THE PROPORTIONAL MEAN DECOMPOSITION: A BRIDGE BETWEEN THE GAUSSIAN AND BERNOULLI ENSEMBLES

Samet Oymak[b] *Babak Hassibi*[c]

[b] University of California, Berkeley
[c] California Institute of Technology

ABSTRACT

We consider ill-posed linear inverse problems involving the estimation of structured sparse signals. When the sensing matrix has i.i.d. standard normal entries, there is a full-fledged theory on the sample complexity and robustness properties. In this work, we propose a way of making use of this theory to get good bounds for the i.i.d. Bernoulli ensemble. We first provide a deterministic relation between the two ensembles that relates the restricted singular values. Then, we show how one can get non-asymptotic results with small constants for the Bernoulli ensemble. While our discussion focuses on Bernoulli measurements, the main idea can be extended to any discrete distribution with little difficulty.

Index Terms— compressed sensing, sample complexity, gaussian processes, restricted singular value

1. INTRODUCTION

Suppose we wish to recover a structured sparse signal $\mathbf{x}_0 \in \mathbb{R}^n$ from underdetermined linear observations $\mathbf{A}\mathbf{x}_0 \in \mathbb{R}^m$. Typically, one can minimize a suitable convex penalty $f(\cdot)$ (such as ℓ_1-norm) and solve

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}_0. \tag{1.1}$$

The question of interest is often the sample complexity, i.e., the number of measurements to reliably estimate the true signal via (1.1). The Gaussian ensemble has particular importance in understanding the behavior of low-dimensional representation problems such as (1.1). Starting from the initial works of Donoho and Tanner [1,2], the recent literature made it clear that one can take advantage of the powerful results on Gaussian processes to precisely predict the performance of (1.1) [3,4]. The results on the noiseless problem were also extended to the related problems such as convergence rates in the noisy setup and recovery from sparse corruption [5–8]. While Gaussian ensemble enjoys an abundance of results the same cannot be said for other ensembles such as the subsampled Discrete Fourier Transform. Our interest in this work is to lay out a framework to obtain results for i.i.d. nongaussian measurements by constructing a proper measure of similarity to the Gaussian measurements. We will restrict our attention to symmetric Bernoulli (i.e. Rademacher) measurements, which are equally likely to be $+1$ and -1. However, the proposed framework can be extended from Bernoulli to other discrete distributions without much effort. Focusing on Bernoulli measurements will make our results cleaner and arguably more elegant. Bernoulli measurement ensemble is interesting in its own right as it is advantageous both from computation and storage points of view [9, 10].

We show that Bernoulli measurements can be used for linear inverse problems in a similar manner to Gaussian's by paying a price of constant multiplier in front of the Gaussian sample complexity. This is along the lines of the recent work of Tropp [11], which provide similar guarantees for the subgaussian ensemble up to unknown constants. [11] uses bounds on nonnegative empirical processes developed by Mendelson et al. as the core technical tool [12, 13]. The reader is also referred to the work by Plan and Vershynin [14] for inherently related results where one observes one-bit measurements $\text{sign}(\mathbf{A}\mathbf{x}_0)$. Unlike our setup, one-bit observations do not allow for the exact recovery of the signal, however, the authors are able to show that good estimation performance is achievable while undersampling proportional to the sparsity. Compared to these, we present a novel strategy that allows us to measure the similarity between two distributions, which yields explicit constants. We are also able to move beyond the standard problem (1.1) to its variations such as the noisy estimation and estimation under outliers. To start the technical discussion, let us define the restricted singular value, which will be crucial for our exposition.

