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 ℓ_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 ℓ_1 analysis technique.**

***Index Terms*—ℓ_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 $\boldsymbol{x} \in \mathbb{R}^n$ from $m \ll n$ noisy linear measurements

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon} \in \mathbb{R}^m \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a known matrix and ϵ is an ℓ_2 bounded noise term, i.e. $\|\epsilon\|_2 \leqslant \eta$ for some $\eta \geqslant 0$. In many scenarios, $\boldsymbol{x}$ is sparse after the application of some analysis operator $\boldsymbol{\Omega}$. Specifically, we say $\boldsymbol{x}$ is s-analysis-sparse with support $\mathcal{S}$ in the analysis domain $\boldsymbol{\Omega} \in \mathbb{R}^{p\times n}$ if $\boldsymbol{\Omega x}$ is s-sparse with support $\mathcal{S}$. Then, the following optimization problem called ℓ_1 analysis is often used (See [3], [4], and [5]) to recover $\boldsymbol{x}$:

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

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 $\boldsymbol{x}$. 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 P_0 for successful recovery and to improve the reconstruction error in P_η for robust and stable recovery. For this purpose, a common way is to use weighted ℓ_1 analysis as follows:

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

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 $\{1, ..., p\}$. Thus, the elements of $\mathcal{P}_i$ are all assigned the same weight (ω_i). Moreover, we define

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

where $|\cdot|$ denotes the cardinality of a set and $\mathbf{1}_{\mathcal{E}}$ is the indicator function of the set $\mathcal{E}$. The parameters α_i and ρ_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 ℓ_1 analysis problem in low-redundant analysis operators in numerical simulations.

A. Related Works

The use of prior information in analysis sparse signal models is initiated by the work [3]. With numerical simulations, it is shown in [3] that by including weights in the ℓ_1 minimization, analysis sparse signals can be recovered with a greater success. In [3], the weights are iteratively determined as $\frac{1}{|\boldsymbol{\Omega}\widehat{\boldsymbol{x}}|+\epsilon}$ where $\widehat{\boldsymbol{x}}$ is the solution of the previous iteration. However, the optimal tuning of the weights is not discussed. Recently, closed-form expressions are derived for the optimal weights in [6] and [7] by minimizing the sample complexity

S. Daei and F. Haddadi are with the School of Electrical Engineering, Iran University of Science & Technology. A. Amini is with EE department, Sharif University of Technology.

functions corresponding to weighted ℓ_1 and $\ell_{1,2}$ minimization problems (for the recovery of sparse and block-sparse signals, respectively). Unfortunately, the approaches in [6] and [7] cannot be extended to analysis sparse signals, mainly because finding the sample complexity function in the case of ℓ_1 analysis is a challenging task [8], [9].

B. Outline and Notations

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 ith element of a vector $\boldsymbol{x}$ 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 $\mathcal{C}$. In this work, $(a)_+$ stands for $\max\{a,0\}$ for a scalar a. For a matrix $\boldsymbol{A}$, the operator norm and the condition number are represented by $\|\boldsymbol{A}\|_{p\to q} = \sup_{\|\boldsymbol{x}\|_p\leqslant 1} \|\boldsymbol{A}\boldsymbol{x}\|_q$ and $\kappa(\boldsymbol{A}) := \|\boldsymbol{A}\|_{2\to 2}\|\boldsymbol{A}^\dagger\|_{2\to 2}$, respectively. We denote i.i.d standard Gaussian random vector by $\boldsymbol{g}$. 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 $f:\mathbb{R}^n\to\mathbb{R}\cup\{\pm\infty\}$ at point $\boldsymbol{x}\in\mathbb{R}^n$ is the set of directions from $\boldsymbol{x}$ that do not increase f:

$$\mathcal{D}(f,\boldsymbol{x}) = \bigcup_{t\geqslant 0}\{\boldsymbol{z}\in\mathbb{R}^n : f(\boldsymbol{x}+t\boldsymbol{z})\leqslant f(\boldsymbol{x})\}. \tag{5}$$

