$$\boxed{\text{JMM}}$$

A survey on compressive sensing: classical results and recent advancements

Ahmad Mousavi[†*] , **Mehdi Rezaee**[‡], **Ramin Ayanzadeh**[‡]

[†]*Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, MN, USA*
[‡]*Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250, USA*

Email(s): amousavi@umn.edu, rezaee1@umbc.edu, ayanzadeh@umbc.edu

Abstract. Recovering sparse signals from linear measurements has demonstrated outstanding utility in a vast variety of real-world applications. Compressive sensing is the topic that studies the associated raised questions for the possibility of a successful recovery. This topic is well-nourished and numerous results are available in the literature. However, their dispersity makes it time-consuming for practitioners to quickly grasp its main ideas and classical algorithms, and further touch upon the recent advancements. In this survey, we overview vital classical tools and algorithms in compressive sensing and describe its significant recent advancements. We conclude by a numerical comparison of the performance of described approaches.

Keywords: Compressive sensing, ℓ_p recovery, greedy algorithms.
AMS Subject Classification 2010: 34A34, 65L05.

1 Introduction

In traditional sensing, i.e., uniform sampling, we need to densify measurements to obtain a higher-resolution representation of physical systems. But in applications like multiband signals with wide spectral ranges, the required sampling rate may exceed the specifications of the best available

analog-to-digital converters (ADCs) [98]. On the other hand, measurements obtained from linear sampling methods approximately carry a close amount of information, which makes them reasonably robust; for example, for packages lost in data streaming applications [93]. Compressive sensing linearly samples sparse signals at a rate lower than the traditional rate provided by the Nyquist-Shannon sampling theorem [135].

Recovering a sparse signal from a linear measurement is of high interest to preserve more storage, to have less computation, energy, and communication time, and to propose effective data compression and transmission methods [8, 9, 16, 25, 46, 53, 69, 98, 99, 104, 122, 144, 147]. There are also specialised motivations behind this interest; for example, the recovery process in compressive sensing requires prior knowledge about original signals, which adds reasonable levels of security and privacy to the compressive sensing based data acquisition framework [140]. Compressive sensing-based face recognition techniques are invariant to rotation, re-scaling and data translation [98]. Besides, compressive sensing has demonstrated outstanding numerical stability in terms of not only noisy measurements but also quantization errors [101]. In machine learning, compressive sensing has demonstrated to boost the performance of pattern recognition methods.

The sparsity assumption as the basis of compressive sensing is not practically restrictive because it is empirically observed that desired signals in an extensive range of applications are sparse, possibly after a change of basis though. However, blind compressive sensing techniques are universal in terms of transfer domains, namely, there is no need to know the sparsifying basis while taking linear measurements [58]. Sparsity notion directly leads to NP-hard problems and thus the original problem of compressive sensing is computationally intractable; see Section 2 for the definition of this problem. Two main strategies to tackle this obstacle are convex or nonconvex relaxations and greedy algorithms; see Sections 3 and 4. There are several other effective approaches that have different ideas than these two categories. Hence, this survey attempts to briefly characterize these promising studies as well.

To relax an ℓ_0-quasi-norm based sparse optimization problem, ℓ_0-quasi-norm is often replaced by ℓ_p-(quasi)-norm with $p > 0$. Then, it is investigated that under which conditions on problem parameters and for which p's both original and approximate problems uniquely obtain the same solution. The efficiency of $p = 1$ is well-documented [22, 23, 27, 36, 42, 44, 54]. But since this choice does not lead into strictly convex programs, the solution uniqueness shortcoming needs to be addressed [81, 138]. This case converts to a linear program and since those matrices involved in sparse optimization

often inherit large dimensions, well-known methods including simplex and interior-point methods or specified algorithms are applied [20, 103]. The shape of unit balls associated with $0 < p < 1$ encourages examining this case as well. It turns out that the recovery results for certain values of $0 < p < 1$ are more robust and stable compared to $p = 1$ [33, 34]. Despite these favorable theoretical results, nonconvex relaxations ask for a global minimizer, which is an intractable task. One numerical way to bypass this is by utilizing classical schemes that obtain a local minimizer with an initial point sufficiently close to a global minimizer. For example, the solution of least squared is an empirical suggestion for this initialization [33], although there is no guarantee that such a solution is close enough. Nonetheless, a recent theoretical study proves that the main representative problems utilized in the realm of sparse optimization, such as the generalized basis pursuit and LASSO, almost always return a full-support solution for $p > 1$ [108].

Greedy algorithms in our second class have low computational complexity, especially for relatively small sparsity levels, and yet they are effective. A well-received algorithm in this class which plays a key role is the orthogonal matching pursuit (OMP). To directly solve the original problem of compressive sensing, it exploits the best local direction in each step and adds its corresponding index to the current support set. Then, it estimates the new iteration as the orthogonal projection of the measurement vector onto the subspace generated by columns in the current support set [116, 118]. There is an extensive literature on the capability of this algorithm in identifying the exact support set that vary based on the required number of steps and accounting for noise in the measurement vector. For example, a sufficient condition for recovering an s sparse vector from an exact linear measurement asks for an $(s+1)$st restricted isometry constant strictly smaller than $(\sqrt{s}+1)^{-1}$ [77]. This upper bound improves to the necessary and sufficient bound of $1/\sqrt{s+1}$. A result with the same spirit for the noisy measurement is available that imposes an extra assumption on the magnitude of the smallest nonzero entry of a desired sparse vector [127]. Further, stability is also achievable with the presence of noisy measurements if the number of required steps is increased [142]. Nevertheless, there are other algorithms in this class that have the same spirit but attempt to resolve issues of the OMP, such as stagewise OMP [43] and multipath matching pursuit [70]. For example, many algorithms allow more indices to enter the current support set for picking a correct index in each step such as the compressive sampling matching pursuit [85] and subspace pursuit [124]. Other effective greedy algorithms extending the OMP are simultaneous, generalized, and grouped

OMP [112, 117, 121, 129]. Nonnegative and more general constrained versions of the OMP obtain similar recovery results based on variations of the restricted isometry and the orthogonality constants [18, 72, 109]. These algorithms are also generalized for block sparse signals [47, 66].

The rest of paper is organized as follows. In Section 2, we first introduce the original problem of compressive sensing and its extensions and variations. Section 3 is devoted to ℓ_p recovery which covers using convex and nonconvex relaxations. In Section 4, we study several well-received greedy algorithms. Finally, we report numerical performance results in Section 5.

Notation. For a natural number N, we let $[N] := \{1, 2, \ldots, N\}$. For a set $S \subseteq [N]$, its number of elements and complement are denoted by $\mathrm{card}(S)$ (or $|S|$) and S^c, respectively.

For a vector $x \in \mathbb{R}^N$ and $p > 0$, let $\|x\|_p := (\sum_{i \in [N]} |x_i|^p)^{1/p}$. For $S = \{i_1, i_2 \ldots, i_{|S|}\} \subseteq [N]$, we let $x_S := z \in \mathbb{R}^N$ such that $z_i = x_i$ for $i \in S$ and $z_i = 0$ whenever $i \in S^c$, that is, x_S denotes the coordinate projection of x onto the subspace generated by $\{e_{i_1}, e_{i_2}, \ldots, e_{i_{|S|}}\}$. Given $x \in \mathbb{R}^N$, the hard thresholding operator of order s is defined as $H_s(x) := x_S$ where S contains s largest absolute entries of x (notation $L_s(x)$ denotes this index set). Further, let $x_+ := \max(x, 0)$. A vector is called s-sparse if it has at most s nonzero entries, i.e., its sparsity level is bounded by s. We use $\langle ., . \rangle$ to denote the standard inner product between two vectors. We employ $\mathbf{1}_N$ to denote a vector of size N with each entry equal to 1, and simply let 0 denote a zero vector, which its dimension should be clear based on the discussed content. For a matrix $A \in \mathbb{R}^{m \times N}$, we denote its transpose by A^T, its adjacent by A^* and its Frobenius norm by $\|A\|_F$. We denote the jth column of A by A_j and, given a subset $S \subseteq [N]$, a corresponding column submatrix is defined as $A_S := [A_{j_1}, A_{j_2}, \ldots, A_{j_{|S|}}]$. The null and range spaces of this matrix are denoted by $\mathcal{K}er(A)$ and $\mathcal{R}(A)$, respectively. The symbol $\mathcal{O}$ stands for the standard big O notation.

2 Compressive sensing problem and its extensions and variations

Compressive sensing (CS), also known as compressed sensing or sparse sampling, is established based on sparsity assumption. Hence, we start with preliminary definitions of sparsity and, a closely related concept, compressibility.

Definition 1. For a vector $x \in \mathbb{R}^N$, let $\mathrm{supp}(x) := \{i \in [N] \,|\, x_i \neq 0\}$ and $\|x\|_0 := \mathrm{card}(\mathrm{supp}(x))$. In particular, for $s \in [N]$, this vector is called

s-sparse if $\|x\|_0 \leq s$.

Definition 2. For a vector $x \in \mathbb{R}^N$, $s \in [N]$, and $p > 0$, let

$$\sigma_s(x)_p := \min_{z \in \mathbb{R}^N} \|x - z\|_p \quad \text{subject to} \quad \|z\|_0 \leq s.$$

This vector is called s-compressible if $\sigma_s(x)_p$ is small for some $p > 0$. This vector is sometimes called nearly s-sparse in ℓ_p-norm.

It is easy to show that $\sigma_s(x)_p = \|x - x_S\|_p$, where $S \subseteq [N]$ contains all the s largest absolute entries of x (consequently, $\sigma_s(x)_p = 0$ when x is s-sparse). Roughly speaking, a vector is called sparse if most of its entries are zero and it is called compressible if it is well-approximated by a sparse vector. Sparsity as a prior assumption for a desired vector is consistent with diverse applications because it is often doable to employ a change of basis technique for finding sparse representations in a transform domain. For instance, wavelet, Radon, discrete cosine, and Fourier transforms are well-known as suitable and efficient choices for natural, medical, and digital images, and speech signals, respectively.

The original problem of compressive sensing aims to recover a sparse signal x from a linear measurement vector $y = Ax$, where $A \in \mathbb{R}^{m \times N}$ (with $m \ll N$) is the so called measurement, coding or design matrix. Hence, the ultimate goal of compressive sensing is reasonably formulated as follows:

$$\min_{x \in \mathbb{R}^N} \|x\|_0 \quad \text{subject to} \quad Ax = y. \tag{CS}$$

To incorporate noisy measurements, one can study this problem:

$$\min_{x \in \mathbb{R}^N} \|x\|_0 \quad \text{subject to} \quad \|Ax - y\|_2 \leq \eta,$$

where $\eta > 0$ controls the noise level. These problems have demonstrated to revolutionize several real-world applications in both science and engineering disciplines, including but not limited to signal processing, imaging, video processing, remote sensing, communication systems, electronics, machine learning, data fusion, manifold processing, natural language and speech processing, and processing biological signals [75, 94, 96, 100, 139].

