Sampling Rates for l1-Synthesis

Maximilian März, Claire Boyer, Jonas Kahn, Pierre Weiss

▶ **To cite this version:**

Maximilian März, Claire Boyer, Jonas Kahn, Pierre Weiss. Sampling Rates for l1-Synthesis. Foundations of Computational Mathematics, 2022. hal-02540711v2

HAL Id: hal-02540711
https://hal.science/hal-02540711v2

Submitted on 7 Dec 2022

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 ℓ^1-Synthesis

Maximilian März[§], Claire Boyer[*], Jonas Kahn[†], Pierre Weiss[‡]
Affiliations: [§]Technische Universität Berlin; [*]LPSM, Sorbonne Université, ENS Paris; [†]Université de Toulouse; [‡]ITAV, CNRS, Université de Toulouse
E-Mail of corresponding author: [§]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 ℓ^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 ℓ^1-descent cone that is linearly transformed by the dictionary. In general, this expression does not allow for a simple calculation by following the polarity-based approach commonly found in the literature. Hence, two upper bounds involving the sparsity structure 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 validated through numerical simulations.

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

1 Introduction

In the last two decades, the methodology of *compressed 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 $\boldsymbol{x}_0 \in \mathbb{R}^n$ be a fixed vector, which is typically referred to as the *signal*. Assume that we are given *m measurements* $\boldsymbol{y} \in \mathbb{R}^m$ of $\boldsymbol{x}_0$ via the linear acquisition model

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{e},$$

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

An important goal is to reconstruct an approximation of the signal $\boldsymbol{x}_0$ from its indirect measurements $\boldsymbol{y}$. Remarkably, even if $m \ll n$, this task can be achieved by incorporating additional information during the reconstruction process. Most classical compressed sensing works directly assume that $\boldsymbol{x}_0$ is *s-sparse*, i.e., that at most $s \ll n$ entries of $\boldsymbol{x}_0$ are nonzero or in symbols $\|\boldsymbol{x}_0\|_0 = \#\operatorname{supp}(\boldsymbol{x}_0) \leq s$. 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 frames or data-adaptive representation systems, 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 $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ and a low-complexity representation $\boldsymbol{z}_0 \in \mathbb{R}^d$ such that $\boldsymbol{x}_0$ can be "synthesized" as

$$\boldsymbol{x}_0 = \boldsymbol{D} \cdot \boldsymbol{z}_0. \tag{1.1}$$

Following the standard terminology of the field, the matrix $\boldsymbol{D} = [\boldsymbol{d}_1, \dots, \boldsymbol{d}_d]$ will be henceforth refered to as *dictionary* and its columns as *dictionary atoms*. Provided that $\boldsymbol{D}$ captures the signal's inherent structure reasonably well, it can be expected that the coefficient vector $\boldsymbol{z}_0$ is dominated by just a few large entries. The resulting sparse synthesis model (1.1) lies at the heart of mathematical signal processing and statistics. It possesses countless applications, ranging from signal compression to the computational foundation of perception in the primary visual cortex [OF96]. The interested reader is for instance referred to [EFM10; Ela10; HTW15; Mal09; MBP14; RBE10] for further details.

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

$$\hat{Z} \coloneqq \operatorname*{argmin}_{\boldsymbol{z} \in \mathbb{R}^d} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2 \leq \eta. \tag{$\mathrm{BP}_\eta^{\mathrm{coef}}$}$$

Under suitable assumptions, one might hope that solutions $\hat{\boldsymbol{z}}$ to this minimization program approximate $\boldsymbol{z}_0$ reasonably well. Indeed, if $\boldsymbol{D} = \mathbf{Id}$, the formulation ($\mathrm{BP}_\eta^{\mathrm{coef}}$) turns into the classical basis pursuit. It allows for the recovery of any s-sparse vector $\boldsymbol{z}_0$ with overwhelming probability, if $\boldsymbol{A}$ 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 $d > n$, 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 $\boldsymbol{D}$ 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 ($\mathrm{BP}_\eta^{\mathrm{coef}}$). However, in many situations of interest, the representation vector itself is irrelevant and a recovery of the actual signal $\boldsymbol{x}_0$ is of primary interest. Thus, one rather cares about the *synthesis basis pursuit for signal recovery*, which amounts to solving

$$\hat{X} \coloneqq \boldsymbol{D} \cdot \left(\operatorname*{argmin}_{\boldsymbol{z} \in \mathbb{R}^d} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2 \leq \eta \right). \tag{$\mathrm{BP}_\eta^{\mathrm{sig}}$}$$

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

1.1 What This Paper Is About

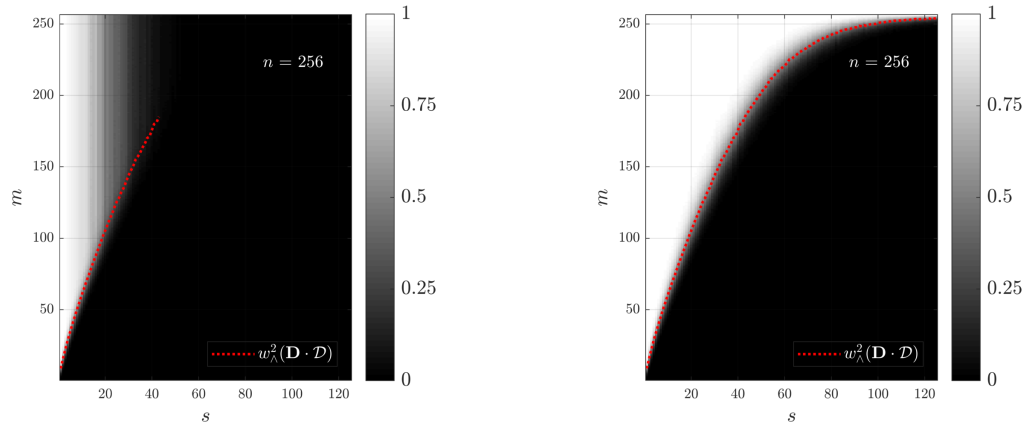
The goal of this work is to broaden the understanding of the conditions that guarantee *coefficient* and *signal recovery* by solving ($\mathrm{BP}_\eta^{\mathrm{coef}}$) and ($\mathrm{BP}_\eta^{\mathrm{sig}}$), respectively. To that end, we believe that addressing the following, non-exhaustive list 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? What structural features of a signal govern this quantity?

(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 restrict ourselves to the following sub-Gaussian measurement model, which will be considered in this work unless stated otherwise:



(a) (b)

Figure 1: **Phase transitions of coefficient and signal recovery by ℓ^1-synthesis.** Subfigure 1(a) shows the empirical probability that atomic coefficient representations are successfully recovered via solving ($\mathrm{BP}^{\mathrm{coef}}_{\eta=0}$), whereas Subfigure 1(b) shows the empirical probability for the associated signal reconstruction by ($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$). 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 3.7 and Theorem 3.9, respectively.

Model 1.2: Sub-Gaussian Measurement Model

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

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

[a] 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 $\boldsymbol{a} \in \mathbb{R}^n$ the sub-Gaussian norm is then given by $\|\boldsymbol{a}\|_{\psi_2} := \sup_{v \in \mathbb{S}^{n-1}} \|\langle \boldsymbol{a}, v \rangle\|_{\psi_2}$ and $\boldsymbol{a}$ is called sub-Gaussian if $\|\boldsymbol{a}\|_{\psi_2} < \infty$; see for instance [Ver12] for further details.

This model has been established as a classical benchmark setup in the context of compressed sensing. It enables 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 ℓ^1-representers* of $\boldsymbol{x}_0$:

$$Z_{\ell^1} := \operatorname*{argmin}_{\boldsymbol{z} \in \mathbb{R}^d} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{z}. \tag{BP_{ℓ^1}}$$

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

(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 $w_\wedge^2(\boldsymbol{D}\cdot\mathcal{D})$, where $\mathcal{D}$ denotes the descent cone of the ℓ^1-norm at any $\boldsymbol{z}_{\ell^1}\in Z_{\ell^1}$ (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 ($\mathrm{BP}_\eta^{\mathrm{sig}}$). For coefficient recovery the additional (but necessary) assumption that Z_{ℓ^1} is a singleton needs to be satisfied.

(C2′) While $w_\wedge^2(\boldsymbol{D}\cdot\mathcal{D})$ provides a tight 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 the width of linearly transformed convex cones.

First, under the assumption that $Z_{\ell^1}=\{\boldsymbol{z}_{\ell^1}\}$ is a singleton and letting $s=\|\boldsymbol{z}_{\ell^1}\|_0$, we obtain a bound for $w_\wedge^2(\boldsymbol{D}\cdot\mathcal{D})$ based on the classical complexity $w_\wedge^2(\mathcal{D})\lesssim s\cdot\log(2n/s)$ multiplied by a certain conic condition number of $\boldsymbol{D}$ (see Section 4.1). Unfortunatley, this bound is overly pessimistic for most applications, and foremost addresses coefficient recovery.

The second upper bound of Section 4.2 is central to our work and is of a more general nature (Z_{ℓ^1} not necessarily a singleton). It is based on a geometric analysis of 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 compressed sensing. Again, $w_\wedge^2(\boldsymbol{D}\cdot\mathcal{D})$ is related to the sparsity of a minimal ℓ^1-representation and a further geometrical parameter (referred to as *circumangle*) that measures the narrowness of the associated cone. An important aspect of this bound is that its computation boils down to a convex optimization problem, which is numerically tractable. In addition, it can be evaluated analytically in some situations of interest. This enables us to demonstrate its usefulness in several examples and to identify non-trivial situations in which such a result is asymptotically near-optimal.

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

1.2 Related Literature and Its Difference to Our Work

The topic of the present paper is closely related to finding sparse decompositions in redundant representation systems — a task that is ubiquitous in signal processing and statistics. In the following, we therefore briefly review some results on the general synthesis sparsity model and then focus on the literature on the actual *synthesis formulation of compressed sensing*, i.e., coefficient/signal-recovery from compressed random measurements by solving ($\mathrm{BP}_\eta^{\mathrm{coef}}$) or ($\mathrm{BP}_\eta^{\mathrm{sig}}$), respectively.

1.2.1 Sparse Representations in Redundant Dictionaries

Apart from solving ill-posed inverse problems as in this work, parsimonious decompositions in redundant systems are a powerful tool for signal compression and denoising [BE08; CDS98; EA06], for classical computer vision and machine learning [MPSZB09; Wri+10; WYGSM08], or for sparse coding in neuroscience [OF96; OF97]. Historically, this research field emerged with the development of (greedy) algorithms for finding expansions in classical time-frequency or wavelet systems [MZ93; PRK93]; see also [FS81; FT74] for prior works in statistics. Initiated[1] by the subsequent work [CDS98], the computation of a minimal ℓ^1-representer via the basis pursuit (BP_{ℓ^1}) became a standard approach to obtain a sparse representation of a given signal $\boldsymbol{x}_0\in\mathbb{R}^n$.

[1] The idea of promoting sparsity in discrete or continuous dictionaries by ℓ^1-norm minimization can be traced back to the works of Beurling [Beu38], Krein [Kre38] and Zuhovickiĭ [Zuh48]. In the 1970's, ℓ^1-regularization was already used 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.

Although the uniqueness of such a representation is a central concern, it may be argued that this aspect is not fully understood in general. Indeed, common theoretical guarantees are for instance based on the assumption of incoherent dictionary atoms [DE03; DH01; Dos05; EB02; GN03; GN08]. In its simplest form, such a result uniformly guarantees that any s-sparse $\boldsymbol{z}_{\ell^1} \in \mathbb{R}^d$ is the unique minimal ℓ^1-representer of the associated signal $\boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{z}_{\ell^1}$ if the *mutual incoherence* of the dictionary $\boldsymbol{D}$ with normalized columns satisfies

$$\mu(\boldsymbol{D}) := \max_{i \neq j} \left| \langle \boldsymbol{d}_i, \boldsymbol{d}_j \rangle \right| \leq 1/(2s-1). \tag{1.2}$$

However, an argumentation based on incoherence is overly pessimistic since it suffers from the *quadratic/square root-bottleneck*: The Welch bound [FR13, Theorem 5.7] reveals that condition (1.2) can only be satisfied for mild sparsity values $s \lesssim \sqrt{n}$, which is often not feasible in practice. Apart from that, popular representation systems (e.g., those based on translation invariant wavelets or learned dictionaries) typically posses linearly dependent and highly coherent atoms. Therefore, the assumption of incoherence is often too restrictive — an observation that also extends to other uniform notions such as the *null space property (NSP)* [CDD09] or the *restricted isometry property (RIP)* [CT05]; see also the discussion in Section 3.1. Next to the previous concepts, also *non-uniform* dual certificates [Fuc04; Fuc05] and exact recovery conditions [Tro04; Tro05] were studied, which concern the uniqueness of specific sparse representations. However, such results come with a limited informative value since they do not provide a simple and explicit criterion for uniqueness (when compared for instance to the sparsity and coherence-based argumentation of (1.2)).

Under the name *compressed 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 $\boldsymbol{D}$ in (BP_{ℓ^1}) is replaced by a random matrix $\boldsymbol{A}$, which follows for instance Model 1.2. This modification is accompanied by interpreting the underlying task as an inverse problem with random measurements rather than a quest for sparse representations in structured dictionaries. From a technical perspective, such a design makes it possible to overcome shortcomings of previous results such as the quadratic/square root-bottleneck. Indeed, for suitable random matrices $\boldsymbol{A}$ 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 highlighted the remarkable potential of sparsity-based methods for signal processing 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 ($\mathrm{BP}_\eta^{\mathrm{sig}}$) and an alternative formulation, which is referred to as ℓ^1-*analysis basis pursuit*:

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

In the previous minimization problem, the analysis operator $\boldsymbol{\Psi} \in \mathbb{R}^{d \times n}$ is chosen in such a way that the coefficient vector $\boldsymbol{\Psi}\boldsymbol{x}_0$ is of low-complexity. It turns out that the latter formulation and the program ($\mathrm{BP}_\eta^{\mathrm{sig}}$) are only equivalent if $\boldsymbol{\Psi}$ (or $\boldsymbol{D}$) forms a basis. In particular for redundant choices of $\boldsymbol{\Psi}$ and $\boldsymbol{D}$, 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 $\boldsymbol{\Psi}\boldsymbol{x}_0$. 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 of compressed sensing for the recovery of coefficient vectors appears to be by Rauhut, Schnass and Vandergheynst [RSV08]. Therein, the formulation ($\mathrm{BP}_\eta^{\mathrm{coef}}$) is studied under a randomized measurement model. The main result roughly reads as follows: Assume that the dictionary $\boldsymbol{D}$ satisfies a RIP with sparsity level s. If the random matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ follows Model 1.2 and $m \gtrsim s \cdot \log(n/s)$, then the composition $\boldsymbol{A}\boldsymbol{D}$ will also satisfy a RIP with sparsity level s with high probability. This property then implies stable

and robust recovery of all s-sparse coefficient vectors by solving ($\mathrm{BP}_\eta^{\mathrm{coef}}$). The assumption that $\boldsymbol{D}$ satisfies a RIP is crucial for the previous result. It can be for instance achieved if the dictionary is sufficiently incoherent, i.e., if it satisfies $\mu(\boldsymbol{D}) \leq 1/(16 \cdot (s-1))$. As mentioned above, such a coherence-based argument is rather crude and suffers from the so-called square-root bottleneck, i.e., it 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 NSP: For a given dictionary $\boldsymbol{D}$, a matrix $\boldsymbol{A}$ is said to satisfy the $\boldsymbol{D}$-NSP of order s, if for any index set $S \subseteq [d]$ with $\#S \leq s$ and any $\boldsymbol{h} \in \boldsymbol{D}^{-1}(\ker \boldsymbol{A} \setminus \{\boldsymbol{0}\})$, there exists $\boldsymbol{z} \in \ker \boldsymbol{D}$, such that $\|\boldsymbol{h}_S + \boldsymbol{z}\|_1 < \|\boldsymbol{h}_{S^c}\|_1$. It can be shown that this condition is necessary and sufficient for the uniform recovery of all signals $\boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{z}_0$ with $\|\boldsymbol{z}_0\|_0 \leq s$ via ($\mathrm{BP}_\eta^{\mathrm{sig}}$). Note that the $\boldsymbol{D}$-NSP is in general weaker than requiring that $\boldsymbol{AD}$ 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 $\boldsymbol{D}$ is of *full spark* (i.e., every n columns of $\boldsymbol{D}$ are linearly independent), both conditions are in fact equivalent. Hence, in this case, signal and coefficient recovery are also equivalent. In the recent work [CCL20], this serves as a motivation to study coefficient recovery by analyzing how many measurements are required in order to guarantee that $\boldsymbol{AD}$ has an NSP. To that end, a result is provided that is conceptually similar to [RSV08], however, it reduces the assumptions on $\boldsymbol{D}$. Instead of requiring that $\boldsymbol{D}$ satisfies a RIP, the authors operate under the weaker assumption that $\boldsymbol{D}$ 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 $\boldsymbol{D}$ satisfies an NSP of order s, a number of $m \gtrsim s \cdot \log(n/s)$ measurements guarantees that also $\boldsymbol{AD}$ satisfies an NSP. This condition then allows for robust recovery of all s-sparse coefficient vectors by solving ($\mathrm{BP}_\eta^{\mathrm{coef}}$).

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 $\|\boldsymbol{d}_i\|_2 \leq 1$, $i \in [d]$ and that $\boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{z}_0$ for an s-sparse representation $\boldsymbol{z}_0 \in \mathbb{R}^d$. For a Gaussian measurement matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, Vershynin establishes the following recovery bound in expectation:

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

where c is a constant and $\hat{\boldsymbol{x}} \in \hat{X}$ is a solution to ($\mathrm{BP}_\eta^{\mathrm{sig}}$). 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 3.9. 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 $K = \boldsymbol{D} \cdot B_1^d$ in [Ver15].

We conclude by mentioning a few more works in the literature on synthesis based compressed sensing that are related to 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 $\boldsymbol{D}$-RIP [CENR11] and on a connection to the analyis formulation with so-called optimal dual frames, [LLMLY12] derives a theorem concerning signal recoery. Finally, [SF09] provides numerical experiments, which empirically compare the analysis and the synthesis formulation of compressed sensing.

1.2.3 The Gap that We Intend to Fill

Although the synthesis formulation of compressed sensing lies at the center of sparse signal processing, it seems to be surprisingly poorly understood. Most of the existing literature is concerned with deriving recovery statements that are *uniform* across all s-sparse coefficient vectors. As such, they resemble classical compressed sensing results in the sense that the sampling rate is determined by the coefficient sparsity alone. This is typically achieved by using strong assumptions on the dictionary, which have been previously used to guarantee uniqueness of ℓ^1-minimization, e.g., incoherent atoms, the NSP or the RIP [CCL20; CWW14; RSV08]. In particular, these assumptions guarantee that the basis pursuit (BP_{ℓ^1}) uniquely identifies every s-sparse coefficient representation in $\boldsymbol{D}$, therefore ensuring signal recovery as well; see [CCL20; CWW14] and the discussion in Section 3.1.

However, as we have argued above, such approaches suffer from severe drawbacks: First, they are typically too restrictive for realistic representation systems with coherent and linearly dependent dictionary atoms. Furthermore, even if a deterministic dictionary satisfied an NSP or RIP, it would be difficult to verify it mathematically [TP13]. Finally, a coherence-based argumentation is feasible, but it has a limited scope due to the square root-bottleneck. In other words, the existing literature on the synthesis formulation of compressed sensing suffers from similar shortcomings as previous results on the uniqueness of sparse representations in redundant dictionaries (see Section 1.2.1).

