

Compressed Sensing With Side Information: Geometrical Interpretation and Performance Bounds

João F. C. Mota, Nikos Deligiannis, and Miguel R. D. Rodrigues
Electronic & Electrical Engineering Department at University College London, UK

Abstract—We address the problem of Compressed Sensing (CS) with side information. Namely, when reconstructing a target CS signal, we assume access to a similar signal. This additional knowledge, the *side information*, is integrated into CS via $\ell_{1-\ell_1}$ and $\ell_{1-\ell_2}$ minimization. We then provide lower bounds on the number of measurements that these problems require for successful reconstruction of the target signal. If the side information has good quality, the number of measurements is significantly reduced via $\ell_{1-\ell_1}$ minimization, but not so much via $\ell_{1-\ell_2}$ minimization. We provide geometrical interpretations and experimental results illustrating our findings.

Index Terms—Compressed sensing, basis pursuit, $\ell_{1-\ell_1}$ minimization, $\ell_{1-\ell_2}$ minimization, Gaussian width.

I. INTRODUCTION

Consider an unknown signal $x^* \in \mathbb{R}^n$ that is s -sparse, i.e., at most s entries are nonzero. Assume we take $m < n$ linear measurements of x^* : $y = Ax^*$, where $A \in \mathbb{R}^{m \times n}$. Compressed Sensing (CS) [1] answers two fundamental questions: *How to reconstruct x^* from y ? And how many measurements m do we need for successful reconstruction?* If A satisfies the RIP [2] or a null-space property [3], CS establishes that x^* can be reconstructed by solving *basis pursuit* [4]:

$$\begin{aligned} & \underset{x}{\text{minimize}} && \|x\|_1 \\ & \text{subject to} && Ax = y. \end{aligned} \quad (1)$$

In particular, when the entries of A are i.i.d. Gaussian, then $m > 2s \log(n/s) + (7/5)s$ measurements guarantee that x^* is the unique solution of (1) with high probability [3], a bound known to be tight for Gaussian matrices [5].

CS with side information. Suppose we have access to *side information*, a vector $w \in \mathbb{R}^n$ that is similar to x^* . This occurs when reconstructing sequences of signals (e.g., video [6] and estimation problems [7]), or when we have access to prior similar signals (e.g., sensor networks [8], multiview cameras [9], and medical imaging [10]). Our goal is to answer the same questions that CS does, but with the additional knowledge of side information: *How to reconstruct x^* from the measurements y and the side information w ? And how many measurements m do we need for successful reconstruction?*

Our strategy. Let $g : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function that models similarity between w and x^* : the smaller $g(x^* - w)$, the higher the similarity. A natural approach to integrate w into (1) is to

$$\begin{aligned} & \underset{x}{\text{minimize}} && \|x\|_1 + \beta g(x - w) \\ & \text{subject to} && Ax = y, \end{aligned} \quad (2)$$

where $\beta > 0$. We consider two models for g : $g_1(\cdot) := \|\cdot\|_1$ and $g_2(\cdot) = (1/2)\|\cdot\|_2^2$; and refer to (2) with $g = g_1$ as $\ell_{1-\ell_1}$ minimization and to (2) with $g = g_2$ as $\ell_{1-\ell_2}$ minimization. Although instances and variations of (2) with g_1 and g_2 have appeared in the literature (see Related work below), to the best of our knowledge, no CS-like recovery guarantees have ever been provided.

Assuming the entries of A are i.i.d. Gaussian, we compute bounds on the number of measurements above which $\ell_{1-\ell_1}$ and $\ell_{1-\ell_2}$ minimization reconstruct x^* perfectly, with high probability. When the side information is “good enough,” our bound for $\ell_{1-\ell_1}$ minimization is much smaller than the bounds both for $\ell_{1-\ell_2}$ minimization and for classical CS. In addition, our experiments confirm that $\ell_{1-\ell_1}$ minimization requires in general less measurements for successful reconstruction than both $\ell_{1-\ell_2}$ minimization and classical CS. We explain this phenomenon using the underlying geometry of the problem. Proofs of the results presented in this paper can be found in [11]. For succinctness, we consider here only the case $\beta = 1$ in (2), but results for $\beta \neq 1$ can be found in [11].

