Corrupted Sensing: Novel Guarantees for Separating Structured Signals

Rina Foygel and Lester Mackey
Department of Statistics, Stanford University

Abstract

We study the problem of corrupted sensing, a generalization of compressed sensing in which one aims to recover a signal from a collection of corrupted or unreliable measurements. While an arbitrary signal cannot be recovered in the face of arbitrary corruption, tractable recovery is possible when both signal and corruption are suitably structured. We quantify the relationship between signal recovery and two geometric measures of structure, the Gaussian width of a tangent cone and the Gaussian distance to a subdifferential. We take a convex programming approach to disentangling signal and corruption, analyzing both penalized programs that trade off between signal and corruption complexity and constrained programs that bound the complexity of signal or corruption when prior information is available. In each case, we provide conditions for exact signal recovery from structured corruption and stable signal recovery from structured corruption with added unstructured noise. Our simulations demonstrate close agreement between our theoretical recovery bounds and the sharp phase transitions observed in practice. In addition, we provide new analytical bounds for the Gaussian width of sparse vectors, block-sparse vectors, and low-rank matrices, which lead to sharper guarantees of recovery when combined with our results and those in the literature.

Index Terms

Corrupted sensing, compressed sensing, deconvolution, error correction, structured signal, sparsity, block sparsity, low rank, atomic norms, $\ell_1$ minimization.

I. INTRODUCTION

In the corrupted sensing problem, our goal is to recover a structured signal from a collection of potentially corrupted measurements. Recent years have seen a flurry of interest in specific instances of this problem, including sparse vector recovery from sparsely corrupted measurements [1] and the recovery of low-rank matrices from sparse corruption [2, 3]. The former arises in applications such as face recognition [4] and in the analysis of sensor network data [5]; the latter arises in problems ranging from latent variable modeling [2] to video background subtraction [3]. In the present work, we are more broadly interested in the deconvolution of an arbitrary signal-corruption pair. While the problem is generally ill-posed, one might hope that recovery is possible when both signal and corruption are suitably structured.

Corrupted sensing can be viewed as a generalization of the compressed sensing problem, in which one aims to recover a structured signal from a relatively small number of measurements. This problem is ubiquitous in modern applications, where one is often interested in estimating a high-dimensional signal from a number of measurements far smaller than the ambient dimension. It is now common practice, when the signal has underlying low-dimensional structure, to promote that structure via a convex penalty and thereby achieve accurate estimation in the face of extreme undersampling. Two examples extensively studied in the literature are the recovery of sparse vectors via $\ell_1$ norm penalization [4, 5, 6] and the recovery of low-rank matrices via trace norm penalization [2, 7, 8, 10, 11, 12].

Recent work by Chandrasekaran et al. [14] formulates a general framework for this compressed sensing problem, in which the complexity of an arbitrary structured signal $x^* \in \mathbb{R}^p$ is encoded in the geometric properties of a norm $\|\cdot\|_{\text{sig}}$ used to estimate the signal. Specifically, given a vector of $n$ noisy measurements $y = \Phi x^* + z$, where $\Phi$ is a Gaussian measurement matrix and $z$ is a bounded noise vector, their work gives conditions for when $x^*$ can be recovered from the convex program

$$
\min_{x} \left\{ \|x\|_{\text{sig}} : \|y - \Phi x\|_2 \leq \delta \right\},
$$

for some bound $\delta$ on the noise level $\|z\|_2$. In the noiseless setting where $\delta = 0$, the authors’ geometric analysis shows that

$$
n = \omega \left( T_{\text{sig}} \cap \mathbb{S}^{p-1} \right)^2 + 1
$$

measurements suffice to recover $x^*$ exactly. Here, $T_{\text{sig}}$ is a convex cone in $\mathbb{R}^p$ induced by $x^*$ and $\|\cdot\|_{\text{sig}}$ and $\omega \left( T_{\text{sig}} \cap \mathbb{S}^{p-1} \right)$ is a specific measure of the width of this cone, defined in Section [11]. In the noisy setting where
δ > 0, the same analysis shows that \( O \left( \omega \left( T_{\text{sig}} \cap \mathbb{S}^{p-1} \right)^2 \right) \) measurements suffice to recover \( x^* \) stably, that is, with error proportional to the noise level.

Our work extends that of Chandrasekaran et al. [14] to a more challenging setting, in which signal measurements may not be trustworthy. Specifically, we allow our measurements

\[
y = \Phi x^* + v^* + z
\]

to be corrupted by an unknown but structured vector \( v^* \), and bound the sample size \( n \) needed to recover \( x^* \) and \( v^* \) exactly or stably, using convex optimization. As an example, if \( z = 0 \) and \( v^* \) is \((n, \gamma)\)-sparse, so that a fraction \( \gamma \) of our linear measurements are arbitrarily corrupted, then our analysis guarantees exact recovery as soon as \( n \) exceeds

\[
\frac{\omega \left( T_{\text{sig}} \cap \mathbb{S}^{p-1} \right)^2}{2 \pi \left( 1 - \gamma \right)^2}
\]

plus an explicit smaller-order term. This provides a close parallel with Chandrasekaran et al. [14]’s result (1) for error proportional to the noise level.

Our analysis characterizes recovery from a wide variety of corruption structures, including block-sparse, low-rank, and binary corruption, by appealing to a unified geometric treatment of the complexity of the vector \( \delta \cdot v \) to be corrupted by an unknown but structured vector \( \omega \cdot \Phi \). This approach is motivated by convex programming approaches to disentangling signal and corruption. The first approach penalizes a combination of signal and corruption complexity, subject to known measurement constraints:

- **The structured signal** \( x^* \in \mathbb{R}^p \), our primary target for recovery. Our recovery ability will depend on the complexity of \( x^* \) with respect to a given norm \( \| \cdot \|_{\text{sig}} \). Common examples of structured signals include sparse vectors, which exhibit low complexity with respect to the \( \ell_1 \) norm, and low-rank matrices which exhibit low complexity under the trace norm (the sum of the matrix singular values). Our specific notion of complexity will be made precise in Section III.

- **The structured corruption** \( v^* \in \mathbb{R}^n \), our secondary recovery target. Our recovery ability will depend on the complexity of \( v^* \) with respect to a second norm, \( \| \cdot \|_{\text{cor}} \).

- **The unstructured noise** \( z \in \mathbb{R}^n \), satisfying \( \| z \|_2 \leq \delta \) for known \( \delta \geq 0 \). We make no additional assumption about the distribution or structure of \( z \). Our estimation error bounds for \( (x^*, v^*) \) will grow in proportion to \( \delta \).

- **The measurement matrix** \( \Phi \in \mathbb{R}^{n \times p} \), consisting of i.i.d. Gaussian entries

\[
\Phi_{ij} \overset{iid}{\sim} N(0, 1/n)
\]

as in [14]. Throughout, we treat \( x^* \) and \( v^* \) as fixed vectors chosen independently of \( \Phi \). However, the unstructured noise \( z \) need not be independent of \( \Phi \) and in particular may be chosen adversarially after \( \Phi \) is generated.

Our goal is tractable estimation of \( x^* \) and \( v^* \) given knowledge of \( y, \delta, \) and \( \Phi \). To this end, we consider two convex programming approaches to disentangling signal and corruption. The first approach penalizes a combination of signal and corruption complexity, subject to known measurement constraints:

\[
\min_{x, v} \left\{ \| x \|_{\text{sig}} + \lambda \| v \|_{\text{cor}} : \| y - (\Phi x + v) \|_2 \leq \delta \right\}.
\]

In Section III we will discuss specific settings of the parameter \( \lambda \), which trades off between the two penalties. The second approach makes use of available prior knowledge of either \( \| x^* \|_{\text{sig}} \) or \( \| v^* \|_{\text{cor}} \) (for example, a binary vector \( x^* \) always satisfies \( \| x^* \|_\infty = 1 \)) to explicitly constrain the signal complexity via

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

or the corruption complexity using

\[
\min_{x, v} \left\{ \| x \|_{\text{sig}} : \| v \|_{\text{cor}} \leq \| v^* \|_{\text{cor}}, \| y - (\Phi x + v) \|_2 \leq \delta \right\}.
\]

\footnote{We focus on norms due to their popularity in structured estimation, but any convex complexity measure would suffice.}

\footnote{In some settings, we can allow for \( x^* \) and \( v^* \) to be selected adversarially after the matrix \( \Phi \) is generated with a similar theoretical analysis but do not present this work here. In the literature, Chandrasekaran et al. [13] also treat the signal \( x^* \) as fixed, while McCoy and Tropp [15] give an additional deconvolution guarantee that holds universally over all low-complexity signals via a union bound argument.}
We note that McCoy and Tropp [15] study a similar, noiseless setting in which \( n = p \) and
\[
y = Ux^* + v^*
\]
for \( U \) a uniformly random orthogonal matrix. The authors assume that \( \|v^*\|_{\text{cor}} \) is known in advance and use spherical integral geometry to characterize the exact recovery of \((x^*, v^*)\) via the convex program
\[
\min_{x,v} \left\{ \|x\|_{\text{sig}} : \|v\|_{\text{sig}} \leq \|v^*\|_{\text{cor}} , y = Ux + v \right\}.
\]
Our novel analysis treats both constrained and penalized optimization, provides stable recovery results in the presence of unstructured noise, and covers both the high-dimensional setting \((n < p)\) and the overcomplete setting \((n \geq p)\).

B. Roadmap

The remainder of the paper is organized as follows. In Section II we review several concepts from convex geometry that are used throughout the paper and discuss the notions of Gaussian width and Gaussian distance that form the main ingredients in our recovery results. Section III presents our main results for both constrained and penalized convex recovery and gives a brief sketch of our proof strategy. We apply these results to several specific problems in Section IV, including the problem of secure and robust channel coding, and compare our results with those in the literature. Our extensive experiments with simulated data, summarized in Section V, demonstrate a close agreement between our theory and the phase transitions for successful signal recovery observed in practice. We conclude in Section VI with a discussion of our results and several directions for future research. All proofs are deferred to the appendices.

C. Notation

Throughout, we write \( \mu_n \) to represent the expected length of an \( n \)-dimensional vector with independent standard normal entries (equivalently, \( \mu_n \) is the mean of the \( \chi_n \) distribution)\(^3\). It is known that \( \mu_1 = \sqrt{2/\pi} \) while \( \mu_n \approx \sqrt{n} \) for large \( n \); in fact, the expectation is tightly bounded by
\[
\sqrt{n - 1/2} < \mu_n < \sqrt{n}
\]
for all \( n \) [16, Thm. 2]. In addition, we let \( g \) represent a random vector in \( \mathbb{R}^p \) with i.i.d. standard Gaussian entries.

II. Convex Geometry

In this section, we review the key concepts from convex geometry that underlie our analysis. Hereafter, \( \mathbb{R}^p \) and \( \mathbb{S}^{p-1} \) will denote the unit ball and the unit sphere in \( \mathbb{R}^p \) under the \( \ell_2 \) norm. Throughout the section, we reference a generic norm \( \|\cdot\| \) on \( \mathbb{R}^p \), evaluated at a generic point \( x \in \mathbb{R}^p \setminus \{0\} \), but we will illustrate each concept with the example of the \( \ell_1 \) norm \( \|\cdot\|_1 \) on \( \mathbb{R}^p \) and an \( s \)-sparse vector \( x_{\text{sparse}} \in \mathbb{R}^p \). Ultimately, we will apply the results of this section to the norms \( \|\cdot\|_{\text{sig}} \) on \( \mathbb{R}^p \) evaluated at the signal vector \( x^* \) and \( \|\cdot\|_{\text{cor}} \) on \( \mathbb{R}^n \) evaluated at the corruption vector \( v^* \).

A. The subdifferential

The subdifferential of \( \|\cdot\| \) at \( x \) is the set of vectors
\[
\partial \|x\| = \{ w \in \mathbb{R}^p : \|x + d\| \geq \|x\| + \langle w, d \rangle \text{ for all } d \in \mathbb{R}^p \}.
\]
In our running example of the \( \ell_1 \) norm and the sparse vector \( x_{\text{sparse}} \), we have
\[
\partial \|x_{\text{sparse}}\|_1 = \text{sign}(x_{\text{sparse}}) + \{ w \in \mathbb{R}^p : \text{support}(w) \cap \text{support}(x_{\text{sparse}}) = \emptyset, \|w\|_{\text{inf}} \leq 1 \}, \quad (5)
\]
\(^3\)The same quantity was represented by \( \lambda_n \) in [14].
B. Tangent cones and normal cones

Our analysis of recovery under corrupted sensing will revolve around two notions from convex geometry: the **tangent cone** and the **normal cone**. We define the tangent cone to \( \| \cdot \| \) at \( x \) as the set of descent (or non-ascent) directions of \( \| \cdot \| \) at \( x \):

\[
T = \{ w : \| x + c \cdot w \| \leq \| x \| \text{ for some } c > 0 \}.
\]

By convexity of the norm \( \| \cdot \| \), \( T \) is a convex cone. In our running example, the tangent cone is given by

\[
T_{\text{sparse}} = \{ w : \| w_{\text{support}(x_{\text{sparse})}} \|_1 \leq -\langle w, \text{sign}(x_{\text{sparse})} \rangle \}.
\]

The normal cone to \( \| \cdot \| \) at \( x \) is the polar of the tangent cone, given by

\[
N = \{ w : \langle w, u \rangle \leq 0 \text{ for all } u \in T \},
\]

and may equivalently be written as the conic hull of the subdifferential \( \partial \| x \| \) [17]:

\[
N = \text{cone} \{ \partial \| x \| \} = \{ w : w \in t \cdot \partial \| x \| \text{ for some } t \geq 0 \}.
\]

Figure 1 illustrates these entities in \( \mathbb{R}^2 \) for the \( \ell_1 \) and \( \ell_2 \) norms.

C. Gaussian width and Gaussian distance

To quantify the complexity of a structured vector, we adopt two geometric measures of size known as the **Gaussian width** and the **Gaussian distance**:

**Definition 1** (Gaussian width). We denote the Gaussian width of a set \( C \subset \mathbb{R}^p \) by

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

**Definition 2** (Gaussian distance). We denote the Gaussian distance of a set \( C \subset \mathbb{R}^p \) by

\[
\eta(C) = \mathbb{E}_{g \sim N(0, I_p)} \left[ \inf_{w \in C} \| g - w \|_2 \right].
\]

Chandrasekaran et al. [14] showed that the Gaussian width of a restricted tangent cone \( \omega\left( T \cap S^{p-1} \right) \) determines a sufficient sample size for signal recovery from uncorrupted measurements. We will establish analogous results for our corrupted sensing setting in Section III.

To obtain interpretable sample size bounds in terms of familiar parameters, it is often necessary to bound \( \omega\left( T \cap S^{p-1} \right) \) analytically. Chandrasekaran et al. [14] describe a variety of methods for obtaining such bounds. One
especially profitable technique relates this Gaussian width to the Gaussian distance of the scaled subdifferential of $\|\cdot\|$ at $x$:

$$\omega \left( T \cap S^{p-1} \right) \leq \omega \left( T \cap B^p \right) \leq \min_{t \geq 0} \eta \left( t \cdot \partial \|x\| \right).$$

Indeed, many of the width bounds in [14] are derived by bounding $\eta \left( t \cdot \partial \|x\| \right)$ at a fixed value of $t$. Furthermore, the best choice of $t$ typically yields a bound within a small additive constant of the Gaussian width:

**Proposition 1.** Suppose that $\partial \|x\|$ satisfies a weak decomposability assumption:

$$\exists w_0 \in \partial \|x\| \text{ s.t. } \langle w - w_0, w_0 \rangle = 0, \forall w \in \partial \|x\|. \quad (7)$$

Then

$$\min_{t \geq 0} \eta \left( t \cdot \partial \|x\| \right) \leq \omega \left( T \cap S^{p-1} \right) + 6. \quad (8)$$

**Remark 1.** The weak decomposability assumption (7) is satisfied in all the examples considered in this paper and is weaker than the decomposability assumptions often found in the literature (see, e.g., [18]).

We prove this result in Appendix A-G.

In the case of our $\ell_1$ norm example, Chandrasekaran et al. [14] choose $t = \sqrt{2 \log(p/s)}$ and show that

$$\omega \left( T_{\text{spars}} \cap S^{p-1} \right)^2 \leq \eta \left( t \cdot \partial \|x_{\text{spars}}\|_1 \right)^2 \leq 2s \log(p/s) + \frac{3}{2}s. \quad (9)$$

We will show that the alternative scaling $t' = \sqrt{2/\pi (1 - s/p)}$ yields the bound

$$\omega \left( T_{\text{spars}} \cap S^{p-1} \right)^2 \leq \eta \left( t' \cdot \partial \|x_{\text{spars}}\|_1 \right)^2 \leq p \cdot (1 - 2/\pi) \cdot (1 - s/p)^2,$$

which is a tighter than (8) whenever $s \geq 0.07 \cdot p$. Moreover, this new bound, unlike (8), is strictly less than $p$ for any $s < p$; this has important implications for the recovery of structured signals from nearly dense corruption.

