



**HAL**  
open science

## Sampling Rates for 11-Synthesis

Maximilian März, Claire Boyer, Jonas Kahn, Pierre Weiss

► **To cite this version:**

Maximilian März, Claire Boyer, Jonas Kahn, Pierre Weiss. Sampling Rates for 11-Synthesis. 2020. hal-02540711v1

**HAL Id: hal-02540711**

**<https://hal.science/hal-02540711v1>**

Preprint submitted on 15 Apr 2020 (v1), last revised 7 Dec 2022 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sampling Rates for $\ell^1$ -Synthesis

Maximilian März<sup>§</sup>, Claire Boyer<sup>\*</sup>, Jonas Kahn<sup>†</sup>, Pierre Weiss<sup>‡</sup>

Affiliations: <sup>§</sup>Technische Universität Berlin; <sup>\*</sup>LPSM, Sorbonne Université, ENS Paris; <sup>†</sup>Université de Toulouse; <sup>‡</sup>ITAV, CNRS, Université de Toulouse

E-Mail of corresponding author: <sup>§</sup>maerz@math.tu-berlin.de

**Abstract.** This work investigates the problem of signal recovery from undersampled noisy sub-Gaussian measurements under the assumption of a synthesis-based sparsity model. Solving the  $\ell^1$ -synthesis basis pursuit allows for a simultaneous estimation of a coefficient representation as well as the sought-for signal. However, due to linear dependencies within redundant dictionary atoms it might be impossible to identify a specific representation vector, although the actual signal is still successfully recovered. The present manuscript studies both estimation problems from a non-uniform, signal-dependent perspective. By utilizing recent results on the convex geometry of linear inverse problems, the sampling rates describing the phase transitions of each formulation are identified. In both cases, they are given by the conic Gaussian mean width of an  $\ell^1$ -descent cone that is linearly transformed by the dictionary. In general, this expression does not allow a simple calculation by following the polarity-based approach commonly found in the literature. Hence, two upper bounds involving the sparsity of coefficient representations are provided: The first one is based on a local condition number and the second one on a geometric analysis that makes use of the thinness of high-dimensional polyhedral cones with not too many generators. It is furthermore revealed that both recovery problems can differ dramatically with respect to robustness to measurement noise – a fact that seems to have gone unnoticed in most of the related literature. All insights are carefully undermined by numerical simulations.

**Key words.** Compressed sensing, inverse problems, sparse representations, redundant dictionaries, non-uniform recovery, Gaussian mean width, circumangle.

## 1 Introduction

In the last two decades, the methodology of *compressive sensing* promoted the use of sparsity based methods for many signal processing tasks. Following the seminal works of Candès, Donoho, Romberg and Tao [CRT06a; CT06; Don06], a vast amount of research has extended the understanding, how additional structure can be exploited for solving ill-posed inverse problems. The classical setup in this area considers a *non-adaptive, linear measurement model*, which reads as follows:

### Model 1.1: Linear Noisy Measurements

Let  $x_0 \in \mathbb{R}^n$  be a fixed vector, which is typically referred to as the *signal*. Assume that we are given  $m$  *measurements*  $y \in \mathbb{R}^m$  of  $x_0$  via the linear acquisition model

$$y = Ax_0 + e,$$

where  $A \in \mathbb{R}^{m \times n}$  is the so-called *measurement matrix* and  $e \in \mathbb{R}^m$  models *measurement noise* with  $\|e\|_2 \leq \eta$  for some  $\eta \geq 0$ .

The goal of compressive sensing is to solve this inverse problem by reconstructing an approximation of the signal  $x_0$  from its indirect measurements  $y$ . Remarkably, even if  $m \ll n$ , this task can be achieved by incorporating additional information during the reconstruction process. Most classical compressive sensing works directly assume that  $x_0$  is *s-sparse*, i.e., that at most  $s \ll n$  entries of  $x_0$  are nonzero or in symbols  $\|x_0\|_0 = \#\text{supp}(x_0) \leq s$ . However, this assumption is hardly satisfied in any real-world application. Nevertheless, many signals allow for sparse representations using specific transforms, such as Gabor dictionaries, wavelet systems or data-adaptive representations, which are inferred from a given set of training samples. Such a model is referred to as *synthesis formulation*, since it assumes that there exists a matrix  $D \in \mathbb{R}^{n \times d}$  and a low-complexity representation  $z_0 \in \mathbb{R}^d$  such that  $x_0$  can be “synthesized” as

$$x_0 = D \cdot z_0. \tag{1.1}$$

Following the standard terminology of the field, the matrix  $D = [d_1, \dots, d_d]$  will be henceforth referred to as *dictionary* and its columns as *dictionary atoms*. It can be expected that the coefficient

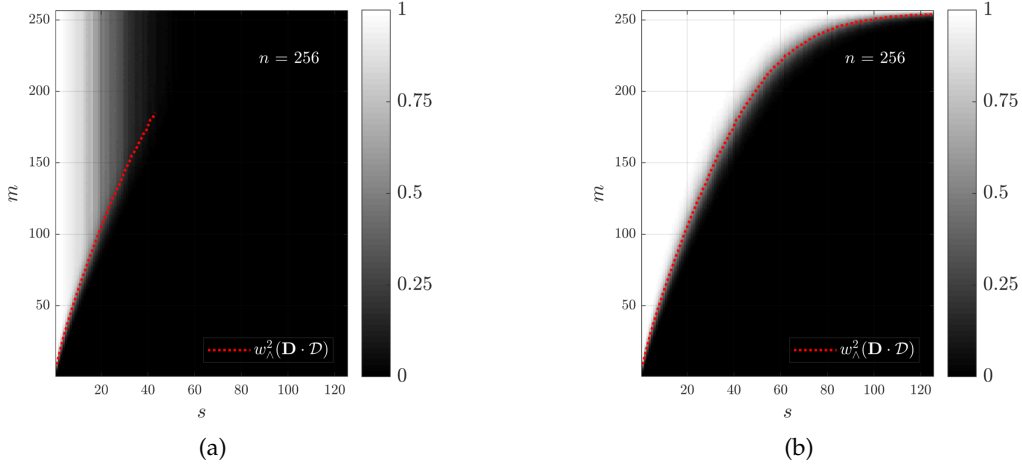


Figure 1: **Phase transitions of coefficient and signal recovery by  $\ell^1$ -synthesis.** Subfigure 1(a) shows the empirical probability that atomic coefficient representations are successfully recovered via solving  $(\text{BP}_{\eta=0}^{\text{coef}})$ , whereas Subfigure 1(b) shows the empirical probability for the associated signal reconstruction by  $(\text{BP}_{\eta=0}^{\text{sig}})$ . The underlying dictionary is a redundant Haar wavelet frame with three decomposition levels and the defining  $s$ -sparse coefficients are chosen at random; see Section 5.3 for a precise documentation of the experiment. The brightness of each pixel reflects the observed probability of success, reaching from certain failure (black) to certain success (white). The dotted line shows our predictions for the location of the phase transitions, see Theorem 3.6 and Theorem 3.8, respectively.

vector  $\mathbf{z}_0$  is dominated by just a few large entries, provided that  $\mathbf{D}$  allows to capture the signal's inherent structure reasonably well.

The synthesis formulation of compressive sensing exploits such a representation model, for instance, by employing greedy-based reconstruction algorithms or by utilizing the sparsity-promoting effect of the  $\ell^1$ -norm. In this work, we will consider the following convex program, which we refer to as *synthesis basis pursuit for coefficient recovery*:

$$\hat{\mathbf{Z}} := \underset{\mathbf{z} \in \mathbb{R}^d}{\text{argmin}} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{D}\mathbf{z}\|_2 \leq \eta. \quad (\text{BP}_{\eta}^{\text{coef}})$$

Under suitable assumptions, one might hope that solutions  $\hat{\mathbf{z}}$  of this minimization program approximate  $\mathbf{z}_0$  reasonably well. Indeed, if  $\mathbf{D} = \mathbf{Id}$ , the formulation  $(\text{BP}_{\eta}^{\text{coef}})$  turns into the classical basis pursuit. It allows to recover any  $s$ -sparse vector  $\mathbf{z}_0$  with overwhelming probability, if  $\mathbf{A}$  additionally follows a suitable random distribution and  $m \gtrsim s \cdot \log(2n/s)$  [FR13].

In many practical and theoretical situations, it turns out that using redundant dictionaries, i.e., choosing  $d \gg n$ , is beneficial. For instance, the stationary wavelet transform overcomes the lack of translation invariance and learned dictionaries typically infer a larger set of convolutional filters, which are adapted to a particular data distribution. If  $\mathbf{D}$  does not form a basis, representations as in (1.1) are not necessarily unique anymore. Hence, it is not to be expected that a specific representation can be identified by solving  $(\text{BP}_{\eta}^{\text{coef}})$ . However, in many situations of interest, the representation vector itself is irrelevant and a recovery of the actual signal  $\mathbf{x}_0$  is of primary interest. Thus, one rather cares about the *synthesis basis pursuit for signal recovery*, which amounts to solving

$$\hat{\mathbf{X}} := \mathbf{D} \cdot \left( \underset{\mathbf{z} \in \mathbb{R}^d}{\text{argmin}} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{D}\mathbf{z}\|_2 \leq \eta \right). \quad (\text{BP}_{\eta}^{\text{sig}})$$

In the noiseless case (i.e., when  $\mathbf{e} = \mathbf{0}$  and  $\eta = 0$ ), it might be the case that  $\hat{\mathbf{Z}} \neq \{\mathbf{z}_0\}$ , but there is still hope that  $\hat{\mathbf{X}} = \mathbf{D} \cdot \hat{\mathbf{Z}} = \{\mathbf{x}_0\}$ . In other words, although solving  $(\text{BP}_{\eta}^{\text{coef}})$  might fail in identifying a specific coefficient representation, it is still possible that the actual signal is successfully recovered by a subsequent synthesis with  $\mathbf{D}$ .

## 1.1 What This Paper Is About

The goal of this work is to broaden the understanding of the conditions that guarantee *coefficient* and *signal recovery* by solving  $(\text{BP}_\eta^{\text{coef}})$  and  $(\text{BP}_\eta^{\text{sig}})$ , respectively. To that end, we believe that addressing the following, non-exhaustive list of questions will be of particular importance:

- (Q1) Under which circumstances does coefficient and signal recovery differ, i.e., when is it impossible to reconstruct a specific coefficient representation although the signal itself might still be identified?
- (Q2) If possible, how many measurements are required to reconstruct a specific coefficient representation? Analogously, how many measurements are required to recover the associated signal?
- (Q3) In case that coefficients and signals can both be identified, are there still differences between the two formulations, for instance with respect to robustness to measurement noise?

Set out to find answers to these questions, we first restrict ourselves to the following sub-Gaussian measurement model, which will be considered in this work unless stated otherwise:

### Model 1.2: Sub-Gaussian Measurement Model

Let  $\mathbf{a} \in \mathbb{R}^n$  be an isotropic ( $\mathbb{E}[\mathbf{a}\mathbf{a}^T] = \mathbf{Id}$ ), zero mean, sub-Gaussian<sup>a</sup> random vector with  $\|\mathbf{a}\|_{\psi_2} \leq \gamma$ . The sampling matrix  $A$  is formed by drawing  $m$  independent copies  $\mathbf{a}_1, \dots, \mathbf{a}_m$  of  $\mathbf{a}$  and setting

$$A = \begin{bmatrix} -\mathbf{a}_1^T & - \\ \vdots & \\ -\mathbf{a}_m^T & - \end{bmatrix}.$$

<sup>a</sup>A random variable  $a$  is *sub-Gaussian* if  $\|a\|_{\psi_2} := \sup_{q \geq 1} q^{-1/2} (\mathbb{E}[|a|^q])^{1/q} < \infty$ , with  $\|\cdot\|_{\psi_2}$  being the *sub-Gaussian norm* of  $a$ . For a random vector  $\mathbf{a} \in \mathbb{R}^n$  the sub-Gaussian norm is then given by  $\|\mathbf{a}\|_{\psi_2} := \sup_{\mathbf{v} \in \mathcal{S}^{n-1}} \|\langle \mathbf{a}, \mathbf{v} \rangle\|_{\psi_2}$  and  $\mathbf{a}$  is called *sub-Gaussian* if  $\|\mathbf{a}\|_{\psi_2} < \infty$ ; see for instance [Ver12] for further details.

This model has been established as a somewhat classical benchmark setup in the context of compressive sensing. It allows us to follow the methodology initiated in [MPT07; RV07] and extended in [ALMT14; CRPW12; Sto09; Tro15]. In a nutshell, the aim is to determine the *sampling rate* of a convex program (i.e., the number of required measurements for successful recovery) by calculating the so-called *Gaussian mean width*.

We now briefly outline our work and summarize its main contributions:

- (C1) A cornerstone of our analysis is formed by the set of *minimal  $\ell^1$ -representers* of  $\mathbf{x}_0$ :

$$Z_{\ell^1} := \underset{\mathbf{z} \in \mathbb{R}^d}{\text{argmin}} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \mathbf{x}_0 = \mathbf{D}\mathbf{z}. \quad (\text{BP}_{\ell^1})$$

Independently of Model 1.2, Section 3.1 reveals that if  $Z_{\ell^1} = \{\mathbf{z}_0\}$ , exact recovery of  $\mathbf{z}_0$  via  $(\text{BP}_{\eta=0}^{\text{coef}})$  is equivalent to perfect recovery of  $\mathbf{x}_0$  by solving  $(\text{BP}_{\eta=0}^{\text{sig}})$ . Furthermore, exact recovery of a coefficient vector  $\mathbf{z}_0$  by  $(\text{BP}_{\eta=0}^{\text{coef}})$  is only possible, if  $\mathbf{z}_0$  is the unique minimal  $\ell^1$ -representer of  $\mathbf{x}_0 = \mathbf{D}\mathbf{z}_0$ , i.e., if  $Z_{\ell^1} = \{\mathbf{z}_0\}$ .

- (C2) In Section 3.2 and Section 3.3, it will be shown that the sampling rate of both formulations can be expressed by the squared conic mean width  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D})$ , where  $\mathcal{D}$  denotes the descent cone of the  $\ell^1$ -norm at any  $\mathbf{z}_{\ell^1} \in Z_{\ell^1}$  (see Section 2 for a brief summary of the general recovery framework and definitions of these notions). This observation holds unconditionally true in the case of signal recovery by  $(\text{BP}_\eta^{\text{sig}})$ . For coefficient recovery the additional assumption that  $Z_{\ell^1}$  is a singleton needs to be satisfied.
- (C2') While  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D})$  forms a precise description of the sampling rate, it is a quantity that is hard to analyze and compute, in general. Therefore, an important goal of our work is to

derive more informative upper bounds for this expression. First, under the assumption that  $Z_{\ell^1} = \{z_{\ell^1}\}$  is a singleton, we show a condition number bound that relates  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D})$  to the classical complexity  $w_\lambda^2(\mathcal{D}) \lesssim s \cdot \log(2n/s)$ , where  $s = \|z_{\ell^1}\|_0$  (see Section 4.1).

The second upper bound of Section 4.2 is central to our work. It is based on a geometric analysis that makes use of generic arguments from high-dimensional convex geometry. In comparison to the first bound, it is more general since we do not assume that  $Z_{\ell^1}$  is a singleton. Hence, it particularly addresses the recovery of signals, without requiring the identification of a coefficient representation. The resulting upper bound on the conic mean width relies on the thinness of high-dimensional polyhedral cones with not exponentially many generators. We believe that such an argument might be of general interest beyond its application to the synthesis formulation of compressive sensing. Again,  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D})$  is related to the sparsity of a minimal  $\ell^1$ -representation and a further geometrical parameter (referred to as *circumangle*) that measures the narrowness of the associated cone.

- (C3) Lastly, our recovery statements reveal that recovery of signals by  $(\text{BP}_\eta^{\text{sig}})$  is robust to measurement noise without any further restrictions. In contrast, the robustness of coefficient recovery via solving  $(\text{BP}_\eta^{\text{coef}})$  is influenced by an additional factor that is related to the convex program  $(\text{BP}_{\ell^1})$ .

All our findings are underpinned by extensive numerical experiments; see Section 5. As a first “teaser” we refer the reader to Figure 1, which displays two *phase transition plots* and our sampling rates for a redundant Haar wavelet system  $\mathbf{D}$ .

## 1.2 Related Literature

In the following, we first briefly discuss some historical references that are of general interest for  $\ell^1$ -norm minimization, sparse representations in redundant dictionaries and compressive sensing. Subsequently, we focus on the existing literature on the synthesis formulation in more depth.

### 1.2.1 Some Historical Landmarks

The idea of promoting sparsity in discrete or continuous dictionaries by  $\ell^1$ -norm minimization can be traced back to the works of Beurling [Beu38] and Krein [Kre38]. Motivated by questions in Banach space geometry, first theorems establishing sparsity of solutions of related minimization problems can be found in the 1940’s [Zuh48]. In his PhD-thesis of 1965, Logan utilized  $\ell^1$ -minimization for sparse frequency estimation [Log65] and in the 1970’s it was employed for solving deconvolution problems in geophysics [CM73; TBM79]. Of particular importance became the so-called *Rudin-Osher-Fatemi-model* [ROF92], which pioneered the use of total variation minimization for image processing tasks.

The field of sparse representations arose with the development of (greedy) algorithms for finding expansions in redundant dictionaries such as time-frequency systems [MZ93; PRK93]. Subsequently, the work [CDS98] triggered notable interest in achieving this task by solving the basis pursuit  $(\text{BP}_{\ell^1})$ ; see for instance [DE03; DH01; EB02]. A special emphasis was thereby given to unions of orthogonal bases [CR06; EB02; GN03]. Next to the classical concepts of coherence and spark, which are *uniform* across all  $s$ -sparse signals, also the *non-uniform* notions of dual certificates and exact recovery conditions were progressively introduced [Fuc04; Fuc05; Tro04; Tro06].

Under the notion of *compressive sensing*, Candès, Romberg and Tao [CRT06a; CRT06b] and Donoho [Don06] first proposed to capitalize on randomized models in the basis pursuit. In these works, the structured dictionary  $\mathbf{D}$  is replaced by a random matrix  $\mathbf{A}$ , which follows for instance Model 1.2. Such a design allows to overcome severe shortcomings of previous results, in particular the *quadratic/square root-bottleneck*; see next subsection or [FR13, Chapter 5.4]. Indeed, under such a randomness assumption, it can be shown that any  $s$ -sparse vector can be recovered with overwhelming probability if the number of measurements obeys  $m \gtrsim s \cdot \log(2n/s)$ . These seminal works can furthermore be acknowledged for highlighting the remarkable potential of sparsity-based methods for many signal recovery tasks.

### 1.2.2 Results on the Synthesis Formulation of Compressed Sensing

An important insight on solving the inverse problem of Model 1.1 by means of redundant dictionaries was provided by Elad, Milanfar and Rubinstein [EMR07]. Therein, the authors compare two different formulations: The synthesis basis pursuit ( $\text{BP}_\eta^{\text{sig}}$ ) and an alternative formulation, which is referred to as  $\ell^1$ -analysis basis pursuit:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\Psi \mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A} \mathbf{x}\|_2 \leq \eta.$$

The analysis operator  $\Psi \in \mathbb{R}^{d \times n}$  is thereby chosen in such a way that the coefficient vector  $\Psi \mathbf{x}_0$  is of low-complexity. It turns out that the latter formulation and the program ( $\text{BP}_\eta^{\text{sig}}$ ) are only equivalent if  $\Psi$  (or  $D$ ) forms a basis. In particular for redundant choices of  $\Psi$  and  $D$ , the geometry of both formulations departs significantly from each other. While the synthesis variant appears to be more natural from a historical perspective, its analysis-based counterpart gained considerable attention in the past years [CENR11; GNEGD14; KNW15; KR15; KRZ15; NDEG13]. Recently, the non-uniform approach of [GKM20] revealed that the measure of “low-complexity” in the analysis model goes beyond pure sparsity of  $\Psi \mathbf{x}_0$ . Instead, a novel sampling-rate bound was proposed that is based on a generalized notion of sparsity, taking the support and the coherence structure of the underlying analysis operator into account.

The earliest reference that deals with the synthesis formulation for the recovery of coefficient vectors appears to be by Rauhut, Schnass and Vandergheynst [RSV08]. Therein, the formulation ( $\text{BP}_\eta^{\text{coef}}$ ) is studied under a randomized measurement model. The main result roughly reads as follows: Assume that the dictionary  $D$  satisfies a *restricted isometry property* (RIP) with sparsity level  $s$ . If the random matrix  $A \in \mathbb{R}^{m \times n}$  follows Model 1.2 and  $m \gtrsim s \cdot \log(n/s)$ , then the composition  $AD$  will also satisfy an RIP with sparsity level  $s$  with high probability. This property then implies stable and robust recovery of all  $s$ -sparse coefficient vectors by solving ( $\text{BP}_\eta^{\text{coef}}$ ). The assumption that  $D$  satisfies an RIP is crucial for the previous result. It can be for instance achieved if the dictionary is sufficiently incoherent, i.e., if it satisfies

$$\mu(D) := \max_{i \neq j} |\langle \mathbf{d}_i, \mathbf{d}_j \rangle| / (\|\mathbf{d}_i\|_2 \cdot \|\mathbf{d}_j\|_2) \leq 1 / (16 \cdot (s - 1)). \quad (1.2)$$

However, as the authors of [RSV08] point out, such a coherence-based estimate is rather crude and suffers from the so-called square-root bottleneck: The Welch bound [FR13, Theorem 5.7] reveals that condition (1.2) can only be satisfied for mild sparsity values  $s \lesssim \sqrt{n}$ .

In [CWW14], Chen, Wang and Wang study conditions for signal recovery via a dictionary-based *null space property* (NSP): For a given dictionary  $D$ , a matrix  $A$  is said to satisfy the  $D$ -NSP of order  $s$ , if for any index set  $S \subseteq [d]$  with  $\#S \leq s$  and any  $\mathbf{h} \in D^{-1}(\ker A \setminus \{\mathbf{0}\})$ , there exists  $\mathbf{z} \in \ker D$ , such that  $\|\mathbf{h}_S + \mathbf{z}\|_1 < \|\mathbf{h}_{S^c}\|_1$ . It can be shown that this condition is necessary and sufficient for the uniform recovery of all signals  $\mathbf{x}_0 = D\mathbf{z}_0$  with  $\|\mathbf{z}_0\|_0 \leq s$  via ( $\text{BP}_\eta^{\text{sig}}$ ). Note that the  $D$ -NSP is in general weaker than requiring that  $AD$  satisfies the standard NSP. This means that the previous result is addressing signal recovery without necessarily requiring coefficient recovery. However, the authors then show that under the additional assumption that  $D$  is of *full spark* (i.e., every  $n$  columns of  $D$  are linearly independent), both conditions are in fact equivalent. Hence, in this case, signal and coefficient recovery are also equivalent. In the recent work [CCL19], this serves as a motivation to study coefficient recovery by analyzing how many measurements are required in order to guarantee that  $AD$  has an NSP. To that end, a result is provided that is conceptually similar to [RSV08], however, it reduces the assumptions on  $D$ . Instead of requiring that  $D$  satisfies an RIP, the authors operate under the weaker assumption that  $D$  satisfies an NSP. The main result essentially reads as follows: Under a sub-Gaussian measurement setup similar to Model 1.2 and under the assumption that  $D$  satisfies an NSP of order  $s$ , a number of  $m \gtrsim s \cdot \log(n/s)$  measurements guarantees that also  $AD$  satisfies an NSP. This condition then allows for robust recovery of all  $s$ -sparse coefficient vectors by solving ( $\text{BP}_\eta^{\text{coef}}$ ).

To the best of our knowledge, the only work that provides a bound on the required number of measurement for signal recovery (without necessarily requiring coefficient recovery) is the tutorial [Ver15, Theorem 7.1]: Assume that  $\|\mathbf{d}_i\|_2 \leq 1$ ,  $i \in [d]$  and that  $\mathbf{x}_0 = D\mathbf{z}_0$  for an  $s$ -sparse representation  $\mathbf{z}_0 \in \mathbb{R}^d$ . For a Gaussian measurement matrix  $A \in \mathbb{R}^{m \times n}$ , Vershynin establishes



the following recovery bound in expectation:

$$\mathbb{E} \|\hat{\mathbf{x}} - \mathbf{x}_0\|_2 \leq c \cdot \sqrt{\frac{s \cdot \log(d)}{m}} \cdot \|\mathbf{z}_0\|_2 + \sqrt{2\pi} \cdot \frac{\eta}{\sqrt{m}},$$

where  $c$  is a constant and  $\hat{\mathbf{x}} \in \hat{X}$  is a solution of  $(\text{BP}_{\eta}^{\text{sig}})$ . Note that we have slightly adapted the statement of [Ver15, Theorem 7.1] for a better match with our setup. Due to the first summand on the right hand side, the previous error bound is suboptimal, cf. Theorem 3.8. In particular, it does not guarantee exact recovery from noiseless measurements. We emphasize that parts of our work are inspired by Vershynin, who also studies the gauge of the set  $K = \mathbf{D} \cdot \mathbf{B}_1^d$  in [Ver15].

We conclude by mentioning a few more works in the literature on synthesis based compressed sensing that appear to be of less relevance for this work. The influential paper [CRPW12] studies signal recovery via atomic minimization, however, it does not provide specific insights when redundant dictionaries are used. In [DNW13], a (theoretical) CoSaMP algorithm is adapted to the recovery of signals with sparse representations in redundant dictionaries. Based on the  $\mathbf{D}$ -RIP [CENR11] and on a connection to the analysis formulation with so-called optimal dual frames, [LLMLY12] derives a theorem concerning signal recovery. Finally, [SF09] provides numerical experiments, which empirically compare the analysis and the synthesis formulation.

### 1.2.3 The Gap that We Intend to Fill

In order to obtain statements that are *uniform* across all  $s$ -sparse signals, most existing results assume that the dictionary  $\mathbf{D}$  satisfies strong assumptions, e.g., incoherent atoms, an NSP or an RIP [CCL19; CWW14; RSV08]. Such notions are well established and allow for appealing results that often resemble known principles of compressive sensing. However, in many situations of interest, these assumptions are too restrictive. In particular, redundant representation systems (such as Gabor systems, wavelets, curvelets, ...) or data-adaptive dictionaries do not satisfy any such property. Their atoms are typically highly coherent and share many linear dependencies. We aim to address this issue by following a local, non-uniform approach, which avoids strong assumptions on the dictionary. We believe that such a signal-dependent refinement is crucial for redundant representation systems, cf. [GKM20].

Similarly, it is occasionally argued that distinguishing signal and coefficient recovery is of minor importance, cf. [CCL19]. This is justified by the observation that exact recovery of coefficients and signals is equivalent if  $\mathbf{D}$  is in general position. However, due to the linear dependencies in many structured representation systems, such an argumentation is often not valid. Indeed, simple numerical experiments with popular dictionaries reveal that signal recovery can be frequently observed without reconstructing a specific coefficient representation, see Figure 1 and Section 5. Hence, we believe that it is important to study both formulations and to identify under which conditions coefficient recovery might be expected, see (C1) above.

To the best of our knowledge, this is the first work that provides a precise description of the phase transition behavior of both formulations, see (C2). While the identified conic mean width of a linearly transformed set  $w_{\lambda}^2(\mathbf{D} \cdot \mathcal{D})$  is a rather implicit quantity, it constitutes an important step towards the understanding of  $\ell^1$ -synthesis. By deriving more explicit upper bounds on the sampling rate, coefficient sparsity is identified as an important factor. However, additional properties that account for the local geometry are also taken into account, see (C2').

Last but not least, we establish that both formulations behave differently with respect to robustness to measurement noise, see (C3). To the best of our knowledge, this aspect has gone unnoticed in the literature so far, although it might have dramatic implications on the reconstruction quality of coefficient representations.

