TomoPIV meets Compressed Sensing

Stefania Petra[§] and Christoph Schnörr[§]

[§] Image and Pattern Analysis Group, University of Heidelberg, Germany; e-mail: {petra, schnoerr}@math.uni-heidelberg.de

Abstract

We study the discrete tomography problem in Experimental Fluid Dynamics – Tomographic Particle Image Velocimetry (TomoPIV) – from the viewpoint of compressed sensing (CS). The CS theory of recoverability and stability of sparse solutions to underdetermined linear inverse problems has rapidly evolved during the last years. We show that all currently available CS concepts predict an extremely poor worst case performance, and a low expected performance of the TomoPIV measurement system, indicating why low particle densities only are currently used by engineers in practice. Simulations demonstrate however that slight random perturbations of the TomoPIV measurement matrix considerably boost both worst-case and expected reconstruction performance. This finding is interesting for CS theory and for the design of TomoPIV measurement systems in practice.

AMS Subject Classifications: 65F22, 68U10

Keywords: compressed sensing, underdetermined systems of linear equations, positivity constraints in ill-posed problems, sparsest solution, TomoPIV

1 Introduction

1.1 TomoPIV

Our research work is motivated by [21]. The authors introduced a new 3D technique, called *Tomographic Particle Image Velocimetry (TomoPIV)* for imaging turbulent fluids with high speed cameras. The technique is based on the instantaneous reconstructions of particle volume functions from few and simultaneous projections (2D images) of tracer particles within the fluid. The reconstruction of the 3D image from 2D images employs a standard algebraic reconstruction algorithm [22].

TomoPIV can use only few projections due to both limited optical access to wind and water tunnels and cost and complexity of the necessary measurement apparatus. As a consequence, the reconstruction problem becomes severely ill-posed, and both the mathematical analysis and the design of algorithms fundamentally differ from the standard scenarios of medical imaging.

A crucial parameter for 3D fluid flow estimation from image measurements is particle density. This parameter also largely influences the tomographical reconstruction problem. Higher densities ease subsequent flow estimation and increase the resolution and measurement accuracy. However, higher densities also aggravate ill-posedness of the reconstruction problem. A thorough investigation of this trade-off is lacking. Our objective is to address this problem taking into account relevant developments in applied mathematics.

TomoPIV adopts a simple discretized model for an image-reconstruction problem known as the *algebraic image reconstruction* model [11], which assumes that the image consists of an array of unknowns (voxels), and sets up algebraic equations for the unknowns in terms of measured projection data. The latter are the pixel entries in the recorded 2D images that represent the integration of the 3D light intensity distribution $I(z)$ along the pixels line-of-sight L_i obtained from a calibration procedure. Thus, the i-th measurement obeys

$$b_i :\approx \int_{L_i} I(z)dz \approx \sum_{j=1}^{n} x_j \int_{L_i} \mathcal{B}_j(z)dz = \sum_{j=1}^{n} x_j a_{ij} \,,$$

where a_{ij} is the value of the i-th pixel if the object to be reconstructed is the j-th basis function. The values a_{ij} depend on the choice of the basis function. Typically, $\mathcal{B}_j$ are cube-shaped uniform basis functions, the classical *voxels*. For simplicity we will adopt this discretization scheme and stress that other choices are possible, see e.g. [27].

The main task is to estimate the weights x_j from the recorded 2D images, corresponding to basis functions and solve $Ax \approx b$. The matrix A has dimensions $(\#\text{ pixel} =: m) \times (\#\text{ basis functions} = n)$, where $m \ll n$. Since each row indicates those basis functions whose support intersect with the corresponding projection ray the projection matrix A will be sparse.

1.2 Compressed Sensing

We study the tomographic problem of reconstructing particle volume functions from the general viewpoint of *Compressed Sensing*, which is a central theme of current research in applied mathematics. Compressed Sensing [9, 10, 17] is a new technique for acquiring a signal $x^* \in \mathbb{R}^n$ by incomplete linear measurement

$$Ax = b \,, \tag{1}$$

where $A \in \mathbb{R}^{m\times n}$, $m < n$, and for reconstructing x^* *exactly* provided that the signal is sparse (or compressible in some basis), i.e. $\|x^*\|_0 := \big|\{i \,|\, x_i^* \neq 0\}\big| \ll n$.

Instead of considering the NP-hard ℓ_0-minimization problem

$$\min \|x\|_0 \quad \text{s.t. } Ax = b \,, \tag{2}$$

it considers the convex ℓ_1-minimization problem

$$\min \|x\|_1 \quad \text{s.t. } Ax = b \,, \tag{3}$$

and investigates the situations when the same x^* solve both problems (2) and (3), coined as ℓ_0/ℓ_1-equivalence.

A remarkable result of Candès and Tao [10] is that if, for example, the rows of A are randomly chosen Gaussian distributed vectors, there is a constant C such that if the signal sparsity level obeys $\|x^*\|_0 < Cm/\log(\frac{n}{m})$, the solution of (3) will be exactly the original signal x^* with overwhelming probability.

In fact, random measurement matrices are optimal [8, 19, 29, 2] in the sense that for a given sparsity level k, the required number of samples is minimal such that ℓ_0/ℓ_1-equivalence holds. On the other hand, for a given number of measurements m the sparsity level k of x^* which allows recovery by ℓ_1-minimization is maximal. The different derivations of ℓ_0/ℓ_1-equivalence are quite involved and are based on the notion of *Restricted Isometry Property (RIP)* [2], see Section 4.3, or on "counting faces" of polytopes [15, 19, 16].

When the solution is known to be sparse and positive then, under a similar assumption on A, k and m, all nonnegative k-sparse vectors x^* are the unique positive solution of $Ax = Ax^*$, [6, 19].

Donoho and Tanner [16, 19] have computed sharp reconstruction thresholds for Gaussian measurements, such that for any choice of sparsity k and signal size n, the required number of measurements m to recover x^* can be determined precisely.

Recent trends [3, 4, 24] tend to replace random dense matrices by adjacency matrices of "high quality" unbalanced expander graphs. Here, the measurement matrices A are sparse binary matrices.

1.3 Stylized Problem

Likewise, we will concentrate on a particular binary measurement matrix. We consider a 3D image I with a cubic domain V discretized in d^3 voxels. Three cameras, with d^2 pixels (L_i rays) each, keep the volume under investigation in focus from three orthogonal directions, compare Fig. 1.3 (left). According to Section 1.1 each entry of the measurement matrix A will be

$$a_{ij} = \int_{L_i} \mathcal{B}_j(x)dz = 1 \ ,$$

if the line of sight L_i of the i-th pixel intersects the j-th voxel $\mathcal{B}_j$, or $a_{ij} = 0$ otherwise. By numbering voxels and pixels according to Fig. 1.3 (left) matrix A can be written in closed form as

$$A = \begin{pmatrix} \mathbf{1}_d^\top \otimes I_d \otimes I_d \\ I_d \otimes \mathbf{1}_d^\top \otimes I_d \\ I_d \otimes I_d \otimes \mathbf{1}_d^\top \end{pmatrix} \ , \tag{4}$$

where $\otimes$ denotes the Kronecker product, see [23]. Notice that A is the adjacency matrix of a bipartite graph with regular left degree 3 and regular right degree d, compare Fig. 1.3 (right). The $n := d^3$ left nodes correspond to the voxels in the d^3 cube and thus to the entries of x. The $m := 3d^2$ right nodes (or measurement nodes) correspond to the camera pixel. In a bipartite graph connections within the variable nodes and within the right nodes do not occur. The existing edges between the left nodes and right nodes are represented by our $m \times n$ matrix A from (4). In particular,

$$a_{ij} = \begin{cases} 1, & \text{if } j\text{-th ray intersects } i\text{-th voxel}, \\ 0, & \text{otherwise}, \end{cases}$$

for all $i \in \{1, \ldots, m\}$, $j \in \{1, \ldots, n\}$.

Throughout this paper we denote by x^* the indicator vector corresponding to the original particle distribution and assume that our measurements based on the sampling matrix A are exact, i.e. $b = Ax^*$. Moreover, we assume that x^* is *positive* and *sparse*.

We investigate the sparsity level of x^* up to which the sparsest solution of $Ax = Ax^*$ is unique. Furthermore, we are interested in recovering x^* as minimizer of the ℓ_1-minimization problem (3) or, as minimizer of the linear program

$$\min \mathbf{1}^\top x \ , \quad Ax = b, x \geq 0 \ . \tag{5}$$

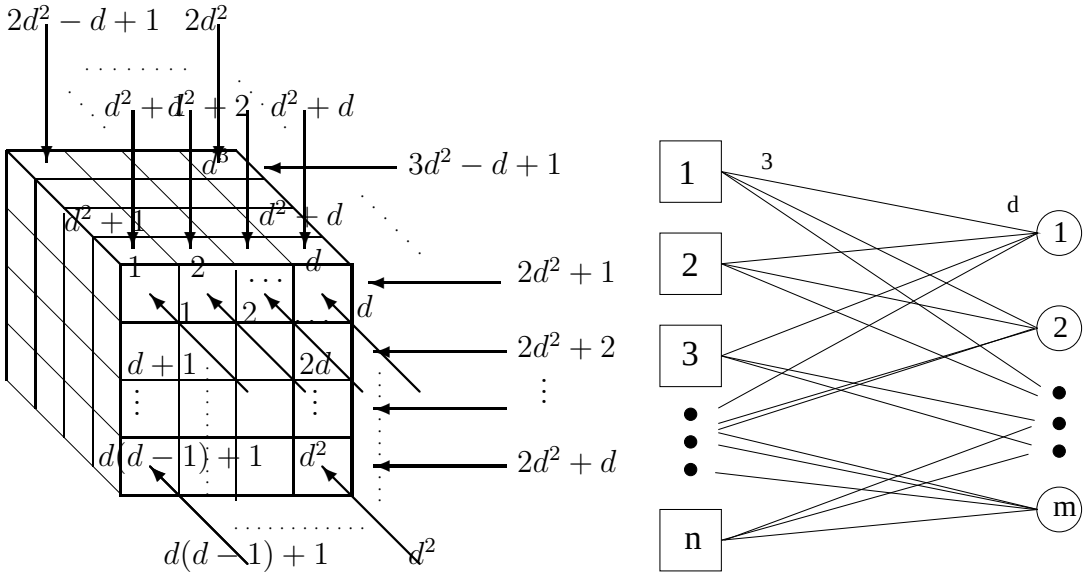

Figure 1: **Left:** Discretization of the $d \times d \times d$ volume and corresponding $3d^2$ rays for the 3 orthogonal projections. **Right:** A is the adjacency matrix of a bipartite graph with regular left degree 3 and regular right degree d.

1.4 Contribution and Organization

We provide a detailed study of the TomoPIV problem from the viewpoint of compressed sensing. We assess the worst-case and average performance of this severely ill-posed reconstruction problem of discrete tomography, based on convex ℓ_1-regularization and on a range of recently established theoretical results.

The critical parameter both in theory and in practice is the particle density of the imaged fluid, that in mathematical terms corresponds to the sparsity of the vector to be reconstructed from observed measurements. Of particular interest are phase transitions of this parameter below of which unique reconstructions can be assumed to hold in practice – an essential requirement for subsequent processing steps for, e.g., estimating fluid flow velocity from a sequence of reconstructed volume functions. On the other hand, using as large as possible particle densities is important in practice too, in order to improve the space-time resolution of observed fluid structures.