The descent cone of a convex function is a convex set. There is a famous duality [10, Ch. 23] between decent cone and subdifferential of a convex function given by:

$$\mathcal{D}^\circ(f,\boldsymbol{x}) = \operatorname{cone}(\partial f(\boldsymbol{x})) := \bigcup_{t\geqslant 0} t.\partial f(\boldsymbol{x}). \tag{6}$$

Definition 1. Statistical Dimension [11]: Let $\mathcal{C}\subseteq\mathbb{R}^n$ be a convex closed cone. Statistical dimension of $\mathcal{C}$ is defined as:

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

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

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 $\boldsymbol{\Omega}$. A lower-bound is also derived for non-singular $\boldsymbol{\Omega}$. The lower-bound is not new and was previously reported in [12, Theorem A], but here we present a simpler approach for the proof.

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

$$\delta(\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_{1,\boldsymbol{w}},\boldsymbol{x}))\leqslant\kappa^2(\boldsymbol{\Omega})\delta(\mathcal{D}(\|\cdot\|_{1,\boldsymbol{w}},\boldsymbol{\Omega}\boldsymbol{x})). \tag{8}$$

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

$$\frac{1}{\kappa^2(\boldsymbol{\Omega})}\delta(\mathcal{D}(\|\cdot\|_{1,\boldsymbol{w}},\boldsymbol{\Omega}\boldsymbol{x}))\leqslant\delta(\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_{1,\boldsymbol{w}},\boldsymbol{x}))\leqslant \kappa^2(\boldsymbol{\Omega})\delta(\mathcal{D}(\|\cdot\|_{1,\boldsymbol{w}},\boldsymbol{\Omega}\boldsymbol{x})) \tag{9}$$

Proof. See Appendix A.

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

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

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

$$\|\widehat{\boldsymbol{x}}-\boldsymbol{x}\|_2\leqslant\frac{2\eta}{\left(\sqrt{m-1}-\kappa(\boldsymbol{\Omega})\sqrt{\delta(\mathcal{D}(\|\cdot\|_{1,\boldsymbol{w}},\boldsymbol{\Omega}\boldsymbol{x}_{ap}))}-t\right)_+}, \tag{11}$$

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

Proof. See Appendix B.

Remark. (Prior works) In the literature, there exist two recovery results for P_η in [3, Theorem 1.4] and [5, Theorem 1]. One could think of generalizing the results regarding P_η to the weighted case of $\mathsf{P}_\eta^{\boldsymbol{w}}$. The main challenge is, however, to incorporate the prior information in the analysis domain. The distinctive advantage of our approach in (11) is that it paves the way for simply exploiting the prior information in the analysis domain. In turn, it allows us to achieve closed-form expressions for optimal weights. Besides the fact that the error bounds in [3, Theorem 1.4] and [5, Theorem 1] for the case of P_η hold only for the exact analysis sparse signals (which are rare in practice), the tightness of the bounds is of question. In contrast, our results in Theorem 1 hold even for approximately analysis sparse signals, and we demonstrate tightness of the error bounds at least for the low-coherence dictionaries. We should highlight that without the tightness result, the weights that minimize the error bound are irrelevant.

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):

$$\boldsymbol{\omega}^* = \underset{\boldsymbol{\nu}\in\mathbb{R}_+^L}{\arg\min}\, \mathbb{E}\mathrm{dist}^2(\boldsymbol{g}, (\boldsymbol{D}\boldsymbol{\nu}) \odot \partial\|\cdot\|_1(\boldsymbol{\Omega}\boldsymbol{x})), \tag{12}$$

where $\boldsymbol{D} := [\mathbf{1}_{\mathcal{P}_1}, ..., \mathbf{1}_{\mathcal{P}_L}] \in \mathbb{R}^{p\times L}$. The latter optimization problem is very similar to the one in weighted ℓ_1 minimization. With the same approach as in [6], one can show that (12) reduces to solving the following equations simultaneously [6, Corollary 11]:

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

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 ℓ_1 analysis phase transition curve for low-redundancy regime (see bottom right image of Figure 1).

