# THE USE OF SIDE INFORMATION IN COMPRESSIVE SENSING: MEASUREMENT DESIGN AND SIGNAL RECONSTRUCTION

*Pingfan Song*[⋆] *João F. C. Mota*[⋆] *Nikos Deligiannis*[†‡] *Miguel R. D. Rodrigues*[⋆]

[⋆] Electronic and Electrical Engineering Department, University College London, UK
[†] Department of Electronics and Informatics, Vrije Universiteit Brussel, Belgium
[‡] iMinds vzw, Ghent, Belgium

## ABSTRACT

We study the problem of sparse signal acquisition and reconstruction known as Compressive Sensing (CS) in the presence of side information, i.e., a class of signals correlated with the target signal. Side-information aided CS has been applied in various fields, such as medical imaging, remote sensing, sensor networks and compressive video. In this context, we consider a setup where the side information is available during both the signal acquisition stage, at the encoder, and the signal reconstruction stage, at the decoder. Our approach leverages side information to construct specific measurement matrices and then integrates side information into signal reconstruction by solving a $\ell_1$-$\ell_1$ minimization problem. The exploitation of side information both at the encoder and decoder allows us to achieve successful signal reconstruction with fewer measurements than just using side information at the decoder. This is shown theoretically, via establishing bounds on the number of measurements, as well as experimentally, via a series of simulations.

***Index Terms*—** Compressive Sensing, side information, measurement matrix, $\ell_1$-$\ell_1$ minimization

## 1. INTRODUCTION

Sparse signal acquisition and reconstruction based on Compressive Sensing (CS) [1, 2] has been applied in various fields, such as medical imaging [3], radar detection [4], sensor networks [5], and compressive video [6]. In many of these scenarios, one has access not only to the signal of interest, but also to side information (SI), i.e., a class of signals correlated with our target signal. For example, in the medical imaging domain, previous MRI scans can serve as side information for current MRI image reconstruction. In sensor networks, the signals captured by nearby sensors can be regarded as side information to assist distributed reconstruction.

The case where the side information is available at the decoder has been extensively analyzed. $\ell_1$-$\ell_1$ and $\ell_1$-$\ell_2$ minimization [7–9] exploit a signal analogous to the signal to be reconstructed. Modified-CS [10, 11] deterministically or probabilistically use estimates of the support of the signal to modify the reconstruction procedure, e.g., Basis Pursuit (BP) [12]. However, to our knowledge, the scenario where side information is available at both the encoder and decoder, has rarely been studied and no CS-type bounds are known. Fig. 1 illustrates such a compressive sensing system where, in addition to the signal of interest, one also has access to side information during both the signal acquisition process and the reconstruction process. In this context, we propose an approach to leverage side information to construct a more effective measurement matrix and also to adopt a more suitable optimization program to reconstruct the original signal. In sum, our goal is to explore the availability of such side information at both the encoder and decoder, by conceiving better sampling procedures and reconstruction procedures. Then, by capitalizing on the up-to-date convex geometry tools (namely the concept of Gaussian width [13]), we establish a tight theoretical bound which indicates that our approach allows to reduce the number of measurements via exploiting appropriate side information.

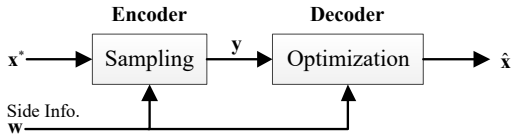



**Fig. 1**. CS with Side Information at both the encoder and decoder.

**Problem statement.** Fig. 1 illustrates the situation where side information is available at both the encoder and decoder. Let $\mathbf{x}^\star \in \mathbb{R}^n$ be the signal of interest with sparsity $s := |\{i : x_i^\star \neq 0\}|$ and $\mathbf{w} \in \mathbb{R}^n$ be the side information, i.e., a signal similar to $\mathbf{x}^\star$, with sparsity $s_w := |\{i : w_i \neq 0\}|$. We take $m$ linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$ from $\mathbf{x}^\star$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the measurement matrix designed with side information. Then the vector of measurements $\mathbf{y}$ is sent to the decoder where the original signal is reconstructed from $\mathbf{y}$ by solving an optimization problem, e.g., $\ell_1$-$\ell_1$ optimization. In this context, we propose a scheme to design a specific measurement matrix $\mathbf{A}$ by leveraging the side information $\mathbf{w}$, and then analyze the reconstruction performance of the resulting system.

