Living near the edge: A lower-bound on the phase transition of total variation minimization

Sajad Daei, Arash Amini, Farzan Haddadi

***Abstract*—This work is about the total variation (TV) minimization which is used for recovering gradient-sparse signals from compressed measurements. Recent studies indicate that TV minimization exhibits a phase transition behavior from failure to success as the number of measurements increases. In fact, in large dimensions, TV minimization succeeds in recovering the gradient-sparse signal with high probability when the number of measurements exceeds a certain threshold; otherwise, it fails almost certainly. Obtaining a closed-form expression that approximates this threshold is a major challenge in this field and has not been appropriately addressed yet. In this work, we derive a tight lower-bound on this threshold in case of any random measurement matrix whose null space is distributed uniformly with respect to the Haar measure. In contrast to the conventional TV phase transition results that depend on the simple gradient-sparsity level, our bound is highly affected by generalized notions of gradient-sparsity. Our proposed bound is very close to the true phase transition of TV minimization confirmed by simulation results.**

***Index Terms*—Sample Complexity, Statistical Dimension, Total Variation Minimization.**

I. Introduction

COMPRESSED sensing (CS) has gained a lot of attention in the past decade as it provides a strategy to recover signals from undersampled measurements [1], [2]. In mathematical terms, the measured data about a signal $\boldsymbol{x} \in \mathbb{R}^n$ is given by a collection of linear projections

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^m, \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ with $m \ll n$ is the measurement matrix. Without having any prior knowledge about the structure of $\boldsymbol{x}$, it is impossible to reconstruct $\boldsymbol{x}$ from $\boldsymbol{y}$. The standard prior knowledge in CS is that the signal of interest is sparse in an orthonormal basis which means that it can be expressed as the linear combination of a few basis elements. However, this simple assumption often seems irrational in practical scenarios as we deal with signals that are sparse in some overcomplete dictionary $\boldsymbol{D} \in \mathbb{R}^{n\times N}$ with $n \ll N$ which means that $\boldsymbol{x}$ can be described as $\boldsymbol{x} = \boldsymbol{D}\boldsymbol{\alpha}$ for some sparse $\boldsymbol{\alpha} \in \mathbb{R}^N$. This setting is known as the synthesis sparsity model since it describes a way to synthesize the signal $\boldsymbol{x}$. Then, for recovering $\boldsymbol{\alpha}$ from $\boldsymbol{y}$, it is common to use the so-called ℓ_1 synthesis problem defined as

$$\begin{aligned} \widehat{\alpha} = \underset{\boldsymbol{z}\in\mathbb{R}^N}{\arg\min}\ & \|\boldsymbol{z}\|_1 \\ \text{s.t.}\ \ & \boldsymbol{y} = \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}. \end{aligned} \tag{2}$$

S. Daei and F. Haddadi are with the School of Electrical Engineering, Iran University of Science & Technology. A. Amini is with EE department, Sharif University of Technology.

The reconstructed signal is then $\widehat{\boldsymbol{x}} = \boldsymbol{D}\widehat{\boldsymbol{\alpha}}$. Interestingly, the synthesis model has an analysis (and more general) counterpart which assumes that the signal of interest i.e. $\boldsymbol{x}$ is sparse after applying an operator called analysis operator $\boldsymbol{\Omega}$. The fact that $\boldsymbol{\Omega}\boldsymbol{x}$ is sparse, motivates the optimization problem

$$\begin{aligned} \widehat{\boldsymbol{x}} = \underset{\boldsymbol{z}\in\mathbb{R}^n}{\arg\min}\ & \|\boldsymbol{\Omega}\boldsymbol{z}\|_1 \\ \text{s.t.}\ \ & \boldsymbol{y} = \boldsymbol{A}\boldsymbol{z}, \end{aligned} \tag{3}$$

which is typically referred to as ℓ_1 analysis problem. Although the two methods ℓ_1 synthesis and ℓ_1 analysis perform very differently on large families of signals, the numerical results of the works [3], [4] show that the analysis formulation outperforms its synthesis-based counterpart in many scenarios of interest.

An important and special case of ℓ_1 analysis formulation (3) is the case where $\boldsymbol{\Omega}$ is the finite difference matrix defined as

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & \cdots & \cdots & 1 & -1 \end{bmatrix} \in \mathbb{R}^{n-1\times n}.$$

By replacing this $\boldsymbol{\Omega}$ in the problem (3), we reach the well-known total variation (TV) minimization:

$$\begin{aligned} \mathsf{P}_{\mathrm{TV}}:\quad \min_{\boldsymbol{z}\in\mathbb{R}^n} \|\boldsymbol{z}\|_{\mathrm{TV}} &:= \|\boldsymbol{\Omega}\boldsymbol{z}\|_1 = \sum_{i=1}^{n-1} |z_{i+1} - z_i| \\ \text{s.t.}\ \ & \boldsymbol{y}_{m\times 1} = \boldsymbol{A}\boldsymbol{z}. \end{aligned} \tag{4}$$

TV minimization has been proven to be very effective in image processing [5]–[10] and other fields [11], [12]. One of the important issues concerning the problem P_{TV} is the required number of measurements i.e. the minimum m that P_{TV} needs for successful recovery. A series of works studying convex geometry approaches have found valuable results regarding the required number of measurements in recovering structured[1] signals [13]–[15]. Specifically, it has been shown in [13] that P_{TV} undergoes a phase transition from failure to success as the number of measurements increases. This means that there exists a certain curve $\Psi(\boldsymbol{\Omega}, \boldsymbol{x})$ (known as statistical dimension) in the boundary of failure and success that P_{TV} succeeds to recover the gradient-sparse vector with probability $\frac{1}{2}$. Obtaining an expression that approximates $\Psi(\boldsymbol{\Omega}, \boldsymbol{x})$ is a fundamental challenge in image processing and has not been

[1]For example, sparse signals.

exactly addressed yet. In this work, we obtain a tight lower-bound on $\Psi(\boldsymbol{\Omega}, \boldsymbol{x})$. As numerical results confirm, our proposed lower-bound follows the true phase transition curve very well.