In practice, one may encounter some inaccuracies in determining $\boldsymbol{\alpha} \in \mathbb{R}^L$. The study of the sensitivity of weights to the inaccuracies in α were previously considered in [7]. Fortunately, small changes in α 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 bottom right image of 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 $\boldsymbol{x} \in \mathbb{R}^{55}$ in two different analysis operators with $\kappa(\boldsymbol{\Omega}) = 1.1$ and $\kappa(\boldsymbol{\Omega}) = 240$. We consider two strategies

1) Choose five random sets $\{\mathcal{P}_i\}_{i=1}^5$ that partition the analysis domain $\{1, ..., p\}$ with $\alpha_1 = \frac{8}{10}$, $\alpha_2 = \frac{2}{10}$, $\alpha_3 = \frac{3}{10}$, $\alpha_4 = \frac{1}{8}$ and $\alpha_5 = 0$ and correspondingly assign five suitable weights.
2) Consider two sets $\mathcal{P}_1 \cup \mathcal{P}_2$ and $\mathcal{P}_3 \cup \mathcal{P}_4 \cup \mathcal{P}_5$ with $\alpha_1 = \frac{4}{10}$ and $\alpha_2 = \frac{4}{45}$ and choose two suitable weights accordingly.

The suitable weights are obtained via equation (13) by MATLAB function fzero. The top and bottom left images of Figure 1 show the success rate of $\mathsf{P}_0^{\boldsymbol{w}^*}$ averaged over 50 Monte Carlo simulations. It is evident that the weighted ℓ_1 analysis (either with $L = 5$ or $L = 2$ estimates) with suitable weights needs less number of measurements than regular ℓ_1 analysis. Moreover, we see that the strategy with $L = 5$ estimators acts clearly better than just considering $L = 2$ estimators. Reweighted ℓ_1 analysis in [3] with 4 steps are also examined. However, its performance is weaker than our weighting strategy.

In the third experiment, we evaluate the relation between accuracies and required number of measurements. In particular, we consider α_1, α_2 corresponding to two sets $\mathcal{P}_1$ and $\mathcal{P}_2$ that partition the analysis domain $[p = 60]$ of a signal. In the top right image of Figure 1, we depict the required number of measurements versus the accuracy of the first set i.e. α_1. As it turns out, the more our analysis prior information is nonuniform (the difference of α_1 and α_2 increases), the required number of measurements of $\mathsf{P}_0^{\boldsymbol{w}^*}$ becomes less. In fact, the choice of suitable weights $\boldsymbol{w}^*$ becomes more important as $|\alpha_1 - \alpha_2|$ increases. In case of $\alpha_1 = \alpha_2$ (uniform analysis prior information), the corresponding optimal weights w_1^* and w_2^* become equal and the required number of measurements in $\mathsf{P}_0^{\boldsymbol{w}^*}$ reduces to the one in P_0.

In the forth experiment shown in the right block of Figure 2, we compare the speed of convergence of our optimal weighting strategy with existing methods. We can see that P_η has the best performance in terms of computational complexity.

