On the Phase Transition of Corrupted Sensing

Huan Zhang*†, Yulong Liu‡, and Hong Lei*
*Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China
†University of Chinese Academy of Sciences, Beijing 100049, China
‡School of Physics, Beijing Institute of Technology, Beijing 100081, China

***Abstract*—In [1], a sharp phase transition has been numerically observed when a constrained convex procedure is used to solve the corrupted sensing problem. In this paper, we present a theoretical analysis for this phenomenon. Specifically, we establish the threshold below which this convex procedure fails to recover signal and corruption with high probability. Together with the work in [1], we prove that a sharp phase transition occurs around the sum of the squares of spherical Gaussian widths of two tangent cones. Numerical experiments are provided to demonstrate the correctness and sharpness of our results.**

***Index Terms*—Corrupted sensing, phase transition, Gaussian width, compressed sensing, signal separation.**

I. Introduction

Corrupted sensing aims to recover a structured signal from a small number of corrupted measurements

$$\boldsymbol{y} = \boldsymbol{\Psi}\boldsymbol{x}^\star + \boldsymbol{v}^\star, \tag{1}$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{m\times n}$ is the sensing measurement matrix which is assumed to have i.i.d. standard Gaussian entries in this paper, $\boldsymbol{x}^\star \in \mathbb{R}^n$ is the unknown signal, and $\boldsymbol{v}^\star \in \mathbb{R}^m$ is an unknown corruption. The goal is to estimate $\boldsymbol{x}^\star$ and $\boldsymbol{v}^\star$ from $\boldsymbol{y}$ and $\boldsymbol{\Psi}$.

This problem is encountered in many practical applications, such as face recognition [2], subspace clustering [3], network data analysis [4], and so on. Theoretical guarantees for this problem include sparse signal recovery from sparse corruption [5]–[11] and structured signal recovery from structured corruption [1], [12], [13].

To make the recovery possible, we will assume that both $\boldsymbol{x}$ and $\boldsymbol{v}$ have some structures which are promoted by the convex functions $f(\cdot)$ and $g(\cdot)$ respectively. When prior information about $f(\boldsymbol{x}^\star)$ or $g(\boldsymbol{v}^\star)$ is available, it is natural to consider the following program to recover the signal and corruption:

$$\min f(\boldsymbol{x}), \quad \text{s.t. } \boldsymbol{y} = \boldsymbol{\Psi}\boldsymbol{x} + \boldsymbol{v}, \quad g(\boldsymbol{v}) \le g(\boldsymbol{v}^\star), \tag{2}$$

or

$$\min g(\boldsymbol{v}), \quad \text{s.t. } \boldsymbol{y} = \boldsymbol{\Psi}\boldsymbol{x} + \boldsymbol{v}, \quad f(\boldsymbol{x}) \le f(\boldsymbol{x}^\star). \tag{3}$$

In [1], Foygel and Mackey provided conditions under which convex program (2) or (3) succeeds with high probability. Numerical experiments in [1] also suggested that there is a sharp phase transition when (2) or (3) is used to solve the corrupted sensing problem. However, little work has devoted to determining the threshold below which (2) or (3) fails with high probability. Therefore, theoretical understanding of the phase transition for program (2) and (3) is far from satisfactory.

In this paper, we present a theoretical analysis for the phase transition of (2) or (3). In particular, we figure out the exact position of phase transition, and demonstrate that the phase transition occurs in a relatively narrow region.

This work was supported by the National Natural Science Foundation of China under Grant 61301188.

II. Preliminaries

In this section, we present some preliminaries which will be used in our analysis.

Our result involves two important concepts: the Gaussian width and the tangent cone. Given a subset T in $\mathbb{R}^n$, the Gaussian width is defined by

$$\omega(T) = \mathbb{E} \sup_{\boldsymbol{t}\in T} \langle \boldsymbol{g}, \boldsymbol{t}\rangle, \text{ where } \boldsymbol{g} \sim N(0, I_n).$$

We also define two tangent cones corresponding to signal and corruption respectively. The tangent cone of $f(\cdot)$ at the true signal $\boldsymbol{x}^\star$ is defined as

$$\mathcal{D}_s = \left\{\boldsymbol{a} \in \mathbb{R}^n : \exists\, t > 0, f(\boldsymbol{x}^\star + \boldsymbol{a}t) \le f(\boldsymbol{x}^\star)\right\}. \tag{4}$$

Similarly, the tangent cone of $g(\cdot)$ at the true corruption $\boldsymbol{v}^\star$ is given by

$$\mathcal{D}_c = \left\{\boldsymbol{b} \in \mathbb{R}^m : \exists\, t > 0, g(\boldsymbol{v}^\star + \boldsymbol{b}t) \le g(\boldsymbol{v}^\star)\right\}. \tag{5}$$

III. Main results

In this section, we state our main results with some discussions.