The goal of the present work is to address these drawbacks by following a signal-dependent approach, which we believe to be crucial for redundant representation systems, cf. [GKM20]. Important differences to existing results are: 1) We avoid strong assumptions on the dictionary $\boldsymbol{D}$. 2) We do not target uniform statements, where the coefficient sparsity is the sole complexity measure. Instead, our work is based on a non-uniform strategy, in which an analysis of the sampling rate identifies relevant structural properties. 3) As a natural consequence of the previous aspects, it becomes necessary to distinguish between coefficient and signal recovery. 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.

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 $w_\wedge^2(\boldsymbol{D}\cdot\mathcal{D})$ is a rather implicit quantity, it constitutes an important step towards the understanding of ℓ^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. Furthermore, 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 Roadmap and Notation

Roadmap. In Section 2, we begin with a short summary of a well-established, non-uniform proof strategy concerning the recovery of structured signals from undersampled random measurements. This approach is based on computing conic mean widths of descent cones and can be skipped by a reader who is already familiar with this methodology. In Section 3, we set the basis of our work by studying coefficient and signal recovery via ($\mathrm{BP}_\eta^{\mathrm{coef}}$) and ($\mathrm{BP}_\eta^{\mathrm{sig}}$), respectively. Based on the previous strategy, novel reconstruction theorems are obtained that characterize the sampling rates in terms of the conic mean width of a linearly transformed descent cone. Section 4 presents two independent possibilities to obtain upper bounds on the previous quantity: The first one (Section 4.1) follows a basic, yet suboptimal conditioning argument and can be skipped on the first reading. The second upper bound (Section 4.2) is of a more geometrical flavor and forms the heartpiece of our work. The paper is written in such a way that the impatient reader can directly jump to Section 4.2.1 to assess one of our main contributions, which is a generic bound on the conic mean width of pointed polyhedral cones. The remainder of Section 4.2 is dedicated to a discussion of this result in the context of the synthesis formulation. Finally, we present detailed numerical experiments in Section 5 and conclude in Section 6.

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 $\boldsymbol{x}=(x_1,\dots,x_n)\in\mathbb{R}^n$. For an index set $\mathcal{I}\subseteq[n]$, we let the vector $\boldsymbol{x}_\mathcal{I}\in\mathbb{R}^{\#\mathcal{I}}$ denote the restriction to the components indexed by $\mathcal{I}$. The *support* of $\boldsymbol{x}$ is defined by the set of its non-zero entries $\operatorname{supp}(\boldsymbol{x}):=\{k\in[n]\mid x_k\neq 0\}$ and the *sparsity* of $\boldsymbol{x}$ is $\|\boldsymbol{x}\|_0:=\#\operatorname{supp}(\boldsymbol{x})$. For $1\le p\le\infty$, $\|\cdot\|_p$ denotes the *ℓ^p-norm* on $\mathbb{R}^n$. The associated *ball with radius $r>0$* is given by $B_p^n(r):=\{\boldsymbol{x}\in\mathbb{R}^n\mid\|\boldsymbol{x}\|_p\le r\}$, the *unit ball* is denoted by B_p^n, and the *Euclidean unit sphere* is $\mathbb{S}^{n-1}:=\{\boldsymbol{x}\in\mathbb{R}^n\mid\|\boldsymbol{x}\|_2=1\}$. The i-th standard basis vector of $\mathbb{R}^n$ is refered to as $\boldsymbol{e}_i$, the identity

Notation	Term
$\boldsymbol{x}_0 \in \mathbb{R}^n$	(ground truth) signal vector
$\boldsymbol{A} \in \mathbb{R}^{m\times n}$	measurement matrix
$\boldsymbol{e} \in \mathbb{R}^m$, with $\|\boldsymbol{e}\|_2 \leq \eta$	(adversarial) noise
$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{e} \in \mathbb{R}^m$	linear, noisy measurements of $\boldsymbol{x}_0$
$\boldsymbol{d}_1, \dots, \boldsymbol{d}_d \in \mathbb{R}^n$	dictionary atoms
$\boldsymbol{D} = [\boldsymbol{d}_1, \dots, \boldsymbol{d}_d] \in \mathbb{R}^{n\times d}$	dictionary
$\hat{\boldsymbol{x}} \in \mathbb{R}^n$	a solution to ($\mathrm{BP}_\eta^{\mathrm{sig}}$)
$\hat{X} = \boldsymbol{D}\cdot\hat{Z} \subseteq \mathbb{R}^n$	solution set of ($\mathrm{BP}_\eta^{\mathrm{sig}}$)
$\boldsymbol{z}_{\ell^1} \in \mathbb{R}^d$	a minimal ℓ^1-decomposition of $\boldsymbol{x}_0$ in $\boldsymbol{D}$, i.e., a solution to (BP_{ℓ^1})
$Z_{\ell^1} \subseteq \mathbb{R}^d$	solution set of (BP_{ℓ^1})
$\boldsymbol{z}_0 \in \mathbb{R}^d$	a sparse representation of $\boldsymbol{x}_0$ in $\boldsymbol{D}$, not necessarily contained in Z_{ℓ^1}
$\hat{\boldsymbol{z}} \in \mathbb{R}^d$	a solution to ($\mathrm{BP}_\eta^{\mathrm{coef}}$)
$\hat{Z} \subseteq \mathbb{R}^d$	solution set of ($\mathrm{BP}_\eta^{\mathrm{coef}}$)

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

matrix is denoted by $\mathbf{Id} \in \mathbb{R}^{n\times n}$, and $\mathbf{1} = (1,\dots,1)^T \in \mathbb{R}^n$. For a set $K \subset \mathbb{R}^n$, we let $\mathrm{conv}(K)$ denote its convex hull, $\mathrm{cone}(K) = \{\sum_{i=1}^k \alpha_i \boldsymbol{x}_i : \boldsymbol{x}_i \in K, \alpha_i \geq 0, k \in \mathbb{N}\}$ its *conical hull*, and $\mathrm{cl}(K)$ its closure. If $L \subset \mathbb{R}^n$ is a linear subspace, the associated *orthogonal projector onto* L is denoted $\boldsymbol{P}_L \in \mathbb{R}^{n\times n}$. Then, we have $\boldsymbol{P}_{L^\perp} = \mathbf{Id} - \boldsymbol{P}_L$, where $L^\perp \subset \mathbb{R}^n$ is the orthogonal complement of L. The letter c is usually reserved for a (generic) constant, with a value that can change at each occurence. 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 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ we let $\mathrm{ran}(A)$ denote its range and $\|\boldsymbol{A}\|_2$ denote its *spectral norm*. For a set $K \subseteq \mathbb{R}^n$, $\lambda \in \mathbb{R}$ and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ we set $\lambda \cdot K := \{\lambda \boldsymbol{k} : \boldsymbol{k} \in K\}$ and $\boldsymbol{A}\cdot K := \{\boldsymbol{A}\boldsymbol{k} : \boldsymbol{k} \in K\}$. 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 ℓ^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 $\boldsymbol{x}_0$ from its linear, noisy measurements $\boldsymbol{y}$, we consider the *generalized basis pursuit*

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

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

The recovery performance of (BP_η^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 $\boldsymbol{A}$ in a neighborhood of the signal vector $\boldsymbol{x}_0$. To that end, we first introduce the following notions of descent cones and minimum conic singular values.

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

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

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

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

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

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

The following result characterizes exact recoverability of the signal $\boldsymbol{x}_0$ and provides a deterministic error bound for the solutions to (BP_η^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 (BP_η^f)) *Assume that $\boldsymbol{x}_0, \boldsymbol{A}, \boldsymbol{y}, \boldsymbol{e}$ and η follow Model 1.1 and let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then the following holds true:*

(a) If $\eta = 0$ and the descent cone $\mathcal{D}_\wedge(f, \boldsymbol{x}_0)$ is closed, exact recovery of $\boldsymbol{x}_0$ by solving ($\mathrm{BP}_{\eta=0}^f$) is equivalent to $\lambda_{min}(\boldsymbol{A}; \mathcal{D}_\wedge(f, \boldsymbol{x}_0)) > 0$.

(b) In addition, any solution $\hat{\boldsymbol{x}}$ of (BP_η^f) satisfies

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

2.2 Conic Mean Width

While Proposition 2.3 provides a deterministic analysis of the solutions to the optimization problem (BP_η^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 $\boldsymbol{A}$ 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 γ_2-functional* in stochastic processes [Tal14], or under the name of *Gaussian complexity* in statistical learning theory [BM02]. Their benefits for compressed 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

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

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

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

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

We refer to $w_\wedge(\mathcal{D}(f, \boldsymbol{x}_0))$ as the *conic mean width* of f at 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 $\boldsymbol{A}$ satisfies the assumption in Model 1.2 and let $K \subseteq \mathbb{S}^{n-1}$ be a set. Then there exists a numerical constant $c > 0$ such that, for every $u > 0$, we have*

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

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

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

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

In the case of Gaussian measurements, it is known that this bound yields a tight description of the so-called *phase transition* of ($\mathrm{BP}^f_{\eta=0}$). Indeed, for a convex cone $C \subseteq \mathbb{R}^n$ it can be shown that $\lambda_{\min}(\boldsymbol{A}; C) = 0$ 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 $\mathcal{D}_\wedge(f, \boldsymbol{x}_0)$ reveals that exact recovery of $\boldsymbol{x}_0$ by solving ($\mathrm{BP}^f_{\eta=0}$) fails with high probability when

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

Hence, exact signal recovery by solving ($\mathrm{BP}^f_{\eta=0}$) obeys a sharp phase transition at $m \approx w_\wedge^2(\mathcal{D}(f, \boldsymbol{x}_0))$ 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 (BP^f_η) is characterized by the minimum conic singular value $\lambda_{min}(\boldsymbol{A}; \mathcal{D}_\wedge(f, \boldsymbol{x}_0))$. The required number of sub-Gaussian random measurements can be determined by the conic mean width of f at $\boldsymbol{x}_0$, in symbols $w_\wedge^2(\mathcal{D}(f, \boldsymbol{x}_0))$.*

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 ℓ^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 ($\mathrm{BP}^{\mathrm{sig}}_\eta$) 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 ℓ^1-Representers

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

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

$$Z_{\ell^1} := \operatorname*{argmin}_{\boldsymbol{z} \in \mathbb{R}^d} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{z}. \tag{BP_{ℓ^1}}$$

In general, Z_{ℓ^1} may not be a singleton. Indeed, a coefficient vector $\boldsymbol{z}_{\ell^1}$ can only be the unique minimal ℓ^1-representer of the associated signal $\boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{z}_{\ell^1}$, if the set of atoms $\{\boldsymbol{d}_i : i \in \operatorname{supp}(\boldsymbol{z}_{\ell^1})\}$ is linearly independent [FR13, Theorem 3.1]. However, many dictionaries of practical interest possess

linearly dependent and coherent atoms. Hence, typical notions that would certify uniqueness for all signals with sparse representations in $\boldsymbol{D}$ (e.g., incoherence, the NSP, or the RIP) are not expected to hold for such dictionaries.

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

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

(a) Assume that $\boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{z}_0$ and that we wish to reconstruct $\boldsymbol{z}_0$. If $Z_{\ell^1} \neq \{\boldsymbol{z}_0\}$, then recovering $\boldsymbol{z}_0$ by solving ($\mathrm{BP}^{\mathrm{coef}}_{\eta=0}$) is impossible.

(b) Signal recovery by solving ($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$), i.e., having $\hat{X} = \{\boldsymbol{x}_0\}$, is equivalent to the condition $Z_{\ell^1} = \hat{Z}$.

Proof. (b) The condition $\hat{X} = \{\boldsymbol{x}_0\}$ imposes $\boldsymbol{D}\hat{\boldsymbol{z}} = \boldsymbol{x}_0$ for any $\hat{\boldsymbol{z}} \in \hat{Z}$. The problem ($\mathrm{BP}^{\mathrm{coef}}_{\eta=0}$) can thus be rewritten by splitting the constraints:

$$\hat{Z} = \operatorname*{argmin}_{\boldsymbol{z}\in\mathbb{R}^d} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{z},\ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0.$$

The second constraint is seen to be superfluous, and we get precisely (BP_{ℓ^1}).

"$\Leftarrow$": If $\hat{Z} = Z_{\ell^1}$, then $\hat{X} = \boldsymbol{D}\cdot\hat{Z} = \boldsymbol{D}\cdot Z_{\ell^1} = \{\boldsymbol{x}_0\}$.

(a) If $\hat{Z} = \{\boldsymbol{z}_0\}$, then also $\hat{X} = \boldsymbol{D}\cdot\hat{Z} = \{\boldsymbol{x}_0\}$ and (b) would imply $Z_{\ell^1} = \hat{Z} = \{\boldsymbol{z}_0\}$.

■

Thus, under the assumption that $\boldsymbol{x}_0$ has a unique minimal ℓ^1-representer, exact coefficient recovery by ($\mathrm{BP}^{\mathrm{coef}}_{\eta=0}$) and signal recovery by ($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$) are equivalent. The following example illustrates Lemma 3.2.

Example 3.3 Let the dictionary $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ be defined by $(\boldsymbol{D}\boldsymbol{x})_i = x_i + x_{i+1}$ for $i \in [n-1]$ and $(\boldsymbol{D}\boldsymbol{x})_n = x_n + x_1$, i.e., $\boldsymbol{D}$ corresponds to a circular convolution with the filter $(1,1)^T$. Consider coefficients $\boldsymbol{z}_0 = 1/2\cdot\boldsymbol{1} \in \mathbb{R}^n$ with the associated signal $\boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{z}_0 = \boldsymbol{1} \in \mathbb{R}^n$. It is then easy to see that $Z_{\ell^1} = \{(\delta, 1-\delta, \delta, 1-\delta, \dots)^T \in \mathbb{R}^n : \delta \in [0,1]\}$. Hence, Lemma 3.2(a) implies that it will be impossible to uniquely recover $\boldsymbol{z}_0$ by solving ($\mathrm{BP}^{\mathrm{coef}}_{\eta}$) for any measurement matrix $\boldsymbol{A}$ (for instance $\boldsymbol{z}_0$ and $(1,0,1,0,\dots)^T$ are both minimal ℓ^1-norm solutions to $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}$). However, if we choose the measurement matrix $\boldsymbol{A} = \boldsymbol{D}$ it is again straightforward to see that $\hat{Z} = Z_{\ell^1}$. In this case, Lemma 3.2(b) guarantees that solving ($\mathrm{BP}^{\mathrm{sig}}_{\eta}$) will recover $\boldsymbol{x}_0$. ◊

3.2 Signal Recovery and the Convex Gauge

The literature on compressed 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 ($\mathrm{BP}^{\mathrm{sig}}_{\eta}$) as a minimization problem over the signal space. In this process, the ℓ^1-ball B_1^d in the coefficient domain is mapped to the convex body $\boldsymbol{D}\cdot B_1^d$, which is referred to as *synthesis defining polytope* in [EMR07]. The formulation ($\mathrm{BP}^{\mathrm{sig}}_{\eta}$) can be equivalently expressed as a constrained minimization of its corresponding *convex gauge*.

Definition 3.4 (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

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

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 ($\mathrm{BP}^{\mathrm{sig}}_{\eta}$).

Lemma 3.5 *Assume that $\boldsymbol{x}_0, \boldsymbol{A}, \boldsymbol{y}, \boldsymbol{e}$ and η follow Model 1.1 and let $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ be a dictionary. Then we have:*

$$\hat{X} = \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{R}^n} p_{\boldsymbol{D}\cdot B_1^d}(\boldsymbol{x}) \quad s.t. \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \eta. \tag{3.1}$$

Proof. By definition,

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

■

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 $p_{\operatorname{conv}(\mathcal{A})}$ in (3.1). It turns out that many popular methods such as classical ℓ^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 ($\mathrm{BP}_\eta^{\mathrm{sig}}$), one would choose the atoms $\mathcal{A} = \{\pm\boldsymbol{d}_i : i \in [d]\}$ to obtain $\operatorname{conv}(\mathcal{A}) = \boldsymbol{D}\cdot B_1^d$. 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 ($\mathrm{BP}_\eta^{\mathrm{sig}}$) and ($\mathrm{BP}_\eta^{\mathrm{coef}}$) 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 $p_{\boldsymbol{D}\cdot B_1^d}$ at $\boldsymbol{x}_0$. The following lemma reveals how this cone is related to descent cones of the ℓ^1-norm in the coefficient space. To make the statement geometrically more accessible, first recall that for a convex set $K \subseteq \mathbb{R}^n$ the *tangent cone* at $\boldsymbol{x}_0 \in K$ is defined by $T_K(\boldsymbol{x}_0) = \operatorname{cl}(\operatorname{cone}(F_K(\boldsymbol{x}_0)))$, where $F_K(\boldsymbol{x}_0) = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x}_0 + \boldsymbol{x} \in K\}$ denotes the *set of feasible directions* at $\boldsymbol{x}_0$. Then the descent cone of the gauge p_K at $\boldsymbol{x}_0$ can be characterized as the *tangent cone* at $\boldsymbol{x}_0$ with respect to the scaled unit ball $p_K(\boldsymbol{x}_0)\cdot K$, i.e., $T_{p_K(\boldsymbol{x}_0)\cdot K}(\boldsymbol{x}_0) = \operatorname{cl}(\mathcal{D}_\wedge(p_K, \boldsymbol{x}_0))$. From this perspective, the next result shows that when transforming the ℓ^1-ball B_1^d into the synthesis defining polytope $\boldsymbol{D}\cdot B_1^d$, small inward pointing directions at a minimal ℓ^1-representer $\boldsymbol{z}_{\ell^1}$ are mapped to small inward pointing vectors at $\boldsymbol{x}_0$, and vice versa.

Lemma 3.6 *Let $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ be a dictionary and let $\boldsymbol{x}_0 \in \operatorname{ran}(\boldsymbol{D})$. For any $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$ we have*

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

The proof is given in Appendix A.1.

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 $\boldsymbol{x}_0$ has a unique minimal ℓ^1-representer $\boldsymbol{z}_{\ell^1}$ with respect to $\boldsymbol{D}$. Proposition 2.3 implies that this condition can be equivalently expressed by

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

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

Theorem 3.7 (Coefficient recovery) *Assume that $\boldsymbol{x}_0, \boldsymbol{A}, \boldsymbol{y}, \boldsymbol{e}$ and η follow Model 1.1, where $\boldsymbol{A}$ is drawn according to the sub-Gaussian Model 1.2 with sub-Gaussian norm γ. Let $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ be a dictionary and $\boldsymbol{z}_{\ell^1} \in \mathbb{R}^d$ be a coefficient vector for the signal $\boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{z}_{\ell^1} \in \mathbb{R}^n$, such that*

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

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

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

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

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

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

The reader is refered to Appendix A.2 for an informative proof of the previous result. It first establishes a deterministic error bound based on minimum conic singular values (see equation (A.3)). From this, the recovery statement (3.3) follows by an application of Gordon's Escape Through a Mesh theorem. Before turning towards signal recovery, let us highlight a few observations regarding the previous theorem.

