On the Error in Phase Transition Computations for Compressed Sensing

Sajad Daei, Farzan Haddadi, Arash Amini, Martin Lotz

Abstract—Evaluating the statistical dimension is a common tool to determine the asymptotic phase transition in compressed sensing problems with Gaussian ensemble. Unfortunately, the exact evaluation of the statistical dimension is very difficult and it has become standard to replace it with an upper-bound. To ensure that this technique is suitable, [1] has introduced an upper-bound on the gap between the statistical dimension and its approximation. In this work, we first show that the error bound in [1] in some low-dimensional models such as total variation and \( \ell_1 \) analysis minimization becomes poorly large. Next, we develop a new error bound which significantly improves the estimation gap compared to [1]. In particular, unlike the bound in [1] that is not applicable to settings with overcomplete dictionaries, our bound exhibits a decaying behavior in such cases.

Index Terms—statistical dimension, error estimate, low-complexity models.

I. INTRODUCTION

UNDERSTANDING the behavior of random compressed sensing problems in transition from absolute failure to success (known as phase transition) has been the subject of research in recent years [1]–[8]. Most of these works concentrate on simple sparse models and do not allude to the challenges in other low-dimensional structures such as low-rank matrices, block-sparse vectors, gradient-sparse vectors and cosparse vectors. For simplicity, we associate such structures with their common recovery techniques and rename the structures accordingly. For instance, total variation (TV), \( \ell_1 \) analysis and \( \ell_1,2 \) minimization refer to both the recovery techniques and the underlying low-dimensional structures. In this work, we revisit linear inverse problems with the aim of recovering a vector \( x \in \mathbb{R}^n \) from a few random linear measurements \( y = Ax \in \mathbb{R}^m \). This is summarized as solving the following convex program:

\[
P_f : \min_{z \in \mathbb{R}^n} f(z) \quad \text{s.t. } y = Az,
\]

where, \( A \in \mathbb{R}^{m \times n} \) is the measurement matrix whose entries are i.i.d. random variables with normal distribution and \( f \) is a convex penalty function that promotes the low-dimensional structure. A major subject of recent research is the number of Gaussian measurements (the number of rows in \( A \)) one needs to recover a structured vector \( x \) from \( y \in \mathbb{R}^m \). In [3], a bound is obtained using polytope angle calculations with asymptotic sharpness in case of \( f = \| \cdot \|_1 \). In [9], it is proved that the minimax MSE for the TV-regularized denoising problem:

\[
\min_{z \in \mathbb{R}^n} \tau \| z \|_{TV} + \frac{1}{2} \| y - z \|_2^2
\]

is an upper-bound for the required number of measurements in \( P_f \) when \( f = \| \cdot \|_{TV} \). Also, in [7], the authors showed that the related minimax MSE is the same as the number of required measurements that TV approximate message passing (TV-AMP) algorithm needs. [2] introduced a general framework for obtaining the number of Gaussian measurements in different low-dimensional structures using Gordon min-max inequality [9] and the concept of atomic norms. Specifically, it was shown that \( \omega^2(D(f, x) \cap \mathbb{B}^n) + 1 \) measurements are sufficient. Here, \( D(f, x) \) is the descent cone of \( f \) at \( x \in \mathbb{R}^n \) and \( \omega^2(D(f, x) \cap \mathbb{B}^n) \) is the squared Gaussian width, which intuitively measures the size of this cone. In [1], it has been shown that the statistical dimension of this cone, which is defined below and differs from the squared Gaussian width above by at most 1, specifies the phase transition of the [random] convex program \( P_f \) from absolute failure to absolute success:

\[
\delta(D(f, x)) := \mathbb{E} \text{dist}^2(g, \text{cone}(\hat{f}(x))).
\]

\( \delta(D(f, x)) \) is the average distance of a standard Gaussian i.i.d. vector \( g \in \mathbb{R}^n \) from non-negative scalings of the sub-differential at point \( x \in \mathbb{R}^n \). So far, we know that a phase transition exists in \( P_f \) and its boundary is interpreted via the statistical dimension. A natural question is how we can find an expression for the phase transition curve. The upper bounds for \( \delta(D(f, x)) \) and \( \omega(D(f, x) \cap \mathbb{B}^n) \), first used in the context of \( \ell_1 \) minimization by Stojnic ( [5]), are given by:

\[
\delta(D(f, x)) \leq \inf_{t \geq 0} \mathbb{E} \text{dist}^2(g, t\hat{f}(x)) := U_\delta,
\]

\[
\omega(D(f, x) \cap \mathbb{B}^n) \leq \inf_{t \geq 0} \mathbb{E} \text{dist}(g, t\hat{f}(x)) := U_.
\]

However, it is still unknown whether \( U_\delta \) and \( U_\omega \) are sharp for different low-dimensional structures. For ease of notation, we define the errors \( E_\omega \) and \( E_\delta \) by:

\[
E_\omega := U_\omega - \omega(D(f, x) \cap \mathbb{B}^n),
\]

\[
E_\omega := U_\delta - \delta(D(f, x)).
\]

Here, \( U_\delta \) represents a sufficient number of measurements that \( P_f \) needs for successful recovery. In [1], explicit formulas are derived for the upper bound \( \delta \) in case of \( \ell_1 \) and nuclear norm. Distance of the upper bound from \( \delta(D(f, x)) \) (known as the error estimate) has also been discussed (see Theorem [1]). In case of \( f = \| \cdot \|_1 \) and in very-low-sparsity regimes, it
is shown that the error estimate is high whereas it is small in other regimes. This puts in doubt whether the upper bound properly describes the statistical dimension. This raises the below questions about the performance of $U_\delta$ (4):

- Does $U_\delta$ provide a fair estimate of the statistical dimension?
- How to quantify the gap between the exact phase transition curve and the one obtained via $U_\delta$?
- Can one extend the previous error bounds obtained for $\ell_1$ minimization in [1] (see [1]) to other low-dimensional structures such as block sparsity, low-rank, TV and $\ell_1$ analysis?

