Binary Sparse Signal Recovery Algorithms Based On Logic Observation

Article *in* Pattern Recognition · June 2019

DOI: 10.1016/j.patcog.2019.01.018

5 authors, including:

Xiao Li Hu
22 PUBLICATIONS 330 CITATIONS

Jiajun Wen
Shenzhen University
33 PUBLICATIONS 392 CITATIONS

Zhihui Lai
Nanjing University of Science and Technology
109 PUBLICATIONS 1,654 CITATIONS

Linlin Shen
Shenzhen University
302 PUBLICATIONS 5,540 CITATIONS

Pattern Recognition 90 (2019) 147–160

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Binary sparse signal recovery algorithms based on logic observation

Xiao-Li Hu[a,c,1], Jiajun Wen[a,c,d,1,*], Zhihui Lai[a,c], Wai Keung Wong[d,e], Linlin Shen[a,b]

[a] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, PR China
[b] Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen 518060, PR China
[c] The National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen 518060, PR China
[d] Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hong Kong
[e] The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518055, PR China

A R T I C L E I N F O

Article history:
Received 28 February 2018
Revised 31 August 2018
Accepted 7 January 2019
Available online 22 January 2019

Keywords:
Binary sparse signal recovery
Logic observation
Matching pursuit method
Bayesian method

A B S T R A C T

Binary observation has been widely reported in the literature to localize or track moving objects due to its simple realization and good performance in improving energy efficiency. However, with the implementation of logic operators, the new observation models are out of the range of standard compressive sensing context, and thus lack of effective recovery algorithm. The purpose of this paper is to develop effective recovery algorithms and analyze their performance. Two kinds of recovery algorithms are developed and they are inspired from the matching pursuit method and Bayesian method, respectively. Theoretical conditions are also formulated to guarantee the successful recovery and the proposed algorithms are verified by a series of numerical experiments. Moreover, a construction method for the measurement matrix is also proposed, which is essential for model design. It is hoped that the proposed theories and algorithms can make contribution to the related applications of pattern recognition.

1. Introduction

Sparse representation has been widely used in pattern recognition [1–9], communication [10–13], compressive sensing (CS) [14–24], etc, due to that sparsity enables effective description and easy implementation of further relevant operations. Based on these efforts, the sparse theory has undergone a fast development in the recent years. In the field of pattern recognition, various advance sparse models [3–5] have been proposed to handle different recognition tasks. In order to enhance the discriminant ability of sparse model, the researchers in [6] added the discriminant information into the original model to greatly enhance the classification performance. Thus a unified version [7] of sparse representation was designed to reveal the underlying relationship of different classifiers. The work in [8] extended the classification task from single view to multi-view by making full use of the complementary properties lying in diverse views while maintaining the view consistency. To further understand the relations between different sparse feature selection methods, recent work [9] was the first time to give a deep insight of the relevant studies. In the recognition task, the sparse parameter vector to be estimated can be determined by different principles to enhance the differences between different classes. Thus, the sparse vector for feature selection is not uniquely prior determined. While, the sparse signal of CS to be estimated is the original unknown signal intended to be uniquely recovered. The accuracy of estimation is a central issue for CS. The estimation issue in CS is based on the observation equation, which is typically a system of linear observations comprising of observation, source sparse signal and measurement matrix. All these values were originally considered in the field of real numbers, e.g., [14–17], and then extended to the field of complex numbers, e.g., [12,18], and also partially to a set of discrete values [11,13,25,26]. The finite range of observation value (e.g., sensor) may improve observation accuracy and reduce relevant circuit disturbance in real application. Recent studies [27–31] show that binary-valued measurement model with linear/nonlinear logic operator can be implemented in real practice to improve energy efficiency. Such a model is all binary-valued for source signal, observation signal and measurement matrix, and thus it is different from the partially discrete-valued CS category [11,13,25,26], where only the observation value or the source sparse signal is discrete. In the last decade, the closely related hashing-based methods have been proposed to solve the image search problem. Literature [32] designed an improved and flexible supervised discrete hashing model to accurately measure the residual of the regression model. The work in [33] developed a fast and effective hashing strategy to map the training samples to the class labels. More related studies can

* Corresponding author at: College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, PR China.
E-mail address: enjoy_world@163.com (J. Wen).
[1] XiaoLi Hu and Jiajun Wen contributed equally to this work.

https://doi.org/10.1016/j.patcog.2019.01.018

a

1: 1000000	4: 0001000	7: 0000001
2: 0100000	5: 0000100	8: 0011111
3: 0010000	6: 0000010	9: 1100111

b

1: 0001	4: 0100	7: 0111
2: 0010	5: 0101	8: 1000
3: 0011	6: 0110	9: 1001

Fig. 1. (a) The space encoding for 2-sparse signal recovery with 7 sensors in 'exclusive or' model (1). (b) The space encoding for 1-sparse signal recovery with 4 sensors in 'or' model (2).

be found in [34]. However, the binary-valued measurement model is different from those hashing-based methods [32,33,34], where only the derived hash representation is binary-valued. The quantities in model equation $y = Ax$ considered in this paper are all (i.e., y, A, x) binary-valued. While the three above-mentioned hashing-based methods for classification are all partially binary-valued, i.e., only the derived hashing representation is binary-valued. As a matter of fact, the recovery algorithms for the all binary-valued models in [27–31] are relatively elementary by mechanically using the existing recovery algorithm from CS. Therefore, a crucial question is how to develop effective recovery algorithm for these new observation models with logic operators.

The binary sensor networks comprised of linear/nonlinear logic operators [27–31] have been recently proposed for target tracking and pattern recognition. One bit of information on targets is collected at each binary sensor indicating whether they are present or not in the sensor's field of view (FOV). The FOV of each binary sensor can be designed by various geometric shapes: disk, line segment, a group of lines, or grids [35]. The advantages of binary sensing include (i) improvement of energy efficiency, (ii) reduction of data throughput and communication overhead, and (iii) tolerance of low signal-to-noise ratios of low-cost sensors. A typical case of targets tracking in a binary sensing network [27] is a sparse signal recovery issue with binary-valued measurement matrix and binary-valued observation based on certain logic operations (not necessarily linear). The recovery algorithms of the above-mentioned literatures are essentially the implementation of standard CS recovery algorithms with corresponding adjustment to binary field by using round operator. Since the new model structure has not been taken into account, the recovery algorithms are generally ineffective. In Fig. 1, two simple examples are respectively presented for two typical logic observation models (1) and (2) respectively. The whole FOV of all sensors are separated by 9 locations, labeled from the natural number's set $\{1, 2, \dots, 9\}$. The corresponding 7 bit binary number in Fig. 1(a) can be viewed as a space encoding for these locations. The binary number in the ith position of these binary numbers stands for that this location is in the FOV of the ith sensor if the number is 1, otherwise the location is out of the detection range of the ith sensor. For instance, the FOV of the 1st sensor in (a) consists of locations labeled by {1, 9}, and the second sensor consists of locations labeled by {2, 9}, etc. The design of space encoding of case (a) in Fig. 1 is for 2 sparse signal recovery and the case (b) is for 1 sparse signal recovery, which will be further specified in Example 2.1.

Let us simply call binary sparse signal recovery (BSSR) to mention that the recovery issue using the above-mentioned logic models. Before introducing the main idea of the proposed recovery algorithms for BSSR, we briefly review the established algorithms for CS and a binary-valued matrix named after Low-Density Parity-Check (LDPC) Code.

The main established algorithms for sparse signal recovery are l_1-norm convex optimization method (e.g., the basis pursuit method [16]) and the greedy search approach, which identifies the support (index set of nonzero elements) of the sparse signal in an iterative fashion. Orthogonal matching pursuits (OMP) [36] is the most popular greedy algorithm, due to its easy implementation and high operation efficiency. However, such kind of algorithm becomes costly when the signal is not much sparse. Improved versions of OMP have been proposed like regularized OMP [37,38], stagewise OMP [39], gradient pursuits [40], compressive sampling matching pursuits (CoSaMP) [41], subspace pursuits [42], and orthogonal multiple matching pursuit [43]. All these matching pursuit algorithms are heuristic and generally require restrictive property of the measurement matrix. To integrate the advantage and idea of matching pursuit method into BSSR, particular design of the measurement matrix is quite necessary, which can simplify the optimization process. The Bayesian framework is an effective tool and has been used for standard CS in [44], and later to name a few [45–50] develop further the Bayesian idea to more topics in practice. These Bayesian methods are all considered for standard CS issues, i.e., the measurements are real-valued and all linearly generated. To integrate the advantage and idea of Bayesian method into BSSR, relevant preliminary calculations of posterior probability are key in the application, which further requires a particular structure of the measurement matrix.

A suitable matrix for BSSR is the Low-Density Parity-Check (LDPC) matrix, which is derived from the LDPC Code for channel coding introduced in [51]. Because the bit error rate decoding performance of LDPC approaches asymptotically the Shannon limit [52], LDPC code attracts much research and application attention. The relevant Parity-Check LDPC matrix is binary-valued with regular/irregular forms, which is typically characterized by the numbers of 1 located in each column and the maximal number of 1 for any two different columns with same row index. The LDPC matrix is generally sparse with most entries valued zero and a few valued 1. Thus it is convenient to roughly assume that a system of a few row equations are independent from each other, which can be used to simplify the calculation of relevant joint probability.

In this paper we merge the OMP idea and Bayesian idea with the logic binary model structure to develop effective recovery algorithms for relevant logic observation models. The LDPC matrix structure is used for the construction of measurement matrix, which essentially determines the model structure. It is analyzed that the algorithms inspired by OMP idea can guarantee exact recovery even for the noise case under suitable conditions. It is also shown by simulation that the proposed recovery algorithms outperform much than the state-of-the-art algorithms, such as the typical one in [27].

The contributions of this paper are summarized as follows:

- By integrating the idea of matching pursuit (MP) method into the two kinds of logic binary observation models, two relevant effective MP-type algorithms are developed and analyzed.
- By using the Bayesian idea in the two kinds of logic binary observation models, four relevant effective Bayesian type algorithms are developed corresponding to noiseless and noise cases.
- The measurement matrix is designed in the form of LDPC matrix to facilitate theoretical analysis and an effective construction algorithm is proposed.

The rest of the paper is organized as follows. Section 2 describes the framework of binary sparse signal recovery. Section 3 develops the proposed matching pursuit-based algorithms and Bayesian-based algorithms for binary sparse signal recovery issues. Section 4 presents theoretical analysis for the two matching pursuit algorithms and discussions for the proposed Bayesian-based algorithms. Numerical examples and verification are provided in Section 5. Section 6 concludes the paper and outlines future works.

2. Framework and basic notations

To address the binary sparse signal recovery problem, the basic models and the corresponding supporting theory will be introduced in this Section.

Let F_2 be a {0, 1}-valued finite field and F_2^n be the n-dimension linear space. A vector $v = (v_i) \in F_2^n$ is k-sparse if $|\text{supp}(v)| \le k$, where $\text{supp}(v) = \{i : v_i \neq 0\}$ and $|S|$ counts the number of elements in set S. Let $x \in F_2^n$ represent a k-sparse signal. Considering the logic 'exclusive or' observation model [27] as following

$$y = A \otimes x, \tag{1}$$

where matrix $A \in F_2^{m\times n}$ and $A \otimes x = Ax \pmod 2$. Here the equality $\xi = \eta \pmod 2$ under modulo operation means that the integer ξ and integer η's remainders after division by 2 are the same. BSSR concerns the recovery issue of a sparse signal x based on y and A under the case $k < m < n$.

