

# Stable Recovery of Structured Signals From Corrupted Sub-Gaussian Measurements

Jinchi Chen\* and Yulong Liu\*

\*School of Physics, Beijing Institute of Technology, Beijing 100081, China

## Abstract

This paper studies the problem of accurately recovering a structured signal from a small number of corrupted sub-Gaussian measurements. We consider three different procedures to reconstruct signal and corruption when different kinds of prior knowledge are available. In each case, we provide conditions (in terms of the number of measurements) for stable signal recovery from structured corruption with added unstructured noise. Our results theoretically demonstrate how to choose the regularization parameters in both partially and fully penalized recovery procedures and shed some light on the relationships among the three procedures. The key ingredient in our analysis is an extended matrix deviation inequality for isotropic sub-Gaussian matrices, which implies a tight lower bound for the restricted singular value of the extended sensing matrix. Numerical experiments are presented to verify our theoretical results.

## Index Terms

Corrupted sensing, compressed sensing, signal separation, signal demixing, sub-Gaussian, Gaussian width, Gaussian complexity, Gaussian squared distance, extended matrix deviation inequality.

## I. INTRODUCTION

Corrupted sensing concerns the problem of recovering a structured signal from a relatively small number of corrupted measurements

$$\mathbf{y} = \Phi \mathbf{x}^* + \mathbf{v}^* + \mathbf{z}, \quad (1)$$

where  $\Phi \in \mathbb{R}^{m \times n}$  is the sensing matrix,  $\mathbf{x}^* \in \mathbb{R}^n$  is the structured signal,  $\mathbf{v}^* \in \mathbb{R}^m$  is the structured corruption, and  $\mathbf{z} \in \mathbb{R}^m$  is the unstructured observation noise. The goal is to estimate  $\mathbf{x}^*$  and  $\mathbf{v}^*$  from given knowledge of  $\mathbf{y}$  and  $\Phi$ .

This problem has received increasing attention recently with many interesting practical applications as well as theoretical consideration. Examples of applications include face recognition [2], subspace clustering [3], sensor network [4], latent variable modeling [5], and so on. On the theoretical side, performance guarantees include sparse signal recovery from sparse corruption [6], [7], [8], [9], [10], [11], [12], [13], low-rank matrix recovery from sparse corruption [5], [14], [15], [16], and structured signal recovery from structured corruption [17]. It is worth noting that this model (1) also includes the signal separation (or demixing) problem [18] in which  $\mathbf{v}^*$  might actually contain useful information and thus be necessary to be recovered. In particular, if there is no corruption ( $\mathbf{v}^* = \mathbf{0}$ ), this model (1) reduces to the standard compressed sensing problem.

Since this problem is generally ill-posed, tractable recovery is possible when both signal and corruption are suitably structured. Typical examples of structured signals (or corruptions) include sparse vectors, which exhibit low complexity under the  $\ell_1$  norm, and low-rank matrices which exhibit low complexity with respect to the nuclear norm (more examples can be found in [19]). Let  $f(\cdot)$  and  $g(\cdot)$  be suitable norms which promote structures for signal and corruption respectively. We consider three different convex optimization approaches to disentangle signal and corruption when different kinds of prior information are available. Specifically, when prior knowledge of either signal  $f(\mathbf{x}^*)$  or corruption  $g(\mathbf{v}^*)$  is available and the noise level  $\delta$  (in terms of the  $\ell_2$  norm) is known, it is natural to consider the following constrained convex recovery procedures

$$\min_{\mathbf{x}, \mathbf{v}} f(\mathbf{x}), \quad \text{s.t. } g(\mathbf{v}) \leq g(\mathbf{v}^*), \quad \|\mathbf{y} - \Phi \mathbf{x} - \mathbf{v}\|_2 \leq \delta \quad (2)$$

and

$$\min_{\mathbf{x}, \mathbf{v}} g(\mathbf{v}), \quad \text{s.t. } f(\mathbf{x}) \leq f(\mathbf{x}^*), \quad \|\mathbf{y} - \Phi \mathbf{x} - \mathbf{v}\|_2 \leq \delta. \quad (3)$$

When only the noise level  $\delta$  is known, it is convenient to use the partially penalized convex recovery procedure

$$\min_{\mathbf{x}, \mathbf{v}} f(\mathbf{x}) + \lambda \cdot g(\mathbf{v}), \quad \text{s.t. } \|\mathbf{y} - \Phi \mathbf{x} - \mathbf{v}\|_2 \leq \delta, \quad (4)$$

where  $\lambda > 0$  is a regularization parameter. When there is no prior knowledge available, it is practical to utilize the fully penalized convex recovery procedure

$$\min_{\mathbf{x}, \mathbf{v}} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x} - \mathbf{v}\|_2^2 + \tau_1 \cdot f(\mathbf{x}) + \tau_2 \cdot g(\mathbf{v}), \quad (5)$$

where  $\tau_1, \tau_2 > 0$  are some regularization parameters.

### A. Contribution

This paper considers the problem of recovering structured signals from corrupted sub-Gaussian measurements. More precisely, we consider the following observation model:

- Sub-Gaussian measurements: we assume that each row  $\Phi_i$  of the sensing matrix  $\Phi$  is independent, centered, and sub-Gaussian random vector with

$$\|\Phi_i\|_{\psi_2} \leq K/\sqrt{m} \quad \text{and} \quad \mathbb{E} \Phi_i^T \Phi_i = \mathbf{I}_n/m, \quad (6)$$

where  $\|\cdot\|_{\psi_2}$  denotes the sub-Gaussian norm (defined in (9)) and  $\mathbf{I}_n$  is the  $n$ -dimensional identity matrix;

- Bounded or sub-Gaussian noise: the unstructured noise  $\mathbf{z}$  is assumed to be bounded ( $\|\mathbf{z}\|_2 \leq \delta$ ) or be a random vector with independent centered sub-Gaussian entries satisfying

$$\|\mathbf{z}_i\|_{\psi_2} \leq L \quad \text{and} \quad \mathbb{E} \mathbf{z}_i^2 = 1. \quad (7)$$

Under the above model assumptions, the contribution of this paper is twofold:

- First, we establish an extended matrix deviation inequality for isotropic sub-Gaussian matrices, which provides a powerful tool for analyzing the corrupted sensing problem;
- Second, based on the deviation inequality, we present performance guarantees for all three convex recovery procedures. These results theoretically demonstrate how to select the regularization parameters in both partially and fully penalized recovery procedures and shed some light on the relationships among the three approaches.

### B. Related Work

Within the past few years, there have been numerous studies to understand the theoretical aspect of the corrupted sensing problem. These works might be roughly put into the following three groups by the type of structures considered.

1) *Sparse Signal Recovery from Sparse Corruption*: The specific problem of recovering a sparse vector from sparsely corrupted measurements has been analyzed under a variety of different model assumptions. For instance, Wright and Ma [6] study exact recovery by the partially penalized recovery procedure (with  $f(\cdot) = g(\cdot) = \|\cdot\|_1$ ,  $\lambda = 1$ , and  $\delta = 0$ ) under a bouquet model:  $\mathbf{y} = \Phi \mathbf{x}^* + \mathbf{v}^*$ , where the columns of  $\Phi$  are independent identically distributed (i.i.d.) samples from a Gaussian distribution  $\mathcal{N}(\boldsymbol{\mu}, \nu^2 \mathbf{I}_m/m)$  with  $\|\boldsymbol{\mu}\|_2 = 1$  and  $\|\boldsymbol{\mu}\|_\infty \leq C/\sqrt{m}$ , and  $\mathbf{v}^*$  has uniformly random signs. Li [7] considers stable recovery of  $(\mathbf{x}^*, \mathbf{v}^*)$  from  $\mathbf{y} = \Phi \mathbf{x}^* + \mathbf{v}^* + \mathbf{z}$  by the partially penalized recovery procedure (with  $f(\cdot) = g(\cdot) = \|\cdot\|_1$  and  $\lambda = 1/\sqrt{\log(n/m) + 1}$ ), where  $\Phi$  is a Gaussian matrix with i.i.d. entries and  $\mathbf{z}$  is a bounded noise vector. In the absence of noise, [7] also considers a class of sensing matrices, in which the rows of  $\Phi$  are sampled independently from a population obeying  $\mathbb{E} \Phi_i^T \Phi_i = \mathbf{I}_n$  and  $\|\Phi_i\|_\infty \leq c$ . In contrast to [7], Nguyen and Tran [8] also establish stable or exact recovery of  $(\mathbf{x}^*, \mathbf{v}^*)$  from  $\mathbf{y} = \Phi \mathbf{x}^* + \mathbf{v}^* + \mathbf{z}$  by the partially penalized recovery procedure, but with a different model assumption in which  $\Phi$  has columns sampled uniformly from an orthonormal matrix,  $\mathbf{x}^*$  has uniformly random signs,  $\mathbf{v}^*$  has uniformly distributed support, and  $\mathbf{z}$  is a bounded noise vector. In a subsequent paper [9], Nguyen and Tran study stable recovery of  $(\mathbf{x}^*, \mathbf{v}^*)$  from  $\mathbf{y} = \Phi \mathbf{x}^* + \mathbf{v}^* + \mathbf{z}$  by the fully penalized recovery procedure (with  $f(\cdot) = g(\cdot) = \|\cdot\|_1$ ), where  $\Phi$  is an Gaussian sensing matrix with i.i.d. entries and  $\mathbf{z}$  is a Gaussian noise vector. Pope *et al.* [12] investigate the exact recovery from the observation model  $\mathbf{y} = \Phi \mathbf{x}^* + \Psi \mathbf{v}^*$ , where  $\Phi$  and  $\Psi$  are deterministic matrices. The authors require uniform randomness in the support set of  $\mathbf{x}^*$  or  $\mathbf{v}^*$  and rely on certain incoherence properties of  $\Phi$  and  $\Psi$ . These probabilistic recovery guarantees improve or refine the previous results in [10], [11], [13].

2) *Low-rank Matrix Recovery from Sparse Corruption*: Another line of work considers the problem of recovering a low-rank matrix from sparse corruption. Chandrasekaran *et al.* [5] study the exact decomposition of a given matrix into its sparse and low-rank components by the partially penalized recovery procedure (with  $f(\cdot) = \|\cdot\|_*$ ,  $g(\cdot) = \|\cdot\|_1$  and  $\delta = 0$ ), where  $\|\mathbf{D}\|_*$  denotes the nuclear norm of  $\mathbf{D}$  and  $\|\mathbf{D}\|_1 = \sum_{i,j} |\mathbf{D}_{i,j}|$ . The work of Candès *et al.* [14] uses this model for robust principal component analysis and image processing applications. Modifications to the rank-sparsity model also find applications in robust statistics, see e.g., [15], [16].

3) *Structured Signal Recovery from Structured Corruption*: The most closely related to our current paper are the results by [18], [17], [20], [21], in which the problem of recovering a structured signal from structured corruption is considered. In [18], McCoy and Tropp analyze the constrained recovery procedures ((2) and (3)) in the setting where  $\Phi$  is a random orthogonal matrix ( $m = n$ ) and measurements are noiseless ( $\delta = 0$ ). In the same setting, the work of Amelunxen *et al.* [20] establishes the sharpness of their recovery results, by specifying an appropriate threshold under which the constrained procedures fail with high probability. Foygel and Mackey [17] study both the constrained recovery procedures ((2) and (3)) and the partially penalized recovery procedure (4) in the setting where  $\Phi$  is a Gaussian matrix ( $m \leq n$  or  $m > n$ ) with i.i.d. entries and the noise is bounded. A recent work of Zhang *et al.* [21] establishes the phase transition theory of the constrained recovery procedures under the setting of [17]. In the present paper, we establish performance guarantees for all three convex recovery procedures in the setting where  $\Phi$  is a sub-Gaussian matrix and the noise is bounded or sub-Gaussian. These results solve a series of open problems in [17], for example, allowing non-Gaussian sensing matrix and stochastic unstructured noise in observation model (1) and analyzing the fully penalized convex recovery procedure (5) for arbitrary structured signals and structured corruptions.

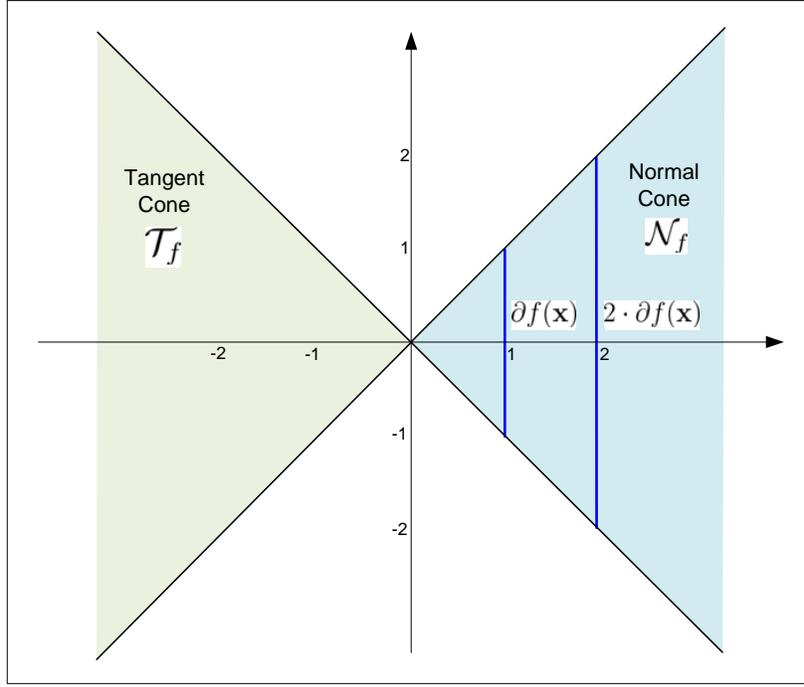


Fig. 1. Illustrations of the subdifferential, scaled subdifferential, normal cone, and tangent cone for  $f(\mathbf{x}) = \|\mathbf{x}\|_1$  at the point  $\mathbf{x} = (1, 0)$ . (For clarity, cones are shifted to the original point.)

### C. Organization

The remainder of the paper is organized as follows. We begin by reviewing some preliminaries that are necessary for our subsequent analysis in Section II. In Section III, we establish a mathematical tool which is critical for analyzing the corrupted sensing problem. Section IV is devoting to presenting the main theoretical results of the paper. We then present results from numerical simulations in Section V. We conclude the paper in Section VI.

## II. PRELIMINARIES

In this section, we review some useful concepts from convex geometry and high-dimensional probability that underlie our analysis. Throughout the paper,  $\mathbb{S}^{n-1}$  and  $\mathbb{B}_2^n$  denote the unit sphere and ball in  $\mathbb{R}^n$  under the  $\ell_2$  norm respectively, while  $\mathbb{B}_f^n := \{\mathbf{u} \in \mathbb{R}^n : f(\mathbf{u}) \leq 1\}$  is the unit ball in  $\mathbb{R}^n$  under the norm defined by  $f$ . The compatibility constant between  $f$  and the  $\ell_2$  norm is defined as  $\alpha_f := \sup_{\mathbf{u} \neq 0} f(\mathbf{u})/\|\mathbf{u}\|_2$ . We use the notation  $C, C', c, c', \dots$ , to refer to positive constants, whose value may change from line to line.

### A. Convex Geometry

1) *Subdifferential*: The *subdifferential* of  $f$  at  $\mathbf{x}$  is the set of vectors

$$\partial f(\mathbf{x}) = \{\mathbf{u} \in \mathbb{R}^n : f(\mathbf{x} + \mathbf{d}) \geq f(\mathbf{x}) + \langle \mathbf{u}, \mathbf{d} \rangle \text{ for all } \mathbf{d} \in \mathbb{R}^n\}.$$

For any number  $\kappa \geq 0$ , we denote the scaled (by  $\kappa$ ) subdifferential as  $\kappa \cdot \partial f(\mathbf{x}) = \{\kappa \mathbf{u} : \mathbf{u} \in \partial f(\mathbf{x})\}$ .

2) *Tangent Cone and Normal Cone*: A cone is a set that is closed under multiplication by positive scalars. The *tangent cone* of  $f$  at  $\mathbf{x}$  is defined as the set of descent directions of  $f$  at  $\mathbf{x}$

$$\mathcal{T}_f = \{\mathbf{u} \in \mathbb{R}^n : f(\mathbf{x} + t \cdot \mathbf{u}) \leq f(\mathbf{x}) \text{ for some } t > 0\}.$$

The *normal cone* of  $f$  at  $\mathbf{x}$  is the polar of the tangent cone, given by

$$\mathcal{N}_f = \{\mathbf{u} \in \mathbb{R}^n : \langle \mathbf{u}, \mathbf{d} \rangle \leq 0 \text{ for all } \mathbf{d} \in \mathcal{T}_f\},$$

which may be written as the cone hull of the subdifferential (if  $0 \notin \partial f(\mathbf{x})$ ) [22, Theorem 1.3.5]

$$\mathcal{N}_f = \text{cone}\{\partial f(\mathbf{x})\} = \{\mathbf{u} \in \mathbb{R}^n : \mathbf{u} \in t \cdot \partial f(\mathbf{x}) \text{ for some } t > 0\}.$$

Fig. 1 illustrates related concepts in  $\mathbb{R}^2$  for the  $\ell_1$  norm.

