Stable Recovery and Separation of Structured Signals by Convex Geometry

NINGNING LI AND WENGU CHEN
Institute of Applied Physics and Computational Mathematics, Beijing 100088, China
Corresponding author: Wengu Chen (chenwg@iapcm.ac.cn)

This work was supported in part by NSF of China under Grant 11871109, in part by CAEP Foundation under Grant CX20200027, and in part by the Key Laboratory of Computational Physics Foundation under Grant 6142A05210502.

ABSTRACT We consider the ill-posed problem of identifying the signal that is structured in some general dictionary (i.e., possibly redundant or over-complete matrix) from corrupted measurements, where the corruption is structured in another general dictionary. We formulate the problem by applying appropriate convex constraints to the signal and corruption according to their structures and provide conditions for exact recovery from structured corruption and stable recovery from structured corruption with added stochastic bounded noise. In addition, this paper provides estimates of the number of the measurements needed for recovery. These estimates are based on computing the Gaussian complexity of a tangent cone and the Gaussian distance to a subdifferential. Applications covered by the proposed programs include the recovery of signals that is disturbed by impulse noise, missing entries, sparse outliers, random bounded noise, and signal separation. Numerical simulation results are presented to verify and complement the theoretical results.

INDEX TERMS Convex optimization, corrupted sensing, geometry measure, signal restoration, signal separation, tangent cone.

I. INTRODUCTION
We study the recovery problem of a signal $\mathbf{x} \in \mathbb{R}^p$ from the corrupted measurements

$$\mathbf{y} = \Phi \mathbf{x} + \Psi \mathbf{v} + \mathbf{e},$$

where $\mathbf{e} \in \mathbb{R}^n$ is a bounded noise with $\|\mathbf{e}\|_2 \leq \delta$, $\Phi \in \mathbb{R}^{n \times p}$ and $\Psi \in \mathbb{R}^{n \times q}$ are general dictionaries, such as bases, redundant or over-complete matrices. In the corrupted sensing, the vector $\mathbf{x}$ is a structured signal that we desire to recover, $\mathbf{v} \in \mathbb{R}^q$ represents structured noise. The model (1) also allows us to consider signal separation. In the signal separation setting, $\Phi \mathbf{x}$ and $\Psi \mathbf{v}$ are two distinct features from noisy measurements $\mathbf{y}$ and vectors $\mathbf{x}$ and $\mathbf{v}$ need to be simultaneously recovered.

In the classical compressed sensing, the fundamental problem is to consider the recovery of sparse signals from a relatively small number of linear measurements

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{e},$$

where $\Phi$ is a known sensing matrix, $\mathbf{x}$ is an unknown sparse signal and $\mathbf{e}$ is a measurement error with $\|\mathbf{e}\|_2 \leq \delta$. Recovery guarantees for the model (2) have been successfully developed in [7], [8], [10], [12], [15], [19], [27], [46], [47], [49]. A well studied example in the literature is the recovery of sparse vectors via $l_1$ norm minimization and we can refer to [6], [11], [16], [18], [20], [21], [39]. The work in [13] provided a general framework for recovering an arbitrary structured vector $\mathbf{x} \in \mathbb{R}^p$ from $n$ unreliable measurements $\mathbf{y} = \Phi \mathbf{x} + \mathbf{e}$ via convex regularizer, where $\Phi$ is a Gaussian matrix and $\|\mathbf{e}\|_2 \leq \delta$. The convex regularizer formulation is based on minimizing the norm induced by the convex hull of the atomic set and this norm is referred to as the atomic norm. They also obtained estimates of the number of generic measurements required for exact and robust recovery. These estimates are based on the geometric properties of the norm used for $\mathbf{x}$.

In practical settings, the signal can be impaired by impulse noise, narrow-band interference or contaminated during transmission. Therefore, the noise may not be bounded. The authors in [10] proved that as the noise energy gets larger, the solution might be very different from the ground truth.
To overcome this limitation, many specific corrupted sensing instances have been studied in recent years, for example, signal recovery and low-rank matrix recovery from corrupted observations [9], [14], [34], [36], [37], [40], [48]. In the context of signal contamination, we have observation

\[ y = \Phi x + v + e, \] (3)

where \( v \) is an unknown but structured noise vector. For measurements \( y = \Phi x + v \), Donoho and Stark in [22] gave a guarantee for the exact recovery of \( x \) when \( \Phi \) is the discrete Fourier transform matrix and the support of sparse corruption \( v \) is known. Li in [36] showed that if the \( m \times n \) sensing matrix has independent Gaussian entries, then one can recover a sparse signal \( x \) exactly by \( l_1 \) minimization

\[ \min_{x,v} \|x\|_1 + \lambda \|v\|_1 \quad s.t. \quad y = \Phi x + v, \] (4)

provided the number of nonzero elements in \( x \) is \( O(m/\log(n/m) + 1) \). Foygel and Mackey in [26] extended the work of [13] to a more challenging setting, i.e., they took into account unreliable measurements \( y = \Phi x + v + e \). They used a convex programming method to recover \( x \) and \( v \) with or without prior information and provided new bounds for the Gaussian complexity of structured signals, leading to a sharper restoration guarantee.

In addition, a generalized corrupted model is as follows

\[ y = \Phi x + \Psi v, \] (5)

where \( \Phi \in \mathbb{R}^{n \times p} \) and \( \Psi \in \mathbb{R}^{n \times q} \) are general dictionaries and \( v \in \mathbb{R}^q \) is a structured corruption. The model (5) enables us to consider signal separation, the recovery of saturated or clipped signals [11], [2], [33], sparsity-based super-resolution and in-painting [23], [38], etc. In signal separation problems, the dictionaries \( \Phi \) and \( \Psi \) are chosen such that they allow for sparse or approximately sparse representation of the two distinct features. Signal separation problems arise in applications such as the separation of textures from cartoon parts in images [4], [24] and the separation of neuronal calcium transients from smooth signals caused by astrocytes in calcium imaging [30]. Recovery guarantees in the deterministic setting for noiseless measurements \( y = \Phi x + \Psi v \) have been studied in [3], [32], [41], [44], [45]. The work in [32], [45] considered the problem of recovering the sparse signal \( x \) from \( n \) linear corrupted measurements \( y \) in (5). They provided deterministic recovery guarantees based on an uncertainty relation for pairs of general dictionaries and presented corresponding practicable recovery algorithms. Their recovery guarantees depend on the coherence parameters of \( \Phi \) and \( \Psi \) and the prior knowledge about the support sets of the signal and noise. Studer and Baraniuk in [44] presented efficient algorithms for signal restoration and separation and derived corresponding recovery guarantees that depend on the knowledge of the supports of \( x \) and \( v \) and the coherence parameters of the dictionaries \( \Phi \) and \( \Psi \).

In this paper, we consider the model (1), which is a more general model than that in [26], where \( x \) is structured under a general dictionary, the unbounded noise vector \( v \) is structured under another general dictionary and \( e \) is a stochastic bounded noise. We use convex programming approach to recover both \( x \) and \( v \) with or without prior information and provide recovery conditions through the geometry measures of signal and corruption.

The rest of the paper is arranged as follows. In Section II, we recall some concepts from analysis needed in this paper. We state the main results with or without prior information and give some applications for special signal and corruption in Section III. In Section IV, we conduct a series of simulation experiments to reinforce our theoretical results. The proofs of main results will be detailed in Section V. Conclusion and future work are presented in Section VI.