Denote $I_{[\eta\ge 1]}$ as a componentwise saturation function for a vector η and 1 is a vector of same size of η with all 1 components. When a component $\eta_i \ge 1$, the resultant ith component of $I_{[\eta\ge 1]}$ is 1, and when $\eta_j < 1$, the jth component is 0. The logic 'or' observation model [27] can be represented in the form of multiplication with saturation as

$$y = A \odot x, \tag{2}$$

where $M \odot \xi = I_{[M\xi\ge 1]}$ for a matrix $M_{m\times n}$ and vector $\xi_{n\times 1}$.

Next we show two simple examples for models (1) and (2) respectively.

Example 2.1. (i). Let the measurement matrix be

$$A = \begin{bmatrix} 1&0&0&0&0&0&0&0&1\\ 0&1&0&0&0&0&0&0&1\\ 0&0&1&0&0&0&0&1&0\\ 0&0&0&1&0&0&0&1&0\\ 0&0&0&0&1&0&0&1&1\\ 0&0&0&0&0&1&0&1&1\\ 0&0&0&0&0&0&1&1&1 \end{bmatrix}, \tag{3}$$

which corresponds to the space encoding of case (a) in Fig. 1. The ith row of matrix A is corresponding to the ith digits listed in the ordered sub-locations in Fig. 1(a). Specifically, when the (i, j)th element of the above matrix is 1, it means that the jth location belongs to the FOV of the ith sensor. For instance, the FOV of the first sensor is the locations labeled by {1, 9}. This model is designed for 2 object-tracking in a region divided into 9 locations. Based on Theorem 2.1, any 2-sparse binary signal in terms of model (1) can be uniquely recovered, due to that the spark of the above matrix A is greater than 4.

(ii). Let the measurement matrix be

$$A = \begin{bmatrix} 0&0&0&0&0&0&0&1&1\\ 0&0&0&1&1&1&1&0&0\\ 0&1&1&0&0&1&1&0&0\\ 1&0&1&0&1&0&1&0&1 \end{bmatrix}, \tag{4}$$

which corresponds to the space encoding of case (b) in Fig. 1. The ith row of matrix A is also corresponding to the ith digits listed in the ordered sub-locations in Fig. 1(b). This model is designed for 1 object-localization in a region divided into 9 locations. In terms of model (2), any 1-sparse binary sparse signal can be uniquely recovered, due to that each column of the above matrix A represents different binary number.

However, noises generally exist in real measurement. For a simple start, we consider noise in the field of F_2, i.e., the value of noise is binary-valued. Thus the corresponding additive noise model can also be presented. Let v be binary valued noise. The noiseless model Eqs. (1) and (2) can be respectively extended to

$$y = (A \otimes x) \oplus v, \tag{5}$$

$$y = (A \odot x) \oplus v, \tag{6}$$

where $\oplus$ is module 2 summation. Let the number of nonzero components of v be n_v.

It is natural to find the k-sparse solutions of (1) and (2) respectively by solving combinatorial optimization problems:

$$(P_1) \quad \min_x \|x\|_0 \text{ s.t. } y = A \otimes x, \tag{7}$$

$$(P_2) \quad \min_x \|x\|_0 \text{ s.t. } y = A \odot x, \tag{8}$$

where $\|x\|_0$ denotes the number of non-zeros in x. The optimization (7) or (8) is actually conceptual framework to recover the source sparse signal, which is the starting point of the existing algorithms for the models (1) and (2) proposed by Lu and Co-authors [27–31]. However, the algorithms in these references are essentially the relevant standard CS recovery algorithms combined with rough replacement of logic operation by common multiplication. These raw algorithms are very ineffective and lack of analysis.

In this paper, we intend to develop recovery algorithms by using the subtle structure of the measurement matrix A and the corresponding logic operations, rather than from the optimization framework (7) or (8). Specifically, it is intended to develop effective heuristic algorithms both in matching pursuit-type and the Bayesian-type.

Next, we consider the uniqueness recovery issue of (P_1), which depends on suitable property of matrix A. The usual properties for guaranteeing the uniqueness of the corresponding l_1-norm optimization in the continuous real number space are spark, RIP, NSP and MC in the literature [24]. Note that the operator $\otimes$ defined above is a linear operator over F_2^n, the idea of spark in traditional CS theory can be applied for analyzing the issue (P_1). For the operator $\odot$, it is a nonlinear operator over F_2^n. Therefore, the analysis of (P_2) is relatively complicated.

Next we present the conditions for uniqueness of solution for optimizing (P_1) using the spark concept, which is effective for uniqueness of sparse solution with linear model structure.

2.1. Spark

The uniqueness of solution for (P_1) can be analyzed by the concept 'spark', which is introduced in [14] for the common CS theory using l_0-norm framework. Similarly, the notion 'spark' can be defined for binary-valued matrix $A \in F_2^{m\times n}$ as follows.

Definition 2.1. The spark of a matrix $A \in F_2^{m\times n}$ denoted by spark(A) is the minimal number of linearly dependent columns of A.

The difference between the spark of A and its traditional version in [14] is that the elements of matrix A are all binary value with {0, 1} and the relevant operations are now taking module 2 after usual implementation. The spark of a matrix A is a whole requirement of all columns of the matrix, and clearly $\text{spark}(A) \le 1 + \text{rank}(A)$, where rank($A$) is the rank of A. The notion spark has a natural connection with the matrix A's null space, which is denoted by $\mathcal{N}(A)$, i.e.,

$$\mathcal{N}(A) = \{z \in F_2^n : \ A \otimes z = 0\}. \tag{9}$$

The spark is equivalent to the minimum number of nonzero components of all nonzero vector in $\mathcal{N}(A)$.

A sufficient and necessary condition in terms of spark condition for the uniqueness of k-sparse solution of (P_1) is presented below, which is similar to Donoho and Co-authors [14,24] for the CS common case.

Theorem 2.1. *The following statements are equivalent for the optimization problem* (P_1)*:*

(i) *If a solution* x *of* (P_1) *satisfies* $\|x\|_0 \le k$*, then it is the unique solution.*
(ii) $\mathrm{spark}(A) > 2k$.

Proof. (i)→(ii). Suppose $\mathrm{spark}(A) \le 2k$, there exists a vector z satisfying $A \otimes z = 0$ and $\|z\|_0 \le 2k$. Thus, the vector z can be divided into two separate parts, i.e., $z = z^1 - z^2$ with $\|z^1\|_0 \le k$ and $\|z^2\|_0 \le k$. By linear property of operator ⊗, it follows that $A \otimes z^1 = A \otimes z^2$, which implies the contrary of statement (i). This completes the assertion (i)→(ii).

(ii)→(i). Suppose the contrary of (i), i.e., let x^1 and x^2 be two at most k-sparse solutions of (P_1). Based on the fact that $A \otimes x^1 = A \otimes x^2$, it follows that $A \otimes (x^1 - x^2) = 0$, which means that the vector $x^1 - x^2$ is in the null space of matrix A. Note that $x^1 - x^2$ is at most $2k$-sparse, which implies that $\mathrm{spark}(A) \le 2k$. The assertion (ii)→(i) follows. □

The condition (ii) tells us that the number of rows of the measurement matrix $A_{m\times n}$ should not be less than $2k$ for a k-sparse signal to be uniquely recovered by (P_1). Thus, for the design of matrix A for a k-sparse signal recovery, the matrix parameters should be $2k \le m < n$.

Next, a brief introduction of LDPC matrix is presented to facilitate the construction of effective measurement matrix.

2.2. LDPC matrix

The measurement matrix of models (1) or (2) is of key importance for algorithm construction and analysis. The LDPC matrix is sparse with binary values (most of the entries are 0) and particular structure, which is convenient for algorithm development and analysis. Next we briefly review the LDPC matrix and present basic concept ‘girth’ of LDPC matrix.

The Low-Density Parity-Check (LDPC) Code was introduced in [51] for channel encoding. Much research interest has been attracted for LDPC codes because its bit error rate decoding performance approaches asymptotically the Shannon limit [52]. A LDPC code is constructed by a generation matrix, which is called a LDPC matrix. To make use of the advantages of LDPC codes for binary sparse recovery problem, in this paper, we mainly consider the optimization problems (P_1) and (P_2) under a LDPC matrix.

The LDPC matrix can be divided into two classes: regular and irregular. The regular one means that the number of 1 in each column and each row are respectively the same, while the irregular one requires the number of 1 in each column to be the same. For simplicity, we specify the irregular LDPC matrix as follows.

Definition 2.2. A {0, 1}-valued matrix $A_{m\times n}$ is called an irregular LDPC matrix with parameter (n_c, n_s) for $0 < n_s < n_c < m < n$, if the number of 1 in each column of matrix A is n_c, and for any two columns the number of 1 with same row index is at most n_s.

The notion ‘girth’ is defined as follows.

Definition 2.3. A simple circle of {0, 1}-valued matrix $A_{m\times n}$ is a sequence of nonzero entries $\{a_{i_1,j_1}, a_{i_1,j_2}, a_{i_2,j_2}, a_{i_2,j_3}, \ldots, a_{i_r,j_c}, a_{i_r,j_1}\}$ located in r rows and c columns such that it is comprised of alternatively an edge connecting two entries in a same row and then an edge in a same column, the last entry is in the same column of the first entry. The matrix A is called with girth g, if the number of edges of the minimum circle of matrix A is g.

It is found that girth is important for probabilistic decoding methods for LDPC code. A relationship between girth and spark will be proposed in Section 4.

3. The proposed algorithms

The existing algorithms for the models (1) and (2) proposed by Lu and Co-authors [27–31] are essentially the relevant standard CS recovery algorithms combined with rough replacement (e.g., replacement of $y = A \odot x$ by $y = Ax$) or refined replacement (e.g., replacement of $y = A \otimes x$ by $y = [A - 2I][x\ z]'$ with z a suitable vector), and round operator to derive {0, 1}-valued solution. The relevant optimization issue with the replacement model equations can be typically solved by convex method (e.g., BP algorithm) or linear program method, etc. These raw algorithms are very ineffective and lack of analysis. In this section, we develop precise recovery algorithms by using the subtle structure of LDPC matrix and the ideas from matching pursuit method and Bayesian method respectively.

3.1. Matching pursuit (MP)-based recovery algorithms

By observing the binary structure of models (1) and (2), using the heuristic idea of Matching Pursuit (MP) method for standard CS, two MP algorithms for (1) and (2) are developed to recursively identify the support index set of the source binary sparse signal.

To facilitate the presentation of MP algorithms, denote Ω as the whole index set {1, 2, ⋯, n} of signal x. Introduce $\|x\|_0^+$ as the summation of its positive components, i.e.,

$$\|x\|_0^+ = \sum_{x_j>0} x_j. \tag{10}$$

In a simple example, let $\xi = [1, 0, 1, 0]^T$ and $\eta = [1, 0, 0, 1]^T$, we have $\|\xi - \eta\|_0^+ = 1$, which counts the half difference describing how many 1 elements of ξ has been erased by η. We use this half difference operator to construct the relevant matching pursuit-based algorithms for models (1) and (2). It is the heuristic scheme to select the best possible atom at each step by the minimum difference between the rest observation and all columns.

