $R_{i-1}^0$  after removing these rays. Then

$$Y_i = \sum_{r \in R_i^0} p_d^{k-i}.$$

Furthermore, since  $R_i^0 \subset R_{i-1}^0$  and  $|R_{i-1}^0 \setminus R_i^0| \leq D$ ,

<sup>3</sup> Personal communication.

$$\begin{aligned}
 Y_{i-1} - Y_i &= \sum_{r \in R_{i-1}^0 \setminus R_i^0} p_d^{k-(i-1)} - \sum_{r \in R_i^0} (p_d^{k-i} - p_d^{k-(i-1)}) \\
 &\leq D p_d^{k-i+1} - \sum_{r \in R_i^0} p_d^{k-i} (1 - p_d) \leq D p_d^{k-i+1}.
 \end{aligned}
 \tag{5.15}$$

And similar to (5.14),

$$\begin{aligned}
 Y_{i-1} - Y_i &\geq - \sum_{r \in R_i^0} (p_d^{k-i} - p_d^{k-(i-1)}) \geq -|R_i^0| p_d^{k-i} (1 - p_d) \\
 &\geq -|R| p_d^{k-i} q_d.
 \end{aligned}
 \tag{5.16}$$

Comparing the bounds (5.14), (5.15) and (5.16), we have with  $|R|q_d = D$ ,

$$\begin{aligned}
 |Y_i - Y_{i-1}| &\leq \max\{|R| p_d^{k-i} q_d, D p_d^{k-i+1}\} = \max\{D p_d^{k-i}, D(1 - q_d) p_d^{k-i}\} \\
 &= D p_d^{k-i}.
 \end{aligned}
 \tag{5.17}$$

Thus, in view of (2.4) and the bound that follows, we compute

$$\sum_{i=1}^k (D p_d^{k-i})^2 = D^2 \frac{1 - p_d^{2k}}{1 - p_d^2}.$$

Inserting  $p_d$  from (5.4) and expanding in terms of  $d^{-1}$  at 0, we obtain

$$\frac{1 - p_d^{2k}}{1 - p_d^2} = \begin{cases} k + (k - k^2)d^{-1} + \mathcal{O}(d^{-2}), & \text{in 2D } d \rightarrow \infty \\ k + (k - k^2)d^{-2} + \mathcal{O}(d^{-4}), & \text{in 3D } \end{cases} k.$$

Applying Theorem 2.1 completes the proof.  $\square$

### 5.1.2. Expected number of cells

In the previous section, we computed the expected number of measurements  $N_R = \mathbb{E}[|\text{supp}(b)|]$  induced by a random unknown  $k$ -sparse vector  $x$  (Lemma 5.1) along with a tail bound for  $N_R^0 = |R| - N_R$  (Proposition 5.2).

In the present section, we determine the expected number of cells corresponding to  $N_R$ , denoted by  $N_C$ . We confine ourselves to the practically more relevant 3D case.

As in the previous section,  $X \in \{0, 1\}^{|R|}$  denotes a random vector indicating subsets of projection rays.  $X_r = 1, r \in R$ , corresponds to a zero observation along ray  $r$ . For a subset of rays  $R_b \subset R$ , we say that the corresponding subset of cells  $C_b$  in (3.6) supports  $R_b$ .

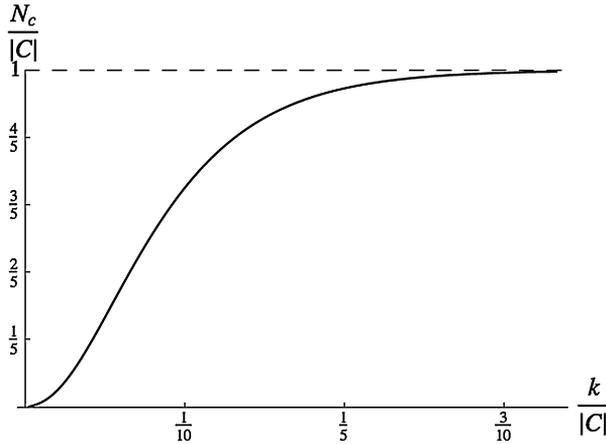
**Proposition 5.3.** For a given value of the sparsity parameter  $k$ , the expected size of subsets of cells that support random subsets  $R_b \subset R$  of observed non-zero measurements, is

$$N_C = N_C(k) = d^3 \left( 1 - 3 \left( 1 - \frac{1}{d^2} \right)^k + 3 \left( 1 - \frac{2d - 1}{d^3} \right)^k - \left( 1 - \frac{3d - 2}{d^3} \right)^k \right).
 \tag{5.18}$$

**Proof.** We partition the set of rays  $R = R_1 \cup R_2 \cup R_3$  according to the three projection images (Fig. 2) and associate with the cells  $C$  the corresponding set of triples of projection rays

$$R_{1,2,3} = \left\{ (r_1, r_2, r_3) : \bigcap_{i=1}^3 r_i \neq \emptyset, r_i \in R_i, i = 1, 2, 3 \right\},$$

with each triple intersecting in a single cell. Thus, we have  $|R_{1,2,3}| = |C| = d^3$ , and each cell  $c_{ijk}$  at the intersection of  $(r_i, r_j, r_k)$  belongs to the set  $C_b$  supporting  $R_b$  if  $(r_i \cup r_j \cup r_k) \subset R_b$ . In terms of random



**Fig. 7.** The expected number  $N_C = \mathbb{E}[|C_b|]$  of cells supporting observed measurement vectors  $b$ , given by (5.18). Starting with rate  $N_C \propto k$  for very small values of  $k$ , it quickly increases and exceeds  $N_R$  (Fig. 6), thus leading to underdetermined reduced systems (3.5).

variables  $X_r$  indicating zero measurements by  $X_r = 1$ , this means that  $c_{ijk} \in C_b$  if  $X_{r_i} = X_{r_j} = X_{r_k} = 0$ . Thus,

$$\begin{aligned}
 N_C &= \mathbb{E} \left[ \sum_{R_{1,2,3}} (1 - X_{r_1})(1 - X_{r_2})(1 - X_{r_3}) \right] \\
 &= \sum_{R_{1,2,3}} \left( 1 - (\mathbb{E}[X_{r_1}] + \mathbb{E}[X_{r_2}] + \mathbb{E}[X_{r_3}]) + \sum_{1 \leq i < j \leq 3} \mathbb{E}[X_{r_i} X_{r_j}] - \mathbb{E}[X_{r_1} X_{r_2} X_{r_3}] \right).
 \end{aligned}$$

This expression takes into account the intersection of projection rays  $r_i, r_j$  (inclusion–exclusion principle) in order not to overcount the number of supporting cells.

We have  $\mathbb{E}[X_{r_i}] = p_d^k = (1 - d^{-2})^k$  by (5.5) and (5.4). The event  $X_{r_i} X_{r_j} = 1$  means that both rays correspond to zero measurements, which happens with probability

$$\left( 1 - \frac{|r_i \cup r_j|}{|C|} \right)^k = \left( 1 - \frac{2d - 1}{d^3} \right)^k.$$

We have three pairs of sets of rays from  $R = R_1 \cup R_2 \cup R_3$ , and each of the  $d^2$  rays  $r_i \in R_i$  intersects with  $d$  rays  $r_j \in R_j$ . Finally, three intersecting rays correspond to zero measurements with probability

$$\left( 1 - \frac{|r_1 \cup r_2 \cup r_3|}{|C|} \right)^k = \left( 1 - \frac{3d - 2}{d^3} \right)^k,$$

for each of the  $d^3$  cells  $c \in C$ . □

**Remark 5.3.** Note that  $N_C$  specifies the expected value of  $n_{red}$  in (3.7) induced by random  $k$ -sparse vectors  $x \in \mathbb{R}_{k,+}^n$ . See Fig. 7 for an illustration.

