1

Phase Transitions and Cosparse Tomographic Recovery of Compound Solid Bodies from Few Projections

Andreea Denițiu[*]

Image and Pattern Analysis Group
University of Heidelberg
Speyerer Str. 6, 69115 Heidelberg, Germany
denitiu@math.uni-heidelberg.de

Claudius Schnörr

University of Applied Sciences
Lothstr. 64, 80335 Munich, Germany
schnoerr@cs.hm.edu

Stefania Petra[†]

Image and Pattern Analysis Group
University of Heidelberg
Speyerer Str. 6, 69115 Heidelberg, Germany
petra@math.uni-heidelberg.de

Christoph Schnörr

Image and Pattern Analysis Group
University of Heidelberg
Speyerer Str. 6, 69115 Heidelberg, Germany
schnoerr@math.uni-heidelberg.de

Abstract. We study unique recovery of cosparse signals from limited-view tomographic measurements of two- and three-dimensional domains. Admissible signals belong to the union of subspaces defined by all cosupports of maximal cardinality ℓ with respect to the discrete gradient operator. We relate ℓ both to the number of measurements and to a nullspace condition with respect to the measurement matrix, so as to achieve unique recovery by linear programming. These results are supported by comprehensive numerical experiments that show a high correlation of performance in practice and theoretical predictions. Despite poor properties of the measurement matrix from the viewpoint of compressed sensing, the class of uniquely recoverable signals basically seems large enough to cover practical applications, like contactless quality inspection of compound solid bodies composed of few materials.

Keywords: compressed sensing, underdetermined systems of linear equations, cosparsity, total variation, discrete and limited-view tomography.

1. Introduction

1.1. Overview, Motivation

Discrete tomography [17] is concerned with the recovery of functions from few tomographic projections. Feasibility of this severely ill-posed problem rests upon assumptions that restrict the degrees of freedom of functions to be reconstructed. The canonical assumption is that functions only attain values from a finite set. According to [17], the field of discrete tomography deals with the reconstruction of images from few projections (limited views, limited data). Limited-view and -data tomography on the other hand, has shown potential for large-scale applications in various areas [31, 15] and this also stimulates theoretical research.

As advocated in [29], considering the problem of discrete tomography from the broader viewpoint of *compressive sensing* [1, 9] enables to consider more general scenarios and to employ additional methods for investigating theoretical and practical aspects of discrete tomography. While the set of measurements (tomographic projections) is still "discrete" as opposed to the "continuous" theory of established tomographic settings [25], functions to be reconstructed are required to be compressible: a *sparse* representation exists such that few measurements of any function of some admissible class capture the degrees of freedom and enable recovery. Establishing corresponding sampling rates in connection with a given sparse representation and a model of the imaging sensor constitutes the main problem of mathematical research.

In this paper, we focus on the problem of reconstructing compound solid bodies in two or three dimensions, as illustrated in Figure 1. However, the scope of our approach is not limited to this particular application of reconstructing compound bodies. More generally, we consider functions that are represented by vectors $u \in \mathbb{R}^n$ in a high-dimensional Euclidean space, with components $u(v_i)$ indexed by vertices $v_i \in V$ of a regular grid graph $G = (V, E)$ corresponding to pixels in 2D ($q = 2$) and to voxels in 3D ($q = 3$). The key assumption is that the functions we wish to reconstruct have sufficiently *sparse gradients*.
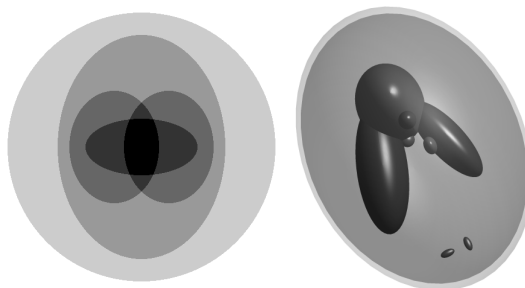

Figure 1.   The left figure sketches the class of compound solid bodies considered in this paper for reconstruction from few tomographic projections. These objects are similar to the 3D Shepp-Logan phantom (right) and are composed of different materials in a homogeneous way, but with unknown geometry. The gradient of the piecewise constant intensity function is sparse. Recovery conditions depending on this property and the number of measurements are studied in this paper. For the present example, the cosparsity ℓ (defined by(4.2)) of the 3D Shepp-Logan phantom with 128^3 voxels equals $\ell = 3(d-1)d^2 - 109930 = 6132374$ with $d = 128$. As we will show in connection with eqn. (4.17), the $\approx 2 \cdot 10^6$ voxel values of the 3D Shepp-Logan phantom can be recovered *exactly* from tomographic projections along $4 \times (2d-1)d = 130560$ parallel rays via the projecting matrix from four directions, see Section 2.1.2, Fig. 5.
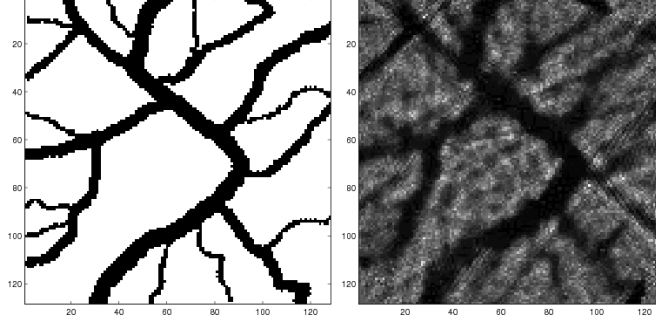
Figure 2. Two experimental results are shown that demonstrate the effect of regularization – total variation minimization recovery (left) versus ℓ_1-minimization recovery (right). Our recovery analysis considered in the present paper applies also to the 128×128 binary image on the left. This image has cosparsity $\ell = 2(d-1)d - 2251 = 30261$, while the sparsity of the gradient equals 2251. As a consequence, the left image can be reconstructed *exactly* via (1.3) from 18 projections. Uniqueness of the ℓ-cosparse solution is provided for at least 2263 measurements according to (4.17). On the other hand, taking into account that the image by itself (rather than its gradient) is 12648-sparse, we can reconstruct it exactly by (1.1), but from 63 projections, following the analysis from [30, 32]. Thus, about $3.5\times$ more measurements are needed than in the previous case. Using the same number of 18 projections that suffice for exact reconstruction via (1.3), the reconstruction via (1.1), yields the poor result shown on the right.

As a consequence, if the linear system $Au = b$ represents the tomographic imaging set-up with given measurements $b \in \mathbb{R}^m$, then the standard ℓ_1-minimization approach

$$\min \|u\|_1 \qquad \text{s.t.} \quad Au = b, \tag{1.1}$$

does not apply, because u itself is *not* sparse. We consider instead, the *total variation* criterion

$$\min_u \text{TV}(u) \qquad \text{s.t.} \quad Au = b, \tag{1.2}$$

and its *nonnegative* counterpart

$$\min_u \text{TV}(u) \qquad \text{s.t.} \quad Au = b, \quad u \geq 0, \tag{1.3}$$

that in the continuous case returns the $(q-1)$-dimensional Hausdorff measure of discontinuities of indicator functions [35], with numerous applications in mathematical imaging [33]. This provides the natural sparse representation of the class of functions considered in this paper (cf. Fig. 1 & Fig. 2). Our objective in this paper is to establish sampling rates that enable the recovery of u as solution to the optimization problem (1.2) or (1.3). Thus, we will concentrate strictly on the noiseless scenarios.

For industrial applications that motivate our work, we refer to e.g. [4, 14]. In this context, scenarios of *limited-data tomography* are relevant to our work as they enable minimization of acquisition time and related errors, affecting the quality of projection measurements and, in turn, object reconstruction. However, the problems in (1.2) or (1.3) have to be extended when dealing with noisy measurements in practice.

1.2.   Related Work and Contribution

Theoretical recovery guarantees, expressed as thresholds on the critical parameters of problem (1.1), relate the solution sparsity to the solution degrees of freedom and to the number of measurements. Recent work illustrates that the focus of corresponding research in *compressed sensing (CS)* is shifting - in contrast to discrete tomography [17] - from a *worst-case analysis* [13, 29] towards an *average-case analysis* [21, 22, 18]. As for many other difficult combinatorial problems, the probabilistic approach is a plausible, and often the only possible way, to make well-founded statements that go beyond idealistic mathematical assumptions and that are also relevant for real-world applications.

In discrete tomography, images to be reconstructed are sampled along lines. Thus, sampling patterns are quite different from random and non-adaptive measurements that are favourable from the viewpoint of compressed sensing. In [29], we showed that structured sampling patterns as used in commercial *computed tomography (CT)* scanners do not satisfy the CS conditions, like the nullspace property and the *restricted isometry property (RIP)*, that guarantee accurate recovery of sparse (or compressible) signals. In fact, these recovery conditions predict a quite poor worst-case performance of tomographic measurements, due to the high nullspace sparsity of a tomographic projection matrix A. Moreover, the gap between available *worst-case* recovery results of CS [10] and *worst-case* results from tomographic projections in [29] is dramatic.

In [32, 30], we presented an average-case relation between image sparsity and sufficient number of measurements for recovery, and we showed that the transition from non-recovery to recovery is sharp for specific sparse images. The analysis is based on the non-negativity of the coefficient matrix and of the signal itself and utilizes new mathematical tools from CS via expander graphs.

However, due to the unrestricted sign patterns of the sparse vector ∇u and of the corresponding coefficient matrix, compare Section 5, we cannot transfer the recovery results established in [32] to the problem (1.2) and (1.3).

We overcome this difficulty by adopting the recently introduced *cosparse analysis model* from [24], that provides an alternative viewpoint to the classical *synthesis model* and is more suitable to the problem class considered in this paper. Our present work applies and extends the results from [24] to the 3D recovery problem from few tomographic projections of three-dimensional images consisting of few homogeneous regions. We give a theoretical relation between the image *cosparsity* and sufficient sampling, validate it empirically and conclude that TV-reconstructions of a class of synthetic phantoms exhibit a well-defined recovery curve similar to the study in [30, 32].

Empirical evidence for the recovery of piecewise constant functions from few tomographic measurements was already observed in [34, 16, 19]. The first theoretical guarantees that have been obtained for recovery from noiseless samples of images with exactly sparse gradients via total variation minimization, date back to the beginnings of CS [8, 7]. However, the measurements considered were incomplete Fourier samples, and images were not sampled along lines in the spatial domain, but along few radial lines in the frequency domain. Such measurement ensembles are known to have good CS properties as opposed to the CT setup, and are almost isometric on sparse signals for a sufficient number of samples. As a result, recovery is stable in such scenarios. Stable recovery of the image gradient from incomplete Fourier samples was shown in [28], while Needell [27] showed that stable image reconstruction via total variation minimization is possible also beyond the Fourier setup, provided the measurement ensemble satisfies the RIP condition.

1.3. Organization

Section 2 collects basic definitions from compressed sensing and characterizes accordingly the imaging scenarios considered in this paper. We work out in more detail in Section 3 that the required assumptions in [27] do not imply relevant recovery guarantees for the discrete tomography set-ups considered here. In Section 4, we adopt the cosparse analysis model [24] and generalize corresponding results to the practically relevant three-dimensional case. Aspects of the linear programming formulation used to solve problem (1.3), are examined in Section 5. A comprehensive numerical study underpinning our results is reported in Section 6. We conclude in Section 7.

1.4. Basic Notation

For $n \in \mathbb{N}$, we use the shorthands $[n] = \{1, 2, \ldots, n\}$ and $[n]_0 = \{0, 1, \ldots, n-1\}$. For a subset $\Gamma \subset [n]$, the complement is denoted by $\Gamma^c = [n] \setminus \Gamma$. For some matrix A and a vector z, A_Γ denotes the submatrix of rows indexed by Γ, and z_Γ the corresponding subvector. Thus, $A_\Gamma z = (Az)_\Gamma$. $\mathcal{N}(A)$ denotes the nullspace of A. Vectors are column vectors and indexed by superscripts. $z^\top$ denotes the transposed vector z and $\langle z^1, z^2 \rangle$ the Euclidean inner product. To save space, however, we will sometimes simply write e.g. $z = (z^1, z^2)$ instead of correctly denoting $z = \left((z^1)^\top, (z^2)^\top\right)^\top$, for $z = \begin{pmatrix} z^1 \\ z^2 \end{pmatrix}$. $\mathbb{1} = (1, 1, \ldots, 1)^\top$ denotes the one-vector whose dimension will always be clear from the context. The dimension of a vector z we denote by $\dim(z)$.