Remark 3.8 (a) Note that Theorem 3.7 does not assume anything on the dictionary $\boldsymbol{D}$ and the coefficient representation $\boldsymbol{z}_{\ell^1}$, except for $\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})\right) > 0$, which is a necessary condition for the theorem to hold true. As pointed out above, it reflects that $\boldsymbol{z}_{\ell^1}$ is a unique ℓ^1-representer of $\boldsymbol{x}_0$ with respect to $\boldsymbol{D}$, i.e., that $\boldsymbol{z}_{\ell^1}$ is the unique solution to (BP_{ℓ^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 3.7 is that it enables us to isolate the minimum prerequisite of a unique ℓ^1-representer in $\boldsymbol{D}$ from the actual task of compressive coefficient recovery.

(b) Equation (3.2) identifies $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; \boldsymbol{z}_{\ell^1}))$ as the essential component of the sampling rate for coefficient recovery by ($\mathrm{BP}_\eta^{\mathrm{coef}}$). 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 $\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})\right) \gg 0$; 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.2) in the proof, which is not necessarily sharp).

$\lozenge$

Signal recovery Considering signal recovery by ($\mathrm{BP}_\eta^{\mathrm{sig}}$), a combination of the gauge formulation (3.1), its description of the descent cone in Lemma 3.6, and Theorem 2.5 directly yields the next result.

Theorem 3.9 (Signal recovery) *Assume that $\boldsymbol{x}_0, \boldsymbol{A}, \boldsymbol{y}, \boldsymbol{e}$ and η follow the measurement Model 1.1, where $\boldsymbol{A}$ is drawn according to the sub-Gaussian Model 1.2 with sub-Gaussian norm γ. Let $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ be a dictionary with $\boldsymbol{x}_0 \in \mathrm{ran}(\boldsymbol{D})$ and pick any $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$.*
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

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

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

$$\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\|_2 \leq \frac{2\eta}{\sqrt{m-1} - \sqrt{m_0 - 1}}. \tag{3.5}$$

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

Let us discuss the previous result in view of its counterpart for coefficient recovery, Theorem 3.7.

Remark 3.10 (a) Similarly as for coefficient recovery, (3.4) identifies $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; \boldsymbol{z}_{\ell^1}))$ as the main quantity of the sampling rate for signal recovery by ($\mathrm{BP}_\eta^{\mathrm{sig}}$). An important difference is that the set minimal of ℓ^1-representers is not required to be a singleton: The descent cone in the signal space may be evaluated at any possible $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$ 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 coincide, the error bounds of the two theorems differ. The inequality (3.5) does not involve the minimal conic singular value as in Theorem 3.7. This suggests the following noteworthy consequence: In the case of simultaneous coefficient and signal recovery, the robustness to noise of ($\mathrm{BP}_\eta^{\mathrm{coef}}$) and ($\mathrm{BP}_\eta^{\mathrm{sig}}$) might still be different. Indeed, while a reconstruction of $\boldsymbol{x}_0$ is independent of the value of $\lambda_{\min}(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1}))$ — in fact, even 0 is allowed—, the error with respect to $\boldsymbol{z}_{\ell^1}$ 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 3.7 only implies that

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

which is worse than (3.5), in general.

♢

While the bound (3.4) is accurate for an exact recovery from noiseless measurements, it can be improved when an approximate recovery of $\boldsymbol{x}_0$ 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.3 for a proof. Note that such an argumentation does not allow for a similar statement about stable coefficient recovery, due to the product $\boldsymbol{AD}$ in ($\mathrm{BP}_\eta^{\mathrm{coef}}$).

Proposition 3.11 (Stable signal recovery) *Assume that* $\boldsymbol{x}_0, \boldsymbol{A}, \boldsymbol{y}, \boldsymbol{e}$ *and* η *follow the measurement Model 1.1, where* $\boldsymbol{A}$ *is drawn according to the sub-Gaussian Model 1.2 with sub-Gaussian norm* γ. *Let* $\boldsymbol{D} \in \mathbb{R}^{n \times d}$ *be a dictionary with* $\boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{z}_0$. *For a desired precision* $\varepsilon > 0$ *let*

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

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*

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

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

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

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

The previous result extends Theorem 3.9 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 $\boldsymbol{z}^*$ of $\boldsymbol{x}_0$ in $\boldsymbol{D}$ that yield a minimal sampling rate. Note that the original coefficient vector $\boldsymbol{z}_0$ is not required to be a minimal ℓ^1-representer of $\boldsymbol{x}_0$ with respect to $\boldsymbol{D}$. Thus, Proposition 3.11 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.

We conclude this section with an illustration of stable recovery via a simple example.

Example 3.12 Assume that $\boldsymbol{D} = \mathbf{Id}$ and let $\boldsymbol{x}_0 \in \mathbb{R}^n$ 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 $w_\wedge^2(\mathcal{D}(\|\cdot\|_1; \boldsymbol{x}_0)) = n$. Hence, it is impossible to exactly recover $\boldsymbol{x}_0$ from noiseless compressive measurements. However, if we are satisfied with an approximate recovery of $\boldsymbol{x}_0$, we can set the precision for instance to $\varepsilon^2 = \sigma_s(\boldsymbol{x}_0)_1^2/s + \sigma_s(\boldsymbol{x}_0)_2^2$, where $\sigma_s(\boldsymbol{x}_0)_p$ denotes the ℓ^p-error of the best s-term approximation to $\boldsymbol{x}_0$. Then the surrogate vector $\boldsymbol{x}^* \in \mathbb{R}^n$ defined as $x_i^* \coloneqq x_{0,i} + \sigma_s(\boldsymbol{x}_0)_1/s$ for $i = 1, \dots, s$ and $x_i^* \coloneqq 0$ for $i = s+1, \dots, n$ satisfies $\|\boldsymbol{x}^*\|_1 = \|\boldsymbol{x}_0\|_1$ and $\|\boldsymbol{x}^* - \boldsymbol{x}_0\|_2 = \varepsilon$. Furthermore, a computation of the conic mean width yields that $w_\wedge^2(\mathcal{D}(\|\cdot\|_1; \boldsymbol{x}^*)) \lesssim s \log(n/s)$. Hence, Proposition 3.11 shows that ($\mathrm{BP}_{\eta=0}^{\mathrm{sig}}$) allows for the reconstruction of an approximation $\hat{\boldsymbol{x}}$ from $m \gtrsim s \log(n/s)$ noiseless sub-Gaussian measurements that satisfies $\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\|_2 \lesssim \varepsilon$. Note that the appearance of $\sigma_s(\boldsymbol{x}_0)_1^2/s$ is caused by the assumption $\|\boldsymbol{z}_0\|_1 = \|\boldsymbol{z}\|_1$ in (3.6), which is believed to be an artefact of the proof. Nevertheless, this term typically shows up in stability bounds (e.g., see [FR13, Thm. 6.12]). It serves as a proxy for the desired error bound $\sigma_s(\boldsymbol{x}_0)_2$, which cannot be achieved for uniform recovery results; see [FR13, Chap. 11]. Hence, the stability result of this example is near-optimal; cf. [FR13, Rem. 4.23 & Chap. 11].

$\lozenge$

4 Upper Bounds on the Conic Gaussian Width

The previous results identify the conic mean width $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; \boldsymbol{z}_{\ell^1}))$ as the key quantity that controls coefficient and signal recovery by ℓ^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 $\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; \boldsymbol{z}_{\ell^1})$ is given by $(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; \boldsymbol{z}_{\ell^1}))^\circ = (\boldsymbol{D}^T)^{-1}(\mathcal{D}(\|\cdot\|_1; \boldsymbol{z}_{\ell^1})^\circ)$. The appearance of the preimage $(\boldsymbol{D}^T)^{-1}$ 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 $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; \boldsymbol{z}_{\ell^1}))$ that are more accessible and intuitive: Section 4.1 is based on a local conditioning argument, and addresses recovery when a unique minimal ℓ^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 $\boldsymbol{D}$ "out of" the expression $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; \boldsymbol{z}_{\ell^1}))$, in order to make use of the fact that $w_\wedge^2(\mathcal{D}(\|\cdot\|_1; \boldsymbol{z}_{\ell^1}))$ is well understood. We begin by introducing the following notation of a local condition number.

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

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

with the convention $\kappa_{\boldsymbol{D},C} = +\infty$ if $\lambda_{\min}(\boldsymbol{D};C) = 0$. We also use the notation $\kappa_{\boldsymbol{D},z_0} := \kappa_{\boldsymbol{D},\mathcal{D}_\wedge(\|\cdot\|_1;z_0)}$, which we refer to as *local condition number of* $\boldsymbol{D}$ *at* $\boldsymbol{z}_0$ *with respect to the* ℓ^1*-norm*.

Before the previous quantity will be used to simplify $w_\wedge^2(\boldsymbol{D}\cdot\mathcal{D}(\|\cdot\|_1;z_{\ell^1}))$, 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 $\kappa_{\boldsymbol{D},C}$ as a local condition number is motivated by the fact that it can also be written as

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

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

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

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

satisfies

$$\|\boldsymbol{z}_0 - \hat{\boldsymbol{z}}\|_2 \le \frac{2\cdot\|\boldsymbol{D}\|_2\cdot\hat{\eta}}{\lambda_{\min}(\boldsymbol{D};\mathcal{D}_\wedge(\|\cdot\|_1;z_0))} \lesssim \kappa_{\boldsymbol{D},z_0}\cdot\hat{\eta},$$

which shows that $\kappa_{\boldsymbol{D},z_0}$ can be seen as a measure for the stability of $\boldsymbol{z}_0$ with respect to ℓ^1-minimization with $\boldsymbol{D}$.

◊

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* $\boldsymbol{D} \in \mathbb{R}^{n\times d}$*, we have*

$$w_\wedge^2(\boldsymbol{D}\cdot C) \le \kappa_{\boldsymbol{D},C}^2\cdot\left(w_\wedge^2(C)+1\right). \tag{4.1}$$

Proof. See Appendix B. ■

Note that for sparse coefficient vectors $\boldsymbol{z}_0$ the quantity $w_\wedge^2(\mathcal{D}(\|\cdot\|_1;z_0))$ 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

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

where $s = \#\operatorname{supp}(\boldsymbol{z}_0)$. Hence, we directly obtain the following corollary.

Corollary 4.4 *If* $\boldsymbol{z}_{\ell^1}$ *is the unique minimal* ℓ^1*-representer of the associated signal* $\boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{z}_{\ell^1}$*, the critical number of measurements* m_0 *in* (3.2) *and* (3.4) *satisfies*

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

where $s = \#\operatorname{supp}(\boldsymbol{z}_{\ell^1})$.

We have assumed z_{ℓ^1} to be a unique minimal ℓ^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 ($\mathrm{BP}_\eta^{\mathrm{coef}}$), as well as a reconstruction of signals with unique minimal ℓ^1-representers in D by ($\mathrm{BP}_\eta^{\mathrm{sig}}$). 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_{ℓ^1} as the minimal ℓ^1-representer of $x_0 = Dz_{\ell^1}$;

(ii) The complexity of z_{ℓ^1} with respect to ℓ^1-norm, which is measured by $w_\wedge^2(\mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$, or by its sparsity $s = \#\operatorname{supp}(z_{\ell^1})$;

(iii) The quantity $\kappa_{D,z_{\ell^1}}$, which resembles a local measure for the conditioning of D at z_{ℓ^1}.

While the conditioning strategy of this section is tight for the best possible case of an orthogonal dictionary D, it is too pessimistic in general. The following example demonstrates that the factor $\kappa_{D,C}^2$ on the right hand side of (4.1) can become larger than the trivial bound on $w_\wedge^2(D \cdot C)$ by the ambient dimension.

Example 4.5 Let $D \in \mathbb{R}^{n\times n}$ be defined as a discrete gradient operator with $(Dx)_i = x_{i+1} - x_i$ for $i = 1, \dots, n-1$ and $(Dx)_n = -x_n$. Furthermore, consider the closed convex cone $C = t \cdot \mathbf{1}$ for $t \geq 0$, which yields $D \cdot C = t \cdot (0, \dots, 0, -1)$ for $t \geq 0$. For both cones, it is clear that $w_\wedge^2(C), w_\wedge^2(D \cdot C) \leq 1$. However, $Dx = (0, \dots, 0, n^{-1/2})$ for $x = n^{-1/2} \cdot \mathbf{1} \in C \cap \mathbb{S}^{n-1}$, so that $\lambda_{\min}(D; C) \leq n^{-1/2}$. Since $\|D\|_2 \geq 1$, we obtain that $\kappa_{D,C}^2 \geq n$. ◊

Similar findings regarding the accuracy of Corollary 4.4 are obtained in the numerical experiments of Section 5.1. Figure 6 shows examples with $w_\wedge^2(D \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1})) \leq w_\wedge^2(\mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$, yet $\kappa_{D,z_{\ell^1}} \gg 1$. We suspect that a more accurate description might require a detailed analysis of random conic spectra [ST03].

Remark 4.6 (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

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

where $C \subseteq \mathbb{R}^d$ is a closed, convex cone, $\sigma_{\mathbb{R}^n\to C}(-D^T) := \min_{x\in\mathbb{S}^{n-1}} \left\|\Pi_C(-D^T x)\right\|_2$ and Π_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 δ 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 ν and the condition on m for simplicity, the bound on the required samples of Corollary 4.4 could also be formulated with $\mathcal{R}^2_{\mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1})}(D)$ or $\mathcal{R}^2_{\mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}),m}(D)$ instead of $\kappa_{D,z_{\ell^1}}$.

◊

4.2 A Geometric Bound

In this section, we derive an upper bound for $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$ that is based on generic arguments from high-dimensional convex geometry. We exploit the fact that the cone $\boldsymbol{D} \cdot \mathcal{D}_\wedge(\|\cdot\|_1; z_{\ell^1})$ is finitely generated by at most $2d$ vectors (see the proof of Proposition 4.18 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_{ℓ^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 $\mathcal{D}_\wedge(p_{\boldsymbol{D}\cdot B_1^d}, x_0)$ (see Section 4.2.2) in order to derive the desired upper bound on the expression $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$ 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* $\boldsymbol{\theta} \in \mathbb{S}^{n-1}$ and *(half-aperture) angle* $\alpha \in [0, \pi/2]$ is defined as

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

The conic mean width of a circular cone depends linearly on the ambient dimension n, i.e., $w_\wedge^2(C(\alpha, \boldsymbol{\theta})) = n \cdot \sin^2(\alpha) + O(1)$, see for instance [ALMT14, Prop. 3.4]. Although not directly related, it will be insightful to compare this result with the subsequent upper bound of Theorem 4.10.

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

Definition 4.7 (Circumangle) Let $C \subset \mathbb{R}^n$ denote a nontrivial closed convex cone. Its *circumangle* α is defined by

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

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

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

where a vector $\boldsymbol{\theta}$ that maximizes the right hand side is referred to as *circumcenter* (or *outer center*[1]) 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 $\boldsymbol{\theta} \mapsto \inf_{\boldsymbol{x} \in C \cap \mathbb{S}^{n-1}} \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle$ 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:

$$\cos(\alpha) = \sup_{\boldsymbol{\theta} \in B_2^n} \inf_{\boldsymbol{x} \in C \cap \mathbb{S}^{n-1}} \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle.$$

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.

[1] 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].

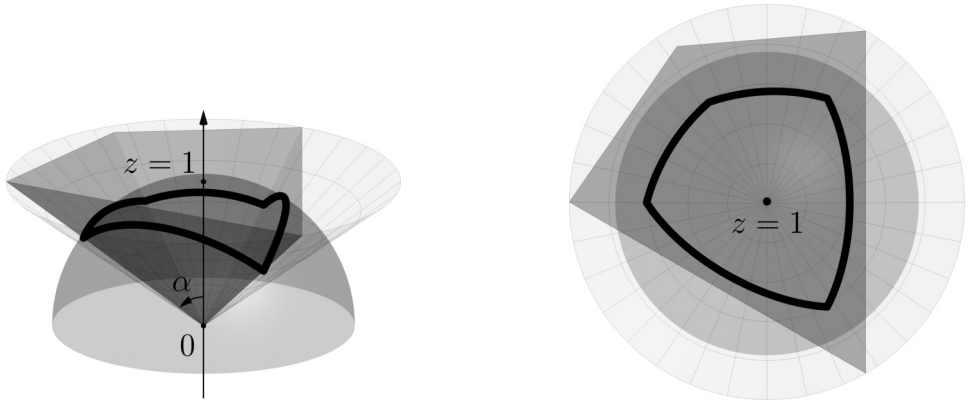


Figure 2: **Geometry of Theorem 4.10.** The figure shows a polyhedral cone (transparent gray) truncated at $z = 1$ and the corresponding circumscribed circular cone with circumangle α (wire-frame). The thick line is the intersection of the unit sphere with the faces of the polyhedral cone. Right view 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, the mean width is small if i) the polyhedral cone does not have overwhelmingly many extremal rays and ii) the circumangle is small.

Proposition 4.8 (Circumangle and circumcenter of polyhedral cones) *Let $\boldsymbol{x}_i \in \mathbb{S}^{n-1}$ for $i \in [k]$ and let $C = \operatorname{cone}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)$ be a nontrivial pointed polyhedral cone. Finding the circumcenter and circumangle of C amounts to solving the convex problem:*

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

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 α. To that end, we first introduce the following notation:

Definition 4.9 A *k-polyhedral α-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 α-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

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

Theorem 4.10 *The conic mean width of a k-polyhedral α-cone C in $\mathbb{R}^n$ is bounded by*

$$W(\alpha, k, n) \leq \tan(\alpha) \cdot \sqrt{2 \log(k)}. \tag{4.3}$$

The underlying geometric idea of the previous result 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.

Remark 4.11 (a) The result of Theorem 4.10 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 a tighter estimate there (possibly an implicit description as in [ALMT14, Prop. 4.5]) would in turn also improve (4.3).

(b) The circumangle of polyhedral cones α already appeared in a different context, yet with no mention of the number k of generating vectors. Under the notion *smallest including cap*, it was studied in the context of conditioning for linear programming. It turns out that the inverse of the *GCC condition number* (for Goffin, Cheaun and Cucker) for the *polyhedral cone feasibility problem* can be expressed by $|\cos(\alpha)|$. This condition number is for instance used to analyze the complexity of linear programming algorithms. We refer the interested reader to [BC13, Section 6.5] and the references therein for further details on this subject.

$\Diamond$

This section is concluded by providing two examples of Theorem 4.10.

Example 4.12 (a) For the non-negative orthant $C := \{\boldsymbol{x} \in \mathbb{R}^n, x_i \geq 0 \text{ for } i \in [n]\}$ it is known that $n/2 - 1 \leq w_\wedge^2(C) \leq n/2$, cf. [ALMT14, Prop. 3.2]. In this case, the circumcenter is the constant vector $[1, 1, \ldots, 1]^T/\sqrt{n}$, and it is easy to see that $\tan(\alpha) = \sqrt{n-1}$. Theorem 4.10 therefore yields $w_\wedge^2(C) \leq 2(n-1)\log(n)$. Unfortunately, this bound is vacuous, since the conic mean width is always bounded by the ambient dimension n. Nevertheless, we still see that it provides the right order of magnitude up to a logarithmic factor.

(b) Consider the cone $C := \{\boldsymbol{x} \in \mathbb{R}^n : x_1 \geq x_2 \geq \cdots \geq x_n \geq 0\}$. It is known that C is isometric to the normal cones of certain permutahedra, e.g. see [ALMT14, Fact D.2] for details. By relying on the intrinsic formulation of the statistical dimension, it was shown in [ALMT14, Prop. 3.5] that $\frac{1}{2}\log(n) - 1 \leq w_\wedge^2(C) \leq \frac{1}{2}\log(n) + 1$. Again, we show below that Theorem 4.10 provides the right order of magnitude up to a logarithmic factor.