*Bounding the deviation of  $N_C$ .* We wish to bound the deviation of the random number of cells supporting a random subset of observed non-zero measurements, from its expected value  $N_C$  given by (5.18). To this end, we set with  $\mathbf{X} = (X_1, \dots, X_r, \dots, X_m)$ ,  $r \in R$ ,  $|R| = m$ ,

$$f(\mathbf{X}) = \sum_{R_{1,2,3}} (1 - X_{r_1})(1 - X_{r_2})(1 - X_{r_3}) \tag{5.19}$$

such that  $\mu := \mathbb{E}[f(\mathbf{X})] = N_C$ . The objective is to apply [Theorem 2.2](#), based on the relevant quantities [\(2.7\)](#), [\(2.8\)](#) and [\(2.9\)](#). See [Proposition 5.8](#) below.

The set  $B$  [\(2.6\)](#) takes in our setting the specific form of a subset  $B \subseteq R$  of  $|B| = r - 1$  rays, partitioned  $B = B_1 \cup B_2 \cup B_3$  according to the three projection directions  $R = R_1 \cup R_2 \cup R_3$ . Recall that  $R_{1,2,3}$  denotes the set of all triplets of rays  $(r_1, r_2, r_3)$ , each corresponding to a cell  $c_{ijl}$ ,  $i, j, l \in [d]$ , in which they intersect. We denote<sup>4</sup> the characteristic function of a set  $B$  by

$$\mathbf{1}_B(r) = \begin{cases} 1, & r \in B, \\ 0, & r \notin B. \end{cases} \tag{5.20}$$

We identify  $B$  with the event that  $r - 1$  random variables  $X_1 = x_1, \dots, X_{r-1} = x_{r-1}$  have been observed, with  $x_i \in \{0, 1\}$ ,  $i \in [r - 1]$  indicating non-zero measurements ( $x_i = 0$ ) and a zero measurements ( $x_i = 1$ ), respectively.

We will use the shorthands for mutually intersecting rays  $r, r', r''$ ,

$$\begin{aligned} p_0 &:= 1 - p_1, \\ p_1 &:= \mathbb{E}[X_r] = p_d^k = \left(1 - \frac{1}{d^2}\right)^k, \\ p_2 &:= \mathbb{E}[X_r X_{r'}] = \left(1 - \frac{2d - 1}{d^3}\right)^k, \\ p_3 &:= \mathbb{E}[X_r X_{r'} X_{r''}] = \left(1 - \frac{3d - 2}{d^3}\right)^k. \end{aligned} \tag{5.21}$$

Note that

$$p_1 > p_2 > p_3. \tag{5.22}$$

Furthermore, summation over  $R_{1,2,3}$  is carried out by using double indices for rays as follows.

$$(r_1, r_2, r_3) \leftrightarrow c_{ijl} \Leftrightarrow r_1 \leftrightarrow jl, \quad r_2 \leftrightarrow il, \quad r_3 \leftrightarrow ij, \quad i, j, l \in [d]. \tag{5.23}$$

This just identifies each ray with one of the  $d^2$  pixels in the corresponding projection image.

**Lemma 5.4.** Consider an arbitrary event  $B$  of the form [\(2.6\)](#) and assume  $r = j_r l_r \in R_1$ . Then  $g(x)$  [\(2.7\)](#) is given by

$$\begin{aligned} g(x_r) &= \mathbb{E}[f(\mathbf{X})|X_r = x_r, B] - \mathbb{E}[f(\mathbf{X})|B] \\ &= \mathbb{E}\left[\sum_{R_{1,2,3}} \mathbf{1}_{\{r\}}(r_1)(X_{r_1} - x_{r_1})(\mathbf{1}_{B_2}(r_2)(X_{r_2} - x_{r_2}) + \mathbf{1}_{R_2}(r_2)(1 - X_{r_2})) \right. \\ &\quad \left. \times (\mathbf{1}_{B_3}(r_3)(X_{r_3} - x_{r_3}) + \mathbf{1}_{R_3}(r_3)(1 - X_{r_3}))\right], \end{aligned} \tag{5.24}$$

and the expression [\(2.8b\)](#) evaluates to

$$\begin{aligned} \text{ran}(x_1, \dots, x_{r-1}) &= |g(0) - g(1)| \\ &= \sum_{i \in [d]} \left[ \mathbf{1}_{B_2}(il_r) \mathbf{1}_{B_3}(ij_r) (x_{il_r} x_{ij_r} - (x_{il_r} + x_{ij_r}) p_1 + p_2) \right. \\ &\quad \left. - \mathbf{1}_{B_2}(il_r) (x_{il_r} - (1 + x_{il_r}) p_1 + p_2) - \mathbf{1}_{B_3}(ij_r) (x_{ij_r} - (1 + x_{ij_r}) p_1 + p_2) \right. \\ &\quad \left. + 1 - 2p_1 + p_2 \right]. \end{aligned} \tag{5.25}$$

<sup>4</sup> We economize notation here by re-using the symbol  $r$  indexing rays  $r \in R$ . Due to the context, there should be no danger of confusion with  $r = |B| + 1$ .

**Proof.** See [Appendix A](#). □

**Remark 5.4.** The assumption  $r \in R_1$  in [Lemma 5.4](#) is not essential due to the symmetry of the imaging set-up. Assuming  $r \in R_2$  instead, for instance, results in expressions (5.24), (5.25) with all quantities related to direction  $R_2$  replaced by those related to direction  $R_1$ .

We further analyze expression (5.25).

**Lemma 5.5.** Expression (5.25) has the form

$$\text{ran}(x_1, \dots, x_{r-1}) = \sum_{i \in [d]} s_i \tag{5.26}$$

with  $s_i = 0$  except for the cases:

(i) both rays  $il_r, ij_r$  are not contained in  $B$ , in which case

$$s_i = \varepsilon_2(d, k) := 1 - 2p_1 + p_2; \tag{5.27}$$

(ii) either ray  $r' = il_r$  or  $r'' = ij_r$  is contained in  $B$  and corresponds to a non-zero measurement, i.e. either  $x_{r'} = 0$  or  $x_{r''} = 0$ , in which case

$$s_i = \varepsilon_1(d, k) := 1 - p_1; \tag{5.28}$$

(iii) both rays  $il_r, ij_r$  are contained in  $B$  and correspond to non-zero measurements, in which case

$$s_i = 1. \tag{5.29}$$

**Proof.** See [Appendix A](#). □

We next consider an upper bound for  $\hat{r}^2$  defined by (2.9).

**Lemma 5.6.** For a given sparsity parameter  $k$ , assume

$$3k \leq sd^2, \quad s \leq 1, \tag{5.30}$$

and set

$$n_d = \max\{i \in \mathbb{N} : i \cdot d \leq 3k\}. \tag{5.31}$$

Then

$$\hat{r}^2 \leq 3p_0^2 d^4 + (1 - p_0^2)n_d d^2 \tag{5.32}$$

with  $p_0$  given by (5.21).

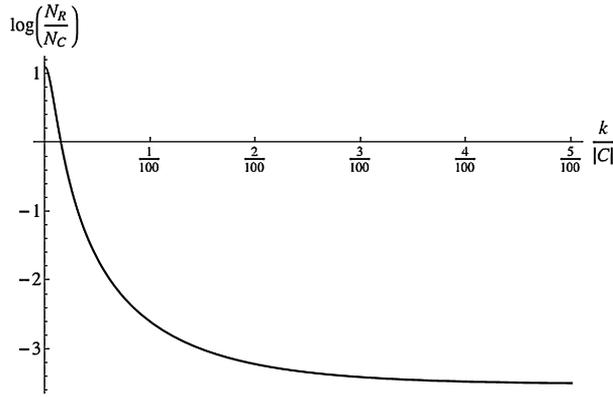
**Proof.** See [Appendix A](#). □

**Remark 5.5.** Assumption (5.30) is not restrictive from the viewpoint of applications. For example, the representative 3D case (5.9) discussed above yields  $s = 0.15$ .

We consider the last quantity needed to apply [Theorem 2.2](#).