Both bounds (8) and (9) are ultimately derived from the expected squared distance relation

$$\eta \left( t \cdot \partial \|x\| \right)^2 \leq \mathbb{E} \left[ \inf_{w \in \partial \|x\|} \|g - w\|_2^2 \right].$$

In the $\ell_1$ setting, this expected squared distance has a manageable representation (see [14, Sec. C]) that can be directly minimized over $t$:

$$\min_{t \geq 0} \mathbb{E} \left[ \inf_{w \in \partial \|x\|_1} \|g - w\|_2^2 \right] = \min_{t \geq 0} \left( 1 + t^2 \right) + \frac{2(p - s)}{\sqrt{2\pi}} \left( 1 + t^2 \right) \int_{t}^{\infty} e^{-c^2/2} dc - te^{-t^2/2}. \quad (10)$$

The resulting bound on the Gaussian width is tighter than either (8) or (9), albeit without an interpretable closed form. We compare the distance bounds arising from these three calculations in Figure 2. Notice that the new analytic bound (9) closely approximates the optimal squared distance bound (10) for a wide range of sparsity levels, while the prior analytic bound (8) only dominates (9) in the extreme sparsity regime.

It is instructive to contrast the width of $T_{\text{spars}} \cap S^{p-1}$ in our example with that induced by the $\ell_2$ norm on $\mathbb{R}^p$. The $\ell_2$ norm tangent cone at any point $x$ will be a half-space (as illustrated in Figure 1), and its Gaussian width will be approximately $\sqrt{p}$, irrespective of the structure of $x$. Meanwhile, any sparse vector will exhibit a much smaller Gaussian width under the $\ell_1$ norm tangent cone, as illustrated in (8) and (9). We will see that the choice of norm and the attendant reduction in geometric complexity are critical for achieving signal recovery in the presence of corruption.

---

5Chandrasekaran et al. [13] report a slightly smaller bound based on a minor miscalculation.
Figure 2: Three upper bounds on the rescaled Gaussian distance $\min_{t \geq 0} \eta (t \cdot \partial \|x_{\text{sparse}}\|_1) / \sqrt{p}$ for an $s$-sparse vector $x_{\text{sparse}} \in \mathbb{R}^p$ as a function of the sparsity proportion $\gamma = s/p$. The prior bound comes from (8), the new bound from (9), and the optimal bound comes from the exact expression for expected squared distance given in (10).

D. Structured vectors and structure-inducing norms

While our analysis is applicable to any signal coupled with any norm, it is most profitable when the signal exhibits low complexity under its associated norm. In this section, we will highlight a number of important structured signal classes and norms under which these signals have low complexity (additional examples can be found in [14]):

- **Sparse vectors**: As discussed throughout the section, an appropriate norm for recovering sparse vectors in $\mathbb{R}^p$ is the $\ell_1$ norm, $\|\cdot\|_1$.
- **Low-rank matrices**: We will see that low-rank matrices in $\mathbb{R}^{m_1 \times m_2}$ exhibit low geometric complexity under the trace norm $\|\cdot\|_*$, given by the sum of the singular values of a matrix.
- **Binary vectors**: A binary vector of length $p$ has relatively low complexity under the $\ell_\infty$ norm, $\|\cdot\|_\infty$, which has a known value of 1 for any such vector.
- **Block-sparse vectors**: We will show that a block-sparse vector in $\mathbb{R}^p$, supported on a partition of $\{1, \ldots, p\}$ into $m$ blocks of size $p/m$, has low complexity under the $\ell_1/\ell_2$ norm

$$\|x\|_{\ell_1/\ell_2} = \sum_{b=1}^m \|x_{V_b}\|_2,$$

where $V_b \subset \{1, \ldots, p\}$ is the $b$th block.

Table II gives bounds on the tangent cone width for each of these structured settings (recall that $\mu_k$ is the mean of a $\chi_k$ distribution). These estimates will be useful for studying signal recovery via the constrained convex optimization problems (3) and (4). Table II highlights those results obtained by bounding the larger quantity $\eta (t \cdot \partial \|x\|)$ for a fixed scaling $t$ of the subdifferential. We will make use of these Gaussian distance estimates and their associated settings of $t$ when studying penalized convex recovery via (2). The new results in Tables II and III are derived in Appendix A.
Table I: Upper bounds on the restricted tangent cone Gaussian width $\omega(T \cap S^{p-1})$ for various structured vectors and structure-inducing norms. Prior bounds on Gaussian width are due to [12].

| Structure                      | $\|\cdot\|$ | Prior Gaussian width bound                          | New Gaussian width bound                        |
|-------------------------------|--------------|---------------------------------------------------|-------------------------------------------------|
| $s$-sparse $p$-dim. vector     | $\|\cdot\|_1$ | $\sqrt{2s \log(\frac{p}{s}) + \frac{3}{2}s}$     | $\sqrt{2 \log(n/s)}$                            |
| $s$-block-sparse vector        | $\|\cdot\|_1$ | $\sqrt{2s \log(\frac{p}{s}) + \frac{3}{2}s}$     | $\sqrt{2 \log(n/s)}$                            |
| (m blocks of size $k = p/m$)   | $\|\cdot\|_1$ | $\sqrt{2s \log(\frac{p}{s}) + \frac{3}{2}s}$     | $\sqrt{2 \log(n/s)}$                            |
| $r$-rank $m_1 \times m_2$ matrix | $\|\cdot\|_*$ | $\sqrt{3r(m_1 + m_2 - r)}$                        | $\frac{1}{\sqrt{m_1 m_2}} \frac{1}{1 - \frac{r}{m_1}}$ |
| $m_1 \geq m_2$                |              |                                                   | $\frac{1}{\sqrt{m_1 m_2}} \frac{1}{1 - \frac{r}{m_1}}$ |

Table II: Upper bounds on the scaled subdifferential Gaussian distance $\eta(t \cdot \|x\|)$ for various structured vectors, structure-inducing norms, and settings of $t$.

| Structure                      | $\|\cdot\|$ | Gaussian distance bound                          | Setting of $t$ used to achieve bound |
|-------------------------------|--------------|---------------------------------------------------|--------------------------------------|
| $s$-sparse $p$-dim. vector     | $\|\cdot\|_1$ | $\sqrt{2s \log(\frac{p}{s}) + \frac{3}{2}s}$     | $\sqrt{2 \log(n/s)}$                |
| $s$-block-sparse vector        | $\|\cdot\|_1$ | $\sqrt{2s \log(\frac{p}{s}) + \frac{3}{2}s}$     | $\sqrt{2 \log(n/s)}$                |
| (m blocks of size $k = p/m$)   | $\|\cdot\|_1$ | $\sqrt{2s \log(\frac{p}{s}) + \frac{3}{2}s}$     | $\sqrt{2 \log(n/s)}$                |
| $r$-rank $m_1 \times m_2$ matrix | $\|\cdot\|_*$ | $\sqrt{3r(m_1 + m_2 - r)}$                        | $\frac{1}{\sqrt{m_1 m_2}} \frac{1}{1 - \frac{r}{m_1}}$ |
| $m_1 \geq m_2$                |              |                                                   | $\frac{1}{\sqrt{m_1 m_2}} \frac{1}{1 - \frac{r}{m_1}}$ |

### III. Recovery from Corrupted Gaussian Measurements

We now present our theoretical results on the recovery of structured signals from corrupted measurements via convex programming. Both main theorems are proved in Appendix [C]. Below, $T_{\text{sig}}$ denotes the tangent cone for $\|\cdot\|_{\text{sig}}$ at the true structured signal vector $x^*$, and $T_{\text{cor}}$ designates the tangent cone for $\|\cdot\|_{\text{cor}}$ at the true structured corruption vector $v^*$.

#### A. Recovery via constrained optimization

We begin by analyzing the constrained convex recovery procedures

$$\min \left\{ \|x\|_{\text{sig}} : \|v\|_{\text{cor}} \leq \|v^*\|_{\text{cor}}, \|y - (\Phi x + v)\|_2 \leq \delta \right\}$$

(11)

and

$$\min \left\{ \|v\|_{\text{cor}} : \|x\|_{\text{sig}} \leq \|x^*\|_{\text{sig}}, \|y - (\Phi x + v)\|_2 \leq \delta \right\},$$

(12)

which are natural candidates for estimating $(x^*, v^*)$ whenever prior knowledge of $\|x^*\|_{\text{sig}}$ or $\|v^*\|_{\text{cor}}$ is available. Our first result shows that, with high probability, approximately

$$\omega(T_{\text{sig}} \cap S^{p-1})^2 + \omega(T_{\text{cor}} \cap S^{n-1})^2$$

corrupted measurements suffice to recover $(x^*, v^*)$ exactly in the absence of noise ($\delta = 0$) and stably in the presence of noise ($\delta \neq 0$), via either of the procedures (11) or (12).

**Theorem 1** (General constrained recovery). If $(\hat{x}, \hat{v})$ solves either of the constrained optimization problems (11) or (12), then

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

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

$$\tau = \sqrt{\omega(T_{\text{sig}} \cap S^{p-1})^2 + \omega(T_{\text{cor}} \cap S^{n-1})^2} + 8 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2\pi}}.$$  

**Remark 2.** In the noiseless setting where $\delta = 0$, Theorem 1 entails exact recovery of $(x^*, v^*)$ with probability at least $1 - e^{-(\mu_n - \tau)^2/2}$ as long as $\mu_n \geq \tau$. 

B. Recovery via penalized optimization

When prior knowledge of \( \|x^*\|_{\text{sig}} \) or \( \|v^*\|_{\text{cor}} \) is unavailable, one can instead rely on the penalized convex recovery procedure

\[
\min \left\{ \|x\|_{\text{sig}} + \lambda \|v\|_{\text{cor}} : \|y - (\Phi x + v)\|_2 \leq \delta \right\} \tag{13}
\]

with penalty parameter \( \lambda \). Our next result provides sufficient conditions for exact and stable penalized recovery and demonstrates how to set the penalty parameter \( \lambda \) in practice.

**Theorem 2** (General penalized recovery). Fix any \( t_{\text{sig}}, t_{\text{cor}} \geq 0 \). If \( (\hat{x}, \hat{v}) \) solves the penalized optimization problem

\[
\min \left\{ \|x\|_{\text{sig}} + \lambda \|v\|_{\text{cor}} : \|y - (\Phi x + v)\|_2 \leq \delta \right\}
\]

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

\[
\tau = 2\eta \left( t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}} \right) + \eta \left( t_{\text{cor}} \cdot \partial \|v^*\|_{\text{cor}} \right) + 3\sqrt{2\pi} \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}.
\]

**Remark 3.** In the noiseless setting (\( \delta = 0 \)), Theorem 2 entails exact recovery of \( (x^*, v^*) \) with probability at least \( 1 - e^{-\left(\mu_n - \tau\right)^2/2} \) as long as \( \mu_n \geq \tau \).

**Remark 4.** We suspect that the factor of 2 multiplying the first Gaussian distance term in Theorem 2 is an artifact of the proof technique and not essential.

To make use of Theorem 2 it suffices to bound the Gaussian distance terms

\[
\eta \left( t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}} \right) \quad \text{and} \quad \eta \left( t_{\text{cor}} \cdot \partial \|v^*\|_{\text{cor}} \right)
\]

for suitably chosen subdifferential scalings \( t_{\text{sig}} \) and \( t_{\text{cor}} \). Recall, from Section II-C that these Gaussian distances represent practical upper bounds on the tangent cone Gaussian widths, \( \omega \left( T_{\text{sig}} \cap S^{p-1} \right) \) and \( \omega \left( T_{\text{cor}} \cap S^{n-1} \right) \); in fact, Proposition 1 showed that a well-chosen distance bound closely approximates the tangent cone width for most norms of interest. In practice we can choose the distance scalings \( t_{\text{sig}} \) and \( t_{\text{cor}} \) based on known upper bounds on the Gaussian distance, such as those found in Table II.

In the application section, we will see that different choices of \( t_{\text{sig}}, t_{\text{cor}}, \) and \( \lambda \) in Theorem 2 allow us to recover or improve upon existing recovery results for specialized signal and corruption structure. In the experiments section, we will choose \( t_{\text{sig}}, t_{\text{cor}}, \) and \( \lambda \) based on practical considerations and demonstrate recovery performance that nearly matches recovery rates obtained with prior knowledge of \( \|x^*\|_{\text{sig}} \) or \( \|v^*\|_{\text{cor}} \).

C. Proof overview

In each of the optimization problems above, the recovery target \( (x^*, v^*) \) and estimate \( (\hat{x}, \hat{v}) \) both lie in the feasible set \( \{ (x, v) : \|y - (\Phi x + v)\|_2 \leq \delta \} \), and hence the error \( (\hat{x} - x^*, \hat{v} - v^*) \) satisfies \( \|\Phi(\hat{x} - x^*) + (\hat{v} - v^*)\|_2 \leq 2\delta \). In Appendix C we bound the size of this error with high probability by lower bounding

\[
\min \frac{\|\Phi a + b\|_2}{\sqrt{||a||^2_2 + ||b||^2_2}}, \tag{14}
\]

where the minimum is taken uniformly over the class of nonzero perturbation vectors \( (a, b) \in \mathbb{R}^p \times \mathbb{R}^n \) satisfying either the stronger requirement

\[
\|x^* + a\|_{\text{sig}} \leq \|x^*\|_{\text{sig}} \quad \text{and} \quad \|v^* + b\|_{\text{cor}} \leq \|v^*\|_{\text{cor}}
\]

for either of the two constrained recovery procedures, or the weaker requirement

\[
\|x^* + a\|_{\text{sig}} + \lambda \|v^* + b\|_{\text{cor}} \leq \|x^*\|_{\text{sig}} + \lambda \|v^*\|_{\text{cor}}
\]

for the penalized recovery procedure. The lower bound on (14) can then be applied to the error vector \( (a, b) = (\hat{x} - x^*, \hat{v} - v^*) \) to bound its \( \ell_2 \) norm.
IV. APPLICATIONS AND RELATED WORK

In this section, we apply the general theory of Section III to specific signal and corruption structures of interest, drawing comparisons to existing literature where relevant. All results in this section are proved in Appendix B.

A. Binary recovery from sparse corruption and unstructured noise

To exemplify the constrained recovery setting, we analyze a communications protocol for secure and robust channel coding proposed by Wyner [19, 20] and studied by McCoy and Tropp [15]. The aim is to transmit a binary signal securely across a communications channel while tolerating sparse communication errors. The original protocol applied a random rotation to the binary signal and thereby constrained the length of the transmitted message to equal the length of the input message. Here, we take \( n \) random Gaussian measurements \( \Phi \) for \( n \) not necessarily equal to \( p \). This offers the flexibility of transmitting a shorter, cheaper message with reduced tolerance to corruption, or a longer message with increased corruption tolerance. In addition, unlike past work, we allow our measurements to be perturbed by dense, unstructured noise.

To recover the signal \( x^* \) after observing the sparsely corrupted and densely perturbed message \( y = \Phi x^* + v^* + z \) with \( \|z\|_2 \leq \delta \), we solve the constrained optimization problem

\[
\min_{x,v} \{ \|v\|_1 : \|x\|_\infty \leq 1, \|y - (\Phi x + v)\|_2 \leq \delta \},
\]

where we take advantage of the prior knowledge that \( \|x\|_\infty = 1 \) for any binary signal. The following result, which follows from Theorem 1, provides sufficient conditions for exactly reconstructing \( x^* \) using the convex program (15).

**Corollary 1** (Binary signal, sparse corruption). Suppose that \( x^* \in \{+1, -1\}^p \) and that \( v^* \) is supported on at most \( n \cdot \gamma \) entries. Suppose the number of measurements satisfies

\[
\mu_n - 2\delta > \sqrt{\frac{p}{2}} + \omega \left( T_{\text{sparse}(n,\gamma)} \cap \mathbb{S}^{n-1} \right)^2 + 8 + \frac{1}{\sqrt{2\pi}} \leq \tau,
\]

where \( T_{\text{sparse}(n,\gamma)} \) is the tangent cone of a \( (n \cdot \gamma) \)-sparse \( n \)-dimensional vector under the \( \ell_1 \) norm, known to satisfy

\[
\omega \left( T_{\text{sparse}(n,\gamma)} \cap \mathbb{S}^{n-1} \right) \leq \sqrt{n} \cdot \min \left\{ \sqrt{2\gamma \log (\gamma^{-1}) + \frac{3}{2} \gamma}, \sqrt{1 - \frac{2}{\pi} (1 - \gamma)^2} \right\}.
\]

Then with probability at least \( 1 - e^{-(\mu_n - 2\delta - \tau)^2/2} \), the binary vector \( x^* \) is exactly equal to \( \text{sign}(\hat{x}) \), where \((\hat{x}, \hat{v})\) is any solution to the constrained recovery procedure (15).

**Remark 5.** In addition to the analytical bound (16), \( \omega \left( T_{\text{sparse}(n,\gamma)} \cap \mathbb{S}^{n-1} \right) \) admits the sharp, computable bound described in (10).