## 1.3 Notation

For the convenience of the reader, we have collected the most important and frequently used objects in Table 1.

Throughout this manuscript we will use the following notation and conventions: for an integer  $n \in \mathbb{N}$  we set  $[n] := \{1, 2, \dots, n\}$ . If  $\mathcal{I} \subseteq [n]$ , we let  $\mathcal{I}^c := [n] \setminus \mathcal{I}$  denote the complement of  $\mathcal{I}$  in  $[n]$ . Vectors and matrices are symbolized by lower- and uppercase bold letters, respectively. Let  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ . For an index set  $\mathcal{I} \subseteq [n]$ , we let the vector  $\mathbf{x}_{\mathcal{I}} \in \mathbb{R}^{\#\mathcal{I}}$  denote the restriction

Notation	Term
$\mathbf{x}_0 \in \mathbb{R}^n$	(ground truth) signal vector
$\mathbf{A} \in \mathbb{R}^{m \times n}$	measurement matrix
$\mathbf{e} \in \mathbb{R}^m$ , with $\ \mathbf{e}\ _2 \leq \eta$	(adversarial) noise
$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{e} \in \mathbb{R}^m$	linear, noisy measurements of $\mathbf{x}_0$
$\mathbf{d}_1, \dots, \mathbf{d}_d \in \mathbb{R}^n$	dictionary atoms
$\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_d] \in \mathbb{R}^{n \times d}$	dictionary
$\hat{\mathbf{x}} \in \mathbb{R}^n$	a solution of $(\text{BP}_\eta^{\text{sig}})$
$\hat{X} = \mathbf{D} \cdot \hat{Z} \subseteq \mathbb{R}^n$	solution set of $(\text{BP}_\eta^{\text{sig}})$
$z_{\ell^1} \in \mathbb{R}^d$	a minimal $\ell^1$ -decomposition of $\mathbf{x}_0$ in $\mathbf{D}$ , i.e., a solution of $(\text{BP}_{\ell^1})$
$Z_{\ell^1} \subseteq \mathbb{R}^d$	solution set of $(\text{BP}_{\ell^1})$
$\mathbf{z}_0 \in \mathbb{R}^d$	a sparse representation of $\mathbf{x}_0$ in $\mathbf{D}$ , without $\mathbf{z}_0 \in Z_{\ell^1}$ , in general
$\hat{\mathbf{z}} \in \mathbb{R}^d$	a solution of $(\text{BP}_\eta^{\text{coef}})$
$\hat{Z} \subseteq \mathbb{R}^d$	solution set of $(\text{BP}_\eta^{\text{coef}})$

Table 1: A summary of the central notations used in this work.

to the components indexed by  $\mathcal{I}$ . The *support* of  $\mathbf{x}$  is defined by the set of its non-zero entries  $\text{supp}(\mathbf{x}) := \{k \in [n] \mid x_k \neq 0\}$  and the *sparsity* of  $\mathbf{x}$  is  $\|\mathbf{x}\|_0 := \#\text{supp}(\mathbf{x})$ . For  $1 \leq p \leq \infty$ ,  $\|\cdot\|_p$  denotes the  $\ell^p$ -norm on  $\mathbb{R}^n$ . The associated *unit ball* is given by  $B_p^n := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_p \leq 1\}$  and the *Euclidean unit sphere* is  $S^{n-1} := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_2 = 1\}$ . The  $i$ -th standard basis vector of  $\mathbb{R}^n$  is referred to as  $\mathbf{e}_i$  and  $\mathbf{Id} \in \mathbb{R}^{n \times n}$  denotes the identity matrix. Furthermore, let  $\text{cone}(K)$  denote the *conic hull* of a set  $K \subset \mathbb{R}^n$ . If  $L \subset \mathbb{R}^n$  is a linear subspace, the associated *orthogonal projection onto  $L$*  is given by  $\mathbf{P}_L \in \mathbb{R}^{n \times n}$ . Then, we have  $\mathbf{P}_{L^\perp} = \mathbf{Id} - \mathbf{P}_L$ , where  $L^\perp \subset \mathbb{R}^n$  is the orthogonal complement of  $L$ . The letter  $c$  is usually reserved for a (generic) constant, whose value could change from time to time. We refer to  $c$  as a *numerical constant* if its value does not depend on any other involved parameter. If an (in-)equality holds true up to a numerical constant  $c$ , we sometimes write  $a \lesssim b$  instead of  $a \leq c \cdot b$ . For a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  we let  $\|\mathbf{A}\|_2$  denote its *spectral norm*. For a set  $K \subseteq \mathbb{R}^n$ ,  $\lambda \in \mathbb{R}$  and  $\mathbf{A} \in \mathbb{R}^{m \times n}$  we set  $\lambda \cdot K := \{\lambda \mathbf{k} \mid \mathbf{k} \in K\}$  and  $\mathbf{A} \cdot K := \{\mathbf{A}\mathbf{k} \mid \mathbf{k} \in K\}$ . Lastly, the term *orthonormal basis* is abbreviated by ONB.

## 2 A Primer on the Convex Geometry of Linear Inverse Problems

In this section, we give a brief introduction to a well-established methodology that addresses the recovery of structured signals from independent linear random measurements. This summary mainly serves the purpose of introducing the required technical notions for our subsequent analysis of the  $\ell^1$ -synthesis formulation. It is inspired by [ALMT14; CRPW12; Tro15] and we refer the interested reader to these works for a more detailed discussion of the presented material.

### 2.1 Minimum Conic Singular Value

Assume that Model 1.1 is satisfied. For a robust recovery of  $\mathbf{x}_0$  from its linear, noisy measurements  $\mathbf{y}$ , we consider the *generalized basis pursuit*

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \eta, \quad (\text{BP}_\eta^f)$$

where  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is a convex function that is supposed to reflect the “low complexity” of the signal  $\mathbf{x}_0$ . Hence, the previous minimization problem searches for the most structured signal that is still consistent with the given measurements  $\mathbf{y}$ .

The recovery performance of  $(\text{BP}_\eta^f)$  can be understood by a fairly standard geometric analysis. It seeks to understand the geometric interplay of the structure-promoting functional  $f$  and the



measurement matrix  $A$  in a neighborhood of the signal vector  $\mathbf{x}_0$ . To that end, we first introduce the following notions of descent cones and minimum conic singular values.

**Definition 2.1** (Descent cone) Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be a convex function and let  $\mathbf{x}_0 \in \mathbb{R}^n$ . The *descent set* of  $f$  at  $\mathbf{x}_0$  is given by

$$\mathcal{D}(f, \mathbf{x}_0) := \{\mathbf{h} \in \mathbb{R}^n : f(\mathbf{x}_0 + \mathbf{h}) \leq f(\mathbf{x}_0)\},$$

and the corresponding *descent cone* is defined by  $\mathcal{D}_\wedge(f, \mathbf{x}_0) := \text{cone}(\mathcal{D}(f, \mathbf{x}_0))$ .

The notion of minimum conic singular values describes the behavior of a matrix  $A$  when it is restricted to a cone  $C \subseteq \mathbb{R}^n$ .

**Definition 2.2** (Minimum conic singular value) Consider a matrix  $A \in \mathbb{R}^{m \times n}$  and a cone  $C \subseteq \mathbb{R}^n$ . The minimum conic singular value of  $A$  with respect to the cone  $C$  is defined by

$$\lambda_{\min}(A; C) := \inf_{\mathbf{x} \in C \cap \mathcal{S}^{n-1}} \|A\mathbf{x}\|_2.$$

The following result characterizes exact recoverability of the signal  $\mathbf{x}_0$  and provides a deterministic error bound for the solutions to  $(\text{BP}_\eta^f)$ . The statement is an adapted version of Proposition 2.1 and Proposition 2.2 in [CRPW12]; see also Proposition 2.6 in [Tro15].

**Proposition 2.3** (A deterministic error bound for  $(\text{BP}_\eta^f)$ ) Assume that  $\mathbf{x}_0, A, \mathbf{y}, \mathbf{e}$  and  $\eta$  follow Model 1.1 and let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be a convex function. Then the following holds true:

- (a) If  $\eta = 0$ , exact recovery of  $\mathbf{x}_0$  by solving  $(\text{BP}_{\eta=0}^f)$  is equivalent to  $\lambda_{\min}(A; \mathcal{D}_\wedge(f, \mathbf{x}_0)) > 0$ .
- (b) In addition, any solution  $\hat{\mathbf{x}}$  of  $(\text{BP}_\eta^f)$  satisfies

$$\|\mathbf{x}_0 - \hat{\mathbf{x}}\|_2 \leq \frac{2\eta}{\lambda_{\min}(A; \mathcal{D}_\wedge(f, \mathbf{x}_0))}. \quad (2.1)$$

## 2.2 Conic Mean Width

While Proposition 2.3 provides an elegant analysis of the solutions to the optimization problem  $(\text{BP}_\eta^f)$ , it can be difficult to apply. The notion of a minimum conic singular value is related to the concept of co-positivity [HS10c] and its computation is known to be an NP-hard task for general matrices and cones [HS10c; MK87].

However, when  $A$  is chosen at random, sharp estimates can be obtained by exploring a connection to the *statistical dimension* or *Gaussian mean width*. These geometric parameters stem from geometric functional analysis and convex geometry (e.g., see [GM04; Gor85; Gor88; Mil85]), but they also show up in Talagrand's  $\gamma_2$ -functional in stochastic processes [Tal14], or under the name of *Gaussian complexity* in statistical learning theory [BM02]. Their benefits for compressive sensing have first been exploited in [MPT07; RV07]. More important for our work is their use in the more recent line of research [ALMT14; CRPW12; Sto09; Tro15], which aims for non-uniform signal recovery statements.

**Definition 2.4** Let  $K \subseteq \mathbb{R}^n$  be a set.

- (a) The *(global) mean width* of  $K$  is defined as

$$w(K) := \mathbb{E} \left[ \sup_{\mathbf{h} \in K} \langle \mathbf{g}, \mathbf{h} \rangle \right],$$

where  $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{Id})$  is a standard Gaussian random vector.

- (b) The *conic mean width* of  $K$  is given by

$$w_\wedge(K) := w(\text{cone}(K) \cap \mathcal{S}^{n-1}).$$

We refer to  $w_\wedge(\mathcal{D}(f, \mathbf{x}_0))$  as the *conic mean width* of  $f$  at  $\mathbf{x}_0$ .

The next theorem is known as *Gordon's Escape Through a Mesh* and dates back to [Gor88]. The version presented here follows from [LMPV17].

**Theorem 2.5** (Theorem 3 in [LMPV17]) *Assume that  $\mathbf{A}$  satisfies the assumption in Model 1.2 and let  $K \subseteq \mathcal{S}^{n-1}$  be a set. Then there exists a numerical constant  $c > 0$  such that, for every  $u > 0$ , we have*

$$\inf_{\mathbf{x} \in K} \|\mathbf{A}\mathbf{x}\|_2 > \sqrt{m-1} - c \cdot \gamma^2 \cdot (w(K) + u), \quad (2.2)$$

with probability at least  $1 - e^{-u^2/2}$ . If  $\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \mathbf{Id})$ , we have  $c = \gamma = 1$ .

Thus, a straightforward combination the error bound in (2.1) and the estimate in (2.2) for the set  $K = \mathcal{D}_\wedge(f, \mathbf{x}_0) \cap \mathcal{S}^{n-1}$  reveals that robust recovery via  $(\text{BP}_\eta^f)$  is possible if the number of sub-Gaussian measurements obeys

$$m \geq c^2 \cdot \gamma^4 \cdot w_\wedge^2(\mathcal{D}(f, \mathbf{x}_0)) + 1.$$

In the case of Gaussian measurements, it is known that this bound yields a tight description of the so-called *phase transition* of  $(\text{BP}_{\eta=0}^f)$ . Indeed, for a convex cone  $C \subseteq \mathbb{R}^n$  it can be shown that  $\lambda_{\min}(\mathbf{A}; C) = 0$  with high probability when  $m \leq w_\wedge^2(C) - c \cdot w_\wedge(C)$ , where  $c > 0$  denotes a numerical constant. Applying this statement to the descent cone  $\mathcal{D}_\wedge(f, \mathbf{x}_0)$  reveals that exact recovery of  $\mathbf{x}_0$  by solving  $(\text{BP}_{\eta=0}^f)$  fails with high probability when

$$m \leq w_\wedge^2(\mathcal{D}(f, \mathbf{x}_0)) - c \cdot w_\wedge(\mathcal{D}(f, \mathbf{x}_0)).$$

Hence, exact signal recovery by solving  $(\text{BP}_{\eta=0}^f)$  obeys a sharp phase transition at  $m \approx w_\wedge^2(\mathcal{D}(f, \mathbf{x}_0))$  Gaussian measurements. We refer to [ALMT14] and [Tro15, Remark 3.4] for more details on this matter and conclude our discussion by the following summary:

*Robust signal recovery via the generalized basis pursuit  $(\text{BP}_\eta^f)$  is characterized by the minimum conic singular value  $\lambda_{\min}(\mathbf{A}; \mathcal{D}_\wedge(f, \mathbf{x}_0))$ . The required number of sub-Gaussian random measurements can be determined by the conic mean width of  $f$  at  $\mathbf{x}_0$ , in symbols  $w_\wedge(\mathcal{D}(f, \mathbf{x}_0))$ .*

### 3 Coefficient and Signal Recovery

Our study of the synthesis formulation in this section is based on the differentiation between coefficient and signal recovery. First, we introduce the set of minimal  $\ell^1$ -representers in Section 3.1 and discuss its importance for the relationship between both formulations. Section 3.2 is then dedicated to the fact that signal recovery via  $(\text{BP}_\eta^{\text{sig}})$  can be cast as an instance of atomic norm minimization, in which the gauge of the synthesis defining polytope is minimized. Finally, in Section 3.3, we derive two non-uniform recovery theorems that determine the sampling rates of robust coefficient and signal recovery, respectively.

#### 3.1 Recovery and Minimal $\ell^1$ -Representers

In this section, we discuss how the uniqueness of a minimal  $\ell^1$ -representer impacts coefficient and signal recovery.

**Definition 3.1** (Minimal  $\ell^1$ -representers) The set of *minimal  $\ell^1$ -representers* of a signal  $\mathbf{x}_0$  with respect to a dictionary  $\mathbf{D}$  is defined by

$$Z_{\ell^1} := \underset{\mathbf{z} \in \mathbb{R}^d}{\text{argmin}} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \mathbf{x}_0 = \mathbf{D}\mathbf{z}. \quad (\text{BP}_{\ell^1})$$

In general,  $Z_{\ell^1}$  may not be a singleton. Indeed, a coefficient vector  $\mathbf{z}_{\ell^1}$  can only be the unique minimal  $\ell^1$ -representer of the associated signal  $\mathbf{x}_0 = \mathbf{D}\mathbf{z}_{\ell^1}$ , if the set of atoms  $\{\mathbf{d}_i : i \in \text{supp}(\mathbf{z}_{\ell^1})\}$

is linearly independent [FR13, Theorem 3.1]. However, many dictionaries of practical interest possess linear dependent and coherent atoms. Hence, typical notions that would certify uniqueness for all signals with sparse representations in  $\mathbf{D}$  (e.g., the NSP [FR13, Theorem 4.5]) are not expected to hold for such dictionaries.

The following simple lemma shows that exact coefficient recovery by solving  $(\text{BP}_{\eta=0}^{\text{coef}})$  requires  $Z_{\ell^1}$  to be a singleton. Otherwise, it is impossible to recover a specific coefficient representation, while a retrieval of the signal by  $(\text{BP}_{\eta=0}^{\text{sig}})$  might still be possible.

**Lemma 3.2** *Assume that  $\mathbf{x}_0, \mathbf{A}$  and  $\mathbf{y}$  follow Model 1.1 with  $\eta = 0$ . Let  $\mathbf{D} \in \mathbb{R}^{n \times d}$  be a dictionary such that  $\mathbf{x}_0 \in \text{ran}(\mathbf{D})$ .*

- (a) *Assume that  $\mathbf{x}_0 = \mathbf{D}\mathbf{z}_0$  and that we wish to reconstruct  $\mathbf{z}_0$ . If  $Z_{\ell^1} \neq \{\mathbf{z}_0\}$ , then recovering  $\mathbf{z}_0$  by solving  $(\text{BP}_{\eta=0}^{\text{coef}})$  is impossible.*
- (b) *Signal recovery by solving  $(\text{BP}_{\eta=0}^{\text{sig}})$ , i.e., having  $\hat{X} = \{\mathbf{x}_0\}$ , is equivalent to the condition  $Z_{\ell^1} = \hat{Z}$ .*

A short proof of the previous result is given in Appendix A.1. Thus, under the assumption that  $\mathbf{x}_0$  has a unique minimal  $\ell^1$ -representer, exact coefficient recovery by  $(\text{BP}_{\eta=0}^{\text{coef}})$  and signal recovery by  $(\text{BP}_{\eta=0}^{\text{sig}})$  are equivalent.

### 3.2 Signal Recovery and the Convex Gauge

The literature on compressive sensing predominantly focuses on a recovery of coefficient representations. However, if the goal is to recover the associated signal, this approach may be insufficient for structured dictionaries, as argued previously. In this section, we express the initial optimization problem over the coefficient domain  $(\text{BP}_{\eta}^{\text{sig}})$  as a minimization problem over the signal space. The  $\ell^1$ -ball  $\mathbb{B}_1^d$  in the coefficient domain is thereby mapped to the convex body  $\mathbf{D} \cdot \mathbb{B}_1^d$ , which is referred to as *synthesis defining polytope* in [EMR07]. The formulation  $(\text{BP}_{\eta}^{\text{sig}})$  can be equivalently expressed as a constrained minimization of its corresponding *convex gauge*.

**Definition 3.3** (Convex gauge) Let  $K \subseteq \mathbb{R}^n$  be a closed convex set that contains the origin. The *gauge* of  $K$  (also referred to as *Minkowski functional*) is defined as

$$p_K(\mathbf{x}) := \inf \{ \lambda > 0 : \mathbf{x} \in \lambda \cdot K \}.$$

For a symmetric set (i.e.,  $-K = K$ ) the gauge defines a semi-norm on  $\mathbb{R}^n$ , which becomes a norm if  $K$  is additionally bounded.

The following lemma provides an alternative characterization of the solutions  $\hat{X}$  to  $(\text{BP}_{\eta}^{\text{sig}})$ .

**Lemma 3.4** *Assume that  $\mathbf{x}_0, \mathbf{A}, \mathbf{y}, e$  and  $\eta$  follow Model 1.1 and let  $\mathbf{D} \in \mathbb{R}^{n \times d}$  be a dictionary. Then we have:*

$$\hat{X} = \underset{\mathbf{x} \in \mathbb{R}^n}{\text{argmin}} p_{\mathbf{D} \cdot \mathbb{B}_1^d}(\mathbf{x}) \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \eta. \quad (3.1)$$

A short proof for his equivalence is given in Appendix A.2. Under the heading of *atomic norm minimization*, problems of the form (3.1) were previously considered in greater generality in [CRPW12]: Given a collection of atoms  $\mathcal{A} \subseteq \mathbb{R}^n$ , Chandrasekaran et al. study the geometry of signal recovery based on minimizing the associated gauge  $p_{\text{conv}(\mathcal{A})}$  in (3.1). It turns out that many popular methods such as classical  $\ell^1$ -, or nuclear norm-minimization can be cast in such a form, e.g., by choosing  $\mathcal{A}$  as the set of one-sparse unit-norm vectors, or the set of rank-one matrices with unit-Euclidean-norm. Note that in the considered case of signal recovery via  $(\text{BP}_{\eta}^{\text{sig}})$ , one would choose the atoms  $\mathcal{A} = \{\pm \mathbf{d}_i : i \in [d]\}$  to obtain  $\text{conv}(\mathcal{A}) = \mathbf{D} \cdot \mathbb{B}_1^d$ . With this reformulation it is evident that dictionary atoms that are convex combinations of the remaining atoms in  $\mathcal{A}$  can be removed without altering  $\hat{X}$  [EMR07, Corollary 1].

For some specific problem instances novel sampling rate bounds are derived in [CRPW12]. Although this work plays a key role for the foundation of our work, we wish to emphasize that no explicit insights or bounds are derived in the case of signal recovery with dictionaries. In particular, the connection between  $(\text{BP}_{\eta}^{\text{sig}})$  and  $(\text{BP}_{\eta}^{\text{coef}})$  has not been studied.

With regard to the recovery framework of Section 2.1, it is of interest to determine the descent cone of the functional  $p_{D \cdot B_1^d}$  at  $\mathbf{x}_0$ . The following lemma shows how this cone is related to descent cones of the  $\ell^1$ -norm in the coefficient space.

**Lemma 3.5** *Let  $D \in \mathbb{R}^{n \times d}$  be a dictionary and let  $\mathbf{x}_0 \in \text{ran}(D)$ . For any  $\mathbf{z}_{\ell^1} \in Z_{\ell^1}$  we have*

$$\mathcal{D}_{\wedge}(p_{D \cdot B_1^d}, \mathbf{x}_0) = D \cdot \mathcal{D}_{\wedge}(\|\cdot\|_1, \mathbf{z}_{\ell^1}) \quad \text{and} \quad \mathcal{D}(p_{D \cdot B_1^d}, \mathbf{x}_0) = D \cdot \mathcal{D}(\|\cdot\|_1, \mathbf{z}_{\ell^1}).$$

The proof is given in Appendix A.3.

### 3.3 Sampling Rates for Signal and Coefficient Recovery

The purpose of this section is to determine the sampling rates for robust coefficient and signal recovery from sub-Gaussian measurements.

**Coefficient recovery** With Lemma 3.2 in mind, studying coefficient recovery is meaningful only if the signal  $\mathbf{x}_0$  has a unique minimal  $\ell^1$ -representer  $\mathbf{z}_{\ell^1}$  with respect to  $D$ . Proposition 2.3 implies that this condition can be equivalently expressed by

$$\lambda_{\min}(D; \mathcal{D}_{\wedge}(\|\cdot\|_1, \mathbf{z}_{\ell^1})) > 0.$$

Equipped with this assumption, we now state our main theorem regarding the recovery of coefficient vectors via  $(\text{BP}_{\eta}^{\text{coef}})$ .

**Theorem 3.6** (Coefficient recovery) *Assume that  $\mathbf{x}_0, A, \mathbf{y}, \mathbf{e}$  and  $\eta$  follow Model 1.1, where  $A$  is drawn according to the sub-Gaussian Model 1.2 with sub-Gaussian norm  $\gamma$ . Let  $D \in \mathbb{R}^{n \times d}$  be a dictionary and  $\mathbf{z}_{\ell^1} \in \mathbb{R}^d$  be a coefficient vector for the signal  $\mathbf{x}_0 = D\mathbf{z}_{\ell^1} \in \mathbb{R}^n$ , such that*

$$\lambda_{\min}(D; \mathcal{D}_{\wedge}(\|\cdot\|_1, \mathbf{z}_{\ell^1})) > 0.$$

*Then there exists a numerical constant  $c > 0$  such that for every  $u > 0$ , the following holds true with probability at least  $1 - e^{-u^2/2}$ : If the number of measurements obeys*

$$m > m_0 := c^2 \cdot \gamma^4 \cdot (w_{\wedge}(D \cdot \mathcal{D}(\|\cdot\|_1; \mathbf{z}_{\ell^1})) + u)^2 + 1, \quad (3.2)$$

*then any solution  $\hat{\mathbf{z}}$  to the program  $(\text{BP}_{\eta}^{\text{coef}})$  satisfies*

$$\|\mathbf{z}_{\ell^1} - \hat{\mathbf{z}}\|_2 \leq \frac{2\eta}{\lambda_{\min}(D; \mathcal{D}_{\wedge}(\|\cdot\|_1; \mathbf{z}_{\ell^1})) \cdot (\sqrt{m-1} - \sqrt{m_0-1})}. \quad (3.3)$$

*If  $\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \text{Id})$ , then  $c = \gamma = 1$ .*

A proof is given in Appendix A.4. Before turning towards signal recovery, let us highlight a few observations regarding the previous theorem.

**Remark 3.7** (a) Note that Theorem 3.6 does not assume anything on the dictionary  $D$  and the coefficient representation  $\mathbf{z}_{\ell^1}$ , except for  $\lambda_{\min}(D; \mathcal{D}_{\wedge}(\|\cdot\|_1, \mathbf{z}_{\ell^1})) > 0$ , which is a necessary condition for the theorem to hold true. As pointed out above, it reflects that  $\mathbf{z}_{\ell^1}$  is a unique  $\ell^1$ -representer of  $\mathbf{x}_0$  with respect to  $D$ , i.e., that  $\mathbf{z}_{\ell^1}$  is the unique solution to  $(\text{BP}_{\ell^1})$ . In general, verifying this property is involved (cf. the discussion in Section 2.2) and forms a trail of research on its own, e.g., see [Mal09, Chapter 12] or [CK13, Chapter 9]. In this regard, we think that an important contribution of Theorem 3.6 is that it allows to isolate the minimum prerequisite of a unique  $\ell^1$ -representer in  $D$  from the actual task of compressive coefficient recovery.

(b) Equation (3.2) identifies  $w_{\wedge}(D \cdot \mathcal{D}(\|\cdot\|_1; \mathbf{z}_{\ell^1}))$  as the essential component of the sampling rate for coefficient recovery by  $(\text{BP}_{\eta}^{\text{coef}})$ . Indeed, the proof reveals (in combination with the discussion subsequent to Theorem 2.5) that  $m_0$  is a tight description of the required number of noiseless Gaussian measurements for exact recovery.

- (c) Lastly, the error bound (3.3) shows that coefficient recovery is robust to measurement noise, provided that  $\lambda_{\min}(\mathbf{D}; \mathcal{D}_{\wedge}(\|\cdot\|_1, \mathbf{z}_{\ell^1})) \gg 0$ ; cf. the numerical experiments in Section 5, which confirm this observation. However, we note that this bound might not be tight, in general (cf. the intermediate inequality (A.1) in the proof, which is not necessarily sharp).  $\diamond$

**Signal recovery** Considering signal recovery by  $(\text{BP}_{\eta}^{\text{sig}})$ , a combination of the gauge formulation (3.1), its description of the descent cone in Lemma 3.5, and Theorem 2.5 directly yields the next result.

**Theorem 3.8** (Signal recovery) *Assume that  $\mathbf{x}_0, \mathbf{A}, \mathbf{y}, \mathbf{e}$  and  $\eta$  follow the measurement Model 1.1, where  $\mathbf{A}$  is drawn according to the sub-Gaussian Model 1.2 with sub-Gaussian norm  $\gamma$ . Let  $\mathbf{D} \in \mathbb{R}^{n \times d}$  be a dictionary with  $\mathbf{x}_0 \in \text{ran}(\mathbf{D})$  and pick any  $\mathbf{z}_{\ell^1} \in Z_{\ell^1}$ .*

*Then there exists a numerical constant  $c > 0$  such that for every  $u > 0$ , the following holds true with probability at least  $1 - e^{-u^2/2}$ : If the number of measurements obeys*

$$m > m_0 := c^2 \cdot \gamma^4 \cdot (w_{\wedge}(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; \mathbf{z}_{\ell^1})) + u)^2 + 1, \quad (3.4)$$

*then any solution  $\hat{\mathbf{x}}$  to the program  $(\text{BP}_{\eta}^{\text{sig}})$  satisfies*

$$\|\mathbf{x}_0 - \hat{\mathbf{x}}\|_2 \leq \frac{2\eta}{\sqrt{m-1} - \sqrt{m_0-1}}. \quad (3.5)$$

*If  $\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \text{Id})$ , then  $c = \gamma = 1$ .*

Let us discuss the previous result in view of its counterpart for coefficient recovery, Theorem 3.6.