**Lemma 5.7.** For the expression  $\max_{\text{dev}}$  defined by (2.9), it holds that

$$\max_{\text{dev}} = p_1 d = p_d^k d. \tag{5.33}$$



**Fig. 8.** The ratio  $\frac{m_{red}}{n_{red}}(k) = \frac{N_R(k)}{N_C(k)}$  of the expected dimensions of the reduced system, given by (5.6) and (5.18), is strictly monotonically decreasing with the sparsity parameter  $k$ .

**Proof.** See Appendix A. □

We finally state the desired result of this section.

**Proposition 5.8.** Let  $N_C = \mathbb{E}[|C_b|]$  be the expected number of cells supporting the non-zero measurements  $R_b \subset R$  for a given sparsity parameter  $k$ , given by (5.18). Then, for any  $\tau > 0$ ,

$$\Pr(|C_b| - N_C \geq \tau) \leq 2 \exp\left(-\frac{\tau^2}{2p_0\hat{r}^2(1 + \frac{\tau p_1 d}{3p_0\hat{r}^2})}\right), \tag{5.34}$$

with  $\hat{r}^2$  given by (5.32) and  $p_0, p_1$  given by (5.21).

**Proof.** Based on the preceding lemmata, apply Theorem 2.2 with  $p = p_0$  being the smaller probability of (5.5). □

We inspect the bound (5.34) for large problem sizes  $d \rightarrow \infty$  that occur in 3D applications. We then have

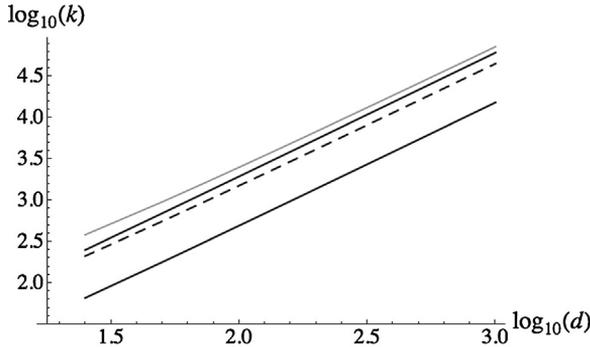
$$p_0 \approx kd^{-2}, \quad \hat{r}^2 \approx 3k^2 + 3kd\left(1 - \frac{k^2}{d^4}\right), \quad p_1 \approx 1 - kd^{-2}. \tag{5.35}$$

Setting  $k = s \cdot d^2$ ,  $s < 1$  (cf. (5.9)), we obtain for the denominator of (5.34) approximately  $6k(k^2 + (1 - s^2)kd)/d^2 + 2/3(1 - s)d\tau$ , which enables control via the sparsity parameter  $k$  and grows slower with  $\tau$  than the numerator of (5.34). Note that we do not claim this bound to be tight. We merely point out that concentration happens in principle.

As a result, Proposition (5.8) complements Proposition (5.2) concerning the concentration of the expected reduced system dimensions (3.7) on their expected values. This substantiates the validity “with high probability” of the results derived in this paper based on these expected values.

5.1.3. Average dimensions ratio of reduced systems

For small value of  $k$ , that is for highly sparse scenarios, the expected value  $N_R(k) \approx 3k$  grows faster than  $N_C(k) \approx k$ . Consequently, the expected reduced system due to Definition 3.5 will be overdetermined. This holds up to a critical value  $k \leq k_{crit}$  because for increasing values of  $k$  it is more likely that several particles are incident with some projection ray, making  $N_C$  increasing faster than  $N_R$ . Below this value reduced systems are underdetermined. Moreover, the average dimensions ratio  $N_R(k)/N_C(k)$  of the reduced system decreases with  $k$  – see Fig. 8.



**Fig. 9.** Critical upper bound sparsity values  $k = k(d)$  that guarantee unique recovery of  $k$ -sparse vectors  $x$  on average with high probability. From bottom to top:  $k_\delta$  (5.36b) for unperturbed matrices  $A$ ,  $k_{crit}$  (5.37) resulting in overdetermined reduced systems,  $k_{max}$  (5.39) for underdetermined perturbed matrices  $A$ , and fully random measurement matrices.

### 5.2. Unperturbed systems

We consider the recovery properties of the 3D setup depicted in Fig. 2, based on Theorem 3.5 and on the expected quantities involved in the corresponding condition (3.12), as worked out in Section 3.3. Concerning the interpretation of the following claims, we refer to Remark 5.1.

**Proposition 5.9.** *The unperturbed system  $Ax = b$ , with measurement matrix  $A$  given by (2.2), admits unique recovery of  $k$ -sparse nonnegative vectors  $x$  with high probability, if*

$$k \leq \frac{N_C(k_\delta)}{1 + \delta} = \frac{N_R(k_\delta)}{\ell}, \quad \delta = \frac{\sqrt{5} - 1}{2}, \tag{5.36a}$$

where  $k_\delta$  solves

$$N_R(k_\delta) = \ell \delta N_C(k_\delta) \tag{5.36b}$$

and  $N_R(k)$ ,  $N_C(k)$  are given by (5.6) and (5.18).

**Proof.** Recall, that  $\delta(1 + \delta) = 1$  holds. The assertion now follows from replacing the quantities forming condition (3.12) by their expected values, due to Remarks 5.2 and 5.3, and taking into account that  $N_R(k) \leq 3\delta N_C(k)$  for  $k \geq k_\delta$  and  $k_\delta$  solving (5.36b), due to the monotonicity of the average ratio  $N_R(k)/N_C(k)$  – compare Section 5.1.3.  $\square$

**Remark 5.6.** Eq. (5.36b) shows that unique recovery of a  $k$ -sparse,  $k \leq \frac{n_{red}}{(1+\delta)}$ , nonnegative vector can be expected using the unperturbed measurement matrix provided the reduced system (3.5) is by a factor  $m_{red} \geq 1.854n_{red}$  overdetermined. See Fig. 9 for an illustration. Moreover, for small values of  $k$ , each particle generates on average  $\ell$  non-zero measurements, and  $k_\delta \approx \frac{N_R(k_\delta)}{\ell}$ .

### 5.3. Overdetermined perturbed systems

Analogously to the previous section, we evaluate the average recovery performance using perturbed systems based on Theorem 3.6.

**Proposition 5.10.** *The perturbed system  $\tilde{A}x = \tilde{b}$ , admits unique recovery of  $k$ -sparse nonnegative vectors  $x$  with high probability, if  $k$  satisfies condition  $k \leq k_{crit}$ , where  $k_{crit}$  solves*

$$N_R(k_{crit}) = N_C(k_{crit}) \tag{5.37}$$

and  $N_R(k_{crit})$ ,  $N_C(k_{crit})$  are given by (5.6) and (5.18).

**Proof.** Immediate from the monotonicity of the average ratio  $N_R(k)/N_C(k)$  and [Theorem 3.6](#), replacing the quantities forming condition [\(3.13\)](#) by their expected values, and taking into account  $\ell = 3$  for the measurement matrix [\(2.2\)](#) and the case  $D = 3$ .  $\square$

**Remark 5.7.** In view of this assertion and [Remark 5.6](#), it is remarkable that a significant gain of recovery performance can be obtained by a simple device: structure-preserving perturbation of the measurement matrix. See [Fig. 9](#) for an illustration.

5.4. Underdetermined perturbed systems

Unlike [Propositions 5.9](#) and [5.10](#), we specifically consider here less sparse scenarios that result in underdetermined reduced systems [\(3.5\)](#).

**Proposition 5.11.** Let  $\tilde{A}$  be a perturbation of  $A$  from [\(2.2\)](#) satisfying the assumptions of [Lemma 3.3](#) with complete rank  $r_0(\tilde{A}) =: \tilde{r}_0 = N_R(\tilde{k}_{max})$ , where  $\tilde{k}_{max}$  solves

$$N_R(\tilde{k}_{max}) = \delta N_C(\tilde{k}_{max}), \quad \delta = \frac{\sqrt{5} - 1}{2}, \tag{5.38}$$

with  $N_R(k)$ ,  $N_C(k)$  given by [\(5.6\)](#) and [\(5.18\)](#). Then a  $k$ -sparse nonnegative vector  $x$  can be uniquely recovered via  $\tilde{A}$  with high probability, if

$$k \leq k_{max} := \left\lceil \frac{N_R(\tilde{k}_{max})}{\ell} - 1 \right\rceil. \tag{5.39}$$

**Proof.** By assumption and [Lemma 3.3](#), [Theorem 3.2](#) (see also [Remark 3.1](#)) implies [\(5.39\)](#), thereby taking into account that  $\ell = 3$  for the measurement matrix [\(2.2\)](#) and the case  $D = 3$ .  $\square$

[Fig. 9](#) illustrates the value  $k_{max}$  [\(5.39\)](#) and compares it to the previous results.