After establishing basic properties of the measurement matrix (4) in Section 2, we clarify in Sections 3 and 4 the relationship between the regularized reconstruction problems (2), (3) and (5) and assess the worst-case and average performance by applying recently established results from the theory of compressed sensing to the TomoPIV problem. Taking into account that sparse volume functions generate sparse observations, we provide in Section 5 a probabilistic analysis of TomoPIV reconstructions based on systems (1) that have been *reduced* accordingly in a preprocessing step. Finally, we discuss in Section 6 the statistics of numerical simulations based on slightly and randomly perturbed measurement matrices A.

In a nutshell, we show that the TomoPIV problem is quite degenerate from the viewpoint of compressed sensing, thus leading to poor performance guarantees (Sections 3, 4). On the other hand, the probabilistic analysis of Section 5 yields average performance bounds that back up current rules of thumb of engineers for choosing particle densities in practice. Finally, Section 6 indicates a dramatic performance boost based on only slightly modified measurement systems,

raising novel problems for theory and implications for the improved design of real TomoPIV measurement systems.

While Section 3 is mainly based on established theoretical concepts, all remaining sections – and Section 3 too – contain novel material from the specific viewpoint of TomoPIV and also from the more general viewpoint of discrete tomography. In particular, our papers aims at pointing out connections between the fields of compressed sensing and discrete tomography in order to stimulate further research.

1.5 Notation

$|X|$ denotes the cardinality of a finite set X. We will denote by $\|\cdot\|$ and $\|\cdot\|_1$ the Euclidean ℓ_2-norm and the ℓ_1-norm in the n-dimensional real vector space $\mathbb{R}^n$. We already introduced the pseudo-norm $\|x\|_0 = |\{i \,|\, x_i \neq 0\}|$ and denote the set of k-sparse vectors by $\mathbb{R}^n_k = \{x \in \mathbb{R}^n \,|\, \|x\|_0 \leq k\}$. The support of a vector $x \in \mathbb{R}^n$, $\mathrm{supp}(x) \subseteq \{1, 2, \ldots, n\}$, denotes the set of indices of nonvanishing 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 $\mathrm{supp}(x) = I^+(x) \cup I^-(x)$ and $\|x\|_0 = |\mathrm{supp}(x)|$.

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

$A_{\bullet,i}$ 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.

2 Preliminaries

The aim of this section is an examination of the properties of the system (1) for this simple prototype of data-collection geometry. Such properties will be also relevant for other regular imaging geometries, e.g. when additionally using a fourth camera (projection direction).

By the nature of the problem the coefficient matrix A is very sparse, in contrast to most compressed sensing measurement ensembles. This together with the sparsity of the original signal x^* induces a sparsity also in the measurement vector b which in more classic scenarios is not given. As a consequence, we can remove equations with zero right-hand side leading us to a feasible set of reduced dimensionality as will be detailed next.

Consider the feasible polyhedral set with respect to A and b

$$\mathcal{F} := \{x \mid Ax = b, x \geq 0\}\,, \tag{6}$$

where all entries a_{ij} in A are nonnegative. Let us introduce the following partitions of the right and left nodes

$$\begin{aligned} I &:= I^0(b) = \{i \in \{1, \ldots, m\} \mid b_i = 0\} && \text{and} && I^c\,, \\ J &:= \mathcal{N}(I) = \{j \in \{1, \ldots, n\} \mid \exists i \in I : a_{ij} > 0\} && \text{and} && J^c\,. \end{aligned}$$

Further define

$$\mathcal{F}_{red} := \{x \mid A_{I^c J^c} x = b_{I^c}, x \geq 0\}. \tag{7}$$

Then we can make the simple, compare [27, Prop.1], but important observation.

Proposition 2.1. *Let $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ have all nonnegative entries and $\mathcal{F}$ and $\mathcal{F}_{red}$ be defined as in* (6) *and* (7) *respectively. Then*

$$\mathcal{F} = \{x \in \mathbb{R}^n \mid x_J = 0 \text{ and } x_{J^c} \in \mathcal{F}_{red}\}. \tag{8}$$

Remark 2.1. Assume that for a particular measurement vector b, which induces the partitions I, I^c and J, J^c of the right and left nodes as defined above, we obtained an overdetermined and full rank submatrix $A_{I^cJ^c}$. Then the vector $x^*_{J^c}$ is the unique solution of $A_{I^cJ^c}x = b_{I^c}$ and $x^* \in \mathbb{R}^n$, where $x^*_J = 0$, is the unique positive solution of $Ax = b$.

Clearly, when the above situation occurs solving the ℓ_0-problem (2) amounts to solve a feasibility problem. Moreover, any method which solves

$$\min_{x\in\mathcal{F}} f(x)$$

for an arbitrary objective function f will lead to the same *correct* result.

Let us assume for the time being that we have a sufficiently sparse vector x^* and a sufficiently sparse measurement vector $b = Ax^*$ such that $A_{I^cJ^c}$ is overdetermined, i.e.

$$|\mathcal{N}(I^0(b))| - |I^0(b)| \geq n - m\ .$$

The rank of $A_{I^cJ^c}$ will equal the rank of A_{J^c}, while the latter cannot be full if it contains a subset of linearly dependent columns.

This observation motivates us to find an upper bound to the maximal number s of columns such that all s (or less) column combinations are linearly independent. A useful tool for achieving this task, which is in general of combinatorial nature, is to investigate the null space of A.

Proposition 2.2. *[26, Prop. 2.2] Let $d \in \mathbb{N}$, $d \geq 3$, A from* (4) *and $N \in \mathbb{R}^{d^3\times(d-1)^3}$ defined as*

$$N := \begin{pmatrix} -\mathbf{1}^\top_{d-1} \\ I_{d-1} \end{pmatrix} \otimes \begin{pmatrix} -\mathbf{1}^\top_{d-1} \\ I_{d-1} \end{pmatrix} \otimes \begin{pmatrix} -\mathbf{1}^\top_{d-1} \\ I_{d-1} \end{pmatrix}. \tag{9}$$

Then the following statements hold

(a) $AN = 0$, *with A from* (4).

(b) Every column in N has exactly 8 *nonzero elements.*

(c) N is a full rank matrix and $\operatorname{rank}(N) = (d-1)^3$.

(d) $\operatorname{rank}(A) = 3d^2 - 3d + 1$.

(e) $\ker(A) = span\{N\}$, *i.e. the columns of N provide a basis for the null space of A.*

(f) $\sum_{i=1}^n v_i = 0$ *holds for all* $v \in \ker(A)$.

3 Unique Sparsest Solution

In order to study ℓ_0/ℓ_1-equivalence for A from (4) we decompose this problem in two separate conditions:

1. ℓ_0-unique-optimality: x^* is the unique optimum of (2) ;
2. ℓ_1-unique-optimality: x^* is the unique optimum of (3) .

In this section we investigate the first subproblem, while the second one will be addressed in the next section.

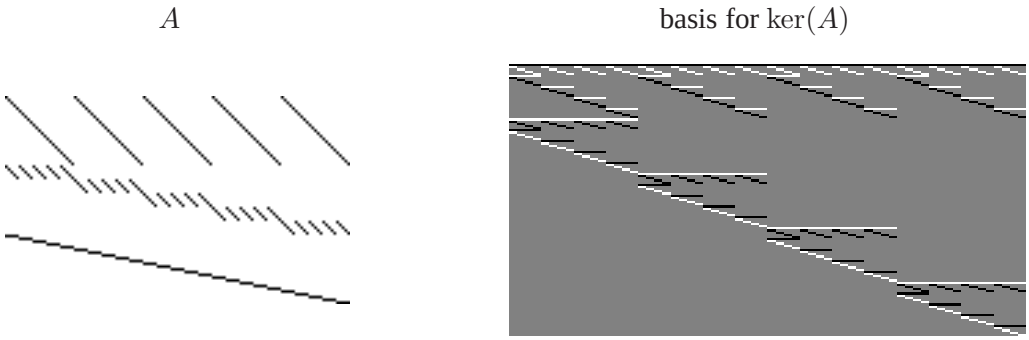

Figure 2: Matrix A from (4) for $d = 5$ (left) along with a *sparse* basis for its null space, the columns of N from (9) (right).

3.1 Spark

Besides being one of the classical NP-hard problems, see [25] for this NP-hardness result, problem (2) has a highly nonconvex objective function and thus many local optima may occur. Fortunately previous work has shown that if a sparse enough solution to (2) exists than it will be necessarily unique. The analysis in [20] involves the measure $\mathrm{spark}(A)$ which equals the minimal number of linearly dependent columns of A, see [18, 20]. In contrast to $\mathrm{rank}(A)$, $\mathrm{spark}(A)$ is NP-hard to compute. Fortunately bounds on this measures can be derived, see [18] and Section 4.2.

The following result is surprisingly elementary and can be found in [18].

Theorem 3.1. *(Uniqueness) Let x^* be a solution of* (1) *with $\|x^*\|_0 < \frac{\mathrm{spark}(A)}{2}$. Then x^* is the unique solution of* (2)*.*

Clearly, $2 \leq \mathrm{spark}(A) \leq \mathrm{rank}(A) + 1$. Again, Gaussian matrices $A \in \mathbb{R}^{m\times n}$, $m < n$, are optimal in the sense that $\mathrm{spark}(A)$ is maximal and equals $\mathrm{rank}(A)+1 = m+1$. Unfortunately, with A from (4) we come off badly.

Proposition 3.2. *[26, Prop. 3.2] For all $d \in \mathbb{N}, d \geq 3$ the minimal number of linearly dependent columns of matrix A from* (4) *equals* 8*, i.e.* $\mathrm{spark}(A) = 8$*.*

Hence, Thm. 3.1 and Prop. 3.2 yield *guaranteed* uniqueness of *every* 3-sparse vector x^* only. This bound is tight, since we can construct two 4-sparse solutions x^1 and x^2 such that $Ax^1 = Ax^2$, compare Fig. 3.

3.2 Signature

In [20] Elad adopts a probabilistic point of view to study uniqueness of sparse solutions of (2) beyond the worst-case scenario based on the *signature* of a matrix $A \in \mathbb{R}^{m\times n}$. This is defined as the discrete function $\mathrm{sig}_A(k) \in [0, 1]$, for $k \in \{2, \ldots, n\}$, that equals the number of k column combinations in A which are *linearly dependent* divided by the number of all k columns from the n existing ones. By definition $\mathrm{sig}_A(k) = 0$, for all $k < \mathrm{spark}(A)$.

Theorem 3.3. *[20, Thm. 6,Thm. 5] Let $\sigma := \mathrm{spark}(A) \leq \mathrm{rank}(A) =: r$ and x^* be a solution $Ax = b$. Assume the locations of the nonzero entries in x^* are chosen at random with equal and independent probability. If $1/2\sigma \leq \|x^*\|_0 =: k \leq r$, then the probability that x^* is the*

sparsest solution of $Ax = b$ is approximately $1 - \mathrm{sig}_A(k)$ and the probability to find a solution of $Ax = b$ of the same cardinality k is approximately

(a) $\sum_{j=0}^{k-\sigma}(k-j)(n-k+j)\begin{pmatrix} k \\ j \end{pmatrix}\mathrm{sig}_A(k-j)$ *or lower, if* $\|x^*\|_0 \geq \sigma$*;*

(b) 0, if $1/2\sigma \leq \|x^*\|_0 < \sigma$.

An upper bound on the signature was derived via arguments from matroid theory [5], under the assumption that the spark is known.