We consider signals $u(x)$, $x \in \Omega \subset \mathbb{R}^q$, $q \in \{2, 3\}$ discretized as follows. Ω is assumed to be a rectangular cuboid covered by a *regular grid graph* $G = (V, E)$ of size $|V| = n$. Accordingly, we identify $V = \prod_{i \in [d]} [n_i]_0 \subset \mathbb{Z}^q$, $n_i \in \mathbb{N}$. Thus, vertices $v \in V$ are indexed by $(i, j)^\top \in \mathbb{Z}^2$ and $(i, j, k)^\top \in \mathbb{Z}^3$ in the case $q = 2$ and $q = 3$, respectively, with ranges $i \in [n_1]_0, j \in [n_2]_0, k \in [n_3]_0$, and

$$n = n_1 n_2 n_3. \tag{1.4}$$

As a result, discretization of $u(x)$, $x \in \Omega$, yields the vector $u \in \mathbb{R}^n$, where we keep the symbol u for simplicity.

Two vertices $v_1, v_2 \in V$ are adjacent, i.e. form an edge $e = (v_1, v_2) \in E$, if $\|v_1 - v_2\|_1 = 1$. We also denote this by $v_1 \sim v_2$.

Remark 1.1. Informally speaking, G corresponds to the *regular pixel or voxel grid* in 2D and 3D, respectively, and should not be confused with the general notion of a *regular graph*, defined by equal valency $\left|\{v' \in V : v' \sim v\}\right|$ for every $v \in V$. In this sense, the regular grid graphs G considered here are *not* regular graphs.

Consider the one-dimensional discrete derivative operator

$$\partial \colon \mathbb{R}^m \to \mathbb{R}^{m-1}, \qquad \partial_{i,j} = \begin{cases} -1, & i = j, \\ +1, & j = i+1, \\ 0, & \text{otherwise.} \end{cases} \tag{1.5}$$

Forming corresponding operators $\partial_1, \partial_2, \partial_3$ for each coordinate, conforming to the ranges of i, j, k such that $(i, j, k) \in V$, we obtain the discrete gradient operator

$$\nabla = \begin{pmatrix} \partial_1 \otimes I_2 \otimes I_3 \\ I_1 \otimes \partial_2 \otimes I_3 \\ I_1 \otimes I_2 \otimes \partial_3 \end{pmatrix} \in \mathbb{R}^{p \times n}, \tag{1.6}$$

where $\otimes$ denotes the Kronecker product and I_i, $i = 1, 2, 3$, are identity matrices with appropriate dimensions. The *anisotropic* discretized TV-measure is given by

$$\mathrm{TV}(u) := \|\nabla u\|_1. \tag{1.7}$$

2.   Properties of Tomographic Sensing Matrices

Depending on the application, different scanning geometries are used in CT imaging. In the present study, we adopt a simple discretized model based on an image $u(x)$, $x \in \Omega \subset \mathbb{R}^q$, $q \in \{2, 3\}$, that represents the inhomogeneity of Ω and consists of an array of unknown densities $u_j, j \in [n]$ as defined in Section 1.4. The model comprises algebraic equations for these unknowns in terms of measured projection data. To set up these equations, the sensing device measures line integrals of the object attenuation coefficient along X-rays $L_i, i \in [m]$, along some known orientations. The i-th corresponding measurement obeys

$$b_i :\approx \int_{L_i} u(x) dx \approx \sum_{j=1}^n u_j \int_{L_i} \mathcal{B}_j(x) dx = \sum_{j=1}^n u_j A_{ij}. \tag{2.1}$$

The values A_{ij} which form the *measurement or projecton matrix* A depend on the choice of the basis function. We assume $\mathcal{B}_j$ are cube- or square-shaped uniform basis functions, the classical *voxel* in 3D or *pixel* in 2D.

The main task studied in this paper concerns estimation of the weights u_j from the recorded measurements b_i and solving the noiseless setting $Au = b$. The matrix A has dimensions (# rays =: m) $\times$ (# voxel/pixel =: n), where $m \ll n$. Since the projection matrix encodes the incident relation between rays and voxels/pixels, the projection matrix A will be sparse. Based on additional assumptions on u, we will devise in this paper conditions for exact recovery of u from the *underdetermined* linear system $Au = b$.

2.1.   Imaging Set-Up

For simplicity, we will assume that Ω is a cube in 3D or a square in 2D and that $\Omega = [0, d]^3$ is discretized into d^3 voxels, while $\Omega = [0, d]^2$ is discretized into d^2 pixels. We consider a parallel ray geometry and choose the projection angles such that the intersection of each line with all adjacent cells is constant, thus yielding binary projection matrices after scaling. This simplification is merely made in order to obtain a structure in the projection matrix which allows to compute relevant combinatorial measures. We stress however that other discretization choices are possible. As shown in the numerical section, perturbed versions of these matrices lead to similar results in terms of performance.

2.1.1. 2D Case: $3, \ldots, 8$ Views

We set $\Omega = [0, d]^2$ and obtain the binary projection matrices according to (2.1) from few projecting directions (three to eight), compare Fig. 3. We summarize the used parameters in Table 1.
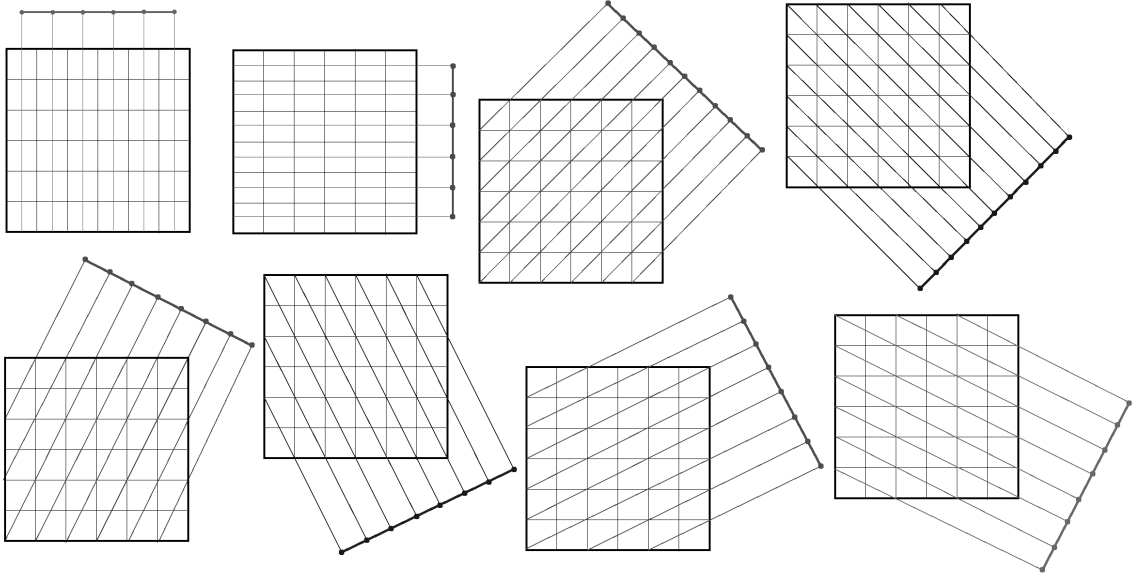

Figure 3. Eight different projecting directions and corresponding projecting rays for $90°, 0°, \mp 45°, \mp \arctan(2)$ and $\mp \arctan(0.5)$ (from left to right, top to bottom). Note that the intersection segments for each projection ray with all adjacent pixel are equal. As a consequence, we obtain after appropriate scaling, binary projection matrices. Each sensor resolution varies with the projection angle, however. The illustration above depicts $\Omega = [0, d]^2$ with $d = 6$.

Table 1. Dimensions of projection matrices in 2D.

# views	m	n	projection angles
3	$4d - 1$	d^2	$0°, 90°, 45°$
4	$6d - 2$	d^2	$0°, 90°, \mp 45°$
5	$7d + \lfloor \frac{d}{2} \rfloor - 2$	d^2	$0°, 90°, \mp 45°, \arctan(2)$
6	$8d + 2\lfloor \frac{d}{2} \rfloor - 2$	d^2	$0°, 90°, \mp 45°, \mp \arctan(2)$
7	$9d + 3\lfloor \frac{d}{2} \rfloor - 2$	d^2	$0°, 90°, \mp 45°, \mp \arctan(2), \arctan(0.5)$
8	$10d + 4\lfloor \frac{d}{2} \rfloor - 2$	d^2	$0°, 90°, \mp 45°, \mp \arctan(2), \mp \arctan(0.5)$

2.1.2.   **3D Case: 3 or 4 Views**

We consider the imaging set-up depicted by Fig. 4 and Fig. 5. The projection angles were chosen again such that the intersection of each ray with all adjacent voxels is constant. After appropriate scaling, the resulting measurement matrices are binary as well.
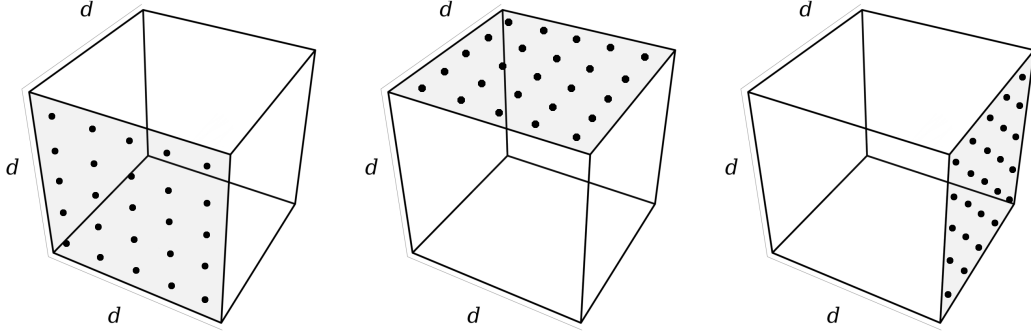

Figure 4.   Imaging set-up for three orthogonal projections corresponding to each shaded plane of the cube. From left to right: Cell centers projected along each direction are shown as dots for the case $d = 5$. The cube $\Omega = [0, d]^3$ is discretized into d^3 cells and projected along $3 \cdot d^2$ rays.
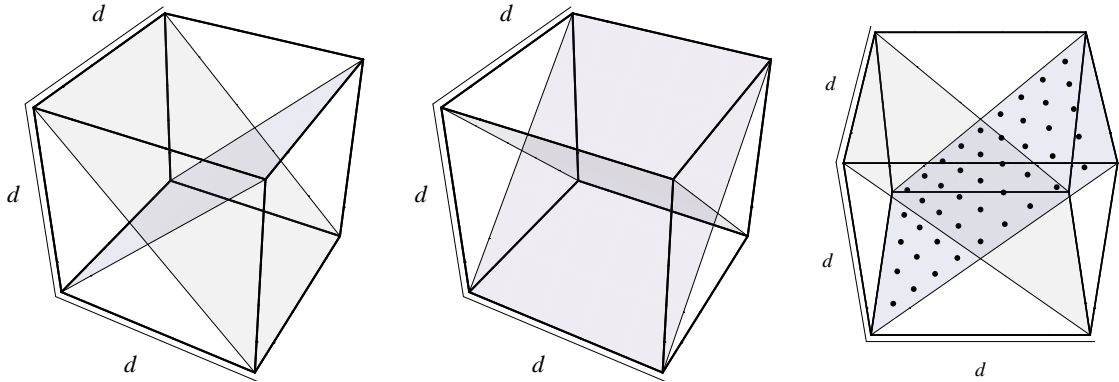

Figure 5.   Imaging set-up for four projecting directions corresponding to the image planes shown as two pairs in the left and center panel, respectively. Right panel: Voxel centers projected onto the first image plane are shown as dots for the case $d = 5$. The cube $\Omega = [0, d]^3$ is discretized into d^3 voxel and projected along $4 \cdot d(2d - 1)$ rays.

2.2.   **Complete Rank and RIP**

For recovery of k-sparse signals by *compressed sensing (CS)* both necessary and sufficient conditions have been provided, which not only depend on the sparsity k of the original signal, but also on the conditions of the sensing matrix A. In particular, compressed sensing aims for a matrix which has high spark, also known as complete rank, high nullspace property order, and a small RIP constant, as detailed next.

Definition 2.1. ([12])
Let $A \in \mathbb{R}^{m \times n}$ be an arbitrary matrix. Then the *spark* of A denoted by $\mathrm{spark}(A)$ is the minimal number of linearly dependent columns of A.

Any k-sparse solution $\overline{u}$ of a linear system $Au = b$ is unique if $\|\overline{u}\|_0 = k < \mathrm{spark}(A)/2$.

Due to the fact that A is underdetermined, the nullspace of A also plays a particular role in the analysis of uniqueness of the minimization problem (1.1). The related so-called *nullspace property (NSP)* is defined as follows.

Definition 2.2. Let $A \in \mathbb{R}^{m \times n}$ be an arbitrary matrix. Then A has the *nullspace property (NSP) of order* k if, for all $v \in \mathcal{N}(A) \setminus \{0\}$ and for all index sets $|S| \le k$, $\|v_S\|_1 < \frac{1}{2}\|v\|_1$.

Any k-sparse solution $\overline{u}$ of a linear system $Au = b$ is the unique solution of (1.1), if A satisfies the nullspace property of order k. For nonnegative signals, the NSP can be characterized in terms of the minimal number of negative components in the sparsest nullspace vector.