In a broad range of real-world applications, we utilize ADCs to map the real-valued measurements of physical phenomena (over a potentially infinite range) to discrete values (over a finite range). As an illustration, when we use b bits for representing the measurements digitally, the quantization module in an ADC maps measurements to one of the 2^b distinct values that introduces error in measurements. Quantization module plays

a bottleneck role in restricting the sampling rate of the ADCs because the maximum sampling rate decreases exponentially when the number of bits per measurement increases linearly. Quantization module also is the main source of energy consumption in ADCs. Single-bit (or 1-bit) compressive sensing enables us to reduce the number of bits per measurements to one and introduces a proper model for successful recovery of the original signal. This extreme quantization approach only retains the sign of measurements, i.e., we have $y_i \in \{1, -1\}$ for all $i \in [m]$, which results in a significantly efficient, simple, and fast quantization [17, 65].

A recent generalization of (CS) is the so called compressive sensing with matrix uncertainty. This uncertainty finds two formulations that incorporate measurement errors as well. The first formulation is the following:

$$\min_{x \in \mathbb{R}^N, \ E \in \mathbb{R}^{m \times N}} \|(A + E)x - y\|_2^2 + \lambda_E \|E\|_F^2 + \lambda \|x\|_0,$$

where the matrix E is the perturbation matrix. The second one is the following:

$$\min_{x \in \mathbb{R}^N, \ d \in \mathbb{R}^r} \left\|(A^{(0)} + \sum_{i \in [r]} d_i A^{(i)})x - y\right\|_2^2 + \lambda_d \|d\|_2^2 + \lambda \|x\|_0,$$

where the measurement matrices $A^{(i)}$'s for $i = 0, 1, \ldots, r$ are known and the unknown vector d is the uncertainty vector. In quantized compressed sensing, we have either $y = \mathcal{Q}Ax$ or $y = \mathcal{Q}(Ax + v)$ where v is the noise such that $v \sim \mathcal{N}(0, \sigma^2 I)$ and $\mathcal{Q} : \mathbb{R}^m \to \mathcal{A} \subseteq \mathbb{R}^m$ is the set-valued quantization function. Well-studied examples of such function, map Ax or $Ax + v$ into $\mathcal{A} = \{x \in \mathbb{R}^m \mid l \leq x \leq u\}$ or $\mathcal{A} = \{+1, -1\}^m$ [65, 132, 149]. The former case leads to:

$$\min_{x \in \mathbb{R}^N} \|x\|_0 \quad \text{subject to} \quad l \leq Ax + v \leq u,$$

and the latter one yields the following problem:

$$\min_{x \in \mathbb{R}^N} \|x\|_0 \quad \text{subject to} \quad y = \text{sign}(Ax + v).$$

In some recent applications, we may have nonlinear measurements that encourage the so called cardinality constrained optimization:

$$\min_{x \in \mathbb{R}^N} f(x) \quad \text{subject to} \quad \|x\|_0 \leq s.$$

Two important examples are $f(x) = \|Ax - y\|^2$ (motivated by linear measurements $y = Ax$) and $f(x) = \sum_i^m (y_i - x^T A^{(i)} x)^2$ (proposed for quadratic

measurements of $y_i = x^T A^{(i)} x$ for $i = 1, 2, \ldots, m$ and symmetric matrices $A^{(i)}$'s) [107, 133]. Many algorithms have been proposed for solving this cardinality constrained problem with specific and general smooth functions [12–14, 19, 106]. In matrix setting, this problem converts to rank minimization, which finds numerous applications [48]. Although we listed some vital variations of compressive sensing above, our focus in this survey is on the original problem to motivate the important tools, algorithms and results in compressive sensing for a linear measurement possibly contaminated with some error or noise.

3 ℓ_p recovery with $0 < p \le 1$: main formulations and results

Since the problem (CS) is NP-hard [82], it is not tractable and it must be handled indirectly. Because ℓ_p-norm approximates ℓ_0-norm as p goes to zero, one approach to tackle (CS) is exploiting the so called Basis Pursuit (BS) problem:

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{subject to} \quad Ax = y. \tag{BP}$$

Then, one investigates under which conditions (CS) and (BP) appoint the same solution. This equivalence property occurs when the solution is highly sparse and the measurement matrix has sufficiently small mutual incoherence (defined below) [44].

Definition 3. For a matrix $A \in \mathbb{R}^{m \times N}$, the mutual coherence is defined as

$$\mu(A) := \max_{i \neq j} \frac{|\langle A_i, A_j \rangle|}{\|A_i\|_2 . \|A_j\|_2},$$

where A_i denotes the ith column vector of the matrix A.

The mutual coherence simply seeks the largest correlation between two different columns. The following global 2-coherence is a generalization of the mutual coherence ($\mu = \nu_1$):

Definition 4. For a matrix $A \in \mathbb{R}^{m \times N}$, its kth global 2-coherence is defined as below:

$$\nu_k(A) := \max_{i \in [N]} \max_{\Lambda \subseteq [N] \setminus \{i\}, |\Lambda| \le k} \left(\sum_{j \in \Lambda} \frac{\langle A_i, A_j \rangle^2}{\|A_i\|_2^2 . \|A_j\|_2^2} \right)^{1/2}.$$

This tool is useful for a successful recovery of a weak version of orthogonal matching pursuit [134], while we focus on the standard mutual coherence here to convey the main idea.

Theorem 1. *[45] Assume that $\|x\|_0 < 0.5(1 + \mu(A)^{-1})$. Then, problems* CS *and* BP *uniquely obtain the signal* x.

Sparsity is an inevitable assumption in the recovery process. This assumption accompanied by other conditions on a measurement matrix such as the mutual coherence or restricted isometry property (RIP) guarantee a successful recovery through the problem BP. For example, the bound in Theorem 1 is improved using the RIP tool defined below.

Definition 5. Let A be an $m \times N$ matrix. Then, this matrix has the restricted isometry property of order s provided that there exists $\delta_s \in (0, 1)$ such that

$$(1 - \delta_s)\|x\|_2^2 \ \leq \ \|Ax\|_2^2 \ \leq \ (1 + \delta_s)\|x\|_2^2 \qquad \forall x; \ \|x\|_0 \leq s.$$

This definition demands each column submatrix A_S with card$(S) \leq s$ to have singular values in $[1 - \delta_s, 1 + \delta_s]$. Consequently, such submatrices have linearly independent columns if $\delta_s < 1$. Since this definition involves all the s-tuples of columns, it is more rigorous than the mutual coherence ($\delta_2 = \mu$). Thus, the RIP tool derives to better upper bounds on the sparsity level of a vector to be recovered, that is, less sparse vectors are properly handled.

Theorem 2. *[78] Problems* CS *and* BP *uniquely obtain the same s-sparse solution if $\delta_{2s} < 0.4931$.*

There are numerous similar results for recovering s-sparse signals in compressive sensing literature based on the RIP constants, which are often in the form of $\delta_{ks} \leq \delta$ for some numbers $k > 0$ and $\delta \in (0, 1)$. For example, the BP recovers all the s-sparse signals if $\delta_s < 1/3$ and this bound is sharp [24]. Since the condition $\delta_{ks} \leq \delta$ is not practically verifiable due to its computational complexity [74], a breakthrough in the compressive sensing field took place when random matrices, such as Gaussian and Bernoulli, demonstrated their capability to satisfy such appealing inequalities with a high probability. In fact, for an $m \times N$ random matrix, where each entry is independently drawn from Gaussian or Bernoulli distributions, we have $\delta_s \leq \delta$ for $m \geq C\delta^{-2}s\ln(\frac{eN}{s})$ in which $C > 0$ does not depend on s, m and N, and e is the natural number [11]. It must be mentioned that most results for constructing deterministic matrices with the RIP property depend on the mutual coherence, leading to $m \geq cs^2$ with $c > 0$. This bound is not practical because the number of measurements m scales quadratically in sparsity level s for deterministic matrices versus linearly for random matrices. This confirms the desire of using random matrices

in compressive sensing. On the other hand, there are recent promising results toward the practicality of deterministic matrices in terms of not only the compression ratio of m/N but also the complexity of the recovery process [3, 28, 59, 87]. For example, deterministic sparse coding matrices have demonstrated to outperform both binary tree recovery and binary ℓ_1-norm recovery methods in the so called binary compressive sensing [84]. The majority of analyses within compressive sensing relies on the RIP but this tool has minor drawbacks that are partially rectified via the NSP defined below.

Definition 6. Let A be an $m \times N$ matrix. This matrix has the null space property of order s if

$$\|v_S\|_1 < \|v_{S^c}\|_1; \quad \forall v \in \mathcal{K}er(A) \setminus \{0\} \text{ and } \forall S \subseteq [N] \text{ with } |S| \leq s.$$

In principle, the measurement pair (A, y) carries all the required information for the recovery process. Since linear systems $Ax = y$ and $PAx = Py$ have identical solution sets for a nonsingular matrix P, the measurement pair (PA, Py) has the same information but it is numerically a better choice for a suitable conditioning matrix P. Although the RIP constants of matrices A and PA can vastly differ [141], the NSP is preserved if either one holds it. In addition, this property is necessary and sufficient for the solution uniqueness of problem BP (in the uniform sense).

Theorem 3. *[53] Every s-sparse vector x is the unique solution to BP with $y = Ax$ if and only the measurement matrix A holds the NSP of order s.*

The above theorem implies that under the NSP of order s, the convex program BP as a matter of fact solves the NP-hard problem CS.

The Range Space Property (RSP) is another useful condition in compressive sensing given by Zhao [146].

Definition 7. The matrix A^T has the range space property of order s if for any disjoint subsets S_1 and $S_2 \subseteq [N]$ with $|S_1| + |S_2| \leq s$ there exists a vector $\eta \in \mathcal{R}(A^T)$ such that

$$\eta_i = 1 \quad \forall i \in S_1; \quad \eta_i = -1 \quad \forall i \in S_2 \quad \text{and} \quad \|\eta_{(S_1 \cup S_2)^c}\|_\infty < 1.$$

Theorem 4. *[146] Every s-sparse vector x is exactly recovered by the problem BP with $y = Ax$ if and only if A^T has the RSP of order s.*

The matrix A^T has the range space property of order s under several assumptions such as $s < 0.5(1 + \mu(A)^{-1})$, $\delta_{2s}(A) < \sqrt{2} - 1$ and the NSP

of order $2s$ [146]. Further, an extended version of this property is useful to provide a similar result to the above theorem for the nonnegative sparse recovery.