First, observe that C is a pointed polyhedral cone that is generated by the vectors

$$\boldsymbol{r}_i := i^{-1/2} \cdot [\underbrace{1, \ldots, 1}_{i\text{-times}}, 0, \ldots, 0]^T \in \mathbb{S}^{n-1},$$

where $i \in [n]$. Next, let us define an axis vector by setting

$$\boldsymbol{\theta} := N \cdot [1, \sqrt{2}-1, \sqrt{3}-\sqrt{2}, \ldots, \sqrt{n}-\sqrt{n-1}]^T,$$

where $N > 0$ is a normalization constant that is chosen such that $\boldsymbol{\theta} \in \mathbb{S}^{n-1}$. Thus, we obtain that $\langle \boldsymbol{\theta}, \boldsymbol{r}_i \rangle = N$ for all $i \in [n]$, i.e., C is contained in a circular cone $C(\alpha, \boldsymbol{\theta})$ with $\cos(\alpha) = N$. An upper-bound for the normalization constant N is:

$$N^{-2} = \sum_{j=1}^{n} (\sqrt{j} - \sqrt{j-1})^2 \leq \sum_{j=1}^{n} j^{-1} \leq \log(n) + 1,$$

where the first estimate follows from the inequality $x^{-1} \geq (\sqrt{x} - \sqrt{x-1})^2$ for $x \geq 1$, and the second estimate is a standard bound on the n-th harmonic number. Therefore, we obtain $\tan(\alpha) \leq \sqrt{\log(n)}$ and Theorem 4.10 implies that $w_\wedge^2(C) \leq 2\log^2(n)$.

$\Diamond$

4.2.2 Geometry of the Descent Cone

In order to derive an upper bound for the quantity $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; \boldsymbol{z}_{\ell^1}))$ based on the previous result, we first need a more geometrical description of the descent cones of the ℓ^1-norm and of the gauge $p_{\boldsymbol{D}\cdot B_1^d}$.

Lemma 4.13 *Let $\boldsymbol{z} \in \mathbb{R}^d$ with support $\operatorname{supp} \boldsymbol{z} = \mathcal{S}$ and $\#\mathcal{S} = s$. Then*

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

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

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 $\boldsymbol{z}$ along the edge linking $\boldsymbol{e}_2$ with $\boldsymbol{e}_3$ leaves the descent cone unchanged.

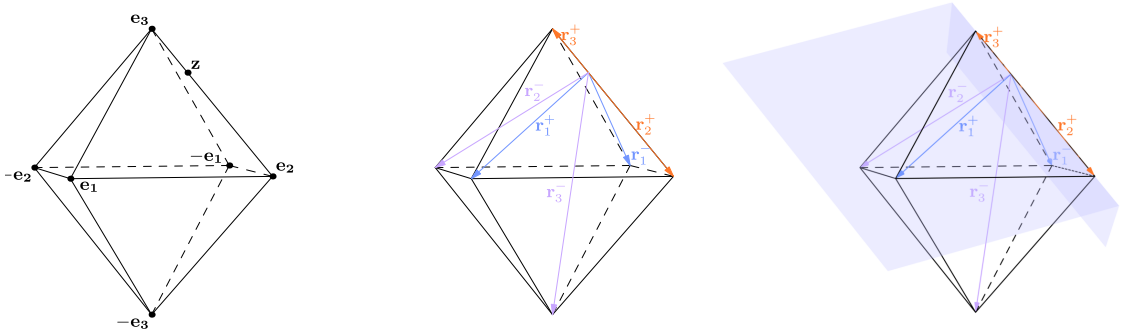


Figure 3: **Illustration of Lemma 4.13.** Left: ℓ^1-ball in $\mathbb{R}^3$ and a 2-sparse vector z. Center: The rays of the descent cone are supported by the vectors $r_i^{\pm} = \pm e_i - z$, which corresponds to the generators of Lemma 4.13 with $v = s \cdot z / \|z\|_1$, multiplied by $1/2$. Right: The resulting descent cone (shifted by z). Note that it contains a linear subspace spanned by r_2^+ and r_3^+.

Whenever a convex cone contains a subspace, its circumangle is $\pi/2$ and the bound of Theorem 4.10 is not applicable. As can be seen in Figure 3, the descent cone of the ℓ^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 4.14 (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 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

$$C = C_L \oplus C_R \text{ with } C_R := P_{C_L^\perp}(C), \tag{4.4}$$

where $P_{C_L^\perp}$ 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 Theorem 4.10 to the latter set only. Figure 4 illustrates the orthogonal decomposition of (4.4) for the descent cone of Figure 3.

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

Lemma 4.15 *Let $z = (z_1, \ldots, z_d)$ be a vector with support $\operatorname{supp} z = \mathcal{S}$ and $\#\mathcal{S} = s \geq 1$. The lineality space of $C = \mathcal{D}_\wedge(\|\cdot\|_1, z)$ is then given by*

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

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

Notice that the lineality space is nothing but the span of the face of the ℓ^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 its lineality space and range. To that end, let us first make the following simple observation.

Lemma 4.16 (Sign pattern of ℓ^1-representers) *All minimal ℓ^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.*

Proof. Let $z_{\ell^1}^1, z_{\ell^1}^2 \in Z_{\ell^1}$ and towards a contradiction assume that there exists an index $i \in [d]$ such

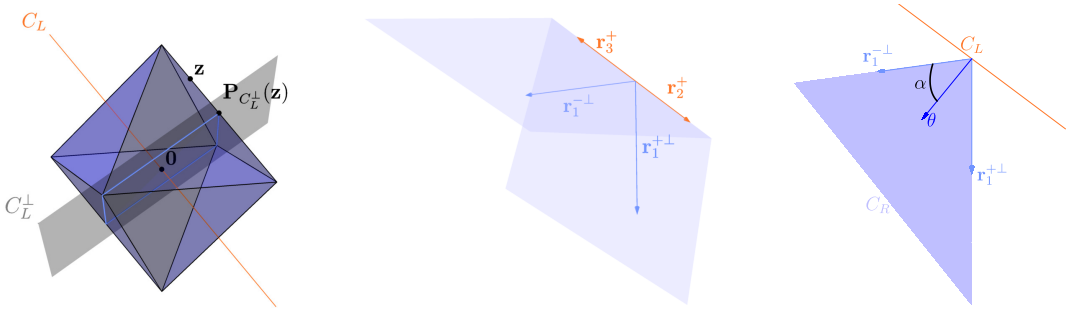


Figure 4: **Decomposition into the direct sum of the lineality space and range**. Left: Decomposition of $\mathbb{R}^3$ into the lineality space C_L and its orthogonal complement $C_L^\perp$, where C is the descent cone of Figure 3. Middle: Different view of C, where $\boldsymbol{r}_i^{\pm\perp} := \boldsymbol{P}_{C_L^\perp}(\boldsymbol{r}_i^{\pm})$.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

that $z^1_{\ell^1,i} \cdot z^2_{\ell^1,i} < 0$. Thus, there exists a $t \in (0,1)$ such that $t \cdot z^1_{\ell^1,i} + (1-t) \cdot z^2_{\ell^1,i} = 0$ and we have

$$\left\| t \cdot \boldsymbol{z}^1_{\ell^1} + (1-t) \cdot \boldsymbol{z}^2_{\ell^1} \right\|_1 = \sum_{\substack{j=1 \\ j \neq i}}^{d} |t \cdot z^1_{\ell^1,j} + (1-t) \cdot z^2_{\ell^1,j}| \leq \sum_{\substack{j=1 \\ j \neq i}}^{d} t \cdot |z^1_{\ell^1,j}| + (1-t) \cdot |z^2_{\ell^1,j}|$$

$$< \sum_{j=1}^{d} t \cdot |z^1_{\ell^1,j}| + (1-t) \cdot |z^2_{\ell^1,j}| = t \cdot \left\| \boldsymbol{z}^1_{\ell^1} \right\|_1 + (1-t) \cdot \left\| \boldsymbol{z}^2_{\ell^1} \right\|_1 = \left\| \boldsymbol{z}^1_{\ell^1} \right\|_1 .$$

This is in contradiction to $t \cdot \boldsymbol{z}^1_{\ell^1} + (1-t) \cdot \boldsymbol{z}^2_{\ell^1} \in Z_{\ell^1}$. ∎

This lemma allows us to define the maximal ℓ^1-support.

Definition 4.17 (Maximal ℓ^1-support) Let $\boldsymbol{x}_0 \in \mathbb{R}^n$ and $\boldsymbol{D} \in \mathbb{R}^{n \times d}$ be a dictionary. The maximal ℓ^1-support $\bar{\mathcal{S}}$ of $\boldsymbol{x}_0$ in $\boldsymbol{D}$ (or simply maximal support) is defined as $\bar{\mathcal{S}} := \cup_{\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}} \operatorname{supp}(\boldsymbol{z}_{\ell^1})$. In what follows, we let $\bar{s} = \#\bar{\mathcal{S}}$ denote its cardinality.

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

Proposition 4.18 *Let $\boldsymbol{D} \in \mathbb{R}^{n \times d}$ be a dictionary and let $\boldsymbol{x}_0 \in \operatorname{ran}(\boldsymbol{D}) \setminus \{\boldsymbol{0}\}$. Let $C = \mathcal{D}_\wedge(p_{\boldsymbol{D} \cdot B_1^d}, \boldsymbol{x}_0)$ denote the descent cone of the gauge at $\boldsymbol{x}_0$. Let $\boldsymbol{z}_{\ell^1} \in \operatorname{ri}(Z_{\ell^1})$ be any minimal ℓ^1-representer of $\boldsymbol{x}_0$ in $\boldsymbol{D}$ with maximal support and set $\bar{\mathcal{S}} = \operatorname{supp}(\boldsymbol{z}_{\ell^1})$ as well as $\bar{s} = \#\bar{\mathcal{S}}$. Assume $\bar{s} < d$. Then we have:*

(a) The lineality space of C has a dimension not larger than $\bar{s} - 1$ and is given by

$$C_L = \operatorname{span}(\bar{s} \cdot \operatorname{sign}(z_{\ell^1,i}) \cdot \boldsymbol{d}_i - \boldsymbol{D} \cdot \operatorname{sign}(\boldsymbol{z}_{\ell^1}) : i \in \bar{\mathcal{S}}).$$

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

$$C_R = \operatorname{cone}(\boldsymbol{r}_j^{\pm\perp} : j \in \bar{\mathcal{S}}^c) \text{ with } \boldsymbol{r}_j^{\pm\perp} := \boldsymbol{P}_{C_L^\perp}\left(\pm\bar{s} \cdot \boldsymbol{d}_j - \boldsymbol{D} \cdot \operatorname{sign}(\boldsymbol{z}_{\ell^1})\right).$$

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 $\mathcal{D}_\wedge(p_{\boldsymbol{D} \cdot B_1^d}, \boldsymbol{x}_0)$.

Theorem 4.19 *We obtain that*

$$w^2_\wedge(\mathcal{D}_\wedge(p_{\boldsymbol{D} \cdot B_1^d}, \boldsymbol{x}_0)) \leq \bar{s} + 2 \cdot \tan^2(\alpha) \cdot \log(2(d-\bar{s})),$$

where we have used the same notation and assumptions as in Proposition 4.18.

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 4.20 *The critical number of measurements m_0 in (3.2) and (3.4) satisfies*

$$m_0 \leq \bar{s} + 2 \cdot \tan^2(\alpha) \cdot \log(2(d - \bar{s})).$$

This result shows that robust coefficient and signal recovery is possible, when the number of measurements obeys $m \geq \bar{s} + 2 \cdot \tan^2(\alpha) \cdot \log((d - \bar{s}))$. Hence, the sampling rate is mainly governed by the sparsity of maximal support ℓ^1-representations of $\boldsymbol{x}_0$ in $\boldsymbol{D}$ 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 4.21 (a) 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 4.19 reveals that $\dim(C_L)$ is a necessary component in the required number of measurements. Indeed, since $w^2_\wedge(\mathcal{D}_\wedge(p_{\boldsymbol{D}\cdot B_1^d}, \boldsymbol{x}_0))$ is a sharp description for the required number of measurements, equation (C.3) shows that the number of measurements for successful recovery is lower bounded by $\dim(C_L)$.

(b) The previous level of detail might not always be needed to compute an upper bound on the conic mean width $w^2_\wedge(\boldsymbol{D} \cdot C)$ of a linearly transformed polyhedral cone $C \subseteq \mathbb{R}^d$. Indeed, assume that $\ker(\boldsymbol{D}) \cap C = \{\boldsymbol{0}\}$, which would correspond to assuming the existence of a unique minimal ℓ^1-representer in the context of this work. It is then straightforward to see that $\dim((\boldsymbol{D} \cdot C)_L) \leq \dim(C_L)$, where $C = C_L \oplus C_R$ denotes the decomposition into the lineality space and range. Therefore, we obtain that

$$w^2_\wedge(\boldsymbol{D} \cdot C) \leq \delta(\boldsymbol{D} \cdot C) = \delta((\boldsymbol{D} \cdot C)_L) + \delta((\boldsymbol{D} \cdot C)_R) = \dim(C_L) + \delta((\boldsymbol{D} \cdot C)_R),$$

where δ denotes the *statistical dimension*, which is essentially equivalent to the conic mean width; see the proof of Proposition 4.3 in Appendix B for details. If C_R is generated by $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$, then the generators of $(\boldsymbol{D} \cdot C)_R$ can be found among $\{\boldsymbol{D}\boldsymbol{x}_1, \ldots, \boldsymbol{D}\boldsymbol{x}_k\}$, and an application of Theorem 4.10 eventually leads to the bound

$$w^2_\wedge(\boldsymbol{D} \cdot C) \leq \dim(C_L) + 2 \cdot \tan^2(\alpha) \cdot \log(k),$$

where α denotes the circumangle of $(\boldsymbol{D} \cdot C)_R$.

$\lozenge$

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 4.20 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 4.20 is consistent with the standard theory of compressed sensing when $\boldsymbol{D} = \mathbf{Id}$. Hence, assume that we are given a sparse vector $\boldsymbol{x}_0 \in \mathbb{R}^n$ with $\mathcal{S} = \operatorname{supp}(\boldsymbol{x}_0)$ and $s = \#\mathcal{S} \geq 1$. Trivially, $\boldsymbol{x}_0$ is then its own, unique ℓ^1-representation with respect to $\mathbf{Id}$. According to Lemma 4.15, the $(s-1)$-dimensional lineality space of $C = \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{x}_0)$ is given by

$$C_L = \operatorname{span}\left(\boldsymbol{r}_i^+ : i \in \mathcal{S}\right),$$

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

$$\langle \boldsymbol{\theta}, \boldsymbol{r}_i^\pm \rangle = \sqrt{s} = (1/\sqrt{s+1}) \cdot \|\boldsymbol{r}_i^\pm\|_2 ,$$

so that the vectors $\boldsymbol{r}_i^\pm$ generate a $2(n-s)$-polyhedral α-cone C_R with $\tan^2(\alpha) = s$. Hence, Corollary 4.20 states that robust recovery of $\boldsymbol{x}_0$ is possible for $m \geq s + 2s\log(2(n-s))$ measurements. This bound is to be compared with the classical compressed sensing result, which prescribes to take $m \gtrsim s\log(n/s)$ measurements. Note that the slight difference in the logarithmic factor is due to our simple bound on $W(\alpha, k, n)$, cf. Remark 4.11(a).

A Convolutional Dictionary Consider a dictionary $\boldsymbol{D}$ defined by the concatenation of two convolution matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ with convolution kernels $\boldsymbol{h}_1 = [1, 1]$ and $\boldsymbol{h}_2 = [1, -1]$, respectively. For instance, in dimension $n = 4$, this would yield the following matrix:

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

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 ℓ^1-representer by ($\mathrm{BP}_\eta^{\mathrm{coef}}$) 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 ($\mathrm{BP}_\eta^{\mathrm{sig}}$).

To that end, we consider the recovery of a simple vector $\boldsymbol{x}_0 \in \mathbb{R}^n$ 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 ℓ^1-representers can be completely characterized. Assuming additionally that $x_{0,1} > x_{0,n} > 0$, one can show that

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

Let $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$ denote any representer with maximal support $S = \operatorname{supp}(\boldsymbol{z}_{\ell^1}) = \{1, 2, n+1, n+2\}$ and set $\boldsymbol{v} = \boldsymbol{D} \cdot \operatorname{sign}(\boldsymbol{z}_{\ell^1})$. According to Proposition 4.18, we then decompose the descent cone $C = \mathcal{D}_\wedge(p_{\boldsymbol{D} \cdot B_1^d}, \boldsymbol{x}_0)$ into $C = C_L \oplus C_R$, where C_L is the lineality space given by

$$C_L = \operatorname{span}(4 \cdot \operatorname{sign}(z_{\ell^1,i}) \cdot \boldsymbol{d}_i - \boldsymbol{v} : i \in S),$$

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

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

The goal is now to show that C_R is contained in a circular cone with angle $\alpha = \arccos(1/\sqrt{3})$ and axis $\boldsymbol{\theta} = -\boldsymbol{P}_{C_L^\perp}(\boldsymbol{v}) / \|\boldsymbol{P}_{C_L^\perp}(\boldsymbol{v})\|_2 = -\boldsymbol{e}_1$. Indeed, a straightforward computation shows that for

$j \in S^c$ we have

$$\left(\boldsymbol{P}_{C_L^\perp}(\pm 4 \cdot \boldsymbol{d}_j - \boldsymbol{v}) / \left\|\boldsymbol{P}_{C_L^\perp}(\pm 4 \cdot \boldsymbol{d}_j - \boldsymbol{v})\right\|_2\right)_1 \in \left\{-1/\sqrt{2}, -1/\sqrt{3}\right\}.$$

Hence, Corollary 4.20 implies that robust recovery of $\boldsymbol{x}_0$ is possible for $m \geq 2 + 4\log(4n)$ measurements.

1D Total Variation As a third example we consider the problem of total variation minimization in 1D. Assume that $\boldsymbol{x}_0, \boldsymbol{A}, \boldsymbol{y}, \boldsymbol{e}$ and η follow Model 1.1 with $\eta = 0$ and that $\boldsymbol{A}$ obeys Model 1.2. Total variation minimization is based on the assumption that $\boldsymbol{x}_0$ is gradient-sparse, i.e., that $\#\operatorname{supp}(\nabla \boldsymbol{x}_0) \leq s \ll n$, 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 $\boldsymbol{x}_0$ from its noiseless, compressed measurements $\boldsymbol{y}$, one solves the program

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

For signals with $\mathbf{1}^T \cdot \boldsymbol{x}_0 = 0$ it is easy to see that the previous formulation is equivalent to solving the synthesis basis pursuit ($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$) with $\boldsymbol{D} = \nabla^\dagger$, where $\nabla^\dagger \in \mathbb{R}^{n \times n-1}$ denotes the Moore-Penrose inverse of ∇.