A. Prior Works

In the last few years, great works have been established for obtaining the required number of measurements in P_{TV} (see e.g. [9], [10], [16]–[23]).

Needell et al. in [17], [24] obtain recovery guarantees for two-dimensional TV minimization. Their result is based on transforming gradient-sparse images into compressible signals in a Haar wavelet basis. Then, a modified restricted isometry property (RIP) is developed to guarantee the image recovery. Their approach does not hold for one-dimensional gradient-sparse signals. Besides, their sample complexity result is only designed for the asymptotic setting.

Nam et al. in [20] consider the original problem of minimizing the number of variations (NV) under some affine constraints defined as

$$\mathsf{P}_{\mathrm{NV}}: \quad \min_{\boldsymbol{z}\in\mathbb{R}^n} \|\boldsymbol{z}\|_{\mathrm{NV}} := \|\boldsymbol{\Omega z}\|_0 := \sum_{i=1}^{n-1} 1_{|z_{i+1}-z_i|>0} \quad \text{s.t. } \boldsymbol{y}_{m\times 1} = \boldsymbol{A z}, \tag{5}$$

where $1_{\mathcal{E}}$ denotes the indicator function of a set $\mathcal{E}$. Unfortunately, P_{NV} is known to be NP-complete [20, Section 4.1] and P_{TV} is the closest tractable problem to P_{NV}. In fact, the TV norm sums the amplitudes of variations and in some sense, is to the NV function what the ℓ_1 norm is to the ℓ_0 function in the area of sparse recovery. For a s-gradient-sparse vector $\boldsymbol{x} \in \mathbb{R}^n$, it has been shown in [20] that

$$m > 2\kappa_{\boldsymbol{\Omega}}(s) \tag{6}$$

measurements suffice for exact recovery via P_{NV}. Here, $\kappa_{\boldsymbol{\Omega}}(s)$ is a special function describing the signal manifold dimension and is obtained via a combinatorial search over all s-gradient-sparse signals. This result matches with the well-known fact in conventional CS (in particular sparse recovery and matrix completion) where the number of needed measurements is proportional to the signal's manifold dimension. However, the authors of [25] have numerically shown that when we are dealing with the recovery of gradient-sparse signals via P_{TV}, the required number of measurements is not explained by the manifold dimension anymore. Somewhat interestingly, our proposed bound in Section III is consistent with this fact.

Donoho et al. in [21] obtain the asymptotic minimax mean square error (MSE) of a TV denoiser and using numerical simulations, conjecture that it matches with the phase transition curve of P_{TV}. However, this conjecture is not proved. Moreover, the result holds only in the asymptotic case.

Cai et al. in [9] show that $\mathcal{O}(\sqrt{sn}\log(n))$ and $\mathcal{O}(\sqrt{sn})$ measurements are respectively sufficient and necessary for exact recovery of a given s-gradient-sparse vector via P_{TV}. The order of their bounds seems to be optimal. However, their (necessary and sufficient) bounds do not provide a good prediction for sample complexity in the non-asymptotic case.

In [26, Theorem 4], an upper-bound is derived for the statistical dimension (in case of TV minimization) using the results of [27]. Their bound depends on the gradient-sparsity level and provides an inaccurate prediction for the required sample complexity in low-sparsity regimes.

Recently, the authors of [22] obtain a non-asymptotic sample complexity bound for P_{TV}. In contrast to the previous bounds that would depend on the gradient-sparsity s, their non-asymptotic bound is highly affected by the number of consecutive variations (adjacent pairs in the gradient support). Their proof approach is based on a refined analysis of [27]. While their bound outperforms the previous bounds, it is still far from the statistical dimension in low-sparsity regimes.

By using a more general analysis of [27], the authors in [19] obtain an explicit formula describing the required number of measurements. Their bound depends on the coherence structure of $\boldsymbol{\Omega}$. Again, the proposed bound does not provide an accurate prediction of the true sample complexity in low-sparsity regimes.

B. Contribution

In this work, we obtain a lower-bound on the statistical dimension that provides the necessary number of measurements that P_{TV} needs for exact recovery. Our bound is very close to the true sample complexity even in low-sparsity levels and numerical experiments in SectionIV show that it is tight in the asymptotic case. Our bound depends on the number of the consecutive, individual and tail-end variations of the interested signal. We hope that our bound sheds more light on the effective parameters in the statistical dimension in case of TV minimization. It is worth mentioning that the only lower-bound on the statistical dimension in the literature is established in [9, Theorem 2.1 Part b], and therefore served as an object of comparison in Sections III and IV.

C. Outline of the paper

The paper is organized as follows. First, we review a few basic concepts in convex geometry. Section III amounts to our main result. Section IV is about numerical experiments that verify our theoretical findings. Finally, the paper is concluded in Section V.

D. Notation

Throughout the paper, scalars are denoted by lowercase letters, vectors by lowercase boldface letters, and matrices by uppercase boldface letters. The ith element of a vector $\boldsymbol{x}$ is given either by $x(i)$ or x_i. $(\cdot)^\dagger$ denotes pseudo inverse. We reserve calligraphic uppercase letters for sets (e.g. $\mathcal{S}$). The cardinality of a set $\mathcal{S}$ is denoted by $|\mathcal{S}|$. The complement of a set $\mathcal{S}$ in $\{1, ..., n\}$ is denoted by $\overline{\mathcal{S}}$. $\mathcal{C}^\circ$ denotes the polar of a cone $\mathcal{C}$. The symbol $\mathrm{cone}(\cdot)$ signifies the conic hull of a set. Null space of a matrix is shown by $\mathrm{null}(\cdot)$. For a matrix $\boldsymbol{A}$, $\boldsymbol{A}_{\mathcal{S}}$ means the matrix $\boldsymbol{A}$ restricted to the rows indexed by $\mathcal{S}$. We denote i.i.d. standard Gaussian random vector by $\boldsymbol{g}$. Lastly, $\vee$ means logical "or ".