So far, we have only discussed noiseless measurements in this section, although a practical situation definitely imposes investigating noisy scenario as well. Taking into account noisy measurements yields in the following Quadratically Constrained Basis Pursuit (QCBP) problem (after replacing ℓ_0 by ℓ_1):

$$\min_{x\in\mathbb{R}^N} \ \|x\|_1 \quad \text{subject to} \quad \|Ax - y\|_2 \leq \eta. \qquad \text{(QCBP)}$$

Theorem 5. *[53] Given $x \in \mathbb{R}^N$ and a matrix $A \in \mathbb{R}^{m\times N}$ such that $\delta_{2s} < 4/\sqrt{41} \approx 0.6246$, every minimizer x^* of QCBP satisfies $\|x^* - x\|_2 \leq Cs^{-1/2}\sigma_s(x)_1 + D\eta$ with $C > 0$ and $D > 0$.*

The following so called Least Absolute Shrinkage and Selection Operator (LASSO) is practically more efficient [50, 64, 128]:

$$\min_{x\in\mathbb{R}^N} \quad \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1. \qquad \text{(LASSO)}$$

There is a trade-off between feasibility and sparsity in the nature of this problem, which is controlled by the regularization parameter λ. Not only the LASSO has been extensively studied [113], but also there are various results for it in terms of its effectiveness in sparse optimization. For instance, we bring the following theorem for deterministic measurement matrices.

Theorem 6. *[64] Let $y = Ax + e$ where x is supported on $S \subseteq [N]$ with $|S| \leq s$ and e is a zero-mean additive observation noise. Assume that the measurement matrix A satisfies $\|A_{S^c}^T A_S (A_S^T A_S)^{-1}\|_\infty < 1$, $\lambda_{\min}(A_S^T A_S) > m$ and $\max_{j\in S^C}\|A_j\|_2 \leq \sqrt{m}$. Further, assume that $N = \mathcal{O}(\exp(m^\gamma))$, $s = \mathcal{O}(m^\alpha)$, $\min_{i\in S} x_i > 1/m^{\frac{1-\beta}{2}}$ with $0 < \alpha + \gamma < \beta < 1$. For $\lambda = m^{\frac{1-\delta}{2}}$ such that $\delta \in (\gamma, \beta - \alpha)$, the LASSO problem recovers the sparsity pattern S with probability $1 - \exp(-cm^\delta)$ for a constant $c > 0$.*

For the case of random measurement matrices, this problem needs a sample size $m > 2s\ln(N - s)$ to achieve exact recovery with a high probability. This probability converges to one for larger problems [64]. An efficient algorithm for solving the BP utilizes a sequence of the LASSO problems [136]. The following so called Basis Pursuit Denoising (BPD) minimizes the feasibility violation while implicitly bounding the sparsity level with a parameter $\tau \geq 0$:

$$\min_{x\in\mathbb{R}^N} \ \|y - Ax\|_2 \quad \text{subject to} \quad \|x\|_1 \leq \tau. \qquad \text{(BPD)}$$

There are known relations among the optimums of the QCBP, LASSO, and BPD problems, which can be found in [53]. The Dantzig Selector (DS) problem arises in several statistical applications, so it has been also employed in sparse recovery [23, 26]:

$$\min_{x \in \mathbb{R}^N} \ \|x\|_1 \quad \text{subject to} \quad \|A^T(Ax - y)\|_\infty \leq \sigma. \tag{DS}$$

This problem manages noisy measurements and reduces to a linear programming problem. To bring a pertaining result next, we need the following restricted orthogonality quantity.

Definition 8. Given a matrix $A \in \mathbb{R}^{m \times N}$, its s, s'-restricted orthogonality (RO) constant $\theta_{s,s'}$ is defined as the smallest $\theta > 0$ such that

$$|\langle Ax, Ax' \rangle| \leq \theta \|x\|_2 \|x'\|_2 \quad \forall x \text{ and } x'; \ \|x\|_0 \leq s \text{ and } \|x'\|_0 \leq s'.$$

According to [23], for $x \in \mathbb{R}^N$, and a noisy y such that $\|A^T(Ax-y)\|_\infty \leq \sigma$ and $\delta_{1.5s} + \theta_{s,1.5s} < 1$, an optimal solution x^* to DS obeys $\|x^* - x\|_2 \leq Cs^{\frac{1}{2}}\sigma + Ds^{\frac{-1}{2}}\sigma_s(x)_1$, where C and D are two constants depending on $\delta_{1.5s}$ and $\theta_{s,1.5s}$. In particular, if x is an s-sparse vector, then $\|x^* - x\|_2 \leq Cs^{\frac{1}{2}}\sigma$. A very recent result related to the recovery of this problem as reported in Corollary 5.6.1 of [144] is brought below.

Theorem 7. *[144] Let $A \in \mathbb{R}^{m \times N}$ with $m < N$ and $rank(A) = m$. Consider the DS problem. If A^T satisfies the RSP of order k, then for any $x \in \mathbb{R}^n$, there is a solution x^* of the DS approximating x with error*

$$\|x - x^*\|_2 \leq \gamma\{2\sigma_k(x)_1 + (\|A^T(Ax - y)\|_\infty - \sigma)_+ + c_A\sigma + c_A\|A^T(Ax-y)\|_\infty\},$$

where γ is a constant and c_A is a constant given as

$$c_A = \max_{G \subseteq \{1,\ldots,n\},|G|=m} \|A_G^{-1}(AA^T)^{-1}A\|_{\infty \to 1},$$

where A_G stands for all positive $m \times m$ invertible submatrix of A. In particular, for any x satisfying $\|A^T(Ax - y)\|_\infty \leq \sigma$, there is a solution x^* of the DS approximating x with error

$$\|x - x^*\|_2 \leq \gamma\{2\sigma_k(x)_1 + c_A\sigma + c_A\|A^T(Ax - y)\|_\infty\} \leq 2\gamma\{\sigma_k(x)_1 + c_A\sigma\}.$$

However the choice of $p = 1$ is the most interesting case, as ℓ_1-norm is the closest convex norm to ℓ_0-quasi-norm [97] and convex optimization is extremely well-nourished, the shape of a unit ball associated with ℓ_p-norm

for $0 < p < 1$ motivates researchers to explore the following nonconvex problem:

$$\min_{x \in \mathbb{R}^N} \ \|x\|_p^p \quad \text{subject to} \quad Ax = y.$$

For certain values of p where $0 < p < 1$, the above scheme leads to more robust and stable theoretical guarantees compared to the case of $p = 1$ [34,102,130]. In other words, much less restrictive recoverability conditions are achievable for certain p's with $0 < p < 1$ compared to ℓ_1-norm recovery [148]. For instance, the following result implies that a sufficient condition for recovering an s-sparse vector in the noiseless case via $\ell_{0.5}$ minimization is $\delta_{3s} + 27\delta_{4s} < 26$, where an analogous result for ℓ_1-norm recovery requires $\delta_{2s} + 2\delta_{3s} < 1$. We bring the following related theorem, which discusses compressible vectors. Since it is practically stringent to impose sparsity, it is desired to study compressible vectors.

Theorem 8. *[102] Assume that $A \in \mathbb{R}^{m \times N}$ satisfies*

$$\delta_{ks}(A) + k^{\frac{2}{p}-1}\delta_{(k+1)s}(A) < k^{\frac{2}{p}-1} - 1,$$

for $k > 1$ such that ks is a natural number. Given $x \in \mathbb{R}^N$, let $y = Ax + \boldsymbol{e}$ with $\|\boldsymbol{e}\|_2 \leq \epsilon$. Then, an optimal solution x^* of

$$\min_{x \in \mathbb{R}^N} \|x\|_p^p \quad \text{subject to} \quad \|Ax - y\|_2 \leq \epsilon, \tag{1}$$

obeys

$$\|x^* - x\|_2^p \leq C\epsilon^p + Ds^{\frac{p}{2}-1}[\sigma_s(x)_p]^p,$$

where

$$C = 2^p \left[\frac{1 + k^{p/2-1}(2/p - 1)^{-p/2}}{(1 - \delta_{(k+1)s})^{\frac{p}{2}} - (1 + \delta_{ks})^{p/2}k^{p/2-1}} \right], \tag{2}$$

$$D = \frac{2(\frac{p}{2-p})^{p/2}}{k^{1-p/2}} \left[1 + \frac{(1 + k^{p/2-1})(1 + \delta_{ks})^{p/2}}{(1 - \delta_{(k+1)s})^{p/2} - (1 + \delta_{ks})^{p/2}k^{p/2-1}} \right]. \tag{3}$$

In particular, if x is an s-sparse vector, then $\|x^* - x\|_2^p \leq C\epsilon^p$.

The main issue with this result is its demand for a global minimizer of a nonconvex function for which there is no known theoretical guarantee. One way to bypass this is by utilizing classical schemes that obtain a local minimizer with an initial point that is sufficiently close to a global minimizer. For example, the solution of least-squares is experimentally suggested for this initialization [33], while there is no guarantee that such solution is in

general close enough to a global minimizer. Since solving an unconstrained problem is simpler, another approach for taking advantage of nonconvex ℓ_p-minimization (1) with $0 < p < 1$ is tackling the following ℓ_p-regularized least squares problem:

$$\min_{x \in \mathbb{R}^N} \ \frac{1}{2}\|Ax - y\|_2^2 + \lambda\|x\|_p^p, \tag{4}$$

where $\lambda > 0$ is the penalty parameter. Because problems (1) and (4) are equivalent in limit, this parameter must be selected meticulously to obtain an approximate solution for the original problem. Once $\lambda > 0$ is fixed, algorithms such as iteratively-reweighted least squares explained in [105] are beneficial. Furthermore, the choice of p plays a key role in the efficiency of this approach; for example, the best choices in image deconvolution are $p = 1/2$ and $p = 2/3$. The case of $p = 1/2$ is a critical choice because it provides sparser solutions among $p \in [1/2, 1)$ and any $p \in (0, 1/2)$ does not show significantly better performance [119]. As a result, it is essential to have a specialized algorithm for this choice [131].

Despite $p > 1$ leads to strictly convex programming, it is theoretically shown that in this case not only the solution is not sparse, but also each entry is almost always nonzero. The same result holds for all the primary problems mentioned above.