In the last experiment, we investigate a more practical scenario where a Shepp-Logan image (denoted by $\boldsymbol{X}$ 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 2 illustrates the recovery of this image ($n_1 = n_2 = 128$) by solving (2), (3), and reweighted ℓ_1 analysis proposed in [3] when $\boldsymbol{\Omega} \in \mathbb{R}^{114688\times 16384}$ is a redundant wavelet matrix from haar family (the weights in (3) are obtained via 13) and in case of $m = 6554$. The recovery problems (2), (3), and reweighed strategy [3] are all carried out using TFOCS algorithm [13]. The quality of each method is reported in terms of the Peak SNR (PSNR) given by:

$$\mathrm{PSNR}(\boldsymbol{X}, \widehat{\boldsymbol{X}}) := 20\log_{10}\left(\frac{\|\boldsymbol{X}\|_\infty\sqrt{n_1 n_2}}{\|\boldsymbol{X} - \widehat{\boldsymbol{X}}\|_F}\right). \tag{14}$$

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 2, while $\mathsf{P}_\eta^{\boldsymbol{w}^*}$ (bottom right image of Figure 2) has an acceptable performance with PSNR=19 dB, P_η (top right image of Figure 2) clearly fails with a poor performance PSNR=10 dB. Moreover, since the selection mechanism of the weights in reweighted ℓ_1 strategy (bottom left image of Figure 2) is heuristic, its performance (PSNR=13.1 dB) is weaker than our theoretical weighting strategy.

Appendix

A. Proof of Proposition 1

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

$$\begin{aligned}
&\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1, \boldsymbol{x})^{\circ\circ} = \mathrm{closure}(\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1, \boldsymbol{x}))\\
&\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1, \boldsymbol{x}) = \mathrm{cone}^\circ(\boldsymbol{\Omega}^T\partial\|\cdot\|_1(\boldsymbol{\Omega}\boldsymbol{x}))\\
&\{\boldsymbol{w}\in\mathbb{R}^n : \langle\boldsymbol{w}, \boldsymbol{\Omega}^T\boldsymbol{v}\rangle \leqslant 0 \;:\; \forall\boldsymbol{v}\in\mathrm{cone}(\partial\|\cdot\|_1(\boldsymbol{\Omega}\boldsymbol{x}))\} =\\
&\{\boldsymbol{w}\in\mathbb{R}^n :\; \boldsymbol{\Omega}\boldsymbol{w}\in\mathrm{cone}^\circ(\partial\|\cdot\|_1(\boldsymbol{\Omega}\boldsymbol{x}))\} =\\
&\{\boldsymbol{w}\in\mathbb{R}^n :\; \boldsymbol{\Omega}\boldsymbol{w}\in\mathcal{D}(\|\cdot\|_1, \boldsymbol{\Omega}\boldsymbol{x})\}.
\end{aligned} \tag{15}$$

Therefore,

$$\boldsymbol{\Omega}\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1, \boldsymbol{x}) \subset \mathcal{D}(\|\cdot\|_1, \boldsymbol{\Omega}\boldsymbol{x}). \tag{16}$$

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

$$\boldsymbol{\Omega}\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1, \boldsymbol{x}) = \mathcal{D}(\|\cdot\|_1, \boldsymbol{\Omega}\boldsymbol{x}), \tag{17}$$

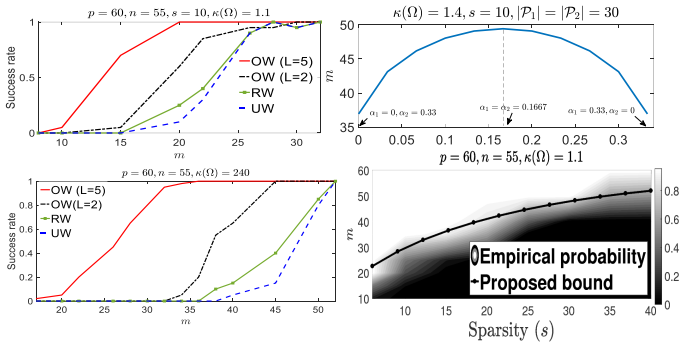

Fig. 1. The top and bottom left images show the success rate of P_0, $\mathsf{P}_0^{\boldsymbol{w}*}$ and reweighted ℓ_1 analysis with parameters $p = 60$, $n = 55$, $s = 10$, $\kappa(\Omega) = 1.1$. OW, RW, and UW denote optimal weighting strategy, reweighted ℓ_1 analysis with 4 steps and unweighted case, respectively. The top right image shows the relation between accuracies and the required number of measurements. The bottom right image shows empirical probability that problem (2) recovers $\boldsymbol{x} \in \mathbb{R}^{55}$ that has s non-zero entries after applying a redundant analysis operator with $\kappa(\boldsymbol{\Omega}) = 1.1$. The black line shows the number of measurements obtained by (8).