Owing to the linearity of the operator ⊗, the corresponding MP-based method for model (1) reserves the main style of traditional OMP methods except binary operators, as shown in Algorithm MP1.

By contrast, for the operator ⊙, it is inconvenient to separate the residual for each step as usual due to its nonlinearity. Thus, for the solution of model (2), we compare the difference between the observation and the variable selected atoms at each step. The variable atom sets at each step are constructed by the union of the fixed atom set determined by the former step and another variable atom changing from 1 to n. In other words, the former atom set is expanded by including another varying atom. Then the next optimal atom will be selected to be the minimal difference between the observation and resultant observation caused by the expanding index set. The detailed steps for the solution of model (2) are shown in Algorithm MP2.

Algorithms MP1 and MP2 can be effectively used to find the sparse binary solutions of the corresponding noise models (5) and (6) respectively without any modification if the noise is also sufficiently sparse.

The conditions for the uniqueness of Algorithms MP1 and MP2 for both noiseless and noise observation models will be analyzed in Section 4.

3.2. Bayesian-based recovery algorithms

The matching pursuit algorithms are heuristic method, which uses the sketchy properties of the measurement matrix and model

Algorithm MP1 for model (1)

Task: recover k-sparse signal x by observation $y = A \otimes x$ for all binary-valued signal, observation and matrix setting.

1. *Initialization*. Set residual $y_r^0 = y$, sparse signal $\hat{x}^0 = 0$, optimal atom set $T^0 = \emptyset$, iteration $j = 0$.
2. *Main iteration*. While $j < k$
   Set $j \longleftarrow j + 1$;
   Calculate $t_j = \arg\min_{i \in \Omega - T^{j-1}} \|y_r^{j-1} - A(:, i)\|_0^+$;
   Set jth index set $T^j = T^{j-1} \cup \{t_j\}$;
   Set jth estimate $\hat{x}^j_{T^j} = 1$;
   Calculate jth residual $y_r^j = y - A \otimes \hat{x}^j$.
3. *Output*. The recovered signal $\hat{x}^k$ and optimal atom set T^k.

Algorithm MP2 for model (2)

Task: recover k-sparse signal x by observation $y = A \odot x$ for all binary-valued signal, observation and matrix setting.

1. *Initialization*. Set sparse signal $\hat{x}^0 = 0$, optimal atom set $T^0 = \emptyset$, iteration $j = 0$.
2. *Main iteration*. While $j < k$
   Set $j \longleftarrow j + 1$, $T_i = \emptyset$;
   For $i = 1 : n$, set index set $T_i = T^{j-1} \cup \{i\}$, let $x_{T_i} = 1$, calculate $y_0^i = \|y - A \odot x_{T_i}\|_0^+$;
   Set $t_j = \arg\min_i y_0^i$;
   Set jth index set $T^j = T^{j-1} \cup \{t_j\}$;
   Set jth estimate $\hat{x}^j_{T^j} = 1$.
3. *Output*. The recovered signal $\hat{x}^k$ and optimal atom set T^k.

structural information. Hence, more refined methods using much observation information may be developed. Recent advances show that the Bayesian framework can discover the majority of information under the model structure. Moreover, the sparse structure (most are zero-valued and few are one-valued) of the sparse LDPC matrix also contributes to the recovery problem, such a structure is convenient to treat a system of a few row equations as independent random variables, which largely simplifies the calculation of relevant joint probabilities comprising of a few row equations. Let us call 'approximate independence property of LDPC matrix' for later reference to mention that a few row equations derived from LDPC measurement matrix can be viewed as a set of independent random variables. This structural property supports the calculation of joint probability by a product:

$$P[y_{i_1} = b_1, \ldots, y_{i_t} = b_t] = P[y_{i_1} = b_1] \cdots P[y_{i_t} = b_t], \tag{11}$$

where t is a small finite integer, the sub-index $i_j \in \{1, 2, \ldots, n\}$ and b_i is a binary value. This approximate independence property will be frequently used for the calculation of relevant joint probability in the following Bayesian recovery algorithms. However, this property is approximately true, due to that the system of observations may not be independent under certain cases. This is a reason why the proposed Bayesian algorithms based on LDPC measurement matrix are not easy to be theoretically analyzed.

Next, we briefly review the famous Bayesian framework, which converts a prior probability into a posterior probability by incorporating the evidence derived by the observation data. Based on the given prior distribution, the likelihood function is deduced to express how probable the observation data set is for different settings of the relevant parameters. In the Bayesian framework, the posterior and prior probability are connected by

$$\text{posterior} \propto \text{prior} \times \text{likelihood}, \tag{12}$$

where $\propto$ means a proportional relation between the two sides. For a discrete distributed random variable, whose distribution can be calculated by the whole proportional values for each possible value of the random parameter. Typically, for a {0, 1}-valued random variable θ and observation data Y, once the posterior ration is calculated to be $r = P[\theta = 0|Y]/P[\theta = 1|Y]$, the posterior distribution can be derived as $P[\theta = 0|Y] = r/(r+1)$ and $P[\theta = 1|Y] = 1/(r+1)$.

The feasibility of Bayesian method mainly depends on the easy calculation of likelihood function and the resultant posterior probability. Owing to the sparse structure of the LDPC matrix, the likelihood function (probability of observation) can be approximately calculated by assuming that all the row vectors of the measurement matrix are independent.

3.2.1. Method for model (1)

To calculate the observation probability $P[y_j = 0]$ and $P[y_j = 1]$ in model (1), recalling the linear operation in model (1) under module 2, i.e.,

$$y = x_1 + x_2 + \cdots + x_s \pmod 2, \tag{13}$$

which implies that the observation value $y = 0$ if there are even number of 1 occurred in $\{x_i, i = 1, \ldots, s\}$ and $y = 1$ if there are odd number of 1 occurred in $\{x_i, i = 1, \ldots, s\}$. Let us cite a lemma from Gallager [51] as follows, which plays a key role in the calculation of likelihood function for model (1).

Lemma 3.1. *Considering a sequence of s independent binary digits in which the lth digit is 1 with probability p_l, then the probability that an even number of digit are 1 is*

$$p_e = \frac{1}{2}\left[1 + \prod_{i=1}^{s}(1 - 2p_l)\right]. \tag{14}$$

Based on this lemma, it is clear in Eq. (13) that $P[y = 0] = p_e$ and $P[y = 1] = 1 - p_e$. These facts are used in the following Bayesian-based algorithm for model (1).

The so-called sum-product decoding idea for LDPC code in [51] is first to calculate the posterior probability for each row equation, and then based on the independent assumption for a few row equations to derive the combined (product) posterior probability for the whole observation vector. Finally, the k support indices of x are selected by choosing the k largest posterior probability $P[x_i = 1|y]$ for $i = 1, \dots, n$. Next, the detailed Bayesian-based algorithm for binary sparse signal recovery of model (1) is listed.

(i) First, setting prior distribution of x by a row vector

$$p = [p_1, p_2, \dots, p_n] \tag{15}$$

to indicate that the probability of the jth component of x being 1 is p_j, i.e., $P[x_j = 1] = p_j$. With the prior knowledge of k-sparsity of source signal x, one can set k components of p with values in (0.5,1) and others in (0,0.5). While, it is found in simulation that the equivalent setting of prior distribution by letting $p = 0.1 * \text{ones}(1, n)$ is a good choice.

Let L_i denote the set of column subsripts of nonzero entries of matrix A in the ith row, $i = 1, \dots, m$, and M_j the set of row subsripts of nonzero entries of matrix A in the jth column, $j = 1, \dots, n$. Denote $L_i^0 = L_i - \{j\}$ and $P_{i,j} = \prod_{l \in L_i^0} (1 - 2p_l)$.

(ii). For $j = 1 : n$, for any $i \in M_j$, based on Lemma 3.1, the (i, j)th likelihood ratio is calculated by

$$c_{i,j} = \frac{P[y_i | x_j = 0]}{P[y_i | x_j = 1]} = \begin{cases} \frac{1+P_{i,j}}{1-P_{i,j}}, & \text{if } y_i = 0, \\ \frac{1-P_{i,j}}{1+P_{i,j}}, & \text{if } y_i = 1. \end{cases} \tag{16}$$

Then, based on the 'approximate independence property' of LDPC matrix (assumption that a few row equations are independent), the jth column likelihood ratio is approximated by a product of ratios from the relevant row equation

$$c_j = \frac{P[y_i, i \in M_j | x_j = 0]}{P[y_i, i \in M_j | x_j = 1]} = \prod_{i \in M_j} c_{i,j} \tag{17}$$

for $j = 1, \dots, n$.

(iii). The posterior ratio of the jth component of signal x is

$$\bar{c}_j = \frac{P[x_j = 0 | y_i, i \in M_j]}{P[x_j = 1 | y_i, i \in M_j]} = \frac{P[x_j = 0] P[y_i, i \in M_j | x_j = 0]}{P[x_j = 1] P[y_i, i \in M_j | x_j = 1]} = \frac{(1 - p_j) c_j}{p_j}, \tag{18}$$

which indicates the post-probability

$$\bar{p}_j = P[x_j = 1 | y_i, i \in M_j] = \frac{p_j}{p_j + (1 - p_j) c_j} \tag{19}$$

for $j = 1, \dots, n$. This is an updated probability distribution of each components of the signal x. The process can be iteratively updated until certain criterion is fulfilled. Finally, it is natural to select the k largest components of the posterior distributions as the k support components of x.

We outline the whole steps of Bayesian algorithm for model (1) in Algorithm B1.

3.2.2. Method for model (2)

The Bayesian method for model (2) is mainly the same ratio calculation of $c_{i,j}$, which is calculated based on the following Lemma 3.2, due to the different nonlinear operator $\odot$. For a single equation of model (2), it is equivalent to be written as

$$y = I_{[x_1 + x_2 + \cdots + x_s \geq 1]}, \tag{20}$$

where $I_{[\xi \geq 1]}$ is the indicating function of set $[\xi \geq 1]$. Then, the observation value $y = 0$ occurs if all $\{x_i, i = 1, \dots, s\}$ are zeros and $y = 1$ occurs if there are some 1 in $\{x_i, i = 1, \dots, s\}$. Therefore, we can derive the following probabilities of events $[y = 0]$ and $[y = 1]$, which are used for the calculation of likelihood function for model (2).