## 2. BACKGROUND

Given $m$ linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$ of a signal of interest $\mathbf{x}^\star$ and side information $\mathbf{w}$, $\mathbf{x}^\star$ can be reconstructed from $\mathbf{y}$ via solving a $\ell_1$-$\ell_1$ minimization problem [8, 9, 14, 15], formulated as follows:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \|\mathbf{x}\|_1 + \|\mathbf{x} - \mathbf{w}\|_1 \\ \text{subject to} \quad & \mathbf{y} = \mathbf{A}\mathbf{x}\,. \end{aligned} \tag{1}$$

where $\|\cdot\|_1$ denotes the $\ell_1$-norm.

Assuming that the measurement matrix $\mathbf{A}$ is composed of i.i.d. Gaussian entries, [9] establishes a bound for the minimum number of measurements required for successful signal reconstruction via

This work is supported by China Scholarship Council(CSC), UCL Overseas Research Scholarship (UCL-ORS), the VUB-UGent-UCL-Duke International Joint Research Group grant (VUB: DEFIS41010), and by EPSRC grant EP/K033166/1.

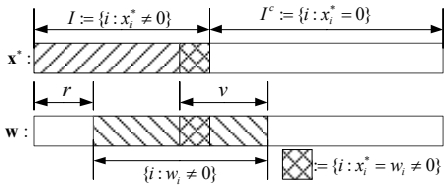



**Fig. 2**. Visualization of the mismatch parameters $r$ and $v$ for SI at both the encoder and decoder.

solving (1). Let us define

$$\overline{h} := |\{i : x_i^\star > 0,\, x_i^\star > w_i\} \cup \{i : x_i^\star < 0,\, x_i^\star < w_i\}|\,, \quad (2)$$
$$\xi := |\{i : x_i^\star = 0,\, w_i \neq 0\}| - |\{i : w_i = x_i^\star \neq 0\}|\,, \quad (3)$$

where $|\cdot|$ denotes the cardinality of a set. Note that, $\overline{h}$ is defined on the support $I := \{i : x_i^\star \neq 0\}$ of $\mathbf{x}^\star$ and thus $\overline{h} \leq s$.

**Theorem 1** (Theorem 1 in [9])**.** *Let $\mathbf{x}^\star \in \mathbb{R}^n$ be the signal of interest with sparsity $s := |\{i : x_i^\star \neq 0\}|$, and $\mathbf{w} \in \mathbb{R}^n$ be the side information. Given a vector of measurements $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$, where the matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is composed of i.i.d. Gaussian entries with zero mean and variance $1/m$, $\mathbf{x}^\star$ is the unique optimal solution of* (1) *with probability at least $1 - \exp\left(-\frac{1}{2}\left(\sqrt{m} - \sqrt{m-1}\right)^2\right)$, provided*

$$m \geq 2\overline{h} \ln\left(\frac{n}{s + \frac{\xi}{2}}\right) + \frac{7}{10}\left(s + \frac{\xi}{2}\right) + 1\,. \quad (4)$$

The theorem shows that the number of measurements required for signal reconstruction via solving (1) is $O(\overline{h} \ln\, n)$. In the next section, we will demonstrate how to reduce (4) via integrating side information into the design of measurement matrix.

## 3. SIDE INFORMATION AIDED CS SYSTEM

In this section, we consider the case where side information is available at both the encoder and the decoder. We start by presenting our measurement matrix design scheme with side information and the corresponding reconstruction program; then, we analyse the resulting system and establish a theoretical bound for the number of measurements required for perfect reconstruction in subsection 3.2, and the proof outline for the bound is given in subsection 3.3.