The goal of this work is to find answers to these questions. Specifically, we want to study how well $U_\delta$ describes $\delta(D(f, x))$ in low-dimensional structures represented by $\| \cdot \|_1$, $\| \cdot \|_1, \| \cdot \|_{\alpha}$, $\| \cdot \|_1$ and $\| \cdot \|_{TV} := \| \Omega \cdot \|_1$, where

$$\Omega_d = \begin{bmatrix}
1 & -1 & 0 & \cdots & 0 \\
0 & 1 & -1 & \cdots & 0 \\
& & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & 1 & -1
\end{bmatrix} \in \mathbb{R}^{n-1 \times n} \quad (8)$$

is the difference operator used in TV-norm minimization.

A. Motivation

Tables I and II present the results of a computer experiment designed to evaluate the error of $U_\delta$ in estimating the statistical dimension. In two experiments shown in Tables I and II we test the error bound for $\ell_1$ and TV minimization. To compute $\delta(D(f, x))$, we find the number of measurements that $P_f$ succeeds with probability $\frac{1}{2}$ from 100 trials. In the first experiment, for each sparsity level, we construct a sparse vector $x \in \mathbb{R}^{1000}$ with random non-zero values (distributed as $\mathcal{N}(0, \sqrt{1000})$) at uniformly random locations. In the second experiment, we generate a gradient sparse vector $x \in \mathbb{R}^{1000}$ with large $\ell_2$ norm and small variation. For Tables I and II the upper bound (4) is obtained by [1] Equation D.6 and numerical optimization, respectively. As shown in Table I there exists a large gap between the numerically obtained error and the state of the art theoretical error estimate (23). This gap is more pronounced in TV minimization (Table II). Now, a natural question that arises is: can we find a better bound that reduces the gap?

B. Contributions

In this work, we rigorously analyze the error of estimating the phase transition. The significance of this error is to have a good understanding about the required number of measurements that $P_f$ needs to recover a structured vector from under-sampled measurements. Moreover, we study the effect of the condition number of the analysis operator on the error bounds. Our analysis is general and holds for a variety of low-dimensional structures including sparse, block-sparse, analysis sparse [1] and gradient-sparse vectors, as well as low-rank matrices. In brief, the contributions of this work can be listed as follows.

1) How the conditioning affects the error estimate: For $f = \| \Omega \cdot \|_1$ in $P_f$, the error estimate from [1] (see (23)) explicitly depends on $\kappa(\Omega)$ as

$$\frac{E_{h_\beta}}{\text{ambient dimension}} \leq \frac{2\kappa(\Omega)}{\text{sparsity} \times \text{ambient dimension}}. \quad \text{(9)}$$

The above inequality includes two special cases that are of considerable interest in their own right, namely gradient and analysis sparsity. In the former, when $\Omega$ is the difference operator, $\kappa(\Omega)$ changes linearly with the ambient dimension. Besides, in the latter, when $\Omega$ is a highly redundant and coherent analysis operator, $\kappa(\Omega)$ can be arbitrary large. These two cases make the right-hand side of (9) in some sense trivial and suggest that $U_\delta$ does not approximate $\delta(D(\| \Omega \cdot \|_1, x))$ well. We precisely investigate this in Section IV.

2) Obtaining an error bound for $\omega(D(f, x) \cap \mathbb{B}^n)$ with rather general $f(\cdot)$: $\omega_\beta(D(f, x) \cap \mathbb{B}^n)$ itself is a lower bound for the required number of measurements [2], [11]. The significance of this error bound lies in the fact that it shows where we can use $U_\omega$ instead of $\omega(D(f, x) \cap \mathbb{B}^n)$. Our bound states that

$$\frac{E_{h}}{\sqrt{\text{ambient dimension}}} \leq h_1(\beta), \quad \text{(10)}$$

where $\beta$ is some parameter that depends implicitly on $\delta(f(x))$. Furthermore, $h_1(\beta)$ vanishes as ambient dimension grows large. This shows that $U_\omega$ describes $\omega(D(f, x) \cap \mathbb{B}^n)$ asymptotically well. To a great extent, the setting considered for $f$ (see (33)) is unrestricted. In particular, it includes the important special cases of $\| \cdot \|_{TV}$ and $\| \Omega \cdot \|_1$ for a highly coherent and redundant analysis operator.

3) Obtaining an error bound for $\delta(D(f, x))$ with rather general $f(\cdot)$: $\delta(D(f, x))$ precisely determines the boundary of failure and success of $P_f$. However, exact computation of $\delta(D(f, x))$ is very difficult. It is common to approximate $\delta(D(f, x))$ with $U_\delta$. By providing an error bound, we formally show that this approximation is good. More precisely, we show that

$$\frac{E_h}{\text{ambient dimension}} \leq h_2(\beta, \omega), \quad \text{(11)}$$

1Cosparse [10] Definition 1.
where \( \beta \) depends on \( \partial f(x) \), and \( h_2(\beta, \omega) \) is a function of \( \beta \) and \( \omega(D(f, x) \cap \mathbb{R}^n) \) that is succinctly shown by \( \omega \).
In addition, \( h_2(\beta, \omega) \) vanishes as the ambient dimension grows sufficiently large. Again, the setting considered for \( f \) (see (33)) holds for a broad class of low-dimensional structures such as gradient sparsity and cosparse with a highly redundant and coherent analysis operator. This bound, unlike the error bound in (2), reveals that \( U_\delta \) is a good measure for the number of measurements that \( P_f \) needs. It is also worth mentioning that the dependence of \( \beta \) on \( \partial f(x) \) somewhat limits the applicability of the method.

