

# MEASUREMENT MATRIX DESIGN FOR COMPRESSIVE SENSING WITH SIDE INFORMATION AT THE ENCODER

Pingfan Song<sup>\*</sup>    João F. C. Mota<sup>\*</sup>    Nikos Deligiannis<sup>†‡</sup>    Miguel Raul Dias Rodrigues<sup>\*</sup>

<sup>\*</sup> Electronic and Electrical Engineering Department, University College London, UK

<sup>†</sup> Department of Electronics and Informatics, Vrije Universiteit Brussel, Belgium

<sup>‡</sup> iMinds vzw, Ghent, Belgium

## ABSTRACT

We study the problem of measurement matrix design for Compressive Sensing (CS) when the encoder has access to side information, a signal analogous to the signal of interest. In particular, we propose to incorporate this extra information into the signal acquisition stage via a new design for the measurement matrix. The goal is to reduce the number of encoding measurements, while still allowing perfect signal reconstruction at the decoder. Then, the reconstruction performance of the resulting CS system is analysed in detail assuming the decoder reconstructs the original signal via Basis Pursuit. Finally, Gaussian width tools are exploited to establish a tight theoretical bound for the number of required measurements. Extensive numerical experiments not only validate our approach, but also demonstrate that our design requires fewer measurements for successful signal reconstruction compared with alternative designs, such as an i.i.d. Gaussian matrix.

**Index Terms**— measurement matrix design, side information, Compressive Sensing, Basis Pursuit

## 1. INTRODUCTION

Compressive Sensing (CS) [1, 2] is a signal acquisition paradigm that leverages signal sparsity to acquire signals with far less measurements than classical sampling schemes. It has been applied, for example, in medical imaging [3], radar detection [4], sensor networks [5], and compressive video [6]. In many of these applications, one has access to *side information*, a signal analogous to the signal we want to reconstruct. For example, in compressive video past reconstructed frames can be used as side information to reconstruct the current frame [7, 8]; in medical imaging, prior patient scans can also be used as side information in image reconstruction [9]. One of the main benefits of using side information in these applications, where measurements are expensive, is that it allows reducing the number of required measurements for reconstruction.

Much research has used side or prior information to obtain even lower acquisition rates in CS. This typically involves modifying the reconstruction procedure, e.g., Basis Pursuit (BP) [10] by using estimates of the support of the signal, either deterministically, as in modified-CS [11–15], probabilistically, as in [16, 17], or in the form of a signal analogous to the signal to be reconstructed, as in  $\ell_1$ - $\ell_1$  and  $\ell_1$ - $\ell_2$  minimization [18–21]. Other work, such as [22], has used side information to perform classification and reconstruction under

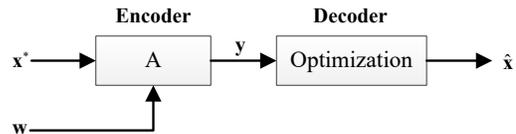


Fig. 1. CS with Side Information at the encoder.

probabilistic models. In all these work, however, side information is used to aid only the sparse reconstruction process, not the acquisition one.

In practice, side information may be available both at the encoder and decoder. In this paper, the goal is to integrate side information into the measurement matrix to aid the signal acquisition process. Note that our approach differs from those where the encoder uses richer models to improve performance, e.g., [23, 24]. We show that, conceptually, our measurement matrix design is equivalent to solving a weighted  $\ell_1$  minimization problem at the decoder. In that sense, our work is closely related to [25], which establishes bounds on the number of measurements required to reconstruct sparse signals using weighted  $\ell_1$  minimization. Although we use tools similar to the ones in [25] (namely the concept of Gaussian width [26–29]) to establish related bounds, our bounds are much simpler and, as our experiments show, also much tighter.

