A New Perspective on the Sample Complexity of the Analysis Basis Pursuit

Martin Genzel
Institut für Mathematik
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
Email: genzel@math.tu-berlin.de

Gitta Kutyniok
Institut für Mathematik
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
Email: kutyniok@math.tu-berlin.de

Maximilian März
Institut für Mathematik
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
Email: maerz@math.tu-berlin.de

***Abstract*—In this article, we summarize our findings concerning signal estimation from undersampled Gaussian measurements under the assumption of a cosparse model. Based on generalized notions of sparsity, we obtain a novel recovery guarantee for the ℓ^1-analysis basis pursuit, enabling accurate predictions of its sample complexity. The corresponding bounds on the number of required measurements do explicitly depend on the Gram matrix of the analysis operator and therefore particularly account for its mutual coherence structure. Our findings defy conventional wisdom which promotes the sparsity of analysis coefficients as the crucial quantity to study. In fact, this common paradigm breaks down completely in many situations of practical interest, for instance, when applying a redundant (multilevel) frame as analysis prior. A detailed exposition of the results can be found in our original work, which contains proofs for all statements and extensive numerical experiments.**

I. Introduction

Ever since Candès, Donoho, Romberg, and Tao initiated the field of *compressed sensing* ([1], [2]), sparsity has become a fundamental model in many different signal processing tasks. The standard setup in this field considers the problem of reconstructing an unknown *sparse signal vector* $\boldsymbol{x}^* \in \mathbb{R}^n$ from *non-adaptive, linear measurements* of the form

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

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a known *measurement matrix* and $\boldsymbol{e} \in \mathbb{R}^m$ captures potential distortions, typically due to noise. Even in highly underdetermined situations where $m \ll n$, the methodology of compressed sensing allows for a perfect recovery of $\boldsymbol{x}^*$ by employing sparsity promoting priors. A widely-used algorithmic approach suggests to solve the so-called *basis pursuit*

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

where $\eta \geq 0$ is chosen such that $\|\boldsymbol{e}\|_2 \leq \eta$. The distinguished feature of (BP$_\eta$) is the use of the ℓ^1-norm as objective function, which is known to encourage sparse solutions. One of the key findings in compressed sensing is that (BP$_\eta$) indeed recovers an S-sparse vector $\boldsymbol{x}^*$ (i.e., $\|\boldsymbol{x}^*\|_0 \coloneqq |\mathrm{supp}(\boldsymbol{x}^*)| \leq S$) with the optimal sampling rate of $m = O(S\cdot\log(2S/n))$, provided that the measurement design $\boldsymbol{A}$ is generated according to an appropriate random distribution.

Unfortunately, the assumption that $\boldsymbol{x}^*$ is sparse by itself is typically not satisfied in most real-world applications. However, it is widely accepted that many signals of interest can be sparsely represented with regard to a suitable transformation, i.e., $\boldsymbol{x}^* = \boldsymbol{D}\boldsymbol{z}$ for a (possibly redundant) dictionary $\boldsymbol{D} \in \mathbb{R}^{n\times N}$ and a sparse coefficient vector $\boldsymbol{z} \in \mathbb{R}^N$. This prior information is then utilized in the modified basis pursuit

$$\min_{\boldsymbol{z}\in\mathbb{R}^N} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{A}\boldsymbol{D}\boldsymbol{z} - \boldsymbol{y}\|_2 \leq \eta, \tag{BPS$_\eta^{\boldsymbol{D}}$}$$

which is typically referred to as the *synthesis formulation of compressed sensing*.

In this work, we study the related, yet fundamentally different *analysis sparsity model* (also known as *cosparse model*), which has gained increasing attention within the past years ([3], [4], [5]). Instead of asking for a synthesis sparse representation of $\boldsymbol{x}^*$, we test ("analyze") the signal with a collection of *analysis vectors* $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_N \in \mathbb{R}^n$, i.e., one computes

$$\boldsymbol{\Psi}\boldsymbol{x}^* = (\langle\boldsymbol{\psi}_1, \boldsymbol{x}^*\rangle, \dots, \langle\boldsymbol{\psi}_N, \boldsymbol{x}^*\rangle) \in \mathbb{R}^N,$$

where the matrix $\boldsymbol{\Psi} \coloneqq [\boldsymbol{\psi}_1 \dots \boldsymbol{\psi}_N]^\mathsf{T} \in \mathbb{R}^{N\times n}$ is called the *analysis operator*. If $\boldsymbol{\Psi}$ is able to reflect the underlying structure of $\boldsymbol{x}^*$, one might expect that these *analysis coefficients* are dominated by only a few large entries. This assumption is incorporated in the following generalization of (BP$_\eta$) that is typically referred to as the *analysis basis pursuit* (or ℓ^1-*analysis minimization*):

