

# Improved Recovery of Analysis Sparse Vectors in Presence of Prior Information

Sajad Daei, Farzan Haddadi, Arash Amini

**Abstract**—In this work, we consider the problem of recovering analysis-sparse signals from under-sampled measurements when some prior information about the support is available. We incorporate such information in the recovery stage by suitably tuning the weights in a weighted  $\ell_1$  analysis optimization problem. Indeed, we try to set the weights such that the method succeeds with minimum number of measurements. For this purpose, we exploit the upper-bound on the statistical dimension of a certain cone to determine the weights. Our numerical simulations confirm that the introduced method with tuned weights outperforms the standard  $\ell_1$  analysis technique.

**Index Terms**— $\ell_1$  analysis, prior information, conic integral geometry.

## I. INTRODUCTION

COMPRESSED sensing (CS), initiated by [1], [2], has been the focus of many research works for more than a decade. Briefly, CS, in its general form, investigates the reconstruction of a sparse vector  $\mathbf{x} \in \mathbb{R}^n$  from  $m \ll n$  noisy linear measurements

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon} \in \mathbb{R}^m \quad (1)$$

where  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is a known matrix and  $\boldsymbol{\epsilon}$  is an  $\ell_2$  bounded noise term, i.e.  $\|\boldsymbol{\epsilon}\|_2 \leq \eta$  for some  $\eta \geq 0$ . In many scenarios,  $\mathbf{x}$  is sparse after the application of some analysis operator  $\boldsymbol{\Omega}$ . Specifically, we say  $\mathbf{x}$  is  $s$ -analysis-sparse with support  $\mathcal{S}$  in the analysis domain  $\boldsymbol{\Omega} \in \mathbb{R}^{p \times n}$  if  $\boldsymbol{\Omega}\mathbf{x}$  is  $s$ -sparse with support  $\mathcal{S}$ . Then, the following optimization problem called  $\ell_1$  analysis is often used (See [3], [4], and [5]) to recover  $\mathbf{x}$ :

$$\mathbf{P}_\eta : \min_{\mathbf{z} \in \mathbb{R}^n} \|\boldsymbol{\Omega}\mathbf{z}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2 \leq \eta \quad (2)$$

In many applications, there is some additional information in the analysis domain. For instance, consider the line spectral estimation where the signal of interest is sparse after applying the Discrete Fourier Transform. In some applications, one might a priori know the probability with which a set in the spectral domain contributes to the true line spectra. The extra information about the probability of contribution of certain subsets could be beneficial in the recovery of  $\mathbf{x}$ . For example, for channel estimation in communication systems or in remote sensing, the availability of previous estimates builds a history that can specify the intersection probability of any given set with the true support. Also, natural images often tend to have larger values in lower frequencies after applying Fourier or wavelet transforms; therefore, subsets composed of low-frequencies have higher probabilities of appearing in the support. In these cases, we intend to exploit these additional information. This work analyses possible benefits of this extra

information to reduce the required number of measurements of  $\mathbf{P}_0$  for successful recovery and to improve the reconstruction error in  $\mathbf{P}_\eta$  for robust and stable recovery. For this purpose, a common way is to use weighted  $\ell_1$  analysis as follows:

$$\mathbf{P}_\eta^w : \min_{\mathbf{z} \in \mathbb{R}^n} \|\boldsymbol{\Omega}\mathbf{z}\|_{1,w} := \sum_{i=1}^p w_i |\boldsymbol{\Omega}\mathbf{z}|_i \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2 \leq \eta, \quad (3)$$

where  $w_i$  represents the weight associated with the  $i$ 'th element of the coefficient vector in the analysis domain. In this work, we assume that the available prior information is about the subsets  $\{\mathcal{P}_i\}_{i=1}^L$  that partition  $\{1, \dots, p\}$ . Thus, the elements of  $\mathcal{P}_i$  are all assigned the same weight ( $\omega_i$ ). Moreover, we define

$$\mathbf{w} = \sum_{i=1}^L \omega_i \mathbf{1}_{\mathcal{P}_i}, \quad \alpha_i = \frac{|\mathcal{P}_i \cap \mathcal{S}|}{|\mathcal{P}_i|}, \quad \rho_i = \frac{|\mathcal{P}_i|}{p}, \quad (4)$$