**Remark 3.9** (a) Similarly as for coefficient recovery, (3.4) identifies  $w_{\wedge}^2(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; \mathbf{z}_{\ell^1}))$  as the main quantity of the sampling rate for signal recovery by  $(\text{BP}_{\eta}^{\text{sig}})$ . An important difference is that the set minimal of  $\ell^1$ -representers is not required to be a singleton: The descent cone in the signal space may be evaluated at any possible  $\mathbf{z}_{\ell^1} \in Z_{\ell^1}$  and the resulting sampling rate for signal recovery does not depend on this choice.

- (b) In the case of noiseless Gaussian measurements, the number  $m_0$  is a tight description of the phase transition of signal recovery, cf. the discussion subsequent to Theorem 2.5.
- (c) While the sampling rates for coefficient and signal recovery are similar, the error bounds of the two theorems differ. The inequality (3.5) does not involve the minimal conic singular value as in Theorem 3.6. This suggests the following noteworthy consequence: In the case of simultaneous coefficient and signal recovery, the robustness to noise of  $(\text{BP}_{\eta}^{\text{coef}})$  and  $(\text{BP}_{\eta}^{\text{sig}})$  might still be different. Indeed, while a reconstruction of  $\mathbf{x}_0$  is independent of the value of  $\lambda_{\min}(\mathbf{D}; \mathcal{D}_{\wedge}(\|\cdot\|_1, \mathbf{z}_{\ell^1}))$  – in fact, even 0 is allowed –, the error with respect to  $\mathbf{z}_{\ell^1}$  is directly influenced by it. We emphasize that the bound (3.5) cannot be retrieved from the analysis conducted for coefficient recovery. Indeed, the estimate (3.3) of Theorem 3.6 only implies that

$$\|\mathbf{x}_0 - \hat{\mathbf{x}}\|_2 = \|\mathbf{D}(\mathbf{z}_{\ell^1} - \hat{\mathbf{z}})\|_2 \leq \frac{\|\mathbf{D}\|_2}{\lambda_{\min}(\mathbf{D}; \mathcal{D}_{\wedge}(\|\cdot\|_1, \mathbf{z}_{\ell^1}))} \cdot \frac{2\eta}{\sqrt{m-1} - \sqrt{m_0-1}},$$

which is worse than (3.5), in general.  $\diamond$

While the bound (3.4) is accurate for an exact recovery from noiseless measurements, it can be improved when an approximate recovery of  $\mathbf{x}_0$  is already sufficient. This is reflected by the following proposition on *stable recovery*, which is an adaptation of a result in [GKM20]; see Appendix A.5 for a proof. Note that such an argumentation does not allow for a similar statement about stable coefficient recovery, due to the product  $\mathbf{A}\mathbf{D}$  in  $(\text{BP}_{\eta}^{\text{coef}})$ .

**Proposition 3.10** (Stable signal recovery) *Assume that  $\mathbf{x}_0, \mathbf{A}, \mathbf{y}, \mathbf{e}$  and  $\eta$  follow the measurement Model 1.1, where  $\mathbf{A}$  is drawn according to the sub-Gaussian Model 1.2 with sub-Gaussian norm  $\gamma$ . Let  $\mathbf{D} \in \mathbb{R}^{n \times d}$  be a dictionary with  $\mathbf{x}_0 = \mathbf{D}\mathbf{z}_0$ . For a desired precision  $\varepsilon > 0$  let*

$$\mathbf{z}^* \in \underset{\substack{\mathbf{z}: \|\mathbf{x}_0 - \mathbf{D}\mathbf{z}\|_2 \leq \varepsilon \\ \|\mathbf{z}_0\|_1 = \|\mathbf{z}\|_1}}{\operatorname{argmin}} w_\wedge(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; \mathbf{z})). \quad (3.6)$$

*Then there exists a numerical constant  $c > 0$  such that for every  $r > 0$  and  $u > 0$  the following holds true with probability at least  $1 - e^{-u^2/2}$ : If the number of measurements obeys*

$$m > \tilde{m}_0 := c^2 \cdot \gamma^4 \cdot \left( \frac{r+1}{r} \cdot [w_\wedge(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; \mathbf{z}^*)) + 1] + u \right)^2 + 1,$$

*then any solution  $\hat{\mathbf{x}}$  to (BP $_{\eta}^{\text{sig}}$ ) satisfies*

$$\|\mathbf{x}_0 - \hat{\mathbf{x}}\|_2 \leq \max \left( r\varepsilon, \frac{2\eta}{\sqrt{m-1} - \sqrt{\tilde{m}_0-1}} \right).$$

*If  $\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \mathbf{Id})$ , then  $c = \gamma = 1$ .*

The previous result extends Theorem 3.8 by an intuitive trade-off regarding stable signal recovery: By allowing for a lower recovery precision  $\varepsilon > 0$ , the number of required measurements  $\tilde{m}_0$  can be significantly lowered in comparison to  $m_0$  in (3.4). Indeed, (3.6) searches for surrogate representations  $\mathbf{z}^*$  of  $\mathbf{x}_0$  in  $\mathbf{D}$  that yield a minimal sampling rate. Note that the original coefficient vector  $\mathbf{z}_0$  is not required to be a minimal  $\ell^1$ -representer of  $\mathbf{x}_0$  with respect to  $\mathbf{D}$ . Thus, Proposition 3.10 enables to trade off the required number of measurements against the desired recovery accuracy. The factor  $r > 0$  is an additional oversampling parameter that may assist in balancing out this trade-off.

We emphasize that this approach to stability is centered around a Euclidean approximation in the signal domain  $\mathbb{R}^n$ . This is in stark contrast to a stability theory in the coefficient domain, which is typically based on an approximation of compressible vectors by ordinary *best s-term approximations*. We refer to Section 2.4 and 6.1 in [GKM20] as well as Section 2.4 in [GMS20] for more details on the presented approach to stable recovery and related results in the literature.

**Remark 3.11** The normalization condition  $\|\mathbf{z}_0\|_1 = \|\mathbf{z}\|_1$  in (3.6) can be discarded at the expense of a slightly worse error bound. Under the same conditions as in Proposition 3.10, we can also derive the following result. Set  $\varepsilon > 0$  and let

$$\mathbf{x}^* \in \underset{\|\mathbf{x} - \mathbf{x}_0\|_2 \leq \varepsilon}{\operatorname{argmin}} w_\wedge(\mathcal{D}(p_{\mathbf{D} \cdot \mathbb{B}_1^d}; \mathbf{x})) \quad \text{and} \quad K = \mathcal{D}(p_{\mathbf{D} \cdot \mathbb{B}_1^d}, \mathbf{x}^*).$$

For every  $u > 0$  and

$$m > \tilde{m}_0 := c^2 \cdot \gamma^4 \cdot (w_\wedge(K) + u)^2 + 1,$$

the inequality

$$\|\mathbf{x}_0 - \hat{\mathbf{x}}\|_2 \leq \frac{4(\eta + \varepsilon(\sqrt{m} + C\gamma^2 + u))}{\sqrt{m-1} - \sqrt{\tilde{m}_0-1}} + \varepsilon \quad (3.7)$$

holds true with probability larger than  $1 - 2e^{-u^2}$  for some constant  $C$  depending only on the distribution of  $\mathbf{A}$ . For instance, letting  $m = r\tilde{m}_0$  with  $r > 1$  as an oversampling factor, the latter bound essentially becomes:

$$\|\mathbf{x}_0 - \hat{\mathbf{x}}\|_2 \lesssim \frac{4\eta}{(\sqrt{r}-1)\tilde{m}_0} + \frac{4\varepsilon\sqrt{r}}{(\sqrt{r}-1)} + \varepsilon.$$

◇

We conclude this section by an illustration of stable recovery in two simple examples.



**Example 3.12** (a) Assume that  $\mathbf{D} = \mathbf{Id}$  and let  $\mathbf{x}_0 \in \mathbb{R}^n$  denote a fully populated vector, which is without loss of generality assumed to be positive and nonincreasing. Standard results on the computation of the conic mean width (see for instance [Tro15, Example 4.3]) stipulate that  $w_\lambda^2(\mathcal{D}(\|\cdot\|_1; \mathbf{x}_0)) = n$ . Hence, it is impossible to exactly recover  $\mathbf{x}_0$  from noiseless compressive measurements. However, if we are satisfied with an approximate recovery of  $\mathbf{x}_0$ , we can set the precision for instance to  $\varepsilon = 3 \cdot \sigma_s(\mathbf{x}_0)_1 / \sqrt{s}$ , where  $\sigma_s(\mathbf{x}_0)_p$  denotes the  $\ell^p$ -error of the best  $s$ -term approximation to  $\mathbf{x}_0$ . Then, the surrogate vector  $\mathbf{x}^* \in \mathbb{R}^n$  defined as  $x_i^* := x_{0,i} + \sigma_s(\mathbf{x}_0)_1 / s$  for  $i = 1, \dots, s$  and  $x_i^* := 0$  for  $i = s + 1, \dots, n$ , satisfies  $\|\mathbf{x}^*\|_1 = \|\mathbf{x}_0\|_1$ . A straightforward calculation shows that  $\|\mathbf{x}^* - \mathbf{x}_0\|_2^2 \leq \sigma_s(\mathbf{x}_0)_1^2 / s + \sigma_s(\mathbf{x}_0)_2^2$ , which eventually leads to  $\|\mathbf{x}^* - \mathbf{x}_0\|_2 \leq 3 \cdot \sigma_s(\mathbf{x}_0)_1 / \sqrt{s}$ . Furthermore, a computation of the conic mean width yields that  $w_\lambda^2(\mathcal{D}(\|\cdot\|_1; \mathbf{x}^*)) \lesssim 2s \log(n/s)$ . Hence, Proposition 3.10 shows that  $(\text{BP}_{\eta=0}^{\text{sig}})$  allows for the reconstruction of an approximation  $\hat{\mathbf{x}}$  from  $m \gtrsim 2s \log(n/s)$  noiseless sub-Gaussian measurements that satisfies  $\|\mathbf{x}_0 - \hat{\mathbf{x}}\|_2 \lesssim 3 \cdot \sigma_s(\mathbf{x}_0)_1 / \sqrt{s}$ . A comparison with [FR13, Theorem 4.22 and Section 11.1] shows that such a stability result is essentially optimal.

(b) Let us provide another simple result highlighting the important difference between signal and coefficient recovery. Consider a dictionary consisting of a convolution with a low pass filter  $\mathbf{h} \in \mathbb{R}^n$ , i.e., for any  $\mathbf{z} \in \mathbb{R}^d$ ,  $\mathbf{D}\mathbf{z} = \mathbf{h} \star \mathbf{z}$ . The problem  $(\text{BP}_\eta^{\text{sig}})$  then becomes a *deconvolution* problem and it could be turned to a super-resolution problem by considering non integer shifts of the kernel. In this setting, it is well known [CF14] that coefficients of the form  $\mathbf{z}_0 = [1, -1, 0, \dots, 0]$  are hard to recover by solving  $(\text{BP}_\eta^{\text{coef}})$  since  $\mathbf{h} \star \mathbf{z}_0 \simeq 0$ . There is a minimum separation distance to respect to guarantee the recovery of sparse spikes with arbitrary signs. We can however use the result (3.7) by setting  $\varepsilon = \|\mathbf{D}\mathbf{z}_0\|_2$ . In that case, we obtain  $\hat{m}_0 = c^2 \gamma^4 u^2 + 1$  by picking  $\mathbf{x}^* = 0$ . Hence, we can recover an  $\varepsilon$ -approximation of  $\mathbf{x}_0 = \mathbf{D}\mathbf{z}_0$  with a few measurements. ◇

## 4 Upper Bounds on the Conic Gaussian Width

The previous results identify the conic mean width  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; \mathbf{z}_{\ell^1}))$  as the key quantity that controls coefficient and signal recovery by  $\ell^1$ -synthesis. However, this expression does not convey an immediate understanding without further simplification. While tight and informative upper bounds are available for simple dictionaries such as orthogonal matrices, the situation becomes significantly more involved for general, possibly redundant transforms. Indeed, note that the polar cone of  $\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; \mathbf{z}_{\ell^1})$  is given by  $(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; \mathbf{z}_{\ell^1}))^\circ = (\mathbf{D}^T)^{-1}(\mathcal{D}(\|\cdot\|_1; \mathbf{z}_{\ell^1}))^\circ$ . The appearance of the preimage  $(\mathbf{D}^T)^{-1}$  hinders the application of the standard approach based on polarity; see for instance [ALMT14, Recipe 4.1].

Hence, the goal of this section is to provide two upper bounds for  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; \mathbf{z}_{\ell^1}))$  that are more accessible and intuitive: Section 4.1 is based on a local conditioning argument, and addresses recovery when a unique minimal  $\ell^1$ -representer exists. The second bound of Section 4.2 follows a geometric analysis that explores the thinness of high-dimensional polyhedral cones with not too many generators. This approach possesses a broader scope and plays a central role in our work.

### 4.1 A Condition Number Bound

In this section, we aim at “pulling” the dictionary  $\mathbf{D}$  “out of” the expression  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; \mathbf{z}_{\ell^1}))$ , in order to make use of the fact that  $w_\lambda^2(\mathcal{D}(\|\cdot\|_1; \mathbf{z}_{\ell^1}))$  is well understood. We begin by introducing the following notation of a local condition number.

**Definition 4.1** (Local condition number) Let  $\mathbf{D} \in \mathbb{R}^{n \times d}$  be a dictionary and let  $C \subseteq \mathbb{R}^d$  be closed convex cone. Then, we define the *local condition number of  $\mathbf{D}$  with respect to  $C$*  by

$$\kappa_{\mathbf{D}, C} := \frac{\|\mathbf{D}\|_2}{\lambda_{\min}(\mathbf{D}; C)},$$

with the convention  $\kappa_{D,C} = +\infty$  if  $\lambda_{\min}(D; C) = 0$ . We also use the notation  $\kappa_{D,z_0} := \kappa_{D, \mathcal{D} \wedge (\|\cdot\|_1; z_0)}$  which we refer to as *local condition number of  $D$  at  $z_0$  with respect to the  $\ell^1$ -norm*.

Before the previous quantity will be used to simplify  $w_\lambda^2(D \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$ , we first comment on the origin of its name and give an intuitive interpretation of its meaning in the following remark.

**Remark 4.2** (a) First, recall that the classical, generalized condition number of a matrix is defined as the ratio of the largest and the smallest nonzero singular value. Hence, referring to  $\kappa_{D,C}$  as a local condition number is motivated by the fact that it can also be written as

$$\kappa_{D,C} = \frac{\|D\|_2}{\lambda_{\min}(D; C)} = \frac{\lambda_{\max}(D; \mathbb{R}^d)}{\lambda_{\min}(D; C)},$$

where  $\lambda_{\max}(D; \mathbb{R}^d) := \max_{z \in \mathbb{R}^d \cap \mathcal{S}^{d-1}} \|Dz\|_2 = \|D\|_2$  is the largest singular value of  $D$ .

(b) Furthermore, note that  $\kappa_{D,z_0}$  acts as a local measure for the conditioning of  $D$  at  $z_0$  with respect to the  $\ell^1$ -norm. It quantifies how robustly  $z_0$  can be recovered as the minimal  $\ell^1$ -representer of  $x_0 = Dz_0$ : Consider the perturbation  $\hat{z}_0 = z_0 + \hat{e}$ , where  $\hat{e} \in \mathbb{R}^d$  with  $\|\hat{e}\|_2 \leq \hat{\eta}$ . Thus, in the signal domain we obtain  $\|Dz_0 - D \cdot \hat{z}_0\|_2 = \|D\hat{e}\|_2 \leq \|D\|_2 \cdot \hat{\eta}$ . Proposition 2.3 then yields that any solution  $\hat{z}$  of the program

$$\min_{z \in \mathbb{R}^d} \|z\|_1 \quad \text{s.t.} \quad \|D\hat{z}_0 - Dz\|_2 \leq \|D\|_2 \cdot \hat{\eta}$$

satisfies

$$\|z_0 - \hat{z}\|_2 \leq \frac{2 \cdot \|D\|_2 \cdot \hat{\eta}}{\lambda_{\min}(D; \mathcal{D} \wedge (\|\cdot\|_1; z_0))} \lesssim \kappa_{D,z_0} \cdot \hat{\eta},$$

which shows that  $\kappa_{D,z_0}$  can be seen as a measure for the stability of  $z_0$  with respect to  $\ell^1$ -minimization with  $D$ . ◇

The following proposition provides a generic upper bound for the conic mean width of a linearly transformed cone.

**Proposition 4.3** *Let  $C \subseteq \mathbb{R}^d$  denote a closed convex cone. For any dictionary  $D \in \mathbb{R}^{n \times d}$ , we have*

$$w_\lambda^2(D \cdot C) \leq \kappa_{D,C}^2 \cdot (w_\lambda^2(C) + 1). \quad (4.1)$$

*Proof.* See Appendix B. ■

Note that for sparse coefficient vectors  $z_0$  the quantity  $w_\lambda^2(\mathcal{D}(\|\cdot\|_1; z_0))$  is well understood and has been frequently calculated in the literature, see for instance [Tro15, Example 4.3]. It turns out that it can be bounded from above by

$$w_\lambda^2(\mathcal{D}(\|\cdot\|_1; z_0)) \leq 2s \log(d/s) + 2s,$$

where  $s = \#\text{supp}(z_0)$ . Hence, we directly obtain the following corollary.

**Corollary 4.4** *If  $z_{\ell^1}$  is the unique minimal  $\ell^1$ -representer of the associated signal  $x_0 = Dz_{\ell^1}$ , the critical number of measurements  $m_0$  in (3.2) and (3.4) satisfies*

$$m_0 \leq c^2 \cdot \gamma^4 \cdot \left( \kappa_{D,z_{\ell^1}} \cdot (w_\lambda^2(\mathcal{D}(\|\cdot\|_1; z_{\ell^1})) + 1)^{1/2} + u \right)^2 + 1 \lesssim \kappa_{D,z_{\ell^1}}^2 \cdot s \log(d/s),$$

where  $s = \#\text{supp}(z_{\ell^1})$ .

We have assumed  $z_{\ell^1}$  to be a unique minimal  $\ell^1$ -representer since otherwise the previous statement becomes meaningless due to  $\kappa_{D,z_{\ell^1}} = +\infty$ . Thus, the condition number bound of Corollary 4.4 foremost addresses coefficient recovery via  $(\text{BP}_\gamma^{\text{coef}})$ , as well as a reconstruction of signals

with unique minimal  $\ell^1$ -representers in  $D$  by (B <sub>$\eta$</sub> <sup>psig</sup>). In both cases,  $m \gtrsim \kappa_{D, z_{\ell^1}}^2 \cdot s \log(d/s)$  sub-Gaussian measurements are sufficient (recall that the two formulations might nevertheless differ with respect to robustness to measurement noise). Hence, the results of this section identify the following three decisive factors for successful recovery:

- (i) The uniqueness of  $z_{\ell^1}$  as the minimal  $\ell^1$ -representer of  $x_0 = Dz_{\ell^1}$ ;
- (ii) The complexity of  $z_{\ell^1}$  with respect to  $\ell^1$ -norm, which is measured by  $w_\lambda^2(D(\|\cdot\|_1; z_{\ell^1}))$ , or by its sparsity  $s = \#\text{supp}(z_{\ell^1})$ ;
- (iii) The quantity  $\kappa_{D, z_{\ell^1}}$ , which resembles a local measure for the conditioning of  $D$  at  $z_{\ell^1}$ .

We demonstrate in the numerical experiments of Section 5.1 that such a condition number approach might be accurate for some specific problems, however, it is overly pessimistic in general. Indeed, it is possible that  $w_\lambda^2(D \cdot D(\|\cdot\|_1; z_{\ell^1})) \leq w_\lambda^2(D(\|\cdot\|_1; z_{\ell^1}))$ , but  $\kappa_{D, z_{\ell^1}} \gg 1$ . We suspect that a more accurate description might require a detailed analysis of random conic spectra [ST03].

**Remark 4.5** (a) In the case  $D = \mathbf{Id}$ , observe that Corollary 4.4 is consistent with standard compressed sensing results. Indeed, in this situation, it holds true that

$$\kappa_{\mathbf{Id}, z_{\ell^1}} = 1 = \|\mathbf{Id}\|_2 = \lambda_{\min}(\mathbf{Id}; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1})),$$

implying that  $m \gtrsim s \log(n/s)$  measurements are sufficient for robust recovery of  $s$ -sparse signals.

- (b) During completion of this work, we discovered that similar bounds as (4.1) were recently derived in [ALW20]. Amelunxen et al. do not address the synthesis formulation of compressed sensing, but they study the statistical dimension of linearly transformed cones in a general setting. Their results are based on a notion of *Renegar's condition number*, which can be defined as

$$\mathcal{R}_C(D) = \min \left\{ \frac{\|D\|_2}{\lambda_{\min}(D; C)}, \frac{\|D\|_2}{\sigma_{\mathbb{R}^n \rightarrow C}(-D^T)} \right\}, \quad (4.2)$$

where  $C \subseteq \mathbb{R}^d$  is a closed, convex cone,  $\sigma_{\mathbb{R}^n \rightarrow C}(-D^T) := \min_{x \in \mathcal{S}^{n-1}} \|\Pi_C(-D^T x)\|_2$  and  $\Pi_C$  denotes the orthogonal projection on  $C$ . [ALW20, Theorem A] then establishes the bound  $\delta(D \cdot C) \leq \mathcal{R}_C^2(D) \cdot \delta(C)$ , where  $\delta$  denotes the *statistical dimension*, which is essentially equivalent to the conic mean width; see proof of Proposition 4.3 in Appendix B for details.

Additionally, the authors of [ALW20] provide a “preconditioned”, probabilistic version of the latter bound: For  $m \leq n$  let  $P_m$  denote the projection onto the first  $m$  coordinates and define the quantity  $\mathcal{R}_{C, m}^2(D) := \mathbb{E}_Q[\mathcal{R}_C(P_m Q D)^2]$ , where the expectation is with respect to a random orthogonal matrix  $Q$ , distributed according to the normalized Haar measure on the orthogonal group. [ALW20, Theorem B] then states that for any parameter  $\nu \in (0, 1)$  and  $m \geq \delta(C) + 2\sqrt{\log(2/\nu)m}$ , we have that  $\delta(D \cdot C) \leq \mathcal{R}_{C, m}^2(D) \cdot \delta(C) + (n - m) \cdot \nu$ . Due to the second term in (4.2), both versions of Renegar's condition number will be not greater than  $\kappa_{D, C}$ , in general. Hence, ignoring the dependence on  $\nu$  and the condition on  $m$  for simplicity, the bound on the required samples of Corollary 4.4 could also be formulated with  $\mathcal{R}_{\mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1})}^2(D)$  or  $\mathcal{R}_{\mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}), m}^2(D)$  instead of  $\kappa_{D, z_{\ell^1}}$ . ◇

## 4.2 A Geometric Bound

In this section, we derive an upper bound for  $w_\lambda^2(D \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$  that is based on generic arguments from high-dimensional convex geometry. We exploit the fact that the cone  $D \cdot \mathcal{D}_\wedge(\|\cdot\|_1; z_{\ell^1})$  is finitely generated by at most  $2d$  vectors (see the proof of Proposition 4.16 in Appendix C.3) – a number that is typically significantly smaller than exponential in the ambient dimension  $n$ . The resulting upper bound depends on the maximal sparsity of elements in  $Z_{\ell^1}$  and on a single geometric parameter that we refer to as *circumangle*, whereas the number of generators only has a logarithmic influence. This is comparable to the mean width of a convex polytope, which is

mainly determined by its diameter (cf. Lemma C.1) and by the logarithm of its number of vertices.

In Section 4.2.1, we first introduce the required notation and show an upper bound on the conic mean width of pointed polyhedral cones. We then focus on the geometry of the descent cone  $\mathcal{D}_\wedge(p_{D \cdot \mathbb{B}_1^d}, x_0)$  (see Section 4.2.2) in order to derive the desired upper bound on the expression  $w_\lambda^2(D \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$  in Section 4.2.3. Finally, we show how this bound can be used in practical examples; see Section 4.2.4.

#### 4.2.1 The Circumangle

The goal of this section is to relate the conic mean width of a pointed polyhedral cone to its *circumangle*, which describes the angle of an enclosing circular cone. To that end, recall that a *circular cone* (also referred to as *revolution cone*) with axis  $\theta \in \mathcal{S}^{n-1}$  and (half-aperture) angle  $\alpha \in [0, \pi/2]$  is defined as

$$C(\alpha, \theta) := \{x \in \mathbb{R}^n, \langle x, \theta \rangle \geq \|x\|_2 \cdot \cos(\alpha)\}.$$

The conic mean width of a circular cone depends linearly on the ambient dimension  $n$ , i.e.,  $w_\lambda^2(C(\alpha, \theta)) = n \cdot \sin^2(\alpha) + O(1)$ , see for instance [ALMT14, Proposition 3.4]. Although not directly related, it will be insightful to compare this result with the subsequent upper bound of Proposition 4.9.

The following definition introduces the so-called circumangle of a nontrivial (different from  $\{0\}$  and  $\mathbb{R}^n$ ) closed convex cone  $C$ . It describes the angle of the smallest circular cone that contains  $C$ .

**Definition 4.6** (Circumangle) Let  $C \subset \mathbb{R}^n$  denote a nontrivial closed convex cone. Its *circumangle*  $\alpha$  is defined by

$$\alpha := \inf \{\hat{\alpha} \in [0, \pi/2] : \exists \theta \in \mathcal{S}^{n-1}, C \subseteq C(\hat{\alpha}, \theta)\}.$$

The previous notion can be found under various names in the literature, see for instance [FV99; HS10b; IS08; Ren95]. In particular, the previous quantity arises in the definition of an outer center of a cone [HS10a]. It turns out that the circumangle satisfies

$$\cos(\alpha) = \sup_{\theta \in \mathcal{S}^{n-1}} \inf_{x \in C \cap \mathcal{S}^{n-1}} \langle \theta, x \rangle,$$

where a vector  $\theta$  that maximizes the right hand side is referred to as *circumcenter* (or *outer center*<sup>1</sup>) of  $C$  [HS10a; IS08]. Furthermore, if  $C$  is pointed (i.e., if it does not contain a line), the circumcenter is unique and  $\alpha \in [0, \pi/2)$  [HS10a].

Note that the function  $\theta \mapsto \inf_{x \in C \cap \mathcal{S}^{n-1}} \langle \theta, x \rangle$  is concave as a minimum of concave functions. Hence, if  $C$  is pointed, it is easy to see that determining the circumcenter and the circumangle amounts to solving the following convex optimization problem:

$$\cos(\alpha) = \sup_{\theta \in \mathbb{B}_2^n} \inf_{x \in C \cap \mathcal{S}^{n-1}} \langle \theta, x \rangle.$$

We now show that this characterization can be further simplified for pointed polyhedral cones. The simple characterization of the following proposition makes it possible to numerically compute the circumangle of such cones. We emphasize that this stands in contrast to previously discussed notions such as the minimum conic singular value, which is intractable in general. A short proof is included in Appendix C.1.

**Proposition 4.7** (Circumangle and circumcenter of polyhedral cones) Let  $x_i \in \mathcal{S}^{n-1}$  for  $i \in [k]$  and let  $C = \text{cone}(x_1, \dots, x_k)$  be a nontrivial pointed polyhedral cone. Finding the circumcenter and

<sup>1</sup>Note that the notions of circumcenter and outer centers generally differ, however, in the Euclidean setting of this work they are equivalent [HS10a, Section 5].