3) *Gaussian Complexity, Gaussian Width, and Gaussian Distance*: For any  $\mathcal{T} \subseteq \mathbb{R}^n$ , a simple way to quantify the ‘‘size’’ of  $\mathcal{T}$  is through the *Gaussian complexity*

$$\gamma(\mathcal{T}) := \mathbb{E} \sup_{\mathbf{x} \in \mathcal{T}} |\langle \mathbf{g}, \mathbf{x} \rangle|, \quad \text{where } \mathbf{g} \sim \mathcal{N}(0, \mathbf{I}_n).$$

Another popular geometric quantity closely related to Gaussian complexity is the *Gaussian width*

$$\omega(\mathcal{T}) := \mathbb{E} \sup_{\mathbf{x} \in \mathcal{T}} \langle \mathbf{g}, \mathbf{x} \rangle, \quad \text{where } \mathbf{g} \sim \mathcal{N}(0, \mathbf{I}_n).$$

In particular<sup>1</sup> [23],

$$(\omega(\mathcal{T}) + \|\mathbf{y}\|_2) / 3 \leq \gamma(\mathcal{T}) \leq 2\omega(\mathcal{T}) + \|\mathbf{y}\|_2 \quad \text{for every } \mathbf{y} \in \mathcal{T}. \quad (8)$$

The *Gaussian squared distance*  $\eta^2(\mathcal{T})$  of a subset  $\mathcal{T} \subset \mathbb{R}^n$  is defined as

$$\eta^2(\mathcal{T}) := \mathbb{E} \inf_{\mathbf{u} \in \mathcal{T}} \|\mathbf{g} - \mathbf{u}\|_2^2, \quad \text{where } \mathbf{g} \sim \mathcal{N}(0, \mathbf{I}_n).$$

## B. High-Dimensional Probability

1) *Sub-Gaussian and Sub-exponential Random Variables*: A random variable  $X$  is sub-Gaussian if its distribution is dominated by a Gaussian distribution. Formally,  $X$  is called a *sub-Gaussian* random variable if the Orlicz norm

$$\|X\|_{\psi_2} = \inf\{K > 0 : \mathbb{E} \psi_2(|X|/K) \leq 1\} \quad (9)$$

is finite for the Orlicz function  $\psi_2(x) = \exp(x^2) - 1$ . The *sub-Gaussian norm* of  $X$ , denoted  $\|X\|_{\psi_2}$ , is defined to be the smallest  $K$  in (9). There are several equivalent definitions used in the literature. In particular,  $X$  is sub-Gaussian if there exist absolute constants  $C > 1$  and  $K$  such that

$$[\mathbb{E} |X|^p]^{1/p} \leq K\sqrt{p}, \quad \text{for all } p \geq 1 \quad (10)$$

or

$$\mathbb{P}\{|X| \geq t\} \leq C \exp\{-t^2/K^2\}, \quad \text{for all } t \geq 0. \quad (11)$$

Thus we can redefine the sub-Gaussian norm of  $X$  as the smallest  $K$  such that (10) or (11) holds. One can show that  $\|X\|_{\psi_2}$  defined this way is within an absolute constant factor from that defined in (9), see [24, Section 5.2.3]. Classical examples of sub-Gaussian random variables include Gaussian, Bernoulli, and all bounded random variables. Similarly, a random variable  $X$  is sub-exponential if its distribution is dominated by an exponential distribution. More precisely,  $X$  is called a *sub-exponential* random variable if the Orlicz norm

$$\|X\|_{\psi_1} = \inf\{K > 0 : \mathbb{E} \psi_1(|X|/K) \leq 1\} \quad (12)$$

is finite for the Orlicz function  $\psi_1(x) = \exp(x) - 1$ . The *sub-exponential norm* of  $X$ , denoted  $\|X\|_{\psi_1}$ , is defined to be the smallest  $K$  in (12).

2) *Sub-Gaussian and Sub-exponential Random Vectors*: A random vector  $\mathbf{x}$  in  $\mathbb{R}^n$  is sub-Gaussian (or sub-exponential) if all of its one-dimensional marginals are sub-Gaussian (or sub-exponential) random variables and its  $\psi_2$ -norm (or  $\psi_1$ -norm) is defined as

$$\|\mathbf{x}\|_{\psi_2} := \sup_{\mathbf{y} \in \mathbb{S}^{n-1}} \|\langle \mathbf{x}, \mathbf{y} \rangle\|_{\psi_2} \quad \text{or} \quad \|\mathbf{x}\|_{\psi_1} := \sup_{\mathbf{y} \in \mathbb{S}^{n-1}} \|\langle \mathbf{x}, \mathbf{y} \rangle\|_{\psi_1}. \quad (13)$$

A random vector  $\mathbf{x}$  in  $\mathbb{R}^n$  is called *isotropic* if

$$\mathbb{E}(\mathbf{x}\mathbf{x}^T) = \mathbf{I}_n.$$

For basic properties of sub-Gaussian and sub-exponential random variables (vectors), see e.g., [24], [25, Chapter 2]. Some properties which will be used in our proofs are included in Appendix A.

<sup>1</sup>Here we use a slightly sharper upper bound than that in [23]. Indeed, since  $\mathcal{T} - \mathcal{T}$  is original symmetric, then  $2\omega(\mathcal{T}) = \gamma(\mathcal{T} - \mathcal{T}) = \mathbb{E} \sup_{\mathbf{x}, \mathbf{y} \in \mathcal{T}} |\langle \mathbf{g}, \mathbf{x} - \mathbf{y} \rangle| \geq \mathbb{E} \sup_{\mathbf{x} \in \mathcal{T}} |\langle \mathbf{g}, \mathbf{x} - \mathbf{y} \rangle| \geq \mathbb{E} \sup_{\mathbf{x} \in \mathcal{T}} [|\langle \mathbf{g}, \mathbf{x} \rangle| - |\langle \mathbf{g}, \mathbf{y} \rangle|] \geq \gamma(\mathcal{T}) - (\mathbb{E} |\langle \mathbf{g}, \mathbf{y} \rangle|^2)^{\frac{1}{2}} = \gamma(\mathcal{T}) - \|\mathbf{y}\|_2, \forall \mathbf{y} \in \mathcal{T}$ . Rearranging yields the desired result.

3) *Talagrand's Majorizing Measure Theorem*: The key ingredient in the proofs of our mathematical tools is the Talagrand's Majorizing Measure Theorem which states that any sub-Gaussian process is dominated by a Gaussian process with the same (or larger) increments.

**Fact 1** (Talagrand's Majorizing Measure Theorem). *Let  $(X_u)_{u \in \mathcal{T}}$  be a random process indexed by points in a bounded set  $\mathcal{T} \subset \mathbb{R}^n$ . Assume that the process has sub-Gaussian increments, that is, there exists  $M \geq 0$  such that*

$$\|X_u - X_v\|_{\psi_2} \leq M\|u - v\|_2 \quad \text{for every } u, v \in \mathcal{T}.$$

Then

$$\mathbb{E} \sup_{u, v \in \mathcal{T}} |X_u - X_v| \leq CM\omega(\mathcal{T}). \quad (14)$$

Moreover, for any  $t \geq 0$ , the event

$$\sup_{u, v \in \mathcal{T}} |X_u - X_v| \leq CM[\omega(\mathcal{T}) + t \cdot \text{diam}(\mathcal{T})] \quad (15)$$

holds with probability at least  $1 - \exp(-t^2)$ , where  $\text{diam}(\mathcal{T}) := \sup_{\mathbf{x}, \mathbf{y} \in \mathcal{T}} \|\mathbf{x} - \mathbf{y}\|_2$  denotes the diameter of  $\mathcal{T}$ .

*Remark 1.* The expectation bound (14) of this theorem can be found e.g. in [26, Theorem 2.1.1, 2.1.5]. The high probability bound (15) can be found e.g. in [27, Theorem 3.2] or [23, Theorem 4.1].

### III. EXTENDED MATRIX DEVIATION INEQUALITY

In this section, we establish a mathematical tool which provides a unified framework for analyzing all three recovery procedures. Note that if we denote  $\mathbf{Y} = [\Phi, \mathbf{I}_m]$  and  $\mathbf{s}^* = [(\mathbf{x}^*)^T, (\mathbf{v}^*)^T]^T$ , then the observation model (1) can be reformulated as  $\mathbf{y} = \mathbf{Y}\mathbf{s}^* + \mathbf{z}$ , which is the same as the standard compressed sensing model. It is now well known that the restricted singular value of the sensing matrix  $\mathbf{Y}$  (see, e.g., [28], [19]) plays a key role in analyzing the compressed sensing problem. With this observation in mind, we establish an extended matrix deviation inequality which implies a tight lower bound for the restricted singular value of the extended sensing matrix  $[\Phi, \mathbf{I}_m]$ . This result states that for any isotropic sub-Gaussian matrix  $\mathbf{A}$  and any bounded subset  $\mathcal{T} \subseteq \mathbb{R}^n \times \mathbb{R}^m$ , the deviation of  $\|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2$  around  $\sqrt{m} \cdot \sqrt{\|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2}$  over vectors  $(\mathbf{a}, \mathbf{b}) \in \mathcal{T}$  is uniformly bounded by the Gaussian complexity of  $\mathcal{T}$ .

**Theorem 1** (Extended Matrix Deviation Inequality). *Let  $\mathbf{A}$  be an  $m \times n$  matrix whose rows  $\mathbf{A}_i$  are independent centered isotropic sub-Gaussian vectors with  $K = \max_i \|\mathbf{A}_i\|_{\psi_2}$ , and  $\mathcal{T}$  be a bounded subset of  $\mathbb{R}^n \times \mathbb{R}^m$ . Then*

$$\mathbb{E} \sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{T}} \left| \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \sqrt{m} \cdot \sqrt{\|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2} \right| \leq CK^2 \cdot \gamma(\mathcal{T}).$$

For any  $t \geq 0$ , the event

$$\sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{T}} \left| \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \sqrt{m} \cdot \sqrt{\|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2} \right| \leq CK^2[\gamma(\mathcal{T}) + t \cdot \text{rad}(\mathcal{T})]$$

holds with probability at least  $1 - \exp(-t^2)$ , where  $\text{rad}(\mathcal{T}) := \sup_{\mathbf{x} \in \mathcal{T}} \|\mathbf{x}\|_2$  denotes the radius of  $\mathcal{T}$ .

*Proof.* See Appendix B. □

*Remark 2* (Anisotropic case). By using a simple linear transform, Theorem 1 can be extended to the anisotropic case in which each row of  $\mathbf{A}$  satisfies  $\mathbb{E} \mathbf{A}_i^T \mathbf{A}_i = \Sigma$  for some invertible covariance matrix  $\Sigma$ . To see this, consider the whitened version  $\mathbf{B} = \mathbf{A}\Sigma^{-1/2}$ . Note that

$$\begin{aligned} \|\mathbf{B}_i\|_{\psi_2} &= \sup_{\mathbf{y} \in \mathbb{S}^{n-1}} \|\langle \mathbf{B}_i^T, \mathbf{y} \rangle\|_{\psi_2} = \sup_{\mathbf{y} \in \mathbb{S}^{n-1}} \left\| \left\langle \mathbf{A}_i^T, \Sigma^{-1/2} \mathbf{y} \right\rangle \right\|_{\psi_2} \\ &= \sup_{\mathbf{y} \in \mathbb{S}^{n-1}} \|\Sigma^{-1/2} \mathbf{y}\|_2 \cdot \left\| \left\langle \mathbf{A}_i^T, \Sigma^{-1/2} \mathbf{y} / \|\Sigma^{-1/2} \mathbf{y}\|_2 \right\rangle \right\|_{\psi_2} \\ &\leq \|\Sigma^{-1/2}\| \cdot K. \end{aligned}$$

Define the block diagonal matrix  $\tilde{\Sigma} = \text{diag}\{\Sigma^{1/2}, \mathbf{I}_m\}$ , then we have

$$\begin{aligned} \mathbb{E} \sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{T}} \left| \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \sqrt{m} \cdot \sqrt{\|\Sigma^{1/2} \mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2} \right| &= \mathbb{E} \sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{T}} \left| \|\mathbf{B}\Sigma^{1/2} \mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \sqrt{m} \cdot \sqrt{\|\Sigma^{1/2} \mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2} \right| \\ &= \mathbb{E} \sup_{(\mathbf{a}, \mathbf{b}) \in \tilde{\Sigma}\mathcal{T}} \left| \|\mathbf{B}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \sqrt{m} \cdot \sqrt{\|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2} \right| \\ &\leq C\|\Sigma^{-1}\| K^2 \gamma(\tilde{\Sigma}\mathcal{T}) \\ &\leq C\|\Sigma^{-1}\| \max\{\|\Sigma\|^{1/2}, 1\} K^2 \gamma(\mathcal{T}). \end{aligned}$$

The first inequality follows from Theorem 1. The last inequality holds because  $\gamma(\tilde{\Sigma}\mathcal{T}) \leq \|\tilde{\Sigma}\|\gamma(\mathcal{T}) = \max\{\|\Sigma\|^{1/2}, 1\}\gamma(\mathcal{T})$ . Similarly, we can obtain the anisotropic version for the second part of Theorem 1.

*Remark 3.* Very recently, Liaw *et al.* [23] have shown that if  $\mathbf{A}$  is an isotropic (or anisotropic) sub-Gaussian matrix, then

$$\mathbb{E} \sup_{\mathbf{a} \in \mathcal{T}} \left| \|\mathbf{A}\mathbf{a}\|_2 - \sqrt{m}\|\mathbf{a}\|_2 \right| \leq CK^2\gamma(\mathcal{T})$$

and the event

$$\sup_{\mathbf{a} \in \mathcal{T}} \left| \|\mathbf{A}\mathbf{a}\|_2 - \sqrt{m}\|\mathbf{a}\|_2 \right| \leq CK^2[\gamma(\mathcal{T}) + t \text{rad}(\mathcal{T})] \quad (16)$$

holds with probability at least  $1 - \exp(-t^2)$ . It is not hard to see that these results can not be directly applied to the corrupted sensing problem, because the covariance matrix of each row of  $[\Phi, \mathbf{I}_m]$  is singular. However,  $[\Phi, \mathbf{I}_m]$  still has independent sub-Gaussian rows. This fact allows us to generalize the analysis in [23] to the extended case. It is worth noting that there are several earlier variants of (16) proved in [29], [30], [27]. See [23] for their comparisons.

When  $\mathcal{T}$  is a subset of the unit sphere, we have the following corollary.

**Corollary 1.** *Under the assumptions of Theorem 1, for any  $t \geq 0$ , the event*

$$\sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{T} \cap \mathbb{S}^{n+m-1}} \left| \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \sqrt{m} \right| \leq CK^2[\gamma(\mathcal{T} \cap \mathbb{S}^{n+m-1}) + t] \quad (17)$$

holds with probability at least  $1 - \exp(-t^2)$ .

Corollary 1 may be specialized to bound to the restricted singular value of  $[\mathbf{A}, \sqrt{m}\mathbf{I}_m]$ . Indeed, let  $\mathcal{T} \subset \mathbb{S}^{n+m-1}$ . Then the corollary states that, with high probability (e.g.,  $1 - \exp\{-\gamma^2(\mathcal{T} \cap \mathbb{S}^{n+m-1})\}$ ),

$$\inf_{(\mathbf{a}, \mathbf{b}) \in \mathcal{T} \cap \mathbb{S}^{n+m-1}} \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 \geq \sqrt{m} - CK^2\gamma(\mathcal{T} \cap \mathbb{S}^{n+m-1}) \quad (18)$$

and

$$\sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{T} \cap \mathbb{S}^{n+m-1}} \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 \leq \sqrt{m} + CK^2\gamma(\mathcal{T} \cap \mathbb{S}^{n+m-1}) \quad (19)$$

for all  $(\mathbf{a}, \mathbf{b}) \in \mathcal{T}$ . As we will see in the next section, (18) plays a key role in analyzing the corrupted sensing problem.

#### IV. RECOVERY FROM CORRUPTED SUB-GAUSSIAN MEASUREMENTS

In this section, we present our theoretical results on the recovery of structured signals from corrupted sub-Gaussian measurements via three different convex recovery procedures.

##### A. Recovery via Constrained Optimization

We start with analyzing the constrained convex recovery procedures (2) and (3). Our first result shows that, with high probability, approximately

$$CK^4 [\omega^2(\mathcal{T}_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1}) + \omega^2(\mathcal{T}_g(\mathbf{v}^*) \cap \mathbb{S}^{m-1})] \quad (20)$$

corrupted measurements suffice to recover  $(\mathbf{x}^*, \mathbf{v}^*)$  exactly in the absence of noise and stably in the presence of noise, via either of the procedures (2) or (3).

Before stating our result, we need to define the error set

$$\mathcal{E}_1(\mathbf{x}^*, \mathbf{v}^*) := \{(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^m : f(\mathbf{x}^* + \mathbf{a}) \leq f(\mathbf{x}^*) \text{ and } g(\mathbf{v}^* + \mathbf{b}) \leq g(\mathbf{v}^*)\},$$

in which the error vector  $(\hat{\mathbf{x}} - \mathbf{x}^*, \hat{\mathbf{v}} - \mathbf{v}^*)$  lives. By the convexity of  $f$  and  $g$ ,  $\mathcal{E}_1(\mathbf{x}^*, \mathbf{v}^*)$  belongs to the following convex cone

$$\mathcal{C}_1(\mathbf{x}^*, \mathbf{v}^*) := \{(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^m : \langle \mathbf{a}, \mathbf{u} \rangle \leq 0 \text{ and } \langle \mathbf{b}, \mathbf{s} \rangle \leq 0 \text{ for any } \mathbf{u} \in \partial f(\mathbf{x}^*) \text{ and } \mathbf{s} \in \partial g(\mathbf{v}^*)\},$$

which is equivalent to

$$\{(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^m : \mathbf{a} \in \mathcal{T}_f(\mathbf{x}^*) \text{ and } \mathbf{b} \in \mathcal{T}_g(\mathbf{v}^*)\}.$$

Fig. 2 illustrates the error set  $\mathcal{E}_1$ , error cone  $\mathcal{C}_1$ , and corresponding spherical part of  $\mathcal{C}_1$ . As we will see in this section, the Gaussian complexity of the spherical part of the error cone is closely related to the number of measurements to guarantee successful recovery.

Then we have the following result.

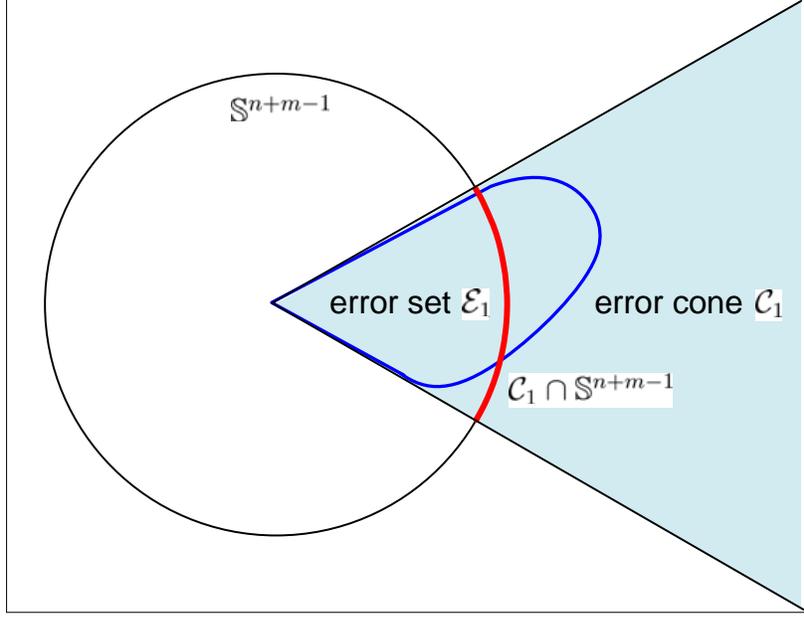


Fig. 2. Illustrations of the error set  $\mathcal{E}_1$ , error cone  $\mathcal{C}_1$ , and corresponding spherical part of  $\mathcal{C}_1$ :  $\mathcal{C}_1 \cap \mathbb{S}^{n+m-1}$  (denoted by the red arc).