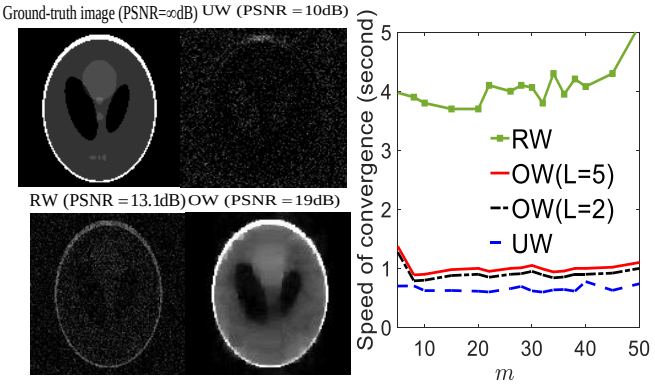

Fig. 2. Left block: Effect of suitable weights in analysis sparse recovery with prior information. The associated parameters are $n_1 = n_2 = 128$, and $m = 6554$. Right block: Comparing our optimal weighting (OW) strategy with reweighted (RW) method (4 steps) and unweighting (UW) strategy.

where in the last line of (15), we used the fact that $\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1, \boldsymbol{x})$ 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 $\delta(\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1, \boldsymbol{x}))$.

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 $\mathbb{E}|X_t - X_s|^2 \leqslant \mathbb{E}|Y_t - Y_s|^2 \;:\; \forall s, t \in T$, then*

$$\mathbb{E}\sup_{t\in T} X_t^2 \leqslant \mathbb{E}\sup_{t\in T} Y_t^2. \tag{18}$$

$$\delta(\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1, \boldsymbol{x})) := \mathbb{E}\Big(\sup_{\substack{\boldsymbol{w}\in\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1,\boldsymbol{x})\\ \|\boldsymbol{w}\|_2\leqslant 1}} \langle \boldsymbol{g}, \boldsymbol{w}\rangle\Big)^2 \leqslant \|\boldsymbol{\Omega}\|_{2\to 2}^2$$

$$\mathbb{E}\Big(\sup_{\substack{\boldsymbol{w}\in\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1,x)\\ \|\boldsymbol{\Omega}\|_{2\to 2}\|\boldsymbol{v}\|_2\leqslant 1}} \langle \boldsymbol{g}, \boldsymbol{v}\rangle\Big)^2 \leqslant \|\boldsymbol{\Omega}\|_{2\to 2}^2\|\boldsymbol{\Omega}^\dagger\|_{2\to 2}^2$$

$$\mathbb{E}\Big(\sup_{\substack{\boldsymbol{w}\in\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1,x)\\ \|\boldsymbol{\Omega}\|_{2\to 2}\|\boldsymbol{v}\|_2\leqslant 1}} \langle \boldsymbol{h}, \boldsymbol{\Omega}\boldsymbol{v}\rangle\Big)^2 \kappa^2(\boldsymbol{\Omega})\mathbb{E}\Big(\sup_{\substack{\boldsymbol{w}\in\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1,\boldsymbol{x})\\ \|\boldsymbol{\Omega}\boldsymbol{v}\|_2\leqslant 1}} \langle \boldsymbol{h}, \boldsymbol{\Omega}\boldsymbol{v}\rangle\Big)^2$$

