


RESEARCH

Open Access



Exact recovery of sparse signals with side information

Xiaohu Luo^{1,3} , Nianci Feng², Xuhui Guo⁴ and Zili Zhang^{1,5*}

*Correspondence:
zhangzl@swu.edu.cn

¹ Faculty of Computer and Information Science, Southwest University, Chongqing 400715, China

² School of Mathematics and Statistics, Southwest University, Chongqing 400715, China

³ School of Information Engineering, Ningxia University, Yinchuan 750021, China

⁴ School of Intelligent Engineering and Technology, Ningxia University (Zhongwei), Zhongwei 755000, China

⁵ School of Information Technology, Deakin University, Melbourne 3217, Australia

Abstract

Compressed sensing has captured considerable attention of researchers in the past decades. In this paper, with the aid of the powerful null space property, some deterministic recovery conditions are established for the previous ℓ_1 - ℓ_1 method and the ℓ_1 - ℓ_2 method to guarantee the exact sparse recovery when the side information of the desired signal is available. These obtained results provide a useful and necessary complement to the previous investigation of the ℓ_1 - ℓ_1 and ℓ_1 - ℓ_2 methods that are based on the statistical analysis. Moreover, one of our theoretical findings also shows that the sharp conditions previously established for the classical ℓ_1 method remain suitable for the ℓ_1 - ℓ_1 method to guarantee the exact sparse recovery. Numerical experiments on both the synthetic signals and the real-world images are also carried out to further test the recovery performance of the above two methods.

Keywords: Compressed sensing, Side information, ℓ_1 - ℓ_1 method, ℓ_1 - ℓ_2 method, Null space property

1 Introduction

Over the past decades, the problem of sparse signal recovery, now termed as the compressed sensing (CS) [1–3], has been greatly developed and widely used in many fields such as the pattern recognition [4], image processing [5–7], medical image [8] and also the camera design [9], to name a few. Simply speaking, we say a signal $\hat{\mathbf{x}} \in \mathbb{R}^n$ is sparse if and only if it has fewer nonzero components than its length, and if there are at most $k (\ll n)$ nonzero entries in $\hat{\mathbf{x}}$, $\hat{\mathbf{x}}$ is said to be a k -sparse signal. In fact, one of the key goals of CS is the recovery of such a k -sparse signal $\hat{\mathbf{x}}$ from its as few as observations $\mathbf{b} = A\hat{\mathbf{x}}$, where $A \in \mathbb{R}^{m \times n} (m < n)$ is a pre-designed measurement matrix and $\mathbf{b} \in \mathbb{R}^m$ is the resultant observed signal. To realize this goal, it is often suggested to solve the following ℓ_1 method [10, 11]

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i| \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b}. \quad (1.1)$$

Since model (1.1) is convex, it can be efficiently solved by lots of convex optimization algorithms [12, 13], and many recovery guarantees, including the recovery conditions as

well as their resultant recovery error estimates, have also been obtained for (1.1) before, see, e.g., [14–16].

Unfortunately, the ℓ_1 method do not incorporate any side information of $\hat{\mathbf{x}}$ due to the equal treatment of the ℓ_1 norm for the components of the variable \mathbf{x} . Considering that such side information is often available in many real-world applications, it is naturally expected that the performance of model (1.1) can be further improved if the side information is well integrated. In general, there are two types of common side information in CS filed. The first one takes the form of a known support estimate. To deal with this type of side information, the authors in [17] first modeled the known support as a set T , and then integrated it into the ℓ_1 norm, leading to the model

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}_{K^c}\|_1 \triangleq \sum_{i \in K^c} |x_i| \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b}, \tag{1.2}$$

where K^c models the complement set of K in $\{1, 2, 3, \dots, n\}$. Their work also showed that the resultant recovery conditions are weaker than those without side information. In [18], the authors considered a more general weight rather than a constant weight in the known support estimate. In [19], a variant iterative hard thresholding (IHT) algorithm was proposed by incorporating the partially known support information, and some theoretical analysis was also established for this algorithm. In [20], the orthogonal matching pursuit (OMP), as an iterative greedy algorithm, was extended by using the partially known support. The authors of [21] also considered embedding the known support information into the iterative reweighted least squares (IRLS) algorithm at each iteration, leading a reduction of the number of the measurements as well as the computational cost. Recently, some new recovery conditions were obtained by Ge et al. in [22].

Another type of the side information takes the form of a referenced similar to the original signal $\hat{\mathbf{x}}$. The side information of this type usually comes from applications such as the magnetic resonance imaging (MRI) [23], video processing [24, 25] and estimate problems [26]. For example, when faced with some video processing problems, we usually know some previous video frames before we cope with the next video frames. These video frames known in advance, to some degree, can be viewed as the side information of the next video frames. By introducing two ℓ_1 norm and ℓ_2 norm approximation terms to model (1.1), respectively, Mota et al. [27] proposed to solve an ℓ_1 - ℓ_1 method

$$\mathbf{x}^\sharp = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_1 + \beta \|\mathbf{x} - \mathbf{w}\|_1 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b} \tag{1.3}$$

and the ℓ_1 - ℓ_2 method

$$\mathbf{x}^\diamond = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_1 + \frac{\beta}{2} \|\mathbf{x} - \mathbf{w}\|_2^2 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b}, \tag{1.4}$$

where β is a positive parameter and $\mathbf{w} \in \mathbb{R}^n$ is the referenced signal that models the side information. For simple, \mathbf{w} is assumed to obey $\text{Supp}(\mathbf{w}) \subset \text{Supp}(\hat{\mathbf{x}})$, where $\text{Supp}(\mathbf{w}) = \{i : |\mathbf{w}_i| \neq 0, i = 1, 2, \dots, n\}$. Based on some statistical tools, the authors affirmatively answer how many measurements one required to ensure the exact recovery of any k -sparse signal $\hat{\mathbf{x}}$. Some convincing experiments are also conducted to support their claims. Note that there also exist some works which embeds the prior information

of the desired signals into some other models. For example, Zhang et al. [28] recently proposed to use an ℓ_{1-2} model to deal with the signal recovery with prior information. We refer the interested readers to [28–30] and the references within for more details.