Theorem 9. *[108] Let $p > 1, N \geq m, \lambda > 0,$ and $\tau > 0$. For almost all $(A, y) \in \mathbb{R}^{m \times N}$, a unique optimal solution $x^*_{(A,y)}$ to the any of the problems BP, QCBP, LASSO and BPD, when $\|.\|_1$ is replaced by $\|.\|_p^p$, satisfies $|\text{supp}(x^*_{(A,y)})| = N$.*

Recent applications in image processing, statistics and data science lead to another class of sparse optimization problems subject to the nonnegative constraint, i.e.,

$$\min_{x \in \mathbb{R}^N} \ \|x\|_0 \quad \text{subject to} \quad Ax = y, \quad x \geq 0. \tag{NCS}$$

This problem is similar to CS but their optimums differ in general. Similar convex/nonconvex relaxation approaches are advantageous to tackle NCS problem. Its convex relaxation leads to the so called Nonnegative Basis Pursuit (NBP) [42, 68, 145]:

$$\min_{x \in \mathbb{R}^N} \ \|x\|_1 \quad \text{subject to} \quad Ax = y, \quad x \geq 0. \tag{NBP}$$

Theorem 10. *[145] Any $x \geq 0$ such that $\|x\|_0 \leq s$ is recovered by NBP with $y = Ax$ if and only if for any index subset $S \subseteq [N]$ such that $|S| \leq s$ there exists $\eta \in \mathcal{R}(A^T)$ such that $\eta_i = 1$ for $i \in S$ and $\|\eta_{S^c}\|_\infty < 1$.*

For a theoretical study on the following nonnegative version of LASSO formulation, see [63]:

$$\min_{x \in \mathbb{R}^N} \quad \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1 \quad \text{subject to} \quad x \geq 0. \qquad \text{(NLASSO)}$$

Nonnegativity is an example of possible available prior information in the recovery process, which enables us to employ sparse measurement matrices for the recovery of remarkably larger signals [68]. In applications like infrared absorption spectroscopy, the non-zero elements of the original sparse signal are bounded. Imposing this boundary condition reduces the optimal set, and therefore its convex relaxation leads to a bounded/boxed basis pursuit that obtains better recovery properties [40, 41, 73].

Majority of studies in compressive sensing focuses on continues signals where $x \in \mathbb{R}^N$, even though there are several real-world applications that include discrete signals [73]. Examples are discrete control signal design, black-and-white or gray-scale sparse image reconstruction, machine-type multi-user communications, and blind estimation in digital communications. Here, we aim to reconstruct a discrete signal whose elements take their values from a finite set of alphabet. It is worth highlighting that when x is discrete, even the ℓ_1-norm recovery method leads to an NP-hard problem [71]. Nevertheless, the study [71] reveals this additional prior knowledge about the original signal enhances the performance of both ℓ_0 and ℓ_1 norm recovery methods via imposing new constraints. For example, the sum of absolute values is designed to recover discrete sparse signals whose non-zero elements are generated from a finite set of alphabets with a known probability distribution. In binary compressive sensing, we can represent this model as follows:

$$\min_{x \in \mathbb{R}^N} \quad (1-p)\|x\|_1 + p\,\|x - \mathbf{1}_N\|_1 \quad \text{subject to} \quad Ax = y, \qquad (5)$$

where p is the probability of each component being one; analogous to the sparsity rate [67, 73, 83]. In this approach, when p goes to zero (x is sparse), the problem (5) behaves like the BP. The sum of norms represents the problem of binary compressive sensing as follows:

$$\min_{x \in \mathbb{R}^N} \quad \|x\|_1 + \lambda\,\|x - \tfrac{1}{2}\mathbf{1}_N\|_\infty \quad \text{subject to} \quad Ax = y,$$

where the parameter $\lambda > 0$ keeps ℓ_1 and ℓ_∞ balanced [120]. It is observed that the ℓ_∞-norm minimization tends to capture a representation whose coefficients have roughly the same absolute value [56, 111]. Hence, the objective function $\|x\|_1 + \lambda\,\|x - \tfrac{1}{2}\mathbf{1}_N\|_\infty$ is two fold. First, the sparsity is

imposed via the ℓ_1-norm term. Second, for those remaining coefficients that are deviated from 0, the binary property is explored through the ℓ_∞ term. By the observation that solution of the ℓ_∞-norm minimization favors to achieve a solution with the components of the same magnitude, the ℓ_∞ term encourages those (deviated from 0) nonzero coefficients to be centered at 1 (because $|x_i - \frac{1}{2}|$ obtains the same value for $x_i = 0$ or 1). Nonetheless, the smoothed ℓ_0 gradient descent technique has demonstrated to outperform such approaches in terms of recovery rate and time [73].

Finally, we emphasize on two important points. First, that most results in sparse recovery are based on asymptotic properties of random matrices but sometimes an application specifies a deterministic measurement matrix. Thus, it is vital to have tractable schemes for testing properties like the NSP, RIP, and etc. in potential situations; see, e.g., [37, 115]. In fact, the nature of this requirement, encourages RIPless conditions [29, 141]. Second, all the introduced convex minimization problems can be solved by general purpose interior-point methods, however specified algorithms designed for these problems exist as well; see, e.g., [20, 32, 103, 137].

4   Greedy algorithms: main variations and results

Greedy algorithms are iterative approaches that take local optimal decisions in each step to eventually obtain a global solution. Hence, their success is due to some conditions on problem parameters. Greedy algorithms are mostly simple and fast, which find numerous applications in various contemporary fields, including biology, applied mathematics, engineering, and operations research. This supports the emerging interest in their performance analysis.

Greedy algorithms are efficient in tackling those problems in compressive sensing as well. Finding an s-sparse solution for an underdetermined linear system casts as the following sparsity constrained problem:

$$\min_{x \in \mathbb{R}^N} \ \|Ax - y\|_2^2 \quad \text{subject to} \quad \|x\|_0 \leq s. \tag{SC}$$

The most popular greedy algorithm to solve this problem is the Orthogonal Matching Pursuit (OMP). To tackle the problem SC, starting from zero as the initial iteration, the following OMP algorithm picks an appropriate index in each step to add its current support set and it estimates the new iteration as the orthogonal projection of the measurement vector onto the subspace generated by corresponding columns of the current support set [118, 123].

Orthogonal Matching Pursuit

Input:	$A \in \mathbb{R}^{m \times N}, y \in \mathbb{R}^m$, and initialize with $x_0 = 0 \in \mathbb{R}^N$, and $S_0 = \emptyset$.
Iteration:	repeat until convergence:
	$S_{n+1} = S_n \cup \{j_{n+1}\}$ s.t. $j_{n+1} = \arg\max_{j \in S_n^c} |\langle y - Ax_n, A_j \rangle|$,
	$x_{n+1} = \arg\min \|Ax - y\|_2^2$ subject to $\text{supp}(x) \subseteq S_{n+1}$.

There is an extensive literature on the capability of the OMP in identifying the exact support set either in at most s steps or at arbitrary many steps [21, 116, 118].

Theorem 11. *[77, 127] Assume that $A \in \mathbb{R}^{m \times N}$ satisfies the RIP of order $s + 1$ such that $\delta_{s+1} \leq \frac{1}{\sqrt{s+1}}$. Then, the OMP recovers any s-sparse vector $x \in \mathbb{R}^N$ using $y = Ax$ in at most s iterations.*

Further, it is proved that, when $s \geq 2$, for any δ such that $1/\sqrt{s+1} \leq \delta < 1$, any matrix with $\delta_{s+1} = \delta$ fails to recover all the s-sparse vectors [80, 125]. It is still an open question that in terms of the OMP performance what happens if $\delta_{s+1} \in (\frac{1}{\sqrt{s+1}}, \frac{1}{\sqrt{s+1}})$. A result with the same spirit of Theorem 11 for noisy measurements requires an extra assumption on the magnitude of the smallest nonzero entry of the desired vector [126]. Here, we bring the next theorem that assumes $\delta_s + 2\delta_{31s} \leq 1$. This inequality holds under the condition $\delta_{31s} \leq 1/3$, which implies the number of required measurements to successfully recover any s-sparse signal via the OMP is $\mathcal{O}(s \ln N)$, that is, the number of measurements m must be in the worst case linear in sparsity level s.

Theorem 12. *[142] Assume that $A \in \mathbb{R}^{m \times N}$ satisfies $\delta_s + 2\delta_{31s} \leq 1$. Then, given an s-sparse vector $x \in \mathbb{R}^N$, the OMP obeys*

$$\left\|x^{(30s)} - x\right\|_2 \leq \frac{2\sqrt{6}(1 + \delta_{31s})^{\frac{1}{2}}}{1 - \delta_{31s}} \|Ax - y\|_2.$$

There are also results for the success of this algorithm in case of availability of partial information on the optimal support set [57, 60]. This partial information is of the form of either a subset of the optimal support or an approximate subset with possibly wrong indices.

In many practical situations, there is no prior information available about the sparsity of a desired signal but generalized versions of the OMP are still enabled to recover it under a more stringent condition [38].

The OMP is also beneficial in recovering block sparse signals [47, 62, 66]. To elaborate on this block version of the OMP, assume that $x = [x[1]^T, x[2]^T, \ldots, x[L]^T]^T$ with $x[i] = [x_{d(i-1)+1}, x_{d(i-1)+2}, \ldots, x_{di}]^T, 1 \leq i \leq$

L (we assume the number of nonzero elements in different blocks are equal only for the sake of simplicity of exposition). Then, x is called a block s-sparse signal if there are at most s blocks (indices) i such that $x[i] = 0 \in \mathbb{R}^d$ [47]. The support of this block sparse vector is defined as $\Omega := \{i \mid x[i] \neq 0\}$. This definition reduces to the standard sparsity definition for $d = 1$. We also partition the measurement matrix A as $A = [A[1], A[2], \ldots, A[L]]$, where $A[i] = [A_{d(i-1)+1}, A_{d(i-1)+2}, \ldots, A_{di}]$, $1 \leq i \leq L$. Then, for a set of indices $S = \{i_1, i_2, \ldots, i_{|S|}\}$, we have $x[S] := [x[i_1]^T, x[i_2]^T, \ldots, x[i_{|S|}]^T]^T$, and $A[S] := [A[i_1], A[i_2], \ldots, A[i_{|S|}]]$. The block orthogonal matching pursuit (BOMP) is as follows.

Block Orthogonal Matching Pursuit

Input:	$A \in \mathbb{R}^{m \times Ld}, y \in \mathbb{R}^m$, and initialize with $x_0 = 0 \in \mathbb{R}^{Ld}$, and $S_0 = \emptyset$.
Iteration:	repeat until convergence:
	$S_{n+1} = S_n \cup \{j_{n+1}\}$ s.t. $j_{n+1} = \arg\max_{j \in [L]} \|A[j]^T(y - A[S_n]x[S_n])\|_2$,
	$x[S_{n+1}] = \arg\min_{x \in \mathbb{R}^{|S_{n+1}|d}} \|A[S_{n+1}]x - y\|_2^2$.

To present the exact recovery result of this algorithm, we need two other tools.

