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 $x \in \mathbb{R}^n$ from $m < n$ noisy linear measurements

$$ y = Ax + \epsilon \in \mathbb{R}^m $$

(1)

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

$$ P_\eta : \min_{z \in \mathbb{R}^n} \|\Omega z\|_1 \text{ s.t. } \|y - Az\|_2 \leq \eta $$

(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 $x$. 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 $P_\eta$ for robust and stable recovery. For this purpose, a common way is to use weighted $\ell_1$ analysis as follows:

$$ P^w_\eta : \min_{z \in \mathbb{R}^n} \|\Omega z\|_1, w := \sum_{i=1}^p w_i \|\Omega z_i\|_1 \text{ s.t. } \|y - Az\|_2 \leq \eta, $$

(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 $\{P_i\}_{i=1}^L$ that partition $\{1, \ldots, p\}$. Thus, the elements of $P_i$ are all assigned the same weight ($\omega_i$). Moreover, we define

$$ w = \sum_{i=1}^L \omega_i 1_{P_i}, \alpha_i = \frac{|P_i \cap S|}{|P_i|}, \rho_i = \frac{|P_i|}{p}, $$

(4)

where $|\cdot|$ denotes the cardinality of a set and $1_{E}$ is the indicator function of the set $E$. The parameters $\alpha_i$ and $\rho_i$ are commonly called the accuracy and the normalized size of the subsets, respectively. Alternatively, $\{P_i\}_{i=1}^L$ 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 cannot 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.

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 $\ell_1$ minimization, analysis sparse signals can be recovered with a greater success. In [3], the weights are iteratively determined as $\frac{1}{\|Hx + \epsilon\|_2}$ where $\hat{x}$ 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.
functions corresponding to weighted $\ell_1$ 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 $\ell_1$ 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 $i$th element of a vector $\mathbf{x}$ is shown either by $x(i)$ or $x_i$. $(\cdot)^\top$ denotes the pseudo inverse operator. We reserve the calligraphic uppercase letters $\mathcal{C}$ for sets (e.g. $\mathcal{S}$). The cardinality of a set $\mathcal{S}$ is denoted by $|\mathcal{S}|$. $C^\circ$ represents the polar of a cone $C$. In this work, $(a)_+$ stands for $\max\{a, 0\}$ for a scalar $a$. For a matrix $\mathbf{A}$, the operator norm and the condition number := $\|\mathbf{A}\|_{2\rightarrow 2} := \|\mathbf{A}\|_{2\rightarrow 2}$, respectively. We denote i.i.d standard Gaussian random vector by $\mathbf{g}$. Lastly, $\| \cdot \|$ 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 $f$:

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

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:

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

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

$$\delta(\mathcal{C}) := E[\|P_{\mathcal{C}}(\mathbf{g})\|^2] = \text{Edist}^2(\mathbf{g}, \mathcal{C}^\circ), \tag{7}$$

where $\mathbf{g}$ has i.i.d. standard normal distribution, and $P_{\mathcal{C}}(\mathbf{x})$ is the orthogonal projection of $\mathbf{x} \in \mathbb{R}^n$ onto the set $\mathcal{C}$ defined as:

$$P_{\mathcal{C}}(\mathbf{x}) = \arg\min_{\mathbf{z} \in \mathcal{C}} \|\mathbf{z} - \mathbf{x}\|_2.$$  

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 $\Omega$. A lower-bound is also derived for non-singular $\Omega$. 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 $x \in \mathbb{R}^n$ be a $s$-analysis sparse vector with redundant analysis operator $\Omega \in \mathbb{R}^{p \times n}(p \geq n)$. Then,

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

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

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

Proof. See Appendix A.