Related work. Several methods improve the performance of CS by assuming access to side (or prior) information. The majority, however, uses concepts of side information different from ours, for example, estimates on the support of x^* [12], or its probability distribution [13]. The first work using side information in our sense, namely $\ell_{1-\ell_1}$ minimization, appears to be [10]. That work focuses on the application of computed tomography and does not provide either any type of analysis or a comparison with standard CS; see [14] for a recent related approach. In [12], a problem similar to $\ell_{1-\ell_2}$ minimization appears as an extension of the main problem studied in that paper. Although experimental results are presented, no analysis is provided for the $\ell_{1-\ell_2}$ -type problem. Prior work also has considered the Lagrangian version of (2) where there are no constraints, but the extra term $\lambda\|y - Ax\|_2^2$ is added to the objective, with $\lambda > 0$. For example, [7] estimates the state of a dynamical system using the previous instant’s state as side information. The estimation problem is posed as the Lagrangian version of (2) with both g_1 and g_2 . Although the experimental results in [7] indicate that $\ell_{1-\ell_1}$ -type of minimization requires less measurements than $\ell_{1-\ell_2}$, no rationale is given. Our theoretical results and geometrical interpretations explain this phenomenon in the context of (2). Finally, the work in [15] analyzes the performance of a message passing algorithm to solve the Lagrangian version of (2) with g_2 .

II. DEFINITIONS AND GEOMETRICAL INTERPRETATIONS

As mentioned before, [3] establishes tight bounds for CS. The main tool is the concept of *Gaussian width of a cone* $C \subset \mathbb{R}^n$, given by $w(C) := \mathbb{E}_g[\sup_z \{g^\top z : z \in C \cap B_n(0, 1)\}]$, where $g \in \mathbb{R}^n$ has i.i.d. zero-mean, unit variance Gaussian entries, and $\mathbb{E}_g[\cdot]$ is the expected value w.r.t. g . We use $B_n(0, 1) := \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$ to denote the unit ℓ_2 -norm ball in \mathbb{R}^n . The Gaussian width was originally proposed in [16] for measuring the “width” (aperture) of a cone. Related work using this concept includes [5], [17]–[27].

Theorem 1 (Corollary 3.3 in [3]). *Let $A \in \mathbb{R}^{m \times n}$ be a matrix whose entries are i.i.d. zero-mean Gaussian random variables with variance $1/m$, and let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function. Assume $y = Ax^*$ and that $m \geq w(T_f(x^*))^2 + 1$, where $T_f(x^*)$ denotes the tangent cone of f at x^* . Consider*

$$\hat{x} \in \underset{x}{\operatorname{arg\,min}} \quad f(x) \quad (3)$$

$$\text{s.t.} \quad Ax = y.$$

Then, $\hat{x} = x^*$ is the unique solution of (3) with probability at least $1 - \exp(-\frac{1}{2}[w(T_f(x^*)) - \lambda_m]^2)$, where λ_m is the expected length of a zero-mean, unit-variance Gaussian vector in \mathbb{R}^m .

Recall that the tangent cone of a convex function f at a point x^* is $T_f(x^*) := \operatorname{cone}\{x - x^* : f(x) \leq f(x^*)\}$, where $\operatorname{cone} C := \{\alpha c : \alpha \geq 0, c \in C\}$ is the cone generated by the set C . In other words, $T_f(x^*)$ is the cone generated by the *sublevel set* $S_f(x^*) := \{x \in \mathbb{R}^n : f(x) \leq f(x^*)\}$ from the point x^* , i.e., $T_f(x^*) = \operatorname{cone}(S_f(x^*) - x^*)$. Note that (3) becomes (1) when $f(x) = \|x\|_1$ and becomes (2) when $f(x) = \|x\|_1 + g(x, w)$. Note also that $m/\sqrt{m+1} \leq \lambda_m \leq \sqrt{m}$ [3]. Theorem 1 states that (3) recovers x^* with high probability if the number of measurements is larger than the squared Gaussian width of the tangent cone of f at x^* . The work in [3] then establishes:

Proposition 2 (Proposition 3.10 in [3]). *Let $x^* \neq 0$ be an s -sparse vector in \mathbb{R}^n . Then,*

$$w(T_{\|\cdot\|_1}(x^*))^2 \leq 2s \log\left(\frac{n}{s}\right) + \frac{7}{5}s. \quad (4)$$

By upper bounding the squared Gaussian width of $T_{\|\cdot\|_1}(x^*)$, Proposition 2 establishes a lower bound on the number of measurements that (1) requires to recover x^* with high probability. Note that, since $w(T_f(x^*))$ is usually unknown, Theorem 1 is not very informative in practice. Proposition 2 instills it with operational significance by upper bounding $w(T_f(x^*))^2$ in terms of the key signal parameters s and n . Our goal is to do the same for the functions $f_1(x) := \|x\|_1 + \|x - w\|_1$ and $f_2(x) := \|x\|_1 + \frac{1}{2}\|x - w\|_2^2$.