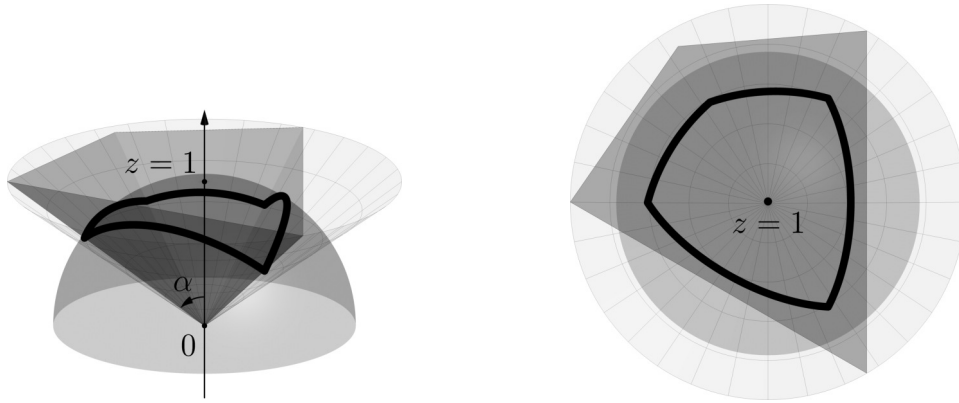


Figure 2: **Geometry of Proposition 4.9.** The figure shows a polyhedral cone (transparent gray) truncated at  $z = 1$  and included in a circular cone with angle  $\alpha$  (wire-frame). The standard basis vector  $e_z$  corresponds to the circumcenter of the polyhedral cone and  $\alpha$  is chosen as its circumangle. The thick line is the intersection of the unit sphere with the faces of the polyhedral cone. Right view is from above, or equivalently, the projection on the plane  $z = 1$ . The conic mean width of the polyhedral cone can be bounded by evaluating the mean width of any set containing the thick line plus 1 (as a subset of the plane). The proposed bound is based on using the intersection of the polyhedral cone and the plane  $z = 1$ . Notice that this convex body is included in the disk with radius  $\tan \alpha$ . In high dimensions, it will be thin if the polyhedral cone does not have overwhelmingly many extremal rays.

circumangle of  $C$  amounts to solving the convex problem:

$$\cos(\alpha) = \sup_{\theta \in B_2^n} \inf_{i \in [k]} \langle \theta, x_i \rangle.$$

The goal of this section is to upper bound the conic mean width of all polyhedral cones  $C \subset \mathbb{R}^n$  with  $k$  generators that are contained in a circular cone of angle  $\alpha$ . To that end, we first introduce the following notation:

**Definition 4.8** A  $k$ -polyhedral  $\alpha$ -cone  $C \subset \mathbb{R}^n$  is a nontrivial pointed polyhedral cone generated by  $k$  vectors that is included in a circular cone with angle  $\alpha \in [0, \pi/2)$ . Furthermore, we let  $\mathcal{C}_k^\alpha$  denote the set of all  $k$ -polyhedral  $\alpha$ -cones.

Note that  $C$  being a nontrivial pointed polyhedral cone implies that such an encompassing circular cone with angle  $\alpha \in [0, \pi/2)$  exists. The next result provides a simple upper bound on the quantity

$$W(\alpha, k, n) := \sup_{C \in \mathcal{C}_k^\alpha, C \subset \mathbb{R}^n} w_\wedge(C).$$

The underlying geometric idea is explained in Figure 2 and its proof is detailed in Appendix C.2. Note that the bound does not depend on the ambient dimension  $n$ , which is in contrast to the conic width of a circular cone.

**Proposition 4.9** For  $k \geq 5$ , the conic mean width of a  $k$ -polyhedral  $\alpha$ -cone  $C$  in  $\mathbb{R}^n$  is bounded by

$$W(\alpha, k, n) \leq \tan \alpha \cdot \left( \sqrt{2 \log(k/\sqrt{2\pi})} + \frac{1}{\sqrt{2 \log(k/\sqrt{2\pi})}} \right) + \frac{1}{\sqrt{2\pi}}.$$

We conclude this section with the following remark:

**Remark 4.10** The previous upper bound is based on Lemma C.1, which provides a basic bound on the Gaussian mean width of a convex polytope; see also [Ver18, Ex. 7.5.10 & Prop. 7.5.2]. Using

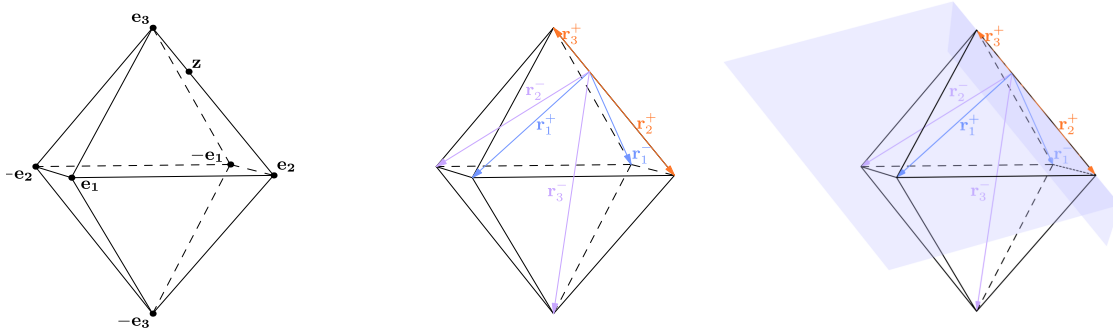


Figure 3: **Illustration of Lemma 4.11.** Left:  $\ell^1$ -ball in  $\mathbb{R}^3$  and a 2-sparse vector  $z$ . Center: The rays of the descent cone are supported by the vectors  $r_i^\pm = \pm e_i - z$ , which corresponds to the generators of Lemma 4.11 with  $v = s \cdot z / \|z\|_1$ , multiplied by  $1/2$ . Right: The resulting descent cone (shifted by  $z$ ). Note that it contains a linear subspace spanned by  $r_2^+$  and  $r_3^+$ .

a tighter estimate there (possibly an implicit description as in [ALMT14, Proposition 4.5]) would in turn also improve Proposition 4.9.  $\diamond$

#### 4.2.2 Geometry of the Descent Cone

In order to derive an upper bound for the quantity  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$  based on the previous result, we first need a more geometrical description of the descent cones of the  $\ell^1$ -norm and of the gauge  $p_{\mathbf{D} \cdot \mathbf{B}_1^d}$ .

**Lemma 4.11** *Let  $z \in \mathbb{R}^d$  with support  $\text{supp } z = S$  and  $\#S = s$ . Then,*

$$\mathcal{D}_\wedge(\|\cdot\|_1, z) = \text{cone}(\pm s \cdot e_i - v : i \in [d]).$$

where  $v$  is any vector such that  $\|v\|_1 = s$  and  $\text{sign } v = \text{sign } z$ , e.g.,  $v = \text{sign } z$  or  $v = s \cdot z / \|z\|_1$ .

A proof of the previous lemma can be found in Appendix C.3. The statement is illustrated in Figure 3 for dimension  $d = 3$ . Observe that sliding  $z$  along the edge linking  $e_2$  with  $e_3$  leaves the descent cone unchanged.

Whenever a convex cone contains a subspace, its circumangle is  $\pi/2$  and the bound of Proposition 4.9 is not applicable. As can be seen in Figure 3, the descent cone of the  $\ell^1$ -norm at  $z$  contains the subspace spanned by the face of minimal dimension containing  $z$ . To avoid this pitfall, let us recall the notion of lineality.

**Definition 4.12** (Lineality [Roc70]) For a non-empty convex set  $C \subseteq \mathbb{R}^n$ , the *lineality space*  $C_L$  of  $C$  is defined as

$$C_L := \{x \in \mathbb{R}^n : \forall \tilde{x} \in C : \{\tilde{x} + \alpha \cdot x : \alpha \in \mathbb{R}\} \subseteq C\}.$$

It defines a subspace of  $\mathbb{R}^n$  and its dimension is referred to as the *lineality* of  $C$ .

Any non-empty convex set  $C$  can be expressed as the direct sum

$$C = C_L \oplus C_R \text{ with } C_R := \mathbf{P}_{C_L^\perp}(C), \quad (4.3)$$

where  $\mathbf{P}_{C_L^\perp}$  is the orthogonal projection onto  $C_L^\perp$ . The notation  $C_R$  is used in analogy to the *range* in linear algebra. The set  $C_R$  is “line-free” (i.e., its lineality is  $\{0\}$ ) and its dimension is called the *rank* of  $C$ . If  $C$  is a convex cone, the lineality space is the largest subspace contained in  $C$  and the circumangle of the range  $C_R$  is less than  $\pi/2$ . In the following, we will therefore apply Proposition 4.9 to the latter set only. Figure 4 illustrates the orthogonal decomposition of (4.3) for the descent cone of Figure 3.

The following lemma characterizes the lineality space and the range for descent cones of the  $\ell^1$ -norm. A proof is given in Appendix C.3.



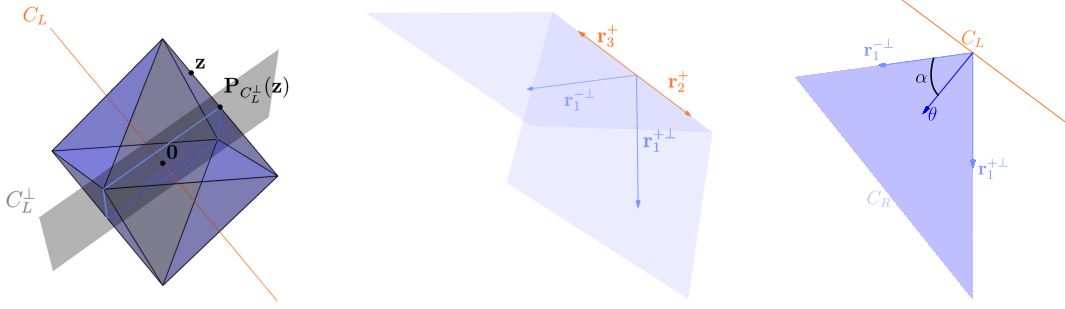


Figure 4: **Decomposition into the direct sum of the lineality space and range.** Left: Decomposition of  $\mathbb{R}^3$  into the lineality space  $C_L$  and its orthogonal complement  $C_L^\perp$ , where  $C$  is the descent cone of Figure 3. Middle: Different view of  $C$ , where  $r_i^{\pm\perp} := P_{C_L^\perp}(r_i^\pm)$ . Right: Visualization of the orthogonal decomposition of  $C$  into its lineality space and range  $C_R = P_{C_L^\perp}(C)$ . The angle  $\alpha$  corresponds to the circumangle of  $C_R$  and  $\theta$  denotes its circumcenter.

**Lemma 4.13** Let  $z = (z_1, \dots, z_d)$  be a vector with support  $\text{supp } z = \mathcal{S}$  and  $\#\mathcal{S} = s \geq 1$ . The lineality space of  $C = \mathcal{D}_\wedge(\|\cdot\|_1, z)$  is then given by

$$C_L = \text{span}(s \cdot \text{sign}(z_i) \cdot e_i - \text{sign}(z) : i \in \mathcal{S}),$$

with lineality  $\dim(C_L) = s - 1$ .

Notice that the lineality space is nothing but the face of the  $\ell^1$ -ball of smallest dimension containing  $z$ .

We now turn to the decomposition of the descent cone of the gauge  $p_{D \cdot B_1^d}$  into lineality space and range. To that end, let us first make the following simple observation.

**Lemma 4.14** (Sign pattern of  $\ell^1$ -representers) All minimal  $\ell^1$ -representers of  $x_0$  with respect to  $D$  share the same sign pattern, in the sense that for all  $z_{\ell^1}^1, z_{\ell^1}^2 \in Z_{\ell^1}$ , the coordinate-wise product  $z_{\ell^1}^1 \cdot z_{\ell^1}^2$  is nonnegative.

This lemma allows us to define the maximal  $\ell^1$ -support.

**Definition 4.15** (Maximal  $\ell^1$ -support) Let  $x_0 \in \mathbb{R}^n$  and  $D \in \mathbb{R}^{n \times d}$  be a dictionary. The maximal  $\ell^1$ -support  $\bar{\mathcal{S}}$  of  $x_0$  in  $D$  (or simply maximal support) is defined as  $\bar{\mathcal{S}} := \cup_{z_{\ell^1} \in Z_{\ell^1}} \text{supp}(z_{\ell^1})$ . In what follows, we let  $\bar{s} = \#\bar{\mathcal{S}}$  denote its cardinality.

Since all solutions  $z_{\ell^1} \in Z_{\ell^1}$  have the same sign pattern, any point  $z_{\ell^1}$  in the relative interior of  $Z_{\ell^1}$  has maximal support. The next decomposition forms the main result of this section; see Appendix C.3 for a proof.

**Proposition 4.16** Let  $D \in \mathbb{R}^{n \times d}$  be a dictionary and let  $x_0 \in \text{ran}(D) \setminus \{0\}$ . Let  $C = \mathcal{D}_\wedge(p_{D \cdot B_1^d}, x_0)$  denote the descent cone of the gauge at  $x_0$ . Let  $z_{\ell^1} \in \text{ri}(Z_{\ell^1})$  be any minimal  $\ell^1$ -representer of  $x_0$  in  $D$  with maximal support and set  $\bar{\mathcal{S}} = \text{supp}(z_{\ell^1})$  as well as  $\bar{s} = \#\bar{\mathcal{S}}$ . Assume  $\bar{s} < d$ . Then we have:

(a) The lineality space of  $C$  has a dimension less than  $\bar{s} - 1$  and is given by

$$C_L = \text{span}(\bar{s} \cdot \text{sign}(z_{\ell^1, i}) \cdot d_i - D \cdot \text{sign}(z_{\ell^1}) : i \in \bar{\mathcal{S}}).$$

(b) The range of  $C$  is a  $2(d - \bar{s})$ -polyhedral  $\alpha$ -cone given by:

$$C_R = \text{cone}(r_j^{\pm\perp} : j \in \bar{\mathcal{S}}^c) \text{ with } r_j^{\pm\perp} := P_{C_L^\perp}(\pm \bar{s} \cdot d_j - D \cdot \text{sign}(z_{\ell^1})).$$

### 4.2.3 Consequence for the Sampling Rates

We now combine the main results of the previous two sections to derive an upper bound on the conic mean width of  $\mathcal{D}_\wedge(p_{D \cdot B_1^d}, x_0)$ .

**Theorem 4.17** *If  $\bar{s} \leq d - 3$ , we obtain that*

$$w_\lambda^2(\mathcal{D}_\wedge(p_{D \cdot B_1^d}, \mathbf{x}_0)) \leq \bar{s} + \left( \tan \alpha \cdot \left( \sqrt{2 \log \left( \frac{2(d - \bar{s})}{\sqrt{2\pi}} \right) + 1} + \frac{1}{\sqrt{2\pi}} \right) \right)^2,$$

where we have used the same notation and assumptions as in Proposition 4.16.

A proof of the previous result is given in Appendix C.4. As a direct consequence, we get the following upper bound on the sampling rates for coefficient and signal recovery.

**Corollary 4.18** *The critical number of measurements  $m_0$  in (3.2) and (3.4) satisfies*

$$m_0 \lesssim \bar{s} + \tan^2 \alpha \cdot \log(2(d - \bar{s}) / \sqrt{2\pi}).$$

This result shows that robust coefficient and signal recovery is possible, when the number of measurements obeys  $m \gtrsim \bar{s} + \tan^2 \alpha \cdot \log(d)$ . Hence, the sampling rate is mainly governed by the sparsity of maximal support  $\ell^1$ -representations of  $\mathbf{x}_0$  in  $D$  and the “narrowness” of the remaining cone  $C_R$ , which is captured by its circumangle  $\alpha \in [0, \pi/2)$ . The number of dictionary atoms only has a logarithmic influence. The next section is devoted to applying the previous result to various examples.

**Remark 4.19** For the sake of clarity, the previous results are given in terms of the maximal sparsity. However, (potentially) more precise bounds can be achieved when replacing  $\bar{s}$  by  $\dim(C_L)$ . Furthermore note, that the proof of Theorem 4.17 reveals that  $\dim(C_L)$  is a necessary component in the required number of measurements. Indeed, since  $w_\lambda^2(\mathcal{D}_\wedge(p_{D \cdot B_1^d}, \mathbf{x}_0))$  is a sharp description for the required number of measurements, equation (C.5) shows that the number of measurements for successful recovery is lower bounded by  $\dim(C_L)$ .  $\diamond$

#### 4.2.4 Examples

In this section, we discuss four applications of the previous upper bound on the conic mean width. First, we show that prediction for the required number of measurements agrees with the standard theory of compressed sensing. We then analytically compute the sampling rate of Corollary 4.18 for a specific scenario, in which the dictionary is formed by a concatenation of convolutions. The third example focuses on a numerical simulation in the case of 1D total variation. Lastly, we demonstrate how the circumangle can be controlled by the classical notion of coherence.

**The Standard Basis** Our first example is dedicated to showing that the result of Corollary 4.18 is consistent with the standard theory of compressed sensing when  $D = \mathbf{Id}$ . Hence, assume that we are given a sparse vector  $\mathbf{x}_0 \in \mathbb{R}^n$  with  $\mathcal{S} = \text{supp}(\mathbf{x}_0)$  and  $s = \#\mathcal{S} \leq n - 3$ . Trivially,  $\mathbf{x}_0$  is then its own, unique  $\ell^1$ -representation with respect to  $\mathbf{Id}$ . According to Lemma 4.13, the  $(s - 1)$ -dimensional lineality space of  $C = \mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{x}_0)$  is given by

$$C_L = \text{span}(\mathbf{r}_i^\pm : i \in \mathcal{S}),$$

where  $\mathbf{r}_i^\pm = s \cdot \text{sign}(x_{0,i}) \cdot \mathbf{e}_i - \text{sign}(\mathbf{x}_0)$ . For  $i \in \mathcal{S}^c$  a simple calculation shows that  $\boldsymbol{\theta}, \mathbf{r}_i^\pm \in C_L^\perp$ , where  $\boldsymbol{\theta} = -\text{sign}(\mathbf{x}_0) / \sqrt{s} \in \mathcal{S}^{n-1}$  and  $\mathbf{r}_i^\pm = \pm s \cdot \mathbf{e}_i - \text{sign}(\mathbf{x}_0)$ . Furthermore, for  $i \in \mathcal{S}^c$  it holds true that

$$\langle \boldsymbol{\theta}, \mathbf{r}_i^\pm \rangle = \sqrt{s} = (1/\sqrt{s+1}) \cdot \|\mathbf{r}_i^\pm\|,$$

so that the vectors  $\mathbf{r}_i^\pm$  generate a  $2(n - s)$ -polyhedral  $\alpha$ -cone  $C_R$  with  $\tan^2 \alpha = s$ . Hence, Corollary 4.18 states that robust recovery of  $\mathbf{x}_0$  is possible for  $m \gtrsim 2s \log(2(n - s) / \sqrt{2\pi})$  measurements. This bound is to be compared with the classical compressed sensing result, which prescribes to take  $m \gtrsim 2s \log(n/s)$  measurements. Note that the slight difference in the logarithmic factor is due to our simple bound on  $W(\alpha, k, n)$ , cf. Remark 4.10.

**A Convolutional Dictionary** Consider a dictionary  $D$  defined by the concatenation of two convolution matrices  $H_1$  and  $H_2$  with convolution kernels  $h_1 = [1, 1]$  and  $h_2 = [1, -1]$ , respectively. For instance, in dimension  $n = 4$ , this would yield the following matrix:

$$D = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \end{pmatrix}.$$

In particular for imaging applications, popular signal models are based on sparsity in such concatenations of convolutional matrices, e.g., translation invariant wavelets [Mal09] or learned filters in the convolutional sparse coding model [BEL13; Woh14]. Note that the resulting dictionary is highly redundant and correlated, so that existing coherence- and RIP-based arguments cannot provide satisfactory recovery guarantees. For the same reason, a recovery of a unique minimal  $\ell^1$ -representer by  $(\text{BP}_\eta^{\text{coef}})$  is unlikely, cf. the numerical simulation in Section 5.2. However, in the following, we will show how the previous upper bound based on the circumangle can be used in order to analyze signal recovery by  $(\text{BP}_\eta^{\text{sig}})$ .

To that end, we consider the recovery of a simple vector  $x_0 \in \mathbb{R}^n$  supported on the first and the last component only, i.e.,  $x_{0,i} = 0$  for all  $2 \leq i \leq n-1$ . A generalization to vectors supported on supports made of pairs of contiguous indices separated by pairs of contiguous zeros is doable, but we prefer this simple setting for didactic reasons. In this case, the set of minimal  $\ell^1$ -representers can be completely characterized. Assuming additionally that  $x_{0,1} > x_{0,n} > 0$ , one can show that

$$\begin{aligned} Z_{\ell^1} &= \{z_{\ell^1} = [z^{(1)}; z^{(2)}] \in \mathbb{R}^{2n}, \text{ with } \text{supp}(z^{(1)}) = \text{supp}(z^{(2)}) = \{1, 2\}, \\ & z_1^{(1)} = \frac{x_{0,1} + x_{0,n}}{2} - \delta, z_1^{(2)} = \delta, z_2^{(1)} = \frac{x_{0,1} - x_{0,n}}{2} - \delta, z_2^{(2)} = -\delta, \\ & 0 \leq \delta \leq \frac{x_{0,1} - x_{0,n}}{2}\}. \end{aligned}$$

Let  $z_{\ell^1} \in Z_{\ell^1}$  denote any representer with maximal support  $S = \text{supp}(z_{\ell^1}) = \{1, 2, n+1, n+2\}$  and set  $v = D \cdot \text{sign}(z_{\ell^1})$ . According to Proposition 4.16, we then decompose the descent cone  $C = \mathcal{D}_\wedge(p_{D, B_1^d}, x_0)$  into  $C = C_L \oplus C_R$ , where  $C_L$  is the lineality space given by

$$C_L = \text{span}(4 \cdot \text{sign}(z_{\ell^1, i}) \cdot d_i - v : i \in S),$$

and the range is given by  $C_R = \text{cone}(P_{C_L^\perp}(\pm 4 \cdot d_j - v) : j \in S^c)$ . It is easy to see that  $\dim(C_L) = 2$ , and that the projection on  $C_L^\perp$  can be expressed as

$$(P_{C_L^\perp}(x))_i = \begin{cases} 0, & \text{if } i \in \{2, n\}, \\ x_i, & \text{otherwise.} \end{cases}$$

The goal is now to show that  $C_R$  is contained in a circular cone with angle  $\alpha = \arccos(1/\sqrt{3})$  and axis  $\theta = -P_{C_L^\perp}(v) / \|P_{C_L^\perp}(v)\|_2 = -e_1$ . Indeed, a straightforward computation shows that for  $j \in S^c$  we have

$$\left( P_{C_L^\perp}(\pm 4 \cdot d_j - v) / \|P_{C_L^\perp}(\pm 4 \cdot d_j - v)\|_2 \right)_1 \in \left\{ -1/\sqrt{2}, -1/\sqrt{3} \right\}.$$

Hence, Corollary 4.18 implies that robust recovery of  $x_0$  is possible for  $m \gtrsim 2 + 2 \log(4n)$  measurements.

**1D Total Variation** As a third example we consider the problem of total variation minimization in 1D. Assume that  $x_0, A, y, e$  and  $\eta$  follow Model 1.1 with  $\eta = 0$  and that  $A$  obeys Model 1.2. Total variation minimization is based on the assumption that  $x_0$  is gradient-sparse, i.e., that  $\#\text{supp}(\nabla x_0) \leq s \ll n$ , where  $\nabla \in \mathbb{R}^{(n-1) \times n}$  denotes a discrete gradient operator, which is for instance based on forward differences with von Neumann boundary conditions. In order to recover

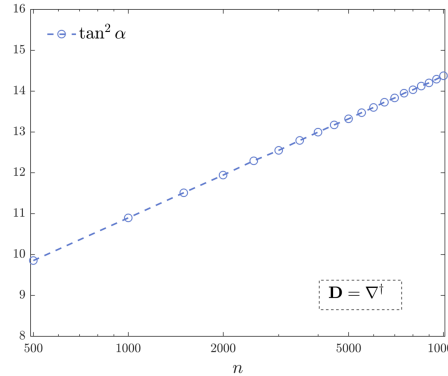


Figure 5: **Logarithmic scaling of  $\tan^2 \alpha$ .** The figure displays the behavior of  $\tan^2 \alpha$  for increasing ambient dimension  $n$ , where  $\alpha$  denotes the circumangle of the range  $C_R$  in Proposition 4.9. Here, the dictionary is chosen as  $D = \nabla^\dagger$  and the considered signals  $x_0$  have equidistant jump discontinuities and zero average. The plot indicates that the upper bound based on the circumangle is sharp enough to break the square-root bottleneck of [CX15].

$x_0$  from its noiseless, compressed measurements  $y$ , one solves the program

$$\min_{x \in \mathbb{R}^n} \|\nabla x\|_1 \quad \text{s.t.} \quad y = Ax.$$

For signals with  $\mathbf{1}^T \cdot x_0 = 0$  it is easy to see that the previous formulation is equivalent to solving the synthesis basis pursuit ( $\text{BP}_{\eta=0}^{\text{sig}}$ ) with  $D = \nabla^\dagger$ , where  $\nabla^\dagger \in \mathbb{R}^{n \times n-1}$  denotes the Moore-Penrose inverse of  $\nabla$ .

The research of the past three decades demonstrates that encouraging a small total variation norm often efficaciously reflects the inherent structure of real-world signals. Although not as popular as its counterpart in 2D, total variation methods in one spatial dimension find application in many practical applications, see for instance [LJ11]. Somewhat surprisingly, Cai and Xu have shown that a *uniform* recovery of all  $s$ -gradient-sparse signals is possible if and only if the number of (Gaussian) measurements obeys  $m \gtrsim \sqrt{sn} \cdot \log(n)$ ; see [CX15]. Recently, [GMS20] has proven that this square-root bottleneck can be broken for signals with well separated jump discontinuities. This result is also based on establishing a non-trivial upper bound on the conic mean width. For such “natural” signals,  $m \gtrsim s \cdot \log^2(n)$  measurements are already sufficient for exact recovery. See also [DHL17; GLCS20] for closely related results in a denoising context.

We want to demonstrate that the upper bound on the conic mean width based on the circumangle is capable of breaking the square-root bottleneck of the synthesis-based reformulation above. A theoretical analysis appears to be doable, however, it is beyond the scope of this work. Instead, we restrict ourselves to a simple numerical simulation. We consider signals that are defined by the pointwise discretization of a function on an interval with a few equidistant discontinuities and zero average. Note that for such a signal  $x_0$  the unique minimal  $\ell^1$ -representer with respect to  $\nabla^\dagger$  is simply given by  $\nabla x_0$ . Hence, we are only left with numerically computing the circumangle  $\alpha$  of the range  $C_R$  in Proposition 4.16, which is done by means of Proposition 4.7. In order to confirm that required number of measurements scales logarithmically in the ambient dimension  $n$ , we analyze the behavior of  $\tan^2 \alpha$  when the resolution is increased, i.e., for  $n = 500, 1000, \dots, 10000$ . The result is shown in Figure 5. The logarithmic scaling of  $\tan^2 \alpha$  (note that the  $n$ -axis is logarithmic) indeed suggests that the bound of Corollary 4.18 predicts that  $m \gtrsim s \cdot \log^2(n)$  measurements suffice for exact recovery. Hence, the presented upper bound based on the circumangle appears to be sharp enough to break the square-root bottleneck of total variation minimization in 1D.

**Coherence and Circumangle** In our last example, we show that the circumangle of  $C_R$  of Proposition 4.16 can be controlled in terms of the *mutual coherence* of the dictionary (see Equation (1.2)). This notion is a classical concept in the literature on sparse representations, which is frequently used to derive uniform recovery statements; see for instance [FR13, Chapter 5] for an overview.