where  $|\cdot|$  denotes the cardinality of a set and  $\mathbf{1}_{\mathcal{E}}$  is the indicator function of the set  $\mathcal{E}$ . The parameters  $\alpha_i$  and  $\rho_i$  are commonly called the accuracy and the normalized size of the subsets, respectively. Alternatively,  $\{\mathcal{P}_i\}_{i=1}^L$  can be considered as  $L$  analysis support estimators with different accuracies  $\{\alpha_i\}_{i=1}^L$ . Our goal is to find the weights that minimize the required number of measurements. To this end, we first find an upper-bound for the required number of measurements in Proposition 1. Then, we minimize the upper-bound with respect to the weights. Since the bound is not tight (especially in redundant and coherent dictionaries), we can not claim optimality of the weights. However, with the obtained weights, we almost achieve the optimal phase transition curve of  $\ell_1$  analysis problem in low-redundant analysis operators in numerical simulations. The paper is organized as follows: a brief overview of convex geometry is given in Section II. We explain our main contribution in Section III followed by numerical experiments in Section IV. Indeed, the experiments confirm the theoretical results.

Throughout the paper, scalars are denoted by lowercase letters, vectors by lowercase boldface letters, and matrices by uppercase boldface letters. The  $i$ th element of a vector  $\mathbf{x}$  is shown either by  $x(i)$  or  $x_i$ .  $(\cdot)^\dagger$  denotes the pseudo inverse operator. We reserve the calligraphic uppercase letters for sets (e.g.  $\mathcal{S}$ ). The cardinality of a set  $\mathcal{S}$  is denoted by  $|\mathcal{S}|$ .  $\mathcal{C}^\circ$  represents the polar of a cone  $\mathcal{C}$ . Given a vector  $\mathbf{x} \in \mathbb{R}^n$  and a set  $\mathcal{C} \subseteq \mathbb{R}^n$ ,  $\mathbf{x} \odot \mathcal{C}$  denotes the set  $\mathcal{C}$  which is scaled by the elements of  $\mathbf{x}$ . In this work,  $\mathbf{1}_{\mathcal{E}}$  denotes the indicator of the set  $\mathcal{E}$ .  $(a)_+$  stands for  $\max\{a, 0\}$  for a scalar  $a$ . Null space and range of linear operators are denoted by  $\text{null}(\cdot)$ , and  $\text{range}(\cdot)$ ,

respectively. For a matrix  $\mathbf{A}$ , the operator norm is defined as  $\|\mathbf{A}\|_{p \rightarrow q} = \sup_{\|\mathbf{x}\|_p \leq 1} \|\mathbf{A}\mathbf{x}\|_q$ . We denote i.i.d standard Gaussian random vector by  $\mathbf{g}$ . Lastly,  $\|\cdot\|_\infty$  returns the maximum absolute value of the elements of a vector or matrix.

## II. CONVEX GEOMETRY

In this section, basic concepts of conic integral geometry are reviewed.

### A. Descent Cones and Statistical dimension

The descent cone of a proper convex function  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$  at point  $\mathbf{x} \in \mathbb{R}^n$  is the set of directions from  $\mathbf{x}$  that do not increase  $f$ :

$$\mathcal{D}(f, \mathbf{x}) = \bigcup_{t \geq 0} \{\mathbf{z} \in \mathbb{R}^n : f(\mathbf{x} + t\mathbf{z}) \leq f(\mathbf{x})\}. \quad (5)$$

The descent cone of a convex function is a convex set. There is a famous duality [6, Ch. 23] between decent cone and subdifferential of a convex function given by:

$$\mathcal{D}^\circ(f, \mathbf{x}) = \text{cone}(\partial f(\mathbf{x})) := \bigcup_{t \geq 0} t \cdot \partial f(\mathbf{x}). \quad (6)$$

**Definition 1.** Statistical Dimension [7]: Let  $\mathcal{C} \subseteq \mathbb{R}^n$  be a convex closed cone. Statistical dimension of  $\mathcal{C}$  is defined as:

$$\delta(\mathcal{C}) := \mathbb{E}\|\mathcal{P}_{\mathcal{C}}(\mathbf{g})\|_2^2 = \mathbb{E}\text{dist}^2(\mathbf{g}, \mathcal{C}^\circ), \quad (7)$$

where  $\mathbf{g}$  has i.i.d. standard normal distribution, and  $\mathcal{P}_{\mathcal{C}}(\mathbf{x})$  is the orthogonal projection of  $\mathbf{x} \in \mathbb{R}^n$  onto the set  $\mathcal{C}$  defined as:  $\mathcal{P}_{\mathcal{C}}(\mathbf{x}) = \arg \min_{\mathbf{z} \in \mathcal{C}} \|\mathbf{z} - \mathbf{x}\|_2$ .