Good and bad components. Naturally, our bounds for ℓ_1 - ℓ_1 and ℓ_1 - ℓ_2 minimization are a function of the “quality” of the side information w . A way to measure the quality of each component of w arises naturally in the proofs of our results, but it can be motivated geometrically, as we do next.

First, recall that the relation $T_f(x^*) = \operatorname{cone}(S_f(x^*) - x^*)$ means that $T_f(x^*)$ is composed of all the half-lines that join x^* to a point of the sublevel set $S_f(x^*)$. Therefore, the width of $T_f(x^*)$ can be estimated by looking at the sublevel set $S_f(x^*)$. Fig. 1 shows the sublevel sets of f_1 and f_2 for $n = 2$. In those plots, x^* is always $x^* = (0, 1)$, and we consider four different w 's: $w^{(a)} = (0, 1.6)$, $w^{(b)} = (0, 1.3)$, $w^{(c)} = (0, 0.5)$, and $w^{(d)} = (0, -0.5)$. In Figs. 1(a) and 1(b) the side information is $w^{(a)}$ and $w^{(b)}$, and in Figs. 1(c) and 1(d) it is $w^{(c)}$ and $w^{(d)}$. Figs. 1(a) and 1(c) show the sublevel sets of f_1 , while Figs. 1(b) and 1(d) show the sublevel sets of f_2 . For reference, we show in all plots the sublevel set of the ℓ_1 -norm ball $S_{\|\cdot\|_1} := B_2(0, \|x^*\|_1)$. To represent all the other sublevel sets, we use the notation $S_{f_i}^{(j)} := \{x : \|x\|_1 + g_i(x - w^{(j)}) \leq \|x^*\|_1 + g_i(x^* - w^{(j)})\}$, for $i = 1, 2$ and $j = a, b, c, d$. For example, the sublevel sets in Fig. 1(a) are the line segments $S_{f_1}^{(a)} = \{(0, x_2) : 0 \leq x_2 \leq 1.6\}$ and $S_{f_1}^{(b)} = \{(0, x_2) : 0 \leq x_2 \leq 1.3\}$. The tangent cone they generate is the line $\{(0, x_2) : x_2 \in \mathbb{R}\}$, which has zero Gaussian width. This means that the nonzero components of $w^{(a)}$ and $w^{(b)}$ do not contribute “any width” to $T_{f_1}(x^*)$. A careful inspection of the remaining figures reveals that the tangent cones in Figs. 1(a) and 1(b) have smaller “geometrical widths” (and thus Gaussian widths) than the cone generated by $S_{\|\cdot\|_1}$. In Figs. 1(c) and 1(d), in contrast, the tangent cones have either the same width as the cone generated by $S_{\|\cdot\|_1}$ (Fig. 1(c)), or larger widths (Fig. 1(d)). Note, in particular that, in Fig. 1(c), $S_{f_1}^{(c)}$, $S_{f_1}^{(d)}$, and $S_{\|\cdot\|_1}$ all generate the same tangent cone. In Fig. 1(d), $S_{f_2}^{(c)}$ and $S_{f_2}^{(d)}$ generate tangent cones with widths larger than the cone generated by $S_{\|\cdot\|_1}$. Since we want small widths, we say that w_2 , the nonzero component of w , is a *good component* in Figs. 1(a) and 1(b) and is a *bad component* in Figs. 1(c) and 1(d). The generic definition is:

Definition 1 (Good and bad components). *Let $x^* \in \mathbb{R}^n$ be the vector to reconstruct and let $w \in \mathbb{R}^n$ be the side information. For $i = 1, \dots, n$, a component w_i is considered good if*

$$x_i^* > 0 \text{ and } x_i^* < w_i \quad \text{or} \quad x_i^* < 0 \text{ and } x_i^* > w_i,$$

and w_i is considered bad if

$$x_i^* > 0 \text{ and } x_i^* > w_i \quad \text{or} \quad x_i^* < 0 \text{ and } x_i^* < w_i.$$

Fig. 1 gives the intuition why ℓ_1 - ℓ_1 minimization requires less measurements than standard CS and ℓ_1 - ℓ_2 minimization: its good components induce less width in $T_{f_1}(x^*)$, and its bad components never induce more width in $T_{f_1}(x^*)$ than the absence of side information. Although Fig. 1 shows the impact only of the components w_i for which $x_i^* \neq 0$ and $x_i^* \neq w_i$, the components for which $w_i = x_i^* \neq 0$ and for which $w_i \neq x_i^* = 0$ also impact the Gaussian width, as shown next.