Finally, we comment on the uniqueness condition established in [\[14\]](#) which corresponds to the top  $k(d)$  curve in [Fig. 9](#). This result does not apply to our setting. The reason is that a basic assumption underlying the application of [\(2.12\)](#) does not hold. While after some perturbation the points corresponding to the columns of  $\tilde{A}$  and the sparsity value  $|I^-(x)| = k$  are in general position, the underlying distribution lacks symmetry with respect to the origin. As a result, we cannot establish the superior performance of “fully” random sensors considered in [\[14\]](#).

5.5. Two cameras are not enough

In the present section, we briefly discuss how the previously obtained bounds on sparsity apply in the 2D scenario. To this end, we first compute the expected value of nonempty cells connected to  $R_b$  measurements generated by a  $k$ -sparse nonnegative vector.

**Proposition 5.12.** In 2D, the expected size of subsets of cells that support random subsets  $R_b \subset R$  of observed non-zero measurements, is

$$N_C = N_C(k) = d^2 \left( 1 - \left( 1 - \frac{1}{d} \right)^k \right)^2, \tag{5.40}$$

for a given sparsity parameter  $k$ ,

**Proof.** We partition the set of rays  $R = R_1 \cup R_2$  according to the two projection images ([Fig. 2](#)), left, and associate with the cells  $C$  the corresponding set of pairs of projection rays

$$R_{1,2} = \left\{ (r_1, r_2) : \bigcap_{i=1}^2 r_i \neq \emptyset, r_i \in R_i, i = 1, 2 \right\},$$

with each pair intersecting in a single cell. Thus, we have  $|R_{1,2}| = |C| = d^2$ , and each cell  $c_{ij}$  at the intersection of  $(r_i, r_j)$  belongs to the set  $C_b$  supporting  $R_b$  if  $(r_i \cup r_j) \subset R_b$ . In terms of random variables  $X_r$  indicating zero measurements by  $X_r = 1$ , this means that  $c_{ij} \in C_b$  if  $X_{r_i} = X_{r_j} = 0$ . Thus,

$$N_C = \mathbb{E} \left[ \sum_{R_{1,2}} (1 - X_{r_1})(1 - X_{r_2}) \right] \\ = \sum_{R_{1,2}} \left( 1 - (\mathbb{E}[X_{r_1}] + \mathbb{E}[X_{r_2}]) + \sum_{1 \leq i < j \leq 2} \mathbb{E}[X_{r_i} X_{r_j}] \right),$$

taking the intersection of projection rays  $r_i, r_j$  into account. We obtained  $\mathbb{E}[X_{r_i}] = p_d^k = (1 - \frac{1}{d})^k$  in (5.5) and (5.4). The event that both rays correspond to zero measurements  $X_{r_i} X_{r_j} = 1$  happens with probability

$$\left( 1 - \frac{|r_i \cup r_j|}{|C|} \right)^k = \left( 1 - \frac{2d - 1}{d^2} \right)^k = \left( 1 - \frac{1}{d} \right)^{2k}. \quad \square$$

By Proposition 5.12 and Lemma 5.1 we can now compute the expected ratio of the dimensions of the reduced system, further denoted by  $c$ . We solve the polynomial  $N_R(k) = cN_C(k)$  according to (5.6a) and (5.40). Interesting are the values  $c \in \{2\delta, 1, \delta, \frac{1}{2}\}$ , with  $\delta = \frac{\sqrt{5}-1}{2}$ . For example, if  $c = 2\delta$ , we obtain guaranteed recovery of all 1-sparse vectors, which also equals the strong threshold for the 2D case. If  $c = 1$  we obtain that a  $k$ -sparse nonnegative vector  $x$ , with

$$k \leq k_{crit} = \frac{\log(\frac{d-2}{d})}{\log(\frac{d-1}{d})} \approx 2, \tag{5.41}$$

induces on average reduced overdetermined systems. Thus two particles can always be reconstructed, after perturbation. If  $c = \frac{1}{2}$  the critical sparsity value approximately equals 4 for arbitrary  $d$ . This is the best achievable bound, which is obviously useless for applications. For  $k = 3$  it can be shown that the probability of correct recovery via the perturbed matrix  $A_d^2$  is

$$1 - \frac{2 \cdot 4 \cdot \binom{d}{2} \binom{d}{3} + 4 \cdot \binom{d}{3}^2}{\binom{d^3}{3}} = \frac{d^2 + 6d - 10}{3(d^2 - 2)} \xrightarrow{d \rightarrow \infty} 1/3.$$

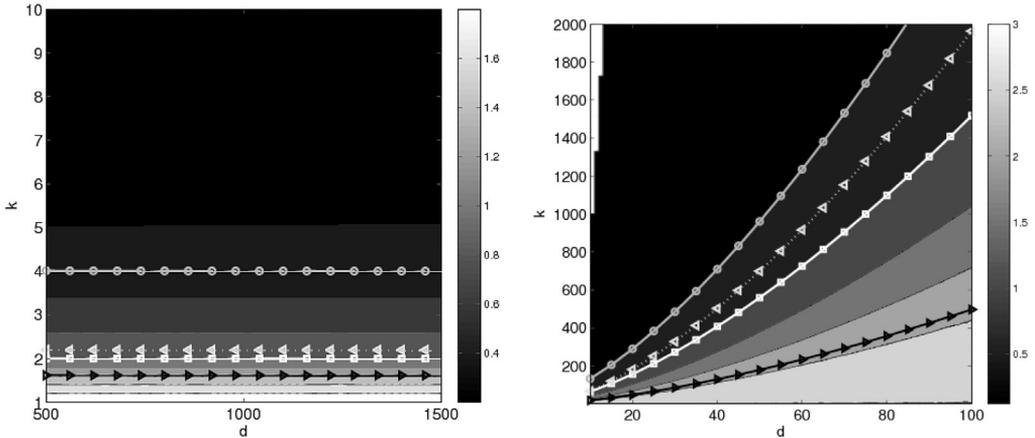
We mention that the value of  $N_R$  and  $N_C$  does not vary much with different 2 camera arrangements. These highly pessimistic results may be explained by the fact that there is no expander with constant left degree  $\ell$  less than 3.

### 6. Numerical experiments and discussion

In this section we empirically investigate bounds on the required sparsity that guarantee unique nonnegative or binary  $k$ -sparse solutions and compare them to the analytically derived bounds on sparsity from previous sections.

