Performance Bounds For Co-/Sparse Box Constrained Signal Recovery

Jan Kuske and Stefania Petra

Abstract

The recovery of structured signals from a few linear measurements is a central point in both compressed sensing (CS) and discrete tomography. In CS the signal structure is described by means of a low complexity model e.g. co-/sparsity. The CS theory shows that any signal/image can be undersampled at a rate dependent on its intrinsic complexity. Moreover, in such undersampling regimes, the signal can be recovered by sparsity promoting convex regularization like ℓ_1- or total variation (TV-) minimization. Precise relations between many low complexity measures and the sufficient number of random measurements are known for many sparsity promoting norms. However, a precise estimate of the undersampling rate for the TV seminorm is still lacking. We address this issue by: a) providing dual certificates testing uniqueness of a given cosparse signal with bounded signal values, b) approximating the undersampling rates via the statistical dimension of the TV descent cone and c) showing empirically that the provided rates also hold for tomographic measurements.

1 Introduction

The current work is motivated by the image reconstruction problem from few linear measurements, e.g. discrete tomography [HK99], where typically bounds on the image values are available. This explains why we are concerned with the recovery

of *box constrained* signals. Moreover, we aim to reconstruct an image that exhibits some structure, e.g. sparsity/gradient-sparsity.

Recovering a structured signal/image from few linear measurements is a central point in both compressed sensing (CS) [FR13] and discrete tomography [HK99]. In CS the signal structure is described by means of a low complexity model. For instance, signal sparsity $s := \|x\|_0$, that equals the number of non-zero components in x, and the associated ℓ_0-minimization is typically relaxed to ℓ_1-minimization. The theory of CS implies that if the number of measurements $m \geq C \cdot s \log(n/s)$ and the entries of the measurement matrix A follow a normal distribution, i.e. $a_{ij} \sim \mathcal{N}(0, 1)$, the solution of ℓ_0-minimization is *equal* to the solution of ℓ_1-minimization. Thus, a combinatorial problem can be solved by a convex optimization problem.

In the present work we consider the noiseless setting

$$A\overline{x} = b, \tag{1.1}$$

where the unknown (structured) signal $\overline{x} \in \mathbb{R}^n$ is undersampled either by a Gaussian matrix $A \in \mathbb{R}^{m \times n}$, i.e. A has independent standard normal entries, or by a tomographic projection matrix A.

For the reconstruction of the sparse or gradient-sparse box constrained signal of interest we consider some structure enforcing regularizer $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, that is a proper convex function, and solve

$$\min_x f(x) \qquad \text{subject to} \quad Ax = A\overline{x}, \tag{1.2}$$

where f is specialized to one of the following functions

$$f_1(x) = \|x\|_1, \tag{1.3a}$$

$$f_2(x) = \|x\|_1 + \delta_{\mathbb{R}_+^n}(x), \tag{1.3b}$$

$$f_3(x) = \|x\|_1 + \delta_{[0,1]^n}(x), \tag{1.3c}$$

$$f_4(x) = \|Dx\|_1, \tag{1.3d}$$

$$f_5(x) = \|Dx\|_1 + \delta_{\mathbb{R}_+^n}(x), \tag{1.3e}$$

$$f_6(x) = \|Dx\|_1 + \delta_{[0,1]^n}(x). \tag{1.3f}$$

In the first part of the paper $D \in \mathbb{R}^{p \times n}$ is an arbitrary matrix, while in the latter part D is specialized to the discrete gradient operator, in view of the (anisotropic) TV-regularizer we are interested in.

1.1 Contribution, Related Work, Organization

Our main objectives are:

Individual Uniqueness We derive practical tests for individual uniqueness: Given a signal $\overline{x}$ and an instance of the optimization problem above, can we efficiently verify whether the signal is the unique solution? Our approach will be to apply the uniqueness condition $0 \in \mathrm{int}\, \partial f(\overline{x})$ [Gil17] to several types of polyhedral functions f including ℓ_1- and TV-minimization with and without box constraints, as detailed in Sect. 2. Such testable uniqueness conditions for TV-minimization were also developed in [JKL15, ZMY16], but these use different mathematical tools. Our derivation is shorter and additionally considers box constrained TV-minimization.

Probabilistic Uniqueness We wish to accurately describe the relation between the signal sparsity and the number of measurements that guarantee uniqueness with high probability. We build on recent CS theory [ALMT14, OH16, DJM13, FM14] that upper bounds the statistical dimension of the descent cone of the structure enforcing regularization. The difficulty lies in the fact that certain CS guarantees are missing for TV-minimization. Moreover, the CS recovery performance bounds based on the statistical dimension (also called phase transitions) do not apply directly to tomography, since tomographic measurements are known to be ill-conditioned on the common structured signal classes from CS. Phase transitions in CS are discussed in Section 3.

Empirical validation We verify the theoretical probabilistic results for Gaussian matrices along with the theoretical uniqueness tools and compare them to the recovery and uniqueness performance of typical tomographic measurements. We stress that tomographic projections fall short of the common assumptions (e.g. restricted isometry property) underlying compressed sensing, see [PS14], and CS recovery guarantees *do not apply* for tomography. In view of its practical importance (reducing radiation exposure) we emphasize the need of accurately describing how much undersampling is tolerable in tomographic recovery. We empirically validate our theoretical results in Sect. 4. Hence, our work is closely related to the work in [JKL15], but additionally provides approximations to the *theoretical* phase transition curves that are validated empirically.

1.2 Notation

For $n \in \mathbb{N}$, we use the shorthands $[n] = \{1, 2, \ldots, n\}$. We define $\mathbb{R}_+^n := \{x \in \mathbb{R}^n : x_i \geq 0, \forall i \in [n]\}$ and $\mathbb{R}_{++}^n := \{x \in \mathbb{R}^n : x_i > 0, \forall i \in [n]\}$. Analogously, we define $\mathbb{R}_-^n$ and $\mathbb{R}_{--}^n$. The extended real line is denoted by $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. Vectors $x \in \mathbb{R}^n$ are column vectors and indexed by superscripts. $x^\top$ denotes the transposed vector x and $\langle x, y \rangle$ or $x^\top y$ the Euclidean inner product. $\top$ stands for the transpose. For a vector $x \in \mathbb{R}^n$, $\mathrm{sign}(x) \in \mathbb{R}^n$ is the *sign vector* of x defined component-wise as $\mathrm{sign}(x)_i = 1$ if $x_i > 0$ and $\mathrm{sign}(x)_i = -1$ if $x_i < 0$. For some matrix $A \in \mathbb{R}^{m \times n}$,

the nullspace is denoted by $\mathcal{N}(A)$ and its range by $\mathcal{R}(A)$. For some matrix $\Omega \in \mathbb{R}^{p \times n}$ we denote the submatrix indexed by the rows in $\Lambda \subset [p]$ by $\Omega_{\Lambda,\bullet}$. Similarly, for some matrix $A \in \mathbb{R}^{m \times n}$ the columns indexed by $S \subset [n]$ are denoted by $A_{\bullet,S}$. The complement of an index set $S \subset [n]$ will be denoted by $S^c := [n] \setminus S$. For a subspace X, $X^\perp$ will denote its orthogonal complement. The affine hull of a set X, see Appendix, is denoted by $\mathrm{aff}(X)$. The conic hull is denoted by $\mathrm{cone}\, X$. We use the ℓ_1-norm $\|x\|_1 = \sum_{i \in [n]} |x_i|$ and the maximum norm $\|x\|_\infty := \max_{i \in [n]} |x_i|$. The ℓ_0-measure (not a norm!) stands for the cardinality of the support of x, i.e. $\|x\|_0 := |\mathrm{supp}(x)|$, with $\mathrm{supp}(x) = \{i \in [n]\colon x_i \neq 0\}$. The cosupport of x is denoted by $\mathrm{cosupp}(x) = \{i \in [n]\colon x_i = 0\}$.

We define the one-dimensional discrete derivative operator $\partial_n \colon \mathbb{R}^n \to \mathbb{R}^{n-1}$, $(\partial_n)_{ij} = -1$ if $i = j$, $(\partial_n)_{ij} = 1$ if $j = i + 1$ and $(\partial_n)_{ij} = 0$ otherwise. For $m \times n$ images the discrete gradient operator is defined as

$$\nabla = \begin{pmatrix} I_n \otimes \partial_m \\ \partial_n \otimes I_m \end{pmatrix} \in \mathbb{R}^{p \times n}, \tag{1.4}$$

where $\otimes$ denotes the Kronecker product and I_m, I_n, are identity matrices. The *anisotropic* discretized total variation (TV) is given by $\mathrm{TV}(x) := \|\nabla x\|_1$. The image gradient sparsity is given by $\|\nabla x\|_0$.

We further denote the indicator function of a convex set as $\delta_C : \mathbb{R}^n \to \overline{\mathbb{R}}$. Recall, $\delta_C(x) = 0$ if $x \in C$ and $\delta_C(x) = \infty$, otherwise. The subdifferential of a function f at x will be denoted by $\partial f(x)$, see Appendix for the definition. We will use that the subdifferential of the indicator function equals the normal cone $\partial \delta_C(x) = N_C(x) := \{v \in \mathbb{R}^n \colon v^\top (y - x) \leq 0, \forall y \in C\}$. The distance of a vector x to C is denoted by $\mathrm{dist}(x, C)$ and defined as $\mathrm{dist}(x, C) = \min_{y \in C} \|x - y\|$. Finally, $\mathbb{E}(X)$ will denote the expectation of the random variable X.

2 Individual Uniqueness and Dual Certificates

In this section we derive testable uniqueness conditions for our six problems of interest. We will first provide some recovery conditions called *dual certificates*. More precisely, consider that we are given a specific vector $\overline{x}$ and we have to decide whether it is the *unique* solution of (1.2). We will provide *necessary and sufficient* conditions which certify the existence and uniqueness of a solution (1.2) for the case of a polyhedral function f. Our analysis closely follows [Gil17]. These conditions are formulated in terms of a solution to the dual problem of (1.2). For the cases enumerated in (1.3), we will see that it is possible to test uniqueness by simply solving a linear program.