III. BOUNDS FOR ℓ_1 - ℓ_1 AND ℓ_1 - ℓ_2 MINIMIZATION

To state our results for ℓ_1 - ℓ_1 minimization, we need to define

$$\bar{h} := |\{i : x_i^* > 0, x_i^* > w_i\} \cup \{i : x_i^* < 0, x_i^* < w_i\}|$$

$$\xi := |\{i : w_i \neq x_i^* = 0\}| - |\{i : w_i = x_i^* \neq 0\}|,$$

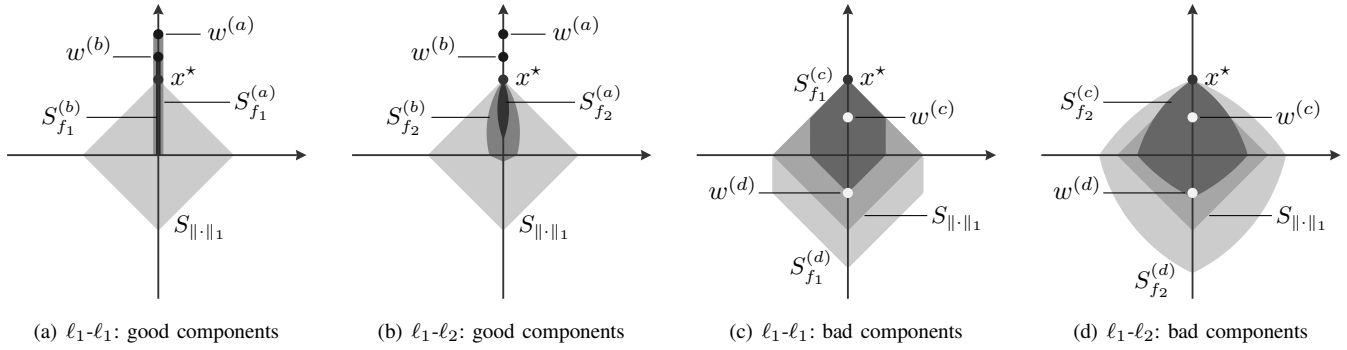


Figure 1. Sublevel sets $S_{f_i}^{(j)} := \{x : \|x\|_1 + g_i(x - w^{(j)}) \leq \|x^*\|_1 + g_i(x^* - w^{(j)})\}$, for $i = 1, 2$ and $j = a, b, c, d$, where $w^{(a)} = (0, 1.6)$, $w^{(b)} = (0, 1.3)$, $w^{(c)} = (0, 0.5)$, and $w^{(d)} = (0, -0.5)$. In both (a) and (b), w is $w^{(a)}$ and $w^{(b)}$, whose nonzero components are good. In both (c) and (d), w is $w^{(c)}$ and $w^{(d)}$, whose nonzero components are bad. The ℓ_1 -norm sublevel set $S_{\|\cdot\|_1}$ at x^* is also shown in all figures, and is associated with (1).

where $|\cdot|$ denotes the cardinality of a set. Note that \bar{h} is the number of bad components of w . Naturally, $\bar{h} \leq s$, where the difference $s - \bar{h} = h + r$ is the number of good components h plus $r := |\{i : w_i = x_i^* \neq 0\}|$. The quantity ξ is the number of components where w overestimates the support of x^* minus r . Our bound for ℓ_1 -minimization depends on these two key parameters.