### 3.1. Measurements Acquisition and Signal Reconstruction

The measurement matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is designed by taking each row vector of the measurement matrix to be independently drawn from a Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, where the covariance matrix $\mathbf{\Sigma}$, assumed diagonal, is designed according to the side information $\mathbf{w}$. That is, $\mathbf{\Sigma} = \text{diag}(\sigma_1, \ldots, \sigma_i, \ldots, \sigma_n) \in \mathbb{R}^{n\times n}$, with

$$\sigma_i = \begin{cases} 1 & \text{if } w_i \neq 0 \\ \varepsilon \in (0, 1] & \text{if } w_i = 0 \end{cases} \quad (5)$$

The intuition of setting $\varepsilon \leq 1$ stems from the consideration of spending less energy acquiring those components that the side information indicates being close to zero. Based on the designed measurement matrix, we solve a $\ell_1$-$\ell_1$ minimization problem (1) to reconstruct the signal of interest.

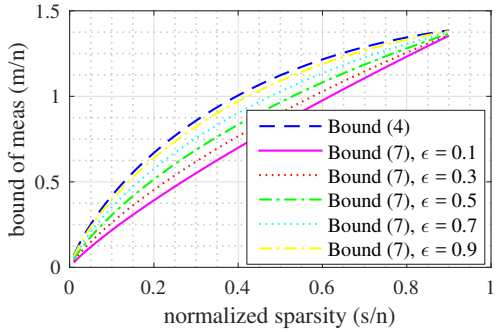



**Fig. 3**. Comparison of our theoretical bound (7) with the classic $\ell_1$-$\ell_1$ bound (4) with $r/s = 0.1$, $v/s = 0.1$. It shows that our bound (7) is lower than bound (4) and our bound (7) asymptotically approaches (4) as $\varepsilon$ increases toward to 1.

### 3.2. Theoretical Bound

In this section, we analyse the performance of the resulting CS system and present a new tight bound on the number of measurements required for successful reconstruction.

To present our result, we define two parameters that capture the amount of mismatch between $\mathbf{x}^\star$ and $\mathbf{w}$:

$$r := |\{i : x_i^\star \neq 0,\, w_i = 0\}|\,, \quad (6a)$$
$$v := |\{i : x_i^\star = 0, w_i \neq 0\}| + c\,, \quad (6b)$$

where, $c := |\{i : w_i = x_i^\star \neq 0\}|$ denotes the non-zero components commonly shared by both $\mathbf{x}^\star$ and $\mathbf{w}$, which is usually very small in practice. In words, $r$ counts the number of components missed by $\mathbf{w}$, and $v - c$ counts the number of components that $\mathbf{w}$ overestimates, as shown in Fig. 2. There holds $0 \leq r \leq \overline{h}$ and $0 \leq v \leq \min\{s_w, n - s - c\}$.

**Proposition 1.** *Let $\mathbf{x}^\star \in \mathbb{R}^n$ be the signal of interest with sparsity $s := |\{i : x_i^\star \neq 0\}|$, and $\mathbf{w} \in \mathbb{R}^n$ be the side information. Let $r > 0$ and $v > 0$, defined as* (6a) *and* (6b)*, denote the two types of mismatch between $\mathbf{x}^\star$ and $\mathbf{w}$. Given a vector of measurements $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is designed as* (5)*, then $\mathbf{x}^\star$ is the unique optimal solution of the $\ell_1$-$\ell_1$ minimization* (1) *with probability at least $1 - \exp\left(-\frac{1}{2}\left(\sqrt{m} - \sqrt{m-1}\right)^2\right)$, provided*

$$m \geq 2\big(\varepsilon\overline{h} + (1-\varepsilon)r\big) \ln\left(\frac{n}{s + \frac{\xi}{2}}\right) + \frac{7}{10}\left(s + \frac{\xi}{2}\right) + \frac{1 - \sqrt{\varepsilon}}{2} v + 1\,. \quad (7)$$

**Remark 1.** Proposition 1 establishes that in the case where side information is available at both the encoder and the decoder, our approach allows to reduce the number of required measurements from $O(\overline{h} \ln\, n)$ [cf. (4)] to $O((\varepsilon\overline{h} + (1-\varepsilon)r) \ln\, n)$. In particular, if the number of components missed by $\mathbf{w}$, namely $r$, is small and the dimension $n$ is large, the reduction in the number of measurements can be significant.