Statistical dimension specifies the boundary of success and failure in random convex programs with affine constraints.

## III. MAIN RESULTS

In this section, we first present an upper-bound for the required number of Gaussian measurements for the case of a redundant analysis operator  $\mathbf{\Omega}$ . A lower-bound is also derived for non-singular  $\mathbf{\Omega}$ . The lower-bound is not new and was previously reported in [8, Theorem A], but here we present a simpler approach for the proof.

**Proposition 1.** Let  $\mathbf{x} \in \mathbb{R}^n$  be a  $s$ -analysis sparse vector with redundant analysis operator  $\mathbf{\Omega} \in \mathbb{R}^{p \times n}$  ( $p \geq n$ ). Then,

$$\delta(\mathcal{D}(\|\mathbf{\Omega} \cdot\|_{1,w}, \mathbf{x})) \leq \kappa^2(\mathbf{\Omega}) \delta(\mathcal{D}(\|\cdot\|_{1,w}, \mathbf{\Omega}\mathbf{x})). \quad (8)$$

Moreover, if  $\mathbf{\Omega}$  is non-singular and  $p = n$ ,

$$\frac{1}{\kappa^2(\mathbf{\Omega})} \delta(\mathcal{D}(\|\cdot\|_{1,w}, \mathbf{\Omega}\mathbf{x})) \leq \delta(\mathcal{D}(\|\mathbf{\Omega} \cdot\|_{1,w}, \mathbf{x})) \leq \kappa^2(\mathbf{\Omega}) \delta(\mathcal{D}(\|\cdot\|_{1,w}, \mathbf{\Omega}\mathbf{x})) \quad (9)$$

Proof. See Appendix A.

**Theorem 1.** Let  $\mathbf{x} \in \mathbb{R}^n$ . Let the entries of  $\mathbf{A} \in \mathbb{R}^{m \times n}$  be a random matrix with entries drawn from an i.i.d. standard normal distribution. If  $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m$ , and

$$m > \left( \kappa(\mathbf{\Omega}) \sqrt{\delta(\mathcal{D}(\|\cdot\|_{1,w}, \mathbf{\Omega}\mathbf{x}) + t)} \right)^2 + 1, \quad (10)$$

for a given  $t > 0$ , then,  $\mathbf{P}_0^w$  recovers  $\mathbf{x}$  with probability at least  $1 - e^{-\frac{t^2}{2}}$ . Also, if  $\mathbf{y} = \mathbf{A}\mathbf{x} + \epsilon$  and  $\mathbf{\Omega}\mathbf{x}_{ap}$  is the best  $\tilde{s}$ -term approximation of the  $s$ -sparse vector  $\mathbf{\Omega}\mathbf{x}$  ( $s \geq \tilde{s}$ ), then any solution  $\hat{\mathbf{x}}$  of  $\mathbf{P}_\eta^w$  satisfies

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{2\eta}{(\sqrt{m-1} - \kappa(\mathbf{\Omega})\sqrt{\delta(\mathcal{D}(\|\cdot\|_{1,w}, \mathbf{\Omega}\mathbf{x}_{ap}))} - t)_+}, \quad (11)$$

with probability at least  $1 - e^{-\frac{t^2}{2}}$ .

Proof. See Appendix B.