**Theorem 2** (Constrained Recovery). *Let  $(\hat{\mathbf{x}}, \hat{\mathbf{v}})$  be the solution to either of the constrained optimization problems (2) or (3). If the number of measurements*

$$\sqrt{m} \geq CK^2\gamma(\mathcal{C}_1(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}) + \epsilon, \quad (21)$$

then

$$\sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2} \leq \frac{2\delta\sqrt{m}}{\epsilon}$$

with probability at least  $1 - \exp\{-\gamma^2(\mathcal{C}_1 \cap \mathbb{S}^{n+m-1})\}$ .

*Remark 4.* If we consider the observation model  $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \sqrt{m}\mathbf{v}^* + \mathbf{z}$  as that in [9], then the factor  $\sqrt{m}$  can be removed from the error bound. This change of the model seems to make our result more interpretable, but we analyze the original observation model (1) here.

*Proof.* Since  $(\hat{\mathbf{x}}, \hat{\mathbf{v}})$  solves (2) or (3), we have  $f(\hat{\mathbf{x}}) \leq f(\mathbf{x}^*)$  and  $g(\hat{\mathbf{v}}) \leq g(\mathbf{v}^*)$ . This implies  $(\hat{\mathbf{x}} - \mathbf{x}^*, \hat{\mathbf{v}} - \mathbf{v}^*) \in \mathcal{E}_1(\mathbf{x}^*, \mathbf{v}^*) \subset \mathcal{C}_1(\mathbf{x}^*, \mathbf{v}^*)$ . It then follows from (18) and (21) that the event

$$\min_{(\mathbf{a}, \mathbf{b}) \in \mathcal{C}_1(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}} \sqrt{m}\|\Phi\mathbf{a} + \mathbf{b}\|_2 \geq \sqrt{m} - CK^2\gamma(\mathcal{C}_1(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}) \geq \epsilon \quad (22)$$

holds with probability at least  $1 - \exp\{-\gamma^2(\mathcal{C}_1(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1})\}$ .

On the other hand, since both  $(\hat{\mathbf{x}}, \hat{\mathbf{v}})$  and  $(\mathbf{x}^*, \mathbf{v}^*)$  are feasible, by the triangle inequality, we have

$$\|\Phi(\hat{\mathbf{x}} - \mathbf{x}^*) + (\hat{\mathbf{v}} - \mathbf{v}^*)\|_2 \leq \|\mathbf{y} - \Phi\hat{\mathbf{x}} - \hat{\mathbf{v}}\|_2 + \|\mathbf{y} - \Phi\mathbf{x}^* - \mathbf{v}^*\|_2 \leq 2\delta. \quad (23)$$

Combining (22) and (23) yields

$$\begin{aligned} 2\delta &\geq \|\Phi(\hat{\mathbf{x}} - \mathbf{x}^*) + (\hat{\mathbf{v}} - \mathbf{v}^*)\|_2 \\ &= \sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2} \cdot \left\| \frac{\Phi(\hat{\mathbf{x}} - \mathbf{x}^*)}{\sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2}} + \frac{(\hat{\mathbf{v}} - \mathbf{v}^*)}{\sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2}} \right\|_2 \\ &\geq \sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2} \cdot \frac{\epsilon}{\sqrt{m}}. \end{aligned}$$

Rearranging completes the proof.  $\square$

To make use of Theorem 2, it suffices to bound  $\gamma(\mathcal{C}_1(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1})$ . Since a number of upper bounds on Gaussian width and Gaussian squared distance for different structured settings are available in the literature (see, e.g., [19], [20], [17],

[31]), it is highly desirable to bound  $\gamma(\mathcal{C}_1(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1})$  in terms of these familiar parameters. Then we have the following result.

**Lemma 1.** *The Gaussian complexity of  $\mathcal{C}_1(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}$  satisfies*

$$\gamma(\mathcal{C}_1(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}) \leq 2[\omega(\mathcal{T}_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1}) + \omega(\mathcal{T}_g(\mathbf{v}^*) \cap \mathbb{S}^{m-1}) + 1].$$

*Proof.* Let  $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$  and  $\mathbf{h} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$ , it follows from the definition of Gaussian complexity that

$$\begin{aligned} \gamma(\mathcal{C}_1(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}) &= \mathbb{E} \sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{C}_1(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}} |\langle \mathbf{g}, \mathbf{a} \rangle + \langle \mathbf{h}, \mathbf{b} \rangle| \\ &= \mathbb{E} \sup_{\substack{c \in (0,1) \\ \mathbf{a} \in \mathcal{T}_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1} \\ \mathbf{b} \in \mathcal{T}_g(\mathbf{v}^*) \cap \mathbb{S}^{m-1}}} |\langle \mathbf{g}, \mathbf{a} \rangle \cdot c + \langle \mathbf{h}, \mathbf{b} \rangle \cdot \sqrt{1-c^2}| \\ &\leq \mathbb{E} \sup_{\substack{c \in (0,1) \\ \mathbf{a} \in \mathcal{T}_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1} \\ \mathbf{b} \in \mathcal{T}_g(\mathbf{v}^*) \cap \mathbb{S}^{m-1}}} c \cdot |\langle \mathbf{g}, \mathbf{a} \rangle| + \sqrt{1-c^2} |\langle \mathbf{h}, \mathbf{b} \rangle| \\ &\leq \mathbb{E} \sup_{\mathbf{a} \in \mathcal{T}_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1}} |\langle \mathbf{g}, \mathbf{a} \rangle| + \mathbb{E} \sup_{\mathbf{b} \in \mathcal{T}_g(\mathbf{v}^*) \cap \mathbb{S}^{m-1}} |\langle \mathbf{h}, \mathbf{b} \rangle| \\ &= \gamma(\mathcal{T}_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1}) + \gamma(\mathcal{T}_g(\mathbf{v}^*) \cap \mathbb{S}^{m-1}) \\ &\leq 2[\omega(\mathcal{T}_f \cap \mathbb{S}^{n-1}) + \omega(\mathcal{T}_g \cap \mathbb{S}^{m-1}) + 1]. \end{aligned}$$

The last inequality is due to (8). □

Thus, (20) follows from Theorem 2 and Lemma 1.

*Remark 5.* When  $\Phi$  is a random orthogonal matrix ( $m = n$ ) and measurements are noiseless ( $\delta = 0$ ), Amelunxen *et al.* [20] show that if  $m \geq \iota(\mathcal{T}_f(\mathbf{x}^*)) + \iota(\mathcal{T}_g(\mathbf{v}^*))$ , then the constrained procedures (2) and (3) succeed with high probability, where the statistical dimension  $\iota(\mathcal{T})$  of a closed convex cone  $\mathcal{T} \in \mathbb{R}^n$  is defined as  $\iota(\mathcal{T}) = \mathbb{E} \left( \sup_{\mathbf{u} \in \mathcal{T} \cap \mathbb{B}_2} \langle \mathbf{u}, \mathbf{g} \rangle \right)^2$  with  $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ . They also establish the sharpness of their result. Our result (20) essentially coincides with this result because the statistical dimension of a closed convex cone  $\mathcal{T} \in \mathbb{R}^n$  is closely related to its Gaussian width [20, Proposition 10.2]:  $\omega^2(\mathcal{T} \cap \mathbb{S}^{n-1}) \leq \iota(\mathcal{T}) \leq \omega^2(\mathcal{T} \cap \mathbb{S}^{n-1}) + 1$ .

*Remark 6.* When  $\Phi$  is a Gaussian matrix ( $m > n$  or  $m \leq n$ ), Foygel and Mackey [17] establish similar result as (20) in terms of Gaussian squared complexity<sup>2</sup> under bounded noise ( $\|\mathbf{z}\|_2 \leq \delta$ ). They also numerically show that their result is sharp. In the absence of noise, Zhang *et al.* [21] recently establish that the phase transition of the constrained procedure under Gaussian measurements occurs around  $\omega^2(\mathcal{T}_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1}) + \omega^2(\mathcal{T}_g(\mathbf{v}^*) \cap \mathbb{S}^{m-1})$ .

### B. Recovery via Partially Penalized Optimization

We next present performance analysis for the partially penalized optimization problem (4). Our second result shows that, with high probability, approximately

$$CK^4 [\eta^2(\lambda_1 \cdot \partial f(\mathbf{x}^*)) + \eta^2(\lambda_2 \cdot \partial g(\mathbf{v}^*))] \quad (24)$$

corrupted measurements suffice to recover  $(\mathbf{x}^*, \mathbf{v}^*)$  exactly in the absence of noise and stably in the presence of noise, via the procedure (4). Here,  $\lambda_1$  and  $\lambda_2$  are absolute constants such that the regularization parameter  $\lambda = \lambda_2/\lambda_1$ .

In this case, it is natural to define the following error set

$$\mathcal{E}_2(\mathbf{x}^*, \mathbf{v}^*) := \{(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^m : f(\mathbf{x}^* + \mathbf{a}) + \lambda \cdot g(\mathbf{v}^* + \mathbf{b}) \leq f(\mathbf{x}^*) + \lambda \cdot g(\mathbf{v}^*)\}.$$

By the convexity of  $f$  and  $g$ ,  $\mathcal{E}_2(\mathbf{x}^*, \mathbf{v}^*)$  belongs to the following convex cone

$$\mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*) := \{(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^m : \langle \mathbf{a}, \mathbf{u} \rangle + \lambda \langle \mathbf{b}, \mathbf{s} \rangle \leq 0 \text{ for any } \mathbf{u} \in \partial f(\mathbf{x}^*) \text{ and } \mathbf{s} \in \partial g(\mathbf{v}^*)\}.$$

Then we have the following result.

**Theorem 3** (Partially Penalized Recovery). *Let  $(\hat{\mathbf{x}}, \hat{\mathbf{v}})$  be the solution to the partially penalized optimization problem (4). If the number of measurements*

$$\sqrt{m} \geq CK^2 \gamma(\mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}) + \epsilon, \quad (25)$$

<sup>2</sup>The Gaussian squared complexity of a set  $\mathcal{T} \in \mathbb{R}^n$  is defined as  $\zeta^2(\mathcal{T}) = \mathbb{E} (\sup_{\mathbf{u} \in \mathcal{T}} \langle \mathbf{u}, \mathbf{g} \rangle)_+^2$ , where  $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$  and  $(a)_+ = \max\{a, 0\}$ . Note that the Gaussian squared complexity  $\zeta^2(\mathcal{T}_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1})$  is very slightly bigger than the squared Gaussian width  $\omega^2(\mathcal{T}_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1})$ .

then

$$\sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2} \leq \frac{2\delta\sqrt{m}}{\epsilon}$$

with probability at least  $1 - \exp\{-\gamma^2(\mathcal{C}_2 \cap \mathbb{S}^{n+m-1})\}$ .

*Proof.* The proof is similar to that of Theorem 2. Since  $(\hat{\mathbf{x}}, \hat{\mathbf{v}})$  solves (4), we have  $(\hat{\mathbf{x}} - \mathbf{x}^*, \hat{\mathbf{v}} - \mathbf{v}^*) \in \mathcal{E}_2(\mathbf{x}^*, \mathbf{v}^*) \subset \mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*)$ . It then follows from (18) and (25) that the event

$$\min_{(\mathbf{a}, \mathbf{b}) \in \mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}} \sqrt{m} \|\Phi \mathbf{a} + \mathbf{b}\|_2 \geq \sqrt{m} - CK^2 \gamma (\mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}) \geq \epsilon \quad (26)$$

holds with probability at least  $1 - \exp\{-\gamma^2(\mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1})\}$ .

On the other hand, since both  $(\hat{\mathbf{x}}, \hat{\mathbf{v}})$  and  $(\mathbf{x}^*, \mathbf{v}^*)$  are feasible, by the triangle inequality, we have

$$\|\Phi(\hat{\mathbf{x}} - \mathbf{x}^*) + (\hat{\mathbf{v}} - \mathbf{v}^*)\|_2 \leq \|\mathbf{y} - \Phi \hat{\mathbf{x}} - \hat{\mathbf{v}}\|_2 + \|\mathbf{y} - \Phi \mathbf{x}^* - \mathbf{v}^*\|_2 \leq 2\delta. \quad (27)$$

Combining (26) and (27) completes the proof.  $\square$

To bound  $\gamma(\mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1})$  in terms of familiar parameters, we have the following result.

**Lemma 2.** *The Gaussian complexity of  $\mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}$  satisfies*

$$\gamma(\mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}) \leq 2\sqrt{\eta^2(\lambda_1 \cdot \partial f(\mathbf{x}^*)) + \eta^2(\lambda_2 \cdot \partial g(\mathbf{v}^*))} + 1.$$

*Proof.* By the definition of  $\mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*)$ , for any point  $(\mathbf{a}, \mathbf{b}) \in \mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*)$ , we have

$$\langle \mathbf{a}, \mathbf{u} \rangle + \lambda \langle \mathbf{b}, \mathbf{s} \rangle \leq 0$$

for any  $\mathbf{u} \in \partial f(\mathbf{x}^*)$  and  $\mathbf{s} \in \partial g(\mathbf{v}^*)$ . Multiplying both sides by  $\lambda_1$  yields

$$\langle \mathbf{a}, \lambda_1 \mathbf{u} \rangle + \langle \mathbf{b}, \lambda_2 \mathbf{s} \rangle \leq 0.$$

For any fixed  $\mathbf{g} \in \mathbb{R}^n$  and  $\mathbf{h} \in \mathbb{R}^m$ , it follows from the Cauchy-Schwarz inequality that

$$\begin{aligned} \langle \mathbf{a}, \mathbf{g} \rangle + \langle \mathbf{b}, \mathbf{h} \rangle &\leq \langle \mathbf{a}, \mathbf{g} - \lambda_1 \mathbf{u} \rangle + \langle \mathbf{b}, \mathbf{h} - \lambda_2 \mathbf{s} \rangle \\ &\leq \|\mathbf{a}\|_2 \|\mathbf{g} - \lambda_1 \mathbf{u}\|_2 + \|\mathbf{b}\|_2 \|\mathbf{h} - \lambda_2 \mathbf{s}\|_2. \end{aligned}$$

Choosing suitable  $\bar{\mathbf{u}} \in \partial f(\mathbf{x}^*)$  and  $\bar{\mathbf{s}} \in \partial g(\mathbf{v}^*)$  such that

$$\|\mathbf{g} - \lambda_1 \cdot \bar{\mathbf{u}}\|_2 = \text{dist}(\mathbf{g}, \lambda_1 \cdot \partial f(\mathbf{x}^*)) := d_f$$

and

$$\|\mathbf{h} - \lambda_2 \cdot \bar{\mathbf{s}}\|_2 = \text{dist}(\mathbf{h}, \lambda_2 \cdot \partial g(\mathbf{v}^*)) := d_g,$$

we obtain

$$\begin{aligned} \langle \mathbf{a}, \mathbf{g} \rangle + \langle \mathbf{b}, \mathbf{h} \rangle &\leq \|\mathbf{a}\|_2 \cdot \text{dist}(\mathbf{g}, \lambda_1 \cdot \partial f(\mathbf{x}^*)) + \|\mathbf{b}\|_2 \cdot \text{dist}(\mathbf{h}, \lambda_2 \cdot \partial g(\mathbf{v}^*)) \\ &= d_f \cdot \|\mathbf{a}\|_2 + d_g \cdot \|\mathbf{b}\|_2. \end{aligned} \quad (28)$$

Then, by the definition of Gaussian width,

$$\begin{aligned} \omega(\mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}) &= \mathbb{E} \sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}} [\langle \mathbf{g}, \mathbf{a} \rangle + \langle \mathbf{h}, \mathbf{b} \rangle] \\ &\leq \mathbb{E} \sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}} [\|\mathbf{a}\|_2 \cdot d_f + \|\mathbf{b}\|_2 \cdot d_g] \\ &\leq \mathbb{E} \sqrt{d_f^2 + d_g^2} \leq \sqrt{\mathbb{E} d_f^2 + \mathbb{E} d_g^2} = \sqrt{\eta^2(\lambda_1 \cdot \partial f(\mathbf{x}^*)) + \eta^2(\lambda_2 \cdot \partial g(\mathbf{v}^*))}, \end{aligned}$$

where  $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$  and  $\mathbf{h} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$ . The second and the third inequalities follow from the Cauchy-Schwarz and Jensen's inequalities, respectively. By (8), we complete the proof.  $\square$

Thus, combining Theorem 3 and Lemma 2 yields (24).

1) *How to Choose the Regularization Parameter  $\lambda$ ?*: The result (24) also suggests a simple strategy to choose  $\lambda_1$  and  $\lambda_2$ , and hence the regularization parameter  $\lambda = \lambda_2/\lambda_1$ . Observe that the number of observations to guarantee successful recovery by (4) depends on  $\lambda_1$  and  $\lambda_2$  through  $\eta^2(\lambda_1 \cdot \partial f(\mathbf{x}^*))$  and  $\eta^2(\lambda_2 \cdot \partial g(\mathbf{v}^*))$  respectively. In order to achieve the possibly smallest number of observations, it is natural to choose  $\lambda_1$  and  $\lambda_2$  which make the two Gaussian squared distances as small as possible. Thus we can pick

$$\lambda_1^* = \arg \min_{\lambda_1 > 0} \eta^2(\lambda_1 \cdot \partial f(\mathbf{x}^*)), \quad \lambda_2^* = \arg \min_{\lambda_2 > 0} \eta^2(\lambda_2 \cdot \partial g(\mathbf{v}^*)), \quad \text{and} \quad \lambda^* = \lambda_2^*/\lambda_1^*. \quad (29)$$

Furthermore, it has been shown in [20, Proposition 4.1] that if the subdifferential  $\partial f(\mathbf{x}^*)$  (or  $\partial g(\mathbf{v}^*)$ ) is nonempty, compact, and does not contain the original, then the function  $J(\kappa) = \eta^2(\kappa \cdot \partial f(\mathbf{x}^*))$  (or  $\eta^2(\kappa \cdot \partial g(\mathbf{v}^*))$ ) is strictly convex for  $\kappa > 0$ , so it achieves its minimum at a unique point. Thus, under mild conditions, the optimal regularization parameter  $\lambda^*$  is always unique.

*Remark 7.* In the setting of Gaussian measurements, Foygel and Mackey [17] establish similar result as (24) in terms of Gaussian squared distance under bounded noise ( $\|\mathbf{z}\|_2 \leq \delta$ ). Their result also indicates the same way to choose the regularization parameter  $\lambda$ . However, in the sub-Gaussian case, our proof is totally different from theirs. Because many useful properties of Gaussian distribution (e.g., the rotation invariance property, Gordon's comparison lemma [32]) which are heavily used in [17] do not hold in the sub-Gaussian case.

### C. Recovery via Fully Penalized Optimization

Finally, we analyze the fully penalized optimization problem (5). In this case, due to the presence of noise, we require regularization parameters  $\tau_1$  and  $\tau_2$  to satisfy the following condition.