In this paper, we revisit the above $\ell_{1-\ell_1}$ and $\ell_{1-\ell_2}$ methods for exact sparse recovery with side information. Different from the pioneering work of [27], this paper aims at investigating both the $\ell_{1-\ell_1}$ and $\ell_{1-\ell_2}$ methods in a deterministic way. To do so, by means of the powerful null space property (NSP), we established two kind of deterministic sufficient and necessary condition for these two methods. Our obtained theoretical results not only well complement the previous work [27] that was based on the statistical analysis, but also surprisingly find that the sharp exact recovery conditions of model (1.1) are still suitable for the $\ell_{1-\ell_1}$ model (1.3). Moreover, the resultant numerical experiments show that the recovery performance of the $\ell_{1-\ell_1}$ method is superior to other methods in terms of the number of the measurements required by incorporating the side information.

The rest of this paper is organized as follows the main theoretical results are presented in Sect. 2, and the resultant numerical experiments are provided in Sect. 3. Finally, we conclude this paper in Sect. 4.

2 The deterministic analysis of $\ell_{1-\ell_1}$ and $\ell_{1-\ell_2}$ methods

Our main results will be presented in this section, which include the exact recovery guarantees of $\ell_{1-\ell_1}$ and $\ell_{1-\ell_2}$ methods. Before moving on, we first introduce the following two key definitions.

Definition 2.1 (NSP, see, e.g., [31]) For any subsets $K \subset \{1, 2, \dots, n\}$ with $|K| \leq k$ and any $\mathbf{h} \in \text{Ker}(A) \setminus \{0\} \triangleq \{\mathbf{h} : A\mathbf{h} = 0, \mathbf{h} \neq 0\}$, we say $A \in \mathbb{R}^{m \times n}$ satisfies the k -order NSP if it holds that

$$\|\mathbf{h}_K\|_1 < \|\mathbf{h}_{K^c}\|_1. \quad (2.5)$$

Furthermore, if it holds that

$$\|\mathbf{h}_K\|_1 \leq \alpha \|\mathbf{h}_{K^c}\|_1 \quad (2.6)$$

for certain $0 < \alpha < 1$, then we say A satisfies the k -order stable NSP with constant α .

Definition 2.2 (Restricted isometry property, see, e.g., [3]) A matrix A is said to satisfy the k -order restricted isometry property (RIP) if there exists $0 < \delta < 1$ such that

$$(1 - \delta)\|\mathbf{h}_K\|_2^2 \leq \|A\mathbf{h}_K\|_2^2 \leq (1 + \delta)\|\mathbf{h}_K\|_2^2, \quad (2.7)$$

holds for all k -sparse signals $\mathbf{h} \in \mathbb{R}^n$ and subsets $K \subset \{1, 2, \dots, n\}$ with $|K| \leq k$. Moreover, the smallest δ obeying (2.7) is denoted by δ_k , i.e., the known k -order restricted isometry constant (RIC).

Theorem 2.3 *The $\ell_{1-\ell_1}$ model (1.3) with $\text{Supp}(w) \subset \text{Supp}(\hat{x})$ has a unique k -sparse solution if and only if A obeys the k -order NSP.*

1 Remark 2.4

The k -order NSP has been demonstrated to be a necessary and sufficient condition for the classical ℓ_1 method to ensure the exactly k -sparse signal recovery. However, according to our Theorem 2.3, this condition also applies to the ℓ_1 - ℓ_1 model (1.3). On the other hand, it has also been shown in [32, 33] that if A obeys the k -order stable NSP with constant α , then α can be expressed by tk -order RIC δ_{tk} with $t > 1$ as follows:

$$\alpha = \frac{\delta_{tk}}{\sqrt{(1 - (\delta_{tk})^2)(t - 1)}}.$$

If one further restricts $\alpha < 1$, then we will get

$$\delta_{tk} < \sqrt{\frac{t - 1}{t}}. \tag{2.8}$$

Note that condition (2.8) has been proved to sharp for the classical ℓ_1 to exactly recover any k -sparse signal. Again, condition (2.8) is also suitable to the ℓ_1 - ℓ_1 model (1.3). As far as we know, the RIC-based sufficient conditions have not been established for the ℓ_1 - ℓ_1 method before.

1 Proof of Theorem 2.3

First, we prove the sufficiency. Pick any feasible k -sparse vectors $\hat{\mathbf{x}}$. Let $K = \text{Supp}(\hat{\mathbf{x}})$. Since $\mathbf{h} \in \text{Ker}(A)$, it holds that $A(\hat{\mathbf{x}} + \mathbf{h}) = A\hat{\mathbf{x}} = \mathbf{b}$. And

$$\begin{aligned} \|\hat{\mathbf{x}} + \mathbf{h}\|_1 + \beta \|\hat{\mathbf{x}} - \mathbf{w} + \mathbf{h}\|_1 &= \|\hat{\mathbf{x}}_K + \mathbf{h}_K\|_1 + \beta \|\hat{\mathbf{x}}_K + \mathbf{h}_K - \mathbf{w}\|_1 + \|\mathbf{h}_{K^c}\|_1 + \beta \|\mathbf{h}_{K^c}\|_1 \\ &\geq \|\hat{\mathbf{x}}_K\|_1 - \|\mathbf{h}_K\|_1 + \beta \|\hat{\mathbf{x}}_K - \mathbf{w}\|_1 - \beta \|\mathbf{h}_K\|_1 + \|\mathbf{h}_{K^c}\|_1 + \beta \|\mathbf{h}_{K^c}\|_1 \\ &= \|\hat{\mathbf{x}}_K\|_1 + \beta \|\hat{\mathbf{x}}_K - \mathbf{w}\|_1 + (1 + \beta)(\|\mathbf{h}_{K^c}\|_1 - \|\mathbf{h}_K\|_1), \end{aligned} \tag{2.9}$$

where we have used the triangle inequality in the first inequality.