Proposition 2.3. Every k-sparse nonnegative vector $\overline{u}$ is the unique positive solution of $Au = A\overline{u}$ if and only if every nonzero nullspace vector has at least $k + 1$ negative (and positive) entries.

Table 2.   Properties of projection matrices in 2D and 3D.

dim.	# views	m	n	$\mathrm{rank}(A)$	$\mathrm{spark}(A)$	NSP
$q = 2$	3	$4d - 1$	d^2	$4d - 4$	6	2
	4	$6d - 2$	d^2	$6d - 9$	8	3
	5	$7d + \lfloor \frac{d}{2} \rfloor - 2$	d^2	$7d + \lfloor \frac{d}{2} \rfloor - 13$	12	5
	6	$8d + 2\lfloor \frac{d}{2} \rfloor - 2$	d^2	$8d + 2\lfloor \frac{d}{2} \rfloor - 19$	16	7
	7	$9d + 3\lfloor \frac{d}{2} \rfloor - 2$	d^2	$9d + 3\lfloor \frac{d}{2} \rfloor - 23$	16	7
	8	$10d + 4\lfloor \frac{d}{2} \rfloor - 2$	d^2	$10d + 4\lfloor \frac{d}{2} \rfloor - 29$	16	7
$q = 3$	3	$3d^2$	d^3	$3d^2 - 3d + 1$	8	3
	4	$8d^2 - 4d$	d^3	$8d^2 - 20d + 16$	15	6

The *restricted isometry property (RIP)*, defined next, characterizes matrices which are well conditioned when operating on sparse vectors. This is probably the most popular CS condition since it also enables *stable* recovery.

Definition 2.4. A matrix A is said to have the *Restricted Isometry Property* $RIP_{\delta,k}$ if, for any k-sparse vector u, the relation

$$(1 - \delta)\|u\|^2 \le \|Au\|^2 \le (1 + \delta)\|u\|^2, \quad \delta \in (0, 1) \tag{2.2}$$

holds.

This property implies that every submatrix $(A^{i_1}, \dots, A^{i_k})$ formed by keeping at most k-columns of A has nonzero singular values bounded from above by $1 + \delta$ and from below by $1 - \delta$. In particular, (2.2) implies that a matrix A cannot satisfy $RIP_{\delta,k}$ if $k \geq \mathrm{spark}(A)$.

Candès has shown [3, Thm. 1.1] that if $A \in RIP_{\delta,2k}$ with $\delta < \sqrt{2} - 1$, then all k-sparse solutions $\overline{u}$ of (1.1) are unique. Moreover, when measurements are corrupted with noise

$$b = A\overline{u} + \nu,$$

where ν is an unknown noise term, recovery of $\overline{u}$ is usually performed by

$$\min \|u\|_1 \qquad \text{s.t.} \quad \|Au - b\|_2 \leq \varepsilon \tag{2.3}$$

where ε is an upper bound on the size of ν. As shown in [3, Thm. 1.2], the RIP condition also implies *stable* recovery even in case of observation errors. Provided that the observation error is small enough, $\|\nu\|_2 \leq \varepsilon$, and $A \in RIP_{\delta,2k}$ with $\delta < \sqrt{2} - 1$, then the solution u of (2.3) obeys

$$\|u - \overline{u}\|_2 \leq C_0 k^{-\frac{1}{2}} \|\overline{u} - (\overline{u})_k\|_2 + C_1 \varepsilon,$$

where C_0 and C_1 are explicit constants depending on δ, and $(\overline{u})_k$ is the vector $\overline{u}$ with all but the k-largest entries set to zero.

It has been shown in [6] that binary matrices cannot satisfy $RIP_{\delta,k}$ unless the numbers of rows is $\Omega(k^2)$.

Theorem 2.5. [6, Thm. 1] Let $A \in \mathbb{R}^{m \times n}$ be any $0/1$-matrix that satisfies $RIP_{\delta,k}$. Then

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

Taking into account that there exists $\mathrm{spark}(A)$-columns in A which are linearly dependent, we obtain, together with $m \ll n$, the following result.

Corollary 2.6. Let $\delta \in (0, 1)$. A necessary condition for A to satisfy the $RIP_{\delta,k}$ for all k-sparse vectors is that

$$k \leq \min \left\{ \frac{1+\delta}{1-\delta} m^{\frac{1}{2}}, \mathrm{spark}(A) - 1 \right\}.$$

Application to our scenarios. For our particular matrices A defined in Sections 2.1.1 and 2.1.2 we obtain, along the lines of [29, Prop. 3.2], that $\mathrm{spark}(A)$ is a constant for all dimensions d with $m < n$, while the number of measurements obeys $O(d^{q-1})$, $q \in \{2, 3\}$, see Table 2. Compare also Fig. 6, left. However, we cannot be sure that A possesses the $RIP_{\sqrt{2}-1,\sigma}$, with $\sigma = \mathrm{spark}(A) - 1$, unless we compute the singular values of all submatrices containing σ or less columns of A.

The previous results show that the poor properties of A, from the viewpoint of CS, rest upon the small spark of A. In order to increase the maximal number of columns such that all column collections of size k (or less) are linearly independent, we can add to the entries of A small random numbers. Due to the fact that $\mathrm{rank}(A)$ almost equals m in all considered situations, the probability that k-arbitrary columns are linearly independent slowly decreases from 1, when $k < \mathrm{spark}(A)$, to 0, when $k > \mathrm{rank}(A)$.
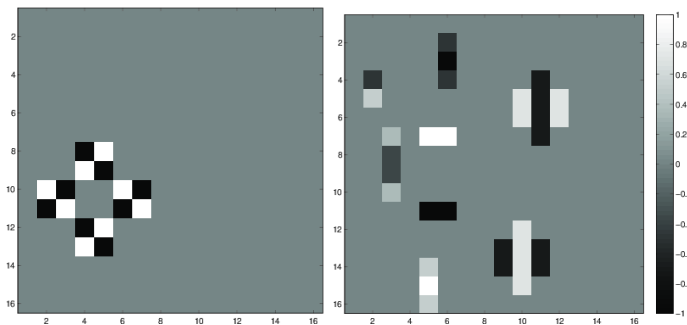
Figure 6.   The *sparsest* nullspace vector $u_{\mathcal{N}} \in \mathcal{N}(A) \setminus \{0\}$ is shown on the left as a 16×16 image, for matrices A from Section 2.1.1 with 6,7 or 8 projecting directions, where $d = 16$, compare Table 1. Gray indicates components with value 0, white the value 1 and black the value -1. Projections along all rays depicted in Fig. 3 sum up to zero. This shows that $\mathrm{spark}(A) = 16$ and the matrix has a NSP of order 7. These numbers do not change with the problem's size for any $d \geq 16$. The image on the right depicts the bivariate Haar-transformed nullspace basis vector $Hu_{\mathcal{N}}$, which has 32 nonzero elements.

The perturbed matrix $\tilde{A}$ is computed by uniformly perturbing the non-zero entries $A_{ij} > 0$ to obtain $\tilde{A}_{ij} \in [A_{ij} - \epsilon, A_{ij} + \epsilon]$, and by normalizing subsequently all column vectors of $\tilde{A}$. In practice, such perturbations can be implemented by discretizing the image by different basis functions or choose their locations on an irregular grid.

As argued in Section 1.1 and illustrated by Figures 1 and 2, considering functions $u(x)$ with sparse gradients and the corresponding optimization criterion $\mathrm{TV}(u)$ for recovery (1.2), we may boost recovery performance in severely undersampled tomographic scenarios, despite the poor properties of measurement matrices A.

3.   Sparsity and TV-Based Reconstruction

As discussed in Section 1.2, it has been well known empirically that solving the problem

$$\min \|\nabla u\|_1 \qquad \text{s.t.} \quad \|Au - b\|_2 \leq \varepsilon \qquad (3.1)$$

can provide high-quality and stable image recovery. Until the recent work [27], however, it had been an open problem to provide provable *theoretical* guarantees, beyond incomplete Fourier measurements [8, 7]. The gradient operator ∇ is not an orthonormal basis or a tight frame, thus neither the standard theory of CS nor the theoretical extensions in [5] concerning the *analysis model* apply to (3.1), even for images with truly sparse gradient ∇u.

The recent work [27, 26] *proves* that stable recovery is possible via the convex program (3.1) and considers a general matrix A which is incoherent with the multidimensional Haar wavelet transform and satisfies a RIP condition. The Haar wavelet transform provides a sparsifying basis for 2D and 3D images and is closely related to the discrete gradient operator. In the remainder of this section, we denote the discrete multidimensional Haar wavelet transform by H and refer the reader to the definition of [26, p. 6]. The following theorem summarizes the main results of [27, Thm. 5, Thm. 6], and [26] and specializes them to the case of *anisotropic* TV (1.7) as considered in the present paper, see also the Remarks following [27, Thm. 6] and [26, Main Thm.].

Theorem 3.1. Let $d = 2^N$ be a power of two and $n = d^q$, $q \in \{2, 3\}$. Further, let H be the discrete multidimensional Haar transform, and let $A \in \mathbb{R}^{m \times n}$ such that AH^{-1} satisfies $RIP_{\delta,5k}$ with $\delta < \frac{1}{3}$. Then for any $\overline{u} \in \mathbb{R}^n$ with $b = A\overline{u} + \nu$ and $\|\nu\|_2 \leq \varepsilon$, the solution u of (3.1) satisfies the *gradient error bound*

$$\|\nabla u - \nabla \overline{u}\|_1 \leq \|\nabla \overline{u} - (\nabla \overline{u})_k\|_1 + \sqrt{k}\varepsilon,$$

and the *signal error bound*

$$\|u - \overline{u}\|_2 \leq \log\left(\frac{n}{k}\right) \frac{\|\nabla \overline{u} - (\nabla \overline{u})_k\|_1}{\sqrt{k}} + \varepsilon.$$

Note that recovery is exact when $\nabla \overline{u}$ is exactly k-sparse and $\|\nu\|_2 = 0$.

The RIP assumption on $AH^{-1} = AH^\top$ implies that $\mathcal{N}(A)$ cannot contain any signals admitting a k-sparse wavelet expansion, apart from the zero vector, since $\|Av\|_2 = \|AH^{-1}Hv\|_2 \approx (1 \pm \delta)\|Hv\|_2 = (1 \pm \delta)\|v\|_2$, with the last equality holding because $H^\top H = I$.

There exist sensing matrices $A \in \mathbb{R}^{m \times n}$ which satisfy the above conditions, e.g. $RIP_{1/3,k}$, 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 and other orthogonal matrices.

In our scenario, however, due to the low RIP order of the tomographic projection matrix A, for any image dimension d, the RIP order of AH^{-1} does not improve significantly. To illustrate this point, let us consider further the bivariate discrete Haar transform H and a sparse nullspace vector $u_\mathcal{N}$ with $\|u_\mathcal{N}\|_0 = 16$, depicted in Fig. 6, left panel, of the 2D projection matrix A from 6, 7 or 8 projections. Then $0 = \|Au_\mathcal{N}\|_2 = \|AH^{-1}Hu_\mathcal{N}\|_2$ holds with $\|Hu_\mathcal{N}\|_0 = 32$, see Fig. 6, right. Thus, the matrix AH^{-1} cannot satisfy RIP of an order larger than $32 - 1$, and this holds for any $d \geq 16$. Consequently, suppose it does accordingly satisfy $RIP_{1/3,31}$, then Thm. 3.1 would imply exact recovery of any image with a 6-sparse image gradient. Unfortunately, such an extremely low sparsity is of limited use for practical applications.

4. Cosparsity and TV-Based Reconstruction

We introduce some basic definitions related to the cosparsity of a given analysis operator B in Section 4.1 followed by uniqueness results from [20, 24] in Section 4.2. The corresponding conditions imply bounds for the number m of measurements, depending on the cosparsity of the vector u, that should be reconstructed, with respect to B. We apply these results in Section 4.3 to the discrete gradient operator $B = \nabla$ given by (1.6). This requires to estimate the dimension of the subspace of ℓ-cosparse vectors. We relate this problem to the isoperimetric problem on grid graphs studied by [2]. In this more general way, we reproduce the estimate proved differently in [24] for the 2D case and additionally provide an estimate for the 3D case.

4.1. Definitions

Let B be any given analysis operator.