C. Notation

Throughout the paper, scalars are denoted by lowercase letters, vectors by lowercase boldface letters, and matrices by uppercase boldface letters. The \( i \)th element of the vector \( x \) is given either by \( x(i) \) or \( x_i \). The notation \( \langle \cdot \rangle^\dagger \) stands for the pseudo-inverse operator. We reserve calligraphic uppercase letters, vectors by lowercase boldface letters, and matrices by \( \mathbb{C} \) letters, vectors by lowercase boldface letters, and matrices by \( \mathbb{C} \).

D. Outline

The paper is organized as follows. The required concepts from convex geometry are reviewed in Section II. Section III discusses two approaches in obtaining the error estimate. Section IV is dedicated to present our main contributions. In Section V, we investigate the estimate in [1] and introduce some examples for which the error estimate does not work. In Section VI numerical experiments are presented which confirm our theory. Finally, the paper is concluded in Section VII.

II. CONVEX GEOMETRY

In this section, a review of basic concepts of convex geometry is provided.

A. Descent Cones

The descent cone \( D(f, x) \) at a point \( x \in \mathbb{R}^n \) consists of the set of directions that do not increase \( f \) and is given by:

\[
D(f, x) = \bigcup_{t > 0} \{ z \in \mathbb{R}^n : f(x + tz) \leq f(x) \}.
\] (13)

The descent cone reveals the local behavior of \( f \) near \( x \) and is a convex set. There is also a relation between descent cone and subdifferential [12] Chapter 23 given by:

\[
D(f, x) = \text{cone}(\partial f(x)) := \bigcup_{t > 0} t \partial f(x).
\] (14)

B. Statistical Dimension

Definition 1. Statistical Dimension [1]: Let \( C \subseteq \mathbb{R}^n \) be a closed convex cone. Statistical dimension of \( C \) is defined as:

\[
\delta(C) := E[\|P_C(g)\|_2^2] = E \text{ dist}^2(g, C^c),
\] (15)

where, \( P_C(x) \) is the projection of \( x \in \mathbb{R}^n \) onto the set \( C \) defined as: \( P_C(x) = \arg\min_{z \in C} \|z - x\|_2 \).

The statistical dimension extends the concept of linear subspaces to convex cones. Intuitively, it measures the size of a cone. Furthermore, \( \delta(D(f, x)) \) determines the precise location of transition from failure to success in \( P_f \).

C. Gaussian width

Definition 2. The Gaussian width of a set \( C \) is defined as:

\[
\omega(C) := E \sup_{y \in C} \langle y, g \rangle.
\] (16)

The relation between statistical dimension and Gaussian width which is summarized in the following [2] Proposition 3.6. [1] Proposition 10.2.

\[
\omega(C \cap \mathbb{S}^{n-1}) \leq \omega(C \cap \mathbb{R}^n) = E[\|P_C(g)\|_2^2] = E \text{ dist}(g, C^c),
\] (17)

\[
\omega^2(C \cap \mathbb{R}^n) = (E[\|P_C(g)\|_2^2])^2 \leq \delta(C).
\] (18)

In fact, these concepts differ numerically by at most 1.

D. Optimality Condition

In the following, we characterize when \( P_f \) succeeds in noise-free case.

Proposition 1. [2] Proposition 2.1] Optimality condition: Let \( f \) be a proper convex function. The vector \( x \in \mathbb{R}^n \) is the unique optimal point of \( P_f \) if and only if \( D(f, x) \cap \text{null}(A) = \{0\} \).

The next theorem determines the number of measurements needed for successful recovery of \( P_f \) for any proper convex function \( f \).

Theorem 1. [7] Theorem 2: Let \( f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm \infty\} \) be a proper convex function and \( x \in \mathbb{R}^n \) a fixed sparse vector. Suppose that \( m \) independent Gaussian linear measurements of \( x \) are observed via \( y = Ax \in \mathbb{R}^m \). If

\[
m \geq \delta(D(f, x)) + \sqrt{8 \log \frac{4}{\eta}} n,
\] (19)

for a given probability of failure (tolerance) \( \eta \in [0, 1] \), then, we have

\[
P(D(f, x) \cap \text{null}(A) = \{0\}) \geq 1 - \eta.
\] (20)
Besides, if
\[ m \leq \delta(D(f, x)) - \sqrt{8 \log(\frac{4}{\eta}) n} \]  
(21)
then,
\[ D(f, x) \cap \text{null}(A) = \{0\} \leq \eta \]  
(22)

III. RELATED WORKS IN ERROR ESTIMATION

For bounding the distance between \( \omega(D(f, x) \cap \mathbb{R}^n) \) and \( U_\omega \) (alternatively, between \( \delta(D(f, x)) \) and \( \delta_\omega \)), two different approaches are proposed in [1], [11]. In the following, we briefly describe these methods.

Result 1. [7] Theorem 4.3] Let \( f \) be a norm. Then, for any \( x \in \mathbb{R}^n \setminus \{0\} \):
\[ 0 \leq \inf_{\tilde{x} > 0} \text{dist}^2(g, \tilde{f}(x)) - \delta(D(f, x)) \leq \frac{2 \sup_{x \in \tilde{f}(x)} \|s\|_2}{f(\|x\|_2)} \]  
(23)

Result 2. [11] Proposition 1] Suppose that for \( x \in \mathbb{R}^n \setminus \{0\} \), \( \tilde{f}(x) \) satisfies a weak decomposability assumption:
\[ \exists z_0 \in \tilde{f}(x) \text{ s.t. } \langle x - z_0, z_0 \rangle = 0, \forall z \in \tilde{f}(x). \]  
(24)
Then,
\[ \inf_{\tilde{x} > 0} \text{dist}(g, \tilde{f}(x)) \leq \omega(D(f, x) \cap \mathbb{R}^n) + 6. \]  
(25)

In [12], it is shown that \( f = \| \cdot \|_{\text{TV}} \) satisfies the weak decomposability assumption (24).