Definition 1.1 (Restricted Singular Value)**.** *Given a nonempty and nonzero cone $\mathcal{C} \subset \mathbb{R}^n$ and a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, the minimum and maximum restricted singular values of $\mathbf{A}$ at $\mathcal{C}$ are respectively defined as*

$$\sigma_{\mathcal{C}}(\mathbf{A}) = \min_{\mathbf{v}\in\mathcal{C}, \|\mathbf{v}\|=1} \|\mathbf{A}\mathbf{v}\|_2, \quad \Sigma_{\mathcal{C}}(\mathbf{A}) = \max_{\mathbf{v}\in\mathcal{C}, \|\mathbf{v}\|=1} \|\mathbf{A}\mathbf{v}\|_2.$$

Observe that these quantities reduces to usual minimum and maximum singular value when $\mathcal{C} = \mathbb{R}^n$. To give an initial intuition, we start with a basic comparison between a matrix with independent $\mathcal{N}(0, 1)$ entries and one with symmetric Bernoulli entries.

Proposition 1.1. *Let $\mathcal{C} \subset \mathbb{R}^n$ be a nonempty, closed cone. Let $\mathbf{G} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{m\times n}$ be matrices with independent $\mathcal{N}(0,1)$ and symmetric Bernoulli entries respectively. Suppose, for some $\varepsilon_1, \varepsilon_2 > 0$, we have*

$$(1-\varepsilon_1)m \le \mathbb{E}\,\sigma_{\mathcal{C}}^2(\mathbf{G}) \le \mathbb{E}\,\Sigma_{\mathcal{C}}^2(\mathbf{G}) \le (1+\varepsilon_2)m. \quad (1.2)$$

Then, we also have

$$(1-\frac{\pi}{2}\varepsilon_1)m \le \mathbb{E}\,\sigma_{\mathcal{C}}^2(\mathbf{B}) \le \mathbb{E}\,\Sigma_{\mathcal{C}}^2(\mathbf{B}) \le (1+\frac{\pi}{2}\varepsilon_2)m.$$

Observe that, if $\mathcal{C}$ is a line i.e. $\mathcal{C} = \{\alpha\mathbf{v} \mid \alpha \in \mathbb{R}\}$ for some $\mathbf{v} \in \mathbb{R}^n$, the RSV's are trivially equal to m as we have

$$\mathbb{E}\,\|\mathbf{G}\mathbf{v}\|_2^2 = \mathbb{E}\,\|\mathbf{B}\mathbf{v}\|_2^2 = m\|\mathbf{v}\|_2^2.$$

On the other hand, in general, it is nontrivial to estimate these quantities. The RSV plays an important role in the analysis of (1.1) and has been the subject of interest recently. A standard example is when we let $\mathcal{C}$ to be the set of at most k sparse vectors, i.e.

$$\mathcal{C} = \{\mathbf{v} \in \mathbb{R}^n \mid \|\mathbf{v}\|_0 \le k\}.$$

In this case, the smallest possible ε in (1.2) effectively corresponds to the k-Restricted Isometry Constant (RIC) $\delta_k(\mathbf{G})$ of the Gaussian matrix $\mathbf{G}$[1]. Hence, Proposition 1.1 deterministically relates the $\delta_k(\mathbf{G})$ and $\delta_k(\mathbf{B})$, namely, $\delta_k(\mathbf{B}) \le \frac{\pi}{2}\delta_k(\mathbf{G})$. Following from this initial observation, in the remainder of the paper, we will outline the main idea and illustrate its applications on (1.1) and its extensions. We note that detailed results and some of the proofs will be deferred to the technical report [15].

Notation: Given a probability density function (p.d.f) $f(\cdot)$, mean(f) and var(f) will correspond to the mean and variance of the associated random variable. The symbol $\sim$ should be read "is distributed as". The unit ℓ_2 ball and sphere will be denoted by $\mathcal{B}^{n-1}$ and $\mathcal{S}^{n-1}$. If $\mathcal{C}$ is a cone in $\mathbb{R}^n$, $\bar{\mathcal{C}}$ will denote $\mathcal{C} \cap \mathcal{B}^{n-1}$.

We will now describe how to establish a similarity between the symmetric Bernoulli and the standard normal distribution.