II. Convex Geometry

A. Statistical Dimension

Definition 1. Statistical Dimension [13]: Statistical dimension is a generalized concept of subspace dimension in the class of convex cones. Intuitively, it measures the size of a cone. More precisely, for a closed convex cone $\mathcal{C} \subseteq \mathbb{R}^n$, statistical dimension of $\mathcal{C}$ is defined as:

$$\delta(\mathcal{C}) := \mathbb{E} \inf_{\boldsymbol{z} \in \mathcal{C}^\circ} \|\boldsymbol{g} - \boldsymbol{z}\|_2^2, \tag{7}$$

where $\boldsymbol{g}$ is a random vector in $\mathbb{R}^n$ chosen from Gaussian ensemble with independent entries.

B. Linear Inverse Problems and Sample Complexity

In [13], it is proved that any random convex optimization problem P_f:

$$\mathsf{P}_f: \quad \min_{\boldsymbol{x} \in \mathbb{R}^n} \ f(\boldsymbol{x}) \quad \text{s.t.} \ \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \tag{8}$$

undergoes a phase transition between success and failure as the number of measurements increases. The location of this transition (the boundary of success and failure) is specified by the statistical dimension of the descent cone of f at $\boldsymbol{x} \in \mathbb{R}^n$ i.e. $\delta(\mathcal{D}(f, \boldsymbol{x}))$ [13]. Here, $\mathcal{D}(f, \boldsymbol{x})$ is defined as the set of directions toward which f is decreased and is given by

$$\mathcal{D}(f, \boldsymbol{x}) = \bigcup_{t \geqslant 0} \{\boldsymbol{z} \in \mathbb{R}^n : f(\boldsymbol{x} + t\boldsymbol{z}) \leqslant f(\boldsymbol{x})\}. \tag{9}$$

There is also a well-known fact between the descent cone and the subdifferential (c.f. [28, Chapter 23]) given by

$$\mathcal{D}(f, \boldsymbol{x})^\circ = \operatorname{cone}(\partial f(\boldsymbol{x})). \tag{10}$$

By using this fact and (7), one may write:

$$\delta(\mathcal{D}(f, \boldsymbol{x})) = \mathbb{E} \inf_{\boldsymbol{z} \in \operatorname{cone}(\partial f(\boldsymbol{x}))} \|\boldsymbol{g} - \boldsymbol{z}\|_2^2 = \mathbb{E} \inf_{t \geqslant 0} \inf_{\boldsymbol{z} \in \partial f(\boldsymbol{x})} \|\boldsymbol{g} - t\boldsymbol{z}\|_2^2. \tag{11}$$

Calculating statistical dimension has been a difficult task in the literature and therefore, it is commonly approximated with the expression

$$B_u := \inf_{t \geqslant 0} \mathbb{E} \inf_{\boldsymbol{z} \in \partial f(\boldsymbol{x})} \|\boldsymbol{g} - t\boldsymbol{z}\|_2^2, \tag{12}$$

which is first proposed by Stojnic [29] in the context of ℓ_1 minimization. In [30, Theorem 2], it has been shown that B_u approximates the statistical dimension well for a large class of structure inducing functions in particular $f = \|\cdot\|_{\mathrm{TV}}$.

III. Main result

Before stating our main result, we need to define some parameters regarding gradient-sparse signals which are required in our analysis.

Definition 2. (Consecutive, individual and tail-end variations) Consider a gradient-sparse signal $\boldsymbol{x} \in \mathbb{R}^n$ with gradient $\boldsymbol{d} := \boldsymbol{\Omega}\boldsymbol{x}$ and gradient support $\mathcal{S}$. The consecutive variations correspond to the adjacent pairs in $\mathcal{S}$ defined as

$$\mathcal{S}_1 := \{i \in [n-1] : i \in \mathcal{S}, \ i-1 \in \mathcal{S}\}. \tag{13}$$

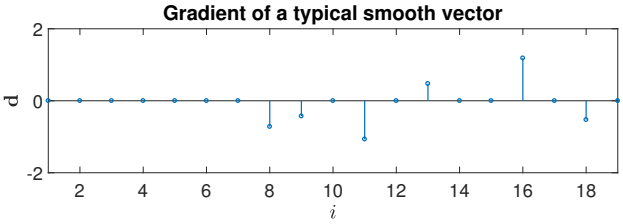

Figure 1: This plot shows the discrete gradient of a typical gradient-sparse vector $\boldsymbol{x} \in \mathbb{R}^{20}$ with parameters $s_1^+ = 1$, $s_1^- = 0$, $s_2 = 10$, and $s_3 = 0$.

The set $\mathcal{S}_1$ can be divided into two sets:

$$\begin{aligned} &\mathcal{S}_1^+ := \\ &\{i \in [n-1] : i \in \mathcal{S}, \ i-1 \in \mathcal{S}, \operatorname{sgn}(d_i)\operatorname{sgn}(d_{i-1}) > 0\}, \\ &\mathcal{S}_1^- := \\ &\{i \in [n-1] : i \in \mathcal{S}, \ i-1 \in \mathcal{S}, \operatorname{sgn}(d_i)\operatorname{sgn}(d_{i-1}) < 0\}, \end{aligned} \tag{14}$$

which are interpreted as the consecutive variations (adjacent pairs in $\mathcal{S}$) with the same and opposite signs, respectively. We define the individual variations by the sets

$$\begin{aligned} \mathcal{S}_2 &:= \{i \in [n-1] : i \in \mathcal{S}, \ i-1 \in \overline{\mathcal{S}}\}, \\ \mathcal{S}_2' &:= \{i \in [n-1] : i \in \overline{\mathcal{S}}, \ i-1 \in \mathcal{S}\}. \end{aligned} \tag{15}$$

The variations in the left and right ends of the signal are called tail-end variations which are defined by the set

$$\begin{aligned} &\mathcal{S}_3 := \\ &\{i \in [n-1] : i \in \mathcal{S}, \ i-1 \notin [n-1] \vee i \in \mathcal{S}, \ i+1 \notin [n-1]\}. \end{aligned} \tag{16}$$