II. PRELIMINARIES

A. NOTATIONS

Boldface lowercase and uppercase letters stand for vectors and matrices, respectively. We use \( \mu_m \) to represent the expected length of an \( m \)-dimensional Gaussian random vector. Further, one can show that the expectation is tightly bounded as \( \sqrt{m - 1/2} < \mu_m < \sqrt{m} \) for all \( m \) and it is known that \( \mu_1 = \sqrt{2/\pi} \) while \( \mu_m \approx \sqrt{m} \) for large \( m \) [17], [26]. \( \mathbb{E}^m \) and \( \mathcal{S}^{m-1} \) represent the unit ball and unit sphere in \( \mathbb{R}^m \) under the \( l_2 \) norm. \( (c)_+ \) means taking the positive part of \( c \).

B. PROBLEM FORMULATION

Let the matrix \( \Phi : \mathbb{R}^p \rightarrow \mathbb{R}^n \) be a random map with i.i.d. zero-mean Gaussian entries having variance 1/\( n \), the matrix \( \Psi : \mathbb{R}^q \rightarrow \mathbb{R}^n \) also be a random map with i.i.d. zero-mean Gaussian entries having variance 1/\( n \). We use appropriate norms \( \| \cdot \|_{\text{sig}} \) and \( \| \cdot \|_{\text{cor}} \) for vectors \( x \) and \( v \), respectively, according to their structures. And any convex complexity measure can be used for \( \| \cdot \|_{\text{sig}} \) and \( \| \cdot \|_{\text{cor}} \). For example, \( l_1 \) norm for sparse vectors, trace norm \( \| \cdot \|_* \) (the sum of the singular values of a matrix) for low-rank matrices, \( l_\infty \) norm for binary vectors.

When there is no prior knowledge about the ground truth \( x^* \) and \( v^* \) in model (1), we consider the following minimization program

\[ \min_{x \in \mathbb{R}^p, v \in \mathbb{R}^q} \|x\|_{\text{sig}} + \lambda \|v\|_{\text{cor}} \quad s.t. \quad y - (\Phi x + \Psi v) \|_2 \leq \delta, \] (6)

where \( \lambda > 0 \) is a regularization parameter that provides a trade-off between \( \|x\|_{\text{sig}} \) and \( \|v\|_{\text{cor}} \).

When some prior knowledge of \( x^* \) or \( v^* \) is available, we take prior information into account in the following optimization programs

\[ \min_{x \in \mathbb{R}^p, v \in \mathbb{R}^q} \left\{ \|v\|_{\text{cor}} : \|x\|_{\text{sig}} \leq \|x^*\|_{\text{sig}}, \|y - (\Phi x + \Psi v)\|_2 \leq \delta \right\} \] (7)

and

\[ \min_{x \in \mathbb{R}^p, v \in \mathbb{R}^q} \left\{ \|x\|_{\text{sig}} : \|v\|_{\text{cor}} \leq \|v^*\|_{\text{cor}}, \|y - (\Phi x + \Psi v)\|_2 \leq \delta \right\}. \] (8)
Our results depend on two notions of convex geometry. Next, in order to present our results, we recall some concepts from convex analysis with respect to a generic norm $\| \cdot \|$ on $\mathbb{R}^m$ as in [25, 26].

The subdifferential of $\| \cdot \|$ at $x \in \mathbb{R}^m \setminus \{0\}$ is the set of vectors

$$\partial \|x\| = \{ f \in \mathbb{R}^m : \|x + d\| \geq \|x\| + \langle f, d \rangle \text{ for all } d \in \mathbb{R}^m \}.$$ 

Given $x \in \mathbb{R}^m \setminus \{0\}$, we define the tangent cone of $\| \cdot \|$ at $x$ as

$$T = \{ f : \|x + c \cdot f\| \leq \|x\| \text{ for some } c > 0 \} = \text{cone}\{z - x : \|z\| \leq \|x\| \}.$$ 

The tangent cone $T$ is equal to the set of descent (or non-ascent) directions of norm $\| \cdot \|$ at the point $x$. The normal cone to $\| \cdot \|$ at $x$ is defined to be the set of all directions $s$ that form obtuse angles with each element in the tangent cone at $x$, given by

$$N = \{ s : \langle s, f \rangle \leq 0 \text{ for all } f \in T \}.$$ 

In this paper, we use $T_{\text{sig}}$ and $T_{\text{cor}}$ to represent the tangent cones of norms $\| \cdot \|_{\text{sig}}$ and $\| \cdot \|_{\text{cor}}$ at corresponding points, respectively.

To quantify the complexity of a structured vector, we review two geometric measures of structure, the Gaussian complexity and the Gaussian distance.

**Definition 1** [26]: The Gaussian squared complexity $\omega^2(C)$ of a set $C \subset \mathbb{R}^p$ is given by

$$\omega^2(C) = \mathbb{E}_{g \sim N(0, I_p)} \left[ \left( \sup_{w \in C} \langle g, w \rangle \right)^2 \right].$$

Then Gaussian complexity is the square root of Gaussian squared complexity $\omega^2(C)$, i.e., $\omega(C) = \sqrt{\omega^2(C)}$.

**Definition 2** [26]: The Gaussian squared distance $\eta^2(C)$ of a set $C \subset \mathbb{R}^p$ is given by

$$\eta^2(C) = \mathbb{E}_{g \sim N(0, I_p)} \left[ \inf_{w \in C} \| g - w \|^2 \right].$$

Similarly, the Gaussian distance is the square root of Gaussian squared distance $\eta^2(C)$, $\eta(C) = \sqrt{\eta^2(C)}$.

We know that there is a relationship between Gaussian complexity and Gaussian distance of the scaled subdifferential of $\|x\|$ from [42, 43]

$$\omega^2(T \cap \mathbb{R}^p) = \eta^2(N) = \mathbb{E}_{g \sim N(0, I_p)} \left[ \inf_{w \in N} \| g - w \|^2 \right] = \mathbb{E}_{t \geq 0} \left[ \inf_{t \geq 0} \| g - w \|^2 \right] \leq \min_{t \geq 0} \eta^2(t \cdot \partial \|x\|).$$

So we can find the bound of Gaussian complexity by bounding $\eta(t \cdot \partial \|x\|)$ at a value of $t$. And different choices of $t$ can produce different bounds.

### III. MAIN RESULTS

In this section, we will present the main results with or without prior information and give some applications for special signal and corruption. The proofs of the theorems can be found in Section V.

#### A. RECOVERY GUARANTEE WITH PRIOR INFORMATION

We give sufficient conditions for exact and stable recovery whenever prior information of $x$ and $v$ is available by the following result.

**Theorem 1**: If $(\hat{x}, \hat{v})$ selects either of the optimization problems (7) or (8), then

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

with probability at least $1 - e^{-\left(\mu_n - \tau - \sqrt{n}\epsilon\right)^2/2}$, as long as $\mu_n - \sqrt{n}\epsilon$ exceeds the threshold

$$\tau = \sqrt{\omega^2(T_{\text{sig}} \cap \mathbb{R}^p) + \omega^2(T_{\text{cor}} \cap \mathbb{R}^p) + \frac{1}{\sqrt{2\pi}}}.$$

**Remark 1**: In the absence of stochastic noise, i.e., $\delta = 0$, Theorem 1 gives an exact solution of problems (7) or (8) with probability at least $1 - e^{-\left(\mu_n - \tau\right)^2/2}$, as long as $\mu_n \geq \tau$.

#### B. RECOVERY GUARANTEE WITHOUT PRIOR INFORMATION

Next we present sufficient conditions for exact and stable recovery when prior information of $x$ or $v$ is unavailable.