A. Explanations

Proposition 2 presents an error estimate for the Gaussian width of the descent cone (restricted to unit ball) that is used to upper bound the number of Gaussian measurements in various low-dimensional structures [2] Section 3.1. The condition (24) holds for most of the structures except \( \ell_1 \) analysis (See V-A). The error estimate (24), however, depends on the function \( f \) and \( \tilde{f}(x) \). Although [7] Theorem 4.3 restricts \( f \) to be a norm, the provided proof also supports semi-norms such as TV. In some cases, the error bound in (23) becomes large, particularly for \( f = \| \cdot \|_{\text{TV}} \) and a vector \( x \) with large elements but small variations, is an illustrative example. In Proposition 2 we study further examples. A naive interpretation of this fact is that \( U_\delta \) is a poor approximation of \( \delta(D(f, x)) \) in such cases. Fortunately, as we show in Section V this argument is invalid which in turn suggests that (23) is a loose bound in those cases.

IV. THE EXISTING APPROACH

The cases of sparse vectors and low-rank matrices are well-studied in the literature. In particular, for these two cases, some error bounds are already available. In this section, we extend the general result of [1] (see Result 1) to block-sparse, cosparse and low-variational vectors. Next, we discuss the shortcomings of this approach.

A. Summary of bounds

In the following proposition, we evaluate the error bound in Result 1 for any of \( \ell_1, \ell_{1,2}, \) nuclear norm, \( \ell_1 \) analysis and TV.

Proposition 2. Define the error as (7). The normalized error in different low-dimensional models is bounded as follows.

Sparse Vectors: Let \( x \in \mathbb{R}^n \) be a s-sparse vector and \( f = \| \cdot \|_s \), then:
\[ E_\delta_{\text{n}} \leq \frac{2}{\sqrt{n/s}} \]  
(26)

Block-sparse Vectors: Let \( x \in \mathbb{R}^n \) be an s-block-sparse vector that consists of q blocks of equal length k and \( f = \| \cdot \|_{1,2} \), then,
\[ E_\delta_{\text{q}} \leq \frac{2}{\sqrt{q/\sqrt{k}}} \]  
(27)

Low-rank Matrices: Let \( X \in \mathbb{R}^{n_1 \times n_2} \) be a matrix with rank \( r < \min\{n_1, n_2\} \) and \( f = \| \cdot \|_{\text{TV}} \), then,
\[ E_\delta_{\text{TV}} \leq \frac{2}{n_2 \sqrt{r n_1}} \]  
(28)

Cosparse Vectors: Let the analysis coefficients of \( x \in \mathbb{R}^n \) in the analysis domain \( \Omega \in \mathbb{R}^{p \times n} (p \geq n) \) form a s-sparse vector. Also assume \( f = \| \Omega \cdot \|_{1} \), then,
\[ E_\delta_{\text{p}} \leq \frac{2n(k\Omega)}{\sqrt{s p}} \]  
(29)

Gradient-sparse Vectors: Let \( x \in \mathbb{R}^n \) be a s-sparse vector in the gradient domain \( \Omega_{\text{grad}} \in \mathbb{R}^{n-1 \times n} \) and \( f = \| \cdot \|_{\text{TV}} \). Then,
\[ E_\delta_{\text{TV}} \leq \frac{2n(k\Omega_{\text{grad}})}{\sqrt{(n-1)s}} \]  
(30)

Proof. See Appendix A-D.

B. Discussion

Before we describe our main contribution in Section V, we first examine the above bounds case by case.

• Sparse vectors.
  The error estimate in [26] becomes large in low-sparsity regimes. In [1] Comentary 10.1.2], the use of \( U_\delta \) instead of \( \delta(D(\| \cdot \|_1, x)) \) is recommended only for \( s > \sqrt{n} + 1, \) for which the error bound on \( E_\delta \) becomes small. In particular, the bound in (23) is inclusive whether (4) in case of \( f = \| \cdot \|_1 \) determines the phase transition when \( s \leq \sqrt{n} \) (See Figure 1).

• Block-sparse Vectors.
  The error estimate (27) works well when \( s > \sqrt{q} + 1. \) However, similar to the previous case for \( s \leq \sqrt{q} \) the error bound is large (See Figure 2).

• Low-rank Matrices.
  [28] implies that the upper bound (3) in case of \( f = \| \cdot \|_{\text{TV}} \) does not specify the location of the phase transition.
when \( r \ll \min \{n_1, n_2\} \). The latter situation happens, for example, in very tall matrices (See Figure 5). Meanwhile, when \( r \) is sufficiently large, the error vanishes asymptotically.

- **Cosparse Vectors.**
  For highly coherent analysis operator \( \Omega \in \mathbb{R}^{p \times n} \), the condition number \( \kappa(\Omega) \) can become arbitrarily large. In turn, the upper-bound in (29) becomes large (See for example Figure 6). The bound becomes even worse at low sparsity regimes. In redundant analysis operators, however, \( s \) is lower bounded by \( p - n \) [10, Section 4]. (29) implies that the upper bound is not sharp in highly coherent analysis operators (See for example Figure 6) since \( \kappa(\Omega) \) can become arbitrarily large and in particular gets worse in the low sparsity regimes however \( s \) must be larger than \( p - n \) in redundant analysis operator \( \Omega \) [10, Section 4]. As an example of a highly coherent dictionary, consider the below Hilbert matrix \( \Omega \in \mathbb{R}^{n \times n} \)

\[
\Omega(i,j) = \frac{1}{i + j - 1}.
\]

(31)

For this matrix, \( \kappa(\Omega) \) grows like \( O\left(\frac{\ln n}{\sqrt{n}}\right) \) and the upper-bound in (29) gets very large as \( n \) grows.