It is enlightening to compare Corollary 1 with the binary recovery from sparse corruption results of McCoy and Tropp [15]. Assuming that \( n = p \) and that the measurement matrix is a uniformly random rotation drawn independently of \((x^*, v^*)\), McCoy and Tropp [15] proved that signal recovery is possible with up to 19.3\% of measurements corrupted. In their work, the threshold value of \( \gamma_0 = 19.3\% \) was obtained through a numerical estimate of a geometric parameter analogous to the Gaussian width \( \omega \left( T_{\text{sparse}(n,\gamma)} \cap \mathbb{S}^{n-1} \right) \). For large \( n = p \), the additive constants in Corollary 1 become negligible, and our result implies exact recovery with some probability when

\[
\omega \left( T_{\text{sparse}(n,\gamma)} \cap \mathbb{S}^{n-1} \right) < \sqrt{(1 - 2\delta)^2 - 1/2} \cdot \sqrt{n}.
\]

In the noiseless setting (\( \delta = 0 \)), this precondition for exact recovery becomes \( \omega \left( T_{\text{sparse}(n,\gamma)} \cap \mathbb{S}^{n-1} \right) < \sqrt{n}/2 \). Our plot of \( \ell_1 \) width bounds (Figure 2) shows that this indeed holds when \( \gamma < 19.3\% \). Thus, our result parallels that of McCoy and Tropp [15] in the noiseless setting. Unlike the work of McCoy and Tropp [15], Corollary 1 also provides for exact recovery from dense, unstructured noise, and additionally allows for \( n \neq p \).

B. General penalized recovery from block-sparse corruption and unstructured noise

As a first application of our penalized recovery result, Theorem 2, we consider a setting in which the corruption is entry-wise or block-wise sparse, and the signal exhibits an arbitrary structure. While past work has analyzed corruption-free block-sparse signal recovery in the setting of compressed sensing [21, 22, 23, 24], to our knowledge, Corollary 2 is the first result to analyze structured signal recovery from block-sparse corruption.
Let \( m \) denote the number of disjoint measurement blocks and \( k = n/m \) represent the common block size. Suppose that no more than \( s \) measurement blocks have been corrupted and that the identities of the corrupted blocks are unknown. We consider a “dense” corruption regime where the proportion of corruptions \( \gamma = \frac{m}{m} \) may lie anywhere in \((0,1)\). The following corollary bounds the number of measurements needed for exact or stable recovery, using the penalized program

\[
\min_{x,v} \left\{ \|x\|_{\text{sig}} + \lambda \|v\|_{\ell_1/\ell_2} : \|y - (\Phi x + v)\|_2 \leq \delta \right\}
\]

with \( \lambda = \mu_k(1-\gamma)/t_{\text{sig}} \) for any choice of \( t_{\text{sig}} \geq 0 \). In particular, when \( \delta = 0 \), we find that

\[
O\left( \left( \frac{\eta t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}}}{(1-\gamma)^4} \right)^{2} \right)
\]

measurements suffice for exact recovery of a structured signal \( x^* \).

**Corollary 2** (Penalized recovery, dense block or entrywise corruption). Fix any \( t_{\text{sig}} \geq 0 \), and suppose that \( v^* \) exhibits block-sparse structure with blocks of size \( k \). If the fraction of nonzero blocks of \( v^* \) is at most \( \gamma \), \((\widehat{x}, \widehat{v})\) solves (17) with the penalty parameter \( \lambda = \frac{\mu_k(1-\gamma)}{t_{\text{sig}}} \), and the number of measurements \( n \) satisfies

\[
\sqrt{n} \geq \frac{2\eta \left( t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}} \right) + \sqrt{2\log(1/\beta) + 3\sqrt{2} \pi + 1 + \frac{1}{\sqrt{2} \pi}}}{\alpha_k \left( (1-\gamma)^2 - \epsilon \right)}
\]

for \( \alpha_k = 1 - \sqrt{1 - \mu_k^2}/k \), then with probability at least \( 1 - \beta \),

\[
\sqrt{\|\widehat{x} - x^*\|_2^2 + \|\widehat{v} - v^*\|_2^2} \leq \frac{2\delta}{\epsilon}.
\]

**Remark 6.** In the noiseless setting, where \( \delta = 0 \), Corollary 2 entails exact recovery with probability \( 1 - \beta \) whenever

\[
\sqrt{n} \geq \frac{2\eta \left( t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}} \right) + \sqrt{2\log(1/\beta) + 3\sqrt{2} \pi + 1 + \frac{1}{\sqrt{2} \pi}}}{\alpha_k \left( (1-\gamma)^2 - \epsilon \right)}
\]

In the important special case where \( x^* \) is a sparse vector with at most \( s \) non-zero entries and \( v^* \) is entrywise sparse \((k = 1)\), the \( \ell_1 \) distance bound in Table I and Corollary 2 together imply that

\[
O\left( \frac{s \log(p/s) + s}{(1-\gamma)^4} \right) \quad \text{or} \quad O\left( \frac{s \log(p/s) + s}{(\alpha_1(1-\gamma)^2 - \epsilon)^2} \right)
\]

measurements suffice to recover \( x^* \) exactly or stably, respectively, with high probability, using the convex program

\[
\min_{x,v} \left\{ \|x\|_1 + \lambda \|v\|_1 : \|y - (\Phi x + v)\|_2 \leq \delta \right\},
\]

with \( \lambda = (1-\gamma)/\sqrt{\pi \log(p/s)} \). This result is particularly well suited to recovery from dense entrywise corruption and accords with the best sparse recovery from sparse corruption results in the literature. For example, Li [1] establishes stable recovery via (18) using the penalty parameter \( \lambda = 1/\sqrt{\log(p/n)} + 1 \), whenever the signal and corruption sparsity levels satisfy

\[
\|x^*\|_0 \leq C \cdot \frac{n}{\log(p/n)} + 1 \quad \text{and} \quad \|v^*\|_0 \leq C \cdot n,
\]

for an unspecified universal constant \( C > 0 \). Our result admits a guarantee of this form, while providing an explicit trade-off between the sparsity levels of the signal and the corruption.

Specifically, Corollary 2 yields a stable recovery result (derived in Appendix B) under the conditions

\[
\|x^*\|_0 \leq s_\gamma := \left| \frac{A_\gamma n}{2 \log(\frac{p}{x_\gamma n}) + 3/2} \right| \quad \text{and} \quad \|v^*\|_0 \leq n \cdot \gamma,
\]

for any corruption proportion bound \( \gamma < 1 \) and

\[
A_\gamma := (\alpha_1(1-\gamma)^2 - \epsilon)^2/144.
\]
where the corruption proportion parameter extremely sparse case could be attained with a similar analysis as in Corollary 2; we do not state such a result here rather than using the second distance bound in this table as for the high-corruption setting. A general result for the but instead give a specific example where both the signal and corruption exhibit extreme element-wise sparsity.

C. Penalized recovery when signal and corruption are both extremely sparse

Corollary 2 is well-suited to recovery from moderate or frequent corruption, but in the extreme sparsity setting, their analysis requires exact recovery of a sparse signal from sparsely corrupted measurements via the convex program

\[ \min_{x, v} \{ \|x\|_1 + \|v\|_1 : \|y - (\Phi x + v)\|_2 \leq \delta \} \]

and the number of measurements \( n \) satisfies

\[ \mu_n - \epsilon \sqrt{n} \geq \tau \triangleq 3 \sqrt{(s_{\text{sig}} + s_{\text{cor}}) \cdot \left( 2 \log \left( \frac{p + n}{s_{\text{sig}} + s_{\text{cor}}} \right) + \frac{3}{2} \right) + 3 \sqrt{2 \pi} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2\pi}}}, \]

then with probability at least \( 1 - e^{-\left( \mu_n - \tau - \epsilon \sqrt{n} \right)^2 / 2} \),

\[ \sqrt{\|\hat{x} - x^*\|_2^2 + \|\hat{v} - v^*\|_2^2} \leq \frac{2\delta}{\epsilon}. \]

When the noise level equals zero, we recover a result due to Laska et al. [25], which establishes high probability exact recovery of a sparse signal from sparsely corrupted measurements via the convex program

\[ \min_{x, v} \{ \|x\|_1 + \|v\|_1 : y = \Phi x + v \} . \]

Their analysis requires \( n \geq C(s_{\text{sig}} + s_{\text{cor}}) \log \left( \frac{p + n}{s_{\text{sig}} + s_{\text{cor}}} \right) \) measurements for recovery, where \( C \) is an unspecified constant. In comparison, the present analysis features small, explicit constants, provides for stable recovery in the presence of unstructured noise, and follows from a more general treatment of signal and corruption structure.

We remark that a similar result to Corollary 3 can be stated for block-sparse signals with \( m_{\text{sig}} \) blocks of size \( k_{\text{sig}} \) and block-sparse corruption vectors with \( m_{\text{cor}} \) blocks of size \( k_{\text{cor}} \). In this case,

\[ \mathcal{O} \left( s_{\text{sig}} + s_{\text{cor}} \left( \log \left( \frac{m_{\text{sig}} + m_{\text{cor}}}{s_{\text{sig}} + s_{\text{cor}}} \right) + k_{\text{sig}} + k_{\text{cor}} \right) \right) \]

measurements suffice for high probability recovery. We omit the details here.

D. Additional related work

We conclude this section with a summary of some existing results in the vicinity of this work.
1) Sparse signal recovery from sparse corruption: The specific problem of recovering a sparse vector $x^* \in \mathbb{R}^p$ from linear measurements with sparse corruption $v^* \in \mathbb{R}^n$ has been analyzed under a variety of different modeling assumptions. For example, Wright and Ma [26] study exact recovery under a cross-and-bouquet model:

$$y = Ax^* + v^*$$

for $A^{(i)} \overset{i.i.d.}{\sim} N(\mu, \frac{\mu^2}{n} I_p), \|\mu\|_2 = 1, \|\mu\|_\infty \leq \frac{C}{\sqrt{n}},$

where $A^{(i)}$ is the $i$th column of $A$, and $v^*$ has uniformly random signs. Nguyen and Tran [27] establish stable or exact recovery of $(x^*, v^*)$ from $y = Ax^* + v^* + z$, where $A$ has columns sampled uniformly from an orthonormal matrix, $x^*$ has uniformly random signs, $v^*$ has uniformly distributed support, and $z$ is a bounded noise vector. Pope et al. [28] analyze the exact recovery of $x^*$ and $v^*$ from $y = Ax^* + Bv^*$, where $A$ and $B$ are known. The authors require uniform randomness in the support set of $x^*$ or $v^*$ and rely on certain incoherence properties of $A$ and $B$. In contrast to these works, we analyze the recovery of deterministic signal and corruption vectors from Gaussian measurements, derive small, explicit constants, and generalize to arbitrary structured signals and structured corruption vectors.

2) Structured signal recovery from structured corruption: Hegde and Baraniuk [29] analyze a nonconvex procedure for deconvolving a pair of signals lying on incoherent manifolds when the measurement matrix satisfies a restricted isometry property. Here, we focus on convex procedures for which minimum is known to be found in polynomial time.

In addition to their work on the binary signal plus sparse corruption problem, McCoy and Tropp [15] give recovery results for arbitrary structured signals and structured corruptions, in the setting where $n = p$, observations are noiseless ($\delta = 0$ in our notation), and either $\|x^*\|_{\text{sig}}$ or $\|v^*\|_{\text{cor}}$ is known exactly. In comparison, our work allows us to consider $n$ smaller or larger than $p$, to handle stable recovery in the presence of noise, and to use penalized rather than constrained optimization where practical.

V. Experimental Evaluation

In this section, we verify and complement the theoretical results of Section III with a series of synthetic corrupted sensing experiments. We present both constrained and penalized recovery experiments for several types of structured signals and corruptions in the presence or absence of noise. Our goals are twofold: first, to test the agreement between our constrained recovery theory and empirical recovery behavior and, second, to evaluate the utility of the penalty. Procedure for deconvolving a pair of signals lying on incoherent manifolds when the measurement matrix satisfies polynomial time.

A. Phase transitions from constrained recovery

We begin by investigating the empirical behavior of the constrained recovery program (12) when the noise level $\delta = 0$ and the norm of the true signal $\|x^*\|_{\text{sig}}$ are known exactly. In this setting, our constrained recovery result (Theorem 1) guarantees exact recovery with some probability once $\mu_n^2 \approx n$ exceeds $\omega \left( T_{\text{sig}} \cap \mathbb{S}^{p-1} \right)^2 + \omega \left( T_{\text{cor}} \cap \mathbb{S}^{n-1} \right)^2$. Using new and existing bounds on the relevant Gaussian widths, our aim is to determine how conservative this theoretical recovery guarantee is in practice. To this end, we will apply our convex recovery procedures to synthetic problem instances and compare the empirical probability of successful recovery to the recovery behavior predicted by our theory. We will see that our theoretical recovery thresholds closely align with observed phase transitions in three different settings: binary signal recovery from sparse corruption, sparse signal recovery from sparse corruption, and sparse signal recovery from block-wise sparse corruption.

Binary signal, sparse corruption: We first consider the secure and robust communications protocol discussed in Section IV-A, where we aim to recover a binary message from noiseless ($\delta = 0$) but sparsely corrupted Gaussian measurements. Fixing the message length $p = 1000$, we vary the number of measurements $n \in [750, 1250]$ and the number of corruptions $s_{\text{cor}} \in [50, 350]$ and perform the following experiment 10 times for each $(n, s_{\text{cor}})$ pair:

1) Sample a binary vector $x^*$ uniformly from $\{\pm 1\}^p$.
2) Draw a Gaussian matrix $\Phi \in \mathbb{R}^{n \times p}$ with independent $N(0, 1/n)$ entries.
3) Generate a corruption vector $v^*$ with $s_{\text{cor}}$ independent standard normal entries and $n - s_{\text{cor}}$ entries set to 0.
4) Solve the constrained optimization problem (13) with $y = \Phi x^* + v^*$ and $\delta = 0$.
5) Declare success if $\|\hat{x} - x^*\|_2 / \|x^*\|_2 < 10^{-3}$.
independent standard normal entries and optimization problem follow the experimental setup of the previous experiment with two substitutions: we generate $\mu = 10$. This theoretical recovery threshold reflects the observed empirical phase transition accurately.

Sparse signal, block-sparse corruption: We next consider the problem of recovering a sparse signal vector of known $\ell_1$ norm from sparsely corrupted Gaussian measurements. We fix the signal length and sample size $p = n = 1000$, and vary the sparsity levels $(s_{\text{sig}}, s_{\text{cor}}) \in [1, 860]^2$. We perform the following experiment 10 times for each $(s_{\text{sig}}, s_{\text{cor}})$ pair:

1. Generate a signal vector $x^*$ with $s_{\text{sig}}$ independent standard normal entries and $p-s_{\text{sig}}$ entries set to 0.
2. Generate a corruption vector $v^*$ with $s_{\text{cor}}$ independent standard normal entries and $n-s_{\text{cor}}$ entries set to 0.
3. Draw a Gaussian matrix $\Phi \in \mathbb{R}^{n \times p}$ with independent $N(0, 1/n)$ entries.
4. Solve the following constrained optimization problem with $y = \Phi x^* + v^*$:

   $$(\hat{x}, \hat{v}) \in \arg\min_{x,v} \left\{ \|v\|_1 : \|x\|_1 \leq \|x^*\|_1, y = \Phi x + v \right\}.$$ 

5. Declare success if $\|\hat{x} - x^*\|_2 / \|x^*\|_2 < 10^{-3}$.

Figure 3 reports the average empirical probability of success for each setting of $(n, s_{\text{cor}})$ averaged over the 10 iterations. To compare these empirical results with our theory, we overlay the theoretical recovery thresholds

$$\mu_n^2 = \omega (T_{\text{sig}} \cap S^{p-1})^2 + \omega (T_{\text{cor}} \cap S^{n-1})^2$$

where the Gaussian width for the binary signal is estimated with the bound given in Table I, while the width for the sparse corruption is given by the optimal squared distance bound for the $\ell_1$ norm, given in (10). This threshold shows the parameter regions where our theory gives any positive probability of successful recovery (we have ignored the small additive constants, which are likely artifacts of the proof and are negligible for large $n$). We find that our theory provides a practicable phase transition that reflects empirical behavior.

Sparse signal, sparse corruption: We next consider the problem of recovering a sparse signal vector of known $\ell_1$ norm from sparsely corrupted Gaussian measurements. We fix the signal length and sample size $p = n = 1000$, and partition the indices of $\Phi$ into $m = 100$ blocks of size $k = 10$. We test sparsity and block-sparsity levels $(s_{\text{sig}}, s_{\text{cor}}) \in [1, 860] \times [1, 86]$. For each $(s_{\text{sig}}, s_{\text{cor}})$ pair, we follow the experimental setup of the previous experiment with two substitutions: we generate $v^*$ with $s_{\text{cor}}$ blocks of independent standard normal entries and $m-s_{\text{cor}}$ blocks with all entries set to zero, and we solve the constrained optimization problem

$$(\hat{x}, \hat{v}) \in \arg\min_{x,v} \left\{ \|v\|_{\ell_1/\ell_2} : \|x\|_1 \leq \|x^*\|_1, y = \Phi x + v \right\}.$$
Sparsity level of signal vector

Sparsity level of corruption vector

Empirical probability of recovery

Figure 4: Phase transition for sparse signal recovery under sparse corruption, using constrained recovery (see Section V-A). The red curve plots the recovery threshold predicted by Theorem 1 (ignoring small additive constants), where Gaussian width is estimated by minimizing (10).

Results from this simulation are displayed in Figure 5, with the overlaid threshold coming from the optimal squared distance bound for the $\ell_1$ norm, given in (10), and for the $\ell_1/\ell_2$ norm, as follows (see (25) in Appendix A-C for the derivation):

$$\min_{t \geq 0} \mathbb{E} \left[ \inf_{w \in t \cdot \partial \|v^*\|_{\ell_1/\ell_2}} \|g - w\|_2^2 \right] = \min_{t \geq 0} s_{\text{cor}}(t^2 + k) + \frac{2^{1-k/2}(m - s_{\text{cor}})}{\Gamma(k/2)} \int_1^\infty (c - t)^{k-1} e^{-c^2/2} dc.$$

Again, we see that our theory accurately matches the observed phase transition of successful recovery. For comparison, we also plot the theoretical recovery threshold from the previous experiment (sparse signal and sparse corruption), which would be the recovery threshold for optimization using the $\ell_1$ norm on both signal and corruption, rather than leveraging the additional block structure via the $\ell_1/\ell_2$ norm.