Definition 4.1. (cosparsity, cosupport)
The *cosparsity* of $u \in \mathbb{R}^n$ *with respect to* $B \in \mathbb{R}^{p \times n}$ is

$$\ell := p - \|Bu\|_0, \tag{4.1}$$

and the *cosupport* of *u with respect to B* is

$$\Lambda := \{r \in [p] \colon (Bu)_r = 0\}, \qquad |\Lambda| = \ell. \tag{4.2}$$

We denote by B_r the r-th row of the matrix B and by B_Λ the submatrix of B formed by the *rows* indexed by $\Lambda \subset [p]$. Thus, a ℓ-cosparse vector u satifies $B_\Lambda u = 0$, hence is contained in the subspace

$$\mathcal{W}_\Lambda := \mathcal{N}(B_\Lambda). \tag{4.3}$$

In this connection, we define the basic function

$$\kappa_B(\ell) := \max_{|\Lambda| \geq \ell} \dim \mathcal{W}_\Lambda. \tag{4.4}$$

4.2. Basic Uniqueness Results

This section collects some results from [24] that were derived based on [20].

Proposition 4.2. (Uniqueness with known cosupport)
Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times n}$ be given measurement and analysis operators and assume the rows of the matrix $\left(\begin{smallmatrix} A \\ B \end{smallmatrix} \right)$ are linearly independent. Then, if the cosupport $\Lambda \subset [p]$ of the ℓ-cosparse vector $u \in \mathbb{R}^n$ is known, the condition

$$\kappa_B(\ell) \leq m \tag{4.5}$$

is necessary and sufficient for recovery of every such vector from the measurements $b = Au$.

Proposition 4.2 says that if the dimension of the subspace $\mathcal{W}_\Lambda$ increases, then more measurements are needed for recovery of ℓ-cosparse vectors $u \in \mathcal{W}_\Lambda$. The dimension $\dim \mathcal{W}_\Lambda$ increases for decreasing ℓ.

Proposition 4.3. (Uniqueness with unknown cosupport)
Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times n}$ be given measurement and analysis operators, and assume the rows of the matrix $\left(\begin{smallmatrix} A \\ B \end{smallmatrix} \right)$ are linearly independent. Then a necessary condition for uniqueness of a ℓ-cosparse solution u to the measurement equations $Au = b$ is

$$\tilde{\kappa}_B(\ell) \leq m, \qquad \tilde{\kappa}_B(\ell) := \max \big\{ \dim(\mathcal{W}_{\Lambda_1} + \mathcal{W}_{\Lambda_2}) \colon |\Lambda_i| \geq \ell,\, i = 1, 2 \big\}, \tag{4.6}$$

whereas a sufficient such condition is

$$\kappa_B(\ell) \leq \frac{m}{2}, \tag{4.7}$$

with κ_B from (4.4).

Roughly speaking, lack of knowledge of Λ implies the need of twice the number of measurements for unique recovery.

Remark 4.4. Both propositions assume the rows of A and B to be independent. This is neither the case for typical sensor matrices A used in discrete tomography nor in the specific case $B = \nabla$ considered next.

Our experimental results will show, however, that the estimates of $\kappa_B(\ell) = \kappa_\nabla(\ell)$ derived in Section 4.3 correctly display the relationship between the basic parameters involved, up to some scale factor discussed in Section 6.

4.3. Application to the Analysis Operator ∇

In order to apply the results of Section 4.2, the function (4.4) has to be evaluated, or estimated, in the case $B = \nabla$.

For a given cosupport $\Lambda \subset E$, define the set of vertices covered by Λ,

$$V(\Lambda) = \{v \in V : v \in e \text{ for some } e \in \Lambda\}, \tag{4.8}$$

and denote the number of connected components of $V(\Lambda)$ by $|V(\Lambda)|_\sim$. Due to definition (1.6) of the analysis operator ∇ and (4.2), each component $(\nabla_\Lambda u)_i$ corresponds to an edge $e = (v_1, v_2)$ with $u(v_1) = u(v_2)$. Therefore, following the reasoning in [24], $u \in \mathcal{W}_\Lambda = \mathcal{N}(\nabla_\Lambda)$ if and only if u is constant on each connected component of $V(\Lambda)$. Hence $\dim \mathcal{W}_\Lambda$ equals the size of the remaining vertices $|V \setminus V(\Lambda)|$ plus the degree of freedom for each connected component,

$$\dim \mathcal{W}_\Lambda = |V| - |V(\Lambda)| + |V(\Lambda)|_\sim. \tag{4.9}$$

Now, in view of (4.4), consider some Λ with $|\Lambda| = \ell$ and the problem

$$\max_{\Lambda : |\Lambda| = \ell} \dim \mathcal{W}_\Lambda = |V| - \min_{\Lambda : |\Lambda| = \ell} (|V(\Lambda)| - |V(\Lambda)|_\sim). \tag{4.10}$$

Clearly, the minimal value of the last term is $|V(\Lambda)|_\sim = 1$. It will turn out below that this value is attained for extremal sets Λ and that the maximum in (4.4) is achieved for $|\Lambda| = \ell$.

We therefore temporarily ignore the last term and focus on the second term. The problem is to *minimize* over all subsets $\Lambda \subset E$ of cardinality $|\Lambda| = \ell$ the number $|V(\Lambda)|$ of vertices covered by Λ. We establish this relationship by considering instead the problem of *maximizing* the set Λ over all sets $S := V(\Lambda) \subseteq V$ of fixed cardinality $s = |S(\Lambda)|$. This problem was studied in [2] for regular grid graphs $G = (V, E)$ with vertex set $V = [d]_0^q$ with equal dimension along each coordinate, in terms of the problem

$$\max_{S : |S| = s} |\text{Int}_e(S)|, \tag{4.11}$$

where $\text{Int}_e(S)$ denotes the *edge interior* of a set $S \subset V(G)$,

$$\text{Int}_e(S) := \{(v_1, v_2) \in E : v_1, v_2 \in S\}, \tag{4.12}$$

which equals $\text{Int}_e(S) = \Lambda$ for our definition $S = V(\Lambda)$.

Theorem 4.5. ([2, Thm. 13])
Let S be a subset of $[d]_0^q$, with $s = |S|$, $d \geq 3$, $q \geq 2$. Then

$$|\text{Int}_e(S)| \leq \max \left\{ qs(1 - s^{-1/q}), qd^q(1 - 1/d)\left(1 - (1 - s/d^q)^{1 - 1/q}\right) \right\}. \tag{4.13}$$

Some sets $S = V(\Lambda) \subseteq V$ corresponding to maximal sets $\Lambda = \text{Int}_e(S)$ are also determined in [2]. The following corresponding assertion is based on the *cube order* or vertices $v \in V = [d]_0^q$ (identified with grid vectors $v = (v_1, \ldots, v_q)^\top$, cf. Section 1.4): $v \prec v' \Leftrightarrow w(v) < w(v')$, where $w(v) = \sum_{i \in [q]} 2^{i+qv_i}$. See Figure 7 for an illustration.
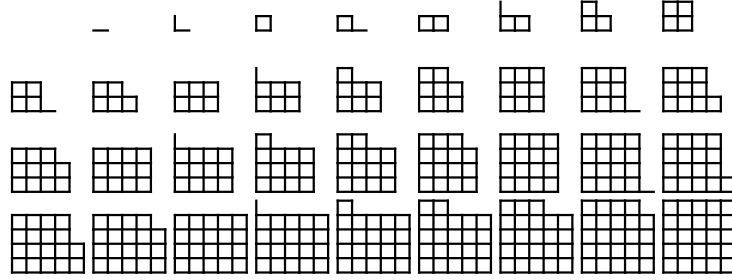
Figure 7. From left to right, top to bottom: Edge sets Λ corresponding to the subsets $S = V(\Lambda) \subseteq V = [d]_0^q$, $d = 5$, $q = 2$, of cube-ordered vertices of cardinalities $s = |S| = 1, 2, \ldots, |V|$. According to Thm. 4.6, these sets belong to the maximizers of $|\Lambda|$ among all subsets $V(\Lambda) \subseteq V$ with fixed $s = |V(\Lambda)|$.

Theorem 4.6. ([2, Thm. 15])
Let $S' \subset V = [d]^q$, and let S be the set of the first $s = |S'|$ vertices in the cube order on V. Then $|\text{Int}_e(S')| \leq |\text{Int}_e(S)|$.

Thm. 4.6 says (cf. Fig. 7) that singly connected mimimal sets $V(\Lambda)$ in (4.10) are achieved, that is $|V(\Lambda)|_\sim = 1$. Furthermore, these sets $\{\Lambda\}_{|\Lambda| \geq \ell}$ are nested. Hence the maximum in (4.4) is achieved for $|\Lambda| = \ell$.

A closer inspection of the two terms defining the upper bound (4.13) shows that the first term of the r.h.s. is larger if $s \geq q^q = 4$ in the 2D case $q = 2$, respectively, if $s \geq q^q = 27$ in the 3D case $q = 3$. The corresponding values of the bound are $|\text{Int}_e(S)| \leq 4$ and $|\text{Int}_e(S)| \leq 54$, respectively. As a consequence, we consider the practically relevant first term. Setting $\ell = |\text{Int}_e(S)| = |V(\Lambda)|$ and solving the equality for s (due to Thm. 4.6) yields

$$s = \frac{1}{2}(1 + \ell + \sqrt{1 + 2\ell}) \qquad\qquad (q = 2) \qquad (4.14a)$$

$$s = \frac{1}{3}\left(2^{1/3}\frac{1 + 2\ell}{t(\ell)} + \left(1 + \ell + \frac{1}{2^{1/3}}t(\ell)\right)\right) \qquad\qquad (q = 3) \qquad (4.14b)$$

$$t(\ell) = \left(2 + 6\ell + 3\ell^2 + \sqrt{(4 + 9\ell)\ell^3}\right)^{1/3} \qquad\qquad (4.14c)$$

$$\geq \frac{1}{3}\left(1 + \ell + (3\ell^2)^{1/3} + 2(\ell/3)^{1/3}\right) + O(\ell^{-1/3}). \qquad\qquad (4.14d)$$

Inserting s, or simpler terms lower bounding s, for $|V(\Lambda)|$ in (4.10) and putting all conclusions together, yields for (4.4) and $B = \nabla$:

Lemma 4.7. Let $G = (V, E)$ be a regular grid graph with $V = [d]_0^q$, $n = |V| = d^3$. Then

$$\forall \ell > 4, \qquad \kappa_\nabla(\ell) \leq n - \frac{1}{2}(\ell + \sqrt{1 + 2\ell}) + \frac{1}{2}, \qquad (q = 2), \qquad (4.15a)$$

$$\forall \ell > 54, \qquad \kappa_\nabla(\ell) \leq n - \frac{1}{3}\left(\ell + \sqrt[3]{3\ell^2} + 2\sqrt[3]{\frac{\ell}{3}}\right) + \frac{2}{3}, \qquad (q = 3). \qquad (4.15b)$$
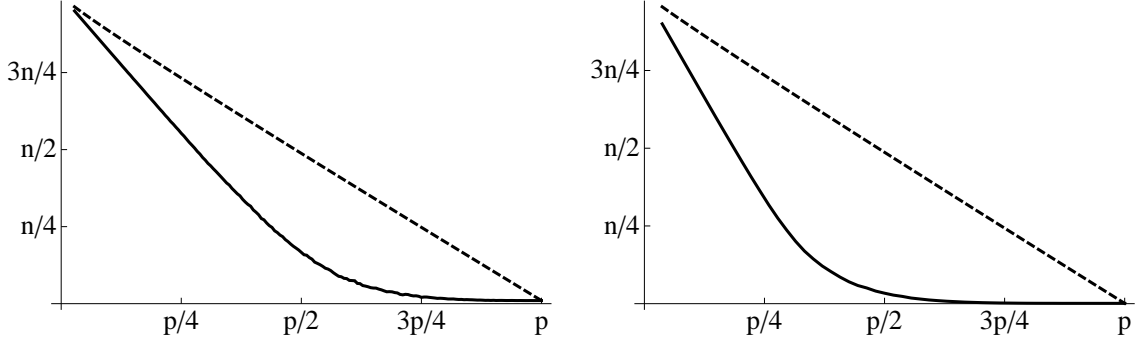
Figure 8 illustrates these bounds.

Figure 8. The bounds (4.15a) and (4.15b) shown in the left and right panel respectively, as dashed lines, for $d = 10$, as a function of ℓ. The solid lines show empirical expected values of $\kappa_\nabla(\ell)$ computed by averaging over 100 analysis matrices B_Λ for each value $\ell = |\Lambda|$. The gap to the upper bound simply shows that the extremal sets discussed in connection with Theorem 4.6 are not observed in random experiments. For example, it is quite unlikely that random cosupports Λ are singly connected.

Remark 4.8. Up to the constant 1 under the square root, the bound (4.15a) for the 2D case equals the bound derived in a different way in [24]. We provided the bounds (4.15) based on general results in [2] that apply to grid graphs in any dimension $q \geq 2$.

We conclude this section by applying Propositions 4.2 and 4.3.