**Problem statement.** Fig. 1 illustrates the situation where side information is available at the encoder. Let  $\mathbf{x}^* \in \mathbb{R}^n$  be the signal of interest, from which we take  $m$  linear measurements  $\mathbf{y} = \mathbf{A}\mathbf{x}^*$ , where  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is the measurement matrix. Then the vector of measurements  $\mathbf{y}$  is sent to the decoder. In turn, the decoder reconstructs a vector  $\hat{\mathbf{x}} \in \mathbb{R}^n$  from  $\mathbf{y}$  by, e.g., solving an optimization problem like BP. If there are enough measurements,  $\hat{\mathbf{x}}$  should be a good approximation to  $\mathbf{x}^*$ . In the figure,  $\mathbf{w} \in \mathbb{R}^n$  comprises the side information, a vector that we assume to be similar to  $\mathbf{x}^*$ . In this context, we study the problem of designing the measurement matrix  $\mathbf{A}$  at the encoder using the side information  $\mathbf{w}$ , and then analyze the reconstruction performance of the resulting system.

**Contributions.** In the presence of side information at the encoder, we propose a new design scheme for the measurement matrix  $\mathbf{A}$  in which each row vector is independently drawn from the Gaussian distribution  $\mathcal{N}(\mathbf{0}, \Sigma)$ , where the covariance matrix  $\Sigma \in \mathbb{R}^{n \times n}$  is, assumed diagonal, is designed according to the side information  $\mathbf{w}$ . Then, based on this design, we establish a bound on the number of measurements that guarantees perfect reconstruction of the original signal via solving BP. Experimental results illustrate the gains of the proposed measurement matrix design with respect to the scenarios where no side information is used.

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.

## 2. BACKGROUND

Let  $\mathbf{x}^* \in \mathbb{R}^n$  be the sparse signal of interest. When there is no side information available, it is a standard CS problem to reconstruct  $\mathbf{x}^*$  from a given vector of measurements  $\mathbf{y} = \mathbf{A}\mathbf{x}^*$ . One common approach is to solve Basis Pursuit [10]:

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && f(\mathbf{x}) = \|\mathbf{x}\|_1 \\ & \text{subject to} && \mathbf{y} = \mathbf{A}\mathbf{x}, \end{aligned} \quad (1)$$

where  $\|\mathbf{x}\|_1 := \sum_{i=1}^n |x_i|$  denotes the  $\ell_1$ -norm of  $\mathbf{x}$ . Assuming  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is composed of i.i.d. Gaussian entries with zero mean and variance  $1/m$ , there are several different tools to analyse (1), for example, the Restricted Isometry Property (RIP) [30] and the Gaussian width/distance [26–29]. We will use the latter, since it allows providing sharper performance characterizations, although with the shortcoming of being applicable only to Gaussian matrices. Concretely, [28] establishes the following theorem.

**Theorem 1** (Corollary 3.3 and Proposition 3.10 in [28]). *Let  $\mathbf{x}^* \in \mathbb{R}^n$  be the signal of interest with sparsity  $s := |\{i : x_i^* \neq 0\}|$ . Given a vector of measurements  $\mathbf{y} = \mathbf{A}\mathbf{x}^*$ , where matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is composed of i.i.d. zero-mean Gaussian random variables with variance  $1/m$ , then,  $\mathbf{x}^*$  is the unique optimal solution of (1) with probability at least  $1 - \exp(-\frac{1}{2}(\lambda_m - \omega(\Omega))^2)$ , provided*

$$m \geq 2s \ln\left(\frac{n}{s}\right) + \frac{7}{5}s + 1. \quad (2)$$

In the above theorem,  $\lambda_m := \mathbb{E}_{\mathbf{g}}[\|\mathbf{g}\|_2]$  denotes the expected length of a zero-mean, unit-variance Gaussian vector  $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$  in  $\mathbb{R}^m$ . Moreover,  $\omega(\Omega) := \mathbb{E}_{\mathbf{g}}\left[\sup_{\mathbf{z} \in \Omega} \mathbf{g}^T \mathbf{z}\right]$  denotes the Gaussian width of the set  $\Omega$ , where  $\Omega = T_f(\mathbf{x}^*) \cap \mathbb{S}^{n-1}$  is the spherical part of the tangent cone  $T_f(\mathbf{x}^*)$  and  $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n) \in \mathbb{R}^n$  is a zero-mean, unit-variance Gaussian vector. Thus, when no prior information is available, the number of measurements necessary to recover  $\mathbf{x}^*$  is  $O(s \ln n)$ . Next, we will see that providing side information to the encoder allows obtaining a bound smaller than (2).