Recall that $\|\mathbf{h}_K\|_1 < \|\mathbf{h}_{K^c}\|_1$, $\beta > 0$, and we have assumed that A obeys the k -order NSP, we get $\|\hat{\mathbf{x}} + \mathbf{h}\|_1 + \beta \|\hat{\mathbf{x}} + \mathbf{h} - \mathbf{w}\|_1 > \|\hat{\mathbf{x}}\|_1 + \beta \|\hat{\mathbf{x}} - \mathbf{w}\|_1$. Hence, the sufficiency is proved.

Now, we prove the necessity. To do so, we first assume that the i th component of $\hat{\mathbf{x}}$ obeys $\hat{\mathbf{x}}_i = -\text{sign}(\|\mathbf{h}\|_\infty)$ for all $i \in K$. Then, we can obtain the following properties $\|\hat{\mathbf{x}}\|_0 \leq k$, $\|\hat{\mathbf{x}}_K + \tau \mathbf{h}_K\|_1 = \|\hat{\mathbf{x}}_K\|_1 - \|\tau \mathbf{h}_K\|_1$ holds for all $0 < \tau \leq 1$. Now, By replacing \mathbf{h} in (2.9) with $\tau \mathbf{h}$, and noting that $\|\hat{\mathbf{x}} + \tau \mathbf{h}\|_1 + \beta \|\hat{\mathbf{x}} + \tau \mathbf{h} - \mathbf{w}\|_1 > \|\hat{\mathbf{x}}\|_1 + \beta \|\hat{\mathbf{x}} - \mathbf{w}\|_1$ since $\hat{\mathbf{x}}$ is the exact solution, we can easily deduce that

$$(1 + \beta)(\|\tau \mathbf{h}_{K^c}\|_1 - \|\tau \mathbf{h}_K\|_1) > 0$$

for any $0 < \tau \leq 1$ and $\beta > 0$, which requires (2.5) to hold. \square

In what follows, we establish the stable NSP condition of order s for the ℓ_1 - ℓ_2 model (1.4).

Theorem 2.5 Assume that \mathbf{w} is a side information with $\text{Supp}(\mathbf{w}) \subset \text{Supp}(\hat{\mathbf{x}})$. The ℓ_1 - ℓ_2 model (1.4) has a unique k -sparse solution $\hat{\mathbf{x}} (\neq 0)$ if and only if A obeys k -order stable NSP with α being

$$\alpha = \frac{1}{1 + \beta(\|\hat{\mathbf{x}}\|_\infty + \|\mathbf{w}\|_\infty)} \tag{2.10}$$

1 Remark 2.6

Compared with the previous NSP condition for the ℓ_1 - ℓ_1 model (1.3), the obtained stable NSP for ℓ_1 - ℓ_2 model (1.4) performs a bit loose. Besides, it is also affected by the infinite norm of the desired k -sparse signal $\hat{\mathbf{x}}$ and the referenced signal \mathbf{w} . From this point of view, the ℓ_1 - ℓ_2 model is less effective than the ℓ_1 - ℓ_1 model in theoretical aspect.

1 Proof of Theorem 2.5

Our proof is partially inspired by [34]. We start with proving the sufficiency. Pick any feasible k -sparse vectors $\hat{\mathbf{x}}$. Let $K = \text{Supp}(\hat{\mathbf{x}})$. Since $\mathbf{h} \in \text{Ker}(A)$, we first have $A(\hat{\mathbf{x}} + \mathbf{h}) = A\hat{\mathbf{x}} = \mathbf{b}$, and

$$\begin{aligned} \|\hat{\mathbf{x}} + \mathbf{h}\|_1 + \frac{\beta}{2} \|\hat{\mathbf{x}} + \mathbf{h} - \mathbf{w}\|_2^2 &= \|\hat{\mathbf{x}}_K + \mathbf{h}_K\|_1 + \frac{\beta}{2} \|\hat{\mathbf{x}}_K + \mathbf{h}_K - \mathbf{w}\|_2^2 + \|\mathbf{h}_{K^c}\|_1 + \frac{\beta}{2} \|\mathbf{h}_{K^c}\|_2^2 \\ &\geq \|\hat{\mathbf{x}}_K\|_1 - \|\mathbf{h}_K\|_1 + \frac{\beta}{2} \|\hat{\mathbf{x}}_K - \mathbf{w}\|_2^2 + \frac{\beta}{2} \|\mathbf{h}_K\|_2^2 + \beta \langle \hat{\mathbf{x}}_K - \mathbf{w}, \mathbf{h}_K \rangle + \|\mathbf{h}_{K^c}\|_1 + \frac{\beta}{2} \|\mathbf{h}_{K^c}\|_2^2 \\ &\geq \|\hat{\mathbf{x}}_K\|_1 + \frac{\beta}{2} \|\hat{\mathbf{x}}_K - \mathbf{w}\|_2^2 + \|\mathbf{h}_{K^c}\|_1 - \|\mathbf{h}_K\|_1 + \beta \langle \hat{\mathbf{x}}_K - \mathbf{w}, \mathbf{h}_K \rangle + \frac{\beta}{2} \|\mathbf{h}\|_2^2 \\ &\geq \|\hat{\mathbf{x}}_K\|_1 + \frac{\beta}{2} \|\hat{\mathbf{x}}_K - \mathbf{w}\|_2^2 + \|\mathbf{h}_{K^c}\|_1 - \|\mathbf{h}_K\|_1 - \beta(\|\hat{\mathbf{x}}\|_\infty + \|\mathbf{w}\|_\infty) \|\mathbf{h}_K\|_1 + \frac{\beta}{2} \|\mathbf{h}\|_2^2 \end{aligned} \tag{2.11}$$

where we have used $\|\mathbf{h}_K\|_2^2 + \|\mathbf{h}_{K^c}\|_2^2 = \|\mathbf{h}\|_2^2$ and $\langle \hat{\mathbf{x}}_K - \mathbf{w}, \mathbf{h}_K \rangle \geq -(\|\hat{\mathbf{x}}\|_\infty + \|\mathbf{w}\|_\infty) \|\mathbf{h}_K\|_1$ in the second inequality. Since $\|\mathbf{h}\|_2^2 > 0$ and $[1 + \beta(\|\hat{\mathbf{x}}\|_\infty + \|\mathbf{w}\|_\infty)] \|\mathbf{h}_K\|_1 < \|\mathbf{h}_{K^c}\|_1$, we get $\|\hat{\mathbf{x}} + \mathbf{h}\|_1 + \frac{\beta}{2} \|\hat{\mathbf{x}} + \mathbf{h} - \mathbf{w}\|_2^2 > \|\hat{\mathbf{x}}\|_1 + \frac{\beta}{2} \|\hat{\mathbf{x}} - \mathbf{w}\|_2^2$. Hence, we prove that $\hat{\mathbf{x}}$ is the unique minimizer of (1.4).