Theorem 1 (Failure of convex program (2) or (3))**.** *Consider convex program (2) or (3). Assume that both tangent cones $\mathcal{D}_s$ and $\mathcal{D}_c$ are closed. For any $t \ge 0$, if the measurement number m satisfies*

$$\sqrt{m} < \sqrt{\omega^2\left(\mathcal{D}_s \cap S^{n-1}\right) + \omega^2\left(\mathcal{D}_c \cap S^{m-1}\right)} - t,$$

then the constrained convex program (2) or (3) fails with probability at least $1 - \exp(-t^2/2)$, where S^{n-1} and S^{m-1} are the unit sphere of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively.

Proof. See Appendix A. □

Remark 1 (Phase transition of corrupted sensing)**.** *Recall Theorem* 1 *and Remark* 2 *in [1], which stated that* [1] [2] *when*

$$\sqrt{m} \geq \sqrt{\omega^2(\mathcal{D}_s \cap S^{n-1}) + \omega^2(\mathcal{D}_c \cap S^{m-1})} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2\pi}} + t,$$

the constrained convex program (2) or (3) succeeds with probability at least $1 - \exp(-t^2/2)$*. This, together with our result Theorem 1, demonstrate that the phase transition of corrupted sensing occurs around*

$$\omega^2\big(\mathcal{D}_s \cap S^{n-1}\big) + \omega^2\big(\mathcal{D}_c \cap S^{m-1}\big),$$

and the width of phase transition area is about

$$C\sqrt{\omega^2\big(\mathcal{D}_s \cap S^{n-1}\big) + \omega^2\big(\mathcal{D}_c \cap S^{m-1}\big)},$$

where C *is an absolute constant.*

Remark 2. *Our result also agrees with the result of Amelunxen el al. [14]. Indeed, by Proposition 10.2 and Proposition 3.1 (9) in [14], we have*

$$\omega^2\big(\mathcal{D}_s \cap S^{n-1}\big) + \omega^2\big(\mathcal{D}_c \cap S^{m-1}\big) \approx \delta(\mathcal{D}_s) + \delta(\mathcal{D}_c) = \delta(\mathcal{D}_s \times \mathcal{D}_c),$$

where $\delta(\mathcal{D})$ *denotes the statistical dimension of a convex cone* $\mathcal{D}$*.*

Remark 3. *In [14], Amelunxen et al. considered the phase transition of the following demixing problem:*

$$\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{U}\boldsymbol{y},$$

where $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ *are unknown signals and* $\boldsymbol{U} \in \mathbb{R}^{n \times n}$ *is a random orthogonal matrix. This model is different from ours since we have random Gaussian measurement matrix with* $m \ll n$*.*

Remark 4. *In [15], Oymak and Tropp considered the phase transition of the following demixing model:*

$$\boldsymbol{y} = \boldsymbol{\Psi}_0 \boldsymbol{x}_0 + \boldsymbol{\Psi}_1 \boldsymbol{x}_1,$$

where $\boldsymbol{x}_0, \boldsymbol{x}_1 \in \mathbb{R}^n$ *are two signals and* $\boldsymbol{\Psi}_0, \boldsymbol{\Psi}_1 \in \mathbb{R}^{m \times n}$ *are some random transformation matrices. This model is also different from ours since* $\boldsymbol{\Psi}_1$ *is a deterministic matrix in our case. This makes the problem more difficult to analyze.*

IV. Simulation Results

In this section, we employ a numerical experiment to verify our theoretical guarantees (Theorem 1). In the experiment, both signal and corruption are designed to be sparse vectors. We use CVX [16] [17] to solve the convex program (2) or (3).

In the experiment, we assume that the prior information of $f(\boldsymbol{x}^\star)$ is known exactly, and solve program (3). The experiment settings are as follows: the ambient dimension n is set to 128, the measurement number $m = n = 128$, the sparsity level of signal changes from 1 to n with step 1, and the same for corruption. For every sparsity level of signal and corruption, we run and solve (3) 20 times. We declare success if the solution to (3), denoted by $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{v}})$, satisfies $\|\hat{\boldsymbol{x}} - \boldsymbol{x}^\star\|_2 \leq 10^{-3}$. Then we get the empirical probability of successful recovery. At last, we plot the theoretical curve predicted by Theorem 1.