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

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

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

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

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 $P_{\Omega}$ in [3, Theorem 1.4] and [5, Theorem 1. One could think of generalizing the results regarding $P_{\Omega}$ to the weighted case of $P^w_{\Omega}$. 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 $P_{\Omega}$ 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]):

$$\omega^* = \arg \min_{\omega \in \mathbb{R}^2} \text{Edist}^2(g_i(Du) \odot \hat{c} \cdot \|_{1}(\Omega x)), \quad (12)$$

where $D := [1_{p_1}, \ldots, 1_{p_L}] \in \mathbb{R}^{p \times L}$. The latter optimization problem is very similar to the one in weighted $\ell_1$ minimization. With the same approach as in [6], one can show that (12) reduces to solving the following equations simultaneously [6, Corollary 11]:

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

It is not obvious whether the inequality [6] 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 bottom right image of 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 [7]. 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 (12) on the number of required measurements. First, we consider the scaling of the required number of measurements for successful recovery of $\mathbb{C}$ 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 $x \in \mathbb{R}^{55}$ in two different analysis operators with $\kappa = 1.1$ and $\kappa(\Omega) = 240$. We consider two strategies

1) Choose five random sets $\{P_1\}_{i=1}^{5}$ that partition the analysis domain $\{1, \ldots, p\}$ with $\alpha_1 = \frac{8}{10}$, $\alpha_2 = \frac{2}{10}$, $\alpha_3 = \frac{3}{8}$, $\alpha_4 = \frac{3}{8}$, and $\alpha_5 = 0$ and correspondingly assign five suitable weights.

2) Consider two sets $P_1 \cup P_2$ and $P_3 \cup P_4 \cup P_5$ with $\alpha_1 = \frac{1}{10}$ and $\alpha_2 = \frac{4}{25}$ 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 $P_{0^w}^*$ averaged over 50 Monte Carlo simulations. It is evident that the weighted $\ell_1$ analysis (either with $L = 5$ or $L = 2$ estimates) with suitable weights needs less number of measurements than regular $\ell_1$ analysis. Moreover, we see that the strategy with $L = 5$ estimators acts clearly better than just considering $L = 2$ estimators. Reweighted $\ell_1$ 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 $\alpha_1$, $\alpha_2$ corresponding to two sets $P_1$ and $P_2$ 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. $\alpha_1$. As it turns out, the more our analysis prior information is nonuniform (the difference of $\alpha_1$ and $\alpha_2$ increases), the required number of measurements of $P_{0^w}^*$ becomes less. In fact, the choice of suitable weights $w^*$ becomes more important as $|\alpha_1 - \alpha_2|$ increases. In case of $\alpha_1 = \alpha_2$ (uniform analysis prior information), the corresponding optimal weights $w^*$ become equal and the required number of measurements in $P_{0^w}^*$ 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 $P_0$ 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 $X$ 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 $\text{SNR} = 10 \, \text{dB}$. Figure 2 illustrates the recovery of this image ($n_1 = n_2 = 128$) by solving (2), (3), and reweighed $\ell_1$ analysis proposed in [3] when $\Omega \in \mathbb{R}^{11488 \times 16384}$ is a redundant wavelet matrix from haar family (the weights in $\mathbb{C}$ 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:

$$\text{PSNR}(X, \hat{X}) := 20 \log_{10} \left( \frac{\|X\|_2 \sqrt{n_1 n_2}}{\|X - \hat{X}\|_2} \right). \quad (14)$$

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 $\hat{s}$ in Theorem 1) of wavelet coefficients. As shown by Figure 2 while $P_{0^w}$ (bottom right image of Figure 2) has an acceptable performance with PSNR$=19 \, \text{dB}$, $P_\eta$ (top right image of Figure 2) clearly fails with a poor performance PSNR$=10 \, \text{dB}$. Moreover, since the selection mechanism of the weights in reweighted $\ell_1$ strategy (bottom left image of Figure 2) is heuristic, its performance (PSNR$=13.1 \, \text{dB}$) is weaker than our theoretical weighting strategy.

APPENDIX

A. Proof of Proposition 4

Proof. In the following, we relate $D(\{\Omega \cdot \|_{1}, x\})$ to $D(\| \cdot \|_{1}, \Omega x)$.

$$D(\{\Omega \cdot \|_{1}, x\}) = \text{closure}(D(\| \cdot \|_{1}, x))$$

$$D(\| \cdot \|_{1}, x) = \text{cone}(\| \Omega^T \hat{c} \cdot \|_{1}(\Omega x))$$

$$\{w \in \mathbb{R}^n : \langle w, \Omega^T \hat{c} \rangle \leq 0 \quad \forall w \in \text{cone}(\| \hat{c} \cdot \|_{1}(\Omega x))\} =$$

$$\{w \in \mathbb{R}^n : \Omega w \in \text{cone}(\| \hat{c} \cdot \|_{1}(\Omega x))\} =$$

$$\{w \in \mathbb{R}^n : \Omega w \in D(\| \cdot \|_{1}, \Omega x)\}. \quad (15)$$

