# Dynamic Recovery for Block Sparse Signals

Junying Ren[a], Chen Wei[a], Lei Yu[a,b,*], Haijian Zhang[a], Hong Sun[a]

[a]*Signal Processing Laboratory, School of Electronic Information, Wuhan University, Wuhan 430072, China*
[b]*ECE, Duke University, Durham, NC*

**Abstract**

As an important extension to sparsity, Block Structured Sparsity (BSS) has been widely investigated and a large set of algorithms have been designed to recover signals with BSS. In this paper, a dynamic system is proposed to recover signals with BSS, namely, D-BSS. Particularly, D-BSS turns to exploiting the dynamic systems governed by an $\ell_{2,1}$ norm constraint. It can be proved that the equilibrium of D-BSS is equivalent to optimum of traditional BSS algorithms, e.g., Group-lasso. Simulation results are given to illustrate the desirable performance of the D-BSS system, especially the improvement of convergence rate.

*Keywords:* Sparsity, Block Structured Sparsity, Dynamic System

## 1. Introduction

The technique of sparse recovery has been exploited in many applications recently [1][2][3], including computer vision, image processing (denoising, detection, recognition, classification and so on), e.g., the recent algorithms DTDML [4], LM3FE [5], mHDSC [6]. Sparse representation has shown huge potential capabilities in processing these problems. As a consequence, it is not redundant at all to further research and improve sparse recovery methods, and the problem of sparse signal recovery from

☆This work is supported by NSFC Grant 61401315, China Scholarship Council and the Project-sponsored by SRF for ROCS, SEM, under Grant 230303.

*Corresponding author

*Email addresses:* rjy@whu.edu.cn (Junying Ren), weichen@whu.edu.cn (Chen Wei), ly94@duke.edu (Lei Yu), haijian.zhang@whu.edu.cn (Haijian Zhang), hongsun@whu.edu.cn (Hong Sun)

insufficient measurements can be expressed via the following canonical formula [7]

$$\boldsymbol{y} = \Phi\boldsymbol{a} + \boldsymbol{\varepsilon} \tag{1}$$

with $\boldsymbol{y} \in \mathbb{R}^M$ the observed measurements corrupted by noise term $\boldsymbol{\varepsilon}$, $\Phi \in \mathbb{R}^{M\times N}$ the representation matrix (or measurement matrix), and $\boldsymbol{a} = [a_1, \cdots, a_N]^T \in \mathbb{R}^N$ an unknown sparse signal to be recovered. In common case, $M \ll N$, i.e., the recovery of sparse signal $\boldsymbol{a}$ based on $\boldsymbol{y}$ is an underdetermined problem. Generally, the solver to inverse (1) can be casted as the minimizer of the following cost function

$$V(\boldsymbol{a}) = \frac{1}{2}\|\boldsymbol{y} - \Phi\boldsymbol{a}\|_2^2 + \lambda\rho(\boldsymbol{a}) \tag{2}$$

where $\lambda > 0$ is a balance parameter and $\rho(\boldsymbol{a}) : \mathbb{R}^N \to \mathbb{R}_+$ is a function that considers the signal priori, e.g., sparsity or beyond sparsity.

Many different methods have been investigated for the inversion of (1). And theories and algorithms proposed recently are primarily inspired from either of the following aspects [3]: **Signal Sparsity** that investigates various sparsity promoting functions $\rho$, for instance $\ell_0$-norm [8], $\ell_1$-norm [7] or $\ell_p$-norm [9], and **Optimization Approaches** that investigate the method of minimizing (2), i.e., greedy approaches [10], constrained optimization [11, 12], homotopy based algorithms [13] and so on.

However, these aforementioned algorithms need high computational complexity and large storage requirements caused by the digital computation in discrete time. Thus, it might be impractical for engineering application that is time-varying and with large data set, such as radar imaging[14], face recognition[15] and DOA estimation[16] and so on. Inspired by the analog circuit (continuous dynamic system), which has faster response, some algorithms turn to continuous system for the optimization problem. The recent developed algorithm called *Locally Competitive Algorithm* (LCA) [17][18][19] is to make up for this defect. It is armed with the concepts of continuous dynamic system, where the sparse signals can be solved via a dynamic system that is highly parallel, rapid, and power-efficient.

On the other hand, considering more information of the signals, exploiting structure of signals for better performance is an important extension and a trend in sparse signal recovery. Thus block sparsity has been widely applied naturally. For example,

multi-band signals[20] [21] [22], gene expression [23], equalization of sparse communication channels [24], and source localization [25]. The concept of block-sparsity was previously introduced and investigated in many classic literatures, including BMP and BOMP [10], block-CoSaMP [26], Group-Lasso [27], CluSS-MCMC [28], MBCS-LBP [29] and so on [30] [31] [32] [33]. It can be theoretically and empirically proved that introducing block (or group) structure could largely improve the recovery performance of sparse signals[33, 26]. Consequently, a block awarded dynamic recovery algorithm for sparse signal will be applausive in the aforementioned applications where the intrinsic block structures are embedded as specific patterns.

In conclusion, the biggest difficulty for D-BSS is the analysis of the convergence, the accuracy of the recovery results and the convergence rate after adding the structure information. The main contributions of this paper can be clarified as follows. Firstly, the theoretical analysis of the proposed dynamical system is addressed. Particularly, providing that RIP condition is verified for the measurement matrix $\Phi$, it is guaranteed that the proposed dynamic system with block structured "active function"will converge with exponential rate. On the other hand, it can be shown theoretically that due to the consideration of the block structures, the convergence rate of the proposed dynamical system will be improved. Secondly, performances of the proposed D-BSS are compared to the state-of-the-art block aware algorithms, which conveys the superiority of D-BSS. Meanwhile, we exploited D-BSS to cope with the DOA problem.