B. Phase transitions from penalized recovery

Next, we will consider the penalized recovery program (13) in the noiseless setting where neither $\|x^*\|_{\text{sig}}$ nor $\|v^*\|_{\text{cor}}$ is known a priori. We focus on a sparse signal $x^*$ plus sparse corruption $v^*$ model—note that for this type of structure, in practice neither $\|x^*\|_{\text{sig}} = \|x^*\|_1$ nor $\|v^*\|_{\text{cor}} = \|v^*\|_1$ will be known exactly a priori, and therefore penalized recovery is often more practical than constrained recovery.

Since constrained recovery is guaranteed to be successful at a sample size of approximately $\omega \left( T_{\text{sig}} \cap S^{p-1} \right)^2 + \omega \left( T_{\text{cor}} \cap S^{n-1} \right)^2$, the known correspondence between constrained and penalized optimization implies that for some (unknown) value of $\lambda$, the penalized recovery program should also yield exact recovery at this sample size. The primary difficulty in the penalized setting, however, lies in choosing a penalty parameter that leads to good recovery
1) $\lambda_{\text{sparse}} = \sqrt{\log(n/s_{\text{cor}})}/\sqrt{\log(p/s_{\text{sig}})}$, based on the analytic bound (8) that is nearly optimal for highly sparse vectors,

2) $\lambda_{\text{dense}} = (1 - s_{\text{cor}}/n)/(1 - s_{\text{sig}}/p)$, based on the analytic bound (9) that is nearly optimal for all but the sparsest vectors,

3) $\lambda_{\text{opt}}$, which chooses $t_{\text{sig}}$ and $t_{\text{cor}}$ to minimize the respective expected squared distances exactly via (10), and

4) $\lambda_{\text{const}} = 1$, for which we expect recovery when both $s_{\text{sig}}$ and $s_{\text{cor}}$ are sufficiently sparse (see Corollary 3).

For direct comparison with the constrained recovery case, we generate sparse signal, sparse corruption pairs and noiseless Gaussian measurements with $p = n = 1000$, precisely as in Section V-A. To recover each $(x^*, v^*)$ pair, we solve the penalized optimization problem (18) with each penalty parameter setting and declare success if $\|\hat{x} - x^*\|_2 / \|x^*\|_2 < 10^{-3}$. Figure 5 displays the empirical probability of success as the signal and corruption sparsity levels $s_{\text{sig}}$ and $s_{\text{cor}}$ vary. For reference, the theoretical recovery curve of Figure 4 has been overlaid. Remarkably, setting $\lambda$ to equal any of $\lambda_{\text{sparse}}$, $\lambda_{\text{dense}}$, or $\lambda_{\text{best}}$ yields empirical performance nearly as good that obtained in the constrained setting (Figure 4) but without knowledge of $\|x^*\|_1$ or $\|v^*\|_1$. In other words, although
Figure 6: Phase transitions for sparse signal recovery under sparse corruption, using penalized recovery with varied settings of the penalty parameter $\lambda$ (see Section V-B). The red curve in each figure is the theoretical recovery threshold predicted by Theorem 1 for the constrained recovery problem (same curve as in Figure 4). Grayscale representation of success probability is same as in Figure 4.
Theorem 2 requires a larger number of measurements than Theorem 1 to guarantee success, the penalized recovery program offers nearly the same phase transition as the constrained program in practice. Moreover, as predicted in Corollary 3 when estimates of the sparsity levels \( s_{\text{sig}} \) and \( s_{\text{cor}} \) are unavailable, the setting \( \lambda = \lambda_{\text{const}} = 1 \) yields high probability recovery, provided that neither signal nor corruption is overly dense.

\section*{C. Stable recovery error}

Finally, we study the empirical behavior of the constrained recovery program \((12)\) in the noisy measurement setting, where \( \delta \neq 0 \). We will compare the achieved error in recovering \((x^*, v^*)\) to the error bounds predicted by Theorem 1. We focus on the recovery of a sparse signal from sparsely corrupted measurements perturbed by dense, unstructured noise. For fixed noise level \( \delta = 1 \) and sparsity fractions \((\gamma_{\text{sig}}, \gamma_{\text{cor}}) = (0.01, 0.4)\), we vary the signal length \( p \in \{100, 300, 500\} \) and the number of measurements \( n \in [100, 6000] \) and perform the following experiment 20 times for each \((p, n)\) pair:

1) Generate a signal vector \( x^* \) with \( p \cdot \gamma_{\text{sig}} \) i.i.d. standard normal entries and \( p \cdot (1 - \gamma_{\text{sig}}) \) entries set to 0.
2) Draw a Gaussian matrix \( \Phi \in \mathbb{R}^{n \times p} \) with independent \( N(0, 1/n) \) entries.
3) Generate a corruption vector \( v^* \) with \( n \cdot \gamma_{\text{cor}} \) i.i.d. standard normal entries and \( n \cdot (1 - \gamma_{\text{cor}}) \) entries set to 0.
4) Solve the following constrained optimization problem with \( y = \Phi x^* + v^* \):
\[
(\tilde{x}, \tilde{v}) \in \arg \min_{x,v} \{||v||_1 : ||x||_1 \leq ||x^*||_1, ||y - \Phi x - v||_2 \leq \delta \},
\]

5) Record the \( \ell_2 \) recovery error \[ \sqrt{||\tilde{x} - x^*||^2_2 + ||\tilde{v} - v^*||^2_2} \].

Our theory suggests that rescaling the \( \ell_2 \) recovery error by a factor of
\[
n^{-1/2} \left( \mu_n - \sqrt{\omega (T_{\text{sig}} \cap S^{p-1})^2 + \omega (T_{\text{cor}} \cap S^{n-1})^2} \right),
\]
will give rise to quantity bounded by a universal constant, independent of \( n \) or \( p \). This is precisely what we observe in Figure 7. Here we have plotted, for each setting of \( p \) and \( n \), the average recovery error and the average rescaled recovery error across 20 iterations. We have used the optimal squared distance bound \((10)\) to estimate the widths in the scaling factor \((20)\). We see that while the absolute error curves vary with the choice of \( p \), the rescaled error curves converge to a common value, as predicted by the results of Theorem 1.
VI. CONCLUSIONS AND FUTURE WORK

We have presented a geometric approach to analyzing the recovery of structured signals from corrupted and noisy measurements. We analyzed both penalized and constrained convex programs and, in each case, provided conditions for exact signal recovery from structured corruption and stable signal recovery from structured corruption with added unstructured noise. Our analysis revolved around two geometric measure of signal complexity, the Gaussian width and the Gaussian distance, for which we developed new analytical bounds. The utility of our theory was borne out in our simulations, which demonstrated close agreement between our theoretical recovery bounds and the sharp phase transitions observed in practice. We envision several interesting directions for future work:

Matching theory in penalized and constrained settings: In Section V, we observed that the penalized convex program (1) with canonical choice of penalty parameter performed nearly as well empirically as the constrained convex program (2) with side information. This suggests that our penalized recovery theory could be sharpened to more closely match that obtainable in the constrained recovery setting.

Penalized noise term: It would be of great practical interest to analyze the fully penalized convex program

\[
\min_{x,v} \left\{ \|x\|_{\text{sig}} + \lambda \|v\|_{\text{cor}} + \nu \|y - (\Phi x + v)\|_2 \right\} \quad \text{or} \quad \min_{x,v} \left\{ \|x\|_{\text{sig}} + \lambda \|v\|_{\text{cor}} + \nu \|y - (\Phi x + v)\|_2 \right\}
\]

when no prior bound \(\delta\) on the noise level is available.

Stochastic noise: Our analysis of stable recovery in the presence of unstructured noise assumes only that the noise is bounded in the \(\ell_2\) norm. We anticipate improved bounds on estimation error under more specific, stochastic assumptions such as sub-Gaussian or sub-exponential noise.

Non-Gaussian measurements: Finally, an important open question is to what extent the results of this work extend to corrupted sensing problems with non-Gaussian measurements, either stochastic or deterministic, with suitable incoherence conditions.
**APPENDIX A**

**TANGENT CONE COMPLEXITY BOUNDS**

In this section, we characterize the complexity of the tangent cones generated by the structured vectors and structure-inducing norms introduced in Section II-D. For each tangent cone, we will present new and existing bounds on the Gaussian width $\omega(T \cap S^{p-1})$ and the Gaussian distance $\eta(t \cdot \partial \|x\|)$. The results are summarized in Table I and Table II. Throughout, we use the notation
\[
\text{dist}(x, C) = \inf_{w \in C} \|x - w\|_2
\]
to refer to the Euclidean distance between a vector $x \in \mathbb{R}^p$ and a set $C \subset \mathbb{R}^p$.

**A. Fundamentals**

Our width and bounds will be based on two fundamental relationships between Gaussian widths and Gaussian distances. The first relation, established in [14], upper bounds the Gaussian width of a constrained tangent cone in terms of the Gaussian distance and the expected squared distance to the scaled subdifferential
\[
\omega(T \cap S^{p-1}) \leq \omega(T \cap B^p) \leq \min_{t \geq 0} \eta(t \cdot \partial \|x\|) \leq \min_{t \geq 0} \sqrt{\mathbb{E} \left[ \text{dist}(g, t \cdot \partial \|x\|)^2 \right]}, \tag{21}
\]
Our bounds on the right-hand side of (21) will provide specific settings of the subdifferential scale $t$ that can then be used in choosing a penalty parameter for penalized optimization (see Theorem 2).

In fact, the second fundamental relation provides such a bound on the expected squared distance, in terms of the Gaussian width of the subdifferential, via the following result (proved in Appendix A-F):

**Proposition 2.** Let $\tilde{\omega}$ be any lower bound on $\omega(\partial \|x\|)$. For $t = \frac{\omega}{\max_{\|z\|=1} \|z\|}$,
\[
\mathbb{E} \left[ \text{dist}(g, t \cdot \partial \|x\|)^2 \right] \leq p - \left( \frac{\omega}{\max_{\|z\|=1} \|z\|} \right)^2.
\]
We will see that the subdifferentials of many structured norms admit simple lower bounds that lead to tight upper bounds on the Gaussian distance and Gaussian width via Proposition 2.

**B. Sparse vectors**

We begin by considering an $s$-sparse vector $x \in \mathbb{R}^p$ and the sparsity inducing norm $\|\cdot\|_1$. The subdifferential and tangent cone were given in (5) and (6) respectively. In this setting, Chandrasekaran et al. [14] established the squared bound
\[
\omega(T \cap S^{p-1}) \leq \eta(t \cdot \partial \|x\|) \leq \sqrt{\mathbb{E} \left[ \text{dist}(g, t \cdot \partial \|x\|_1)^2 \right]} \leq 2s \log \left( \frac{p}{s} \right) + \frac{3}{2}s,
\]
for $t = \sqrt{2 \log \left( \frac{p}{s} \right)}$. Using Proposition 2, we obtain a second bound
\[
\omega(T \cap S^{p-1}) \leq \eta(t' \cdot \partial \|x\|) \leq \sqrt{\mathbb{E} \left[ \text{dist}(g, t' \cdot \partial \|x\|_1)^2 \right]} \leq \sqrt{p} \sqrt{1 - 2/\pi (1 - s/p)}, \tag{22}
\]
for $t' = \sqrt{2/\pi (1 - s/p)}$. This follows from our more general treatment of block-sparse vectors in Proposition 3 below.

**C. Block-sparse vectors**

Suppose that the indices $\{1, \ldots, p\}$ have been partitioned into disjoint blocks $V_1, \ldots, V_m$ and that $x$ is supported only on $s$ of these blocks. Then it is natural to consider a norm that encourages block sparsity, e.g.,
\[
\|x\|_{t_1/t_2} = \sum_{b=1}^{m} \|x_{V_b}\|_2.
\]
Proposition 3 presents our two new bounds on Gaussian distance and hence on Gaussian width in this setting.
Proposition 3 (Block-sparse vector Gaussian distance). Partition the indices \( \{1, \ldots, p\} \) into blocks \( V_1, \ldots, V_m \) of size \( k = p/m \), and let \( \|x\|_{\ell_1/\ell_2} = \sum_{b=1}^m \|x_{V_b}\|_2 \). If \( x^* \) is supported on at most \( s \) of these blocks, then both of the following estimates hold:

\[
\eta \left( t \cdot \partial \|x\|_{\ell_1/\ell_2} \right) \leq \sqrt{4s \log(m/s) + (0.5 + 3k)s} \quad \text{for} \quad t = \sqrt{2 \log(m/s) + \sqrt{E}},
\]

\[
\eta \left( t' \cdot \partial \|x\|_{\ell_1/\ell_2} \right) \leq \sqrt{\pi} \sqrt{1 - \frac{\mu_k^2}{s}} (1 - s/m) \quad \text{for} \quad t' = \mu_k (1 - s/m).
\]

For element-wise sparsity \( (k = 1) \), the bound (24) specializes to the bound (22) given above. The advantage of exploiting block-sparse structure is evident when one compares the factor of \( \frac{\mu_k^2}{s} \) in the bound (24) to the term \( \frac{\sqrt{m}}{s} \) found in (22). The former is larger and approaches 1 as the block size \( k = p/m \) grows, and hence fewer measurements will suffice to recover from these block-sparse corruptions.

Proof. Throughout, let \( B \) denote the block indices on which \( x \) is supported, with \( |B| = s' \leq s \), let \( T \) be the space of vectors with support only on those blocks in \( B \), and let \( a \in T \) be the vector satisfying

\[
a_{V_b} = \begin{cases} x_{V_b}/\|x_{V_b}\|_2 & \text{if } b \in B \\ 0 & \text{otherwise.} \end{cases}
\]

Then, we have the subdifferential (32)

\[
\partial \|x\|_{\ell_1/\ell_2} = a + \{ w \in T^\perp : \max_{b \in B} \|w_{V_b}\|_2 \leq 1 \}.
\]

We begin by establishing the bound (24) based on the width of the subdifferential. For each \( b, \|P_{V_b}(g)\|_2 \) is \( \chi_k \) distributed. Using the form of the subdifferential stated above,

\[
\omega \left( \partial \|x\|_{\ell_1/\ell_2} \right) = \mathbb{E} \left[ \langle g, a \rangle + \sum_{b \in B} \sup_{w \in \mathbb{R}^p : \|w_{V_b}\|_2 \leq 1} \langle g, w_{V_b} \rangle \right] = \sum_{b \in B} \mathbb{E} \left[ \sup_{w \in \mathbb{R}^p : \|w\|_2 \leq 1} \langle P_{V_b}(g), w \rangle \right] = \sum_{b \in B} \mathbb{E} \left[ \|P_{V_b}(g)\|_2 \right] = (m - s') \mu_k \geq (m - s) \mu_k.
\]

Since \( \sup_{g \in \mathbb{R}^p : \|g\|_2 = 1} \|g\|_{\ell_1/\ell_2} = \sqrt{m} \), Proposition 2 now implies the result (24).