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

It was first noted in the seminal work [6] that, despite their algebraic similarities, the synthesis- and analysis-based programs are in fact very different if the involved operators $\boldsymbol{\Psi}$ and $\boldsymbol{D}$ are *redundant*. While the intuitive synthesis formulation (BPS$_\eta^{\boldsymbol{D}}$) has established itself as the predominant methodology, it turned out that (BP$_\eta^{\boldsymbol{\Psi}}$) is surprisingly effective for numerous problem settings, such as in *total variation minimization* [7] or for *multiscale transforms* in classical signal and image processing tasks ([8]). In many situations of interest, one can even observe that the analysis formulation outperforms its synthesis-based counterpart ([6], [9]). However, despite this empirical success, theoretical properties of the analysis basis pursuit remain largely unexplored and "its rigorous understanding is still in its infancy" [5, p. 174].

In this work, we summarize some of our recent advances on the sample complexity of ($\mathrm{BP}_\eta^{\boldsymbol{\Psi}}$) as presented in the original paper [10]. In Section 2, we will first discuss the dominant sparsity-based interpretation of the analysis basis pursuit ($\mathrm{BP}_\eta^{\boldsymbol{\Psi}}$) and show a numerical example indicting that this common paradigm does not lead to a sound theoretical foundation of redundant transforms. Our main result is detailed in Section 3 together with a numerical verification of its predictive power. In the last section, we finally conclude with a brief discussion of available extensions in [10].

II. Previous Approaches and Open Problems

In order to achieve recovery of $\boldsymbol{x}^*$ via ($\mathrm{BP}_\eta^{\boldsymbol{\Psi}}$), the essential task is to come up with an analysis operator $\boldsymbol{\Psi}$ that results in a coefficient vector $\boldsymbol{\Psi}\boldsymbol{x}^*$ of "low complexity". Inspired by the classical setup of compressed sensing, where $\boldsymbol{\Psi}$ is simply the identity matrix, a large part of the pertinent literature is build upon the assumption that $\boldsymbol{\Psi}\boldsymbol{x}^*$ is (almost) *sparse*. Although many of these approaches rely on different proof strategies, e.g., the $\boldsymbol{D}$-RIP [3] or conic geometry [5], they eventually promote results of a very similar type: Recovery of $\boldsymbol{x}^* \in \mathbb{R}^n$ via ($\mathrm{BP}_\eta^{\boldsymbol{\Psi}}$) succeeds if the number of measurements obeys

$$m \geq C \cdot S \cdot \mathrm{PolyLog}(\tfrac{2N}{S}), \tag{1}$$

where $S \coloneqq \|\boldsymbol{\Psi}\boldsymbol{x}^*\|_0$ and $C > 0$ is a constant that might depend on $\boldsymbol{\Psi}$. Although such a sparsity-based approach is appealing due to its simplicity and the fact that it is consistent with the classical setup, it remains unclear whether it is sufficient for a general foundation of ($\mathrm{BP}_\eta^{\boldsymbol{\Psi}}$). The notion of analysis sparsity is merely determined by the support of $\boldsymbol{\Psi}\boldsymbol{x}^*$, which in turn does not take into account the coherence pattern of the individual analysis vectors $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_N$; in other words, the underlying "geometry" of $\boldsymbol{\Psi}$ is ignored.