## 3. MEASUREMENT DESIGN WITH SIDE INFORMATION

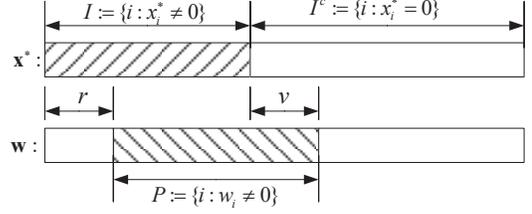
In this section, we consider the case where side information is available at the encoder. We start by presenting our design scheme for the measurement matrix; then, we analyse the resulting scheme and establish a theoretical bound for the number of measurements required for perfect reconstruction in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

### 3.1. Design Scheme

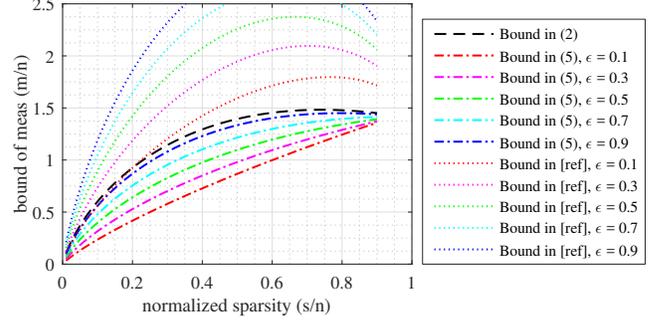
Assuming the encoder has access to  $\mathbf{w}$ , the side information, we design the measurement matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  as follows: each row of  $\mathbf{A}$  is generated independently as a realization of a random Gaussian vector with zero mean and covariance matrix  $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_i, \dots, \sigma_n) \in \mathbb{R}^{n \times n}$ , where each variance  $\sigma_i$  is given by

$$\sigma_i = \begin{cases} 1 & \text{if } w_i \neq 0 \\ \varepsilon \in (0, 1] & \text{if } w_i = 0. \end{cases} \quad (3)$$

In (3),  $\varepsilon$  is a predefined parameter that determines the gains with respect to the no-side-information case. Driven by the side information  $\mathbf{w}$ , the intuition of setting  $\varepsilon \leq 1$  comes from energy considerations: less energy should be spent on acquiring the components of  $\mathbf{x}^*$  that the side information indicates to be zero.



**Fig. 2.** Visualization of the mismatch parameters  $r$  and  $v$ .



**Fig. 3.** Comparison of our theoretical bound (5) with the classic CS bound (2), and Mansour and Saab's bound [25], with  $r/s = 0.1$ ,  $v/s = 0.1$ .

### 3.2. Main result: CS with side information at the encoder

In this section, we analyse the performance of the resulting CS system and present a new bound on the number of measurements required for successful reconstruction when the measurement matrix is designed as above.

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

$$r := |\{i : x_i^* \neq 0, w_i = 0\}|, \quad (4a)$$

$$v := |\{i : x_i^* = 0, w_i \neq 0\}|. \quad (4b)$$

In other words,  $r$  counts the number of components missed by  $\mathbf{w}$ , and  $v$  counts the number of components that  $\mathbf{w}$  overestimates as shown in Fig. 2. Denoting the sparsity of  $\mathbf{x}^*$  by  $s$  and the sparsity of  $\mathbf{w}$  by  $s_w$ , there holds  $0 \leq r \leq s$  and  $0 \leq v \leq \min\{s_w, n - s\}$ .