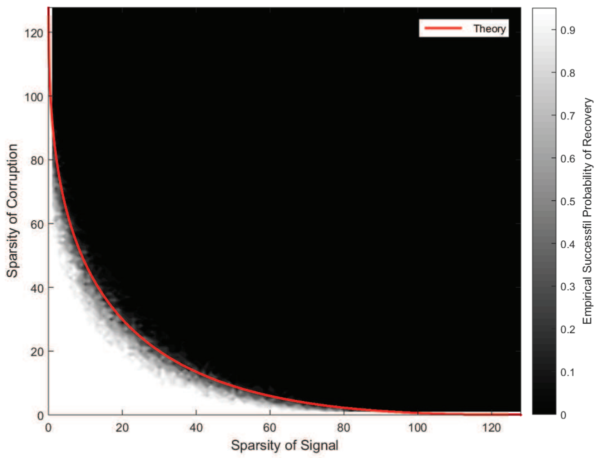

Fig. 1. Phase transition for constrained convex program 3. The red curve plots the phase transition threshold predicted by Theorem 1.

Our numerical experiment result is shown in Fig. 1. We can see that the theoretical threshold given by Theorem 1 is closely matched with the empirical phase transition. It means that our theory can give a reliable prediction of the phase transition curve.

V. Conclusion

This paper studied the problem of phase transition when we use convex program to solve corrupted sensing problem. Our results, together with previous work [1], gave the exact location of phase transition and the size of transition region. Simulations were provided to verify the correctness of our results. Our ongoing work is to establish a general framework to analyze the phase transition of various convex programs with noise-free or noisy data.

Appendix A
Proof of Main Results

In this section, we present proof for our main result (Theorem 1). First, we will establish a sufficient condition under which convex program (2) or (3) fails, then some necessary tools are introduced, and at last, we give the proof for Theorem 1.

A. Sufficient Condition for failure

In this subsection, we establish an easy-to-handle sufficient condition under which program (2) or (3) fails.

Lemma 1. *Let* $\mathcal{D}_s$ *and* $\mathcal{D}_c$ *denote the signal and the corruption tangent cones defined in (4) and (5) respectively. Then a sufficient condition under which constrained convex program (2) or (3) fails is*

$$\min_{(\boldsymbol{a},\boldsymbol{b}) \in (\mathcal{D}_s \times \mathcal{D}_c) \cap S^{n+m-1}} \big\|\boldsymbol{\Psi}\boldsymbol{a} + \boldsymbol{b}\big\| = 0. \qquad (6)$$

[1]The authors believe that the small additive constants are artifacts of the proof technique.

[2]The original result is stated in terms of Gaussian complexity $\gamma(\mathcal{D}_s \cap B^n)$, difined as $\gamma^2(\mathcal{D}_s \cap B^n) = \mathbb{E}\big(\sup_{t \in \mathcal{D}_s \cap B^n} \langle g, t\rangle\big)^2$, where B^n denotes the ℓ_2 unit ball in $\mathbb{R}^n$. However, as the author stated, the Gaussian complexity $\gamma(\mathcal{D}_s \cap B^n)$ is only very slightly larger than $\omega(\mathcal{D}_s \cap S^{n-1})$.

In other words, the subset $\mathcal{D}_s \times \mathcal{D}_c \cap S^{n+m-1}$ intersects the null space of matrix $\begin{bmatrix}\boldsymbol{\Psi} & \boldsymbol{I}\end{bmatrix}$.

Proof. Lemma 1 is a generalization of Proposition 2.1 of [18]. The proof is similar, and hence is omitted. ☐

Although Lemma 1 gives a sufficient condition for failure, it is difficult to check when (6) holds. The following lemma can overcome this drawback.

Lemma 2 (Sufficient condition for failure, Proposition 3.8, [15])**.** *Under the condition of Lemma 1, if both $\mathcal{D}_s$ and $\mathcal{D}_c$ are closed, a sufficient condition for (6) to hold is*

$$\min_{\|\boldsymbol{r}\|=1} \min_{\boldsymbol{s}\in(\mathcal{D}_s\times\mathcal{D}_c)^\circ} \left\|\boldsymbol{s}-\boldsymbol{A}^*\boldsymbol{r}\right\| > 0, \tag{7}$$

where $(\mathcal{D}_s \times \mathcal{D}_c)^\circ$ denotes the polar cone of $\mathcal{D}_s \times \mathcal{D}_c$, $\boldsymbol{A} = \begin{bmatrix}\boldsymbol{\Psi} & \boldsymbol{I}\end{bmatrix}$, and $\boldsymbol{I}$ denotes the identity matrix.

Remark 5. *One can easily check that*

$$(\mathcal{D}_s \times \mathcal{D}_c)^\circ = \mathcal{D}_s^\circ \times \mathcal{D}_c^\circ.$$

Thus, the sufficient condition under which convex program (2) or (3) fails can be rewritten as

$$\min_{\|\boldsymbol{r}\|=1} \min_{\boldsymbol{s}\in\mathcal{D}_s^\circ\times\mathcal{D}_c^\circ} \left\|\boldsymbol{s}-\boldsymbol{A}^*\boldsymbol{r}\right\| > 0. \tag{8}$$

In the following parts, we will prove that (8) holds with high probability when the condition of Theorem 1 is satisfied. Before this, let's state some tools that will be used in our proof.