**Condition 1.** For any  $\beta > 1$ ,

$$\tau_1 \geq \beta f^*(\Phi^T \mathbf{z}) \quad \text{and} \quad \tau_2 \geq \beta g^*(\mathbf{z}).$$

This condition is a natural extension to that in [33, Lemma 1], where only the regularization parameter for signal ( $\tau_1$ ) is considered and  $\beta = 2$ . Then our third result shows that, with high probability, approximately

$$CK^4 \left[ \eta^2(\tau_1 \cdot \partial f(\mathbf{x}^*)) + \eta^2(\tau_2 \cdot \partial g(\mathbf{v}^*)) + \frac{(\tau_1 \alpha_f)^2 + (\tau_2 \alpha_g)^2}{\beta^2} \right] \quad (30)$$

corrupted measurements suffice to recover  $(\mathbf{x}^*, \mathbf{v}^*)$  exactly in the absence of noise and stably in the presence of noise, via the procedure (5).

Under the above condition, we similarly define the error set

$$\mathcal{E}_3(\mathbf{x}^*, \mathbf{v}^*) := \{(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^m : \tau_1 f(\mathbf{x}^* + \mathbf{a}) + \tau_2 g(\mathbf{v}^* + \mathbf{b}) \leq \tau_1 f(\mathbf{x}^*) + \tau_2 g(\mathbf{v}^*) + \frac{1}{\beta} [\tau_1 f(\mathbf{a}) + \tau_2 g(\mathbf{b})]\}.$$

Note that if  $0 < \beta \leq 1$ , it then follows from the triangle inequality that the inequality defining  $\mathcal{E}_3(\mathbf{x}^*, \mathbf{v}^*)$  holds for all  $(\mathbf{a}, \mathbf{b})$ . Thus we require  $\beta > 1$ , which will restrict the set of  $\mathcal{E}_3(\mathbf{x}^*, \mathbf{v}^*)$  and yield the restricted error set. By the convexity of  $f$  and  $g$ ,  $\mathcal{E}_3(\mathbf{x}^*, \mathbf{v}^*)$  belongs to the following convex cone

$$\mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*) := \{(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^m : \tau_1 \langle \mathbf{a}, \mathbf{u} \rangle + \tau_2 \langle \mathbf{b}, \mathbf{s} \rangle \leq \frac{1}{\beta} [\tau_1 f(\mathbf{a}) + \tau_2 g(\mathbf{b})] \text{ for any } \mathbf{u} \in \partial f(\mathbf{x}^*) \text{ and } \mathbf{s} \in \partial g(\mathbf{v}^*)\}.$$

Then we have the following result.

**Theorem 4** (Fully Penalized Recovery). *Let  $(\hat{\mathbf{x}}, \hat{\mathbf{v}})$  be the solution to the fully penalized optimization problem (5) with  $\tau_1$  and  $\tau_2$  satisfying Condition 1. If the number of measurements*

$$\sqrt{m} \geq CK^2 \gamma (\mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}) + \epsilon, \quad (31)$$

then

$$\sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2} \leq 2m \cdot \frac{\beta + 1}{\beta} \cdot \frac{(\tau_1 \alpha_f + \tau_2 \alpha_g)}{\epsilon^2} \quad (32)$$

with probability at least  $1 - \exp\{-\gamma^2(\mathcal{C}_3 \cap \mathbb{S}^{n+m-1})\}$ .

*Proof.* Since  $(\hat{\mathbf{x}}, \hat{\mathbf{v}})$  solves (5), we have

$$\frac{1}{2} \|\mathbf{y} - \Phi \hat{\mathbf{x}} - \hat{\mathbf{v}}\|_2^2 + \tau_1 f(\hat{\mathbf{x}}) + \tau_2 g(\hat{\mathbf{v}}) \leq \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x}^* - \mathbf{v}^*\|_2^2 + \tau_1 f(\mathbf{x}^*) + \tau_2 g(\mathbf{v}^*). \quad (33)$$

Observe that

$$\frac{1}{2} \|\mathbf{y} - \Phi \hat{\mathbf{x}} - \hat{\mathbf{v}}\|_2^2 = \frac{1}{2} \|\Phi(\hat{\mathbf{x}} - \mathbf{x}^*) + (\hat{\mathbf{v}} - \mathbf{v}^*)\|_2^2 + \frac{1}{2} \|\mathbf{z}\|_2^2 - \langle \Phi(\hat{\mathbf{x}} - \mathbf{x}^*), \mathbf{z} \rangle - \langle \hat{\mathbf{v}} - \mathbf{v}^*, \mathbf{z} \rangle.$$

Substituting this into (33) yields

$$\frac{1}{2} \|\Phi(\hat{\mathbf{x}} - \mathbf{x}^*) + (\hat{\mathbf{v}} - \mathbf{v}^*)\|_2^2 \leq \tau_1 [f(\mathbf{x}^*) - f(\hat{\mathbf{x}})] + \tau_2 [g(\mathbf{v}^*) - g(\hat{\mathbf{v}})] + \langle \Phi(\hat{\mathbf{x}} - \mathbf{x}^*), \mathbf{z} \rangle + \langle \hat{\mathbf{v}} - \mathbf{v}^*, \mathbf{z} \rangle. \quad (34)$$

On the one hand, since  $\|\Phi(\hat{\mathbf{x}} - \mathbf{x}^*) + (\hat{\mathbf{v}} - \mathbf{v}^*)\|_2^2 \geq 0$ , we have

$$\begin{aligned} \tau_1 f(\hat{\mathbf{x}}) + \tau_2 g(\hat{\mathbf{v}}) &\leq \tau_1 f(\mathbf{x}^*) + \tau_2 g(\mathbf{v}^*) + \langle \Phi(\hat{\mathbf{x}} - \mathbf{x}^*), \mathbf{z} \rangle + \langle \hat{\mathbf{v}} - \mathbf{v}^*, \mathbf{z} \rangle \\ &\leq \tau_1 f(\mathbf{x}^*) + \tau_2 g(\mathbf{v}^*) + f^*(\Phi^T \mathbf{z}) \cdot f(\hat{\mathbf{x}} - \mathbf{x}^*) + g^*(\mathbf{z}) \cdot g(\hat{\mathbf{v}} - \mathbf{v}^*) \\ &\leq \tau_1 f(\mathbf{x}^*) + \tau_2 g(\mathbf{v}^*) + \frac{\tau_1}{\beta} \cdot f(\hat{\mathbf{x}} - \mathbf{x}^*) + \frac{\tau_2}{\beta} \cdot g(\hat{\mathbf{v}} - \mathbf{v}^*), \end{aligned}$$

where  $f^*(\cdot)$  and  $g^*(\cdot)$  denote the dual norm<sup>3</sup> of  $f(\cdot)$  and  $g(\cdot)$  respectively. The second inequality follows from generalized Hölder's inequality. The last inequality is due to Condition 1. This implies  $(\hat{\mathbf{x}} - \mathbf{x}^*, \hat{\mathbf{v}} - \mathbf{v}^*) \in \mathcal{E}_3(\mathbf{x}^*, \mathbf{v}^*) \subset \mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*)$ . It then follows from (18) and (31) that the event

$$\min_{(\mathbf{a}, \mathbf{b}) \in \mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}} \sqrt{m} \|\Phi \mathbf{a} + \mathbf{b}\|_2 \geq \sqrt{m} - CK^2 \gamma (\mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}) \geq \epsilon \quad (35)$$

holds with probability at least  $1 - \exp\{-\gamma^2 (\mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1})\}$ .

On the other hand, it follows from (34) that

$$\begin{aligned} \frac{1}{2} \|\Phi(\hat{\mathbf{x}} - \mathbf{x}^*) + (\hat{\mathbf{v}} - \mathbf{v}^*)\|_2^2 &\leq \tau_1 \cdot f(\hat{\mathbf{x}} - \mathbf{x}^*) + \tau_2 \cdot g(\hat{\mathbf{v}} - \mathbf{v}^*) + \frac{\tau_1}{\beta} \cdot f(\hat{\mathbf{x}} - \mathbf{x}^*) + \frac{\tau_2}{\beta} \cdot g(\hat{\mathbf{v}} - \mathbf{v}^*) \\ &= \frac{\beta + 1}{\beta} (\tau_1 \cdot f(\hat{\mathbf{x}} - \mathbf{x}^*) + \tau_2 \cdot g(\hat{\mathbf{v}} - \mathbf{v}^*)) \\ &= \frac{\beta + 1}{\beta} (\alpha_f \tau_1 \cdot \|\hat{\mathbf{x}} - \mathbf{x}^*\|_2 + \alpha_g \tau_2 \cdot \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2) \\ &\leq \frac{\beta + 1}{\beta} \cdot \sqrt{(\tau_1 \alpha_f)^2 + (\tau_2 \alpha_g)^2} \cdot \sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2} \\ &\leq \frac{\beta + 1}{\beta} \cdot (\tau_1 \alpha_f + \tau_2 \alpha_g) \cdot \sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2}, \end{aligned} \quad (36)$$

where  $\alpha_f$  and  $\alpha_g$  are compatibility constants. The first inequality follows from the triangle inequality. In the last two inequalities, we have used the Cauchy-Schwarz inequality and the fact that  $\sqrt{a^2 + b^2} \leq a + b$  for  $a, b \geq 0$ , respectively.

Combining (35) and (36) yields

$$\begin{aligned} \frac{2(\beta + 1)}{\beta} \cdot (\tau_1 \alpha_f + \tau_2 \alpha_g) &\geq \sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2} \cdot \left\| \frac{\Phi(\hat{\mathbf{x}} - \mathbf{x}^*)}{\sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2}} + \frac{(\hat{\mathbf{v}} - \mathbf{v}^*)}{\sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2}} \right\|_2^2 \\ &\geq \sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2} \cdot \frac{\epsilon^2}{m}. \end{aligned}$$

Rearranging yields the desired result.  $\square$

To bound the Gaussian complexity of  $\mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}$  in terms of familiar parameters, we have the following result.

**Lemma 3.**

$$\gamma(\mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}) \leq 2 \left[ \sqrt{\eta^2 (\tau_1 \cdot \partial f(\mathbf{x}^*)) + \eta^2 (\tau_2 \cdot \partial g(\mathbf{v}^*))} + \frac{\sqrt{(\tau_1 \alpha_f)^2 + (\tau_2 \alpha_g)^2}}{\beta} \right] + 1.$$

*Proof.* The proof is similar to that of lemma 2. By the definition of  $\mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*)$ , for any point  $(\mathbf{a}, \mathbf{b}) \in \mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*)$ , we have

$$\langle \mathbf{a}, \tau_1 \mathbf{u} \rangle + \langle \mathbf{b}, \tau_2 \mathbf{s} \rangle - \frac{1}{\beta} [\tau_1 f(\mathbf{a}) + \tau_2 g(\mathbf{b})] \leq 0$$

for any  $\mathbf{u} \in \partial f(\mathbf{x}^*)$  and  $\mathbf{s} \in \partial g(\mathbf{v}^*)$ . It then follows from the Cauchy-Schwarz inequality that

$$\begin{aligned} \langle \mathbf{a}, \mathbf{g} \rangle + \langle \mathbf{b}, \mathbf{h} \rangle &\leq \langle \mathbf{a}, \mathbf{g} - \tau_1 \mathbf{u} \rangle + \langle \mathbf{b}, \mathbf{h} - \tau_2 \mathbf{s} \rangle + \frac{1}{\beta} [\tau_1 f(\mathbf{a}) + \tau_2 g(\mathbf{b})] \\ &\leq \|\mathbf{a}\|_2 \|\mathbf{g} - \tau_1 \mathbf{u}\|_2 + \|\mathbf{b}\|_2 \|\mathbf{h} - \tau_2 \mathbf{s}\|_2 + \frac{1}{\beta} [\tau_1 f(\mathbf{a}) + \tau_2 g(\mathbf{b})], \end{aligned}$$

<sup>3</sup>The dual norm of  $f$  is defined as  $f^*(\mathbf{d}) = \sup_{\mathbf{u} \in \mathbb{B}_f^1} \langle \mathbf{u}, \mathbf{d} \rangle$ , where  $\mathbb{B}_f^1 = \{\mathbf{u} \in \mathbb{R}^n : f(\mathbf{u}) \leq 1\}$ .

where  $\mathbf{g} \in \mathbb{R}^n$  and  $\mathbf{h} \in \mathbb{R}^m$  are arbitrary vectors. Choosing suitable  $\tilde{\mathbf{u}} \in \partial f(\mathbf{x}^*)$  and  $\tilde{\mathbf{s}} \in \partial g(\mathbf{v}^*)$  such that

$$\|\mathbf{g} - \tau_1 \cdot \tilde{\mathbf{u}}\|_2 = \text{dist}(\mathbf{g}, \tau_1 \cdot \partial f(\mathbf{x}^*)) := d'_f$$

and

$$\|\mathbf{h} - \tau_2 \cdot \tilde{\mathbf{s}}\|_2 = \text{dist}(\mathbf{h}, \tau_2 \cdot \partial g(\mathbf{v}^*)) := d'_g,$$

we obtain

$$\begin{aligned} \langle \mathbf{a}, \mathbf{g} \rangle + \langle \mathbf{b}, \mathbf{h} \rangle &\leq \|\mathbf{a}\|_2 \cdot \text{dist}(\mathbf{g}, \tau_1 \cdot \partial f(\mathbf{x}^*)) + \|\mathbf{b}\|_2 \cdot \text{dist}(\mathbf{h}, \tau_2 \cdot \partial g(\mathbf{v}^*)) + \frac{1}{\beta} [\tau_1 f(\mathbf{a}) + \tau_2 g(\mathbf{b})] \\ &\leq d'_f \cdot \|\mathbf{a}\|_2 + d'_g \cdot \|\mathbf{b}\|_2 + \frac{1}{\beta} [\tau_1 \alpha_f \|\mathbf{a}\|_2 + \tau_2 \alpha_g \|\mathbf{b}\|_2], \end{aligned} \quad (37)$$

where the last inequality follows from the definition of the compatibility constant.

Then, by the definition of Gaussian width,

$$\begin{aligned} \omega(\mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}) &= \mathbb{E} \sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}} [\langle \mathbf{g}, \mathbf{a} \rangle + \langle \mathbf{h}, \mathbf{b} \rangle] \\ &\leq \mathbb{E} \sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}} \left[ d'_f \cdot \|\mathbf{a}\|_2 + d'_g \cdot \|\mathbf{b}\|_2 + \frac{\tau_1 \alpha_f}{\beta} \cdot \|\mathbf{a}\|_2 + \frac{\tau_2 \alpha_g}{\beta} \cdot \|\mathbf{b}\|_2 \right] \\ &\leq \mathbb{E} \sqrt{d_f'^2 + d_g'^2} + \frac{\sqrt{(\tau_1 \alpha_f)^2 + (\tau_2 \alpha_g)^2}}{\beta} \\ &\leq \sqrt{\eta^2 (\tau_1 \cdot \partial f(\mathbf{x}^*)) + \eta^2 (\tau_2 \cdot \partial g(\mathbf{v}^*))} + \frac{\sqrt{(\tau_1 \alpha_f)^2 + (\tau_2 \alpha_g)^2}}{\beta}, \end{aligned}$$

where  $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$  and  $\mathbf{h} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$ . The second and the third inequalities follow from the Cauchy-Schwarz and Jensen's inequalities respectively. By (8), we complete the proof.  $\square$

Clearly, (30) comes from Theorem 4 and Lemma 3.

1) *Identify the Range of  $\tau_1$  and  $\tau_2$ :* Since our result relies on Condition 1, it is necessary to identify the range of the regularization parameters  $\tau_1$  and  $\tau_2$  under different kinds of noise. In the absence of noise ( $\mathbf{z} = \mathbf{0}$ ), we can easily see that Condition 1 holds with  $\tau_1 \geq 0$  and  $\tau_2 \geq 0$ . In general, we establish the following Chevet-type inequality which indicates Condition 1 holds with high probability under both bounded and stochastic noise scenarios.

**Lemma 4.** *Let  $\mathbf{A}$  be an  $m \times n$  matrix whose rows  $\mathbf{A}_i$  are independent centered isotropic sub-Gaussian vectors with  $\max_i \|\mathbf{A}_i\|_{\psi_2} \leq K$ , and  $\mathbf{w}$  be any fixed vector. Let  $\mathcal{T}$  be any bounded subset  $\mathbb{R}^n$ . Then, for any  $t \geq 0$ , the event*

$$\sup_{\mathbf{u} \in \mathcal{T}} \langle \mathbf{A}\mathbf{u}, \mathbf{w} \rangle \leq CK \|\mathbf{w}\|_2 [\gamma(\mathcal{T}) + t \cdot \text{rad}(\mathcal{T})]$$

holds with probability at least  $1 - \exp\{-t^2\}$ .