Now we turn to the bound (23). By (21), it suffices to bound the expected squared distance

\[
\mathbb{E} \left[ \text{dist} \left( g, t \cdot \partial \|x\|_{\ell_1/\ell_2} \right)^2 \right]
\]

for the given value of \( t \). Note that

\[
\mathbb{E} \left[ \text{dist} \left( g, t \cdot \partial \|x\|_{\ell_1/\ell_2} \right)^2 \right] = \mathbb{E} \left[ \|P_T(g) - ta\|_2^2 + \frac{\inf_{w_{V_b} \in \mathbb{R}^p : \|w_{V_b}\|_2 \leq t} \|P_{V_b}(g) - \omega_b\|_2^2} \right] = \mathbb{E} \left[ \|P_T(g) - ta\|_2^2 + \frac{\sum_{b \in B} \|P_{V_b}(g) - \omega_b\|_2^2} \right] = s'(t^2 + p/m) + \sum_{b \in B} \mathbb{E} \left[ \|P_{V_b}(g) - \omega_b\|_2^2 \right] \leq s'(t^2 + p/m) + (m - s') \cdot \mathbb{E} \left[ (\xi - t)^2 \right] \leq s'(t^2 + p/m) + (m - s) \cdot \mathbb{E} \left[ (\xi - t)^2 \right],
\]

where \( \xi \) is a \( \chi_k \) random variable. We will bound each summand in this expression.
Letting \( d = t - \sqrt{k} \), we have
\[
\mathbb{E} \left[ (\xi - t)^2 \right] = \int_0^\infty \mathbb{P} \left\{ (\xi - t)^2 \geq a \right\} \, da \\
= \int_0^\infty \mathbb{P} \left\{ \xi^2 \geq (t + \sqrt{a})^2 \right\} \, da \\
= \int_0^\infty \mathbb{P} \left\{ \xi^2 - k \geq 2\sqrt{k}(d + \sqrt{a}) + (d + \sqrt{a})^2 \right\} \, da
\]
Applying the change of variables \( c = (d + \sqrt{a}) \),
\[
= 2\int_d^\infty \mathbb{P} \left\{ \xi^2 - k \geq 2\sqrt{k}c + c^2 \right\} (c - d) \, dc \\
\leq 2\int_d^\infty e^{-c^2/2}(c - d) \, dc \\
= 2e^{-d^2/2} - 2d\int_d^\infty e^{-c^2/2} \, dc \\
\leq 2e^{-d^2/2} - 2d^2 e^{-d^2/2} = \frac{2}{d^2 + 1}e^{-d^2/2},
\]
where the penultimate inequality follows from the chi-squared tail bound [33, Lem. 1],
\[
\mathbb{P} \left\{ \xi^2 - k \geq 2\sqrt{k}c + c^2 \right\} \leq e^{-c^2/2} \quad \text{for all} \quad c > 0,
\]
and the final inequality follows from a bound on the Gaussian Q-function
\[
Q(d) = \frac{1}{\sqrt{2\pi}} \int_d^\infty e^{-c^2/2} \, dc \geq \frac{1}{\sqrt{2\pi}} \frac{d^2}{d^2 + 1} \int_d^\infty \frac{c^2 + 1}{c^2} e^{-c^2/2} \, dc = \frac{1}{\sqrt{2\pi}} \frac{d}{d^2 + 1} e^{-d^2/2} \quad \text{for all} \quad d > 0.
\]
Hence we have
\[
\mathbb{E} \left[ \text{dist} \left( g, t \cdot \partial \|x\|_{\ell_1/t_2} \right)^2 \right] \leq s(t^2 + p/m) + \frac{2(m-s)}{d^2 + 1} e^{-d^2/2} \\
= s(d^2 + 2d\sqrt{k} + 2k) + \frac{2(m-s)}{d^2 + 1} e^{-d^2/2} \\
\leq s(2d^2 + 3k) + \frac{2(m-s)}{d^2 + 1} e^{-d^2/2}.
\]
Since \( d = \sqrt{2\log(m/s)} \), we have
\[
\mathbb{E} \left[ \text{dist} \left( g, t \cdot \partial \|x\|_{\ell_1/t_2} \right)^2 \right] \leq s(4\log(m/s) + 3k) + \frac{2s(1-s/m)}{2\log(m/s) + 1} \\
\leq (0.5 + 3k)s + 4s\log(m/s),
\]
as desired, since
\[
\frac{2(1-s/m)}{2\log(m/s) + 1} < 0.5
\]
whenever \( 1 \leq s \leq m \).

\textbf{D. Binary vectors}

The convex hull of the set of binary vectors \( \{ \pm 1 \}^p \) is the unit ball of the \( \ell_1 \) norm, so we choose \( \| \cdot \| = \| \cdot \|_{\ell_\infty} \) when \( x \in \{ \pm 1 \}^p \). This choice yields
\[
\partial \|x\|_{\ell_\infty} = \{ w \in \mathbb{R}^p : \|w\|_1 = 1, w_ix_i \geq 0 \ \forall i \}
\]
and therefore
\[
T = \{ w \in \mathbb{R}^p : w_ix_i \leq 0 \ \forall i \} \quad \text{and} \quad N = \{ w \in \mathbb{R}^p : w_ix_i \geq 0 \ \forall i \}.
\]
Thus, \( \omega(T \cap S^{p-1}) = \omega(N \cap S^{p-1}) \), and \([14, \text{Lem. 3.7}]\) implies that
\[
\omega(T \cap S^{p-1}) \leq \sqrt{\frac{p}{2}}.
\]

We note that a bound based on Proposition\([2] is typically looser in this setting.\]

**E. Low-rank matrices**

When \( x \in \mathbb{R}^{m_1 \times m_2} \) is a rank \( r \) matrix, we consider the trace norm \( \| \cdot \|_\tau \). Chandrasekaran et al. \([14]\) established the bound \( \omega(T \cap S^{m_1 \times m_2}) \leq \frac{3}{r}(m_1 + m_2 - r) \). Proposition\([4] presents our new estimate based on the Gaussian width of the subdifferential.

**Proposition 4** (Low-rank matrix Gaussian width). Let \( x \in \mathbb{R}^{m_1 \times m_2} \) have rank at most \( r \), and let \( T \) be the tangent cone associated with \( \| \cdot \|_\tau \) at \( x \). If \( m_1 \geq m_2 \), then setting \( t = \frac{1}{27}(m_2 - r)\sqrt{m_1 - r/m_2} \),
\[
\eta(t \cdot \partial \| x \|_\tau) \leq \sqrt{m_1 m_2} \cdot \sqrt{1 - \left( \frac{4}{27} \right)^2 \left( 1 - \frac{r}{m_1} \right) \left( 1 - \frac{r}{m_2} \right)^2}.
\]

**Proof.** Let \( U \Sigma V^T \) be the compact singular value decomposition of \( x \), and let \( T \) be the space of matrices in the column or row space of \( x \). The subdifferential is given by \([34]\)
\[
\partial \| x \|_\tau = U V^T + \{ w \in T^\perp : \| w \|_{op} \leq 1 \},
\]
where \( \| w \|_{op} \) is the operator norm, i.e., the largest singular value, of \( w \).

We begin by bounding the Gaussian width of \( \partial \| x \|_\tau \). Let \( \Gamma \in \mathbb{R}^{m_1 \times m_2} \) have i.i.d. standard Gaussian entries. Since orthogonal projection cannot increase the operator norm, we have
\[
\omega(\partial \| x \|_\tau) = E \left[ \langle \Gamma, U V^T \rangle + \sup_{w: \| \mathcal{P}_\Gamma^T (w) \|_{op} \leq 1} \langle \Gamma, \mathcal{P}_\Gamma^T (w) \rangle \right] \\
\geq E \left[ \sup_{w: \| w \|_{op} \leq 1} \langle \mathcal{P}_\Gamma^T (\Gamma), w \rangle \right] = E \left[ \| \mathcal{P}_\Gamma^T (\Gamma) \|_\tau \right].
\]

Now, suppose that \( \tilde{U} = [U \ U'] \) and \( \tilde{V} = [V \ V'] \) are orthonormal for \( U' \in \mathbb{R}^{m_1 \times (m_1 - r)} \) and \( V' \in \mathbb{R}^{m_2 \times (m_2 - r)} \). By definition of \( T \), we know that \( \mathcal{P}_\Gamma^T (\Gamma) = U'U'^T \Gamma V'V'^T \), and so
\[
E \left[ \| \mathcal{P}_\Gamma^T (\Gamma) \|_\tau \right] = E \left[ \| U'U'^T \Gamma V'V'^T \|_\tau \right] = E \left[ \| U'^T \Gamma V' \|_\tau \right],
\]

since the trace norm is unitarily invariant. Furthermore, since \( U' \) and \( V' \) each have orthonormal columns, the entries of \( U'^T \Gamma V' \) are i.i.d. standard Gaussian. Therefore, applying Lemma\([11]\)
\[
\omega(\partial \| x \|_\tau) = E \left[ \| U'^T \Gamma V' \|_\tau \right] \geq \frac{4}{27}(m_2 - r)\sqrt{m_1 - r}.
\]

Finally, we apply Proposition\([2] to obtain the desired bound. \( \square \)

**F. Proof of subdifferential distance bound**

We begin with a definition and a lemma (proved in Appendix\([D]\). The **dual norm** to \( \| \cdot \| \) is defined as
\[
\| w \|^* = \sup_{\| x \| \leq 1} \langle w, x \rangle,
\]
and satisfies
\[
\langle w, x \rangle \leq \| x \| \cdot \| w \|^* \text{ for all } w, x \in \mathbb{R}^p.
\]

**Lemma 1.** For any norm \( \| \cdot \| \) on \( \mathbb{R}^p \) with dual norm \( \| \cdot \|^* \),
\[
\max_{x \in \mathbb{R}^p} \| x \| = \max_{x \in \mathbb{R}^p} \| x \|^*.
\]

Now we prove our Gaussian distance bound that is based on the subdifferential. We use the fact that \( \| w \|^* = 1 \) for all \( w \in \partial \| x \| \).
Proof of Proposition 2} Fix any \( g \in \mathbb{R}^p \), and choose any

\[
    w_0 \in \arg \max_{w \in \partial \|x\|} \langle g, w \rangle .
\]

(Since \( \partial \|x\| \) is closed and is in the unit sphere of the norm \( \|\cdot\|^* \), it is compact, and so the maximum must be attained at some \( w_0 \).) Then for any \( t \),

\[
    \text{dist}(g, t \cdot \partial \|x\|)^2 \leq \|g - t \cdot w_0\|_2^2 = \|g\|_2^2 - 2t \langle g, w_0 \rangle + t^2 \|w_0\|_2^2
\]

Since \( \|w\|^* = 1 \) for all \( w \in \partial \|x\| \),

\[
    \leq \|g\|_2^2 - 2t \max_{w \in \partial \|x\|} \langle g, w \rangle + t^2 \max_{\|w\|=1} \|w\|_2^2
\]

\[
    \leq \|g\|_2^2 - 2t \max_{w \in \partial \|x\|} \langle g, w \rangle + t^2 \max_{\|w\|=1} \|w\|_2^2 ,
\]

where the last step comes from Lemma 1. Taking expectations,

\[
    \mathbb{E} \left[ \text{dist}(g, t \cdot \partial \|x\|)^2 \right] \leq p - 2t \cdot \omega (\partial \|x\|) + t^2 \cdot \max_{\|w\|=1} \|w\|_2^2 \leq p - 2t \cdot \omega + t^2 \cdot \max_{\|w\|=1} \|w\|_2^2 .
\]

Now we can minimize this quadratic in \( t \) by setting \( t = \frac{\omega}{\max_{\|w\|=1} \|w\|_2^2} \), which yields the desired bound. \( \square \)

G. Relating Gaussian distance and Gaussian width

Proposition 7} Suppose that \( \partial \|x\| \) satisfies a weak decomposability assumption:

\[
    \exists w_0 \in \partial \|x\| \text{ s.t. } \langle w - w_0, w_0 \rangle = 0 \quad \forall w \in \partial \|x\| .
\]

Then

\[
    \min_{t \geq 0} \eta (t \cdot \partial \|x\|) \leq \omega (T \cap \mathbb{S}^{p-1}) + 6 .
\]

Proof. For \( g \in \mathbb{R}^p \), define\[7\]

\[
    t_g = \arg \min_{t \geq 0} \text{dist}(g, t \cdot \partial \|x\|) .
\]

Define the event

\[
    \mathcal{E} = \left\{ |t_g - \mathbb{E} [t_g]| < \frac{2}{\|w_0\|_2} \right\} .
\]

Lemma 3 below shows that, when assumption (26) is satisfied, \( g \mapsto t_g \) is a \( \frac{1}{\|w_0\|_2} \) Lipschitz function of \( g \). Therefore, applying a bound due to Ledoux [35, (2.8)],

\[
    \mathbb{P} \{ \mathcal{E} \} \geq 1 - 2e^{-2^2/2} .
\]

Now suppose that \( \mathcal{E} \) holds. Find \( w \in \partial \|x\| \) such that \( t_g \cdot w \) is the projection of \( g \) to \( N \), that is,

\[
    \text{dist}(g, N) = \text{dist}(g, t_g \cdot \partial \|x\|) = \|g - t_g \cdot w\|_2 .
\]

Since the subdifferential \( \partial \|x\| \) is convex, we have

\[
    t_g \frac{\mathbb{E} [t_g]}{\|w_0\|_2} \cdot w + \frac{\mathbb{E} [t_g]}{\|w_0\|_2} + \frac{\|w_0\|_2}{\|w_0\|_2} - t_g \cdot w_0 \in \partial \|x\| ,
\]

\[6\] Note that \( t_g \) is unique because \( N = \cup_{t \geq 0} t \cdot \partial \|x\| \) is convex, and so there is a unique projection of \( g \) onto this cone; since \( \|w\|^* = 1 \) for all \( w \in t \cdot \partial \|x\| \), there is no overlap between \( t \cdot \partial \|x\| \) and \( t' \cdot \partial \|x\| \) for any \( t \neq t' \), and so \( t_g \) is defined uniquely from the projection of \( g \) onto \( N \).
and so
\[
\text{dist} \left( g, (E[t_g] + \frac{2}{\|w_0\|_2}) \cdot \|x\| \right) \leq \left\| g - \left[ t_g \cdot w + (E[t_g] + \frac{2}{\|w_0\|_2}) \cdot w_0 \right] \right\|_2 \\
\leq \left\| g - t_g \cdot w \right\|_2 + (E[t_g] + \frac{2}{\|w_0\|_2}) \cdot \|w_0\|_2 \\
= \text{dist}(g, N) + (E[t_g] + \frac{2}{\|w_0\|_2}) \cdot \|w_0\|_2 \\
< \text{dist}(g, N) + 4 ,
\]
where the last step comes from the definition of the event \( \mathcal{E} \). Therefore,
\[
P \left\{ \frac{1}{2} \text{dist} \left( g, (E[t_g] + \frac{2}{\|w_0\|_2}) \cdot \|x\| \right) - \frac{1}{2} \text{dist}(g, N) < 2 \right\} \geq P \{ \mathcal{E} \} \geq 1 - 2e^{-2/2} .
\]
Since \( g \mapsto \text{dist}(g, A) \) is clearly a 1-Lipschitz function of \( g \) for any convex set \( A \), we see that
\[
g \mapsto \frac{1}{2} \text{dist} \left( g, (E[t_g] + \frac{2}{\|w_0\|_2}) \cdot \|x\| \right) - \frac{1}{2} \text{dist}(g, N)
\]
is a 1-Lipschitz function of \( g \). Therefore, applying Lemma 4 below,
\[
\eta \left( (E[t_g] + \frac{2}{\|w_0\|_2}) \cdot \|x\| \right) - \eta (N) = \mathbb{E} \left[ \frac{1}{2} \text{dist} \left( g, (E[t_g] + \frac{2}{\|w_0\|_2}) \cdot \|x\| \right) - \frac{1}{2} \text{dist}(g, N) \right] \\
\leq 2 + \sqrt{2 \log \left( \frac{1}{1 - 2e^{-2/2}} \right)} .
\]
Since \( \eta (N) = \omega (T \cap B^p) \leq \omega (T \cap S^{p-1}) + \frac{1}{\sqrt{2\pi}} \) by Lemma 6 we see that
\[
\eta \left( (E[t_g] + \frac{2}{\|w_0\|_2}) \cdot \|x\| \right) \leq \omega (T \cap S^{p-1}) + \frac{1}{\sqrt{2\pi}} + 2 \cdot \left( 2 + \sqrt{2 \log \left( \frac{1}{1 - 2e^{-2/2}} \right)} \right) \\
\leq \omega (T \cap S^{p-1}) + 6 .
\]

\[\Box\]

**APPENDIX B**

**Proofs of Application Results**

**A. Corollaries**

First we prove our result on constrained recovery of a binary signal corrupted with sparse noise.

**Proof of Corollary 7** The bound 16 follows from the \( \ell_1 \) tangent cone width bounds of Table 1. Next, applying Theorem 1 with any \( \epsilon \in (2\delta, \frac{1}{\sqrt{2\pi}}) \) (and making use of the \( \ell_\infty \) tangent cone width bound of Table 1), we see that with probability at least \( 1 - e^{-\beta(\mu_n - \epsilon/\sqrt{\pi-\tau})^2/2} \),
\[
\|\widehat{x} - x^*\|_\infty \leq \sqrt{\|\widehat{x} - x^*\|^2 + \|\widehat{v} - v^*\|_2^2} \leq \frac{2\delta}{\epsilon} < 1 .
\]
This implies that \( x^* \) is the nearest binary vector to \( \widehat{x} \), that is, we can exactly recover \( x^* \) by setting \( x^* = \text{sign}(\widehat{x}) \). Letting \( \epsilon \) approach \( 2\delta \), we see that this is true with probability at least \( 1 - e^{-\beta(\mu_n - 2\delta/\sqrt{\pi-\tau})^2/2} \). \[\Box\]

Next we prove our result on penalized recovery of a structured signal observed with a high frequency of block-wise corruptions.

**Proof of Corollary 2** Theorem 2 implies stable recovery with error at most \( 2\delta/\epsilon \) and probability at least \( 1 - \beta \) once
\[
\mu_n - \sqrt{n} \epsilon \geq 2\eta \left( \ell_{\text{sig}} \cdot \|x^*_\text{sig}\|_{\ell_{\text{sig}}} \right) + \eta \left( \mu_k (1 - \gamma) \cdot \|v^*_\ell_{\text{sig}}\|_{\ell_{\text{sig}}} \right) + \sqrt{2 \log(1/\beta)} + 3\sqrt{2\pi} + \frac{1}{\sqrt{2\pi}} .
\]
Since \( \mu_n > \sqrt{n - 1/2} \geq \sqrt{n} - (1 - 1/\sqrt{2}) \) for all \( n \geq 1 \), it is sufficient to have
\[
\sqrt{n}(1 - \epsilon) \geq 2\eta \left( t_{\text{sig}} \cdot \partial \|x^*\|_1 \right) + \eta \left( \mu_k (1 - \gamma) \cdot \partial \|v^*\|_{\ell_2} \right) + \sqrt{2 \log(1/\beta)} + 3\sqrt{2\pi} + 1 + \frac{1}{\sqrt{2\pi}}.
\]
Table II bounds \( \eta \left( \mu_k (1 - \gamma) \cdot \partial \|v^*\|_{\ell_2} \right) \) by \( \sqrt{n} \sqrt{1 - \mu^2_k/(1 - \gamma)^2} \), and hence the condition
\[
\sqrt{n} \geq \frac{2\eta \left( t_{\text{sig}} \cdot \partial \|x^*\|_1 \right) + \sqrt{2 \log(1/\beta)} + 3\sqrt{2\pi} + 1 + \frac{1}{\sqrt{2\pi}}}{1 - \sqrt{1 - \mu^2_k/(1 - \gamma)^2} - \epsilon}
\]
is sufficient for recovery. To obtain the final form of our statement, we note that
\[
1 - \sqrt{1 - \mu^2_k/(1 - \gamma)^2} = \frac{\mu^2_k/(1 - \gamma)^2}{1 + \sqrt{1 - \mu^2_k/(1 - \gamma)^2}} \geq \frac{\mu^2_k}{1 + \sqrt{1 - \mu^2_k/(1 - \gamma)^2}} (1 - \gamma)^2 = 1 - \sqrt{1 - \mu^2_k/(1 - \gamma)^2} (1 - \gamma)^2.
\]

Finally we prove our result on penalized recovery of an extremely sparse signal observed under extremely sparse corruptions.