B. Other Useful Tools

Lemma 3 (Gordon's inequality, Theorem 3.16, [19])**.** *Let $(X_{\boldsymbol{ut}})_{\boldsymbol{u}\in U,\boldsymbol{t}\in T}$ and $(Y_{\boldsymbol{ut}})_{\boldsymbol{u}\in U,\boldsymbol{t}\in T}$ be two Gaussian processes indexed by pairs of points $(\boldsymbol{u},\boldsymbol{t})$ in a product set $U \times T$. Assume that*

$$\mathbb{E}(X_{\boldsymbol{ut}} - X_{\boldsymbol{us}})^2 \le \mathbb{E}(Y_{\boldsymbol{ut}} - Y_{\boldsymbol{us}})^2 \quad \text{for all } \boldsymbol{u},\boldsymbol{t},\boldsymbol{s};$$

$$\mathbb{E}(X_{\boldsymbol{ut}} - X_{\boldsymbol{vs}})^2 \ge \mathbb{E}(Y_{\boldsymbol{ut}} - Y_{\boldsymbol{vs}})^2 \quad \text{for all } \boldsymbol{u}\neq\boldsymbol{v} \text{ and all } \boldsymbol{t},\boldsymbol{s}.$$

Then we have

$$\mathbb{E} \inf_{\boldsymbol{u}\in U}\sup_{\boldsymbol{t}\in T} X_{\boldsymbol{ut}} \le \mathbb{E} \inf_{\boldsymbol{u}\in U}\sup_{\boldsymbol{t}\in T} Y_{\boldsymbol{ut}}.$$

Lemma 4 (Concentration of measure, Theorem 5.6, [20])**.** *Let $X = (X_1, \dots, X_n)$ be a vector of n independent standard normal random variables. Let $f : \mathbb{R}^n \to \mathbb{R}$ denotes an L-Lipschitz function. Then, for all $t \ge 0$,*

$$\mathbb{P}\{f(X) - \mathbb{E}f(X) \ge t\} \le e^{-t^2/(2L^2)}.$$

Lemma 5 (Lemma 3.7, [18])**.** *Let $\mathcal{D} \subset \mathbb{R}^n$ be a non-empty closed, convex cone. Then we have that*

$$\omega^2(\mathcal{D}\cap S^{n-1}) + \omega^2(\mathcal{D}^\circ\cap S^{n-1}) \le n.$$

Lemma 6. *Let Ω_1 and Ω_2 be subsets of S^{m-1} and S^{n-1} respectively. Then the function*

$$F(\boldsymbol{\Psi}) = \min_{\boldsymbol{t}\in\Omega_1}\max_{\boldsymbol{u}\in\Omega_2} \langle \boldsymbol{\Psi u}, \boldsymbol{t}\rangle$$

is a 1-Lipschitz function, where $\boldsymbol{\Psi}$ is the same as in (1).

Proof. See Appendix B. ☐

C. Proof of Main Results

According to Remark 5, we only need to prove that when

$$\sqrt{m} < \sqrt{\omega^2\big(\mathcal{D}_s\cap S^{n-1}\big) + \omega^2\big(\mathcal{D}_c\cap S^{m-1}\big)} - t,$$

the following event

$$\min_{\|\boldsymbol{r}\|=1} \min_{\boldsymbol{s}\in\mathcal{D}_s^\circ\times\mathcal{D}_c^\circ} \left\|\boldsymbol{s}-\boldsymbol{A}^*\boldsymbol{r}\right\| > 0$$

holds with probability at least $1 - e^{-t^2/2}$. Moreover, a simple calculation verifies that this inequality is equivalent to

$$\begin{aligned}
&\min_{\|\boldsymbol{r}\|=1} \min_{\boldsymbol{s}\in\mathcal{D}_s^\circ\times\mathcal{D}_c^\circ} \|\boldsymbol{s}-\boldsymbol{A}^*\boldsymbol{r}\|_2 > 0\\
\iff\ & \min_{\|\boldsymbol{r}\|=1} \min_{\boldsymbol{s}\in\mathcal{D}_s^\circ\times\mathcal{D}_c^\circ} \|\boldsymbol{s}-\boldsymbol{A}^*\boldsymbol{r}\|_2^2 > 0\\
\iff\ & \min_{\|\boldsymbol{r}\|=1} \min_{\substack{\boldsymbol{s}_1\in\mathcal{D}_s^\circ\\ \boldsymbol{s}_2\in\mathcal{D}_c^\circ}} \left[\left\|\boldsymbol{s}_1-\boldsymbol{\Psi}^*\boldsymbol{r}\right\|_2^2 + \left\|\boldsymbol{s}_2-\boldsymbol{r}\right\|_2^2\right] > 0.
\end{aligned} \tag{9}$$