$$\leqslant \kappa^2(\boldsymbol{\Omega})\mathbb{E}\Big(\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\Omega}\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1,\boldsymbol{x})\\ \|\boldsymbol{z}\|_2\leqslant 1}} \langle \boldsymbol{h}, \boldsymbol{z}\rangle\Big)^2 \leqslant$$

$$\kappa^2(\boldsymbol{\Omega})\mathbb{E}\Big(\sup_{\substack{\boldsymbol{z}\in\mathcal{D}(\|\cdot\|_1,\boldsymbol{\Omega}\boldsymbol{x})\\ \|\boldsymbol{z}\|_2\leqslant 1}} \langle \boldsymbol{h}, \boldsymbol{z}\rangle\Big)^2 = \kappa^2(\boldsymbol{\Omega})\delta(\mathcal{D}(\|\cdot\|_1, \boldsymbol{\Omega}\boldsymbol{x})), \tag{19}$$

where in (19), $\boldsymbol{h} \in \mathbb{R}^p$ is a standard normal vector with i.i.d components. In the first inequality of (19), we used the change of variable $\boldsymbol{v} = \|\boldsymbol{\Omega}\|_{2\to 2}^{-1}\boldsymbol{w}$. The second inequality comes from Theorem 2 with $X_{\boldsymbol{v}} = \langle \boldsymbol{g}, \boldsymbol{v}\rangle$ and $Y_{\boldsymbol{v}} = \|\boldsymbol{\Omega}^\dagger\|_{2\to 2}\langle \boldsymbol{h}, \boldsymbol{\Omega}\boldsymbol{v}\rangle$ and the fact that:

$$\mathbb{E}|X_{\boldsymbol{v}} - X_{\boldsymbol{w}}|^2 = \|\boldsymbol{v} - \boldsymbol{w}\|_2^2 \leqslant \|\boldsymbol{\Omega}^\dagger\|_{2\to 2}^2\|\boldsymbol{\Omega}(\boldsymbol{v} - \boldsymbol{w})\|_2^2 =$$
$$= \mathbb{E}|Y_{\boldsymbol{v}} - Y_{\boldsymbol{w}}|^2 \;:\; \forall \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n. \tag{20}$$

The last inequality comes from (16). In the special case $p = n$ and $\boldsymbol{\Omega}$ is non-singular we have:

$$\delta(\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1, \boldsymbol{x})) := \mathbb{E}\Big(\sup_{\substack{\boldsymbol{w}\in\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1,\boldsymbol{x})\\ \|\boldsymbol{w}\|_2\leqslant 1}} \langle \boldsymbol{g}, \boldsymbol{w}\rangle\Big)^2 =$$

$$\mathbb{E}\Big(\sup_{\substack{\boldsymbol{v}\in\mathcal{D}(\|\cdot\|_1,\boldsymbol{\Omega}\boldsymbol{x})\\ \|\boldsymbol{\Omega}^\dagger\boldsymbol{v}\|_2\leqslant 1}} \langle \boldsymbol{g}, \boldsymbol{\Omega}^\dagger\boldsymbol{v}\rangle\Big)^2 \geqslant$$

$$\|\boldsymbol{\Omega}\|_{2\to 2}^{-2}\mathbb{E}\Big(\sup_{\substack{\boldsymbol{v}\in\mathcal{D}(\|\cdot\|_1,\boldsymbol{\Omega}\boldsymbol{x})\\ \|\boldsymbol{\Omega}^\dagger\boldsymbol{v}\|_2\leqslant 1}} \langle \boldsymbol{h}, \boldsymbol{v}\rangle\Big)^2 \geqslant$$