In this section, it will be useful to recast (1.2) as an unconstrained optimization

problem

$$\min f(x) \quad \text{subject to} \quad Ax = b \quad \Longleftrightarrow \quad \min_{x \in \mathbb{R}^n} g(x), \quad g(x) := f(x) + \delta_{\mathcal{X}}(x), \tag{2.1}$$

where $\delta_{\mathcal{X}}$ denotes the indicator function of the *feasible set* of (1.2) with

$$\mathcal{X} := \{x \in \mathbb{R}^n : Ax = b\}, \tag{2.2}$$

f the objective in (1.2) and

$$b := A\overline{x}. \tag{2.3}$$

Hence, the function g considered in this section is *convex polyhedral*, meaning that its epigraph is a convex polyhedron.

2.1 Uniqueness of a Minimizer of a Polyhedral Function

We will use that the condition $0 \in \text{int } \partial g(\overline{x})$ is equivalent to the uniqueness of the solution $\overline{x}$, in the case of a polyhedral convex function g. The following result gives us a necessary and sufficient condition for the uniqueness of a polyhedral function minimizer and was recently emphasized in [Gil17].

Lemma 2.1. *[Gil17, Lem 2.2] Let $g \in \mathcal{F}_0(\mathbb{R}^n)$ be any proper, convex and lsc function that is in addition polyhedral. Then*

$$\arg\min g = \{\overline{x}\} \iff 0 \in \text{int } \partial g(\overline{x}). \tag{2.4}$$

We note: the necessity part uses polyhedrality, the sufficiency part is straightforward and does not require polyhedrality. We will see, in Sect. 3, Prop. 3.1, an alternative uniqueness condition for the problem (1.2). The advantage of (2.4) is that it directly allows the derivation of testable uniqueness conditions as detailed next.

2.2 Certifying Recovery of Individual Vectors via Dual Certificates

In this section we show how to decide if a given solution for (1.3) is unique. To this end, we derive uniqueness conditions for the problem

$$\min_{x \in \mathbb{R}^n} \|Dx\|_1 + \delta_{\mathcal{X}}(x) + \delta_{[l,u]}(x), \tag{2.5}$$

where $l_i \in \mathbb{R} \cup \{-\infty\}$, $u_i \in \mathbb{R} \cup \{+\infty\}$, $l_i < u_i$, $i \in [n]$. Hence, (2.5) can be regarded as a generalization of (1.3a)-(1.3f). For example (1.3b) can be obtained from (2.5) by setting $D := I_n$, $l := 0 \in \mathbb{R}^n$ and $u_i := +\infty$, $i \in [n]$. In order to apply the uniqueness criterion of Lem. (2.1) we will need to calculate the subdifferential of

the objective function in (2.5). To compute the subdifferentials at a feasible point $\overline{x}$ one can apply the sum rule [RW09, Cor. 10.9]

$$\partial\left(\|D\overline{x}\|_1 + \delta_{\mathcal{X}}(\overline{x}) + \delta_{[l,u]^n}(\overline{x})\right) = \partial\|D\overline{x}\|_1 + \partial\delta_{\mathcal{X}}(\overline{x}) + \partial\delta_{[l,u]^n}(\overline{x}). \qquad (2.6)$$

Since all individual components are polyhedral functions and $\overline{x} \in \operatorname{dom}\|D\cdot\|_1 \cap \mathcal{X} \cap [l,u]^n$ holds by construction, we can simply add the subdifferentials of the individual terms

$$\partial\|D\overline{x}\|_1 = \{D^\top\alpha\colon \alpha_{\Lambda^c} = \operatorname{sign}(D_{\Lambda^c,\bullet}\overline{x}), \|\alpha_\Lambda\|_\infty \le 1\}, \qquad (2.7a)$$

$$\partial\delta_{\mathcal{X}}(\overline{x}) = N_{\mathcal{X}}(\overline{x}) = \mathcal{N}(A)^\perp, \qquad (2.7b)$$

$$\partial\delta_{[l,u]}(x) = N_{[l,u]}(\overline{x}). \qquad (2.7c)$$

The normal cone corresponding to the box constraints reads

$$N_{[l,u]}(\overline{x}) = N_{[l_1,u_1]}(\overline{x}_1) \times \cdots \times N_{[l_n,u_n]}(\overline{x}_n) = \mathbb{R}_-^{|S_l|} \times \{0\}^{|S_{lu}|} \times \mathbb{R}_+^{|S_u|}, \qquad (2.8)$$

since

$$N_{[l_i,u_i]}(\overline{x}_i) = \begin{cases} 0, & \overline{x}_i \in (l_i, u_i), \\ (-\infty, 0], & \overline{x}_i = l_i, \\ [0, \infty), & \overline{x}_i = u_i, \end{cases} \qquad (2.9)$$

where $S_l = \{i \in [n]\colon x_i = l_i\}$, $S_u = \{i \in [n]\colon x_i = u_i\}$ and $S_{lu} = (S_l \cup S_u)^c = \{i \in [n]\colon x_i \in (l_i, u_i)\}$. If for example we specialize (2.5) to (1.3b), the normal cone of the nonnegative orthant becomes relevant and specializes to

$$\partial\delta_{\mathbb{R}_+^n}(\overline{x}) = N_{\mathbb{R}_+^n}(\overline{x}) = N_{\mathbb{R}_+}(\overline{x}_1) \times \cdots \times N_{\mathbb{R}_+}(\overline{x}_n) = \{0\}^{|S|} \times \mathbb{R}_-^{|S^c|}, \qquad (2.10)$$

since $S_u = \emptyset$, $S_{lu} = \{i \in [n]\colon \overline{x}_i \in (0, +\infty)\} = \operatorname{supp}(\overline{x}) =: S$ and $S_l = S^c$. Based on the considerations above we now derive the main result of this section.

Theorem 2.2 (Dual Certificate). *A solution $\overline{x}$ to (2.5) is unique, if and only if the following conditions hold*

(i) $\mathcal{N}(A) \cap \mathcal{N}(D_{\Lambda,\bullet}) \cap \mathcal{N}(\Psi) = \{0\}$,

(ii) $\exists\alpha \in \mathbb{R}^p, \exists\mu \in \mathbb{R}^n :$
$D^\top\alpha + \Psi\mu \in \mathcal{R}(A^\top), \alpha_{\Lambda^c} = \operatorname{sign}(D_{\Lambda^c,\bullet}\overline{x}), \|\alpha_\Lambda\|_\infty < 1, \mu > 0,$

where $\Lambda := \operatorname{cosupp}(D\overline{x})$ and $\Psi = \Psi(\overline{x})$ is a diagonal matrix that depends on $\overline{x}$ with entries

$$\Psi_{ii} = \begin{cases} -1, & \overline{x}_i = l_i, \\ 1, & \overline{x}_i = u_i, \\ 0, & \text{otherwise.} \end{cases} \qquad (2.11)$$

84

Proof. By Lem. 2.1 the uniqueness of $\overline{x}$ as the unique solution of (2.5) is equivalent to

$$0 \in \text{int}\left(\partial\left(\|D\overline{x}\|_1 + \delta_{\mathcal{X}}(\overline{x}) + \delta_{[l,u]^n}(\overline{x})\right)\right). \tag{2.12}$$

We rewrite (2.12) in the form of the two conditions

(a) $\mathbb{R}^n = \text{aff}\left(\partial\left(\|D\overline{x}\|_1 + \partial\delta_{\mathcal{X}}(\overline{x}) + \partial\delta_{[l,u]}(\overline{x})\right)\right)$,

(b) $0 \in \text{rint}\left(\partial\left(\|D\overline{x}\|_1 + \partial\delta_{\mathcal{X}}(\overline{x}) + \partial\delta_{[l,u]}(\overline{x})\right)\right)$.

Condition (a) ensures that the interior of the subdifferential is not-empty, and can be recast as

$$\mathbb{R}^n = \text{aff}(\partial\|D\overline{x}\|_1) + \text{aff}(\partial\delta_{\mathcal{X}}(\overline{x})) + \text{aff}(\partial\delta_{[l,u]}(\overline{x})). \tag{2.13}$$

By (2.7a) we get

$$\begin{aligned}
\text{aff}(\partial\|D\overline{x}\|_1) &= \text{aff}(\{D^\top\alpha \colon \alpha_{\Lambda^c} = \text{sign}(D_{\Lambda^c,\bullet}\overline{x}), \|\alpha_\Lambda\|_\infty \le 1\}) \tag{2.14}\\
&= \text{aff}(D_{\Lambda^c,\bullet}{}^\top \text{sign}(D_{\Lambda^c,\bullet}\overline{x}) + \{D_{\Lambda,\bullet}{}^\top\alpha \colon \|\alpha\|_\infty \le 1\}) \tag{2.15}\\
&= \underbrace{D_{\Lambda^c,\bullet}{}^\top \text{sign}(D_{\Lambda^c,\bullet}\overline{x})}_{=:y_0} + \text{aff}(\{D_{\Lambda,\bullet}^\top\alpha \colon \|\alpha\|_\infty \le 1\}) \tag{2.16}\\
&= y_0 + \mathcal{R}(D_{\Lambda,\bullet}^\top). \tag{2.17}
\end{aligned}$$

By (2.7b) we obtain

$$\text{aff}(\partial\delta_{\mathcal{X}}(\overline{x})) = \text{aff}(N_{\mathcal{X}}(\overline{x})) = \text{aff}(\mathcal{N}(A)^\perp) = \mathcal{N}(A)^\perp. \tag{2.18}$$

And finally, by (2.8) we write

$$\begin{aligned}
\text{aff}(\partial\delta_{[l,u]}(\overline{x})) &= \text{aff}(N_{[l,u]}(\overline{x})) = \text{aff}(\mathbb{R}_-^{|S_l|} \times \{0\}^{|S_{lu}|} \times \mathbb{R}_+^{|S_u|}) \tag{2.19}\\
&= \mathbb{R}^{|S_l|} \times \{0\}^{|S_{lu}|} \times \mathbb{R}^{|S_u|} = \mathcal{R}(\Psi). \tag{2.20}
\end{aligned}$$

Hence, (2.13) is equivalent to $\mathbb{R}^n = \mathcal{R}(D_{\Lambda,\bullet}^\top) + \mathcal{N}(A)^\perp + \mathcal{R}(\Psi)$. Taking the orthogonal, condition (a) now becomes

$$\begin{aligned}
&\mathcal{R}(D_{\Lambda,\bullet}^\top)^\perp \cap (\mathcal{N}(A)^\perp)^\perp \cap \mathcal{R}(\Psi)^\perp \tag{2.21}\\
&= \mathcal{N}(D_{\Lambda,\bullet}) \cap \mathcal{N}(A) \cap \mathcal{N}(\Psi) = \{0\}, \tag{2.22}
\end{aligned}$$

that yields (i).