Now, we will consider two cases for $\boldsymbol{r}$:
Case I: $\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}$. In this case, when we minimize over $\boldsymbol{s}_2$, the second term $\left\|\boldsymbol{s}_2-\boldsymbol{r}\right\|_2^2$ will be zero. Thus, the above inequality (9) is equivalent to

$$\begin{aligned}
&\min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \min_{\substack{\boldsymbol{s}_1\in\mathcal{D}_s^\circ\\ \boldsymbol{s}_2\in\mathcal{D}_c^\circ}} \left[\left\|\boldsymbol{s}_1-\boldsymbol{\Psi}^*\boldsymbol{r}\right\|_2^2 + \left\|\boldsymbol{s}_2-\boldsymbol{r}\right\|_2^2\right] > 0\\
&\iff \min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \min_{\boldsymbol{s}_1\in\mathcal{D}_s^\circ} \left\|\boldsymbol{s}_1-\boldsymbol{\Psi}^*\boldsymbol{r}\right\|_2^2 > 0\\
&\iff \min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \min_{\boldsymbol{s}_1\in\mathcal{D}_s^\circ} \left\|\boldsymbol{s}_1-\boldsymbol{\Psi}^*\boldsymbol{r}\right\|_2 > 0.
\end{aligned} \tag{10}$$

For our purpose, we need to lower bound the left side of (10). Note that for any fixed $\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}$, we have

$$\begin{aligned}
\min_{\boldsymbol{s}_1\in\mathcal{D}_s^\circ} \|\boldsymbol{s}_1-\boldsymbol{\Psi}^*\boldsymbol{r}\|_2 &= \min_{\boldsymbol{s}_1\in\mathcal{D}_s^\circ} \max_{\boldsymbol{u}\in S^{n-1}} \langle \boldsymbol{u}, \boldsymbol{\Psi}^*\boldsymbol{r}-\boldsymbol{s}_1\rangle\\
&\ge \max_{\boldsymbol{u}\in S^{n-1}} \min_{\boldsymbol{s}_1\in\mathcal{D}_s^\circ} \langle \boldsymbol{u}, \boldsymbol{\Psi}^*\boldsymbol{r}-\boldsymbol{s}_1\rangle\\
&= \max_{\boldsymbol{u}\in S^{n-1}} \left[\langle \boldsymbol{u}, \boldsymbol{\Psi}^*\boldsymbol{r}\rangle - \max_{\boldsymbol{s}\in\mathcal{D}_s^\circ}\langle \boldsymbol{u}, \boldsymbol{s}\rangle\right]\\
&= \max_{\boldsymbol{u}\in\mathcal{D}_s\cap S^{n-1}} \langle \boldsymbol{u}, \boldsymbol{\Psi}^*\boldsymbol{r}\rangle\\
&= \max_{\boldsymbol{u}\in\mathcal{D}_s\cap S^{n-1}} \langle \boldsymbol{\Psi u}, \boldsymbol{r}\rangle.
\end{aligned}$$

The first equality is due to the definition of ℓ_2-norm. The first inequality is because of the minimax inequality. The second equality comes from the linear property of inner product. The third equality uses the fact that $\max_{\boldsymbol{s}\in\mathcal{D}_s^\circ}\langle \boldsymbol{u}, \boldsymbol{s}\rangle = 0$ when $\boldsymbol{u}\in\mathcal{D}_s$, otherwise it equals ∞. The last equality can be derived by a simple transformation. As the above inequality holds for any $\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}$, we have

$$\begin{aligned}
&\min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \min_{\boldsymbol{s}_1\in\mathcal{D}_s^\circ} \|\boldsymbol{s}_1-\boldsymbol{\Psi}^*\boldsymbol{r}\|_2\\
&\qquad \ge \min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \max_{\boldsymbol{u}\in\mathcal{D}_s\cap S^{n-1}} \langle \boldsymbol{\Psi u}, \boldsymbol{r}\rangle.
\end{aligned} \tag{11}$$