2. PROPORTIONAL MEAN DECOMPOSITION

Given a continuous density function f_C and a discrete distribution f_D, we propose the following partitioning of the continuous distribution in terms of the discrete one.

Definition 2.1 (Proportional mean decomposition (PMD))**.** *Let f_C and f_D be probability distributions with zero-mean and unit-variance. Suppose f_D is a discrete distribution with alphabet size of K, taking increasingly-ordered values $\{a_i\}_{i=1}^K \in \mathbb{R}$ with probabilities $\{p_i\}_{i=1}^K$ where $\sum_{i=1}^K p_i = 1$. We say $\{f_i\}_{i=1}^K$ is a proportional mean decomposition of f_C with respect to f_D with the similarity constant c_S, if $\{f_i\}_{i=1}^K$'s are probability distributions satisfying,*

- $f_C = \sum_{i=1}^K p_i f_i$.
- *mean*$(f_i) = c_S a_i$ *for all* $1 \le i \le K$.

Additionally, define $\nu_i = \sqrt{var(f_i)}$, $\nu_{\max} = \max_{1\le i\le K}\nu_i$ and $\nu_{\min} = \min_{1\le i\le K}\nu_i$.

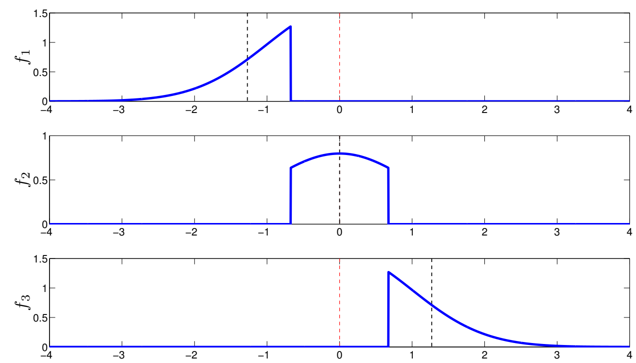

Fig. 1: PMD for the symmetric ternary distribution. In this case, $\nu_{\max}^2 = \nu_1^2 = \nu_3^2 \approx 0.242$, $\nu_2^2 \approx 0.143$ and $c_S^2 \approx 0.808$. Dashed black lines correspond to mean(f_i).

2.1. Examples

To provide a better intuition, we provide examples on PMD when $f_C \sim \mathcal{N}(0,1)$.

- Suppose f_D is symmetric Bernoulli ± 1. Let

  $$f_1(x) = \sqrt{\frac{2}{\pi}}\exp(-\frac{x^2}{2}) \text{ for } x \ge 0 \text{ and } = 0 \text{ otherwise,}$$

  $$f_2(x) = \sqrt{\frac{2}{\pi}}\exp(-\frac{x^2}{2}) \text{ for } x < 0 \text{ and } = 0 \text{ otherwise.}$$

  Then $c_S^2 = \frac{2}{\pi}$ and $\nu_{\max}^2 = \nu_1^2 = \nu_2^2 = 1 - \frac{2}{\pi}$.

- Figure 1 describes the symmetric ternary distribution $f_D = \frac{1}{4}\delta(x+\sqrt{2}) + \frac{1}{2}\delta(x) + \frac{1}{4}\delta(x-\sqrt{2})$ with maximal similarity c_S. Here $\delta(\cdot)$ is the Dirac delta function.

PMD satisfies the following properties [15].

Lemma 2.1. *Consider the setup in Definition 2.1.*

- *The set of achievable similarity constants c_S is convex and contains* 0.
- $\sum_{i=1}^K p_i \nu_i^2 = 1 - c_S^2$.

[1] RIC is a function of the matrix $\mathbf{G}$ and hence is a random variable. However, RIC of a Gaussian matrix would concentrate around its mean due to Lipschitzness of RSV; hence it is safe to say $\delta_k(\mathbf{G}) \approx \mathbb{E}[\delta_k(\mathbf{G})]$.