Theorem 3.4. *[20, Th. 7] Let $A \in \mathbb{R}^{m\times n}$ with the signature function sig_A, $\mathrm{spark}(A) = \sigma$ and $\mathrm{rank}(A) = r$. Then*

$$\mathrm{sig}_A(k) \leq 1 - \frac{\sum_{i=1}^{\sigma-1}\binom{n-r+i-1}{i}\binom{r-i}{k-i}}{\binom{n}{k}}, \quad 0 \leq k \leq r.$$

The computation of the signature seems even harder than computing the spark. However, the signature will be close to zero for k small enough, but growing with the dimension of A. If $spark(A) = 8$ it does not necessarily mean that every 8 or more column combination are linearly dependent. In fact, only a limited number of k column combinations can be dependent without violating $\mathrm{rank}(A) = 3d^2 - 3d + 1$. It turns out that this number is tiny for smaller k when compared to $\binom{n}{k}$. As k increases this number also grows and equals one only when $k > r$, compare Fig. 6.1 and Fig. 7 (left). Numerical experiments suggest that most $0.9d^2$ column combinations in A are linearly independent.

4 Unique Positive Solution

This section might seem useless from a practical point of view since ℓ_0/ℓ_1-equivalence cannot be claimed *for all* k-sparse vectors when k exceeds 3 in view of the nonuniqueness of the ℓ_0-minimizer in this case. However, we trace relevant conditions yielding ℓ_0/ℓ_1-equivalence, review known connections between different concepts and establish some new ones.

4.1 Relations between problems (3) and (5)

Most Compressed Sensing results explore conditions under which the minimum of the ℓ_1-minimization problem (3) is unique (and identical to the ℓ_0-minimization problem (2)). We note in this section that if the measurement matrix A has equal column sum and if a positive solution x^* to $Ax = b$ exists, then a unique ℓ_1-minimizer must equal x^*. Conversely, if the solution of (5) x^* is unique then also the ℓ_1-minimizer must be unique.

Proposition 4.1. *Assume there is a positive solution x^* to $Ax = b$, with A from (4) and let x^1 be the unique solution of the ℓ_1-minimization problem (3). Then x^1 must equal x^*.*

Proof. Denote by x^1 a (unique) solution to the ℓ_1-minimization problem (3). In view of Prop. 2.2 (f), $\mathbf{1}^\top x = \mathbf{1}^\top b/3$ holds for all solutions x of $Ax = b$. Thus we obtain

$$\mathbf{1}^\top x^* = \mathbf{1}^\top x^1 \leq \|x^1\|_1 \leq \mathbf{1}^\top x^* \,,$$

where the last inequality holds since x^* is feasible. Thus equality must hold. □

On the other hand, we have

Proposition 4.2. *Consider A from* (4) *and assume that the positive solution x^* to $Ax = b$ is unique. Then x^* will be also the unique minimizer of the ℓ_1-minimization problem* (3).

Proof. Note that if x^* is the unique minimizer of (5) then x^* is necessarily k-sparse for some $k < n$. Otherwise, it cannot be unique since for x^* with no vanishing entries $x^* + tv$ will also solve (5) for t sufficiently small and $v \in \ker(A)$. Hence $S := I^0(x^*) \neq \emptyset$. Moreover, $S \cap I^-(v) \neq \emptyset$ or $S^c \cap I^-(v) \neq \emptyset$ hold for all $v \in \ker(A) \setminus \{0\}$, in view of the uniqueness of x^*. From $\sum_{i\in S} v_i + \sum_{i\in S^c} v_i = 0$ we now obtain

$$|\sum_{i\in S^c} \operatorname{sign}(x_i^*)v_i| = |\sum_{i\in S^c} v_i| < \sum_{i\in S} |v_i| \,,$$

for all $v \in \ker(A) \setminus \{0\}$. This is a well known condition for the uniqueness of the ℓ_1-minimizer, see e.g. [28]. □

Note, that the above results hold for all matrices A with equal column sum.

4.2 Mutual incoherence

The mutual coherence of a matrix A, denoted by $\mu(A)$, is defined as the maximal absolute scalar product between two different normalized columns of A,

$$\mu(A) = \max_{\substack{i,j \\ i\neq j}} \frac{\langle A_{\bullet,i}, A_{\bullet,j}\rangle}{\|A_{\bullet,i}\|\|A_{\bullet,j}\|} \,. \tag{10}$$

It measures the similarity between the matrix's columns. For an orthogonal matrix A, $\mu(A) = 0$. For an $m < n$ we necessarily have $\mu(A) > 0$. Uniqueness of the sparsest solution and exact recovery of x^* via ℓ_1-minimization can be guaranteed [18] if

$$\|x^*\|_0 \leq 0.5\left(1 + \frac{1}{\mu(A)}\right) \,.$$

Hence, there is an interest in matrices with $\mu(A)$ as small as possible. $\mu(A) = 1$ implies the existence of two "parallel" columns, and this causes confusion in the construction of a sparse representation of the measurement vector b. In [30] it was shown that for a full rank matrix of size $m \times n$

$$\mu(A) \geq \sqrt{\frac{n-m}{m(n-1)}}$$

and equality is obtained for a family of matrices called *Grassmanian frames*.

The mutual coherence is often used to lower bound the spark, since the following relationship always holds

$$\operatorname{spark}(A) \geq 1 + \frac{1}{\mu(A)} \,.$$

In [6], nonnegativity is taken into account. Here a one-sided coherence is considered

$$\nu(A) = \max_{\substack{i,j \\ i\neq j}} \frac{\langle A_{\bullet,i}, A_{\bullet,j}\rangle}{\|A_{\bullet,i}\|^2} \,. \tag{11}$$

The authors obtained the following result.

Theorem 4.3. *[6, Thm. 2] Let $A \in \mathbb{R}^{m\times n}$ be a matrix with nonnegative entries such that all solutions of $Ax = b$ satisfy $\mathbf{1}^\top x = c$, where c is some constant. If there is a* nonnegative *sparse solution x^* to this system with $\|x^*\|_0 < 0.5(1 + \frac{1}{\nu(A)})$, then it is a unique solution of this problem.*

For our particular matrix A we obtain

Proposition 4.4. *For all $d \in \mathbb{N}, d \geq 3$ and A defined in* (4)

$$\mu(A) = \nu(A) = \frac{1}{3} .$$

Proof. Since every column contains exactly 3 ones we obtain $\|A_{\bullet,i}\|^2 = 3$ for all $i \in \{1, ..., n\}$. Thus $\mu(A) = \nu(A)$. Since two different voxels can both be intersected by at most one ray the maximal common support length of two different columns is one. This immediately implies the result. □

However, recovery bounds based on this bound are too pessimistic since, due to the above result we obtain guaranteed recovery via ℓ_1-minimization *for all* k-sparse vectors if $k < 2$, which is (needless to say) useless. Derivation of stronger results that refer to specific matrices and bypass the use of the mutual coherence should be attempted.

4.3 Restricted Isometry Property

In order to prove that there exist matrices A with only $m = O(k \log(n/k))$ rows such that for all k-sparse x ℓ_0/ℓ_1-equivalence holds, Candès and Tao [9] introduced a concept that outranks the coherence measure $\mu(A)$. A matrix A is said to have the *Restricted Isometry Property* $RIP_{2,k,\delta}$ if for any k-sparse vector x, the following expression is verified

$$(1 - \delta)\|x\|^2 \leq \|Ax\|^2 \leq (1 + \delta)\|x\|^2 , \quad \delta \in (0, 1) . \tag{12}$$

This property implies that every submatrix A_S formed by combining at most k-columns in A has its nonzero singular values bounded above by $1 + \delta$ and below by $1 - \delta$. In particular, (12) implies that a matrix A cannot satisfy $RIP_{2,k,\delta}$ if $k \geq \operatorname{spark}(A)$.

Candès has shown, see [7, Thm. 1.1], that if $A \in RIP_{2,2k,\delta}$ with $\delta < \sqrt{2} - 1$ all $x \in \mathbb{R}^n_k$ solve both (2) and (3). Moreover, there exist sensing matrices $A \in \mathbb{R}^{m\times n}$ which satisfy e.g. the $RIP_{2,k,1/4}$, where k can be as large as $O(m/\log(m/n))$. This class includes matrices with i.i.d. standard Gaussian or ± 1 entries, random submatrices of the Fourier transform or other orthogonal matrices.

It has been shown recently [12] that binary matrices cannot satisfy $RIP_{2,k,\delta}$, unless the numbers of rows is $\Omega(k^2)$. Note that the best known explicit construction of (binary) compressed sensing matrices due to DeVore [14] yields $\Omega(k^2)$ measurements, which is worse than the bound $m = O(k \log(n/k))$.

Theorem 4.5. *[12, Thm. 1] Let $A \in \mathbb{R}^{m\times n}$ be any $0/1$-matrix that satisfies $RIP_{2,k,\delta}$. Then*

$$m \geq \min\left\{ \left(\frac{1-\delta}{1+\delta}\right)^2 k^2, \frac{1-\delta}{1+\delta}\, n \right\} .$$

For our particular A defined in (4) with $\operatorname{spark}(A) = 8$ we therefore obtain taking $m = 3d^2$ into account

Corollary 4.6. *Let $\delta \in (0,1)$. Then a necessary condition for A to satisfy the $RIP_{2,k,\delta}$ for all k-sparse vectors is that*

$$k \leq \min\left\{\frac{\sqrt{3}}{2}\cdot\frac{1+\delta}{1-\delta}\,d, 7\right\}\ .$$

However, we cannot be sure that A possesses the $RIP_{2,7,\sqrt{2}-1}$, unless we compute all singular values of all A_S, $|S| \leq 7$ matrices. In case of a positive result we obtain ℓ_0/ℓ_1-equivalence for *all* less then 3-sparse particle distributions, even in case of observation errors, since RIP also implies *stable* recovery, provided that the observation error is small enough, compare [7, Thm. 1.2] for the "noisy" version result.

4.4 Binary Matrices with $RIP_{1,k,\delta}$

In [3] it is shown that a particular class of *binary* measurement matrices $A \in \{0,1\}^{m\times m}$, namely adjacency matrices of expander graphs, see the following Def. 4.1, satisfy a different form of restricted isometry property, the so-called $RIP_{1,k,\delta}$

$$\forall x \in \mathbb{R}^n_k\ , \qquad (1-\delta)\|x\|_1 \leq \|Ax\|_1 = (1+\delta)\|x\|_1\ , \quad \delta \in (0,1)\ . \tag{13}$$

Definition 4.1. A (k,ϵ)-unbalanced expander is a bipartite simple graph $G = (A, B, E)$ with left degree ℓ such that for any $X \subset A$ with $|X| \leq k$, the set of neighbors $\mathcal{N}(X) \subset B$ of X has size $|\mathcal{N}(X)| \geq (1-\epsilon)\ell|X|$.

The existence of expander graphs can be shown using the probabilistic method without explicitly constructing them, see [3, 24] and the references therein.

Conversely, any binary matrix with ℓ ones in each column and satisfying $RIP_{1,k,\delta}$ must be the adjacency matrix of an unbalanced expander graph, compare [3, Thm. 2].

The significance of $RIP_{1,k,\delta}$ is the following performance guarantee when reconstructing an *arbitrary* (not necessarily k-sparse) vector x by solving (3).

Theorem 4.7. *[3, Thm. 3] Let $A \in \{0,1\}^{m\times n}$ be the adjacency matrix of an unbalanced $(2k,\epsilon)$-expander. Let $\alpha(\epsilon) = 2\epsilon/(1-2\epsilon)$. Consider any two vectors x, x^* such that $Ax = Ax^*$ and $\|x\|_1 \leq \|x^*\|_1$. Then*

$$\|x^* - x\|_1 \leq \frac{2}{1-2\alpha(\epsilon)}\|x^* - x^k\|_1\ ,$$

where $x^k \in \mathbb{R}^n_k$.