It remains to bound the right side. To this end, we will first use Gordon's inequality (Lemma 3) to derive a lower bound for the expectation, and then concentration of measure (Lemma 4) to obtain the desired result. Let $X_{\boldsymbol{ru}} := \langle \boldsymbol{\Psi u}, \boldsymbol{r}\rangle$ and $Y_{\boldsymbol{ru}} := \langle \boldsymbol{g}, \boldsymbol{r}\rangle + \langle \boldsymbol{h}, \boldsymbol{u}\rangle$ be two Gaussian processes, where $\boldsymbol{g} \sim N(\boldsymbol{0}, \boldsymbol{I}_{m\times m})$ and $\boldsymbol{h} \sim N(\boldsymbol{0}, \boldsymbol{I}_{n\times n})$ are independent standard Gaussian random vectors. It can be easily checked that the increments satisfy

$$\mathbb{E}(X_{\boldsymbol{ru}} - X_{\boldsymbol{ru}'})^2 = \left\|\boldsymbol{u} - \boldsymbol{u}'\right\|_2^2 = \mathbb{E}(Y_{\boldsymbol{ru}} - Y_{\boldsymbol{ru}'})^2,$$

$$\begin{aligned}\mathbb{E}(X_{\boldsymbol{ru}} - X_{\boldsymbol{r}'\boldsymbol{u}'})^2 &= \left\|\boldsymbol{u}\boldsymbol{r}^T - \boldsymbol{u}'\boldsymbol{r}'^T\right\|_F^2 \\ &\le \left\|\boldsymbol{u} - \boldsymbol{u}'\right\|_2^2 + \left\|\boldsymbol{r} - \boldsymbol{r}'\right\|_2^2 \\ &= \mathbb{E}(Y_{\boldsymbol{ru}} - Y_{\boldsymbol{r}'\boldsymbol{u}'})^2.\end{aligned}$$

Therefore, Gordon's inequality (Lemma 3) gives us:

$$\begin{aligned}&\mathbb{E} \min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \max_{\boldsymbol{u}\in\mathcal{D}_s\cap S^{n-1}} X_{\boldsymbol{ru}} \\ &\quad\ge \mathbb{E} \min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \max_{\boldsymbol{u}\in\mathcal{D}_s\cap S^{n-1}} Y_{\boldsymbol{ru}} \\ &\quad= \mathbb{E} \min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \langle \boldsymbol{g}, \boldsymbol{r}\rangle + \mathbb{E} \max_{\boldsymbol{u}\in\mathcal{D}_s\cap S^{n-1}} \langle \boldsymbol{h}, \boldsymbol{u}\rangle. \quad (12)\end{aligned}$$

Since $\boldsymbol{g}$ is a symmetric random vector, we have

$$\begin{aligned}\mathbb{E} \min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \langle \boldsymbol{g}, \boldsymbol{r}\rangle &= \mathbb{E} \min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \langle -\boldsymbol{g}, \boldsymbol{r}\rangle \\ &= -\mathbb{E} \max_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \langle \boldsymbol{g}, \boldsymbol{r}\rangle \\ &= -\omega(\mathcal{D}_c^\circ \cap S^{m-1}).\end{aligned}$$

Substituting this into (12), we get

$$\mathbb{E} \min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \max_{\boldsymbol{u}\in\mathcal{D}_s\cap S^{n-1}} X_{\boldsymbol{ru}} \ge \omega(\mathcal{D}_s\cap S^{n-1}) - \omega(\mathcal{D}_c^\circ\cap S^{m-1}). \quad (13)$$

As $\mathcal{D}_c$ is a closed convex cone, by Lemma 5, we know that

$$\omega^2\big(\mathcal{D}_c^\circ \cap S^{m-1}\big) + \omega^2\big(\mathcal{D}_c \cap S^{m-1}\big) \le m,$$

which implies

$$\omega(\mathcal{D}_c^\circ \cap S^{m-1}) \le \sqrt{m - \omega^2(\mathcal{D}_c \cap S^{m-1})}.$$

Substituting this into (13), we get the following result:

$$\begin{aligned}&\mathbb{E} \min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \max_{\boldsymbol{u}\in\mathcal{D}_s\cap S^{n-1}} \langle \boldsymbol{\Psi u}, \boldsymbol{r}\rangle \\ &\quad\ge \omega(\mathcal{D}_s \cap S^{n-1}) - \sqrt{m - \omega^2(\mathcal{D}_c \cap S^{m-1})} \\ &\quad\ge \sqrt{\omega^2(\mathcal{D}_s \cap S^{n-1}) + \omega^2(\mathcal{D}_c \cap S^{m-1})} - \sqrt{m}. \quad (14)\end{aligned}$$

In the last inequality, we have used the assumption that $\omega^2(\mathcal{D}_s \cap S^{n-1}) + \omega^2(\mathcal{D}_c \cap S^{m-1}) > m$.

