

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 complexity 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 interpretable bounds for the Gaussian complexity 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, ℓ_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 ℓ_1 norm penalization [6, 7, 8] and the recovery of low-rank matrices via trace norm penalization [9, 10, 11, 12, 13].

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 Φ 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 δ on the noise level $\|z\|_2$. In the noiseless setting where $\delta = 0$, the authors' geometric analysis shows that

$$n = \omega^2 (T_{\text{sig}} \cap \mathbb{B}^p) + 1 \quad (1)$$

measurements suffice to recover x^* exactly. Here, T_{sig} is a convex cone in \mathbb{R}^p induced by x^* and $\|\cdot\|_{\text{sig}}$, and $\omega^2 (T_{\text{sig}} \cap \mathbb{B}^p)$ is a specific measure of the size of this cone, defined in Section II. In the noisy setting where $\delta > 0$, the same analysis shows that $\mathcal{O}(\omega^2 (T_{\text{sig}} \cap \mathbb{B}^p))$ 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 $(n \cdot \gamma)$ -sparse, so that a fraction γ of our linear measurements are arbitrarily corrupted, then our analysis guarantees exact recovery as soon as n exceeds

$$\frac{\omega^2 (T_{\text{sig}} \cap \mathbb{B}^p)}{2/\pi \cdot (1 - \gamma)^2}$$

plus an explicit smaller-order term. This provides a close parallel with Chandrasekaran et al. [14]’s result (1) for the corruption-free setting and gives an explicit scaling in terms of the corruption complexity γ . More generally, 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 v^* .

A. Problem formulation and methodology

In the formal corrupted sensing problem, we observe a measurement vector $y = \Phi x^* + v^* + z$ comprised of four components:

- **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 $\|\cdot\|_{\text{sig}}$.¹ Common examples of structured signals include sparse vectors, which exhibit low complexity with respect to the ℓ_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 II.
- **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 δ .
- **The measurement matrix** $\Phi \in \mathbb{R}^{n \times p}$, consisting of i.i.d. Gaussian entries

$$\Phi_{ij} \stackrel{\text{iid}}{\sim} N(0, 1/n)$$

as in [14]. Throughout, we treat x^* and v^* as fixed vectors chosen independently of Φ .² However, the

¹We focus on norms due to their popularity in structured estimation, but any convex complexity measure would suffice.

²In some settings, we can allow for x^* and v^* to be selected adversarially after the matrix Φ is generated with a similar theoretical analysis but do not present this work here. In the literature, Chandrasekaran et al. [14] 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.

unstructured noise z need not be independent of Φ and in particular may be chosen adversarially after Φ is generated.

Our goal is tractable estimation of x^* and v^* given knowledge of y , δ , and Φ . 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:

$$\min_{x,v} \left\{ \|x\|_{\text{sig}} + \lambda \|v\|_{\text{cor}} : \|y - (\Phi x + v)\|_2 \leq \delta \right\}. \quad (2)$$

In Section III, we will discuss specific settings of the parameter λ , 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

$$\min_{x,v} \left\{ \|v\|_{\text{cor}} : \|x\|_{\text{sig}} \leq \|x^*\|_{\text{sig}}, \|y - (\Phi x + v)\|_2 \leq \delta \right\} \quad (3)$$

or the corruption complexity using

$$\min_{x,v} \left\{ \|x\|_{\text{sig}} : \|v\|_{\text{cor}} \leq \|v^*\|_{\text{cor}}, \|y - (\Phi x + v)\|_2 \leq \delta \right\}. \quad (4)$$

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 complexity* 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 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 μ_n to represent the expected length of an n -dimensional vector with independent standard normal entries (equivalently, μ_n is the mean of the χ_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{B}^p and \mathbb{S}^{p-1} will denote the unit ball and the unit sphere in \mathbb{R}^p under the ℓ_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 ℓ_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 ℓ_1 norm and the sparse vector x_{sparse} , we have

$$\begin{aligned} \partial \|x_{\text{sparse}}\|_1 &= \text{sign}(x_{\text{sparse}}) + \{w \in \mathbb{R}^p : \\ &\text{support}(w) \cap \text{support}(x_{\text{sparse}}) = \emptyset, \|w\|_\infty \leq 1\}, \end{aligned} \quad (5)$$

³The same quantity was represented by λ_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 cone⁴ 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

$$T_{\text{sparse}} = \left\{ w : \left\| w_{\text{support}(x_{\text{sparse}})^c} \right\|_1 \leq -\langle w, \text{sign}(x_{\text{sparse}}) \rangle \right\}. \quad (6)$$

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 ℓ_1 and ℓ_2 norms.

C. Gaussian complexity and Gaussian distance

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

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

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

We call the square root of this quantity, $\omega(C) = \sqrt{\omega^2(C)}$, the *Gaussian complexity*.

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

$$\eta^2(C) = \mathbb{E}_{g \sim N(0, I_p)} \left[\inf_{w \in C} \|g - w\|_2^2 \right].$$

We call the square root of this quantity, $\eta(C) = \sqrt{\eta^2(C)}$, the *Gaussian distance*.

Chandrasekaran et al. [14] showed that the Gaussian squared complexity of a restricted tangent cone

⁴We adopt the same terminology for this cone as used by Chandrasekaran et al. [14]; McCoy and Tropp [15] call it the “feasible cone” since it is the cone of feasible directions under the constraint $\|x + w\| \leq \|x\|$. This definition of “tangent cone” nearly coincides with the use of the term in convex geometry—considering the convex subset $\{z \mid \|z\| \leq \|x\|\}$ (that is, the unit ball of the norm $\|\cdot\|$, rescaled by $\|x\|$), the tangent cone as defined in convex geometry is given by \overline{T} , the closure of T .

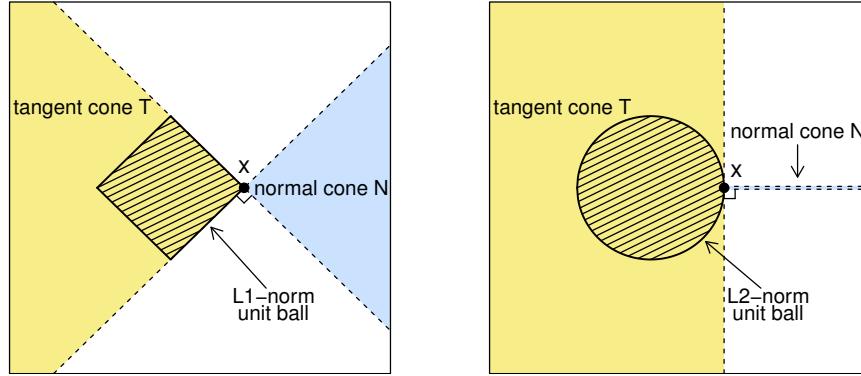


Figure 1: Illustrations of the tangent cone and normal cone for the ℓ_1 norm (left) and the ℓ_2 norm (right) at the point $x = (1, 0)$. (For visual clarity, cones are shifted to originate at the point x .)

$\omega^2(T \cap \mathbb{B}^p)$ determines a sufficient sample size for signal recovery from uncorrupted measurements.⁵ 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 $\omega(T \cap \mathbb{B}^p)$. Chandrasekaran et al. [14] describe a variety of methods for obtaining such bounds. One especially profitable technique, which can be traced back to [18, 19], relates this Gaussian complexity to the Gaussian distance of the scaled subdifferential of $\|\cdot\|$ at x , by way of the normal cone N :

$$\begin{aligned}\omega^2(T \cap \mathbb{B}^p) &= \eta^2(N) = \mathbb{E} \left[\inf_{w \in N} \|g - w\|_2^2 \right] \\ &= \mathbb{E} \left[\min_{t \geq 0} \inf_{w \in t \cdot \partial \|x\|} \|g - w\|_2^2 \right] \leq \min_{t \geq 0} \eta^2(t \cdot \partial \|x\|) .\end{aligned}$$

Indeed, many of the complexity bounds in [14] are derived by bounding $\eta(t \cdot \partial \|x\|)$ at a fixed value of t .

In Appendix A-G, we show that the best choice of t typically yields a bound within a small additive constant of the Gaussian complexity:

Proposition 1. Suppose that, for $x \neq 0$, $\partial \|x\|$ 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

$$\min_{t \geq 0} \eta(t \cdot \partial \|x\|) \leq \omega(T \cap \mathbb{B}^p) + 6 .$$

Remark 1. The weak decomposability assumption (7) is satisfied in all the examples considered in this paper and

⁵Chandrasekaran et al. [14] express their results in terms of the complexity measure $\nu(T \cap \mathbb{S}^{p-1}) := \mathbb{E} [\sup_{w \in T \cap \mathbb{S}^{p-1}} \langle g, w \rangle]$ (known as the Gaussian width), which is very slightly smaller than $\omega(T \cap \mathbb{B}^p)$. Nevertheless, the upper bounds developed in [14] for $\nu(T \cap \mathbb{S}^{p-1})$ hold also for $\omega(T \cap \mathbb{B}^p)$, a quantity which arises more naturally from our theory.

is weaker than the decomposability assumptions often found in the literature (see, e.g., [20]).

A complementary result relating Gaussian distance and Gaussian complexity appears in Amelunxen et al. [21, Thm. 4.5]:

Proposition 2 (Amelunxen et al. [21, Thm. 4.5]). *For any $x \neq 0$,*

$$\min_{t \geq 0} \eta^2(t \cdot \partial \|x\|) \leq \omega^2(T \cap \mathbb{B}^p) + \frac{2 \sup_{s \in \partial \|x\|} \|s\|}{\|x\| / \|x\|_2} .$$

Together, Propositions 1 and 2 demonstrate that the optimized Gaussian distance $\min_{t \geq 0} \eta(t \cdot \partial \|x\|)$ offers a faithful approximation to the Gaussian complexity $\omega(T \cap \mathbb{B}^p)$.

In the case of our ℓ_1 norm example, Chandrasekaran et al. [14] choose $t = \sqrt{2 \log(p/s)}$ and show that⁶

$$\begin{aligned}\omega^2(T_{\text{sparse}} \cap \mathbb{B}^p) &\leq \eta^2(t \cdot \partial \|x_{\text{sparse}}\|_1) \\ &\leq 2s \log(p/s) + \frac{3}{2}s .\end{aligned} \quad (8)$$

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

$$\begin{aligned}\omega^2(T_{\text{sparse}} \cap \mathbb{B}^p) &\leq \eta^2(t' \cdot \partial \|x_{\text{sparse}}\|_1) \\ &\leq p \cdot (1 - 2/\pi \cdot (1 - s/p)^2) ,\end{aligned} \quad (9)$$

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 definition

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

⁶Chandrasekaran et al. [14] report a slightly smaller bound based on a minor miscalculation.

In the ℓ_1 setting, this expected squared distance has a manageable representation (see [14, App. C]) that can be directly minimized over t :

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

The resulting bound on the Gaussian complexity 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 closed-form bound (9) closely approximates the optimal squared distance bound (10) for a wide range of sparsity levels, while the prior closed-form bound (8) only dominates (9) in the extreme sparsity regime.

It is instructive to contrast the complexity of $T_{\text{sparse}} \cap \mathbb{B}^p$ in our example with that induced by the ℓ_2 norm on \mathbb{R}^p . The ℓ_2 norm tangent cone at any point x will be a half-space (as illustrated in Figure 1), and its Gaussian complexity will be approximately \sqrt{p} , irrespective of the structure of x . Meanwhile, any sparse vector will exhibit a much smaller Gaussian complexity under the ℓ_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.

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 ℓ_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 ℓ_∞ 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 $\{1, \dots, p\}$ into m blocks of size p/m , has low complexity under the ℓ_1/ℓ_2 norm

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

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

Table I gives bounds on the squared complexity of the tangent cone for each of these structured settings (recall that μ_k is the mean of a χ_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 $\eta^2(t \cdot \partial \|\boldsymbol{x}\|)$ for a fixed scaling t of the subdifferential. We will make use of these Gaussian squared distance estimates and their associated settings of t when studying penalized convex recovery via (2). The new results in Tables I and II are derived in Appendix A.