Further we reformulate condition (b). We use that for two convex sets C_1, C_2 it holds

$$\text{rint}(C_1 + C_2) = \text{rint}(C_1) + \text{rint}(C_2),$$

see (5.2), and so we get

$$0 = \text{rint}(\partial\|D\overline{x}\|_1 + \partial\delta_{\mathcal{X}}(\overline{x}) + \partial\delta_{[l,u]}(\overline{x})) \tag{2.23}$$

$$\neq\text{rint}(\partial\|D\overline{x}\|_1) + \text{rint}(\partial\delta_{\mathcal{X}}(\overline{x})) + \text{rint}(\partial\delta_{[l,u]}(\overline{x})) \tag{2.24}$$

$$\neq\text{rint}(y_0 + \{D_{\Lambda,\bullet}^\top\alpha \colon \|\alpha\|_\infty \le 1\}) + \text{rint}(\mathcal{R}(A^\top)) + \text{rint}(\mathbb{R}_-^{|S_l|} \times \{0\}^{|S_{lu}|} \times \mathbb{R}_+^{|S_u|}) \tag{2.25}$$

$$\neq y_0 + \{D_{\Lambda,\bullet}^\top\alpha \colon \|\alpha\|_\infty < 1\} + \mathcal{R}(A^\top) + \mathbb{R}_{--}^{|S_l|} \times \{0\}^{|S_{lu}|} \times \mathbb{R}_{++}^{|S_u|} \tag{2.26}$$

$$\neq y_0 + \{D_{\Lambda,\bullet}^\top\alpha \colon \|\alpha\|_\infty < 1\} + \mathcal{R}(A^\top) + \{\Psi\mu \colon \mu > 0\}. \tag{2.27}$$

This shows (ii) and concludes the proof. $\qquad\square$

Remark 2.1. By Thm. 2.2 we can show that $\overline{x}$ is the *unique* solution of (2.5) if and only if (i) and (ii) holds. Condition (i) is tested by evaluating whether $\left(A^\top, \ D_{\Lambda,\bullet}^\top, \ \Psi^\top\right)^\top$ has full column rank. For verifying (ii) we need to test whether there exist $y \in \mathbb{R}^m, \ \alpha \in \mathbb{R}^p, \ \mu \in \mathbb{R}^n : D^\top\alpha + \Psi\mu = A^\top y, \alpha_{\Lambda^c} = \text{sign}(D_{\Lambda^c,\bullet}\overline{x}), \|\alpha_\Lambda\|_\infty < 1, \mu > 0$.

The second condition (ii) could be verified in practice by minimizing $\|\alpha_\Lambda\|_\infty$ w.r.t. y, α, μ while respecting the equality constraints and converting the strict inequality constraint $\mu > 0$ to $\mu >= \varepsilon\mathbb{1}$ for a small ε, e.g. $\varepsilon = 10^{-8}$, to ensure that the inequality is satisfied strictly. This results in a linear program (LP),

$$\min_{t,y,\alpha,\mu} t \qquad \text{s.t.} \quad -t\mathbb{1} \le \alpha_\Lambda \le t\mathbb{1}, \tag{2.28}$$

$$A^\top y = D^\top\alpha + \Psi\mu \tag{2.29}$$

$$\mu \ge \varepsilon\mathbb{1}. \tag{2.30}$$

For the optimal solution $(t^*, y^*, \alpha^*, \mu^*)$ we have by definition the smallest possible $t := \|\alpha_\Lambda\|_\infty$. If t^* is not smaller than one, then no y exists with smaller t. We therefore declare $\overline{x}$ the unique minimizer if $t^* < 1$, and if $t^* \ge 1$, $\overline{x}$ cannot be the unique minimizer. Numerically, we would test whether $t^* \le 1 - \varepsilon$. Technically, by applying the above procedure we risk rejecting a unique solution $\overline{x}$, for which $1 - \varepsilon < t^* < 1$. Recall that the choice for ε is ad hoc.

In order to resolve this issue we provide next a theoretically well-founded methodology to deal with the *strict* feasibility problem

$$D^\top \alpha + \Psi\mu = A^\top y, \tag{2.31}$$

$$\alpha_{\Lambda^c} = \text{sign}(D_{\Lambda^c, \bullet}\overline{x}), \tag{2.32}$$

$$\|\alpha_\Lambda\|_\infty < 1, \tag{2.33}$$

$$\mu > 0. \tag{2.34}$$

The above feasibility problem can be recast as a linear system of inequalities, as we will see next. For this purpose we transform the above problem into the form $Mz = q, Pz < p$.

Theorem 2.3. *Let M be a matrix so that $\mathcal{N}(M^\top) = \{0\}$. Then there is a point $\overline{z}$ with $M\overline{z} = q, P\overline{z} < d$ if and only if $v = 0$ is the only feasible solution of the problem*

$$q^\top u + p^\top v \leq 0, \quad M^\top u + P^\top v = 0, \quad v \geq 0. \tag{2.35}$$

Proof. We consider the following pair of linear programs

$$\begin{array}{ll} \max 0 & \text{(P)} \\ \text{s.t. } Mz = q & \\ \quad Pz \leq d, & \end{array} \qquad \begin{array}{ll} \min q^\top u + d^\top v & \text{(DP)} \\ \text{s.t. } M^\top u + P^\top v = 0 & \\ \quad v \geq 0 & \end{array}$$

where (P) is the primal problem and (DP) its dual.

First, note that under the assumption $\mathcal{N}(M^\top) = \{0\}$, vector $v = 0$ is the only feasible solution of (2.35) if and only if $v = 0$ is the only solution of (DP). Indeed, the common constraint $M^\top u + P^\top v = 0$ implies due to the assumptions $v = 0$ and $\mathcal{N}(M^\top) = \{0\}$ that $(u, v) = (0, 0)$.

Hence, we can show the statement of the theorem by showing that (P) is strictly feasible if and only if $v = 0$ is the only solution of (DP).

Assume on one hand that (P) is strictly feasible. Since any feasible solution is also a solution of (P) we can deduce from the existence of a primal solution the existence of a dual solution. The optimality conditions yield in particular $(d - P\overline{z})^\top v^* = 0$ for such a primal solution $\overline{z}$ of (P) and any dual solution (u^*, v^*) of (DP). Since $P\overline{z} < d$ holds, it follows that $v^* = 0$ is the only solution.

Assume on the other hand that for each dual solution (u^*, v^*) it holds $v^* = 0$. In view of the strict complementarity condition [Van14, Thm 10.7] it exists a pair of primal and dual solutions such that

$$(d - Pz^*) + v^* > 0. \tag{2.36}$$

Since by assumption $v^* = 0$, (2.36) implies the existence of a primal solution z^* with $Pz^* < d$.

This concludes the proof. $\square$

Remark 2.2. In view of Thm. 2.3 we can check feasibility of the system (2.31) – (2.34) by linear programming. We first associate to the system (2.31) – (2.34) with strict inequality constraints the primal linear program

$$\max_{\alpha,\mu,y} 0 \tag{$P_<$}$$

$$\text{s.t. } \underbrace{\begin{pmatrix} D^\top_{\Lambda,\bullet} & \Psi & A^\top \end{pmatrix}}_{=:M} \underbrace{\begin{pmatrix} \alpha \\ \mu \\ y \end{pmatrix}}_{=:z} = \underbrace{-D^\top_{\Lambda^c,\bullet} \operatorname{sign}(D_{\Lambda^c,\bullet}\overline{x})}_{=:q}$$

$$\underbrace{\begin{pmatrix} -I & 0 & 0 \\ I & 0 & 0 \\ 0 & -I & 0 \end{pmatrix}}_{=:P} \underbrace{\begin{pmatrix} \alpha \\ \mu \\ y \end{pmatrix}}_{=:z} \le \underbrace{\begin{pmatrix} \mathbb{1} \\ \mathbb{1} \\ 0 \end{pmatrix}}_{=:d},$$

as in the Thm. 2.3. Assumption $\mathcal{N}(M^\top) = \{0\}$ in Thm. 2.3 now corresponds to the first condition from Thm. 2.2. Assuming that it holds, we can now verify system (2.31) – (2.34) and strict feasibility in ($P_<$) and if

$$\max \langle \mathbb{1}, u^+ + u^- + w \rangle \tag{$DP_<$}$$
$$\text{s.t. } \langle \mathbb{1}, u^+ + u^- \rangle \le \langle \operatorname{sign}(D_{\Lambda^c,\bullet}\overline{x}), D_{\Lambda^c,\bullet}v \rangle$$
$$D_{\Lambda,\bullet}v = u^+ - u^-$$
$$\Psi v = w$$
$$Av = 0$$
$$u^+, u^-, w \ge 0$$

has an optimal objective value equal to zero.