Definition 9. *[47]* Consider a block measurement matrix A introduced above. The block mutual coherence of A is defined as

$$\mu_B(A) := \max_{1 \leq l \neq r \leq L} \frac{1}{d} \|A[l]^T A[r]\|_2,$$

and its block subcoherence is defined as

$$\nu_B(A) := \max_{1 \leq l \leq L} \max_{1 \leq i \neq j \leq d} \frac{1}{d} |A_i[l]^T A_j[l]|,$$

where $A_i[l]$ and $A_j[l]$ are the ith and jth columns of the block $A[l]$.

Theorem 13. *[66] Let x be a block sparse vector supported on Ω and $y = Ax + e$ where $\|e\|_2 \leq \eta$ and the measurement matrix satisfies $(2s - 1)d\mu_B(A) + (d-1)\nu_B(A) < 1$. Further, suppose that*

$$\min_{i \in \Omega} \|x[i]\|_2 \geq \frac{\eta\sqrt{2(1 + (2d-1)\nu_B(A))}}{1 - (2|\Omega| - 1)d\mu_B(A) - (d-1)\nu_B(A)}.$$

Then, the BOMP algorithm recovers block sparse vector x in $|\Omega|$ iterations.

A similar result for the capability of another extended version of the OMP in recovering joint block sparse matrices is found in [55, 110]. Despite

many interesting features of the OMP, its index selection is problematic. Precisely, if a wrong index is chosen, the OMP fails to expel this index so that the exact support cannot be found in this situation. The Compressive Sampling Matching Pursuit (CoSaPM) is an algorithm to resolve this drawback [85]. To find an s-sparse feasible vector for an underdetermined linear system, this algorithm allows $2s$ best potential indices enter the current support set. Then, the following CoSaMP keeps s entries that play the key role in the pertaining projection in the sense that their corresponding entries have most magnitude.

<div align="center">

Compressive Sampling Matching Pursuit

</div>

Input:	$A \in \mathbb{R}^{m \times N}, y \in \mathbb{R}^m$, and initialize with $x_0 = 0 \in \mathbb{R}^N$, and $S_0 = \emptyset$.
Iteration:	repeat until convergence:
	$U_{n+1} = \operatorname{supp}(x_n) \cup L_{2s}(A^T(y - Ax_n))$,
	$u_{n+1} = \arg\min \|Ax - y\|_2^2 \quad \text{subject to} \quad \operatorname{supp}(x) \subseteq U_{n+1}$,
	$x_{n+1} = H_s(u_{n+1})$.

In the noiseless measurement case, the following theorem sates that any s-sparse signal, is recovered as the limit point of a sequence generated by the CoSaMP.

Theorem 14. *[53] Assume that $A \in \mathbb{R}^{m \times N}$ satisfies $\delta_{8s} < 0.4782$. Then, for any $x \in \mathbb{R}^N$ and $e \in \mathbb{R}^m$, the sequence x_n generated by CoSaMP using $y = Ax + e$, where s is replaced by $2s$, obeys*

$$\|x - x_n\|_2 \leq Cs^{-1/2}\sigma_s(x)_1 + D\|e\|_2 + 2\rho^n\|x\|_2,$$

where constants C, D and $\rho \in (0,1)$ only depend on δ_{8s}. In particular, if $\tilde{x}$ denotes a cluster point of this sequence, then

$$\|x - \tilde{x}\|_2 \leq Cs^{-1/2}\sigma_s(x)_1 + D\|e\|_2.$$

The following generalized orthogonal matching pursuit (gOMP) allows a given number of indices $t \geq 1$ enter a current support set. Consequently, a faster exact recovery under the RI constants involved conditions is attained [121, 129]. A recent algorithm combines the BOMP and the gOMP to propose a block generalized orthogonal matching pursuit for recovering block sparse vectors with possibly different number of nonzero elements in each block [95]. A closely related task to the problem SC is the following nonnegative sparsity constrained (NSC) problem

$$\min_{x \in \mathbb{R}^{\mathbb{N}}} \|Ax - y\|_2^2 \quad \text{subject to} \quad \|x\|_0 \leq s \quad \text{and} \quad x \geq 0. \qquad \text{(NSC)}$$

Generalized Orthogonal Matching Pursuit

Input:	$A \in \mathbb{R}^{m \times N}, y \in \mathbb{R}^m, t \in \mathbb{N}$, and initialize with $x_0 = 0 \in \mathbb{R}^N$, and $S_0 = \emptyset$.
Iteration:	repeat until convergence: $U_{n+1} = U_n \cup L_t\big(A^T(y - Ax_n)\big),$ $x_{n+1} = \arg\min \|Ax - y\|_2^2 \quad$ subject to $\quad \text{supp}(x) \subseteq U_{n+1}.$

Then, its solution may completely differ from the one to SC. Bruckstein et al. [18] presented an adapted version of the OMP for finding nonnegative sparse vectors of an underdetermined system, namely, the Nonnegative Orthogonal Matching Pursuit (NOMP) below. They demonstrated its capability to find sufficiently sparse vectors. In practice, it is probable not to have sparsity level in hand, this challenge is also doable [86].

Nonnegative Orthogonal Matching Pursuit

Input:	$A \in \mathbb{R}^{m \times N}, y \in \mathbb{R}^m$, and initialize with $x_0 = 0 \in \mathbb{R}^N$, and $S_0 = \emptyset$.
Iteration:	repeat until convergence: $S_{n+1} = S_n \cup \{j_{n+1}\} \quad$ s.t. $\quad j_{n+1} = \arg\max_{j \in S_n^c} \langle y - Ax_n, A_j \rangle_+,$ $x_{n+1} = \arg\min \|Ax - y\|_2^2 \quad$ subject to $\quad \text{supp}(x) \subseteq S_{n+1}$ and $x \geq 0.$

There is an emerging interest to explore novel greedy algorithms to find a sparsest feasible point of a set [7,12]. This casts as a minimization problem under sparsity constrained possibly intersecting a desired set. Nevertheless, the main idea of a greedy algorithm here is to start with a feasible sparse point (possibly zero) and add one or several candidate indices to the current support set. Once the support set is updated, a new iteration is obtained via a projection step. The Constrained Matching Pursuit (CMP) [109] investigates the following constrained sparse problem:

$$\min_{x \in \mathbb{R}^N} \|x\|_0 \quad \text{subject to} \quad Ax = y \quad \text{and} \quad x \in \mathcal{P},$$

where $\mathcal{P} \subseteq \mathbb{R}^N$ is a closed constraint set containing the origin. One can see that the OMP and NOMP can be employed for solving the above problem with the constraint set $\mathcal{P}$ with $\mathbb{R}^N$ and $\mathbb{R}_+^N$, respectively.

The performance of the CMP relies on the measurement matrix A as well as the involved constraint set $\mathcal{P}$. In fact, there exists a set $\mathcal{P}$ for which there is no measurement matrix A such that the CMP is successful [109]. Hence, a class of convex coordinate-projection (CP) admissible sets is of interest. A nonempty set $\mathcal{P} \subseteq \mathbb{R}^N$ is called CP admissible if for any $x \in \mathcal{P}$ and any index set $\mathcal{J} \subseteq \text{supp}(x)$, we have $x_{\mathcal{J}} \in \mathcal{P}$. The conic hull structure of the CP sets can be used to extend the RIP and the RO

Constrained Matching Pursuit

Input:	$A \in \mathbb{R}^{m \times N}, y \in \mathbb{R}^m, \mathcal{P} \subseteq \mathbb{R}^N$, and initialize with $x_0 = 0$, and $S_0 = \emptyset$.
Iteration:	repeat until convergence:
	$S_{n+1} = S_n \cup \{j_{n+1}\}$ s.t. $j_{n+1} = \arg\min_{j \in S_n^c} g_j^*$, where
	$g_j^* = \min_{t \in \mathbb{R}} \|y - A(x_n + t\,\mathbf{e}_j)\|_2^2$ subject to $x_n + t\,\mathbf{e}_j \in \mathcal{P}$,
	$x_{n+1} = \arg\min \|Ax - y\|_2^2$ subject to $\operatorname{supp}(x) \subseteq S_{n+1}$ and $x \in \mathcal{P}$.

constants over these sets. A verifiable exact recovery condition is then developed based on such constants for the closed convex CP admissible cones [109]. (Consequently, a verifiable condition for the NOMP is in hand because the nonnegative orthant is a closed convex CP admissible cone). These verifiable conditions are often NP-hard to check, like almost all the introduced tools in this survey such as the NSP, and RIP. But random measurement matrices with the right size hold these verifiable conditions with a high probability. This fact confirms the significance of random measurements in sparse optimization [114].

There are many other effective algorithms in sparse optimization that are applicable for the compressive sensing setting as well. For example, threshholding based algorithms use the adjacent matrix A^* for approximating inversion action and exploit the hard thresholding operator to solve the square system $A^*Ax = A^*y$ via a fixed-point method [15, 31, 39, 49, 51]. Iterative reweighted algorithms often outperform algorithms presented in Section 3. The iterative reweighted ℓ_p-algorithm is as follows:

$$x_{n+1} \in \arg\min_{x \in \mathbb{R}^N} \|W_n x\|_p \quad \text{subject to} \quad \|Ax - y\|_2 \le \epsilon,$$

where W_n is a diagonal weight matrix defined based on the current iteration x_n. For example, $W^k = \operatorname{diag}(w_n^1, w_n^2, \ldots, w_n^N)$ where $w_n^j = 1/(|x_n^j| + \gamma)^t$ with t and $\gamma > 0$, which encourages small elements to approach zero quickly [1, 30, 35]. The so called smoothed ℓ_0-norm takes advantage of smooth approximations of the ℓ_0 quasi-norm to tackle the CS and NCS [61, 79].

In traditional compressive sensing approaches, we recover sparse signals from $m = \mathcal{O}(s \log N/s)$ linear measurements. Model-based compressive sensing enables us to substantially reduce the number of measurements to $m = \mathcal{O}(s)$ without loosing the robustness of the recovery process [10]. In fact, model based compressive sensing includes the structural dependencies between the values and locations of the signal coefficients. For example, modeling the problem of binary compressive sensing with a bi-partite graph and representing the recovery process as an edge recovery scheme recovers binary signals more accurately, compared to binary ℓ_1-norm and tree-based binary compressive sensing methods [84].

On the other hand, quantum annealers are type of adiabatic quantum computers that tackle computationally intensive problems, which are intractable in the realm of classical computing. From a problem solving point of view, quantum annealers receive coefficients of an Ising Hamiltonian as input and return a solution that minimizes the given energy function in a fraction of a second [88]. Recent studies have revealed that well-posed binary compressive sensing and binary compressive sensing with matrix uncertainty problems are tractable in the realm of quantum computing [5,6].