$$\|\boldsymbol{\Omega}\|_{2\to 2}^{-2}\mathbb{E}\Big(\sup_{\substack{\boldsymbol{v}\in\mathcal{D}(\|\cdot\|_1,\boldsymbol{\Omega}\boldsymbol{x})\\ \|\boldsymbol{v}\|_2\leqslant\|\boldsymbol{\Omega}^\dagger\|_{2\to 2}^{-1}}} \langle \boldsymbol{h}, \boldsymbol{v}\rangle\Big)^2 = \frac{1}{\kappa^2(\boldsymbol{\Omega})}\delta(\mathcal{D}(\|\cdot\|_1, \boldsymbol{\Omega}\boldsymbol{x})), \tag{21}$$

where the first inequality comes from $\boldsymbol{\Omega}^\dagger\boldsymbol{\Omega} = \boldsymbol{I}$ and (17). The second inequality comes from Theorem 2 with $X_v = \langle \boldsymbol{g}, \boldsymbol{\Omega}^\dagger\boldsymbol{v}\rangle$ and $Y_v = \|\boldsymbol{\Omega}\|_{2\to 2}^{-1}\langle \boldsymbol{h}, \boldsymbol{v}\rangle$ and the fact that

$$\mathbb{E}|X_{\boldsymbol{v}} - X_{\boldsymbol{w}}|^2 = \|\boldsymbol{\Omega}^\dagger(\boldsymbol{v} - \boldsymbol{w})\|_2^2 \geqslant \|\boldsymbol{\Omega}\|_{2\to 2}^{-2}\|\boldsymbol{v} - \boldsymbol{w}\|_2^2 =$$
$$= \mathbb{E}|Y_{\boldsymbol{v}} - Y_{\boldsymbol{w}}|^2 \;:\; \forall \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n, \tag{22}$$

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,

$$\|\boldsymbol{\Omega}\boldsymbol{x}_{ap}\|_1 := \|(\boldsymbol{\Omega}\boldsymbol{x})_{T_0}\|_1 \geqslant \frac{\tilde{s}\|\boldsymbol{\Omega}\boldsymbol{x}\|_1}{s}, \tag{23}$$

and as a result, we have, $\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1, \boldsymbol{x}) \subseteq \mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1, \frac{s}{\tilde{s}}\boldsymbol{x})$ and thus $\delta(\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1, \boldsymbol{x})) \leqslant \delta(\mathcal{D}(\|\boldsymbol{\Omega}\cdot\|_1, \frac{s}{\tilde{s}}\boldsymbol{x}))$. The result in Theorem 1 follows from [14, Corollary 3.5], Proposition 1, and the fact that $\delta(\mathcal{D}(\|\cdot\|_1, \boldsymbol{\Omega}\boldsymbol{x}))$ only depends on the support of $\boldsymbol{\Omega}\boldsymbol{x}$.

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 ℓ_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 ℓ_1-recovery with frames and gaussian measurements," *Acta Applicandae Mathematicae*, vol. 140, no. 1, pp. 173–195, 2015.

[6] A. Flinth, "Optimal choice of weights for sparse recovery with prior information," *IEEE Transactions on Information Theory*, vol. 62, no. 7, pp. 4276–4284, 2016.

[7] S. Daei, F. Haddadi, and A. Amini, "Exploiting prior information in block sparse signals," *arXiv preprint arXiv:1804.08444*, 2018.

[8] S. Daei, F. Haddadi, and A. Amini, "Sample complexity of total variation minimization," *IEEE Signal Processing Letters*, vol. 25, pp. 1151–1155, Aug 2018.

[9] S. Daei, F. Haddadi, A. Amini, and M. Lotz, "On the error in phase transition computations for compressed sensing," *arXiv preprint arXiv:1806.10583*, 2018.

[10] R. T. Rockafellar, *Convex analysis*. Princeton university press, 2015.

[11] 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.

[12] D. Amelunxen, M. Lotz, and J. Walvin, "Effective condition number bounds for convex regularization," *arXiv preprint arXiv:1707.01775*, 2017.

[13] 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.

[14] J. A. Tropp, "Convex recovery of a structured signal from independent random linear measurements," in *Sampling Theory, a Renaissance*, pp. 67–101, Springer, 2015.