Corollary 4.9. Under the assumptions of Propositions 4.2 and 4.3, a ℓ-cosparse solution u to the measurement equations $Au = b$ will be unique if the number of measurements satisfies

$$m \geq n - \frac{1}{2}\Big(\ell + \sqrt{2\ell + 1} - 1\Big) \qquad (q = 2), \qquad (4.16a)$$

$$m \geq n - \frac{1}{3}\Big(\ell + \sqrt[3]{3\ell^2} + 2\sqrt[3]{\frac{\ell}{3}} - 2\Big) \qquad (q = 3) \qquad (4.16b)$$

in case the cosupport Λ is known, and

$$m \geq 2n - (\ell + \sqrt{2\ell + 1} - 1) \qquad (q = 2), \qquad (4.17a)$$

$$m \geq 2n - \frac{2}{3}\Big(\ell + \sqrt[3]{3\ell^2} + 2\sqrt[3]{\frac{\ell}{3}} - 2\Big) \qquad (q = 3) \qquad (4.17b)$$

in case the cosupport Λ is unknown.

The above derived bounds on the required image cosparsity guarantees uniqueness in case of known or unknown cosupport, and imply that recovery can be carried out via

$$\min_u \|Bu\|_0 \quad \text{s.t.} \quad Au = b, B_\Lambda u = 0, \qquad (4.18)$$

when the cosupport is known, or via

$$\min_u \|Bu\|_0 \quad \text{s.t.} \quad Au = b, \qquad (4.19)$$

when the cosupport is unknown. In Section 6, we compare these relationships to numerical results involving convex relaxations of (4.18) and (4.19), studied in Section 5.

5. Recovery by Linear Programming

In this section, uniqueness of the optimum $\overline{u}$ solving problem (1.3) is studied. The resulting condition is necessary and sufficient for unique recovery $\overline{u} = u^*$ of any ℓ-cosparse vector u^* that satisfies $Au^* = b$ and has cosupport Λ, $|\Lambda| = \ell$, with respect to the analysis operator $B = \nabla$.

We turn problem (1.3) into a standard linear programming formulation. Defining

$$M := \begin{pmatrix} B & -I & I \\ A & 0 & 0 \end{pmatrix}, \quad q := \begin{pmatrix} 0 \\ b \end{pmatrix} \tag{5.1}$$

and the polyhedral set

$$\mathcal{P} := \{w \in \mathbb{R}^{n+2p} : Mw = z, \ w \geq 0\}, \quad w := \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u \\ v^1 \\ v^2 \end{pmatrix}, \tag{5.2}$$

problem (1.3) equals the linear program (LP)

$$\min_{w \in \mathcal{P}} \langle c, w \rangle = \min_{(u,v^1,v^2) \in \mathcal{P}} \langle \mathbb{1}, v^1 + v^2 \rangle, \quad c = \begin{pmatrix} 0 \\ \mathbb{1} \\ \mathbb{1} \end{pmatrix}. \tag{5.3}$$

Let $\overline{w} = (\overline{u}, \overline{v}) = (\overline{u}, \overline{v}^1, \overline{v}^2)$ solve (5.3). We assume throughout

$$\overline{u}_i > 0, \ i \in [n] \tag{5.4}$$

which is not restrictive with respect to applications (u may e.g. represent strictly positive material densities). Based on $\overline{w}$, we define the corresponding index sets

$$J := \{i \in [\dim(w)] : \overline{w}_i = 0\}, \quad \overline{J} := \{i \in [\dim(v)] : \overline{v}_i = 0\}, \quad w_J = v_{\overline{J}}, \ \forall w = \begin{pmatrix} u \\ v \end{pmatrix}. \tag{5.5}$$

Theorem 5.1. ([23, Thm. 2(iii)])
Let $\overline{w}$ be a solution of the linear program (5.3). The following statements are equivalent:

(i) $\overline{w}$ is unique.

(ii) There exists no w satisfying

$$Mw = 0, \quad w_J \geq 0, \quad \langle c, w \rangle \leq 0, \quad w \neq 0. \tag{5.6}$$

We turn Theorem (5.1) into a *nullspace condition* w.r.t. the sensor matrix A, for the unique solvability of problems (5.3) and (1.3). This condition is stated as Corollary 5.3 below, after a preparatory Lemma.

Lemma 5.2. Let $\overline{w}$ be a solution of the LP (5.3). Then the cardinality of the index set $\overline{J}$ defined by (5.5) is

$$|\overline{J}| = 2\ell + k = p + \ell, \quad |\overline{J}^c| = 2p - |\overline{J}| = k, \quad k := p - \ell. \tag{5.7}$$

Proof:
The minimal objective function value (5.3) is $\sum_{i\in[p]} \overline{v}_i^1 + \overline{v}_i^2$ with all summands being non-negative. Since $B\overline{u} = \overline{v}^1 - \overline{v}^2$, $(B\overline{u})_\Lambda = 0$ and optimality of $\overline{v}$ imply $\overline{v}_\Lambda^1 = \overline{v}_\Lambda^2 = 0$, which contributes $2|\Lambda| = 2\ell$ indices to $\overline{J}$. Furthermore, if $(B\overline{u})_i = \overline{v}_i^1 - \overline{v}_i^2 < 0$, then optimality of $\overline{v}$ implies $\overline{v}_i^1 = 0$, $\overline{v}_i^2 > 0$ and vice versa if $(B\overline{u})_i > 0$. Hence Λ^c supports $|\Lambda^c| = p - \ell = k$ vanishing components of $\overline{v}$. $\qquad\square$

Corollary 5.3. Let $\overline{w} = (\overline{u}, \overline{v}^1, \overline{v}^2)$ be a solution of the linear program (5.3) with corresponding index sets $J, \overline{J}$ given by (5.5), and with component $\overline{u}$ that solves problem (1.3) and has cosupport Λ with respect to B. Then $\overline{w}$ resp. $\overline{u}$ are unique if and only if

$$\forall w = \begin{pmatrix} u \\ v \end{pmatrix}, \ v = \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} \quad \text{s.t.} \quad u \in \mathcal{N}(A) \setminus \{0\} \quad \text{and} \quad Bu = v^1 - v^2 \tag{5.8}$$

the condition

$$\|(Bu)_\Lambda\|_1 > \langle (Bu)_{\Lambda^c}, \mathrm{sign}(B\overline{u})_{\Lambda^c} \rangle \tag{5.9}$$

holds. Furthermore, any unknown ℓ-cosparse vector u^* with $Au^* = b$ can be uniquely recovered as solution $\overline{u} = u^*$ to (1.3) if and only if, for all vectors u conforming to (5.8), the condition

$$\|(Bu)_\Lambda\|_1 > \sup_{\Lambda \subset [p]:\, |\Lambda| = \ell} \ \sup_{\overline{u} \in \mathcal{W}_\Lambda} \ \langle (Bu)_{\Lambda^c}, \mathrm{sign}(B\overline{u})_{\Lambda^c} \rangle \tag{5.10}$$

holds.

Remark 5.4. Condition (5.9) corresponds up to a magnitude $|\cdot|$ operation applied to the right-hand side to the statement of [24, Thm. 7]. The authors do not present an explicit proof, but mention in [24, App. A] that the result follows by combining a strictly local minimum condition with convexity of the optimization problem for recovery.

  Our subsequent explicit proof elaborates basic LP-theory due to [23] and Thm. 5.1.

Proof:
[Proof of Corollary 5.3] Theorem (5.1) asserts that $\overline{w}$ is unique if and only if for every $w \in \mathcal{N}(M) \setminus \{0\}$ with $w_J \geq 0$ the condition $\langle c, w \rangle > 0$ holds. In view of the definition (5.1) of M, vectors $w \in \mathcal{N}(M) \setminus \{0\}$ are determined by (5.8). Condition (5.8) excludes vectors $0 \neq w = (0, v^1, v^2) \in \mathcal{N}(M)$ because then $v^1 = v^2$ and $w_J \geq 0$ implies exclusion of those w by $\langle c, w \rangle \leq 0$ in (5.6).

  It remains to turn the condition (5.6) into a condition for vectors u given by vectors $w = (u, v^1, v^2)$ satisfying (5.8). To this end, we focus on such vectors w with $w_J \geq 0$ that minimize $\langle c, w \rangle$. We have $w_J = v_{\overline{J}}$ by (5.5), and the proof of Lemma 5.2 shows that $v_{\overline{J}} \geq 0$ decomposes into

- 2ℓ conditions $v_\Lambda^1, v_\Lambda^2 \geq 0$ leading to the choice

$$\begin{cases} v_i^1 = (Bu)_i \geq 0, \quad v_i^2 = 0, & \text{if } (Bu)_i \geq 0, \\ v_i^1 = 0, \quad v_i^2 = -(Bu)_i \geq 0, & \text{if } (Bu)_i \leq 0, \end{cases} \quad i \in \Lambda \tag{5.11}$$

  minimizing $\langle c, w \rangle$;

- k conditions supported by Λ^c of the form: either $v_i^1 \geq 0$ or $v_i^2 \geq 0$ depending on $(B\overline{u})_i > 0$ or $(B\overline{u})_i < 0$, $i \in \Lambda^c$. In order to minimize $\langle c, w \rangle$, this leads to the choice

$$\begin{cases} v_i^1 = 0, & v_i^2 = -(Bu)_i \leq 0, & \text{if } (Bu)_i \geq 0, (B\overline{u})_i > 0, \\ v_i^1 = 0, & v_i^2 = (Bu)_i \geq 0, & \text{if } (Bu)_i \leq 0, (B\overline{u})_i > 0, \\ v_i^1 = (Bu)_i \geq 0, & v_i^2 = 0, & \text{if } (Bu)_i \geq 0, (B\overline{u})_i < 0, \\ v_i^1 = (Bu)_i \leq 0, & v_i^2 = 0, & \text{if } (Bu)_i \leq 0, (B\overline{u})_i < 0, \end{cases} \quad i \in \Lambda^c. \quad (5.12)$$

By (5.3), $\langle c, w \rangle = \langle \mathbb{1}, v^1 + v^2 \rangle = \langle \mathbb{1}, (v^1 + v^2)_\Lambda \rangle + \langle \mathbb{1}, (v^1 + v^2)_{\Lambda^c} \rangle$, and (5.11) shows that $\langle \mathbb{1}, (v^1 + v^2)_\Lambda \rangle = \|(Bu)_\Lambda\|_1$ whereas (5.12) shows that $\langle \mathbb{1}, (v^1 + v^2)_{\Lambda^c} \rangle = \langle (Bu)_{\Lambda^c}, -\operatorname{sign}(B\overline{u})_{\Lambda^c} \rangle$. Thus $\langle c, w \rangle \leq 0 \quad \Leftrightarrow \quad \|(Bu)_\Lambda\|_1 - \langle (Bu)_{\Lambda^c}, \operatorname{sign}(B\overline{u})_{\Lambda^c} \rangle \leq 0$, and non-existence of such w means $\langle c, w \rangle > 0$ for every such w, which equals (5.8) and (5.9).

Finally, generalizing condition (5.9) to all vectors $u^* \in \mathcal{W}_\Lambda$ and all possible cosupports Λ leads to (5.10). $\qquad \square$

Conditions (5.9) and (5.10) clearly indicate the direct influence of cosparsity on the recovery performance: If $\ell = |\Lambda|$ increases, then these conditions will more likely hold.

On the other hand, these results are mainly theoretical since numerically checking (5.10) is infeasible. This motivates the comprehensive experimental assessment of recovery properties reported in Section 6.

6. Numerical Experiments

In this section, we relate the previously derived bounds on the required image cosparsity that guarantees uniqueness in case of known or unknown cosupport Λ to numerical experiments.

6.1. Set-Up

This section describes how we generate 2D or 3D images for a given cosparsity ℓ and how we acquire measurements.

6.1.1. Test Images

Recall from Section 4.1 that the sparsity of the image gradient is denoted by k and the cosparsity by ℓ,

$$k = \|Bu\|_0 = |\operatorname{supp}(Bu)|, \quad B \in \mathbb{R}^{p \times n}, \tag{6.1a}$$

$$\ell = p - \|Bu\|_0 = p - k, \tag{6.1b}$$

$\Lambda := \{r \in [p] : (Bu)_r = 0\}$ denotes the cosupport of the input image u with respect to the analysis operator B, and $\Lambda^c = [p] \setminus \Lambda$ denotes the complement of the set Λ.