**Proof of Corollary 3.** To establish the result with \( \lambda = 1 \), we choose
\[
t_{\text{sig}} = t_{\text{cor}} = \sqrt{2 \log \left( \frac{p + n}{s_{\text{sig}} + s_{\text{cor}}} \right)}.
\]
To compute the relevant expected squared distances, we use the fact that, by (10),
\[
\mathbb{E} \left[ \text{dist}(g, t_{\text{sig}} \cdot \partial \|x^*\|_1)^2 \right] \leq s_{\text{sig}} (1 + t_{\text{sig}}^2) + \frac{2(p - s_{\text{sig}})}{\sqrt{2\pi}} (1 + t_{\text{sig}}^2) \int_{t_{\text{sig}}}^{\infty} \exp(-c^2/2) \ dc - t_{\text{sig}} \exp(-t_{\text{sig}}^2/2)
\]
Bounding the integral as in [14, Proof of Prop. 3.10],
\[
\leq s_{\text{sig}} (1 + t_{\text{sig}}^2) + \frac{2(p - s_{\text{sig}})}{\sqrt{2\pi}} \cdot \frac{1}{t_{\text{sig}}} \exp(-t_{\text{sig}}^2/2)
\]
\[
\leq s_{\text{sig}} \left( 1 + 2 \log \left( \frac{p + n}{s_{\text{sig}} + s_{\text{cor}}} \right) \right) + \frac{2(p - s_{\text{sig}} + n - s_{\text{cor}})}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2 \log \left( \frac{p + n}{s_{\text{sig}} + s_{\text{cor}}} \right)}} \cdot \frac{s_{\text{sig}} + s_{\text{cor}}}{p + n}
\]
\[
\leq s_{\text{sig}} \left( 1 + 2 \log \left( \frac{p + n}{s_{\text{sig}} + s_{\text{cor}}} \right) \right) + \frac{1}{2} (s_{\text{sig}} + s_{\text{cor}})
\]
\[
\leq (s_{\text{sig}} + s_{\text{cor}}) \cdot \left( 2 \log \left( \frac{p + n}{s_{\text{sig}} + s_{\text{cor}}} \right) + \frac{3}{2} \right),
\]

since
\[
\frac{1 - (s_{\text{sig}} + s_{\text{cor}})/(p + n)}{\sqrt{\pi \log((p + n)/(s_{\text{sig}} + s_{\text{cor}}))}} \leq 1/2
\]
whenever \( 1 \leq s_{\text{sig}} + s_{\text{cor}} \leq p + n \). We also have the analogous bound for the expected Gaussian distance for the corruption, and combining the two, we get
\[
2\eta (t_{\text{sig}} \cdot \partial \|x^*\|_1) + \eta (t_{\text{cor}} \cdot \partial \|v^*\|_1) \leq 3 \left( s_{\text{sig}} + s_{\text{cor}} \right) \cdot \left( 2 \log \left( \frac{p + n}{s_{\text{sig}} + s_{\text{cor}}} \right) + \frac{3}{2} \right).
\]
The result then follows by applying Theorem 2.
B. Scaling for sparse signal and dense corruption

**Corollary 4.** Fix any corruption proportion \( \gamma < 1 \) and let \( s_\gamma \) be defined as in (19). If \( x^* \) and \( v^* \) have at most \( s_\gamma \) and \( n \cdot \gamma \) nonzero entries, respectively, and \( (x, \hat{v}) \) is a solution to

\[
\arg \min_{(x,v)} \{ \|x\|_1 + \lambda \|v\|_1 : \|y - (\Phi x + v)\|_2 \leq \delta \}
\]

with

\[
\lambda = \frac{1 - \gamma}{\sqrt{\pi \log(p/s_\gamma)}},
\]

then with probability at least 0.9, the recovery error satisfies

\[
\sqrt{\|\hat{x} - x^*\|_2^2 + \|\hat{v} - v^*\|_2^2} \leq \frac{2\delta}{\epsilon}.
\]

**Proof.** If \( s_\gamma = 0 \), then \( \lambda = 0 \), \( \hat{x} = x^* = 0 \), and \( \|y - (\Phi \hat{x} + \hat{v})\|_2 = \|v^* + z - \hat{v}\|_2 \leq \delta \). Hence, with probability 1,

\[
\sqrt{\|\hat{x} - x^*\|_2^2 + \|\hat{v} - v^*\|_2^2} \leq 2\delta \leq 2\delta/\epsilon.
\]

Otherwise, \( s_\gamma \geq 1 \). With the choice \( t_{\text{sig}} = \sqrt{2 \log(p/s_\gamma)} \), Corollary 2 and the prior sparse distance bound of Table II together imply that

\[
\sqrt{n} \geq \frac{2 \sqrt{2 s_\gamma \log(p/s_\gamma) + \frac{3}{2} s_\gamma} + 2 \log(1/p) + 3 \sqrt{2 \pi} + 1 + \frac{1}{\sqrt{2 \pi}}}{\alpha_1 (1 - \gamma)^2 - \epsilon}
\]

suffices to achieve the desired recovery guarantee. Since \( s_\gamma \geq 1 \), we have

\[
\sqrt{2 \log(10)} + 3 \sqrt{2 \pi} + 1 + \frac{1}{\sqrt{2 \pi}} \leq 11.1 \leq \frac{1}{10} \sqrt{2 s_\gamma \log(p/s_\gamma) + \frac{3}{2} s_\gamma}
\]

and hence

\[
\sqrt{n} \geq \frac{12 \sqrt{2 s_\gamma \log(p/s_\gamma) + \frac{3}{2} s_\gamma}}{\alpha_1 (1 - \gamma)^2 - \epsilon}
\]

is sufficient for recovery. One can check that the chosen \( s_\gamma \) always satisfies this bound.

\( \square \)

**Appendix C**

**Proofs of Main Results**

**A. Main idea**

Under the assumptions of Section [LA] we have an \( \ell_2 \) bound on the noise level of our problem,

\[
\|y - (\Phi x^* + v^*)\|_2 \leq \delta,
\]

and hence our feasible set is

\[
\{(x, v) \in \mathbb{R}^{p \times n} : \|y - (\Phi x + v)\|_2 \leq \delta \}.
\]

**Tangent cones:** We will refer to the signal and corruption tangent cones

\[
T_{\text{sig}} = \{a \in \mathbb{R}^p : \|x^* + t \cdot a\|_{\text{sig}} \leq \|x^*\|_{\text{sig}} \text{ for some } t > 0\}
\]

and

\[
T_{\text{cor}} = \{b \in \mathbb{R}^n : \|v^* + t \cdot b\|_{\text{cor}} \leq \|v^*\|_{\text{cor}} \text{ for some } t > 0\}.
\]

Given a penalty parameter \( \lambda \in (0, \infty) \), we write a joint penalty function

\[
\text{Pen}_\lambda(x, v) = \|x\|_{\text{sig}} + \lambda \|v\|_{\text{cor}}
\]

and define the *joint tangent cone* given by

\[
T^\lambda_{\text{joint}} = \{(a, b) \in \mathbb{R}^p \times \mathbb{R}^n : \text{Pen}_\lambda(x^* + t \cdot a, v^* + t \cdot b) \leq \text{Pen}_\lambda(x^*, v^*) \text{ for some } t > 0\}.
\]

By definition,

\[
T_{\text{sig}} \times T_{\text{cor}} \subseteq T^\lambda_{\text{joint}} \subseteq (T_{\text{sig}} \times \mathbb{R}^n) \cup (\mathbb{R}^p \times T_{\text{cor}}).
\]
Constrained problem: Consider either version of the constrained estimation problem:

\[ (\hat{x}, \hat{v}) = \arg \min \left\{ \|x\|_{\text{sig}} : \|v\|_{\text{cor}} \leq \|v^*\|_{\text{cor}}, \|y - (\Phi x + v)\|_2 \leq \delta \right\} \]

or

\[ (\hat{x}, \hat{v}) = \arg \min \left\{ \|v\|_{\text{cor}} : \|x\|_{\text{sig}} \leq \|x^*\|_{\text{sig}}, \|y - (\Phi x + v)\|_2 \leq \delta \right\} \]

In both optimization problems, our solution \((\hat{x}, \hat{v})\) will necessarily satisfy

\[ \|\hat{x}\|_{\text{sig}} \leq \|x^*\|_{\text{sig}} \quad \text{and} \quad \|\hat{v}\|_{\text{cor}} \leq \|v^*\|_{\text{cor}}. \]

We know therefore that

\[ (\hat{x} - x^*, \hat{v} - v^*) \in T_{\text{sig}} \times T_{\text{cor}}. \]

Below, we will derive a high probability lower bound on

\[
\min_{(a,b) \in (T_{\text{sig}} \times T_{\text{cor}}) \setminus \{(0,0)\}} \frac{\|\Phi a + b\|_2}{\sqrt{\|a\|_2^2 + \|b\|_2^2}},
\]

or, equivalently after rescaling,

\[
\min_{(a,b) \in (T_{\text{sig}} \times T_{\text{cor}}) \cap \mathbb{R}^{p+n-1}} \|\Phi a + b\|_2.
\]

Given this bound, since we know that

\[ \|\Phi (\hat{x} - x^*) + (\hat{v} - v^*)\|_2 \leq \|y - (\Phi \hat{x} + \hat{v})\|_2 + \|y - (\Phi x^* + v^*)\|_2 \leq 2\delta, \]

we obtain the estimation error bound

\[ \sqrt{\|\hat{x} - x\|_2^2 + \|\hat{v} - v\|_2^2} \leq 2\delta \cdot \left( \min_{(a,b) \in (T_{\text{sig}} \times T_{\text{cor}}) \cap \mathbb{R}^{p+n-1}} \|\Phi a + b\|_2 \right)^{-1}. \]

Penalized problem: Consider the penalized optimization problem

\[ (\hat{x}, \hat{v}) = \arg \min \{ \text{Pen}_\lambda(x, v) : \|y - (\Phi x + v)\|_2 \leq \delta \}. \]

Our solution \((\hat{x}, \hat{v})\) will satisfy the weaker condition

\[ \text{Pen}_\lambda(\hat{x}, \hat{v}) \leq \text{Pen}_\lambda(x^*, v^*), \]

and so we will have the weaker inclusion

\[ (\hat{x} - x^*, \hat{v} - v^*) \in T_{\text{joint}}^\lambda. \]

Following the same reasoning as in the constrained case, we obtain the estimation error bound

\[ \sqrt{\|\hat{x} - x\|_2^2 + \|\hat{v} - v\|_2^2} \leq 2\delta \cdot \left( \min_{(a,b) \in T_{\text{joint}}^\lambda \cap \mathbb{R}^{p+n-1}} \|\Phi a + b\|_2 \right)^{-1}. \]

B. Preliminaries

In both the constrained and the penalized setting, we begin by relating \(\min_{(a,b) \in \Omega} \|\Phi a + b\|_2\) to a second Gaussian functional via the following lemma (proved in Appendix D):

**Lemma 2.** Take any \(\Omega \subset \mathbb{R}^p \times \mathbb{R}^n\). Let \(\Phi \in \mathbb{R}^{n \times p}\) have i.i.d. \(\mathcal{N}(0, \frac{1}{n})\) entries, and let \(g \in \mathbb{R}^p\) and \(h \in \mathbb{R}^n\) have i.i.d. \(\mathcal{N}(0,1)\) entries. Then

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

Next, consider the matrix \(\sqrt{n} \cdot \Phi\), which has i.i.d. standard Gaussian entries. We know that \(\sqrt{n} \cdot \min_{(a,b) \in \Omega} \|\Phi a + b\|_2\) is a 1-Lipschitz function of the matrix \(\sqrt{n} \cdot \Phi\), because of the assumption that \(\|a\|_2 \leq 1\) for all \(a \in \Omega\). Therefore, applying a bound of Ledoux [23], we see that with probability at least \(1 - \beta\),

\[
\sqrt{n} \cdot \min_{(a,b) \in \Omega} \|\Phi a + b\|_2 \geq \sqrt{n} \cdot \mathbb{E} \left[ \min_{(a,b) \in \Omega} \|\Phi a + b\|_2 \right] - 2\log(\frac{1}{\beta}).
\]
for any $\beta \in (0, 1]$.

These two bounds form the main ingredients of our proofs. In the next two sections, we will lower-bound the expectation in the right-hand side of (28) for $\Omega = (T_{\text{sig}} \times T_{\text{cor}}) \cap \mathbb{S}^{p+n-1}$ and for $\Omega = T_{\text{joint}} \cap \mathbb{S}^{p+n-1}$, to handle the constrained and penalized problems, respectively. Combining these results with the probability bound in (29), our main results are obtained.

C. Lower bound: constrained setting

In this section we derive a lower bound on

$$\min_{(a,b) \in (T_{\text{sig}} \times T_{\text{cor}}) \cap \mathbb{S}^{p+n-1}} \| \Phi a + b \|_2.$$ 

Let

$$\omega = \omega \left( T_{\text{sig}} \cap \mathbb{S}^{p-1} \right) \quad \text{and} \quad \omega' = \omega \left( T_{\text{cor}} \cap \mathbb{S}^{n-1} \right)$$

be shorthand for the two relevant Gaussian widths.

Theorem 3. Then

$$\sqrt{n} \mathbb{E} \left[ \min_{(a,b) \in (T_{\text{sig}} \times T_{\text{cor}}) \cap \mathbb{S}^{p+n-1}} \| \Phi a + b \|_2 \right] \geq \mu_n - \sqrt{\omega^2 + \omega'^2 + 8 - \frac{1}{\sqrt{2}}} - \frac{1}{\sqrt{2\pi}}.$$

Proof. We will combine Lemma 2 with a lower bound for

$$\mathbb{E} \left[ \min_{(a,b) \in (T_{\text{sig}} \times T_{\text{cor}}) \cap \mathbb{S}^{p+n-1}} \| h \cdot \| a \|_2 + \sqrt{n} \cdot b \|_2 + \langle g, a \rangle \right].$$

Let

$$\hat{\omega} = \max_{a \in T_{\text{sig}} \cap \mathbb{S}^{p-1}} \langle -g, a \rangle \quad \text{and} \quad \hat{\omega}' = \max_{b \in T_{\text{cor}} \cap \mathbb{S}^{n-1}} \langle -h, b \rangle$$

be the ‘observed’ Gaussian widths, with expectations $\omega$ and $\omega'$, where $g \in \mathbb{R}^p$ and $h \in \mathbb{R}^n$ are independent vectors with i.i.d. standard normal entries.