*Proof.* Define the random process

$$X_{\mathbf{u}} := \langle \mathbf{A}\mathbf{u}, \mathbf{w} \rangle, \text{ for any } \mathbf{u} \in \mathcal{T},$$

which has sub-Gaussian increments

$$\begin{aligned} \|X_{\mathbf{u}} - X_{\mathbf{u}'}\|_{\psi_2} &= \|\langle \mathbf{A}(\mathbf{u} - \mathbf{u}'), \mathbf{w} \rangle\|_{\psi_2} \\ &\leq \|\mathbf{w}\|_2 \|\mathbf{A}(\mathbf{u} - \mathbf{u}')\|_{\psi_2} \\ &\leq CK \|\mathbf{w}\|_2 \|\mathbf{u} - \mathbf{u}'\|_2 \end{aligned}$$

for any  $\mathbf{u}, \mathbf{u}' \in \mathcal{T}$ . The first inequality follows from the definition of the  $\psi_2$ -norm of a sub-Gaussian random vector (13) and the last inequality holds because  $\{\langle \mathbf{A}_i^T, (\mathbf{u} - \mathbf{u}') \rangle\}_{i=1}^m$  are independent centered sub-Gaussian random variables with  $\|\langle \mathbf{A}_i^T, (\mathbf{u} - \mathbf{u}') \rangle\|_{\psi_2} \leq K \|\mathbf{u} - \mathbf{u}'\|_2$  and hence  $\|\mathbf{A}(\mathbf{u} - \mathbf{u}')\|_{\psi_2} \leq CK \|\mathbf{u} - \mathbf{u}'\|_2$ . Define  $\bar{\mathcal{T}} = \mathcal{T} \cup \{\mathbf{0}\}$ . It follows from Talagrand's Majorizing Measure Theorem that the event

$$\begin{aligned} \sup_{\mathbf{u} \in \mathcal{T}} \langle \mathbf{A}\mathbf{u}, \mathbf{w} \rangle &\leq \sup_{\mathbf{u} \in \bar{\mathcal{T}}} |\langle \mathbf{A}\mathbf{u}, \mathbf{w} \rangle| = \sup_{\mathbf{u} \in \bar{\mathcal{T}}} |\langle \mathbf{A}\mathbf{u}, \mathbf{w} \rangle| \\ &= \sup_{\mathbf{u} \in \bar{\mathcal{T}}} |\langle \mathbf{A}\mathbf{u}, \mathbf{w} \rangle - \langle \mathbf{A}\mathbf{0}, \mathbf{w} \rangle| \\ &\leq \sup_{\mathbf{u}, \mathbf{u}' \in \bar{\mathcal{T}}} |\langle \mathbf{A}\mathbf{u}, \mathbf{w} \rangle - \langle \mathbf{A}\mathbf{u}', \mathbf{w} \rangle| \\ &\leq C' K \|\mathbf{w}\|_2 (\omega(\bar{\mathcal{T}}) + t \text{diam}(\bar{\mathcal{T}})) \\ &\leq C'' K \|\mathbf{w}\|_2 (\gamma(\mathcal{T}) + t \text{rad}(\mathcal{T})) \end{aligned}$$

holds with probability at least  $1 - \exp\{-t^2\}$ . In the last inequality, we have used the facts that  $\omega(\bar{\mathcal{T}}) \leq \gamma(\bar{\mathcal{T}}) = \gamma(\mathcal{T})$  and  $\text{diam}(\mathcal{T}) \leq 2\text{rad}(\mathcal{T})$ . This completes the proof.  $\square$

**Bounded Noise Case:** When the noise is bounded ( $\|\mathbf{z}\|_2 \leq \delta$ ), it follows from Lemma 4 that the event (choosing  $t = \sqrt{m}$ )

$$f^*(\Phi^T \mathbf{z}) = \sup_{\mathbf{u} \in \mathbb{B}_f^n} \langle \Phi \mathbf{u}, \mathbf{z} \rangle \leq \frac{CK\delta}{\sqrt{m}} [\gamma(\mathbb{B}_f^n) + \sqrt{m} \cdot r_f]$$

holds with probability at least  $1 - \exp(-m)$ , where  $\mathbb{B}_f^n = \{\mathbf{u} \in \mathbb{R}^n : f(\mathbf{u}) \leq 1\}$  and  $r_f = \sup\{\|\mathbf{u}\|_2 : \mathbf{u} \in \mathbb{B}_f^n\}$ . Thus it is natural to choose

$$\tau_1 \geq \frac{CK\delta\beta}{\sqrt{m}} [\gamma(\mathbb{B}_f^n) + \sqrt{m} \cdot r_f] := \tau_{1B}, \quad (38)$$

which implies that the first part of Condition 1 holds with high probability. In addition, note that  $g^*(\mathbf{z}) = \sup_{\mathbf{u} \in \mathbb{B}_g^m} \langle \mathbf{z}, \mathbf{u} \rangle \leq \delta \sup_{\mathbf{u} \in \mathbb{B}_g^m} \|\mathbf{u}\|_2 = \delta \cdot r_g$ , where  $\mathbb{B}_g^m = \{\mathbf{u} \in \mathbb{R}^m : g(\mathbf{u}) \leq 1\}$  and  $r_g = \sup\{\|\mathbf{u}\|_2 : \mathbf{u} \in \mathbb{B}_g^m\}$ . Therefore, we can choose

$$\tau_2 \geq \beta\delta \cdot r_g := \tau_{2B}. \quad (39)$$

**Sub-Gaussian Noise Case:** When  $\mathbf{z}$  is a sub-Gaussian random vector satisfying (7), then  $\|\mathbf{z}\|_2$  concentrates near the value  $\sqrt{m}$  with high probability [25, Theorem 3.1.1], namely  $\|\|\mathbf{z}\|_2 - \sqrt{m}\|_{\psi_2} \leq CK^2$ . This implies

$$\mathbb{P}\{\|\mathbf{z}\|_2 \geq (L^2 + 1)\sqrt{m}\} \leq \mathbb{P}\{|\|\mathbf{z}\|_2 - \sqrt{m}| \geq L^2\sqrt{m}\} \leq 2e^{-cm}.$$

Combining this with Lemma 4 and taking union bound yields

$$f^*(\Phi^T \mathbf{z}) = \sup_{\mathbf{u} \in \mathbb{B}_f^n} \langle \Phi \mathbf{u}, \mathbf{z} \rangle \leq CK(1 + L^2) [\gamma(\mathbb{B}_f^n) + \sqrt{m} \cdot r_f]$$

with probability at least  $1 - 3e^{-cm}$ . Thus we can choose

$$\tau_1 \geq CK(1 + L^2)\beta [\gamma(\mathbb{B}_f^n) + \sqrt{m} \cdot r_f] := \tau_{1S}, \quad (40)$$

which means that the first part of Condition 1 holds with high probability in the sub-Gaussian noise case. For the second part of Condition 1, note that the random process

$$X_{\mathbf{a}} := \langle \mathbf{a}, \mathbf{z} \rangle, \quad \text{for any } \mathbf{a} \in \mathcal{T}$$

has sub-Gaussian increments

$$\|X_{\mathbf{a}} - X_{\mathbf{a}'}\|_{\psi_2} = \|\langle \mathbf{a} - \mathbf{a}', \mathbf{z} \rangle\|_{\psi_2} \leq L \cdot \|\mathbf{a} - \mathbf{a}'\|_2 \quad \text{for any } \mathbf{a}, \mathbf{a}' \in \mathcal{T}.$$

It follows from Talagrand's Majorizing Measure Theorem that the event (similar arguments to that of Lemma 4)

$$\sup_{\mathbf{a} \in \mathcal{T}} \langle \mathbf{a}, \mathbf{z} \rangle \leq CL[\gamma(\mathcal{T}) + t\text{rad}(\mathcal{T})]$$

holds with probability at least  $1 - \exp(-t^2)$ . Choosing  $t = \sqrt{m}$ , we obtain that the event

$$g^*(\mathbf{z}) = \sup_{\mathbf{u} \in \mathbb{B}_g^m} \langle \mathbf{z}, \mathbf{u} \rangle \leq CL[\gamma(\mathbb{B}_g^m) + \sqrt{m} \cdot r_g]$$

holds with probability at least  $1 - \exp\{-m\}$ . Thus, in the sub-Gaussian noise case, we can choose

$$\tau_2 \geq CL\beta [\gamma(\mathbb{B}_g^m) + \sqrt{m} \cdot r_g] := \tau_{2S}. \quad (41)$$

2) *How to Choose the Regularization Parameters  $\tau_1$  and  $\tau_2$ ?:* Once the range of  $\tau_1$  and  $\tau_2$  is determined, then we can choose the ‘‘best’’ regularization parameters under certain criteria in this range. Our theoretical results show that both the number of observations (30) and the recovery error bound (32) rely on the regularization parameters  $\tau_1$  and  $\tau_2$ . Then one might expect to select  $\tau_1$  and  $\tau_2$  in the specified range such that these two quantities are as small as possible. However, it is not hard to see that these two quantities do not achieve their minima at the same time in general, so it is practical to choose some criteria which make a tradeoff between the number of observations and the recovery error bound. Specifically, if one wants to choose  $\tau_1$  and  $\tau_2$  such that the number of observations (30) which guarantees successful recovery for (5) is as small as possible, then one can pick

$$\tau_1^* = \arg \min_{\tau_1 \geq \tau_{1B}} \left\{ \eta^2(\tau_1 \cdot \partial f(\mathbf{x}^*)) + \frac{\alpha_f^2}{\beta^2} \tau_1^2 \right\} \quad \text{and} \quad \tau_2^* = \arg \min_{\tau_2 \geq \tau_{2B}} \left\{ \eta^2(\tau_2 \cdot \partial g(\mathbf{v}^*)) + \frac{\alpha_g^2}{\beta^2} \tau_2^2 \right\}, \quad (42)$$

where  $\tau_1$ . (or  $\tau_2$ .) denotes  $\tau_{1B}$  (or  $\tau_{2B}$ ) in the bounded noise case and  $\tau_{1S}$  (or  $\tau_{2S}$ ) in the sub-Gaussian noise case. Moreover, when both  $\partial f(\mathbf{x}^*)$  and  $\partial g(\mathbf{v}^*)$  are nonempty, compact, and do not contain the original, by the strict convexity of objective

functions, these “best” parameters are uniquely attained. If one wants to select  $\tau_1$  and  $\tau_2$  which make the recovery error bound (32) as small as possible, then one can pick

$$\tau_1^* = \arg \min_{\tau_1 \geq \tau_1} \frac{2m(\beta+1)\alpha_f}{\beta\epsilon^2} \tau_1 = \tau_1. \quad \text{and} \quad \tau_2^* = \arg \min_{\tau_2 \geq \tau_2} \frac{2m(\beta+1)\alpha_g}{\beta\epsilon^2} \tau_2 = \tau_2. \quad (43)$$

This result demonstrates that if we want to recover signal and corruption accurately by the fully penalized procedure in the noise-free case, we require

$$\tau_1 \rightarrow 0_+ \quad \text{and} \quad \tau_2 \rightarrow 0_+, \quad (44)$$

which might be regarded as promoting the equality constraint  $\mathbf{y} = \Phi \mathbf{x} + \mathbf{v}$  in the fully penalized optimization problem.

#### D. Relationships Among Three Procedures

The theory of Lagrange multipliers [34, Section 28] asserts that the three procedures are essentially equivalent when the regularization parameters  $\lambda$ ,  $\tau_1$ , and  $\tau_2$  are chosen correctly. However, the primary difficulty lies in determining these parameters when prior information is unavailable. Our results, (24), (30), and (32), suggest some useful strategies to choose these parameters and shed some light on the relationships among these approaches.

Let  $m_1$ ,  $m_2$ , and  $m_3$  be the necessary numbers of observations which guarantee successful recovery for constrained, partially penalized, and fully penalized recovery procedures, respectively. Since  $\mathcal{C}_1(\mathbf{x}^*, \mathbf{v}^*) \subseteq \mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*) \subseteq \mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*)$ , by the definition of Gaussian complexity, we have  $\gamma(\mathcal{C}_1(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}) \leq \gamma(\mathcal{C}_2(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1}) \leq \gamma(\mathcal{C}_3(\mathbf{x}^*, \mathbf{v}^*) \cap \mathbb{S}^{n+m-1})$  and hence  $m_1 \leq m_2 \leq m_3$ . This relation seems natural, because one might expect that the more prior information, the less number of measurements to guarantee success.

On the other hand, it follows from [20, Theorem 4.3, Proposition 10.2] and [17, Proposition 1] that, under a weak decomposition assumption<sup>4</sup>, the optimized Gaussian squared distance  $\min_{\kappa \geq 0} \eta^2(\kappa \cdot \partial f(\mathbf{x}^*))$  provides a faithful approximation to the statistical dimension  $\iota(\mathcal{T}_f(\mathbf{x}^*))$ , and hence the squared Gaussian width  $\omega^2(\mathcal{T}_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1})$ , i.e.,  $\min_{\kappa \geq 0} \eta^2(\kappa \cdot \partial f(\mathbf{x}^*)) \approx \omega^2(\mathcal{T}_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1})$ . This implies that

$$\eta^2(\lambda_1^* \cdot \partial f(\mathbf{x}^*)) \approx \omega^2(\mathcal{T}_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1}) \quad \text{and} \quad \eta^2(\lambda_2^* \cdot \partial g(\mathbf{v}^*)) \approx \omega^2(\mathcal{T}_g(\mathbf{v}^*) \cap \mathbb{S}^{m-1}),$$

provided that both  $\partial f(\mathbf{x}^*)$  and  $\partial g(\mathbf{v}^*)$  are nonempty, compact, and do not contain the original. It then follows from (20) and (24) that  $m_1 \approx m_2$  if we choose  $\lambda$  according to (29). For the fully penalized procedure, if we pick  $\tau_1$  and  $\tau_2$  according to (42) under the noise-free case, i.e.,

$$\tau_1^* = \arg \min_{\tau_1 > 0} \left\{ \eta^2(\tau_1 \cdot \partial f(\mathbf{x}^*)) + \frac{\alpha_f^2}{\beta^2} \tau_1^2 \right\} \quad \text{and} \quad \tau_2^* = \arg \min_{\tau_2 > 0} \left\{ \eta^2(\tau_2 \cdot \partial g(\mathbf{v}^*)) + \frac{\alpha_g^2}{\beta^2} \tau_2^2 \right\},$$

and if  $\beta$  is chosen large enough such that both  $\alpha_f^2 \tau_1^2 / \beta^2$  and  $\alpha_g^2 \tau_2^2 / \beta^2$  are negligible, then we have  $\tau_1^* \approx \lambda_1^*$  and  $\tau_2^* \approx \lambda_2^*$ , and hence  $m_2 \approx m_3$ . Thus we might conclude that under proper parameter selection strategies, the three approaches are approximately equivalent in terms of the necessary number of observations to guarantee success.

## V. NUMERICAL SIMULATIONS

In this section, we provide a series of numerical tests to verify our theoretical results in Section IV. We present constrained, partially penalized, and fully penalized recovery experiments for sparse signals and sparse corruptions in the absence or presence of noise. In each experiment, we employ the CVX Matlab package [35], [36] to solve our convex optimization problems.

### A. Phase Transition of Constrained Recovery Procedure

We first investigate the empirical behavior of the constrained recovery procedure when the noise level  $\delta = 0$  and the norm of the true signal  $f(\mathbf{x}^*) = \|\mathbf{x}^*\|_1$  are known exactly. We fix the sample size and signal length  $m = n = 128$ , and vary the sparsity levels  $(s_{\text{sig}}, s_{\text{cor}}) \in [1, 128] \times [1, 128]$ . We implement the following experiment 20 times for each  $(s_{\text{sig}}, s_{\text{cor}})$  pair:

- (1) Generate a signal vector  $\mathbf{x}^*$  with  $s_{\text{sig}}$  independent standard normal entries and set the other  $n - s_{\text{sig}}$  entries to 0.
- (2) Generate a corruption vector  $\mathbf{v}^*$  with  $s_{\text{cor}}$  independent standard normal entries and set the other  $m - s_{\text{cor}}$  entries to 0.
- (3) For Gaussian measurements, draw a sensing matrix  $\Phi \in \mathbb{R}^{m \times n}$  with independent  $\mathcal{N}(0, 1/m)$  entries; for scaled symmetric Bernoulli measurements, draw a sensing matrix  $\Phi \in \mathbb{R}^{m \times n}$  with independent entries obeying  $\mathbb{P}\{\Phi_{i,j} = 1/\sqrt{m}\} = 1/2$  and  $\mathbb{P}\{\Phi_{i,j} = -1/\sqrt{m}\} = 1/2$ .
- (4) Solve the following constrained optimization problem:

$$(\hat{\mathbf{x}}, \hat{\mathbf{v}}) \in \arg \min_{\mathbf{x}, \mathbf{v}} \|\mathbf{v}\|_1, \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \leq \|\mathbf{x}^*\|_1, \quad \mathbf{y} = \Phi \mathbf{x} + \mathbf{v}. \quad (45)$$

<sup>4</sup>For  $\mathbf{x} \neq \mathbf{0}$ ,  $\partial f(\mathbf{x})$  satisfies the weak decomposition assumption:  $\exists \mathbf{w}_0 \in \partial f(\mathbf{x})$  s.t.  $\langle \mathbf{w} - \mathbf{w}_0, \mathbf{w}_0 \rangle = 0, \forall \mathbf{w} \in \partial f(\mathbf{x})$ .

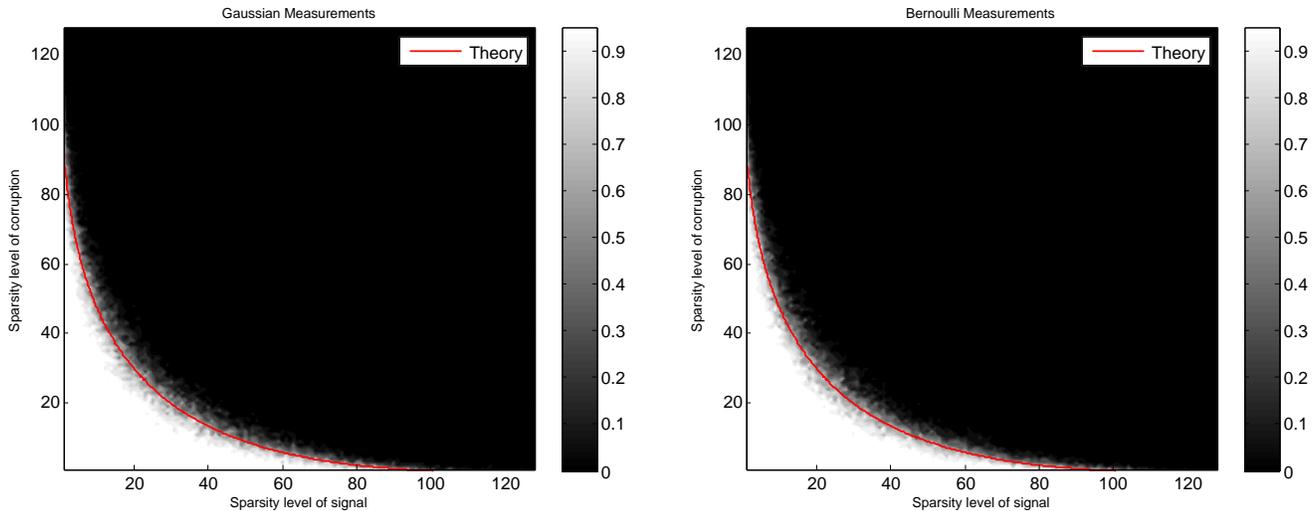


Fig. 3. Phase transition of the constrained recovery procedure (45) under Gaussian and Bernoulli measurements. The red curve plots the phase transition threshold predicted by (46).

(5) Declare success if  $\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2 / \|\mathbf{x}^*\|_2 \leq 10^{-3}$ .

To compare these empirical results with our theory, we overlay the theoretical recovery threshold suggested by (20)<sup>5</sup>

$$\omega^2(\mathcal{T}_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1}) + \omega^2(\mathcal{T}_g(\mathbf{v}^*) \cap \mathbb{S}^{m-1}), \quad (46)$$

where the squared Gaussian widths of signals and corruptions can be accurately estimated by (see, e.g., [19, Appendix C] or [17, eq.(10)])

$$\min_{t \geq 0} \mathbb{E} \left[ \inf_{\mathbf{u} \in t \cdot \partial \|\mathbf{x}^*\|_1} \|\mathbf{g} - \mathbf{u}\|_2^2 \right] = \min_{t \geq 0} s_{\text{sig}}(1+t^2) + \frac{2(n-s_{\text{sig}})}{\sqrt{2\pi}} \left( (1+t^2) \int_t^\infty e^{-x^2/2} dx - te^{-t^2/2} \right)$$

and

$$\min_{t \geq 0} \mathbb{E} \left[ \inf_{\mathbf{u} \in t \cdot \partial \|\mathbf{v}^*\|_1} \|\mathbf{g} - \mathbf{u}\|_2^2 \right] = \min_{t \geq 0} s_{\text{cor}}(1+t^2) + \frac{2(m-s_{\text{cor}})}{\sqrt{2\pi}} \left( (1+t^2) \int_t^\infty e^{-x^2/2} dx - te^{-t^2/2} \right),$$

respectively.