In the exact recovery case, we determine the suitable weights by minimizing the right-hand side of (10). In the noisy setting, for stable and robust recovery, we determine the weights by minimizing the reconstruction error (the right-hand side of 11):

$$\omega^* = \arg \min_{\nu \in \mathbb{R}_+^L} \text{Edist}^2(\mathbf{g}, (\mathbf{D}\nu) \odot \partial\|\cdot\|_1(\mathbf{\Omega}\mathbf{x})), \quad (12)$$

where  $\mathbf{D} := [\mathbf{1}_{\mathcal{P}_1}, \dots, \mathbf{1}_{\mathcal{P}_L}] \in \mathbb{R}^{p \times L}$ . The latter optimization problem is very similar to the one in weighted  $\ell_1$  minimization. With the same approach as in [9], one can show that (12) reduces to solving the following equations simultaneously [9, Corollary 11]:

$$\alpha_i \omega_i^* = (1 - \alpha_i) \sqrt{\frac{2}{\pi}} \int_{\omega_i^*}^{\infty} (u - \omega_i^*) e^{-\frac{u^2}{2}} du : i = 1, \dots, L. \quad (13)$$

It is not obvious whether the inequality (8) in Proposition 1 is tight for highly redundant and coherent analysis operators. However, numerical evidence suggests that the obtained bound is close to the  $\ell_1$  analysis phase transition curve for low-redundancy regime (See Figure 1).

In practice, one may encounter some inaccuracies in determining  $\alpha \in \mathbb{R}^L$ . The study of the sensitivity of weights to the inaccuracies in  $\alpha$  were previously considered in [10]. Fortunately, small changes in  $\alpha$  are shown to have insignificant impact on the derived weights.

## IV. SIMULATION RESULTS

In this section, we numerically study the effect of weights obtained by (13) on the number of required measurements. First, we consider the scaling of the required number of measurements for successful recovery of (2) with analysis sparsity. The heatmap in Figure 1 shows the empirical probability of success. Indeed, the results are consistent with (8). In the second experiment, we generate a  $s = 10$ -analysis sparse random vector  $\mathbf{x} \in \mathbb{R}^{55}$  in two different analysis operators with  $\kappa(\mathbf{\Omega}) = 1.1$  and  $\kappa(\mathbf{\Omega}) = 230$ . We consider two random sets  $\mathcal{P}_1$  and  $\mathcal{P}_2$  that partition the analysis domain  $\{1, \dots, p\}$  with  $\alpha_1 = \frac{7}{10}$  and  $\alpha_2 = \frac{3}{50}$ . The suitable weights are obtained via equation (13) by MATLAB function `fzero`. Figures 2 and 3 show the success rate of  $\mathbf{P}_0^w$  averaged over 50 Monte Carlo simulations. It is evident that the weighted  $\ell_1$  analysis with suitable weights needs less number of measurements than regular  $\ell_1$  analysis.

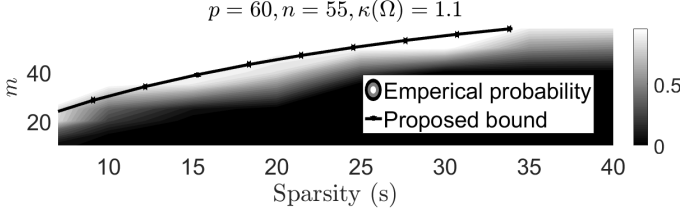


Fig. 1. Empirical probability that problem (2) recovers  $\mathbf{x} \in \mathbb{R}^{55}$  that has  $s$  non-zero entries after applying a redundant analysis operator with  $\kappa(\Omega) = 1.1$ . The black line shows the number of measurements obtained by (8).

In a separate scenario, we repeat the second experiment for redundant Fourier analysis operator which is widely used in line spectral and direction of arrival estimation. In this experiment, the measurements are contaminated with additive noise (SNR = 30 dB). For the recovery, both  $P_\eta$  and  $P_\eta^*$  are implemented. Also, the random analysis support estimates  $\mathcal{P}_1$  and  $\mathcal{P}_2$  are such made that  $\alpha_1 = \frac{8}{10}$  and  $\alpha_2 = \frac{2}{45}$ . The normalized mean square error

$$\text{NMSE} = \frac{\|\hat{\mathbf{x}} - \mathbf{x}\|_2}{\|\mathbf{x}\|_2}, \quad (14)$$

is averaged out of 50 trials. From Figure 5, it is clear that weighted  $\ell_1$  analysis with suitable weights obtained from (13) needs less measurements than regular  $\ell_1$  analysis in a fixed NMSE.