5   Numerical performance analysis

In this section, we numerically examine the performance of the most prevalent sparse recovery methods introduced in this survey in recovering unigram text representation from their linear embedding (measurement). To describe the sparse recovery task of interest, we start by an introductory explanation on document (text) unigram representations and embeddings. Suppose that a vocabulary $\mathcal{V}$ is available, namely, a set of all the considered words in a context. Then, a unigram representation of a document is simply a vector in $\mathbb{R}^{|\mathcal{V}|}$, where its ith entry counts the occurrences of the ith word of $\mathcal{V}$ in that document [89, 91]. Unigram representations are highly sparse as the size of a vocabulary is often too large. These representations are expensive to store and further may not be able to effectively capture the semantic relations among the words of a vocabulary, so text embeddings are naturally favorable.

We start by defining word embeddings and discuss text embeddings afterward. The goal in word embedding is to encode each word of a vocabulary into a vector representation in a much lower dimensional space $m \ll |\mathcal{V}|$ such that the semantic relations among words are preserved. Word embeddings have lately gained much attentions in broad applications of the natural language processing such as classification, question answering and part of speech tagging. There are different linear and nonlinear effective approaches for word embedding, for example, pretrained models Global Vectors For Word Representation (GloVe) [92], word2vec [76] and Rademacher embeddings. The GloVe is a neural (network) based word embedding, where the word2vec is mainly based on matrix decomposition. Assuming that word embeddings are available for all the words of a vocabulary, a text embedding is simply a linear combination of them with the coefficients coming from its unigram representation. In other words, we have $Ax_T^{\text{unigram}} = y_T^{\text{embedding}}$, where $A \in \mathbb{R}^{m \times |\mathcal{V}|}$ is an embedding (or measurement matrix) generated from a specific methodology, $x_T^{\text{unigram}} \in \mathbb{R}^{|\mathcal{V}|}$ is

the unigram representation of a text, and $y_T^{\text{embedding}} \in \mathbb{R}^m$ is its (unigram) embedding. The measurement matrices in our experiments are generated via GloVe embedding method and the Radamacher distribution with the following probability mass function:

$$f(k) = \begin{cases} 1/2 & \text{if } k = -1, \\ 1/2 & \text{if } k = +1. \end{cases}$$

These matrices are: $m \times 17000$ for MR movie reviews [90] and $m \times 20000$ for SUBJ subjectivity dataset [89], respectively. Embedding sizes for our experiments are $m = 50, 100, 200, 300$ and 1600.
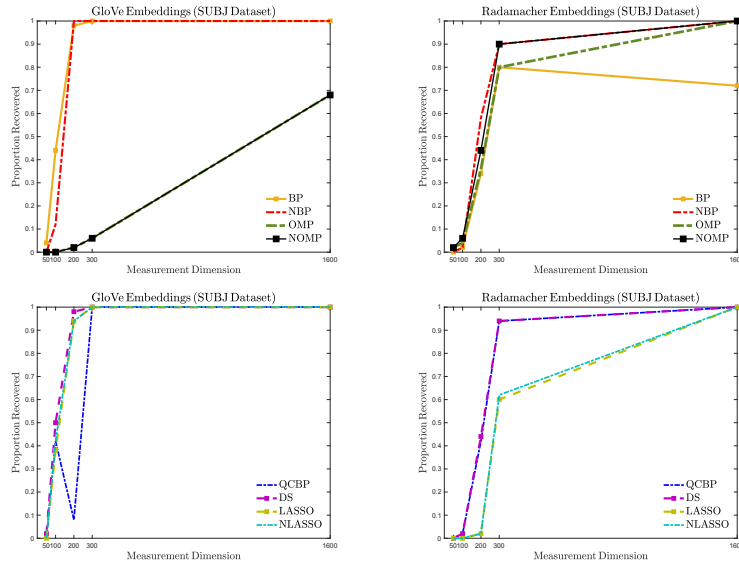

Figure 1: SUBJ dataset recovery.

These figures confirm the efficiency of ℓ_1 recovery and greedy algorithms in recovering unigram representations of 50 documents, where success is achieved if the relative error is smaller than 10^{-7}. We avoid a detailed explanation on the implemented algorithms for this specific application rather finish this survey with a more general conclusion as follows. The ℓ_p recovery with $0 < p \le 1$, where $p < 1$ promises better theoretical results than $p = 1$, and greedy algorithms are both effective in recovering sparse vectors in various applications if the measurement matrix inherits several properties explained in this survey. In the constrained case, the constraint set also plays a crucial role in the sparse recovery conditions as well [109]. Nevertheless, these properties often provide sufficient conditions and for
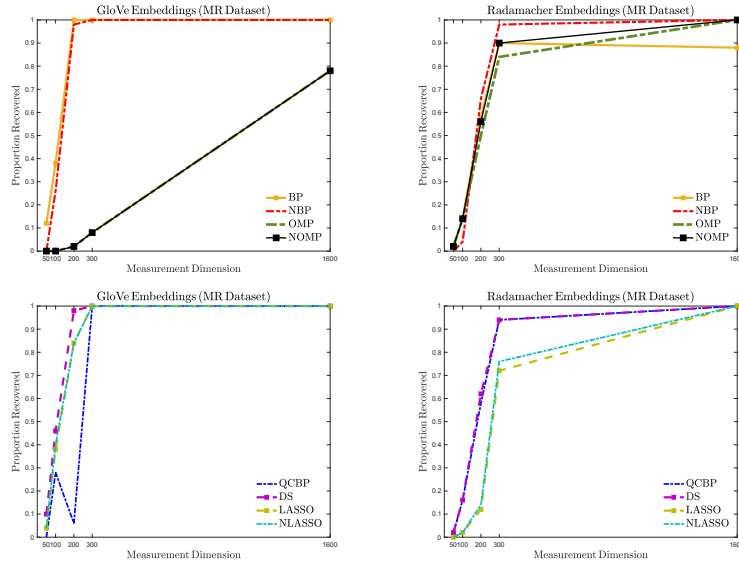
Figure 2: MR dataset recovery.

some applications specified measurement matrices can work properly as well and even better than random matrices [4]. Further, except when the sparsity level is relatively small, the ℓ_p recovery with $0 < p \le 1$ is superior to the greedy approach. For a more comprehensive study on the computational complexities and efficiency of these algorithms or related studies, see, e.g., [2, 52, 94, 143].

Acknowledgements

References

[1] M.J. Abdi, *Cardinality optimization problems*, University of Birmingham, 2013.

[2] L. Adhikari, *Nonconvex sparse recovery methods*, Uinversity of California, Merced, 2017.

[3] A. Amini and F. Marvasti, *Deterministic construction of binary, bipolar, and ternary compressed sensing matrices*, IEEE Trans. Inform. Theory. **57** (4) IEEE (2011) 2360–2370.

[4] S. Arora, M. Khodak, and N. Saunshi, and K. Vodrahalli, *A compressed sensing view of unsupervised text embeddings, bag-of-n-grams, and LSTMs*, 2018.

[5] R. Ayanzadeh, M. Halem and T. Finin, *SAT-based compressive sensing*, arXiv:1903.03650, 2019.

[6] R. Ayanzadeh, S. Mousavi, M. Halem and T. Finin, *Quantum annealing based binary compressive sensing with matrix uncertainty*, arXiv:1901.00088, 2019.

[7] S. Bahmani, B. Raj and P.T. Boufounos, *Greedy sparsity-constrained optimization*, J. Mach. Learn. Res. **14** (2013) 807–841.

[8] L. Bai, H. Lu and Y. Liu, *High-efficiency observations: compressive sensing and recovery of seismic waveform data*, Pure Appl. Geophys. **177** (1) (2020) 469–485.

[9] R.G. Baraniuk, *Compressive sensing, IEEE signal processing magazine*, **24** (2007) 118–121.

[10] R.G. Baraniuk, V. Cevher, M F. Duarte and C. Hegde, *Model-based compressive sensing*, arXiv:0808.3572, 2008.

[11] R.G. Baraniuk, M. Davenport, R. DeVore and M. Wakin, *A simple proof of the restricted isometry property for random matrices*, Constr. Approx. **28** (3) (2008) 253–263.

[12] A. Beck and N. Hallak, *On the minimization over sparse symmetric sets: projections, optimality conditions, and algorithms*, Math. Oper. Res. **41** (1) 196–223.

[13] A. Beck and Y.C. Eldar, *Sparse signal recovery from nonlinear measurements*, IEEE International Conference on Acoustics, Speech and Signal Processing, (2013) 5464–5468.

[14] A. Beck, and Y.C. Eldar, *Sparsity constrained nonlinear optimization: Optimality conditions and algorithms*, SIAM J. Optim. **23** (3) (2013) 1480–1509.

[15] T. Blumensath, and M.E. Davies, *Iterative hard thresholding for compressed sensing*, Appl. Comput. Harmon. Anal. **27** (3) (2009) 265–274.

[16] H. Boche, R. Calderbank, G. Kutyniok and J. Vybíral, *Compressed sensing and its applications: MATHEON workshop 2013*, Springer, 2015.

[17] P.T. Boufounos and R.G. Baraniuk, *1-bit compressive sensing*, IEEE CISS (2008) 16–21.

[18] A.M. Bruckstein, and M. Elad and M. Zibulevsky, *On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations*, IEEE Trans. Inform. Theory. **54** (11) (2008) 4813–4820.

[19] O.P. Burdakov, C. Kanzow and A. Schwartz, *Mathematical programs with cardinality constraints: reformulation by complementarity-type conditions and a regularization method*, SIAM J. Optim. **26** (1) (2016) 397–425.

[20] M. Burger, and M. Möller, M. Benning and S. Osher, *An adaptive inverse scale space method for compressed sensing*, Math. Comput. **82** (281) (2013) 269–299.

[21] T.T. Cai and L. Wang, *Orthogonal matching pursuit for sparse signal recovery with noise*, Institute of Electrical and Electronics Engineers, 2011.

[22] T.T. Cai, L. Wang and G. Xu, *Shifting inequality and recovery of sparse signals*, IEEE Trans. Signal Process **58** (3) (2009) 1300–1308.

[23] T.T. Cai, G. Xu and J. Zhang, *On recovery of sparse signals via ℓ_1 Minimization*, IEEE Trans. Inform. Theory **55** (7) (2009) 3388–3397.

[24] T. Cai and A. Zhang, *Sharp RIP bound for sparse signal and low-rank matrix recovery*, Appl. Comput. Harmon. Anal. **35**(1) (2013) 74–93.

[25] E.J. Candes, *Compressive sampling*, Proceedings of the International Congress of Mathematicians **3** (2006) 1433–1452.

[26] E.J. Candes and T. Tao, *The Dantzig selector: statistical estimation when p is much larger than n*, Ann. Statist **35** (2007) 2313–2351.