Fig. 3 displays the empirical probability of success for each setting of  $(s_{\text{sig}}, s_{\text{cor}})$  averaged over the 20 runs. We can see that our theoretical recovery threshold closely aligns with observed phase transition under both Gaussian and Bernoulli measurements.

### B. Phase Transition of Partially Penalized Recovery Procedure

We next consider the partially penalized recovery procedure in the noiseless setting where neither  $f(\mathbf{x}) = \|\mathbf{x}^*\|_1$  nor  $g(\mathbf{v}) = \|\mathbf{v}^*\|_1$  is known beforehand. The experiment settings are almost the same as in the constrained case except that we recover signals and corruptions via the following procedure in step (4):

$$\min_{\mathbf{x}, \mathbf{v}} \|\mathbf{x}\|_1 + \lambda \|\mathbf{v}\|_1, \text{ s.t. } \mathbf{y} = \Phi \mathbf{x} + \mathbf{v}. \quad (47)$$

We test two kinds of the regularization parameter: 1)  $\lambda = \lambda^*$ , which is chosen according to (29) and depends on the signal and corruption sparsity levels  $s_{\text{sig}}$  and  $s_{\text{cor}}$ , and 2)  $\lambda = 1$ . As discussed in Section IV-D, when we choose  $\lambda = \lambda^*$ , the necessary number of measurements to guarantee success by the partially penalized procedure is nearly the same as that of the constrained one. Thus, we also overlay the curve of (46) to compare with the empirical results.

Fig. 4 shows the empirical probability of success as the signal and corruption sparsity levels  $s_{\text{sig}}$  and  $s_{\text{cor}}$  vary. We can find that the recovery performance in the case of  $\lambda = \lambda^*$  is better than that in the case of  $\lambda = 1$ . This demonstrates that the choice of the regularization parameter suggested by (29) is effective. Moreover, signal recovery with  $\lambda = \lambda^*$  is nearly as good as in the constrained case, but without any prior knowledge of  $\|\mathbf{x}^*\|_1$ .

<sup>5</sup>In [21], the second author of this paper and his collaborators have shown that the phase transition of the constrained procedure under Gaussian measurements occurs around  $\omega^2(\mathcal{T}_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1}) + \omega^2(\mathcal{T}_g(\mathbf{v}^*) \cap \mathbb{S}^{m-1})$ . A large number of numerical experiments indicate that this theoretical recovery threshold also holds for sub-Gaussian measurements. This observed universality phenomenon suggests us to set  $CK^4 = 1$  in (20) for our numerical simulations.

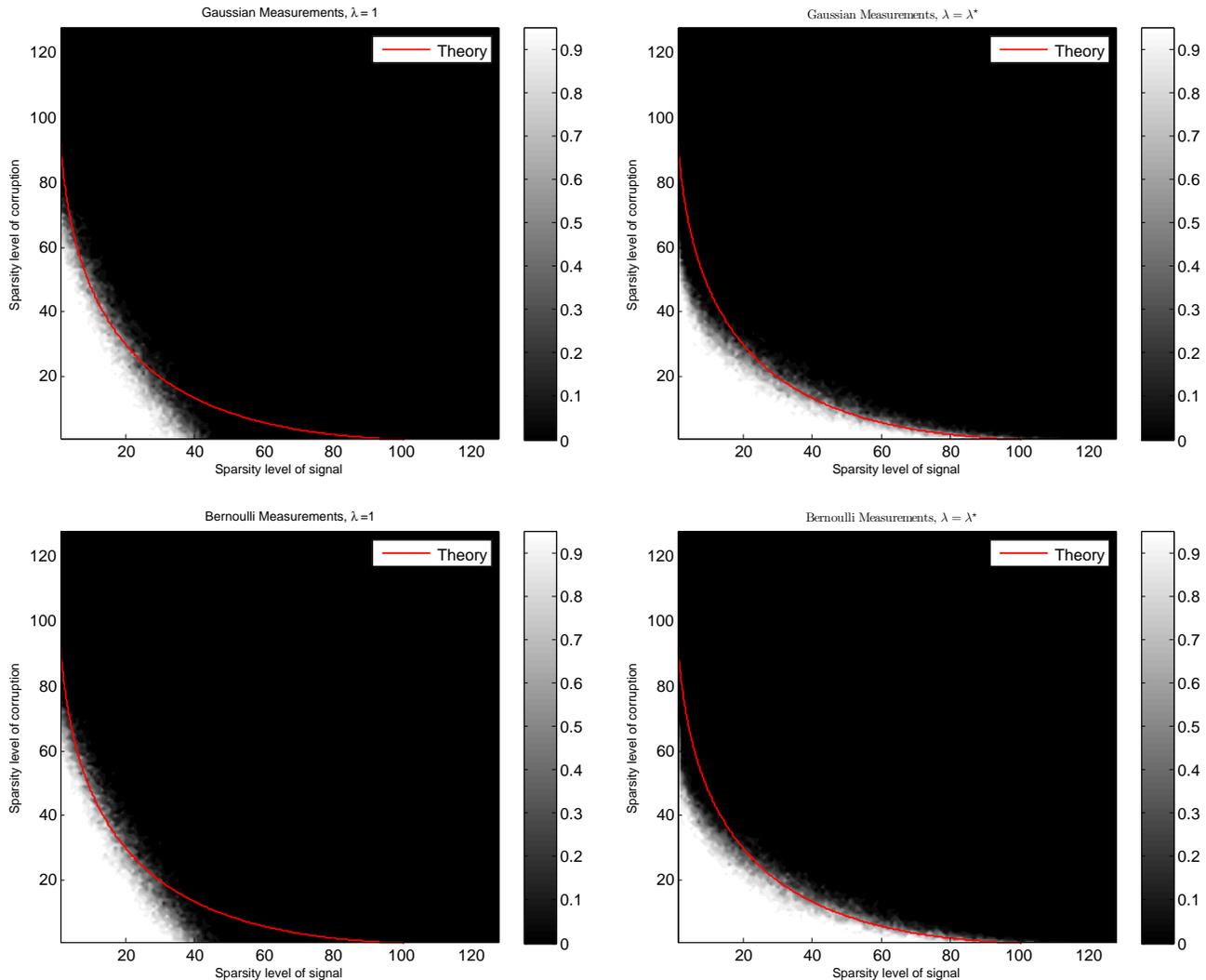


Fig. 4. Phase transition of the partially penalized recovery procedure (47) with different kinds of the regularization parameter under Gaussian and Bernoulli measurements. The red curve plots the phase transition threshold predicted by (46).

### C. Phase Transition of Fully Penalized Recovery Procedure

Thirdly, we study the empirical behavior of the fully penalized recovery procedure in the noiseless setting where neither  $f(\mathbf{x}) = \|\mathbf{x}^*\|_1$  nor  $g(\mathbf{v}) = \|\mathbf{v}^*\|_1$  is known a priori. Similarly, the experiment settings are nearly the same as in the constrained case except that we reconstruct signals and corruptions via the following procedure in step (4):

$$\min_{\mathbf{x}, \mathbf{v}} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x} - \mathbf{v}\|_2^2 + \tau_1 \|\mathbf{x}\|_1 + \tau_2 \|\mathbf{v}\|_1. \quad (48)$$

As suggested in Section IV-C2, in order to recover signals and corruptions faithfully by this procedure, we require that the regularization parameters  $\tau_1$  and  $\tau_2$  tend to zero. Thus we set  $\tau_1 = \tau_2 = 10^{-5}$ . For reference, we also overlay the curve of (46) to compare with the empirical results.

Fig. 5 illustrates the average empirical probability of success for each setting  $(s_{\text{sig}}, s_{\text{cor}})$ . We can see that the theoretical recovery threshold (46) roughly predict observed phase transition with this choice of the regularization parameters.

### D. Stable Recovery

Finally, we investigate the empirical behavior of the three procedures under noisy measurements. We fix  $m = n = 128$  and  $s_{\text{sig}} = s_{\text{cor}} = 20$ . The first three steps of this experiment are the same as that in the constrained case and the other steps are as follows:

- (4) Generate a Gaussian noise vector  $\mathbf{z}$  and scale  $\mathbf{z}$  such that  $\|\mathbf{z}\|_2 = \delta$ .
- (5) Solve constrained, partially penalized, and fully penalized procedures under different noise levels with different regularization parameters. In the partially penalized case, we choose  $\lambda = \lambda^*$  according to (29) or  $\lambda = 1$ . In the fully penalized

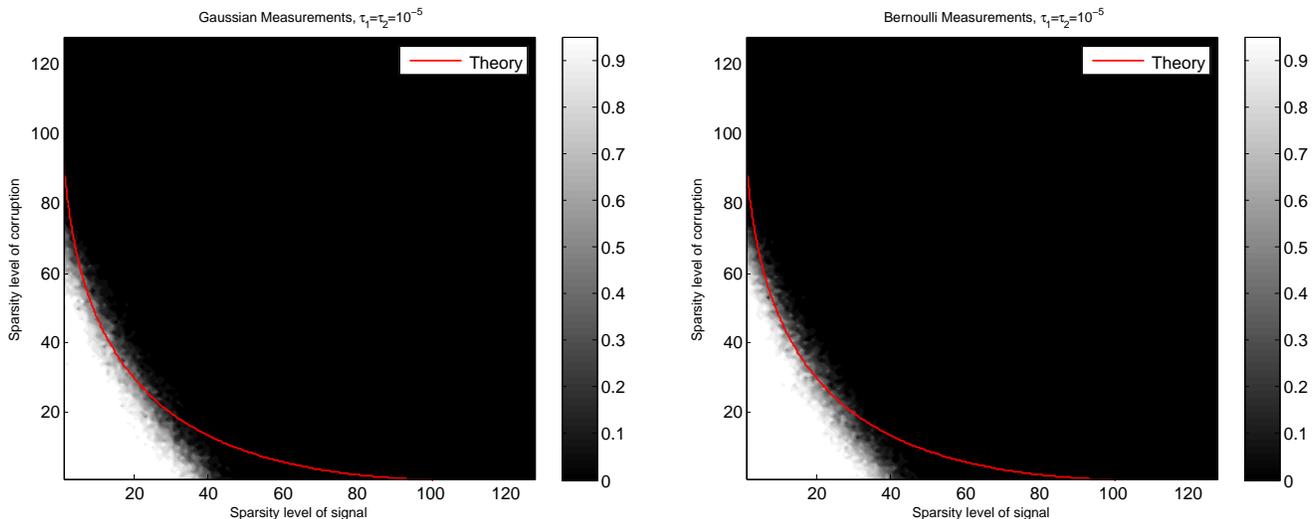


Fig. 5. Phase transition of the fully penalized recovery procedure (48) with  $\tau_1 = \tau_2 = 10^{-5}$  under Gaussian and Bernoulli measurements. The red curve plots the phase transition threshold predicted by (46).

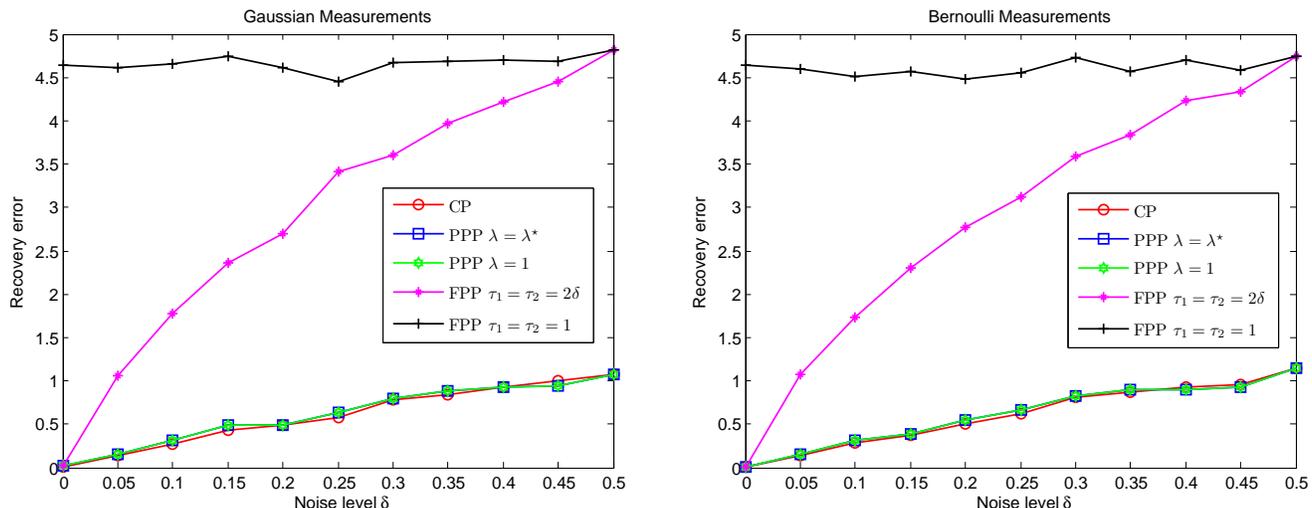


Fig. 6. Recovery error  $\sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2}$  by constrained, partially penalized, and fully penalized procedures (abbreviated as CP, PPP, and FPP, respectively) under different noise levels with different regularization parameters. Left: Gaussian measurements, right: Bernoulli measurements.

case, we choose  $\tau_1$  and  $\tau_2$  according (43) or  $\tau_1 = \tau_2 = 1$ . Here, for sparse signal and sparse corruption, (43) implies  $\tau_1 = \tau_{1B} \approx CK\beta\delta(\sqrt{n} + \sqrt{m})/\sqrt{m} = 2CK\beta\delta = C'K\beta\delta$  and  $\tau_2 = \tau_{2B} = \beta\delta$ . In our simulations, we choose  $C'K = 1$  and  $\beta = 2$ .

(6) Record the  $\ell_2$  recovery error  $\sqrt{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 + \|\hat{\mathbf{v}} - \mathbf{v}^*\|_2^2}$  in each case.

Fig. 6 shows the average recovery error across 20 runs. We can see that the recovery error increases linearly as the noise level increases in both constrained and partially penalized cases, which is consistent with our theoretical results (Theorems 2 and 3). Moreover, the partially penalized procedure with different regularization parameters shows almost the same performance as the constrained one when the sparsity levels of signal and corruption are not too high ( $s_{\text{sig}} = s_{\text{cor}} = 20$ ). This implies that when the estimate of the sparsity levels  $s_{\text{sig}}$  and  $s_{\text{cor}}$  is unavailable, the setting  $\lambda = 1$  yields high probability recovery, provided that the sparsity levels of signal and corruption are relatively low. In the fully penalized case, the recovery performance with  $\tau_1$  and  $\tau_2$  chosen according to (43) is much better than that with  $\tau_1 = \tau_2 = 1$ . This demonstrates the effectiveness of the parameter selection strategy (43).

## VI. CONCLUSION

In this paper, we have established an extended matrix deviation inequality for sub-Gaussian matrices, which provides a powerful tool to analyze the corrupted sensing problem. We then presented performance analysis for three convex recovery

procedures which are used to recover structured signals from corrupted sub-Gaussian measurements when different kinds of prior information are available. We considered both bounded and stochastic noise. Our results have shown that, under proper conditions, these approaches reconstruct both signal and corruption exactly in the absence of noise and stably in the presence of noise. Moreover, our results also indicate how to pick the regularization parameters in both partially and fully penalized recovery procedures and reveal the relationships among these methods. For future work, it would be of great interest to establish the phase transition theory for both constrained and partially penalized procedures under sub-Gaussian measurements.

## APPENDIX A

### SOME PROPERTIES OF SUB-GAUSSIAN AND SUB-EXPONENTIAL RANDOM VARIABLES AND VECTORS

**Fact 2** (Product of sub-Gaussians is sub-exponential). [25, Lemma 2.7.7] *Let  $X$  and  $Y$  be sub-Gaussian random variables (not necessarily independent). Then  $XY$  is sub-exponential. Moreover,*

$$\|XY\|_{\psi_1} \leq \|X\|_{\psi_2} \|Y\|_{\psi_2}. \quad (49)$$

**Fact 3** (Centering). [25, Lemma 2.6.8 and Exercise 2.7.10] *If  $X$  is sub-Gaussian (or sub-exponential), then so is  $X - \mathbb{E}X$ . Moreover,*

$$\|X - \mathbb{E}X\|_{\psi_2} \leq C\|X\|_{\psi_2} \quad \text{and} \quad \|X - \mathbb{E}X\|_{\psi_1} \leq C\|X\|_{\psi_1}.$$

**Fact 4** (Hoeffding-type Inequality). [25, Theorem 2.6.3] *Let  $X_1, \dots, X_m$  be independent, mean-zero, sub-Gaussian random variables, and  $\mathbf{a} = (a_1, a_2, \dots, a_m)^T \in \mathbb{R}^m$ . Then, for any  $t \geq 0$ , we have*

$$\mathbb{P} \left\{ \left| \sum_{i=1}^m a_i X_i \right| \geq t \right\} \leq 2 \exp \left\{ -\frac{ct^2}{K^2 \|\mathbf{a}\|_2^2} \right\}, \quad (50)$$

where  $K = \max_i \|X_i\|_{\psi_2}$ .

**Fact 5** (Bernstein-type Inequality). [25, Theorem 2.8.2] *Let  $X_1, X_2, \dots, X_m$  be independent, mean-zero, sub-exponential random variables, and  $\mathbf{a} = (a_1, a_2, \dots, a_m)^T \in \mathbb{R}^m$ . Then, for any  $t \geq 0$ , we have*

$$\mathbb{P} \left\{ \left| \sum_{i=1}^m a_i X_i \right| \geq t \right\} \leq 2 \exp \left\{ -c \min \left( \frac{t^2}{K^2 \|\mathbf{a}\|_2^2}, \frac{t}{K \|\mathbf{a}\|_\infty} \right) \right\}, \quad (51)$$

where  $K = \max_i \|X_i\|_{\psi_1}$ .

## APPENDIX B

### PROOF OF THEOREM 1

In order to prove Theorem 1, we need the following key lemma which states that the random process  $X_{\mathbf{a}, \mathbf{b}} := \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \sqrt{m} \cdot \sqrt{\|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2}$  has sub-Gaussian increments.