As an illustration of our concern, let us consider a simple example: Figure 1 shows the results of a numerical simulation that reconstructs a block-signal (see Figure 1(a)) using three different analysis operators. The plot of Figure 1(b) exhibits the phase transition behavior of ($\mathrm{BP}_{\eta=0}^{\boldsymbol{\Psi}}$) for a *redundant, discrete Haar wavelet transform* $\boldsymbol{\Psi}_{\mathtt{rdwt}}$ and the analysis operator $\boldsymbol{\Psi}_{\mathtt{irdwt}}$ associated with the inverse wavelet transform. Although the sparsity $S = 906$ is exactly the same in both cases, their recovery capability indeed differs dramatically! We conclude that just investigating the sparsity does by far not explain why the transition of $\boldsymbol{\Psi}_{\mathtt{irdwt}}$ happens much earlier ($m \approx 85$) than the one of $\boldsymbol{\Psi}_{\mathtt{rdwt}}$ ($m \approx 240$). Even more striking, the prediction of (1) deviates from the truth by orders of magnitudes. The plot of Figure 1(c) reveals another insight: While (1) is reliable for an *orthonormal Haar wavelet transform* $\boldsymbol{\Psi}_{\mathtt{dwt}} \in \mathbb{R}^{n\times n}$, a comparison with the actual recovery rate of ($\mathrm{BP}_{\eta=0}^{\boldsymbol{\Psi}_{\mathtt{irdwt}}}$) indicates that *redundancy* can be beneficial in the analysis model. Finally, it is also worth mentioning that a *compressibility* argument does not "save the day" here: Figure 1(d) demonstrates that the transition takes place far before the remaining coefficients would be negligibly small. We emphasize that the above example is not too specific or artificial, but rather illustrates a scenario that often occurs in applications: Due to linear dependencies within $\boldsymbol{\Psi}$, the analysis sparsity oftentimes cannot become arbitrarily small. For instance, if $\boldsymbol{\Psi}$ corresponds to a highly redundant dictionary, one typically has $S \gg n$, whereas the true sample complexity is relatively small (below the space dimension n).

Finally, let us point out that another influential approach to the analysis formulation is given by [4], where similar concerns are raised and the *cosparse analysis model* is introduced. A remarkable observation of [4] is that the location of vanishing coefficients in $\boldsymbol{\Psi}\boldsymbol{x}^*$ is the driving force behind the analysis formulation, rather than the number of non-zero components. This viewpoint naturally leads to the so-called *cosparse signal model*, which is typically described by *union-of-subspaces*. However, this methodology is also not capable of explaining the example of Figure 1(b), since for both frames the cosparsity $L \coloneqq N - S$ and the underlying cosparse structure coincide; see [10] for details. Therefore [4] neither provide an accurate sampling rate for ($\mathrm{BP}_\eta^{\boldsymbol{\Psi}}$) nor can it explain the performance gap between both examples. The interested reader is referred to [10, Section 4] for an extensive discussion of the related literature. We conclude this section by raising the following fundamental question which we aimed to answer in [10]:

> *If (co-)sparsity does not fully explain what is happening, which general principles lead to success or failure of the analysis basis pursuit* ($\mathrm{BP}_\eta^{\boldsymbol{\Psi}}$)*?*

III. Main Result

Let us begin by stating the general assumptions for our recovery analysis of ($\mathrm{BP}_\eta^{\boldsymbol{\Psi}}$). For the sake of brevity, we restrict ourselves to the prototypical case of noiseless Gaussian measurements, but remark that the results presented below also hold true for noisy sub-Gaussian measurements [10].

Model 1 (Linear Gaussian Measurements)**.** *Let $\boldsymbol{x}^* \in \mathbb{R}^n$ be a fixed* signal vector*. The* measurement vectors *$\boldsymbol{a}_1, \dots, \boldsymbol{a}_m \in \mathbb{R}^n$ are assumed to be independent copies of a standard Gaussian random vector $\boldsymbol{a} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$. These vectors form the rows of the* measurement matrix *$\boldsymbol{A} \coloneqq [\boldsymbol{a}_1 \dots \boldsymbol{a}_m]^\intercal \in \mathbb{R}^{m\times n}$. The actual* measurements *of $\boldsymbol{x}^*$ are then given by*

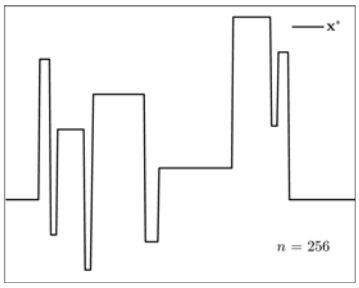
$$\boldsymbol{y} \coloneqq \boldsymbol{A}\boldsymbol{x}^* \in \mathbb{R}^m.$$