#### 6.1. Reduced systems versus analytical sparsity thresholds

The workhorse of the previous theoretical average case performance analysis of the discrete tomography matrix from (4.1) is the derivation of the expected number of non-zero rows  $N_R(k)$  induced by the  $k$ -sparse nonnegative vector along with the number  $N_C(k)$  of “active” cells which cannot be empty. This can be done also empirically, see Fig. 10, left, for the 2D case and right, for the 3D case. To generate the figures we varied  $k \in \{1, 2, \dots, 10\}$  and  $d \in \{500, 501, \dots, 1500\}$  in 2D and  $k \in \{1, 2, \dots, 2000\}$  and  $d \in \{10, 11, \dots, 100\}$  in 3D, and generated for each point  $(k, d)$  500 problem instances. The plots



**Fig. 10.** The contourplots of the average fraction of the reduced systems as a function of the resolution parameter  $d$  and the sparsity parameter  $k$ . **Left:** The plots for the 2D case shows that level lines of  $N_R(k, d)/N_C(k, d)$  are constant with varying  $d$ . **Right:** In 3D the situation dramatically changes. Higher sparsity values are allowed for increasing values of  $d$ , as the derived threshold curves show. Under the  $\triangleright$ -marked curve  $k_\delta$  (5.36b) reconstruction for unperturbed systems is guaranteed with high probability. Under the  $\square$ -marked curve  $k_{crit}$  (5.37) reduced systems are overdetermined. For points under the dotted  $\triangleleft$ -marked curve  $k_{max}$  (5.39) reconstruction is guaranteed for perturbed systems. And finally, problem instances under the  $\circ$ -marked curve  $k_{opt}$  (6.1) would be reconstructible if the reduced matrices would follow a symmetrical distribution with respect to the origin.

show  $N_R(k, d)/N_C(k, d)$  along with the curves:  $k_\delta$  (5.36b) for unperturbed matrices  $A$ ,  $k_{crit}$  (5.37) resulting in overdetermined reduced systems,  $k_{max}$  (5.39) for underdetermined perturbed matrices  $A$ , and  $k_{opt}$  which solves

$$N_R(k_{opt}) = 0.5N_C(k_{opt}). \tag{6.1}$$

Throughout this section we set  $\delta = \frac{\sqrt{5}-1}{2}$ .

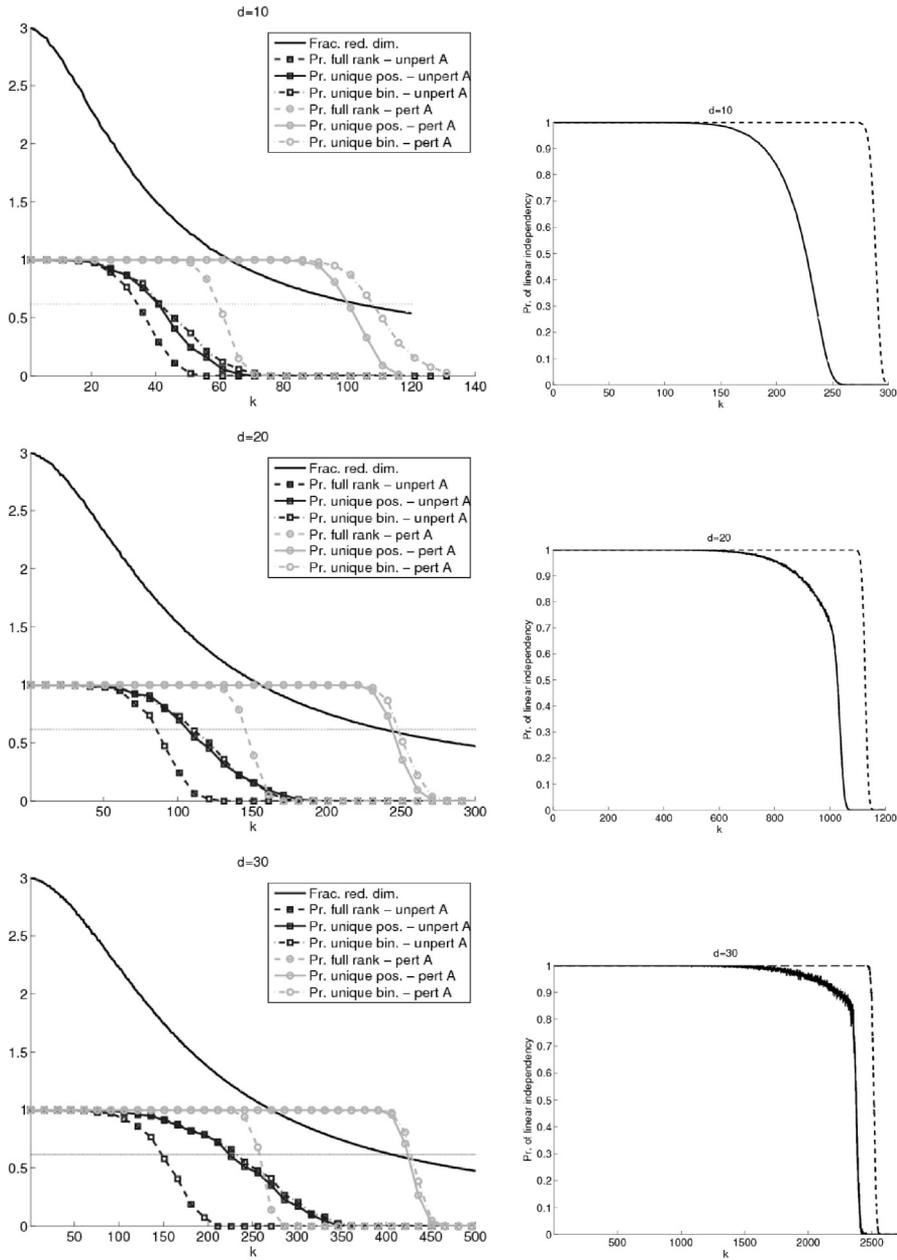
### 6.2. Empirical phase transitions

We further concentrate on the 3D case. In analogy to [10] we assess the so called *phase transition*  $\rho$  as a function of  $d$ , which is reciprocally proportional to the undersampling ratio  $\frac{m}{n} \in (0, 1)$ . We consider  $d \in \{10, 11, \dots, 100\}$ , the corresponding matrix  $A_d^3 \in \mathbb{R}^{3d^2 \times d^3}$  from (4.1) and its perturbed version  $\tilde{A}$  and the sparsity as a fraction of  $d^2$ ,  $k = \rho d^2$ , for  $\rho \in (0, 1)$ .

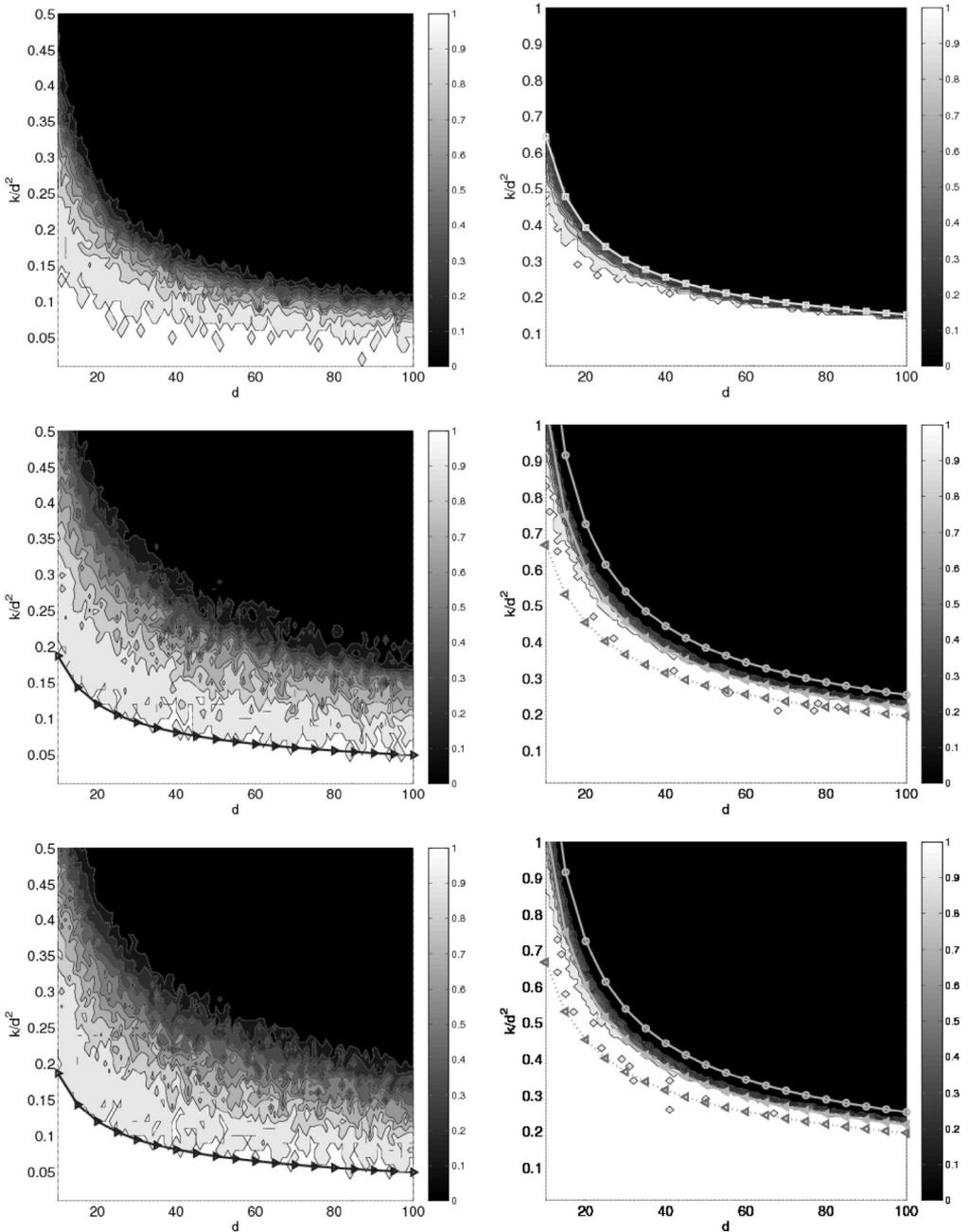
This phase transition  $\rho(d)$  indicates the necessary relative sparsity to recover a  $k$ -sparse solution with overwhelming probability. More precisely, if  $\|x\|_0 \leq \rho(d) \cdot d^2$ , then with overwhelming probability a random  $k$ -sparse nonnegative (or binary) vector  $x^*$  is the unique solution in  $\mathcal{S}_+ := \{x: Ax = Ax^*, x \geq 0\}$  or  $\mathcal{S}_{[0,1]} := \{x: Ax = Ax^*, x \in [0, 1]^n\}$ , respectively. Uniqueness can be “verified” by minimizing and maximizing the same objective  $f^T x$  over  $\mathcal{S}_+$  or  $\mathcal{S}_{[0,1]}$ , respectively. If the minimizers coincide for several random vectors  $f$  we claim uniqueness. As shown in Fig. 11 the threshold for a unique nonnegative solution and a unique 0/1-bounded solution are quite close.