[27] E.J. Candes and T. Tao, *Decoding by linear programming*, arXiv:math/0502327, 2005.

[28] E. Candes and J. Romberg, *Sparsity and incoherence in compressive sampling*, Inverse Problems **23** (2007) 969–985.

[29] E.J. Candes and Y. Plan, *A probabilistic and RIPless theory of compressed sensing*, IEEE Trans. Inform. Theory **57** (11) (2011) 7235–7254.

[30] E. Candes, M.B. Wakin and S.P. Boyd, *Enhancing sparsity by Reweighted ℓ_1 minimization*, J. Fourier Anal. Appl. **14** (5) 877–905.

[31] R.E. Carrillo, L.F. Polania and K.E. Barner, *Iterative hard thresholding for compressed sensing with partially known support*, ICASSP, (2011) 4028–4031.

[32] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision **40** (1) (2011) 120–145.

[33] R. Chartrand, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Process. Lett. **14** (10) (2007) 707–710.

[34] R. Chartrand, *Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data,* 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, (2009) 262-265.

[35] R. Chartrand and W. Yin, *Iteratively reweighted algorithms for compressive sensing*, ICASSP, (2008) 3869–3872.

[36] S.S. Chen, D.L. Donoho and M.L. Saunders, *Atomic decomposition by basis pursuit*, SIAM Rev. **43** (1) (2001) 129–159.

[37] A. d'Aspremont and L. El Ghaoui, *Testing the nullspace property using semidefinite programming*, Springer Math. Program. **127** (1) (2011) 123–144.

[38] T.T. Do, L. Gan, N. Nguyen and T.D. Tran, *Sparsity adaptive matching pursuit algorithm for practical compressed sensing*, 42nd Asilomar Conference on Signals, Systems and Computers, (2008) 581–587.

[39] I. Daubechies, M. Defrise and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math. **57** (11) (2004) 1413–1457.

[40] D.L. Donoho and J. Tanner, *Precise undersampling theorems*, IEEE. **98** (6) (2010) 913–924.

[41] D.L. Donoho and J. Tanner, *Counting the faces of randomly-projected hypercubes and orthants, with applications*, Discrete Comput. Geom. **43** (3) (2010) 522–541.

[42] D.L. Donoho and J. Tanner, *Sparse nonnegative solution of underdetermined linear equations by linear programming*, Proc. Natl. Acad. Sci. **102** (27) (2005).

[43] D.L. Donoho, I. Drori, Y. Tsaig and J.L. Starck, *Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit*, IEEE Trans. Inform. Theory **58** (2) (2012) 1094-1121.

[44] D.L. Donoho, and M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization*, Proc. Natl. Acad. Sci. **100** (5) (2003) 2197–2202.

[45] M. Elad and A.M. Bruckstein, *A generalized uncertainty principle and sparse representation in pairs of bases*, IEEE Trans. Inform. Theory **48** (9) (2002) 2558-2567.

[46] Y.C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*, Cambridge Univ. Press, 2012.

[47] Y.C. Eldar, P. Kuppinger and H. Bolcskei, *Block-sparse signals: Uncertainty relations and efficient recovery*, IEEE Trans. Signal Process **58** (6) (2010) 3042–3054.

[48] M. Fazel, H. Hindi, S.P. Boyd and others, *A rank minimization heuristic with application to minimum order system approximation*, Proc. Am. Control Conf. **6** (2001) 4734–4739.

[49] M. Figueiredo and R.D. Nowak, *An EM algorithm for wavelet-based image restoration*, IEEE Trans. Image Process. **12** (8) (2003) 906–916.

[50] M. Figueiredo, R.D. Nowak and S.J. Wright, *Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems*, IEEE J-STSP **1** (4) (2007) 586–597.

[51] M. Fornasier and H. Rauhut, *Iterative thresholding algorithms*, Appl. Comput. Harmon. Anal. **25** (2) (2008) 187–208.

[52] M. Fornasier, *Numerical methods for sparse recovery*, pages 93–200, Radon Series on Computational and Applied Mathematics. De Gruyter, 2010.

[53] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, Springer, 2013.

[54] S. Foucart, *A note on guaranteed sparse recovery via ℓ_1-minimization*, Appl. Comput. Harmon. Anal. **29** (1) (2010) 97–103.

[55] Y. Fu, H. Li, Q. Zhang and J. Zou, *Block-sparse recovery via redundant block OMP*, Signal Process. **97** (2014) 162–171.

[56] J. Fuchs, *Spread representations,* Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers, IEEE. (2011) 814-817.

[57] H. Ge and W. Chen, *An Optimal Recovery Condition for Sparse Signals with Partial Support Information via OMP*, Circuits Systems Signal Process. (2019) 1–26.

[58] S. Gleichman and Y.C. Eldar *Blind compressed sensing*, IEEE Trans. Inform. Theory. **57** (10) (2011) 6958–6975.

[59] Z. Gu, and Z. Zhou, and Y. Yang, and A.R. Adhikary and X. Cai, *Deterministic compressed sensing matrices from sequences with optimal correlation*, IEEE Access. **7** (2019) 16704–16710.

[60] C. Herzet, C. Soussen, J. Idier and R. Gribonval, *Exact recovery conditions for sparse representations with partial support information*, IEEE Trans. Inform. Theory **59** (11) (2013) 7509–7524.

[61] H. Mohimani, M. Babaie-Zadeh and C. Jutten, *A fast approach for overcomplete sparse decomposition based on smoothed ℓ_0 norm*, IEEE Trans. Signal Process. **57** (1) (2008) 289–301.

[62] J. Huang, J. Wang, W. Wang and F. Zhang, *Sharp sufficient condition of block signal recovery via ℓ_2/ℓ_1-minimisation*, IEEE Trans. Signal Process. 2019.

[63] Y. Itoh, M.F. Duarte and M. Parente, *Perfect recovery conditions for non-negative sparse modeling*, IEEE Trans. Signal Process **65** (1) (2016) 69–80.

[64] M.J. Wainwright, *Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso)*, IEEE Trans. Inform. Theory. **55** (5) (2009) 2183–2202.

[65] L. Jacques, J.N. Laska, P.T. Boufounos and R.G. Baraniuk, *Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors*, IEEE Trans. Inform. Theory **59** (4) (2013) 2082–2102.

[66] Wen, Jinming and H. Chen and Z. Zhou, *An optimal condition for the block orthogonal matching pursuit algorithm*, IEEE Access **6** (2018) 38179–38185.

[67] S. Keiper, G. Kutyniok, D.G. Lee and G.E. Pfander, *Compressed sensing for finite-valued signals*, Linear Algebra Appl. **532** (2017) 570–613.

[68] M.A. Khajehnejad, A.G. Dimakis, W. Xu and B. Hassibi, *Sparse recovery of nonnegative signals with minimal expansion*, IEEE Trans. Signal Process. **59** (1) (2010) 196–208.

[69] G. Kutyniok, *Theory and applications of compressed sensing*, Wiley Online Library **36** (1) (2013), 79–101.

[70] S. Kwon, J. Wang and B. Shim, *Multipath matching pursuit*, IEEE Trans. Inform. Theory **60** (5) (2014) 2986–3001.

[71] J. Lange, M.E. Pfetsch, B.M. Seib and A.M. Tillmann, *Sparse recovery with integrality constraints*, arXiv:1608.08678, 2016.

[72] T.H. Lin and H.T. Kung, *Stable and efficient representation learning with nonnegativity constraints*, ICML, (2014) 1323–1331.

[73] T. Liu and D.G. Lee, *Binary compressive sensing via smoothed ℓ_0 gradient descent*, arXiv:1801.09937, 2018.

[74] A.M. Tillmann and M.E. Pfetsch, *The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing*, IEEE Trans. Inform. Theory. **60** (2) (2013) 1248–1259.

[75] E.C. Marques, N. Maciel, L. Naviner, H. Cai and J. Yang, *A review of sparse recovery algorithms*, IEEE Access. **7** (2018) 1300–1322.

[76] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado and J. Dean, *Distributed representations of words and phrases and their compositionality*, Adv. Neural Inf. Process Syst. (2013) 3111–3119.

[77] Q. Mo and Y. Shen, *A remark on the restricted isometry property in orthogonal matching pursuit*, IEEE Trans. Inform. Theory. **58** (6) (2012) 3654–3656.

[78] Q. Mo and S. Li, *New bounds on the restricted isometry constant δ_{2k}*, Appl. Comput. Harmon. Anal. **31** (3) (2011) 460–468.

[79] M.R. Mohammadi, E. Fatemizadeh and M.H. Mahoor, *Non-negative sparse decomposition based on constrained smoothed ℓ_0 norm*, Signal Process. **100** (2014) 42–50.

[80] Q. Mo, *A sharp restricted isometry constant bound of orthogonal matching pursuit*, arXiv:1501.01708, 2015.

[81] S. Mousavi and J. Shen, *Solution uniqueness of convex piecewise affine functions based optimization with applications to constrained ℓ_1 minimization*, ESAIM Control Optim. Calc. Var. **25** (2019) 56.

[82] S. Muthukrishnan, *Data streams: algorithms and applications*, Found. Trends Theor. Comput. Sci. **1** (2) (2005) 117–236.

[83] M. Nagahara, *Discrete signal reconstruction by sum of absolute values*, IEEE Signal Process. Lett. **22** (10) (2015) 1575–1579.

[84] U. Nakarmi and N. Rahnavard, *BCS: Compressive sensing for binary sparse signals*, MILCOM (2012) 1–5.

[85] D. Needell and J.A. Tropp, *CoSaMP: Iterative signal recovery from incomplete and inaccurate samples*, Appl. Comput. Harmon. Anal. **26** (3) (2009) 301–321.

[86] D. Needell and R. Vershynin, *Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit*, arXiv:0712.1360, 2007.

[87] T.L. Nguyen and Y. Shin, *Deterministic sensing matrices in compressive sensing: a survey*, Sci. World J., Volume 2013, Article ID 192795, 6 pages.

[88] H. Nishimori and K. Takada, *Exponential enhancement of the efficiency of quantum annealing by non-stoquastic hamiltonians*, Front. ICT. **4** (2017) 2.

[89] B. Pang and L. Lee, *A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts,* in Proceedings of the Association for Computational Linguistics (ACL), (2004) 271–278.

[90] B. Pang and L. Lee, *Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales*, in Proceedings of the Association for Computational Linguistics (ACL), (2005) 115–124.

[91] H.S. Paskov, R. West, J.C. Mitchell, T.J. Hastie, *Compressive feature learning*, In: Advances in Neural Information Processing Systems **26** (2013) 2931–2939.

[92] J. Pennington, R. Socher and C. Manning, *Glove: Global vectors for word representation*, EMNLP. (2014) 1532–1543.