In the noiseless case, the analysis basis pursuit then takes the form

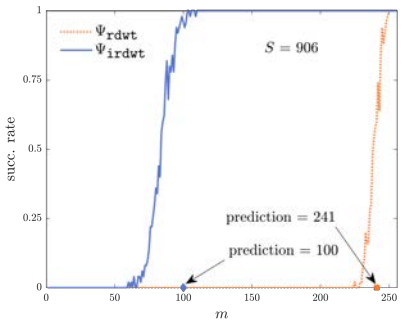
$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{\Psi}\boldsymbol{x}\|_1 \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}, \tag{$\mathrm{BP}_{\eta=0}^{\boldsymbol{\Psi}}$}$$

so that we can even hope for an *exact* retrieval of $\boldsymbol{x}^* \in \mathbb{R}^n$ from $\boldsymbol{y}$. Indeed, the recent work of [11] has made the remarkable observation that a convex program of the above type typically undergoes a sharp *phase transition* as m varies: Recovery of $\boldsymbol{x}^*$ fails with overwhelmingly high probability if m is below a certain threshold. But once m exceeds a small transition region, recovery succeeds with overwhelmingly high probability; see Figure 1(b) for an example. This minimal number of required measurements (also depending on the desired probability of success) is often referred to as the *sample complexity* (or *optimal sampling rate*) of an estimation problem. Before stating an upper bound on the sample complexity of ($\mathrm{BP}_{\eta=0}^{\boldsymbol{\Psi}}$), we first need to introduce more notation related to the analysis operator:

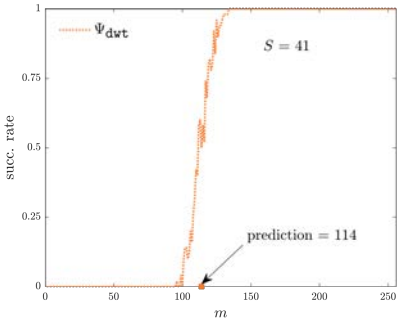
Definition 2. *(1) The matrix $\boldsymbol{\Psi} = [\psi_{k,j}] \in \mathbb{R}^{N\times n}$ is called an* analysis operator *(or* analysis matrix*) if none of its rows equals the zero vector. The rows of $\boldsymbol{\Psi}$, denoted by $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_N \in \mathbb{R}^n$,*



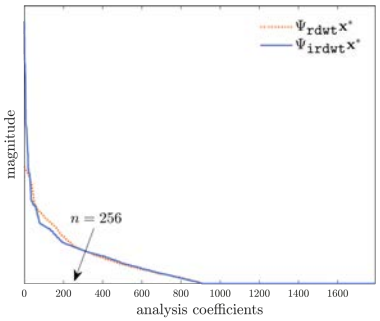
(a) Signal vector $\boldsymbol{x}^* \in \mathbb{R}^{256}$.



(b) Success rate of exact recovery via $(\mathrm{BP}_{\eta=0}^{\boldsymbol{\Psi}_{\mathtt{rdwt}}})$ and $(\mathrm{BP}_{\eta=0}^{\boldsymbol{\Psi}_{\mathtt{irdwt}}})$, respectively.



(c) Success rate of exact recovery via $(\mathrm{BP}_{\eta=0}^{\boldsymbol{\Psi}_{\mathtt{dwt}}})$.



(d) Sorted and ℓ^1-normalized magnitudes of the coefficients $\boldsymbol{\Psi}_{\mathtt{rdwt}}\boldsymbol{x}^*$ and $\boldsymbol{\Psi}_{\mathtt{irdwt}}\boldsymbol{x}^*$.

Figure 1. The phase transition of $(\mathrm{BP}_{\eta=0}^{\boldsymbol{\Psi}})$ for wavelets in 1D with noiseless Gaussian measurements ($\eta = 0$), see [10] for further details.

are called the analysis vectors. *Moreover, we define the* Gram matrix *of* $\boldsymbol{\Psi}$ *as*

$$\boldsymbol{G} = [g_{k,k'}] \coloneqq \boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathsf{T}} = [\langle \boldsymbol{\psi}_k, \boldsymbol{\psi}_{k'} \rangle] \in \mathbb{R}^{N \times N}.$$

(2) The analysis coefficients *of a vector* $\boldsymbol{x}^* \in \mathbb{R}^n$ *(with respect to* $\boldsymbol{\Psi}$*) are given by*

$$\boldsymbol{\Psi}\boldsymbol{x}^* = (\langle \boldsymbol{\psi}_1, \boldsymbol{x}^* \rangle, \dots, \langle \boldsymbol{\psi}_N, \boldsymbol{x}^* \rangle) \in \mathbb{R}^N.$$

The analysis support *of* $\boldsymbol{x}^*$ *is denoted by* $\mathcal{S} \coloneqq \operatorname{supp}(\boldsymbol{\Psi}\boldsymbol{x}^*) \subset [N]$*, and if* $S = |\mathcal{S}|$*, we say that* $\boldsymbol{x}^*$ *is* S-analysis-sparse. *Analogously, we call the complement* $\mathcal{S}^c \coloneqq \operatorname{supp}(\boldsymbol{\Psi}\boldsymbol{x}^*)^c \subset [N]$ *the* analysis cosupport *of* $\boldsymbol{x}^*$ *and speak of an* L-analysis-cosparse *vector if* $L = |\mathcal{S}^c| = N - S$.