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_{sig} denotes the tangent cone for $\|\cdot\|_{\text{sig}}$ at the true structured signal vector \boldsymbol{x}^* , and T_{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

$$\min \left\{ \|\boldsymbol{x}\|_{\text{sig}} : \|v\|_{\text{cor}} \leq \|v^*\|_{\text{cor}}, \|y - (\Phi \boldsymbol{x} + v)\|_2 \leq \delta \right\} \quad (11)$$

and

$$\min \left\{ \|v\|_{\text{cor}} : \|\boldsymbol{x}\|_{\text{sig}} \leq \|x^*\|_{\text{sig}}, \|y - (\Phi \boldsymbol{x} + v)\|_2 \leq \delta \right\}, \quad (12)$$

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

$$\omega^2(T_{\text{sig}} \cap \mathbb{B}^p) + \omega^2(T_{\text{cor}} \cap \mathbb{B}^n)$$

corrupted measurements suffice to recover (\boldsymbol{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 $(\widehat{\boldsymbol{x}}, \widehat{v})$ solves either of the constrained optimization problems (11) or (12), then*

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

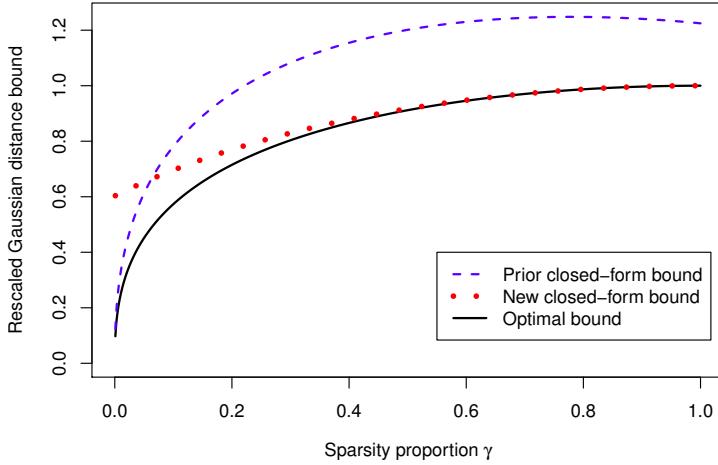


Figure 2: Three upper bounds on the rescaled Gaussian distance $\min_{t \geq 0} \eta(t \cdot \partial \|x_{\text{sparse}}\|_1) / \sqrt{p}$ for an s -sparse vector $x_{\text{sparse}} \in \mathbb{R}^p$ as a function of the sparsity proportion $\gamma = s/p$. The prior bound comes from (8), the new bound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

Table I: Upper bounds on the restricted tangent cone Gaussian squared complexity $\omega^2(T \cap \mathbb{B}^p)$ for various structured vectors and structure-inducing norms. Prior bounds on Gaussian complexity are due to [14].

Structure	$\ \cdot\ $	Prior Gaussian sq. complexity bound	New Gaussian sq. complexity bound
s -sparse p -dim. vector	$\ \cdot\ _1$	$2s \log(\frac{p}{s}) + \frac{3}{2}s$	$p \left(1 - \frac{2}{\pi}(1 - \frac{s}{p})^2\right)$
s -block-sparse vector (m blocks of size $k = p/m$)	$\ \cdot\ _{\ell_1/\ell_2}$	-	$4s \log(\frac{m}{s}) + (\frac{1}{2} + 3k)s,$ $p \left(1 - \frac{\mu_k^2}{k}(1 - \frac{s}{m})^2\right)$
r -rank $m_1 \times m_2$ matrix ($m_1 \geq m_2$)	$\ \cdot\ _*$	$3r(m_1 + m_2 - r)$	$m_1 m_2 \left(1 - (\frac{4}{27})^2 \left(1 - \frac{r}{m_1}\right) \left(1 - \frac{r}{m_2}\right)^2\right)$
binary p -dim. vector	$\ \cdot\ _\infty$	$\frac{p}{2}$	-

Table II: Upper bounds on the scaled subdifferential Gaussian squared distance $\eta^2(t \cdot \partial \|x\|)$ for various structured vectors, structure-inducing norms, and settings of t .

Structure	$\ \cdot\ $	Gaussian squared distance bound	Setting of t used to achieve bound
s -sparse p -dim. vector	$\ \cdot\ _1$	$2s \log(\frac{p}{s}) + \frac{3}{2}s,$ $p \left(1 - \frac{2}{\pi}(1 - \frac{s}{p})^2\right)$	$\sqrt{2 \log(p/s)},$ $\sqrt{\frac{2}{\pi}}(1 - s/p)$
s -block-sparse vector (m blocks of size $k = p/m$)	$\ \cdot\ _{\ell_1/\ell_2}$	$4s \log(\frac{m}{s}) + (\frac{1}{2} + 3k)s,$ $p \left(1 - \frac{\mu_k^2}{k}(1 - \frac{s}{m})^2\right)$	$\sqrt{2 \log(m/s)} + \sqrt{k},$ $\frac{\mu_k}{\sqrt{k}}(1 - s/m)$
r -rank $m_1 \times m_2$ matrix ($m_1 \geq m_2$)	$\ \cdot\ _*$	$m_1 m_2 \left(1 - (\frac{4}{27})^2 \left(1 - \frac{r}{m_1}\right) \left(1 - \frac{r}{m_2}\right)^2\right)$	$\frac{4}{27}(m_2 - r)\sqrt{m_1 - r}/m_2$

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

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

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 $\mu_n \geq \tau$. When $n = p$ and $\delta = 0$, this conclusion closely resembles the recent results of [15, 21]. See Section IV-D for a more detailed discussion of these related works.

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

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

with penalty parameter λ . Our next result provides sufficient conditions for exact and stable penalized recovery and demonstrates how to set the penalty parameter λ 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 (13) with penalty parameter $\lambda = t_{\text{cor}}/t_{\text{sig}}$, 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

$$\begin{aligned} \tau &= 2\eta \left(t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}} \right) + \eta \left(t_{\text{cor}} \cdot \partial \|v^*\|_{\text{cor}} \right) \\ &\quad + 3\sqrt{2\pi} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2\pi}}. \end{aligned}$$

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

Remark 4. In analogy to Theorem 1, one might expect that

$$\mu_n^2 \geq \eta^2 \left(t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}} \right) + \eta^2 \left(t_{\text{cor}} \cdot \partial \|v^*\|_{\text{cor}} \right) \quad (14)$$

would suffice for high probability penalized recovery. Indeed, our recovery result would have this form under a more symmetric observation model in which the corruption vector is also multiplied by a random Gaussian matrix (i.e., $y = \Phi x^* + \Psi v^* + z$, where Φ and Ψ are independent). Our experimental results (see Figure 6) suggest that while (14) is nearly the right threshold for recovery, a visible asymmetry exists between signal and corruption, stemming from the fact that only the signal vector is modulated by a Gaussian matrix under our observation model.

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

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

for suitably chosen subdifferential scalings t_{sig} and t_{cor} . Recall, from Section II-C, that these Gaussian distances represent practical upper bounds on the tangent cone Gaussian complexities, $\omega(T_{\text{sig}} \cap \mathbb{B}^p)$ and $\omega(T_{\text{cor}} \cap \mathbb{B}^n)$; in fact, Proposition 1 showed that a well-chosen distance bound closely approximates the tangent cone complexity for most norms of interest. In practice we can choose the distance scalings t_{sig} and t_{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 λ 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 λ 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

$$\min \frac{\|\Phi a + b\|_2}{\sqrt{\|a\|_2^2 + \|b\|_2^2}}, \quad (15)$$

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 (15) can then be applied to the error vector $(a, b) = (\hat{x} - x^*, \hat{v} - v^*)$ to bound its ℓ_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 [22, 23] 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 Φ 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

$$\min_{x,v} \{ \|v\|_1 : \|x\|_\infty \leq 1, \|y - (\Phi x + v)\|_2 \leq \delta \}, \quad (16)$$

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 (16).

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*

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

where $T_{\text{sparse}(n,\gamma)}$ is the tangent cone of a $(n \cdot \gamma)$ -sparse n -dimensional vector under the ℓ_1 norm, known to satisfy

$$\begin{aligned} \omega^2(T_{\text{sparse}(n,\gamma)} \cap \mathbb{B}^n) &\leq n \cdot \\ \min \left\{ 2\gamma \log(\gamma^{-1}) + \frac{3}{2}\gamma, 1 - \frac{2}{\pi}(1-\gamma)^2 \right\}. \end{aligned} \quad (17)$$

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 (16).

Remark 5. In addition to the interpretable bound (17), $\omega(T_{\text{sparse}(n,\gamma)} \cap \mathbb{B}^n)$ 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 $n = p$, that $\delta = 0$, 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 complexity $\omega(T_{\text{sparse}(n,\gamma)} \cap \mathbb{B}^n)$. For large $n = p$, the additive constants in Corollary 1 become negligible, and our result implies exact recovery with some probability when

$$\omega^2(T_{\text{sparse}(n,\gamma)} \cap \mathbb{B}^n) < ((1 - 2\delta)^2 - 1/2) \cdot n.$$

In the noiseless setting ($\delta = 0$), this precondition for exact recovery becomes $\omega^2(T_{\text{sparse}(n,\gamma)} \cap \mathbb{B}^n) < n/2$. Our plot of ℓ_1 complexity 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 [19, 24, 25, 26], 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 $\gamma = \frac{s}{m}$ may lie anywhere in $(0, 1)$. The following corollary bounds the number of measurements needed for exact or stable recovery, using the penalized program

$$\min_{x,v} \left\{ \|x\|_{\text{sig}} + \lambda \|v\|_{\ell_1/\ell_2} : \|y - (\Phi x + v)\|_2 \leq \delta \right\} \quad (18)$$

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

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

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 γ , (\hat{x}, \hat{v}) solves (18) with the penalty parameter $\lambda = \frac{\mu_k(1-\gamma)}{t_{\text{sig}}}$, and \sqrt{n} is at least as large as*

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

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

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

Remark 6. In the noiseless setting, where $\delta = 0$, Corollary 2 entails exact recovery with probability $1 - \beta$ whenever \sqrt{n} is at least as large as

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

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 ℓ_1 distance bound in Table I and Corollary 2 together imply that

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

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

$$\min_{x,v} \{ \|x\|_1 + \lambda \|v\|_1 : \|y - (\Phi x + v)\|_2 \leq \delta \}, \quad (19)$$

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, Thm. 1.1] establishes stable recovery via (19) using the penalty parameter $\lambda = 1/\sqrt{\log(p/n) + 1}$, whenever the signal and corruption sparsity levels satisfy

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

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-B) under the conditions

$$\|x^*\|_0 \leq s_\gamma := \left\lfloor \frac{A_\gamma n}{2 \log(\frac{p}{A_\gamma n}) + 3/2} \right\rfloor \quad \text{and} \quad \|v^*\|_0 \leq n \cdot \gamma, \quad (20)$$

for any corruption proportion bound $\gamma < 1$ and

$$A_\gamma := (\alpha_1(1-\gamma)^2 - \epsilon)_+^2 / 144.$$

(Recall that $\alpha_1 = 1 - \sqrt{1 - 2/\pi} \approx 0.4$.) The associated setting of the penalty parameter $\lambda = (1-\gamma)/\sqrt{\pi \log(p/s_\gamma)}$ is parameterized only by γ and, like Li's result, requires no prior knowledge of the signal sparsity level $\|x^*\|_0$. Our setting has the added advantage of allowing the user to specify an arbitrary upper bound on the corruption level.

It should be noted that Li [1] addresses the adversarial setting in which x^* and v^* may be selected given knowledge of the measurement matrix Φ . Our work can be adapted to the adversarial setting by measuring the Gaussian complexity of the ℓ_1 unit ball in place of the associated tangent cones; we save such an extension for future work.

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, where the corruption proportion γ is nearly zero, the result can be improved with a different choice of penalty parameter λ ; the reason for this distinction is that, when choosing t_{cor} , it is more advantageous to use the first distance bound that is given in Table II for the block-wise sparse setting when the corruptions are extremely sparse, rather than using the second distance bound in this table as for the high-corruption setting. A general result for the extremely sparse case could be attained with a similar analysis as in Corollary 2; we do not state such a result here but instead give a specific example where both the signal and corruption exhibit extreme element-wise sparsity.

In the following corollary, we show that

$$\mathcal{O}\left((s_{\text{sig}} + s_{\text{cor}}) \left(\log\left(\frac{p+n}{s_{\text{sig}} + s_{\text{cor}}}\right) + 1 \right)\right)$$

measurements suffice to ensure recovery even without any knowledge of the sparsity levels s_{sig} and s_{cor} , by selecting the penalty parameter $\lambda = 1$.

Corollary 3 (Extremely sparse signal and corruption). *Suppose that x^* is supported on at most s_{sig} of p entries, and that v^* is supported on at most s_{cor} of n entries. If (\hat{x}, \hat{v}) solves*

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

and the number of measurements n satisfies

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

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

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

When the noise level equals zero, we recover a result due to Laska et al. [27], 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 $n \geq C(s_{\text{sig}} + s_{\text{cor}}) \log(\frac{p+n}{s_{\text{sig}} + s_{\text{cor}}})$ 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_{sig} blocks of size k_{sig} and block-sparse corruption vectors with m_{cor} blocks of size k_{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 [28] study exact recovery under a *cross-and-bouquet* model: $y = Ax^* + v^*$ for

$$A^{(i)} \stackrel{\text{iid}}{\sim} N(\mu, \frac{\nu^2}{n} \mathbf{I}_p), \quad \|\mu\|_2 = 1, \quad \|\mu\|_\infty \leq \frac{C}{\sqrt{n}},$$

where $A^{(i)}$ is the i th column of A , and v^* has uniformly random signs. Nguyen and Tran [29] 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. [30] 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 [31] 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 global minima can 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 the same setting, the recent work of Amelunxen et al. [21] independently relates the notion of Gaussian (squared) distance to the probability of signal recovery. These works also establish the sharpness of their recovery results, by demonstrating a decay in success probability whenever the sample size falls below an appropriate threshold. In contrast, our work does not discuss high probability failure bounds but 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 parameter settings suggested in Theorem 2 when using penalized recovery. In each experiment, we employ the CVX Matlab package [32, 33] to specify and solve our convex recovery programs.