We also show the number of consecutive, individual and tail-end variations respectively by

$$\begin{aligned} s_1 &:= |\mathcal{S}_1| = |\mathcal{S}_1^+| + |\mathcal{S}_1^-| := s_1^+ + s_1^- \\ s_2 &:= |\mathcal{S}_2| + |\mathcal{S}_2'|, \\ s_3 &:= |\mathcal{S}_3|. \end{aligned}$$

As an illustrative example, in Figure 1, the discrete gradient $\boldsymbol{d} := \boldsymbol{\Omega}\boldsymbol{x}$ of a typical gradient-sparse signal $\boldsymbol{x} \in \mathbb{R}^{20}$ is depicted. There is one pair of elements in $\boldsymbol{d}$ both of which belong to the gradient support (alternatively representing consecutive variations of $\boldsymbol{x}$) and have the same sign. Thus, $s_1^+ = 1$ and $s_2^- = 0$. Also, there are 10 elements for which $i \in \mathcal{S}, i-1 \notin \overline{\mathcal{S}}$ or $i \in \overline{\mathcal{S}}, i-1 \notin \mathcal{S}$ (alternatively representing individual variations in $\boldsymbol{x}$). As a result, $s_2 = 10$. Lastly, there are no elements in the tail-ends of $\boldsymbol{d}$. This means that the first and last elements of $\boldsymbol{x}$ include no variations and $s_3 = 0$.

In the following theorem, we propose a lower-bound on $\delta(\mathcal{D}(\|\cdot\|_{\mathrm{TV}}, \boldsymbol{x}))$. This lower-bound depends on the associated parameters in Definition 2

Theorem 1. *Let $\boldsymbol{x} \in \mathbb{R}^n$ be a gradient-sparse signal with gradient $\boldsymbol{d} := \boldsymbol{\Omega}\boldsymbol{x}$, and gradient support $\mathcal{S}$. Assume that $\boldsymbol{x}$ has $s_1 = s_1^+ + s_1^-$, s_2, and s_3 consecutive, individual and tail-end variations, respectively as defined in Definition 2. Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ be a random matrix whose null space*

is uniformly distributed with respect to the Haar measure. Consider $\boldsymbol{y}_{m\times 1} = \boldsymbol{A}\boldsymbol{x}$ *as the vector of measurements. Then,*

$$\delta(\mathcal{D}(\|\cdot\|_{\mathrm{TV}}, \boldsymbol{x})) \geqslant \inf_{t\geqslant 0} \Psi_t(s_1^+, s_1^-, s_2, s_3) := \widehat{m}_{\mathrm{TV}}, \quad (17)$$

where

$$\begin{aligned}\Psi_t(s_1^+, s_1^-, s_2, s_3) &= s_1^+ + s_1^-(1+4t^2)\\ &+ s_2\phi_1(t,t) + (n-2-s_1^+-s_1^--s_2)\phi_2(2t)+\\ &s_3(1+t^2) + (2-s_3)\phi_2(t),\end{aligned} \quad (18)$$

and

$$\begin{aligned}\phi_1(a,b) &:= \frac{1}{\sqrt{2\pi}}\int_b^\infty (u-b)^2[e^{-\frac{(u-a)^2}{2}} + e^{-\frac{(u+a)^2}{2}}]du,\\ \phi_2(z) &:= \sqrt{\frac{2}{\pi}}\int_z^\infty (u-z)^2 e^{-\frac{u^2}{2}}du,\end{aligned} \quad (19)$$

and thus if $m \leqslant \widehat{m}_{\mathrm{TV}}$*, then with probability at least* $1 - 4\mathrm{e}^{-\frac{(\widehat{m}_{\mathrm{TV}}-m)^2}{16\widehat{m}_{\mathrm{TV}}}}$*,* P_{TV} *fails to recover* $\boldsymbol{x}$*.*

Proof sketch . As discussed in (12), we intend to find a lower-bound for

$$\inf_{t\geqslant 0} \mathbb{E} \inf_{\boldsymbol{z}\in\partial\|\cdot\|_{\mathrm{TV}}(\boldsymbol{x})} \|\boldsymbol{g} - t\boldsymbol{z}\|_2^2. \quad (20)$$

The expression inside the latter infimum is formed of a few summands. Each summand separately has a unique minimizer over the set $\partial\|\cdot\|_{\mathrm{TV}}(\boldsymbol{x})$. By passing this infimum through each summand (this leads to a lower-bound) and then applying expectation, we reach a closed-form expression for each summand and thereby a strictly convex function with respect to t for the expression inside the former infimum.

Discussion. One of the key properties of Theorem 1 is that, unlike the most works in TV minimization, the widely-used concept of gradient sparsity $s := |\mathcal{S}|$ does not directly explain our proposed bound. In fact, it seems that the statistical dimension in case of TV minimization i.e. $\delta(\mathcal{D}(\|\cdot\|_{\mathrm{TV}}, \boldsymbol{x}))$, depends on generalized concepts of gradient-sparsity: namely the number of consecutive, individual, and tail-end variations. To examine how these parameters affect our bound $\widehat{m}_{\mathrm{TV}}$, we designed a numerical experiment in Table I and considered different values for s_1^+, s_1^-, s_2 and s_3. We observe from Table I that the number of consecutive variations with negative signs, i.e. s_1^-, has the highest impact on $\widehat{m}_{\mathrm{TV}}$. In other words, for a fixed number of variations ($s = \|\boldsymbol{\Omega x}\|_0$), recovering a highly oscillating signal is harder (alternatively needs more measurements) than the one with well-separated variations.