Let us continue by defining three generalized notions of (co-)sparsity, which form the key ingredients of our bounds on the sampling rate:

Definition 3 (Generalized (Co-)Sparsity)**.** *Let* $\boldsymbol{\Psi} \in \mathbb{R}^{N \times n}$ *be an analysis operator and let* $\boldsymbol{x}^* \in \mathbb{R}^n$*. We define the* generalized sparsity *of* $\boldsymbol{x}^*$ *(with respect to* $\boldsymbol{\Psi}$*) by*

$$\tilde{S} \coloneqq \sum_{k,k' \in \mathcal{S}} \operatorname{sign}(\langle \boldsymbol{\psi}_k, \boldsymbol{x}^* \rangle) \cdot \operatorname{sign}(\langle \boldsymbol{\psi}_{k'}, \boldsymbol{x}^* \rangle) \cdot g_{k,k'}.$$

Moreover, we introduce the terms

$$\tilde{L} \coloneqq \sum_{k,k' \in \mathcal{S}^c} \frac{g_{k,k'}^2}{\sqrt{g_{k,k} \cdot g_{k',k'}}}, \quad \tilde{L}_d \coloneqq \sum_{k \in \mathcal{S}^c} \sqrt{g_{k,k}},$$

which are both referred to as the generalized cosparsity *of* $\boldsymbol{x}^*$.

Considering the canonical choice of an orthonormal basis, it becomes actually clear why we speak of *generalized* sparsity: Since $\boldsymbol{G} = \boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathsf{T}} = \boldsymbol{I}_n$ in this case, one obtains $\tilde{S} = S = \|\boldsymbol{\Psi}\boldsymbol{x}^*\|_0$ and $\tilde{L} = \tilde{L}_d = L = n - S$. Hence, the respective parameters of Definition 3 precisely coincide with their traditional counterparts. In general, this correspondence is more complicated. The definition of the generalized sparsity $\tilde{S}$ still operates on the analysis support of $\boldsymbol{x}^*$, but also involves a weighted sum over all Gram matrix entries associated with $\mathcal{S}$. The same holds true for the generalized cosparsity term $\tilde{L}$, respectively. Such an incorporation of the off-diagonal entries of $\boldsymbol{G}$ is in fact quite natural because, to a certain extent, it reflects the mutual coherence structure of the analysis vectors $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_N$.

Our actual goal is to come up with an upper bound on the sample complexity of $(\mathrm{BP}_{\eta=0}^{\boldsymbol{\Psi}})$ that is tight in many situations. For this purpose, let us introduce the following function, which essentially determines the sampling rate proposed by Theorem 5 below:

Definition 4. *Let* $\boldsymbol{\Psi} \in \mathbb{R}^{N \times n}$ *be an analysis operator and let* $\boldsymbol{x}^* \in \mathbb{R}^n$ *with* $\boldsymbol{x}^* \notin \ker \boldsymbol{\Psi}$*. Then, we define the* sampling-rate function *of* $\boldsymbol{\Psi}$ *and* $\boldsymbol{x}^*$ *by*

$$\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{x}^*) \coloneqq \begin{cases} n - \frac{(\tilde{L}_d)^2}{\tilde{L}} \cdot \Phi\left(\frac{\tilde{S}}{\tilde{L}}\right), & \text{if } \mathcal{S}^c \neq \emptyset, \\ n, & \text{otherwise}, \end{cases}$$

where

$$\Phi(\rho) \coloneqq \operatorname{erf}\left(\frac{h^{-1}(\rho)}{\sqrt{2}}\right), \quad \rho > 0,$$

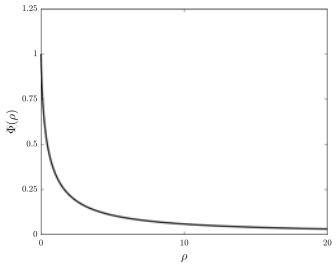


Figure 2. The graph of Φ. This function is strictly monotonically decreasing (in particular, $\lim_{\rho\searrow 0}\Phi'(\rho)=-\infty$), and at its boundary points, we have $\lim_{\rho\searrow 0}\Phi(\rho)=1$ and $\lim_{\rho\to\infty}\Phi(\rho)=0$.