Lemma 3.2. *Considering a sequence of s independent binary digits in which the lth digit is 1 with probability p_l, then the probabilities of the events $[y = 0]$ and $[y = 1]$ given by (20) are respectively*

$$P[y = 0] = \prod_{j=1}^{s} (1 - p_j), \tag{21}$$

$$P[y = 1] = 1 - \prod_{j=1}^{s} (1 - p_j). \tag{22}$$

Denote $Q_{i,j} = \prod_{l \in L_i^0} (1 - p_l)$. Based on this lemma, the calculation of $c_{i,j}$ given by (16) is separated by

$$c_{i,j}^0 = P[y_i | x_j = 0] = Q_{i,j}, \tag{23a}$$

$$c_{i,j}^1 = P[y_i | x_j = 1] = 0 \tag{23b}$$

for $y_i = 0$, and

$$c_{i,j}^0 = P[y_i | x_j = 0] = 1 - Q_{i,j}, \tag{24a}$$

$$c_{i,j}^1 = P[y_i | x_j = 1] = 1 \tag{24b}$$

for $y_i = 1$. Then, based on the 'approximate independence property' of LDPC matrix, we approximate the jointed condition probability as

$$c_j^0 = \prod_{i \in M_j} c_{i,j}^0, \tag{25}$$

$$c_j^1 = \prod_{i \in M_j} c_{i,j}^1 \tag{26}$$

The post-probability of x is described by

$$\bar{p}_j = P[x_j = 1 | y_i, i \in M_j] = \frac{p_j c_j^1}{p_j c_j^1 + (1 - p_j) c_j^0} \tag{27}$$

for $j = 1, \dots, n$.

The brief steps of Bayesian algorithm for model (2) are shown in Algorithm B2.

3.2.3. Method for noise model (5)

With the noise observation by model Eq. (5), Algorithm 3 is not directly applicable to this generalized situation. The corresponding Bayesian-based method for model (5) should extend the random vector to include the noise part to achieve accurate posterior estimation.

First let $p_i = P[x_i = 1]$, $i = 1, \dots, n$, and $p_{n+j} = P[v_j = 1]$, $j = 1, \dots, m$. The vector $p = [p_1, p_2, \dots, p_n, p_{n+1}, \dots, p_{n+m}]$ describes the prior distribution of signal x and noise v, since the signal x and noise v are binary-valued. For simplicity, set p_i with equivalent small values, i.e., $p_1 = \cdots = p_{n+m} = p_0$, where $p_0 \in (0, 1)$ is a small positive real number.

For $j = 1 : n$, for any $i \in M_j$, by considering the augmented noise component, the (i, j)th likelihood ratio is calculated by

$$c_{i,j} = \frac{P[y_i | x_j = 0]}{P[y_i | x_j = 1]} = \begin{cases} \frac{(1+P_{i,j})(1-p_{n+i}) + (1-P_{i,j}) p_{n+i}}{(1+P_{i,j}) p_{n+i} + (1-P_{i,j})(1-p_{n+i})}, & \text{if } y_i = 0, \\ \frac{(1+P_{i,j}) p_{n+i} + (1-P_{i,j})(1-p_{n+i})}{(1+P_{i,j})(1-p_{n+i}) + (1-P_{i,j}) p_{n+i}}, & \text{if } y_i = 1. \end{cases} \tag{28}$$

Algorithm B1 for model (1)
Task: recover k-sparse signal x by observation $y = A \otimes x$ for all binary-valued signal, observation and matrix setting.

1. *Initialization*. Set prior distribution $p = [p_1, p_2, \dots, p_n]$ of signal x with equivalent small values. SetNum=10 and $t = 0$.
2. *Main iteration*. While $t < \text{Num}$
   $t \leftarrow t + 1$.
   for $j = 1 : n$
   for $i \in M_j$, calculate $c_{i,j}$ by (16); endfor
   Calculate c_j by (17);
   endfor
   Calculate the post-probability $\bar{p}_j$ by (19).
   Replace p by the posterior probability $[\bar{p}_1, \dots, \bar{p}_n]$.
3. *Output*. Denote the index set T as the first k large components of the posterior probabilities and let $x_T = 1$ and $x_{T^c} = 0$.

Algorithm B2 for model (2)
Task: recover k-sparse signal x by observation $y = A \odot x$ for all binary-valued signal, observation and matrix setting.

1. *Initialization*. Set prior distribution $p = [p_1, p_2, \dots, p_n]$ of signal x with equivalent small values. SetNum=10 and $t = 0$.
2. *Main iteration*. While $t < \text{Num}$
   $t \leftarrow t + 1$.
   for $j = 1 : n$
   for $i \in M_j$, calculate $c_{i,j}^0$ and $c_{i,j}^1$ by (23) and (24) respectively; endfor
   Calculate c_j^0 and c_j^1 by (25) and (26) respectively;
   endfor
   Calculate the post-probability $\bar{p}_j$ by (27).
   Replace p by the posterior probability $[\bar{p}_1, \dots, \bar{p}_n]$.
3. *Output*. Denote the index set T as the first k large probabilities of the vector p. Then let $x_T = 1$ and $x_{T^c} = 0$.

Then, the jth column likelihood ratio c_j is calculated by (17) for $j = 1, \dots, n$, under the assumption that there are nearly no common nonzero entries in any two rows. Similarly, the post-probability $\bar{c}_j$ is calculated by (19) for $j = 1, \dots, n$. This is the updated probability distribution of each components of the signal x.

The calculation of posterior probability for noise part is simple, since the probability is now determined by x part. Denote $\bar{P}_i = \prod_{s \in L_i}(1 - 2p_s)$. For $i = 1 : m$,

$$c_{n+i} = \frac{P[y_i | v_i = 0]}{P[y_i | v_i = 1]} = \begin{cases} (1 + \bar{P}_i)/(1 - \bar{P}_i), & \text{if } y_i = 0, \\ (1 - \bar{P}_i)/(1 + \bar{P}_i), & \text{if } y_i = 1. \end{cases} \tag{29}$$

Then, the post-probability of noise is calculated by

$$\bar{p}_{n+i} = P[v_i = 1 | y_i] = \frac{p_{n+i}}{p_{n+i} + (1 - p_{n+i})c_{n+i}} \tag{30}$$

for $i = 1, \dots, m$.

The process can be iteratively updated until certain criterion is fulfilled.

We outline the whole steps in Algorithm B3.

3.2.4. Method for noise model (6)

Similarly, Algorithm B2 is not directly applicable to noise observation by model Eq. (6). The corresponding Bayesian method for model (6) should also consider the noise affection for obtaining accurate posterior estimation.

First let vector $p = [p_1, p_2, \dots, p_n, p_{n+1}, \dots, p_{n+m}]$ to describe the prior distribution of signal x and noise v as in Section 3.2.3 and set each $p_i \in (0, 1)$ with equivalent small values for simplicity.

For $j = 1 : n$, for any $i \in M_j$, by considering the augmented noise component, the calculation of $c_{i,j}^0$ and $c_{i,j}^1$ based on (23) and (24) should be replaced by

$$c_{i,j}^0 = P[y_i | x_j = 0] = Q_{i,j}(1 - p_{n+i}) + (1 - Q_{i,j})p_{n+i}, \tag{31a}$$

$$c_{i,j}^1 = P[y_i | x_j = 1] = p_{n+i} \tag{31b}$$

for $y_i = 0$, and

$$c_{i,j}^0 = P[y_i | x_j = 0] = Q_{i,j}p_{n+i} + (1 - Q_{i,j})(1 - p_{n+i}), \tag{32a}$$

$$c_{i,j}^1 = P[y_i | x_j = 1] = 1 - p_{n+i} \tag{32b}$$

for $y_i = 1$. Then, we calculate c_j^0 and c_j^1 by (25). and the post-probability $\bar{p}_j$ of x is calculated by (27) for $j = 1, \dots, n$.

The calculation of posterior probability for noise part is simple, since the probability is now determined by x part. Denote $\bar{Q}_i = \prod_{s \in L_i}(1 - p_s)$. For $i = 1 : m$,

$$c_{n+i}^0 = \bar{Q}_i, \ c_{n+i}^1 = 1 - \bar{Q}_i \text{ if } y_i = 0; \tag{33}$$

$$c_{n+i}^0 = 1 - \bar{Q}_i, \ c_{n+i}^1 = \bar{Q}_i \text{ if } y_i = 1. \tag{34}$$

Then, the post-probability of noise is calculated by (27) for $j = n + 1, \dots, n + m$.

The process can be iteratively updated until the stopping criterion is fulfilled.

We outline the whole steps in Algorithm B4.

Algorithm B3 for model (5)

Task: recover k-sparse signal x by observation $y = A \otimes x \oplus v$ for all binary-valued signal, observation and matrix setting.

1. *Initialization*. Set prior distribution $p = [p_1, p_2, \dots, p_n, p_{n+1}, \dots, p_{n+m}]$ of signal x and noise v with equivalent small values. SetNum=10 and $t = 0$.
2. *Main iteration*. While $t < \text{Num}$
   $t \leftarrow t + 1$.
   for $j = 1 : n$
   for $i \in M_j$, calculate $c_{i,j}$ by (28); endfor
   Calculate c_j by (17);
   endfor
   Calculate the post-probability $\bar{p}_j$ by (16) for $j = 1, \dots, n$, and by (29)(30) for $j = n+1, \dots, n+m$.
   Replace p by the posterior probability $[\bar{p}_1, \dots, \bar{p}_{n+m}]$.
3. *Output*. Denote the index set T as the first k large components of the posterior probabilities and let $x_T = 1$ and $x_{T^c} = 0$.

Algorithm B4 for model (6)

Task: recover k-sparse signal x by observation $y = A \odot x \oplus v$ for all binary-valued signal, observation and matrix setting.

1. *Initialization*. Set prior distribution $p = [p_1, p_2, \dots, p_n, p_{n+1}, \dots, p_{n+m}]$ of signal x and noise v with equivalent small values. SetNum=10 and $t = 0$.
2. *Main iteration*. While $t < \text{Num}$
   $t \leftarrow t + 1$.
   for $j = 1 : n$
   for $i \in M_j$, calculate $c^0_{i,j}$ and $c^1_{i,j}$ by (31) and (32) respectively; endfor
   Calculate c^0_j and c^1_j by (25);
   endfor
   Calculate the post-probability $\bar{p}_j$ by (27) for $j = 1, \dots, n$ and by (33)(34)(27) for $j = n+1, \dots, n+m$.
   Replace p by the posterior probability $[\bar{p}_1, \dots, \bar{p}_{n+m}]$.
3. *Output*. Denote the index set T as the first k large probabilities of the vector p. Then let $x_T = 1$ and $x_{T^c} = 0$.

4. Analysis and discussions of algorithms

We investigate here the performance of Matching Pursuit-based algorithms MP1, MP2 and Bayesian-based algorithms B1-B4 theoretically. Some sufficient conditions for successful recovery of the Algorithms MP1 and MP2 for noiseless and noise case are developed. These conditions can be described by LDPC matrix parameters and the sparsity of the source signal. We provide discussion on the performance of the Bayesian-based algorithms, since the ‘approximate independence property’ of LDPC matrix does not always in accord with the truth. Moreover, a relationship between concepts spark and girth is analyzed and a random construction algorithm is presented for generating the LDPC matrix.

4.1. Analysis of algorithms MP1 and MP2

The proposed Algorithms MP1 and MP2 for model (1) and model (2) respectively utilize the structural properties and the corresponding logic operation, and thus the conditions for successful recovery of these algorithms can be described by the relation between the LDPC matrix parameters n_c, n_s and sparsity k. Next we develop the conditions for the uniqueness of solution of Algorithms MP1 and MP2 under both noiseless and noise cases.

First, for the noiseless case, the uniqueness condition can be presented as the following theorem for model (1).