**Remark 2.** Proposition 1 also shows that $\varepsilon$ provides a trade-off on how much we trust the side information in the measurement matrix design. For high quality side information with small mismatch $r$ and $v$, a smaller $\varepsilon$ should be selected to reduce the dominant logarithmic term. Otherwise, a larger $\varepsilon$ is more favourable.

**Remark 3.** It also shows that our bound (7) generalizes the bound (4). Concretely, as shown in Fig. 3, bound (7) asymptotically approaches (4) as $\varepsilon$ increases. Note that taking $\varepsilon = 1$ leads to (7) simplifying to (4), as $\varepsilon\overline{h} + (1-\varepsilon)r = \overline{h}$ and $(1 - \sqrt{\varepsilon})v/2 = 0$.

### 3.3. Outline of the proof of Proposition 1

This section gives the proof outline of Proposition 1, which involves mainly two stages: converting the $\ell_1$-$\ell_1$ minimization with our measurement matrix design to a weighted $\ell_1$-$\ell_1$ minimization with an i.i.d. Gaussian matrix and computing the upper bound.

**Conversion to weighted $\ell_1$-$\ell_1$ minimization.** We notice the equivalence between $\ell_1$-$\ell_1$ minimization with our measurement matrix design $\mathbf{A}$ and a weighted $\ell_1$-$\ell_1$ minimization with an i.i.d. Gaussian matrix $\widetilde{\mathbf{A}}$. Concretely, problem (1) with the measurement matrix $\mathbf{A}$ as proposed in §3.1 can be formulated as

$$\begin{array}{lll} \min\limits_{\mathbf{x}} & \|\mathbf{x}\|_1 + \|\mathbf{x}-\mathbf{w}\|_1 & \Longleftrightarrow \quad \min\limits_{\mathbf{x}} \quad \|\mathbf{x}\|_1 + \|\mathbf{x}-\mathbf{w}\|_1 \\ \text{s.t.} & \mathbf{y} = \mathbf{A}\mathbf{x} & \quad\ \ \text{s.t.} \quad \mathbf{y} = (\mathbf{A}\mathbf{D})(\mathbf{D}^{-1}\mathbf{x}) \\ & & \Longleftrightarrow \quad \min\limits_{\mathbf{z}} \quad \|\mathbf{D}\mathbf{z}\|_1 + \|\mathbf{D}\mathbf{z}-\mathbf{w}\|_1 \\ & & \quad\ \ \text{s.t.} \quad \mathbf{y} = \widetilde{\mathbf{A}}\mathbf{z} \end{array} \tag{8}$$

where $\widetilde{\mathbf{A}} := \mathbf{A}\mathbf{D}$ and $\mathbf{z} = \mathbf{D}^{-1}\mathbf{x}$ for $\mathbf{D} := \mathrm{diag}(d_1, \ldots, d_n)$, with

$$d_i = \frac{1}{\sqrt{\sigma_i}} = \begin{cases} 1; & \text{if } w_i \neq 0 \\ 1/\sqrt{\varepsilon}; & \text{if } w_i = 0 \end{cases} \qquad 0 < \varepsilon \leq 1\,. \tag{9}$$

Note that from (1) to (8), the optimization variable is changed from $\mathbf{x}$ to $\mathbf{z} = \mathbf{D}^{-1}\mathbf{x}$, where the matrix $\mathbf{D}$ is invertible because $\varepsilon > 0$. The goal of $\mathbf{D}$ is to transform the current non-i.i.d. (anisotropic) Gaussian measurement matrix $\mathbf{A}$ to an i.i.d. (isotropic) Gaussian matrix $\widetilde{\mathbf{A}}$. In addition, the diagonal form of $\mathbf{D}$ ensures that the optimal solution $\mathbf{z}^\star$ of (8) has the same support as $\mathbf{x}^\star$.