Proposition 4.8. *Let A be the adjacency matrix of an arbitrary bipartite graph with regular left degree ℓ and denote $\sigma = \operatorname{spark}(A)$. Then A is the adjacency matrix of a $(\sigma - 1, 1 - \frac{1}{\ell})$–unbalanced expander.*

Proof. Let $X \subset \{1,\ldots,n\}$ and $|X| = k \leq \sigma - 1$. Then A_X is an overdetermined full rank matrix. In particular, there exist k linearly independent rows in A_X, each of them having at least one nonzero entry. Hence $|\mathcal{N}(X)| \geq |X| = \ell(1-(1-\frac{1}{\ell}))|X|$. □

The above proposition implies that A from (4) is the adjacency matrix of a $(7, \frac{2}{3})$–unbalanced expander. As a consequence we obtain exact recovery for every 3-sparse particle distribution in view of Thm. 4.7 and Prop. 4.1.

4.5 Neighborly Polytopes

Donoho and Tanner [15, 19] explained the ℓ_0/ℓ_1-equivalence phenomenon from the viewpoint of convex neighborly polytopes. In contrast to sufficient conditions for exact recovery, this theory provides necessary conditions additionally.

A polytope is said to be *outwardly k-neighborly* if every subset of k vertices not including 0 span a $(k-1)$-face, see [19], thus a outwardly k-neighborly polytope behaves like a simplex, at least from the viewpoint of it's lowdimensional faces (not including 0), since every p-dimensional face (not including 0) is simplicial, for $0 \leq p < k$.

The main result in [19] connects outward neighborliness to the question of uniqueness of *any* k-sparse nonnegative vector. Such a k-sparse vector x^* will "live" on a k-face of the convex hull of the standard simplex in $\mathbb{R}^n$ and the origin, denoted by Δ_0^{n-1}. If Ax^* will "survive" on a k-face of $A\Delta_0^{n-1} = \operatorname{conv}\{A_{\bullet,1}, \ldots, A_{\bullet,n}, 0\}$ then it will be the unique positive solution satisfying $Ax = Ax^*$. If Ax^* falls "inside" the "transformed" polytope $A\Delta_0^{n-1}$ then x^* cannot be recovered by (5). For a outwardly k-neighborly polytope AT_0^{n-1} this will never happen.

We will extend this result by the following simple observation.

Theorem 4.9. *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. $\operatorname{conv}\{A_{\bullet,1}, \ldots, A_{\bullet,n}, 0\}$ is outwardly k-neighborly.

(c) Every nonzero null space vector has at least $k+1$ negative entries.

Proof. The equivalence of (a) and (b) is the main result in [19, Thm. 1].
$(c) \Rightarrow (a)$: Now, let x^* be a k-sparse vector. Any other (different) positive solution of $Ax = Ax^*$ must be of the form $x^* + v$ such that $x^* + v \geq 0$ and $v \in \ker(A)\setminus\{0\}$. Hence $I^-(x^* + v) = \emptyset$. This contradicts $|I^-(x^* + v)| \geq 1$ as claimed by (c).
$(a) \Rightarrow (c)$: Conversely, lets assume that there exist a nonzero null space vector v with $|I^-(v)| \leq k$. We now define two nonnegative vectors x^1 and x^2 in the following way

$$x_i^1 = \begin{cases} v_i, & \text{if } i \notin I^-(v) \\ 0, & \text{otherwise} \end{cases}$$

and

$$x_i^2 = \begin{cases} -v_i, & \text{if } i \notin I^-(v) \\ 0, & \text{otherwise .} \end{cases}$$

Since $x^1 - x^2 = v \neq 0$ we obtain two different solutions to $Ax^1 = Ax^2$ although x^2 is k-sparse. This completes the proof. □

From Prop. 2.2 we know the existence of null space vectors with only 4 negative entries. This together with Thm 4.9 now yields

Corollary 4.10. *The convex hull of the columns in matrix A defined in* (4) *and the zero vector, i.e. $\operatorname{conv}\{A_{\bullet,1}, \ldots, A_{\bullet,n}, 0\}$ is outwardly 3-neighborly.*

Hence, the maximal sparsity level k such that ℓ_0/ℓ_1-equivalence holds *for all* k-sparse nonnegative vectors holds is 3. Indeed, in a $d \times d \times d$ volume there are $\binom{d}{2}^3$ 4-sparse vector pairs with equal projections, compare Fig. (3).

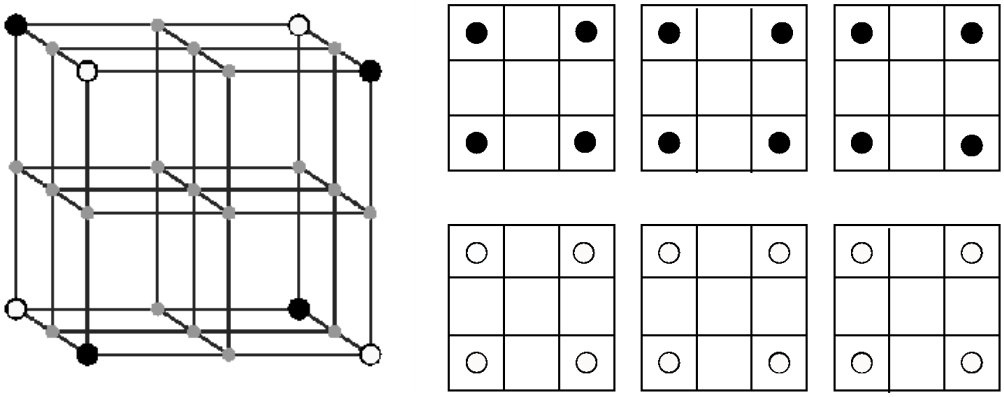

Figure 3: Two different *non unique* 4-sparse "particle" distributions in a $3 \times 3 \times 3$ volume, based on the construction in the proof of Thm. 4.9. Both configurations (represented by black and white dots) yield identical projections in all three directions.

4.6 Null Space Property

Similar to our simple observation in the previous section the authors in [13] derived sparse recovery conditions based on properties of the null space of A. In particular, they say that a matrix A has the *Null Space Property*[1] of order k for $\gamma > 0$ if

$$\|v_S\|_1 \leq \gamma \|v_{S^c}\|_1 \tag{14}$$

holds for all sets S of cardinality less then k and $v \in \ker(A)$. In [13, Thm. 4.3] it is shown that if A has the null space property of order $\geq k$ and $\gamma < 1$, it is guaranteed that every k-sparse vector is the unique ℓ_1-minimizer of (3). The null space property is a weaker version of the restricted ℓ_2-isometry property. Indeed, Cohen et al showed [13, Lem. 4.1] that if A satisfies the $RIP_{2,3k,\delta}$, then A satisfies the null space property of order $2k$ and $\gamma = \frac{\sqrt{2}}{2}\sqrt{\frac{1+\delta}{1-\delta}}$.

Independently, Zhang [31] used the general concept of *k-balanceness* to study uniqueness of the ℓ_1-minimizer. A subspace X is k-balanced (in ℓ_1-norm) if for any S with $|S| \leq k$

$$\|v_S\|_1 \leq \|v_{S^c}\|_1$$

holds for all $v \in X$. X is called *strictly* k-balanced if the strict inequality holds. Hence, strict k-balanceness of the null space of A implies the null space property of order k with $\gamma < 1$, thus, exact recovery. In fact, the author shows in [31] that k-balanceness of $\ker(A)$ is equivalent to $\mathrm{conv}\{\pm A_{\bullet,1}, \ldots, \pm A_{\bullet,n}, 0\}$ being (outwardly) k-neighborly. The latter is the analogous sufficient and necessary condition for recovery of all k-sparse vector when the vector might have different signs, compare [15].

In the nonnegative case Zhang showed [32] the equivalence of the (outwardly) k-neighborliness of the polytope $\mathrm{conv}\{A_{\bullet,1}, \ldots, A_{\bullet,n}, 0\}$ and the notion of *half k-balanceness* of the null space of A. A subspace X is half k-balanced (in ℓ_1-norm) if for any S with $|S| \leq k$

$$\sum_{i \in S} v_i \leq \|v_{S^c}\|_1$$

holds for all $v \in X$. X is called *strictly* half k-balanced if the strict inequality holds. Hence, this different form of null space property for nonnegative vectors turns out to be sufficient and

[1] For convenience we slightly modified the original definition of the Null Space Property given in [13].

necessary condition for uniqueness of every k-sparse nonnegative vector, in view of the first part of Thm. 4.9. However, testing the null space property conditions on generic matrices is potentially harder than solving the combinatorial ℓ_0-problem in (2) as it implies solving a combinatorial problem to compute γ. However, we can conclude that A from (4) has the null space property of order 3 with $\gamma < 1$, due to the previous observations. This ends the series of highly pessimistic conclusions concerning our particular A.

5 Most Probably Unique Positive Solution

5.1 Weak Equivalence

The concept of ℓ_0/ℓ_1-equivalence demands that for a given measurement matrix A, equivalence for *all* instances (A, b) generated by *any* k-sparse vector holds. A weaker form of equivalence considers equivalence for *most* problem instances (A, b). In [19] it is shown that a weaker form of neighborliness implies weak equivalence. The authors define a polytope P to be (k, ϵ)-weakly (outwardly) neighborly if, among all k-subsets of vertices (resp. among those not including 0), all except a fraction ϵ span $k-1$-faces of P.

The columns of A are in *general position* if all $subsets$ of m columns of A are linearly independent, thus $\operatorname{spark}(A) = m + 1$. It is shown in [19] that if the columns of A are in general position, weak neighborliness of $A\Delta_0^{n-1} = \operatorname{conv}\{A_{\bullet,1}, \ldots, A_{\bullet,n}, 0\}$ is the same thing as saying that $A\Delta_0^{n-1}$ has at least $(1-\epsilon)$-times as many $(k-1)$-faces as Δ_0^{n-1}. Thm. 2 in [19] shows the equivalence between (k, ϵ)-weakly (outwardly) neighborliness and weak equivalence, i.e. uniqueness of all except a fraction ϵ of k-sparse nonnegative vectors.

However, the columns of A from (4) are not in general position. Besides, counting faces of polytopes is again a combinatorial problem.

To overcome this difficulty we appeal to the observation already made in Section 2. If the matrix obtained by reducing zero measurements and corresponding adjacent voxels is overdetermined and of full rank, then the underlying solution which generates the sparse measurement vector must be unique. This is also a criterion of individual equivalence for a given problem instance (A, b). Moreover, a critical sparsity level k yielding weak equivalence for A of most k-sparse nonnegative vectors can be derived by estimating the probability that k-columns are linearly independent with probability close to one, i.e. $\operatorname{sig}_k(A) \approx 0$, and estimating the probability that the induced reduced matrix is overdetermined.

5.2 Probability of $m_{red}(k) \geq n_{red}(k)$

Sparse vectors x give rise to sparse vectors $b = Ax$. Based on the zero components of b corresponding rows and columns can be removed from A, leading to a *reduced matrix* $A_{red} \in \mathbb{R}^{m_{red}(k) \times n_{red}(k)}$. In this section, we estimate the expected dimension of the reduced matrix depending on the sparsity k of x.