Remark 1. (Prior work) In [9, Theorem 2.1], it has been proved that

$$\begin{aligned}&\frac{9\sqrt{sn}}{50\pi} - \frac{12}{5\pi} \leqslant \delta(\mathcal{D}(\|\cdot\|_{\mathrm{TV}}, \boldsymbol{x})) \leqslant\\ &\sqrt{32}(2\sqrt{5}+\sqrt{10})^2\sqrt{ns}\log(2n) + 1.\end{aligned} \quad (21)$$

Their proof approach is based on estimating the statistical dimension of a certain set using a wavelet-based argument. Since the lower and upper-bound are in the same order (up to a log factor), their approach is optimal in the asymptotic case ($n \to \infty$). However, in the non-asymptotic case, it demonstrates poor prediction of the true sample complexity.

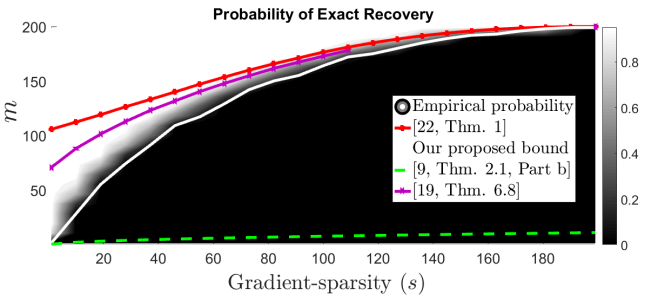

Figure 2: This figure shows the phase transition of P_{TV} in case of $n = 200$. The white curve is our proposed bound. The bounds in [22, Theorem 1] and [9, Theorem 2.1 part b] are depicted by dashed and dotted line, respectively. The white, purple, and red curves are obtained by computing the empirical mean of the sample complexity for each s. The brightness of figure in each pair (s, m), shows the empirical probability of success (black=0%, white=100%).

n	s	s_1^+	s_1^-	s_2	s_3	$\widehat{m}_{\mathrm{TV}}$
100	10	9	0	2	0	10
100	10	0	9	2	0	32.04
100	10	0	8	2	2	31.74
100	10	0	9	1	1	32.584
100	10	9	0	1	1	12.33
100	10	0	0	20	0	10
100	10	0	1	17	1	16.53
100	10	4	5	2	0	25.54

Table I: In this table, we examine the impact of s_1^+, s_1^-, s_2 and s_3 on our bound $\widehat{m}_{\mathrm{TV}}$ in Theorem 1. We observe that s_1^- has the highest impact.

IV. Simulations

In this section, we evaluate how our proposed bound in Theorem 1 predicts the phase transition of P_{TV}. For each m and s, we repeat the following steps 50 times:

- Select a random subset $\mathcal{S} \subseteq \{1, ..., n-1\}$ with $|\mathcal{S}| = s$.
- Generate a vector $\boldsymbol{x} \in \mathrm{null}(\boldsymbol{\Omega}_{\overline{\mathcal{S}}})$ whose gradient is supported on $\mathcal{S}$.
- Construct the vector $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ where $\boldsymbol{A}$ is an i.i.d. Gaussian matrix (the null space of Gaussian matrices is distributed uniformly with respect to the Haar measure).
- Solve P_{TV} to obtain an estimate $\widehat{\boldsymbol{x}}$.
- Declare success if $\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2 \leqslant 10^{-6}$.

Figures 2, and 3 show the empirical probability of successful recovery obtained from 50 Monte Carlo iterations in case of $n = 200$ and $n = 400$, respectively. Notice that since the sample complexity bounds in [22, Theorem 1], [19, Theorem 1] and $\widehat{m}_{\mathrm{TV}}$ do not directly depend on the gradient-sparsity s, we depict the empirical mean of the bounds over 300 iterations, for each s. As it turns out from Figures 2, and 3 our proposed bound in Theorem 1 predicts the statistical dimension well; the lower-bound [9, Theorem 2.1] does not seem to be exact; and the upper-bounds [22, Theorem 1] and [19, Theorem 6.8] fail to explain the true sample complexity in low-sparsity levels.

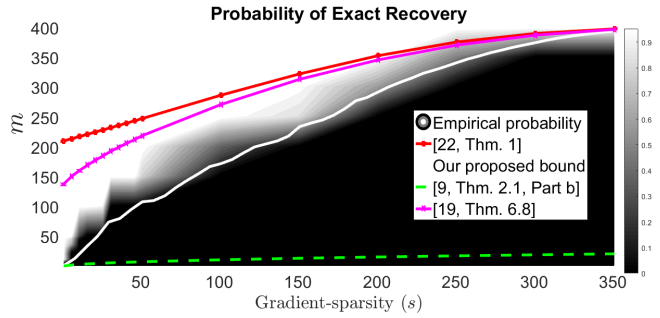

Figure 3: This figure shows the phase transition of P_{TV} in case of $n = 400$. The white curve is our proposed bound. The bounds in [22, Theorem 1] and [9, Theorem 2.1 part b] are depicted by dashed and dotted line, respectively. The white, purple, and red curves are obtained by computing the empirical mean of the sample complexity for each s. The brightness of figure in each pair (s, m), shows the empirical probability of success (black=0%, white=100%).

V. Conclusion

In this work, we obtained a lower-bound on the statistical dimension in case of TV minimization. This lower-bound provides the necessary number of measurements that P_{TV} needs for successful recovery. Our bound depends on the number of consecutive, individual and tail-end variations of the signal and precisely captures the location of the TV phase transition. In fact, it seems that these quantities specify the effective parameters of the statistical dimension for gradient-sparse signals.

Appendix A
Proof of Theorem 1

Proof. Recall that the sets $\mathcal{S}_1^+$, $\mathcal{S}_1^-$, $\mathcal{S}_2$, $\mathcal{S}_2'$, $\mathcal{S}_3$ are defined in Definition 2. Moreover, define the set of adjacent pairs in $\overline{\mathcal{S}}$ as

$$\mathcal{S}_4 := \{i \in [n-1] : \ i \in \overline{\mathcal{S}}, \ i-1 \in \overline{\mathcal{S}}\}, \tag{22}$$

which are used in the proof.