Note that the assumption  $s < \frac{1}{2}(1 + \mu^{-1})$  of the following result guarantees that every  $s$ -sparse  $\mathbf{z}_{\ell^1}$  is the unique minimal  $\ell^1$ -representer of its associated signal  $\mathbf{D}\mathbf{z}_{\ell^1}$  [DE03; GN03]. Hence, in this situation, coefficient and signal recovery are equivalent, and both formulations are governed by the conic mean width of the cone  $C = \mathcal{D}_{\wedge}(p_{\mathbf{D} \cdot \mathbf{B}_1^d}, \mathbf{D}\mathbf{z}_{\ell^1}) = \mathbf{D} \cdot \mathcal{D}_{\wedge}(\|\cdot\|_1, \mathbf{z}_{\ell^1})$ .

**Proposition 4.20** *Let  $\mathbf{D} \in \mathbb{R}^{n \times d}$  be a dictionary that spans  $\mathbb{R}^n$  with  $\|\mathbf{d}_i\|_2 \leq 1$  for  $i \in [d]$  and mutual coherence  $\mu = \mu(\mathbf{D})$ . Let  $\mathbf{z}_{\ell^1} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$  denote an arbitrary  $s$ -sparse vector with  $s < \frac{1}{2}(1 + \mu^{-1})$ . Then the circumangle  $\alpha$  of the range  $C_R$  of the descent cone  $C = \mathcal{D}_{\wedge}(p_{\mathbf{D} \cdot \mathbf{B}_1^d}, \mathbf{D}\mathbf{z}_{\ell^1})$  obeys:*

$$\tan^2 \alpha \leq \frac{s(1 - s\mu)}{(1 - 2s\mu)^2}.$$

A proof of the previous result can be found in Appendix C.5. The statement allows us to retrieve a bound of the order  $m \gtrsim s \log(d)$  for the needed number of measurements. For example, with  $s\mu = 1/10$ , the bound of Corollary 4.18 results in a sampling rate of  $m_0 \leq 4s \log(d)$ . Observe that this is essentially the same result as [RSV08, Corollary II.4], however, the constants of Proposition 4.20 are better controlled.

Note that the mutual coherence of a dictionary (sometimes also referred to as *worst-case coherence* [CK13, Chapter 9]) is a global quantity that is usually used to derive recovery guarantees that are uniform across all  $s$ -sparse signals. Approaches based on this notion suffer from the so-called square-root bottleneck: The Welch bound [FR13, Theorem 5.7] reveals that the condition  $s < \frac{1}{2}(1 + \mu^{-1})$  can only be satisfied for mild sparsity values  $s \lesssim \sqrt{n}$ . We emphasize that this is in contrast to the strategy of this work, which is tailored for a non-uniform recovery of individual signals. Indeed, the circumangle is a signal-dependent notion that allows to describe the local geometry.

## 5 Numerical Experiments

In this section, we illustrate our main findings regarding the  $\ell^1$ -synthesis formulation by performing numerical simulations. First, we study the recovery of coefficient representations by  $(\text{BP}_{\eta=0}^{\text{coef}})$ ; see Section 5.1. In Section 5.2, we then focus on signal recovery by  $(\text{BP}_{\eta=0}^{\text{sig}})$  in situations, where the identification of a coefficient representation is impossible. Section 5.3 is dedicated to the experiment of Figure 1, in which both formulations are compared. Lastly, we investigate the differences concerning robustness to measurement noise in Section 5.4. For general design principles and more details on our simulations we refer the interested reader to Appendix D.

To the best of our knowledge, all other compressed sensing results on the  $\ell^1$ -synthesis formulation with redundant dictionaries describe the sampling rate as an asymptotic order bound. Hence, these results are not compatible with the experiments in this section and will not be further considered.

### 5.1 Sampling Rates for Coefficient Recovery

In order to study coefficient recovery by  $(\text{BP}_{\eta=0}^{\text{coef}})$ , we create *phase transition plots* by running Experiment 5.1 for different dictionary and signal combinations reported below.

**Experiment 5.1 (Phase transition for a fixed coefficient vector)**

**Input:** Dictionary  $D \in \mathbb{R}^{n \times d}$ , coefficient vector  $z_{\ell^1} \in \mathbb{R}^d$ .

**Compute:** Repeat the following procedure 100 times for every  $m = 1, 2, \dots, n$ :

- ▶ Draw a standard i.i.d. Gaussian random matrix  $A \in \mathbb{R}^{m \times n}$  and determine the measurement vector  $y = ADz_{\ell^1}$ .
- ▶ Solve the program  $(\text{BP}_{\eta=0}^{\text{coef}})$  to obtain an estimator  $\hat{z} \in \mathbb{R}^d$ .
- ▶ Compute and store the recovery error  $\|z_{\ell^1} - \hat{z}\|_2$ . Declare success if  $\|z_{\ell^1} - \hat{z}\|_2 < 10^{-5}$ .

**Simulation Settings** Our first two examples are based on a redundant Haar wavelet frame, which can be seen as a typical representation system in the field of applied harmonic analysis, see [Mal09] for more details on wavelets and Section 3.1 in [GKM20] for a short discussion in the context of compressed sensing. As a back-end for defining the wavelet transform, we are using the Matlab software package `spot` [BF13], which is in turn based on the Rice Wavelet Toolbox [BCN+17]. We set the ambient dimension to  $n = 256$  and consider a Haar system with 3 decomposition levels and normalized atoms. The resulting dictionary is denoted by  $D = D_{\text{Haar}} \in \mathbb{R}^{256 \times 1024}$ . The first coefficient vector  $z_{\ell^1}^1 \in \mathbb{R}^{1024}$  is obtained by selecting a random support set of cardinality  $s = 16$ , together with random coefficients; see Subfigure 6(c) for a visualization of  $z_{\ell^1}^1$  and Subfigure 6(b) for the resulting signal  $x_1 = D_{\text{Haar}} \cdot z_{\ell^1}^1$ . The second coefficient vector  $z_{\ell^1}^2$  is created by defining two contiguous blocks of non-zero coefficients in the low frequency part, again with  $s = 16$ ; see Subfigure 6(f) for a plot of  $z_{\ell^1}^2$  and Subfigure 6(e) for the resulting signal  $x_2 = D_{\text{Haar}} \cdot z_{\ell^1}^2$ . For each signal we run Experiment 5.1 and report the empirical success rate in the Subfigures 6(a), 6(d), respectively.

In the third example, the dictionary is chosen as a Gaussian random matrix, which is a typical benchmark system for compressed sensing with redundant frames, see for instance [CWW14; GKM20; KR15]. Also in this case, we set  $n = 256$ , but we choose  $d = 512$ . The resulting dictionary is denoted by  $D_{\text{rand}} \in \mathbb{R}^{256 \times 512}$ . The coefficient vector  $z_{\ell^1}^3$  is defined in the same manner as  $z_{\ell^1}^1$  above (see Subfigure 6(i)), where we again have  $\|z_{\ell^1}^3\|_0 = 16$ . The resulting signal  $x_3$  is shown in Subfigure 6(h) and the empirical success rate in Subfigure 6(g).

Our fourth and last dictionary is inspired by super-resolution; see for instance [CF14]. We again set  $n = 256$  and choose the dictionary  $D_{\text{super}} \in \mathbb{R}^{256 \times 256}$  as a convolution with a discrete Gaussian function of large variance. This example can therefore be considered as a finely discretized super-resolution problem. The coefficient vector  $z_{\ell^1}^4$  is then chosen as a sparse vector with  $z_{\ell^1,128}^4 = 1$  and  $z_{\ell^1,129}^4 = -1$ , see Subfigure 6(l). Hence, in the signal  $x_4$ , the two neighboring peaks almost cancel out and result in the low amplitude signal shown in Subfigure 6(k). Finally, the empirical success rate is depicted in Subfigure 6(j). Note that for each example we have verified the condition  $\lambda_{\min}(D; \mathcal{D}_{\wedge}(\|\cdot\|_1, z_{\ell^1}^i)) > 0$  heuristically by verifying that  $Z_{\ell^1} = \{z_{\ell^1}^i\}$ , respectively.

**Results** Let us now analyze the empirical success rates of Figure 6 and compare them with the estimates of  $w_{\wedge}^2(D \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}^i))$  and  $w_{\wedge}^2(\mathcal{D}(\|\cdot\|_1; z_{\ell^1}^i))$ . Our findings are summarized in the following:

- (i) The convex program  $(\text{BP}_{\eta=0}^{\text{coef}})$  obeys a sharp phase transition in the number of measurements  $m$ : Recovery of a coefficient vector fails if  $m$  is below a certain threshold and succeeds with overwhelming probability once a small transition region is surpassed. This observation could have been anticipated, given for instance the works [ALMT14; Tro15]. However, note that the product structure of the matrix  $AD$  in  $(\text{BP}_{\eta=0}^{\text{coef}})$  does not allow for a direct application of these results.
- (ii) The quantity  $w_{\wedge}^2(D \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}^i))$  accurately describes the sampling rate of  $(\text{BP}_{\eta=0}^{\text{coef}})$ . Indeed, in all four simulation settings of Figure 6, the phase transition occurs near the estimated conic mean width of  $D \cdot \mathcal{D}_{\wedge}(\|\cdot\|_1; z_{\ell^1}^i)$ , as predicted by Theorem 3.6.



- (iii) In contrast,  $w_\lambda^2(\mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$  does not describe the sampling rate of  $(\text{BP}_{\eta=0}^{\text{coef}})$ , in general. Indeed,  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}^2)) \ll w_\lambda^2(\mathcal{D}(\|\cdot\|_1; z_{\ell^1}^2))$  and  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}^4)) \gg w_\lambda^2(\mathcal{D}(\|\cdot\|_1; z_{\ell^1}^4))$ ; see Subfigure 6(d) and 6(j), respectively.
- (iv) For two minimal  $\ell^1$ -representations  $z_{\ell^1}^1, z_{\ell^1}^2$  with the same sparsity, but with different supports, the quantities  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}^1))$  and  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}^2))$  might differ significantly, while  $w_\lambda^2(\mathcal{D}(\|\cdot\|_1; z_{\ell^1}^1)) = w_\lambda^2(\mathcal{D}(\|\cdot\|_1; z_{\ell^1}^2))$ ; see Subfigures 6(a) and 6(d). Hence, sparsity alone does not appear to be a good proxy for the sampling complexity of  $(\text{BP}_{\eta=0}^{\text{coef}})$ . A refined understanding of coefficient recovery requires a theory that is non-uniform across the class of all  $s$ -sparse signals.
- (v) The local condition number  $\kappa_{\mathbf{D}, z_{\ell^1}}$  might explode, which often renders a condition bound as in Proposition 4.3 unusable. Indeed, we report upper bounds for  $\lambda_{\min}(\mathbf{D}; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}^i))$  in the first column of Figure 6. Since the norms of each dictionary are well controlled, this quantity is responsible for the large values of the local condition number.

## 5.2 Sampling Rates for Signal Recovery

For the investigation of signal recovery via  $(\text{BP}_{\eta=0}^{\text{sig}})$ , we also create phase transition plots by running Experiment 5.1 for different dictionary and signal combinations. Note that we also compute and store the signal error  $\|x_i - \hat{x}\|_2 = \|\mathbf{D}z_{\ell^1}^i - \mathbf{D}\hat{z}\|_2$  in the third step of the experiment (in addition to  $\|z_{\ell^1}^i - \hat{z}\|_2$ ). Recovery is declared successful if  $\|x_i - \hat{x}\|_2 < 10^{-5}$ .

**Simulation Settings** Our first two examples are based on the same Haar wavelet system with 3 decomposition levels that is used in Section 5.1. The first signal is constructed by defining a coefficient vector  $z_1 \in \mathbb{R}^{1024}$  with a random support set of cardinality  $s = 35$  and random coefficients; see Subfigure 7(j) for a visualization of  $z_1$  and Subfigure 7(d) for the resulting signal  $x_1 = \mathbf{D}_{\text{Haar}} \cdot z_1$ . In order to apply the result of Theorem 3.8, we compute a minimal  $\ell^1$ -representer  $z_{\ell^1}^1 \in Z_{\ell^1}$  of  $x_1$ ; see Subfigure 7(g). The second coefficient vector  $z_2$  is created by defining two contiguous blocks of non-zero coefficients in a lower frequency decomposition scale, again with  $s = 35$ ; see Subfigure 7(k) for a plot of  $z_2$  and Subfigure 7(e) for the resulting signal  $x_2 = \mathbf{D}_{\text{Haar}} \cdot z_2$ . A corresponding minimal  $\ell^1$ -representer  $z_{\ell^1}^2 \in Z_{\ell^1}$  of  $x_2$  is shown in Subfigure 7(h).

Finally, for the third setup we are choosing a simple example in 2D. In order to keep the computational burden manageable, we restrict ourselves to a  $28 \times 28$ -dimensional digit from the MNIST data set [LBBH98], i.e., the vectorized image is of size  $n = 28^2 = 784$ . As a sparsifying system we utilize a dictionary that is based on the 2D discrete cosine transform (dct-2). It makes use of Matlab's standard dct-2 transform as convolution filters on  $3 \times 3$  patches. The resulting operator is denoted by  $\mathbf{D} = \mathbf{D}_{\text{dct-2}} \in \mathbb{R}^{n \times 9n}$ , i.e.,  $d = 9n$ . Note that such a convolution sparsity model is frequently used in the literature, in particular also with learned filters, e.g., see *convolutional sparse coding* in [BEL13]. Although the dct-2 filters might not be a perfect match for MNIST digits, we consider them as a classical representative that is well suited to demonstrate the predictive power of our results. In order to construct a suitable sparse representation  $z_3 \in \mathbb{R}^d$  of an arbitrarily picked digit in the database, we make use of the orthogonal matching pursuit algorithm [PRK93]; see Subfigure 7(l) for a visualization of  $z_3$ . The resulting digit  $x_3 = \mathbf{D}_{\text{dct-2}} \cdot z_3$  is displayed in Figure 7(f). A minimal  $\ell^1$ -representer  $z_{\ell^1}^3$ , which is needed to apply Theorem 3.8, is shown in Subfigure 7(i).

**Results** Our conclusions on the numerical experiments shown in Figure 7 are reported in the following:

- (vi) The convex program  $(\text{BP}_{\eta=0}^{\text{sig}})$  obeys a sharp phase transition in the number of measurements. Due to the equivalent, gauge-based reformulation (3.1) of  $(\text{BP}_{\eta=0}^{\text{sig}})$ , this observation is predicted by the works [ALMT14; Tro15]. However, a recovery of a coefficient representation via solving  $(\text{BP}_{\eta=0}^{\text{coef}})$  is impossible in all three examples, even for  $m = n$ .

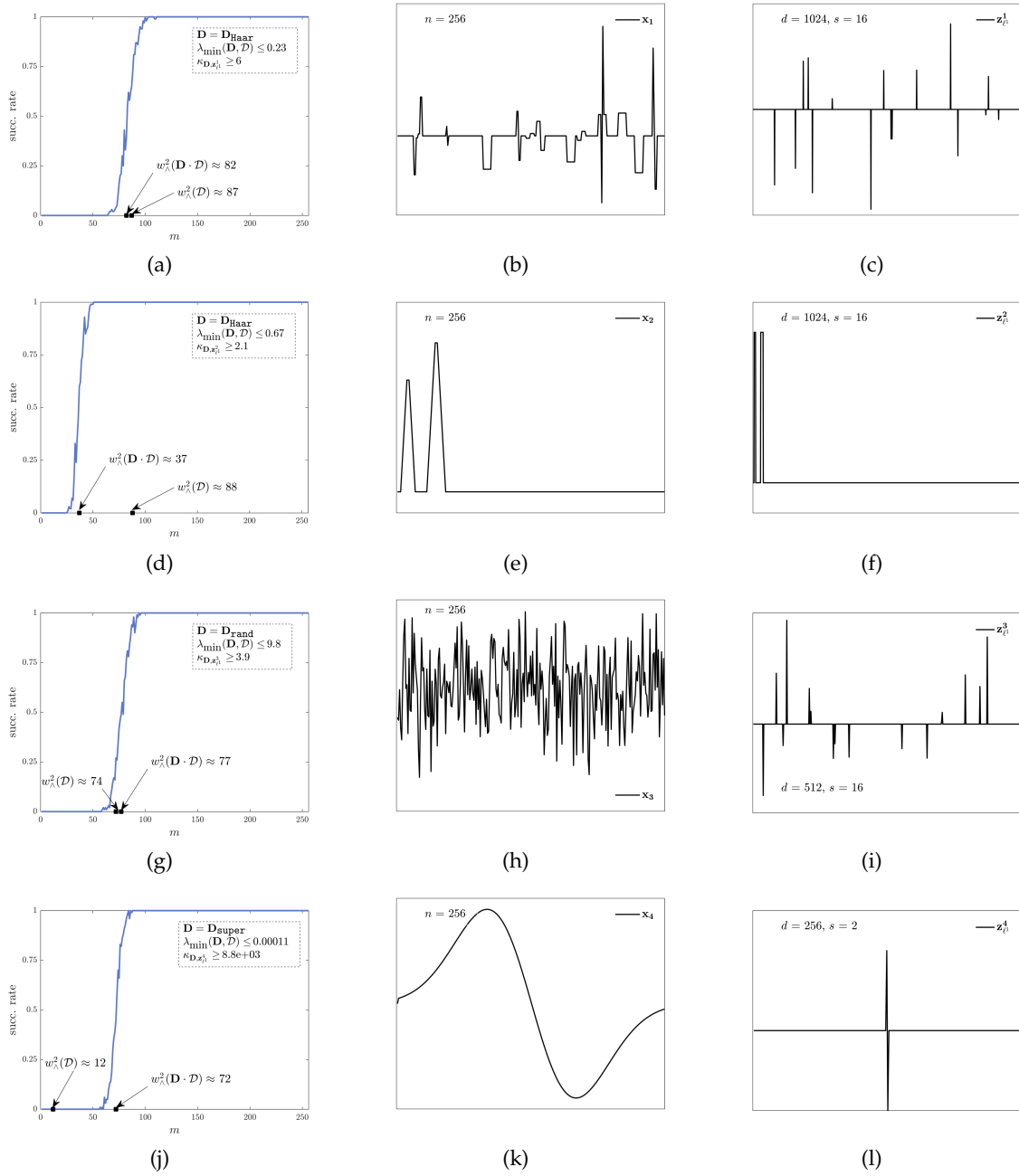


Figure 6: Phase transitions of coefficient recovery by solving  $(\text{BP}_{\eta=0}^{\text{coef}})$ . Empirical success rates and other key figures are reported in the first column, where we use the notation  $\mathcal{D} = \mathcal{D}(\|\cdot\|_1, z_{\ell^1}^i)$ . The coefficient vectors  $z_{\ell^1}^i$  that are used in each experiment are shown in the third column. The associated signal vectors  $x_i = \mathcal{D}z_{\ell^1}^i$  are displayed in the second column. The first two rows are relying on a redundant Haar wavelet frame, the third row is based on a Gaussian random matrix and the last row is using a dictionary inspired by super-resolution.

- (vii) For any  $z_{\ell^1} \in Z_{\ell^1}$ , the conic mean width  $w_{\lambda}^2(D \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$  accurately describes the sampling rate of  $(\text{BP}_{\eta=0}^{\text{sig}})$ , as predicted by Theorem 3.8. Indeed, in all three cases of Figure 7, the estimated  $w_{\lambda}^2(D \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$  matches precisely the 50% recovery rate.
- (viii) In contrast, for any other sparse representation  $z \notin Z_{\ell^1}$ , the conic width  $w_{\lambda}^2(D \cdot \mathcal{D}(\|\cdot\|_1; z))$  does not describe the sampling rate of  $(\text{BP}_{\eta=0}^{\text{sig}})$ , in general. Indeed, observe that we have  $w_{\lambda}^2(D \cdot \mathcal{D}(\|\cdot\|_1; z_i)) \approx n$  in all three examples. Also note that  $\|z_1\|_0 = 35 = \|z_2\|_0$ , however,

the locations of the phase transitions deviate considerably. Similarly, although  $\|z_{\ell^1}^1\|_0 < \|z_{\ell^1}^2\|_0$ , we have that  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}^1)) > w_\lambda^2(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}^2))$ . This observation is yet another indication that sparsity alone is not a good proxy for the sampling rate of  $\ell^1$ -synthesis, in general.

### 5.3 Creating a “Full” Phase Transition

Let us now focus on the phase transition plots shown in Figure 1. Up to now, we have only considered one specific signal at a time. However, it is also of interest to assess the quality of our results if the “complexity” of the underlying signals is varied. In the classical situation of  $\mathbf{D}$  being an ONB, the location of the phase transition is entirely determined by the sparsity of the underlying signal. Hence, it is a natural choice to create phase transitions over the sparsity, as it is for instance done in [ALMT14]. Recalling Claims (iv) and (viii), it might appear odd to do the same in the case of a redundant dictionary. However, as the result of Figure 1 shows, if the support is chosen uniformly at random, sparsity is still a somewhat reasonable proxy for the sampling rate. Indeed, these plots are created by running Experiment 5.2 with  $\mathbf{D} = \mathbf{D}_{\text{Haar}} \in \mathbb{R}^{256 \times 1024}$ , maximal sparsity  $s_0 = 125$  and displaying the empirical success rates of coefficient and signal recovery, respectively. Additionally the dotted line shows the averaged conic mean width values  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$ .

#### Experiment 5.2 (Phase transition of Figure 1)

**Input:** Dictionary  $\mathbf{D} \in \mathbb{R}^{n \times d}$ , maximal sparsity  $s_0 \in [n]$ .

**Compute:** Repeat the following procedure 500 times for each  $s \in \{1, \dots, s_0\}$ :

- ▶ Select a set  $S \subset [n]$  uniformly at random with  $\#S = s$ . Then draw a standard Gaussian random vector  $c \in \mathbb{R}^s$  and define  $z_0$  by setting  $(z_0)_S = c$  and  $(z_0)_{S^c} = \mathbf{0}$ .
- ▶ Define the signal  $x_0 = \mathbf{D}z_0$  and compute a minimal  $\ell^1$ -representation  $z_{\ell^1} \in Z_{\ell^1}$  of  $x_0$ . Compute the conic mean width  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$ .
- ▶ Run Experiment 5.1 with  $\mathbf{D}$  and  $z_{\ell^1}$  as input, where the number of repetitions is lowered to 5. In the third step, coefficient/signal recovery is declared successful if  $\|z_{\ell^1} - \hat{z}\|_2 < 10^{-5}$  or if  $\|\mathbf{D}z_{\ell^1} - \hat{x}\|_2 = \|x_0 - \mathbf{D}\hat{z}\|_2 < 10^{-5}$ , respectively.

First note, that the averaged values of the conic mean width perfectly match the center of the phase transition of  $(\text{BP}_{\eta=0}^{\text{sig}})$  in Subfigure 1(b), as it is predicted by Theorem 3.8. However, observe that for sparsity values between  $s \approx 20$  and  $s \approx 80$  the transition region is spread out in the vertical direction, cf. [ALMT14]. This phenomenon can be related to Claim (viii): Given that sparsity alone is not a good proxy for the sample complexity of  $(\text{BP}_{\eta=0}^{\text{sig}})$ , averaging over different instances with the same sparsity necessarily results in a smeared out transition area.

Regarding the phase transition in Subfigure 1(a), we observe that the location of the phase transition is also determined by  $w_\lambda^2(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$ ; provided that coefficient recovery is possible, i.e., if  $\lambda_{\min}(\mathbf{D}; \mathcal{D}_\wedge(\|\cdot\|_1; z_{\ell^1})) > 0$ . For sparsity values  $s \geq 50$ , recovery of coefficients appears to be impossible, whereas signal and coefficient recovery seems to be equivalent for very small sparsity values (i.e.,  $s \leq 5$ ). The interval in between forms a transition region, in which coefficient recovery becomes gradually less likely. We suspect that with more repetitions in Experiment 5.2, the empirical success rates on this interval would eventually smooth out.

We conclude that:

- (ix) Whether  $\lambda_{\min}(\mathbf{D}; \mathcal{D}_\wedge(\|\cdot\|_1; z_{\ell^1})) \neq 0$ , i.e., whether coefficient recovery is possible, is again a property that is non-uniform in the sparsity of  $z_{\ell^1}$ .

### 5.4 Robustness to Noise

The purpose of this last numerical simulation is to analyze coefficient and signal recovery with respect to robustness to measurement noise. To that end, we run Experiment 5.3 for different

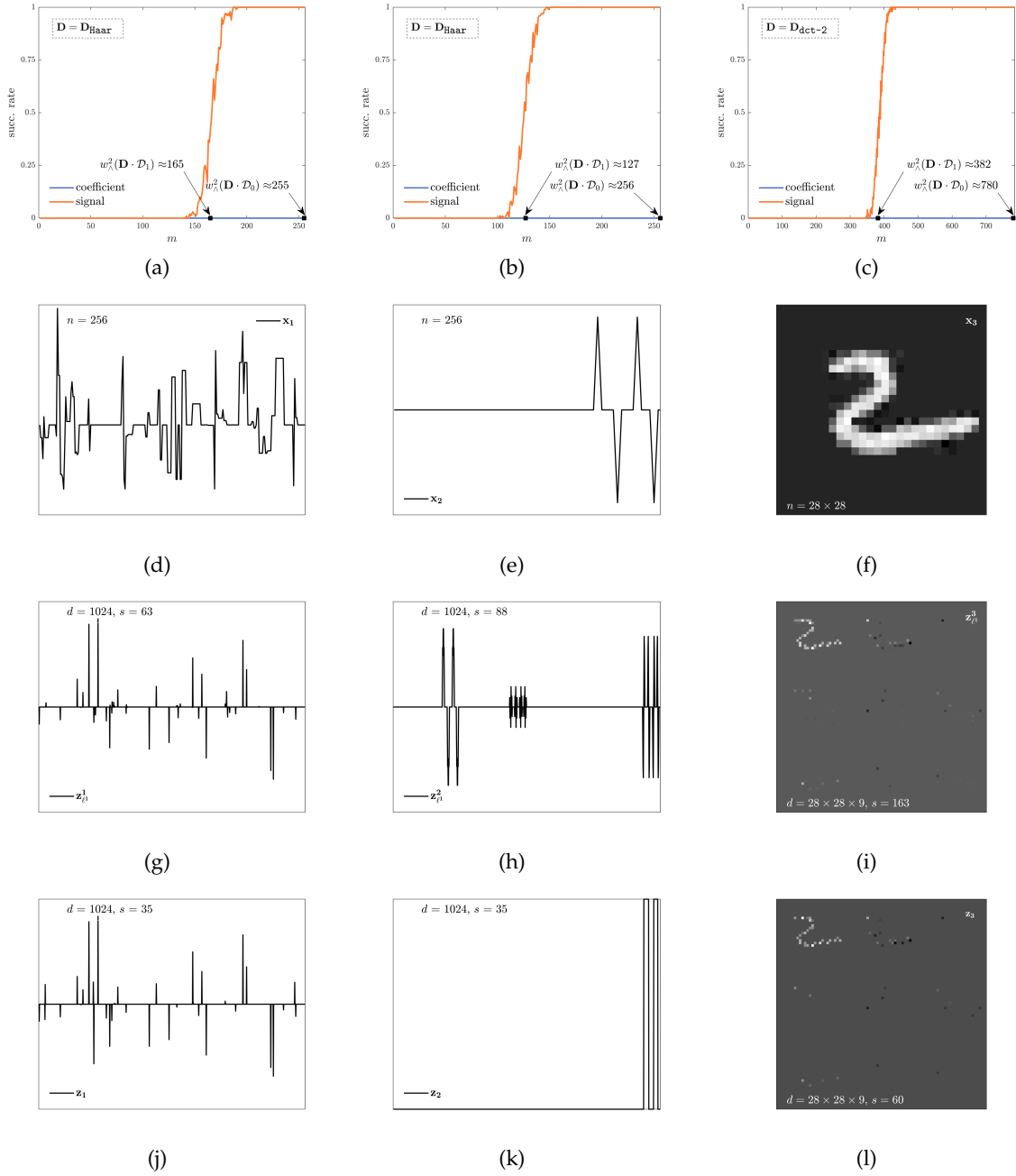


Figure 7: Phase transitions of signal recovery by solving  $(\text{BP}_{\eta=0}^{\text{sig}})$ . Empirical success rates and other key figures are reported in the first row, where we use the notation  $\mathcal{D}_1 = \mathcal{D}(\|\cdot\|_1, z_{\ell^1}^i)$  and  $\mathcal{D}_0 = \mathcal{D}(\|\cdot\|_1, z_i)$ . The signals  $x_i$  that are used in each experiment are shown in the second row. The associated minimal  $\ell^1$ -representers  $z_{\ell^1}^i$  are displayed in the third row and the original coefficient representations  $z_i$  are shown in the fourth row. The first two columns are relying on a redundant Haar wavelet frame and the third column is based on the dct-2. Note that in all three examples, coefficient recovery is not possible since  $\lambda_{\min}(\mathbf{D}; \mathcal{D}_{\wedge}(\|\cdot\|_1, z_{\ell^1}^i)) = 0$ .

setups, which are reported below.