2.3 Case Studies: (1.3a)-(1.3f)

In the previous section we provided verifiable uniqueness conditions for problem (2.5). In this section we specialize these conditions to verifiable uniqueness optimality conditions for our objective functions (1.3a)-(1.3f).

Corollary 2.4. *(Case f_1) A feasible point $\overline{x}$ of problem* (1.2) *with the objective function defined in* (1.3a) *is the unique solution if and only if*

$$\exists y \in \mathbb{R}^m : A^\top_{\bullet,S}y = \operatorname{sign}(\overline{x}_S) \wedge \|A^\top_{\bullet,S^c}y\|_\infty < 1 \text{ and}$$
$$A_{\bullet,S} \text{ is injective,}$$

where $S := \operatorname{supp}(\overline{x})$ and $A_{\bullet,S}$ selects the columns indexed by S.

Proof. We derive the above claim from Thm. 2.2. In view of the definition in (1.3a) we set $D = I$ the identity matrix and $\Psi = 0$ the zero matrix. The latter is due to the fact that $\mathbb{R}^n$ can be seen as $\delta_{[l,u]}(\overline{x})$ with $l_i = -\infty$, $u_i = \infty$, for all $i \in [n]$. Hence $\overline{x}$ always lies in the interior of $\mathbb{R}^n$ and $N_{\mathbb{R}^n}(\overline{x}) = 0$, compare (2.8). These choices specialize the conditions of Thm. 2.2 to the following conditions

(i) $\mathcal{N}(A) \cap \mathcal{N}(I_{S^c,\bullet}) = \{0\}$,

(ii) $\exists \alpha \in \mathbb{R}^n : \alpha \in \mathcal{R}(A^\top), \alpha_S = \text{sign}(\overline{x}_S), \|\alpha_{S^c}\|_\infty < 1$,

taking into account also that $\Lambda = S^c$ and $\mathcal{N}(\Psi) = \mathbb{R}^n$. Condition (i) can be further simplified to

$$\mathcal{N}(A) \cap \mathcal{N}(I_{S^c,\bullet}) = \mathcal{N}(A) \cap \mathcal{N}(I_{\bullet,S^c}) = \mathcal{N}(A) \cap \{x \colon x_{S^c} = 0\} = \mathcal{N}(A_{\bullet,S}) = \{0\},$$
$$(2.37)$$

which yields the condition that $A_{\bullet,S}$ has to be injective. We conclude the proof by transforming the condition (ii) to

$$\exists \alpha \in \mathbb{R}^n, y \in \mathbb{R}^m : \alpha = A^\top y \wedge \alpha_S = \text{sign}(\overline{x}_S) \wedge \|\alpha_{S^c}\|_\infty < 1 \qquad (2.38)$$
$$\Leftrightarrow \exists y \in \mathbb{R}^m : A_{\bullet,S}^\top y = \text{sign}(\overline{x}_S) \wedge \|A_{\bullet,S^c}^\top y\|_\infty < 1. \qquad (2.39)$$

$\square$

Corollary 2.5. *(Case f_2) A feasible point $\overline{x}$ of problem* (1.2) *with the objective function defined in* (1.3b) *is the unique solution if and only if*

$$\exists y \in \mathbb{R}^m : A_{\bullet,S}^\top y = \mathbb{1}_S \wedge A_{\bullet,S^c}^\top y < \mathbb{1}_{S^c} \text{ and}$$
$$A_{\bullet,S} \text{ is injective,}$$

with $S := \text{supp}(\overline{x})$.

Proof. As in the proof of Cor. 2.4 we apply Thm. 2.2 with $D = I$ and $\Psi_{ii} = 0$, for $i \in S$ and $\Psi_{ii} = -1$, for $i \in S^c$, since $N_{R_+^n}(\overline{x}) = \{0\}^{|S|} \times \mathbb{R}_-^{|S^c|}$. Conditions (i) and (ii) in Thm. 2.2 become

(i) $\mathcal{N}(A) \cap \mathcal{N}(I_{S^c,\bullet}) \cap \mathcal{N}(\Psi) = \{0\}$,

(ii) $\exists \alpha \in \mathbb{R}^n, \exists \mu \in \mathbb{R}^n :$
$\alpha + \Psi\mu \in \mathcal{R}(A^\top), \alpha_S = \text{sign}(\overline{x}_S), \|\alpha_{S^c}\|_\infty < 1, \mu > 0$.

Since $\mathcal{N}(I_{S^c, \bullet}) = \mathcal{N}(\Psi)$ condition (i) above simplifies as in the proof of Cor. 2.4 to condition: $A_{\bullet, S}$ is injective.

Since $\{\Psi\mu : \mu \in \mathbb{R}^n_{++}\} = \Psi\mathbb{R}^n_{++} = \{0\}^{|S|} \times \mathbb{R}^{|S^c|}_{--}$ and

$$\{\alpha_{S^c} : \|\alpha_{S^c}\|_\infty < 1\} + \mathbb{R}^{|S^c|}_{--} = \mathbb{1}_{S^c} + \mathbb{R}^{|S^c|}_{--},$$

condition (ii) above is equivalent to the existence of a vector $y \in \mathbb{R}^m$ such that

$$A^\top y \in \mathbb{1} + \{0\}^{|S|} \times \mathbb{R}^{|S^c|}_{--},$$

that can be written as

$$A^\top_{\bullet, S} y = \mathbb{1}_S, \quad A^\top_{\bullet, S^c} y < \mathbb{1}_{S^c}. \tag{2.40}$$

This completes the proof. □

Corollary 2.6. *(Case f_3) A binary feasible point $\overline{x} \in \{0, 1\}^n$ of problem* (1.2) *with the objective function defined in* (1.3c) *is the unique solution if and only if*

$$\exists y \in \mathbb{R}^m \; : \; A^\top_{\bullet, S} y > \mathbb{1} \wedge A^\top_{\bullet, S^c} y < \mathbb{1}, \tag{2.41}$$

with $S := \text{supp}(\overline{x})$.

Proof. As above we specialize Thm. 2.2 with $D = I$ and $\Psi_{ii} = -1$, for $i \in S^c$ and $\Psi_{ii} = 1$, for $i \in S$, since $N_{[0,1]^n}(\overline{x}) = \mathbb{R}^{|S|}_+ \times R^{|S^c|}_-$ for $\overline{x} \in \{0, 1\}^n$. Hence $\mathcal{N}(\Psi) = \{0\}$. As a consequence condition (i) Thm. 2.2 is automatically fulfilled. We now focus on condition (ii) that now reads

$$\begin{pmatrix} \alpha_S \\ \alpha_{S^c} \end{pmatrix} + \begin{pmatrix} \mu^+ \\ \mu^- \end{pmatrix} \in \mathcal{R}\left(\begin{pmatrix} A^\top_{\bullet, S} \\ A^\top_{\bullet, S^c} \end{pmatrix}\right), \text{ with } \mu^- < 0, \mu^+ > 0, \alpha_S = \mathbb{1}, \|\alpha_{S^c}\|_\infty < 1, \tag{2.42}$$

since for $\overline{x} \in \{0, 1\}^n$ we have $\alpha_S = \text{sign}(\overline{x}_S) = \mathbb{1}$. In view of

$$\{\alpha_{S^c} : \|\alpha_{S^c}\|_\infty < 1\} + \mathbb{R}^{|S^c|}_{--} = \mathbb{1}_{S^c} + \mathbb{R}^{|S^c|}_{--},$$

condition (2.42) is equivalent to the existence of a vector $y \in \mathbb{R}^m$ such that

$$A^\top y \in \mathbb{1} + \mathbb{R}^{|S|}_{++} \times \mathbb{R}^{|S^c|}_{--},$$

that can be written as

$$A^\top_{\bullet, S} y > \mathbb{1}_S, \quad A^\top_{\bullet, S^c} y < \mathbb{1}_{S^c}. \tag{2.43}$$

□

Corollary 2.7. *(Case f_4) A feasible point $\overline{x} \in \mathbb{R}^n$ of problem* (1.2) *with the objective function defined in* (1.3d) *is the unique solution if and only if*

(i) $\mathcal{N}(A) \cap \mathcal{N}(D_{\Lambda,\bullet}) = \{0\}$,

(ii) $\exists \alpha : D^\top \alpha \in \mathcal{R}(A^\top), \alpha_{\Lambda^c} = \operatorname{sign}(D_{\Lambda^c,\bullet}\overline{x}), \|\alpha_\Lambda\|_\infty < 1$,

where $\Lambda = \operatorname{cosupp}(D\overline{x})$.