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 $\boldsymbol{x}_0$ the unique minimal ℓ^1-representer with respect to $\nabla^\dagger$ is simply given by $\nabla \boldsymbol{x}_0$. Hence, we are only left with numerically computing the circumangle α of the range C_R in Proposition 4.18, which is done by means of Proposition 4.8. 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 $n = 500, 1000, \ldots 10000$. 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 4.20 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 4.18 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 $\boldsymbol{z}_{\ell^1}$ is the unique minimal ℓ^1-representer of its associated signal $\boldsymbol{D}\boldsymbol{z}_{\ell^1}$ [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 $C = \mathcal{D}_\wedge(p_{\boldsymbol{D} \cdot B_1^d}, \boldsymbol{D}\boldsymbol{z}_{\ell^1}) = \boldsymbol{D} \cdot \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})$.

Proposition 4.22 *Let $\boldsymbol{D} \in \mathbb{R}^{n \times d}$ be a dictionary that spans $\mathbb{R}^n$ with $\|\boldsymbol{d}_i\|_2 = 1$ for $i \in [d]$ and mutual coherence $\mu = \mu(\boldsymbol{D})$. Let $\boldsymbol{z}_{\ell^1} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ denote an arbitrary s-sparse vector with $s < \frac{1}{2}(1 + \mu^{-1})$. Then*

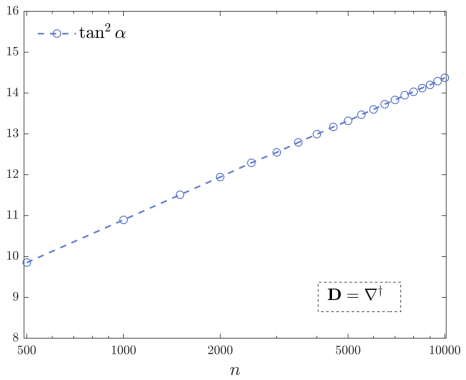


Figure 5: **Logarithmic scaling of $\tan^2(\alpha)$.** The figure displays the behavior of $\tan^2(\alpha)$ for increasing ambient dimension n, where α denotes the circumangle of the range C_R in Theorem 4.10. Here, the dictionary is chosen as $\boldsymbol{D} = \nabla^\dagger$ and the considered signals $\boldsymbol{x}_0$ 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].

the circumangle α of the range C_R of the descent cone $C = \mathcal{D}_\wedge(p_{\boldsymbol{D}\cdot B_1^d}, \boldsymbol{D}\boldsymbol{z}_{\ell^1})$ 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 makes it possible 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 4.20 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 4.22 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 for a description of the local geometry.

5 Numerical Experiments

In this section, we illustrate our main findings regarding the ℓ^1-synthesis formulation by performing numerical simulations. First, we study the recovery of coefficient representations by ($\mathrm{BP}^{\mathrm{coef}}_{\eta=0}$); see Section 5.1. In Section 5.2, we then focus on signal recovery by ($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$) 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 ℓ^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 ($\mathrm{BP}^{\mathrm{coef}}_{\eta=0}$), 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 $\boldsymbol{D} \in \mathbb{R}^{n\times d}$, coefficient vector $\boldsymbol{z}_{\ell^1} \in \mathbb{R}^d$.

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

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

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 $\boldsymbol{D} = \boldsymbol{D}_{\mathtt{Haar}} \in \mathbb{R}^{256\times 1024}$. The first coefficient vector $\boldsymbol{z}^1_{\ell^1} \in \mathbb{R}^{1024}$ is obtained by selecting a random support set of cardinality $s = 16$, together with random coefficients; see Subfigure 6(c) for a visualization of $\boldsymbol{z}^1_{\ell^1}$ and Subfigure 6(b) for the resulting signal $\boldsymbol{x}_1 = \boldsymbol{D}_{\mathtt{Haar}} \cdot \boldsymbol{z}^1_{\ell^1}$. The second coefficient vector $\boldsymbol{z}^2_{\ell^1}$ 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 $\boldsymbol{z}^2_{\ell^1}$ and Subfigure 6(e) for the resulting signal $\boldsymbol{x}_2 = \boldsymbol{D}_{\mathtt{Haar}} \cdot \boldsymbol{z}^2_{\ell^1}$. 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 $\boldsymbol{D}_{\mathtt{rand}} \in \mathbb{R}^{256\times 512}$. The coefficient vector $\boldsymbol{z}^3_{\ell^1}$ is defined in the same manner as $\boldsymbol{z}^1_{\ell^1}$ above (see Subfigure 6(i)), where we again have $\|\boldsymbol{z}^3_{\ell^1}\|_0 = 16$. The resulting signal $\boldsymbol{x}_3$ 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 $\boldsymbol{D}_{\mathtt{super}} \in \mathbb{R}^{256\times 256}$ 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 $\boldsymbol{z}^4_{\ell^1}$ is then chosen as a sparse vector with $z^4_{\ell^1,128} = 1$ and $z^4_{\ell^1,129} = -1$, see Subfigure 6(l). Hence, in the signal $\boldsymbol{x}_4$, 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 $\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}^i_{\ell^1})\right) > 0$ heuristically by verifying that $Z_{\ell^1} = \{\boldsymbol{z}^i_{\ell^1}\}$, respectively.

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

(i) The convex program ($\mathrm{BP}^{\mathrm{coef}}_{\eta=0}$) 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 $\boldsymbol{AD}$ in ($\mathrm{BP}^{\mathrm{coef}}_{\eta=0}$) does not allow for a direct application of these results.

(ii) The quantity $w_\wedge^2(\boldsymbol{D}\cdot\mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}_{\ell^1}))$ accurately describes the sampling rate of ($\mathrm{BP}_{\eta=0}^{\mathrm{coef}}$). Indeed, in all four simulation settings of Figure 6, the phase transition occurs near the estimated conic mean width of $\boldsymbol{D}\cdot\mathcal{D}_\wedge(\|\cdot\|_1\,;\boldsymbol{z}_{\ell^1}^i)$, as predicted by Theorem 3.7.

(iii) In contrast, $w_\wedge^2(\mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}_{\ell^1}))$ does not describe the sampling rate of ($\mathrm{BP}_{\eta=0}^{\mathrm{coef}}$), in general. Indeed, $w_\wedge^2(\boldsymbol{D}\cdot\mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}_{\ell^1}^2)) \ll w_\wedge^2(\mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}_{\ell^1}^2))$ and $w_\wedge^2(\boldsymbol{D}\cdot\mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}_{\ell^1}^4)) \gg w_\wedge^2(\mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}_{\ell^1}^4))$; see Subfigure 6(d) and 6(j), respectively. This underlines the suboptimality of the condition number bounds in Section 4.1; see also observation (v) below.

(iv) For two minimal ℓ^1-representations $\boldsymbol{z}_{\ell^1}^1, \boldsymbol{z}_{\ell^1}^2$ with the same sparsity, but with different supports, the quantities $w_\wedge^2(\boldsymbol{D}\cdot\mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}_{\ell^1}^1))$ and $w_\wedge^2(\boldsymbol{D}\cdot\mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}_{\ell^1}^2))$ might differ significantly, while $w_\wedge^2(\mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}_{\ell^1}^1)) = w_\wedge^2(\mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}_{\ell^1}^2))$; see Subfigures 6(a) and 6(d). Hence, sparsity alone does not appear to be a good proxy for the sampling complexity of ($\mathrm{BP}_{\eta=0}^{\mathrm{coef}}$). 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 $\kappa_{\boldsymbol{D},\boldsymbol{z}_{\ell^1}}$ might explode, which often renders a condition bound as in Proposition 4.3 unusable. Indeed, we report upper bounds for $\lambda_{\min}\left(\boldsymbol{D};\mathcal{D}_\wedge(\|\cdot\|_1\,,\boldsymbol{z}_{\ell^1}^i)\right)$ 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 ($\mathrm{BP}_{\eta=0}^{\mathrm{sig}}$), 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 $\|\boldsymbol{x}_i-\hat{\boldsymbol{x}}\|_2 = \|\boldsymbol{D}\boldsymbol{z}_{\ell^1}^i - \boldsymbol{D}\hat{\boldsymbol{z}}\|_2$ in the third step of the experiment (in addition to $\|\boldsymbol{z}_{\ell^1}^i-\hat{\boldsymbol{z}}\|_2$). Recovery is declared successful if $\|\boldsymbol{x}_i-\hat{\boldsymbol{x}}\|_2 < 10^{-5}$.

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 $\boldsymbol{z}_1 \in \mathbb{R}^{1024}$ with a random support set of cardinality $s = 35$ and random coefficients; see Subfigure 7(j) for a visualization of $\boldsymbol{z}_1$ and Subfigure 7(d) for the resulting signal $\boldsymbol{x}_1 = \boldsymbol{D}_{\texttt{Haar}}\cdot\boldsymbol{z}_1$. In order to apply the result of Theorem 3.9, we compute a minimal ℓ^1-representer $\boldsymbol{z}_{\ell^1}^1 \in Z_{\ell^1}$ of $\boldsymbol{x}_1$; see Subfigure 7(g). The second coefficient vector $\boldsymbol{z}_2$ 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 $\boldsymbol{z}_2$ and Subfigure 7(e) for the resulting signal $\boldsymbol{x}_2 = \boldsymbol{D}_{\texttt{Haar}}\cdot\boldsymbol{z}_2$. A corresponding minimal ℓ^1-representer $\boldsymbol{z}_{\ell^1}^2 \in Z_{\ell^1}$ of $\boldsymbol{x}_2$ 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 28×28-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 3×3 patches. The resulting operator is denoted by $\boldsymbol{D} = \boldsymbol{D}_{\text{dct-2}} \in \mathbb{R}^{n\times 9n}$, 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 $\boldsymbol{z}_3 \in \mathbb{R}^d$ 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 $\boldsymbol{z}_3$. The resulting digit $\boldsymbol{x}_3 = \boldsymbol{D}_{\text{dct-2}}\cdot\boldsymbol{z}_3$ is displayed in Figure 7(f). A minimal ℓ^1-representer $\boldsymbol{z}_{\ell^1}^3$, which is needed to apply Theorem 3.9, is shown in Subfigure 7(i).

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

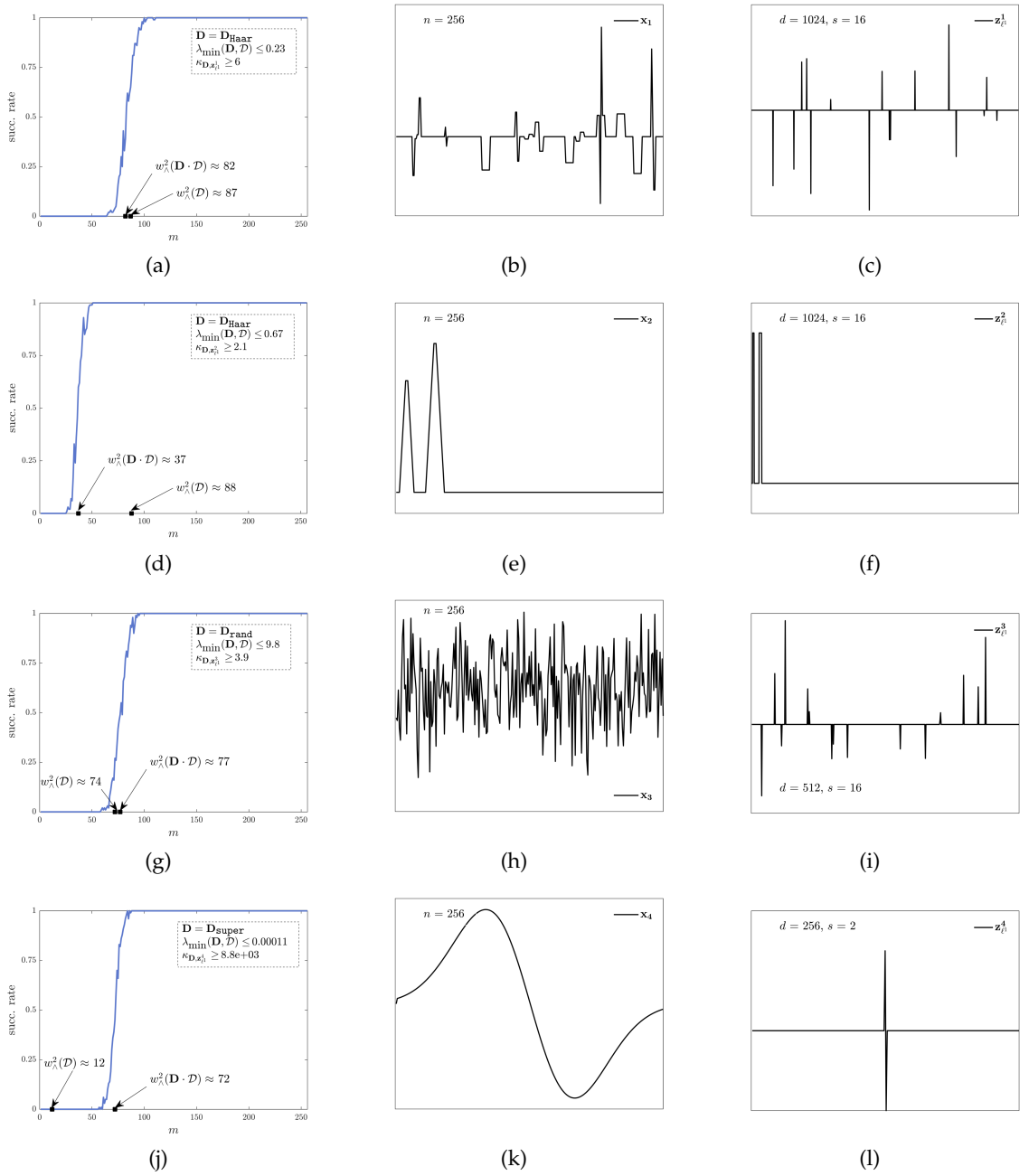


Figure 6: **Phase transitions of coefficient recovery by solving ($\text{BP}^{\text{coef}}_{\eta=0}$).** Empirical success rates and other key figures are reported in the first column, where we use the notation $\mathcal{D} = \mathcal{D}(\|\cdot\|_1, \boldsymbol{z}^i_{\ell^1})$. The coefficient vectors $\boldsymbol{z}^i_{\ell^1}$ that are used in each experiment are shown in the third column. The associated signal vectors $\boldsymbol{x}_i = \boldsymbol{D}\boldsymbol{z}^i_{\ell^1}$ 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.

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

(vii) For any $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$, the conic mean width $w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; \boldsymbol{z}_{\ell^1}))$ accurately describes the sampling rate of ($\text{BP}^{\text{sig}}_{\eta=0}$), as predicted by Theorem 3.9. Indeed, in all three cases of Figure 7, the estimated $w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; \boldsymbol{z}^i_{\ell^1}))$ matches precisely the 50% recovery rate.

(viii) In contrast, for any other sparse representation $z \notin Z_{\ell^1}$, the conic width $w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1 ; z))$ does not describe the sampling rate of ($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$), in general. Indeed, observe that we have $w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1 ; z_i)) \approx n$ 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 $\|z^1_{\ell^1}\|_0 < \|z^2_{\ell^1}\|_0$, we have that $w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1 ; z^1_{\ell^1})) > w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1 ; z^2_{\ell^1}))$. This observation is yet another indication that sparsity alone is not a good proxy for the sampling rate of ℓ^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 $\boldsymbol{D}$ 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 $\boldsymbol{D} = \boldsymbol{D}_{\mathtt{Haar}} \in \mathbb{R}^{256\times 1024}$, 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 $w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1 ; z_{\ell^1}))$.

Experiment 5.2 (Phase transition of Figure 1)

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

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

- Select a set $S \subset [d]$ uniformly at random with $\#S = s$. Then draw a standard Gaussian random vector $\boldsymbol{c} \in \mathbb{R}^s$ and define z_0 by setting $(z_0)_S = \boldsymbol{c}$ and $(z_0)_{S^c} = \boldsymbol{0}$.
- Define the signal $\boldsymbol{x}_0 = \boldsymbol{D}z_0$ and compute a minimal ℓ^1-representation $z_{\ell^1} \in Z_{\ell^1}$ of $\boldsymbol{x}_0$. Compute the conic mean width $w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1 ; z_{\ell^1}))$.
- Run Experiment 5.1 with $\boldsymbol{D}$ and z_{ℓ^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 $\|\boldsymbol{D}z_{\ell^1} - \hat{\boldsymbol{x}}\|_2 = \|\boldsymbol{x}_0 - \boldsymbol{D}\hat{z}\|_2 < 10^{-5}$, respectively.

First note, that the averaged values of the conic mean width perfectly match the center of the phase transition of($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$) in Subfigure 1(b), as it is predicted by Theorem 3.9. 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 ($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$), 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 its location is also determined by $w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1 ; z_{\ell^1}))$. This applies under the condition that coefficient recovery is possible, i.e., that $\lambda_{\min}(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1 ; z_{\ell^1})) > 0$. This property appears to be impossible to satisfy for sparsity values $s \geq 50$, whereas it seems to always hold true for very small sparsity values (i.e., $s \leq 5$). The interval in between forms a wide transition region, in which the possibility for 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.

Hence, we conclude that:

(ix) For redundant and structured dictionaries the ability of ($\mathrm{BP}^{\mathrm{coef}}_{\eta}$) to reconstruct coefficients may not be uniform in the coefficient-sparsity. Put differently, other structural properties besides $\|z_{\ell^1}\|_0$ determine if z_{ℓ^1} is recoverable by ($\mathrm{BP}^{\mathrm{coef}}_{\eta}$), i.e., if $\lambda_{\min}(\boldsymbol{D}; \mathcal{D}_\wedge(\|\|_1 ; z_{\ell^1})) > 0$.