**Theorem 2**: If $(\hat{x}, \hat{v})$ solves the optimization problem (6) for any fixed $t_{\text{sig}} > 0$, $t_{\text{cor}} > 0$ with parameter $\lambda = t_{\text{cor}}/t_{\text{sig}}$, then

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

with probability at least $1 - e^{-\left(\mu_n - \tau - \sqrt{n}\epsilon\right)^2/2}$, as long as $\mu_n - \sqrt{n}\epsilon$ exceeds the threshold

$$\tau = \eta(t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}}) + \eta(t_{\text{cor}} \cdot \partial \|v^*\|_{\text{cor}}) + 2\sqrt{2\pi} + \frac{1}{\sqrt{2\pi}}.$$

**Remark 2**: In the absence of stochastic noise ($\delta = 0$), Theorem 2 gives an exact solution of problems (6) with probability at least $1 - e^{-\left(\mu_n - \tau\right)^2/2}$, as long as $\mu_n \geq \tau$.

#### C. COMPARISON WITH RELATED LITERATURE

For Gaussian matrices $\Phi \in \mathbb{R}^{n \times p}$ and $\Psi \in \mathbb{R}^{n \times n}$, we can write

$$y = \Phi x + \Psi v = [\Phi \, \Psi] \begin{bmatrix} x \\ v \end{bmatrix}. \quad (9)$$

In this sense, $[\Phi \, \Psi] \in \mathbb{R}^{n \times (p + n)}$ is a large Gaussian matrix and $\begin{bmatrix} x \\ v \end{bmatrix} \in \mathbb{R}^{p+n}$ is a large vector. When $x$ is $s_1$-sparse and $v$ is $s_2$-sparse, then $\begin{bmatrix} x \\ v \end{bmatrix}$ is $(s_1 + s_2)$-sparse. In this subsection, we use $T_1$, $T_2$ and $T_U$ to represent the tangent cones of norms $\|x\|_1$, $\|v\|_1$ and $\|x\|_1 + \|v\|_1$ at the corresponding points, respectively.
According to the work in [13], we can recover \( \begin{bmatrix} x \\ y \end{bmatrix} \) from (9) provided \( n \gg \omega^2(T_U \cap \mathbb{B}^{p+q}) \) (ignoring the small constant). On the other hand, the theoretical recovery bounds obtained from our models can all be approximated as \( n \gg \omega^2(T_1 \cap \mathbb{B}^p) + \omega^2(T_2 \cap \mathbb{B}^q) \). We will show that the theoretical recovery threshold obtained by our results for \( \| \cdot \|_{Sig} = \| \cdot \|_{Cor} = \| \cdot \|_1 \) is less than the theoretical recovery threshold obtained by the work [13] for \( \| \cdot \|_A = \| \cdot \|_1 \) in some cases, albeit the main contribution of our results is for different measures for the signal and the corruption.

For an \( s \)-sparse \( p \)-dimensional vector, there are two ways to calculate its Gaussian squared complexity: \( 2s \log(\frac{\sqrt{s}}{\pi}) + \frac{3}{2}s \) and \( p\left(1 - \frac{2}{\pi\sqrt{n}}\right)^2 \) according to [26]. The threshold required to recover \( x \) form model (9) according to [13] is denoted as \( b_o \) and \( b_n \) designates the threshold needed by our methods. Next, we compare \( b_o \) and \( b_n \) in some cases.

First, we take Gaussian squared complexity \( p\left(1 - \frac{2}{\pi\sqrt{n}}\right)^2 \).

1): When \( n = p \) and \( s_1 = s_2 = \alpha p < \alpha < 1 \),
\[
b_o = 2p\left(1 - \frac{2}{\pi}(1 - \alpha)^2\right),
b_n = 2p\left(1 - \frac{2}{\pi}(1 - \alpha)^2\right).
\]
We have \( b_o = b_n \).

2): When \( n = p, s_1 = \alpha p, s_2 = \beta n \) and \( s_1 \neq s_2 \),
\[
b_o = 2p - \frac{4p}{\pi}\left(1 - \frac{2}{\pi}(1 - \alpha + \beta)^2\right),
b_n = 2p - \frac{2p}{\pi}\left[1 - \left(\frac{\alpha - \beta}{2}\right)^2\right],
b_n - b_o = \frac{4p}{\pi}\left(1 - \frac{2}{\pi}(1 - \alpha + \beta)^2\right) - \frac{2p}{\pi}\left(1 - \left(\frac{\alpha - \beta}{2}\right)^2\right) \quad \text{implies} \quad \frac{\pi}{2}(\alpha - \beta)^2 < 0.<n>

So we have \( b_n \) is always less than \( b_o \). This means that our condition is weaker than that in [13].

3): When \( n \neq p, s_1 = \alpha p, s_2 = \beta n \),
\[
b_o = (p + n)\left(1 - \frac{2}{\pi}(1 - \frac{s_1 + s_2}{p + n})^2\right),
b_n = p\left(1 - \frac{2}{\pi}(1 - \alpha)^2\right) + n\left(1 - \frac{2}{\pi}(1 - \beta)^2\right).
\]

We cannot directly get the size relationship between \( b_o \) and \( b_n \) from their expressions. However, for some special values of \( \alpha \) and \( \beta \), we can visually compare the size of \( b_o \) and \( b_n \). For example, we let \( \alpha = 0.01, \beta = 0.4 \) in Figure 1. We can see that in this case \( b_n \) is always less than \( b_o \).

Second, we take Gaussian squared complexity \( 2s \log(\frac{\sqrt{s}}{\pi}) + \frac{3}{2}s \).

1): When \( n = p \) and \( s_1 = s_2 = \alpha p < \alpha < 1 \),
\[
b_o = 2(s_1 + s_2)\log\left(\frac{p + n}{s_1 + s_2}\right) + \frac{3}{2}(s_1 + s_2) = 4ap \log\left(\frac{1}{\alpha}\right) + 3\alpha p.
\]

For \( \alpha = 0.1, \beta = 0.01 \), we can see that in this case \( b_n \) is always less than \( b_o \) from Figure 2.
By comparing the magnitudes of $b_0$ and $b_n$, we show that our results are meaningful. Interested readers can also try other values.

**D. PROOF OUTLINE**

The ground truth $(x^*, v^*)$ and the solution $(\hat{x}, \hat{v})$ obtained from the above two cases lie in the feasible set $(x, v) : \|y - (\Phi x + \Psi v)\|_2 \leq \delta$. So we have a bound on the noise level $\|\Phi(\hat{x} - x^*) + \Psi(\hat{v} - v^*)\|_2 \leq 2\delta$. For problems with prior information, we have the following two tangent cones

$$T_{\text{sig}} = \{a \in \mathbb{R}^p : \exists t > 0, \|x^* + t \cdot a\|_{\text{sig}} \leq \|x^*\|_{\text{sig}}\}$$

and

$$T_{\text{cor}} = \{b \in \mathbb{R}^q : \exists t > 0, \|v^* + t \cdot b\|_{\text{cor}} \leq \|v^*\|_{\text{cor}}\}.$$

If $(\hat{x}, \hat{v})$ is the solution of program (7) or (8), we know that $(\hat{x} - x^*, \hat{v} - v^*) \in T_{\text{sig}} \times T_{\text{cor}}$ from the definitions of $T_{\text{sig}}$ and $T_{\text{cor}}$.

For problems without prior information, we have a joint tangent cone from [26]

$$T_{\text{joint}} = \{(a, b) \in \mathbb{R}^p \times \mathbb{R}^q : \exists t > 0, \text{Pen}_a(x^* + t \cdot a, v^* + t \cdot b) \leq \text{Pen}_a(x^*, v^*)\},$$

where $\text{Pen}_a(x, v) = \|x\|_{\text{sig}} + \lambda\|v\|_{\text{cor}}$, $\lambda > 0$ is a parameter. Similarly, if $(\hat{x}, \hat{v})$ is the solution of program (6), we have $(\hat{x} - x^*, \hat{v} - v^*) \in T_{\text{joint}}^\lambda$.