### Experiment 5.3 (Robustness to Measurement Noise)

**Input:** Dictionary  $D \in \mathbb{R}^{n \times d}$ , number of measurements  $m$ , minimal  $\ell^1$ -representation  $z_{\ell^1}$  of the signal  $x_0 \in \mathbb{R}^n$ , range of noise levels  $H$ .

**Compute:** Repeat the following procedure 100 times for every  $\eta \in H$ :

- ▶ Draw a standard i.i.d. Gaussian random matrix  $A \in \mathbb{R}^{m \times n}$  and determine the noisy measurement vector  $y = Ax_0 + \eta \cdot e$ , where  $\|e\|_2 = 1$ .
- ▶ Solve the program  $(\text{BP}_\eta^{\text{coef}})$  to obtain an estimator  $\hat{z} \in \mathbb{R}^d$ .
- ▶ Compute and store the recovery errors  $\|z_{\ell^1} - \hat{z}\|_2$  and  $\|x_0 - \hat{x}\|_2 = \|Dz_{\ell^1} - D\hat{z}\|_2$ .

**Simulation settings** First, we choose the same dictionary and signal combination as in Section 5.1 and restrict the noise level to  $H = \{0, 0.05, 0.1, 0.15, \dots, 1\}$ . Furthermore, we consider the 1D examples of Section 5.2, together with the noise range  $H = \{0, 0.005, 0.01, \dots, 0.1\}$ . Recall that the difference of these two setups is that  $\lambda_{\min}(D; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}^i)) > 0$  in the first case, whereas  $\lambda_{\min}(D; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}^i)) = 0$  in the second case. In all experiments, we roughly pick the number of measurements as  $m \approx w_\wedge^2(D \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}^i)) + 40$  to ensure that Theorem 3.6 (or Theorem 3.8, respectively) is applicable. The averaged coefficient and signal recovery errors are displayed in Figure 8, together with the theoretical upper bound on the signal error of Equation (3.5). Note that it is not possible to show the corresponding error bound for coefficient recovery. In the first set of examples, we do not have access to  $\lambda_{\min}(D; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}^i))$  and in the last two cases,  $\lambda_{\min}(D; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}^i)) = 0$  and therefore Theorem 3.6 is not applicable. Nevertheless, it is possible to obtain an upper bound for the latter quantity, as outlined in the Appendix D. If  $D = D_{\text{rand}}$ , it is additionally possible to use the result on minimum conic singular values of Gaussian matrices to get a lower bound with high probability [Tro15, Proposition 3.3].

**Results** We summarize the findings of the results shown in Figure 8 below:

- (x) If the number of measurements exceeds the sampling rate in Theorem 3.8, signal recovery via solving the Program  $(\text{BP}_\eta^{\text{sig}})$  is robust to measurement noise. This phenomenon holds true without any further assumptions. Indeed, observe that in all simulations of Figure 8, the signal error  $\|x_0 - \hat{x}\|$  lies below its theoretical upper bound of Equation (3.5).
- (xi) If  $z_{\ell^1}$  is the unique minimal  $\ell^1$ -representation of  $x_0$  (i.e., if  $\lambda_{\min}(D; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1})) > 0$ ) and if the number of measurements exceeds the sampling rate in Theorem 3.6, it is possible to robustly recovery  $z_{\ell^1}$ . However, in contrast to signal recovery, the robustness is influenced by the “stability” of the minimal  $\ell^1$ -representation of  $z_{\ell^1}$  in  $D$ , i.e., by the value of  $\lambda_{\min}(D; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}))$  in the error bound (3.3). Indeed, it is possible that the signal  $x_0$  is more robustly recovered than its coefficients  $z_{\ell^1}$ , or vice versa. This can be seen by comparing coefficient and signal recovery in the Subfigures 8(a)-8(d).<sup>1</sup> If  $\lambda_{\min}(D; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1})) \ll 1$ , coefficient recovery is less robust than signal recovery, see Subfigures 8(a), 8(b) and 8(d). However, if  $\lambda_{\min}(D; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1})) \gg 1$ , the contrary holds true, see Subfigure 8(c).

## Acknowledgments

Acknowledgements: M.M. is supported by the DFG Priority Programme DFG-SPP 1798. He wishes to thank Martin Genzel for inspiring discussions. C.B. have been supported by a PEPS « Jeunes chercheuses et jeunes chercheurs » funding in 2017 and 2018 for this work. P.W. and J.K. are supported by the ANR JCJC OMS.

<sup>1</sup>Note that the quantity  $\lambda_{\min}(D_{\text{super}}; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}^i))$  is very small. Hence, even for a small amount of noise the error of Equation (3.3) explodes. For  $\eta > 0.1$  the error stays roughly constant since the solution  $\hat{z}$  of  $(\text{BP}_\eta^{\text{coef}})$  is always close to 0.

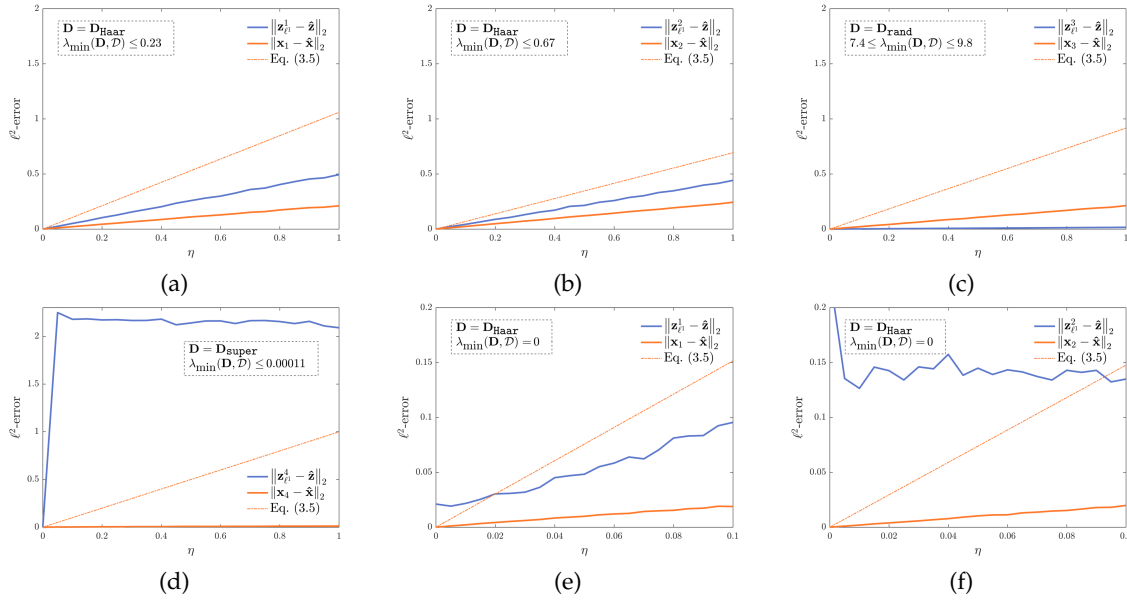


Figure 8: **Robustness to measurement noise.** We display the reconstruction errors for a recovery from noisy measurements with an increasing noise level. The first four Subfigures are based on the examples for coefficient recovery of Section 5.1, and the last two on the examples based on Haar wavelet of Section 5.2. We use the notation  $\mathcal{D} = \mathcal{D}(\|\cdot\|_1, z_{\ell^1}^i)$ , where  $z_{\ell^1}^i \in Z_{\ell^1}$ .

## 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* 3.3 (2014), 224–294.
- [ALW20] D. Amelunxen, M. Lotz, and J. Walvin. “Effective condition number bounds for convex regularization”. *IEEE Trans. Inf. Theory* (2020). in press.
- [BCN+17] R. Baraniuk, H. Choi, R. Neelamani, et al. *Rice Wavelet Toolbox, Version 3*. URL: <https://github.com/ricedsp/rwt>. Aug. 2017.
- [BEL13] H. Bristow, A. Eriksson, and S. Lucey. “Fast Convolutional Sparse Coding”. *2013 IEEE Conference on Computer Vision and Pattern Recognition*. 2013.
- [Beu38] A. Beurling. “Sur les intégrales de Fourier absolument convergentes et leur application à une transformation fonctionnelle”. *Ninth Scandinavian Mathematical Congress*. 1938, 345–366.
- [BF13] E. van den Berg and M. P. Friedlander. *Spot – A Linear-Operator Toolbox*. URL: <http://www.cs.ubc.ca/labs/scl/spot/index.html>. Aug. 2013.
- [BM02] P. L. Bartlett and S. Mendelson. “Rademacher and Gaussian complexities: Risk bounds and structural results”. *J. Mach. Learn. Res.* 3 (2002), 463–482.
- [CCL19] P. G. Casazza, X. Chen, and R. G. Lynch. “Preserving Injectivity under Subgaussian Mappings and Its Application to Compressed Sensing”. arXiv preprint: 1710.09972. 2019.
- [CDS98] S. Chen, D. Donoho, and M. Saunders. “Atomic Decomposition by Basis Pursuit”. *SIAM J. Sci. Comput.* 20.1 (1998), 33–61.
- [CENR11] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall. “Compressed sensing with coherent and redundant dictionaries”. *Appl. Comput. Harmon. Anal.* 31.1 (2011), 59–73.
- [CF14] E. J. Candès and C. Fernandez-Granda. “Towards a Mathematical Theory of Super-resolution”. *Commun. Pur. Appl. Math.* 67.6 (2014), 906–956.
- [CK13] P. G. Casazza and G. Kutyniok, eds. *Finite Frames: Theory and Applications*. Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.
- [CM73] J. F. Claerbout and F. Muir. “Robust Modeling With Erratic Data”. *Geophysics* 38.5 (1973), 826–844.
- [CR06] E. J. Candès and J. Romberg. “Quantitative robust uncertainty principles and optimally sparse decompositions”. *Found. Comput. Math.* 6.2 (2006), 227–254.
- [CRPW12] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. “The convex geometry of linear inverse problems”. *Found. Comput. Math.* 12.6 (2012), 805–849.
- [CRT06a] E. J. Candès, J. Romberg, and T. Tao. “Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information”. *IEEE Trans. Inf. Theor.* 52.2 (2006), 489–509.



- [CRT06b] E. J. Candès, J. K. Romberg, and T. Tao. “Stable signal recovery from incomplete and inaccurate measurements”. *Comm. Pure Appl. Math.* 59.8 (2006), 1207–1223.
- [CT06] E. J. Candès and T. Tao. “Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?” *IEEE Trans. Inf. Theor.* 52.12 (2006), 5406–5425.
- [CWW14] X. Chen, H. Wang, and R. Wang. “A null space analysis of the  $\ell_1$ -synthesis method in dictionary-based compressed sensing”. *Appl. Comput. Harmon. Anal.* 37.3 (2014), 492–515.
- [CX15] J.-F. Cai and W. Xu. “Guarantees of total variation minimization for signal recovery”. *Inf. Inference* 4.4 (2015), 328–353.
- [DE03] D. L. Donoho and M. Elad. “Optimally sparse representation in general (nonorthogonal) dictionaries via  $\ell^1$  minimization”. *Proc. Natl. Acad. Sci.* 100.5 (2003), 2197–2202.
- [DH01] D. L. Donoho and X. Huo. “Uncertainty principles and ideal atomic decomposition”. *IEEE Trans. Inf. Theory* 47.7 (2001), 2845–2862.
- [DHL17] A. S. Dalalyan, M. Hebiri, and J. Lederer. “On the prediction performance of the Lasso”. *Bernoulli* 23.1 (Feb. 2017), 552–581.
- [DNW13] M. A. Davenport, D. Needell, and M. B. Wakin. “Signal Space CoSaMP for Sparse Recovery With Redundant Dictionaries”. *IEEE Trans. Inf. Theory* 59.10 (2013), 6820–6829.
- [Don06] D. L. Donoho. “Compressed sensing”. *IEEE Trans. Inf. Theory* 52.4 (2006), 1289–1306.
- [EB02] M. Elad and A. M. Bruckstein. “A Generalized Uncertainty Principle and Sparse Representation in Pairs of Bases”. *IEEE Trans. Inf. Theory* 48.9 (2002).
- [EMR07] M. Elad, P. Milanfar, and R. Rubinstein. “Analysis versus synthesis in signal priors”. *Inverse Probl.* 23.3 (2007), 947–968.
- [FR13] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.
- [Fuc04] J.-J. Fuchs. “On sparse representations in arbitrary redundant bases”. *IEEE Trans. Inf. Theory* 50.6 (2004), 1341–1344.
- [Fuc05] J.-J. Fuchs. “Recovery of exact sparse representations in the presence of bounded noise”. *IEEE Trans. Inf. Theory* 51.10 (2005), 3601–3608.
- [FV99] R. M. Freund and J. R. Vera. “Condition-Based Complexity of Convex Optimization in Conic Linear Form via the Ellipsoid Algorithm”. *SIAM J. Optim.* 10.1 (1999), 155–176.
- [GB08] M. Grant and S. Boyd. “Graph implementations for nonsmooth convex programs”. *Recent Advances in Learning and Control*. Ed. by V. Blondel, S. Boyd, and H. Kimura. Vol. 371. Lecture Notes in Control and Information Sciences. Springer London, 2008, 95–110.
- [GB14] M. Grant and S. Boyd. *CVX: Matlab Software for Disciplined Convex Programming, version 2.1*. URL: <http://cvxr.com/cvx>. Mar. 2014.
- [GKM20] M. Genzel, G. Kutyniok, and M. März. “ $\ell^1$ -Analysis Minimization and Generalized (Co-)Sparsity: When Does Recovery Succeed?” *Appl. Comput. Harmon. Anal.* (2020). Accepted, arXiv preprint: 1710.04952.
- [GLCS20] A. Guntuboyina, D. Lieu, S. Chatterjee, and B. Sen. “Adaptive risk bounds in univariate total variation denoising and trend filtering”. *Ann. Statist.* 48.1 (Feb. 2020), 205–229.
- [GM04] A. A. Giannopoulos and V. D. Milman. “Asymptotic Convex Geometry Short Overview”. *Different Faces of Geometry*. Ed. by S. Donaldson, Y. Eliashberg, and M. Gromov. Springer, 2004, 87–162.
- [GMS20] M. Genzel, M. März, and R. Seidel. “Compressed Sensing with 1D Total Variation: Breaking Sample Complexity Barriers via Non-Uniform Recovery”. arXiv preprint: 2001.09952. (2020).
- [GN03] R. Gribonval and M. Nielsen. “Sparse representations in unions of bases”. *IEEE Trans. Inf. Theory* 49.12 (2003), 3320–3325.
- [GNEGD14] R. Giryes, S. Nam, M. Elad, R. Gribonval, and M. E. Davies. “Greedy-like algorithms for the cosparsity analysis model”. *Linear Algebra Appl.* 441 (2014), 22–60.
- [Gor85] Y. Gordon. “Some inequalities for Gaussian processes and applications”. *Isr. J. Math.* 50.4 (1985), 265–289.
- [Gor88] Y. Gordon. “On Milman’s inequality and random subspaces which escape through a mesh in  $\mathbb{R}^n$ ”. *Geometric aspects of functional analysis*. Ed. by J. Lindenstrauss and V. D. Milman. Vol. 1317. Lecture Notes in Mathematics. Springer, 1988, 84–106.
- [HS10a] R. Henrion and A. Seeger. “On Properties of Different Notions of Centers for Convex Cones”. *Set-Valued Anal.* 18 (2010), 205–231.
- [HS10b] R. Henrion and A. Seeger. “Inradius and Circumradius of Various Convex Cones Arising in Applications”. *Set-Valued Anal.* 18 (2010), 483–511.
- [HS10c] J.-B. Hiriart-Urruty and A. Seeger. “A variational approach to copositive matrices”. *SIAM Rev.* 52.4 (2010), 593–629.
- [IS08] A. Iusem and A. Seeger. “Normality and modulability indices. Part I: Convex cones in normed spaces”. *J. Math. Anal. Appl.* 338.1 (2008), 365–391.

- [KNW15] F. Krahmer, D. Needell, and R. Ward. “Compressive Sensing with Redundant Dictionaries and Structured Measurements”. *SIAM J. Math. Anal.* 47.6 (2015), 4606–4629.
- [KR15] M. Kabanava and H. Rauhut. “Analysis  $\ell_1$ -recovery with Frames and Gaussian Measurements”. *Acta Appl. Math.* 140.1 (2015), 173–195.
- [Kre38] M. Kreĭn. “The L-problem in an abstract normed linear space”. *Some questions in the theory of moments*. Ed. by I. Ahiezer and M. Kreĭn. English Transl. Amer. Math. Soc., Providence, R.I., 1962. MR 29 # 5073. Gos. Naučno-Tehn. Izdat. Ukraine, 1938. Chap. 4.
- [KRZ15] M. Kabanava, H. Rauhut, and H. Zhang. “Robust analysis  $\ell_1$ -recovery from Gaussian measurements and total variation minimization”. *Eur. J. Appl. Math.* 26.6 (2015), 917–929.
- [LBBH98] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. “Gradient-based learning applied to document recognition”. *Proc. IEEE* 86.11 (1998), 2278–2324.
- [LJ11] M. A. Little and N. S. Jones. “Generalized methods and solvers for noise removal from piecewise constant signals. I. Background theory”. *Proc. Royal Soc. Lond. A* 467.2135 (2011), 3088–3114.
- [LLMLY12] Y. Liu, S. Li, T. Mi, H. Lei, and W. Yu. “Performance analysis of  $\ell_1$ -synthesis with coherent frames”. *2012 IEEE International Symposium on Information Theory Proceedings*. July 2012, 2042–2046.
- [LMPV17] C. Liaw, A. Mehrabian, Y. Plan, and R. Vershynin. “A Simple Tool for Bounding the Deviation of Random Matrices on Geometric Sets”. *Geometric Aspects of Functional Analysis*. Ed. by B. Klartag and E. Milman. Vol. 2169. Lecture Notes in Mathematics. Springer, 2017, 277–299.
- [Log65] B. Logan. “Properties of High-Pass Signals”. PhD thesis. New York: Columbia University, 1965.
- [Mal09] S. Mallat. *Cover image A Wavelet Tour of Signal Processing: The Sparse Way*. 3rd Edition. Elsevier, 2009.
- [Mil85] V. D. Milman. “Random subspaces of proportional dimension of finite dimensional normed spaces: Approach through the isoperimetric inequality”. *Banach Spaces*. Ed. by N. J. Kalton and E. Saab. Vol. 1166. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 1985, 106–115.
- [MK87] K. G. Murty and S. N. Kabadi. “Some NP-complete problems in quadratic and nonlinear programming”. *Math. Program.* 39.2 (1987), 117–129.
- [MPT07] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. “Reconstruction and subgaussian operators in asymptotic geometric analysis”. *Geom. Funct. Anal.* 17.4 (2007), 1248–1282.
- [MZ93] S. G. Mallat and Zhifeng Zhang. “Matching pursuits with time-frequency dictionaries”. *IEEE Trans. Signal Process.* 41.12 (1993), 3397–3415.
- [NDEG13] S. Nam, M. E. Davies, M. Elad, and R. Gribonval. “The cosparsity analysis model and algorithms”. *Appl. Comput. Harmon. Anal.* 34.1 (2013), 30–56.
- [PRK93] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad. “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition”. *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*. 1993, 40–44 vol.1.
- [Ren95] J. Renegar. “Linear programming, complexity theory and elementary functional analysis”. *Math. Program.* 70 (1995), 279–351.
- [Roc70] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [ROF92] L. Rudin, S. Osher, and E. Fatemi. “Nonlinear total variation based noise removal algorithms”. *Physica D* 60.1–4 (1992), 259–268.
- [RSV08] H. Rauhut, K. Schnass, and P. Vandergheynst. “Compressed Sensing and Redundant Dictionaries”. *IEEE Trans. Inf. Theory* 54.5 (2008), 2210–2219.
- [RV07] M. Rudelson and R. Vershynin. *Comm. Pure Appl. Math.* 61.8 (2007), 1025–1045.
- [SF09] I. W. Selesnick and M. A. T. Figueiredo. “Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors”. *Proceedings of SPIE, Wavelets XIII*. Ed. by V. K. Goyal, M. Papadakis, and D. V. D. Ville. Vol. 7446. 2009.
- [ST03] A. Seeger and M. Torki. “On eigenvalues induced by a cone constraint”. *Linear Algebra Appl.* 372 (2003), 181–206.
- [Sto09] M. Stojnic. “Various thresholds for  $\ell_1$ -optimization in compressed sensing”. Preprint arXiv:0907.3666. 2009.
- [Syl57] J. J. Sylvester. “A question in the geometry of situation”. *Quarterly Journal of Pure and Applied Mathematics* 1.1 (1857), 79–80.
- [Tal14] M. Talagrand. *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*. Springer, 2014.
- [TBM79] H. L. Taylor, S. C. Banks, and J. F. McCoy. “Deconvolution with the  $\ell_1$  norm”. *Geophysics* 44.1 (1979), 39–52.
- [Tro04] J. A. Tropp. “Greed is good: algorithmic results for sparse approximation”. *IEEE Trans. Inf. Theory* 50.10 (2004), 2231–2242.
- [Tro06] J. A. Tropp. “Just relax: Convex programming methods for identifying sparse signals in noise”. *IEEE Trans. Inf. Theory* 52.3 (2006), 1030–1051.

- [Tro15] J. A. Tropp. “Convex Recovery of a Structured Signal from Independent Random Linear Measurements”. *Sampling Theory, a Renaissance*. Ed. by G. E. Pfander. Applied and Numerical Harmonic Analysis. Birkhäuser, 2015, 67–101.
- [Ver12] R. Vershynin. “Introduction to the non-asymptotic analysis of random matrices”. *Compressed Sensing Theory and Applications*. Ed. by Y. C. Eldar and G. Kutyniok. Cambridge University Press, 2012, 210–268.
- [Ver15] R. Vershynin. “Estimation in High Dimensions: A Geometric Perspective”. *Sampling Theory, a Renaissance: Compressive Sensing and Other Developments*. Ed. by G. E. Pfander. Cham: Springer International Publishing, 2015, 3–66.
- [Ver18] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.
- [Woh14] B. Wohlberg. “Efficient convolutional sparse coding”. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2014, 7173–7177.
- [Zuh48] S. Zuhovickii. “Remarks on problems in approximation theory”. *Mat. Zbirnik KDU* (1948). (Ukrainian), 169–183.

## A Proofs of Section 3

### A.1 Proof of Lemma 3.2 (Minimal $\ell^1$ -Representers)

(b) “ $\Rightarrow$ ”: Assume that  $\hat{X} = \{x_0\}$ . First, we show that  $Z_{\ell^1} \subseteq \hat{Z}$ : Let  $z_{\ell^1} \in Z_{\ell^1}$ . Since  $Dz_{\ell^1} = x_0$ ,  $z_{\ell^1}$  is in the feasible set of  $(\text{BP}_{\eta=0}^{\text{coef}})$ . Furthermore, due to  $\hat{X} = \{x_0\}$  it holds true that  $D\hat{z} = x_0$  for each  $\hat{z} \in \hat{Z}$ . Thus also  $\|z_{\ell^1}\|_1 \leq \|\hat{z}\|_1$  and therefore  $z_{\ell^1} \in \hat{Z}$ . Secondly, we show  $\hat{Z} \subseteq Z_{\ell^1}$ . For a  $\hat{z} \in \hat{Z}$  it holds true that  $D\hat{z} = x_0$  since we have assumed that  $\hat{X} = \{x_0\}$ . Thus,  $\hat{z}$  is also feasible for  $Z_{\ell^1}$ . Since each  $z_{\ell^1} \in Z_{\ell^1}$  is in turn feasible for  $\hat{Z}$ , we obtain that  $\|\hat{z}\|_1 \leq \|z_{\ell^1}\|_1$  and therefore  $\hat{z} \in Z_{\ell^1}$ .

“ $\Leftarrow$ ”: If  $\hat{Z} = Z_{\ell^1}$ , then  $\hat{X} = D \cdot \hat{Z} = D \cdot Z_{\ell^1} = \{x_0\}$ .

(a) If  $\hat{Z} = \{z_0\}$ , then also  $\hat{X} = D \cdot \hat{Z} = \{x_0\}$  and (b) would imply  $Z_{\ell^1} = \hat{Z} = \{z_0\}$ .

### A.2 Proof of Lemma 3.4 (The Gauge Formulation)

By definition,

$$\begin{aligned} \hat{X} &= D \cdot \left( \underset{z \in \mathbb{R}^d}{\operatorname{argmin}} \|z\|_1 \quad \text{s.t.} \quad \|y - ADz\|_2 \leq \eta \right) \\ &= D \cdot \left( \underset{z \in \mathbb{R}^d}{\operatorname{argmin}} \inf \{ \lambda > 0 : z \in \lambda \cdot B_1^d \} \quad \text{s.t.} \quad \|y - ADz\|_2 \leq \eta \right) \\ &= \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \inf \{ \lambda > 0 : x \in \lambda \cdot D \cdot B_1^d \} \quad \text{s.t.} \quad \|y - Ax\|_2 \leq \eta. \end{aligned}$$

### A.3 Proof of Lemma 3.5 (Descent Cone of the Gauge)

We will only prove the first equality and note that the other one follows essentially the same argumentation. Pick any  $z_{\ell^1} \in Z_{\ell^1}$  and note that  $p_{D \cdot B_1^d}(x_0) = \|z_{\ell^1}\|_1$ .

“ $\supseteq$ ”: Let  $h \in \mathcal{D}_{\wedge}(\|\cdot\|_1, z_{\ell^1})$ , i.e., there exists a  $\tau > 0$  such that  $\|z_{\ell^1} + \tau h\|_1 \leq \|z_{\ell^1}\|_1$ . Hence,

$$p_{D \cdot B_1^d}(Dz_{\ell^1} + \tau Dh) = p_{D \cdot B_1^d}(D \cdot (z_{\ell^1} + \tau h)) \leq \|z_{\ell^1} + \tau \cdot h\|_1 \leq \|z_{\ell^1}\|_1 = p_{D \cdot B_1^d}(x_0),$$

and therefore  $Dh \in \mathcal{D}_{\wedge}(p_{D \cdot B_1^d}, x_0)$ .