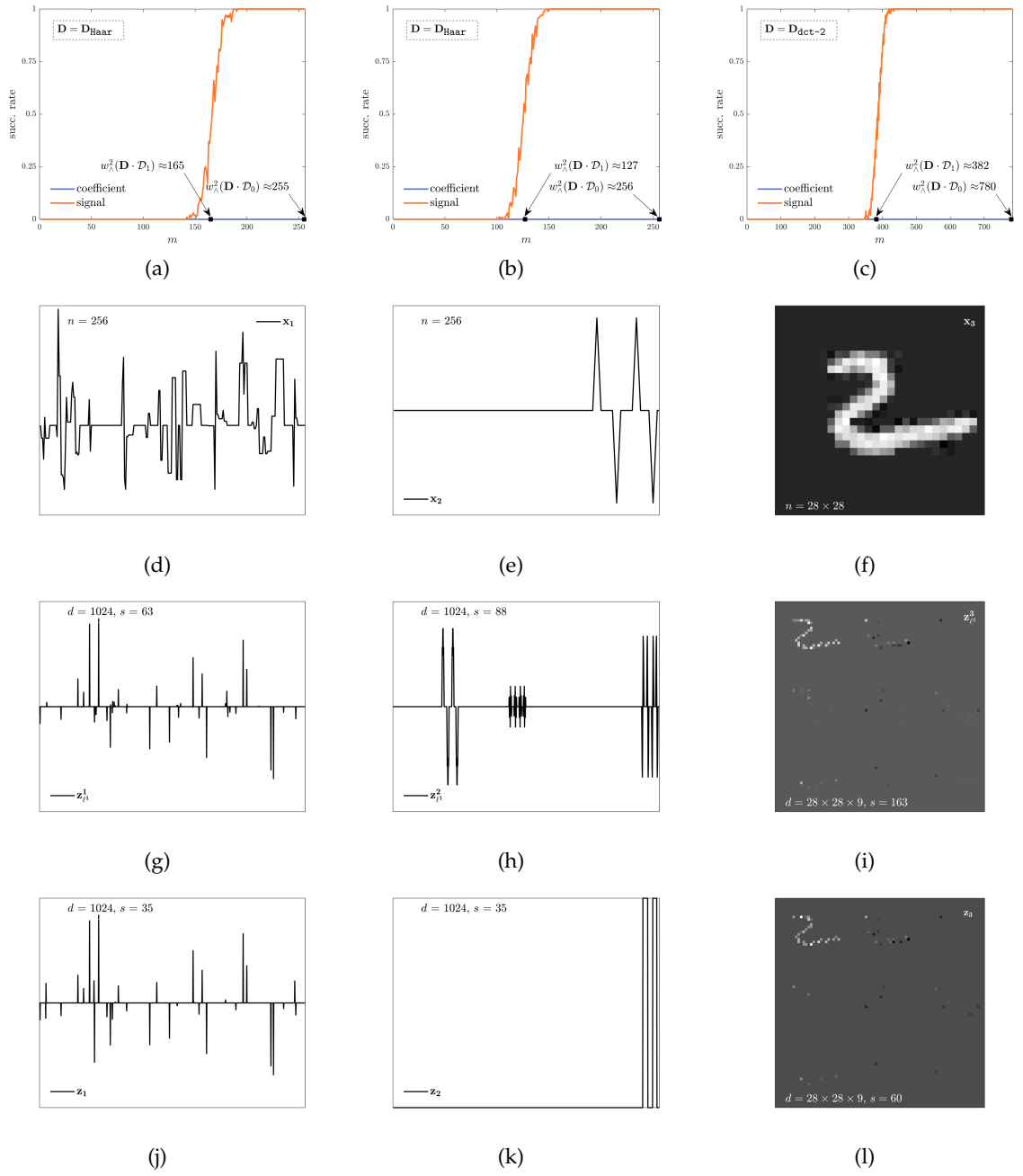


Figure 7: **Phase transitions of signal recovery by solving ($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$).** Empirical success rates and other key quantities are reported in the first row, where we use the notation $\mathcal{D}_1 = \mathcal{D}(\|\cdot\|_1, \boldsymbol{z}^i_{\ell^1})$ and $\mathcal{D}_0 = \mathcal{D}(\|\cdot\|_1, \boldsymbol{z}_i)$. The signals $\boldsymbol{x}_i$ that are used in each experiment are shown in the second row. The associated minimal ℓ^1-representers $\boldsymbol{z}^i_{\ell^1}$ are displayed in the third row and the original coefficient representations $\boldsymbol{z}_i$ 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 $\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}^i_{\ell^1})\right) = 0$.

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 setups, which are reported below.

Experiment 5.3 (Robustness to Measurement Noise)

Input: Dictionary $\boldsymbol{D} \in \mathbb{R}^{n \times d}$, number of measurements m, minimal ℓ^1-representation $\boldsymbol{z}_{\ell^1}$ of the signal $\boldsymbol{x}_0 \in \mathbb{R}^n$, 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 $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and determine the noisy measurement vector $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \eta \cdot \boldsymbol{e}$, where $\|\boldsymbol{e}\|_2 = 1$.
- Solve the program ($\mathrm{BP}_\eta^{\mathrm{coef}}$) to obtain an estimator $\hat{\boldsymbol{z}} \in \mathbb{R}^d$.
- Compute and store the recovery errors $\|\boldsymbol{z}_{\ell^1} - \hat{\boldsymbol{z}}\|_2$ and $\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\|_2 = \|\boldsymbol{D}\boldsymbol{z}_{\ell^1} - \boldsymbol{D}\hat{\boldsymbol{z}}\|_2$.

Simulation settings First, we choose the same dictionary and signal combination as in Section 5.1 and restrict the noise level to $H = \{0, 0.05, 0.1, 0.15, \ldots, 1\}$. Furthermore, we consider the 1D examples of Section 5.2, together with the noise range $H = \{0, 0.005, 0.01, \ldots, 0.1\}$. Recall that the difference of these two setups is that $\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}^i_{\ell^1})\right) > 0$ in the first case, whereas $\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}^i_{\ell^1})\right) = 0$ in the second case. In all experiments, we roughly pick the number of measurements as $m \approx w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; \boldsymbol{z}^i_{\ell^1})) + 40$ to ensure that Theorem 3.7 (or Theorem 3.9, 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 $\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}^i_{\ell^1})\right)$ and in the last two cases, $\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}^i_{\ell^1})\right) = 0$ and therefore Theorem 3.7 is not applicable. Nevertheless, it is possible to obtain an upper bound for the latter quantity, as outlined in the Appendix D. If $\boldsymbol{D} = \boldsymbol{D}_{\mathtt{rand}}$, 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, Prop. 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 3.9, signal recovery via solving the Program ($\mathrm{BP}_\eta^{\mathrm{sig}}$) 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 $\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\|_2$ lies below its theoretical upper bound of Equation (3.5).

(xi) If $\boldsymbol{z}_{\ell^1}$ is the unique minimal ℓ^1-representation of $\boldsymbol{x}_0$ (i.e., if $\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})\right) > 0$) and if the number of measurements exceeds the sampling rate in Theorem 3.7, it is possible to robustly recovery $\boldsymbol{z}_{\ell^1}$. However, in contrast to signal recovery, the robustness is influenced by the "stability" of the minimal ℓ^1-representation of $\boldsymbol{z}_{\ell^1}$ in $\boldsymbol{D}$, i.e., by the value of $\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})\right)$ in the error bound (3.3). Indeed, it is possible that the signal $\boldsymbol{x}_0$ is more robustly recovered than its coefficients $\boldsymbol{z}_{\ell^1}$, or vice versa. This can be seen by comparing coefficient and signal recovery in the Subfigures 8(a)-8(d).[1] If $\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})\right) \ll 1$, coefficient recovery is less robust than signal recovery, see Subfigures 8(a), 8(b) and 8(d). However, if $\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})\right) \gg 1$, the contrary holds true, see Subfigure 8(c).

6 Conclusion

Most of the existing works on compressed sensing with redundant systems state recovery results in terms of sparsity of the underlying signal only. This approach typically leads to statements where the coherence and the restricted isometry property play a critical role. In this paper, we illustrated through a few examples that this approach has severe limitations when linear dependencies between the atoms occur. It then becomes vital to account for the *sparsity structure* as well.

[1] Note that the quantity $\lambda_{\min}\left(\boldsymbol{D}_{\mathtt{super}}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}^4_{\ell^1})\right)$ 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 $\hat{\boldsymbol{z}}$ of ($\mathrm{BP}_\eta^{\mathrm{coef}}$) is always close to $\boldsymbol{0}$.

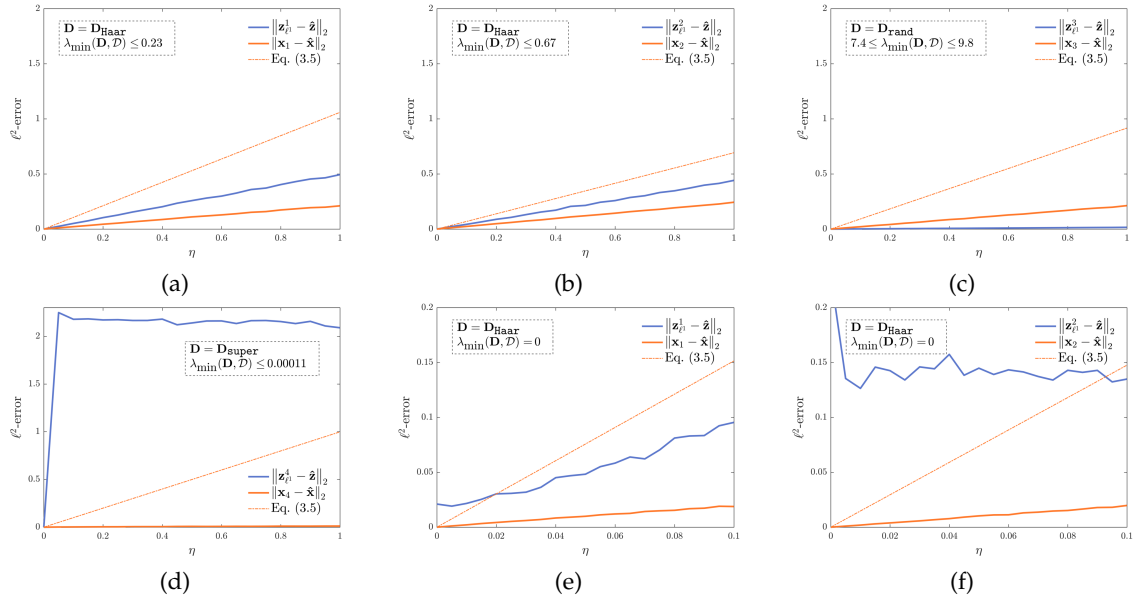


Figure 8: **Robustness to measurement noise.** We display the reconstruction errors for a recovery from noisy measurements with an increasing noise level. The first four Subfigures are based on the examples for coefficient recovery of Section 5.1, and the the last two on the examples based on Haar wavelet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

This motivated us to conduct a basic but fundamental study of recovery conditions for the synthesis formulation with random sub-Gaussian matrices. We precisely characterized the minimal number of measurements needed for stable recovery of both, the signal's coefficients and the signal itself. We concluded that this sampling rate is identical in both cases, but that the stability to noise may differ significantly. The key quantity to control is the conic Gaussian width of a linearly transformed ℓ^1-descent cone. We then derived two upper-bounds for this expression. In particular, we brought to light a geometric feature called circumangle, which already appears in the complexity analysis of linear programming problems. For a given signal, this quantity can be computed efficiently by solving a convex problem. We showed through a few simple analytical and numerical examples that these results make it possible to break some of the limitations currently faced with uniform approaches.

This works opens up many promising perspectives. It seems worth exploring the value of the circumangle for more elaborate and popular representation systems (such as redundant wavelets), or for total variation minimization. Our initial numerical experiments suggest that it yields near optimal guarantees. It could therefore enable us to unravel non-trivial signal classes that can be efficiently coded and recovered by a given dictionary. Alternatively, it could provide new principled approaches for the problem of dictionary learning.

To finish, let us discuss a few limitations of this work. The proposed bounds depend logarithmically on the number of atoms in the dictionary. The latter should therefore not be exponential in the ambient dimension. While this might seem harmless for the synthesis model, it gets problematic for the analysis formulation. Indeed, it is known that a problem in an analysis-based form can be cast as a synthesis-based version, at the expense of a combinatorial explosion of the number of atoms [EMR07]. Finding ways to break this limitation would therefore provide a unified way to treat the analysis and synthesis formulation, which are typically studied separately. Similarly, continuous dictionaries currently gain more attention since they make it possible to model important problems such as super-resolution imaging or two-layer neural networks. The dictionaries then possess an infinite number of atoms and our results do not apply directly. However, in both cases, the proposed geometrical approach, the stable versions of our results (e.g., Proposition 3.11) and specific discretization procedures appear to be promising tools.

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 Optimization on Measures Spaces and the ANR MicroBlind for blind inverse problems. The authors wish to thank the anonymous reviewers for their careful revision and comments, which helped improving the presentation significantly.

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.

[BC13] P. Bürgisser and F. Cucker. *Condition: The geometry of numerical algorithms*. Vol. 349. Springer Science & Business Media, 2013.

[BCN+17] R. Baraniuk, H. Choi, R. Neelamani, et al. *Rice Wavelet Toolbox, Version 3*. URL: https://github.com/ricedsp/rwt. Aug. 2017.

[BE08] O. Bryt and M. Elad. "Compression of facial images using the K-SVD algorithm". *J. Vis. Commun. Image Represent.* 19.4 (2008), 270–282.

[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.

[BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 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.

[CCL20] P. G. Casazza, X. Chen, and R. G. Lynch. "Preserving injectivity under subgaussian mappings and its application to compressed sensing". *Appl. Comput. Harmon. Anal.* 49.2 (2020), 451–470.

[CDD09] A. Cohen, W. Dahmen, and R. DeVore. "Compressed sensing and best k-term approximation". *J. Am. Math. Soc.* 22.1 (2009), 211–231.

[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.

[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.

[CT05] E. J. Candès and T. Tao. "Decoding by linear programming". *IEEE Trans. Inf. Theory* 51.12 (2005), 4203–4215.

[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 ℓ_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 ℓ^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.

[Dos05] C. Dossal. "Estimation de fonctions géométriques et déconvolution". https://tel.archives-ouvertes.fr/tel-00855128/. PhD thesis. 2005.

[EA06] M. Elad and M. Aharon. "Image denoising via sparse and redundant representations over learned dictionaries". *IEEE Trans. Imag. Proc.* 15.12 (2006), 3736–3745.

[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).

[EFM10] M. Elad, M. A. T. Figueiredo, and Y. Ma. "On the Role of Sparse and Redundant Representations in Image Processing". *Proc. IEEE* 98.6 (2010), 972–982.

[Ela10] M. Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer-Verlag, New York, 2010.

[EMR07] M. Elad, P. Milanfar, and R. Rubinstein. "Analysis versus synthesis in signal priors". *Inverse Probl.* 23.3 (2007), 947–968.

[Fer75] X. Fernique. "Regularité des trajectoires des fonctions aléatoires gaussiennes". *Ecole d'Eté de Probabilités de Saint-Flour IV—1974*. Springer, 1975, 1–96.

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

[FS81] J. H. Friedman and W. Stuetzle. "Projection pursuit regression". *J. Am. Stat. Assoc.* 76.376 (1981), 817–823.

[FT74] J. H. Friedman and J. W. Tukey. "A projection pursuit algorithm for exploratory data analysis". *IEEE Trans. Comput.* 100.9 (1974), 881–890.

[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. "ℓ^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". *Inf. Inference* (2020). Accpeted for publication.

[GN03] R. Gribonval and M. Nielsen. "Sparse representations in unions of bases". *IEEE Trans. Inf. Theory* 49.12 (2003), 3320–3325.

[GN08] R. Gribonval and M. Nielsen. "Beyond sparsity: Recovering structured representations by ℓ^1 minimization and greedy algorithms". *Adv. Comput. Math.* 28.1 (2008), 23–41.

[GNEGD14] R. Giryes, S. Nam, M. Elad, R. Gribonval, and M. E. Davies. "Greedy-like algorithms for the cosparse 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.

[HTW15] T. Hastie, R. Tibshirani, and M. Wainwright. *Statistical learning with sparsity: the lasso and generalizations*. CRC press, 2015.

[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 ℓ_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. Kabanva, H. Rauhut, and H. Zhang. "Robust analysis ℓ_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 ℓ_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.

[Mal09] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. 3rd Edition. Elsevier, 2009.

[MBP14] J. Mairal, F. Bach, and J. Ponce. "Sparse Modeling for Image and Vision Processing". *Found. Trends. Comput. Graph. Vis.* 8.2-3 (2014), 85–283.

[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.

[MPSZB09] J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, and F. Bach. "Supervised Dictionary Learning". *Advances in Neural Information Processing Systems*. Ed. by D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou. Vol. 21. 2009.

[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 cosparse analysis model and algorithms". *Appl. Comput. Harmon. Anal.* 34.1 (2013), 30–56.

[OF96] B. A. Olshausen and D. J. Field. "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". *Nature* 381.6583 (1996), 607–609.

[OF97] B. A. Olshausen and D. J. Field. "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vis. Res.* 37.23 (1997), 3311–3325.

[PRK93] Y. C. Pati, R. Rezaiifar, 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.

[RBE10] R. Rubinstein, A. M. Bruckstein, and M. Elad. "Dictionaries for sparse representation modeling". *Proc. IEEE* 98.6 (2010), 1045–1057.

[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 ℓ_1-optimization in compressed sensing". Preprint arXiv:0907.3666. 2009.

[Sud71] V. N. Sudakov. "Gaussian random processes and measures of solid angles in Hilbert space". *Doklady Akademii Nauk*. Vol. 197. 1. Russian Academy of Sciences. 1971, 43–45. English translation: *Soviet Math. Dokl.* Vol. 12. 1971, 412–415, p. 546.

[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.

[TP13] A. M. Tillmann and M. E. Pfetsch. "The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing". *IEEE Trans. Inf. Theory* 60.2 (2013), 1248–1259.

[Tro04] J. A. Tropp. "Greed is good: algorithmic results for sparse approximation". *IEEE Trans. Inf. Theory* 50.10 (2004), 2231–2242.

[Tro05] J. A. Tropp. "Recovery of Short, Complex Linear Combinations Via ℓ^1 Minimization". *IEEE Trans. Inf. Theory* 51.4 (2005), 1568–1570.

[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.

[Wri+10] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan. "Sparse representation for computer vision and pattern recognition". *Proc. IEEE* 98.6 (2010), 1031–1044.

[WYGSM08] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. "Robust face recognition via sparse representation". *IEEE Trans. Pattern Anal. Mach. Intell.* 31.2 (2008), 210–227.

[Zuh48] S. Zuhovickiĭ. "Remarks on problems in approximation theory". *Mat. Zbirnik KDU* (1948). (Ukrainian), 169–183.

A Proofs of Section 3

A.1 Proof of Lemma 3.6 (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 $p_{\boldsymbol{D}\cdot B_1^d}(\boldsymbol{x}_0) = \|z_{\ell^1}\|_1$.
"$\supseteq$": Let $\boldsymbol{h} \in \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1})$, i.e., there exists a $\tau > 0$ such that $\|z_{\ell^1} + \tau \boldsymbol{h}\|_1 \leq \|z_{\ell^1}\|_1$. Hence,

$$p_{\boldsymbol{D}\cdot B_1^d}(\boldsymbol{D}z_{\ell^1} + \tau \boldsymbol{D}\boldsymbol{h}) = p_{\boldsymbol{D}\cdot B_1^d}(\boldsymbol{D}\cdot(z_{\ell^1} + \tau \boldsymbol{h})) \leq \|z_{\ell^1} + \tau \cdot \boldsymbol{h}\|_1 \leq \|z_{\ell^1}\|_1 = p_{\boldsymbol{D}\cdot B_1^d}(\boldsymbol{x}_0),$$

and therefore $\boldsymbol{D}\boldsymbol{h} \in \mathcal{D}_\wedge(p_{\boldsymbol{D}\cdot B_1^d}, \boldsymbol{x}_0)$.
"$\subseteq$": Let $\boldsymbol{x} \in \mathcal{D}_\wedge(p_{\boldsymbol{D}\cdot B_1^d}, \boldsymbol{x}_0)$, i.e., there exists $\tau > 0$ such that

$$R := p_{\boldsymbol{D}\cdot B_1^d}(\boldsymbol{x}_0 + \tau \boldsymbol{x}) \leq p_{\boldsymbol{D}\cdot B_1^d}(\boldsymbol{x}_0) = \|z_{\ell^1}\|_1 .$$

Now, choose $\boldsymbol{h} \in B_1^d$ such that $R \cdot \boldsymbol{D}\boldsymbol{h} = \boldsymbol{x}_0 + \tau \boldsymbol{x}$ and write $\boldsymbol{x} = \boldsymbol{D}\cdot(R/\tau \cdot \boldsymbol{h} - 1/\tau \cdot z_{\ell^1}) =: \boldsymbol{D}(\bar{\boldsymbol{z}})$. Observe that

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

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

Remark A.1 The proof of "$\subseteq$" shows that z_{ℓ^1} could be replaced by any other z_0 with $\boldsymbol{x}_0 = \boldsymbol{D}z_0$, which is not necessarily a minimal ℓ^1-representer of $\boldsymbol{x}_0$. Hence, $\mathcal{D}_\wedge(p_{\boldsymbol{D}\cdot B_1^d}, \boldsymbol{x}_0) \subseteq \boldsymbol{D}\cdot\mathcal{D}_\wedge(\|\cdot\|_1, z_0)$ and $\mathcal{D}(p_{\boldsymbol{D}\cdot B_1^d}, \boldsymbol{x}_0) \subseteq \boldsymbol{D}\cdot\mathcal{D}(\|\cdot\|_1, z_0)$ for any z_0 with $\boldsymbol{x}_0 = \boldsymbol{D}z_0$, with equality if $z_0 \in Z_{\ell^1}$. ♢

A.2 Proof of Theorem 3.7 (Coefficient Recovery)

Since $\boldsymbol{x}_0, \boldsymbol{A}, \boldsymbol{y}, \boldsymbol{e}, \eta$ follow Model 1.1 and $\boldsymbol{x}_0 = \boldsymbol{D}z_{\ell^1}$, Proposition 2.3 guarantees that any solution $\hat{\boldsymbol{z}}$ to the program ($\mathrm{BP}_\eta^{\mathrm{coef}}$) obeys

$$\|z_{\ell^1} - \hat{\boldsymbol{z}}\|_2 \leq \frac{2\eta}{\lambda_{\min}(\boldsymbol{A}\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1; z_{\ell^1}))}. \tag{A.1}$$

Hence, the goal of the proof is to find a lower bound for the minimum conic singular value $\lambda_{\min}(\boldsymbol{A}\boldsymbol{D}; C) = \inf\{\|\boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2 : \boldsymbol{z} \in C \cap \mathbb{S}^{d-1}\}$, where $C := \mathcal{D}_\wedge(\|\cdot\|_1; z_{\ell^1})$.