If we can get lower bounds on

$$\min_{(a, b) \in T_{\text{sig}} \times T_{\text{cor}}} \frac{\|\Phi a + \Psi b\|_2}{\sqrt{\|a\|_2^2 + \|b\|_2^2}}$$

and

$$\min_{(a, b) \in T_{\text{joint}}} \frac{\|\Phi a + \Psi b\|_2}{\sqrt{\|a\|_2^2 + \|b\|_2^2}}.$$  \hspace{5.7cm} (10) \hspace{5.7cm} (11)

Then the lower bounds of (10) and (11) can be applied to the error vector $(a, b) = (\hat{x} - x^*, \hat{v} - v^*)$ to bound its $l_2$ norm.

**E. SOME EXAMPLES**

For binary signal and sparse corruption, we can formulate the recovery problem as follows

$$\min_{x, v} \{\|v\|_1 : \|x\|_\infty \leq 1, \|y - (\Phi x + \Psi v)\|_2 \leq \delta\}. \hspace{5.7cm} (12)$$

Following from Theorem 1, we present the exact recovery of binary signal $x^*$ via program (12).

**Corollary 1**: If $x^* \in \{+1, -1\}^p$, $v^*$ is an $s$-sparse vector. Suppose the number of measurements satisfies

$$\mu_n - 2\delta \sqrt{n} > \sqrt{\frac{p}{2}} + \omega^2(T_{\text{spar}(q, s)} \cap B^q) + \frac{1}{\sqrt{2\pi}} := \tau,$$

where $T_{\text{spar}(q, s)}$ is the tangent cone of an $s$-sparse $q$-dimensional vector under the $l_1$ norm, satisfying

$$\omega^2(T_{\text{spar}(q, s)} \cap B^q) \leq \min \{2s \log(q) + \frac{3}{2}s, q(1 - \frac{2}{\pi}(1 - \frac{s}{q})^2)\}. \hspace{5.7cm} (13)$$

Then the binary signal $x^*$ is exactly recovered by the sign of $\hat{x}$ with probability at least $1 - e^{-\mu_n - \delta \sqrt{n} - \frac{1}{2}}$, where $(\hat{x}, \hat{v})$ is any solution of optimization problem (12).

**Proof**: Using the upper bounds of the Gaussian complexity in [26] and Theorem 1 with any $\varepsilon \in (2s, \frac{\mu_n}{\sqrt{\pi}})$ and the tangent cone complexity bound in [26], we have

$$\|\hat{x} - x^*\|_\infty \leq \sqrt{\|\hat{x} - x^*\|_2^2 + \|v^* - v^*\|_2^2} \leq \frac{2\delta}{\varepsilon} < 1$$

with probability at least $1 - e^{-\mu_n - \delta \sqrt{n} - \frac{1}{2}}$. Therefore, letting $\varepsilon$ approach $2\delta$, we can exactly recover $x^*$ by setting $x^* = \text{sign}(\hat{x})$ with probability at least $1 - e^{-\mu_n - \frac{1}{2}}$.

If both the signal $x$ and corruption $v$ are sparse vectors, we consider the recovery via program

$$\min_{x \in \mathbb{R}^p, v \in \mathbb{R}^q} \{\|x\|_1 + \|v\|_1 : \|y - (\Phi x + \Psi v)\|_2 \leq \delta\}.$$  \hspace{5.7cm} (14)

where we set $\lambda = 1$.

**Corollary 2**: If $x \in \mathbb{R}^p$ has at most $s_{\text{sig}}$ nonzero entries, $v \in \mathbb{R}^q$ has at most $s_{\text{cor}}$ nonzero entries. Suppose that $(\hat{x}, \hat{v})$ is the solution of program (14), then

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

with probability at least $1 - e^{-\mu_n - \varepsilon \sqrt{\mu_n}}$, as long as the number of measurements $n$ satisfies

$$\mu_n - \sqrt{n\varepsilon} \geq \tau$$

where

$$\mu_n = \sqrt{2s_{\text{sig}} \log \left( \frac{p}{s_{\text{sig}}} \right) + \frac{3}{2}s_{\text{sig}}} + \sqrt{2s_{\text{cor}} \log \left( \frac{q}{s_{\text{cor}}} \right) + \frac{3}{2}s_{\text{cor}} + 2\sqrt{2\pi} + \frac{1}{\sqrt{2\pi}}}.$$  \hspace{5.7cm} (15)

**Proof**: Using the upper bounds of the Gaussian distance in [26] and Theorem 2, we can easily get the result. □
IV. NUMERICAL EXPERIMENTS

In this section, we verify and reinforce the theoretical results of Section III with a series of simulations. We present experiments for both problem models with and without prior information. In each experiment, we used the CVX Matlab package [28], [29] to specify and solve the convex recovery programs. Figures 4-7 report the empirical probability of success for different structured vectors and different problem models. Figure 8 plots the recovery error ∥x̂ − x∗∥2 + ∥ŷ − v∗∥2 for sparse signal recovery under noise and sparse corruption with prior information.

In Figure 4, we consider binary signal and sparse corruption in the absence of noise (δ = 0). Fixing the message length p = 300, we vary the number of measurements n ∈ [200, 300] and the number of corruptions s_cor ∈ [2, 82] and perform the following experiment 10 times for each (n, s_cor) pair:

1): Generate matrices Φ ∈ ℜnxp and Ψ ∈ ℜnxq with independent N(0, 1/n) entries, respectively.
2): Generate a binary signal vector x∗ uniformly from {±1}p.
3): Generate a corruption vector v∗ whose s_cor nonzero entries are generated from a standard Gaussian distribution.
4): Solve the optimization problem (12) with y = Φx + Ψv and δ = 0.
5): Declare success if ∥x̂ − x∗∥2/∥x∗∥2 < 0.001.

In Figure 5, we consider sparse signal and sparse corruption in the presence of prior information. Fixing the signal length and the measurement size n = p = 200, we vary the sparsity levels (s_sig, s_cor) ∈ [1, 101]2. Then we perform the following experiment 10 times for each (s_sig, s_cor) pair:

1): Generate matrices Φ ∈ ℜnxp and Ψ ∈ ℜnxq with independent N(0, 1/n) entries, respectively.
2): Generate a sparse signal vector x∗ whose s_sig nonzero entries are generated from a standard Gaussian distribution.
3): Generate a corruption vector v∗ whose s_cor nonzero entries are generated from a standard Gaussian distribution.
4): For y = Φx + Ψv and δ = 0, solve the optimization problem as follows

(̂x, ̂v) ∈ arg min{∥v∥1 : ∥x∥1 ≤ ∥x∗∥1, y = Φx + Ψv}.

5): Declare success if ∥x̂ − x∗∥2/∥x∗∥2 < 0.001.

In Figure 6 and Figure 7, we consider recovery problem (6) in the absence of noise (δ = 0) without any prior information. We focus on the case where both the signal and the corruption vector have sparse structure, i.e., ∥x∥some = ∥x∥1, ∥v∥some = ∥v∥1. Fixing the signal length and the measurement size n = p = 200, we vary the sparsity levels (s_sig, s_cor) ∈ [1, 101]2. Set λ = λ_den = (1 − s_cor/n)/(1 − s_sig/p) in Figure 6 and λ = λ_spa = √log(n/s_cor)/√log(p/s_sig) in Figure 7. Then we perform the following experiment 10 times for each (s_sig, s_cor) pair:

1): Generate matrices Φ ∈ ℜnxp and Ψ ∈ ℜnxq with independent N(0, 1/n) entries, respectively.
We also focus on the case where both the signal and the entries are generated from a standard Gaussian distribution.