As for the necessity, it is sufficient to show that for any given nonzero $\mathbf{h} \in \text{Ker}(A)$ and K with $|K| \leq k$ the stable NSP of order k with α given by (2.10) holds. Similarly with Theorem 2.3, we can obtain $\|\hat{\mathbf{x}}\|_0 \leq k$, $\|\hat{\mathbf{x}}_K + \tau \mathbf{h}_K\|_1 = \|\hat{\mathbf{x}}_K\|_1 - \|\tau \mathbf{h}_K\|_1$. Furthermore, assume that the scales $\hat{\mathbf{x}}$ and \mathbf{w} have fixed values of $\|\hat{\mathbf{x}}\|_\infty$ and $\|\mathbf{w}\|_\infty$, respectively, then we get $\langle \hat{\mathbf{x}}_K - \mathbf{w}, \tau \mathbf{h}_K \rangle = -(\|\hat{\mathbf{x}}\|_\infty + \|\mathbf{w}\|_\infty) \|\mathbf{h}_K\|_1$ for any $0 < \tau \leq 1$. Now, we replace $\tau \mathbf{h}$ with \mathbf{h} and observe that both of inequalities of (2.11) now hold with equality. Since $\hat{\mathbf{x}}$ is the exact recovery, it requires $\|\hat{\mathbf{x}} + \tau \mathbf{h}\|_1 + \frac{\beta}{2} \|\hat{\mathbf{x}} + \tau \mathbf{h} - \mathbf{w}\|_2^2 > \|\hat{\mathbf{x}}\|_1 + \frac{\beta}{2} \|\hat{\mathbf{x}} - \mathbf{w}\|_2^2$, so we get

$$\|\tau \mathbf{h}_{K^c}\|_1 - \|\tau \mathbf{h}_K\|_1 - \beta(\|\hat{\mathbf{x}}\|_\infty + \|\mathbf{w}\|_\infty) \|\tau \mathbf{h}_K\|_1 + \frac{\beta}{2} \|\tau \mathbf{h}\|_2^2 > 0$$

for any $0 < 1 \leq \tau$, which proves the necessity. \square

3 Numerical simulations

As can be seen in the previous sections, our goal in this paper is to provide some deterministic recovery conditions for the ℓ_1 - ℓ_1 model and the ℓ_1 - ℓ_2 model to guarantee the exact sparse recovery, and the obtained theoretical results for these models can be found in Theorem 2.3 and Theorem 2.5, respectively. On the other hand, it is still a difficult problem to find the desired measurement matrices according to the conditions stated in two obtained theorems. Nevertheless, we still hope to provide some numerical simulations to testify the efficiency of two models with the side information. Since both the ℓ_1 - ℓ_1 model, ℓ_1 - ℓ_2 model and the ℓ_1 model are convex, in this paper, we resort to the popular and easy-implemented CVX¹ (with an SeDuMi solver) to solve them.

3.1 Experiments on the synthetic signals

We start with the experiments on synthetic signals. For simplicity, in all experiments, we assume that the length of the desired signal $\hat{\mathbf{x}}$ is $n = 256$, and its sparsity is set to be k . We generate such a k -sparse signal $\hat{\mathbf{x}}$ as follows. The location of the nonzero components in $\hat{\mathbf{x}}$ is generated at random and their corresponding values are chosen from a standard normal distribution. We assume that the referenced signal \mathbf{w} is k_w -sparse with $k_w \leq k$, and its nonzero entries are chosen at random from nonzero entries in $\hat{\mathbf{x}}$ with $\text{Supp}(\mathbf{w}) \subseteq \text{Supp}(\hat{\mathbf{x}})$. Obviously, when $k_w < k$, \mathbf{w} contains part of the side information of $\hat{\mathbf{x}}$, and when $k_w = k$, \mathbf{w} contains all the information of $\hat{\mathbf{x}}$, i.e., $\mathbf{w} = \hat{\mathbf{x}}$. Besides, we generate the measurement matrix $A \in \mathbb{R}^{m \times n}$ by drawing it from a standard Gaussian distribution. To judge the recovery performance of the completing methods, we adopt the signal-to-noise (SNR), see, e.g., [35], given by

$$\text{SNR}(\hat{\mathbf{x}}, \tilde{\mathbf{x}}) = 20 \log_{10} \frac{\|\hat{\mathbf{x}}\|_2}{\|\hat{\mathbf{x}} - \tilde{\mathbf{x}}\|_2},$$

where $\hat{\mathbf{x}}$ and $\tilde{\mathbf{x}}$ are denoted by the original signal and the recovered signal by certain model. If there is no specific description, the average SNR results over independent 50 trails are used as the final results.

We first conduct a simple experiment for the ℓ_1 - ℓ_1 model and the ℓ_1 - ℓ_2 model to test their recovery performance on the signal with side information. In this sort of experiments, we set $m = 70$, $k = 20$, $k_w = 10$, and $\beta = 10^4$ for the ℓ_1 - ℓ_1 model and $\beta = 10^{-4}$ for the ℓ_1 - ℓ_2 model.

Figure 1 plots the resultant recovery performance. One can easily observe that both the ℓ_1 - ℓ_1 model and the ℓ_1 - ℓ_2 model well finish the recovery task with the two recovered signals being almost same with the original signal. In the above experiments, we only set $\beta = 10^4$ for the ℓ_1 - ℓ_1 model and $\beta = 10^{-4}$ for the ℓ_1 - ℓ_2 model for simplicity. Obviously, if these parameters can be further optimized, the resultant SNR performance will be correspondingly improved. To select the proper β for the models (1.3) and (1.4). We let the parameter β be chosen from $\{10^{-8}, 10^{-7}, \dots, 10^8\}$, and set the other parameters be same as before.