To generate the success and failure transition plots we generated  $A$  according to (4.1) and  $\tilde{A}$  by slightly perturbing its entries and varying  $d \in \{10, 11, \dots, 100\}$ .  $\tilde{A}$  has the same sparsity structure as  $A$ , but with random entries drawn from the standard uniform distribution on the open interval  $(0.9, 1.1)$ . We tried different perturbation levels, all leading to similar results. Thus we adopted this interval for all presented results.

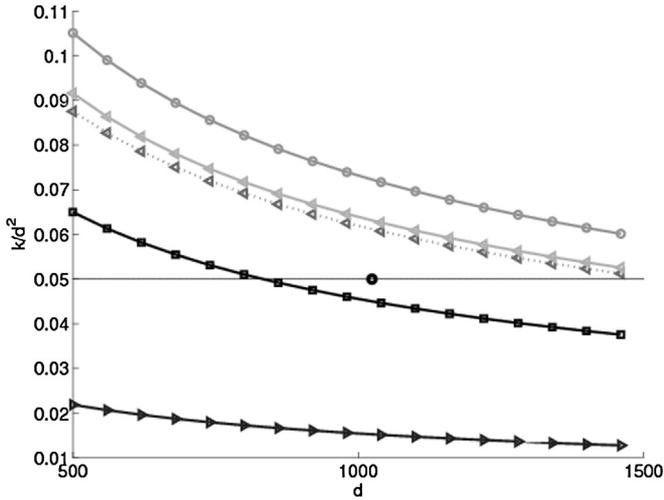
Then for  $\rho \in [0, 1]$  a  $\rho d^2$ -sparse nonnegative or binary vector was generated to compute the right hand side measurement vector and for each  $(d, \rho)$ -point 50 random problem instances were generated. A threshold effect is clearly visible in all figures (see Fig. 12) exhibiting parameter regions where the probability of exact reconstruction is close to one and it is much stronger for the perturbed systems. The results are in excellent agreement with the derived analytical thresholds. We refer to the



**Fig. 11. Left:** Recovery via the unperturbed matrix  $A_d^3$  (dark gray  $\square$ -marked curves),  $d \in \{10, 20, 30\}$  (from top to down) versus the perturbed counterpart (light gray  $\circ$ -marked curves). The dash-dot line depicts the empirical probability (500 trials) that reduced systems are overdetermined and of full rank. The solid line (dark gray: unperturbed, light gray: perturbed) shows the probability that a  $k$ -sparse nonnegative vector is unique. The dashed curve shows the probability that a  $k$ -sparse binary solution is the unique solution in  $[0, 1]^p$ . Additional information like binarity gives only a slight performance boost. The curve  $k_S$  (5.36b) correctly predicts that 18 ( $d=10$ ), 48 ( $d=20$ ), and 85 ( $d=30$ ) particles are reconstructed with high probability via the unperturbed systems and 66 ( $d=10$ ), 181 ( $d=20$ ), 328 ( $d=30$ ) particles, via the perturbed systems according to  $k_{max}$  (5.39). However, 105 ( $d=10$ ), 241 ( $d=20$ ), 408 ( $d=30$ ), by  $k_{max}$  from (5.38) are more accurate. Division by three doesn't seem necessary. **Right:** Empirical probability obtained from 10000 trials that  $k$  random columns of the unperturbed matrix (solid black line) or of the perturbed matrix (dashed black line) are linearly independent.



**Fig. 12.** **Left:** Success and failure empirical phase transitions  $\rho(d)$  for unperturbed and perturbed systems **right. Top:** Probability that the reduced matrices are overdetermined and of full rank, along **(right)** with the estimated relative critical sparsity level  $k_{crit}$  ( $\square$ -marked curve) which induces overdetermined reduced matrices. **Middle:** Probability of uniqueness of a  $k = \rho(d) \cdot d^2$ -sparse nonnegative vector. **Bottom:** Probability of uniqueness in  $[0, 1]^n$  of a  $k = \rho(d) \cdot d^2$ -sparse binary vector. The  $\triangleright$ -marked curve depicts again  $k_s/d^2$  (5.36b), the  $\square$ -marked curve  $k_{crit}/d^2$  (5.37), the dotted  $\triangleleft$ -marked curve  $k_{max}/d^2$  (5.39), the solid  $\triangleleft$ -marked curve  $\hat{k}_{max}/d^2$  (5.39) and the  $\circ$ -marked curve  $k_{opt}/d^2$  (6.1). In case of the perturbed matrix  $\tilde{A}$  exact recovery is possible beyond overdetermined reduced matrices. Moreover,  $\hat{k}_{max}$  follows most accurately the empirical phase transition for perturbed systems.



**Fig. 13.** Relative critical upper bound sparsity values  $k(d)/d^2$  in the practical relevant domain  $d \in (500, 1500)$  that guarantee unique recovery of  $k$ -sparse nonnegative vectors  $x$  on average with high probability. From bottom to top:  $k_\delta/d^2$  (5.36b) for unperturbed matrices  $A$  ( $\triangleright$ -marked),  $k_{crit}/d^2$  (5.37) resulting in overdetermined reduced systems ( $\square$ -marked line),  $k_{max}/d^2$  (5.39) and  $\tilde{k}_{max}/d^2$  (5.38) for square perturbed matrices  $A$  (dotted and solid  $\triangleleft$ -marked lines), and ideal random measurement matrices  $k_{opt}/d^2$  ( $\circ$ -marked line). The thin black line depicts the particle density used by engineers in practice, while the black spot corresponds to the typical resolution parameter  $d = 1024$ . The results demonstrate that specific slight random perturbations of the TomoPIV measurement matrix considerably boost the expected reconstruction performance by at least 150%.

figure captions for detailed explanations. Finally, we refer to the summary in Fig. 13 for the computed sharp sparsity thresholds, which accurately follow the empirical thresholds on sparsity.