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 $\omega^2(T_{\text{sig}} \cap \mathbb{B}^p) + \omega^2(T_{\text{cor}} \cap \mathbb{B}^n)$. Using new and existing bounds on the relevant Gaussian complexities, 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_{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_{cor} independent standard normal entries and $n - s_{\text{cor}}$ entries set to 0.
- 4) Solve the constrained optimization problem (16) with $y = \Phi x^* + v^*$ and $\delta = 0$.
- 5) Declare success if $\|\hat{x} - x^*\|_2 / \|x^*\|_2 < 10^{-3}$.

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

$$\mu_n^2 = \omega^2(T_{\text{sig}} \cap \mathbb{B}^p) + \omega^2(T_{\text{cor}} \cap \mathbb{B}^n)$$

where the Gaussian complexity for the binary signal is estimated with the bound given in Table I, while the complexity for the sparse corruption is given by the optimal squared distance bound for the ℓ_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 ℓ_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_{sig} independent standard normal entries and $p - s_{\text{sig}}$ entries set to 0.
- 2) Generate a corruption vector v^* with s_{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 $y = \Phi x^* + v^*$:

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

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

Figure 4 reports the average empirical probability of success for each setting of $(s_{\text{sig}}, s_{\text{cor}})$. We again overlay the theoretical recovery threshold suggested by Theorem 1, where the Gaussian complexities for both signal and corruption are estimated with the tight upper bounds given by the optimal squared distance bound for the ℓ_1 norm, given in (10). This theoretical recovery threshold reflects the observed empirical phase transition accurately.

Sparse signal, block-sparse corruption: To further test the resilience of our theory to a change in structure type, we next consider the problem of recovering a sparse signal vector of known ℓ_1 norm from measurements with block-sparse corruptions. We fix $p = n = 1000$ and partition the indices of v^* 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_{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

$$\begin{aligned} (\hat{x}, \hat{v}) \in \arg \min_{x, v} \{ \|v\|_{\ell_1/\ell_2} : \\ \|x\|_1 \leq \|x^*\|_1, y = \Phi x + v \}. \end{aligned}$$

Results from this simulation are displayed in Figure 5, with the overlaid threshold coming from the optimal squared distance bound for the ℓ_1 norm, given in (10), and for the ℓ_1/ℓ_2 norm, as follows (see (26) in Ap-

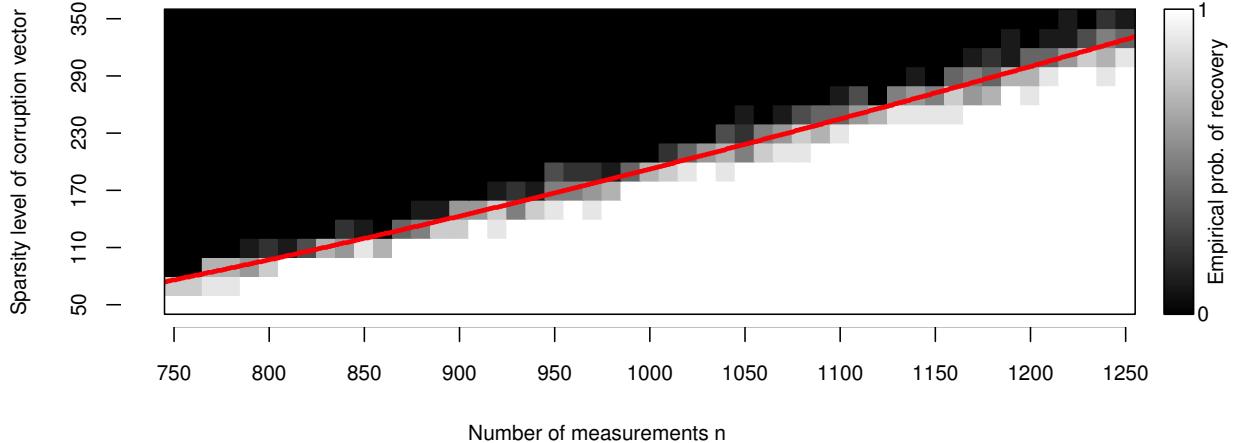


Figure 3: Phase transition for binary signal recovery under sparse corruption, using constrained recovery (see Section V-A). The red curve plots the recovery threshold predicted by Theorem 1 (ignoring the small additive constants that we believe are artifacts of the proof technique), where Gaussian squared complexity for the binary signal is given in Table I, and Gaussian squared complexity for the sparse corruption is estimated by minimizing (10).

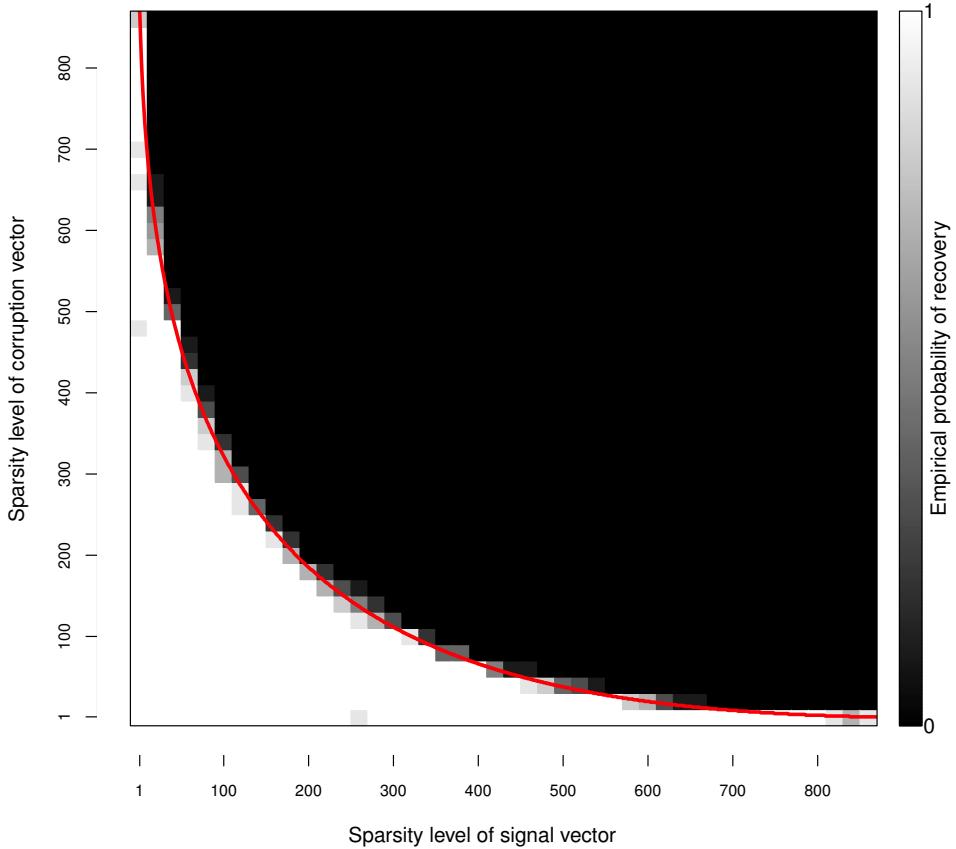


Figure 4: Phase transition for sparse signal recovery under sparse corruption, using constrained recovery (see Section V-A). The red curve plots the recovery threshold predicted by Theorem 1 (ignoring the small additive constants that we believe are artifacts of the proof technique), where Gaussian complexity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

pendix A-C for the derivation):

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

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 ℓ_1 norm on both signal and corruption, rather than leveraging the additional block structure via the ℓ_1/ℓ_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 $\omega^2(T_{\text{sig}} \cap \mathbb{B}^p) + \omega^2(T_{\text{cor}} \cap \mathbb{B}^n)$, the known correspondence between constrained and penalized optimization implies that for some (unknown) value of λ , 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 behavior without prior knowledge of the signal or corruption norm. Our penalized recovery result, Theorem 2, suggest a simple strategy for picking the penalty: set $\lambda = t_{\text{cor}}/t_{\text{sig}}$ where t_{sig} is a scaling that leads to a small bound on the signal Gaussian distance $\eta(t_{\text{sig}} \cdot \partial \|x^*\|_1)$ and t_{cor} is an analogous scaling for a corruption Gaussian distance. We will test four settings of the penalty parameter, three of which depend on the signal and corruption sparsity levels s_{sig} and s_{cor} :

- 1) $\lambda_{\text{sparse}} = \sqrt{\log(n/s_{\text{cor}})/\log(p/s_{\text{sig}})}$, based on the closed-form 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 closed-form bound (9) that is nearly optimal for all but the sparsest vectors,
- 3) λ_{opt} , which chooses t_{sig} and t_{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_{sig} and s_{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 (19) with each penalty parameter setting and declare success if $\|\hat{x} - x^*\|_2 / \|x^*\|_2 < 10^{-3}$. Figure 6 displays the empirical probability of success as the signal and corruption sparsity levels s_{sig} and s_{cor} vary. For reference, the theoretical recovery curve of Figure 4 has been overlaid. Remarkably, setting λ to equal any of λ_{sparse} , λ_{dense} , or λ_{opt} 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 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_{sig} and s_{cor} are unavailable, the setting $\lambda = \lambda_{\text{const}} = 1$ yields high probability recovery, provided that neither signal nor corruption is overly dense.

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 $y = \Phi x^* + v^*$:

$$\begin{aligned} (\hat{x}, \hat{v}) \in \arg \min_{x, v} & \left\{ \|v\|_1 : \right. \\ & \left. \|x\|_1 \leq \|x^*\|_1, \|y - \Phi x - v\|_2 \leq \delta \right\}, \end{aligned}$$

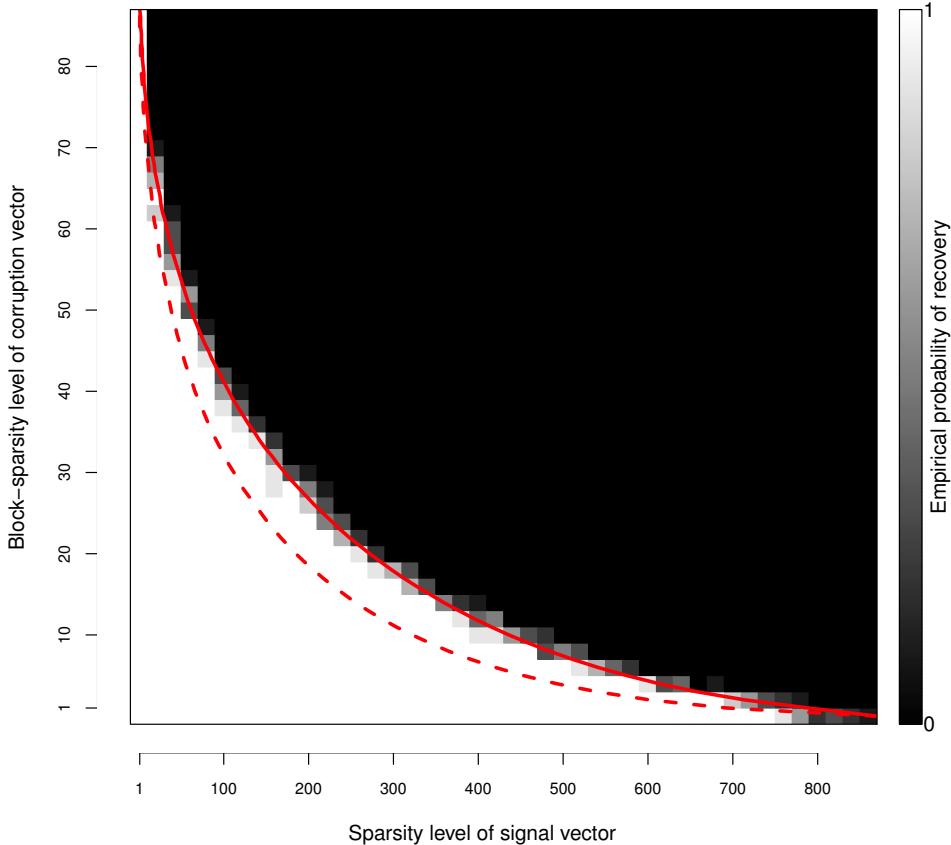


Figure 5: Phase transition for sparse signal recovery under block-wise sparse corruption, using constrained recovery (see Section V-A). The solid red curve plots the recovery threshold predicted by Theorem 1 (ignoring the small additive constants that we believe are artifacts of the proof technique), where Gaussian complexity is estimated analogously to the sparse signal plus sparse corruption problem. The dashed red curve plots the recovery threshold that would be predicted if the ℓ_1 norm, rather than the ℓ_1/ℓ_2 norm, were used on the corruption term (the same curve that appears in Figure 4); this shows the benefit of leveraging the block structure in the corruption.