**Computation of the bounds.** The proof of Proposition 1 relies on a result from [13], that establishes a bound for the minimal number of measurements that guarantees the successful reconstruction of a structured vector $\mathbf{x}^\star$ from a set of linear measurements $\mathbf{y}$. Specifically, Corollary 3.3 in [13] states that given a matrix $\mathbf{\Phi} \in \mathbb{R}^{m\times n}$ whose entries are i.i.d. Gaussian random variables with zero-mean and variance $1/m$, and a set of linear measurements $\mathbf{y} = \mathbf{\Phi}\mathbf{x}^\star$, then $\mathbf{x}^\star$ is the unique optimum solution of convex program $\min\{f(\mathbf{x}) : \mathbf{y} = \mathbf{\Phi}\mathbf{x}\}$ provided $m \geq w(\Omega)^2 + 1$, where $w(\Omega)$ is referred as Gaussian width of the set $\Omega$ [1].

As it is difficult to compute the Gaussian width $w(\Omega)$ in closed form, we compute instead the Gaussian distance as an upper bound. Concretely, the objective function in our case is $f(\mathbf{z}) = \|\mathbf{D}\mathbf{z}\|_1 + \|\mathbf{D}\mathbf{z}-\mathbf{w}\|_1$. Suppose $\mathbf{0} \notin \partial f(\mathbf{z}^\star)$ holds for a given $\mathbf{z}^\star \in \mathbb{R}^n$, [9, 13, 14] show that there holds

$$\mathbb{E}_{\mathbf{g}}\left[\mathrm{dist}\big(\mathbf{g}, \mathrm{cone}\,\partial f(\mathbf{z}^\star)\big)^2\right] \geq w(\Omega)^2 \tag{10}$$

where $\mathrm{cone}\,\partial f(\mathbf{z}^\star)$ is the cone generated by the subdifferential of the objective function $f(\mathbf{z})$ at the point $\mathbf{z}^\star$, and $\mathrm{dist}(\mathbf{g}, S) := \min\{\|\mathbf{z}-\mathbf{g}\|_2 : \mathbf{z} \in S\}$ denotes the Euclidean distance between a point $\mathbf{g}$ and the set $S$.

Before we describe how to compute the Gaussian distance, we introduce some useful notations. Define $I := \{i : z_i^\star \neq 0\}$, $I^c := \{i : z_i^\star = 0\}$, $J := \{i : z_i^\star \neq w_i\}$, and $J^c := \{i : z_i^\star = w_i\}$. To compute an upper bound for the Gaussian distance in (10), the objective is decomposed as $f(\mathbf{z}) = \|\mathbf{D}\mathbf{z}\|_1 + \|\mathbf{D}\mathbf{z}-\mathbf{w}\|_1 = \sum_{i=1}^n f^{(i)}(z_i) = \sum_{i=1}^n |d_i\, z_i| + |d_i\, z_i - w_i|$ and the corresponding cone $\partial f(\mathbf{z}^\star)$ generated by $f(\mathbf{z})$ at $\mathbf{z}^\star$ is computed as

$$\mathrm{cone}\,\partial f(\mathbf{z}^\star) = \Big(t\cdot\partial f^{(1)}(z_1^\star), t\cdot\partial f^{(2)}(z_2^\star), \cdots, t\cdot\partial f^{(n)}(z_n^\star)\Big)\,,$$

$$t\cdot\partial f^{(i)}(z_i^\star) = \begin{cases} t\,d_i\,\mathrm{sign}(d_i\,z_i^\star) + t\,d_i\,\mathrm{sign}(d_i\,z_i^\star - w_i) & \text{if } i \in IJ \\ t\,d_i\,\mathrm{sign}(d_i\,z_i^\star) + \mathcal{I}\big(0\,, t\,d_i\big) & \text{if } i \in IJ^c \\ \mathcal{I}\big(0\,, t\,d_i\big) + t\,d_i\,\mathrm{sign}(d_i\,z_i^\star - w_i) & \text{if } i \in I^c\,J \\ \mathcal{I}\big(0\,, 2\,t\,d_i\big) & \text{if } i \in I^c\,J^c \end{cases}$$

where $i = 1, \ldots, n$ and $\mathcal{I}(a, b)$ denotes an interval with centre $a$ and length $b$. Then, Jensen's inequality [2] is applied to derive (11).