¹ <http://cvxr.com/cvx/>.

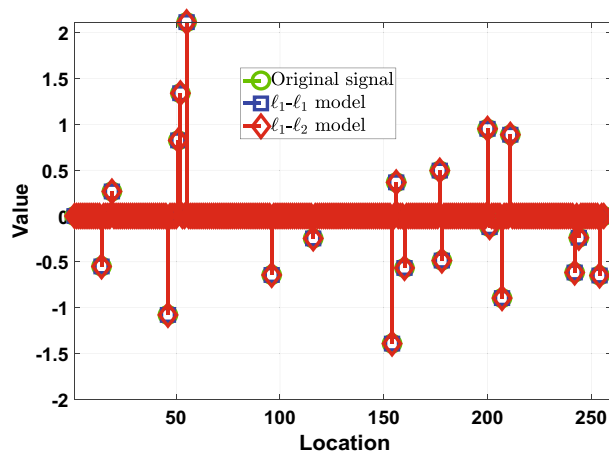


Fig. 1 Recovery performance of the models (1.3) and (1.4) on a given sparse signal with side information

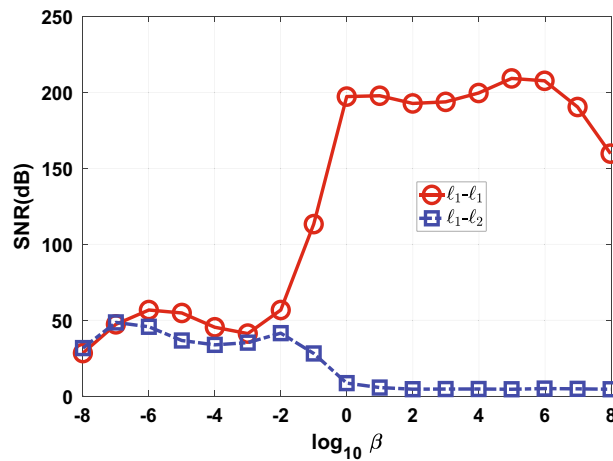


Fig. 2 Selection of the best β 's for both the ℓ_1 - ℓ_1 model and the ℓ_1 - ℓ_2 model

Figure 2 plots the obtained results. According to Fig. 2, in what follows we set $\beta = 10^5$ for the model (1.3) and $\beta = 10^{-6}$ for the model (1.4).

Furthermore, by fixing $k = 20$, $k_w = 10$, we testify the recovery performance of the two models under different number of measurements. In this sort of experiments, we also consider comparing these two models with the classical ℓ_1 model (1.1) and the weighted ℓ_1 model (without noise) in [36], i.e.,

$$\min_{\mathbf{x} \in \mathbb{R}^n} \sum_{i=1}^n w_i |x_i| \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{3.12}$$

where

$$\mathbf{w}_i = \begin{cases} \omega \in [0, 1], & i \in K \\ 1, & i \in K^c \end{cases},$$

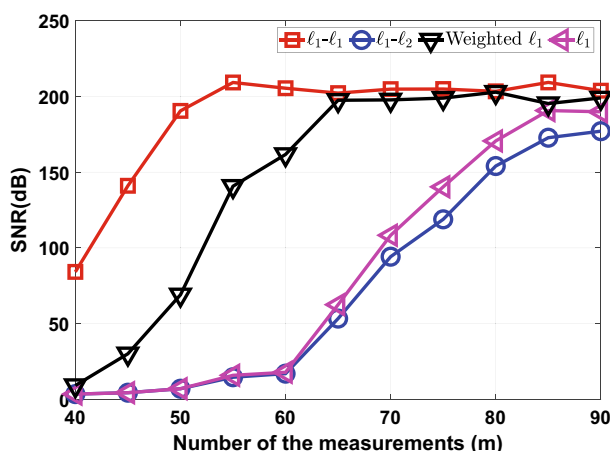


Fig. 3 Recovery performance of four models under different number of the measurements

and $K \subseteq \{1, 2, 3, \dots, n\}$ models the known support set of the original signal \mathbf{x} . For simplicity, we set $K = \text{Supp}(\mathbf{w})$. Since (3.12) is also convex we can solve it easily by means of the CVX. Obviously, the above weighted ℓ_1 model will reduce to (1.2) when one sets $\omega = 0$. According to [36], it is suggested to set ω as closer as 0 when the value of $|K \cap \text{Supp}(\hat{\mathbf{x}})|/|K|$ is as closer as 1. Considering that in our experiments K is strictly included in $\text{Supp}(\hat{\mathbf{x}})$, we set $\omega = 0$ for the weighted ℓ_1 model to boost its best recovery performance. Figure 3 shows the obtained results.

It first shows that an increasing m leads to a better recovery for all three models. However, among these models, the $\ell_1-\ell_1$ model performs best, followed by the weighted ℓ_1 model. The $\ell_1-\ell_2$ model and the classical ℓ_1 model perform worst. It first indicates that a good selection of the constraint on the error of the true signal and its referenced signal plays a key role in enhancing the recovery performance of the models. It also shows that the $\ell_1-\ell_1$ model is better than the weighted ℓ_1 model in taking good advantage of the side information. When it comes to the (classical) ℓ_1 model itself, it is suggested to add an ℓ_1 -norm based error, rather than an ℓ_2 -norm based error, into the objective function of the ℓ_1 model to boost its recovery performance when the side information of the signals becomes available. This observation is also consistent with the conclusion drawn in [21].

Moreover, to further testify the performance of the $\ell_1-\ell_1$ model and the $\ell_1-\ell_2$ model affected by the sparsity of the signals with side information, we consider using these two models to recover the k -sparse signals under different kinds of k_w -sparse referenced signals. Figure 4 first plots the recovery results with k changing from $\{10, 12, 14, \dots, 28\}$ where k_w is set to be $k_w = \lceil k/2 \rceil$.