Theorem 3 (ℓ_1 - ℓ_1 minimization). *Let $x^* \in \mathbb{R}^n$ be the vector to reconstruct and let $w \in \mathbb{R}^n$ be the side information. Let $f_1(x) = \|x\|_1 + \|x - w\|_1$, and assume $\bar{h} > 0$ and that there exists at least one index i for which $x_i^* = w_i = 0$. Let the entries of $A \in \mathbb{R}^{m \times n}$ be i.i.d. Gaussian with zero mean and variance $1/m$. Then,*

$$w(T_{f_1}(x^*))^2 \leq 2\bar{h} \log\left(\frac{n}{s + \xi/2}\right) + \frac{7}{5}\left(s + \frac{\xi}{2}\right). \quad (5)$$

Namely, if $m \geq 2\bar{h} \log(n/(s + \xi/2)) + (7/5)(s + \xi/2) + 1$, then x^* is the unique solution of (2) with $g = g_1$ and $\beta = 1$, with probability at least $1 - \exp(-\frac{1}{2}(\lambda_m - m)^2)$.

By upper bounding the squared Gaussian width of the tangent cone of f_1 , Theorem 3 provides a number of measurements above which ℓ_1 - ℓ_1 minimization reconstructs x^* with high probability. The assumption that there is at least one bad component, $\bar{h} > 0$, is necessary to guarantee that the subdifferential $\partial f_1(x^*)$ equals the normal cone of f_1 at x^* [28], a relation used in the proof. When $\beta \neq 1$, the assumption $\bar{h} > 0$ can be relaxed; see [11]. It can be shown that if, contrary to the theorem's assumptions, there is no index i for which $x_i^* = w_i = 0$, we can have $n = s + \xi/2$, making the right-hand side of (5) evaluate to $-\infty$ [11].

Notice that (4) and (5) have the same format and both provide reconstruction guarantees with probability at least $1 - \exp(-\frac{1}{2}(\lambda_m - m)^2)$. To compare (4) and (5), assume first that $\xi = 0$. In that case, both bounds are equal, apart from the terms multiplying the log's: $2s$ in (4) and $2\bar{h}$ in (5). Since $s - \bar{h} = h + r \geq 0$, the larger the number of good components, h , and the larger r (number of components where x^* and w coincide on the support of x^*), the smaller is (5) with respect to (4). This confirms the interpretation

given in Fig. 1 and complements it with the intuitive fact that a large r should decrease the number of measurements.¹ In general, however, $\xi \neq 0$. In that case, if n is much larger than ξ and s , if $h, r > 0$, and if ξ is larger than a small negative number, then $2\bar{h} \log(n/(s + \xi/2))$, the dominant term of (5), is smaller than $2s \log(n/s)$, the dominant term of (4). That is, (5) is asymptotically smaller than (4).

To present our results on ℓ_1 - ℓ_2 minimization, we define

$$\begin{aligned} I &:= \{i : x_i^* \neq 0\} & J &:= \{j : x_j^* \neq w_j\} \\ I_+ &:= \{i : x_i^* > 0\} & I_- &:= \{i : x_i^* < 0\}, \end{aligned}$$

and $K := |\{i \in I^c \cap J : |w_i| \geq 1\}|$, where I^c is the complement of I . We also define $q := |I \cup J|$ and $\bar{w} := \max_{i \in I^c} |w_i|$. Note that $\bar{w} \leq \|w\|_\infty$.

Theorem 4 (ℓ_1 - ℓ_2 minimization). *Let $x^*, w \in \mathbb{R}^n$ be as in Theorem 3. Let $f_2(x) = \|x\|_1 + \frac{1}{2}\|x - w\|_2^2$ and assume $x^* \neq 0$, $q < n$, and that either $\bar{w} < 1$ or that there exists $i \in I \cap J$ such that $\beta \neq \text{sign}(x_i^*)/(w_i - x_i^*)$. Assume also that*

$$\frac{q - s}{n - q} \leq |1 - \bar{w}| \exp\left(2\bar{w} \log\left(\frac{n}{q}\right)\left(\frac{\bar{w}}{2} - 1\right)\right). \quad (6)$$

Then,

$$w(T_{f_2}(x^*))^2 \leq 2v \log\left(\frac{n}{q}\right) + s + 2K + \frac{4}{5}q, \quad (7)$$

where

$$v := \sum_{i \in I_+} (1 + x_i^* - w_i)^2 + \sum_{i \in I_-} (1 + w_i - x_i^*)^2 + \sum_{i \in I \cap J^c} (|w_i| - 1)^2.$$

Theorem 4 not only requires assumptions stronger than the ones in Theorem 3, but also provides a larger bound. The assumption $q < n$ makes the right-hand sides of (6) and (7) finite. The assumption that $\bar{w} < 1$ or that there exists $i \in I \cap J$ such that $\beta \neq \text{sign}(x_i^*)/(w_i - x_i^*)$ guarantees that $\partial f_2(x^*)$ equals the normal cone of f_2 at x^* [28]. The case where assumption (6) does not hold is also addressed in [11]. Note,

¹Given that $s - \bar{h} = h + r$, we could have defined the good components as the components i for which $x_i^* > 0$ and $x_i^* \leq w_i$, or $x_i^* < 0$ and $x_i^* \geq w_i$. In that case, $s - \bar{h}$ would be exactly the number of good components. This was not done in [11] for technical reasons, and we kept the same notation here.