- *Suppose $x_D \in \mathbb{R}$ is distributed with f_D. Define x_C conditioned on x_D as follows,*

$$x_C \sim f_i \quad \textit{iff} \quad x_D = a_i \quad \textit{for} \quad 1 \le i \le K.$$

Then $x_C \sim f_C$ almost everywhere. Furthermore, $x_C - c_S x_D$ has variance $1-c_S^2$ and conditioned on x_D, $x_C - c_S x_D$ is zero-mean.

2.2. From scalar variables to i.i.d matrices

Our aim is to use PMD to obtain results on random matrices.

Definition 2.2 (Sensing matrices)**.** *Consider Definition 2.1. Let $\mathbf{D} \in \mathbb{R}^{m\times n}$ be a matrix with i.i.d entries distributed as f_D. Let $\mathbf{C}$ be a matrix satisfying,*

$$\mathbf{C}_{i,j} \sim f_k \quad \textit{if} \quad \mathbf{D}_{i,j} = a_k, \quad \forall \quad k \le K,\ i \le m,\ j \le n.$$

The following proposition provides an initial motivation for the PMD.

Proposition 2.1 (Bound in Expectation)**.** *Suppose $\mathbf{D}$ and $\mathbf{C}$ are as defined above. Then, for any nonempty and closed cone $\mathcal{C} \subset \mathbb{R}^n$*

- $\mathbb{E}[\sigma_{\mathcal{C}}^2(\mathbf{D})] \ge \frac{1}{c_S^2}\left(\mathbb{E}[\sigma_{\mathcal{C}}^2(\mathbf{C})] - \nu_{\max}^2 m\right)$,
- $\mathbb{E}[\Sigma_{\mathcal{C}}^2(\mathbf{D})] \le \frac{1}{c_S^2}\left(\mathbb{E}[\Sigma_{\mathcal{C}}^2(\mathbf{C})] - \nu_{\min}^2 m\right)$.

Proof. To prove the first statement, given $\mathbf{D}$ and $\mathcal{C}$, let

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}\in\mathcal{C}\cap\mathcal{S}^{n-1}} \|\mathbf{D}\mathbf{v}\|_2,$$

where $\mathcal{S}^{n-1}$ is the unit ℓ_2 sphere. Conditioned on $\mathbf{D}$, $\hat{\mathbf{v}}$ is fixed and $\mathbf{C} - c_S\mathbf{D}$ has independent, zero-mean entries. Hence,

$$\mathbb{E}_{\mathbf{C}|\mathbf{D}}[\|\mathbf{C}\hat{\mathbf{v}}\|_2^2] = \|c_S\mathbf{D}\hat{\mathbf{v}}\|_2^2 + \mathbb{E}_{\mathbf{C}|\mathbf{D}}[\|(\mathbf{C}-c_S\mathbf{D})\hat{\mathbf{v}}\|_2^2]. \quad (2.1)$$

Since $\hat{\mathbf{v}}$ has unit length and the entries of $\mathbf{C} - c_S\mathbf{D}$ has variance at most $\nu_{\max}^2$, $\mathbb{E}_{\mathbf{C}|\mathbf{D}}[\|(\mathbf{C}-c_S\mathbf{D})\hat{\mathbf{v}}\|_2^2] \le \nu_{\max}^2 m$. Hence, taking the expectation over $\mathbf{D}$, we find,

$$\mathbb{E}[\sigma_{\mathcal{C}}^2(\mathbf{C})] \le \mathbb{E}[\|\mathbf{C}\hat{\mathbf{v}}\|_2^2] \le c_S^2\,\mathbb{E}[\sigma_{\mathcal{C}}^2(\mathbf{D})] + \nu_{\max}^2 m. \quad (2.2)$$