Lemma 5.1. *Let $x \in \{0,1\}^{d^3}$ be a uniformly drawn k-sparse binary vector. Then the expected number of zero measurements in any of the three projection images approximately is*

$$\mathbb{E}[k, d] := \frac{1}{d^{2k}} \sum_{r=0}^{d^2} r \binom{d^2}{r} (d^2 - r)!\, S_{k,d^2-r} \,, \tag{15}$$

where $S_{n,k}$ denotes the Stirling number of the second kind.

Proof. Let $p\colon K \to R$ be any of the three projection directions considered as a function mapping $|K| = k$ particles onto $|R| = d^2$ pixels. We wish to determine the probability that r pixels, corresponding to r rows in the measurement vector b, remain "empty".

This probability is given by $|\Omega_r|/|\Omega|$, where Ω denotes the set of all projections p, i.e. $|\Omega| = |R|^{|K|}$, and where $\Omega_r \subset \Omega$ contains functions p mapping k particles to $|R| - r$ pixels.

Assume r "empty" pixels are fixed. Then only *surjective* mappings p assign k particles to *all* remaining $|R| - r$ pixels without leaving any additional pixel empty. The size of this set is $(|R| - r)!\, S_{k,|R|-r}$, see [1]. Because there are $\binom{|R|}{r}$ ways to locate the r zero pixels, we obtain

$$|\Omega_r| = \binom{|R|}{r} (|R| - r)!\, S_{k,|R|-r} \,. \tag{16}$$

Clearly, $|\Omega| = \sum_{r=0}^{|R|} |\Omega_r|$, and the expected number of zero pixels is $\mathbb{E}[k,d] = \sum_{r=0}^{|R|} r \frac{|\Omega_r|}{|\Omega|}$. □

Remark 5.1. We point out that (15) is just an approximation, because we ignored the dependencies between particles due to the third dimension. Consequently, the numbers (17) determined below as a function of $\mathbb{E}[k,d]$ are approximations as well.

Proposition 5.2. *Let $x \in \{0,1\}^{d^3}$ be a uniformly drawn k-sparse binary vector. Then the expected values of the dimension of the reduced matrix A_{red} approximately are:*

$$m_{red}(k) \approx m - 3\mathbb{E}[k,d] \,, \tag{17a}$$

$$n_{red}(k) \approx d^3 - 3\mathbb{E}[k,d] \cdot d + 3\frac{\mathbb{E}[k,d]^2}{d} - \left(\frac{\mathbb{E}[k,d]}{d^2}\right)^3 d^3 \,. \tag{17b}$$

Figure 4 illustrates that these estimates are reasonably tight.

Proof. The estimate (17a) is based on our assumption that x is uniformly distributed. We simply subtract from the total number of pixels (rows) the expected number of zero measurements in all three projections due to Lemma 5.1, thus obtaining the expected number of zero components of the observed vector b.

Concerning (17b), any zero component of the vector b marks voxels in the volume along the corresponding projection ray, and corresponding columns in A, to be removed from A. $n_{red}(k)$ is the number of voxels (columns) *not* removed by any projection. To estimate the expected value of this number, we have to take into account that projection rays intersect.

Based on the expected number $\mathbb{E}[k,d]$ of zero pixels in any of the three projections – see (15), we compute:

1. Each *single projection* removes $\mathbb{E}[k,d] \cdot d$ voxels.

2. Consider a *pair of projections*, e.g. the x/z-projection and the y/z-projection. Fix the common z-coordinate. There are $\mathbb{E}[k,d]/d$ zero pixels in each of the two corresponding rows of the two projection images, eliminating together $(\mathbb{E}[k,d]/d)^2$ voxels because all projection rays corresponding to the two sets of zeros mutually intersect. As there are d possible values of z, it follows that each pair of projections removes $d(\mathbb{E}[k,d]/d)^2 = (\mathbb{E}[z])^2/d$ voxels.

3. The probability that any fixed voxel projects to a zero in any fixed projection is $\mathbb{E}[k,d]/d^2$, due to Lemma 5.1. Consequently, the expected number of voxels removed by *all three projections* is $\left(\frac{\mathbb{E}[k,d]}{d^2}\right)^3 d^3$.

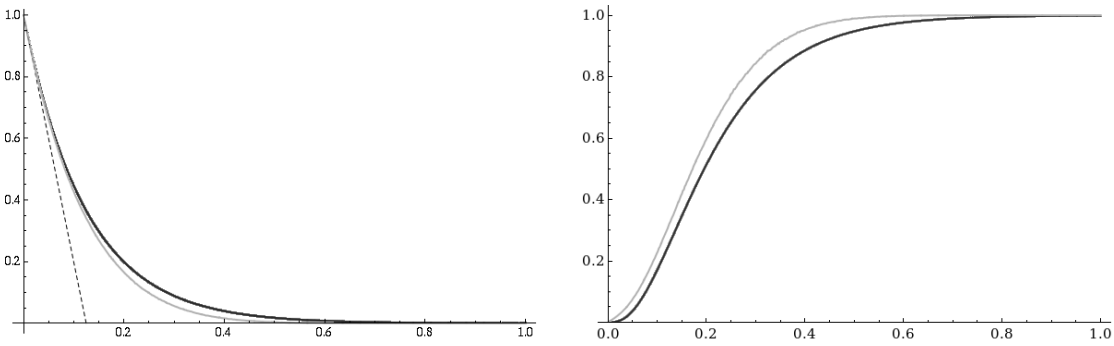

Figure 4: **Left:** Expected number of zero components of the measurement vector $b = Ax$, generated by a k-sparse random vector x. The black curve shows $3\mathbb{E}[k,d]/m$ as a function of $\frac{k}{d^3}$ due to (15). These numbers are related to the expected number $m_{red}(k)$ of *rows of the reduced matrix* A_{red} by (17a). The gray curve shows the corresponding empirical means computed for $d = 8$, i.e. for the matrix $A \in \mathbb{R}^{192\times 512}$, and 1000 trials for each value of k. The dashed curve shows the asymptotic $1 - \frac{k}{m}$ for small values of k. **Right:** Expected number of *columns of the reduced matrix* A_{red}. The black curve shows $\frac{n_{red}(k)}{n}$ as a function of $\frac{k}{n}$, with $n_{red}(k)$ given by (17b). The gray curve shows the corresponding empirical curve obtained by simulations as described above.

$n_{red}(k)$ corresponds to the number of voxels for which all three conditions above do *not* hold, which due to the inclusion-exclusion principle is given by (17b). □

Comparing (17a) and (17b) shows that more columns are removed than rows, depending on the expected number $\mathbb{E}[k,d]$ of vanishing components of $b = Ax$. Hence for a sufficiently k-sparse vector x the reduced matrix A_{red} leads to an *overdetermined* system with $m_{red}(k) \geq n_{red}(k)$. Solving the polynomial $m_{red}(k) = n_{red}(k)$ according to (17) in the variable $\mathbb{E}[k,d]$ for the root in the admissible interval $[1, d^2]$, we find that this will hold on the average for k-sparse vectors x that generate at least

$$\mathbb{E}[k,d] \approx \left(1 - \sqrt{\frac{3}{d}}\right) d^2 \tag{18}$$

zero entries in each projection. Figure 5 shows the corresponding critical values of the sparsity parameter $k = k(d)$, numerically determined by solving $m_{red}(k) = n_{red}(k)$ resp. (18), as a function of the problem size d. The log-log plot in the right panel of Figure 5 indicates quite accurately the power law

$$k(d) \approx 3.54 d^{1.34} \ . \tag{19}$$

6 Towards an improvement – Perturbation of A

6.1 Increasing Spark

Having the previous results in mind we further address the question of improving the properties of A from (4) with respect to the overall objective: ℓ_0/ℓ_1-equivalence. The weak performance of A rests upon the small spark of A. In order to increase the maximal number s of columns

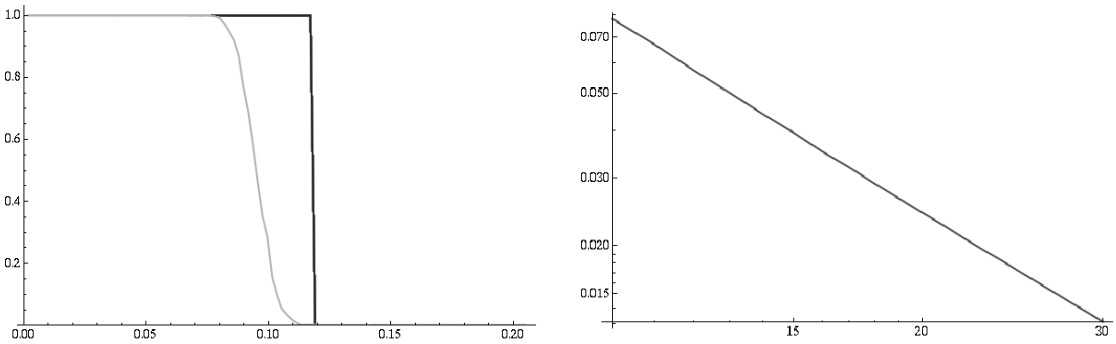

Figure 5: **Left:** Critical value of k/d^3, for $d = 8$, below of which the reduced matrix A_{red} satisfies $m_{red}(k) \geq n_{red}(k)$ with high probability. The black curve shows the estimate based on Prop. 5.2. The gray curve shows the empirical probability based on simulations as described in the caption of Fig. 4. **Right:** Critical value of $k/d^3 = k(d)/d^3$ as a function of the problem size d, according to (17). The log-log plot indicates the power law (19).

such that all s (or less) column combinations are linearly independent we add to the entries of A a small perturbation.

We will keep in mind the following result which might be well known.

Lemma 6.1. *Let $B \in \mathbb{R}^{m\times n}$ be any matrix of rank r, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ its singular values and $B = U\Sigma V^\top$ is singular value decomposition, where*

$$\Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix}$$

with $\Sigma_r = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$. If $\|E\| < \sigma_r$ then $rank(B + E) \geq \mathrm{rank}(B)$. Moreover, if we denote by

$$U^\top EV =: \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}$$

then

$$\mathrm{rank}(B + E) = \mathrm{rank}(A) + \mathrm{rank}(S) \tag{20}$$

where S is the Schur complement $E_{22} - E_{21}(\Sigma_r + E_{11})^{-1}E_{12}$ of

$$\begin{pmatrix} \Sigma_r + E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}.$$

Proof. In view of our assumption we also have $\|E_{11}\| < \sigma_r$ since $U, V^\top$ are orthogonal. Hence $\|\Sigma_r^{-1}E_{11}\| < 1$ holds, which also implies the nonsingularity of $\Sigma_r + E_{11}$. By writing

$$\begin{pmatrix} I & 0 \\ -E_{21}(\Sigma_r + E_{11})^{-1} & I \end{pmatrix} \begin{pmatrix} \Sigma_r + E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix} \begin{pmatrix} I & -(\Sigma_r + E_{11})^{-1}E_{12} & I \\ 0 & I & \end{pmatrix} = \begin{pmatrix} \Sigma_r + E_{11} & 0 \\ 0 & S \end{pmatrix}$$

we obtain he desired result (20). □

We stress that the above result holds for every matrix E. However, we are interested in matrices E having the same sparsity structure like A.