Next, Lemma 6 confirms that the following function

$$\min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \max_{\boldsymbol{u}\in\mathcal{D}_s\cap S^{n-1}} \langle \boldsymbol{\Psi u}, \boldsymbol{r}\rangle$$

is a 1-Lipschitz function. Thus, concentration of measure (Lemma 4) gives us that for any $t \ge 0$,

$$\begin{aligned}&\mathbb{P}\Big\{ \min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \max_{\boldsymbol{u}\in\mathcal{D}_s\cap S^{n-1}} \langle \boldsymbol{\Psi u}, \boldsymbol{r}\rangle - \\ &\qquad \mathbb{E} \min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \max_{\boldsymbol{u}\in\mathcal{D}_s\cap S^{n-1}} \langle \boldsymbol{\Psi u}, \boldsymbol{r}\rangle \ge -t\Big\} \\ &\qquad \ge 1 - \exp(-t^2/2).\end{aligned}$$

Putting the above inequality and (14), (11), (9), (10) together, we eventually get that when

$$\sqrt{m} < \sqrt{\omega^2\big(\mathcal{D}_s \cap S^{n-1}\big) + \omega^2\big(\mathcal{D}_c \cap S^{m-1}\big)} - t,$$

we have

$$\mathbb{P}\Big\{ \min_{\boldsymbol{r}\in\mathcal{D}_c^\circ\cap S^{m-1}} \min_{\boldsymbol{s}\in\mathcal{D}_s\times\mathcal{D}_s^\circ} \|\boldsymbol{s} - \boldsymbol{A}^*\boldsymbol{r}\|_2 > 0\Big\} \ge 1 - \exp(-t^2/2).$$

Case II: $\boldsymbol{r} \notin \mathcal{D}_c^\circ \cap S^{m-1}$. In this case, it is clear that no matter what $\boldsymbol{r}$ and $\boldsymbol{s}_2$ takes value, it is always holds that

$$\left\|\boldsymbol{s}_2 - \boldsymbol{r}\right\|_2^2 > 0.$$

Thus,

$$\mathbb{P}\Big\{ \min_{\boldsymbol{r}\in S^{m-1}\setminus(\mathcal{D}_c^\circ\cap S^{m-1})} \min_{\boldsymbol{s}_1\in\mathcal{D}_s^\circ} \|\boldsymbol{s}_1 - \boldsymbol{\Psi}^*\boldsymbol{r}\|_2 > 0\Big\} = 1,$$

which, by (9) and (10), implies that

$$\mathbb{P}\Big\{ \min_{\boldsymbol{r}\in S^{m-1}\setminus(\mathcal{D}_c^\circ\cap S^{m-1})} \min_{\boldsymbol{s}\in\mathcal{D}_s\times\mathcal{D}_s^\circ} \|\boldsymbol{s} - \boldsymbol{A}^*\boldsymbol{r}\|_2 > 0\Big\} = 1.$$

Union bound. Combining case I and case II and taking a union bound, we have

$$\mathbb{P}\Big\{ \min_{\|\boldsymbol{r}\|_2=1} \min_{\boldsymbol{s}\in\mathcal{D}_s\times\mathcal{D}_s^\circ} \|\boldsymbol{s} - \boldsymbol{A}^*\boldsymbol{r}\|_2 > 0\Big\} \ge 1 - \exp(-t^2/2),$$

provided

$$\sqrt{m} < \sqrt{\omega^2\big(\mathcal{D}_s \cap S^{n-1}\big) + \omega^2\big(\mathcal{D}_c \cap S^{m-1}\big)} - t.$$

By Lemma 1 and Lemma 2, it means that when

$$\sqrt{m} < \sqrt{\omega^2\big(\mathcal{D}_s \cap S^{n-1}\big) + \omega^2\big(\mathcal{D}_c \cap S^{m-1}\big)} - t,$$

the convex program (2) or (3) fails with probability at least $1 - \exp(-t^2/2)$. This completes the proof.

Appendix B
Proof of Lemma 6

To prove Lemma 6, we only need to show that for any $\boldsymbol{C}, \boldsymbol{D} \in \mathbb{R}^{m\times n}$

$$\begin{aligned}\big|F(\boldsymbol{C}) - F(\boldsymbol{D})\big| &= \Big|\min_{\boldsymbol{t}\in\Omega_1} \max_{\boldsymbol{u}\in\Omega_2} \langle \boldsymbol{Cu}, \boldsymbol{t}\rangle - \min_{\boldsymbol{t}\in\Omega_1} \max_{\boldsymbol{u}\in\Omega_2} \langle \boldsymbol{Du}, \boldsymbol{t}\rangle\Big| \\ &\le \|\boldsymbol{C} - \boldsymbol{D}\|_F.\end{aligned}$$

For any fixed $\boldsymbol{t} \in \Omega_1$, let

$$\boldsymbol{u}_0(\boldsymbol{t}) \in \arg\max_{\boldsymbol{u}\in\Omega_2} \langle \boldsymbol{Cu}, \boldsymbol{t}\rangle.$$