Therefore,

$$\Omega D(\{\Omega \cdot \|_{1}, x\}) \subset D(\| \cdot \|_{1}, \Omega x). \quad (16)$$

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

$$\Omega D(\| \cdot \|_{1}, x) = D(\| \cdot \|_{1}, \Omega x), \quad (17)$$
where in the last line of (15), we used the fact that \( \ell \) with reweighted (RW) method (4 steps) and unweighting (UW) strategy. Theorem 2 follows from \([14, Corollary 3.5]\), Proposition 1, where the last inequality is a result of norm properties.

### Proof of Theorem 1

Let \( D(\| \Omega \cdot \|_1, x) \) be a closed convex set. In the following, we state Sudakov-Fernique inequality which helps to control the supremum of a random process by that a simpler random process and is used to find an upper-bound for \( \delta(D(\| \Omega \cdot \|_1, x)) \).

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

\[
E \sup_{t \in T} X_t^2 \leq E \sup_{t \in T} Y_t^2. \tag{18}
\]

Hence, \( \delta(D(\| \Omega \cdot \|_1, x)) := E \left( \sup_{w \in D(\| \Omega \cdot \|_1, x)} \langle g, w \rangle \right)^2 \leq \| \Omega \|_{2 \rightarrow 2}^2 \)

\[
\kappa^2(\Omega) E \left( \sup_{w \in D(\| \Omega \cdot \|_1, x)} \langle g, w \rangle \right)^2 \leq \kappa^2(\Omega) E \left( \sup_{z \in D(\| \Omega \cdot \|_1, x)} \langle h, z \rangle \right)^2 \leq \kappa^2(\Omega) \delta(D(\| \cdot \|_1, \Omega x)), \tag{19}
\]

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

\[
E|X_v - X_w|^2 = \| v - w \|^2_2 \leq \| \Omega \|^2_{2 \rightarrow 2} \| v - w \|^2_2 = E|Y_v - Y_w|^2 : \forall v, w \in \mathbb{R}^n. \tag{20}
\]

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

\[
\delta(D(\| \Omega \cdot \|_1, x)) := E \left( \sup_{w \in D(\| \Omega \cdot \|_1, x)} \langle g, w \rangle \right)^2 = E \left( \sup_{v \in D(\| \Omega \cdot \|_1, \Omega x)} \langle g, \Omega^2 v \rangle \right)^2 \geq \| \Omega \|_{2 \rightarrow 2}^{-2} E \left( \sup_{v \in D(\| \Omega \cdot \|_1, \Omega x)} \langle \Omega^2 v, v \rangle \right)^2 = \frac{1}{\kappa^2(\Omega)} \delta(D(\| \cdot \|_1, \Omega x)), \tag{21}
\]

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

\[
E|X_v - X_w|^2 = \| \Omega^2 v - w \|^2_2 \geq \| \Omega \|^2_{2 \rightarrow 2} \| v - w \|^2_2 = E|Y_v - Y_w|^2 : \forall v, w \in \mathbb{R}^n, \tag{22}
\]

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

### B. Proof of Theorem 4

Let \( T_0 \) be the index set of \( s \) largest analysis coefficients. Then, it holds that,

\[
\| \Omega x_{ap} \|_1 := \| (\Omega x)_{T_0} \|_1 \geq \frac{s \| \Omega x \|_1}{s}, \tag{23}
\]

and as a result, we have, \( D(\| \Omega \cdot \|_1, x) \subseteq D(\| \Omega \cdot \|_1, \frac{2}{s} x) \) and thus \( \delta(D(\| \Omega \cdot \|_1, x)) \leq \delta(D(\| \Omega \cdot \|_1, \frac{2}{s} x)) \). The result in Theorem 4 follows from [14, Corollary 3.5], Proposition 1 and the fact that \( \delta(D(\| \cdot \|_1, \Omega x)) \) only depends on the support of \( \Omega x \).
REFERENCES


[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.