**Proposition 1.** *Let  $\mathbf{x}^* \in \mathbb{R}^n$  be the signal of interest with sparsity  $s := |\{i : x_i^* \neq 0\}|$ , and  $\mathbf{w} \in \mathbb{R}^n$  be the side information. Let  $r > 0$  and  $v > 0$ , defined as (4a) and (4b), denote the two types of mismatch between  $\mathbf{x}^*$  and  $\mathbf{w}$ . Given a vector of measurements  $\mathbf{y} = \mathbf{A}\mathbf{x}^*$ , where  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is designed as in Section 3.1, then  $\mathbf{x}^*$  is the unique optimal solution of (1) with probability at least  $1 - \exp(-\frac{1}{2}(\lambda_m - \omega(\Omega))^2)$ , provided*

$$m \geq 2(\varepsilon s + (1 - \varepsilon)r) \ln\left(\frac{n}{s}\right) + \frac{7}{5}s + (1 - \sqrt{\varepsilon})v + 1. \quad (5)$$

**Remark 1.** Proposition 1 establishes that our measurement matrix design scheme reduces the number of required measurements from  $O(s \ln n)$  [cf.(2)] to  $O((\varepsilon s + (1 - \varepsilon)r) \ln n)$ . 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, as the dominant logarithmic term in (5) is reduced remarkably compared to the counterpart term in the CS bound (2), as shown in (6a). On the other hand, (6b) ensures that the unfavourable

increase from the non-dominant linear term in (5) is limited to no more than the mismatch  $v$ .

$$0 \leq \varepsilon \leq 1; r \leq s \implies \varepsilon s + (1 - \varepsilon)r \leq s. \quad (6a)$$

$$0 \leq \varepsilon \leq 1; v \geq 0 \implies v \geq (1 - \sqrt{\varepsilon})v \geq 0. \quad (6b)$$

**Remark 2.** Proposition 1 also shows that  $\varepsilon$  determines the sensitivity of the bound to the quality of the side information. For high quality side information with small mismatch  $r$  and  $v$ , a smaller  $\varepsilon$  should be selected. Otherwise, a larger  $\varepsilon$  is more favourable.

**Remark 3.** It can also be seen that our bound (5) generalizes the classical CS bound (2). Concretely, as shown in Fig. 3, bound (5) asymptotically approaches the classical CS bound (2) with the increase of  $\varepsilon$ . Note that, taking  $\varepsilon = 1$ , (5) simplifies to (2), as  $\varepsilon \cdot s + (1 - \varepsilon) \cdot r = s$  and  $(1 - \sqrt{\varepsilon}) \cdot v = 0$ .

**Remark 4.** Section 3.3 will show that solving BP with our measurement matrix design as in Proposition 1 is equivalent to solving a weighted  $\ell_1$  minimization problem at the decoder. In that context, Mansour and Saab's work [25] established a result similar to Proposition 1; see Theorem 5 in [25]. Their result is, however, considerably more complicated and looser than ours, as shown in Fig. 3. To compare them roughly, assume  $n$  is large enough and set  $\varepsilon \rightarrow 0$ . The dominant term of the bound in (5) becomes  $2r \ln n$ , and the dominant term of the bound in [25, Thm.5] becomes  $2 \left[ (r+v) \ln n + (r+v) \sqrt{\ln n} + \sqrt{(r+v) \ln n} \right]$ , where  $r$  and  $v$  are as in (4).

### 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 BP optimization with our measurement matrix design to a weighted  $\ell_1$  minimization with an i.i.d. Gaussian matrix and computing the upper bound for the Gaussian width/distance.

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

$$\begin{aligned} \min_{\mathbf{x}} \|\mathbf{x}\|_1 &\iff \min_{\mathbf{x}} \|\mathbf{x}\|_1 \\ \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x} &\quad \text{s.t. } \mathbf{y} = (\mathbf{A}\mathbf{D})(\mathbf{D}^{-1}\mathbf{x}) \\ &\iff \min_{\mathbf{z}} \|\mathbf{D}\mathbf{z}\|_1 \\ &\quad \text{s.t. } \mathbf{y} = \tilde{\mathbf{A}}\mathbf{z} \end{aligned} \quad (7)$$

where  $\tilde{\mathbf{A}} := \mathbf{A}\mathbf{D}$  and  $\mathbf{z} = \mathbf{D}^{-1}\mathbf{x}$  for  $\mathbf{D} := \text{diag}(d_1, \dots, d_n)$ .