- **Gradient-sparse vectors.**
  Because of the regular structure of the finite difference matrix, we are able to express its condition number via the closed form:

\[
\kappa(\Omega_d) = \sqrt{\frac{1 - \cos(\frac{\pi(n-1)}{n})}{1 - \cos(\frac{\pi}{n})}}.
\]

(32)

For large \( n \), we have \( \kappa(\Omega_d) \approx \frac{2n}{\pi} \). Thus, \( \kappa(\Omega_d) \) and consequently, the upper-bound in (30) grow linearly with \( n \). Again, (30) is inconclusive whether \( \delta \) could be approximated with \( U_\delta \) in case of gradient-sparse vectors (See Figure 4).

**V. MAIN RESULTS**

Our main results which are stated in the following theorem, estimate the distance between \( \delta(D(f,x)) \) and \( \omega(D(f,x) \cap B^n) \) from their corresponding upper bounds.

**Theorem 2.** Let \( f \) be a proper convex function that promotes the structure of \( x \neq 0 \in \mathbb{R}^n \) and let \( g \in \mathbb{R}^n \) be a standard i.i.d Gaussian vector. Suppose \( \partial f(x) \) satisfies

\[
\exists z_0 \neq 0 \text{ s.t. } \langle z - z_0, z_0 \rangle = 0, \forall z \in \partial f(x).
\]

(33)

Then for any positive values of \( \lambda, \zeta \), we have that

\[
0 \leq \inf_{t \geq 0} \mathbb{E} \text{ dist}^2(g, \text{cone}(\partial f(x))) - \delta(D(f,x)) \leq (4\lambda \beta + \gamma) \omega(D(f,x) \cap B^n) + \gamma(\zeta + 2\lambda \beta) + 4\lambda^2 \beta^2,
\]

(34)

and

\[
0 \leq \inf_{t \geq 0} \mathbb{E} \text{ dist}(g, t\partial f(x)) - \omega(D(f,x) \cap B^n) \leq 1.6 + 4\beta,
\]

(35)

where \( \gamma \) is the constant

\[
\gamma = \sqrt{\frac{3}{\ln(1 - 4e^{-2\zeta/2} - 2e^{-2\gamma/2})}}.
\]

(36)

and \( \beta = 1 \) for \( f = \{ \| \cdot \|_1, \| \cdot \|_{1,2}, \| \cdot \|_4, \| \cdot \|_{TV} \} \).

When \( f = \| \cdot \|_1 \), \( \beta \) is given by

\[
\beta = \frac{\| z_1 \|_2}{\| z_0 \|_2},
\]

(37)

where \( z_1 = \arg \min_{z \in \partial f(x)} \| z \|_2 \).

Proof. See Appendix A-E.

Regardless of the choice of \( f \) and \( x \), the error bounds (35) and (34) vanish as dimension \( n \) grows, as long as the condition (33) holds (which is weaker than (24)). In fact, \( 1.6 + 4\beta \to 0 \) as \( n \to \infty \), and

\[
(4\lambda \beta + \gamma) \omega(D(f,x) \cap B^n) + \gamma(\zeta + 2\lambda \beta) + 4\lambda^2 \beta^2 \leq \frac{n}{\sqrt{n}} \to 0 \text{ as } n \to \infty,
\]

(39)

where the inequality comes from [18] and [11, Equation 3.7].

**A. DISCUSSION**

All the low-dimensional structures that we consider in this work, satisfy the condition (33) which is less restrictive than the weak decomposability assumption of [11] (see, [11, Remark 1]), as \( z_0 \) is not required to be contained in the subdifferential. For \( \| \cdot \|_1, \| \cdot \|_{1,2}, \| \cdot \|_4, \| \cdot \|_{TV} \), one can always find \( z_1 = z_0 \in \partial f(x) \) such that (33) holds (see [14, Definition 2] and [13, Lemma 1]). Consequently, \( \beta \) could be set to 1. In the following proposition, we discuss the error estimate of \( \delta(D(\Omega : \| \cdot \|_1, x)) \) and \( \omega(D(\| \cdot \|_1, x) \cap B^n) \) which is not analyzed in the literature yet. The bounds (34) and (35) show that \( U_\omega \) and \( U_\delta \) well represent \( \omega(D(f,x)) \) and \( \delta(D(f,x)) \) asymptotically in case of \( \ell_1 \) analysis even in a highly coherent and redundant dictionary.

**Proposition 3.** Consider the cosparse vector \( x \in \mathbb{R}^n \) in the analysis domain \( \Omega \in \mathbb{R}^{p \times n} \) with support \( S \). Then

\[
z_0 = \Omega_S^T \text{sgn}(\Omega x)_S - \Omega_S^T \Omega_S \Omega_S^T \Omega_S^T \Omega_S^T \text{sgn}(\Omega x)_S
\]

(40)

satisfies (33) and \( z_0 \neq 0 \).

Proof. See Appendix A-E.

In \( \ell_1 \) analysis, \( \beta \) is small even in a highly redundant and coherent dictionary (See Figs. 10, 8, 9). This shows that the bounds (34) and (35) are not very sensitive to \( \beta \) when dimensions \( p \) and \( n \) are large.