5) Record the ℓ_2 recovery error $\sqrt{\|\hat{x} - x^*\|_2^2 + \|\hat{v} - v^*\|_2^2}$.

Our theory suggests that rescaling the ℓ_2 recovery error by a factor of

$$n^{-1/2} \left(\mu_n - \sqrt{\omega^2 (T_{\text{sig}} \cap \mathbb{B}^p) + \omega^2 (T_{\text{cor}} \cap \mathbb{B}^n)} \right), \quad (21)$$

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 complexities in the scaling factor (21). 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 complexity and the Gaussian distance, for which we developed new interpretable 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

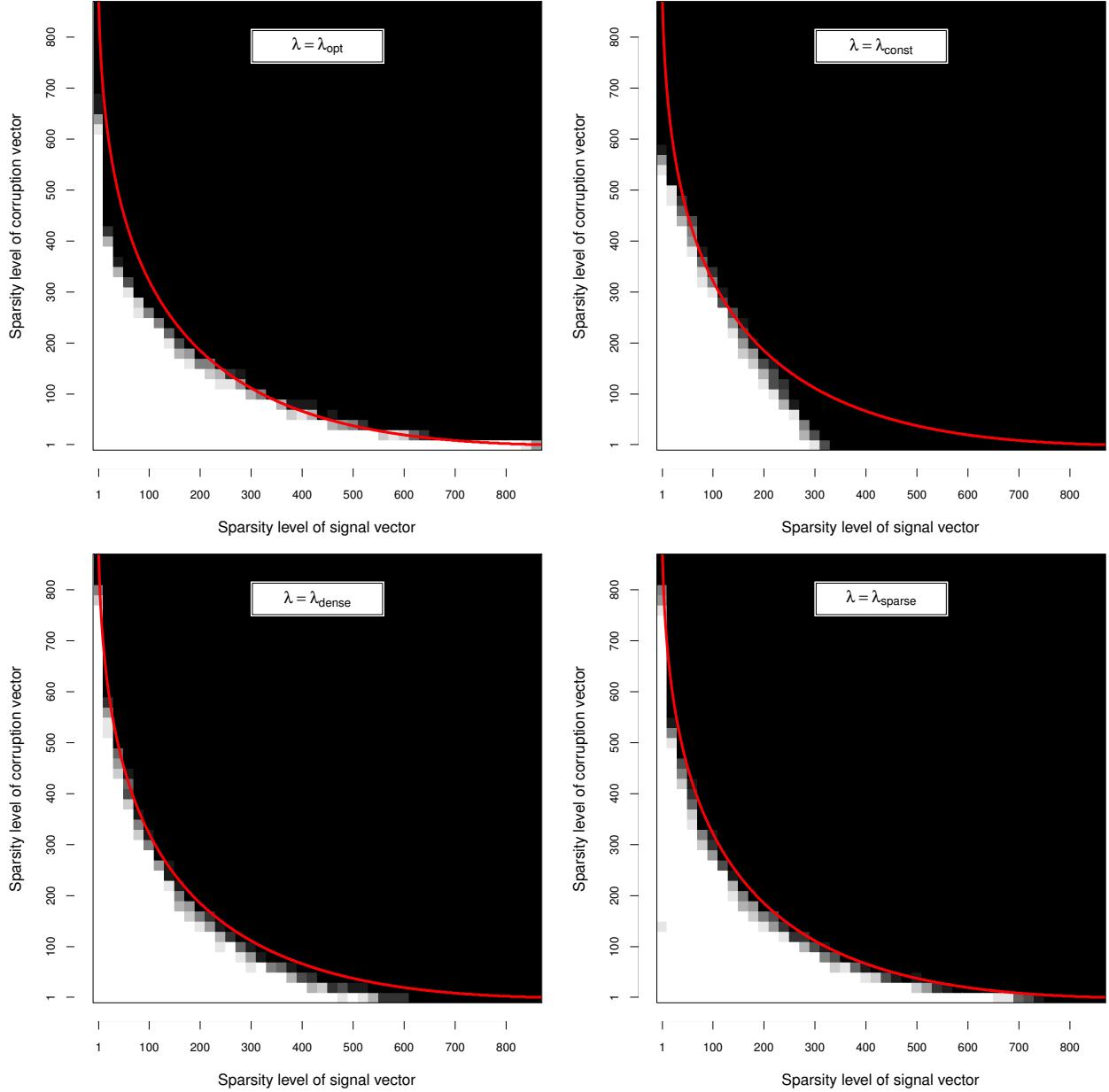


Figure 6: Phase transitions for sparse signal recovery under sparse corruption, using penalized recovery with varied settings of the penalty parameter λ (see Section V-B). Grayscale representation of success probability is same as in Figure 4. The red curve in each figure is the theoretical recovery threshold predicted by Theorem 1 for the constrained recovery problem (same curve as in Figure 4); we display this curve to show that signal recovery is nearly as good as in the constrained case, but without any prior knowledge of $\|x^*\|_1$ or $\|v^*\|_1$.

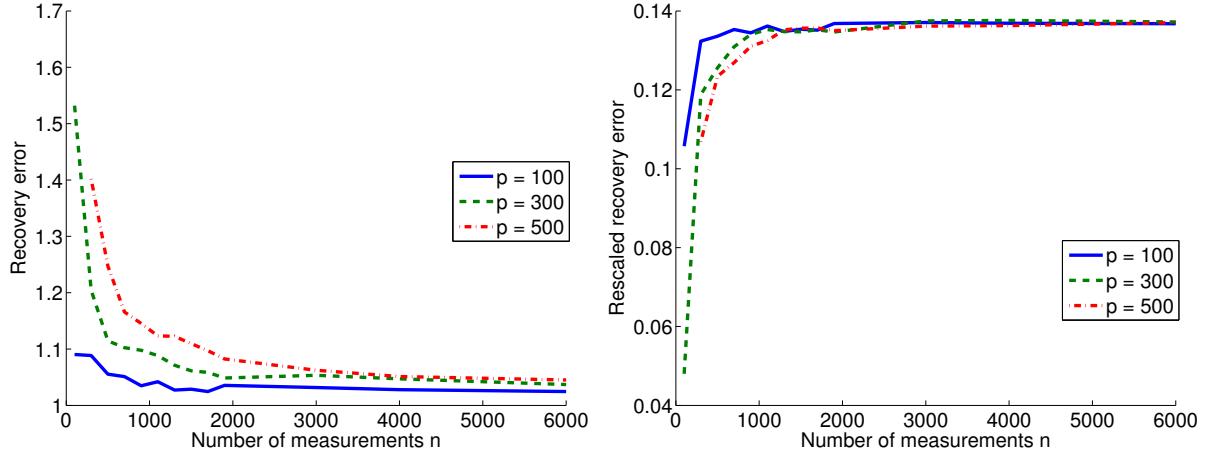


Figure 7: (Left) Recovery error $\sqrt{\|\hat{x} - x^*\|_2^2 + \|\hat{v} - v^*\|_2^2}$ for sparse signal recovery ($s_{\text{sig}} = 0.01p$) under noise and sparse corruption ($s_{\text{cor}} = 0.4n$), using constrained recovery. (Right) Same recovery error rescaled by the optimal squared distance estimate (10) of $(\mu_n - \sqrt{\omega^2(T_{\text{sig}} \cap \mathbb{B}^p) + \omega^2(T_{\text{cor}} \cap \mathbb{B}^n)}) / \sqrt{n}$, the scaling recommended by Theorem 1.

convex program (13) with canonical choice of penalty parameter performed nearly as well empirically as the constrained convex program (12) 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

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

or

$$\min_{x,v} \left\{ \|x\|_{\text{sig}} + \lambda \|v\|_{\text{cor}} + \nu \|y - (\Phi x + v)\|_2^2 \right\}$$

when no prior bound δ 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 ℓ_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 squared Gaussian complexity $\omega^2(T \cap \mathbb{B}^p)$ and the Gaussian squared distance $\eta^2(t \cdot \partial \|x\|)$. 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 complexity and distance bounds will be based on two fundamental relationships among Gaussian complexity, Gaussian distance, and a third measure of size known as the *Gaussian width*:

$$\nu(C) \triangleq \mathbb{E} \left[\max_{w \in C} \langle g, w \rangle \right].$$

The first relation, established in [14], upper bounds the Gaussian complexity of a constrained tangent cone in terms of the Gaussian distance and the expected squared distance to the scaled subdifferential

$$\begin{aligned} \omega^2(T \cap \mathbb{B}^p) &\leq \min_{t \geq 0} \eta^2(t \cdot \partial \|x\|) \\ &= \min_{t \geq 0} \mathbb{E} \left[\text{dist}(g, t \cdot \partial \|x\|)^2 \right]. \end{aligned} \quad (22)$$

Our bounds on the right-hand side of (22) 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 3. *Let $\tilde{\omega}$ be any lower bound on $\nu(\partial \|x\|)$. For $t = \frac{\tilde{\omega}}{\max_{\|z\|_2=1} \|z\|}$,*

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

We will see that the subdifferentials of many structured norms admit simple lower bounds that lead to tight upper bounds on the Gaussian squared distance and Gaussian squared complexity via Proposition 3.

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 distance bound

$$\begin{aligned} \omega^2(T \cap \mathbb{B}^p) &\leq \eta^2(t \cdot \partial \|x\|) = \mathbb{E} \left[\text{dist}(g, t \cdot \partial \|x\|_1)^2 \right] \\ &\leq 2s \log \left(\frac{p}{s} \right) + \frac{3}{2}s, \end{aligned}$$

for $t = \sqrt{2 \log(p/s)}$. Using Proposition 3, we obtain a second bound

$$\begin{aligned} \omega^2(T \cap \mathbb{B}^p) &\leq \eta^2(t' \cdot \partial \|x\|) \\ &= \mathbb{E} \left[\text{dist}(g, t' \cdot \partial \|x\|_1)^2 \right] \leq p \left(1 - \frac{2}{\pi} (1 - s/p)^2 \right), \end{aligned} \quad (23)$$

for $t' = \sqrt{2/\pi} (1 - s/p)$. This follows from our more general treatment of block-sparse vectors in Proposition 4 below.

C. Block-sparse vectors

Suppose that the indices $\{1, \dots, p\}$ have been partitioned into disjoint blocks V_1, \dots, 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.,

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

Proposition 4 presents our two new bounds on Gaussian distance and hence on Gaussian complexity in this setting.

Proposition 4 (Block-sparse vector Gaussian distance). *Partition the indices $\{1, \dots, p\}$ into blocks V_1, \dots, 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:*

$$\eta^2(t \cdot \partial \|x\|_{\ell_1/\ell_2}) \leq 4s \log(m/s) + (0.5 + 3k)s \quad (24)$$

$$\text{for } t = \sqrt{2 \log(m/s)} + \sqrt{k},$$

$$\eta^2(t' \cdot \partial \|x\|_{\ell_1/\ell_2}) \leq p \left(1 - \frac{\mu_k^2}{k} (1 - s/m)^2 \right) \quad (25)$$

$$\text{for } t' = \mu_k (1 - s/m).$$

For element-wise sparsity ($k = 1$), the bound (25) specializes to the bound (23) given above. The advantage of exploiting block-sparse structure is evident when one compares the factor of $\frac{\mu_k^2}{k}$ in the bound (25) to the term $\frac{2}{\pi} = \mu_1^2$ found in (23). 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 \mathcal{B} denote the block indices on which x is supported, with $|\mathcal{B}| = s' \leq s$, let \mathcal{T} be the space of vectors with support only on those blocks in \mathcal{B} , and let $a \in \mathcal{T}$ be the vector satisfying

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

Then, we have the subdifferential [34]

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

We begin by establishing the bound (25) based on the Gaussian width of the subdifferential. For each b , $\|\mathcal{P}_{V_b}(g)\|_2$ is χ_k distributed. Using the form of the subdifferential stated above,

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

Since $\sup_{y \in \mathbb{R}^p : \|y\|_2=1} \|y\|_{\ell_1/\ell_2} = \sqrt{m}$, Proposition 3 now implies the result (25).