In general, if the number of the measurements is fixed, a larger k always leads to a poorer recovery. Obviously, this conclusion can be easily drawn from Fig. 4. However, once the side information of the desired signals is available and is also well modeled the recovery performance can be further improved. Since the classical ℓ_1 model do not take the side information into consideration, its performance is weaker than the $\ell_1-\ell_1$ model and the weighted ℓ_1 model. In this experiment, the $\ell_1-\ell_2$ model performs poor again. Moreover, it reconfirms again a fact that if one can take good advantage of the

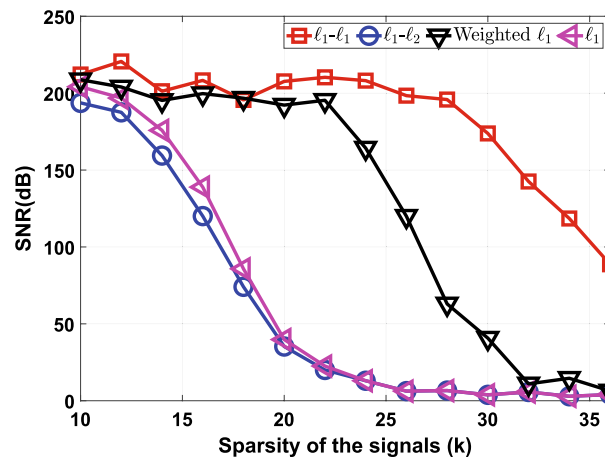


Fig. 4 Recovery performance of four models under different sparsity of the desired signals

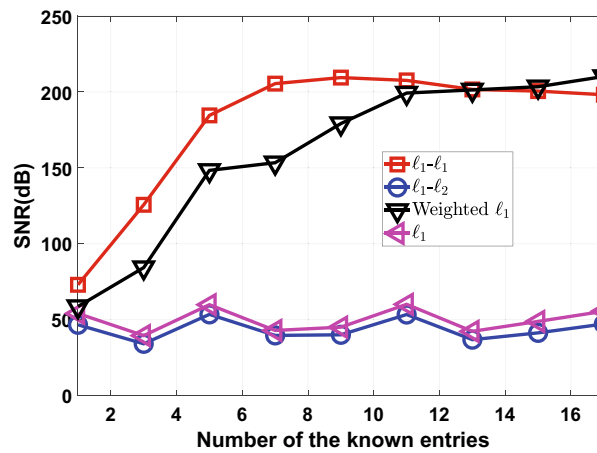


Fig. 5 Recovery performance of four models under different “quality/quantity” of side information

side information and also well model the side information, the recovery performance can be further improved. In Fig. 5, we investigate the recovery performance affected by the “quality/quantity” of the side information. To be specific, we fix the sparsity of the original signals as $k = 20$, and then let k_w change from $\{1, 3, 5, 7, \dots, 17\}$. Obviously, the larger the value of k_w , the higher “quality/quantity” the side information. It is also expected that the recovery performance will be largely improved once the value of k_w is increasing.

Obviously, it is easy to see from Fig. 5 that both the recovery performance of the $\ell_1-\ell_2$ model and the weighted ℓ_1 model is consistent with our expectations, which is far better than the rest two models. Note that in this sort of experiments, we also fix $m = 64$.

At the end of this part, we will conduct a special experiment, in which the referenced signal w is set as $w = \hat{x}$. Under such setting it becomes very important to investigate how the parameter β affects the recovery performance of both the $\ell_1-\ell_1$ model and the $\ell_1-\ell_2$ model. To conduct this experiment, we set $m = 64$, $k = k_w = 20$, and let the parameter β change from $\{10^{-8}, 10^{-7}, \dots, 10^8\}$. Figure 6 plots the obtained results.

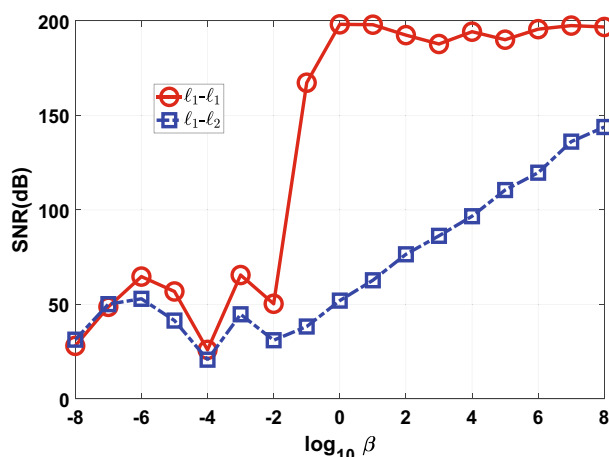


Fig. 6 Recovery performance of the $\ell_1-\ell_1$ model and the $\ell_1-\ell_2$ model under different β 's



Fig. 7 The 128 × 128 test images. They are numbered in order from 1 to 10, from left to right, and from top to bottom

Obviously, a larger β , a better recovery performance of both these two models. When β is relatively small, both the models perform similar. However, when β increases, the $\ell_1-\ell_1$ model performs much better than the $\ell_1-\ell_2$ model. It should also be noted that such a assumption that $w = \hat{x}$ is usually impractical in real-world applications. However, we can rough conclude that a relative bigger β helps the $\ell_1-\ell_1$ model to yield a better recovery performance.

3.2 Experiments on the real-world images

In this part, we consider applying the above-mentioned four models to deal with the real-world image recovery problem.

Figure 7 shows ten real-world images that we will use in the following experiments. As is known to all, the real-world images are generally not nearly sparse themselves, but can be transformed to be nearly sparse by using some sparse dictionaries such as the discrete cosine transform (DCT). On the other hand, almost all the real-world images usually have local smoothness, which indicates that one can use some known information of the original image to help recover some unknown neighboring information of the original image. Therefore, we will take the nearly sparse vectors (generated by applying