Since statistical dimension is approximately equal to B_u in (12), we find a lower-bound for

$$B_u := \inf_{t \geqslant 0} \mathbb{E}\mathrm{dist}^2(\boldsymbol{g}, t\partial\|\cdot\|_{\mathrm{TV}}(\boldsymbol{x})). \tag{23}$$

The first step is to calculate $\partial\|\cdot\|_{\mathrm{TV}}(\boldsymbol{x})$. From the chain rule lemma of subdifferential [28, Chapter 23], we have:

$$\partial\|\cdot\|_{\mathrm{TV}}(\boldsymbol{x}) = \boldsymbol{\Omega}^T \partial\|\cdot\|_1(\boldsymbol{d}) = \boldsymbol{\Omega}^T \left\{ \boldsymbol{z} \in \mathbb{R}^{n-1} : \begin{array}{ll} z_i = \mathrm{sgn}(d_i), & i \in \mathcal{S} \\ |z_i| \leqslant 1, & \text{o.w.} \end{array} \right\}. \tag{24}$$

To calculate (23), regarding (24), we compute the distance of the dilated subdifferential of the descent cone of TV norm at $\boldsymbol{x} \in \mathbb{R}^n$ from a standard Gaussian vector $\boldsymbol{g} \in \mathbb{R}^n$ which is given by:

$$\begin{aligned} &\mathrm{dist}^2(\boldsymbol{g}, t\partial\|\cdot\|_{\mathrm{TV}}(\boldsymbol{x})) = \inf_{\boldsymbol{z} \in \partial\|\cdot\|_{\mathrm{TV}}(\boldsymbol{x})} \|\boldsymbol{g} - t\boldsymbol{z}\|_2^2 = \\ &\inf_{\boldsymbol{z} \in \partial\|\cdot\|_{\mathrm{TV}}(\boldsymbol{x})} \sum_{i=1}^n (g_i - t(\Omega^T z)_i)^2 = \\ &\inf_{\|\boldsymbol{z}\|_\infty \leqslant 1} \sum_{i=1}^n \big(g_i - t \sum_{j \in \mathcal{S}} \Omega(j,i)\mathrm{sgn}(d_j) \\ &- t \sum_{j \in \overline{\mathcal{S}}} \Omega(j,i) z(j)\big)^2. \end{aligned} \tag{25}$$

Each row of the finite difference operator $\boldsymbol{\Omega}$ includes a pair of $\{+1, -1\}$ and is zero elsewhere, i.e.,

$$\Omega(j,i) = \left\{ \begin{array}{ll} 1, & j = i \\ -1, & j = i-1 \end{array} \right\} \tag{26}$$

By using the latter property, the relation (25) reads

$$\begin{aligned} &\inf_{\|\boldsymbol{z}\|_\infty \leqslant 1} \sum_{i=1}^n \Big(g_i - t\mathrm{sgn}(d_i) 1_{i \in \mathcal{S}} + t\mathrm{sgn}(d_{i-1}) 1_{i-1 \in \mathcal{S}} \\ &- tz(i) 1_{i \in \overline{\mathcal{S}}} + tz(i-1) 1_{i-1 \in \overline{\mathcal{S}}} \Big)^2. \end{aligned} \tag{27}$$

By passing the infimum through the summation, we have a lower-bound on (27) as follows:

$$\begin{aligned} &\mathrm{dist}^2(\boldsymbol{g}, t\partial\|\cdot\|_{\mathrm{TV}}(\boldsymbol{x})) \geqslant \\ &\sum_{i \in \mathcal{S}_1^+ \cup \mathcal{S}_1^-} (g_i - t\mathrm{sgn}(d_i) + t\mathrm{sgn}(d_{i-1}))^2 \\ &+ \sum_{i \in \mathcal{S}_2} \inf_{\|\boldsymbol{z}\|_\infty \leqslant 1} (g_i - t\mathrm{sgn}(d_i) + tz(i-1))^2 + \\ &\sum_{i \in \mathcal{S}_2'} \inf_{\|\boldsymbol{z}\|_\infty \leqslant 1} (g_i + t\mathrm{sgn}(d_{i-1}) - tz(i))^2 + \sum_{i \in \mathcal{S}_4} \inf_{\|\boldsymbol{z}\|_\infty \leqslant 1} (g_i - tz(i) \\ &+ tz(i-1))^2 + (g_1 - t\mathrm{sgn}(d_1))^2 1_{1 \in \mathcal{S}} + \\ &\inf_{\|\boldsymbol{z}\|_\infty \leqslant 1} (g_1 - tz(1))^2 1_{1 \in \overline{\mathcal{S}}} + (g_n - t\mathrm{sgn}(d_{n-1}))^2 1_{n-1 \in \mathcal{S}} + \\ &\inf_{\|\boldsymbol{z}\|_\infty \leqslant 1} (g_n - tz(n-1))^2 1_{n-1 \in \overline{\mathcal{S}}}. \end{aligned} \tag{28}$$

Now, we investigate the minimizations in (28), one by one. First, it holds that

$$\begin{aligned} &\inf_{\|\boldsymbol{z}\|_\infty \leqslant 1} (g_i - t\mathrm{sgn}(d_i) + tz(i-1))^2 = \\ &\inf_{|z(i-1)| \leqslant 1} (g_i - t\mathrm{sgn}(d_i) + tz(i-1))^2 \\ &\overset{\text{(I)}}{=} (|g_i - t\mathrm{sgn}(d_i)| - t)_+^2, \end{aligned} \tag{29}$$

where the equality (I) is since the optimal objective function of the problem

$$\inf_{|z(i-1)| \leqslant 1} (g_i - t\mathrm{sgn}(d_i) + tz(i-1))^2,$$

occurs either by the boundaries imposed by the feasible set $|z(i-1)| \leqslant 1$ or equals zero. With a similar reason, we have

$$\inf_{\|\boldsymbol{z}\|_\infty \leqslant 1} (g_i + t\mathrm{sgn}(d_{i-1}) - tz(i))^2 = (|g_i + t\mathrm{sgn}(d_{i-1})| - t)_+^2, \tag{30}$$

and

$$\inf_{\|\boldsymbol{z}\|_\infty\leqslant 1}(g_i - tz(i) + tz(i-1))^2 = \inf_{|z(i)|\leqslant 1, |z(i-1)|\leqslant 1}(g_i - tz(i) + tz(i-1))^2 = (|g_i| - 2t)_+^2. \tag{31}$$