Note that from (1) to (7), the optimization variable is changed from  $\mathbf{x}$  to  $\mathbf{z} = \mathbf{D}^{-1}\mathbf{x}$ , where the weights 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. Gaussian matrix  $\tilde{\mathbf{A}}$ , that is isotropic in this case. In addition, the diagonal form of  $\mathbf{D}$  ensures that the optimal solution  $\mathbf{z}^*$  of (7) has the same support as  $\mathbf{x}^*$ . Recall that the random column vector  $\mathbf{X}$  is called isotropic if  $\mathbb{E}\mathbf{X}\mathbf{X}^T = \mathbf{I}_n$ . In order to make  $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{D}$  isotropic, each row of  $\tilde{\mathbf{A}}$  needs to satisfy  $\mathbb{E}\tilde{\mathbf{a}}_i^T \tilde{\mathbf{a}}_i = \mathbf{I}_n$ , where  $\tilde{\mathbf{a}}_i = [\tilde{a}_{i1}, \dots, \tilde{a}_{ij}, \dots, \tilde{a}_{in}]$  is the  $i$ -th row vector of  $\tilde{\mathbf{A}}$ . Thus, for each element, there holds

$$\mathbb{E}[\tilde{a}_{ij}^2] = \mathbb{E}[a_{ij}^2 \cdot d_i^2] = d_i^2 \cdot [\mathbb{D}a_{ij} + (\mathbb{E}a_{ij})^2] = d_i^2 \cdot \sigma_i = 1,$$

since  $\mathbb{E}a_{ij} = 0$ . The variance  $\sigma_i = \mathbb{D}a_{ij}$  is the  $i$ -th element in the diagonal of the covariance matrix  $\Sigma$ , defined as in (3). Finally, we

take  $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$  with

$$d_i = \frac{1}{\sqrt{\sigma_i}} = \begin{cases} 1; & \text{if } w_i \neq 0 \\ 1/\sqrt{\varepsilon}; & \text{if } w_i = 0 \end{cases} \quad 0 < \varepsilon \leq 1. \quad (8)$$

**Computation of the bounds.** The required number of measurements  $m$  that guarantees successful reconstruction is upper bounded by the Gaussian width [28, 29]. As it is difficult to compute the Gaussian width in a closed form, an upper bound based on Gaussian distance is computed instead. Concretely,  $f(\mathbf{z}) = \|\mathbf{D}\mathbf{z}\|_1$  is a convex function. Suppose  $\mathbf{0} \notin \partial f(\mathbf{z}^*)$  for a given  $\mathbf{z}^* \in \mathbb{R}^n$ , then

$$m \geq \mathbb{E}_{\mathbf{g}} \left[ \text{dist}(\mathbf{g}, \text{cone } \partial f(\mathbf{z}^*))^2 \right] + 1 \implies m \geq w(\Omega)^2 + 1 \quad (9)$$

guarantees perfect recovery with high probability, where  $\text{cone } \partial f(\mathbf{z}^*)$  is the cone generated by the subdifferential of the objective function  $f(\mathbf{z})$  at the optimal point  $\mathbf{z}^*$ . Let  $\text{dist}(\mathbf{g}, S) := \min\{\|\mathbf{z} - \mathbf{g}\|_2 : \mathbf{z} \in S\}$  denote the Euclidean distance between a point  $\mathbf{g}$  and the set  $S$ . To compute an upper bound for the Gaussian distance in (9), the objective is decomposed as  $f(\mathbf{z}) = \|\mathbf{D}\mathbf{z}\|_1 = \sum_{i=1}^n f^{(i)}(z_i) = \sum_{i=1}^n |d_i z_i|$  and the corresponding cone is computed as

$$\begin{aligned} \text{cone } \partial f(\mathbf{z}^*) &= \left( t \cdot \partial f^{(1)}(z_1^*), t \cdot \partial f^{(2)}(z_2^*), \dots, t \cdot \partial f^{(n)}(z_n^*) \right), \\ t \cdot \partial f^{(i)}(z_i^*) &= \begin{cases} t d_i \text{sign}(d_i z_i^*); & \text{if } i \in I \\ \mathcal{I}(0, t d_i) =: [-t d_i, t d_i]; & \text{if } i \notin I \end{cases} \end{aligned}$$