Now we turn to the bound (24). By (22), 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

$$\begin{aligned} & \mathbb{E} \left[\text{dist} \left(g, t \cdot \partial \|x\|_{\ell_1/\ell_2} \right)^2 \right] \\ &= \mathbb{E} \left[\|\mathcal{P}_T(g) - ta\|_2^2 + \sum_{b \notin \mathcal{B}} \inf_{\|z_b\|_2 \leq t} \|\mathcal{P}_{V_b}(g) - z_b\|_2^2 \right] \\ &= \mathbb{E} \left[\|\mathcal{P}_T(g) - ta\|_2^2 + \sum_{b \notin \mathcal{B}} (\|\mathcal{P}_{V_b}(g)\|_2 - t)_+^2 \right] \\ &= s'(t^2 + p/m) + \sum_{b \notin \mathcal{B}} \mathbb{E} [(\|\mathcal{P}_{V_b}(g)\|_2 - t)_+^2] \\ &= s'(t^2 + p/m) + (m - s') \cdot \mathbb{E} [(\xi - t)_+^2] \\ &\leq s(t^2 + p/m) + (m - s) \cdot \mathbb{E} [(\xi - t)_+^2], \end{aligned} \tag{26}$$

where ξ is a χ_k random variable. We will bound each summand in this expression.

Letting $d = t - \sqrt{k}$, we have

$$\begin{aligned} & \mathbb{E} [(\xi - t)_+^2] \\ &= \int_0^\infty \mathbb{P} \{(\xi - t)_+^2 \geq a\} da \\ &= \int_0^\infty \mathbb{P} \{\xi^2 \geq (t + \sqrt{a})^2\} da \\ &= \int_0^\infty \mathbb{P} \{\xi^2 - k \geq 2\sqrt{k}(d + \sqrt{a}) + (d + \sqrt{a})^2\} da \end{aligned}$$

Applying the change of variables $c = (d + \sqrt{a})$,

$$\begin{aligned} &= 2 \int_d^\infty \mathbb{P} \{\xi^2 - k \geq 2\sqrt{kc} + c^2\} (c - d) dc \\ &\leq 2 \int_d^\infty e^{-c^2/2} (c - d) dc \\ &= 2e^{-d^2/2} - 2d \int_d^\infty e^{-c^2/2} dc \\ &\leq 2e^{-d^2/2} - 2 \frac{d^2}{d^2 + 1} e^{-d^2/2} = \frac{2}{d^2 + 1} e^{-d^2/2}, \end{aligned}$$

where the penultimate inequality follows from the chi-squared tail bound [35, Lem. 1],

$$\mathbb{P} \{\xi^2 - k \geq \sqrt{2k}c + c^2\} \leq e^{-c^2/2} \quad \text{for all } c > 0,$$

and the final inequality follows from a bound on the Gaussian Q-function

$$\begin{aligned} Q(d) &= \frac{1}{\sqrt{2\pi}} \int_d^\infty e^{-c^2/2} dc \\ &\geq \frac{1}{\sqrt{2\pi}} \frac{d^2}{d^2 + 1} \int_d^\infty \frac{c^2 + 1}{c^2} e^{-c^2/2} dc \\ &= \frac{1}{\sqrt{2\pi}} \frac{d}{d^2 + 1} e^{-d^2/2} \quad \text{for all } d > 0. \end{aligned}$$

Hence we have

$$\begin{aligned} & \mathbb{E} \left[\text{dist} \left(g, t \cdot \partial \|x\|_{\ell_1/\ell_2} \right)^2 \right] \\ &\leq s(t^2 + p/m) + \frac{2(m - s)}{d^2 + 1} e^{-d^2/2} \\ &= s(d^2 + 2d\sqrt{k} + 2k) + \frac{2(m - s)}{d^2 + 1} e^{-d^2/2} \\ &\leq s(2d^2 + 3k) + \frac{2(m - s)}{d^2 + 1} e^{-d^2/2}. \end{aligned}$$

Since $d = \sqrt{2 \log(m/s)}$, we have

$$\begin{aligned} & \mathbb{E} \left[\text{dist} \left(g, t \cdot \partial \|x\|_{\ell_1/\ell_2} \right)^2 \right] \\ &\leq s(4 \log(m/s) + 3k) + \frac{2s(1 - s/m)}{2 \log(m/s) + 1} \\ &\leq (0.5 + 3k)s + 4s \log(m/s), \end{aligned}$$

as desired, since

$$\frac{2(1 - s/m)}{2 \log(m/s) + 1} < 0.5$$

whenever $1 \leq s \leq m$.

D. Binary vectors

The convex hull of the set of binary vectors $\{\pm 1\}^p$ is the unit ball of the ℓ_∞ norm, so we choose $\|\cdot\| = \|\cdot\|_\infty$ when $x \in \{\pm 1\}^p$. This choice yields

$$\partial \|x\|_\infty = \{w \in \mathbb{R}^p : \|w\|_1 = 1, w_i x_i \geq 0 \forall i\}$$

and therefore

$$T = \{w \in \mathbb{R}^p : w_i x_i \leq 0 \forall i\}$$

and

$$N = \{w \in \mathbb{R}^p : w_i x_i \geq 0 \forall i\}.$$

Thus, $\omega(T \cap \mathbb{B}^p) = \omega(N \cap \mathbb{B}^p)$, and [14, Lem. 3.7] implies that⁷

$$\omega(T \cap \mathbb{B}^p) \leq \sqrt{\frac{p}{2}}.$$

We note that a bound based on Proposition 3 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 $\|\cdot\|_*$. Chandrasekaran et al. [14] established the bound $\omega^2(T \cap \mathbb{B}^{m_1 \times m_2}) \leq 3r(m_1 + m_2 - r)$. Proposition 5 presents our new estimate based on the Gaussian width of the subdifferential.

Proposition 5 (Low-rank matrix Gaussian distance). *Let $x \in \mathbb{R}^{m_1 \times m_2}$ have rank at most r , and let T be the tangent cone associated with $\|\cdot\|_*$ at x . If $m_1 \geq m_2$, then setting $t = \frac{4}{27}(m_2 - r)\sqrt{m_1 - r}/m_2$,*

$$\eta^2(t \cdot \partial \|x\|_*) \leq m_1 m_2 \cdot \left(1 - \left(\frac{4}{27}\right)^2 \left(1 - \frac{r}{m_1}\right) \left(1 - \frac{r}{m_2}\right)^2\right).$$

Proof: Let $U\Sigma V^\top$ be the compact singular value decomposition of x , and let \mathcal{T} be the space of matrices in the column or row space of x . The subdifferential is given by [36]

$$\partial \|x\|_* = UV^\top + \{w \in \mathcal{T}^\perp : \|w\|_{\text{op}} \leq 1\},$$

where $\|w\|_{\text{op}}$ is the operator norm, i.e. the largest singular value of w .

We begin by bounding the Gaussian width of $\partial \|x\|_*$. Let $\Gamma \in \mathbb{R}^{m_1 \times m_2}$ have i.i.d. standard Gaussian entries.

⁷ While this lemma is stated in terms of Chandrasekaran et al. [14] definitions of width (as discussed in Section II-C), examining the proof shows that it holds in this case as well.

- Since orthogonal projection cannot increase the operator norm, we have

$$\begin{aligned} & \mathbb{E} \left[\max_{w \in \partial \|x\|_*} \langle g, w \rangle \right] \\ &= \mathbb{E} \left[\langle \Gamma, UV^\top \rangle + \sup_{w: \|\mathcal{P}_{\mathcal{T}}^\perp(w)\|_{\text{op}} \leq 1} \langle \Gamma, \mathcal{P}_{\mathcal{T}}^\perp(w) \rangle \right] \\ &\geq \mathbb{E} \left[\sup_{w: \|w\|_{\text{op}} \leq 1} \langle \mathcal{P}_{\mathcal{T}}^\perp(\Gamma), w \rangle \right] = \mathbb{E} [\|\mathcal{P}_{\mathcal{T}}^\perp(\Gamma)\|_*]. \end{aligned}$$

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 \mathcal{T} , we know that $\mathcal{P}_{\mathcal{T}}^\perp(\Gamma) = U'U'^\top \Gamma V'V'^\top$, and so

$$\begin{aligned} \mathbb{E} [\|\mathcal{P}_{\mathcal{T}}^\perp(\Gamma)\|_*] &= \mathbb{E} [\|U'U'^\top \Gamma V'V'^\top\|_*] \\ &= \mathbb{E} [\|U'^\top \Gamma V'\|_*], \end{aligned}$$

since the trace norm is unitarily invariant. Furthermore, since U' and V' each have orthonormal columns, the entries of $U'^\top \Gamma V'$ are i.i.d. standard Gaussian. We now apply the following lemma (proved in Appendix D):

Lemma 1. *Let $\Gamma \in \mathbb{R}^{m_1 \times m_2}$ have i.i.d. standard Gaussian entries, with $m_1 \geq m_2$. Then*

$$\mathbb{E} [\|\Gamma\|_*] \geq \frac{4}{27} m_2 \sqrt{m_1}.$$

Therefore,

$$\mathbb{E} \left[\max_{w \in \partial \|x\|_*} \langle g, w \rangle \right] \geq \frac{4}{27} (m_2 - r) \sqrt{m_1 - r}.$$

Finally, we apply Proposition 3 to obtain the desired bound. ■

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 2. *For any norm $\|\cdot\|$ on \mathbb{R}^p with dual norm $\|\cdot\|^*$,*

$$\max_{x \in \mathbb{R}^p} \frac{\|x\|}{\|x\|_2} = \max_{x \in \mathbb{R}^p} \frac{\|x\|_2}{\|x\|^*}.$$

Now we prove our Gaussian distance bound that is based on the subdifferential. We use the fact that $\|w\|^* = 1$ for all $w \in \partial \|x\|$.

Proposition 3. Let $\tilde{\omega}$ be any lower bound on $\mathbb{E} [\max_{w \in \partial \|x\|} \langle g, w \rangle]$. For $t = \frac{\tilde{\omega}}{\max_{\|z\|_2=1} \|z\|}$,

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

Proof: 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 ,

$$\begin{aligned} & \text{dist}(g, t \cdot \partial \|x\|)^2 \\ & \leq \|g - t \cdot w_0\|_2^2 = \|g\|_2^2 - 2t \langle g, w_0 \rangle + t^2 \|w_0\|_2^2 \\ & = \|g\|_2^2 - 2t \max_{w \in \partial \|x\|} \langle g, w \rangle + t^2 \|w_0\|_2^2 \end{aligned}$$

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

$$\begin{aligned} & \leq \|g\|_2^2 - 2t \max_{w \in \partial \|x\|} \langle g, w \rangle + t^2 \max_{\|w\|^*=1} \|w\|_2^2 \\ & \leq \|g\|_2^2 - 2t \max_{w \in \partial \|x\|} \langle g, w \rangle + t^2 \max_{\|w\|_2=1} \|w\|^2, \end{aligned}$$

where the last step comes from Lemma 2. Taking expectations,

$$\begin{aligned} & \mathbb{E} [\text{dist}(g, t \cdot \partial \|x\|)^2] \leq \\ & p - 2t \cdot \mathbb{E} \left[\max_{w \in \partial \|x\|} \langle g, w \rangle \right] + t^2 \cdot \max_{\|w\|_2=1} \|w\|^2 \\ & \leq p - 2t \cdot \tilde{\omega} + t^2 \cdot \max_{\|w\|_2=1} \|w\|^2. \end{aligned}$$

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

G. Relating Gaussian distance and Gaussian complexity

Proposition 1. Suppose that, for $x \neq 0$, $\partial \|x\|$ satisfies a weak decomposability assumption:

$$\exists w_0 \in \partial \|x\| \text{ s.t. } \langle w - w_0, w_0 \rangle = 0 \quad \forall w \in \partial \|x\|. \quad (27)$$

Then

$$\min_{t \geq 0} \eta(t \cdot \partial \|x\|) \leq \omega(T \cap \mathbb{B}^p) + 6.$$

Proof: For $g \in \mathbb{R}^p$, define⁸

$$t_g = \arg \min_{t \geq 0} \text{dist}(g, t \cdot \partial \|x\|). \quad (28)$$

⁸ 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 .

Define the event

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

We will use the following lemma (proved in Appendix D):

Lemma 3. Suppose that, for $x \neq 0$, $\partial \|x\|$ satisfies (27). Let t_g be defined as in (28). Then $g \mapsto t_g$ is a $\frac{1}{\|w_0\|_2}$ -Lipschitz function of g .