Table 1 PSNR|SSIM performance of ten images by four different models

	ℓ_1 model	Weighted ℓ_1 model	ℓ_1 - ℓ_1 model
1	20.58 0.387	25.43 0.612	26.20 0.642
2	19.04 0.379	22.02 0.512	23.10 0.573
3	18.05 0.347	21.72 0.501	22.73 0.565
4	20.83 0.426	23.85 0.562	24.74 0.622
5	19.84 0.412	23.46 0.572	24.57 0.638
6	18.71 0.383	22.25 0.552	23.51 0.627
7	18.33 0.354	21.30 0.476	22.24 0.527
8	19.74 0.353	22.98 0.486	24.08 0.548
9	18.82 0.419	21.63 0.566	22.70 0.634
10	20.88 0.437	23.84 0.589	25.00 0.656

the DCT to each column of the input images) as the test signals to test the recovery performance of these models. To be specific, let the original image G be denoted by $G = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_d]$ with $\mathbf{g}_i \in \mathbb{R}^n$ for $i = 1, 2, \dots, d$, and the DCT dictionary be denoted by $D \in \mathbb{R}^{n \times n}$, then we can easily get the desired sparse (test) signal \mathbf{x}_i by $\mathbf{x}_i = D\mathbf{g}_i$. To model the side information of \mathbf{x}_i , we first generate $\mathbf{r}_i = \mathbf{x}_i + 0.01 * \|\mathbf{x}_i\|_2 * \xi$, where $\xi \in \mathbb{R}^n$ and its elements are generated independently from the standard norm distribution. The signal \mathbf{r}_i can be viewed as the perturbed version of the signal \mathbf{x}_i . As to the support estimate K in the weighted ℓ_1 model, we set K to be the indices of the $\lceil n * 1\% \rceil$ largest absolute elements in \mathbf{r}_i . As to the ℓ_1 - ℓ_1 model and the ℓ_1 - ℓ_2 model, we set the referenced signal of \mathbf{x}_i by \mathbf{w}_i with $(\mathbf{w}_i)_j = (\mathbf{r}_i)_j$ when $j \in K$ and 0 otherwise. As before, we generate the measurement matrix $A \in \mathbb{R}^{m \times n}$ with $m = \lceil n/4 \rceil$ whose elements are generated independently from the standard norm distribution. As to the other parameters, we set them as we have claimed in Sect. 3.1. Obviously, once we obtain the recovered signal one by one, denoted by $\mathbf{x}_i^\#$ the i th recovered signal for $i = 1, 2, \dots, d$, by any of the four models, we can thus obtain the recovered image $G^\#$ by using

$$G^\# = D^{-1}[\mathbf{x}_1^\#, \mathbf{x}_2^\#, \dots, \mathbf{x}_d^\#].$$

To eliminate the column effect on each recovered image, we consider recovering the original images by column and by row, respectively, and then use their average values as the final output. Moreover, to evaluate the quality of the recovered images, we consider using two popular indices, i.e., the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM), more details on these two indices can be found in [37]. Table 1 lists the obtained PSNR|SSIM results of ten test images recovered by three different models, where the highest PSNR and SSIM values are marked in bold.²

It is easy to see that the ℓ_1 - ℓ_1 model performs best among all the models, followed by the weighted ℓ_1 model. The ℓ_1 model and the ℓ_1 - ℓ_2 model perform almost be same, but are all far worse than the ℓ_1 - ℓ_1 model. There results further confirm the claims we have drawn previously. Note that, in this sort of experiments, we still set $\beta = 10^5$ for the ℓ_1 - ℓ_1

² Here we omit the PSNR|SSIM results obtained by the ℓ_1 - ℓ_2 model since they are almost be same with the ones obtained by the ℓ_1 model.

model and $\beta = 10^{-6}$ for the ℓ_1 - ℓ_2 model. It is believed that if β can be further optimized, the final performance of these two models can be further improved.

4 Conclusion and future work

In this paper, we establish two NSP-based sufficient and necessary conditions for two ℓ_1 - ℓ_1 and ℓ_1 - ℓ_2 methods in the case that the side information of the desired signal becomes available. These deterministic theoretical results provide good complement for the previous work on these two methods that are based on the statistical analysis. Besides, some experiments demonstrate ℓ_1 - ℓ_1 method with side information is better than the ℓ_1 - ℓ_2 method, weighted ℓ_1 method and the traditional ℓ_1 method in terms of exact sparse signal recovery when some side information of the desired signals become available.

Note that there also exist a large number of other methods that have been used to deal with the sparse recovery with the prior information of the signals, such as the weighted ℓ_p method [29] and the weighted ℓ_{1-2} [28, 30]. Unfortunately, when compared with the above methods, the ℓ_1 - ℓ_1 method seems to under-perform. To overcome this bad situation, we suggest replacing the ℓ_1 norm used in the ℓ_1 - ℓ_1 method with ℓ_p quasi-norm with $0 < p < 1$ or the ℓ_{1-2} metric, which directly leads to the following two models

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_p + \beta \|\mathbf{x} - \mathbf{w}\|_1 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b},$$

and

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_1 - \|\mathbf{x}\|_2 + \beta \|\mathbf{x} - \mathbf{w}\|_1 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b}.$$

On the other hand, considering that the unknown noise widely exist in many real-world applications, it is very necessary to extend the models (1.3) and (1.4) to more general cases. Inspired by the recent work [33], it is suggested to consider investigating a regularization version of (1.3), i.e.,

$$\mathbf{x}^\# = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_1 + \beta \|\mathbf{x} - \mathbf{w}\|_1 + \theta \|A\mathbf{x} - \mathbf{b}\|_2^2,$$

where $\theta > 0$ is another trade-off parameter. All the above considerations will be our future work.

Abbreviations

CS	Compressed sensing
IHT	Iterative hard thresholding
NSP	Null space property
RIP	Restricted isometry property
RIC	Restricted isometry constant
SNR	Signal to noise ratio
MRI	Magnetic resonance imaging

Acknowledgements

The authors would like to thank the editors and referees for their valuable comments that improve the presentation of this paper. This paper is subsidized by the project of the Natural Science of China (No. 62063028).

Author contributions

All authors read and approved the final manuscript.