In the absence of prior information, phase transition for sparse signal recovery under sparse corruption, \( \lambda = \lambda_{\text{spa}} = \sqrt{\log(p/\gamma_{\text{cor}})} / \sqrt{\log(p/\gamma_{\text{sig}})} \).

Figure 7: In the absence of prior information, phase transition for sparse signal recovery under sparse corruption, \( \lambda = \lambda_{\text{spa}} = \sqrt{\log(p/\gamma_{\text{cor}})} / \sqrt{\log(p/\gamma_{\text{sig}})} \).

2): Generate a sparse signal vector \( x^* \) whose \( s_{\text{sig}} \) nonzero entries are generated from a standard Gaussian distribution.

3): Generate a corruption vector \( v^* \) whose \( s_{\text{cor}} \) nonzero entries are generated from a standard Gaussian distribution.

4): For \( y = \Phi x + \Psi v + \epsilon \), solve the optimization problem as follows

\[
(\hat{x}, \hat{v}) \in \arg \min_{x,v} \{ \|v\|_1 + \lambda \|x\|_1 : y = \Phi x + \Psi v \}.
\]

5): Declare success if \( \|\hat{x} - x^*\|_2 / \|x^*\|_2 < 0.001 \).

In Figure 8, we consider recovery problem (7) in the presence (\( \delta \neq 0 \)) of noise without any prior information. We also focus on the case where both the signal and the corruption vector have sparse structure. For fixed noise level \( \delta = 0.1 \) and sparsity fractions \( \gamma_{\text{sig}}, \gamma_{\text{cor}} = (0.01, 0.4) \), we vary the message length \( p \in \{200, 300, 400\} \) and the number of measurements \( n \in \{50, 350\} \). Then we perform the following experiment 50 times for each \((p, n)\) pair.

1): Generate matrices \( \Phi \in \mathbb{R}^{p \times p} \) and \( \Psi \in \mathbb{R}^{n \times q} \) with independent \( N(0, 1/n) \) entries.

2): Generate a signal vector \( x^* \) whose \( p \cdot s_{\text{sig}} \) nonzero entries are generated from a standard Gaussian distribution.

3): Generate a corruption vector \( v^* \) whose \( n \cdot s_{\text{cor}} \) nonzero entries are generated from a standard Gaussian distribution.

4): A dense noise \( \epsilon \) is generated from a standard Gaussian distribution with \( \|\epsilon\|_2 = 0.1 \).

5): For \( y = \Phi x + \Psi v + \epsilon \), solve the optimization problem as follows

\[
(\hat{x}, \hat{v}) \in \arg \min_{x,v} \{ \|v\|_1 : \|x\|_1 \leq \|x^*\|_1, \|y - (\Phi x + \Psi v)\|_2 \leq 0.1 \}.
\]

6): Record the recovery error \( \|\hat{x} - x^*\|_2 + \|\hat{v} - v^*\|_2 \).

The red lines in Figures 4–7 are the theory thresholds of theorems in Section III and the values of Gaussian squared complexity and Gaussian squared distance used here come from Chandrasekaran et al. [13] and Foygel and Mackey [26]. These theoretical recovery thresholds reflect the observed empirical phase transition accurately. In the noisy measurement setting, we see as \( n \) increases, the lines for three different values \( p \) all converge to a common value from Figure 8.

V. SUPPORTING LEMMATA AND PROOFS OF MAIN RESULTS

Take any \( \Omega \subset \mathbb{B}^{p+q} \). Let \( a \in \mathbb{R}^p \) and \( b \in \mathbb{R}^q \), we first derive a lower bound on \( \mathbb{E} \min_{[a^T b^T]^T \in \Omega} \|\Phi a + \Psi b\|_2 \) via the following lemma.

**Lemma 1**: Let \( \Phi \in \mathbb{R}^{n \times p} \) and \( \Psi \in \mathbb{R}^{n \times q} \) have i.i.d. \( N(0, 1/n) \) entries, and let \( g \in \mathbb{B}^{p+q} \) and \( h \in \mathbb{R}^n \) have i.i.d. \( N(0, 1) \) entries. Then for any \( \Omega \subset \mathbb{B}^{p+q} \), there holds

\[
\sqrt{n} \mathbb{E} \min_{[a^T b^T]^T \in \Omega} \|\Phi a + \Psi b\|_2 \geq \mathbb{E} \left[ \left( \min_{[a^T b^T]^T \in \Omega} \|h \cdot [a \ b]\|_2 + \|g \cdot [a \ b]\|_2 \right) + \right] - \frac{1}{\sqrt{2\pi}}.
\]

**Proof**: For \([a \ b] \in \Omega \) and \( w \in \mathbb{R}^n \), let

\[
X_{[a^T b]^T, w} = \sqrt{n}(\Phi a + \Psi b, w) + v \cdot \left[ a \ b \right]_2
\]

and

\[
Y_{[a^T b]^T, w} = (h, w) \cdot \left[ a \ b \right]_2 + \left( g \cdot \left[ a \ b \right] \right).
\]

where \( v \sim N(0, 1) \) is independent from the other random variables. Then \( \{X_{[a^T b]^T, w}\} \) and \( \{Y_{[a^T b]^T, w}\} \) are both centered Gaussian processes, with

\[
\mathbb{E}[X_{[a^T b]^T, w}^2] = \mathbb{E}[Y_{[a^T b]^T, w}^2] = \left\| \left[ a \ b \right]_2 \right\|_2 (\|w\|_2^2 + 1)
\]

and

\[
\mathbb{E}[X_{[a^T b]^T, w} X_{[a^T b]^T, w} Y_{[a^T b]^T, w} Y_{[a^T b]^T, w}] = \mathbb{E}[Y_{[a^T b]^T, w} Y_{[a^T b]^T, w}] = 0.
\]

1): Generate matrices \( \Phi \in \mathbb{R}^{p \times p} \) and \( \Psi \in \mathbb{R}^{n \times q} \) with independent \( N(0, 1/n) \) entries.

2): Generate a signal vector \( x^* \) whose \( p \cdot s_{\text{sig}} \) nonzero entries are generated from a standard Gaussian distribution.

3): Generate a corruption vector \( v^* \) whose \( n \cdot s_{\text{cor}} \) nonzero entries are generated from a standard Gaussian distribution.

4): A dense noise \( \epsilon \) is generated from a standard Gaussian distribution with \( \|\epsilon\|_2 = 0.1 \).

5): For \( y = \Phi x + \Psi v + \epsilon \), solve the optimization problem as follows

\[
(\hat{x}, \hat{v}) \in \arg \min_{x,v} \{ \|v\|_1 : \|x\|_1 \leq \|x^*\|_1, \|y - (\Phi x + \Psi v)\|_2 \leq 0.1 \}.
\]

6): Record the recovery error \( \|\hat{x} - x^*\|_2 + \|\hat{v} - v^*\|_2 \).