Note that by assumption $\|\boldsymbol{D}\boldsymbol{z}\|_2 \geq \lambda_{\min}(\boldsymbol{D}; C) > 0$ for all $\boldsymbol{z} \in C \cap \mathbb{S}^{d-1}$. Thus, we obtain that

$$\begin{aligned}
\lambda_{\min}(\boldsymbol{A}\boldsymbol{D}; C) &= \inf\left\{\frac{\|\boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2}{\|\boldsymbol{D}\boldsymbol{z}\|_2} \cdot \|\boldsymbol{D}\boldsymbol{z}\|_2 : \boldsymbol{z} \in C \cap \mathbb{S}^{d-1}\right\} \\
&\geq \lambda_{\min}(\boldsymbol{D}; C) \cdot \inf\{\|\boldsymbol{A}\boldsymbol{x}\|_2 : \boldsymbol{x} \in \boldsymbol{D}C \cap \mathbb{S}^{n-1}\} \\
&\geq \lambda_{\min}(\boldsymbol{D}; C) \cdot \lambda_{\min}(\boldsymbol{A}; \boldsymbol{D}C).
\end{aligned} \tag{A.2}$$

The bound in (A.1) then implies that any solution $\hat{\boldsymbol{z}}$ to ($\mathrm{BP}_\eta^{\mathrm{coef}}$) satisfies the deterministic error bound

$$\|z_{\ell^1} - \hat{\boldsymbol{z}}\|_2 \leq \frac{2\eta}{\lambda_{\min}(\boldsymbol{D}; C) \cdot \lambda_{\min}(\boldsymbol{A}; \boldsymbol{D}C)}. \tag{A.3}$$

Since we have additionally assumed that $\boldsymbol{A}$ is drawn according to the sub-Gaussian Model 1.2, a probabilistic lower bound on $\lambda_{\min}(\boldsymbol{A}; \boldsymbol{D}C)$ by means of Gordon's excape through a Mesh can be established. Indeed, Theorem 2.5 guarantees that there is a numerical constant $c > 0$ such that with probability at least $1 - \mathrm{e}^{-u^2/2}$, we have that

$$\lambda_{\min}(\boldsymbol{A}; \boldsymbol{D}C) = \inf\{\|\boldsymbol{A}\boldsymbol{x}\|_2 : \boldsymbol{x} \in \boldsymbol{D}C \cap \mathbb{S}^{n-1}\} > \sqrt{m-1} - c \cdot \gamma^2 \cdot (w_\wedge(\boldsymbol{D}C) + u).$$

Therefore, with the same probability, if $m > m_0 = c^2 \cdot \gamma^4 \cdot (w_\wedge(\boldsymbol{D}\cdot C) + u)^2 + 1$, any solution $\hat{\boldsymbol{z}}$ to

the program ($\mathrm{BP}_\eta^{\mathrm{coef}}$) obeys the desired bound

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

Remark A.2 The above argumentation may be compared with [CCL20, Theorem 3.1]. We also show that if $\boldsymbol{D}$ is bounded away from 0 on the intersection S of a closed convex cone C and the sphere, then $\boldsymbol{AD}$ also stays away from 0 on S with high probability. However, an important difference is that our result does not involve $\lambda_{\min}(\boldsymbol{D}; C)^{-1}$ as a multiplicative factor in the rate $m_0 \approx w_\wedge^2(\boldsymbol{D} \cdot C)$. 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 $\lambda_{\min}(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1; z_{\ell^1}))$ may be very small, while ($\mathrm{BP}_{\eta=0}^{\mathrm{coef}}$) still allows for exact recovery from $m \approx w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; z_{\ell^1}))$ measurements. ◊

A.3 Proof of Proposition 3.11 (Stable Recovery)

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

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

Thus, we need to verify that t satisfies

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

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

$$\begin{aligned} w_t(\mathcal{D}(p_{\boldsymbol{D} \cdot B_1^d}; x_0)) &\leq w_t(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; z_0) + x_0 - x_0) \\ &\overset{t \geq r \cdot \|x_0 - \boldsymbol{D}z^*\|_2}{\leq} w_{r \cdot \|x_0 - \boldsymbol{D}z^*\|_2}((\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; z_0) + x_0) - x_0), \end{aligned}$$

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

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

where the equality follows from:

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

We conclude that

$$t \geq \frac{2\eta}{\sqrt{m-1} - c \cdot \gamma^2 \cdot \left(\frac{r+1}{r} \cdot (w_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; z^*)) + 1) + u\right)}$$

$$\geq \frac{2\eta}{\sqrt{m-1} - c\cdot\gamma^2\cdot\left(w_t(\mathcal{D}(p_{\boldsymbol{D}\cdot B_1^d};\boldsymbol{x}_0)) + u\right)}.$$

Hence, [GKM20, Theorem 6.4] then implies that any minimizer of ($\mathrm{BP}_\eta^{\mathrm{sig}}$) satisfies $\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\|_2 \leq t$.[1]

B Proof of Proposition 4.3 (Width and Condition Number)

As pointed out in Remark 4.6(b), similar results were previously established in [ALW20]. In order to make our work self-contained, we have nevertheless included a short proof of Proposition 4.3. It first requires 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 $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ and a standard Gaussian vector $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Id}_n)$, we have that*

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

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

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

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

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

where C° denotes the *polar cone* of C. Furthermore, it holds true that

$$\mathbb{E}\left[\left(\sup_{z\in C\cap\mathbb{S}^{d-1}}\langle \boldsymbol{g}, \boldsymbol{D}z\rangle\right)^2\cdot \mathbf{1}_{\mathbb{R}^d\setminus C^\circ}(\boldsymbol{D}^*\boldsymbol{g})\right] = \mathbb{E}\left[\left(\sup_{z\in C\cap B_2^d}\langle \boldsymbol{g}, \boldsymbol{D}z\rangle\right)^2\right].$$

Indeed, for an $\boldsymbol{x}\in\mathbb{R}^n$ such that $\boldsymbol{D}^*\boldsymbol{x}\notin C^\circ$ the equality $\sup_{z\in C\cap\mathbb{S}^{d-1}}\langle \boldsymbol{x}, \boldsymbol{D}z\rangle = \sup_{z\in C\cap B_2^d}\langle \boldsymbol{x}, \boldsymbol{D}z\rangle$ holds true, because the supremum over the ball occurs at a vector of length 1. On the other hand, when $\boldsymbol{D}^*\boldsymbol{x}\in C^\circ$, one has $\sup_{z\in C\cap B_2^d}\langle \boldsymbol{x}, \boldsymbol{D}z\rangle = 0$. Therefore, (B.1) is established.

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

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

and therefore by taking the supremum

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

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

$$\mathbb{E}\left[Z^2\right] - \mathbb{E}\left[Z\right]^2 = \mathbb{E}\left[(Z-\mathbb{E}Z)^2\right] = \mathrm{Var}(Z) \leq \lambda_{\max}^2(\boldsymbol{D}, C),$$

[1] This argument does actually not cover the case of $\eta = 0$, but here we can simply use that $t = r\cdot\|\boldsymbol{x}_0 - \boldsymbol{D}\boldsymbol{z}^*\|_1 > 0$ if $\boldsymbol{x}_0 \neq \boldsymbol{D}\boldsymbol{z}^*$.

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

$$w_\wedge^2(\boldsymbol{D}\cdot C) = w^2(\boldsymbol{D}\cdot C\cap \mathbb{S}^{n-1}) \le \delta(\boldsymbol{D}\cdot C) = \mathbb{E}\left[\left(\sup_{\boldsymbol{x}\in \boldsymbol{D}\cdot C\cap B_2^n}\langle \boldsymbol{g},\boldsymbol{x}\rangle\right)^2\right],$$

where δ denotes the *statistical dimension*[1]. Next, it is straightforward to see that

$$\boldsymbol{D}\cdot C\cap B_2^n \subseteq \frac{\boldsymbol{D}}{\lambda_{\min}(\boldsymbol{D};C)}\cdot (C\cap B_2^d),$$

which immediately implies that

$$\delta(\boldsymbol{D}\cdot C) \le \frac{1}{\lambda_{\min}^2(\boldsymbol{D};C)}\cdot \mathbb{E}\left[\left(\sup_{\boldsymbol{x}\in \boldsymbol{D}\left(C\cap B_2^d\right)}\langle \boldsymbol{g},\boldsymbol{x}\rangle\right)^2\right].$$

Exercise 7.5.4 in [Ver18] and Lemma B.1 now makes it possible to derive the desired bound:

$$\begin{aligned}
w_\wedge^2(\boldsymbol{D}\cdot C) \le \delta(\boldsymbol{D}\cdot C) &\le \frac{1}{\lambda_{\min}^2(\boldsymbol{D};C)}\mathbb{E}\left[\left(\sup_{\boldsymbol{x}\in \boldsymbol{D}\left(C\cap B_2^d\right)}\langle \boldsymbol{g},\boldsymbol{x}\rangle\right)^2\right]\\
&\overset{B.1}{\le} \frac{1}{\lambda_{\min}^2(\boldsymbol{D};C)}\left(\left(\mathbb{E}\left[\sup_{\boldsymbol{x}\in \boldsymbol{D}(C\cap \mathbb{S}^{d-1})}\langle \boldsymbol{g},\boldsymbol{x}\rangle\right]\right)^2 + \lambda_{\max}^2(\boldsymbol{D};C)\right)\\
&\overset{(7.5.4)}{\le} \frac{\|\boldsymbol{D}\|_2^2}{\lambda_{\min}^2(\boldsymbol{D};C)}\left(\left(\mathbb{E}\left[\sup_{\boldsymbol{z}\in C\cap \mathbb{S}^{d-1}}\langle \boldsymbol{g},\boldsymbol{z}\rangle\right]\right)^2 + 1\right)\\
&= \kappa_{\boldsymbol{D},C}^2\cdot\left(w_\wedge^2(C)+1\right).
\end{aligned}$$

C Proofs of Section 4.2

C.1 Proof of Proposition 4.8 (Circumangle of Polyhedral Cones)

Consider a nontrivial pointed polyhedral cone $C = \operatorname{cone}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)$ with $\|\boldsymbol{x}_i\|_2 = 1$ for $i\in[k]$ and let α denote its circumangle. Let $\boldsymbol{X} := [\boldsymbol{x}_1,\dots,\boldsymbol{x}_k]\in\mathbb{R}^{n\times k}$. Since C does not contain a line, its circumcenter $\boldsymbol{\theta}$ is unique, belongs to the cone and $0\le\alpha<\pi/2$, see [HS10a]. This implies that $\boldsymbol{X}^*\boldsymbol{\theta} > \boldsymbol{0}$ (element-wise). We have:

$$\begin{aligned}
\cos(\alpha) &= \sup_{\boldsymbol{v}\in\mathbb{S}^{n-1}}\inf_{\boldsymbol{x}\in C\cap\mathbb{S}^{n-1}}\langle \boldsymbol{x},\boldsymbol{v}\rangle = \sup_{\boldsymbol{v}\in\mathbb{S}^{n-1}}\inf_{\boldsymbol{c}\ge\boldsymbol{0},\|\boldsymbol{X}\boldsymbol{c}\|_2^2=1}\langle \boldsymbol{X}\boldsymbol{c},\boldsymbol{v}\rangle\\
&\le \sup_{\boldsymbol{v}\in\mathbb{S}^{n-1}}\inf_{i\in[k]}\langle \boldsymbol{e}_i,\boldsymbol{X}^*\boldsymbol{v}\rangle = \sup_{\boldsymbol{v}\in\mathbb{S}^{n-1}}\min \boldsymbol{X}^*\boldsymbol{v},
\end{aligned}$$

where $\boldsymbol{e}_i$ 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

[1] The statistical dimension of a convex cone $C\subseteq\mathbb{R}^n$ can be defined as $\delta(C) = \mathbb{E}[(\sup_{\boldsymbol{x}\in C\cap B_2^n}\langle \boldsymbol{g},\boldsymbol{x}\rangle)^2]$; see [ALMT14, Prop. 3.1] for details. It holds true that $w_\wedge^2(C)\le\delta(C)\le w_\wedge^2(C)+1$, which is why both notions are often interchangeable [ALMT14, Prop. 10.2].

for any $v \in \mathbb{S}^{n-1}$ such that $X^*v \geq 0$, we also have that

$$\inf_{c\geq 0, \|Xc\|_2^2=1} \langle c, X^*v\rangle \geq \inf_{c\geq 0, \|Xc\|_2^2\geq 1} \langle c, X^*v\rangle \geq \inf_{c\geq 0, \langle \mathbb{1}, c\rangle \geq 1} \langle c, X^*v\rangle = \min X^*v.$$

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 $\operatorname{Ext}(\{c \geq 0, \langle \mathbb{1}, c\rangle \geq 1\}) = \{e_1, \dots, e_k\}$. Note that the condition $X^*v \geq 0$ ensures the existence of a solution.

Finally the infimum over $\mathbb{S}^{n-1}$ 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 Theorem 4.10 (Maximal Width of Polyhedral Cones)

This section is dedicated to establishing Theorem 4.10. We will first present a proof based on *Sudakov-Fernique's comparison inequality* [Fer75; Sud71]. Afterwards, we outline an alternative strategy in Remark C.2, which closer reflects the geometric idea of Figure 2. Both proofs have in common that they necessitate a basic preliminary result on the Gaussian width of general convex polytopes, which we will provide first.

Bounding the Gaussian Width of Convex Polytopes The following lemma is well known in the literature, e.g., see [Ver18, Ex. 7.5.10 & Prop. 7.5.2]. For the sake of completeness, we still provide a short proof.

Lemma C.1 *Let K be a convex polytope with k vertices that is contained in the unit ball of $\mathbb{R}^n$. Then we have that $w(K) \leq \sqrt{2\log(k)}$.*

Proof. Let $K = \operatorname{conv}(x_1, \dots, x_k) \subseteq B_2^n$. Then it holds true that

$$w(K) = \mathbb{E}\big[\sup_{h\in K} \langle g, h\rangle\big] = \mathbb{E}\big[\sup_{i\in[k]} \langle g, x_i\rangle\big] \leq \sqrt{2\log(k)}.$$

The second equality is satisfied since the maximum of a linear function on a convex set is attained at an extreme point, i.e., at a vertex. The last inequality follows from the well known bound $\mathbb{E}\big[\sup_{i\in[k]} X_i\big] \leq \sup_{i\in[k]} \|x_i\|_2^2 \cdot \sqrt{2\log(k)}$, where $X_i = \langle g, x_i\rangle \sim \mathcal{N}(0, \|x_i\|_2^2)$ for $i \in [k]$, e.g., see [BLM13, Section 2.5]. ∎

Back to the proof of Theorem 4.10 Equipped with the previous lemma, we can now prove Theorem 4.10. To that end, let $C = \operatorname{cone}(x_1, \dots, x_k) \subseteq \mathbb{R}^n$ be a k-polyhedral α-cone and let $\theta \in \mathbb{S}^{n-1}$ be an axis vector such that $C \subseteq C(\alpha, \theta)$. Without loss of generality assume that $\|x_i\|_2 = 1$ for $i \in [k]$. Define the affine hyperplane $\mathcal{H} := \{h \in \mathbb{R}^n : \langle h, \theta\rangle = 1\}$ and let $\tilde{K} := C \cap \mathcal{H}$.

We now consider the following two mean zero Gaussian processes indexed by $x \in \tilde{K}$:

$$X_x := \Big\langle g, \frac{x}{\|x\|_2}\Big\rangle, \quad \text{and} \quad Y_x := \langle g, x\rangle,$$

where $g \sim \mathcal{N}(\mathbf{0}, \mathbf{Id})$. Observe that the map $\tilde{K} \to C \cap \mathbb{S}^{n-1}, x \mapsto x/\|x\|_2$ is a contraction, since it is the projection onto the convex set $C \cap B_2^n$. Hence, for any two points $x, x' \in \tilde{K}$ we have that

$$\mathbb{E}\big[(X_x - X_{x'})^2\big] = \Big\| \frac{x}{\|x\|_2} - \frac{x'}{\|x'\|_2}\Big\|_2^2 \leq \|x - x'\|_2^2 = \mathbb{E}\big[(Y_x - Y_{x'})^2\big].$$

Thus, an application of Sudakov-Fernique's comparison inequality ([Ver18, Theorem 7.2.11]) yields

$$w_\wedge(C) = w(C \cap \mathbb{S}^{n-1}) = \mathbb{E}\big[\sup_{x\in\tilde{K}} X_x\big] \leq \mathbb{E}\big[\sup_{x\in\tilde{K}} Y_x\big] = w(\tilde{K}).$$

To conclude, we observe that $\tilde{K}$ forms a convex polytope with vertices belonging to the set $\{\boldsymbol{x}_i / \langle \boldsymbol{x}_i, \boldsymbol{\theta} \rangle : i \in [k]\} \subseteq B_2^n(\cos(\alpha)^{-1}) \cap \mathcal{H}$. Hence, its translated version $K := \boldsymbol{P}_\perp^{\boldsymbol{\theta}}(\tilde{K}) = \tilde{K} - \boldsymbol{\theta}$ satisfies $K \subseteq B_2^n(\tan(\alpha))$, where $\boldsymbol{P}_\perp^{\boldsymbol{\theta}}$ denotes the orthogonal projection onto span $(\boldsymbol{\theta})^\perp$. Lemma C.1 then yields the desired bound

$$w_\wedge(C) \leq w(\tilde{K}) = w(K) = \tan(\alpha) \cdot w(K/\tan(\alpha)) \leq \tan(\alpha) \cdot \sqrt{2\log(k)},$$

where the first equality follows from the translation invariance of the Gaussian width [Ver18, Prop. 7.5.2].