$$\begin{aligned} &\mathbb{E}_{\mathbf{g}}\Big[\mathrm{dist}\big(\mathbf{g}, \mathrm{cone}\,\partial f(\mathbf{z}^\star)\big)^2\Big] \\ &\leq \sum_{i\in IJ} \mathbb{E}_{g_i}\Big[\mathrm{dist}\Big(g_i\,, t\,d_i\,\mathrm{sign}(d_i\,z_i^\star) + t\,d_i\mathrm{sign}(d_i\,z_i^\star - w_i)\Big)^2\Big] && \text{(11a)} \\ &+ \sum_{i\in IJ^c} \mathbb{E}_{g_i}\Big[\mathrm{dist}\Big(g_i\,, \mathcal{I}\big(t\,d_i\,\mathrm{sign}(d_i\,z_i^\star), t\,d_i\big)\Big)^2\Big] && \text{(11b)} \\ &+ \sum_{i\in I^cJ} \mathbb{E}_{g_i}\Big[\mathrm{dist}\Big(g_i\,, \mathcal{I}\big(t\,d_i\,\mathrm{sign}(d_i\,z_i^\star - w_i), t\,d_i\big)\Big)^2\Big] && \text{(11c)} \\ &+ \sum_{i\in I^cJ^c} \mathbb{E}_{g_i}\Big[\mathrm{dist}\Big(g_i\,, \mathcal{I}\big(0\,, 2t\,d_i\big)\Big)^2\Big] && \text{(11d)} \end{aligned}$$

which holds for arbitrary $t > 0$. In our case, we will set $t = \sqrt{\frac{\varepsilon}{2}\ln\left(\frac{n}{s+\xi/2}\right)}$, which yields an upper bound.

For simplicity, we define

$$A(x) = -x\varphi(x) + (1+x^2)Q(x)\,. \tag{12}$$

where $\varphi(x) = \exp\left(-x^2/2\right)/\sqrt{2\pi}$ is the probability density function of a standard Gaussian random variable, and $Q(x) = \int_x^{+\infty}\varphi(t)\,\mathrm{d}t$ is the $Q$-function. Then, we prove that for a scalar zero-mean Gaussian random variable with unit variance, i.e., $g \sim \mathcal{N}(0,1)$, the Gaussian distance can be expressed as

$$\mathbb{E}_g\Big[\mathrm{dist}\big(g, \mathcal{I}(a,b)\big)^2\Big] = A(b-|a|) + A(b+|a|). \tag{13}$$

Then, we obtain

$$\begin{aligned} \text{(11a)} &= |IJ| + \sum_{i\in I_+J_+\cup I_-J_-} (2td_i)^2 \\ \text{(11b)} &= \sum_{i\in IJ^c}\left[A(2td_i) + \frac{1}{2}\right] = \sum_{i\in IJ^c} A(2td_i) + \frac{1}{2}|IJ^c| \\ \text{(11c)} &= \sum_{i\in I^cJ}\left[A(2td_i) + \frac{1}{2}\right] = \sum_{i\in I^cJ} A(2td_i) + \frac{1}{2}|I^cJ| \\ \text{(11d)} &= 2\sum_{i\in I^cJ^c} A(2td_i) \end{aligned}$$