In the last experiment, we investigate a more practical scenario where a Shepp-Logan image (denoted by  $\mathbf{X}$  of size  $n_1 \times n_2$  pixels) is under-sampled with a fat random Gaussian matrix (of size  $m \times n_1 n_2$ ) and passed through an additive noise with SNR=10 dB. Figure 4 illustrates the recovery of this image ( $n_1 = n_2 = 128$ ) by solving 2 and 3 when  $\Omega \in \mathbb{R}^{114688 \times 16384}$  is a redundant wavelet matrix from daubechies family (the weights in (3) are obtained via 13) and in case of  $m = 6554$ . The recovery problems (2) and (3) are carried out using TFOCS algorithm [11]. The quality of each method is reported in terms of the Peak SNR (PSNR) given by:

$$\text{PSNR}(\mathbf{X}, \hat{\mathbf{X}}) := 20 \log_{10} \left( \frac{\|\mathbf{X}\|_\infty \sqrt{n_1 n_2}}{\|\mathbf{X} - \hat{\mathbf{X}}\|_F} \right). \quad (15)$$

We assume 11 disjoint support estimators in the analysis domain with known level of contributing ( $\{\alpha_i\}_{i=1}^{11}$  in (4)) with top 10% (specifying  $\tilde{s}$  in Theorem 1) of wavelet coefficients. As shown by Figure 4, while  $P_\eta^*$  has an acceptable performance with PSNR=18.53 dB,  $P_\eta$  clearly fails with a poor performance PSNR=10.2 dB.

## APPENDIX

### A. Proof of Proposition 1

*Proof.* In the following, we relate  $\mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x})$  to  $\mathcal{D}(\|\cdot\|_1, \Omega \mathbf{x})$ .

$$\begin{aligned} \mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x})^{\circ\circ} &= \text{closure}(\mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x})) \\ \mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x}) &= \text{cone}^\circ(\Omega^T \partial \|\cdot\|_1(\Omega \mathbf{x})) \\ \{\mathbf{w} \in \mathbb{R}^n : \langle \mathbf{w}, \Omega^T \mathbf{v} \rangle \leq 0 : \forall \mathbf{v} \in \text{cone}(\partial \|\cdot\|_1(\Omega \mathbf{x}))\} &= \end{aligned}$$

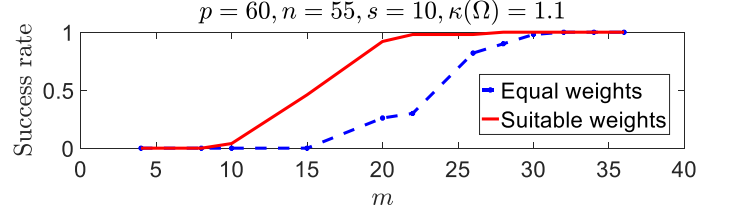


Fig. 2. The probability that (2) and (3) succeed to recover  $\mathbf{x} \in \mathbb{R}^{55}$  from Gaussian linear measurements.  $p = 60, n = 55, s = 10, \kappa(\Omega) = 1.1$ . Suitable weights used in (3), are obtained from (13).

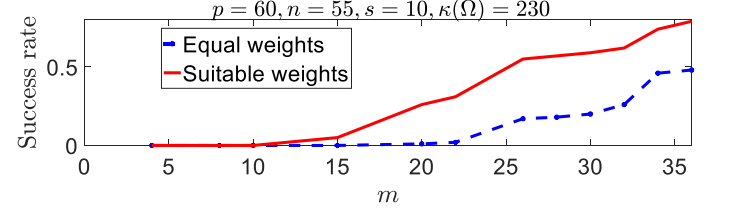


Fig. 3. The probability that (2) and (3) succeed to recover  $\mathbf{x} \in \mathbb{R}^{55}$  from Gaussian linear measurements.  $p = 60, n = 55, s = 10, \kappa(\Omega) = 230$ . Suitable weights used in (3), are obtained from (13).

$$\begin{aligned} \{\mathbf{w} \in \mathbb{R}^n : \Omega \mathbf{w} \in \text{cone}^\circ(\partial \|\cdot\|_1(\Omega \mathbf{x}))\} &= \\ \{\mathbf{w} \in \mathbb{R}^n : \Omega \mathbf{w} \in \mathcal{D}(\|\cdot\|_1, \Omega \mathbf{x})\}. & \end{aligned} \quad (16)$$

Therefore,

$$\Omega \mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x}) \subset \mathcal{D}(\|\cdot\|_1, \Omega \mathbf{x}). \quad (17)$$

In particular, if  $\Omega$  is non-singular and  $p = n$ ,

$$\Omega \mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x}) = \mathcal{D}(\|\cdot\|_1, \Omega \mathbf{x}), \quad (18)$$

where in the last line of (16), we used the fact that  $\mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x})$  is a closed convex set. In the following, we state Sudakov-Fernique inequality which helps to control the supremum of a random process by that of a simpler random process and is used to find an upper-bound for  $\delta(\mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x}))$ .