Therefore, applying a bound due to Ledoux [37, (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

$$\frac{t_g}{\mathbb{E}[t_g] + \frac{2}{\|w_0\|_2}} \cdot w + \frac{\mathbb{E}[t_g] + \frac{2}{\|w_0\|_2} - t_g}{\mathbb{E}[t_g] + \frac{2}{\|w_0\|_2}} \cdot w_0 \in \partial \|x\|,$$

and so

$$\begin{aligned} & \text{dist}\left(g, \left(\mathbb{E}[t_g] + \frac{2}{\|w_0\|_2}\right) \cdot \partial \|x\|\right) \\ & \leq \left\| g - \left[t_g \cdot w + \left(\mathbb{E}[t_g] + \frac{2}{\|w_0\|_2} - t_g\right) \cdot w_0 \right] \right\|_2 \\ & \leq \|g - t_g \cdot w\|_2 + \left(\mathbb{E}[t_g] + \frac{2}{\|w_0\|_2} - t_g\right) \cdot \|w_0\|_2 \\ & = \text{dist}(g, N) + \left(\mathbb{E}[t_g] + \frac{2}{\|w_0\|_2} - t_g\right) \cdot \|w_0\|_2 \\ & < \text{dist}(g, N) + 4, \end{aligned}$$

where the last step comes from the definition of the event \mathcal{E} . Therefore,

$$\begin{aligned} & \mathbb{P} \left\{ \frac{1}{2} \text{dist}\left(g, \left(\mathbb{E}[t_g] + \frac{2}{\|w_0\|_2}\right) \cdot \partial \|x\|\right) - \right. \\ & \quad \left. \frac{1}{2} \text{dist}(g, N) < 2 \right\} \\ & \geq \mathbb{P}\{\mathcal{E}\} \geq 1 - 2e^{-2^2/2}. \end{aligned}$$

Since $g \mapsto \text{dist}(g, A)$ is clearly a 1-Lipschitz function of g for any convex set A , we see that

$$g \mapsto \frac{1}{2} \text{dist}\left(g, \left(\mathbb{E}[t_g] + \frac{2}{\|w_0\|_2}\right) \cdot \partial \|x\|\right) - \frac{1}{2} \text{dist}(g, N)$$

is a 1-Lipschitz function of g . We will now make use of the following lemma (proved in Appendix D):

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

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

Therefore,

$$\begin{aligned} & \frac{\eta\left(\left(\mathbb{E}[t_g] + \frac{2}{\|w_0\|_2}\right) \cdot \partial \|x\|\right) - \eta(N)}{2} \\ &= \mathbb{E}\left[\frac{1}{2}\text{dist}\left(g, \left(\mathbb{E}[t_g] + \frac{2}{\|w_0\|_2}\right) \cdot \partial \|x\|\right)\right] \\ &\quad - \mathbb{E}\left[\frac{1}{2}\text{dist}(g, N)\right] \\ &\leq 2 + \sqrt{2\log\left(\frac{1}{1-2e^{-2^2/2}}\right)}. \end{aligned}$$

Since $\eta(N) = \omega(T \cap \mathbb{B}^p)$, we see that

$$\begin{aligned} & \eta\left(\left(\mathbb{E}[t_g] + \frac{2}{\|w_0\|_2}\right) \cdot \partial \|x\|\right) \\ &\leq \omega(T \cap \mathbb{B}^p) + 2 \cdot \left(2 + \sqrt{2\log\left(\frac{1}{1-2e^{-2^2/2}}\right)}\right) \\ &\leq \omega(T \cap \mathbb{B}^p) + 6. \end{aligned}$$

Since $\mu_n > \sqrt{n-1/2} \geq \sqrt{n} - (1-1/\sqrt{2})$ for all $n \geq 1$, it is sufficient to have

$$\begin{aligned} & \sqrt{n}(1-\epsilon) \geq \\ & 2\eta\left(t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}}\right) + \eta\left(\mu_k(1-\gamma) \cdot \partial \|v^*\|_{\ell_1/\ell_2}\right) \\ & \quad + \sqrt{2\log(1/\beta)} + 3\sqrt{2\pi} + 1 + \frac{1}{\sqrt{2\pi}}. \end{aligned}$$

Table II bounds $\eta\left(\mu_k(1-\gamma) \cdot \partial \|v^*\|_{\ell_1/\ell_2}\right)$ by $\sqrt{n}\sqrt{1-\mu_k^2/k(1-\gamma)^2}$, and hence the condition that \sqrt{n} is at least as large as

$$\frac{2\eta\left(t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}}\right) + \sqrt{2\log(1/\beta)} + 3\sqrt{2\pi} + 1 + \frac{1}{\sqrt{2\pi}}}{\left(1 - \sqrt{1 - \mu_k^2/k(1-\gamma)^2} - \epsilon\right)_+}$$

is sufficient for recovery. To obtain the final form of our statement, we note that

$$\begin{aligned} 1 - \sqrt{1 - \mu_k^2/k(1-\gamma)^2} &= \frac{\mu_k^2/k(1-\gamma)^2}{1 + \sqrt{1 - \mu_k^2/k(1-\gamma)^2}} \\ &\geq \frac{\mu_k^2/k}{1 + \sqrt{1 - \mu_k^2/k}}(1-\gamma)^2 = \left(1 - \sqrt{1 - \mu_k^2/k}\right)(1-\gamma)^2. \end{aligned}$$

■

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),

$$\begin{aligned} & \mathbb{E}\left[\text{dist}(g, t_{\text{sig}} \cdot \partial \|x^*\|_1)^2\right] \\ & \leq s_{\text{sig}}(1 + t_{\text{sig}}^2) + \frac{2(p - s_{\text{sig}})}{\sqrt{2\pi}}. \\ & \quad \left((1 + t_{\text{sig}}^2) \int_{t_{\text{sig}}}^{\infty} \exp(-c^2/2) \, dc - t_{\text{sig}} \exp(-t_{\text{sig}}^2/2) \right) \end{aligned}$$

$$\begin{aligned} \mu_n - \sqrt{n}\epsilon &\geq \\ 2\eta\left(t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}}\right) &+ \eta\left(\mu_k(1-\gamma) \cdot \partial \|v^*\|_{\ell_1/\ell_2}\right) \\ &+ \sqrt{2\log(1/\beta)} + 3\sqrt{2\pi} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2\pi}}. \end{aligned}$$

Bounding the integral as in [14, Proof of Prop. 3.10],

$$\begin{aligned}
&\leq s_{\text{sig}}(1 + t_{\text{sig}}^2) + \frac{2(p - s_{\text{sig}})}{\sqrt{2\pi}} \cdot \frac{1}{t_{\text{sig}}} \exp(-t_{\text{sig}}^2/2) \\
&\leq s_{\text{sig}} \left(1 + 2 \log \left(\frac{p+n}{s_{\text{sig}} + s_{\text{cor}}} \right) \right) \\
&\quad + \frac{2(p - s_{\text{sig}} + n - s_{\text{cor}})}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2 \log \left(\frac{p+n}{s_{\text{sig}} + s_{\text{cor}}} \right)}} \\
&\quad \cdot \frac{s_{\text{sig}} + s_{\text{cor}}}{p+n} \\
&\leq s_{\text{sig}} \left(1 + 2 \log \left(\frac{p+n}{s_{\text{sig}} + s_{\text{cor}}} \right) \right) + \frac{1}{2}(s_{\text{sig}} + s_{\text{cor}}) \\
&\leq (s_{\text{sig}} + s_{\text{cor}}) \cdot \left(2 \log \left(\frac{p+n}{s_{\text{sig}} + s_{\text{cor}}} \right) + \frac{3}{2} \right),
\end{aligned}$$

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

$$\begin{aligned}
&2\eta(t_{\text{sig}} \cdot \partial \|x^*\|_1) + \eta(t_{\text{cor}} \cdot \partial \|v^*\|_1) \\
&\leq 3\sqrt{(s_{\text{sig}} + s_{\text{cor}}) \cdot \left(2 \log \left(\frac{p+n}{s_{\text{sig}} + s_{\text{cor}}} \right) + \frac{3}{2} \right)}.
\end{aligned}$$

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_γ be defined as in (20). If x^* and v^* have at most s_γ and $n \cdot \gamma$ nonzero entries, respectively, and (\hat{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 having \sqrt{n} at least as large as

$$2\sqrt{2s_\gamma \log(p/s_\gamma) + \frac{3}{2}s_\gamma} + \sqrt{2 \log(1/0.1)} + 3\sqrt{2\pi} + 1 + \frac{1}{\sqrt{2\pi}} \left(\alpha_1 (1 - \gamma)^2 - \epsilon \right)_+$$

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

$$\begin{aligned}
&\sqrt{2 \log(10)} + 3\sqrt{2\pi} + 1 + \frac{1}{\sqrt{2\pi}} \leq 11.1 \leq \\
&10\sqrt{2s_\gamma \log(p/s_\gamma) + \frac{3}{2}s_\gamma}
\end{aligned}$$

and hence

$$\sqrt{n} \geq \frac{12\sqrt{2s_\gamma \log(p/s_\gamma) + \frac{3}{2}s_\gamma}}{\left(\alpha_1 (1 - \gamma)^2 - \epsilon \right)_+}$$

is sufficient for recovery. One can check that the chosen s_γ always satisfies this bound. ■

APPENDIX C PROOFS OF MAIN RESULTS

A. Main idea

Under the assumptions of Section I-A, we have an ℓ_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

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

and

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

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

$$\begin{aligned}
T_{\text{joint}}^\lambda &= \{(a, b) \in \mathbb{R}^p \times \mathbb{R}^n : \\
&\text{Pen}_\lambda(x^* + t \cdot a, v^* + t \cdot b) \leq \text{Pen}_\lambda(x^*, v^*) \text{ for some } t > 0\}.
\end{aligned}$$

By definition,

$$T_{\text{sig}} \times T_{\text{cor}} \subsetneq T_{\text{joint}}^\lambda \subsetneq (T_{\text{sig}} \times \mathbb{R}^n) \cup (\mathbb{R}^p \times T_{\text{cor}}).$$

Constrained problem: Consider either version of the constrained estimation problem:

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

or

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

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

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

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,

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

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

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

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

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

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 5. Take any $\Omega \subset \mathbb{B}^p \times \mathbb{R}^n$. 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

$$\sqrt{n} \cdot \mathbb{E} \left[\min_{(a,b) \in \Omega} \|\Phi a + b\|_2 \right] \geq \\ \mathbb{E} \left[\left(\min_{(a,b) \in \Omega} \|h \cdot a + \sqrt{n} \cdot b\|_2 + \langle g, a \rangle \right)_+ \right] - \frac{1}{\sqrt{2\pi}}. \quad (29)$$

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 [37, (2.8)], we see that with probability at least $1 - \beta$,

$$\sqrt{n} \cdot \min_{(a,b) \in \Omega} \|\Phi a + b\|_2 \geq \\ \sqrt{n} \cdot \mathbb{E} \left[\min_{(a,b) \in \Omega} \|\Phi a + b\|_2 \right] - \sqrt{2 \log(1/\beta)}, \quad (30)$$

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 (29) for $\Omega = (T_{\text{sig}} \times T_{\text{cor}}) \cap \mathbb{S}^{p+n-1}$ and for $\Omega = T_{\text{joint}}^\lambda \cap \mathbb{S}^{p+n-1}$, to handle the constrained and penalized problems, respectively. Combining these results with the probability bound in (30), 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

$$\omega^2 = \omega^2(T_{\text{sig}} \cap \mathbb{B}^p) \text{ and } \omega'^2 = \omega^2(T_{\text{cor}} \cap \mathbb{B}^n)$$

be shorthand for the two relevant Gaussian squared complexities.

Theorem 3.