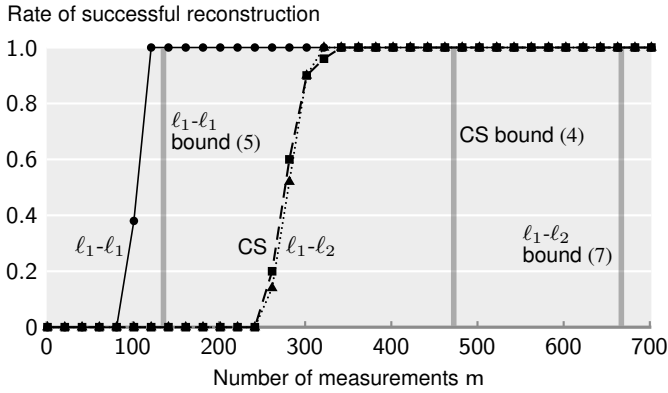


Figure 2. Reconstruction rate of standard CS (1), ℓ_1 - ℓ_1 minimization, and ℓ_1 - ℓ_2 minimization. The vertical lines are the bounds in (4), (5), and (7).

however, that it is easy to satisfy (6) in practice: its left-hand side can be shown to equal $|I^c \cap J|/|I^c \cap J^c|$, i.e., the number of components in which x^* and w differ outside I divided by the number of components in which x^* and w are both zero; if x^* and w are sparse enough, this number is smaller than 1. And the right-hand side of (6) can be large whenever $\bar{w} \neq 1$. The bound in (7) has the same format as (4) and (5). The parameter v in (7), however, depends on the magnitude of the components of both x^* and w . This is contrast with (4) and (5), whose parameters depend only on the signs of x^* and $x^* - w$, but not on the magnitudes of their components. This was expected from the interpretation of Fig. 1: the widths of the tangent cones in Figs. 1(a) and 1(c) do not vary with the magnitude of the nonzero component of w , whereas the widths in Figs. 1(b) and 1(d) do. It is not clear when (7) is smaller than (4), but note that the term inside $\log(\cdot)$ is smaller in (7), and v is always larger than s (e.g., if $w_i = x_i^*$ for all $i \in I$, $v = s + \sum_{I \cap J^c} (|w_i| - 1)^2 \geq s$).

IV. EXPERIMENTAL RESULTS

Our results are illustrated in two types of experiments whose results are shown in Figs. 2 and 3. For Fig. 2, we generated a 70-sparse $x^* \in \mathbb{R}^n$ and a 28-sparse $w \in \mathbb{R}^n$, where $n = 1000$; see [11] for how they were generated. Although the supports of x^* and w coincided in 22 entries (and differed in 6), they were significantly different: $\|w - x^*\|_2/\|x^*\|_2 \simeq 0.45$ and $\|w - x^*\|_1/\|x^*\|_1 \simeq 0.25$. This yielded $\bar{h} = h = 11$, $r = 48$, $\xi = -42$, $v \simeq 103.1$, $q = 76$, and $K = 1$. Replacing these parameters in the bounds (4), (5), and (7), we have that, for perfect recovery with high probability, standard CS requires at least 472 measurements, ℓ_1 - ℓ_1 minimization requires at least 136 measurements, and ℓ_1 - ℓ_2 minimization requires at least 666 measurements, respectively. These values are marked by vertical lines in Fig. 2, which shows the experimental performance of standard CS and ℓ_1 - ℓ_1 and ℓ_1 - ℓ_2 minimization. Specifically, it depicts the success rate of each scheme as a function of the number of measurements m . For a fixed m , we ran each algorithm 50 times, each time for a different (Gaussian) matrix A . The success rate is the number of successful reconstructions over 50, the total number of trials. Successful