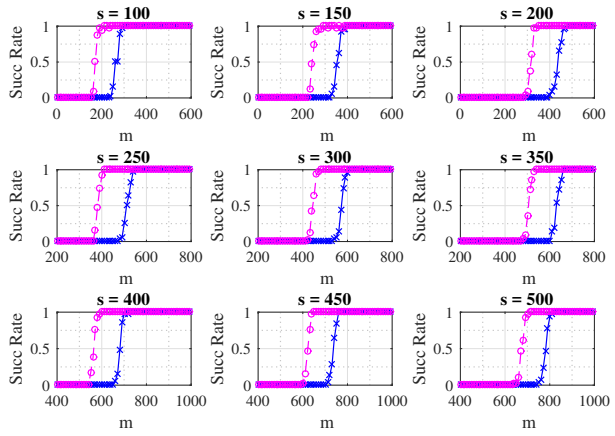
Expression for (11a) + (11b) + (11c) + (11d) is bounded as (14).

$$\begin{aligned} &\mathbb{E}_{\mathbf{g}}\Big[\mathrm{dist}\big(\mathbf{g}, \mathrm{cone}\partial f(\mathbf{z}^\star)\big)^2\Big] \leq |IJ| + \frac{1}{2}\left(|IJ^c| + |I^cJ|\right) \\ &\quad + \sum_{i\in I_+J_+\cup I_-J_-} (2td_i)^2 + \sum_{i\in IJ^c\cup I^cJ} A(2td_i) + 2\sum_{i\in I^cJ^c} A(2td_i) \\ &\leq 2\big(\varepsilon\,\overline{h} + (1-\varepsilon)\,r\big)\ln\left(\frac{n}{s+\frac{\xi}{2}}\right) + \frac{7}{10}(s+\frac{\xi}{2}) + \frac{1-\sqrt{\varepsilon}}{2}\,v \end{aligned}$$

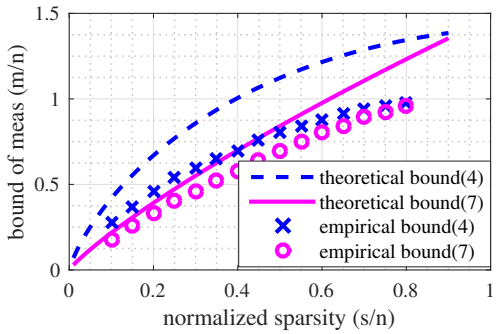
This, together with (10), completes the proof of (7).

[1] Let $\Omega = T_f(\mathbf{x}^\star) \cap \mathbb{S}^{n-1}$ denote the intersection of the tangent cone $T_f(\mathbf{x}^\star) := \mathrm{cone}\,\{\mathbf{x}-\mathbf{x}^\star : f(\mathbf{x}) \leq f(\mathbf{x}^\star)\}$ and the unit sphere $\mathbb{S}^{n-1} := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 = 1\} \subset \mathbb{R}^n$. The Gaussian width of a set $\Omega \subset \mathbb{R}^n$ is defined as: $w(\Omega) := \mathbb{E}_{\mathbf{g}}\left[\sup\limits_{\mathbf{z}\in\Omega} \mathbf{g}^{\mathrm{T}}\mathbf{z}\right]$, where $\mathbf{g} \sim \mathcal{N}(0, \mathbf{I}_n) \in \mathbb{R}^n$ is a vector of independent zero-mean unit-variance Gaussians.

[2] In the context of probability theory, Jensen's inequality is generally stated in the following form: if $\mathbf{X}$ is a random variable and $\varphi$ is a convex function, then $\varphi(\mathbb{E}[\mathbf{X}]) \leq \mathbb{E}[\varphi(\mathbf{X})]$.




(a) Success ratio with measurements for different sparsity levels.




(b) Theoretical (curves) and empirical (markers) bounds.

**Fig. 4**. Experimental results. (a) In each sub-figure, the blue line with marker $\times$ represents the recovery performance of i.i.d. Gaussian matrix and $\ell_1$-$\ell_1$, the red line with marker $\circ$ represents our measurement matrix design with $\ell_1$-$\ell_1$.