with

$$h\colon (0,\infty)\to(0,\infty),\ \tau\mapsto \sqrt{\tfrac{2}{\pi}}\tfrac{e^{-\tau^2/2}}{\tau}+\operatorname{erf}(\tfrac{\tau}{\sqrt{2}})-1$$

and $\operatorname{erf}(\tau):=\frac{2}{\sqrt{\pi}}\int_0^\tau \exp(-x^2)\,dx$ *denoting the* error function*.*

It is not hard to see that the functions h, h^{-1}, and Φ — each one mapping from $(0,\infty)$ to $(0,\infty)$ — are well-defined, which also implies the well-definedness of $\mathcal{M}(\boldsymbol{\Psi},\boldsymbol{x}^*)$ and in particular that $\mathcal{M}(\boldsymbol{\Psi},\boldsymbol{x}^*)\geq 0$ [10, Appendix A.3]. Figure 2 shows the graph of Φ, visualizing how the ratio $\tilde{S}/\tilde{L}$ affects the sampling-rate function.

We are now ready to state our main result:

Theorem 5 (Exact Recovery via $(\mathrm{BP}^{\boldsymbol{\Psi}}_{\boldsymbol{\eta=0}})$)**.** *Assume that Model 1 is fulfilled. Let $\boldsymbol{\Psi}\in\mathbb{R}^{N\times n}$ be an analysis operator such that $\boldsymbol{x}^*\notin\ker\boldsymbol{\Psi}$. Then for every $u>0$, the following holds true with probability at least $1-e^{-u^2/2}$: If the number of measurements obeys*

$$m>\left(\sqrt{\mathcal{M}(\boldsymbol{\Psi},\boldsymbol{x}^*)}+u\right)^2+1, \tag{2}$$

then $(\mathrm{BP}^{\boldsymbol{\Psi}}_{\eta=0})$ *recovers $\boldsymbol{x}^*$ exactly.*

Theorem 5 gives a precise answer to the question raised in the introduction of this work: Roughly speaking, reconstruction succeeds with high probability as m slightly exceeds the sampling-rate function $\mathcal{M}(\boldsymbol{\Psi},\boldsymbol{x}^*)$. Unlike many approaches from the literature, the statement of Theorem 5 is highly non-uniform: the condition of (2) does not only involve the analysis sparsity $S=\|\boldsymbol{\Psi}\boldsymbol{x}^*\|_0$, but explicitly depends on the support $\mathcal{S}$ as well as on the sign vector $\operatorname{sign}(\boldsymbol{\Psi}\boldsymbol{x}^*)$. We refer the interested reader to the experiments in [10], which demonstrate that the performance of $(\mathrm{BP}^{\boldsymbol{\Psi}}_{\eta=0})$ is oftentimes not fully explainable by means of S alone, even if $\boldsymbol{\Psi}$ is fixed.

Remark. (a) The assumption of $\boldsymbol{x}^*\notin\ker\boldsymbol{\Psi}$ in Theorem 5 is not very restrictive and merely of technical nature. If $\boldsymbol{x}^*\in\ker\boldsymbol{\Psi}$, the analysis basis pursuit $(\mathrm{BP}^{\boldsymbol{\Psi}}_{\eta=0})$ uniquely recovers $\boldsymbol{x}^*$ if, and only if, $\ker\boldsymbol{\Psi}\cap(\boldsymbol{x}^*+\ker\boldsymbol{A})=\{\boldsymbol{x}^*\}$.

(b) Since $\mathcal{M}(\boldsymbol{\Psi},\boldsymbol{x}^*)\leq n$, the condition of (2) does not lead to situations where m needs to be much larger than n in order to achieve successful recovery. In contrast, this simple observation is not always reflected by a naive bound of the form (1), at least when the domain of analysis coefficients $\mathbb{R}^N$ is much higher dimensional, i.e., $N\gg n$.