**VI. NUMERICAL EXPERIMENTS**

In this section, we numerically compare the new error bound of (34) against the bound (24) found by the existing approach for various low-dimensional structures. For each test, we optimize \( \lambda \) and \( \zeta \) to minimize the right-hand side of (34).
Figures 1, 2, and 3 show the error bounds (34) and (23) for \( \| \cdot \|_1, \| \cdot \|_{1,2} \) and \( \| \cdot \|_\infty \), respectively. In all cases, the sparsity/rank values are set very small. To compute \( \omega(D(f, x) \cap \mathbb{R}^n) \) in (34), we used its upper-bound obtained via (18), (1) Equations D.6, D.10, and [15, Lemma 1]. It is clear from these figures that the new error bound outperforms the previous error bound (23) in very low sparsity/rank regimes; it should be emphasized that the curves depict the upper-bound of (34). Figure 4 compares the two bounds in case of \( \| \cdot \|_{TV} \) for various sparsity levels. The expectation operator in the upper-bound of \( \omega(D(\| \cdot \|_{TV}, x) \cap \mathbb{R}^n) \) in (34) is computed numerically with Monte Carlo simulations. This figure shows that the new bound is at least 50 times smaller than the previous bound in the considered setting. In Figures 5 and 6, we generated two redundant and coherent dictionaries \( \Omega \in \mathbb{R}^{p \times n} \) with dimensions \( (p = 500, n = 480) \) and \( (p = 1000, n = 500) \), and with condition numbers \( \kappa(\Omega) = 280.1591 \) and \( \kappa(\Omega) = 4118 \), respectively. The relative error \( \frac{\text{New error bound}}{\text{Previous error bound}} \) is almost the same in different analysis sparsity levels. Also, the new error bound outperforms the previous one even in highly redundant and coherent dictionaries. Next we shall determine the value of \( \beta \) in different settings. In Figure 7, we plot the value of \( \beta \) for 50 random realizations of \( x \in \mathbb{R}^n \) with \( s = 800 \) and \( \Omega \in \mathbb{R}^{1000 \times 500} \). It is observed that \( \beta \) takes almost the same value for different \( x \). We examine the dependence of \( \beta \) on the sparsity level in Figures 8 and 9 for two different settings. The shown curves are obtained by averaging the value of \( \beta \) over 10 trials for each sparsity level. The results confirm that \( \beta \) slowly decreases as analysis sparsity level grows.

In Figure 10, we consider the highly coherent Hilbert dictionary with \( \kappa(\Omega) = 1.7 \times 10^{20} \) \( (p = n = 500) \). It is remarkable to see that the value of \( \beta \) in Figure 10 remains bounded below 1 for the whole range of sparsity levels.

Based on Figures 8-10, we conclude that the changes in \( \beta \) is small even in highly redundant and coherent settings. This in turn assures that the bounds (34) and (35) are not severely affected by the change of \( \beta \) values in different settings.

VII. CONCLUSION

In this work, we presented an error estimate bound for the statistical dimension. This new bound together with the bound (23) shows that the statistical dimension is well described by its common upper bound (4) in the structures \( \ell_1, \ell_{1,2}, \) nuclear norm, TV and \( \ell_1 \) analysis.

APPENDIX A

PROOF OF MAIN RESULT AND LEMMAS

A. Proof of Theorem 2 (34)

Before proving the main result, in the following lemma, we obtain a concentration inequality for the function \( \phi(g) := \| g \| \alpha \| \hat{f}(x) \| - \| g \| \alpha \| \hat{f}(x) \| + 2\lambda \| \hat{f}(x) \| \) for various \( \lambda > 0 \) and different \( \| g \| \alpha \| \hat{f}(x) \| \).

Lemma 1. Let \( g \in \mathbb{R}^n \) be a standard normal i.i.d vector and \( \phi(g) \) is defined as (41). Then for given parameters \( \lambda, \zeta > 0 \),

\[
P \left\{ \phi(g) - \mathbb{E}[\phi(g)] \right\} \leq \left( \frac{3}{1 - 4e^{-\frac{3}{2}}} \right) \leq p_0.
\]

Proof. Define the event

\[
\mathcal{E} = \left\{ |t_g - \mathbb{E}[t_g]| \leq \frac{\lambda}{\| z_0 \|_2} \right\}.
\]

Since squaring a function does not change the minimum in a non-negative domain and due to (33), \( t_g \) in (43) is a \( \frac{1}{2} \)-Lipschitz function of \( g \) ([11, Lemma 3]). The proof of ([11, Lemma 3]) does not need \( z_0 \) to be an element of \( \hat{f}(x) \).

Therefore due to concentration inequality:

\[
P(\mathcal{E}) \geq 1 - 2e^{-\frac{\lambda^2}{2}} := p_0.
\]
Suppose $E$ holds. Define $z^*$ such that
\[ \text{dist}^2(g, t_g \hat{c} f(x)) = \|g - t_g z^*\|^2_2. \]  
(47)

Take $\alpha := E[t_g] + \frac{\lambda}{\|z\|_2}$ and
\[ z = t_g \frac{\alpha}{\|g\|} z^* + \left(1 - t_g \frac{\alpha}{\|g\|}\right) z_1 \in \hat{c} f(x). \]  
(48)

Then,
\[ \text{dist}^2(g, \alpha \hat{c} f(x)) \leq \|g - \alpha z\|^2_2 = \|g - t_g z^* + t_g z^* - \alpha z\|^2_2 \]
\[ < \|g - t_g z^*\|^2_2 + \|t_g z^* - \alpha z\|^2_2 + \]
\[ 2 \|g - t_g z^*\| \|z^* - \alpha z\| \leq \text{dist}^2(g, \text{cone}(\hat{c} f(z)))+
\[ (t_g - \alpha) \|z_1\|^2_2 + 2|t_g - \alpha| \|g - t_g z^*\| \|z_1\| \leq \]
\[ \text{dist}^2(g, \text{cone}(\hat{c} f(z)))+4 \lambda^2 \beta^2 + 4 \lambda \beta \text{dist}(g, \text{cone}(\hat{c} f(x))). \]  
(49)

where we used the definition of the event $E$. Therefore
\[ P\{\phi(g) \leq 4 \lambda^2 \beta^2\} > P\{E\} \geq 1 - 2e^{-\frac{\lambda^2}{2}}. \]  
(50)