Introduce the above expressions into (25) to reach

$$\begin{aligned}
&\mathrm{dist}^2(\boldsymbol{g}, t\partial\|\cdot\|_{\mathrm{TV}}(\boldsymbol{x})) \geqslant \\
&= \sum_{i\in\mathcal{S}_1^+\cup\mathcal{S}_1^-}(g_i - t\mathrm{sgn}(d_i) + t\mathrm{sgn}(d_{i-1}))^2 \\
&+ \sum_{i\in\mathcal{S}_2}(\zeta_1 - t)_+^2 + \sum_{i\in\mathcal{S}_2'}(\zeta_2 - t)_+^2 \\
&+ \sum_{i\in\mathcal{S}_4}(|g_i| - t - t)_+^2 + (g_1 - t\mathrm{sgn}(d_1))^2 1_{1\in\mathcal{S}} + \\
&(|g_1| - t)_+^2 1_{1\in\overline{\mathcal{S}}} + (g_n - t\mathrm{sgn}(d_{n-1}))^2 1_{n-1\in\mathcal{S}} + \\
&(|g_n| - t)^2 1_{n-1\in\overline{\mathcal{S}}},
\end{aligned} \tag{32}$$

where $\zeta_1 = |g_i - t\mathrm{sgn}(d_i)|$, $\zeta_2 = |g_i + t\mathrm{sgn}(d_{i-1})|$. By taking expectation from both sides, we reach

$$\begin{aligned}
&\mathbb{E}\mathrm{dist}^2(\boldsymbol{g}, t\partial\|\cdot\|_{\mathrm{TV}}(\boldsymbol{x})) \geqslant \\
&|\mathcal{S}_1^+\cup\mathcal{S}_1^-| + \sum_{i\in\mathcal{S}_1^+\cup\mathcal{S}_1^-} t^2(\mathrm{sgn}(d_i) - \mathrm{sgn}(d_{i-1}))^2 \\
&+ \sum_{i\in\mathcal{S}_2}\mathbb{E}(\zeta_1 - t)_+^2 + \sum_{i\in\mathcal{S}_2'}\mathbb{E}(\zeta_2 - t)_+^2 + \\
&\sum_{i\in\mathcal{S}_4}\mathbb{E}(|g_i| - t - t)_+^2 + (1+t^2)1_{1\in\mathcal{S}} \\
&\mathbb{E}(|g_1| - t)_+^2 + (1+t^2)1_{n-1\in\mathcal{S}} + \mathbb{E}(|g_n| - t)_+^2.
\end{aligned} \tag{33}$$

In what follows, we calculate the expressions within (33). First, consider $\mathbb{E}(\zeta_1 - t)_+^2$, which is calculated as follows:

$$\begin{aligned}
&\mathbb{E}(\zeta_1 - t)_+^2 = 2\int_0^\infty a\mathbb{P}(\zeta_1 \geqslant a + t)da \\
&= 2\frac{1}{\sqrt{2\pi}}\int_0^\infty\int_{t+a}^\infty a\ (e^{-\frac{(u-t)^2}{2}} + e^{-\frac{(u+t)^2}{2}})du\ da \\
&= 2\frac{1}{\sqrt{2\pi}}\int_t^\infty\int_0^{u-t} a\ (e^{-\frac{(u-t)^2}{2}} + e^{-\frac{(u+t)^2}{2}})da\ du \\
&= \phi_1(t,t),
\end{aligned} \tag{34}$$

where in (34), the order of integration is changed together with a change of variable. Similarly, we have: $\mathbb{E}(\zeta_2 - t)_+^2 = \phi_1(t,t)$. Also,

$$\begin{aligned}
&\mathbb{E}(|g_i| - t)_+^2 = 2\int_0^\infty a\mathbb{P}(|g_i| \geqslant a + t)da \\
&= 2\sqrt{\frac{2}{\pi}}\int_0^\infty\int_{t+a}^\infty a\ e^{-\frac{u^2}{2}}du\ da \\
&= 2\sqrt{\frac{2}{\pi}}\int_t^\infty\int_0^{u-t} a\ e^{-\frac{u^2}{2}}da\ du := \phi_2(t),
\end{aligned} \tag{35}$$

where in the third line, the order of integration is changed. Thus, we have, $\mathbb{E}(|g_i| - t - t)_+^2 = \phi_2(2t)$. As a consequence, (33) becomes:

$$\begin{aligned}
&\mathbb{E}\mathrm{dist}^2(\boldsymbol{g}, t\partial\|\cdot\|_{\mathrm{TV}}(\boldsymbol{x})) \geqslant \\
&|\mathcal{S}_1^+\cup\mathcal{S}_1^-| + \sum_{i\in\mathcal{S}_1^+\cup\mathcal{S}_1^-} t^2(\mathrm{sgn}(d_i) - \mathrm{sgn}(d_{i-1}))^2 \\
&+ \sum_{i\in\mathcal{S}_2}\phi_1(t,t) + \sum_{i\in\mathcal{S}_2'}\phi_1(t,t) + \\
&\sum_{i\in\mathcal{S}_4}\phi_2(2t) + (1+t^2)1_{1\in\mathcal{S}} + \phi_2(t)1_{1\in\overline{\mathcal{S}}} \\
&+ (1+t^2)1_{n-1\in\mathcal{S}} + \phi_2(t)1_{n-1\in\overline{\mathcal{S}}} := \Psi_t(s_1^+, s_1^-, s_2, s_3).
\end{aligned} \tag{36}$$