We conjecture that the rank of every perturbed submatrix of A will grow by a factor $O(1)$. By perturbing A we will "eliminate" all 8-column combinations (the column sets corresponding

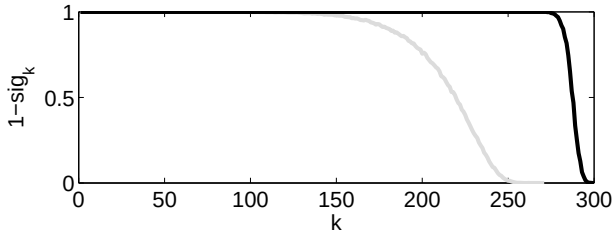

Figure 6: $1 - \mathrm{sig}_k(A)$ versus perturbed $1 - \mathrm{sig}_k(\tilde{A})$ for $d = 10$; Empirical probability obtained from 100000 trials that k columns are linearly independent.

to nonzero entries in the null space basis vectors in N, compare Prop. 2.2). By elimination we mean that the perturbed 8-tuples will have complete rank since the unperturbed clearly have rank 7 since $\mathrm{spark}(A) = 8$. Moreover, all k-linearly independent column sets of A can be obtained by combining linearly independent 8-tuples. By a similar argument most such k-dependent columns in A can be transformed to independent ones by simply perturbing their entries. This suggests that $\mathrm{spark}(\tilde{A})$ will grow proportionally to the rank of A. The numerical results in Section 7 suggest the power law $\mathrm{spark}(\tilde{A}(d)) = O(d^2)$, compare Fig. 7.

Remark 6.1. For $A \in \mathbb{R}^{m\times n}$, let $\{\sigma_1, \sigma_2, \ldots, \sigma_m\}$ and $\{\tilde{\sigma}_1, \tilde{\sigma}_2, \ldots, \tilde{\sigma}_m\}$ be all singular values (nonzero as well as any zero ones) for A and $\tilde{A} = A + E$, respectively. Then

$$|\sigma_i - \tilde{\sigma}_i| \leq \|E\|_2 \quad \text{for each} \quad i = 1, 2, \ldots, m \, .$$

By choosing E properly it seems possible to "adjust" the singular values of A such that $\tilde{A}$ will satisfy the RIP_2 property. We intend to investigate this further in order to obtain recovery results that are stable in the presence of errors in TomoPIV measurements.

6.2 How Neighborly will be the Perturbed Matrix?

In Section 4 we presented several concepts which quantify the recovery performance of a given matrix A. Among these k-neighborliness and the null space property of order k are necessary and sufficient conditions which guarantee uniqueness of every k-sparse positive vectors. In order to address the question of equivalence between (2) and (5) for $\tilde{A}$ we consider neighborliness of $\tilde{A}\Delta_0^{n-1}$.

Assume that by perturbing the nonnegative entries of A we obtained a substantially increased spark $\tilde{\sigma} := \mathrm{spark}(A + E)$. Set $\tilde{A} := A + E$ and note that $\tilde{a}_{ij} = 0$ iff $a_{ij} = 0$.

Theorem 6.2. *The convex hull of the columns in the perturbed matrix $\tilde{A}$ and the zero vector, i.e.* $\mathrm{conv}\{\tilde{A}_{\bullet,1}, \ldots, \tilde{A}_{\bullet,n}, 0\}$ *is* at least *outwardly* $(\frac{\tilde{\sigma}}{3} - 1)$*-neighborly.*

Proof. We will show that every nonzero null space vector has *at least* $\tilde{\sigma}/3$ negative entries. Then Thm. 4.9 will provide the desired result. Let $v \in \ker(\tilde{A}) \setminus \{0\}$ and denote by $S = \mathrm{supp}(v)$. Clearly,

$$|S| \geq \tilde{\sigma} \, , \tag{21}$$

and

$$|\mathcal{N}(S)| \geq \tilde{\sigma} \, , \tag{22}$$

where $\mathcal{N}(S) = \{i \in \{1,\dots,m\} | \tilde{a}_{ij} > 0, j \in S\}$ indexes all neighbors of S. In view of $S = I^{-1}(v) \cup I^{+}(v)$ and $v \in \ker(\tilde{A})$, we have

$$\mathcal{N}(I^{-}(v)) = \mathcal{N}(I^{+}(v)) = \mathcal{N}(S) \tag{23}$$

since it is not possible to find a voxel corresponding to a negative entry in v indexed by $I^{-}(v)$, or a voxel corresponding to a positive entry in v indexed by $I^{+}(v)$, that is not connected to both sets of rows $\mathcal{N}(I^{-}(v))$ and $\mathcal{N}(I^{+}(v))$, since otherwise $\tilde{A}v \neq 0$ in view of $\tilde{a}_{ij} \geq 0$. Summarizing we obtain

$$|\mathcal{N}(I^{-}(v))| = |\mathcal{N}(S)| \geq \tilde{\sigma} \, . \tag{24}$$

On the other hand, since each voxel is connected to exactly 3 rows we have

$$|\mathcal{N}(I^{-}(v))| \leq 3|I^{-}(v)| \, . \tag{25}$$

Combining (24) and (25) we obtain the desired result. □

This guarantees exact recovery by (5) via $\tilde{A}$ for at least all $(\tilde{\sigma}/3 - 1)$-sparse nonnegative vectors.

We stress that it is possible to obtain a good upper bound on the spark of an arbitrary matrix A by computing first its row echelon (which can be done efficiently if A is sparse) and then obtain a sparse null space vector from its row echelon.

6.3 Unique Solution of the Reduced System

Equivalence for most problem instances can be obtained by similar arguments as in Section 5.2. The critical value of k such that a k-sparse vector with uniform distributed nonzero entries induces a overdetermined reduced system is again $k(d) \approx 3.54d^{1.34}$. Then a lower bound to the critical value k such that a k-sparse nonnegative vector with uniformly distributed nonzero entries is most probably unique is

$$k(d) \geq \min\{3.54d^{1.34}, 2.7d^2\} \, ,$$

where we assumed that $2.7d^2$ or less columns combinations are most probably unique based on the results in Fig. 7.

7 Numerical Experiments

7.1 Phase Transitions

In this section we inspect empirical bounds on the required sparsity that guarantee exact reconstruction and critical parameter values that yield a performance similar to the settings considered in compressed sensing (e.g. [15, 19, 16]).

These parameter values allow us to answer the question of how sparse a vector should be (particle density) such that ℓ_0-minimization can be solved by ℓ_1-minimization or simply by the linear program (5).

In analogy to [16] we assess the so called *phase transition* ρ as a function of d, which is reciprocally proportional to the undersampling ratio $\frac{m}{n} \in (0,1)$. We consider $d \in \{3,\dots,55\}$, the corresponding matrix $A \in \mathbb{R}^{3d^2 \times d^3}$ from (4) and it's perturbed version $\tilde{A}$ and the sparsity as a fraction of $m = 3d^2$, $k = \rho m$, for $\rho \in (0,1)$.

This phase transition $\rho(d)$ indicates the necessary ratio $\frac{m}{n}$ to recover a k-sparse solution with overwhelming probability. More precisely, if $\|x\|_0 \leq \rho(d) \cdot m$, then with overwhelming probability the ℓ_0-problem of finding the k-sparsest solution can be solved by the LP (5). For Gaussian matrices there are precise values of $\rho(d)$, see [16, 19], which can be computed analytically.

Relevant for TomoPIV is the setting $d \approx 1024$. In the case of severe undersampling, i.e. as $d \to \infty$, a *strong asymptotic threshold* $\rho_S(d) \approx (2e \log(2\sqrt{\pi}d/3))^{-1}$ and *weak asymptotic threshold*

$$\rho_W(d) \approx \frac{1}{2\log(\frac{d}{3})} \tag{26}$$

holds for Gaussian matrices A_G and nonnegative signals, where we have taken into account $A_G \in \mathbb{R}^{3d^2 \times d^3}$. The weak threshold says that $\rho_W(d) \cdot m$-sparse nonnegative vectors are *typically* the unique solutions of $Ax = b$ while the strong equivalence between (2) and (5) holds *for all* $\rho_S(d) \cdot m$-sparse signals.

In view of Section 4, the strong threshold for A from (4) equals 3 for all d, while for the perturbed matrix it can be lowered according to Thm. 6.2 by

$$\rho_S(d) \geq \frac{\mathrm{spark}(\tilde{A}(d))}{3} - 1 \,.$$

Since $\mathrm{spark}(\tilde{A})$ will grow with d, we obtain an improvement over the constant strong threshold for the unperturbed matrix A. Verifying the strong threshold for $\tilde{A}$ empirically would be NP-hard. However, it is possible to verify the weak thresholds empirically by running tests on a random set of examples.

7.2 Numerical Results

For each $d \in \{3, \ldots, 55\}$ we generated A according to (4) and $\tilde{A}$ by slightly perturbing it's entries. $\tilde{A}$ has the same sparsity structure as A, but random entries drawn from the standard uniform distribution on the open interval $(1, 1.001)$. We have 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 ρm-sparse binary vector was generated to compute the right hand side measurement vector and for each (d, ρ)-point 100 random problem instances were generated.

The empirical probability that $k = \rho m$ columns of A or $\tilde{A}$ are linearly independent for each parameter combination is presented in Fig. 7, while the probability that a $k = \rho m$-sparse vector can be recovered by the LP (5) is illustrated in Fig. 8, Fig. 9 and Fig. 10. Two slices of a phase transition plot for $d = 50$ and $d = 100$ are presented in Fig. 11. A threshold-effect is clearly visible in all figures exhibiting parameter regions where the probability of exact reconstruction is close to one. We refer to the figure captions for detailed explanations.

8 Conclusion and Further Work

The reconstruction of particle volume functions from few projections can be modeled as finding the sparsest solution of an underdetermined linear system of equations, since the original particle distribution can be well approximated with only a very small number of active basis functions relative to the number of possible particle positions in a 3D domain. In general the search for the sparsest solution is intractable (NP-hard), however. The newly developed theory of Compressed Sensing shows that one can compute via ℓ_1-minimization or linear programming