If $\sqrt{\hat{\omega}^2 + \hat{\omega}'^2} > \| h \|_2$, then

$$\left( \min_{(a,b) \in (T_{\text{sig}} \times T_{\text{cor}}) \cap \mathbb{S}^{p+n-1}} \| h \cdot \| a \|_2 + \sqrt{n} \cdot b \|_2 + \langle g, a \rangle \right) \geq 0 \geq \| h \|_2 - \sqrt{\hat{\omega}^2 + \hat{\omega}'^2}.$$

If not, then by definition of $(T_{\text{sig}} \times T_{\text{cor}}) \cap \mathbb{S}^{p+n-1},$

$$\min_{(a,b) \in (T_{\text{sig}} \times T_{\text{cor}}) \cap \mathbb{S}^{p+n-1}} \| h \cdot \| a \|_2 + \sqrt{n} \cdot b \|_2 + \langle g, a \rangle$$

$$\geq \min_{c \in [0,1]: a \in T_{\text{sig}} \cap \mathbb{S}^{p-1}; b \in T_{\text{cor}} \cap \mathbb{S}^{n-1}} \| h \cdot c + b \cdot \sqrt{n} \cdot \sqrt{1 - c^2} \|_2 + c \cdot \langle g, a \rangle$$

$$\geq \min_{c \in [0,1]: a \in T_{\text{sig}} \cap \mathbb{S}^{p-1}; b \in T_{\text{cor}} \cap \mathbb{S}^{n-1}} \| h \cdot c + b \cdot \| h \|_2 \sqrt{1 - c^2} \|_2 + c \cdot \langle g, a \rangle - \sqrt{n} - \| h \|_2$$

$$\geq \min_{c \in [0,1]: a \in T_{\text{sig}} \cap \mathbb{S}^{p-1}; b \in T_{\text{cor}} \cap \mathbb{S}^{n-1}} \sqrt{\| h \|_2^2 \cdot c^2 + \| h \|_2^2 \cdot (1 - c^2) + 2 \| h \|_2 \cdot \langle h, b \rangle \cdot c \sqrt{1 - c^2} + c \cdot \langle g, a \rangle - \sqrt{n} - \| h \|_2}$$

Minimizing over $a$ and $b$ for a fixed $c$,

$$= \min_{c \in [0,1]} \sqrt{\| h \|_2^2 - 2 \| h \|_2 \cdot \hat{\omega}' \cdot c \sqrt{1 - c^2} - c \cdot \hat{\omega} - \sqrt{n} - \| h \|_2}$$

Since $\sqrt{\hat{\omega}^2 + \hat{\omega}'^2} \leq \| h \|_2$, applying Lemma 2

$$\geq \| h \|_2 - \sqrt{\hat{\omega}^2 + \hat{\omega}'^2} - \sqrt{n} - \| h \|_2.$$

In either case, then, we have proved that

$$\left( \min_{(a,b) \in (T_{\text{sig}} \times T_{\text{cor}}) \cap \mathbb{S}^{p+n-1}} \| h \cdot \| a \|_2 + \sqrt{n} \cdot b \|_2 + \langle g, a \rangle \right) \geq \| h \|_2 - \sqrt{\hat{\omega}^2 + \hat{\omega}'^2} - \sqrt{n} - \| h \|_2.$$
Finally, we take expectations: we have $\mathbb{E} \left[ \sqrt{n} - \|h\|_2 \right] \leq \frac{1}{\sqrt{2}}$ by Lemma 5 and by Lemma 8
\[
\mathbb{E} \left[ \sqrt{\omega^2 + \omega'^2} \right] \leq \sqrt{\mathbb{E} \left[ \omega^2 \right] + \mathbb{E} \left[ \omega'^2 \right]} \leq \sqrt{\omega^2 + \omega'^2 + 8}.
\]
Applying Lemma 2 this gives us the desired lower bound. \qed

D. Lower bound: penalized setting

In this section we derive a lower bound on
\[
\min_{(a,b) \in \mathcal{F}_\text{joint} \cap S_{n-1}} \|\Phi a + b\|_2.
\]

**Theorem 4.** Let $\lambda = t_{\text{cor}}/t_{\text{sig}}$ for parameters $t_{\text{sig}}, t_{\text{cor}} \geq 0$. Then
\[
\sqrt{n} \mathbb{E} \left[ \min_{(a,b) \in \mathcal{F}_\text{joint} \cap S_{n-1}} \|h \cdot \|a\|_2 + \sqrt{n} \cdot \|b\|_2 + \langle g, a \rangle \right] +
\]

For $\sigma = +1$ and $\sigma = -1$, define
\[
\begin{align*}
\sigma_{\text{sig}} \colon= \text{dist} \left( \sigma \cdot g, t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}} \right) \quad \text{and} \quad \sigma_{\text{cor}} \colon= \text{dist} \left( \sigma \cdot h, t_{\text{cor}} \cdot \partial \|v^*\|_{\text{cor}} \right),
\end{align*}
\]
and let $u_{\text{sig}}(\sigma) \in \partial \|x^*\|_{\text{sig}}$ and $u_{\text{cor}}(\sigma) \in \partial \|v^*\|_{\text{cor}}$ be such that
\[
\begin{align*}
\| \sigma \cdot g - t_{\text{sig}} \cdot u_{\text{sig}}(\sigma) \|_2 = d(\sigma), \quad \text{and} \quad \| \sigma \cdot h - t_{\text{cor}} \cdot u_{\text{cor}}(\sigma) \|_2 = d(\sigma).
\end{align*}
\]
We know that for each choice of signs $\sigma_1, \sigma_2 \in \{\pm 1\}$,
\[
0 \geq \left( \|x^* + a\|_{\text{sig}} + \lambda \|v^* + b\|_{\text{cor}} \right) - \left( \|x^*\|_{\text{sig}} + \lambda \|v^*\|_{\text{cor}} \right)
\]
\[
\begin{align*}
\geq \langle u_{\text{sig}}(\sigma_1), a \rangle + \lambda \langle u_{\text{cor}}(\sigma_2), b \rangle
\end{align*}
\]
\[
\begin{align*}
= \frac{\langle \sigma_1 \cdot g, a \rangle - \langle \sigma_1 \cdot h - t_{\text{sig}} \cdot u_{\text{sig}}(\sigma_1), a \rangle}{t_{\text{sig}}} + \lambda \frac{\langle \sigma_2 \cdot h, b \rangle - \langle \sigma_2 \cdot h - t_{\text{cor}} \cdot u_{\text{cor}}(\sigma_2), b \rangle}{t_{\text{cor}}}
\end{align*}
\]
\[
\begin{align*}
\geq \langle \sigma_1 \cdot g, a \rangle - d(\sigma_1) \|a\|_2 + \langle \sigma_2 \cdot h, b \rangle - d(\sigma_2) \|b\|_2,
\end{align*}
\]
and so
\[
\begin{align*}
d(\sigma_1) \|a\|_2 + d(\sigma_2) \|b\|_2 \geq \langle \sigma_1 \cdot g, a \rangle + \langle \sigma_2 \cdot h, b \rangle.
\end{align*}
\]
Maximizing the right-hand side over the signs,
\[
d_{\text{sig}} \|a\|_2 + d_{\text{cor}} \|b\|_2 \geq |\langle g, a \rangle| + |\langle h, b \rangle| \geq (|\langle g, a \rangle| + |\langle h, b \rangle|) + (\langle g, a \rangle + \langle h, b \rangle)\geq 
\]
where
\[
d_{\text{sig}} = \max \{d_{\text{sig}}^{(1)}, d_{\text{sig}}^{(-1)}\} \quad \text{and} \quad d_{\text{cor}} = \max \{d_{\text{cor}}^{(1)}, d_{\text{cor}}^{(-1)}\}.
\]
Next, since \( \|b\|_2 \leq 1 \) for all \( (a, b) \in T_{joint}^\lambda \cap S^{p+n-1} \), we have
\[
\min_{(a, b) \in T_{joint}^\lambda \cap S^{p+n-1}} \|h \cdot \|a\|_2 + \sqrt{n} \cdot b\|_2 + \langle g, a \rangle \\
\geq \min_{(a, b) \in T_{joint}^\lambda \cap S^{p+n-1}} \|h \cdot \|a\|_2 + \|h\|_2 \cdot b\|_2 + \langle g, a \rangle - |\sqrt{n} - \|h\|_2| \\
= \min_{(a, b) \in T_{joint}^\lambda \cap S^{p+n-1}} \sqrt{\|h\|_2^2 \left(\|a\|_2^2 + \|b\|_2^2\right) + 2 \|h\|_2 \|a\|_2 \cdot \langle h, b \rangle + \langle g, a \rangle - |\sqrt{n} - \|h\|_2|} \\
= \min_{(a, b) \in T_{joint}^\lambda \cap S^{p+n-1}} \sqrt{1 + 2 \|a\|_2 \cdot \left(\frac{h}{\|h\|_2}, b\right) + \langle g, a \rangle - |\sqrt{n} - \|h\|_2|} \\
\geq \min_{(a, b) \in T_{joint}^\lambda \cap S^{p+n-1}} \sqrt{1 - 2 \|a\|_2 \cdot \left(\frac{h}{\|h\|_2}, b\right) - \langle g, a \rangle - |\sqrt{n} - \|h\|_2|} \\
\geq \min_{(a, b) \in T_{joint}^\lambda \cap S^{p+n-1}} \|h\|_2 - \max\{2 \|a\|_2, 1\} \cdot (d_{sig} \|a\|_2 + d_{cor} \|b\|_2 - (-\langle g, a \rangle)_+) - (-\langle g, a \rangle)_+ - |\sqrt{n} - \|h\|_2| \\
\geq \min_{(a, b) \in T_{joint}^\lambda \cap S^{p+n-1}} \|h\|_2 - 2d_{sig} - d_{cor} - |\sqrt{n} - \|h\|_2| \\
\geq \min_{(a, b) \in T_{joint}^\lambda \cap S^{p+n-1}} \|h\|_2 - 2d_{sig} - d_{cor} - |\sqrt{n} - \|h\|_2| \\
\geq \mu_n - 2\eta \left( t_{sig} \cdot \partial \|x^\ast\|_{sig} \right) - \eta \left( t_{cor} \cdot \partial \|v^\ast\|_{cor} \right) - 3\sqrt{2\pi} - \frac{1}{\sqrt{2}}.
\]

\[
\text{APPENDIX D}
\]

\[
\text{LEMMA 7}
\]

For any norm \( \|\cdot\| \) on \( \mathbb{R}^p \) with dual norm \( \|\cdot\|^\ast \),
\[
\max_{x \in \mathbb{R}^p} \frac{\|x\|}{\|x\|_2} = \max_{x \in \mathbb{R}^p} \frac{\|x\|^\ast}{\|x\|_2^\ast}.
\]

Proof. First, for any \( x \),
\[
\|x\|^2 = \langle x, x \rangle \leq \|x\| \cdot \|x\|^\ast \Rightarrow \frac{\|x\|}{\|x\|_2} \geq \frac{\|x\|^\ast}{\|x\|_2^\ast}.
\]

This proves that the left-hand side is greater than or equal to the right-hand side, in the claim. Now we prove the reverse inequality. Choose any \( x \in \mathbb{R}^p \). Then \( \|x\| = \max_{\|w\|^\ast = 1} \langle x, w \rangle \). Choose some \( w \) with \( \|w\|^\ast = 1 \) so that this maximum is attained. Then \( \langle x, w \rangle \leq \|x\| \cdot \|w\|_2 \), and so
\[
\frac{\|x\|}{\|x\|_2} \geq \frac{\langle x, w \rangle}{\|x\|_2} \leq \|w\|_2 \leq \max_{x' \in \mathbb{R}^p} \frac{\|x'\|^2}{\|x'\|^\ast}.
\]
Since this is true for any $x$, this proves the desired bound.

**Lemma 3.** Suppose that $\vartheta \|x\|$ satisfies (26). Let $t_g$ be defined as in (27). Then $g \mapsto t_g$ is a $\frac{1}{\|w_0\|^2}$-Lipschitz function of $g$. 

**Proof.** Consider any $g$ and $g'$. Find $w, w' \in \vartheta \|x\|$ such that

$$
\text{dist}(g, t_g \cdot \vartheta \|x\|) = \|g - t_g \cdot w\|_2 \quad \text{and} \quad \text{dist}(g', t_g' \cdot \vartheta \|x\|) = \|g' - t_g' \cdot w'\|_2.
$$

Recalling that

$$
N = \bigcup_{t \geq 0} (t \cdot \vartheta \|x\|),
$$

we see that $t_g \cdot w$ and $t_g' \cdot w'$ are the projections of $g$ and of $g'$, respectively, onto the cone $N$. Since $N$ is convex and projection onto a convex set is nonexpansive, it follows that

$$
\|t_g \cdot w - t_g' \cdot w'\|_2 \leq \|g - g'\|_2.
$$

Now we use the assumption (26): we have

$$
\|t_g \cdot w - t_g' \cdot w'\|_2 = \|(t_g - t_g') \cdot w_0 + [t_g \cdot (w - w_0) + t_g' \cdot (w' - w_0)]\|_2 \\
\geq |t_g - t_g'| \cdot \|w_0\|_2,
$$

because $(w - w_0, w_0) = (w' - w_0, w_0) = 0$. 

**Lemma 4.** Let $\phi : \mathbb{R}^p \rightarrow \mathbb{R}$ be a 1-Lipschitz function. For any $a \in \mathbb{R}$ and $p_0 > 0$,

$$
\mathbb{P} \{ \phi(g) < a \} \geq p_0 \quad \Rightarrow \quad \mathbb{E} [\phi(g)] \leq a + \sqrt{2 \log(\frac{1}{p_0})}.
$$

**Proof.** Choose any $\varepsilon$ satisfying $0 < \varepsilon < p_0$. By Ledoux (35, (2.8)), we know that

$$
\mathbb{P} \left\{ \phi(g) < \mathbb{E} [\phi(g)] - \sqrt{2 \log(\frac{1}{p_0 - \varepsilon})} \right\} \leq p_0 - \varepsilon,
$$

which means that we must have $\mathbb{E} [\phi(g)] - \sqrt{2 \log(\frac{1}{p_0 - \varepsilon})} < a$ to avoid contradiction. Since this is true for arbitrary $\varepsilon \in (0, p_0)$, the claim follows.

**Lemma 2.** Take any $\Omega \subset \mathbb{E}^p \times \mathbb{R}^n$. Let $\Phi \in \mathbb{R}^{n \times p}$ have i.i.d. $N(0, \frac{1}{n})$ entries, and let $g \in \mathbb{R}^p$ and $h \in \mathbb{R}^n$ have i.i.d. $N(0, 1)$ entries. Then

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

**Proof.** For $(a,b) \in \Omega$ and $w \in \mathbb{B}^n$, let

$$
X_{(a,b),w} = \sqrt{n} \cdot \langle \Phi a, w \rangle + \nu \|a\|_2 \quad \text{and} \quad Y_{(a,b),w} = \langle h, w \rangle \|a\|_2 + \langle g, a \rangle,
$$

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

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

and

$$
\mathbb{E} \left[ X_{(a,b),w} X_{(a',b'),w'} \right] - \mathbb{E} \left[ Y_{(a,b),w} Y_{(a',b'),w'} \right] = [(a, a') \langle w, w' \rangle + \|a\|_2 \|a'\|_2] - \|a\|_2 \|a'\|_2 \langle w, w' \rangle \geq 0,
$$

with equality if $(a,b) = (a',b')$. Therefore, applying Theorem 1.1 of Gordon (36) to these Gaussian processes, we know that for any scalars $\{c_{(a,b),w}\}$,

$$
\mathbb{P} \left\{ \bigcap_{(a,b) \in \Omega} \bigcup_{w \in \mathbb{B}^n} \left[ X_{(a,b),w} \geq c_{(a,b),w} \right] \right\} \geq \mathbb{P} \left\{ \bigcap_{(a,b) \in \Omega} \bigcup_{w \in \mathbb{B}^n} \left[ Y_{(a,b),w} \geq c_{(a,b),w} \right] \right\}.
$$
Now fix any \( C \in \mathbb{R} \) and let \( c_{(a,b),w} = C - \sqrt{n} \cdot \langle b, w \rangle \). Then we can simplify these events:

\[
\bigcap_{(a,b) \in \Omega} \bigcup_{w \in \mathbb{B}^n} X_{(a,b),w} \geq c_{(a,b),w} = \min_{(a,b) \in \Omega} \max_{w \in \mathbb{B}^n} \left[ X_{(a,b),w} + \sqrt{n} \cdot \langle b, w \rangle \geq C \right]
\]

and same for the \( Y \) process. This proves that

\[
P \left\{ \min_{(a,b) \in \Omega} \max_{w \in \mathbb{B}^n} X_{(a,b),w} + \sqrt{n} \cdot \langle b, w \rangle \geq C \right\} \geq P \left\{ \min_{(a,b) \in \Omega} \max_{w \in \mathbb{B}^n} Y_{(a,b),w} + \sqrt{n} \cdot \langle b, w \rangle \geq C \right\},
\]

or in other words, by maximizing over \( w \) on each side,

\[
P \left\{ \sqrt{n} \cdot \min_{(a,b) \in \Omega} \| \Phi a + b \|_2 + \nu \cdot \| a \|_2 \geq C \right\} \geq P \left\{ \min_{(a,b) \in \Omega} \| h \cdot \| a \|_2 + \sqrt{n} \cdot \| b \|_2 + \langle g, a \rangle \geq C \right\}.
\]

Since this is true for all \( C \in \mathbb{R} \), we can integrate over \( C \in (0, \infty) \) to obtain

\[
E \left[ \left( \min_{(a,b) \in \Omega} \sqrt{n} \cdot \| \Phi a + b \|_2 + \nu \cdot \| a \|_2 \right) \right] \geq E \left[ \left( \min_{(a,b) \in \Omega} \| h \cdot \| a \|_2 + \sqrt{n} \cdot \| b \|_2 + \langle g, a \rangle \right) \right].
\]

Finally, since \( \| a \|_2 \in [0, 1] \) for all \( (a, b) \in \Omega \), we see that

\[
\left( \min_{(a,b) \in \Omega} \sqrt{n} \cdot \| \Phi a + b \|_2 + \nu \cdot \| a \|_2 \right) \leq \sqrt{n} \cdot \min_{(a,b) \in \Omega} \| \Phi a + b \|_2 + (\nu)_+,
\]

and therefore, since \( E[ (\nu)_+ ] = \frac{1}{\sqrt{2\pi}} \),

\[
\sqrt{n} \cdot E \left[ \min_{(a,b) \in \Omega} \| \Phi a + b \|_2 \right] \geq E \left[ \min_{(a,b) \in \Omega} \left( \sqrt{n} \cdot \| \Phi a + b \|_2 + \nu \cdot \| a \|_2 \right) \right] - \frac{1}{\sqrt{2\pi}}.
\]

\[\square\]

**Lemma 5.** For \( h \sim N(0, I_n) \),

\[
E[ \| h \|_2 - \mu_n ] \leq \frac{1}{\sqrt{2}}.
\]