To prove the second statement, let $\hat{\mathbf{v}} = \arg\max_{\mathbf{v}\in\mathcal{C}\cap\mathcal{S}^{n-1}} \|\mathbf{D}\mathbf{v}\|_2$ and observe that $\mathbb{E}_{\mathbf{C}|\mathbf{D}}[\|(\mathbf{C}-c_S\mathbf{D})\hat{\mathbf{v}}\|_2^2] \ge \nu_{\min}^2 m$. Now, we again use (2.1) and replace (2.2) with $\mathbb{E}[\Sigma_{\mathcal{C}}^2(\mathbf{C})] \ge \mathbb{E}[\|\mathbf{C}\hat{\mathbf{v}}\|_2^2] \ge c_S^2\,\mathbb{E}[\Sigma_{\mathcal{C}}^2(\mathbf{D})] + \nu_{\min}^2 m$. $\square$

2.3. Proof of Proposition 1.1

We are in a position to prove Proposition 1.1; which is essentially a corollary of Proposition 2.1. For $f_C \sim \mathcal{N}(0,1)$ and f_D is symmetric Bernoulli, we have $c_S^2 = \frac{2}{\pi}$, $\nu_{\max}^2 = \nu_{\min}^2 = 1 - \frac{2}{\pi}$. Hence, if $\mathbb{E}[\sigma_{\mathcal{C}}^2(\mathbf{C})] \ge (1-\varepsilon_1)m$,

$$\mathbb{E}[\sigma_{\mathcal{C}}^2(\mathbf{D})] \ge \frac{(1-\varepsilon_1)m - (1-\frac{2}{\pi})m}{\frac{2}{\pi}} = (1 - \frac{\pi}{2}\varepsilon_1)m.$$

Similarly, using $\mathbb{E}[\Sigma_{\mathcal{C}}^2(\mathbf{C})] \le (1+\varepsilon_2)m$,

$$\mathbb{E}[\Sigma_{\mathcal{C}}^2(\mathbf{D})] \le \frac{(1+\varepsilon_2)m - (1-\frac{2}{\pi})m}{\frac{2}{\pi}} = (1 + \frac{\pi}{2}\varepsilon_2)m.$$

Remark: Proposition 2.1 considers the crude bounds involving $\nu_{\min}^2$ and $\nu_{\max}^2$ in the statements. In fact, one can always replace them with $1 - c_S^2$ by moving from a deterministic statement to a probabilistic one (conditioned on $\mathbf{D}$). This can be done by arguing that, with high probability (for sufficiently large m), each a_i occurs at most $(1+\varepsilon')mp_i$ times at each column of $\mathbf{D}$. For such $\mathbf{D}$'s, the expected energy of each column of $\mathbf{C} - c_S\mathbf{D}$ can be upper bounded by $(1+\varepsilon')m(1-c_S^2)$. One can similarly obtain lower bounds on the column lengths with $(1-\varepsilon')$ multiplicity and then repeat the argument in Proposition 2.1 to get results that hold with high probability over $\mathbf{D}$. Focusing on Rademacher's shorten our discussion and we end up with cleaner results.

3. ON THE SAMPLE COMPLEXITY OF BERNOULLI ENSEMBLE

We will obtain results for Bernoulli matrices by using Gaussian Min-Max Theorem due to Gordon [16]. We first state a standard result that relates RSV to the sharp recovery conditions of (1.1).

Proposition 3.1. *Let f be a continuous and convex function and $\mathcal{T}_f(\mathbf{x}_0)$ be the tangent cone of f at $\mathbf{x}_0$, i.e., the closure of the set $\{\alpha\mathbf{v} \mid f(\mathbf{x}_0+\mathbf{v}) \le f(\mathbf{x}_0),\ \alpha \ge 0\}$. $\mathbf{x}_0$ is the unique solution to (1.1) if $\sigma_{\mathcal{T}_f(\mathbf{x}_0)}(\mathbf{A}) > 0$.*

The next proposition is essentially identical to Lemma 3.1 of [16] and can be derived from the minimax inequality for the Gaussian processes.