Proof. As in the proof of Cor. 2.4 we conclude that $\Psi = 0$ is the zero matrix. Hence, the two conditions of Thm. 2.2 simplify to the two conditions above, that do not involve the variable μ. □

Corollary 2.8. *(Case f_5) A feasible point $\overline{x} \in \mathbb{R}^n_+$ of problem* (1.2) *with the objective function defined in* (1.3e) *is the unique solution if and only if*

(i) $\mathcal{N}(A) \cap \mathcal{N}(D_{\Lambda,\bullet}) \cap \mathcal{N}(\Psi) = \{0\}$,

(ii) $\exists \alpha, \exists \mu : D^\top \alpha + \Psi \mu \in \mathcal{R}(A^\top), \alpha = \operatorname{sign}(D_{\Lambda^c,\bullet}\overline{x}), \|\alpha_\Lambda\|_\infty < 1, \mu > 0$

where $\Lambda = \operatorname{cosupp}(D\overline{x})$ and

$$\Psi_{ii} = \begin{cases} -1, & \overline{x}_i = 0, \\ 0, & otherwise. \end{cases} \tag{2.44}$$

Proof. Immediate from Thm. 2.2 and the definition of f_5 in (1.3e) and Ψ. □

Corollary 2.9. *(Case f_6) A binary feasible point $\overline{x} \in \{0,1\}^n$ of problem* (1.2) *with the objective function defined in* (1.3f) *is the unique solution if and only if*

(i) $\mathcal{N}(A) \cap \mathcal{N}(D_{\Lambda,\bullet}) \cap \mathcal{N}(\Psi) = \{0\}$,

(ii) $\exists \alpha, \exists \mu : D^\top \alpha + \Psi \mu \in \mathcal{R}(A^\top), \alpha = \operatorname{sign}(D_{\Lambda^c,\bullet}\overline{x}), \|\alpha_\Lambda\|_\infty < 1, \mu > 0$

where $\Lambda = \operatorname{cosupp}(D\overline{x})$ and

$$\Psi_{ii} = \begin{cases} -1, & \overline{x}_i = 0, \\ 1, & \overline{x}_i = 1. \end{cases} \tag{2.45}$$

Proof. In view of the definition of f_6 in (1.3f) and Ψ above $\mathcal{N}(\Psi) = \{0\}$. As a consequence condition (i) Thm. 2.2 is automatically fulfilled. Condition (ii) from Thm. 2.2 is kept unchanged. □

3 Probabilistic Uniqueness

In this section we are concerned with probabilistic solution uniqueness when minimizing some structure enforcing regularizer $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ via (1.2) in the noiseless setting of (1.1). In particular we address the question of predicting uniqueness with high probability for a specific number of random linear measurements, when knowing *only* the solution co-/sparsity.

For $f = \| \cdot \|_1$ and a Gaussian matrix A it is well-known that the recovery of a sparse vector x depends only on its sparsity level, and is independent of the locations or values of the nonzero entries. There are precise relations between the sparsity s, ambient dimension n, and the number of samples m that guarantees success of (1.2). Moreover, several authors have shown that there is a transition from absolute success to absolute failure, and they have *accurately* characterized the location of the threshold, also called *phase transition*, in the ℓ_1-case and other *norms*, e.g. nuclear norm. For the TV-seminorm such a complete analysis is still missing. As mentioned in the introduction, the authors in [ZXCL16] have recently shown that in the case of 1D TV-minimization, case (1.3d), the phase transition can be accurately described by an effective bound for the statistical dimension of a descent cone, which is based on the squared distance of a standard normal vector to the subdifferential of the objective function at the sought solution $\overline{x}$,

$$\min_{\tau \geq 0} \mathbb{E}[\mathrm{dist}(X, \tau \partial f(\overline{x})].$$

We will explore next if the same relation that describes phase transitions for ℓ_1-minimization and 1D TV-minimization also hold for all our objective functions in (1.3a)-(1.3f).

3.1 Phase Transitions in Random Linear Inverse Problems

We collect here some recent results of convex signal reconstruction with a Gaussian sampling model and briefly explain how classical results for Gaussian processes lead to a sharp bound for the number of Gaussian measurements that suffice for exact recovery.

Definition 3.1. (Statistical dimension, [ALMT14]) The *statistical dimension* $\delta(K)$ of a closed, convex cone $K \subset \mathbb{R}^n$ is the quantity

$$\delta(K) = \mathbb{E}[\|\Pi_K(X)\|_2^2], \tag{3.1}$$

where Π_K is the Euclidean projection onto K and X is a standard normal vector, i.e. $X \sim \mathcal{N}(0, I_n)$.

Definition 3.2. (Descent cone) The *descent cone* of a proper convex function $f :$ $\mathbb{R}^n \to \overline{\mathbb{R}}$ at a point x is

$$\mathcal{D}_f(\overline{x}) := \text{cone}\{z - \overline{x} \colon f(z) \leq f(\overline{x})\}, \tag{3.2}$$

i.e. the conic hull of the directions that do not increase f near $\overline{x}$.

Definition 3.3. (Gaussian width, [CRPW12]) The *Gaussian width* of $C \subset \mathbb{R}^n$ is

$$\omega(C) = \mathbb{E}[\sup_{z \in C} \langle X, z \rangle],$$

where X is a standard normal vector, i.e. $X \sim \mathcal{N}(0, I_n)$.

Remark 3.1. One has, see [ALMT14], the relationship

$$\omega(K)^2 \leq \delta(K) \leq \omega(K)^2 + 1. \tag{3.3}$$

A classic result for solution uniqueness, equivalent to the exact recovery of individual vectors [FR13, Thm. 4.35], is given next.

Proposition 3.1. *The vector $\overline{x}$ is the* unique *solution of the convex program* (1.2) *if and only if*
$$\mathcal{D}_f(\overline{x}) \cap \mathcal{N}(A) = \{0\}. \tag{3.4}$$

Condition (3.4) holds with high probability for a random matrix A is equal to the probability that a randomly rotated convex cone, here $\mathcal{N}(A)$, shares a ray with a fixed convex cone $\mathcal{D}_f(\overline{x})$. This probability is bounded in terms of the statistical dimension $\delta(\mathcal{D}_f(\overline{x}))$ as stated next.

Theorem 3.2. *[ALMT14, Thm. II] Fix a tolerance parameter $\varepsilon \in (0, 1)$. Let $\overline{x} \in \mathbb{R}^n$ be a fixed vector, and let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper convex function. Suppose $A \in \mathbb{R}^{m \times n}$ has independent standard normal entries, and let $b = A\overline{x}$. Then*

- $m \leq \delta(\mathcal{D}_f(\overline{x})) - a_n\sqrt{n} \Longrightarrow$ (1.2) *succeeds with probability $\leq \varepsilon$;*
- $m \geq \delta(\mathcal{D}_f(\overline{x})) + a_n\sqrt{n} \Longrightarrow$ (1.2) *succeeds with probability $\geq 1 - \varepsilon$,*

with $a_n := \sqrt{8\log(4/\varepsilon)}$.

There is a useful tool for bounding the statistical dimension in terms of $\partial f(\overline{x})$. Interestingly, this bound is tight for some classes of f, e.g. norms.

Proposition 3.3. *[CRPW12, ALMT14] Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper convex function and let $x \in \mathbb{R}^n$. Assume that the subdifferential $\partial f(\overline{x})$ is nonempty, compact, and does not contain the origin. Define the function*

$$J(\tau) = J(\tau; \partial f(\overline{x})) := \mathbb{E}[\text{dist}(X, \tau\partial f(\overline{x})], \quad \text{for } \tau \geq 0, \tag{3.5}$$

where $X \sim \mathcal{N}(0, I)$. We have

$$\delta(\mathcal{D}_f(\overline{x})) \leq \inf_{\tau \geq 0} J(\tau). \tag{3.6}$$

Furthermore, the function J is strictly convex, continuous at $\tau = 0$, and differentiable for $\tau \geq 0$. It achieves its minimum at a unique point.

The above results provide a recipe to upper bound $\delta(\mathcal{D}_f(\overline{x}))$. This consists of the following steps.

- Compute $\partial f(\overline{x})$;

- For each $\tau > 0$, compute $J(\tau) = \mathbb{E}[\text{dist}(X, \tau \partial f(\overline{x})];$

- Find the unique solution to $J'(\tau) = 0$.

We also have the following error bound.

Theorem 3.4. *[ALMT14, Thm. 4.3.] Let $f : \mathbb{R}^n \to \mathbb{R}$ be a norm on $\mathbb{R}^n$ and let $x \in \mathbb{R}^n \setminus \{0\}$. Then*

$$0 \leq \inf_{\tau \geq 0} J(\tau) - \delta(\mathcal{D}_f(\overline{x})) \leq \frac{2 \sup\{\|p\| : p \in \partial f(\overline{x})\}}{f(\overline{x}/\|\overline{x}\|)} \tag{3.7}$$

For the ℓ_1-case the r.h.s. can be made very accurate for large sparsity parameters s providing an accurate estimate of the statistical dimension of the ℓ_1-descent cone. On the other hand, this error estimate is not very accurate when the sparsity s is small. The work in [FM14] contains a bound that improves this result and also extends to more general functions with some additional properties.

Theorem 3.5. *[FM14, Prop. 1] Suppose that, for $\overline{x} \neq 0$, the subdifferential $\partial f(\overline{x})$ satisfies a weak decomposability assumption, i.e.*

$$\exists \overline{p} \in \partial f(\overline{x}) \quad \textit{such that} \quad \langle p - \overline{p}, \overline{p} \rangle = 0, \qquad \forall p \in \partial f(\overline{x}). \tag{3.8}$$

Then

$$\delta(\mathcal{D}_f(\overline{x})) \leq \inf_{\tau \geq 0} J(\tau) \leq \delta(\mathcal{D}_f(\overline{x})) + 6, \tag{3.9}$$

with $J(\tau) = \mathbb{E}[\text{dist}(X, \tau \partial f(\overline{x})]$.

3.2 Phase Transitions for Case Studies: (1.3a)-(1.3f)

Above we discussed probabilistic uniqueness for the general problem

$$\min_{Ax = A\overline{x}} f(x).$$

In this section we discuss how and if we can specialize these results to predict uniqueness for our objective functions (1.3a)-(1.3f).

Remark 3.2. In the case of $f_1(x) = \|x\|_1$ from (1.3a) and $f_4(x) = \|Dx\|_1$ from (1.3d) the statistical dimension of the descent cone $\mathcal{D}_f(\overline{x})$ can be well approximated in terms of the expected squared distance to the scaled subdifferential. The reason is that the subdifferential $\partial f(\overline{x})$ is compact, see (2.7a), and thus by Prop. 3.3 we know that $J(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ has a unique minimizer. Moreover, $\min_{\tau \geq 0} J(\tau)$ provides a tight approximation in view of Thm. 3.4 and Thm. 3.5, since f_1 is a norm and ∂f_4 is weakly decomposable for the choice D equal to the 1D finite difference operator, as shown in [ZXCL16]. For the 2D finite difference operator the weakly decomposable property has not been previously investigated.