(c) The function Φ is quite simple from an analytical perspective, yet it remains somewhat uninformative. Thus, accepting a certain loss of accuracy, the following upper bound gives further insight into its asymptotical behavior: Under the assumptions Theorem 5, we have that (cf. [10, Proposition 2.7])

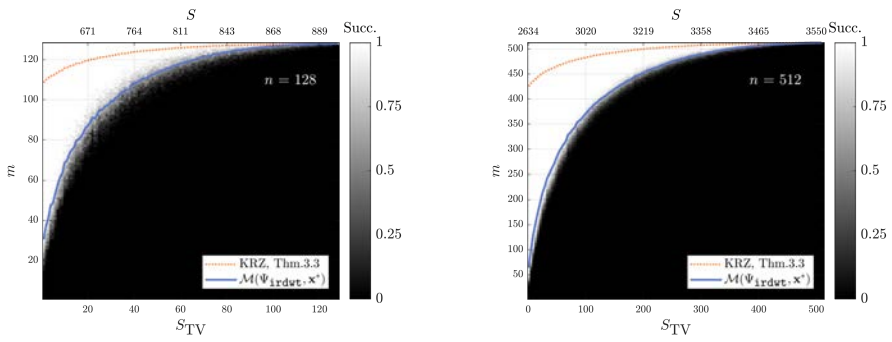
$$\mathcal{M}(\boldsymbol{\Psi},\boldsymbol{x}^*)\leq\min\Big\{n-\tfrac{\tilde{L}_d^2}{\tilde{L}}+(\tfrac{\tilde{L}_d}{\tilde{L}})^2\cdot[2\tilde{S}\cdot\log(\tfrac{\tilde{S}+\tilde{L}}{\tilde{S}})+\tilde{S}],\ n-\tfrac{2}{\pi}\cdot\tfrac{\tilde{L}_d^2}{\tilde{S}+\tilde{L}}\Big\}.$$

(d) The proof strategy of Theorem 5 is based on deriving a sophisticated, non-trivial upper bound on the *Gaussian mean width* of the descent cone of the ℓ^1-analysis functional. We emphasize that our careful analysis is particularly consistent with the standard setup of compressed sensing, where $\boldsymbol{\Psi}$ is an orthonormal basis, which can be easily seen from the bound in (c). This important feature is one of the key differences to the related work of [12], which follows a similar strategy. We refer to [10, Section 4.4] for a detailed comparison of both results.

While the prediction of Theorem 5 is certainly an upper bound for the number of required measurements, it is not clear how tight this estimate is. This issue unfortunately turns out to be very challenging, and in general, we do not have a quantitative error estimate for our prediction. However, we will give numerical evidence for its tightness in the following and refer the interested reader to [10] for a more detailed discussion.

Numerical Examples: To illustrate the predictive power of Theorem 5, we first revisit the example of Figure 1. For each of the three analysis operators, the prediction of the sampling rate is indicated by colored dots on the x-axis. Figure 1(c) confirms that in the case of an orthonormal basis, we reproduce the known results of [11], showing that our approach is consistent with the classical theory of (BP_η). On the other hand, Figure 1(b) reveals that our framework also allows us to predict the phase transition for highly redundant Haar wavelet frames. While the standard theory based on (co-)sparsity is not able to explain the observed performance gap, the bound of Theorem 5 is fairly accurate for each of the two redundant frames.

In the previous example, the signal remained fixed and we evaluated the sampling rate for different frame instances. In a second experiment, we fix the "better" frame (namely the one of the inverse wavelet transform) and consider signals of different complexity. Since the redundant inverse Haar wavelet transform is well suited for piecewise constant signals, the number of their discontinuities serves as measure of complexity, denoted by S_{TV}. Interestingly, the plots of Figure 3 resemble classical phase transitions based on the usual notion of sparsity, e.g., as reported in [11]. This observation is somewhat surprising because the (averaged) coefficient sparsity $S=\|\boldsymbol{\Psi}_{\mathtt{irdwt}}\boldsymbol{x}^*\|_0$, displayed on the top of the plots in Figure 3, appears detached in our setting. Regarding prediction accuracy, we can conclude that $\mathcal{M}(\boldsymbol{\Psi}_{\mathtt{irdwt}},\boldsymbol{x}^*)$ captures the location of the phase transition fairly well. To the best of our knowledge, the only directly comparable result is given by [12, Theorem 3.3], which leads to much worse predictions. We refer to [10, Section 3] for a detailed documentation of both numerical experiments.



(a) $n = 128$ (b) $n = 512$.