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

Proof: We will combine Lemma 5 with a lower bound for

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

Let

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

be the ‘observed’ Gaussian complexities, with expectations ω and ω' , 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

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

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

$$\begin{aligned} & \min_{(a,b) \in (T_{\text{sig}} \times T_{\text{cor}}) \cap \mathbb{S}^{p+n-1}} \|h \cdot \|a\|_2 + \sqrt{n} \cdot b\|_2 + \langle g, a \rangle \\ &= \min_{\substack{c \in [0,1] \\ a \in T_{\text{sig}} \cap \mathbb{S}^{p-1} \\ b \in T_{\text{cor}} \cap \mathbb{S}^{n-1}}} \|h \cdot c + b \cdot \sqrt{n} \cdot \sqrt{1-c^2}\|_2 + c \cdot \langle g, a \rangle \\ &\geq \min_{\substack{c \in [0,1] \\ a \in T_{\text{sig}} \cap \mathbb{S}^{p-1} \\ b \in T_{\text{cor}} \cap \mathbb{S}^{n-1}}} \left\{ \|h \cdot c + b \cdot \|h\|_2 \sqrt{1-c^2}\|_2 \right. \\ &\quad \left. + c \cdot \langle g, a \rangle - |\sqrt{n} - \|h\|_2| \right\} \\ &= \min_{\substack{c \in [0,1] \\ a \in T_{\text{sig}} \cap \mathbb{S}^{p-1} \\ b \in T_{\text{cor}} \cap \mathbb{S}^{n-1}}} \left\{ \sqrt{\|h\|_2^2 + 2 \|h\|_2 \cdot \langle h, b \rangle \cdot c \sqrt{1-c^2}} \right. \\ &\quad \left. + c \cdot \langle g, a \rangle - |\sqrt{n} - \|h\|_2| \right\} \\ &\geq \min_{\substack{c \in [0,1] \\ a \in T_{\text{sig}} \cap \mathbb{B}^p \\ b \in T_{\text{cor}} \cap \mathbb{B}^n}} \left\{ \sqrt{\|h\|_2^2 + 2 \|h\|_2 \cdot \langle h, b \rangle \cdot c \sqrt{1-c^2}} \right. \\ &\quad \left. + c \cdot \langle g, a \rangle - |\sqrt{n} - \|h\|_2| \right\} \end{aligned}$$

Minimizing over a and b for a fixed c ,

$$\begin{aligned} &= \min_{c \in [0,1]} \left\{ \sqrt{\|h\|_2^2 - 2 \|h\|_2 \cdot \hat{\omega}' \cdot c \sqrt{1-c^2}} \right. \\ &\quad \left. - c \cdot \hat{\omega} - |\sqrt{n} - \|h\|_2| \right\} \\ &\geq \|h\|_2 - \sqrt{\hat{\omega}^2 + \hat{\omega}'^2} - |\sqrt{n} - \|h\|_2|, \end{aligned}$$

where for the last step, we use the fact that $\sqrt{\hat{\omega}^2 + \hat{\omega}'^2} \leq \|h\|_2$, and apply the following lemma (proved in Appendix D):

Lemma 6. For any $a, b, C \geq 0$ such that $a^2 + b^2 \leq C^2$,

$$\inf_{u \in [0,1]} \sqrt{C^2 - 2Cu\sqrt{1-u^2} \cdot a} - u \cdot b \geq C - \sqrt{a^2 + b^2}.$$

In either case, then, we have proved that

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

Finally, we take expectations. To bound the first subtracted term, we have

$$\mathbb{E} \left[\sqrt{\hat{\omega}^2 + \hat{\omega}'^2} \right] \leq \sqrt{\mathbb{E} [\hat{\omega}^2 + \hat{\omega}'^2]} = \sqrt{\omega^2 + \omega'^2},$$

by definition of the Gaussian squared complexity. To bound the second subtracted term, we use the following lemma (proved in Appendix D):

Lemma 7. For $h \sim N(0, I_n)$, $\mathbb{E} [\|\|h\|_2 - \sqrt{n}\|] \leq \frac{1}{\sqrt{2}}$.

Finally, applying Lemma 5, this gives us the desired lower bound. ■

D. Lower bound: penalized setting

In this section we derive a lower bound on

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

Theorem 4. Let $\lambda = t_{\text{cor}}/t_{\text{sig}}$ for parameters $t_{\text{sig}}, t_{\text{cor}} \geq 0$. Then

$$\begin{aligned} \sqrt{n} \mathbb{E} \left[\min_{(a,b) \in T_{\text{joint}}^\lambda \cap \mathbb{S}^{p+n-1}} \|\Phi a + b\|_2 \right] &\geq \\ \mu_n - 2\eta \left(t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}} \right) - \eta(t_{\text{cor}} \cdot \partial \|v^*\|_{\text{cor}}) &- 3\sqrt{2\pi} - \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2\pi}}. \end{aligned}$$

Proof:

We will combine Lemma 5 with a lower bound for

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

For $\sigma = +1$ and $\sigma = -1$, define

$$d_{\text{sig}}^{(\sigma)} := \text{dist} \left(\sigma \cdot g, t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}} \right)$$

and

$$d_{\text{cor}}^{(\sigma)} := \text{dist} \left(\sigma \cdot h, t_{\text{cor}} \cdot \partial \|v^*\|_{\text{cor}} \right),$$

and let $w_{\text{sig}}^{(\sigma)} \in \partial \|x^*\|_{\text{sig}}$ and $w_{\text{cor}}^{(\sigma)} \in \partial \|v^*\|_{\text{cor}}$ be such that

$$\left\| \sigma \cdot g - t_{\text{sig}} \cdot w_{\text{sig}}^{(\sigma)} \right\|_2 = d_{\text{sig}}^{(\sigma)}$$

and

$$\left\| \sigma \cdot h - t_{\text{cor}} \cdot w_{\text{cor}}^{(\sigma)} \right\|_2 = d_{\text{cor}}^{(\sigma)}.$$

We know that for each choice of signs $\sigma_1, \sigma_2 \in \{\pm 1\}$,

$$\begin{aligned} 0 &\geq \left(\|x^* + a\|_{\text{sig}} + \lambda \|v^* + b\|_{\text{cor}} \right) \\ &\quad - \left(\|x^*\|_{\text{sig}} + \lambda \|v^*\|_{\text{cor}} \right) \\ &\geq \langle w_{\text{sig}}^{(\sigma_1)}, a \rangle + \lambda \langle w_{\text{cor}}^{(\sigma_2)}, b \rangle \\ &= \frac{\langle \sigma_1 \cdot g, a \rangle - \langle \sigma_1 \cdot g - t_{\text{sig}} \cdot w_{\text{sig}}^{(\sigma_1)}, a \rangle}{t_{\text{sig}}} \\ &\quad + \lambda \frac{\langle \sigma_2 \cdot h, b \rangle - \langle \sigma_2 \cdot h - t_{\text{cor}} \cdot w_{\text{cor}}^{(\sigma_2)}, b \rangle}{t_{\text{cor}}} \\ &\geq \frac{\langle \sigma_1 \cdot g, a \rangle - d_{\text{sig}}^{(\sigma_1)} \|a\|_2 + \langle \sigma_2 \cdot h, b \rangle - d_{\text{cor}}^{(\sigma_2)} \|b\|_2}{t_{\text{sig}}}, \end{aligned}$$

and so

$$d_{\text{sig}}^{(\sigma_1)} \|a\|_2 + d_{\text{cor}}^{(\sigma_2)} \|b\|_2 \geq \langle \sigma_1 \cdot g, a \rangle + \langle \sigma_2 \cdot h, b \rangle.$$

Maximizing the right-hand side over the signs,

$$\begin{aligned} &d_{\text{sig}} \|a\|_2 + d_{\text{cor}} \|b\|_2 \\ &\geq |\langle g, a \rangle| + |\langle h, b \rangle| \geq (\langle -g, a \rangle)_+ + (\langle -h, b \rangle)_+, \quad (31) \end{aligned}$$

where

$$d_{\text{sig}} = \max\{d_{\text{sig}}^{(+1)}, d_{\text{sig}}^{(-1)}\}$$

and

$$d_{\text{cor}} = \max\{d_{\text{cor}}^{(+1)}, d_{\text{cor}}^{(-1)}\}.$$

Next, since $\|b\|_2 \leq 1$ for all $(a, b) \in T_{\text{joint}}^\lambda \cap \mathbb{S}^{p+n-1}$, we have the following (here (a, b) is assumed to be in $T_{\text{joint}}^\lambda \cap \mathbb{S}^{p+n-1}$)

$$\begin{aligned} &\min_{(a,b)} \|h \cdot \|a\|_2 + \sqrt{n} \cdot b\|_2 + \langle g, a \rangle \\ &\geq \min_{(a,b)} \|h \cdot \|a\|_2 + \|h\|_2 \cdot b\|_2 + \langle g, a \rangle - |\sqrt{n} - \|h\|_2| \\ &= \min_{(a,b)} \left\{ \sqrt{\|h\|_2^2 (\|a\|_2^2 + \|b\|_2^2) + 2 \|h\|_2 \|a\|_2 \cdot \langle h, b \rangle} \right. \\ &\quad \left. + \langle g, a \rangle - |\sqrt{n} - \|h\|_2| \right\} \\ &= \min_{(a,b)} \left\{ \|h\|_2 \sqrt{1 + 2 \|a\|_2 \cdot \langle \frac{h}{\|h\|_2}, b \rangle} \right. \\ &\quad \left. + \langle g, a \rangle - |\sqrt{n} - \|h\|_2| \right\} \\ &\geq \min_{(a,b)} \left\{ \|h\|_2 \sqrt{1 - 2 \|a\|_2 \cdot \left(\langle \frac{-h}{\|h\|_2}, b \rangle \right)_+} \right. \\ &\quad \left. - \langle -g, a \rangle - |\sqrt{n} - \|h\|_2| \right\} \end{aligned}$$

Since $\sqrt{1-x} \geq 1-x$ for all $x \in [0, 1]$,

$$\begin{aligned} &\geq \min_{(a,b)} \left\{ \|h\|_2 \left(1 - 2 \|a\|_2 \cdot \left(\langle \frac{-h}{\|h\|_2}, b \rangle \right)_+ \right. \right. \\ &\quad \left. \left. - \langle -g, a \rangle - |\sqrt{n} - \|h\|_2| \right) \right\} \\ &= \min_{(a,b)} \left\{ \|h\|_2 - 2 \|a\|_2 \cdot (\langle -h, b \rangle)_+ \right. \\ &\quad \left. - \langle -g, a \rangle - |\sqrt{n} - \|h\|_2| \right\} \\ &\geq \min_{(a,b)} \left\{ \|h\|_2 - 2 \|a\|_2 \cdot (\langle -h, b \rangle)_+ \right. \\ &\quad \left. - (\langle -g, a \rangle)_+ - |\sqrt{n} - \|h\|_2| \right\} \\ &\text{Applying (31),} \\ &\geq \min_{(a,b)} \left\{ \|h\|_2 - \max\{2 \|a\|_2, 1\} \cdot \right. \\ &\quad (d_{\text{sig}} \|a\|_2 + d_{\text{cor}} \|b\|_2 - (\langle -g, a \rangle)_+) \\ &\quad \left. - (\langle -g, a \rangle)_+ - |\sqrt{n} - \|h\|_2| \right\} \\ &\geq \min_{(a,b)} \left\{ \|h\|_2 - \max\{2 \|a\|_2, 1\} \cdot \right. \\ &\quad (d_{\text{sig}} \|a\|_2 + d_{\text{cor}} \|b\|_2) - |\sqrt{n} - \|h\|_2| \Big\} \\ &= \min_{(a,b)} \left\{ \|h\|_2 - \max\{2 \|a\|_2^2, \|a\|_2\} \cdot d_{\text{sig}} \right. \\ &\quad \left. - \max\{2 \|a\|_2 \|b\|_2, \|b\|_2\} \cdot d_{\text{cor}} - |\sqrt{n} - \|h\|_2| \right\} \end{aligned}$$

Since $2 \|a\|_2 \cdot \|b\|_2 \leq 1$ (because $\|a\|_2^2 + \|b\|_2^2 = 1$),

$$\geq \|h\|_2 - 2d_{\text{sig}} - d_{\text{cor}} - |\sqrt{n} - \|h\|_2|.$$

Taking expectations, we have

$$\begin{aligned} &\mathbb{E} \left[\min_{(a,b) \in T_{\text{joint}}^\lambda \cap \mathbb{S}^{p+n-1}} \|h \cdot \|a\|_2 + \sqrt{n} \cdot b\|_2 + \langle g, a \rangle \right] \\ &\geq \mu_n - 2\eta \left(t_{\text{sig}} \cdot \partial \|x^*\|_{\text{sig}} \right) - \eta \left(t_{\text{cor}} \cdot \partial \|v^*\|_{\text{cor}} \right) \\ &\quad - 3\sqrt{2\pi} - \frac{1}{\sqrt{2}}, \end{aligned}$$

where we use Lemma 7 to bound $\mathbb{E}[\|\sqrt{n} - \|h\|_2\|]$, and to bound $\mathbb{E}[d_{\text{sig}}]$ and $\mathbb{E}[d_{\text{cor}}]$ we use the following lemma (proved in Appendix D):

Lemma 8. For any set $A \subset \mathbb{R}^p$, for $g \in \mathbb{R}^p$ with i.i.d. standard normal entries,

$$\mathbb{E} [\max\{\text{dist}(g, A), \text{dist}(-g, A)\}] \leq \eta(A) + \sqrt{2\pi}.$$

■

APPENDIX D
PROOFS FOR LEMMAS

Lemma 1. 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 $k = \lceil \frac{4}{9} m_2 \rceil \leq m_2$. Let Υ be a $m_1 \times k$ submatrix of Γ ; then Υ 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 $\sigma_1 \geq \dots \geq \sigma_k \geq 0$ be the singular values of Υ . Davidson and Szarek [38, 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 $i = 1, \dots, k$. In the simple case that $\frac{4}{9} m_2$ is an integer, we then have

$$\begin{aligned} \mathbb{E} [\sigma_1 + \dots + \sigma_k] &\geq k(\sqrt{m_1} - \sqrt{k}) \\ &= \frac{4}{9} m_2 \left(\sqrt{m_1} - \frac{2}{3} \sqrt{m_2} \right) \\ &\geq \frac{4}{9} m_2 \left(\sqrt{m_1} - \frac{2}{3} \sqrt{m_1} \right) = \frac{4}{27} m_2 \sqrt{m_1} . \end{aligned}$$

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 Υ . 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. ■

Lemma 2. For any norm $\|\cdot\|$ on \mathbb{R}^p with dual norm $\|\cdot\|^*$,

$$\max_{x \in \mathbb{R}^p} \frac{\|x\|}{\|x\|_2} = \max_{x \in \mathbb{R}^p} \frac{\|x\|_2}{\|x\|^*} .$$

Proof: First, for any x ,

$$\|x\|_2^2 = \langle x, x \rangle \leq \|x\| \cdot \|x\|^* \Rightarrow \frac{\|x\|}{\|x\|_2} \geq \frac{\|x\|_2}{\|x\|^*} .$$

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 $\|x\| = \max_{\|w\|^*=1} \langle x, w \rangle$. Choose some w with $\|w\|^* = 1$ so that this maximum is attained. Then $\langle x, w \rangle \leq \|x\|_2 \cdot \|w\|_2$, and so

$$\frac{\|x\|}{\|x\|_2} = \frac{\langle x, w \rangle}{\|x\|_2} \leq \|w\|_2 \leq \max_{x' \in \mathbb{R}^p} \frac{\|x'\|_2}{\|x'\|^*} .$$

Since this is true for any x , this proves the desired bound. ■

Lemma 3. Suppose that, for $x \neq 0$, $\partial \|x\|$ satisfies (27). Let t_g be defined as in (28). Then $g \mapsto t_g$ is a $\frac{1}{\|w_0\|_2}$ -Lipschitz function of g .