Using the parametrization

$$\rho := \frac{k}{n} \tag{6.2}$$

with

$$k := p - \ell \quad \text{and} \quad n = \begin{cases} d^2 & \text{in 2D} \\ d^3 & \text{in 3D} \end{cases}, \quad p = \begin{cases} 2d(d-1) & \text{in 2D} \\ 3d^2(d-1) & \text{in 3D} \end{cases}, \quad (6.3)$$

we generated random 2D and 3D images composed of randomly located ellipsoids with random radii along the coordinate axes. Figures 9 and 10 depict a small sample of these images for illustration and provide the parameter ranges.
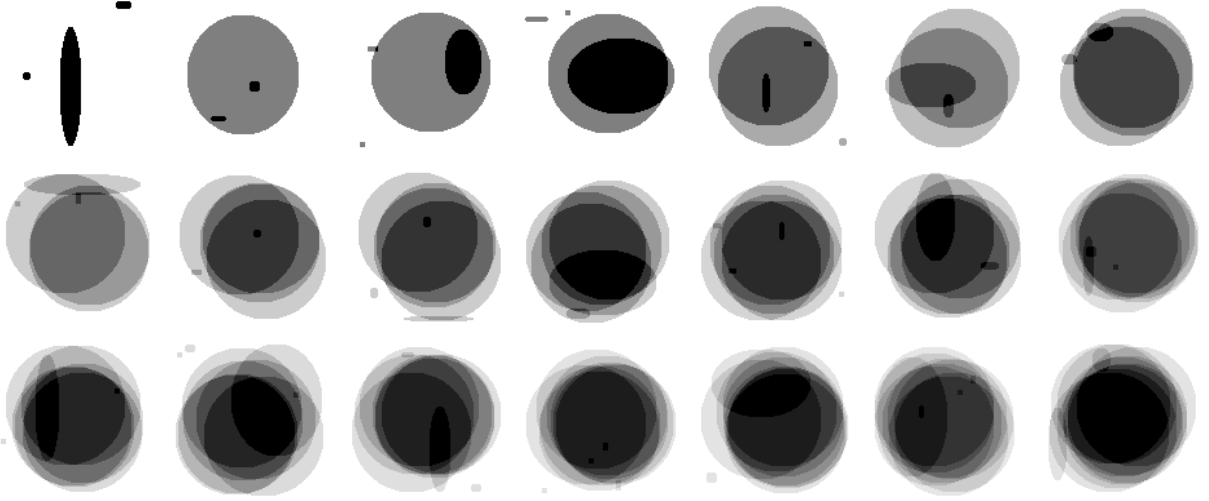

Figure 9. Random images with varying cosparsity $\ell = p - k$, parametrized by ρ (6.2). For each dimension $d = 80 \cdots 180$, random images were generated for $\rho = 0.005 \cdots 0.22$. The figure shows a sample image for a subset of increasing values of ρ and $d = 120$.

6.1.2. Tomographic Projections

Images in 2D are undersampled by the projection matrices from Section 2.1.2, with parameters listed in Table 1. In 3D we consider the two projection matrices from Section 2.1.2, see Fig. 4 and Fig. 5. We also consider a perturbation of each A. Each perturbed matrix $\tilde{A}$ has the same sparsity structure as A, but random entries drawn from the standard uniform distribution on the open interval $(0.9, 1.1)$.

6.2. Optimization

To recover a ℓ-cosparse test image $\overline{u}$, we solve the LP relaxation (5.3) of (4.19), where we take into account the nonnegativity of $\overline{u}$. The relaxation is obtained from (1.3) by considering two additional variables v^1 and v^2 which represent the positive and negative part of Bu. In cases where we assume that Λ is known, we add the constraint $B_\Lambda u = 0$ and solve the LP with the same objective as (5.3), but with the polyhedral feasible set defined by

$$M := \begin{pmatrix} B_{\Lambda^c} & -I_{\Lambda^c} & I_{\Lambda^c} \\ B_\Lambda & 0 & 0 \\ A & 0 & 0 \end{pmatrix} \quad \text{and} \quad q := \begin{pmatrix} 0 \\ 0 \\ b \end{pmatrix}. \quad (6.4)$$
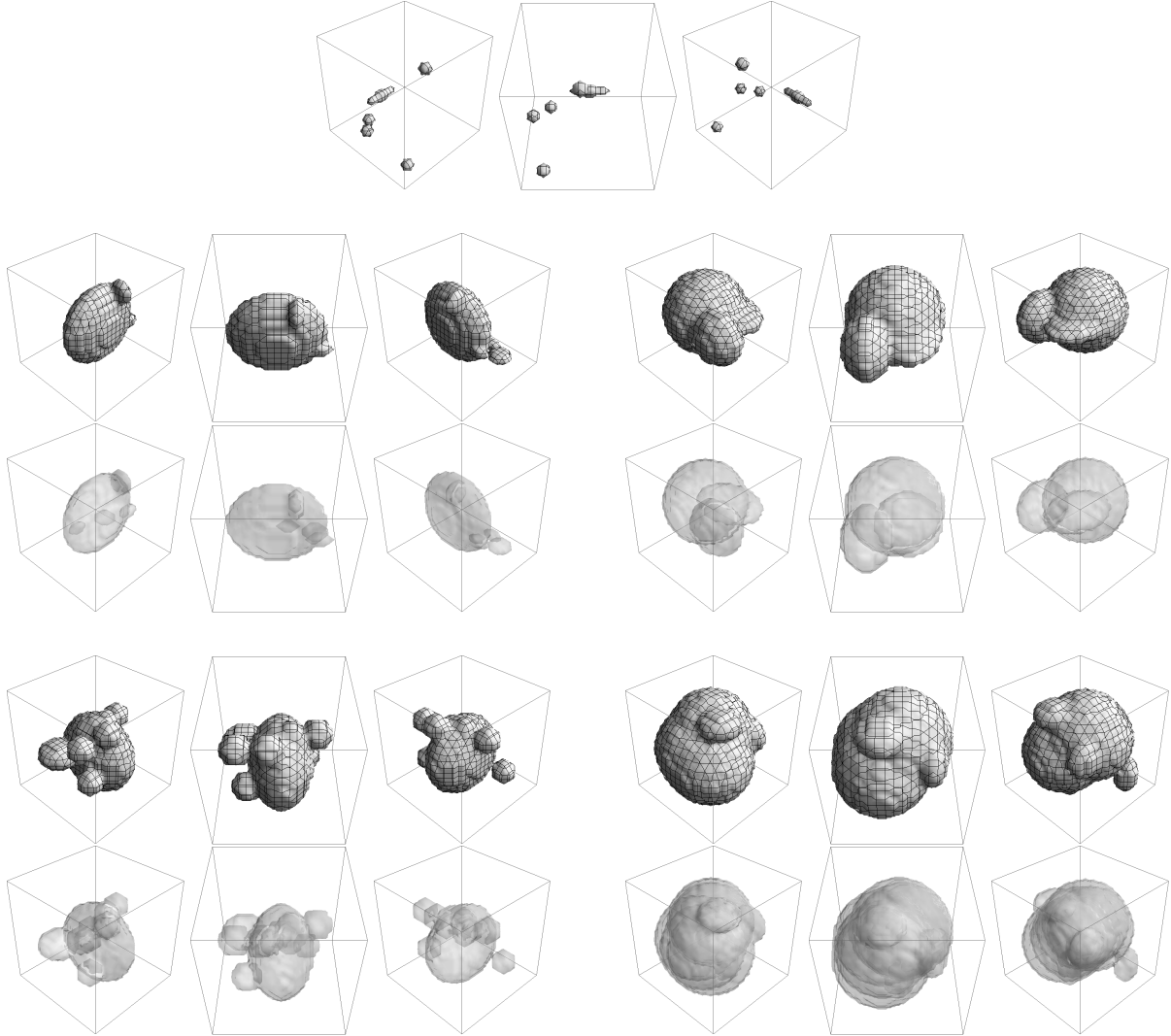
Figure 10.   Random images with varying cosparsity $\ell = p - k$, parametrized by ρ (6.2). For dimension $d = 31$, random images were generated for $\rho = 0.0032 \cdots 1.01$. The figure shows a sample image for five different values of ρ, each plotted from three different viewpoints.

The resulting LPs were solved with the help of a standard LP solver [1]. The reconstruction is considered *successful* if the solution u of the above described LPs is within a small distance from the original ℓ-cosparse $\overline{u}$ generating the data, and $\|u - \overline{u}\|_1 \leq \varepsilon n$ holds, with $\varepsilon = 10^{-6}$ in 2D and $\varepsilon = 10^{-8}$ in 3D.

---

[1] MOSEK http://mosek.com/

6.3. Phase transitions

Phase transitions display the empirical probability of exact recovery over the space of parameters that characterize the problem (cf. [11]). Our parametrization relates to the design of the projection matrices $A \in \mathbb{R}^{m \times n}$. Because both m and n depend on d, we choose d as an *oversampling parameter*, analogously to the *undersampling parameter* $\rho = \frac{m}{n}$ used in [11].

We analyze the influence of the image cosparsity, or equivalently of the image gradient sparsity, on the recovery via (5.3) or (6.4). We assess empirical bounds in relation with the theoretically required sparsity that guarantees exact recovery, described as an empirical phase transition of ρ depending on d. This *phase transition* $\rho(d)$ indicates the necessary relative sparsity ρ to recover a ℓ-cosparse image with overwhelming probability by convex programming.

For each $d \in \{80, 90, \ldots, 170, 180\}$ and for each relative sparsity ρ, we generated 70 images for the 2D case and 50 images for the 3D case, as illustrated in Section 6.1.1, together with corresponding measurements using the matrices from Section 6.1.2. This in turn gave us d, n, m and k, defining a point $(d, \rho) \in [0, 1]^2$. This range was discretized into cells so as to accumulate in a (d, ρ) cell a 1 if the corresponding experiment was successful (exact recovery) and 0 otherwise. In 2D, we performed 30 such runs for each (d, ρ) pair for both cases of known and unknown cosupport. The success rate of image reconstruction is displayed by gray values: black $\leftrightarrow$ 0% recovery rate, white $\leftrightarrow$ 100% recovery rate. In 3D, we analyzed the behavior for two image sizes, $d = 31$ and $d = 41$. The same procedure as in the 2D case was applied, but in order to get better averages the procedure was repeated 6 times. In each repetition, we performed 30 runs per (d, ρ) pair, for both known and unknown cosupport. We show the mean value averaged over the 6 repetitions.

6.3.1. Recovery of 2D Images

The results are shown in Fig. 11 and Fig. 12. The empirical transitions agree with the analytically derived thresholds up to a scaling factor α. The values of α are listed in Table 3. The accordingly rescaled curves are shown as dotted lines in the plots.

All plots display a phase transition and thus exhibit regions where exact image reconstruction has probability equal or close to one.

Table 3. The scaling factors of the theoretical curves (4.16a) and (4.17a) for known and unknown cosupport respectively.

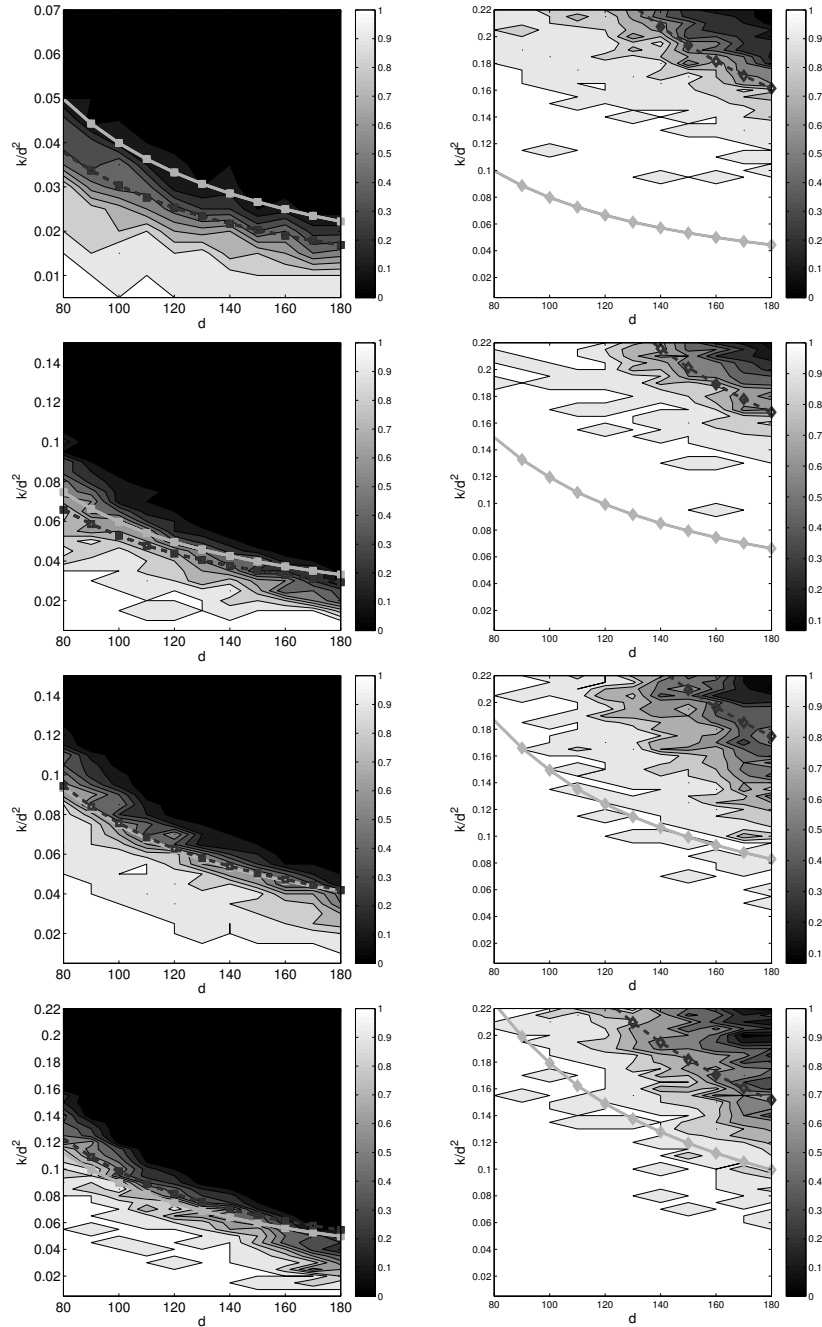| α-values in 2D						
d	cosupport	measurements	3 views	4 views	5 views	6 views
$80 \ldots 180$	known	unperturbed	3.6358	2.5349	2.1073	1.5241
		perturbed	2.9220	2.0614	1.4039	1.2453
	unknown	unperturbed	0.7603	0.8827	1.0115	1.0962
		perturbed	0.6976	1.0744	1.1200	1.2402