where  $i = 1, \dots, n$ ,  $I := \{i : z_i^* \neq 0\}$  is the support of  $\mathbf{z}^*$ , and  $\mathcal{I}(a, b)$  denotes an interval with centre  $a$  and radius  $b$ . Then, Jensen's inequality is applied to derive that

$$\begin{aligned} \mathbb{E}_{\mathbf{g}} \left[ \text{dist}(\mathbf{g}, \text{cone } \partial f(\mathbf{z}^*))^2 \right] &\leq \sum_{i \in I} \mathbb{E}_{g_i} \left[ \text{dist}(g_i, t d_i \text{sign}(d_i z_i^*))^2 \right] \quad (10a) \\ &+ \sum_{i \notin I} \mathbb{E}_{g_i} \left[ \text{dist}(g_i, \mathcal{I}(0, t d_i))^2 \right] \quad (10b) \end{aligned}$$

which holds for arbitrary  $t > 0$ . In our case,  $t = \sqrt{2\varepsilon \ln(n/s)}$  is selected to compute the upper bound.

For simplicity, an auxiliary function  $A(x)$  for  $x \in \mathbb{R}^n$  is defined as

$$A(x) = -x\varphi(x) + (1 + x^2)Q(x). \quad (11)$$

where  $\varphi(x) = \exp(-x^2/2)/\sqrt{2\pi}$  is the probability density function of a scalar Gaussian random variable with zero-mean and unit variance, and  $Q(x) = \int_x^{+\infty} \varphi(t) dt$  is the  $Q$ -function. Then, it is proved 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

$$\mathbb{E}_g \left[ \text{dist}(g, \mathcal{I}(a, b))^2 \right] = A(b - |a|) + A(b + |a|). \quad (12)$$

Specifically, when  $b = 0$ , interval  $\mathcal{I}(a, 0)$  reduces to a point  $a$ , and

$$\mathbb{E}_g \left[ \text{dist}(g, a)^2 \right] = A(-|a|) + A(|a|) = a^2 + 1. \quad (13)$$

When the interval is centered at 0, i.e.,  $a = 0$ , there holds

$$\mathbb{E}_g \left[ \text{dist}(g, \mathcal{I}(0, b))^2 \right] = 2A(b) \quad (14)$$

Then, (13) and (14) are plugged into (10a) and (10b) to obtain

$$(10a) = \sum_{i \in I} [1 + t^2 d_i^2] = s + 2(\varepsilon s + (1 - \varepsilon)r) \ln \left( \frac{n}{s} \right),$$

$$(10b) = \sum_{i \notin I} 2A(t d_i) = 2(n - s - v)A(t) + 2vA(t\sqrt{\varepsilon}) \\ = 2(n - s)A(t) + 2vA(t\sqrt{\varepsilon}) - 2vA(t) \leq \frac{2}{5}s + (1 - \sqrt{\varepsilon})v,$$

where the first term involving  $2s/5$  is derived from Lemma 3 in [20], i.e.,  $\frac{1-1/x}{\sqrt{\pi \ln(x)}} \leq \frac{2}{5}$  for all  $x > 1$ , and the last term is from the property of function  $A(x)$ , that is,  $0 \leq \max_{x \geq 0} (A(\alpha \cdot x) - A(x)) \leq (1 - \alpha)/2$  for  $0 < \alpha \leq 1$ . Finally, (10a) and (10b) are combined to obtain the upper bound (5).

#### 4. EXPERIMENTAL RESULTS

A set of experiments has been conducted to evaluate the efficacy of the proposed measurement matrix design. An i.i.d. Gaussian matrix is used as the benchmark. In the signal reconstruction stage, we use the SPGL1 Toolbox [31] to solve the same BP with both an i.i.d. Gaussian matrix and our measurement matrix design. The parameter setting is shown in Table 1.  $m$  varies from 10 to 1000 with a step of 10 and the relative sparsity  $s/n$  varies from 0.05 to 0.6 with a step of 0.05 so that the number of measurements corresponding to the success rate exceeding 85% can be found as the empirical threshold. To indicate the phase transitions,  $s/m$  varies from 0.1 to 1 with a step of 0.05. In these experiments, 100 instances of  $\mathbf{A}$  are generated for each pair of  $s$  and  $m$ .