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

$$\text{dist}(g, t_g \cdot \partial \|x\|) = \|g - t_g \cdot w\|_2$$

and

$$\text{dist}(g', t_{g'} \cdot \partial \|x\|) = \|g' - t_{g'} \cdot w'\|_2 .$$

Recalling that

$$N = \cup_{t \geq 0} (t \cdot \partial \|x\|) ,$$

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 (27): we have

$$\begin{aligned} &\|t_g \cdot w - t_{g'} \cdot w'\|_2 \\ &= \|(t_g - t_{g'}) \cdot w_0 + [t_g \cdot (w - w_0) - t_{g'} \cdot (w' - w_0)]\|_2 \\ &\geq |t_g - t_{g'}| \cdot \|w_0\|_2 , \end{aligned}$$

because $\langle w - w_0, w_0 \rangle = \langle w' - w_0, w_0 \rangle = 0$. ■

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

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

Proof: Choose any ϵ satisfying $0 < \epsilon < p_0$. By Ledoux [37, (2.8)], we know that

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

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

Lemma 5. Take any $\Omega \subset \mathbb{B}^p \times \mathbb{R}^n$. 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

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

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

$$X_{(a,b),w} = \sqrt{n} \cdot \langle \Phi a, w \rangle + \nu \|a\|_2$$

and

$$Y_{(a,b),w} = \langle h, w \rangle \|a\|_2 + \langle g, a \rangle,$$

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

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

and

$$\begin{aligned} & \mathbb{E}[X_{(a,b),w} X_{(a',b'),w'}] - \mathbb{E}[Y_{(a,b),w} Y_{(a',b'),w'}] \\ &= [\langle a, a' \rangle \langle w, w' \rangle + \|a\|_2 \|a'\|_2] \\ &\quad - [\|a\|_2 \|a'\|_2 \langle w, w' \rangle + \langle a, a' \rangle] \\ &= (\|a\|_2 \|a'\|_2 - \langle a, a' \rangle) \cdot (1 - \langle w, w' \rangle) \geq 0, \end{aligned}$$

with equality if $(a, b) = (a', b')$. Therefore, applying Theorem 1.1 of Gordon [39] to these Gaussian processes, we know that for any scalars $\{c_{(a,b),w}\}$,

$$\begin{aligned} & \mathbb{P}\{\cap_{(a,b)\in\Omega} \cup_{w\in\mathbb{B}^n} [X_{(a,b),w} \geq c_{(a,b),w}]\} \\ & \geq \mathbb{P}\{\cap_{(a,b)\in\Omega} \cup_{w\in\mathbb{B}^n} [Y_{(a,b),w} \geq c_{(a,b),w}]\}. \end{aligned}$$

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:

$$\begin{aligned} & \cap_{(a,b)\in\Omega} \cup_{w\in\mathbb{B}^n} [X_{(a,b),w} \geq c_{(a,b),w}] \\ &= \left\{ \min_{(a,b)\in\Omega} \max_{w\in\mathbb{B}^n} X_{(a,b),w} + \sqrt{n} \cdot \langle b, w \rangle \geq C \right\} \end{aligned}$$

and same for the Y process. This proves that

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

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

$$\begin{aligned} & \mathbb{P}\left\{ \min_{(a,b)\in\Omega} \sqrt{n} \cdot \|\Phi a + b\|_2 + \nu \cdot \|a\|_2 \geq C \right\} \\ & \geq \mathbb{P}\left\{ \min_{(a,b)\in\Omega} \|h \cdot a\|_2 + \sqrt{n} \cdot \|b\|_2 + \langle g, a \rangle \geq C \right\}. \end{aligned}$$

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

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

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

$$\begin{aligned} & \left(\min_{(a,b)\in\Omega} \sqrt{n} \cdot \|\Phi a + b\|_2 + \nu \cdot \|a\|_2\right)_+ \\ & \leq \sqrt{n} \cdot \min_{(a,b)\in\Omega} \|\Phi a + b\|_2 + (\nu)_+, \end{aligned}$$

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

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

\blacksquare

Lemma 6. For any $a, b, C \geq 0$ such that $a^2 + b^2 \leq C^2$,

$$\inf_{u\in[0,1]} \sqrt{C^2 - 2Cu\sqrt{1-u^2} \cdot a - u \cdot b} \geq C - \sqrt{a^2 + b^2}.$$

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

$$\begin{aligned} & (u \cdot a - \sqrt{1-u^2})^2 \geq 0 \\ & \Rightarrow u^2 \cdot a^2 - 2u\sqrt{1-u^2} \cdot a + (1-u^2) \geq 0 \\ & \Rightarrow 1 - 2u\sqrt{1-u^2} \cdot a \geq u^2 \cdot (1-a^2) = u^2 \cdot b^2 \\ & \Rightarrow \sqrt{1 - 2u\sqrt{1-u^2} \cdot a} - u \cdot b \geq 0. \end{aligned}$$

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

$$\begin{aligned} & \sqrt{1 - 2u\sqrt{1-u^2} \cdot a} - u \cdot b \\ &= \sqrt{1 - 2u\sqrt{1-u^2} \cdot a' \cdot c} - u \cdot b' \cdot c \\ &= \left[\sqrt{1 - 2u\sqrt{1-u^2} \cdot a' \cdot c} - c\sqrt{1 - 2u\sqrt{1-u^2} \cdot a'} \right] \\ &\quad + c \cdot \left[\sqrt{1 - 2u\sqrt{1-u^2} \cdot a'} - u \cdot b' \right] \end{aligned}$$

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:

$$(1 - \sqrt{1-d})^2 = 2 - 2\sqrt{1-d} - d$$

Since $c^2 \leq c$,

$$\Rightarrow c^2(1 - \sqrt{1-d})^2 \leq c(2(1 - \sqrt{1-d}) - d)$$

By rearranging terms and adding 1 to both sides,

$$\Rightarrow 1 + c^2(1 - \sqrt{1-d})^2 - 2c(1 - \sqrt{1-d}) \leq 1 - dc$$

By taking the square root,

$$\begin{aligned} \Rightarrow 1 - c(1 - \sqrt{1-d}) &\leq \sqrt{1-dc} \\ \Rightarrow 1 - c &\leq \sqrt{1-dc} - c\sqrt{1-d}. \end{aligned}$$

■

Lemma 7. For $h \sim N(0, I_n)$, $\mathbb{E}[\|\|h\|_2 - \sqrt{n}\|] \leq \frac{1}{\sqrt{2}}$.

Proof: We use the fact that $\mathbb{E}[\|h\|_2] = \mu_n$. We have

$$\begin{aligned} \mathbb{E}[\|\|h\|_2 - \sqrt{n}\|]^2 &\leq \mathbb{E}[(\|h\|_2 - \sqrt{n})^2] \\ &= \mathbb{E}[(\|h\|_2 - \mu_n)^2] + (\mu_n - \sqrt{n})^2 \\ &= (n - 2\mu_n^2 + \mu_n^2) + (\mu_n^2 - 2\mu_n\sqrt{n} + n) \\ &= 2\sqrt{n}(\sqrt{n} - \mu_n) \\ &\leq \frac{1}{2}, \end{aligned}$$

where the last inequality holds for all $n \geq 1$.

■

Lemma 8. For any set $A \subset \mathbb{R}^p$, for $g \in \mathbb{R}^p$ with i.i.d. standard normal entries,

$$\mathbb{E}[\max\{\text{dist}(g, A), \text{dist}(-g, A)\}] \leq \eta(A) + \sqrt{2\pi}.$$

Proof: Let $\alpha = \mathbb{E}[\text{dist}(g, A)] = \mathbb{E}[\text{dist}(-g, A)] \leq \sqrt{\mathbb{E}[\text{dist}(g, A)^2]} = \eta(A)$. Since $g \mapsto \text{dist}(g, A)$ is 1-Lipschitz,

$$\begin{aligned} \mathbb{E}[(\text{dist}(g, A) - \alpha)_+] &= \\ &\int_0^\infty \mathbb{P}\{\text{dist}(g, A) > \alpha + c\} \, dc \\ &= \int_0^\infty e^{-c^2/2} \, dc \leq \sqrt{\pi/2}. \end{aligned}$$

Then

$$\begin{aligned} \mathbb{E}[\max\{\text{dist}(g, A), \text{dist}(-g, A)\}] &= \alpha + \mathbb{E}[\max\{(\text{dist}(g, A) - \alpha), (\text{dist}(-g, A) - \alpha)\}] \\ &\leq \alpha + \mathbb{E}[(\text{dist}(g, A) - \alpha)_+] + \mathbb{E}[(\text{dist}(-g, A) - \alpha)_+] \\ &\leq \alpha + \sqrt{2\pi} \leq \eta(A) + \sqrt{2\pi}. \end{aligned}$$

■

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,” *Constructive Approximation*, vol. 37, no. 1, pp. 73–99, 2013.
- [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 ℓ_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. 717–772, 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. Parrilo, and A. Willsky, “The convex geometry of linear inverse problems,” *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

- [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. pp. 402–404, 1962.
- [17] J. Penot and C. Zalinescu, "Elements of quasi-convex subdifferential calculus," *Journal of Convex Analysis*, vol. 7, no. 2, pp. 243–270, 2000.
- [18] M. Stojnic, "Various thresholds for ℓ_1 -optimization in compressed sensing," vol. abs/0907.3666, 2009.
- [19] 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.
- [20] 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.
- [21] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: A geometric theory of phase transitions in convex optimization," March 2013, available at arXiv:1303.6672.
- [22] 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.
- [23] ———, "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.
- [24] Y. Eldar, P. Kuppinger, and H. Bölskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
- [25] F. Parvaresh and B. Hassibi, "Explicit measurements with almost optimal thresholds for compressed sensing," in *ICASSP*. IEEE, 2008, pp. 3853–3856.
- [26] E. J. Candès and B. Recht, "Simple bounds for low-complexity model reconstruction," *Mathematical Programming Series A*, 2012.
- [27] 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.
- [28] J. Wright and Y. Ma, "Dense error correction via ℓ_1 -minimization," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3540–3560, 2010.
- [29] N. Nguyen and T. Tran, "Exact recoverability from dense corrupted observations via ℓ_1 minimization," *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 2017–2035, 2013.
- [30] G. Pope, A. Bracher, and C. Studer, "Probabilistic recovery guarantees for sparsely corrupted signals," *IEEE Transactions on Information Theory*, 2013.
- [31] C. Hegde and R. G. Baraniuk, "Signal recovery on incoherent manifolds," *IEEE Transactions on Information Theory*, vol. 58, no. 12, pp. 7204–7214, 2012.
- [32] CVX Research Inc., "CVX: Matlab software for disciplined convex programming, version 1.22," <http://cvxr.com/cvx>, Jan. 2013.
- [33] 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.
- [34] J. Friedman, T. Hastie, and R. Tibshirani, "A note on the group lasso and a sparse group lasso," 2010, available at arXiv:1001.0736.
- [35] 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.
- [36] G. Watson, "Characterization of the subdifferential of some matrix norms," *Linear Algebra and its Applications*, vol. 170, pp. 33–45, 1992.
- [37] M. Ledoux, "Isoperimetry and Gaussian analysis," *Lectures on probability theory and statistics*, pp. 165–294, 1996.
- [38] 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.
- [39] Y. Gordon, "Some inequalities for Gaussian processes and applications," *Israel Journal of Mathematics*, vol. 50, no. 4, pp. 265–289, 1985.