**Theorem 2. (Sudakov-Fernique inequality).** Let  $T$  be a set and  $\mathbf{X} = (X_t)_{t \in T}$  and  $\mathbf{Y} = (Y_t)_{t \in T}$  be Gaussian processes satisfying  $\mathbb{E}[X_t] = \mathbb{E}[Y_t] : \forall t \in T$  and  $\mathbb{E}|X_t - X_s|^2 \leq \mathbb{E}|Y_t - Y_s|^2 : \forall s, t \in T$ , then

$$\mathbb{E} \sup_{t \in T} X_t^2 \leq \mathbb{E} \sup_{t \in T} Y_t^2. \quad (19)$$

$$\begin{aligned} \delta(\mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x})) &:= \mathbb{E} \left( \sup_{\substack{\mathbf{w} \in \mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x}) \\ \|\mathbf{w}\|_2 \leq 1}} \langle \mathbf{g}, \mathbf{w} \rangle \right)^2 \leq \|\Omega\|_{2 \rightarrow 2}^2 \\ \mathbb{E} \left( \sup_{\substack{\mathbf{w} \in \mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x}) \\ \|\Omega\|_{2 \rightarrow 2} \|\mathbf{v}\|_2 \leq 1}} \langle \mathbf{g}, \mathbf{v} \rangle \right)^2 &\leq \|\Omega\|_{2 \rightarrow 2}^2 \|\Omega^\dagger\|_{2 \rightarrow 2}^2 \\ \mathbb{E} \left( \sup_{\substack{\mathbf{w} \in \mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x}) \\ \|\Omega\|_{2 \rightarrow 2} \|\mathbf{v}\|_2 \leq 1}} \langle \mathbf{h}, \Omega \mathbf{v} \rangle \right)^2 &\kappa^2(\Omega) \mathbb{E} \left( \sup_{\substack{\mathbf{w} \in \mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x}) \\ \|\Omega \mathbf{v}\|_2 \leq 1}} \langle \mathbf{h}, \Omega \mathbf{v} \rangle \right)^2 \\ &\leq \kappa^2(\Omega) \mathbb{E} \left( \sup_{\substack{\mathbf{z} \in \Omega \mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x}) \\ \|\mathbf{z}\|_2 \leq 1}} \langle \mathbf{h}, \mathbf{z} \rangle \right)^2 \leq \end{aligned}$$