### 7. Conclusions

The main contribution of this work is the transfer of recent results in compressive sensing via expander graphs with bad expansion properties to the discrete tomography problem. In particular, we consider a sparse binary measurement matrix, which encodes the incidence relation between projection rays and image discretization cells, along with its slightly perturbed counterpart. While the expected expansion of the underlying graph does not change with perturbation, the recovery performance can be boosted significantly. We investigate the average performance in recovery of exact sparse nonnegative signals by analyzing the properties of reduced systems obtained by eliminating zero measurements and related redundant discretization cells. We compute sharp sparsity thresholds, such that the maximal sparsity can be determined precisely for both perturbed and unperturbed scenarios. Our theoretical analysis suggests that a similar procedure can be applied to different geometries.

### Acknowledgements

We would like to thank the anonymous reviewers for their very detailed and insightful comments on an earlier draft of the present manuscript. The first author gratefully acknowledges financial support from the Ministry of Science, Research and Arts, Baden-Württemberg within the Margarete von Wrangell postdoctoral lecture qualification program.

### Appendix A

**Proof of Lemma 5.4.** Using the definition (5.19) of  $f$ , we get

$$\mathbb{E}[f(\mathbf{X})|B] = \mathbb{E}\left[\sum_{R_{1,2,3}} \prod_{i \in [3]} (\mathbf{1}_{B_i}(r_i)(1 - x_{r_i}) + \mathbf{1}_{R_i \setminus B_i}(r_i)(1 - X_{r_i}))\right]. \tag{A.1}$$

Thus, we compute (2.7) by factoring out the contribution of ray  $r$  in the first factor so as to subtract  $\mathbb{E}[f(\mathbf{X})|B]$ .

$$\begin{aligned}
 g(x_r) &= \mathbb{E}[f(\mathbf{X})|X_r = x_r, B] - \mathbb{E}[f(\mathbf{X})|B] \\
 &= \mathbb{E}\left[ \sum_{R_{1,2,3}} (\mathbf{1}_{B_1 \cup r}(r_1)(1 - x_{r_1}) + \mathbf{1}_{R_1 \setminus (B_1 \cup r)}(r_1)(1 - X_{r_1})) \right. \\
 &\quad \times \left. \prod_{i \in \{2,3\}} (\mathbf{1}_{B_i}(r_i)(1 - x_{r_i}) + \mathbf{1}_{R_i \setminus B_i}(r_i)(1 - X_{r_i})) \right] - \mathbb{E}[f(\mathbf{X})|B] \\
 &= \mathbb{E}\left[ \sum_{R_{1,2,3}} (\mathbf{1}_{B_1}(r_1)(1 - x_{r_1}) + \mathbf{1}_{\{r\}}(r_1)((1 - x_{r_1}) - (1 - X_{r_1})) + \mathbf{1}_{(R_1 \setminus B_1) \cup r}(r_1)(1 - X_{r_1})) \right. \\
 &\quad \times \left. \prod_{i \in \{2,3\}} (\mathbf{1}_{B_i}(r_i)(1 - x_{r_i}) + \mathbf{1}_{R_i \setminus B_i}(r_i)(1 - X_{r_i})) \right] - \mathbb{E}[f(\mathbf{X})|B] \\
 &= \mathbb{E}\left[ \sum_{R_{1,2,3}} \mathbf{1}_{\{r\}}(r_1)(X_{r_1} - x_{r_1})(\mathbf{1}_{B_2}(r_2)(1 - x_{r_2}) + \mathbf{1}_{R_2 \setminus B_2}(r_2)(1 - X_{r_2})) \right. \\
 &\quad \times \left. (\mathbf{1}_{B_3}(r_3)(1 - x_{r_3}) + \mathbf{1}_{R_3 \setminus B_3}(r_3)(1 - X_{r_3})) \right] \\
 &= \mathbb{E}\left[ \sum_{R_{1,2,3}} \mathbf{1}_{\{r\}}(r_1)(X_{r_1} - x_{r_1})(\mathbf{1}_{B_2}(r_2)(X_{r_2} - x_{r_2}) + \mathbf{1}_{R_2}(r_2)(1 - X_{r_2})) \right. \\
 &\quad \times \left. (\mathbf{1}_{B_3}(r_3)(X_{r_3} - x_{r_3}) + \mathbf{1}_{R_3}(r_3)(1 - X_{r_3})) \right]. \tag{A.2}
 \end{aligned}$$

To compute (2.8b), we use (5.24) and obtain

$$\begin{aligned}
 \text{ran}(x_1, \dots, x_{r-1}) &= |g(0) - g(1)| = g(0) - g(1) \\
 &= \mathbb{E}\left[ \sum_{R_{1,2,3}} \mathbf{1}_{\{r\}}(r_1)(\mathbf{1}_{B_2}(r_2)(X_{r_2} - x_{r_2}) + \mathbf{1}_{R_2}(r_2)(1 - X_{r_2})) \right. \\
 &\quad \times \left. (\mathbf{1}_{B_3}(r_3)(X_3 - x_{r_3}) + \mathbf{1}_{R_3}(r_3)(1 - X_{r_3})) \right] \\
 &= \mathbb{E}\left[ \sum_{i \in [d]} (\mathbf{1}_{B_2}(il_r)(X_{il_r} - x_{il_r}) + (1 - X_{il_r}))(\mathbf{1}_{B_3}(ij_r)(X_{ij_r} - x_{ij_r}) + (1 - X_{ij_r})) \right] \\
 &= \sum_{i \in [d]} [\mathbf{1}_{B_2}(il_r)\mathbf{1}_{B_3}(ij_r)(x_{il_r}x_{ij_r} - (x_{il_r} + x_{ij_r})p_1 + p_2) - \mathbf{1}_{B_2}(il_r)(x_{il_r} - (1 + x_{il_r})p_1 + p_2) \\
 &\quad - \mathbf{1}_{B_3}(ij_r)(x_{ij_r} - (1 + x_{ij_r})p_1 + p_2) + 1 - 2p_1 + p_2]. \quad \square \tag{A.3}
 \end{aligned}$$

**Proof of Lemma 5.5.** Observe that the term  $\mathbf{1}_{B_2}(il_r)$  selects those rays  $il_r \in R_2$  that are contained in  $B_2$  and intersect  $r = j_r l_r \in R_1$ , and similar for  $\mathbf{1}_{B_3}(ij_r)$ . Thus, given  $r = j_r l_r$  and two rays  $il_r, ij_r$ , the summand  $s_i$  of Eq. (5.25) takes the following values.

- $\mathbf{1}_{B_2}(il_r) = 0, \mathbf{1}_{B_3}(ij_r) = 0$ :  $s_i = 1 - 2p_1 + p_2$ .

- $\mathbf{1}_{B_2}(il_r) = 1, \mathbf{1}_{B_3}(ij_r) = 0: s_i = \begin{cases} 1 - p_1, & x_{il_r} = 0, \\ 0, & x_{il_r} = 1. \end{cases}$   
 And similarly for  $\mathbf{1}_{B_2}(il_r) = 0, \mathbf{1}_{B_3}(ij_r) = 1$ .
- $\mathbf{1}_{B_2}(il_r) = 1, \mathbf{1}_{B_3}(ij_r) = 1: s_i = \begin{cases} 1, & x_{il_r} = 0, x_{ij_r} = 0, \\ 0, & \text{otherwise.} \end{cases} \quad \square$

**Proof of Lemma 5.6.** We have  $\text{ran}(x_1, \dots, x_{r-1}) \leq d, \forall r$ , due to (5.26) and (5.29). Because  $R^2$  (2.9) is given by the sum of squares of such terms, we look for a configuration of rays represented by  $\mathbf{x} = (x_1, \dots, x_m) \in \{0, 1\}^m$  such that a maximal number of these terms attain the upper bound  $d$ .