Proposition 3.2 (Gordon's inequality, [4])**.** *Let $\mathbf{G} \in \mathbb{R}^{m\times n}$, $\mathbf{g} \in \mathbb{R}^m, \mathbf{h} \in \mathbb{R}^n$ have i.i.d. standard standard normal entries. Let $\mathcal{C} \subset \mathbb{R}^n$ be a nonempty and closed cone. Then*

$$\mathbb{E}[\min_{\mathbf{x}\in\bar{\mathcal{C}}} \|\mathbf{G}\mathbf{x}\|] \ge \sqrt{m-1} - \omega(\bar{\mathcal{C}}),$$

where $\bar{\mathcal{C}} = \mathcal{C} \cap \mathcal{B}^{n-1}$ and $\omega(\cdot)$ returns the Gaussian width of a set: $\omega(S) = \mathbb{E}_{\mathbf{g}\sim\mathcal{N}(0,\mathbf{I}_n)}[\sup_{\mathbf{v}\in S}\langle\mathbf{v},\mathbf{g}\rangle]$.

In [4], Chandrasekaran et al. used this to conclude that $m^* = (1+o(1))\omega(\bar{\mathcal{T}}_f(\mathbf{x}_0))^2$ Gaussian samples are sufficient for the success of (1.1). More recently, [3] showed the tightness of bound, i.e. if $m \le (1-o(1))\omega(\bar{\mathcal{T}}_f(\mathbf{x}_0))^2$ (1.1) fails. Furthermore, as $\sqrt{m} - \omega(\bar{\mathcal{T}}_f(\mathbf{x}_0))$ grows, lasso variation of (1.1) becomes robuster to noise [4, 5]. Our next result provides a lower bound on the RSV of Bernoulli matrices in terms of $\omega(\bar{\mathcal{C}})$.

Theorem 3.1 (Bounds for Bernoulli RSV). *Let $\mathcal{C} \subset \mathbb{R}^n$ be a nonempty and closed cone. Suppose $\mathbf{B} \in \mathbb{R}^{m\times n}$ is a matrix with i.i.d symmetric Bernoulli entries. Then, whenever*

$$\sqrt{m} \geq \frac{2.6}{1-\sqrt{\varepsilon}}[\omega(\bar{\mathcal{C}})+1],$$

we have that $\mathbb{E}[\sigma^2_{\mathcal{C}}(\mathbf{B})] \geq \varepsilon m$. Furthermore, for $\varepsilon < 0.99$, $\sigma^2_{\mathcal{C}}(\mathbf{B}) \geq \varepsilon m$ with probability $1-\exp(-\frac{t^2}{2})$ if

$$\sqrt{m} \geq \frac{2.6}{1-\sqrt{\varepsilon}}(\omega(\bar{\mathcal{C}})+t+4).$$

Proof. The proof is based on combining the Gordon's inequality with Proposition 2.1. From Proposition 3.2, we have that

$$\mathbb{E}[\sigma^2_{\mathcal{C}}(\mathbf{G})] \geq \mathbb{E}[\sigma_{\mathcal{C}}(\mathbf{G})]^2 \geq (\gamma_m - \omega(\bar{\mathcal{C}}))^2.$$

From Proposition 2.1, we have $[\mathbb{E}[\sigma^2_{\mathcal{C}}(\mathbf{B})]] \geq \varepsilon m$, whenever

$$\mathbb{E}[\sigma^2_{\mathcal{C}}(\mathbf{G})] \geq \frac{2}{\pi}\varepsilon m + (1-\frac{2}{\pi})m$$

Hence, we need to ensure

$$(\sqrt{m-1} - \omega(\bar{\mathcal{C}}))^2 \geq \frac{2}{\pi}\varepsilon m + (1-\frac{2}{\pi})m.$$