**Proof.** We use the fact that \( \mu_n \geq \sqrt{n - 1/2} \) for all \( n \). We have

\[
E[ \| h \|_2 - \mu_n ]^2 \leq E[ (\| h \|_2 - \mu_n )^2 ] = E[ \| h \|_2^2 ] - 2\mu_n \cdot E[ \| h \|_2 ] + \mu_n^2
\]

\[
= n - \mu_n^2 \leq n - (\sqrt{n - 1/2})^2 = \frac{1}{2}.
\]

\[\square\]

**Lemma 6.** Let \( T \) be any nonempty convex cone in \( \mathbb{R}^p \). Then

\[
\omega \left( T \cap \mathbb{S}^{p-1} \right) \leq \omega \left( \{0\} \cup (T \cap \mathbb{S}^{p-1}) \right) = \omega \left( T \cap \mathbb{B}^p \right) \leq \omega \left( T \cap \mathbb{S}^{p-1} \right) + \frac{1}{\sqrt{2\pi}}.
\]

**Proof.** The first inequality is trivial. The equality is proved because \( \{0\} \cup (T \cap \mathbb{S}^{p-1}) \) contains all the extreme points of the convex set \( T \cap \mathbb{B}^p \). Now we prove the second inequality. Take any fixed \( u \in T \cap \mathbb{S}^{p-1} \). Then \( \langle -g, u \rangle \sim N(0, 1) \), and

\[
\omega \left( \{0\} \cup (T \cap \mathbb{S}^{p-1}) \right) - \omega \left( T \cap \mathbb{S}^{p-1} \right) = E \left[ \max_{w \in T \cap \mathbb{S}^{p-1}} \langle g, w \rangle \right] - \max_{w \in T \cap \mathbb{S}^{p-1}} \langle g, w \rangle
\]

\[
= E \left[ \min_{w \in T \cap \mathbb{S}^{p-1}} \langle -g, w \rangle \right] \leq E \left[ \langle -g, u \rangle \right] = \frac{1}{\sqrt{2\pi}}.
\]

\[\square\]
Lemma 7. For any $a, b, C \geq 0$ such that $a^2 + b^2 \leq C^2$, 
\[
\inf_{u \in [0, 1]} \sqrt{C^2 - 2Cu\sqrt{1 - u^2}} a - u \cdot b \geq C - \sqrt{a^2 + b^2}.
\]

Proof. By rescaling, we can assume without loss of generality that $C = 1$ and $a^2 + b^2 \leq 1$. First, suppose $a^2 + b^2 = 1$. Then
\[
(\sqrt{1 - u^2})^2 \leq 0
\]
\[
\Rightarrow u^2 \cdot a^2 - 2u \sqrt{1 - u^2} \cdot a + (1 - u^2) \geq 0
\]
\[
\Rightarrow 1 - 2u \sqrt{1 - u^2} \cdot a \geq u^2 \cdot (1 - a^2) = u^2 \cdot b^2
\]
\[
\Rightarrow \sqrt{1 - 2u \sqrt{1 - u^2} \cdot a} - u \cdot b \geq 0.
\]
This proves the claim whenever $a^2 + b^2 = 1$. Now suppose that $a^2 + b^2 = c^2$ for some $c \in [0, 1]$, and let $a' = \frac{a}{c}$ and $b' = \frac{b}{c}$. Then
\[
\sqrt{1 - 2u \sqrt{1 - u^2} \cdot a} - u \cdot b
\]
\[
= \sqrt{1 - 2u \sqrt{1 - u^2} \cdot a' \cdot c - u \cdot b' \cdot c}
\]
\[
= \sqrt{1 - 2u \sqrt{1 - u^2} \cdot a' \cdot c - c} \cdot \sqrt{1 - 2u \sqrt{1 - u^2} \cdot a'} + c \cdot \sqrt{1 - 2u \sqrt{1 - u^2} \cdot a' - u \cdot b'}
\]
From the work above, since $a^2 + b^2 = 1$,
\[
\geq \sqrt{1 - 2u \sqrt{1 - u^2} \cdot a' \cdot c - c} \cdot \sqrt{1 - 2u \sqrt{1 - u^2} \cdot a'}
\]
Writing $d = 2u \sqrt{1 - u^2} \cdot a' \in [0, 1]$,
\[
= \sqrt{1 - d \cdot c - c} \cdot \sqrt{1 - d}
\]
\[
\geq 1 - c,
\]
where the last step is true for all $c, d \in [0, 1]$ as follows:
\[
(1 - \sqrt{1 - d})^2 = 2 - 2 \sqrt{1 - d} - d
\]
\[
\Rightarrow c^2(1 - \sqrt{1 - d})^2 \leq c(2(1 - \sqrt{1 - d}) - d) \quad \text{since } c^2 \leq c
\]
\[
\Rightarrow 1 + c^2(1 - \sqrt{1 - d})^2 - 2c(1 - \sqrt{1 - d}) \leq 1 - dc \quad \text{by rearranging terms and adding 1 to both sides}
\]
\[
\Rightarrow 1 - c(1 - \sqrt{1 - d}) \leq \sqrt{1 - dc} \quad \text{by taking the square root}
\]
\[
\Rightarrow 1 - c \leq \sqrt{1 - dc - cv \sqrt{1 - d}}.
\]

Lemma 8. For any $\Theta \subset \mathbb{B}^p$, and for $g \in \mathbb{R}^p$ with i.i.d. standard normal entries,
\[
\mathbb{E} \left[ \max_{w \in \Theta} \langle g, w \rangle^2 \right] \leq \omega(\Theta)^2 + 4.
\]

Proof. First, observe that
\[
g \mapsto \max_{w \in \Theta} \langle g, w \rangle
\]

is a 1-Lipschitz function of $g$, because all $w \in \Theta$ have at most unit length. Therefore, since $\omega(\Theta) = \mathbb{E} \left[ \max_{w \in \Theta} \langle g, w \rangle \right]$, by Ledoux [35],
\[
\mathbb{P} \left\{ \max_{w \in \Theta} \langle g, w \rangle \geq \omega(\Theta) + c \right\} \leq e^{-c^2/2}, \quad \mathbb{P} \left\{ \max_{w \in \Theta} \langle g, w \rangle \leq \omega(\Theta) - c \right\} \leq e^{-c^2/2},
\]
for all \( c \geq 0 \). We then have
\[
\mathbb{E} \left[ \max_{u \in \Theta} \langle g, w \rangle^2 \right] = \omega(\Theta)^2 + \mathbb{E} \left[ \max_{u \in \Theta} \left( \frac{\max \langle g, w \rangle - \omega(\Theta)}{c} \right)^2 \right] \\
= \omega(\Theta)^2 + \int_0^\infty \mathbb{P} \left\{ \max_{u \in \Theta} \left( \frac{\max \langle g, w \rangle - \omega(\Theta)}{c} \right)^2 \geq c \right\} \, dc \\
\leq \omega(\Theta)^2 + \int_0^\infty e^{-c/2} \, dc \\
= \omega(\Theta)^2 + 4.
\]

\[ \square \]

**Lemma 9.** For any set \( A \subset \mathbb{R}^p \), for \( g \in \mathbb{R}^p \) with i.i.d. standard normal entries,
\[
\mathbb{E} \left[ \text{dist}(g, A)^2 \right] \leq \mathbb{E} \left[ \text{dist}(g, A)^2 \right] + 4.
\]

**Proof.** The function \( g \mapsto \text{dist}(g, A) \) is clearly a 1-Lipschitz function of \( g \). From this point the proof is the same as for Lemma 8. \[ \square \]

**Lemma 10.** For any set \( A \subset \mathbb{R}^p \), for \( g \in \mathbb{R}^p \) with i.i.d. standard normal entries,
\[
\mathbb{E} \left[ \max \{ \text{dist}(g, A), \text{dist}(-g, A) \} \right] \leq \eta(A) + \sqrt{2\pi}.
\]

**Proof.** Let \( \alpha = \eta(A) = \mathbb{E} [ \text{dist}(g, A) ] = \mathbb{E} [ \text{dist}(-g, A) ] \). Since \( g \mapsto \text{dist}(g, A) \) is 1-Lipschitz,
\[
\mathbb{E} \left[ (\text{dist}(g, A) - \alpha)_+ \right] = \int_0^\infty \mathbb{P} \{ \text{dist}(g, A) > \alpha + c \} \, dc = \int_0^\infty e^{-c^2/2} \, dc \leq \sqrt{\pi/2}.
\]

Then
\[
\mathbb{E} \left[ \max \{ \text{dist}(g, A), \text{dist}(-g, A) \} \right] = \mathbb{E} \left[ \max \{ \alpha + (\text{dist}(g, A) - \alpha), \alpha + (\text{dist}(-g, A) - \alpha) \} \right] \\
\leq \alpha + \mathbb{E} \left[ (\text{dist}(g, A) - \alpha)_+ \right] + \mathbb{E} \left[ (\text{dist}(-g, A) - \alpha)_+ \right] \\
\leq \alpha + \sqrt{2\pi}.
\]

\[ \square \]

**Lemma 11.** Let \( \Gamma \in \mathbb{R}^{m_1 \times m_2} \) have i.i.d. standard Gaussian entries, with \( m_1 \geq m_2 \). Then
\[
\mathbb{E} [ \| \Gamma \|_* ] \geq \frac{4}{27} m_2 \sqrt{m_1}.
\]

**Proof.** Let \( k = \left\lfloor \frac{4}{9} m_2 \right\rfloor \leq m_2 \). Let \( \Upsilon \) be a \( m_1 \times k \) submatrix of \( \Gamma \); then \( \Upsilon \) also has i.i.d. Gaussian entries, and \( \| \Gamma \|_* \geq \| \Upsilon \|_* \), so we only need to find a lower bound for \( \mathbb{E} [ \| \Upsilon \|_* ] \).

Let \( \sigma_1 \geq \cdots \geq \sigma_k \geq 0 \) be the singular values of \( h \). Davidson and Szarek [37, Thm. II.13] establishes that
\[
\mathbb{E} [ \sigma_k ] \geq \sqrt{m_1 - \sqrt{k}}.
\]

This means that \( \mathbb{E} [ \sigma_i ] \geq \sqrt{m_1 - \sqrt{k}} \) for each \( i = 1, \ldots, k \). In the simple case that \( \frac{4}{9} m_2 \) is an integer, we then have
\[
\mathbb{E} [ \sigma_1 + \cdots + \sigma_k ] \geq k \left( \sqrt{m_1 - \sqrt{k}} \right) = \frac{4}{9} m_2 \left( \sqrt{m_1 - \frac{2}{3} \sqrt{m_2}} \right) \geq \frac{4}{9} m_2 \left( \sqrt{m_1 - \frac{2}{3} \sqrt{m_1}} \right) = \frac{4}{27} m_2 \sqrt{m_1}.
\]

In the general case where \( \frac{4}{9} m_2 \) might not be an integer, we also use the fact that
\[
\mathbb{E} [ \sigma_1 ] \geq \mathbb{E} [ \| \Upsilon_{(1)} \|_2 ] = \mu_{m_1},
\]
where \( \Upsilon_{(1)} \) is the first column of \( \Upsilon \). Since \( \mu_{m_1} > \sqrt{m_1 - 1/2} \), which is larger than our previous lower bound \( \sqrt{m_1 - \sqrt{k}} \) on \( \mathbb{E} [ \sigma_1 ] \) used above, this gives us the slack we need in order to handle the slight discrepancy between \( k \) and \( \frac{4}{9} m_2 \). We omit the details. \[ \square \]
ACKNOWLEDGMENTS

The authors thank Emmanuel Candès for helpful suggestions on the presentation of this work. R.F. was supported by NSF grant DMS-1203762.

REFERENCES

[1] X. Li, “Compressed sensing and matrix completion with constant proportion of corruptions,” 2011, available at arXiv:1104.1041.

[2] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, “Rank-sparsity incoherence for matrix decomposition,” SIAM Journal on Optimization, vol. 21, no. 2, pp. 572–596, 2011.

[3] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” Journal of the ACM, vol. 58, no. 3, pp. 1–37, 2011.

[4] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 31, no. 2, pp. 210–227, 2009.

[5] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak, “Compressed sensing for networked data,” IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 92–101, 2008.

[6] E. Candès and T. Tao, “Decoding by linear programming,” IEEE Transactions on Information Theory, vol. 51, no. 12, pp. 4203–4215, 2005.

[7] D. Donoho, “Compressed sensing,” IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, 2006.

[8] T. Cai, G. Xu, and J. Zhang, “On recovery of sparse signals via \( \ell_1 \) minimization,” IEEE Transactions on Information Theory, vol. 55, no. 7, pp. 3388–3397, 2009.

[9] M. Fazel, “Matrix rank minimization with applications,” Ph.D. dissertation, Stanford University, 2002.

[10] M. Fazel, H. Hindi, and S. Boyd, “A rank minimization heuristic with application to minimum order system approximation,” in Proceedings of the 2001 American Control Conference, vol. 6. IEEE, 2002, pp. 4734–4739.

[11] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” Foundations of Computational Mathematics, vol. 9, no. 6, pp. 717–722, 2009.

[12] E. Candès and T. Tao, “The power of convex relaxation: Near-optimal matrix completion,” IEEE Transactions on Information Theory, vol. 56, no. 5, pp. 2053–2080, 2010.

[13] B. Recht, “A simpler approach to matrix completion,” Journal of Machine Learning Research, vol. 12, pp. 3413–3430, 2011.

[14] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, “The convex geometry of linear inverse problems,” Dec. 2010, available at arXiv:1012.0621.

[15] M. B. McCoy and J. A. Tropp, “Sharp recovery bounds for convex deconvolution, with applications,” 2012, available at arXiv:1205.1580.

[16] J. T. Chu, “A modified wallis product and some applications,” The American Mathematical Monthly, vol. 69, no. 5, pp. 402–404, 1962.

[17] J. Penot and C. Zalinescu, “Elements of quasiconvex subdifferential calculus,” Journal of Convex Analysis, vol. 7, no. 2, pp. 243–270, 2000.

[18] S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu, “A unified framework for high-dimensional analysis of \( m \)-estimators with decomposable regularizers,” Statistical Science, vol. 27, no. 4, pp. 538–557, 2012.

[19] A. Wyner, “An analog scrambling scheme which does not expand bandwidth, Part I: Discrete time,” IEEE Transactions on Information Theory, vol. 25, no. 3, pp. 261 – 274, may 1979.

[20] ———, “An analog scrambling scheme which does not expand bandwidth, Part II: Continuous time,” IEEE Transactions on Information Theory, vol. 25, no. 4, pp. 415 – 425, jul 1979.

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

[22] Y. Eldar, P. Kuppinger, and H. Bölcskei, “Block-sparse signals: Uncertainty relations and efficient recovery,” IEEE Transactions on Signal Processing, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[23] F. Parvaresh and B. Hassibi, “Explicit measurements with almost optimal thresholds for compressed sensing,” in ICASSP. IEEE, 2008, pp. 3853–3856.

[24] E. J. Candès and B. Recht, “Simple bounds for low-complexity model reconstruction,” Mathematical Programming Series A, 2012.
[25] J. N. Laska, M. A. Davenport, and R. G. Baraniuk, “Exact signal recovery from sparsely corrupted measurements through the pursuit of justice,” in Proceedings of the 43rd Asilomar conference on Signals, systems and computers, ser. Asilomar’09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 1556–1560.

[26] J. Wright and Y. Ma, “Dense error correction via $\ell_1$-minimization,” IEEE Transactions on Information Theory, vol. 56, no. 7, pp. 3540–3560, 2010.

[27] N. Nguyen and T. Tran, “Exact recoverability from dense corrupted observations via $\ell_1$ minimization,” IEEE Transactions on Information Theory, vol. 59, no. 4, pp. 2017–2035, 2013.

[28] G. Pope, A. Bracher, and C. Studer, “Probabilistic recovery guarantees for sparsely corrupted signals,” 2012, available at arXiv:1203.6001

[29] C. Hegde and R. G. Baraniuk, “Signal recovery on incoherent manifolds,” IEEE Transactions on Information Theory, vol. 58, no. 12, pp. 7204–7214, 2012.

[30] CVX Research Inc., “CVX: Matlab software for disciplined convex programming, version 1.22,” http://cvxr.com/cvx, Jan. 2013.

[31] M. Grant and S. Boyd, “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.

[32] J. Friedman, T. Hastie, and R. Tibshirani, “A note on the group lasso and a sparse group lasso,” 2010, available at arXiv:1001.0736

[33] B. Laurent and P. Massart, “Adaptive estimation of a quadratic functional by model selection,” The Annals of Statistics, vol. 28, no. 5, pp. 1302–1338, 2000.

[34] G. Watson, “Characterization of the subdifferential of some matrix norms,” Linear Algebra and its Applications, vol. 170, pp. 33–45, 1992.

[35] M. Ledoux, “Isoperimetry and Gaussian analysis,” Lectures on probability theory and statistics, pp. 165–294, 1996.

[36] Y. Gordon, “Some inequalities for Gaussian processes and applications,” Israel Journal of Mathematics, vol. 50, no. 4, pp. 265–289, 1985.

[37] K. Davidson and S. Szarek, “Local operator theory, random matrices and Banach spaces,” in Handbook on the Geometry of Banach spaces, Lindenstrauss, Ed. Elsevier Science, 2001, vol. 1, pp. 317–366.