Theorem 4.1. *For an irregular LDPC matrix $A_{m\times n}$ with parameter (n_c, n_s), the solution by Algorithm MP1 of model (1) for k-sparse signal x is unique if $n_c > (2k-1)n_s$.*

Proof. First for the initial step, consider the difference between the observation y and each column of matrix A. Denote the support atom set of source signal x by $S = \text{supp}(x)$. We show that the first atom selected in Algorithm MP1 belongs to the support set S. For an index $i \in S$, note that there are still $(k-1)$ columns with support indices, the column $A(:, i)$ has at most $(k-1)n_s$ number of 1 disturbed by other columns by the definition of LDPC matrix. Thus, by the module multiplication of model Eq. (1), the ith column $A(:, i)$ have contributed at least $n_c - (k-1)n_s$ number of 1 to the observation y, which means that $\|y - A(:, i)\|_0^+$ decreases at least $n_c - (k-1)n_s$. On the other hand, for an atom $j \in \Omega - S$, note that there are total k support columns, the jth column $A(:, j)$ possesses at most kn_s number of 1 belonging to certain support column. Hence, the column $A(:, j)$ has contributed at most kn_s number of 1 to the observation y, which implies that $\|y - A(:, j)\|_0^+$ decreases at most kn_s. Therefore, by the assumption $n_c > (2k-1)n_s$, we have

$$\|y - A(:, i)\|_0^+ < \|y - A(:, j)\|_0^+ \tag{35}$$

for any $i \in S$ and $j \in \Omega - S$. In other words, the first selected atom by Algorithm MP1 belongs to the support set S.

Secondly, for the lth step, $l = 1, \dots, k-1$, it is similar to show that $\|y_r^l - A(:, i)\|_0^+$ decreases at least $n_c - (k-1)n_s$ for $i \in S$, while $\|y_r^l - A(:, j)\|_0^+$ decreases at most $(k-l)n_s$ for $j \in \Omega - S$. Thus, the similar inequality

$$\|y_r^l - A(:, i)\|_0^+ < \|y_r^l - A(:, j)\|_0^+ \tag{36}$$

still holds for any $i \in S - T^l$ and $j \in \Omega - S$. This implies the assertion of the theorem by observing that the resultant atom by Algorithm MP1 belongs to the support set S at each step. □

Remark 4.1. This theorem provides a sufficient condition for successful recovery of MP1 algorithm for model (1) using the LDPC measurement matrix. If the matrix A has sufficiently large row number m compared to the parameters n_s and n_c, the condition is also necessary for uniqueness. Indeed, suppose on the contrary $n_c \le (2k-1)n_s$, two different k-sparse solutions can be constructed similarly to the above proof.

Next we consider sufficient condition of uniqueness condition for the MP2 algorithm with respect to model (2).

Theorem 4.2. *For an irregular LDPC matrix $A_{m\times n}$ with parameter (n_c, n_s), the solution by Algorithm MP2 of model (2) for k-sparse signal x is unique if $n_c > kn_s$.*

Proof. For the initial step, it is to compare the differences between $\|y - A(:, l)\|_0^+$ for $l = 1, \ldots, n$. Notice that the multiplication of $A\odot x$ is an OR logic operator for each component, it is clear that the column $A(:, i)$ has contributed exactly n_c number of 1 to the observation y for an index $i \in S$. This implies that $\|y - A(:, i)\|_0^+$ decreases n_c. On the other hand, for the atom $j \in \Omega - S$, the column $A(:, j)$ has contributed at most kn_s number of 1 to the observation y, since the jth column has at most n_s number of 1 in common with each support column. In other words, $\|y - A(:, j)\|_0^+$ decreases at most kn_s. Then, based on the assumption that $n_c > kn_s$, it follows $\|y - A(:, i)\|_0^+ < \|y - A(:, j)\|_0^+$ for any $i \in S$ and $j \in \Omega - S$. Hence, the first atom selected by Algorithm MP2 belongs to the support set S.

Secondly, for the lth step, $l = 1, \ldots, k-1$, it is to compare $\|y - A \odot x_{T^l \cup \{t\}}\|_0^+$ for $t \in \Omega - T^l$. For $i \in S - T^l$, the constructed $x_{T^l\cup\{i\}}$ by expanding index set $T^l \cup \{i\}$ implies that the included extra ith column has at least contributed $n_c - ln_s$ number of 1 to the observation y, since the ith column has at most the same ln_s number of 1 to the former selected support columns. This means that $\|y - A \odot x_{T^l\cup\{i\}}\|_0^+$ decreases at least $n_c - ln_s$. On the other hand, for $j \in \Omega - S$, the resultant jth column by constructed $x_{T^l\cup\{j\}}$ possesses and contributes at most $(k-l)n_s$ number of 1 to observation y, since there are total $(k-l)$ support columns which have not been identified. It follows that $\|y - A \odot x_{T^l\cup\{j\}}\|_0^+$ decreases at most $(k-l)n_s$ for $j \in \Omega - S$. Therefore, by the assumption $n_c > kn_s$, we have

$$\|y - A \odot x_{T^l\cup\{i\}}\|_0^+ < \|y - A \odot x_{T^l\cup\{j\}}\|_0^+ \tag{37}$$

for any $i \in S - T^l$ and $j \in \Omega - S$. This implies the assertion of the theorem by observing that the resultant atom by Algorithm MP2 belongs to the support set S at each step. □

Remark 4.2. If the matrix A has sufficiently larger row number m compared to the parameters n_s and n_c, the condition is also necessary for uniqueness. Indeed, by supposing $n_c \le kn_s$, an irregular LDPC matrix with parameter (n_c, n_s) can be easily constructed similarly to the above proof such that Algorithm MP2 possesses two different solutions.

Now we consider the uniqueness conditions of the noise case for Algorithms MP1 and MP2.

Theorem 4.3. *For an irregular LDPC matrix $A_{m\times n}$ with parameter (n_c, n_s), the solution by Algorithm MP1 of noise model (5) for k-sparse signal x is unique if $n_c > (2k-1)n_s + 2n_v$.*

Proof. The procedure of proof is similar to Theorem 4.1. The only change is the minimum decreasing of $\|y_r^l - A(:, i)\|_0^+$ which may be fewer than n_v for $i \in S$, while the maximum decreasing of $\|y_r^l - A(:, j)\|_0^+$ may be larger than n_v for $j \in \Omega - S$, due to the disturbance of the binary noise. And based on this fact, the assertion follows. □

Theorem 4.4. *For an irregular LDPC matrix $A_{m\times n}$ with parameter (n_c, n_s), the solution by Algorithm MP2 of noise model (6) for k-sparse signal x is unique if $n_c > kn_s + 2n_v$.*

Proof. The procedure of proof is similar to Theorem 4.2. The only change is the minimum decreasing of $\|y - A \odot x_{T^l\cup\{i\}}\|_0^+$ which may be fewer than n_v for $i \in S$, while the maximum decreasing of $\|y - A \odot x_{T^l\cup\{j\}}\|_0^+$ may be larger than n_v for $j \in \Omega - S$, due to the disturbance of the binary noise. And based on this fact, the assertion follows. □

4.2. Discussion on Bayesian-based algorithms

It is basically believed that Bayesian-based method utilizes the majority of the post-probability information provided by the available observation data. Thus, it is reasonable that Bayesian-based method generally provides much accurate estimation of the source signal, which is demonstrated by the relevant simulation in the next section.

It is interesting to note that Bayesian-based method can be well implemented to the deterministic observation models (1) and (2) without any random variables therein. The idea is typical in Bayesian-based framework to derive post distribution by using the prior distribution and likelihood. The prior is conveniently assumed to be equivalently small positive values, say 0.1, to be close to the sparse source signal. Then, the resultant post probabilities (probably after several iterative steps) is compared to the assumed prior to make a choice of support indices of the source signal, which directly indicates the source binary valued signal. A good criterion is that the most possible support indices are those components with large increasing of post probabilities compared to the corresponding prior probabilities, since the post probabilities is determined by the available observation information and prior. The mentioned criterion actually applies the idea of matching pursuit method to select the most possible support indices.

4.3. Relation between spark and girth

To further understand the relation between the concepts spark in CS context and girth in LDPC code context, we develop an inequality for spark and girth.

Theorem 4.5. *For a LDPC matrix A with parameter $n_c \ge 2$, the spark of A is no less than half of its girth, i.e.,* spark$(A) \ge$ girth$(A)/2$.

Proof. Suppose spark$(A) = s$. Without loss of generality, based on the definition of spark, let the first s columns of matrix A be linearly dependent. It is shown that there exists a closed loop comprising of alternate row and column line segments (connecting two number of 1 in the relevant line) within these s columns. This implies that girth$(A) \le 2s$ and the assertion follows. The loop could start from any selected column of these s columns mentioned. Based on the assumption $n_c \ge 2$, let us start from an entry $a_{i_1,j_1} = 1$ of the j_1th column, $j_1 \in \Omega_s = \{1, 2, \ldots, s\}$, $i_1 \in \Omega_m = \{1, 2, \ldots, m\}$. Then we try to find another entry $a_{i_1,j_2} = 1$ in the rest $s-1$ columns, i.e., $j_2 \in \Omega - \{j_1\}$. With the same row index i_1, the two entries a_{i_1,j_1} and a_{i_1,j_2} construct a horizontal line segment of the loop. Next, within the j_2th column finding another entry $a_{i_2,j_2} = 1$ with $i_2 \in \Omega_m - \{i_1\}$, which is guaranteed by assumption $n_c \ge 2$. The following step is similar to the procedure for finding next column index $j_3 \in \Omega_s - \{j_2\}$ such that the entry $a_{i_2,j_3} = 1$. In the case $j_3 = j_1$, a four-length loop comprised of $a_{i_1,j_1} \to a_{i_1,j_2} \to a_{i_2,j_2} \to a_{i_2,j_1} \to a_{i_1,j_1}$ is found. If $j_3 \ne j_1$, the process continues to find another entry 1 of the same column (guaranteed by $n_c \ge 2$), and then try to find an entry 1 with the same row index in some other column. Once the resultant column index has been already included in the existing line segment, a closed loop is formed by

Table 1
Operation time for model (1): noiseless case.

Method	1	2	3	4	5	6	7	8	9	10
B1	0.1827	0.1757	0.1755	0.1803	0.1801	0.1869	0.1765	0.1717	0.1749	0.1767
MP1	1.0e−4	1.5e−4	1.9e−4	2.3e−4	2.8e−4	3.2e−4	3.5e−4	3.8e−4	4.2e−4	4.5e−4
Ref. [27]	0.0019	0.0021	0.0025	0.0026	0.0027	0.0027	0.0025	0.0024	0.0026	0.0026

Table 2
Operation time for model (5): noise case.