[93] S. Pudlewski, A. Prasanna and T. Melodia, *Compressed-sensing-enabled video streaming for wireless multimedia sensor networks*, IEEE T. Mobile Comput. **6** (2012) 1060–1072.

[94] S. Qaisar, R.M. Bilal, W. Iqbal, M. Naureen and S. Lee, *Compressive sensing: From theory to applications, a survey*, J. Commun. Netw. S. Kor. **15** (5) (2013) 443–456.

[95] R. Qi, D. Yang, Y. Zhang and H. Li, *On recovery of block sparse signals via block generalized orthogonal matching pursuit*, Signal Process. **153** (2018) 34–46.

[96] Z. Qin, J. Fan, Y. Liu, Y. Gao and G. Li Ye, *Sparse representation for wireless communications: A compressive sensing approach*, IEEE Trans. Signal Process. **35** (3) (2018) 40–58.

[97] C. Ramirez, V. Kreinovich and M. Argaez, *Why ℓ_1 is a good approximation to ℓ_0: A geometric explanation*, J. Uncertain Syst. **7** (3) (2013) 203–207.

[98] M. Rani, S.B. Dhok and R.B. Deshmukh, *A systematic review of compressive sensing: Concepts, implementations and applications*, IEEE Access **6** (2018) 4875–4894.

[99] I. Rish and G.G. Grabarnik, *Sparse modeling: theory, algorithms, and applications* CRC press, 2014.

[100] D. Rontani, D. Choi, C.Y. Chang, A. Locquet and D.S. Citrin, *Compressive sensing with optical chaos*, Sci. Rep. **6** (2016) 35206.

[101] R. Saab, R. Wang and O. Yılmaz, *Quantization of compressive samples with stable and robust recovery*, Appl. Comput. Harmon. Anal. **44** (1) (2018) 123–143.

[102] R. Saab, R. Chartrand and Q. Yilmaz, *Stable sparse approximations via nonconvex optimization*, ICASSP (2008) 3885–3888.

[103] M. Salahi, J. Peng and T. Terlaky, *On Mehrotra-type predictor-corrector algorithms*, SIAM J. Optim. **18** (4) (2007) 1377–1397.

[104] A.C. Sankaranarayanan and R.G. Baraniuk, *Compressive Sensing*, Springer Computer Vision: A Reference Guide (2020) 1–5.

[105] J.A. Scales and A. Gersztenkorn, *Robust methods in inverse theory*, Inverse Probl. **4** (4) (1988) 1071.

[106] Y. Shechtman, A. Beck and Y C. Eldar, *GESPAR: Efficient phase retrieval of sparse signals*, IEEE Trans. Signal Process. **62** (4) (2014) 928–938.

[107] Y. Shechtman, Y C. Eldar, A. Szameit and M. Segev, *Sparsity based sub-wavelength imaging with partially incoherent light via quadratic compressed sensing*, J. Opt. Soc. Amer. A. **19** (16) (2011) 14807–14822.

[108] J. Shen and S. Mousavi, *Least sparsity of p-norm based optimization problems with $p > 1$*, SIAM J. Optim. **28** (3) (2018) 2721–2751.

[109] J. Shen and S. Mousavi, *Exact support and vector recovery of constrained sparse vectors via constrained matching pursuit*, arXiv:1903.07236, 2019.

[110] Y. Shi, L. Wang and R. Luo *Sparse Recovery With Block Multiple Measurement Vectors Algorithm*, IEEE Access **7** (2019) 9470–9475.

[111] C. Studer, W. Yin and R. Baraniuk, *Signal representations with minimum ℓ_∞-norm.* Conference on Communication, Control, and Computing, IEEE (2012) 1270-1277.

[112] G. Swirszcz, N. Abe and A.C. Lozano *Grouped orthogonal matching pursuit for variable selection and prediction*, Adv. Neural Inf. Process Syst. (2009) 1150–1158.

[113] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. R. Stat. Soc. Ser. B Stat. Methodol. **58** (1) (1996) 267–288.

[114] A.M. Tillmann and M.E. Pfetsch, *The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing*, IEEE Trans. Inform. Theory. **60** (2) (2013) 1248–1259.

[115] A.M. Tillmann, *Computing the spark: mixed-integer programming for the (vector) matroid girth problem*, Comput. Optim. Appl. **74** (2019) 387441.

[116] J.A. Tropp, *Greed is good: Algorithmic results for sparse approximation*, IEEE Trans. Inform. Theory **50** (10) (2004) 2231–2242.

[117] J.A. Tropp, A. Gilbert and M J. Strauss, *Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit*, Signal Process. **86** (3) (2006) 572–588.

[118] J.A. Tropp and A. Gilbert, *Signal recovery from random measurements via orthogonal matching pursuit*, IEEE Trans. Inform. Theory **53** (12) (2007) 4655–4666.

[119] Z. Xu, H. Guo, Y. Wang and H. Zhang, *Representative of $L_{1/2}$ regularization among $L_q(0 < q < 1)$ regularizations: an Experimental Study Based on Phase Diagram*, Acta Automat. Sinica **38** (7) (2012) 1225–1228.

[120] S. Wang and N. Rahnavard, *Binary compressive sensing via sum of l1-norm and l(infinity)-norm regularization*, MILCOM 2013-2013 IEEE Military Communications Conference, (2013) 1616–1621.

[121] J. Wang, S. Kwon and B. Shim, *Generalized orthogonal matching pursuit*, IEEE Trans. Signal Process. **60** (12) (2012) 6202–6216.

[122] Y. Wang, J. Yang, M. Liu and G. Gui, *LightAMC: Lightweight automatic modulation classification via deep learning and compressive sensing*, IEEE Trans. Veh. Technol. **69** (3) (2020) 3491–3495.

[123] J. Wang, *Support recovery with orthogonal matching pursuit in the presence of noise*, IEEE Trans. Signal Process. **63** (21) (2015) 5868–5877.

[124] Dai, Wei and Milenkovic, Olgica, *Subspace pursuit for compressive sensing signal reconstruction*, IEEE Trans. Inform. Theory **55** (5) (2009) 2230–2249.

[125] W. Dan, *A sharp RIP condition for orthogonal matching pursuit*, Abstr. Appl. Anal., Article ID 482357, 2013.

[126] J. Wen, Z. Zhou, J. Wang, Jian, X. Tang, and Q. Mo, *A sharp condition for exact support recovery with orthogonal matching pursuit*, IEEE Trans. Signal Process. **65** (6) (2016) 1370–1382.

[127] J. Wen, X. Zhu and D. Li, *Improved bounds on restricted isometry constant for orthogonal matching pursuit*, IET Electron. Lett. **49** (23) (2013) 1487–1489.

[128] Z. Wen, W. Yin, D. Goldfarb and Y. Zhang, *A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization, and continuation*, SIAM J. Sci. Comput. **32** (4) (2010) 1832–1857.

[129] J. Wen, Z. Zhou, D. Li and X. Tang, *A novel sufficient condition for generalized orthogonal matching pursuit*, IEEE Commun. Lett. **21** (4) (2016) 805–808.

[130] J. Wen, D. Li and F. Zhu, *Stable recovery of sparse signals via lp-minimization*, Appl. Comput. Harmon. Anal. **38** (1) (2015) 161–176.

[131] Z. Xu, X. Chang, F. Xu, and H. Zhang, $L_{1/2}$ *regularization: A thresholding representation theory and a fast solver*, IEEE Trans. Neur. Net. Learn. **23** (7) (2012) 1013–1027.

[132] C. Xu, and L. Jacques, *Quantized compressive sensing with rip matrices: The benefit of dithering*, arXiv:1801.05870, 2018.

[133] A. Szameit, Y. Shechtman, E. Osherovich, E. Bullkich, P. Sidorenko, H. Dana, S. Steiner, E. Kley, S. Gazit, T. Cohen-Hyams and others, *Sparsity-based single-shot subwavelength coherent diffractive imaging*, Nat. Mater. **11** (5) (2012) 455.

[134] M. Yang and F. de Hoog, *New coherence and RIP analysis for weak orthogonal matching pursuit*, IEEE Workshop on Statistical Signal Processing (SSP), (2014) 376–379.

[135] L. Yaroslavsky, *Is "Compressed Sensing" compressive? Can it beat the Nyquist Sampling Approach?*, arXiv:1501.01811, (2015).

[136] W. Yin, S. Osher, D. Goldfarb and J. Darbon, *Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing*, SIAM J. Imaging Sci. **1** (1) (2008) 143–168.

[137] X. Zhang, M. Burger and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput. **46** (1) (2011) 20–46.

[138] H. Zhang, W. Yin and L. Cheng, *Necessary and sufficient conditions of solution uniqueness in 1-norm minimization*, J. Optim. Theory Appl. **164** (1) (2015) 109–122.

[139] Z. Zhang, X. Liu, S. Wei, H. Gan, F. Liu, Y. Li, C. Liu and F. Liu, *Electrocardiogram reconstruction based on compressed sensing*, IEEE Access, 2019.

[140] P. Zhang, S. Wang, K. Guo and J. Wang, *A secure data collection scheme based on compressive sensing in wireless sensor networks*, Ad Hoc Netw. **70** (2018) 73–84.

[141] Y. Zhang, *Theory of compressive sensing via $\ell 1$-minimization: a non-RIP analysis and extensions*, J. Oper. Res. Soc. China **1** (1) (2013) 79–105.

[142] T. Zhang, *Sparse recovery with orthogonal matching pursuit under RIP*, IEEE Trans. Inform. Theory **57** (9) (2011) 6215–6221.

[143] Y. Zhang, *User's guide for yall1: Your algorithms for L1 optimization*, 2011

[144] Y. Zhao, *Sparse optimization theory and methods*, CRC Press, 2018.

[145] Y. Zhao, *Equivalence and strong equivalence between the sparsest and least ℓ_1-norm nonnegative solutions of linear systems and their applications*, J. Oper. Res. Soc. China. **2** (2) (2014) 171–193.

[146] Y. Zhao, *RSP-based analysis for sparsest and least ℓ_1-norm solutions to underdetermined linear systems*, IEEE Trans. Signal Process. **61** (22) (2013) 5777–5788.

[147] Z. Zhang, Y. Xu, J. Yang, X. Li and D. Zhang *A survey of sparse representation: algorithms and applications*, IEEE Access **3** (2015) 490–530.

[148] L. Zheng, and A. Maleki, H. Weng, X. Wang and T. Long, *Does ℓ_p-minimization outperform ℓ_1-minimization?*, IEEE Trans. on Inform. Theory **63** (11) (2017) 6896–6935.

[149] A. Zymnis, S. Boyd and E. Candes, *Compressed sensing with quantized measurements*, IEEE Signal Process. Lett. **17** (2) (2009) 149–152.