Figure 11. Phase transitions for the unperturbed matrix A, 2D case, 3, 4, 5 and 6 projecting directions (top to bottom), with unknown (left column) and known (right column) cosupport. The continuous gray lines depict the theoretical curves (4.17a) and (4.16a) respectively. The dashed black lines correspond to the empirical threshold, which are all scaled versions of (4.17a) (left column) or (4.16a) (right column) with scaling factors α summarized in Table. 3.
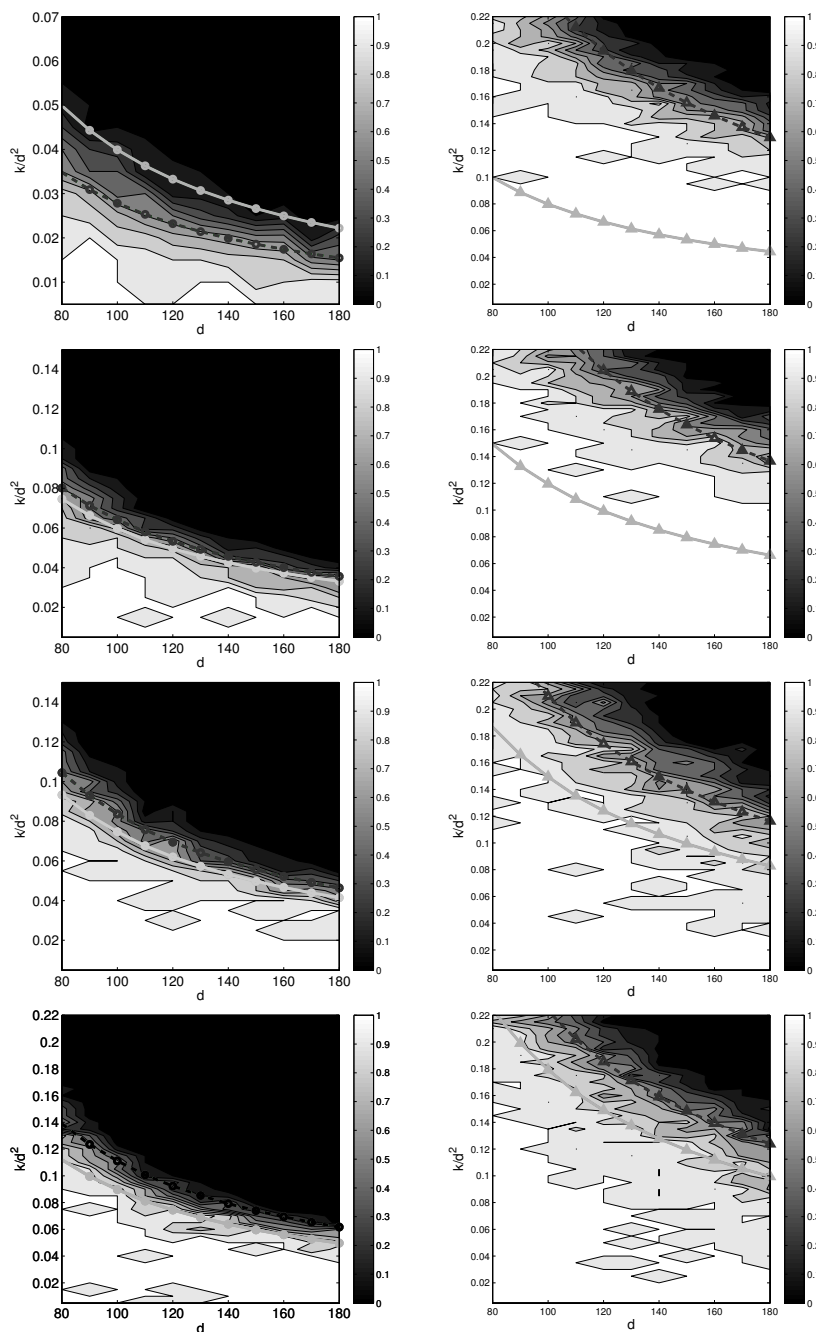
Figure 12. Phase transitions for the perturbed matrix $\tilde{A}$, 2D case, 3, 4, 5 and 6 views (top to bottom), with unknown (left column) and known (right column) cosupport. The continuous gray lines depict the theoretical curves (4.17a) and (4.16a). The dashed black lines correspond to the empirical threshold, which are all scaled versions of (4.17a) (left column) or (4.16a) (right column) with scaling factors α listed in Table. 3. We note the similarity to the unperturbed case in Fig. 11.

6.3.2. Recovery of 3D Images

The results are shown in Fig. 13 and Fig. 14 for $d = 31$ and $d = 41$, and summarized in Fig. 15. The empirical phase transitions differ again from the analytically derived thresholds 4.16b and 4.17b only by a scaling factor α. These values are listed as Table 4. The rescaled curves are shown as dotted lines in the plots. Fig. 15 also relates the critical sparsity of the gradient to the critical sparsity estimated in [30, 32], which in turn implies exact recovery via (1.1).
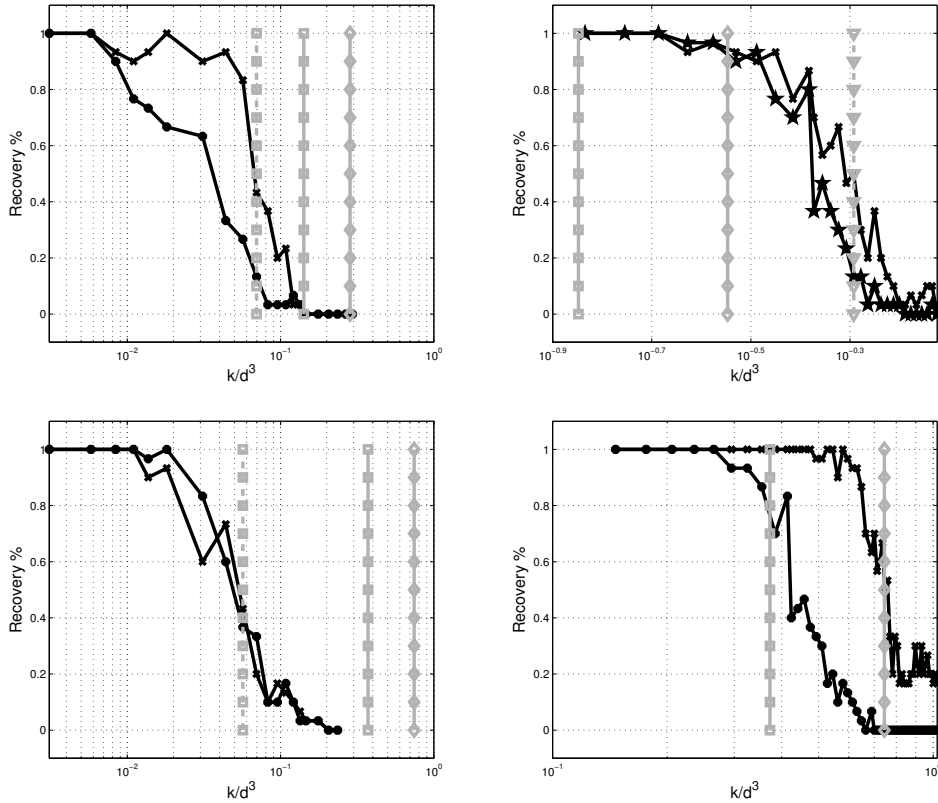

Figure 13. Empirical probability (6×30 trials) of exact recovery by total variation minimization (1.3) via the unperturbed ($\times$-marked curve) and perturbed ($\bullet$-marked curve) matrices from Section 2.1.2, Fig. 4 and Fig. 5, for 3 (top row) and 4 (bottom row) projecting directions, respectively, and $d = 31$. The left column shows the decay of the recovery probability when the cosupport is *unknown* using both perturbed and unperturbed projecting matrices, while the right one, shows results for *known* cosupport. The continuous vertical lines stand for the theoretical thresholds for known (4.16b) ($\diamond$-marked curve) and unknown (4.17b) ($\square$-marked curve) cosupport, while the dotted ∇ and $\square$ vertical lines stand for the empirically estimated threshold for known and unknown cosupport, but for unperturbed matrices only. The deviation of the empirical thresholds from the theoretical curves for known cosupport (4.16b) and unknown cosupport (4.17b) was estimated through least-squares fit and is summarized in Table 4, along with results for the perturbed matrices.
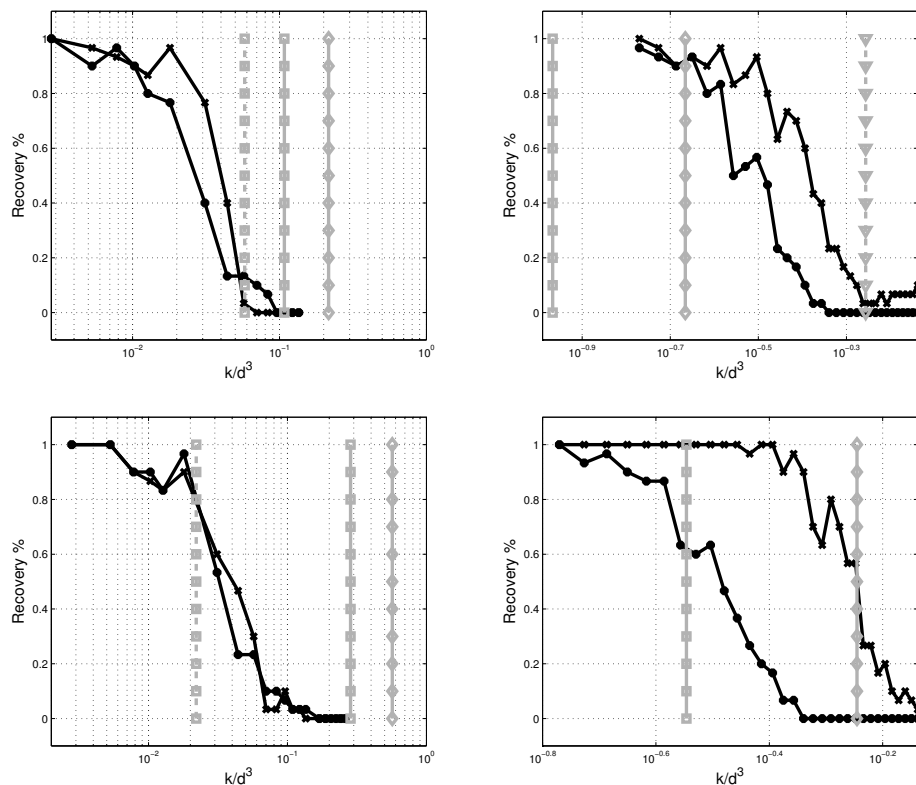
Figure 14. Empirical probability (6 × 30 trials) of exact recovery by total variation minimization (1.3) via the unperturbed (×-marked curve) and perturbed (•-marked curve) matrices from Section 2.1.2, Fig. 4 and Fig. 5, for 3 (top row) and 4 (bottom row) projecting directions respectively. Hereby $d = 41$. The significance of each curve is identically to the one in Fig. 13. Scaling factors are summarized in Table 4,

Table 4. The scaling factors of the theoretical curves (4.16b) and (4.17b) for known and unknown cosupport respectively.