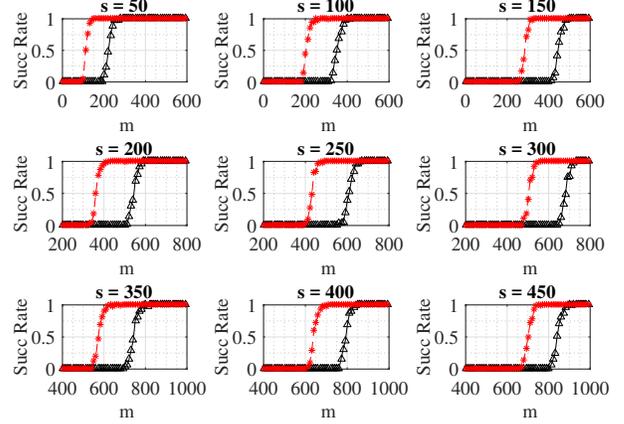
The experiment results are shown in Fig. 4. Fig. 4(a) shows the success rate varying as a function of the number of measurements for different sparsity levels. Fig. 4(b) compares the theoretical bounds and empirical bounds. Fig. 4(c) and Fig. 4(d) compares the phase transition for the two types of measurement matrix. 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 phase transition well.

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

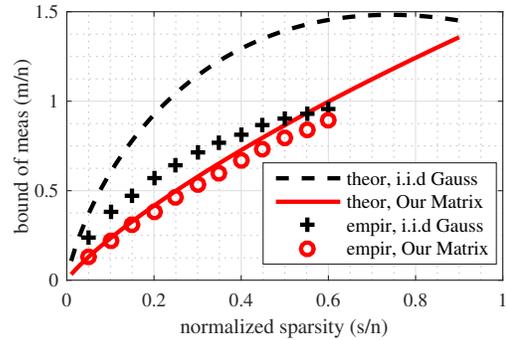
$n$	$m/n$	$s/n$	$s/m$	$r/s$	$v/s$	$\varepsilon$
1000	0.01:1	0.05:0.6	0.1:1	0.1	0.1	0.1

#### 5. CONCLUSIONS

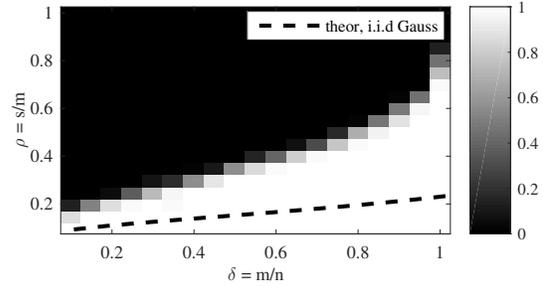
In this paper, a new measurement matrix design scheme is proposed to integrate side information at the encoder into a CS system. Based on the design scheme, the performance of the resulting system is analysed in terms of the number of measurements required for perfect reconstruction. Extensive experiments indicate that our measurement matrix design allows reducing the number of measurements provided the side information has reasonable quality. In addition, it is demonstrated that our theoretical bound is simple and tight. We believe this work can contribute to better design of CS systems in scenarios where side information is available, such as medical imaging, sensor networks, and multi-view camera systems. In the future, we will consider not only the case where the measurements are noisy, but also the scenario where the side information is available at both the encoder and the decoder.



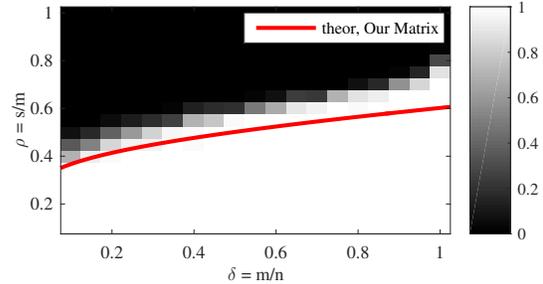
(a) Success ratio with measurements for different sparsity levels.