The red lines in Figures 4–7 are the theory thresholds of theorems in Section III and the values of Gaussian squared complexity and Gaussian squared distance used here come from Chandrasekaran et al. [13] and Foygel and Mackey [26]. These theoretical recovery thresholds reflect the observed empirical phase transition accurately. In the noisy measurement setting, we see as \( n \) increases, the lines for three different values \( p \) all converge to a common value from Figure 8.
\[
\begin{align*}
&\left\langle \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}, \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} \right\rangle (w, \tilde{w}) + \left\| \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} \right\|_2 \left\| \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} \right\|_2 \\
&- (w, \tilde{w}) \left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 \left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 - \left\langle \begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} a \\ b \end{bmatrix} \right\rangle \\
= &\left( \left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 \left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 - \left\langle \begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} a \\ b \end{bmatrix} \right\rangle \right) (1 - (w, \tilde{w}))
\end{align*}
\]
with equality if \( \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} \). Therefore, using Theorem 1.1 of Gordon [31] and simplifying the process, we obtain
\[
P\left\{ \min_{[a^T b^T]^T \in \Omega} \max_{w \in \mathbb{R}^p} X_{[a^T b^T]^T, w} \geq C \right\} \geq P\left\{ \min_{[a^T b^T]^T \in \Omega} \max_{w \in \mathbb{R}^p} Y_{[a^T b^T]^T, w} \geq C \right\}.
\]
by maximizing over \( w \) on each side, we obtain
\[
P\left\{ \min_{[a^T b^T]^T \in \Omega} \sqrt{n} \left\| \phi a + \psi b \right\|_2 + v \left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 \geq C \right\} \\
\geq P\left\{ \min_{[a^T b^T]^T \in \Omega} \left\| h \cdot \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 + \left\langle g, \begin{bmatrix} a \\ b \end{bmatrix} \right\rangle \geq C \right\}.
\]
Since this is true for all \( C \in \mathbb{R} \), we can integrate over \( C \in (-\infty, \infty) \) to obtain
\[
E\left[ \min_{[a^T b^T]^T \in \Omega} \sqrt{n} \left\| \phi a + \psi b \right\|_2 + v \left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 \right]\n\geq E\left[ \min_{[a^T b^T]^T \in \Omega} \left\| h \cdot \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 + \left\langle g, \begin{bmatrix} a \\ b \end{bmatrix} \right\rangle \right].
\]
Since \( \left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 \leq 1 \) for all \( \begin{bmatrix} a \\ b \end{bmatrix} \in \Omega \), we obtain
\[
\left( \min_{[a^T b^T]^T \in \Omega} \sqrt{n} \left\| \phi a + \psi b \right\|_2 + v \left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 \right) \leq \sqrt{n} \min_{[a^T b^T]^T \in \Omega} \left\| \phi a + \psi b \right\|_2 + (v+),
\]
then using \( E[(v+)] = \frac{1}{2\sqrt{\pi}} \),
\[
\sqrt{n} E\left[ \min_{[a^T b^T]^T \in \Omega} \left\| \phi a + \psi b \right\|_2 \right] \\
\geq E\left[ \left( \min_{[a^T b^T]^T \in \Omega} \sqrt{n} \left\| \phi a + \psi b \right\|_2 + v \left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 \right) \right] \\
- \frac{1}{2\sqrt{\pi}} \\
\geq E\left[ \left( \min_{[a^T b^T]^T \in \Omega} \left\| h \cdot \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 + \left\langle g, \begin{bmatrix} a \\ b \end{bmatrix} \right\rangle \right) \right] \\
- \frac{1}{2\sqrt{\pi}}.
\]
This concludes the proof. \( \square \)

Let \( \tilde{\Omega} = (\mathcal{T}_{\text{sig}} \times \mathcal{T}_{\text{cor}}) \cap \mathbb{S}^{p+q-1} \). We use \( \omega^2 \) and \( \omega'^2 \) to represent the relevant Gaussian squared complexities \( \omega^2(\mathcal{T}_{\text{sig}} \cap \mathbb{S}^p) \) and \( \omega'^2(\mathcal{T}_{\text{cor}} \cap \mathbb{S}^q) \), respectively. Next, we relate \( \left\| \phi a + \psi b \right\|_2 \) to the associated Gaussian square complexity by the following lemma.

**Lemma 2:**
\[
\sqrt{n} E\left[ \min_{(a,b) \in \tilde{\Omega}} \left\| \phi a + \psi b \right\|_2 \right] \geq \mu_n - \sqrt{\omega^2 + \omega'^2} - \frac{1}{\sqrt{2\pi}}.
\]

**Proof:** According to Lemma 1, we obtain
\[
\sqrt{n} E\left[ \min_{(a,b) \in \tilde{\Omega}} \left\| \phi a + \psi b \right\|_2 \right] \\
\geq E\left[ \left( \min_{(a,b) \in \tilde{\Omega}} \left\| h \cdot \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 + \left\langle g, \begin{bmatrix} a \\ b \end{bmatrix} \right\rangle \right) \right] - \frac{1}{\sqrt{2\pi}}.
\]

In order to get the desired result, we have to lower bound the right-hand side of (17) for \( \tilde{\Omega} = (\mathcal{T}_{\text{sig}} \times \mathcal{T}_{\text{cor}}) \cap \mathbb{S}^{p+q-1} \). Let \( \tilde{\omega} = \max_{a \in \mathcal{T}_{\text{sig}}} \sqrt{\omega} \) and let \( \tilde{\omega}' = \max_{b \in \mathcal{T}_{\text{cor}}} \sqrt{\omega'} \) be the observed Gaussian complexities with \( E\tilde{\omega} = \omega \) and \( E\tilde{\omega}' = \omega' \), where \( g_1 \in \mathbb{R}^p \) and \( g_2 \in \mathbb{R}^q \) are independent vectors with i.i.d. standard normal entries. The proof consists of the following two cases.

First, if \( \sqrt{\omega^2 + \omega'^2} > \|h\|_2 \), then
\[
E\left[ \left( \min_{(a,b) \in \tilde{\Omega}} \left\| h \cdot \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 + \left\langle g, \begin{bmatrix} a \\ b \end{bmatrix} \right\rangle \right) \right] \\
\geq 0 \\
> |h|_2 - \sqrt{\omega^2 + \omega'^2}.
\]

Second, if \( \sqrt{\omega^2 + \omega'^2} \leq \|h\|_2 \), then
\[
\min_{(a,b) \in \tilde{\Omega}} \|h \cdot \begin{bmatrix} a \\ b \end{bmatrix}\|_2 + \left\langle g, \begin{bmatrix} a \\ b \end{bmatrix} \right\rangle \\
= \min_{c \in [0,1]} \|h\|_2 + c(g_1, a) + (g_2, b) \\
= \min_{c \in [0,1]} \|h\|_2 + c(g_1, a) + \sqrt{1 - c^2}(g_2, b) \\
\geq \min_{c \in [0,1]} \|h\|_2 + c(g_1, a) + \sqrt{1 - c^2}(g_2, b) \\
\geq \|h\|_2 - c\omega - \sqrt{1 - c^2}\omega' \quad \text{(1)} \\
\geq \|h\|_2 - \sqrt{\omega^2 + \omega'^2},
\]
where we made a substitution in the second equality and (1) is minimized over \( a \) and \( b \) for a fixed \( c \). Combining the above two cases, we complete the proof. \( \square \)

Similarly, we want to derive a lower bound on
\[
E\left[ \min_{(a,b) \in \mathcal{T}_{\text{sig}} \cap \mathbb{S}^p} \left\| \phi a + \psi b \right\|_2 \right].
\]