And we have

$$\max_{\boldsymbol{u}\in\Omega_2} \langle \boldsymbol{Du}, \boldsymbol{t}\rangle \ge \langle \boldsymbol{D}\boldsymbol{u}_0(\boldsymbol{t}), \boldsymbol{t}\rangle.$$

Then, let

$$\boldsymbol{t}_0 \in \arg\min_{\boldsymbol{t}\in\Omega_1} \langle \boldsymbol{D}\boldsymbol{u}_0(\boldsymbol{t}), \boldsymbol{t}\rangle,$$

and we have

$$\begin{aligned}F(\boldsymbol{C}) = \min_{\boldsymbol{t}\in\Omega_1} \max_{\boldsymbol{u}\in\Omega_2} \langle \boldsymbol{Cu}, \boldsymbol{t}\rangle &= \min_{\boldsymbol{t}\in\Omega_1} \langle \boldsymbol{C}\boldsymbol{u}_0(\boldsymbol{t}), \boldsymbol{t}\rangle \\ &\le \langle \boldsymbol{C}\boldsymbol{u}_0(\boldsymbol{t}_0), \boldsymbol{t}_0\rangle.\end{aligned}$$

Similarly,

$$\begin{aligned}F(\boldsymbol{D}) = \min_{\boldsymbol{t}\in\Omega_1} \max_{\boldsymbol{u}\in\Omega_2} \langle \boldsymbol{Du}, \boldsymbol{t}\rangle &\ge \min_{\boldsymbol{t}\in\Omega_1} \langle \boldsymbol{D}\boldsymbol{u}_0(\boldsymbol{t}), \boldsymbol{t}\rangle \\ &= \langle \boldsymbol{D}\boldsymbol{u}_0(\boldsymbol{t}_0), \boldsymbol{t}_0\rangle.\end{aligned}$$

Therefore,

$$\begin{aligned} F(\boldsymbol{C}) - F(\boldsymbol{D}) &\leq \langle \boldsymbol{C}\boldsymbol{u}_0(\boldsymbol{t}_0), \boldsymbol{t}_0 \rangle - \langle \boldsymbol{D}\boldsymbol{u}_0(\boldsymbol{t}_0), \boldsymbol{t}_0 \rangle \\ &= \langle (\boldsymbol{C} - \boldsymbol{D})\boldsymbol{u}_0(\boldsymbol{t}_0), \boldsymbol{t}_0 \rangle \\ &\leq \big\| (\boldsymbol{C} - \boldsymbol{D})\boldsymbol{u}_0(\boldsymbol{t}_0) \big\|_2 \big\| \boldsymbol{t}_0 \big\|_2 \\ &\leq \big\| \boldsymbol{C} - \boldsymbol{D} \big\|_2 \leq \| \boldsymbol{C} - \boldsymbol{D} \|_F. \end{aligned} \tag{15}$$

The same argument gives

$$F(\boldsymbol{D}) - F(\boldsymbol{C}) \leq \| \boldsymbol{C} - \boldsymbol{D} \|_F. \tag{16}$$

Thus, combining (15) and (16), we get

$$\big| F(\boldsymbol{C}) - F(\boldsymbol{D}) \big| \leq \| \boldsymbol{C} - \boldsymbol{D} \|_F.$$

The conclusion follows immediately.

REFERENCES

[1] 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.

[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, Feb. 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, Mar. 2008.

[5] J. Wright and Y. Ma, "Dense error correction via ℓ_1-minimization," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3540–3560, Jul. 2010.

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

[7] N. H. Nguyen and T. D. Tran, "Exact recoverability from dense corrupted observations via ℓ_1-minimization," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2017–2035, Jan. 2013.

[8] ——, "Robust lasso with missing and grossly corrupted observations," *IEEE Trans. Inf. Theory*, vol. 4, no. 59, pp. 2036–2058, Apr. 2013.

[9] A. B. G. Pope and C. Studer, "Probabilistic recovery gaurantees for sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3104–3116, Jan. 2013.

[10] G. P. C. Studer, P. Kuppinger and H. Bolcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3115–3130, Dec. 2012.

[11] 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, Jul. 2014.

[12] 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.

[13] J. Chen and Y. Liu, "Corrupted sensing with sub-gaussian measurements," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, 2017, to appear.

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

[15] S. Oymak and J. A. Tropp, "Universality laws for randomized dimension reduction, with applications," 2015, [Online]. Available: https://arxiv.org/abs/1511.09433 preprint.

[16] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[17] ——, "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.

[18] 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, Dec. 2012.

[19] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: isoperimetry and processes*. Berlin Heidelberg: Springer-Verlag, 1991.

[20] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford: Oxford University Press, 2013.