Remark C.2 An alternative, more geometric argumentation allows to show the related bound $w_\wedge(C) \leq w(\boldsymbol{P}_\perp^{\boldsymbol{\theta}}(C \cap \mathbb{S}^{n-1})) + 1/\sqrt{2\pi}$, which is visualized in Figure 2 (the set $\boldsymbol{P}_\perp^{\boldsymbol{\theta}}(C \cap \mathbb{S}^{n-1})$ corresponds to the thick line in the right view). Indeed, we may decompose the mean width into

$$w_\wedge(C) = \mathbb{E}\Big[\sup_{\boldsymbol{h} \in C \cap \mathbb{S}^{n-1}} \langle \boldsymbol{g}, \boldsymbol{h} \rangle\Big] \leq \mathbb{E}\Big[\sup_{\boldsymbol{h} \in C \cap \mathbb{S}^{n-1}} \langle \boldsymbol{P}_\perp^{\boldsymbol{\theta}} \boldsymbol{g}, \boldsymbol{P}_\perp^{\boldsymbol{\theta}} \boldsymbol{h} \rangle\Big] + \mathbb{E}\Big[\sup_{\boldsymbol{h} \in C \cap \mathbb{S}^{n-1}} \langle \boldsymbol{P}^{\boldsymbol{\theta}} \boldsymbol{g}, \boldsymbol{P}^{\boldsymbol{\theta}} \boldsymbol{h} \rangle\Big],$$

where $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \mathbf{Id})$, and $\boldsymbol{P}^{\boldsymbol{\theta}}, \boldsymbol{P}_\perp^{\boldsymbol{\theta}}$ denote the orthogonal projections onto span $(\boldsymbol{\theta})$ and span $(\boldsymbol{\theta})^\perp$, respectively. The first term on the right hand side equals $w(\boldsymbol{P}_\perp^{\boldsymbol{\theta}}(C \cap \mathbb{S}^{n-1}))$. Due to $\boldsymbol{P}^{\boldsymbol{\theta}} \boldsymbol{h} = \lambda \boldsymbol{\theta}$ with $0 \leq \lambda \leq 1$ for $\boldsymbol{h} \in C \cap \mathbb{S}^{n-1}$, the second summand can be bounded by

$$\mathbb{E}\Big[\sup_{\boldsymbol{h} \in C \cap \mathbb{S}^{n-1}} \langle \boldsymbol{P}^{\boldsymbol{\theta}} \boldsymbol{g}, \boldsymbol{P}^{\boldsymbol{\theta}} \boldsymbol{h} \rangle\Big] \leq \mathbb{E}\Big[\max\{0, \langle \boldsymbol{P}^{\boldsymbol{\theta}} \boldsymbol{g}, \boldsymbol{\theta} \rangle\}\Big] = \frac{1}{\sqrt{2\pi}},$$

where the equality follows from $\langle \boldsymbol{P}^{\boldsymbol{\theta}} \boldsymbol{g}, \boldsymbol{\theta} \rangle \sim \mathcal{N}(0,1)$. If $\boldsymbol{\theta} \in C$, we have that $\boldsymbol{P}_\perp^{\boldsymbol{\theta}}(C \cap \mathbb{S}^{n-1}) \subseteq K$, resulting in the desired bound $w_\wedge(C) \leq 1/\sqrt{2\pi} + w(K) \leq 1/\sqrt{2\pi} + \tan(\alpha) \cdot \sqrt{2\log(k)}$. ♢

C.3 Proofs of Section 4.2.2

Descent Cone of ℓ^1-Norm (Lemma 4.13) We begin by showing a polyhedral description of the descent cone of the ℓ^1-norm:

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

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

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

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

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

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

Lineality of Descent Cone of ℓ^1-Norm (Lemma 4.15) Next, we describe the lineality space and lineality of $\mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z})$:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Lineality and Range for Gauge (Proposition 4.18) Lastly, we characterize the range and lineality of $\mathcal{D}_\wedge(p_{\boldsymbol{D} \cdot B_1^d}, \boldsymbol{x}_0)$:

First, observe that a combination of Lemma 3.6 and Lemma 4.13 yields that

$$\begin{aligned}
\mathcal{D}_\wedge(p_{\boldsymbol{D} \cdot B_1^d}, \boldsymbol{x}_0) &= \boldsymbol{D} \cdot \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}) \\
&= \boldsymbol{D} \cdot \operatorname{cone}(\pm \bar{s} \cdot \boldsymbol{e}_i - \operatorname{sign}(z_{\ell^1}) : i \in [d]) \\
&= \operatorname{cone}(\pm \bar{s} \cdot \boldsymbol{d}_i - \boldsymbol{D} \operatorname{sign}(z_{\ell^1}) : i \in [d]).
\end{aligned}$$

By Lemma 4.15, we know how to characterize the lineality of $\mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1})$. Note that for any convex set $C \subseteq \mathbb{R}^d$, it holds true that $(\boldsymbol{D} \cdot C)_L \supseteq \boldsymbol{D} \cdot C_L$, however, the reverse inclusion is not satisfied, in general. Hence, Lemma 3.6 immediately implies $\left(\mathcal{D}_\wedge(p_{\boldsymbol{D} \cdot B_1^d}, \boldsymbol{x}_0)\right)_L \supseteq \boldsymbol{D} \cdot (\mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}))_L$. For proving the reverse inclusion $\left(\mathcal{D}_\wedge(p_{\boldsymbol{D} \cdot B_1^d}, \boldsymbol{x}_0)\right)_L \subseteq \boldsymbol{D} \cdot (\mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}))_L$, we will now show that if $\left(\mathcal{D}_\wedge(p_{\boldsymbol{D} \cdot B_1^d}, \boldsymbol{x}_0)\right)_L \not\subseteq \boldsymbol{D} \cdot (\mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}))_L$, then z_{ℓ^1} did not have maximal support. To that end, pick any vector $\boldsymbol{x} \in \left(\mathcal{D}_\wedge(p_{\boldsymbol{D} \cdot B_1^d}, \boldsymbol{x}_0)\right)_L \setminus \boldsymbol{D} \cdot (\mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}))_L$ and write $\boldsymbol{x} = \boldsymbol{D} \cdot z^1$,

where $z^1 \in \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}) \setminus (\mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}))_L$. Since $x \in \left(\mathcal{D}_\wedge(p_{D\cdot B_1^d}, x_0)\right)_L$, we can also chose a $z^2 \in \mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}) \setminus (\mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}))_L$ with $-x = D \cdot z^2$. Due to $z^i \notin (\mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}))_L$ for $i = 1, 2$, we have that for all $\varepsilon > 0$

$$\left\| z_{\ell^1} - \varepsilon \cdot z^i \right\|_1 > \|z_{\ell^1}\|_1, \tag{C.1}$$

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

$$\left\| z_{\ell^1} + \varepsilon \cdot z^i \right\|_1 \leq \|z_{\ell^1}\|_1. \tag{C.2}$$

For small enough $\varepsilon > 0$, inequality (C.1) implies that

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

whereas (C.2) means that

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

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

$$\left\| z^\delta \right\|_1 = \|z_{\ell^1}\|_1 + \delta \cdot \sum_{j \in \bar{\mathcal{S}}} \operatorname{sign}(z_{\ell^1, j}) \cdot (z_j^1 + z_j^2) + \delta \cdot \sum_{j \in \bar{\mathcal{S}}^c} |z_j^1 + z_j^2|,$$

and we can conclude that $\sum_{j \in \bar{\mathcal{S}}^c} |z_j^1 + z_j^2| > 0$. However, this means that there is at least one $j \in \bar{\mathcal{S}}^c$ such that $z_j^\delta \neq 0$, which shows that z was indeed not maximal. Finally, Lemma 4.15 implies that

$$\dim\left(\left(\mathcal{D}_\wedge(p_{D\cdot B_1^d}, x_0)\right)_L\right) = \dim\left((\mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}))_L\right) - \dim\left(\ker D_{|(\mathcal{D}_\wedge(\|\cdot\|_1, z_{\ell^1}))_L}\right) \leq \bar{s} - 1,$$

which concludes the proof of first part of the proposition concerning the lineality of $\mathcal{D}_\wedge(p_{D\cdot B_1^d}, x_0)$.

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

$$P_{C_L^\perp}(r_i^-) = -2 \cdot P_{C_L^\perp}(D \cdot \operatorname{sign}(z_{\ell^1})) = r_j^{+\perp} + r_j^{-\perp}.$$

Thus, $P_{C_L^\perp}(r_i^-) \in \operatorname{cone}(r_j^{\pm\perp}, j \in \bar{\mathcal{S}}^c)$, which concludes the proof.

C.4 Proof of Theorem 4.19

Let $C = \mathcal{D}_\wedge(p_{D\cdot B_1^d}, x_0)$ and use the orthogonal decomposition provided in Proposition 4.18:

$$\mathcal{D}_\wedge(p_{D\cdot B_1^d}, x_0) = C_L \oplus C_R.$$

This enables us to estimate

$$w_\wedge^2(C) \overset{(1)}{\leq} \delta(C) \overset{(2)}{\leq} \delta(C_L) + \delta(C_R) \overset{(3)}{\leq} \dim(C_L) + w_\wedge^2(C_R) + 1, \tag{C.3}$$

where δ denotes the *statistical dimension*; see [ALMT14, Prop. 10.2] for a justification of (1). Using the statistical dimension as a summary parameter for convex cones brings several advantages; see [ALMT14, Prop. 3.1]: 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.3) is essentially tight.

Proposition 4.18 makes it possible to upper bound $\dim(C_L) + 1$ by $\bar{s}$. The statement then follows by applying Theorem 4.10 to the $2(d - \bar{s})$-polyhedral α-cone C_R.

C.5 Proof of Proposition 4.22 (Coherence Bound)

First, observe that we have

$$\tan^2\left(\angle(\boldsymbol{a}, \boldsymbol{a}+\boldsymbol{b})\right) = \frac{\|\boldsymbol{a} \times (\boldsymbol{a}+\boldsymbol{b})\|_2^2}{\langle \boldsymbol{a}, \boldsymbol{a}+\boldsymbol{b}\rangle^2} = \frac{\|\boldsymbol{a}\|_2^2 \|\boldsymbol{b}\|_2^2 - \langle \boldsymbol{a}, \boldsymbol{b}\rangle^2}{\left(\|\boldsymbol{a}\|_2^2 + \langle \boldsymbol{a}, \boldsymbol{b}\rangle\right)^2} \leq \frac{\|\boldsymbol{a}\|_2^2 \|\boldsymbol{b}\|_2^2}{\left(\|\boldsymbol{a}\|_2^2 + \langle \boldsymbol{a}, \boldsymbol{b}\rangle\right)^2}, \tag{C.4}$$

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

Observe that the assumptions of Proposition 4.18 are satisfied. Indeed, $s < \frac{1}{2}(1 + \mu^{-1}(\boldsymbol{D}))$ guarantees that $\boldsymbol{z}_{\ell^1}$ is the unique minimal ℓ^1-representer of the associated signal $\boldsymbol{D}\boldsymbol{z}_{\ell^1}$ and that $\boldsymbol{D}\boldsymbol{z}_{\ell^1} \neq \boldsymbol{0}$ [DE03; GN03]. Hence, we want to evaluate the circumangle of the cone generated by the vectors $\boldsymbol{r}_j^{\pm\perp} = \boldsymbol{P}_{C_L^\perp}(\pm s \cdot \boldsymbol{d}_j - \boldsymbol{D}\operatorname{sign}(\boldsymbol{z}_{\ell^1}))$ for $j \in S^c$, where $\mathcal{S} = \operatorname{supp}(\boldsymbol{z}_{\ell^1})$. As a proxy for the circumcenter, we can consider the vector $\boldsymbol{v} = -\boldsymbol{P}_{C_L^\perp}(\boldsymbol{D}\operatorname{sign}(\boldsymbol{z}_{\ell^1}))$ and therefore obtain:

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

We can now use the inequality (C.4) with $\boldsymbol{a} = \boldsymbol{v}$ and $\boldsymbol{b} = s \cdot \boldsymbol{P}_{C_L^\perp}(\boldsymbol{d}_j)$; note that $\boldsymbol{v} \neq \boldsymbol{0}$, since otherwise we would have $\boldsymbol{D}\boldsymbol{z}_{\ell^1} = \boldsymbol{0}$. The expression (C.4) is decreasing w.r.t. $\|\boldsymbol{a}\|_2^2$. Hence, we shall find a lower bound for $\|\boldsymbol{v}\|_2^2$. The projection $\boldsymbol{P}_{C_L^\perp}(\boldsymbol{D}\operatorname{sign}(\boldsymbol{z}_{\ell^1}))$ can be written as $\boldsymbol{D}\operatorname{sign}(\boldsymbol{z}_{\ell^1}) + \boldsymbol{w}$ for some vector $\boldsymbol{w} \in C_L$. According to the characterization of the lineality space C_L in Proposition 4.18, this amounts to saying that there exist coefficients $(c_i)_{i \in \mathcal{S}}$ such that

$$\boldsymbol{P}_{C_L^\perp}(\boldsymbol{D}\operatorname{sign}(\boldsymbol{z}_{\ell^1})) = \sum_{i \in \mathcal{S}} c_i \cdot \operatorname{sign}(z_{\ell^1,i}) \cdot \boldsymbol{d}_i, \text{ with } \sum_{i \in \mathcal{S}} c_i = s. \tag{C.5}$$

This yields:

$$\begin{aligned}
\|\boldsymbol{v}\|_2^2 &\geq \inf_{\boldsymbol{c}=(c_i)_{i\in\mathcal{S}}, \sum_i c_i = s} \left\| \sum_{i \in \mathcal{S}} c_i \operatorname{sign}(z_{\ell^1,i}) \boldsymbol{d}_i \right\|_2^2 \\
&= \inf_{\boldsymbol{c}=(c_i)_{i\in\mathcal{S}}, \sum_i c_i = s} \|\boldsymbol{c}\|_2^2 + \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}, j \neq i} c_i c_j \langle \operatorname{sign}(z_{\ell^1,i}) \boldsymbol{d}_i, \operatorname{sign}(z_{\ell^1,j}) \boldsymbol{d}_j \rangle \\
&\geq \inf_{\boldsymbol{c}=(c_i)_{i\in\mathcal{S}}, \sum_i c_i = s} \|\boldsymbol{c}\|_2^2 - \mu \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}, j \neq i} |c_i| \cdot |c_j| \\
&= \inf_{\boldsymbol{c}=(c_i)_{i\in\mathcal{S}}, \sum_i c_i = s} (1+\mu)\|\boldsymbol{c}\|_2^2 - \mu \sum_{i,j \in \mathcal{S}} |c_i||c_j| \\
&= \inf_{\boldsymbol{c}=(c_i)_{i\in\mathcal{S}}, \sum_i c_i = s} (1+\mu)\|\boldsymbol{c}\|_2^2 - \mu \Big(\sum_{i \in \mathcal{S}} |c_i|\Big)^2 \\
&\geq \inf_{\boldsymbol{c}=(c_i)_{i\in\mathcal{S}}, \sum_i c_i = s} (1+\mu)\|\boldsymbol{c}\|_2^2 - \mu(\sqrt{s}\|\boldsymbol{c}\|_2)^2 \\
&= \inf_{\boldsymbol{c}=(c_i)_{i\in\mathcal{S}}, \sum_i c_i = s} \|\boldsymbol{c}\|_2^2 (1 + \mu - \mu s) \\
&= s(1 + \mu - \mu s) \\
&\geq s(1 - \mu s),
\end{aligned}$$

where we have used that $\|\boldsymbol{d}_i\|_2 = 1$ in the first equality. Together with the following inequalities:

$$|\langle \boldsymbol{a}, \boldsymbol{b} \rangle| = s \left| \left\langle \boldsymbol{v}, \boldsymbol{P}_{C_L^\perp} \boldsymbol{d}_j \right\rangle \right| \leq s \left| \langle \boldsymbol{v}, \boldsymbol{d}_j \rangle \right| \overset{(C.5)}{\leq} s^2 \sup_{i \neq j} |\langle \boldsymbol{d}_i, \boldsymbol{d}_j \rangle| = s^2 \mu,$$

$$\|\boldsymbol{b}\|_2^2 = s^2 \left\| \boldsymbol{P}_{C_L^\perp} \boldsymbol{d}_j \right\|_2^2 \leq s^2 \|\boldsymbol{d}_j\|_2^2 = s^2,$$

we obtain the desired bound

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

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 3.7, we only report the quantity $w_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; \boldsymbol{z}_0))$, ignoring for instance the probability parameter u, cf. [ALMT14].

Some Details on Computations Unless stated otherwise, we solve the convex recovery programs such as ($\mathrm{BP}_\eta^{\mathrm{coef}}$) or ($\mathrm{BP}_{\ell^1}$) using the `Matlab` toolbox `cvx` [GB08; GB14]. We employ the default settings and set the precision to `best`. For creating phase transitions, a solution $\hat{\boldsymbol{x}}$ is considered to be "perfectly recovered" if the error to the ground truth vector $\boldsymbol{x}_0$ satisfies $\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\|_2 \leq 10^{-5}$. 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 $w_\wedge^2(C) = w(C \cap \mathbb{S}^{n-1})$ 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, which can be computed as $\delta(C) = \mathbb{E}[\|\Pi_C(\boldsymbol{g})\|_2^2]$; see [ALMT14, Prop. 3.1 & Prop. 10.2]. In order to obtain an approximation of $\delta(C)$, we draw k independent samples $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_k \sim \mathcal{N}(\boldsymbol{0}, \mathbf{Id})$ and for each of them we evaluate the projection $\Pi_C(\boldsymbol{g}_i)$ 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 $\lambda_{\min}(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1}))$ is out of reach in general. In our numerical experiments on coefficient recovery, we nevertheless provide empirical upper bounds on $\lambda_{\min}(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1}))$. Those are obtained as follows: Let $\boldsymbol{x}_0 = \boldsymbol{D} \cdot \boldsymbol{z}_{\ell^1}$ and consider the perturbed $\tilde{\boldsymbol{x}}_{\boldsymbol{0}} = \boldsymbol{x}_0 + \hat{\boldsymbol{e}}$, where $\hat{\boldsymbol{e}} \in \mathbb{R}^n$ such that $\|\hat{\boldsymbol{e}}\|_2 \leq \hat{\eta}$. We then define $\hat{\boldsymbol{z}} \in \mathbb{R}^d$ as a solution to the program

$$\min_{\boldsymbol{z} \in \mathbb{R}^d} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \|\tilde{\boldsymbol{x}}_{\boldsymbol{0}} - \boldsymbol{D}\boldsymbol{z}\|_2 \leq \hat{\eta}.$$

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

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

$$\boldsymbol{\theta} \in \operatorname*{argmin}_{\boldsymbol{v} \in B_2^n} \max_{i \in [k]} \langle -\boldsymbol{v}, \boldsymbol{x}_i \rangle, \tag{D.1}$$

where the vectors $\boldsymbol{x}_i$ are the normalized generators of a nontrivial pointed polyhedral cone; see Proposition 4.8. This problem is closely related to the so-called *smallest bounding sphere problem* [Syl57], 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 $\partial g(\boldsymbol{v}) = \operatorname{conv}(-\boldsymbol{x}_i, i \in I(\boldsymbol{v}))$ and the optimality conditions read

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

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`.