For a given sparsity parameter  $k$ , there cannot be more than  $3k$  active rays with non-zero measurements among the set  $R$  of all  $m = 3d^2 = |R|$  projection rays. We thus look for a configuration of these active rays such that for a maximal number of rays  $r, r \in [m]$ , case (iii) of Lemma 5.5 applies, for  $i \in [d]$ , which implies  $\text{ran}(x_1, \dots, x_{r-1}) = d$ . Case (iii) means that ray  $r$  meets the maximal number  $d$  of cells, each incident with two active rays in  $B = [r - 1]$  from the other two projection directions.

A short reflection of the imaging set-up reveals that  $d$  active rays from each two directions (i.e.  $2d$  active rays in total) are needed to obtain a first such event for some single ray  $r$ , and that each  $d$  further active rays enable another such event for some ray  $r' \neq r$ . Thus, we consider a sequence  $x_1, \dots, x_m$  that includes the upper bound  $(n_d + 1) \cdot d > 3k$  of active rays and correspondingly  $n_d$  further rays  $r \in [m]$  with  $\text{ran}(x_1, \dots, x_{r-1}) = d$ . For all remaining  $m - n_d$  rays  $r$ , case (iii) cannot apply. So we choose case (ii) (5.28) for these rays yielding a larger summand  $s_i = 1 - p_1 > 1 - p_1 - (p_1 - p_2)$ ,  $i \in [d]$ , than for case (i) (5.27).

Summing up yields the upper bound

$$\hat{r}^2 = \max_{\mathbf{x} \in \mathcal{X}} \sum_{r \in [m]} (\text{ran}(x_1, \dots, x_{r-1}))^2 \leq n_d d^2 + (3d^2 - n_d)(1 - p_1)^2 d^2 \tag{A.4}$$

which equals (5.32).  $\square$

**Proof of Lemma 5.7.** In view of (2.8a), we compute

$$\begin{aligned} |g(0)| &= g(0) = \mathbb{E} \left[ \sum_{R_{1,2,3}} \mathbf{1}_{\{r\}}(r_1) X_{r_1} (\mathbf{1}_{B_2}(r_2)(X_{r_2} - x_{r_2}) + \mathbf{1}_{R_2}(r_2)(1 - X_{r_2})) \right. \\ &\quad \left. \times (\mathbf{1}_{B_3}(r_3)(X_{r_3} - x_{r_3}) + \mathbf{1}_{R_3}(r_3)(1 - X_{r_3})) \right] \\ &= \mathbb{E} \left[ \sum_{i \in [d]} X_{j_r l_r} (\mathbf{1}_{B_2}(il_r)(X_{il_r} - x_{il_r}) + (1 - X_{il_r})) (\mathbf{1}_{B_3}(ij_r)(X_{ij_r} - x_{ij_r}) + (1 - X_{ij_r})) \right] \\ &= \sum_{i \in [d]} [\mathbf{1}_{B_2}(il_r) \mathbf{1}_{B_3}(ij_r) (x_{il_r} x_{ij_r} p_1 - (x_{il_r} + x_{ij_r}) p_2 + p_3) \\ &\quad - \mathbf{1}_{B_2}(il_r) (x_{il_r} p_1 - (1 + x_{il_r}) p_2 + p_3) - \mathbf{1}_{B_3}(ij_r) (x_{ij_r} p_1 - (1 + x_{ij_r}) p_2 + p_3) \\ &\quad + p_1 - 2p_2 + p_3] \tag{A.5} \end{aligned}$$

and

$$\begin{aligned} |g(1)| &= -g(1) = \text{ran}(x_1, \dots, x_{r-1}) - g(0) \\ &= \sum_{i \in [d]} [\mathbf{1}_{B_2}(il_r) \mathbf{1}_{B_3}(ij_r) (x_{il_r} x_{ij_r} (1 - p_1) - (x_{il_r} + x_{ij_r})(p_1 - p_2) + p_2 - p_3) \\ &\quad - \mathbf{1}_{B_2}(il_r) (x_{il_r} (1 - p_1) - (1 + x_{il_r})(p_1 - p_2) + p_2 - p_3) \\ &\quad - \mathbf{1}_{B_3}(ij_r) (x_{ij_r} (1 - p_1) - (1 + x_{ij_r})(p_1 - p_2) + p_2 - p_3) \\ &\quad + 1 - p_1 - 2(p_1 - p_2) + p_2 - p_3]. \tag{A.6} \end{aligned}$$

Both expression have the form of (5.25). The coefficients only differ. Thus, the reasoning proving Lemma 5.5 applies, for every  $r \in [m]$ , and the corresponding case (iii) applied to the larger expression  $|g(0)|$  yields the assertion.  $\square$

## References

- [1] K. Azuma, Weighted sums of certain dependent random variables, *Tohoku Math. J.* 19 (3) (1967) 357–367.
- [2] R. Berinde, P. Indyk, Sparse recovery using sparse random matrices, MIT-CSAIL Technical Report, 2008.
- [3] E. Candès, Compressive sampling, in: *Int. Congress of Math.*, vol. 3, Madrid, Spain, 2006.
- [4] E. Candès, M. Rudelson, T. Tao, R. Vershynin, Error correcting via linear programming, in: *46th Ann. IEEE Symp. Found. Computer Science, FOCS'05*, 2005, pp. 295–308.
- [5] T.M. Cover, Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, *IEEE Trans. Electron. Comput.* EC-14 (3) (1965) 326–334.
- [6] A. DasGupta, *Asymptotic Theory of Statistics and Probability*, Springer, 2008.
- [7] L. Devroye, L. Györfi, G. Lugosi, *A Probabilistic Theory of Pattern Recognition*, Springer, 1996.
- [8] D.L. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory* 52 (2006) 1289–1306.
- [9] D.L. Donoho, For most large underdetermined systems of equations, the minimum  $\ell_1$ -norm near-solution approximates the sparsest near-solution, *Comm. Pure Appl. Math.* 59 (7) (2006) 907–934.
- [10] D.L. Donoho, J. Tanner, Sparse nonnegative solution of underdetermined linear equations by linear programming, *Proc. Natl. Acad. Sci. USA* 102 (27) (2005) 9446–9451.
- [11] D.L. Donoho, J. Tanner, Counting the faces of randomly-projected hypercubes and orthants, with applications, *Discrete Comput. Geom.* 43 (3) (2010) 522–541.
- [12] G.E. Elsinga, F. Scarano, B. Wieneke, B.W. van Oudheusden, Tomographic particle image velocimetry, *Exp. Fluids* 41 (2006).
- [13] M.A. Khajehnejad, A.G. Dimakis, W. Xu, B. Hassibi, Sparse recovery of nonnegative signals with minimal expansion, *IEEE Trans. Signal Process.* 59 (2011) 196–208.
- [14] O.L. Mangasarian, B. Recht, Probability of unique integer solution to a system of linear equations, *European J. Oper. Res.* 214 (1) (2011) 27–30.
- [15] C. McDiarmid, Concentration, in: *Probabilistic Methods for Algorithmic Discrete Mathematics*, Springer, 1998, pp. 1–46.
- [16] S. Petra, C. Schnörr, TomoPIV meets compressed sensing, *Pure Math. Appl. (PU.M.A.)* 20 (1–2) (2009) 49–76.
- [17] M. Stojnic,  $l_1$  optimization and its various thresholds in compressed sensing, in: *ICASSP*, 2010, pp. 3910–3913.
- [18] M. Stojnic, Recovery thresholds for  $\ell_1$  optimization in binary compressed sensing, in: *ISIT*, 2010, pp. 1593–1597.
- [19] V.N. Vapnik, A.Y. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, *Theory Probab. Appl.* 16 (1971) 264–280.
- [20] M. Wang, W. Xu, A. Tang, A unique “nonnegative” solution to an underdetermined system: From vectors to matrices, *IEEE Trans. Signal Process.* 59 (3) (2011) 1007–1016.
- [21] J.G. Wendel, A problem in geometric probability, *Math. Scand.* 11 (1962) 109–111.
- [22] W. Xu, B. Hassibi, Efficient compressive sensing with deterministic guarantees using expander graphs, in: *Proc. 2007 IEEE Information Theory Workshop, ITW '07*, 2007, pp. 414–419.