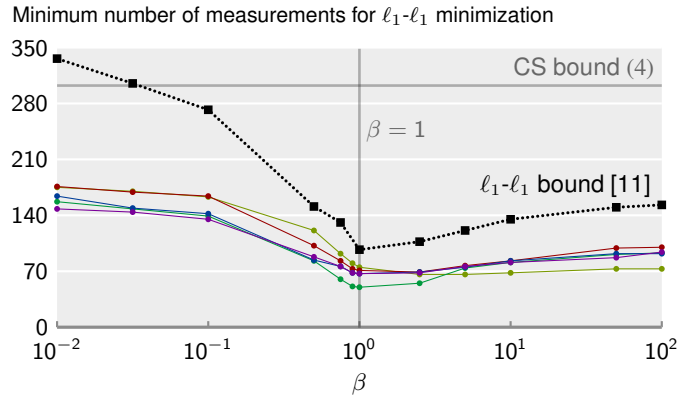


Figure 3. Solid lines: performance of ℓ_1 - ℓ_1 minimization for 5 different matrices versus β . Dotted line: bounds in [11], of which (5) gives the particular case $\beta = 1$ (vertical line). Horizontal line: bound in (4) for classical CS.

reconstruction here means $\|\hat{x} - x^*\|_2/\|x^*\|_2 \leq 10^{-2}$, where \hat{x} is a solution of (1) or (2). We see that ℓ_1 - ℓ_1 minimization required less measurements for successful reconstruction than standard CS or ℓ_1 - ℓ_2 minimization. The performance curves of the last two, in fact, almost coincide, with ℓ_1 - ℓ_2 (line with triangles) having a slightly sharper phase transition. The figure also shows that, while the ℓ_1 - ℓ_2 bound (7) can be quite loose, the ℓ_1 - ℓ_1 bound (5) is quite sharp. In other, unreported experiments, where w was not sparse, but $\|w - x^*\|_2$ was small, a situation apparently very favorable to ℓ_1 - ℓ_2 minimization, we noticed that ℓ_1 - ℓ_2 minimization still has a performance similar to CS; of course, in this case, ℓ_1 - ℓ_1 performs worse than both.

Fig. 3 considers ℓ_1 - ℓ_1 minimization only. There, $x^* \in \mathbb{R}^{500}$ is a 50-sparse vector and w was generated such that $\bar{h} = 11$ and $\xi = -30$; see [11] for details. We proceeded as follows: we generated a Gaussian matrix $\bar{A} \in \mathbb{R}^{500 \times 500}$ and computed $\bar{y} = \bar{A}x^*$. For a fixed β , we solved ℓ_1 - ℓ_1 minimization using only the first row of \bar{A} (and of \bar{y}). If the relative error of the solution was larger than 10^{-2} , we then used the first two rows of \bar{A} , and so on, until we found a minimal number of measurements $m(\beta)$ such that ℓ_1 - ℓ_1 minimization with the first $m(\beta)$ rows of \bar{A} yielded a relative error smaller than 10^{-2} . Fig. 3 shows $m(\beta)$ versus β . The solid lines correspond to 5 different realizations of (\bar{A}, \bar{y}) , and the dotted line corresponds to the theoretical ℓ_1 - ℓ_1 bounds in [11]. Note that (5) is the bound for $\beta = 1$. The plot shows that $\beta = 1$ minimizes both the theoretical curve and the experimental ones. Also, $\beta = 1$ is the value for which the theoretical bound is the sharpest.

V. CONCLUSIONS

We integrate side information in CS via ℓ_1 - ℓ_1 and ℓ_1 - ℓ_2 minimization and establish bounds on the number of measurements that guarantee successful reconstruction, for Gaussian measurement matrices. Our bound for ℓ_1 - ℓ_1 minimization is sharp and indicates that if the side information has reasonable quality, ℓ_1 - ℓ_1 minimization requires much less measurements than both standard CS and ℓ_1 - ℓ_2 minimization. The underlying geometry of the problem provides an explanation of this phenomenon, and our experimental results also confirm it.