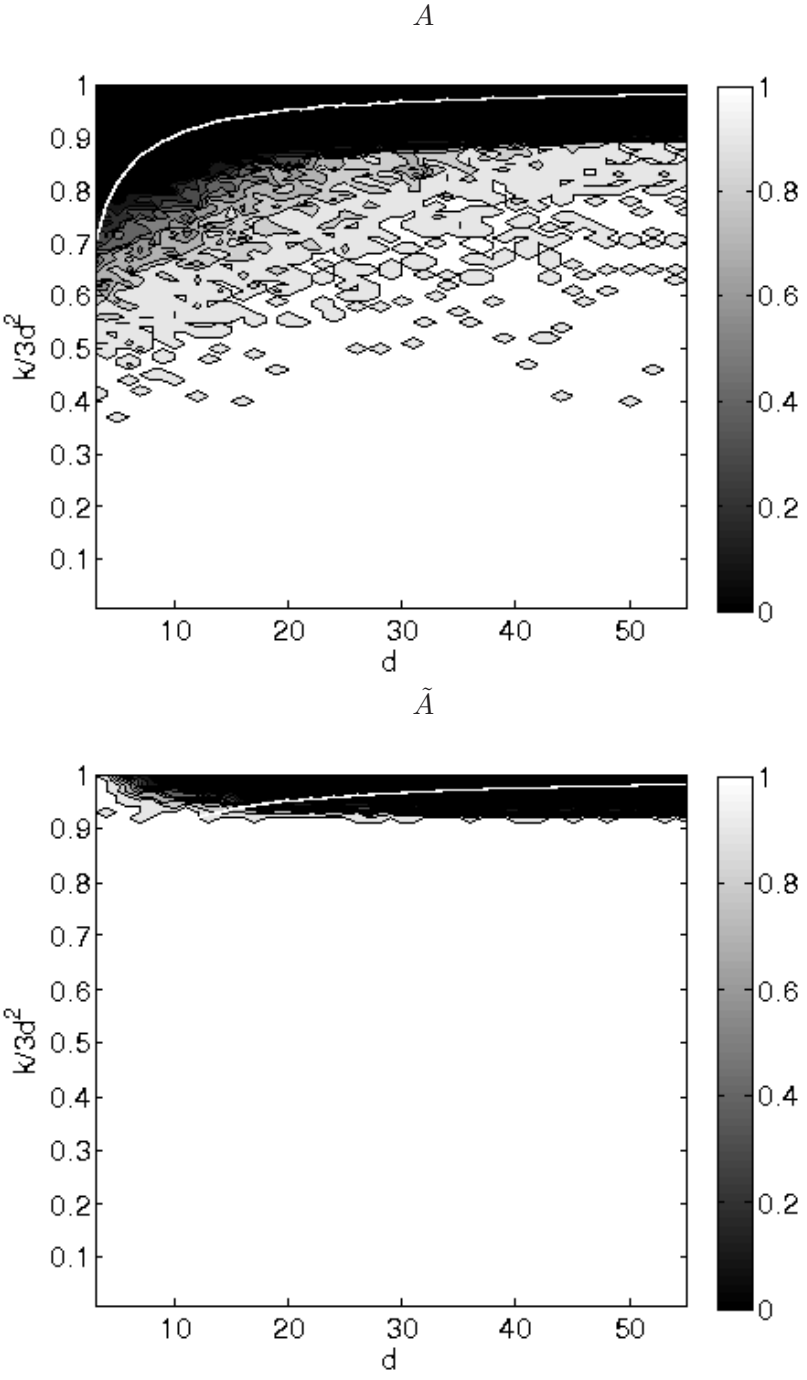

Figure 7: **Top:** Probability that ρm column combinations of A are linearly independent. **Bottom:** Probability that ρm column combinations of the perturbed matrix $\tilde{A}$ are linearly independent. The white curve depicts the scaled rank of matrix A as a function of d. The lower plot suggests that up to $3 \cdot 0.9d^2 = 2.7d^2$ column combinations of the perturbed matrix $\tilde{A}$ are most probably linearly independent. On the other hand, this can be claimed only up to three times less column combinations of A.

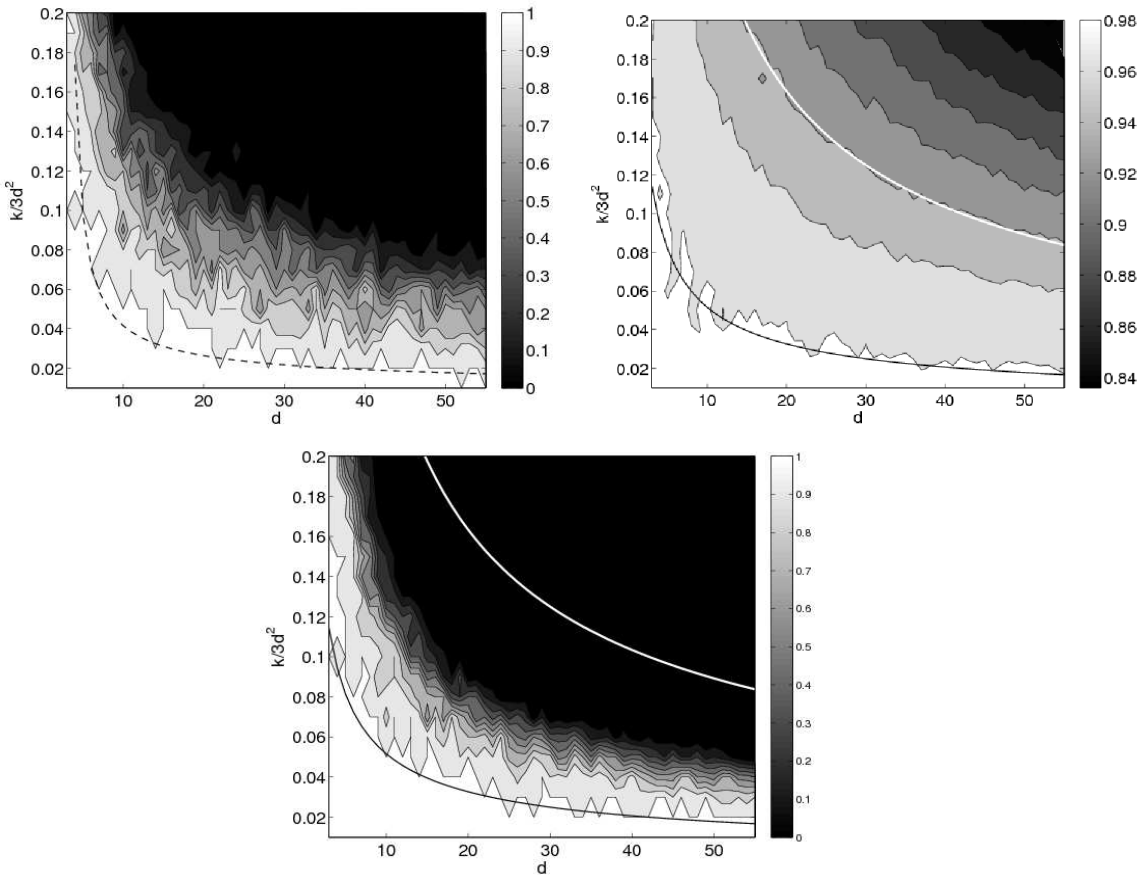

Figure 8: Results for matrix A from (4). **Top left:** Probability of correct recovery by linear programming of a random particle distribution that can be expressed with exactly $k = \rho m$ basis functions as a function of d. The dashed black curve depicts $0.1\rho_W(d)$, where ρ_W is the weak phase transition (26) of linear programming, but for Gaussian random matrices. The results indicate that A from (4) performs ten times worse in recovering *most* sparse nonnegative signals. **Top right:** Probability that the reduced matrix A obtained by eliminating zero measurements and corresponding adjacent voxels is overdetermined along with the estimated critical sparsity level $3.54d^{1.34}$ relative to the number of measurements as a function of d (solid white line), see (19). Five times the black line equals the white one. **Bottom:** Probability that a random $k = \rho m$ particle distribution induces an overdetermined and full rank reduced matrix. The results not only indicate that the reason for successful recovery in case of A are full rank overdetermined reduced matrices, but also that solving just an overdetermined linear system might be more stable than linear programming, when the solution is known to be nonnegative. The white and black solid curve have the same meaning as above.

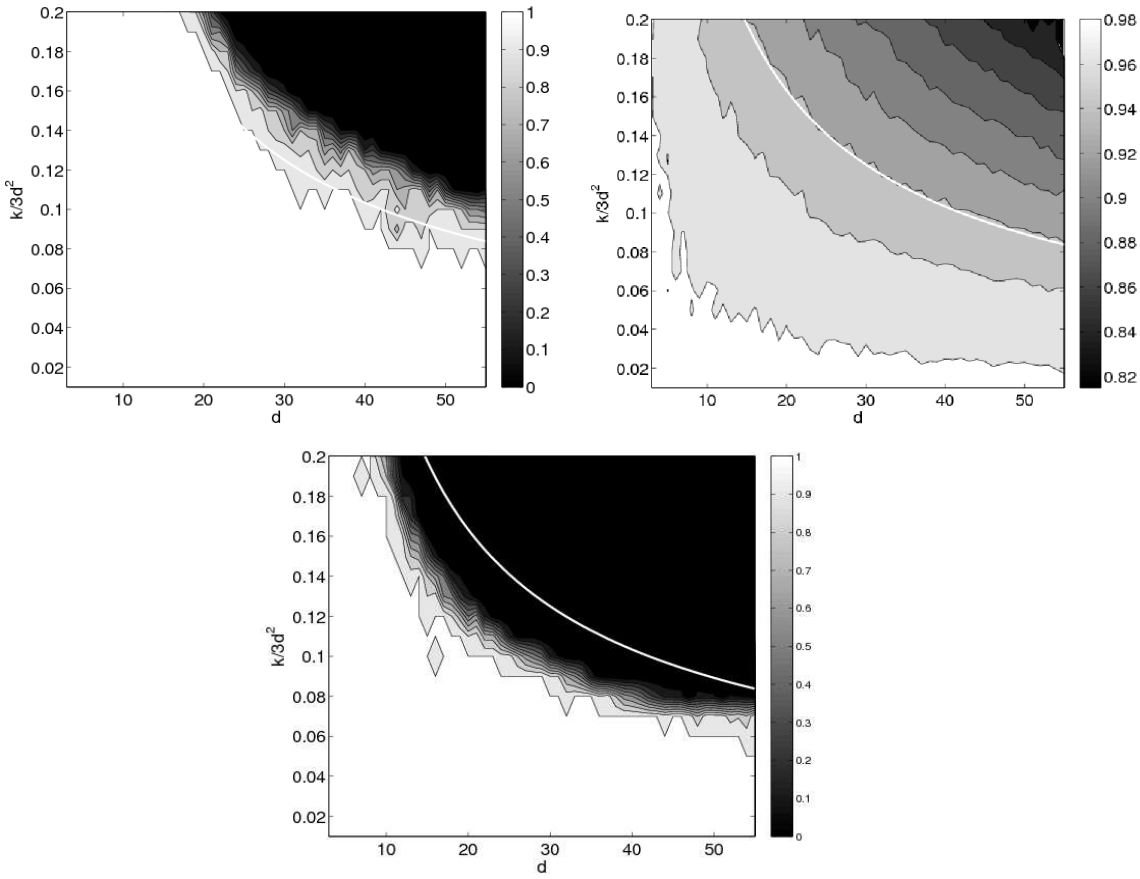

Figure 9: Results for the perturbed matrix $\tilde{A}$. **Top left:** Probability of correct recovery of a $k = \rho m$ sparse binary vector as a function of d. The solid white curve depicts $1.18d^{-0.66}$, compare (19) and right figure. **Top right:** Probability that the reduced matrix A_{red} is overdetermined along with the estimated relative critical sparsity level $1.18d^{-0.66}$ (solid white line) which induces overdetermined reduced matrices $\tilde{A}_{red}$. **Bottom:** Probability that a random $k = \rho m$ particle distribution induces an overdetermined and full rank reduced matrix along with the white curve $1.18d^{-0.66}$. In case of the perturbed matrix $\tilde{A}$ exact recovery is possible *beyond* overdetermined reduced matrices.