Method	1	2	3	4	5	6	7	8	9	10
B3	0.1717	0.1671	0.1633	0.1641	0.1653	0.1627	0.1629	0.1626	0.1628	0.1624
MP1	1.0e−4	1.5e−4	1.8e−4	2.2e−4	2.7e−4	3.0e−4	3.2e−4	3.8e−4	4.3e−4	4.6e−4
Ref. [27]	0.0214	0.0206	0.0243	0.0248	0.0203	0.0203	0.0207	0.0200	0.0197	0.0197

Construction Algorithm for LDPC matrix

1. *Initialization*. Given m, n_c and n_s, generate a random $\{0,1\}$-valued $m\times 1$ vector, α_0, containing n_c 1 value entries independently distributed over the index set $\{1,2,\ldots,n\}$.Let the first column of A be α_0, *i.e.*,$A(1:m,1)=\alpha_0$.
2. *Updating*. Randomly generate a $\{0,1\}$-valued $m\times 1$ vector, α, containing n_c1's entries independently distributed over the index set.
3. *Verification*. Check whether the inner product between α and the existing columnsof A is not greater than n_s. Specifically, if $\max\{\mathrm{abs}(\alpha^T A)\}\le n_s$, thenexpand A to $[A\ \alpha]$. Go to Step 2) for the next possible expansion. Otherwise stop.

connecting the latest entry 1 the and the former entry 1 within the latest column. On the other hand, the process will not stop an entry 1 without finding another entry 1 in some other column by the linearly dependent assumption of these s columns. In other words, the process can always form a closed loop with exact two entries in each relevant row and column. Observing the number of total line segments of the loop is less than 2 s, the assertion follows. □

4.4. Construction of LDPC matrix

Since the measurement matrix plays a key role in the model structure, it is important to find an effective way to construct LDPC matrix for conducting the whole BSSR issue. The LDPC matrix can be constructed in a random generation way by first generating a random candidate column vector and then by checking whether the at most n_s entries with the same row index is obeyed to expand the matrix or not.

5. Numerical examples

In this section, we design several simulations to show the effectiveness of the proposed recovery algorithms (i.e., MP1, MP2, B1, B2, B3, B4) compared with existing algorithms such as the one in literature [27]. Moreover, the theoretical results established in Section 4 are also verified for given model and sparsity k_0 and other corresponding parameters.

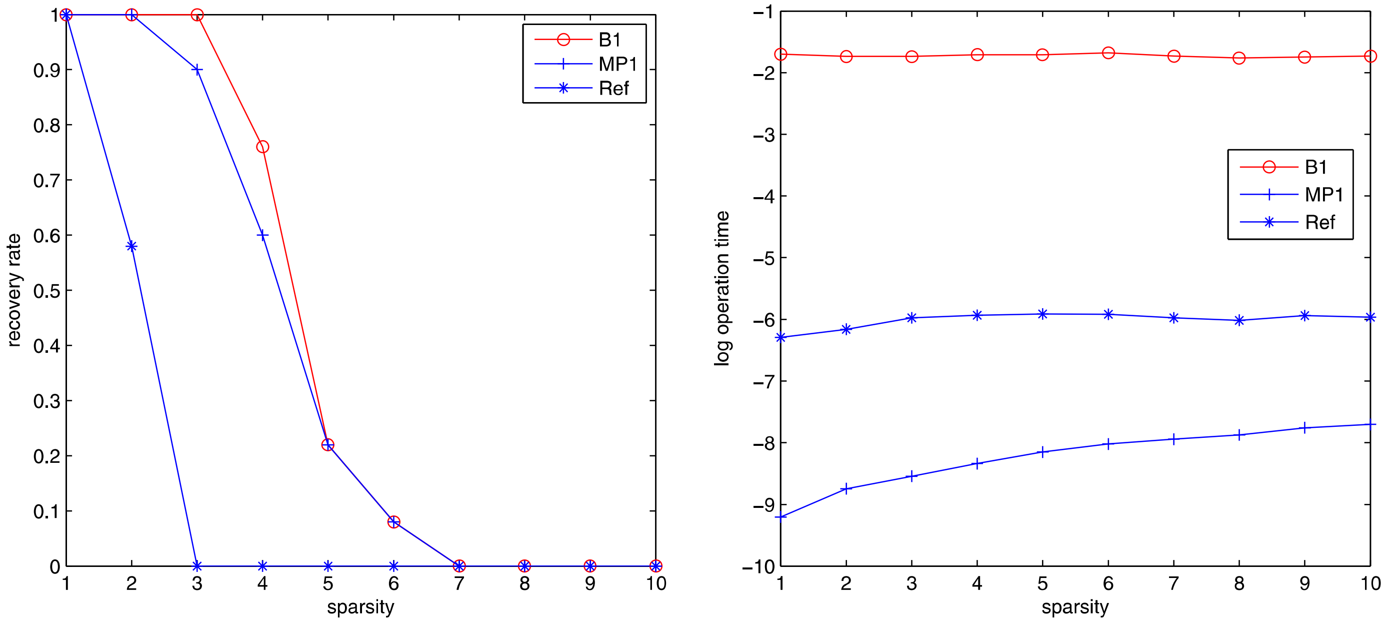
5.1. Models (1) and (5)

We first consider the noiseless Model (1) with given sparsity $k_0=2$. Let $m=30$ and $n_s=1$, based on Theorem 4.1, the number of 1 in each column of the measurement matrix should be at least $n_c=(2k_0-1)n_s+1=4$ to guarantee successful recovery. For the comparison, by using the proposed LDPC matrix construction method, we derived a matrix with the size of 30×39. The experiment checked the sparsity with $k=1,2,\ldots,10$ respectively. For each fixed sparsity, generating 50 random k-sparse signals to recover the source signals by algorithm of Lu et al. [27], Algorithms B1 and MP1 respectively, the resultant recovery rates are plotted in Fig. 2(a), which shows that the proposed algorithms (both MP1 and B1) outperforms the algorithm of Lu et al. [27] (denoted by Ref in the figure), and meanwhile, the Bayesian-based method B1 is slightly better than MP1 in most cases. The simulation also verifies the theoretical assertion of Theorem 4.1 for sparsity $k\le 2$. The corresponding logarithm operation times of these algorithms are plotted in Fig. 2(b), see also in Table 1, which shows that the proposed MP-based method runs the fastest. Even though the efficiency of the proposed BM-based is not competitive, its recovery rate performs the best, which is meaningful to some real-world applications.

Next consider the noise Model (5) with given sparsity $k_0=1$. Let $m=30$, $n_s=1$ and $n_v=1$. According to Theorem 4.3, the number of 1 in each column of the measurement matrix should be at least $n_c=(2k_0-1)n_s+2n_v+1=4$ for successful recovery. By using the proposed LDPC matrix construction method, a matrix with the size of 30×41 is derived as measurement matrix. Checking the sparsity with $k=1,2,\ldots,10$. For each fixed sparsity, generating 50 random k-sparse signals to recover the source signals by algorithm of Lu et al. [27], Algorithms B3 and MP1 respectively, the resultant recovery rates are plotted in Fig. 3(a), which shows that the proposed algorithms (both MP1 and B3) outperforms the algorithm of Lu et al. [27], and meanwhile, the Bayesian method B3 is significantly better than MP1 in most cases. The simulation also verifies the theoretical assertion of Theorem 4.3 for sparsity $k=1$. The corresponding logarithm operation times of these algorithms are plotted in Fig. 3(b), see also in Table 2, which shows almost the same performance as Fig. 2(b) does.

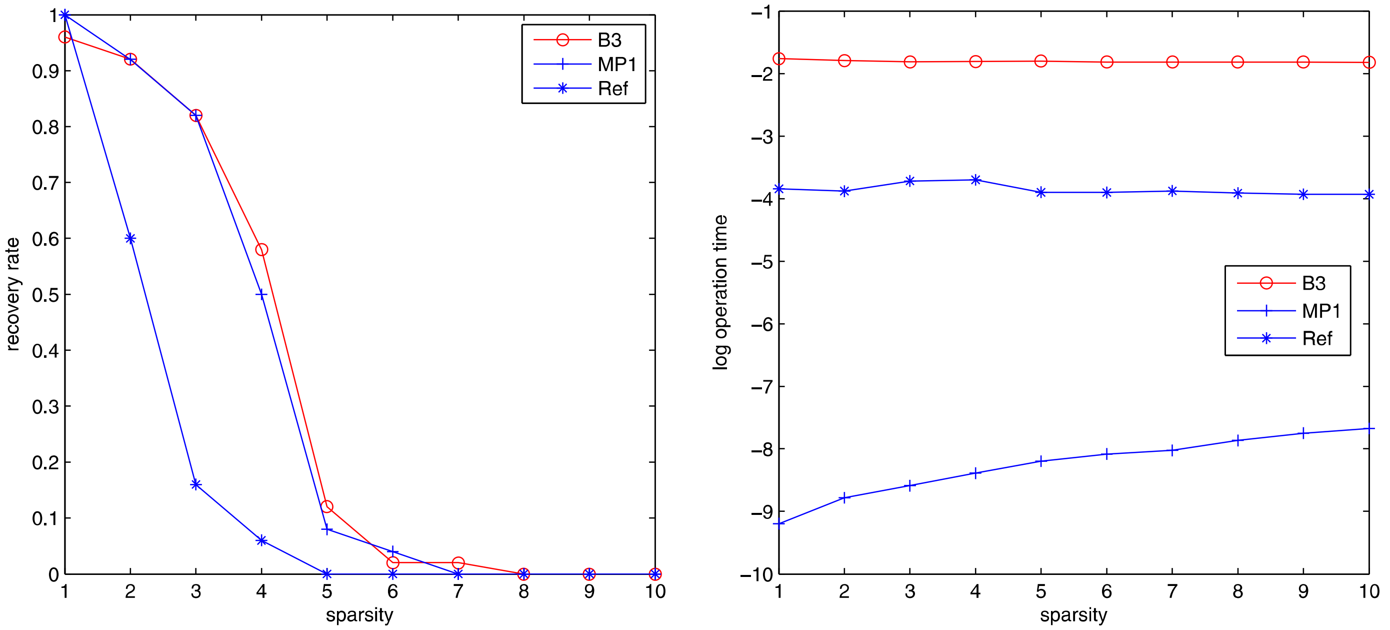
5.2. Models (2) and (6)

We first consider the noiseless model (2) with given sparsity $k_0=3$. Let $m=40$ and $n_s=1$, based on Theorem 4.2, the number of 1 in each column of the measurement matrix should be at least $n_c=k_0n_s+1=4$ to guarantee successful recovery. For the comparison, by using the proposed LDPC matrix construction method, we derived a matrix with the size of 40×71. Let the sparsity be $k=1,2,\ldots,10$, respectively. For each fixed sparsity, generating 50



(a) recovery rate (b) logarithm operation time

Fig. 2. (a) The plot of the recovery rates v.s. the sparsity for Model (1) by algorithm in [27], Algorithms B1 and MP1 under noiseless case using a matrix generated by the proposed LDPC construction method for $n_s = 1$, $n_c = 4$, $k_0 = 2$, $m = 30$ and $n = 39$. (b) The plot of logarithm operation times by algorithm in [27], Algorithms B1 and MP1 under noiseless case.