“ $\subseteq$ ”: Let  $x \in \mathcal{D}_{\wedge}(p_{D \cdot B_1^d}, x_0)$ , i.e., there exists  $\tau > 0$  such that

$$R := p_{D \cdot B_1^d}(x_0 + \tau x) \leq p_{D \cdot B_1^d}(x_0) = \|z_{\ell^1}\|_1.$$

Now, choose  $h \in B_1^d$  such that  $R \cdot Dh = x_0 + \tau x$  and write  $x = D \cdot (R/\tau \cdot h - 1/\tau \cdot z_{\ell^1}) =: D(\bar{z})$ . Observe that

$$\|z_{\ell^1} + \tau \bar{z}\|_1 = R \cdot \|h\|_1 \leq R \leq \|z_{\ell^1}\|_1,$$

and therefore  $\bar{z} \in \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1})$ .

**Remark A.1** The proof of “ $\subseteq$ ” shows that  $z_{\ell^1}$  could be replaced by any other  $z_0$  with  $x_0 = Dz_0$ , which is not necessarily a minimal  $\ell^1$ -representer of  $x_0$ . Hence,  $\mathcal{D}_\wedge(p_{D \cdot B_1^d}, x_0) \subseteq D \cdot \mathcal{D}_\wedge(\|\cdot\|_1, z_0)$  and  $\mathcal{D}(p_{D \cdot B_1^d}, x_0) \subseteq D \cdot \mathcal{D}(\|\cdot\|_1, z_0)$  for any  $z_0$  with  $x_0 = Dz_0$ , with inequality if  $z_0 \in Z_{\ell^1}$ .  $\diamond$

#### A.4 Proof of Theorem 3.6 (Coefficient Recovery)

Recalling Proposition 2.3, the goal of the proof is to find a lower bound for the minimum conic singular value  $\lambda_{\min}(\Phi; C) = \inf\{\|\Phi z\|_2 : z \in C \cap \mathcal{S}^{d-1}\}$ , where we use the abbreviated notation  $C := \mathcal{D}_\wedge(\|\cdot\|_1; z_{\ell^1})$  and  $\Phi := AD$ . Note that by assumption  $\|Dz\|_2 \geq \lambda_{\min}(D; C) > 0$  for all  $z \in C \cap \mathcal{S}^{d-1}$ . Thus, we obtain

$$\begin{aligned} \lambda_{\min}(\Phi; C) &= \inf\left\{\frac{\|ADz\|_2}{\|Dz\|_2} \|Dz\|_2 : z \in C \cap \mathcal{S}^{d-1}\right\} \\ &\geq \lambda_{\min}(D; C) \cdot \inf\{\|Ax\|_2 : x \in DC \cap \mathcal{S}^{n-1}\}. \end{aligned} \quad (\text{A.1})$$

Theorem 2.5 now implies that there is a numerical constant  $c > 0$  such that with probability at least  $1 - e^{-u^2/2}$ , we have

$$\inf\{\|Ax\|_2 : x \in DC \cap \mathcal{S}^{n-1}\} > \sqrt{m-1} - c \cdot \gamma^2 \cdot (w_\wedge(DC) + u).$$

Thus, with probability at least  $1 - e^{-u^2/2}$ , we conclude from the previous steps that

$$\lambda_{\min}(\Phi; C) > \lambda_{\min}(D; C) \cdot \left(\sqrt{m-1} - \sqrt{m_0-1}\right).$$

The claim of the theorem is then a direct consequence of Proposition 2.3.

**Remark A.2** The above argumentation may be compared with [CCL19, Theorem 3.1]. We also show that if  $D$  is bounded away from 0 on the intersection  $S$  of a closed convex cone  $C$  and the sphere, then  $AD$  also stays away from 0 on  $S$  with high probability. However, an important difference is that our result does not involve  $\lambda_{\min}(D; C)^{-1}$  as a multiplicative factor in the rate  $m_0 \approx w_\wedge^2(D \cdot C)$ . It therefore allows for a tight description of the sampling rate in the case of noiseless Gaussian measurements; cf. the discussion subsequent to Theorem 2.5. Indeed, the numerical experiments of Section 5.1 reveal that  $\lambda_{\min}(D; \mathcal{D}_\wedge(\|\cdot\|_1; z_{\ell^1}))$  may be very small, while  $(\text{BP}_{\eta=0}^{\text{coef}})$  still allows for exact recovery from  $m \approx w_\wedge^2(D \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$  measurements.  $\diamond$

#### A.5 Proof of Proposition 3.10 (Stable Recovery)

Let  $z^* \in \mathbb{R}^d$  be chosen according to (3.6), i.e., it satisfies  $\|x_0 - Dz^*\|_2 \leq \varepsilon$  and  $\|z^*\|_1 = \|z_0\|_1$ , where  $z_0$  is any vector with  $x_0 = Dz_0$ . The goal is to invoke [GKM20, Theorem 6.4] with

$$t = \max\left\{r \cdot \|x_0 - Dz^*\|_2, \frac{2\eta}{\sqrt{m-1} - c \cdot \gamma^2 \cdot \left(\frac{r+1}{r} \cdot (w_\wedge(D \cdot \mathcal{D}(\|\cdot\|_1; z^*)) + 1) + u\right)}\right\}.$$

Thus, we need to verify that  $t$  satisfies

$$t \geq \frac{2\eta}{\sqrt{m-1} - c \cdot \gamma^2 \cdot \left(w_t(\mathcal{D}(p_{D \cdot B_1^d}; x_0)) + u\right)},$$

where  $w_t$  denotes the *local mean width* at scale  $t > 0$ ; see [GKM20, Definition 6.1] for details on this notion. To that end, we first observe that

$$w_t(\mathcal{D}(p_{D \cdot B_1^d}; x_0)) \leq w_t(D \cdot \mathcal{D}(\|\cdot\|_1; z_0) + x_0 - x_0)$$

$$t \geq r \cdot \|x_0 - Dz^*\|_2 \leq w_{r, \|x_0 - Dz^*\|_2}((D \cdot \mathcal{D}(\|\cdot\|_1; z_0) + x_0) - x_0),$$

where we have used that  $\mathcal{D}(p_{D \cdot B_1^d}; x_0) \subseteq D \cdot \mathcal{D}(\|\cdot\|_1; z_0)$  for any  $z_0$  with  $x_0 = Dz_0$  in the first step (see Remark A.1), and the monotonicity of the local mean width in the second step. Observe that we have  $Dz^* \in D \cdot \mathcal{D}(\|\cdot\|_1; z_0) + x_0$ , due to the assumption  $\|z^*\|_1 = \|z_0\|_1$ . Hence, we can make use of [GKM20, Lemma A.2] for  $K = D \cdot \mathcal{D}(\|\cdot\|_1; z_0) + x_0$  in order to obtain

$$\begin{aligned} w_{r, \|x_0 - Dz^*\|_2}((D \cdot \mathcal{D}(\|\cdot\|_1; z_0) + x_0) - x_0) &\leq \frac{r+1}{r} \cdot (w_\wedge((D \cdot \mathcal{D}(\|\cdot\|_1; z_0) + x_0) - Dz^*) + 1) \\ &= \frac{r+1}{r} \cdot (w_\wedge(D \cdot \mathcal{D}(\|\cdot\|_1; z^*)) + 1), \end{aligned}$$

where the equality follows from:

$$\begin{aligned} D \cdot \mathcal{D}(\|\cdot\|_1; z_0) + x_0 - Dz^* &= D \cdot \{h \in \mathbb{R}^d : \|z_0 + h\|_1 \leq \underbrace{\|z_0\|_1}_{=\|z^*\|_1}\} + x_0 - Dz^* \\ &\stackrel{h'=h+z_0-z^*}{=} D \cdot \{h' \in \mathbb{R}^d : \|z^* + h'\|_1 \leq \|z^*\|_1\} = D \cdot \mathcal{D}(\|\cdot\|_1; z^*). \end{aligned}$$

We conclude that

$$\begin{aligned} t &\geq \frac{2\eta}{\sqrt{m-1} - c \cdot \gamma^2 \cdot \left(\frac{r+1}{r} \cdot (w_\wedge(D \cdot \mathcal{D}(\|\cdot\|_1; z^*)) + 1) + u\right)} \\ &\geq \frac{2\eta}{\sqrt{m-1} - c \cdot \gamma^2 \cdot (w_t(\mathcal{D}(p_{D \cdot B_1^d}; x_0)) + u)}. \end{aligned}$$

Hence, [GKM20, Theorem 6.4] then implies that any minimizer of  $(BP_\eta^{\text{sig}})$  satisfies  $\|x_0 - \hat{x}\|_2 \leq t$ .<sup>1</sup>

## B Proof of Proposition 4.3 (Width and Condition Number)

Let us start with a preliminary lemma, which generalizes Proposition 10.2 in [ALMT14].

**Lemma B.1** *For a closed convex cone  $C \subseteq \mathbb{R}^d$ , a dictionary  $D \in \mathbb{R}^{n \times d}$  and a standard Gaussian vector  $g \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ , we have that*

$$\mathbb{E} \left[ \left( \sup_{z \in C \cap B_2^d} \langle g, Dz \rangle \right)^2 \right] \leq \left( \mathbb{E} \left[ \sup_{z \in C \cap S^{d-1}} \langle g, Dz \rangle \right] \right)^2 + \lambda_{\max}^2(D; C).$$

*Proof.* Define the random variable  $Z = Z(g) := \sup_{z \in C \cap S^{d-1}} \langle g, Dz \rangle$ . In a first step, we prove that

$$\mathbb{E} \left[ \left( \sup_{z \in C \cap B_2^d} \langle g, Dz \rangle \right)^2 \right] \leq \mathbb{E} [Z^2]. \quad (\text{B.1})$$

Indeed, since  $Z^2$  is a nonnegative random variable, we obtain

$$\mathbb{E}[Z^2] \geq \mathbb{E} \left[ Z^2 \cdot \mathbf{1}_{\mathbb{R}^d \setminus C^\circ}(D^*g) \right] = \mathbb{E} \left[ \left( \sup_{z \in C \cap S^{d-1}} \langle g, Dz \rangle \right)^2 \cdot \mathbf{1}_{\mathbb{R}^d \setminus C^\circ}(D^*g) \right],$$

<sup>1</sup>This argument does actually not cover the case of  $\eta = 0$ , but here we can simply use that  $t = r \cdot \|x_0 - Dz^*\|_1 > 0$  if  $x_0 \neq Dz^*$ .

where  $C^\circ$  denotes the *polar cone* of  $C$ . Furthermore, it holds true that

$$\mathbb{E} \left[ \left( \sup_{z \in C \cap \mathcal{S}^{d-1}} \langle \mathbf{g}, D\mathbf{z} \rangle \right)^2 \cdot \mathbf{1}_{\mathbb{R}^d \setminus C^\circ}(D^* \mathbf{g}) \right] = \mathbb{E} \left[ \left( \sup_{z \in C \cap \mathbb{B}_2^d} \langle \mathbf{g}, D\mathbf{z} \rangle \right)^2 \right].$$

Indeed, for an  $\mathbf{x} \in \mathbb{R}^n$  such that  $D^* \mathbf{x} \notin C^\circ$  the equality  $\sup_{z \in C \cap \mathcal{S}^{d-1}} \langle \mathbf{x}, D\mathbf{z} \rangle = \sup_{z \in C \cap \mathbb{B}_2^d} \langle \mathbf{x}, D\mathbf{z} \rangle$  holds true, because the supremum over the ball occurs at a vector of length 1. On the other hand, when  $D^* \mathbf{x} \in C^\circ$ , one has  $\sup_{z \in C \cap \mathbb{B}_2^d} \langle \mathbf{x}, D\mathbf{z} \rangle = 0$ . Therefore, (B.1) is established.

Moreover, observe that the function  $\mathbf{g} \mapsto Z(\mathbf{g})$  is  $\lambda_{\max}(\mathbf{D}, C)$ -Lipschitz. Indeed, for  $\mathbf{f}, \mathbf{g} \in \mathbb{R}^n$  and  $\mathbf{z} \in C \cap \mathcal{S}^{d-1}$  we obtain that

$$\begin{aligned} \langle \mathbf{g}, D\mathbf{z} \rangle &= \langle \mathbf{f}, D\mathbf{z} \rangle + \langle \mathbf{g}, D\mathbf{z} \rangle - \langle \mathbf{f}, D\mathbf{z} \rangle \leq \langle \mathbf{f}, D\mathbf{z} \rangle + \|\mathbf{f} - \mathbf{g}\|_2 \|D\mathbf{z}\|_2 \\ &\leq \langle \mathbf{f}, D\mathbf{z} \rangle + \lambda_{\max}(\mathbf{D}, C) \|\mathbf{f} - \mathbf{g}\|_2, \end{aligned}$$

and therefore by taking the supremum

$$\sup_{z \in C \cap \mathcal{S}^{d-1}} \langle \mathbf{g}, D\mathbf{z} \rangle \leq \sup_{z \in C \cap \mathcal{S}^{d-1}} \langle \mathbf{f}, D\mathbf{z} \rangle + \lambda_{\max}(\mathbf{D}, C) \|\mathbf{f} - \mathbf{g}\|_2.$$

By swapping the roles of  $\mathbf{f}$  and  $\mathbf{g}$ , an analogue estimate can be obtained, which verifies the claimed Lipschitz continuity. Thus, the fluctuation of  $Z$  can be bounded as follows:

$$\mathbb{E} [Z^2] - \mathbb{E} [Z]^2 = \mathbb{E} [(Z - \mathbb{E}Z)^2] = \text{Var}(Z) \leq \lambda_{\max}^2(\mathbf{D}, C),$$

where the last estimate follows from Fact C.3 in [ALMT14]. ■

**Back to the proof of Proposition 4.3.** In order to prove Proposition 4.3, we continue as follows: First observe that

$$w_\lambda^2(\mathbf{D} \cdot C) = w^2(\mathbf{D} \cdot C \cap \mathcal{S}^{n-1}) \leq \delta(\mathbf{D} \cdot C) = \mathbb{E} \left[ \left( \sup_{x \in \mathbf{D} \cdot C \cap \mathbb{B}_2^d} \langle \mathbf{g}, x \rangle \right)^2 \right],$$

where  $\delta$  denotes the *statistical dimension*<sup>1</sup>. Next, it is straightforward to see that

$$\mathbf{D} \cdot C \cap \mathbb{B}_2^n \subseteq \frac{\mathbf{D}}{\lambda_{\min}(\mathbf{D}; C)} \cdot (C \cap \mathbb{B}_2^d),$$

which immediately implies that

$$\delta(\mathbf{D} \cdot C) \leq \frac{1}{\lambda_{\min}^2(\mathbf{D}; C)} \cdot \mathbb{E} \left[ \left( \sup_{x \in \mathbf{D}(C \cap \mathbb{B}_2^d)} \langle \mathbf{g}, x \rangle \right)^2 \right].$$

Exercise 7.5.4 in [Ver18] and Lemma B.1 now allow us to derive the desired bound:

$$\begin{aligned} w_\lambda^2(\mathbf{D} \cdot C) &\leq \delta(\mathbf{D} \cdot C) \leq \frac{1}{\lambda_{\min}^2(\mathbf{D}; C)} \mathbb{E} \left[ \left( \sup_{x \in \mathbf{D}(C \cap \mathbb{B}_2^d)} \langle \mathbf{g}, x \rangle \right)^2 \right] \\ &\stackrel{\text{(B.1)}}{\leq} \frac{1}{\lambda_{\min}^2(\mathbf{D}; C)} \left( \mathbb{E} \left[ \sup_{x \in \mathbf{D}(C \cap \mathcal{S}^{d-1})} \langle \mathbf{g}, x \rangle \right]^2 + \lambda_{\max}^2(\mathbf{D}; C) \right) \end{aligned}$$

<sup>1</sup>The statistical dimension of a convex cone  $C \subseteq \mathbb{R}^n$  can be defined as  $\delta(C) = \mathbb{E}[(\sup_{x \in C \cap \mathbb{B}_2^n} \langle \mathbf{g}, x \rangle)^2]$ ; see [ALMT14, Prop. 3.1] for details. It holds true that  $w_\lambda^2(C) \leq \delta(C) \leq w_\lambda^2(C) + 1$ , which is why both notions are often interchangeable [ALMT14, Prop. 10.2].



$$\begin{aligned}
&\stackrel{(7.5.4)}{\leq} \frac{\|\mathbf{D}\|_2^2}{\lambda_{\min}^2(\mathbf{D}; C)} \left( \left( \mathbb{E} \left[ \sup_{z \in C \cap \mathcal{S}^{d-1}} \langle \mathbf{g}, z \rangle \right] \right)^2 + 1 \right) \\
&= \kappa_{\mathbf{D}, C}^2 \cdot \left( w_\lambda^2(C) + 1 \right).
\end{aligned}$$

## C Proofs of Section 4.2

### C.1 Proof of Proposition 4.7 (Circumangle of Polyhedral Cones)

Consider a nontrivial pointed polyhedral cone  $C = \text{cone}(x_1, \dots, x_k)$  with  $\|x_i\|_2 = 1$  for  $i \in [k]$  and let  $\alpha$  denote its circumangle. Let  $\mathbf{X} := [x_1, \dots, x_k] \in \mathbb{R}^{n \times k}$ . Since  $C$  does not contain a line, its circumcenter  $\boldsymbol{\theta}$  is unique, belongs to the cone and  $0 \leq \alpha < \pi/2$ , see [HS10a]. This implies that  $\mathbf{X}^* \boldsymbol{\theta} > \mathbf{0}$  (element-wise). We have:

$$\begin{aligned}
\cos(\alpha) &= \sup_{v \in \mathcal{S}^{n-1}} \inf_{x \in C \cap \mathcal{S}^{n-1}} \langle x, v \rangle = \sup_{v \in \mathcal{S}^{n-1}} \inf_{c \geq 0, \|\mathbf{X}c\|_2=1} \langle \mathbf{X}c, v \rangle \\
&\leq \sup_{v \in \mathcal{S}^{n-1}} \inf_{i \in [k]} \langle e_i, \mathbf{X}^* v \rangle = \sup_{v \in \mathcal{S}^{n-1}} \min \mathbf{X}^* v,
\end{aligned}$$

where  $e_i$  denotes  $i$ -th standard basis vector in  $\mathbb{R}^k$  and we have used an inclusion of sets argument in the inequality. We now argue that the inequality is in fact an equality. To that end, observe that for any  $v \in \mathcal{S}^{n-1}$  such that  $\mathbf{X}^* v \geq \mathbf{0}$ , we also have that

$$\inf_{c \geq \mathbf{0}, \|\mathbf{X}c\|_2=1} \langle c, \mathbf{X}^* v \rangle \geq \inf_{c \geq \mathbf{0}, \|\mathbf{X}c\|_2 \geq 1} \langle c, \mathbf{X}^* v \rangle \geq \inf_{c \geq \mathbf{0}, \langle \mathbf{1}, c \rangle \geq 1} \langle c, \mathbf{X}^* v \rangle = \min \mathbf{X}^* v.$$

In this sequence of inequalities, we first used the inclusion of sets, the triangular inequality together with the inclusion of sets and finally the fact that a linear program attains its minimum (if it exists) on an extremal point  $\text{Ext}(\{c \geq \mathbf{0}, \langle \mathbf{1}, c \rangle \geq 1\}) = \{e_1, \dots, e_k\}$ . Note that the condition  $\mathbf{X}^* v \geq \mathbf{0}$  ensures the existence of a solution.

Finally the infimum over  $\mathcal{S}^{n-1}$  can be relaxed to  $B_2^n$ , since the supremum is attained on the boundary of the domain. This concludes the proof.

### C.2 Proof of Proposition 4.9 (Maximal Width of Polyhedral Cones)

The proof of Proposition 4.9 necessitates a basic preliminary result on the Gaussian width of general convex polytopes, which we will prove first.

#### Bounding the Gaussian Width of Convex Polytopes

**Lemma C.1** *Let  $K$  be a convex polytope with  $k \geq 5$  vertices that is contained in the unit ball of  $\mathbb{R}^n$ . Then:*

$$w(K) \leq \sqrt{2 \log \left( \frac{k}{\sqrt{2\pi}} \right)} + \frac{1}{\sqrt{2 \log \left( \frac{k}{\sqrt{2\pi}} \right)}}.$$

*Proof.* Denote by  $\mathcal{V}(K)$  the set of vertices of  $K$ . Since the maximum of the scalar products with points of a convex set is attained on a vertex, we get the following union bound for  $x > 0$  and  $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \text{Id})$ :

$$\begin{aligned}
\mathbb{P} \left[ \sup_{u \in K} \langle u, \mathbf{g} \rangle \geq x \right] &\leq \sum_{v \in \mathcal{V}(K)} \mathbb{P} [\langle v, \mathbf{g} \rangle \geq x] \\
&\leq \sum_{v \in \mathcal{V}(K)} \frac{1}{\sqrt{2\pi} \|v\|_2} \int_x^\infty \exp \left( -\frac{y^2}{2 \|v\|_2^2} \right) dy
\end{aligned}$$

$$\leq \frac{k}{\sqrt{2\pi}} \cdot \int_x^\infty \exp\left(-\frac{y^2}{2}\right) dy. \quad (\text{C.1})$$

Recall the following standard bound on the tail probability of a Gaussian, which will be used in the remainder of the proof:

$$\int_x^\infty \exp(-y^2/2) dy \leq \frac{\exp(-x^2/2)}{x}.$$

Let  $x_0 := \sqrt{2 \log\left(\frac{k}{\sqrt{2\pi}}\right)} > 1$  and note that  $\frac{k}{\sqrt{2\pi}} \exp\left(-\frac{x_0^2}{2}\right) = 1$ . We may now use the bound (C.1) in order to obtain:

$$\begin{aligned} w(K) &\leq \int_{\mathbb{R}_+} \mathbb{P} \left[ \sup_{\mathbf{u} \in K} \langle \mathbf{u}, \mathbf{g} \rangle \geq x \right] dx \\ &\leq \int_{\mathbb{R}_+} \min \left\{ 1, \frac{k}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{y^2}{2}\right) dy \right\} dx \\ &\leq x_0 + \frac{k}{\sqrt{2\pi}} \int_{x_0}^\infty \exp\left(-\frac{x^2}{2}\right) dx \\ &\leq x_0 + \frac{1}{x_0}. \end{aligned}$$

■

**Back to the proof of Proposition 4.9** Equipped with this lemma, we can now prove Proposition 4.9.

Let  $C = \text{cone}(x_1, \dots, x_k) \subset \mathbb{R}^n$  be a  $k$ -polyhedral  $\alpha$ -cone and let  $\boldsymbol{\theta} \in \mathcal{S}^{n-1}$  be an axis vector such that  $C \subseteq C(\alpha, \boldsymbol{\theta})$ . Without loss of generality assume that  $\|x_i\|_2 = 1$  for  $i \in [k]$ . Define the affine hyperplane  $\mathcal{H} := \{\mathbf{h} \in \mathbb{R}^n : \langle \mathbf{h}, \boldsymbol{\theta} \rangle = 1\}$  and let  $K := C \cap \mathcal{H}$ . Observe that  $K$  is a convex polyhedron with vertices belonging to the set  $\{x_i / \langle x_i, \boldsymbol{\theta} \rangle : i \in [k]\} \subseteq B_2^n(\cos(\alpha)^{-1})$ . Since  $\|\mathbf{k}\|_2 \geq 1$  for all  $\mathbf{k} \in K$ , any  $\mathbf{h} \in C \cap \mathcal{S}^{n-1}$  can be written as  $\mathbf{h} = \lambda(\mathbf{k}) \cdot \mathbf{k}$ , where  $\mathbf{k} \in K$  and  $0 < \lambda(\mathbf{k}) \leq 1$ . Hence, for all  $\mathbf{g} \in \mathbb{R}^n$  it holds true that

$$\sup_{\mathbf{h} \in \mathcal{S}^{n-1} \cap C} \langle \mathbf{h}, \mathbf{g} \rangle = \sup_{\mathbf{k} \in K} (\lambda(\mathbf{k}) \cdot \langle \mathbf{k}, \mathbf{g} \rangle) \leq \max \left\{ 0, \sup_{\mathbf{k} \in K} \langle \mathbf{k}, \mathbf{g} \rangle \right\}.$$

Let  $\mathbf{P}^\theta, \mathbf{P}_\perp^\theta$  denote the orthogonal projections onto  $\text{span}(\boldsymbol{\theta})$  and  $\text{span}(\boldsymbol{\theta})^\perp$ , respectively. Observe that for  $\mathbf{k} \in K$  and  $\mathbf{g} \in \mathbb{R}^n$  it holds true that

$$\langle \mathbf{k}, \mathbf{g} \rangle = \langle \boldsymbol{\theta}, \mathbf{P}^\theta(\mathbf{g}) \rangle + \langle \mathbf{P}_\perp^\theta(\mathbf{k}), \mathbf{P}_\perp^\theta(\mathbf{g}) \rangle \leq \max \left\{ 0, \langle \boldsymbol{\theta}, \mathbf{P}^\theta(\mathbf{g}) \rangle \right\} + \langle \mathbf{P}_\perp^\theta(\mathbf{k}), \mathbf{P}_\perp^\theta(\mathbf{g}) \rangle,$$

where the first equality follows from  $\mathbf{P}^\theta(\mathbf{k}) = \boldsymbol{\theta}$ . Furthermore, since  $\boldsymbol{\theta} \in C$ , we have that  $\mathbf{0} \in \mathbf{P}_\perp^\theta(K)$  and hence for all  $\mathbf{g} \in \mathbb{R}^n$ ,

$$\sup_{\mathbf{k} \in K} \langle \mathbf{P}_\perp^\theta(\mathbf{k}), \mathbf{P}_\perp^\theta(\mathbf{g}) \rangle \geq 0.$$

This allows us to conclude that

$$\sup_{\mathbf{h} \in \mathcal{S}^{n-1} \cap C} \langle \mathbf{h}, \mathbf{g} \rangle \leq \max \left\{ 0, \sup_{\mathbf{k} \in K} \langle \mathbf{k}, \mathbf{g} \rangle \right\} \leq \max \left\{ 0, \langle \boldsymbol{\theta}, \mathbf{P}^\theta(\mathbf{g}) \rangle \right\} + \sup_{\mathbf{k} \in K} \langle \mathbf{P}_\perp^\theta(\mathbf{k}), \mathbf{P}_\perp^\theta(\mathbf{g}) \rangle.$$

Hence, we obtain

$$\begin{aligned}
w_\wedge(C) &= \mathbb{E}_n \left[ \sup_{\mathbf{h} \in \mathcal{S}^{n-1} \cap C} \langle \mathbf{h}, \mathbf{g} \rangle \right] \\
&= \mathbb{E}_n \left[ \max \left\{ 0, \langle \boldsymbol{\theta}, \mathbf{P}^\theta(\mathbf{g}) \rangle \right\} + \sup_{\mathbf{k} \in K} \langle \mathbf{P}_\perp^\theta(\mathbf{k}), \mathbf{P}_\perp^\theta(\mathbf{g}) \rangle \right] \\
&= \mathbb{E}_1 \left[ \max \left\{ 0, \langle \boldsymbol{\theta}, \mathbf{P}^\theta(\mathbf{g}) \rangle \right\} \right] + \mathbb{E}_{n-1} \left[ \sup_{\mathbf{k} \in \mathbf{P}_\perp^\theta(K)} \langle \mathbf{k}, \mathbf{P}_\perp^\theta(\mathbf{g}) \rangle \right] \\
&= \frac{1}{\sqrt{2\pi}} + w(\mathbf{P}_\perp^\theta(K)), \tag{C.2}
\end{aligned}$$

where the last equality follows from  $\langle \boldsymbol{\theta}, \mathbf{P}^\theta(\mathbf{g}) \rangle \sim \mathcal{N}(0, 1)$  and the fact that  $\mathbf{P}_\perp^\theta(\mathbf{g})$  is an  $(n-1)$ -dimensional standard Gaussian vector on  $\text{span}(\boldsymbol{\theta})^\perp$ .