REFERENCES

- [1] D. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Info. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [3] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky, "The convex geometry of linear inverse problems," *Found. Computational Mathematics*, vol. 12, pp. 805–849, 2012.
- [4] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.
- [5] D. Amelunxen, M. Lotz, M. McCoy, and J. Tropp, "Living on the edge: A geometric theory of phase transitions in convex optimization," 2013, preprint: <http://arxiv.org/abs/1303.6672>.
- [6] L.-W. Kang and C.-S. Lu, "Distributed compressive video sensing," in *IEEE Intern. Conf. Acoustics, Speech, and Sig. Processing (ICASSP)*, 2009, pp. 1169–1172.
- [7] A. Charles, M. Asif, J. Romberg, and C. Rozell, "Sparsity penalties in dynamical system estimation," in *IEEE Conf. Information Sciences and Systems*, 2011, pp. 1–6.
- [8] D. Baron, M. Wakin, M. Duarte, S. Sarvotham, and R. Baraniuk, "Distributed compressed sensing," ECE06-12, Electrical and Computer Engineering Dept., Rice University, Tech. Rep., 2006.
- [9] M. Trocan, T. Maugey, J. Fowler, and B. Pesquet-Popescu, "Disparity-compensated compressed-sensing reconstruction for multiview images," in *IEEE Intern. Conf. Multimedia and Expo (ICME)*, 2010, pp. 1225–1229.
- [10] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med Phys.*, vol. 35, no. 2, pp. 660–663, 2008.
- [11] J. Mota, N. Deligiannis, and M. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," 2014, submitted to *IEEE Trans. Info. Theory*, preprint: <http://arxiv.org/abs/1408.5250>.
- [12] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Processing*, vol. 58, no. 9, pp. 4595–4607, 2010.
- [13] J. Scarlett, J. Evans, and S. Dey, "Compressed sensing with prior information: Information-theoretic limits and practical decoders," *IEEE Trans. Signal Processing*, vol. 61, no. 2, pp. 427–439, 2013.
- [14] L. Weizman, Y. Eldar, and D. Bashat, "The application of compressed sensing for longitudinal MRI," 2014, preprint: <http://arxiv.org/abs/1407.2602>.
- [15] X. Wang and J. Liang, "Side information-aided compressed sensing reconstruction via approximate message passing," 2013, preprint: <http://arxiv.org/abs/1311.0576v1>.
- [16] Y. Gordon, "On Milman's inequality and random subspaces which escape through a mesh in \mathbb{R}^n ," in *Geometric Aspects of Functional Analysis, Israel Seminar 1986-1987. Lecture Notes in Mathematics*, 1988, pp. 84–106.
- [17] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Communications on Pure and Applied Mathematics*, vol. 61, no. 8, pp. 1025–1045, 2008.
- [18] M. Stojnic, "Various thresholds for ℓ_1 -optimization in compressed sensing," 2009, preprint: <http://arxiv.org/abs/0907.3666>.
- [19] N. Rao, B. Recht, and R. Nowak, "Universal measurement bounds for structured sparse signal recovery," in *Proceedings 15th Inter. Conf. Artificial Intelligence and Statistics*, 2012, pp. 942–950.
- [20] S. Oymak, C. Thrampoulidis, and B. Hassibi, "The squared-error of generalized LASSO: A precise analysis," 2013, preprint: <http://arxiv.org/abs/1311.0830>.
- [21] V. Chandrasekaran and M. Jordan, "Computational and statistical trade-offs via convex relaxation," *Proceedings of the National Academy of Sciences*, vol. 110, no. 13, pp. E1181–E1190, 2013.
- [22] S. Oymak and B. Hassibi, "Sharp MSE bounds for proximal denoising," 2013, preprint: <http://arxiv.org/abs/1305.2714>.
- [23] S. Oymak, A. Jalali, M. Fazel, Y. Eldar, and B. Hassibi, "Simultaneously structured models with application to sparse and low-rank matrices," 2013, preprint: <http://arxiv.org/abs/1212.3753>.
- [24] R. Foygel and L. Mackey, "Corrupted sensing: Novel guarantees for separating structured signals," *IEEE Trans. Info. Theory*, vol. 60, no. 2, pp. 1223–1247, 2014.
- [25] A. Bandeira, D. Mixon, and B. Recht, "Compressive classification and the rare eclipse problem," 2014, preprint: <http://arxiv.org/abs/1404.3203>.
- [26] J. Tropp, "Convex recovery of a structured signal from independent random linear measurements," 2014, preprint: <http://arxiv.org/abs/1405.1102>.
- [27] M. Kabanava, H. Rauhut, and H. Zhang, "Robust analysis ℓ_1 -recovery from Gaussian measurements and total variation minimization," 2014, preprint: <http://arxiv.org/abs/1407.7402>.
- [28] J. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*. Springer, 2004.