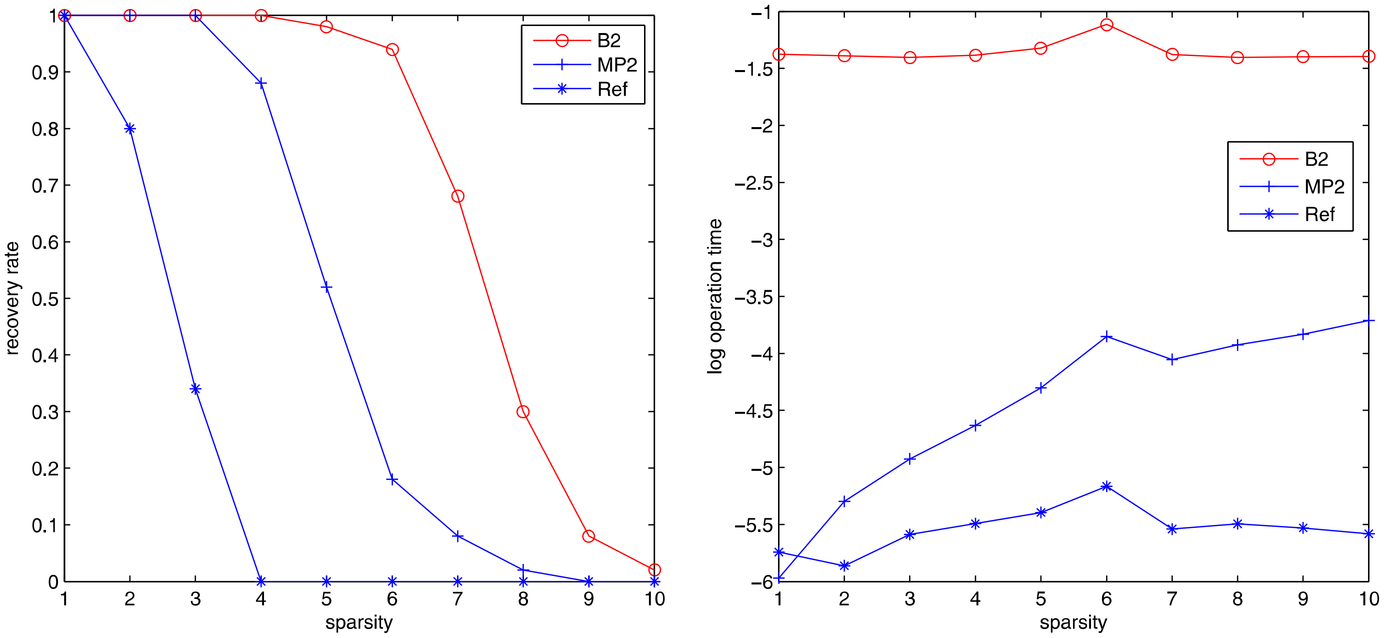


(a) recovery rate (b) logarithm operation time

Fig. 3. (a) The plot of the recovery rates v.s. the sparsity for Model (5) by algorithm in [27], Algorithms B3 and MP1 under noise case with $n_s = 1$, $n_c = 4$, $n_v = 2$, $k_0 = 1$, $m = 30$ and $n = 41$. (b) The plot of logarithm operation times by algorithm in [27], Algorithms B3 and MP1 under noise case.

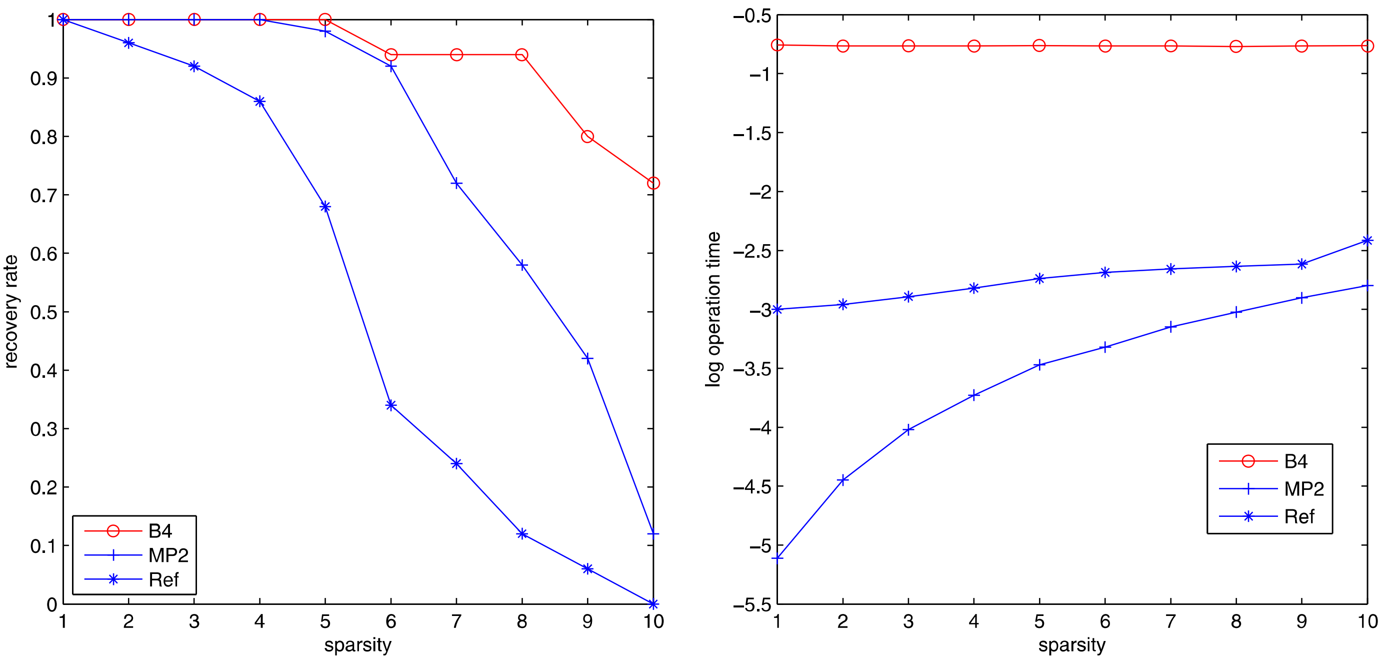
random k-sparse signals to recover the source signals by algorithm of Lu et al. [27], Algorithms B2 and MP2 respectively, the resultant recovery rates are plotted in Fig. 4(a), which shows that the proposed algorithms (both MP2 and B2) outperform algorithm of Lu et al. [27], and meanwhile, the Bayesian method B2 is significantly better than MP2 in most cases. The simulation also verifies the theoretical assertion of Theorem 4.2 for sparsity $k \leq 3$. The corresponding logarithm operation times of these algorithms are plotted in Fig. 4(b), see also in Table 3, which shows almost the same performance as Fig. 2(b) does.

Next, we consider the noise model (6) with given sparsity $k_0 = 2$. Let $m = 80$, $n_s = 1$ and $n_v = 1$, based on Theorem 4.4,



(a) recovery rate (b) logarithm operation time

Fig. 4. (a) The plot of the recovery rates v.s. the sparsity for model (2) by algorithm in [27], Algorithms B2 and MP2 under noiseless case using a matrix generated by the proposed LDPC construction method for $n_s = 1,\ n_c = 4,\ k_0 = 3,\ m = 40$ and $n = 71$. (b) The plot of logarithm operation times by algorithm in [27], Algorithms B2 and MP2 under noiseless case.



(a) recovery rate (b) logarithm operation time

Fig. 5. (a) The plot of the recovery rates v.s. the sparsity for model (6) by algorithm in [27], Algorithms B4 and MP2 under noise case using a matrix generated by the proposed LDPC construction method for $n_s = 1,\ n_c = 5,\ k_0 = 2,\ m = 80$ and $n = 100$. (b) The plot of logarithm operation times by algorithm in [27], Algorithms B4 and MP2 under noise case.

Table 3
Operation time for model (2): noiseless case.

Method	1	2	3	4	5	6	7	8	9	10
B2	0.2528	0.2494	0.2459	0.2505	0.2668	0.3287	0.2521	0.2459	0.2473	0.2482
MP2	0.0026	0.0050	0.0073	0.0097	0.0135	0.0213	0.0174	0.0197	0.0217	0.0245
Ref. [27]	0.0032	0.0028	0.0037	0.0041	0.0045	0.0057	0.0039	0.0041	0.0040	0.0038

Table 4
Operation time for model (6): noise case.

Method	1	2	3	4	5	6	7	8	9	10
B4	0.4695	0.4660	0.4663	0.4660	0.4672	0.4658	0.4653	0.4632	0.4655	0.4664
MP2	0.0060	0.0117	0.0179	0.0240	0.0311	0.0361	0.0429	0.0487	0.0549	0.0610
Ref. [27]	0.0498	0.0519	0.0555	0.0596	0.0647	0.0681	0.0701	0.0717	0.0731	0.0893

the number of 1 in each column of the measurement matrix should be at least $n_c = k_0 n_s + 2n_v + 1 = 5$ to guarantee successful recovery. For the comparison, by using the proposed LDPC matrix construction method, we derived a matrix with the size of 80×100. Let the sparsity be $k = 1, 2, \dots, 10$, respectively. For each fixed sparsity, generating 50 random k-sparse signals to recover the source signals by algorithm of Lu et al. [27], Algorithms B4 and MP2 respectively, the resultant recovery rates are plotted in Fig. 5(a), which shows that the proposed algorithms (both MP2 and B4) outperforms the algorithm of Lu et al. [27], and meanwhile, the Bayesian method B4 is significantly better than MP2 in most cases. The simulation also verifies the theoretical assertion of Theorem 4.4 for sparsity $k \leq 2$. The corresponding logarithm operation times of these algorithms are plotted in Fig. 5(b), see also in Table 4, which shows almost the same performance as Fig. 2(b) does.

6. Conclusion

A new issue of binary sparse signal recovery with underdeterministic binary observation generated from binary measurement matrix is considered. By using the subtle structure of LDPC matrix, and inspired by the matching pursuit method and Bayesian method for compressive sensing issues, effective and efficient recovery algorithms are developed and theoretical conditions are also established to guarantee the successful recovery. Numerical experiments are conducted to demonstrate the effectiveness of these algorithms and verify the four established theorems for matching pursuit-based algorithms. The model structure is mainly determined by its measurement matrix, which is typically in form of LDPC matrix to facilitate algorithms design in this paper. To make the whole framework practical, an effective construction method of LDPC matrix is also presented.

Numerical simulations in the last section show that the recovery rates of the proposed heuristic algorithms are much higher than that of the state-of-the-art algorithms, which is the biggest strength of the proposed methods. However, such a competitive performance is achieved at the cost of the operation times. Simulations in Figs. 2–5(b) show that the proposed Bayesian methods are inevitably time-consuming to certain extend. Even for the matching pursuit methods, the corresponding computation times become much longer as the data size increases. It is of interest in the future to develop high efficient and effective recovery algorithms for these models to overcome the curse of dimensionality.

Acknowledgements

This work was supported in part by the Natural Science Foundation of China under Grant 61703283, 61573248, 61672357, 61802267 and U1713214, in part by the Guangdong Natural Science Foundation under Project 2017A030310067, 2017A030313367, 2018A030310450 and 2018A030310451, in part by the Science and Technology Funding of Guangdong Province (2018A050501014), in part by the Research Grant of The Hong Kong Polytechnic University (Project Code: G-YBD9), in part by the China Postdoctoral Science Foundation under Project 2016M590812, 2017M612736 and 2017T100645, in part by the Shenzhen Municipal Science and Technology Innovation Council under Grant JCYJ20160429182058044 and Grant JCYJ20170302153434048, in part by the Fundamental Research Funds for the Shenzhen University.