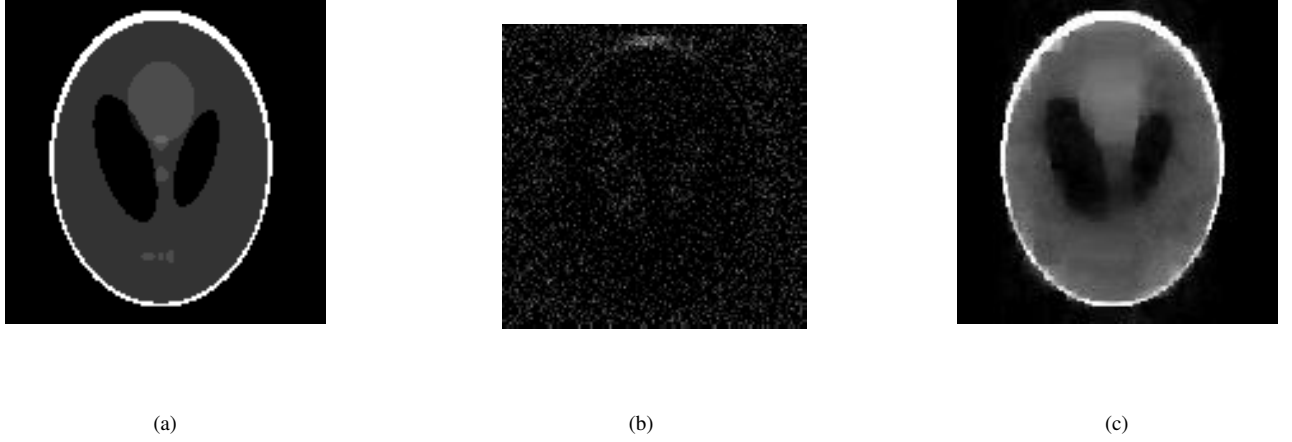


Fig. 4. Effect of suitable weights in analysis sparse recovery with prior information. The associated parameters are  $n_1 = n_2 = 128$ , and  $m = 6554$ . (a) Ground truth image (PSNR= $\infty$  dB). (b) Recovered image via  $\ell_1$  analysis (PSNR=10.2 dB) (c) Recovered image with weighted  $\ell_1$  analysis with suitable weights (PSNR=18.53 dB).

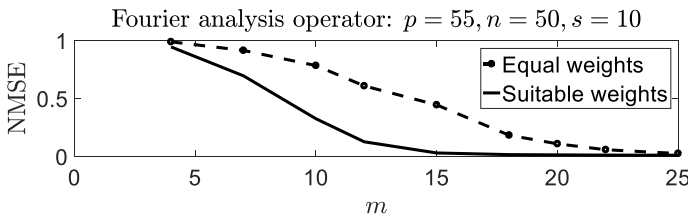


Fig. 5. NMSE versus number of required measurements in problems (2) and (3) in the case that the ground truth signal  $\mathbf{x} \in \mathbb{R}^{50}$  is sparse after applying Fourier analysis operator  $\Omega \in \mathbb{C}^{55 \times 50}$ . NMSE is computed out of 50 Monte Carlo simulations.

$$\kappa^2(\Omega) \mathbb{E} \left( \sup_{\substack{\mathbf{z} \in \mathcal{D}(\|\cdot\|_1, \Omega \mathbf{x}) \\ \|\mathbf{z}\|_2 \leq 1}} \langle \mathbf{h}, \mathbf{z} \rangle \right)^2 = \kappa^2(\Omega) \delta(\mathcal{D}(\|\cdot\|_1, \Omega \mathbf{x})), \quad (20)$$

where in (20),  $\mathbf{h} \in \mathbb{R}^p$  is a standard normal vector with i.i.d components. In the first inequality of (20), we used the change of variable  $\mathbf{v} = \|\Omega\|_{2 \rightarrow 2}^{-1} \mathbf{w}$ . The second inequality comes from Theorem 2 with  $X_v = \langle \mathbf{g}, \mathbf{v} \rangle$  and  $Y_v = \|\Omega^\dagger\|_{2 \rightarrow 2} \langle \mathbf{h}, \Omega \mathbf{v} \rangle$  and the fact that:

$$\begin{aligned} \mathbb{E}|X_v - X_w|^2 &= \|\mathbf{v} - \mathbf{w}\|_2^2 \leq \|\Omega^\dagger\|_{2 \rightarrow 2}^2 \|\Omega(\mathbf{v} - \mathbf{w})\|_2^2 = \\ &= \mathbb{E}|Y_v - Y_w|^2 : \forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^n. \end{aligned} \quad (21)$$

The last inequality comes from (17). In the special case  $p = n$  and  $\Omega$  is non-singular we have:

$$\begin{aligned} \delta(\mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x})) &:= \mathbb{E} \left( \sup_{\substack{\mathbf{w} \in \mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x}) \\ \|\mathbf{w}\|_2 \leq 1}} \langle \mathbf{g}, \mathbf{w} \rangle \right)^2 = \\ \mathbb{E} \left( \sup_{\substack{\mathbf{v} \in \mathcal{D}(\|\cdot\|_1, \Omega \mathbf{x}) \\ \|\Omega^\dagger \mathbf{v}\|_2 \leq 1}} \langle \mathbf{g}, \Omega^\dagger \mathbf{v} \rangle \right)^2 &\geq \end{aligned}$$