This is further implied by $\sqrt{m} \geq \frac{\omega(\bar{\mathcal{C}})+1}{1-\sqrt{1-\frac{2}{\pi}(1-\varepsilon)}}$. We now use the fact that $\frac{1-\sqrt{1-\frac{2}{\pi}(1-\varepsilon)}}{1-\sqrt{\varepsilon}}$ is non-increasing as a function of ε which can be seen by differentiation. Setting $\varepsilon = 0$ returns 2.6^{-1}. The proof of second statement requires extra effort; however follows similar lines. □

Theorem 3.1 compares well with the sample complexity of the Gaussian ensemble. In particular, instead of $\omega(\bar{\mathcal{T}}_f(\mathbf{x}_0))^2$, for a general discrete distribution, our bound on the sample complexity takes the form $(\frac{1}{1-\sqrt{1-c_S^2}})^2\omega(\bar{\mathcal{T}}_f(\mathbf{x}_0))^2$ [15]. For Bernoulli matrices, substituting $c_S = \sqrt{2/\pi}$ we find that $7\cdot\omega(\bar{\mathcal{T}}_f(\mathbf{x}_0))^2$ samples are sufficient for robust recovery ($\varepsilon > 0$).

4. FURTHER EXTENSIONS

4.1. Application to sparse corruption

The usefulness of the PMD is not limited to the bounds on sample complexity and noise robustness levels. For instance, it is possible that, the original measurements $\mathbf{A}\mathbf{x}_0$ are corrupted by a structured sparse signal $\mathbf{s}_0$ [6, 8, 17] so that $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{s}_0$. To address this within our framework, consider the constrained infimal deconvolution

$$(\hat{\mathbf{x}}, \hat{\mathbf{s}}) = \arg\min_{(\mathbf{x},\mathbf{s})} g(\mathbf{s}) \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}+\mathbf{s},\ f(\mathbf{x}) \leq f(\mathbf{x}_0). \tag{4.1}$$

Here the convex function g encourages the structure of $\mathbf{s}_0$ and is usually the ℓ_1 norm. The optimality condition for (4.1) is connected to a generalization of the restricted singular value, where we consider $\sigma_{\mathcal{C},\mathcal{M}}(\mathbf{A}) = \min_{\mathbf{v}\in\mathcal{C}\cap\mathcal{S}^{n-1}} \|\Pi_{\mathcal{M}}(\mathbf{A}\mathbf{v})\|_2$ and $\Pi_{\mathcal{M}}(\cdot)$ is the projection operator to the additional cone $\mathcal{M}$. For this variation, we have the following result.

Theorem 4.1. *Suppose that $m \geq (2.6\cdot\omega(\bar{\mathcal{T}}_f(\mathbf{x}_0)) + t + 4)^2 + 2.6\cdot\omega(\bar{\mathcal{T}}_g(\mathbf{s}_0))^2$. Then, $(\mathbf{x}_0, \mathbf{s}_0)$ is the unique minimizer of (4.1) with probability $1-\exp(-\frac{t^2}{2})$.*

On the other hand, the Gaussian bound requires $\omega(\bar{\mathcal{T}}_f(\mathbf{x}_0))^2 + \omega(\bar{\mathcal{T}}_g(\mathbf{s}_0))^2$ samples [6, 8]. Hence, we pay an extra factor of 7 for the signal $\mathbf{x}_0$ and a factor of 2.6 for the corruption $\mathbf{s}_0$.

4.2. Bounds on the similarity constant

While the results are stated for Rademacher variables which has $c_S = \sqrt{2/\pi}$, not surprisingly, as the similarity to $\mathcal{N}(0,1)$ increases ($c_S \to 1$), our bound on the sample complexity will approach $\omega(\bar{\mathcal{T}}_f(\mathbf{x}_0))^2$. Given the distributions f_C and f_D, calculation of c_S can be cast as an optimization problem, however, it is desirable to have an interpretable upper bound on it. The following result gives such an upper bound in terms of the tails of the distributions.