**Lemma 5** (Sub-Gaussian Process). *Let  $\mathbf{A}$  be an  $m \times n$  matrix whose rows  $\mathbf{A}_i$  are independent centered isotropic sub-Gaussian vectors. Then the random process*

$$\begin{aligned} X_{\mathbf{a}, \mathbf{b}} &= \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - (\mathbb{E} \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2^2)^{1/2} \\ &= \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \sqrt{m} \cdot \sqrt{\|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2} \end{aligned}$$

has sub-Gaussian increments:

$$\|X_{\mathbf{a}, \mathbf{b}} - X_{\mathbf{a}', \mathbf{b}'}\|_{\psi_2} \leq CK^2 \cdot \sqrt{\|\mathbf{a} - \mathbf{a}'\|_2^2 + \|\mathbf{b} - \mathbf{b}'\|_2^2} \quad \text{for every } (\mathbf{a}, \mathbf{b}), (\mathbf{a}', \mathbf{b}') \in \mathbb{R}^n \times \mathbb{R}^m, \quad (52)$$

where  $K = \max_i \|\mathbf{A}_i\|_{\psi_2}$ .

*Proof.* See Appendix C. □

Combing Lemma 5 and Talagrand's Majorizing Measure Theorem yields the proof of Theorem 1.

*Proof.* According to Lemma 5, the random process

$$X_{\mathbf{a}, \mathbf{b}} := \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \sqrt{m} \cdot \sqrt{\|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2}$$

has sub-Gaussian increments, that is

$$\begin{aligned} \|X_{\mathbf{a}, \mathbf{b}} - X_{\mathbf{a}', \mathbf{b}'}\|_{\psi_2} &\leq CK^2 \cdot \sqrt{\|\mathbf{a} - \mathbf{a}'\|_2^2 + \|\mathbf{b} - \mathbf{b}'\|_2^2} \\ &= CK^2 \left\| \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} - \begin{bmatrix} \mathbf{a}' \\ \mathbf{b}' \end{bmatrix} \right\|_2 \end{aligned}$$

for every  $(\mathbf{a}, \mathbf{b}), (\mathbf{a}', \mathbf{b}') \in \mathcal{T}$ . It follows from Talagrand's Majorizing Measure Theorem that

$$\begin{aligned} \mathbb{E} \sup_{(\mathbf{a}, \mathbf{b}), (\mathbf{a}', \mathbf{b}') \in \mathcal{T}} |X_{\mathbf{a}, \mathbf{b}} - X_{\mathbf{a}', \mathbf{b}'}| &\leq CK^2 \mathbb{E} \sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{T}} \left\langle \begin{bmatrix} \mathbf{g} \\ \mathbf{h} \end{bmatrix}, \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \right\rangle \\ &= CK^2 \omega(\mathcal{T}), \end{aligned}$$

where  $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$  and  $\mathbf{h} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$ . Thus, fix an arbitrary point  $(\mathbf{a}_0, \mathbf{b}_0) \in \mathcal{T}$  and use the triangle inequality to obtain

$$\begin{aligned} \mathbb{E} \sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{T}} |X_{\mathbf{a}, \mathbf{b}}| &\leq \mathbb{E} \sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{T}} |X_{\mathbf{a}, \mathbf{b}} - X_{\mathbf{a}_0, \mathbf{b}_0}| + \mathbb{E} |X_{\mathbf{a}_0, \mathbf{b}_0}| \\ &\leq CK^2 \omega(\mathcal{T}) + \mathbb{E} |X_{\mathbf{a}_0, \mathbf{b}_0}|. \end{aligned}$$

The second term can be bounded as follows:

$$\mathbb{E} |X_{\mathbf{a}_0, \mathbf{b}_0}| \leq C' \|X_{\mathbf{a}_0, \mathbf{b}_0}\|_{\psi_2} \leq C'' K^2 \cdot \left\| \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{b}_0 \end{bmatrix} \right\|_2.$$

The first inequality follows from the equivalent definition (10) and the second inequality holds by using Lemma 5 with  $[\mathbf{a}^T, \mathbf{b}^T]^T = \mathbf{0}$ . Therefore, by (8), we have

$$\begin{aligned} \mathbb{E} \sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{T}} |X_{\mathbf{a}, \mathbf{b}}| &\leq CK^2 \omega(\mathcal{T}) + C'' K^2 \cdot \left\| \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{b}_0 \end{bmatrix} \right\|_2 \\ &\leq C_0 K^2 \gamma(\mathcal{T}). \end{aligned}$$

This establishes the expectation bound.

For the high probability bound, define  $\bar{\mathcal{T}} = \mathcal{T} \cup \{\mathbf{0}\}$ . It also follows from Talagrand's Majorizing Measure Theorem that with probability at least  $1 - \exp(-t^2)$ ,

$$\begin{aligned} \sup_{(\mathbf{a}, \mathbf{b}) \in \mathcal{T}} |X_{\mathbf{a}, \mathbf{b}}| &= \sup_{(\mathbf{a}, \mathbf{b}) \in \bar{\mathcal{T}}} |X_{\mathbf{a}, \mathbf{b}}| = \sup_{(\mathbf{a}, \mathbf{b}) \in \bar{\mathcal{T}}} |X_{\mathbf{a}, \mathbf{b}} - X_{\mathbf{0}, \mathbf{0}}| \\ &\leq \sup_{(\mathbf{a}, \mathbf{b}), (\mathbf{a}', \mathbf{b}') \in \bar{\mathcal{T}}} |X_{\mathbf{a}, \mathbf{b}} - X_{\mathbf{a}', \mathbf{b}'}| \\ &\leq C' K^2 [\omega(\bar{\mathcal{T}}) + t \cdot \text{diam}(\bar{\mathcal{T}})] \\ &\leq CK^2 [\gamma(\mathcal{T}) + t \cdot \text{rad}(\mathcal{T})]. \end{aligned}$$

In the last inequality, we have used the facts that  $\omega(\bar{\mathcal{T}}) \leq \gamma(\bar{\mathcal{T}}) = \gamma(\mathcal{T})$  and  $\text{diam}(\bar{\mathcal{T}}) \leq 2 \text{rad}(\mathcal{T})$ . This completes the proof.  $\square$

## APPENDIX C PROOF OF LEMMA 5

Our proof of Lemma 5 is inspired by [37] and [23]. For clarity, the proof is divided into three steps. First, we show a partial case of Lemma 5 in which  $(\mathbf{a}^T, \mathbf{b}^T)^T \in \mathbb{S}^{n+m-1}$  and  $(\mathbf{a}'^T, \mathbf{b}'^T)^T = \mathbf{0}$ . We then extend this by allowing  $(\mathbf{a}'^T, \mathbf{b}'^T)^T$  to be an arbitrary unit vector. Finally, we prove the increment inequality (52) for any  $(\mathbf{a}, \mathbf{b}), (\mathbf{a}', \mathbf{b}') \in \mathbb{R}^n \times \mathbb{R}^m$ .

Since  $\mathbf{A}_i$  are isotropic sub-Gaussian random vectors, it follows from (10) that  $K$  is bounded below by an absolute constant. For simplicity, we will assume that  $K \geq 1$ .

**Step 1:**  $(\mathbf{a}^T, \mathbf{b}^T)^T \in \mathbb{S}^{n+m-1}$  and  $(\mathbf{a}'^T, \mathbf{b}'^T)^T = \mathbf{0}$ . In this case, we have the following result.

**Lemma 6.** *Let  $\mathbf{A}$  be a sub-Gaussian random matrix as in Lemma 5. Then*

$$\left\| \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \sqrt{m} \right\|_{\psi_2} \leq CK^2 \quad \text{for every } [\mathbf{a}^T, \mathbf{b}^T]^T \in \mathbb{S}^{n+m-1}. \quad (53)$$

*Proof.* We begin with establishing a concentration inequality for  $\frac{1}{m}\|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2^2 - 1$ . For any  $t \geq 0$ , we have

$$\begin{aligned}
p &:= \mathbb{P} \left\{ \left| \frac{1}{m}\|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2^2 - 1 \right| \geq t \right\} \\
&= \mathbb{P} \left\{ \left| \frac{1}{m}\|\mathbf{A}\mathbf{a}\|_2^2 + \frac{2}{\sqrt{m}}\langle \mathbf{A}\mathbf{a}, \mathbf{b} \rangle + \|\mathbf{b}\|_2^2 - 1 \right| \geq t \right\} \\
&= \mathbb{P} \left\{ \left| \frac{1}{m}\|\mathbf{A}\mathbf{a}\|_2^2 - \|\mathbf{a}\|_2^2 + \frac{2}{\sqrt{m}}\langle \mathbf{A}\mathbf{a}, \mathbf{b} \rangle \right| \geq t \right\} \\
&\leq \mathbb{P} \left\{ \left| \frac{1}{m}\|\mathbf{A}\mathbf{a}\|_2^2 - \|\mathbf{a}\|_2^2 \right| + \left| \frac{2}{\sqrt{m}}\langle \mathbf{A}\mathbf{a}, \mathbf{b} \rangle \right| \geq t \right\} \\
&\leq \mathbb{P} \left\{ \left| \frac{1}{m}\|\mathbf{A}\mathbf{a}\|_2^2 - \|\mathbf{a}\|_2^2 \right| \geq \frac{t}{2} \right\} + \mathbb{P} \left\{ \left| \frac{2}{\sqrt{m}}\langle \mathbf{A}\mathbf{a}, \mathbf{b} \rangle \right| \geq \frac{t}{2} \right\} \\
&=: p_1 + p_2.
\end{aligned} \tag{54}$$

To bound  $p_1$ , it will be useful to express  $\frac{1}{m}\|\mathbf{A}\mathbf{a}\|_2^2 - \|\mathbf{a}\|_2^2$  as a sum of independent random variables  $\frac{1}{m}\sum_{i=1}^m [\langle \mathbf{A}_i^T, \mathbf{a} \rangle^2 - \|\mathbf{a}\|_2^2] := \frac{1}{m}\sum_{i=1}^m Z_i$ . By assumption,  $\{\langle \mathbf{A}_i^T, \mathbf{a} \rangle\}_{i=1}^m$  are independent centered sub-Gaussian random variables with  $\mathbb{E}\langle \mathbf{A}_i^T, \mathbf{a} \rangle^2 = \|\mathbf{a}\|_2^2$  and  $\|\langle \mathbf{A}_i^T, \mathbf{a} \rangle\|_{\psi_2} \leq K\|\mathbf{a}\|_2 \leq K$ . Therefore, by facts 3 and 2,  $\{Z_i\}_{i=1}^m$  are independent centered sub-exponential random variables with

$$\begin{aligned}
\|Z_i\|_{\psi_1} &= \|\langle \mathbf{A}_i^T, \mathbf{a} \rangle^2 - \mathbb{E}\langle \mathbf{A}_i^T, \mathbf{a} \rangle^2\|_{\psi_1} \\
&\leq C\|\langle \mathbf{A}_i^T, \mathbf{a} \rangle^2\|_{\psi_1} \quad \text{by Fact 3} \\
&\leq C\|\langle \mathbf{A}_i^T, \mathbf{a} \rangle\|_{\psi_2}^2 \quad \text{by Fact 2} \\
&\leq CK^2.
\end{aligned}$$

It follows from Bernstein's inequality (Fact 5) that

$$\begin{aligned}
p_1 &= \mathbb{P} \left\{ \left| \frac{1}{m}\sum_{i=1}^m Z_i \right| \geq \frac{t}{2} \right\} \\
&\leq 2 \exp \left\{ -c \min\left(\frac{t^2}{4C^2K^4/m}, \frac{t}{2CK^2/m}\right) \right\} \\
&= 2 \exp \left\{ -\frac{cm}{4C^2K^4} \min(t^2, 2CK^2t) \right\} \\
&\leq 2 \exp \left\{ -\frac{c_1m}{K^4} \min(t^2, t) \right\}.
\end{aligned} \tag{55}$$

The last inequality holds because we can easily choose  $C$  such that  $2CK^2 \geq 1$ .

To bound  $p_2$ , it will be helpful to write  $\langle \mathbf{A}\mathbf{a}, \mathbf{b} \rangle = \sum_{i=1}^m \mathbf{b}_i \langle \mathbf{A}_i^T, \mathbf{a} \rangle$ , where  $\{\langle \mathbf{A}_i^T, \mathbf{a} \rangle\}_{i=1}^m$  are independent centered sub-Gaussian random variable with  $\|\langle \mathbf{A}_i^T, \mathbf{a} \rangle\|_{\psi_2} \leq K$ . Applying Hoeffding's inequality (Fact 4) yields

$$\begin{aligned}
p_2 &= \mathbb{P} \left\{ \left| \frac{1}{\sqrt{m}}\sum_{i=1}^m \mathbf{b}_i \langle \mathbf{A}_i^T, \mathbf{a} \rangle \right| \geq \frac{t}{4} \right\} \\
&\leq 2 \exp \left\{ -\frac{cmt^2}{16K^2\|\mathbf{b}\|_2^2} \right\} \\
&\leq 2 \exp \left\{ -\frac{c_2mt^2}{K^4} \right\}.
\end{aligned} \tag{56}$$

The last inequality holds because  $\|\mathbf{b}\|_2 \leq 1$  and  $K \geq 1$ .

Combining the upper bounds of  $p_1$  and  $p_2$  yields

$$\begin{aligned}
p &\leq 2 \exp \left\{ -\frac{c_1m}{K^4} \min(t^2, t) \right\} + 2 \exp \left\{ -\frac{c_2mt^2}{K^4} \right\} \\
&\leq 4 \exp \left\{ -\frac{c_0m}{K^4} \min(t^2, t) \right\}.
\end{aligned} \tag{57}$$

We then establish a concentration inequality for  $\frac{1}{\sqrt{m}}\|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\| - 1$ . To this end, we will use the following fact

$$|z - 1| \geq \delta \quad \text{implies} \quad |z^2 - 1| \geq \max(\delta^2, \delta), \quad \text{for any } z \geq 0 \quad \text{and } \delta \geq 0.$$

Using this for  $z = \frac{1}{\sqrt{m}}\|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|$  together with (57) where  $t = \max\{\delta^2, \delta\}$ , we obtain for any  $\delta \geq 0$  that

$$\begin{aligned} \mathbb{P} \left\{ \left| \frac{1}{\sqrt{m}}\|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - 1 \right| \geq \delta \right\} &\leq \mathbb{P} \left\{ \left| \frac{1}{m}\|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2^2 - 1 \right| \geq \max(\delta^2, \delta) \right\} \\ &\leq 4 \exp \left\{ -\frac{c_0 m \delta^2}{K^4} \right\}. \end{aligned}$$

This implies  $\|\|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \sqrt{m}\|_{\psi_2} \leq CK^2$  and completes the proof.  $\square$

**Step 2:**  $(\mathbf{a}^T, \mathbf{b}^T)^T \in \mathbb{S}^{n+m-1}$  and  $(\mathbf{a}'^T, \mathbf{b}'^T)^T \in \mathbb{S}^{n+m-1}$ . In this case, we have the following result.

**Lemma 7.** *Let  $\mathbf{A}$  be a sub-Gaussian random matrix as in Lemma 5. Then*

$$\left\| \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \|\mathbf{A}\mathbf{a}' + \sqrt{m}\mathbf{b}'\|_2 \right\|_{\psi_2} \leq CK^2 \cdot \sqrt{\|\mathbf{a} - \mathbf{a}'\|_2^2 + \|\mathbf{b} - \mathbf{b}'\|_2^2},$$

for every  $(\mathbf{a}^T, \mathbf{b}^T)^T, (\mathbf{a}'^T, \mathbf{b}'^T)^T \in \mathbb{S}^{n+m-1}$ .

*Proof.* Given  $t \geq 0$ , it suffices to show

$$p := \mathbb{P} \left\{ \frac{\left| \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \|\mathbf{A}\mathbf{a}' + \sqrt{m}\mathbf{b}'\|_2 \right|}{\sqrt{\|\mathbf{a} - \mathbf{a}'\|_2^2 + \|\mathbf{b} - \mathbf{b}'\|_2^2}} \geq t \right\} \leq C \exp \left\{ -\frac{ct^2}{K^4} \right\}. \quad (58)$$

We will proceed differently for small and large  $t$ .

**Case 1:**  $t \geq 2\sqrt{m}$ . Denote

$$\mathbf{u} := \frac{\mathbf{a} - \mathbf{a}'}{\sqrt{\|\mathbf{a} - \mathbf{a}'\|_2^2 + \|\mathbf{b} - \mathbf{b}'\|_2^2}} \quad \text{and} \quad \mathbf{v} := \frac{\mathbf{b} - \mathbf{b}'}{\sqrt{\|\mathbf{a} - \mathbf{a}'\|_2^2 + \|\mathbf{b} - \mathbf{b}'\|_2^2}}.$$

It follows from the triangle inequality that

$$\begin{aligned} p &\leq \mathbb{P} \left\{ \frac{\|\mathbf{A}(\mathbf{a} - \mathbf{a}') + \sqrt{m}(\mathbf{b} - \mathbf{b}')\|_2}{\sqrt{\|\mathbf{a} - \mathbf{a}'\|_2^2 + \|\mathbf{b} - \mathbf{b}'\|_2^2}} \geq t \right\} \\ &= \mathbb{P} \left\{ \|\mathbf{A}\mathbf{u} + \sqrt{m}\mathbf{v}\|_2 \geq t \right\} \\ &= \mathbb{P} \left\{ \|\mathbf{A}\mathbf{u} + \sqrt{m}\mathbf{v}\|_2 - \sqrt{m} \geq t - \sqrt{m} \right\} \\ &\leq \mathbb{P} \left\{ \|\mathbf{A}\mathbf{u} + \sqrt{m}\mathbf{v}\|_2 - \sqrt{m} \geq \frac{t}{2} \right\} \\ &\leq 2 \exp \left\{ -\frac{ct^2}{K^4} \right\}. \end{aligned}$$

The second inequality holds because  $t \geq 2\sqrt{m}$  and the last inequality follows from Lemma 6.