Also by Lemma 1, we show that
\[ P\{\phi(g) \leq E[\phi(g)] - 
\sqrt{72}(\zeta + Ef_2(g) + 2\lambda \beta) \sqrt{\ln \frac{3}{1 - 4e^{-\frac{\lambda^2}{2}} - 2e^{-\frac{\lambda^2}{2}}}} \]
\[ < p_0, \]  
(51)

where $f_2(g)$ is defined in (60). By considering (50) and (51), we reach a contradiction unless:
\[ E\{\phi(g)\} \leq \sqrt{72}(\zeta + Ef_2(g) + 2\lambda \beta) \sqrt{\ln \frac{3}{1 - 4e^{-\frac{\lambda^2}{2}} - 2e^{-\frac{\lambda^2}{2}}}} 
+ 4 \lambda^2 \beta^2. \]  
(52)

Thus, we reach the right hand side of (34). The left hand side is obtained by the fact that infimum of an affine function is concave and Jensen’s inequality.

**B. Proof of Theorem 2**

**Proof.** Due to (33) and [11, Lemma 3], $t_g$ is a $\frac{1}{\|z\|_2}$-Lipschitz function of $g$. Now, suppose that [15] holds. Define $z^*$ such that
\[ \text{dist}(g, t_g \hat{c} f(x)) = \|g - t_g z^*\|_2. \]  
(53)

Then due to $z$ in (48) and $\alpha$ in (42), we have:
\[ \text{dist}(g, \alpha \hat{c} f(x)) \leq \|g - \alpha z\|_2 = \|g - t_g z^* + t_g z^* - \alpha z\|_2 \]
\[ \leq \|g - t_g z^*\|_2 + \|t_g z^* - \alpha z\|_2 \leq \|g - t_g z^*\|_2 + \]
\[ t_g - \alpha \|z_1\|_2 \leq \text{dist}(g, \text{cone}(\hat{c} f(x)) + 2 \lambda \beta. \]  
(54)

Define the function
\[ \phi_1(g) := \text{dist}(g, \alpha \hat{c} f(x)) - \text{dist}(g, \text{cone}(\hat{c} f(x))). \]  
(55)

We have:
\[ P\{\phi_1(g) < 2 \lambda \beta\} > P\{E\} > 1 - 2e^{-\frac{\lambda^2}{2}}. \]  
(56)

Since $\phi_1(g)$ is 2 Lipschitz function of $g$, due to concentration inequality we have:
\[ P\{\phi_1(g) - E[\phi_1(g)] \leq -r\} \leq e^{-\frac{r^2}{2}}. \]  
(57)

With a change of variable, we reach:
\[ P\{\phi_1(g) - E[\phi_1(g)] \leq -\sqrt{\frac{8 \ln \frac{1}{p_0}}{p_0}} \} \leq p_0. \]  
(58)

By considering (56) and (58), we reach a contradiction unless:
\[ E\{\phi_1(g)\} \leq \sqrt{\frac{8 \ln \frac{1}{1 - 2e^{-\frac{\lambda^2}{2}}} + 2 \lambda \beta.} \]  
(59)

By setting $\lambda = 2$, we reach (35).
Then, with the same reasoning, we have:

\[ f(\kappa) \text{ and coherent dictionary with } \}

\[ \text{Fig. 9. } \beta \text{ changes versus sparsity in a usual redundant and coherent dictionary with parameters } p = 50, n = 30, \kappa(\Omega) = 110. \]

\[ \text{Fig. 10. } \beta \text{ changes versus sparsity in a non-redundant and highly coherent dictionary with parameters } p = n = 500, \kappa(\Omega) = 1.7 \times 10^{20}. \text{ Below the analysis sparsity 200, } \beta = 0 \text{ and is not shown.} \]

\[ \text{Consequently,} \]

\[ \mathbb{P}(\phi(g) - E[\phi(g)] \leq -r) = \]

\[ \mathbb{P}\left\{ h_1 - E[h_1] - h_2 + E[h_2] - 4\lambda\beta f_2 + 4\lambda\beta E[f_2] \leq -\frac{r}{3}\right\} \leq \]

\[ \mathbb{P}\left\{ h_1 - E[h_1] \leq -\frac{r}{3}\right\} + \mathbb{P}\left\{ h_2 - E[h_2] \geq \frac{r}{3}\right\} + \]

\[ \mathbb{P}\left\{ f_2 - E[f_2] \geq \frac{r}{12\lambda\beta}\right\} \leq \mathbb{P}\left\{ h_1 - E[h_1] \leq -\frac{r}{3}\right\} \mathbb{P}\{\mathcal{E}_1\} + \]

\[ \mathbb{P}\left\{ h_2 - E[h_2] \geq \frac{r}{3}\right\} \mathbb{P}\{\mathcal{E}_1\} + \]

\[ \mathbb{P}\left\{ f_2 - E[f_2] \geq \frac{r}{12\lambda\beta}\right\} \leq e^{-\frac{r^2}{72(\lambda + E[f_2] + 2\lambda\beta)^2}} + 2e^{-\frac{r^2}{2}} + e^{-\frac{r^2}{2}} + e^{-\frac{r^2}{72(\lambda + E[f_2] + 2\lambda\beta)^2}} + e^{-\frac{r^2}{2}} + e^{-\frac{r^2}{2}}, \]

\[ \text{where in the third inequality, we used} \]

\[ \mathbb{P}\left\{ h_1 - E[h_1] \leq -\frac{r}{3}\right\} = \mathbb{P}\left\{ h_1 - E[h_1] \leq -\frac{r}{3}\right\} \mathbb{P}\{\mathcal{E}_1\}, \]