Remark 3.3. The above situation changes for the objectives

$$f_2(x) = \|x\|_1 + \delta_{\mathbb{R}_+^n}(x), \qquad f_3(x) = \|x\|_1 + \delta_{[0,1]^n}(x)$$
$$f_5(x) = \|Dx\|_1 + \delta_{\mathbb{R}_+^n}(x), \qquad f_6(x) = \|Dx\|_1 + \delta_{[0,1]^n}(x),$$

from (1.3b), (1.3c), (1.3e) and (1.3f). It is clear that $\partial f(\overline{x})$ is *not* compact, for all above choices, compare (2.7a) and (2.7c). Thus, we have no guarantee that, J has even a minimizer. Since $\partial f(\overline{x})$ is *not* weakly decomposable, even if $\min_{\tau \geq 0} J(\tau)$ exists, we are not guaranteed a tight upper bound on the statistical dimension $\delta(\mathcal{D}_f(\overline{x}))$ and consequently on the number of sufficient Gaussian measurements m for exact recovery of $\overline{x}$ with high probability.

Explicit curves for $\min_\tau J(\tau)$ can be computed only in the case of f_1, f_2, f_3 along the lines of [ALMT14]. We skip the details here and illustrate this curves in Sect. 4. For f_4, f_5 and f_6 explicit curves for J and for $\min_\tau J(\tau)$ are missing. We use an approximation to the upper bound $J(\tau) = \mathbb{E}\left[\mathrm{dist}^2(X, \tau \partial f(\overline{x}))\right]$ where $X \sim \mathcal{N}(0, I)$. To approximate the expected value we use a very large $k \in \mathbb{N}$ and calculate

$$h_k(\tau) = \frac{1}{k} \sum_{i=1}^{k} \mathrm{dist}^2(X_i, \tau \partial f(\overline{x})) =: J_{\mathrm{approx}}(\tau) \approx J(\tau),$$

where each X_i is sampled from the normal distribution $\mathcal{N}(0, I)$. Note that for computing each $\mathrm{dist}^2(X_i, \tau \partial f(\overline{x}))$ one is faced with solving a quadratic program, i.e.

$$\min \|X_i - \tau y\|_2^2 \qquad \text{subject to} \quad y \in \partial f(\overline{x}).$$

Note that the constraints are linear since $\partial f(\overline{x})$ is described by linear equalities and inequalities, compare (2.7a) and (2.7c).

Based on the empirical results from Sect. 4 we conjecture.

Conjecture 3.1. *For the choices of f in Remark 3.3 the statistical dimension is well approximated similar to* (3.9) *by the squared Euclidean distance to the scaled subdifferential, i.e.*

$$\min_{\tau} J(\tau) := \mathbb{E}\left[\operatorname{dist}^2(X, \tau\partial f(\overline{x}))\right] \approx \delta(\partial f(\overline{x})).$$

4 Experiments

Our objective in this section is to *empirically* verify whether phase transitions occur at

$$\min_{\tau \geq 0} J(\tau),$$

which is a *guaranteed* upper bound to the statistical dimension $\delta(\mathcal{D}_f(\overline{x}))$ *only* in some of the considered cases. We note further, that precise expressions for $J(\tau)$ are only available for ℓ_1-minimization and cases (1.3a)–(1.3c). For the TV-minimization and cases (1.3d)–(1.3f) we construct phase diagrams by *numerically estimating* $\min_{\tau \geq 0} J(\tau)$ and comparing the resulting bounds to the average-case results for both solving the reconstruction problems (1.2) and certifying uniqueness by the procedure derived in Sect. 2.

4.1 1D Empirical Phase Transitions and Theoretical Bounds

In this section we address ℓ_1-minimization and one dimensional TV-minimization without and with constraints. The 1D TV-regularizer reads $\|\partial_n x\|_1$, where $\partial_n \in \mathbb{R}^{n-1 \times n}$ is the one-dimensional discrete derivative operator. Applying $D = \partial_n$ to a signal x gives the the entries' differences for consecutive indices

$$Dx = \partial_n x = \begin{pmatrix} x_1 - x_0 \\ \vdots \\ x_n - x_{n-1} \end{pmatrix}. \tag{4.1}$$

4.1.1 Testset

Our testset consists of several randomly generated signals with specified relative sparsity ρ ranging from 0.05 to 0.95 with step size 0.05. For each of these relative sparsities we create a signal so that

$$\rho \approx \frac{\|\overline{x}\|_0}{n} \qquad \text{resp.} \qquad \rho \approx \frac{\|D\overline{x}\|_0}{p}. \tag{4.2}$$
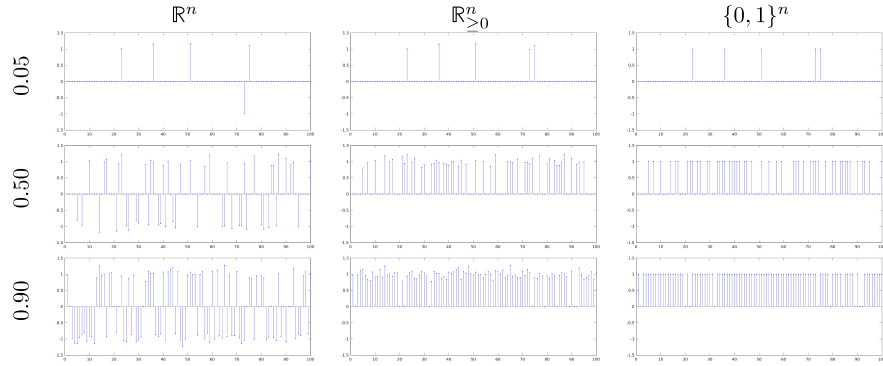
Figure 4.1: *Testset for 1D sparse signals*. The rows display signals $\overline{x}$ with equal number of non-zeros in $\overline{x}$, i.e. $\|\overline{x}\|_0$. The columns display signals that have entries sampled either from $\mathbb{R}, \mathbb{R}_+$ or $\{0, 1\}$ from left to right.
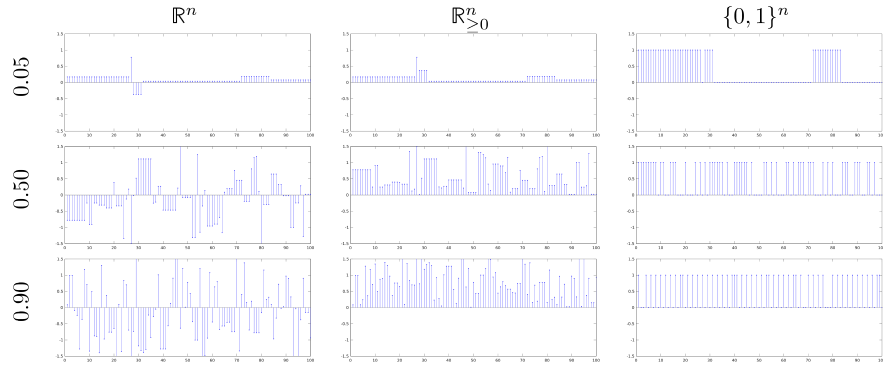

Figure 4.2: *Testset for 1D gradient sparse signals*. The rows display signals $\overline{x}$ with equal number of non-zeros in $D\overline{x}$, i.e. $\|D\overline{x}\|_0$, with respect to the 1D finite difference operator. The columns display signals that have entries sampled either from $\mathbb{R}, \mathbb{R}_+$ or $\{0, 1\}$ from left to right.

97

Algorithm 1: Algorithm for converting a random gradient sparse signal (w.r.t. the 1D finite differences operator) into a binary gradient sparse signal.

    Input: $\Lambda \subseteq [n-1]$
    Output: Binary signal $\overline{x} \in \{0,1\}^n$

$\overline{x}_0 = 1$;
for $i \in [n-1]$ **do**
    if $i \in \Lambda$ **then**
        $\overline{x}_{i+1} \leftarrow \overline{x}_i$;
    else
        $\overline{x}_{i+1} \leftarrow 1 - \overline{x}_i$;
return $\overline{x}$;

Hence, we cover almost the full range from highly sparse to dense signals. In addition, we create three different classes of signals: with real-valued, non-negative or binary entries, i.e. $x_i \in \mathbb{R}$, $x_i \geq 0$ are $x_i \in \{0,1\}$ for all $i \in [n]$.

Creating random real-valued, non-negative or binary signals of a given sparsity is immediate. To create signals that are sparse in a transformed domain we use an idea from of [NDEG13]. Instead of choosing the support of $\overline{x}$ uniformly at random, ones chooses a set $\Lambda \subseteq [p]$ in order to create a signal with $D_{\Lambda,\bullet}\overline{x} = 0$. Having the subset Λ and a randomly created vector v with non-zero entries we obtain the desired signal by calculating

$$\overline{x} = (I - D_{\Lambda,\bullet}^{\top}(D_{\Lambda,\bullet}D_{\Lambda,\bullet}^{\top})^{-1}D_{\Lambda,\bullet})v. \tag{4.3}$$

Since $D\overline{x}$ represents differences of signal entries for consecutive indices, taking the absolute value of each signal entry will most likely not change the cosupport of the resulting signal. Therefore, we can use this method to generate non-negative signals. Binary signals can be obtained by using Algorithm 1.

4.1.2 ℓ_1-**Minimization**

In this section we compare the existing tight upper bounds for the statistical dimension $\delta(\mathcal{D}_f(\overline{x}))$ for f defined in (1.3a)–(1.3c) to the empirical phase transition obtained for the above set of signals.

To this end we first set the ambient dimension $n = 100$ and generate 10 instances of a sparse 1D signal, see Figure 4.1, and a random Gaussian matrix $A \in \mathbb{R}^{m \times}$. For each pair of relative sparsity and given number m of random linear measurements we verify if the signal is the unique solution of (1.2) with f defined in (1.3a)–(1.3c).