References

[1] J. Yang, L. Zhang, Y. Xu, J.Y. Yang, Beyond sparsity: the role of l_1-optimizer in pattern classification, Pattern Recognit. 45 (2012) 1104–1118.
[2] D.S. Pham, On group-wise l_p regularization: theory and efficient algorithms, Pattern Recognit. 48 (2015) 3728–3738.
[3] C.Y. Lu, H. Min, J. Gui, L. Zhu, Y.K. Lei, Face recognition via weighted sparse representation based classification, J. Vis. Commun. Image Represent. 24 (2) (2013) 111–116.
[4] J.X. Mi, D. Lei, J. Gui, A novel method for recognizing face with partial occlusion via sparse representation, Optik 124 (24) (2013) 6786–6789.
[5] J. Gui, Z. Sun, G. Hou, T. Tan, An optimal set of code words and correntropy for rotated least squares regression, Int. Joint Conf. Biom. (2014) 1–6.
[6] J. Gui, Z. Sun, W. Jia, R. Hu, Y. Lei, S. Ji, Discriminant sparse neighborhood preserving embedding for face recognition, Pattern Recognit. 45 (8) (2012) 2884–2893.
[7] J. Gui, T. Liu, D. Tao, Z. Sun, T. Tan, Representative vector machines: a unified framework for classical classifiers, IEEE Trans. Cybern. 46 (8) (2016) 1877–1888.
[8] J. Gui, D. Tao, Z. Sun, Y. Luo, X. You, Y.Y. Tang, Group sparse multiview patch alignment framework with view consistency for image classification, IEEE Trans. Image Process. 23 (7) (2014) 3126–3137.
[9] J. Gui, Z. Sun, S. Ji, D. Tao, T. Tan, Feature selection based on structured sparsity: a comprehensive study, IEEE Trans. Neural Netw. Learn. Syst. 28 (7) (2017) 1490–1507.
[10] A. Ali, A. Al-Rabah, M. Masood, T.Y. Al-Naffouri, Receiver-based recovery of clipped OFDM signals for PAPR reduction: a Bayesian approach, IEEE Access 2 (2014) 1213–1224.
[11] N.M.B. Souto, H.A. Lopes, Efficient recovery algorithm for discrete valued sparse signals using an ADMM approach, IEEE Access 5 (2017) 19562–19569.
[12] Y.V. Zakharov, V.H. Nascimento, R.C. De Lamare, F.G. De Almeida Neto, Low-complexity DCD-based sparse recovery algorithms, IEEE Access 5 (2017) 12737–12750.
[13] T. Wimalajeewa, P.K. Varshney, Performance bounds for sparsity pattern recovery with quantized noisy random projections, IEEE J. Sel. Top. Signal Process. 6 (1) (February 2012) 43–57.
[14] D.L. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization, Proc. Natl. Acad. Sci. 100 (2003) 2197–2202. USA.
[15] E.J. Candes, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory 51 (2005) 4203–4215.
[16] E.J. Candes, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory 52 (2006) 489–509.
[17] E.J. Candes, T. Tao, The dantzig selector: statistical estimation when p is much larger than n, Ann. Statist. 35 (2007) 2313–2351.
[18] L. Baldassarre, Y.H. Li, J. Scarlett, B. Gozcu, I. Bogunovic, V. Cevher, Learning-based compressive subsampling, IEEE J. Sel. Top. Signal Process. 10 (4) (June 2012) 809–822.
[19] Z. Pan, C. Zhang, Relaxed sparse eigenvalue conditions for sparse estimation via non-convex regularized regression, Pattern Recognit. 48 (2015) 231–243.
[20] G. Wang, H. Zhang, S. Yu, S. Ding, A family of the subgradient algorithm with several cosparsity inducing functions to the cosparse recovery problem, Pattern Recognit. Lett. 80 (2016) 64–69.
[21] J. Ma, L. Ping, Orthogonal AMP, IEEE Access 5 (2017) 2020–2033.
[22] A. Cohen, W. Dahmen, R. DeVore, Compressed sensing and best k-term approximation, J. Am. Math. Soc. 22 (2009) 211–231.
[23] R.A. DeVore, Deterministic constructions of compressed sensing matrices, J. Complex. 23 (2007) 918–925.
[24] Y.C. Eldar, G. Kutyniok, Compressed Sensing: Theory and Applications, Cambridge University Press, 2012.
[25] U. Nakarmi, N. Rahnavard, BCS: compressive sensing for binary sparse signal, in: Military Communications Conference Orlando, 2012.
[26] J. Acharya, A. Bhattacharyya, P. Kamath, Improved bounds for universal one-bit compressive sensing, IEEE International Symposium on Information Theory (ISIT 2017), 2017.
[27] J. Lu, T. Zhang, Q. Sun, Q. Hao, F. Hu, Binary compressive tracking, IEEE Trans. Aerosp. Electron. Syst. (2017). In press.

[28] J. Lu, J. Gong, Q. Hao, F. Hu, Multi-agent based wireless pyroelectric infrared sensor networks for multi-human tracking and selfcalibration, in: Proc. IEEE Conf. on Sensors, 2013, pp. 1–4. Baltimore, MD

[29] Q. Hao, F. Hu, J. Lu, Distributed multiple human tracking with wireless binary pyroelectric infrared (PIR) sensor networks, in: Proc. IEEE Sensors, Kona, HI, 2010, pp. 946–950.

[30] J. Lu, J. Gong, Q. Hao, F. Hu, Space encoding based compressive multiple human tracking with distributed binary pyroelectric infrared sensor networks, in: Proc. IEEE Conf. on Multisensor Fusion and Integration for Intelligent Systems, Hamburg, 2012, pp. 180–185.

[31] R. Zheng, K. Vu, A. Pendharkar, G. Song, Obstacle discovery in distributed actuator and sensor networks, ACM Trans. Sens. Netw. 7 (3) (2010) 22.

[32] J. Gui, T. Liu, Z. Sun, D. Tao, T. Tan, Supervised discrete hashing with relaxation, IEEE Trans. Neural Netw. Learn. Syst. 29 (3) (2018) 608–617.

[33] J. Gui, T. Liu, Z. Sun, D. Tao, T. Tan, Fast supervised discrete hashing, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2) (2018) 490–496.

[34] Y. Cao, H. Qi, W. Zhou, J. Kato, K. Li, X. Liu, J. Gui, Binary hashing for approximate nearest neighbor search on big data: a survey, IEEE Access 6 (1) (2018) 2039–2054.

[35] B. Wang, Coverage problems in sensor networks: a survey, ACM Comput. Surv. (CSUR) 43 (4) (2011) 32.

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

[37] D. Needell, R. Vershynin, Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit, Found.Comput.Math. 9 (3) (2009) 317–334.

[38] D. Needell, R. Vershynin, Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit, IEEE J. Sel. Top. Signal Process. 4 (2) (2010) 310–316.

[39] D. Donoho, Y. Tsaig, I. Drori, J. Starck, Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit, 2006,

[40] M. Figueiredo, R. Nowak, S. Wright, Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems, IEEE J. Sel. Top. Signal Process. 1 (4) (2008) 586–597.

[41] D. Needell, J. Tropp, CosaMP: iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. Anal. 26 (3) (2009) 301–321.

[42] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, IEEE Trans. Inf. Theory 55 (5) (2009) 2230–2249.

[43] E. Liu, V. Temlyakov, Orthogonal super greedy algorithm and applications in compressed sensing, IEEE Trans. Inf. Theory 58 (4) (2012) 2040–2047.

[44] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, IEEE Trans. Signal Process. 56 (6) (2008) 2346–2355.

[45] S.D. Babacan, R. Molina, A.K. Katsaggelos, Fast Bayesian compressive sensing using laplace priors, in: IEEE Proc. ICASSP09, 2009, pp. 2873–2876.

[46] D. Yang, H. Li, G.D. Peterson, Decentralized turbo Bayesian compressed sensing with application to UWB systems, EURASIP J. Adv. Signal Process. 2011 (2011). Article ID 817947.

[47] L. Shi, Z. Zhou, L. Tang, Ultra-wideband channel estimation based on distributed Bayesian compressive sensing, JDCTA Int. J. Digital Content Technol.Appl. 5 (2) (2011) 1–8.

[48] L. Shi, Z. Zhou, L. Tang, H. Yao, J. Zhang, Ultra-wideband channel estimation based on Bayesian compressive sensing, in: IEEE Proc. ISCIT10, 2010, pp. 779–782.

[49] S. Zhang, Y. Liu, X. Li, G. Bi, Variational Bayesian sparse signal recovery with LSM prior, IEEE Access 5 (2017) 26690–26702.

[50] Z. Zhang, B.D. Rao, Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning, IEEE J. Sel. Top. Signal Process. 5 (5) (2011) 912–926.

[51] R.G. Gallager, Low-Density Parity-Check Codes, MIT Press, Cambridge, MA, 1963.

[52] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. (1948) 372–423. 623–656.

Xiaoli Hu received Ph.D. from the Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences, in 2006, after which he worked in various universities of Sweden, China, Canada and Australia. He is currently a postdoctoral research fellow of the Research Institute for Future Media Computing at the College of Computer Science & Software Engineering, Shenzhen University, China. He published around 20 refereed research papers and conferences in the field of system identification and signal processing. His research interests include sparse signal recovery and its application, signal detection, system identification. E-mail: hutby@yeah.net.

Jiajun Wen received the Ph.D. degree in computer science and technology from the Harbin Institute of Technology, China, in 2015. He has been a Research Associate and a Post-Doctoral Fellow with The Hong Kong Polytechnic University, Hong Kong, since 2013. He is currently an assistant professor with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. His current research interests include pattern recognition and video analysis.

Zhihui Lai received the B.S. degree in mathematics from South China Normal University, Guangzhou, China, in 2002, the M.S. degree from Jinan University, Guangzhou, in 2007, and the Ph.D. degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology, Nanjing, China, in 2011. He has been a Research Associate, a Post-Doctoral Fellow, and a Research Fellow with Hong Kong Polytechnic University, Hong Kong. He has published over 60 scientific articles, including 30 papers published on top-tier IEEE TRANSACTIONS. His current research interests include face recognition, image processing and content-based image retrieval, pattern recognition, compressive sense, human vision modelization, and applications in the field of intelligent robot research. Dr. Lai is an Associate Editor of the International Journal of Machine Learning and Cybernetics.

Wai Keung Wong received the Ph.D. degree from The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. He is currently a Professor with The Hong Kong Polytechnic University. He has authored or coauthored more than 100 papers in refereed journals and conferences, including the IEEE Transactions on Neural Networks and Learning Systems, the IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics and the IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, *Pattern Recognition, CHAOS, European Journal of Operational Research*, Neural Networks, Applied Soft Computing, Information Science, Decision Support Systems, among others. His current research interests include pattern recognition and feature extraction.

Linlin Shen received the B.Sc. degree from Shanghai Jiaotong University, Shanghai, China, and the Ph.D. degree from the University of Nottingham, Nottingham, U.K. He was a Research Fellow with the University of Nottingham, working on MRI brain image processing. He is currently a Professor and the Director of the Computer Vision Institute, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. His research interests include Gabor wavelets, facial recognition, analysis/synthesis, medical image processing and hyperspectal image classification. Prof. Shen is listed as the Most Cited Chinese Researcher by Elsevier. He received the Most Cited Paper Award from the journal of Image and Vision Computing. His cell classification algorithms were the winners of the International Contest on Pattern Recognition Techniques for Indirect Immunofluorescence Images held by ICIP 2013 and ICPR 2016.