Since  $K \subset \mathcal{H} \cap \mathbb{B}_2^n(\cos(\alpha)^{-1})$ , its  $(n-1)$ -dimensional projection satisfies  $\mathbf{P}_\perp^\theta(K) \subset \mathbb{B}_2^{n-1}(\tan(\alpha))$ . Now, Lemma C.1 yields the following bound on the Gaussian width of a polyhedron included in a ball of radius  $\tan \alpha$  with at most  $k \geq 5$  vertices:

$$w(\mathbf{P}_\perp^\theta(K)) \leq \tan \alpha \cdot \left( \sqrt{2 \log(k/\sqrt{2\pi})} + 1/\sqrt{2 \log(k/\sqrt{2\pi})} \right).$$

The claimed inequality of Proposition 4.9 is then just a consequence of (C.2).

### C.3 Proofs of Section 4.2.2

**Descent Cone of  $\ell^1$ -Norm (Lemma 4.11)** We begin by showing a polyhedral description of the descent cone of the  $\ell^1$ -norm:

Let  $\mathbf{v}$  be any vector such that  $\|\mathbf{v}\|_1 = s$  and  $\text{sign } \mathbf{v} = \text{sign } \mathbf{z}$ . Note that  $\mathbf{v}$  and  $\mathbf{z}$  enjoy the same descent cone associated to the  $\ell^1$ -norm, which is easy to see by observing that

$$\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z}) = \left\{ \mathbf{h} \in \mathbb{R}^d : \sum_{i \in \mathcal{S}^c} |h_i| \leq - \sum_{i \in \mathcal{S}} \text{sign}(z_i) \cdot h_i \right\}.$$

Therefore, the descent set of  $\|\cdot\|_1$  at  $\mathbf{v}$  can be obtained by scaling up the cross-polytope by the factor  $\|\mathbf{v}\|_1 = s$  and shifting it by  $-\mathbf{v}$ , i.e.,

$$\mathcal{D}(\|\cdot\|_1, \mathbf{v}) = \text{conv}(\pm s \cdot \mathbf{e}_i - \mathbf{v} : i \in [d]).$$

We conclude by taking the conic hull of the previous set to obtain

$$\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z}) = \mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{v}) = \text{cone}(\pm s \cdot \mathbf{e}_i - \mathbf{v} : i \in [d]).$$

**Lineality of Descent Cone of  $\ell^1$ -Norm (Lemma 4.13)** Next, we describe the lineality space and lineality of  $\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z})$ :

The lineality space of the descent cone at point  $\mathbf{z}$  corresponds to the span of the face of the  $\ell^1$ -ball of minimal dimension containing  $\mathbf{z}$ . It can therefore be defined as the span of the vectors joining  $\mathbf{z}$  to the vertices of this face, which are exactly the vectors  $\text{sign}(z_i) \cdot \mathbf{e}_i$ .

For a more formal proof for this fact, one could argue as follows: First note that (see for instance Appendix B in [ALMT14])

$$\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z})^\circ = \bigcup_{\tau \geq 0} \tau \cdot \partial \|\mathbf{z}\|_1.$$

Since  $\partial \|\mathbf{z}\|_1 = \left\{ \mathbf{h} \in \mathbb{R}^d : \mathbf{h}_\mathcal{S} = \text{sign}(\mathbf{z})_\mathcal{S}, \mathbf{h}_{\mathcal{S}^c} \in [-1, 1]^{d-s} \right\}$ , it follows that the polar cone is closed, pointed (i.e.,  $\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z})^\circ \cap -\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z})^\circ = \{\mathbf{0}\}$ ) and therefore finitely generated by its extreme

rays

$$\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z})^\circ = \text{cone}(\mathbf{z}^j \in \mathbb{R}^d : j \in [2^{d-s}]),$$

where  $\mathbf{z}_S^j = \text{sign}(\mathbf{z})_S$  and on  $\mathcal{S}^c$  all  $2^{d-s}$  combinations  $\mathbf{z}_{\mathcal{S}^c}^j = \{-1, 1\}^{d-s}$ . Hence, we obtain the following polyhedral description for the descent cone

$$\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z}) = \left\{ \mathbf{h} \in \mathbb{R}^d : \langle \mathbf{h}, \mathbf{z}^j \rangle \leq 0 \text{ for all } j \in [2^{d-s}] \right\}.$$

Using the matrix  $\mathbf{B} := [\mathbf{z}^1, \dots, \mathbf{z}^{2^{d-s}}]^T \in \mathbb{R}^{2^{d-s} \times d}$ , the lineality space can then be conveniently expressed as  $L_{\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z})} = \ker(\mathbf{B})$ .

On the other hand, observe that for any  $\mathbf{h} \in L_{\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z})}$ , we can find  $\tau > 0$  such that  $\|\mathbf{z} + \tau \cdot \mathbf{h}\|_1 \leq \|\mathbf{z}\|_1$  and therefore (by choosing  $\tau > 0$  small enough)

$$\sum_{j \in \mathcal{S}} \text{sign}(z_j) \cdot (z_j + \tau \cdot h_j) + \sum_{i \in \mathcal{S}^c} |h_i| \leq \sum_{j \in \mathcal{S}} |z_j|.$$

Similarly, since also  $-\mathbf{h} \in \mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z})$ , we obtain (again by choosing a small enough  $\tau > 0$ )

$$\sum_{j \in \mathcal{S}} \text{sign}(z_j) \cdot (z_j - \tau \cdot h_j) + \sum_{i \in \mathcal{S}^c} |h_i| \leq \sum_{j \in \mathcal{S}} |z_j|.$$

Adding up these two inequalities, we obtain that  $\sum_{i \in \mathcal{S}^c} |h_i| \leq 0$  and hence  $h_i = 0$  for all  $i \in \mathcal{S}^c$ .

Combining this fact with the previous observation, we obtain that

$$L_{\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z})} = \left\{ \mathbf{h} \in \mathbb{R}^d : \mathbf{h}_{\mathcal{S}^c} = \mathbf{0}, \langle \text{sign}(\mathbf{z}), \mathbf{h} \rangle = 0 \right\},$$

which is of dimension  $s - 1$ . From this description, we can conclude that for each  $i \in \mathcal{S}$  the vector  $s \cdot \text{sign}(z_i) \cdot \mathbf{e}_i - \text{sign}(\mathbf{z})$  is contained in the later space. Hence, if we can show that

$$\dim(\text{span}(s \cdot \text{sign}(z_i) \cdot \mathbf{e}_i - \text{sign}(\mathbf{z}) : i \in \mathcal{S})) = s - 1,$$

we have succeeded in proving the lemma. Indeed, consider the matrix  $\mathbf{C} \in \mathbb{R}^{s \times s-1}$ , where the columns are formed by  $(s \cdot \text{sign}(z_i) \cdot \mathbf{e}_i - \text{sign}(\mathbf{z}))_S$ , for each  $i \in \mathcal{S}$ , except for one. Then, the matrix  $\mathbf{C}^T \cdot \mathbf{C} \in \mathbb{R}^{s-1 \times s-1}$  has the value  $s^2 - s$  on its diagonal and  $-s$  everywhere else. Thus it is strictly diagonal dominant and invertible, implying that  $\mathbf{C}$  is of full rank, as desired.

**Lineality and Range for Gauge (Proposition 4.16)** Lastly, we characterize the range and lineality of  $\mathcal{D}_\wedge(p_{D \cdot \mathbb{B}_1^d}, \mathbf{x}_0)$ :

First, observe that a combination of Lemma 3.5 and Lemma 4.11 yields that

$$\begin{aligned} \mathcal{D}_\wedge(p_{D \cdot \mathbb{B}_1^d}, \mathbf{x}_0) &= \mathbf{D} \cdot \mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z}_{\ell^1}) \\ &= \mathbf{D} \cdot \text{cone}(\pm \bar{s} \cdot \mathbf{e}_i - \text{sign}(\mathbf{z}_{\ell^1}) : i \in [d]) \\ &= \text{cone}(\pm \bar{s} \cdot \mathbf{d}_i - \mathbf{D} \text{sign}(\mathbf{z}_{\ell^1}) : i \in [d]). \end{aligned}$$

By Lemma 4.13, we know how to characterize the lineality of  $\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z}_{\ell^1})$ . Note that for any convex set  $C \subseteq \mathbb{R}^d$ , it holds true that  $(\mathbf{D} \cdot C)_L \supseteq \mathbf{D} \cdot C_L$ , however, the reverse inclusion is not satisfied, in general. Hence, Lemma 3.5 immediately implies  $(\mathcal{D}_\wedge(p_{D \cdot \mathbb{B}_1^d}, \mathbf{x}_0))_L \supseteq \mathbf{D} \cdot (\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z}_{\ell^1}))_L$ . For proving the reverse inclusion  $(\mathcal{D}_\wedge(p_{D \cdot \mathbb{B}_1^d}, \mathbf{x}_0))_L \subseteq \mathbf{D} \cdot (\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z}_{\ell^1}))_L$ , we will now show that if  $(\mathcal{D}_\wedge(p_{D \cdot \mathbb{B}_1^d}, \mathbf{x}_0))_L \not\subseteq \mathbf{D} \cdot (\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z}_{\ell^1}))_L$ , then  $\mathbf{z}_{\ell^1}$  did not have maximal support. To that end, pick any vector  $\mathbf{x} \in (\mathcal{D}_\wedge(p_{D \cdot \mathbb{B}_1^d}, \mathbf{x}_0))_L \setminus \mathbf{D} \cdot (\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z}_{\ell^1}))_L$  and write  $\mathbf{x} = \mathbf{D} \cdot \mathbf{z}^1$ , where  $\mathbf{z}^1 \in \mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z}_{\ell^1}) \setminus (\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z}_{\ell^1}))_L$ . Since  $\mathbf{x} \in (\mathcal{D}_\wedge(p_{D \cdot \mathbb{B}_1^d}, \mathbf{x}_0))_L$ , we can also choose a  $\mathbf{z}^2 \in \mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z}_{\ell^1}) \setminus (\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z}_{\ell^1}))_L$  with  $-\mathbf{x} = \mathbf{D} \cdot \mathbf{z}^2$ . Due to  $\mathbf{z}^i \notin (\mathcal{D}_\wedge(\|\cdot\|_1, \mathbf{z}_{\ell^1}))_L$  for  $i = 1, 2$ , we

have that for all  $\varepsilon > 0$

$$\left\| \mathbf{z}_{\ell^1} - \varepsilon \cdot \mathbf{z}^i \right\|_1 > \|\mathbf{z}_{\ell^1}\|_1, \quad (\text{C.3})$$

however, there exists a small enough  $\varepsilon > 0$  such that

$$\left\| \mathbf{z}_{\ell^1} + \varepsilon \cdot \mathbf{z}^i \right\|_1 \leq \|\mathbf{z}_{\ell^1}\|_1. \quad (\text{C.4})$$

For small enough  $\varepsilon > 0$ , inequality (C.3) implies that

$$\sum_{j \in \mathcal{S}} \text{sign}(z_{\ell^1, j}) \cdot z_j^i - \sum_{j \in \mathcal{S}^c} |z_j^i| < 0,$$

whereas (C.4) means that

$$\sum_{j \in \mathcal{S}} \text{sign}(z_{\ell^1, j}) \cdot z_j^i + \sum_{j \in \mathcal{S}^c} |z_j^i| \leq 0.$$

Summing up the previous two inequalities, we obtain that  $\sum_{j \in \mathcal{S}} \text{sign}(z_{\ell^1, j}) \cdot (z_j^1 + z_j^2) < 0$ . Now, define  $\mathbf{z}^\delta := \mathbf{z}_{\ell^1} + \delta \cdot (\mathbf{z}^1 + \mathbf{z}^2)$  and observe that for all  $\delta > 0$  it holds true that  $\mathbf{x} = \mathbf{D} \cdot \mathbf{z}^\delta$ . Furthermore, for a small enough  $\delta > 0$ , we have that  $\|\mathbf{z}^\delta\|_1 \leq \|\mathbf{z}_{\ell^1}\|_1$ . Hence, we can conclude that  $\mathbf{z}^\delta \in Z_{\ell^1}$  and therefore even  $\|\mathbf{z}^\delta\|_1 = \|\mathbf{z}_{\ell^1}\|_1$ . If  $\delta > 0$  is chosen small enough, this allows us to write

$$\|\mathbf{z}^\delta\|_1 = \|\mathbf{z}_{\ell^1}\|_1 + \delta \cdot \sum_{j \in \mathcal{S}} \text{sign}(z_{\ell^1, j}) \cdot (z_j^1 + z_j^2) + \delta \cdot \sum_{j \in \mathcal{S}^c} |z_j^1 + z_j^2|,$$

and we can conclude that  $\sum_{j \in \mathcal{S}^c} |z_j^1 + z_j^2| > 0$ . However, this means that there is at least one  $j \in \mathcal{S}^c$  such that  $z_j^\delta \neq 0$ , which shows that  $\mathbf{z}$  was indeed not maximal. Finally, Lemma 4.13 implies that

$$\dim \left( \left( \mathcal{D} \wedge (p_{\mathcal{D} \cdot \mathcal{B}_1^d}, \mathbf{x}_0) \right)_L \right) = \dim \left( (\mathcal{D} \wedge (\|\cdot\|_1, \mathbf{z}_{\ell^1}))_L \right) - \dim \left( \ker \mathbf{D}|_{(\mathcal{D} \wedge (\|\cdot\|_1, \mathbf{z}_{\ell^1}))_L} \right) \leq \bar{s} - 1,$$

which concludes the proof of first part of the proposition concerning the lineality of  $\mathcal{D} \wedge (p_{\mathcal{D} \cdot \mathcal{B}_1^d}, \mathbf{x}_0)$ .

The characterization of the range follows easily. Indeed, let  $i \in \mathcal{S}$  and consider the vector  $\mathbf{r}_i^- = -\bar{s} \cdot \text{sign}(z_{\ell^1, i}) \cdot \mathbf{d}_i - \mathbf{D} \cdot \text{sign}(\mathbf{z}_{\ell^1})$ . Observe that we can write  $\mathbf{r}_i^- = -2 \cdot \mathbf{D} \cdot \text{sign}(\mathbf{z}_{\ell^1}) - \mathbf{r}_i^+$ . Hence, for any  $j \in \mathcal{S}^c \neq \emptyset$  we obtain that

$$P_{C_L^\perp}(\mathbf{r}_i^-) = -2 \cdot P_{C_L^\perp}(\mathbf{D} \cdot \text{sign}(\mathbf{z}_{\ell^1})) = \mathbf{r}_j^{\perp\perp} + \mathbf{r}_j^{-\perp}.$$

Thus,  $P_{C_L^\perp}(\mathbf{r}_i^-) \in \text{cone}(\mathbf{r}_j^{\perp\perp}, j \in \mathcal{S}^c)$ , which concludes the proof.

#### C.4 Proof of Theorem 4.17

Let  $C = \mathcal{D} \wedge (p_{\mathcal{D} \cdot \mathcal{B}_1^d}, \mathbf{x}_0)$  and use the orthogonal decomposition provided in Proposition 4.16:

$$\mathcal{D} \wedge (p_{\mathcal{D} \cdot \mathcal{B}_1^d}, \mathbf{x}_0) = C_L \oplus C_R.$$

This allows us to estimate

$$w_\wedge^2(C) \stackrel{(1)}{\leq} \delta(C) \stackrel{(2)}{\leq} \delta(C_L) + \delta(C_R) \stackrel{(3)}{\leq} \dim(C_L) + w_\wedge^2(C_R) + 1, \quad (\text{C.5})$$

where  $\delta$  denotes the *statistical dimension*; see proof of Proposition 4.3 in Appendix B for further details on this notion and a justification of (1). Using the statistical dimension as a summary parameter for convex cones brings several advantages. For a direct sum  $C_1 \oplus C_2$  of two closed convex cones  $C_1, C_2 \subseteq \mathbb{R}^n$  it holds true that  $\delta(C_1 \oplus C_2) = \delta(C_1) + \delta(C_2)$ , explaining (2) in the previous inequalities. Furthermore, for a subspace  $C_L \subseteq \mathbb{R}^n$  we have that  $\delta(C_L) = \dim(C_L)$ , which, together with  $\delta(C_R) \leq w_\wedge^2(C_R) + 1$ , justifies (3). Observe that the estimate of (C.5) is essentially tight.

Proposition 4.16 allows to upper bound  $\dim(C_L) + 1$  by  $\bar{s}$ . The statement then follows by applying Proposition 4.9 to the  $2(\bar{d} - \bar{s})$ -polyhedral  $\alpha$ -cone  $C_R$ .

## C.5 Proof of Proposition 4.20 (Coherence Bound)

First, observe that we have

$$\tan^2(\angle(\mathbf{a}, \mathbf{a} + \mathbf{b})) = \frac{\|\mathbf{a} \times (\mathbf{a} + \mathbf{b})\|_2^2}{\langle \mathbf{a}, \mathbf{a} + \mathbf{b} \rangle^2} = \frac{\|\mathbf{a}\|_2^2 \|\mathbf{b}\|_2^2 - \langle \mathbf{a}, \mathbf{b} \rangle^2}{\left(\|\mathbf{a}\|_2^2 + \langle \mathbf{a}, \mathbf{b} \rangle\right)^2} \leq \frac{\|\mathbf{a}\|_2^2 \|\mathbf{b}\|_2^2}{\left(\|\mathbf{a}\|_2^2 + \langle \mathbf{a}, \mathbf{b} \rangle\right)^2}, \quad (\text{C.6})$$

where  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  with  $\mathbf{a} \neq \mathbf{0}$  and  $\mathbf{a} + \mathbf{b} \neq \mathbf{0}$ .

Observe that the assumptions of Proposition 4.16 are satisfied. Indeed,  $s < \frac{1}{2}(1 + \mu^{-1}(D))$  guarantees that  $\mathbf{z}_{\ell^1}$  is the unique minimal  $\ell^1$ -representer of the associated signal  $D\mathbf{z}_{\ell^1}$  and that  $D\mathbf{z}_{\ell^1} \neq \mathbf{0}$  [DE03; GN03]. Hence, we want to evaluate the circumangle of the cone generated by the vectors  $\mathbf{r}_j^{\pm\perp} = \mathbf{P}_{C_L^\perp}(\pm s \cdot \mathbf{d}_j - D \text{sign}(\mathbf{z}_{\ell^1}))$  for  $j \in S^c$ , where  $S = \text{supp}(\mathbf{z}_{\ell^1})$ . As a proxy for the circumcenter, we can consider the vector  $\mathbf{v} = -\mathbf{P}_{C_L^\perp}(D \text{sign}(\mathbf{z}_{\ell^1}))$  and therefore obtain:

$$\tan^2 \alpha \leq \sup_{j \in S^c} \tan^2(\angle(\mathbf{v}, \mathbf{r}_j^{\pm\perp})) = \sup_{j \in S^c} \tan^2(\angle(\mathbf{v}, \mathbf{v} + \mathbf{P}_{C_L^\perp}(s \cdot \mathbf{d}_j))).$$

We can now use the inequality (C.6) with  $\mathbf{a} = \mathbf{v}$  and  $\mathbf{b} = s \cdot \mathbf{P}_{C_L^\perp}(\mathbf{d}_j)$ ; note that  $\mathbf{v} \neq \mathbf{0}$ , since otherwise we would have  $D\mathbf{z}_{\ell^1} = \mathbf{0}$ . The expression (C.6) is decreasing w.r.t.  $\|\mathbf{a}\|_2^2$ . Hence, we shall find a lower bound for  $\|\mathbf{v}\|_2^2$ . The projection  $\mathbf{P}_{C_L^\perp}(D \text{sign}(\mathbf{z}_{\ell^1}))$  can be written as  $D \text{sign}(\mathbf{z}_{\ell^1}) + \mathbf{w}$  for some vector  $\mathbf{w} \in C_L$ . According to the characterization of the lineality space  $C_L$  in Proposition 4.16, this amounts to saying that

$$\mathbf{P}_{C_L^\perp}(D \text{sign}(\mathbf{z}_{\ell^1})) = \sum_{i \in S} c_i \cdot \text{sign}(z_{\ell^1, i}) \cdot \mathbf{d}_i, \quad \text{with} \quad \sum_{i \in S} c_i = s. \quad (\text{C.7})$$

This yields

$$\begin{aligned} \|\mathbf{v}\|_2^2 &\geq \inf_{\mathbf{c} \in \mathbb{R}^S, \sum_{i \in S} c_i = s} \left\| \sum_{i \in S} c_i \text{sign}(z_{\ell^1, i}) \mathbf{d}_i \right\|_2^2 \\ &= \inf_{\mathbf{c} \in \mathbb{R}^S, \sum_{i \in S} c_i = s} \|\mathbf{c}\|_2^2 + \sum_{i \in S} \sum_{j \in S, j \neq i} c_i c_j \langle \text{sign}(z_{\ell^1, i}) \mathbf{d}_i, \text{sign}(z_{\ell^1, j}) \mathbf{d}_j \rangle \\ &\geq \inf_{\mathbf{c} \in \mathbb{R}^S, \sum_{i \in S} c_i = s} \|\mathbf{c}\|_2^2 - \mu \sum_{i \in S} \sum_{j \in S, j \neq i} c_i c_j. \end{aligned} \quad (\text{C.8})$$

The optimality conditions for this program yield the existence of a Lagrange multiplier  $\lambda \in \mathbb{R}$  such that  $c_i - \mu \sum_{j \neq i} c_j + \lambda = 0$  and  $\sum_{i \in S} c_i = s$ , i.e.,  $c_i = 1$  for all  $i \in S$ . Plugging this expression in (C.8), we obtain that

$$\|\mathbf{v}\|_2^2 \geq s - \mu(s \cdot (s - 1)) \geq s(1 - \mu s).$$

Together with the following inequalities:

$$\begin{aligned} |\langle \mathbf{a}, \mathbf{b} \rangle| &= s \left| \langle \mathbf{v}, \mathbf{P}_{C_L^\perp} \mathbf{d}_j \rangle \right| \leq s |\langle \mathbf{v}, \mathbf{d}_j \rangle| \stackrel{(\text{C.7})}{\leq} s^2 \sup_{i \neq j} |\langle \mathbf{d}_i, \mathbf{d}_j \rangle| = s^2 \mu, \\ \|\mathbf{b}\|_2^2 &= s^2 \left\| \mathbf{P}_{C_L^\perp} \mathbf{d}_j \right\|_2^2 \leq s^2 \|\mathbf{d}_j\|_2^2 = s^2, \end{aligned}$$



we obtain the desired bound

$$\tan^2 \alpha \leq \frac{\|a\|_2^2 \|b\|_2^2}{(\|a\|_2^2 + \langle a, b \rangle)^2} \leq \frac{s(1 - \mu s) \cdot s^2}{(s(1 - \mu s) - s^2 \mu)^2} = \frac{s(1 - \mu s)}{(1 - 2\mu s)^2}.$$

## D Details on Numerical Experiments

In this subsection, we report on the setup that we have used in all our numerical experiments.

**Phase Transition Plots** While our results encompass the more general class of subgaussian measurements, we only consider the benchmark of Gaussian matrices, as it is typically done in the compressed sensing literature. When illustrating the performance of results such as Theorem 3.6, we only report the quantity  $w_\wedge(\mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1; z_0))$ , ignoring for instance the probability parameter  $\mu$ , cf. [ALMT14].

**Some Details on Computations** Unless stated otherwise, we solve the convex recovery programs such as  $(\text{BP}_\eta^{\text{coef}})$  or  $(\text{BP}_{\ell^1})$  using the MatLab toolbox *cvx* [GB08; GB14]. We employ the default settings and set the precision to best. For creating phase transitions, a solution  $\hat{x}$  is considered to be “perfectly recovered” if the error to the ground truth vector  $x_0$  satisfies  $\|x_0 - \hat{x}\|_2 \leq 10^{-5}$ . This threshold produces stable transitions and seems to reflect the numerical accuracy of *cvx*.

**Computing the Statistical Dimension** When analyzing the sampling rate predictions of our results, we often report the conic mean width  $w_\wedge^2(C) = w(C \cap \mathcal{S}^{n-1})$  of a convex cone  $C \subseteq \mathbb{R}^n$ . We will now briefly sketch how this quantity is numerically approximated: First recall that the conic mean width is essentially equivalent to the statistical dimension  $\delta(C) = \mathbb{E}[\sup_{g \in C \cap \mathbb{B}_2^n} \|\Pi_C(g)\|_2^2]$ ; cf. the proof of Proposition 4.3 in Appendix B. Due to the convexity of  $C \cap \mathbb{B}_2^n$ , the statistical dimension is preferred over the conic mean width for numerical simulations. In order to obtain an approximation of  $\delta(C)$ , we draw  $k$  independent samples  $g_1, \dots, g_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Id})$  and for each of them we evaluate the projection  $\Pi_C(g_i)$  using quadratic programming. Due to a concentration phenomenon of empirical Gaussian processes, the arithmetic mean over  $k = 300$  samples yields tight estimates of  $\delta(C)$ .

**Minimal Conic Singular Values** As already mentioned computing  $\lambda_{\min}(\mathbf{D}; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}))$  is out of reach in general. In our numerical experiments on coefficient recovery, we nevertheless provide empirical upper bounds on  $\lambda_{\min}(\mathbf{D}; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}))$ . Those are obtained as follows: Let  $x_0 = \mathbf{D} \cdot z_{\ell^1}$  and consider the perturbed  $\tilde{x}_0 = x_0 + \hat{e}$ , where  $\hat{e} \in \mathbb{R}^n$  such that  $\|\hat{e}\|_2 \leq \hat{\eta}$ . We then define  $\hat{z} \in \mathbb{R}^d$  as a solution of the program

$$\min_{z \in \mathbb{R}^d} \|z\|_1 \quad \text{s.t.} \quad \|\tilde{x}_0 - \mathbf{D}z\|_2 \leq \hat{\eta}.$$

Proposition 2.3 then implies that  $\|z_{\ell^1} - \hat{z}\|_2 \leq 2\hat{\eta} / \lambda_{\min}(\mathbf{D}; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}))$ . Rearranging the terms in the previous inequality then yields an upper bound for  $\lambda_{\min}(\mathbf{D}; \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}))$ . Note thereby that a clever choice of the perturbation  $\hat{e}$  may result in a tighter bound.

**Computing the Circumcenter and the Circumangle** Computing the circumcenter amounts to solving:

$$\theta \in \underset{v \in \mathbb{B}_2^n}{\text{argmin}} \max_{i \in [k]} \langle -v, x_i \rangle, \quad (\text{D.1})$$

where the vectors  $x_i$  are the normalized generators of a nontrivial pointed polyhedral cone; see Proposition 4.7. This problem is closely related to the so-called *smallest bounding sphere problem* [Sy157], which has a long and rich history.

Let  $g(\boldsymbol{v}) = \max_{i \in [k]} \langle -\boldsymbol{v}, \boldsymbol{x}_i \rangle$  and  $I(\boldsymbol{v})$  denote the set of active indices  $i$ , i.e., the indices satisfying  $g(\boldsymbol{v}) = \langle -\boldsymbol{v}, \boldsymbol{x}_i \rangle$ . Then standard convex analysis results state that  $\partial g(\boldsymbol{v}) = \text{conv}(-\boldsymbol{x}_i, i \in I(\boldsymbol{v}))$  and the optimality conditions read

$$\boldsymbol{\theta} \in \text{conv}(\boldsymbol{x}_i, i \in I(\boldsymbol{\theta})) \quad \text{with} \quad \|\boldsymbol{\theta}\|_2 = 1,$$

i.e., the normal cone  $\{-\boldsymbol{\theta}\}$  to the constraint set should intersect the subdifferential  $\partial g(\boldsymbol{\theta})$ .

Problem (D.1) can be solved globally with projected subgradient descents or second order cone programming techniques available in CVX.