## 4. EXPERIMENTAL RESULTS

A set of numerical experiments have been conducted to verify our approach. An i.i.d. Gaussian matrix is used as the benchmark. In the signal reconstruction stage, we solve the same $\ell_1$-$\ell_1$ minimization problem with both an i.i.d. Gaussian matrix and our measurement matrix design. The parameters are shown in Table 1. $m$ varies from 10 to 1000 in steps of 10, and the relative sparsity $s/n$ varies from 0.1 to 0.8 in steps of 0.05. The number of measurements corresponding to the success rate exceeding 85% serves as the empirical threshold. The experiment results are shown in Fig. 4. Fig. 4(a) shows the success rate as a function of the number of measurements for different sparsity levels. Fig. 4(b) compares the theoretical bounds and empirical bounds. The results indicate that our measurement matrix design outperforms the i.i.d. Gaussian matrix for small mismatch, and our bound is tight and practically coincides with the empirical bounds well.

**Table 1**. Parameters setting for the experiments

| $n$ | $m/n$ | $s/n$ | $r/s$ | $v/s$ | $\varepsilon$ |
|---|---|---|---|---|---|
| 1000 | 0.01:0.01:1 | 0.1:0.05:0.8 | 0.1 | 0.1 | 0.1 |

## 5. CONCLUSIONS

In this paper, we studied a side-information aided Compressive Sensing system. We design a specific measurement matrix to integrate side information at the encoder and then exploit $\ell_1$-$\ell_1$ minimization to incorporate side information at the decoder. We established a tight theoretical bound to analyse the performance of the resulting system in terms of the number of measurements required for perfect signal reconstruction. Numerical experiments validate that our theoretical bound practically coincides with empirical bound, and also indicate that our paradigm requires less measurements than conventional approaches. We believe this work can contribute to improve the design of CS systems in scenarios where side information is available, such as medical imaging, sensor networks, and multi-view camera systems.

## 6. REFERENCES


[1] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[3] L. Weizman, Y. C. Eldar, and D. Ben-Bashat, "Compressed sensing for longitudinal MRI: An adaptive-weighted approach," *Medical Physics*, vol. 42, no. 9, pp. 5195–5207, 2015.

[4] Laura Anitori, Arian Maleki, Matern Otten, Richard G Baraniuk, and Peter Hoogeboom, "Design and analysis of compressed sensing radar detectors," *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 813–827, 2013.

[5] W. Bajwa, J. Haupt, A Sayeed, and R Nowak, "Compressive wireless sensing," in *Intern. Conf. Inform. Process. in Sensor Networks (IPSN)*, 2006, pp. 134–142.

[6] J. Mota, N. Deligiannis, A. Sankaranarayanan, V. Cevher, and M. Rodrigues, "Adaptive-rate sparse signal reconstruction with application in compressive background subtraction," *arXiv preprint arXiv:1503.03231*, 2015.

[7] Guang-Hong Chen, Jie Tang, and Shuai Leng, "Prior image constrained compressed sensing (piccs): a method to accurately reconstruct dynamic ct images from highly undersampled projection data sets," *Medical physics*, vol. 35, no. 2, pp. 660–663, 2008.

[8] X. Wang and J. Liang, "Side information-aided compressed sensing reconstruction via approximate message passing," Preprint: `http://arxiv.org/abs/1311.0576v1`, 2013.

[9] J. F. C. Mota, N. Deligiannis, and M. R. D. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," *arXiv preprint arXiv:1408.5250*, 2014.

[10] M. P. Friedlander, H. Mansour, R. Saab, and O. Yilmaz, "Recovering compressively sampled signals using partial support information," *IEEE Trans. Inform. Theory*, vol. 58, no. 2, pp. 1122–1134, 2012.

[11] J. Scarlett, J. Evans, and S. Dey, "Compressed sensing with prior information: Information-theoretic limits and practical decoders," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 427–439, 2013.

[12] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[13] 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.

[14] 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.

[15] J. Mota, N. Deligiannis, and M. Rodrigues, "Compressed sensing with side information: Geometrical interpretation and performance bounds," in *IEEE Global Conf. on Signal and Information Processing (GlobalSIP)*, 2014, pp. 675–679.

[16] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of convex analysis*, Springer Science & Business Media, 2001.