**Lemma 3:** Let \( \lambda = \frac{\omega_{\text{sig}}}{\omega_{\text{cor}}} \) for parameters \( t_{\text{sig}} > 0 \), \( t_{\text{cor}} > 0 \). Then
\[
\sqrt{n} E\left[ \min_{(a,b) \in \mathcal{T}_{\text{sig}} \cap \mathbb{S}^p} \left\| \phi a + \psi b \right\|_2 \right] \\
\geq \mu_n - \eta(t_{\text{sig}} \cdot \partial \|x\|_{\text{sig}}) - \eta(t_{\text{cor}} \cdot \partial \|v\|_{\text{cor}}) \\
- 2\sqrt{2\pi} - \frac{1}{\sqrt{2\pi}}.
\]
We want to derive a lower bound for $\|x\|_{\text{sig}}$, where the last inequality is due to (19). Taking expectations, we obtain
\[
\mathbb{E}\left( \min_{(a,b) \in \Omega} \| h \cdot \left\| a \right\|_2 + \left\| g \cdot \left\| b \right\|_2 \right\| \right) \geq \mu_n - \eta(t_{\text{sig}} \cdot \| x \|_{\text{sig}}) - \eta(t_{\text{cor}} \cdot \| v^* \|_{\text{cor}}) - 2\sqrt{2\pi},
\]
where the last inequality is due to the Lemma 8 in [26]. Combining (18) and (20), we complete the proof. □

**Proof of Theorem 1:** Let $(x^*, v^*)$ be the ground truth, $(\hat{x}, \hat{v})$ be the solution of program (7) or (8), we have
\[
\| \Phi(x^* - x) + \Psi(v^* - \hat{v}) \|_2 \leq \| y - \Phi x^* - \Psi v^* \|_2 \leq 2\delta.
\]
For all $[a, b] \in \Omega \subset \mathbb{R}^{p+q-1}$, the function
\[
\sqrt{n}[\Phi, \Psi] \mapsto \sqrt{n} \min_{(a,b) \in \Omega} \| \Phi a + \Psi b \|_2
\]
is 1-Lipschitz with respect to the Frobenius norm. And note that $(x^* - x, \hat{v} - v^*) \in T_{\text{sig}} \times T_{\text{cor}}$, therefore, we have the following bound with probability at least $1 - \beta$ [35]
\[
\sqrt{n} \min_{(T_{\text{sig}} \times T_{\text{cor}}) \in \mathbb{R}^{p+q-1}} \| \Phi \|_2 \times \| x^* - x \|_2 + \| v^* - \hat{v} \|_2 \leq \| \Phi(x^* - x) + \Psi(v^* - \hat{v}) \|_2 \leq 2\delta.
\]
for any $\beta \in (0, 1]$. Applying Lemma 2, we have
\[
\sqrt{n} \min_{(T_{\text{sig}} \times T_{\text{cor}}) \in \mathbb{R}^{p+q-1}} \| \Phi(x^* - x) + \Psi(v^* - \hat{v}) \|_2 \leq \mu_n - \sqrt{\omega^2 + \omega^2} - \frac{1}{\sqrt{2\pi}} - \sqrt{2\log(1/\beta)}.
\]
The equation above is equivalent to
\[
\| x^* - x \|_2 + \| v^* - \hat{v} \|_2 \leq \frac{1}{\sqrt{n}} \| \Phi(x^* - x) + \Psi(v^* - \hat{v}) \|_2 \leq \frac{2\delta}{\varepsilon}.
\]
Consider the recovery problem without prior information, let
\[
(\hat{x}, \hat{v}) = \arg \min \left\{ \| x \|_{\text{sig}} + \| v \|_{\text{cor}} : \| y - (\Phi x + \Psi v) \|_2 \leq \delta \right\}.
\]
Proof of Theorem 2: Similar to the proof of Theorem 1, let \((x^*, v^*)\) be the ground truth, we have
\[
\|\Phi(\hat{x} - x^*) + \Psi(\hat{v} - v^*)\|_2 \leq \|y - \Phi\hat{x} - \Psi\hat{v}\|_2 + \|y - \Phi x^* - \Psi v^*\|_2 \leq 2\delta.
\]
Note that \((\hat{x} - x^*, \hat{v} - v^*) \in T_{\text{joint}}^\beta\). Therefore, we have the following bound with probability at least \(1 - \beta\) [35]
\[
\sqrt{n} \min_{T_{\text{joint}}^\beta \in \mathbb{G}^n T_{\text{joint}}^\beta + q^{-1}} \left\| \Phi(\hat{x} - x^*) + \Psi(\hat{v} - v^*) \right\|_2
\geq \sqrt{n} \mathbb{E} \left[ \left\{ \min_{T_{\text{joint}}^\beta \in \mathbb{G}^n T_{\text{joint}}^\beta + q^{-1}} \left\| \Phi(\hat{x} - x^*) + \Psi(\hat{v} - v^*) \right\|_2 \right] 
- 2\sqrt{\log(1/\beta)}.
\]
for any \(\beta \in (0, 1]\). Applying Lemma 3, we have
\[
\sqrt{n} \min_{T_{\text{joint}}^\beta \in \mathbb{G}^n T_{\text{joint}}^\beta + q^{-1}} \left\| \Phi(\hat{x} - x^*) + \Psi(\hat{v} - v^*) \right\|_2
\geq \mu_n - \eta(t_{\text{sig}} \cdot \partial ||x^*||_{\text{sig}}) - \eta(t_{\text{cor}} \cdot \partial ||v^*||_{\text{cor}})
- 2\sqrt{2\pi} - \frac{1}{\sqrt{2\pi}} - 2\log(1/\beta).
\]
The equation above is equivalent to
\[
\frac{\sqrt{n}}{\sqrt{n}} \left\| \Phi(\hat{x} - x^*) + \Psi(\hat{v} - v^*) \right\|_2
\leq \frac{\mu_n - \eta||x^*||_{\text{sig}} - \eta||v^*||_{\text{cor}} - 2\sqrt{2\log(1/\beta)} - \delta}{\sqrt{n}}.
\]
\[
\leq \frac{\mu_n - \eta||x^*||_{\text{sig}} - \eta||v^*||_{\text{cor}} - 2\sqrt{2\log(1/\beta)} - \tilde{\delta}}{\sqrt{n} - 2\delta}
\]
\[
= \frac{\delta}{\sqrt{n} - 2\delta}.
\]
where \(\eta_{\text{sig}} = \eta(t_{\text{sig}} \cdot \partial ||x^*||_{\text{sig}}), \eta_{\text{cor}} = \eta(t_{\text{cor}} \cdot \partial ||v^*||_{\text{cor}})\) and \(\tilde{\delta} = 2\sqrt{2\pi} - \frac{1}{\sqrt{2\pi}}\). □

VI. CONCLUSION AND FUTURE WORK
We presented a more general signal recovery model than that in [26]. Our convex programming approaches can recover both \(x\) and \(v\) with or without prior information, and we give recovery guarantees for both the two cases. Using geometry measures of signal and corruption, the number of measurements required for recovery from any convex program can be computed. Our simulations reinforced our theoretical results.

Considering that the constrained models are generally difficult to solve in practice, as pointed out in the future work of [26], it would be of great practical interest to analyze the unconstrained model in our general form and model their performance in theory. This will be included in our future research work.