Finally by setting $s_3 = 1_{1\in\mathcal{S}} + 1_{n-1\in\mathcal{S}}$, $s_2 = |\mathcal{S}_2| + |\mathcal{S}_2'|$, $2 - s_3 = 1_{1\in\overline{\mathcal{S}}} + 1_{n-1\in\overline{\mathcal{S}}}$, $s_1^+ = |\mathcal{S}_1^+|$, $s_1^- = |\mathcal{S}_1^-|$, and $|\mathcal{S}_4| = n - 2 - s_1^+ - s_1^- - s_2$, we reach

$$\delta(\mathcal{D}(\|\cdot\|_{\mathrm{TV}}, \boldsymbol{x})) \geqslant \inf_{t\geqslant 0}\Psi_t(s_1^+, s_1^-, s_2, s_3) := \widehat{m}_{\mathrm{TV}}. \tag{37}$$

We know from [13, Theorems 7.1, 6.1] that if

$$m \leqslant \delta(\mathcal{D}(\|\cdot\|_{\mathrm{TV}}, \boldsymbol{x})) := \delta,$$

then,

$$\mathbb{P}[\boldsymbol{x} \text{ is the unique solution of } \mathsf{P}_{\mathrm{TV}}] \leqslant 4\mathrm{e}^{-\frac{(\delta-m)^2}{16\delta}}. \tag{38}$$

Since the function $f: z \to 4\mathrm{e}^{-\frac{(z-m)^2}{16z}}$ is decreasing, and the fact that $m \leqslant \widehat{m}_{\mathrm{TV}} \leqslant \delta$, it holds that

$$\mathbb{P}[\boldsymbol{x} \text{ is the unique solution of } \mathsf{P}_{\mathrm{TV}}] \leqslant 4\mathrm{e}^{-\frac{(\delta-m)^2}{16\delta}} \leqslant 4\mathrm{e}^{-\frac{(\widehat{m}_{\mathrm{TV}}-m)^2}{16\widehat{m}_{\mathrm{TV}}}}. \tag{39}$$

■

REFERENCES

[1] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE transactions on information theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[2] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution," *Communications on pure and applied mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[3] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.

[4] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse problems*, vol. 23, no. 3, p. 947, 2007.

[5] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, "Image restoration: total variation, wavelet frames, and beyond," *Journal of the American Mathematical Society*, vol. 25, no. 4, pp. 1033–1089, 2012.

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

[7] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine & Biology*, vol. 53, no. 17, p. 4777, 2008.

[8] S. L. Keeling, "Total variation based convex filters for medical imaging," *Applied Mathematics and Computation*, vol. 139, no. 1, pp. 101–119, 2003.

[9] J.-F. Cai and W. Xu, "Guarantees of total variation minimization for signal recovery," *Information and Inference*, vol. 4, no. 4, pp. 328–353, 2015.

[10] F. Krahmer, C. Kruschel, and M. Sandbichler, "Total variation minimization in compressed sensing," in *Compressed Sensing and its Applications*, pp. 333–358, Springer, 2017.

[11] X. Wu, Q. Wang, and M. Liu, "In-situ soil moisture sensing: Measurement scheduling and estimation using sparse sampling," *ACM Transactions on Sensor Networks (TOSN)*, vol. 11, no. 2, p. 26, 2015.

[12] P. M. van den Berg and R. E. Kleinman, "A total variation enhanced modified gradient algorithm for profile reconstruction," *Inverse Problems*, vol. 11, no. 3, p. L5, 1995.

[13] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Information and Inference: A Journal of the IMA*, vol. 3, no. 3, pp. 224–294, 2014.

[14] J. A. Tropp, "Convex recovery of a structured signal from independent random linear measurements," in *Sampling Theory, a Renaissance*, pp. 67–101, Springer, 2015.

[15] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[16] M. Kabanava and H. Rauhut, "Analysis ℓ_1-recovery with frames and gaussian measurements," *Acta Applicandae Mathematicae*, vol. 140, no. 1, pp. 173–195, 2015.

[17] D. Needell and R. Ward, "Stable image reconstruction using total variation minimization," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 1035–1058, 2013.

[18] C. Poon, "On the role of total variation in compressed sensing," *SIAM Journal on Imaging Sciences*, vol. 8, no. 1, pp. 682–720, 2015.

[19] M. Genzel, G. Kutyniok, and M. März, "ℓ_1-analysis minimization and generalized (co-) sparsity: When does recovery succeed?," *arXiv preprint arXiv:1710.04952*, 2017.

[20] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[21] D. L. Donoho, I. Johnstone, and A. Montanari, "Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising," *IEEE transactions on information theory*, vol. 59, no. 6, pp. 3396–3433, 2013.

[22] S. Daei, F. Haddadi, and A. Amini, "Sample complexity of total variation minimization," *IEEE Signal Processing Letters*, vol. 25, pp. 1151–1155, Aug 2018.

[23] B. Zhang, W. Xu, J.-F. Cai, and L. Lai, "Precise phase transition of total variation minimization," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pp. 4518–4522, IEEE, 2016.

[24] D. Needell and R. Ward, "Near-optimal compressed sensing guarantees for total variation minimization," *IEEE transactions on image processing*, vol. 22, no. 10, pp. 3941–3949, 2013.

[25] R. Giryes, Y. Plan, and R. Vershynin, "On the effective measure of dimension in the analysis cosparse model," *IEEE Transactions on Information Theory*, vol. 61, no. 10, pp. 5745–5753, 2015.

[26] M. Kabanava, H. Rauhut, and H. Zhang, "Robust analysis ℓ_1-recovery from gaussian measurements and total variation minimization," *European Journal of Applied Mathematics*, vol. 26, no. 6, pp. 917–929, 2015.

[27] R. Foygel and L. Mackey, "Corrupted sensing: Novel guarantees for separating structured signals," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1223–1247, 2014.

[28] R. T. Rockafellar, *Convex analysis*. Princeton university press, 2015.

[29] M. Stojnic, "Various thresholds for l_1-optimization in compressed sensing," *arXiv preprint arXiv:0907.3666*, 2009.

[30] S. Daei, F. Haddadi, A. Amini, and M. Lotz, "On the error in phase transition computations for compressed sensing," *arXiv preprint arXiv:1806.10583*, 2018.