**Case 2:**  $t \leq 2\sqrt{m}$ . Denote

$$\mathbf{u}' = \mathbf{a} + \mathbf{a}' \quad \text{and} \quad \mathbf{v}' = \mathbf{b} + \mathbf{b}'.$$

Multiplying both sides of the inequality defining  $p$  in (58) by  $\|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 + \|\mathbf{A}\mathbf{a}' + \sqrt{m}\mathbf{b}'\|_2$  yields

$$\begin{aligned} p &= \mathbb{P} \left\{ \left| \frac{\|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2^2 - \|\mathbf{A}\mathbf{a}' + \sqrt{m}\mathbf{b}'\|_2^2}{\sqrt{\|\mathbf{a} - \mathbf{a}'\|_2^2 + \|\mathbf{b} - \mathbf{b}'\|_2^2}} \right| \geq t(\|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 + \|\mathbf{A}\mathbf{a}' + \sqrt{m}\mathbf{b}'\|_2) \right\} \\ &\leq \mathbb{P} \left\{ \left| \frac{\langle \mathbf{A}(\mathbf{a} - \mathbf{a}') + \sqrt{m}(\mathbf{b} - \mathbf{b}'), \mathbf{A}(\mathbf{a} + \mathbf{a}') + \sqrt{m}(\mathbf{b} + \mathbf{b}') \rangle}{\sqrt{\|\mathbf{a} - \mathbf{a}'\|_2^2 + \|\mathbf{b} - \mathbf{b}'\|_2^2}} \right| \geq t \cdot \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 \right\} \\ &= \mathbb{P} \left\{ \left| \langle \mathbf{A}\mathbf{u} + \sqrt{m}\mathbf{v}, \mathbf{A}\mathbf{u}' + \sqrt{m}\mathbf{v}' \rangle \right| \geq t \cdot \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 \right\}. \end{aligned}$$

Define the event  $\Omega_1$  as

$$\Omega_1 := \left\{ \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 \geq \frac{\sqrt{m}}{2} \right\}.$$

By the law of total probability, we have

$$\begin{aligned} p &\leq \mathbb{P} \left\{ \underbrace{\left| \langle \mathbf{A}\mathbf{u} + \sqrt{m}\mathbf{v}, \mathbf{A}\mathbf{u}' + \sqrt{m}\mathbf{v}' \rangle \right|}_{:=\Omega_0} \geq t \cdot \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 \right\} \\ &= \mathbb{P} \left\{ \Omega_0 \mid \Omega_1 \right\} \cdot \mathbb{P} \left\{ \Omega_1 \right\} + \mathbb{P} \left\{ \Omega_0 \mid \Omega_1^c \right\} \cdot \mathbb{P} \left\{ \Omega_1^c \right\} \\ &\leq \mathbb{P} \left\{ \Omega_0 \text{ and } \Omega_1 \right\} + \mathbb{P} \left\{ \Omega_1^c \right\} \\ &:= p_1 + p_2. \end{aligned}$$

We can easily bound  $p_2$  using Lemma 6

$$\begin{aligned}
p_2 &= \mathbb{P} \{ \Omega_1^c \} = \mathbb{P} \left\{ \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 \leq \frac{\sqrt{m}}{2} \right\} \\
&= \mathbb{P} \left\{ \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \sqrt{m} \leq -\frac{\sqrt{m}}{2} \right\} \\
&\leq \mathbb{P} \left\{ \left| \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \sqrt{m} \right| \geq \frac{\sqrt{m}}{2} \right\} \\
&\leq \mathbb{P} \left\{ \left| \|\mathbf{A}\mathbf{a} + \sqrt{m}\mathbf{b}\|_2 - \sqrt{m} \right| \geq \frac{t}{4} \right\} \quad \text{since } t \leq 2\sqrt{m} \\
&\leq 2 \exp \left\{ -\frac{c_1 t^2}{16K^4} \right\} \quad \text{since Lemma 6} \\
&\leq 2 \exp \left\{ -\frac{c'' t^2}{K^4} \right\}.
\end{aligned}$$

To bound  $p_1$ , note that

$$\begin{aligned}
p_1 &:= \mathbb{P} \{ \Omega_0 \text{ and } \Omega_1 \} \\
&\leq \mathbb{P} \left\{ \left| \langle \mathbf{A}\mathbf{u} + \sqrt{m}\mathbf{v}, \mathbf{A}\mathbf{u}' + \sqrt{m}\mathbf{v}' \rangle \right| \geq \frac{t\sqrt{m}}{2} \right\} \\
&= \mathbb{P} \left\{ \left| \langle \mathbf{A}\mathbf{u}, \mathbf{A}\mathbf{u}' \rangle + m \langle \mathbf{v}, \mathbf{v}' \rangle + \sqrt{m} \langle \mathbf{A}\mathbf{u}, \mathbf{v}' \rangle + \sqrt{m} \langle \mathbf{A}\mathbf{u}', \mathbf{v} \rangle \right| \geq \frac{t\sqrt{m}}{2} \right\} \\
&\leq \mathbb{P} \left\{ \left| \langle \mathbf{A}\mathbf{u}, \mathbf{A}\mathbf{u}' \rangle + m \langle \mathbf{v}, \mathbf{v}' \rangle \right| \geq \frac{t\sqrt{m}}{4} \right\} + \mathbb{P} \left\{ \left| \sqrt{m} \langle \mathbf{A}\mathbf{u}, \mathbf{v}' \rangle + \sqrt{m} \langle \mathbf{A}\mathbf{u}', \mathbf{v} \rangle \right| \geq \frac{t\sqrt{m}}{4} \right\} \\
&\leq \mathbb{P} \left\{ \left| \langle \mathbf{A}\mathbf{u}, \mathbf{A}\mathbf{u}' \rangle + m \langle \mathbf{v}, \mathbf{v}' \rangle \right| \geq \frac{t\sqrt{m}}{4} \right\} + \mathbb{P} \left\{ \left| \sqrt{m} \langle \mathbf{A}\mathbf{u}, \mathbf{v}' \rangle \right| \geq \frac{t\sqrt{m}}{8} \right\} + \mathbb{P} \left\{ \left| \sqrt{m} \langle \mathbf{A}\mathbf{u}', \mathbf{v} \rangle \right| \geq \frac{t\sqrt{m}}{8} \right\} \\
&=: p_{1a} + p_{1b} + p_{1c}.
\end{aligned}$$

We bound the three terms separately. To bound  $p_{1a}$ , note that

$$\begin{aligned}
\langle \mathbf{v}, \mathbf{v}' \rangle &= \frac{\langle \mathbf{b} - \mathbf{b}', \mathbf{b} + \mathbf{b}' \rangle}{\sqrt{\|\mathbf{a} - \mathbf{a}'\|_2^2 + \|\mathbf{b} - \mathbf{b}'\|_2^2}} \\
&= \frac{\|\mathbf{b}\|_2^2 - \|\mathbf{b}'\|_2^2}{\sqrt{\|\mathbf{a} - \mathbf{a}'\|_2^2 + \|\mathbf{b} - \mathbf{b}'\|_2^2}} \\
&= -\frac{\|\mathbf{a}\|_2^2 - \|\mathbf{a}'\|_2^2}{\sqrt{\|\mathbf{a} - \mathbf{a}'\|_2^2 + \|\mathbf{b} - \mathbf{b}'\|_2^2}} \\
&= -\langle \mathbf{u}, \mathbf{u}' \rangle.
\end{aligned}$$

Therefore, we can bound  $p_{1a}$  as  $p_1$  (55) in Lemma 6

$$\begin{aligned}
p_{1a} &= \mathbb{P} \left\{ \left| \langle \mathbf{A}\mathbf{u}, \mathbf{A}\mathbf{u}' \rangle + m \langle \mathbf{v}, \mathbf{v}' \rangle \right| \geq \frac{t\sqrt{m}}{4} \right\} \\
&= \mathbb{P} \left\{ \left| \langle \mathbf{A}\mathbf{u}, \mathbf{A}\mathbf{u}' \rangle - m \langle \mathbf{u}, \mathbf{u}' \rangle \right| \geq \frac{t\sqrt{m}}{4} \right\} \\
&= \mathbb{P} \left\{ \left| \sum_{i=1}^m [\langle \mathbf{A}_i^T, \mathbf{u} \rangle \langle \mathbf{A}_i^T, \mathbf{u}' \rangle - \langle \mathbf{u}, \mathbf{u}' \rangle] \right| \geq \frac{t\sqrt{m}}{4} \right\} \\
&= \mathbb{P} \left\{ \left| \sum_{i=1}^m Z_i \right| \geq \frac{t\sqrt{m}}{4} \right\}.
\end{aligned}$$

By assumption, it follows from facts 3 and 2 that  $\{Z_i\}_{i=1}^m$  are independent, mean-zero, sub-exponential random variables with

$$\begin{aligned}
\|Z_i\|_{\psi_1} &= \|\langle \mathbf{A}_i^T, \mathbf{u} \rangle \langle \mathbf{A}_i^T, \mathbf{u}' \rangle - \langle \mathbf{u}, \mathbf{u}' \rangle\|_{\psi_1} \\
&\leq C' \|\langle \mathbf{A}_i^T, \mathbf{u} \rangle \langle \mathbf{A}_i^T, \mathbf{u}' \rangle\|_{\psi_1} \\
&\leq C' \|\langle \mathbf{A}_i^T, \mathbf{u} \rangle\|_{\psi_2} \cdot \|\langle \mathbf{A}_i^T, \mathbf{u}' \rangle\|_{\psi_2} \\
&\leq C' K^2 \|\mathbf{u}\|_2 \cdot \|\mathbf{u}'\|_2 \leq 2C' K^2 = CK^2.
\end{aligned}$$

Using Bernstein's inequality (Fact 5) yields

$$\begin{aligned}
p_{1a} &\leq 2 \exp \left\{ -c \min \left( \frac{t^2}{16C^2K^4}, \frac{\sqrt{mt}}{4CK^2} \right) \right\} \\
&\leq 2 \exp \left\{ -\frac{c}{16C^2K^4} \min (t^2, 4CK^2t\sqrt{m}) \right\} \\
&\leq 2 \exp \left\{ -\frac{c}{16C^2K^4} \min (t^2, 2CK^2t^2) \right\} \quad \text{since } t \leq 2\sqrt{m} \\
&\leq 2 \exp \left\{ -\frac{ct^2}{16C^2K^4} \right\} \quad \text{choose } C \text{ such that } 2CK^2 \geq 1 \\
&\leq 2 \exp \left\{ -\frac{c_1t^2}{K^4} \right\}.
\end{aligned}$$

Similar to  $p_2$  (56) in Lemma 6, we can easily obtain  $p_{1b} \leq 2 \exp(-c_2t^2/K^4)$  and  $p_{1c} \leq 2 \exp(-c_3t^2/K^4)$ . Combining  $p_{1a}, p_{1b}$ , and  $p_{1c}$  yields

$$\begin{aligned}
p_1 &\leq 2 \exp \left\{ -\frac{c_1t^2}{K^4} \right\} + 2 \exp \left\{ -\frac{c_2t^2}{K^4} \right\} + 2 \exp \left\{ -\frac{c_3t^2}{K^4} \right\} \\
&\leq 6 \exp \left\{ -\frac{c't^2}{K^4} \right\}.
\end{aligned}$$

Therefore, we have

$$p \leq p_1 + p_2 \leq C \exp \left\{ -\frac{ct^2}{K^4} \right\}.$$

This completes the proof.  $\square$

**Step 3:**  $(\mathbf{a}, \mathbf{b}), (\mathbf{a}', \mathbf{b}') \in \mathbb{R}^n \times \mathbb{R}^m$ .

Finally, we prove the increment inequality (52) in full generality. Without loss of generality, we assume that  $\|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2 = 1$  and  $\|\mathbf{a}'\|_2^2 + \|\mathbf{b}'\|_2^2 \geq 1$ . Define the unit vector  $(\bar{\mathbf{a}}'^T, \bar{\mathbf{b}}'^T)^T$  with

$$\bar{\mathbf{a}}' := \frac{\mathbf{a}'}{\sqrt{\|\mathbf{a}'\|_2^2 + \|\mathbf{b}'\|_2^2}} \quad \text{and} \quad \bar{\mathbf{b}}' := \frac{\mathbf{b}'}{\sqrt{\|\mathbf{a}'\|_2^2 + \|\mathbf{b}'\|_2^2}}.$$

Then we have

$$\begin{aligned}
\|X_{\mathbf{a}, \mathbf{b}} - X_{\mathbf{a}', \mathbf{b}'}\|_{\psi_2} &\leq \|X_{\mathbf{a}, \mathbf{b}} - X_{\bar{\mathbf{a}}', \bar{\mathbf{b}}'}\|_{\psi_2} + \|X_{\bar{\mathbf{a}}', \bar{\mathbf{b}}'} - X_{\mathbf{a}', \mathbf{b}'}\|_{\psi_2} \\
&=: R_1 + R_2.
\end{aligned}$$

By Lemma 7,  $R_1 \leq CK^2 \cdot \sqrt{\|\mathbf{a} - \bar{\mathbf{a}}'\|_2^2 + \|\mathbf{b} - \bar{\mathbf{b}}'\|_2^2}$ . Since  $(\bar{\mathbf{a}}', \bar{\mathbf{b}}')$  and  $(\mathbf{a}', \mathbf{b}')$  are colinear, we have

$$\begin{aligned}
R_2 &= \|X_{\bar{\mathbf{a}}', \bar{\mathbf{b}}'} - X_{\mathbf{a}', \mathbf{b}'}\|_{\psi_2} \\
&= \sqrt{\|\mathbf{a}' - \bar{\mathbf{a}}'\|_2^2 + \|\mathbf{b}' - \bar{\mathbf{b}}'\|_2^2} \cdot \|X_{\bar{\mathbf{a}}', \bar{\mathbf{b}}'}\|_{\psi_2} \\
&\leq CK^2 \cdot \sqrt{\|\mathbf{a}' - \bar{\mathbf{a}}'\|_2^2 + \|\mathbf{b}' - \bar{\mathbf{b}}'\|_2^2}.
\end{aligned}$$

The last inequality follows from Lemma 6.

Combining  $R_1$  and  $R_2$  yields

$$R_1 + R_2 \leq \sqrt{2}CK^2 \cdot \sqrt{\|\mathbf{a} - \mathbf{a}'\|_2^2 + \|\mathbf{b} - \mathbf{b}'\|_2^2},$$

where we have used the fact [25, Exercise 9.1.7] that

$$\sqrt{\|\mathbf{a} - \bar{\mathbf{a}}'\|_2^2 + \|\mathbf{b} - \bar{\mathbf{b}}'\|_2^2} + \sqrt{\|\bar{\mathbf{a}}' - \mathbf{a}'\|_2^2 + \|\bar{\mathbf{b}}' - \mathbf{b}'\|_2^2} \leq \sqrt{2} \sqrt{\|\mathbf{a} - \mathbf{a}'\|_2^2 + \|\mathbf{b} - \mathbf{b}'\|_2^2}.$$

This completes the proof of Lemma 5.

#### ACKNOWLEDGMENT

The authors thank Huan Zhang and Xu Zhang for helpful discussions.

## REFERENCES

- [1] J. Chen and Y. Liu, "Corrupted sensing with sub-gaussian measurements," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*. Aachen, Germany: IEEE, 2017, pp. 516–520.
- [2] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.
- [3] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami Beach, FL, 2009, pp. 2790–2797.
- [4] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 92–101, 2008.
- [5] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM J. Optim.*, vol. 21, no. 2, pp. 572–596, 2011.
- [6] J. Wright and Y. Ma, "Dense error correction via  $\ell_1$ -minimization," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3540–3560, 2010.
- [7] X. Li, "Compressed sensing and matrix completion with constant proportion of corruptions," *Construct. Approximation*, vol. 37, no. 1, pp. 73–99, 2013.
- [8] N. H. Nguyen and T. D. Tran, "Exact recoverability from dense corrupted observations via-minimization," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2017–2035, 2013.
- [9] —, "Robust lasso with missing and grossly corrupted observations," *IEEE Trans. Inf. Theory*, vol. 4, no. 59, pp. 2036–2058, 2013.
- [10] P. Kuppinger, G. Durisi, and H. Bolcskei, "Uncertainty relations and sparse signal recovery for pairs of general signal sets," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 263–277, 2012.
- [11] C. Studer, P. Kuppinger, G. Pope, and H. Bolcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3115–3130, 2012.
- [12] G. Pope, A. Bracher, and C. Studer, "Probabilistic recovery guarantees for sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3104–3116, 2013.
- [13] C. Studer and R. G. Baraniuk, "Stable restoration and separation of approximately sparse signals," *Appl. Comp. Harmonic Anal.*, vol. 37, no. 1, pp. 12–35, 2014.
- [14] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.
- [15] H. Xu, C. Caramanis, and S. Sanghavi, "Robust pca via outlier pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3047–3064, 2012.
- [16] H. Xu, C. Caramanis, and S. Mannor, "Outlier-robust pca: the high-dimensional case," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 546–572, 2013.
- [17] R. Foygel and L. Mackey, "Corrupted sensing: Novel guarantees for separating structured signals," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1223–1247, 2014.
- [18] M. B. McCoy and J. A. Tropp, "Sharp recovery bounds for convex demixing, with applications," *Found. Comput. Math.*, vol. 14, no. 3, pp. 503–567, 2014.
- [19] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Found. Comput. Math.*, vol. 12, no. 6, pp. 805–849, 2012.
- [20] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Inf. Inference*, no. 3, pp. 224–294.
- [21] H. Zhang, Y. Liu, and H. Lei, "On the phase transition of corrupted sensing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*. Aachen, Germany: IEEE, 2017, pp. 521–525.
- [22] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I: Fundamentals*. New York: Springer-Verlag, 1993.
- [23] C. Liaw, A. Mehrabian, Y. Plan, and R. Vershynin, "A simple tool for bounding the deviation of random matrices on geometric sets," in *Geometric Aspects of Functional Analysis*. Springer, 2017, pp. 277–299.
- [24] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," in *Compressed Sensing: Theory and Applications*, Y. Eldar and G. Kutyniok, Eds. Cambridge, U.K.: Cambridge Univ. Press., 2012, pp. 210–268.
- [25] —, *High-Dimensional Probability An Introduction with Applications in Data Science*. Book draft, available at: <https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.pdf>, 2017.
- [26] M. Talagrand, *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes*. New York, NY, USA: Springer-Verlag, 2005.
- [27] S. Dirksen, "Tail bounds via generic chaining," *Electron. J. Probab.*, vol. 20, no. 53, pp. 1–29, 2015.
- [28] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, "Simultaneous analysis of lasso and dantzig selector," *Ann. Statist.*, vol. 37, no. 4, pp. 1705–1732, 2009.
- [29] B. Klartag and S. Mendelson, "Empirical processes and random projections," *J. Funct. Anal.*, vol. 225, no. 1, pp. 229–245, 2005.
- [30] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Reconstruction and subgaussian operators in asymptotic geometric analysis," *Geom. Funct. Anal.*, vol. 17, no. 4, pp. 1248–1282, 2007.
- [31] C. Thrampoulidis, S. Oymak, and B. Hassibi, "Recovering structured signals in noise: least-squares meets compressed sensing," in *Compressed Sensing and its Applications*. Springer, 2015, pp. 97–141.
- [32] Y. Gordon, "On Milman's inequality and random subspaces which escape through a mesh in  $\mathbb{R}^n$ ." New York, NY, USA: Springer-Verlag, 1988, pp. 84–106.
- [33] S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu, "A unified framework for high-dimensional analysis of  $m$ -estimators with decomposable regularizers," *Statist. Sci.*, vol. 27, no. 4, pp. 538–557, 11 2012.
- [34] R. T. Rockafellar, *Convex analysis*. Princeton university press, 2015.
- [35] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," <http://cvxr.com/cvx>, Mar. 2014.
- [36] —, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110.
- [37] G. Schechtman, "Two observations regarding embedding subsets of euclidean spaces in normed spaces," *Adv. Math.*, vol. 200, pp. 125–135, 2006.