Uniqueness is tested as detailed in Sect. 2. We used Mosek* to solve the optimization problems. The gray value plots in Figure 4.4, first row, show the empirical probability that a given 1D sparse signal is *uniquely* reconstructed by the convex relaxation. White means recovered and unique, and black non recovered. The red curves show $\min_{\tau \geq 0} J(\tau)$, plotted with Mathematica†, and separate these regions accurately.

4.1.3 TV-**Minimization**

To generate the phase diagrams we proceed as above but use the gradient sparse 1D test signals, see Figure 4.2. Since explicit curves for J are missing we use an approximation to the upper bound $J(\tau) = \mathbb{E}\left[\text{dist}^2(X, \tau \partial f(\overline{x}))\right]$, where $X \sim \mathcal{N}(0, I)$. To approximate the expected value we use a large $k = 10000$ and calculate

$$h_k(\tau) = \frac{1}{k} \sum_{i=1}^{k} \text{dist}^2(X_i, \tau \partial f(\overline{x})) =: J_{\text{approx}}(\tau) \approx J(\tau),$$

where each X_i is sampled from the normal distribution $\mathcal{N}(0, I)$. Additionally we know from [ALMT14, Lem C.1] that the minimum is unique and lies in the interval $\left[0, \frac{2\|\overline{x}\|_2}{b}\right]$ with $b \leq \|w\|_2$ for all $w \in \partial f(\overline{x})$, provided $\partial f(\overline{x})$ is bounded. This latter condition is only satisfied for f_4, but not for f_5 and f_6. Nevertheless, we successfully minimize numerically the univariate function J_{approx} above by using the BiSection [BF85] method. Note that for computing each $\text{dist}^2(X_i, \tau \partial f(\overline{x}))$ one is faced with solving a quadratic program. Note that $\partial f(\overline{x})$ is described by linear equalities and inequalities, compare (2.7a) and (2.7c). The empirical probability of uniqueness shown as gray value plots in Figure 4.4, middle row, is plotted along with the curves describing $\min_{\tau \geq 0} J_{\text{approx}}(\tau)$. We emphasize the perfect agreement of the phase transition.

4.2 2D Empirical Phase Transitions and Theoretical Bounds

4.2.1 Testset

Creating gradient sparse images with a given gradient sparsity is more involved than in the one dimensional case. Using random support sets $\Lambda \subseteq [p]$ with $|\Lambda| \geq n$ and the projecting technique used in the 1D case, one would most likely obtain constant 2D images. Thus we use a different approach to construct random gradient sparse images. To this end, we randomly add binary images with homogeneous areas and use the modulo operation to binarize again the result. We show some results in Figure 4.3. Since it is easy to identify all connected components in a binary image, we can assign random real values to the different connected components. The resulting
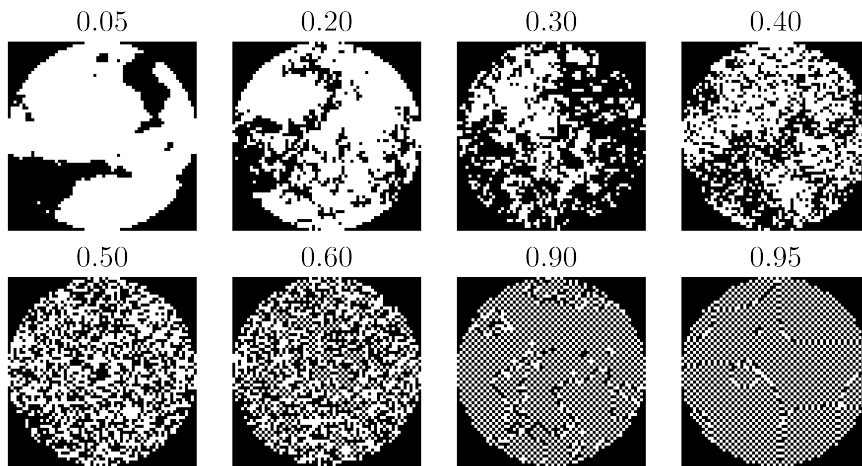
---

*https://www.mosek.com/
†https://www.wolfram.com/mathematica/

Figure 4.3: *Testset for 2D gradient sparse images*. Each image label shows the relative sparsity of $D\overline{x}$, i.e. $\|D\overline{x}\|_0/p$, with respect to the 2D finite difference operator. Our test images are 64×64.

discrete gradient of the new image has the same number of non-zeros, but the image is *non binary*.

4.2.2   Tomographic Measuremnts

Making our experiments more general, we applied the idea from [RLH14] to reduce the influence of angles in tomographic reconstructions. The basic idea is, that each projection should carry the same amount of information independent of the angle. In Figure 4.6 we illustrate the difference the effect on the tomographic projections when sensing an images embedded in a rectangle shape resp. to a circular shape. We emphasize that our testset for 2D contains only images embedded in a circular shape.

4.2.3 ℓ_1-Minimization

The phase diagrams for 2D images coincide with the ones for 1D signals, shown in Figure 4.4, first row, and are omitted here. The two regions are accurately separated by the $\min_\tau J(\tau)$ curve.
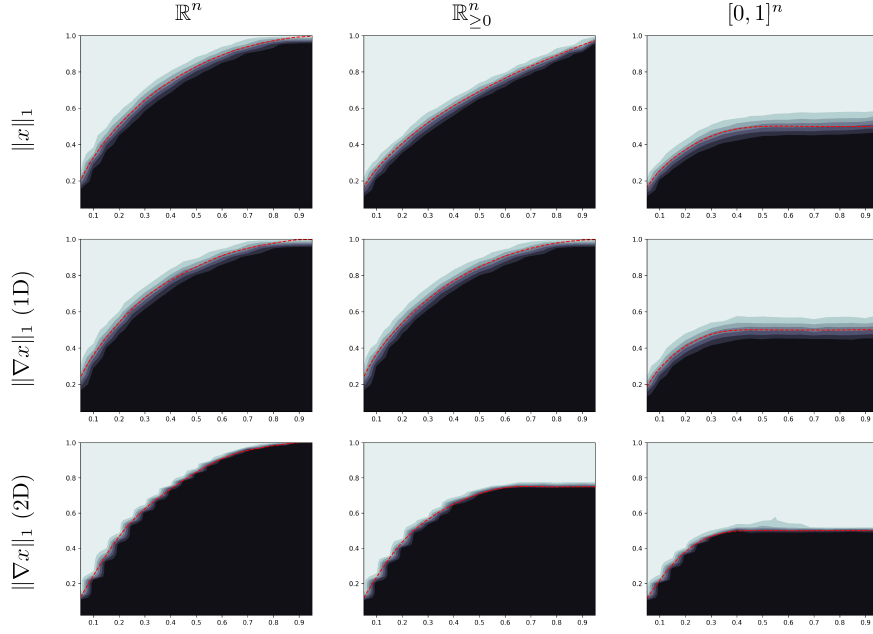
Figure 4.4: *Phase diagrams for random measurements.* The gray value shows the empirical probability of uniqueness for each pair of parameters (relative sparsity or relative gradient sparsity, #measurements): black $\leftrightarrow$ 0% uniqueness rate, white $\leftrightarrow$ 100% uniqueness rate. Both regions are accurately separated by our approximation to the statistical dimension. Rows from top to bottom show results for: ℓ_1-, 1D TV- and 2D TV-minimization. Columns for left to right show the signal/image entries: $\mathbb{R}^n$, $\mathbb{R}^n_+$ or $\{0,1\}^n$.
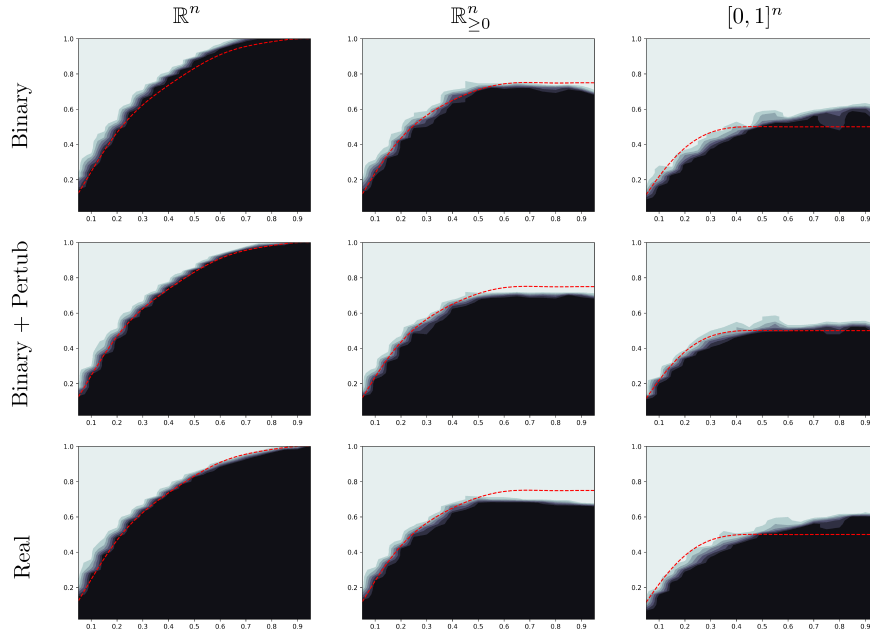
Figure 4.5: *Phase diagrams for tomographic measurements.* As in Figure 4.4 we display the empirical probability of uniqueness in the case of *tomographic measurements* for each pair of parameters (relative sparsity or relative gradient sparsity, #measurements): black $\leftrightarrow 0\%$ uniqueness rate, white $\leftrightarrow 100\%$ uniqueness rate. Rows from top to bottom show results for: binary, perturbed binary and standard matrices, i.e. with nonnegative real entries. The results demonstrate a remarkable agreement of the empirical phase transitions for tomographic recovery with the approximated curve based on the statistical dimension for random measurements.
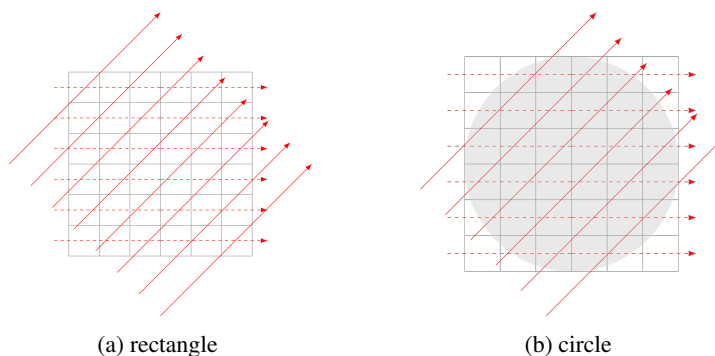
(a) rectangle   (b) circle

Figure 4.6: Illustration of parallel projections of a rectangle image (a) and of an image embedded in a circular mask (a). In (a) one needs at different angles a different amount of parallel rays to cover the whole image. For sensing an image in a circular mask the number of parallel rays covering the whole image would be the same.