\[ \mathbb{P}\left\{ h_2 - E[h_2] \geq \frac{r}{3}\right\} \mathbb{P}\{\mathcal{E}_1\} + \mathbb{P}\left\{ f_2 - E[f_2] \geq \frac{r}{12\lambda\beta}\right\} \leq e^{-\frac{r^2}{72(\lambda + E[f_2] + 2\lambda\beta)^2}} + 2e^{-\frac{r^2}{2}} + e^{-\frac{r^2}{2}}, \]

\[ \text{and (64). With a change of variable, we reach (44).} \]

\[ \text{D. Proof of Proposition 2}\]

\[ \text{Proof. (26) and (28) are proved in [1] Section D.2 and [1] Section D.3 respectively. For (27), see [15] Proposition 3.} \]

Now, we prove (29). By using the error bound (23), we obtain the numerator of the error bound (23) as given by:

\[ 2 \sup_{x \in [\Omega \cdot 1]} \|x\|_2 = \sup_{x \in [\Omega \cdot 1]} 2\|\Omega^T z\|_2, \]

\[ \leq 2\|\Omega\|_{2 \to 2} \|z\|_{\infty, 1} \leq 2\|\Omega\|_{2 \to 2} \leq \sqrt{3}\|\Omega\|_{2 \to 2}^{-1}. \]

\[ \text{Also, since } \Omega \text{ is full column rank, the denominator can be lower bounded as below:} \]

\[ \frac{\|\Omega x\|_1}{\|x\|_2} \geq \frac{\|\Omega x\|_1}{\|\Omega x\|_2} \leq \sqrt{3}\|\Omega\|_{2 \to 2}^{-1}. \]

\[ \text{The error bound in (23) depends only on } \mathcal{D}(\|\Omega \cdot 1\|, x) \text{ and } \hat{c} \cdot \|\Omega x\|_2. \text{ Further, } \|\Omega x\|_2 \text{ depends only on } \text{sgn}(\Omega x) \text{ not the magnitudes of } \Omega x. \text{ So a vector} \]

\[ z = \left\{ \begin{array}{ll} \text{sgn}(\Omega x)_i, & i \in S \\ 0, & i \in S \end{array} \right\} \in \mathbb{R}^p, \]
with \( \text{sgn}(z) = \text{sgn}(\Omega x) \) can be chosen to have equality in the last inequality in (68). Therefore, the error is at most \( \frac{2 \kappa_1}{\Omega \Omega} \). Similar to (29) and with an additional assumption \( x \neq \text{null}(\Omega_d) \), (30) is proved.

E. Proof of Proposition 3

Proof. It is sufficient to prove that there exists a \( w_0 := \text{sgn}(\Omega x)_{\mathcal{S}} + v_{0,\mathcal{S}} \) such that
\[
\langle \Omega^T(w - w_0), \Omega^T w_0 \rangle = 0 : \forall w \in \partial \| \cdot \|_1(\Omega x).
\]
As such, we have:
\[
\langle \Omega^T_{\mathcal{S}}\text{sgn}(\Omega x)_{\mathcal{S}} + \Omega^T_{\mathcal{S}} v_{\mathcal{S}}, \Omega^T w_0 \rangle = \| \Omega^T w_0 \|_2^2,
\forall v_{\mathcal{S}} \text{ with } \| v_{\mathcal{S}} \|_\infty \leq 1,
\]
where reduces to:
\[
\Omega^T_{\mathcal{S}}\text{sgn}(\Omega x)_{\mathcal{S}} + \Omega^T_{\mathcal{S}} v_{0,\mathcal{S}} = 0.
\]
With the assumption that \( \Omega^T_{\mathcal{S}}\Omega_{\mathcal{S}} \) is full rank which is reasonable in our simulations VI the result is obtained.

References

[1] 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.
[2] 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.
[3] D. Donoho and J. Tanner, “Counting faces of randomly projected polytopes when the projection radically lowers dimension,” Journal of the American Mathematical Society, vol. 22, no. 1, pp. 1–53, 2009.
[4] M. Rudelson and R. Vershynin, “On sparse reconstruction from fourier and gaussian measurements,” Communications on Pure and Applied Mathematics, vol. 61, no. 8, pp. 1025–1045, 2008.
[5] M. Bayati, M. Lelarge, and A. Montanari, “Universality in polytope phase transitions and message passing algorithms,” The Annals of Applied Probability, vol. 25, no. 2, pp. 753–822, 2015.
[6] M. Stoicnic, “Various thresholds for \( \ell_1 \)-optimization in compressed sensing,” arXiv preprint arXiv:0907.3666, 2009.
[7] 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.
[8] S. Oymak and B. Hassibi, “Sharp mse bounds for proximal denoising,” Foundations of Computational Mathematics, vol. 16, no. 4, pp. 965–1029, 2016.
[9] Y. Gordon, “On milman’s inequality and random subspaces which escape through a mesh in \( \mathbb{R}^n \),” in Geometric Aspects of Functional Analysis, pp. 84–106, Springer, 1988.
[10] M. Kabanava and H. Rauhut, “Analysis \( \ell_1 \)-recovery with frames and Gaussian measurements,” Acta Applicandae Mathematicae, vol. 140, no. 1, pp. 173–195, 2015.
[11] 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.
[12] R. T. Rockafellar, Convex analysis. Princeton university press, 2015.
[13] 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.
[14] E. Candès and B. Recht, “Simple bounds for recovering low-complexity models,” Mathematical Programming, pp. 1–13.
[15] S. Daei, F. Haddadi, and A. Amini, “Exploiting prior information in block sparse signals,” arXiv preprint arXiv:1804.08444, 2018.