Proposition 4.1. *Suppose f_C and f_D are distributions that are symmetric around 0 and let $x_C \sim f_C$, $x_D \sim f_D$. Then*

$$c_S \geq \inf_{i\geq\lceil\frac{K}{2}\rceil} \frac{\mathbb{P}(x_C \geq a_i)}{\mathbb{P}(x_D \geq a_i)}$$

To prove this, we inductively construct the functions $\{f_i\}_{i=1}^K$ obeying the Definition 2.1. This bound provides a connection between this work and the results of [11–13].

5. CONCLUDING REMARKS

We have introduced the "proportional mean decomposition" as a way of capturing the similarity of one distribution to another and discussed how it can be useful in compressed sensing, especially when the measurement matrix has i.i.d Bernoulli entries. While we are able to obtain small explicit constants in Proposition 1.1 and Theorem 3.1, our basic approach fails to capture the universality phenomenon, which is the common belief that, the sample complexities of various measurement ensembles are asymptotically equal. This remains as an important open question, which is partially answered by Bayati et al. in the case of ℓ_1 minimization [18]. It would be also interesting to see whether the approach presented here can be extended to handle non i.i.d. ensembles such as the subsampled random Fourier transform [9, 19].

6. ACKNOWLEDGMENTS

We thank the anonymous reviewers for their helpful comments and suggestions.

7. REFERENCES

[1] David L Donoho and Jared Tanner, "Neighborliness of randomly projected simplices in high dimensions," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 27, pp. 9452–9457, 2005.

[2] David L Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete & Computational Geometry*, vol. 35, no. 4, pp. 617–652, 2006.

[3] Dennis Amelunxen, Martin Lotz, Michael B McCoy, and Joel A Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Inform. Inference*, 2014.

[4] Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[5] Samet Oymak, Christos Thrampoulidis, and Babak Hassibi, "The squared-error of generalized lasso: A precise analysis," *arXiv preprint arXiv:1311.0830*, 2013.

[6] Michael B McCoy and Joel A Tropp, "Sharp recovery bounds for convex deconvolution, with applications," *arXiv preprint arXiv:1205.1580*, 2012.

[7] Mohsen Bayati and Andrea Montanari, "The lasso risk for gaussian matrices," *Information Theory, IEEE Transactions on*, vol. 58, no. 4, pp. 1997–2017, 2012.

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

[9] Holger Rauhut, "Compressive sensing and structured random matrices," *Theoretical foundations and numerical methods for sparse recovery*, vol. 9, pp. 1–92, 2010.

[10] Gesen Zhang, Shuhong Jiao, Xiaoli Xu, and Lan Wang, "Compressed sensing and reconstruction with bernoulli matrices," in *Information and Automation (ICIA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 455–460.

[11] Joel A Tropp, "Convex recovery of a structured signal from independent random linear measurements," *arXiv preprint arXiv:1405.1102*, 2014.

[12] Shahar Mendelson, "Learning without concentration," *arXiv preprint arXiv:1401.0304*, 2014.

[13] Vladimir Koltchinskii and Shahar Mendelson, "Bounding the smallest singular value of a random matrix without concentration," *arXiv preprint arXiv:1312.3580*, 2013.

[14] Yaniv Plan and Roman Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach," *Information Theory, IEEE Transactions on*, vol. 59, no. 1, pp. 482–494, 2013.

[15] Samet Oymak and Babak Hassibi, "The proportional mean decomposition: a bridge between the gaussian and bernoulli ensembles," *technical report available at* www.its.caltech.edu/~soymak/PMD.pdf, 2014.

[16] Yehoram Gordon, *On Milman's inequality and random subspaces which escape through a mesh in* $\mathbb{R}^n$, Springer, 1988.

[17] Xiaodong Li, "Compressed sensing and matrix completion with constant proportion of corruptions," *Constructive Approximation*, vol. 37, no. 1, pp. 73–99, 2013.

[18] Mohsen Bayati, Marc Lelarge, and Andrea Montanari, "Universality in polytope phase transitions and message passing algorithms," *arXiv preprint arXiv:1207.7321*, 2012.

[19] Emmanuel J Candès, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.