Figure 3. Phase transition plots for piecewise constant signals analyzed with $\mathbf{\Psi}_{\texttt{irdwt}}$ (S_{TV} is the number of discontinuities of $\boldsymbol{x}^* \in \mathbb{R}^n$). The blue and orange curves are obtained by computing the empirical mean of the sampling-rate function $\mathcal{M}(\mathbf{\Psi}_{\texttt{irdwt}}, \boldsymbol{x}^*)$ and the one of [12, Thm. 3.3] for each S_{TV}, respectively.

IV. Conclusion and Outlook

A key insight of this work is that the classical notions of (co-)sparsity are not sufficient for explaining when the analysis basis pursuit ($\mathrm{BP}_\eta^{\mathbf{\Psi}}$) is able to recover a given signal of interest. In contrast, our theoretical framework is based on the novel concept of generalized (co-)sparsity that explicitly takes into account the mutual coherence structure of the frame atoms, which in fact is a missing feature of many traditional approaches. These parameters resemble well-known principles and are particularly consistent with the standard case of orthonormal bases. Our main result, Theorem 5, states a non-uniform and non-asymptotic upper bound on the sampling rate. It is explicit with respect to the analysis operator and simple to compute numerically. The predictive power of this recovery guarantee is demonstrated by two different experiments, considering redundant Haar wavelets. But we wish to emphasize that our analysis is generic in the sense that it is not tailored to a specific type of analysis operator and applies to a general setting.

Our previous work [10] contains several extensions of the prototypical situation of noiseless Gaussian measurements: The presented results also hold true in the presence of measurement noise and for sub-Gaussian measurement ensembles. Furthermore, we derived a simplified bound that is more explicit than the sampling rate $\mathcal{M}(\mathbf{\Psi}, \boldsymbol{x}^*)$; see [10, Proposition 2.7]. Of great importance for practical applications is the extension to *stable recovery*: It is usually not realistic to assume that $\mathbf{\Psi}\boldsymbol{x}^*$ contains more than a very few zero entries. Instead, the coefficient vector is often decaying rapidly, such that many coefficients are close to zero but do never completely vanish. In such a situation, perfect recovery cannot be expected, but one can still hope for an accurate reconstruction from a relatively small number of measurements. However, the bound of Theorem 5 is overly pessimistic in such such situations, yielding an estimate close to the ambient dimension n. In [10, Theorem 2.8], we have relaxed the above framework to account for this phenomenon: It allows us to obtain predictions of the sampling rate based on a nearby surrogate vector with "sparse" analysis coefficients, by accepting a small approximation error. Finally, we point out that this strategy is quite different from previous compressibility arguments which mainly focus on thresholding the analysis coefficients rather than approximating the signal vector itself.

Acknowledgment

M.G. is supported by the European Commission Project DEDALE (contract no. 665044) within the H2020 Framework Program as well as the Einstein Center for Mathematics Berlin. G.K acknowledges partial support by the Einstein Foundation Berlin, the DFG Collaborative Research Center TRR 109 "Discretization in Geometry and Dynamics," and the European Commission Project DEDALE (contract no. 665044) within the H2020 Framework Program, DFG Grant KU 1446/18 as well as DFG-SPP 1798 Grants KU 1446/21 and KU 1446/23. M.M. is supported by the DFG Priority Programme DFG-SPP 1798.

References

[1] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 59–73, 2011.

[4] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Appl. Comput. Harmon. Anal.*, vol. 34, no. 1, pp. 30–56, 2013.

[5] M. Kabanava and H. Rauhut, "Analysis ℓ_1-recovery with frames and Gaussian measurements," *Acta Appl. Math.*, vol. 140, no. 1, pp. 173–195, 2015.

[6] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Probl.*, vol. 23, no. 3, pp. 947–968, 2007.

[7] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, 1992.

[8] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, 2008.

[9] I. W. Selesnick and M. A. T. Figueiredo, "Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors," in *Proceedings of SPIE, Wavelets XIII*, V. K. Goyal, M. Papadakis, and D. V. D. Ville, Eds., vol. 7446, 2009.

[10] G. Kutyniok, M. Genzel, and M. März, "ℓ^1-analysis minimization and generalized (co-)sparsity: When does recovery succeed?" 2017, arXiv preprint: 1710.04952.

[11] 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*, vol. 3, no. 3, pp. 224–294, 2014.

[12] M. Kabanva, H. Rauhut, and H. Zhang, "Robust analysis ℓ_1-recovery from Gaussian measurements and total variation minimization," *Eur. J. Appl. Math.*, vol. 26, no. 6, pp. 917–929, 2015.