REFERENCES
[1] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, “A constrained matching pursuit approach to audio declipping,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Prague, Czech Republic, May 2011, pp. 329–332.
[2] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, “Audio inpainting,” IEEE Trans. Audio, Speech, Lang. Process., vol. 20, no. 3, pp. 922–932, Sep. 2012.
[3] C. Aubel, C. Studer, G. Pope, and H. Bölcskei, “Sparse signal recovery in redundant dictionaries,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Boston, MA, USA, Aug. 2012, pp. 2047–2051.
[4] I.-F. Cai, S. Osher, and Z. Shen, “Split Bregman methods and frame based image restoration,” Multiscale Model. Simul., vol. 8, no. 2, pp. 337–369, 2010.
[5] T. T. Cai, L. Wang, and G. Xu, “Stable recovery of sparse signals and an oracle inequality,” IEEE Trans. Inf. Theory, vol. 56, no. 7, pp. 3516–3522, Jul. 2010.
[6] T. Cai, G. Xu, and J. Zhang, “On recovery of sparse signals via \(\ell_1\) minimization,” IEEE Trans. Inf. Theory, vol. 55, no. 7, pp. 3388–3397, Jun. 2009.
[7] T. T. Cai and A. Zhang, “Sharp RIP bound for sparse signal and low-rank matrix recovery,” Appl. Comput. Harmon. Anal., vol. 35, no. 1, pp. 74–93, Jul. 2013.
[8] T. T. Cai and A. Zhang, “Sparse representation of a polytope and recovery of sparse signals and low-rank matrices,” IEEE Trans. Inf. Theory, vol. 60, no. 1, pp. 122–132, Jan. 2014.
[9] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” J. ACM, vol. 58, no. 3, pp. 1–37, May 2011.
[10] E. J. Candès, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” Commun. Pure Appl. Math., vol. 59, no. 8, pp. 1207–1223, 2006.
[11] E. J. Candes and T. Tao, “Decoding by linear programming,” IEEE Trans. Inf. Theory, vol. 51, no. 12, pp. 4203–4215, Nov. 2005.
[12] E. J. Candes and T. Tao, “The Dantzig selector: Statistical estimation when \(p\) is much large than \(n\),” Ann. Statist., vol. 35, no. 6, pp. 2313–2351, 2007.
[13] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, “The convex geometry of linear inverse problems,” Found. Comput. Math., vol. 12, pp. 805–849, Oct. 2012.
[14] 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.
[15] R. Chartrand, “Exact reconstruction of sparse signals via nonconvex minimization,” IEEE Signal Process. Lett., vol. 14, no. 10, pp. 707–710, Oct. 2007.
[16] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” SIAM Rev., vol. 43, no. 1, pp. 129–159, Feb. 2001.
[17] I. T. Chu, “A modified Wallis product and some applications,” Amer. Math. Monthly, vol. 69, no. 5, pp. 402–404, 1962.
[18] D. L. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[19] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via \(\ell_1\) minimization,” Proc. Nat. Acad. Sci. USA, vol. 100, no. 5, pp. 2197–2202, 2003.
[20] D. L. Donoho, M. Elad, and V. N. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” IEEE Trans. Inf. Theory, vol. 52, no. 1, pp. 6–18, Jan. 2006.
[21] D. L. Donoho and X. Huo, “Uncertainty principles and ideal atomic decomposition,” IEEE Trans. Inf. Theory, vol. 47, no. 7, pp. 2848–2862, Nov. 2001.
[22] D. L. Donoho and P. B. Stark, “Uncertainty principles and signal recovery,” SIAM J. Appl. Math., vol. 49, no. 3, pp. 906–931, Jan. 1989.
[23] M. Elad and Y. Hel-Or, “A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur,” IEEE Trans. Image Process., vol. 10, no. 8, pp. 1187–1193, Aug. 2001.
[24] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, “Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA),” Appl. Comput. Harmon. Anal., vol. 19, no. 3, pp. 340–358, Nov. 2005.
[25] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Basel, Switzerland: Birkhäuser, 2013.

[26] R. Foygel and L. Mackey, “Corrupted sensing: Novel guarantees for separating structured signals,” *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1223–1247, Feb. 2014.

[27] J. J. Fuchs, “Recovery of exact sparse representations in the presence of bounded noise,” *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3601–3608, Oct. 2005.

[28] M. Grant, and S. Boyd. (Sep. 2012). CVX: MATLAB Software for Disciplined Convex Programming, Version 2.0 Beta. [Online]. Available: http://cvxr.com/cvx

[29] M. Grant and S. Boyd, “Graph implementations for nonsmooth convex programs,” in *Recent Advances in Learning and Control* (Lecture Notes in Control and Information Sciences), V. Blondel, S. Boyd, and H. Kimura, Eds. London, U.K.: Springer-Verlag, 2008, pp. 95–110.

[30] W. Göbel and F. Helmchen, “In vivo calcium imaging of neural network function,” *Physiology*, vol. 22, no. 6, pp. 358–365, Dec. 2007.

[31] Y. Gordon, “Some inequalities for Gaussian processes and applications,” *Isr. J. Math.*, vol. 50, no. 4, pp. 265–289, 1985.

[32] P. Kuppinger, G. Durisi, and H. Bölcskei, “Uncertainty relations and sparse signal recovery for pairs of general signal sets,” *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 263–277, Jan. 2012.

[33] J. N. Laska, P. T. Boufounos, M. A. Davenport, and R. G. Baraniuk, “Democracy in action: Quantization, saturation, and compressive sensing,” *Appl. Comput. Harmon. Anal.*, vol. 31, no. 3, pp. 429–443, Nov. 2011.

[34] J. N. Laska, M. A. Davenport, and R. G. Baraniuk, “Exact signal recovery from sparsely corrupted measurements through the pursuit of justice,” in *Proc. 43rd Asilomar Conf. Signals, Syst. Comput.*, Nov. 2009, pp. 1556–1560.

[35] M. Ledoux, “Isoperimetry and Gaussian analysis,” in *Lectures on Probability Theory and Statistics*. New York, NY, USA: Springer-Verlag, Jan. 1996, pp. 165–294.

[36] X. Li, “Compressed sensing and matrix completion with constant proportion of corruptions,” *Constructive Approximation*, vol. 37, no. 1, pp. 73–99, Dec. 2012.

[37] Z. Li, F. Wu, and J. Wright, “On the systematic measurement matrix for compressed sensing in the presence of gross errors,” in *Proc. Data Compress. Conf.*, 2010, pp. 356–365.

[38] S. Mallat and Y. Guoshen, “Super-resolution with sparse mixing estimators,” *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2889–2900, Nov. 2010.

[39] Q. Mo and S. Li, “New bounds on the restricted isometry constant $\delta_2^*$,” *Appl. Comput. Harmon. Anal.*, vol. 31, no. 3, pp. 460–468, Nov. 2011.

[40] N. Nguyen and T. Tran, “Exact recoverability from dense corrupted observations via $l_1$ minimization,” *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2017–2035, Apr. 2013.

[41] 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, May 2013.

[42] M. Stojnic, “Various thresholds for $l_1$-optimization in compressed sensing,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Jul. 2009, vol. 10, no. 22, pp. 3918–3921.

[43] M. Stojnic, F. Parvaresh, and B. Hassibi, “On the reconstruction of block-sparse signals with an optimal number of measurements,” *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.

[44] C. Studer and R. G. Baraniuk, “Stable restoration and separation of approximately sparse signals,” *Appl. Comput. Harmon. Anal.*, vol. 37, no. 1, pp. 12–35, Jul. 2014.

[45] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, “Recovery of sparsely corrupted signals,” *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3115–3130, May 2012.

[46] J. A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[47] J. A. Tropp, “Just relax: Convex programming methods for identifying sparse signals in noise,” *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[48] J. Wright and Y. Ma, “Dense error correction via $l_1$-minimization,” *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3540–3560, Sep. 2008.

[49] R. Zhang and S. Li, “Optimal RIP bounds for sparse signals recovery via $l_1$-minimization,” *Appl. Comput. Harmon. Anal.*, vol. 47, no. 3, pp. 566–584, Nov. 2019.

NINGNING LI received the B.S. degree from the Hebei University of Technology, in 2016, and the M.S. degree from the Institute of Applied Physics and Computational Mathematics, in 2019, where she is currently pursuing the Ph.D. degree. Her research interests include compressed sensing and mathematical image processing.

WENGU CHEN received the Ph.D. degree in mathematics from Beijing Normal University, in 1996. He is currently a Professor with the Institute of Applied Physics and Computational Mathematics. His research interests include harmonic analysis, compressed sensing, and machine learning.

* * *