4.2.4 TV-Minimization

For fixed ambient dimension $n = 64 \cdot 64$ we choose a relative image sparsity $\frac{\|D\bar{x}\|_0}{p} \in [0, 1]$, with $D \in \mathbb{R}^{p \times n}$, corresponding to a gradient sparse test image and also choose the number of measurements m. Measurements can be Gaussian or tomographic. We consider three types of tomographic matrices:

- binary, as described in [RLH14];

- perturbed binary, that is we slightly perturb the nonzero entries above, in order to remove linear dependencies between columns, and

- standard tomographic matrices, with nonnegative real entries. We use the MATLAB routine `paralleltomo.m` from the AIR Tools package [HSH12] that implements such a tomographic matrix for a arbitrary vector of angles. We choose equidistant angles, set `N = 64` the image size and use the default value of p, i.e. the number of parallel rays for each angle `p = round(sqrt(2)*N)` to obtain a tomographic matrix of size $m \times n$.

Results are presented in Figure 4.4, last row, and Figure 4.5. We show the empirical probability over the 10 repetitions. The success rate of image reconstruction *equals* the uniqueness result and is displayed by gray values: black $\leftrightarrow 0\%$ uniqueness rate, white $\leftrightarrow 100\%$ uniqueness rate.

All plots display a phase transition and thus exhibit regions where exact image reconstruction has probability equal or close to one. All regions are accurately separated by our approximation to the statistical dimension.

5 Conclusions

The present work was motivated by the need of accurately describing how much undersampling is tolerable to uniquely recover a sparse or gradient sparse box constrained signal from tomographic measurements. Our results show sharp average-case phase transitions from non-uniqueness and non-recovery to uniqueness and exact recovery as guaranteed by compressed sensing for random measurements. Moreover, we show that the phase transition occurs approximately at the statistical dimension of the descent cone of the structure enforcing objective, irrespectively of the employed measuring device. The approximation of the statistical dimension is computed empirically but cross-checked by the verifiable uniqueness test developed in this work.

Appendix

We collect few results from [Roc70, RW09]. The *subdifferential* of a function $f \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ at a point $\overline{x}$ where $f(\overline{x}) \in \mathbb{R}$, is defined by

$$\partial f(\overline{x}) = \big\{ p \in \mathbb{R}^n \colon f(x) - f(\overline{x}) \geq \langle p, x - \overline{x} \rangle, \quad \forall x \in \mathbb{R}^n \big\}. \tag{5.1}$$

We set $\partial f(\overline{x}) = \emptyset$ if $f(\overline{x})$ is not finite. A function $f \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be *subdifferentiable* at $\overline{x}$ if $\partial f(\overline{x}) \neq \emptyset$. Then the elements of $\partial f(\overline{x})$ are called *subgradients* of f at $\overline{x}$. For a proper, convex function f and $\overline{x} \in \operatorname{rint}(\operatorname{dom} f)$, the subdifferential $\partial f(\overline{x})$ is nonempty. Furthermore, $\partial f(\overline{x})$ is nonempty and bounded if and only if $\overline{x} \in \operatorname{int}(\operatorname{dom} f)$. Hence only for f_1 and f_4 from (1.3a) and (1.3f) $\partial f(\overline{x})$ is compact. For the remaining choices of f in (1.3) $\partial f(\overline{x})$ is unbounded, since in these cases our signals $\overline{x}$ lie on the boundary of $\operatorname{dom} f = \mathbb{R}_+^n$ or of $\operatorname{dom} f = \{0, 1\}^n$.

The *relative interior* of a non-empty convex set C is the interior relative to its affine hull,

$$\operatorname{rint} C = \big\{ x \in \operatorname{aff} C \colon \exists \varepsilon > 0 \text{ such that } (x + \varepsilon \mathbb{B}(0)) \cap \operatorname{aff} C \subset C \big\}.$$

The *relative boundary* of C is

$$\operatorname{rbd} C = \operatorname{cl} C \setminus \operatorname{rint} C.$$

We always have [Roc70, Cor. 6.6.2]

$$\operatorname{rint}(C_1 + C_2) = \operatorname{rint}(C_1) + \operatorname{rint}(C_2). \tag{5.2}$$

Recall that the *affine hull* of some set $C \subset \mathbb{R}^n$ is the set of all affine combinations of its points,

$$\operatorname{aff} C = \big\{ \lambda_1 x^1 + \cdots + \lambda_d x^d \colon \lambda_1 + \cdots + \lambda_d = 1, \ x^1, \dots, x^d \in C \big\},$$

while the *conic hull* equals

$$\operatorname{cone} C = \big\{ \lambda_1 x^1 + \cdots + \lambda_d x^d : \lambda_1, \ldots, \lambda_d \geq 0, \ x^1, \ldots, x^d \in C \big\}.$$

Acknowledgment

References

[ALMT14] D. Amelunxen, M. Lotz, M.B. McCoy, and J.A. Tropp, *Living on the Edge: Phase Transitions in Convex Programs with Random Data*, Inf. Inference **3** (2014), no. 3, 224–294.

[BF85] R.L. Burden and J.D. Faires, *Numerical Analysis (3rd ed.)*, PWS Publishers, 1985.

[CRPW12] V. Chandrasekaran, B. Recht, P.A. Parrilo, and A.S. Willsky, *The Convex Geometry of Linear Inverse Problems*, Found. Comput. Math. **12** (2012), no. 6, 805–849.

[DJM13] D.L. Donoho, I. Johnstone, and A. Montanari, *Accurate Prediction of Phase Transitions in Compressed Sensing via a Connection to Minimax Denoising*, IEEE Trans. Inf, Theory **59** (2013), no. 6, 3396–3433.

[FM14] R. Foygel and L. Mackey, *Corrupted Sensing: Novel Guarantees for Separating Structured Signals*, IEEE Trans. Inf. Theory **60** (2014), no. 2, 1223–1247.

[FR13] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser, 2013.

[Gil17] J.C. Gilbert, *On the Solution Uniqueness Characterization in the L1 Norm and Polyhedral Gauge Recovery*, J Optim. Theory Appl. **172** (2017), no. 1, 70–101.

[HK99] G.T. Herman and A. Kuba, *Discrete Tomography: Foundations, Algorithms, and Applications*, Applied and Numerical Harmonic Analysis, Birkhäuser, 1999.

[HSH12]   P.C. Hansen and M. Saxild-Hansen, *AIR Tools – A MATLAB Package of Algebraic Iterative Reconstruction Methods*, J. Comput. Appl. Math. **236** (2012), no. 8, 2167–2178.

[JKL15]   J.S. Jørgensen, C. Kruschel, and D.A Lorenz, *Testable Uniqueness Conditions for Empirical Assessment of Undersampling Levels in Total Variation-Regularized X-ray CT*, Inverse Probl. Sci. Eng. **23** (2015), no. 8, 1283–1305.

[NDEG13]   S. Nam, M.E. Davies, M. Elad, and R. Gribonval, *The Cosparse Analysis Model and Algorithms*, Appl. Comput. Harmon. Anal. **34** (2013), no. 1, 30–56.

[OH16]   S. Oymak and B. Hassibi, *Sharp MSE Bounds for Proximal Denoising*, Found. Comput. Math. **16** (2016), no. 4, 965–1029.

[PS14]   S. Petra and C. Schnörr, *Average Case Recovery Analysis of Tomographic Compressive Sensing*, Linear Algebra Appl. **441** (2014), 168–198.

[RLH14]   S. Roux, H. Leclerc, and F. Hild, *Efficient Binary Tomographic Reconstruction*, J. Math. Imaging Vision **49** (2014), no. 2, 335–351.

[Roc70]   R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[RW09]   R.T. Rockafellar and R.J.B. Wets, *Variational Analysis*, Springer, 2009.

[Van14]   R.J. Vanderbei, *Linear Programming: Foundations and Extensions*, Springer, 2014.

[ZMY16]   H Zhang, Y. Ming, and W. Yin, *One Condition for Solution Uniqueness and Robustness of both l1-Synthesis and l1-Analysis Minimizations*, Adv. Comput. Math. **42** (2016), no. 6, 1381–1399.

[ZXCL16]   B. Zhang, W. Xu, J.F. Cai, and L. Lai Lai, *Precise Phase Transition of Total Variation Minimization*, 2016 IEEE ICASSP, 2016, pp. 4518–4522.

Jan Kuske
Mathematical Imaging Group,
Heidelberg University, Germany.
Email: jan.kuske@iwr.uni-heidelberg.de

Stefania Petra
Mathematical Imaging Group,
Heidelberg University, Germany.
Email: petra@math.uni-heidelberg.de

106