$$\begin{aligned} \|\Omega\|_{2 \rightarrow 2}^{-2} \mathbb{E} \left( \sup_{\substack{\mathbf{v} \in \mathcal{D}(\|\cdot\|_1, \Omega \mathbf{x}) \\ \|\Omega^\dagger \mathbf{v}\|_2 \leq 1}} \langle \mathbf{h}, \mathbf{v} \rangle \right)^2 &\geq \\ \|\Omega\|_{2 \rightarrow 2}^{-2} \mathbb{E} \left( \sup_{\substack{\mathbf{v} \in \mathcal{D}(\|\cdot\|_1, \Omega \mathbf{x}) \\ \|\mathbf{v}\|_2 \leq \|\Omega^\dagger\|_{2 \rightarrow 2}^{-1}}} \langle \mathbf{h}, \mathbf{v} \rangle \right)^2 &= \frac{1}{\kappa^2(\Omega)} \delta(\mathcal{D}(\|\cdot\|_1, \Omega \mathbf{x})), \end{aligned} \quad (22)$$

where the first inequality comes from  $\Omega^\dagger \Omega = \mathbf{I}$  and (18). The second inequality comes from Theorem 2 with  $X_v = \langle \mathbf{g}, \Omega^\dagger \mathbf{v} \rangle$  and  $Y_v = \|\Omega\|_{2 \rightarrow 2}^{-1} \langle \mathbf{h}, \mathbf{v} \rangle$  and the fact that

$$\begin{aligned} \mathbb{E}|X_v - X_w|^2 &= \|\Omega^\dagger(\mathbf{v} - \mathbf{w})\|_2^2 \geq \|\Omega\|_{2 \rightarrow 2}^{-2} \|\mathbf{v} - \mathbf{w}\|_2^2 = \\ &= \mathbb{E}|Y_v - Y_w|^2 : \forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^n, \end{aligned} \quad (23)$$

where the last inequality is a result of norm properties. ■

### B. Proof of theorem 1

Let  $T_0$  be the index set of  $\tilde{s}$  largest analysis coefficients. Then, it holds that,

$$\|\Omega \mathbf{x}_{ap}\|_1 := \|(\Omega \mathbf{x})_{T_0}\|_1 \geq \frac{\tilde{s} \|\Omega \mathbf{x}\|_1}{s}, \quad (24)$$

and as a result, we have,  $\mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x}) \subseteq \mathcal{D}(\|\Omega \cdot\|_1, \frac{s}{\tilde{s}} \mathbf{x})$  and thus  $\delta(\mathcal{D}(\|\Omega \cdot\|_1, \mathbf{x})) \leq \delta(\mathcal{D}(\|\Omega \cdot\|_1, \frac{s}{\tilde{s}} \mathbf{x}))$ . The result in theorem 1 follows from [12, Corollary 3.5], Proposition 1, and the fact that  $\delta(\mathcal{D}(\|\cdot\|_1, \Omega \mathbf{x}))$  only depends on the support of  $\Omega \mathbf{x}$ .

### REFERENCES

- [1] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE transactions on information theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [2] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal  $\ell_1$ -norm solution is also the sparsest solution," *Communications on pure and applied mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

- [3] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.
- [4] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.
- [5] M. Kabanava and H. Rauhut, "Analysis  $\ell_1$ -recovery with frames and Gaussian measurements," *Acta Applicandae Mathematicae*, vol. 140, no. 1, pp. 173–195, 2015.
- [6] R. T. Rockafellar, *Convex analysis*. Princeton university press, 2015.
- [7] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Information and Inference: A Journal of the IMA*, vol. 3, no. 3, pp. 224–294, 2014.
- [8] D. Amelunxen, M. Lotz, and J. Walvin, "Effective condition number bounds for convex regularization," *arXiv preprint arXiv:1707.01775*, 2017.
- [9] A. Flineth, "Optimal choice of weights for sparse recovery with prior information," *IEEE Transactions on Information Theory*, vol. 62, no. 7, pp. 4276–4284, 2016.
- [10] S. Daei, F. Haddadi, and A. Amini, "Exploiting prior information in block sparse signals," *arXiv preprint arXiv:1804.08444*, 2018.
- [11] S. R. Becker, E. J. Candès, and M. C. Grant, "Templates for convex cone problems with applications to sparse signal recovery," *Mathematical programming computation*, vol. 3, no. 3, p. 165, 2011.
- [12] J. A. Tropp, "Convex recovery of a structured signal from independent random linear measurements," in *Sampling Theory, a Renaissance*, pp. 67–101, Springer, 2015.