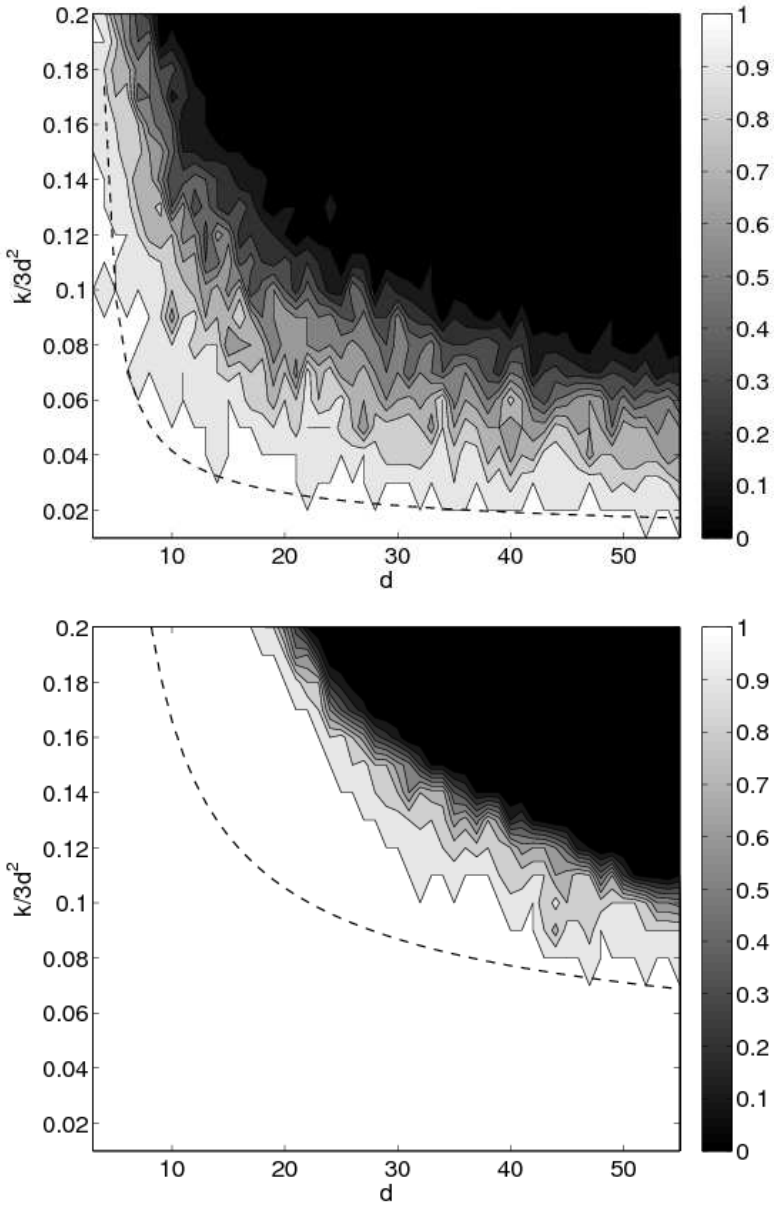

Figure 10: Recovery via A from (4) (top) versus recovery via the perturbed matrix $\tilde{A}$ (bottom). **Top:** Success and failure empirical phase transition for A along with $0.1\rho_W(d)$ (dashed). **Bottom:** Success and failure empirical phase transition for the perturbed matrix $\tilde{A}$ along with $0.4\rho_W(d)$ (dashed), compare (26). The results indicate that $\tilde{A}$ performs at least three times better in recovering ρm sparse vectors by the LP (5) within the considered range of d.

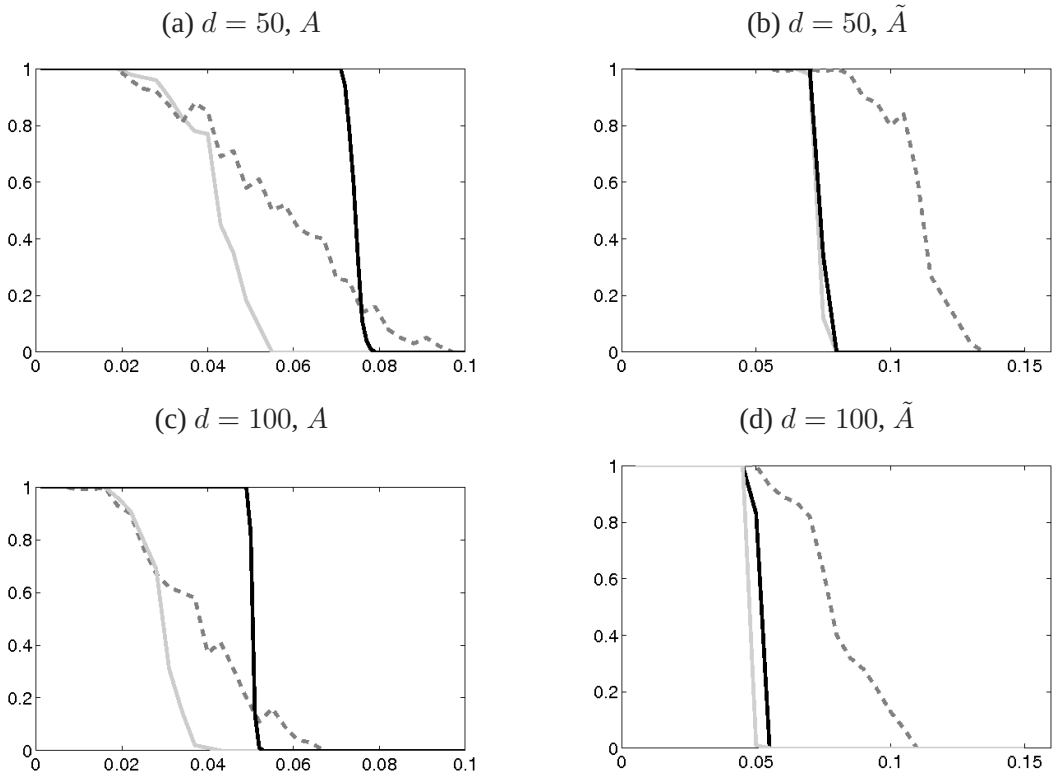

Figure 11: Slices through the contourplots Fig. 8 (top) and Fig. 9 (top); A versus the perturbed matrix $\tilde{A}$ for $d = 50$ and $d = 100$. The dashed gray line depicts the probability (as function of ρ) that a ρm-sparse binary vector is recovered exactly by the LP (5). The gray line illustrates the probability that a ρm-sparse binary vector induces an overdetermined reduced matrix of full rank while the black line plots the probability that the reduced matrix is just overdetermined and not necessary of full rank. Here again $\tilde{A}$ performs three times better.

the sparsest solution for underdetermined systems of equations provided that the coefficient matrix (also called measurement ensemble) satisfies certain conditions. Testing these conditions on generic matrices is often harder than solving the combinatorial ℓ_0-problem in (2) as it also implies solving a combinatorial problem which is intractable given the huge dimensionality of the measurement matrix within the TomoPIV setting. However, we showed in the present work that all currently available recovery conditions predict an extremely poor performance of the TomoPIV measurement ensemble when we restrict to a simple but realistic setup geometry. On average, such matrices perform approximately ten times worse than Gaussian matrices which allow for maximal sparsity such that for all less sparse vectors exact recovery is still guaranteed. However, when we slightly perturb the entries of such a degenerate measurement matrix we can boost both worst case and expected reconstruction performance. Then the particle density can be increased by a factor of three while preserving the number of measurements. The theoretical analysis within this work suggests that a similar procedure can applied to an arbitrary sparse matrix with bad reconstruction performance. We will investigate this issue further for adjacency matrices of expander graphs with bad expansion property.

Acknowledgement

The authors gratefully acknowledge support by the German Science Foundation (DFG), grant SCHN457/10-1.

References

[1] M. Aigner. *Diskrete Mathematik*. Vieweg, Braunschweig/Wiesbaden, Germany, 1993.

[2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx*, 28:253–263, 2008.

[3] R. Berinde, A.C. Gilbert, P. Indyk, H. Karloff, and M.J. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. In *Proc. 46th Ann. Allerton Conf. Communication, Control, and Computing*, pages 798–805, 2008.

[4] R. Berinde and P. Indyk. Sparse recovery using sparse random matrices. 2008. Preprint http://people.csail.mit.edu/indyk/report.pdf.

[5] A. Björner. Some matroid inequalities. *Discrete Mathematics*, 31:101–103, 1980.

[6] A.M. Bruckstein, M. Elad, and M. Zibulevsky. On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations. *IEEE Transactions on Information Theory*, 54:4813–4820, 2008.

[7] E. Candès. The restricted isometry property and its implications for compressed sensing. *Compte Rendus de l'Academie des Sciences, Paris, Series I*, 346:589–592, 2008.

[8] E. Candès, J. Romberg, and Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans on Information Theory*, 52(2):489–509, 2006.

[9] E. Candes, M. Rudelson, T. Tao, and R. Vershynin. Error correcting via linear programming. In *46th Ann. IEEE Symp. Found. Computer Science (FOCS'05)*, pages 295–308, 2005.

[10] E. Candès and Tao T. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans on Information Theory*, 52(12):5406–5425, 2006.

[11] Y. Censor and S.A. Zenios. *Parallel Optimization: Theory, Algorithms and Applications*. Oxford University Press, New York, 1997.

[12] V. Chandar. A negative result concerning explicit matrices with the restricted isometry property. 2008. Preprint http://dsp.rice.edu/files/cs/Venkat_CS.pdf.

[13] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22:211–231, 2009.

[14] R.A. DeVore. Deterministic constructions of compressed sensing matrices. *J. Complexity*, 23:918–925, 2007.

[15] D. Donoho and J. Tanner. Neighborliness of randomly-projected simplices in high dimensions. *Proceedings of the National Academy of Sciences*, 102:9452–9457, 2005.

[16] D. Donoho and J. Tanner. Thresholds for the recovery of sparse solutions via l1 minimization. In *Proc. 40th Ann. Conf. on Information Sciences and Systems*, pages 202–206, 2006.

[17] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[18] D.L. Donoho and M. Elad. Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization. *Proc. Natl. Acad. of Sci. U.S.A.*, 100:2197–2202, 2003.

[19] D.L. Donoho and J. Tanner. Sparse nonnegative solution of underdetermined linear equations by linear programming. *Proc. National Academy of Sciences*, 102(27):9446–9451, 2005.

[20] M. Elad. Sparse representations are most likely to be the sparsest possible. *Journal on Applied Signal Processing*, Vol. 2006:1–12, 2006.

[21] G. Elsinga, F. Scarano, B. Wieneke, and B. van Oudheusden. Tomographic particle image velocimetry. *Exp. Fluids*, 41:933–947, 2007.

[22] R. Gordon, R. Bender, and G.T. Herman. Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography. *J. Theor. Biol*, 29:471–481, 1970.

[23] A. Graham. *Kronecker Products and Matrix Calculus With Applications*. Halsted Press, John. Wiley and Sons, NY., 1981.

[24] P. Indyk. Explicit constructions for compressed sensing of sparse signals. In *Proc. 19th Ann. ACM-SIAM Conf. Discrete Algorithms*, pages 30–33, 2008.

[25] K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24:227–234, 1995.

[26] S. Petra and C. Schnörr. Tomopiv meets compressed sensing. 2009. Preprint http://www.ub.uni-heidelberg.de/archiv/9760.

[27] S. Petra, C. Schnörr, A. Schröder, and B. Wieneke. Tomographic image reconstruction in experimental fluid dynamics: Synopsis and problems. In *In: Mathematical Modelling of Environmental and Life Sciences Problems*. Ed Acad Romane, Bucuresti, 2007.

[28] A.M. Pinkus. *On L_1-Approximation*, volume 93 of *Cambridge Tracts in Math.* Cambridge Univ. Press, 1989.

[29] M. Rudelson and R. Vershynin. Geometric approach to error correcting codes and reconstruction of signals. *Int. Mathematical Research Notices*, 64:4019 – 4041, 2005.

[30] T. Strohmer and R.W. Heath. Grassmannian frames with applications to coding and communication. *Appl. Comp. Harmonic Analysis*, 14(3):257–275, May 2003.

[31] Y. Zhang. A simple proof for recoverability of l1-minimization: Go over or under? 2005. Preprint http://www.caam.rice.edu/ zhang/reports/tr0509.pdf.

[32] Y. Zhang. A simple proof for recoverability of l1-minimization (ii): the nonnegativity case. 2005. Preprint http://www.caam.rice.edu/ zhang/reports/tr0510.pdf.