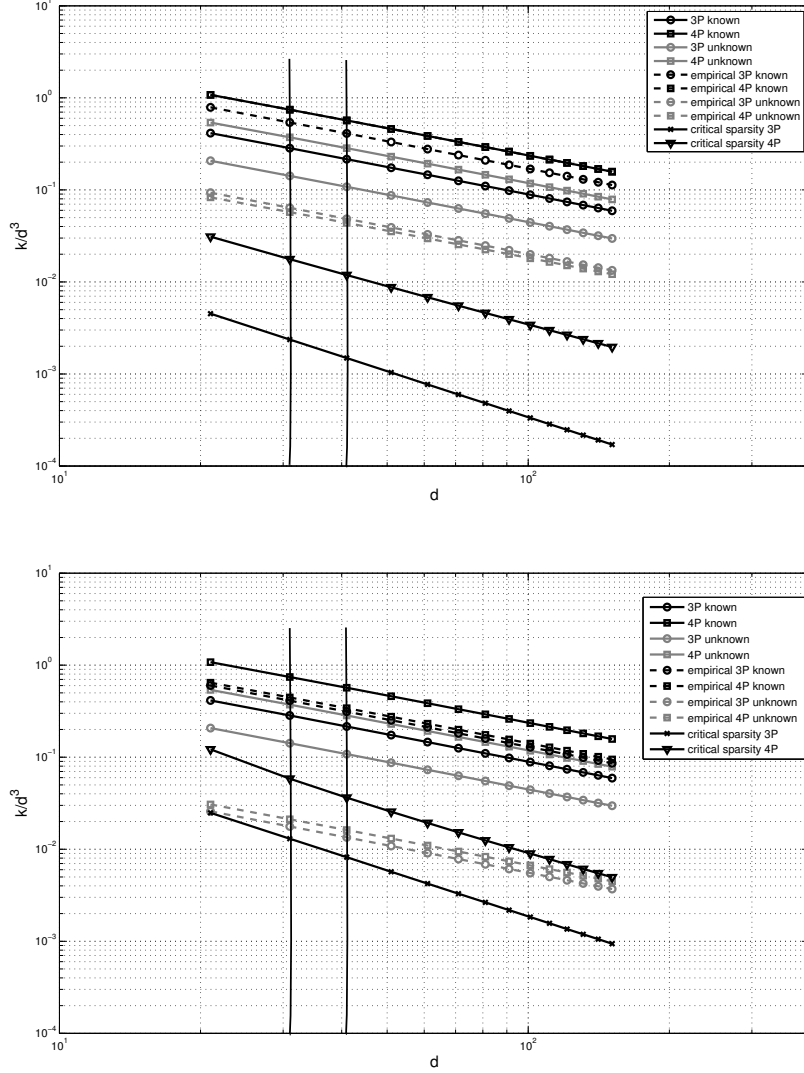| α-values in 3D				
# views	cosupport	measurements	$d = 31$	$d = 41$
3	known	unperturbed	1.8568	1.9527
		perturbed	1.6137	1.2860
	unknown	unperturbed	0.4910	0.4086
		perturbed	0.2098	0.2882
4	known	unperturbed	1	1
		perturbed	0.6153	0.5833
	unknown	unperturbed	0.1526	0.1552
		perturbed	0.1180	0.1094

Figure 15. Log-log plot of phase transitions in 3D for the unperturbed matrix A (top), and perturbed matrix $\tilde{A}$ (bottom) for 3 ($\circ$-marked curve) and 4 ($\square$-marked curve) projecting directions. The continuous gray and black curves depict the theoretical curves (4.17b) and (4.16b) respectively. The dashed lines correspond to the empirical thresholds, which are all scaled versions of (4.17b) or (4.16b) with scaling factors summarized in Table. 4. The black continuous $\times$-marked curve (stands for 3 projecting directions) and black continuous ∇-marked curve (stands for 4 projecting directions) show the relative critical sparsity such that k random points are recovered *exactly* by (1.1). These are the theoretical phase transition ℓ_1-recovery from [30], [32]. The vertical lines correspond to $d = 31$ and $d = 41$, compare with Fig. 13 and Fig. 14.

6.4. Discussion

Several observations are in order.

- Perturbation of projection matrices brings no significant advantage in the practically relevant case of unknown cosupport. The empirical transitions will remain the same for perturbed and unperturbed matrices. This is very different to the ℓ_1-minimization problem (1.1), where perturbation boosts the recovery performance significantly as shown in [30].

- In the case of known cosupport, when $B_\Lambda u = 0$ is added as additional constraint, unperturbed matrices perform better. We notice that the empirical phase transition is *above* the red curve, and deduce that linear dependencies might be beneficial when the cosupport is known.

- When increasing the number of views (4, 5, 6 or more) the differences between estimated (dashed) and theoretical (continuous line) phase transition become smaller. This might be due to the fact that linear dependencies between the columns (and rows) of A become "rare", and the assumptions of Propositions 4.2 and 4.3 are more likely to be satisfied.

- In 3D the difference between empirical phase transitions for 3 and 4 projecting directions is very small, i.e. relative phase transitions are almost equal. This is different to the 2D case above. We currently do not have an explanation for this phenomenon.

- The log-log plot in Figure 15 shows that phase transitions in 3D exhibit a power law behavior, similar to the theoretical phase transitions for ℓ_1-recovery found in [30], [32]. We note that the phase transitions that correspond to TV-minimization (marked with $\circ$ and $\square$) have higher scaling exponents than the phase transitions that correspond to ℓ_1-minimization (marked with $\times$ and $\triangledown$). The plot also tells us that for large d (large volumes) TV-minimization allows higher sparsity thresholds of the image gradient than ℓ_1-minimization does for the image sparsity.

7. Conclusion

We studied the cosparsity model in order to theoretically investigate conditions for unique signal recovery from severely undersampled linear systems, that involve measurement matrices whose properties fall far short of the assumptions commonly made in the compressed sensing literature. Extensive numerical experiments revealed a high accuracy of the theoretical predictions, up to a scale factor caused by slight violations in practice of our mathematical assumptions. Unique recovery can be accomplished by linear programming that in principle copes with large problem sizes. The signal class covered by the cosparsity model seems broad enough to cover relevant industrial applications of non-standard tomography, like contactless quality inspection.

In our future work, we will aim at clarifying quantitatively the above-mentioned scale factor and its origin. In the same context, conducting a probabilistic analysis as in our recent work [30], for the present scenarios, defines an open problem. We expect that the refinement of a probabilistic version of the cosparsity model, in connection with distributions of cosupports learned from relevant collections of signals, may have an impact both theoretically and practically beyond aspects of limited-view tomography.

References

[1] R. Baraniuk, *Compressive Sensing*, IEEE Signal Processing Magazine **24** (2007), no. 4, 118–121.

[2] B. Bollobás and I. Leader, *Edge-Isoperimetric Inequalities in the Grid*, Combinatorica **11** (1991), no. 4, 299–314.

[3] E. J. Candès, *The Restricted Isometry Property and its Implications for Compressed Sensing*, Comptes Rendus Mathematique **346** (2008), no. 9-10, 589–592.

[4] S. Carmignato, *Computed Tomography as a Promising Solution for Industrial Quality Control and Inspection of Castings*, Metallurgical Science and Technology **30-1** (2012), 5–14.

[5] E. J. Candès, Y. C. Eldar, and D. Deanna Needell, *Compressed Sensing with Coherent and Redundant Dictionaries*, Applied and Computational Harmonic Analysis **31** (2010), no. 1, 59–73.

[6] V. Chandar, *A Negative Result Concerning Explicit Matrices with the Restricted Isometry Property*, 2008, Preprint http://dsp.rice.edu/files/cs/Venkat_CS.pdf.

[7] E. J. Candès, J. Romberg, and T. Tao, *Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information*, IEEE Transactions on Information Theory **52** (2006), no. 2, 489–509.

[8] E. J. Candès, J. K. Romberg, and T. Tao, *Stable Signal Recovery from Incomplete and Inaccurate Measurements*, Communications on Pure and Applied Mathematics **59** (2006), no. 8, 1207–1223.

[9] E. J. Candès and M. Wakin, *An Introduction to Compressive Sampling*, IEEE Signal Processing Magazine **25** (2008), no. 2, 21–30.

[10] D.L. Donoho and J. Tanner, *Counting Faces of Randomly Projected Polytopes when the Projection Radically Lowers Dimension*, Journal of American Mathematical Society **22** (2009), 1–53.

[11] D. L. Donoho and J. Tanner, *Counting the Faces of Randomly-Projected Hypercubes and Orthants, with Applications*, Discrete & Computational Geometry **43** (2010), no. 3, 522–541.

[12] M. Elad, *Sparse Representations Are Most Likely to Be the Sparsest Possible*, EURASIP J. Adv. Sig. Proc. **2006** (2006), 1–12.

[13] R.J. Gardner and P. Gritzmann, *Discrete Tomography: Determination of Finite Sets by X-Rays*, Transactions of the American Mathematical Society **349** (1997), no. 6, 2271–2295.

[14] C. Grünzweig, D. Mannes, A. Kaestner, F. Schmid, V. Vontobel, J. Hovind, S. Hartmann, S. Peetermans, and E. Lehmann, *Progress in Industrial Applications using Modern Neutron Imaging Techniques*, Physics Procedia **43** (2013), 231–242.

[15] B. Goris, W. Van den Broek, K.J. Batenburg, H.H. Mezerji, and S. Bals, *Electron Tomography Based on a Total Variation Minimization Reconstruction Techniques*, Ultramicroscopy **113** (2012), 120–130.

[16] G. T. Herman and R. Davidi, *Image Reconstruction from a Small Number of Projections*, Inverse Problems **24** (2008), no. 4, 45011–45028.

[17] G. T. Herman and A. Kuba, *Discrete Tomography: Foundations, Algorithms and Applications*, Birkhäuser, 1999.

[18] S. Jafarpour, M. F. Duarte, and A. R. Calderbank, *Beyond Worst-Case Reconstruction in Deterministic Compressed Sensing.*, ISIT, IEEE, 2012, pp. 1852–1856.

[19] J. H. Jorgensen, E. Y. Sidky, P. C. Hansen, and P. Xiaochuan, *Quantifying Admissible Undersampling for Sparsity-Exploiting Iterative Image Reconstruction in X-Ray CT.*, IEEE Transactions on Medical Imaging **32** (2013), no. 2, 460–473.

[20] Y. M. Lu and M. N. Do, *A Theory for Samping Signals From a Union of Subspaces*, IEEE Transactions on Signal Processing **56** (2008), no. 6, 2334–2345.

[21] F. Lim and V. M. Stojanovic, *On U-Statistics and Compressed Sensing i: Non-Asymptotic Average-Case Analysis*, IEEE Transactions on Signal Processing **61** (2013), no. 10, 2473–2485.

[22] _____ , *On U-Statistics and Compressed Sensing ii: Non-Asymptotic Worst-Case Analysis*, IEEE Transactions on Signal Processing **61** (2013), no. 10, 2486–2497.

[23] O. L. Mangasarian, *Uniqueness of Solution in Linear Programming*, Linear Algebra and its Applications **25** (1979), no. 0, 151–162.

[24] S. Nam, M.E. Davies, M. Elad, and R. Gribonval, *The Cosparse Analysis Model and Algorithms*, Applied and Computational Harmonic Analysis **34** (2013), no. 1, 30–56.

[25] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, SIAM, 2001.

[26] D. Needell and R. Ward, *Near-Optimal Compressed Sensing Guarantees for Total Variation Minimization*, IEEE Transactions on Image Processing **22** (2013), 3941–3949.

[27] D. Needell and R. Ward, *Stable Image Reconstruction Using Total Variation Minimization*, SIAM Journal on Imaging Sciences **6** (2013), no. 2, 1035–1058.

[28] V. M Patel, R. Maleh, Gilbert A. C., and Chellappa R., *Gradient-Based Image Recovery Methods From Incomplete Fourier Measurements*, IEEE Transactions on Image Processing **21** (2012), no. 1, 94–105.

[29] S. Petra and C. Schnörr, *TomoPIV meets Compressed Sensing*, Pure Mathematics and Applications **20** (2009), no. 1-2, 49–76.

[30] _____ , *Average Case Recovery Analysis of Tomographic Compressive Sensing*, Linear Algebra and its Applications **441** (2014), 168–198, Special Issue on *Sparse Approximate Solution of Linear Systems*, in press, http://www.sciencedirect.com/science/article/pii/S0024379513004333.

[31] S. Petra, A. Schröder, and C. Schnörr, *3D Tomography from Few Projections in Experimental Fluid Mechanics*, Imaging Measurement Methods for Flow Analysis (W. Nitsche and C. Dobriloff, eds.), Notes on Numerical Fluid Mechanics and Multidisciplinary Design, vol. 106, Springer, 2009, pp. 63–72.

[32] S. Petra, C. Schnörr, and A. Schröder, *Critical Parameter Values and Reconstruction Properties of Discrete Tomography: Application to Experimental Fluid Dynamics*, Fundamenta Informaticae **125** (2013), 285–312.

[33] O. Scherzer (ed.), *Handbook of Mathematical Methods in Imaging*, Springer, 2011.

[34] E. Y. Sidky and X. Pan, *Image Reconstruction in Circular Cone-Beam Computed Tomography by Constrained, Total-Variation Minimization*, Physics in Medicine and Biology **53** (2008), no. 17, 47–77.

[35] W.P. Ziemer, *Weakly Differentiable Functions*, Springer, 1989.