Availability of data and materials

Please contact any of the authors for data and materials.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 12 July 2021 Accepted: 17 June 2022

Published online: 28 June 2022

References

1. D.L. Donoho, Compressed sensing. *IEEE Trans. Inf. Theory* **52**(4), 1289–1306 (2006)
2. E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **52**(2), 489–509 (2006)
3. E.J. Candès, T. Tao, Decoding by linear programming. *IEEE Trans. Inf. Theory* **51**(12), 4203–4215 (2005)
4. J. Wright, Y. Ma, J. Mairal et al., Sparse representation for computer vision and pattern recognition. *Proc. IEEE* **98**(6), 1031–1044 (2010)
5. R. Baraniuk, P. Steeghs, Compressive radar imaging. *IEEE Radar Conf.* **2007**, 128–133 (2007)
6. S. Archana, K.A. Narayanankutty, A. Kumar, Brain mapping using compressed sensing with graphical connectivity maps. *Int. J. Comput. Appl.* **54**(11), 35–39 (2012)
7. T. Wan, Z.C. Qin, An application of compressive sensing for image fusion. *Int. J. Comput. Math.* **88**(18), 3915–3930 (2011)
8. M. Lustig, D. Donoho, J. Pauly, Sparse MRI: The application of compressed sensing to rapid MR imaging. *Magn. Reson. Med.* **58**(6), 1182–1195 (2007)
9. M. Duarte, M. Davenport, D. Takhar et al., Single-pixel imaging via compressive sampling. *IEEE Signal Process. Mag.* **25**(2), 83–91 (2008)
10. R.G. Baraniuk, Compressive sensing. *IEEE Signal Process. Mag.* **24**(4), 118–121 (2007)
11. E.J. Candès, M.B. Wakin, S.P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.* **14**(5), 877–905 (2008)
12. S. Boyd, L. Vandenberghe, *Convex optimization*, Cambridge University Press, (2004)
13. H. Esmaili, M. Rostami, M. Kimiaei, Combining line search and trust-region methods for ℓ_1 minimization. *Int. J. Comput. Math.* **95**(10), 1950–1972 (2018)
14. T. Cai, L. Wang, G.W. Xu, Stable recovery of sparse signals and an oracle inequality. *IEEE Trans. Inf. Theory* **56**(7), 3516–3522 (2010)
15. T. Cai, A.R. Zhang, Compressed sensing and affine rank minimization under restricted isometry. *IEEE Trans. Signal Process.* **61**(13), 3279–3290 (2013)
16. R. Zhang, S. Li, A proof of conjecture on restricted isometry property constants $\delta_{0 < t < \frac{4}{3}}$. *IEEE Trans. Inf. Theory* **64**(3), 1699–1705 (2018)
17. G.H. Chen, J. Tang, S. Leng, Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. *Med. Phys.* **35**(2), 660–663 (2008)
18. D. Needell, R. Saab, T. Woolf, Weighted ℓ_1 -minimization for sparse recovery under arbitrary prior information. *Inf. Inference A J. IMA* **6**(3), 284–309 (2017)
19. L.F. Polania, K.E. Barner, Iterative hard thresholding for compressed sensing with partially known support, in *IEEE International Conference on Acoustics, Speech and Signal Processing* (2011), pp. 4028–4031
20. R.E. Carrillo, L.F. Polania, K.E. Barner, Iterative algorithms for compressed sensing with partially known support, in *IEEE International Conference on Acoustics, Speech and Signal Processing* vol. 23(no. 3)(2010), pp. 3654–3657
21. C.J. Miosso, R. Borries, M. Argaez et al., Compressive sensing reconstruction with prior information by iteratively reweighted least-square. *IEEE Trans. Signal Process.* **57**(6), 2424–2431 (2009)
22. H.M. Ge, W.G. Chen, M.K. Ng, New RIP bounds for recovery of sparse signals with partial support information via weighted ℓ_p -minimization. *IEEE Trans. Inf. Theory* **66**(6), 3914–3928 (2020)
23. L. Weizman, Y. Eldar, D. Bashat, Compressed sensing for longitudinal MRI: An adaptive-weighted approach. *Med. Phys.* **42**(9), 5195–5208 (2015)
24. V. Stankovic, L. Stankovic, S. Cheng, Compressive image sampling with side information, in *IEEE International Conference on Image Processing* (2009), pp. 3001–3004
25. L.-W. Kang, C.-S. Lu, Distributed compressive video sensing, in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, (2009), pp. 1169–1172
26. A. Charles, M. Asif, J. Romberg, C. Rozell, Sparsity penalties in dynamical system estimation, in *IEEE Conference on Information Sciences and Systems* (2011), pp. 1–6
27. J.F.C. Mota, N. Deligiannis, M.R.D. Rodrigues, Compressed sensing with prior information: Strategies, geometry, and bounds. *IEEE Trans. Inf. Theory* **63**(7), 4472–4496 (2017)
28. J. Zhang, S. Zhang, X. Meng, ℓ_{1-2} minimization for compressed sensing with partially known signal support. *Electron. Lett.* **56**(8), 405–408 (2020)
29. T. Ince, A. Nacaroglu, N. Watsuji, Nonconvex compressed sensing with partially known signal support. *Signal Process.* **93**(1), 338–344 (2013)
30. J. Zhang, S. Zhang, W. Wang, Robust signal recovery for ℓ_{1-2} minimization via prior support information. *Inverse Probl.* **37**(11), 115001 (2021)
31. S. Foucart, H. Rauhut, *A Mathematical Introduction to Compressive Sensing* (Birkhäuser, Basel, 2013)
32. H.M. Ge, J.M. Wen, et al., *Stable Sparse Recovery with Three Unconstrained Analysis Based Approaches*. <http://alpha.math.uga.edu/mjlai/papers/20180126.pdf>

33. W.D. Wang, J.J. Wang, Robust recovery of signals with partially known support information using weighted BPDN. *Anal. Appl.* **18**(6), 1025–1055 (2020)
34. M.J. Lai, W.T. Yin, Augmented ℓ_1 and nuclear-norm models with a globally linearly convergent algorithm. *SIAM J. Imaging Sci.* **6**(2), 1059–1091 (2012)
35. W. Wang, J. Wang, Z. Zhang, Robust signal recovery with highly coherent measurement matrices. *IEEE Signal Process. Lett.* **24**(3), 304–308 (2016)
36. M.P. Friedlander, H. Mansour, R. Saab et al., Recovering compressively sampled signals using partial support information. *IEEE Trans. Inf. Theory* **58**(2), 1122–1134 (2011)
37. Z. Wang, A.C. Bovik, H.R. Sheikh et al., Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