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



(c) Phase transition for i.i.d. Gaussian matrix.



(d) Phase transition for proposed measurement matrix.

**Fig. 4.** Experimental results. (a) In each sub-figure, the black line with marker  $\triangle$  represents the recovery performance of i.i.d. Gaussian matrix and BP, the red line with marker  $*$  represents our measurement matrix design with BP. In (c) and (d), the colorbar represents the success ratio. For each pair of  $m/n$  and  $s/m$ , the higher the success ratio is, the whiter the point is.

## 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] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing to rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, pp. 1182–1195, 2007.
- [4] M. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, 2009.
- [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] V. Cevher, A. Sankaranarayanan, M. Duarte, D. Reddy, R. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," in *European Conf. on Computer Vision (ECCV)*, 2008.
- [7] J. F. C. Mota, N. Deligiannis, A. C. Sankaranarayanan, V. Cevher, and M. R. D. Rodrigues, "Dynamic sparse state estimation using  $\ell_1$ - $\ell_1$  minimization: Adaptive-rate measurement bounds, algorithms, and applications," in *IEEE Intern. Conf. Acoustics, Speech, and Sig. Processing (ICASSP)*, 2015.
- [8] J. F. C. Mota, N. Deligiannis, A. C. Sankaranarayanan, V. Cevher, and M. R. D. Rodrigues, "Adaptive-rate sparse signal reconstruction with application in compressive foreground extraction," *IEEE Trans. Signal Process.*, 2015, (in press).
- [9] 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.
- [10] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.
- [11] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4595–4607, 2010.
- [12] M. Amin Khajehnejad, Weiyu Xu, Babak Hassibi, et al., "Weighted  $\ell_1$  minimization for sparse recovery with prior information," in *IEEE Intern. Symp. Inform. Theory (ISIT)*.
- [13] S. Oymak, M. A. Khajehnejad, and B. Hassibi, "Recovery threshold for optimal weight  $\ell_1$  minimization," in *IEEE Intern. Symp. Inform. Theory (ISIT)*. IEEE, 2012, pp. 2032–2036.
- [14] W. Lu and N. Vaswani, "Exact reconstruction conditions and error bounds for regularized modified basis pursuit (reg-modified-bp)," in *Asilomar Conf. Signals, Systems and Computers (ASILOMAR)*.
- [15] 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.
- [16] 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.
- [17] E. Zimos, J. Mota, M. Rodrigues, and N. Deligiannis, "Bayesian compressed sensing with heterogeneous side information," in *2016 IEEE Data Compression Conference (DCC2016)*, Snowbird, Utah, USA, Apr. 2016.
- [18] 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.
- [19] X. Wang and J. Liang, "Side information-aided compressed sensing reconstruction via approximate message passing," Preprint: <http://arxiv.org/abs/1311.0576v1>, 2013.
- [20] 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.
- [21] 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.
- [22] F. Renna, L. Wang, X. Yuan, J. Yang, G. Reeves, R. Calderbank, L. Carin, and M. R. D. Rodrigues, "Classification and reconstruction of high-dimensional signals from low-dimensional noisy features in the presence of side information," <http://arxiv.org/abs/1412.0614>, 2014.
- [23] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: A new theory for compressed sensing," Preprint: <http://arxiv.org/abs/1302.0561>, 2014.
- [24] W. Chen, M. R. D. Rodrigues, and I. J. Wassell, "Projections design for statistical compressive sensing: A tight frame based approach," *IEEE Trans. Signal Process.*, vol. 61, pp. 2016–2029, 2013.
- [25] H. Mansour and R. Saab, "Recovery analysis for weighted  $\ell_1$ -minimization using a null space property," *arXiv preprint arXiv:1412.1565*, 2014.
- [26] Y. Gordon, *On Milman's inequality and random subspaces which escape through a mesh in  $\mathbb{R}^n$* , Springer, 1988.
- [27] 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.
- [28] 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.
- [29] D. Amelunxen, M. Lotz, M. McCoy, and J. Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Information and Inference: A Journal of the IMA*, pp. 1–71, 2014.
- [30] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [31] Ewout Van Den Berg and Michael P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.