The rest of this paper is organized as follows. In Section 2, the proposed dynamic system for recovery of block structured sparse signals is firstly presented. The mathematical analysis for the main result and the convergence properties are given in Section 3. In Section 4, simulations about the support, amplitude and convergence of D-BSS are presented to verify the theoretical results. Conclusion is given in the last section.

## 2. Dynamic Recovery of Block Sparse Signals

### *2.1. Recovery of Block Sparse Signals*

In this paper, we consider the unknown sparse signals $\boldsymbol{a} \in \mathbb{R}^N$ with block sparsity, i.e.,

$$\boldsymbol{a} = [\underbrace{a_1 \cdots a_m}_{\boldsymbol{a}^T[1]} \ \underbrace{a_{m+1} \cdots a_{2m}}_{\boldsymbol{a}^T[2]} \ \cdots \ \underbrace{a_{N-m+1} \cdots a_N}_{\boldsymbol{a}^T[l]}]^T \tag{3}$$

with $m$ the length of blocks, i.e., $N = lm$, entries of each partition $\boldsymbol{a}[s] = [a_{s,1}, a_{s,2}, ..., a_{s,m}]^T$ are either zero or nonzero simultaneously, where $s \in \{1, ..., l\}$ and $a_{s,i} = a_{(s-1)\cdot m+i}$. Block sparsity has been thoroughly investigated, in general, the estimate of $\boldsymbol{a}$ with block sparsity can be inferred by minimizing the following energy function [27], where $V_b(\boldsymbol{a})$ is differentiable with respect to $\boldsymbol{a}$ everywhere,

$$V_b(\boldsymbol{a}) \triangleq \frac{1}{2} \|\boldsymbol{y} - \Phi\boldsymbol{a}\|_2^2 + \lambda \|\boldsymbol{a}\|_{2,1} \tag{4}$$

where the mixed norm $\|\boldsymbol{a}\|_{2,1} = \sum_{i=1}^{l} \sqrt{\sum_{j=1}^{m} |a_{(i-1)\cdot m+j}|^2}$ is introduced instead of $\ell_1$ term in (2) and $\lambda > 0$ is the balance parameter.

Aimed at the optimization problem of (4), iterative approaches including proximal splitting methods [**?** ], iterative soft-thresholding [12] have been proposed and widely investigated. However, the algorithms exploiting the aforementioned cost function (4) are commonly with iterations composed by several steps and thus lead to heavy burden with high complexity and large storage requirement.

Similar to the method proposed in [17], it seems that we can also draw lessons from the dynamic system to take advantage of its sustainability and fast characteristics. We will exploit a different dynamic system to cope with the block sparse signals, where a new *active function* is designed to solve the sparse signals. In Section 2.2, we will show the state equation and output equation of our dynamic system.

### *2.2. Dynamic Approach*

In this paper, the dynamic system to recover block sparse signals $\boldsymbol{a}$ can be succinctly written as

$$\begin{aligned} \dot{\boldsymbol{x}}(t) &= f(\boldsymbol{x}(t), \boldsymbol{a}(t)) + r(t) \\ \boldsymbol{a}(t) &= h(\boldsymbol{x}(t)) \end{aligned} \tag{5}$$

with $t$ the time instant, $\boldsymbol{x}(t) : \mathbb{R}_+ \to \mathbb{R}^N$ the system states, $r(t) = \tau\Phi^T\boldsymbol{y}$ the inputs with $\boldsymbol{y}$ the measurements from (1), $\boldsymbol{a}(t) : \mathbb{R}_+ \to \mathbb{R}^N$ the outputs, $f : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N$ the vector fields which are defined as, $f(\boldsymbol{x}(t), \boldsymbol{a}(t)) = -\tau(\boldsymbol{x}(t) + (\Phi^T\Phi - \boldsymbol{I})\boldsymbol{a}(t))$, with $\boldsymbol{I}$ the identity matrix, and $\tau > 0$ a time constant that tunes the speed of convergence. For simplicity, the parameter $\tau = 1$ identically in the following. Actually, $\boldsymbol{x}(t)$ is the evolution state of the sparse signal to get, $\boldsymbol{a}(t)$ is the sparse signal which satisfies the active function $h(\cdot)$. The output equation $h(\boldsymbol{x}(t))$ is a point-wise thresholding function defined by

$$h(x_n) = \begin{cases} 0, & \|\boldsymbol{x}[s]\|_2 < \lambda \\ g(x_n), & \|\boldsymbol{x}[s]\|_2 \geq \lambda \end{cases} \tag{6}$$

with $g(x_n) = \left(1 - \frac{\lambda}{\|\boldsymbol{x}[s]\|_2}\right) x_n$, where $x_n$ is the $n$th element of state vector $\boldsymbol{x}$ and $s$ is the block index that $x_n$ belongs to. The estimate $\hat{\boldsymbol{a}}$ of a block sparsity by (4) also can be obtained when the system (5) evolves to steady state, i.e., $\hat{\boldsymbol{a}} = \lim_{t \to +\infty} \boldsymbol{a}(t)$. We call this method Dynamic Recovery of Block Sparse Signals (D-BSS).

Exploiting block structure information, D-BSS converges faster and more accurately than LCA, both for support or amplitude. Theorem 1 (section 3.1) will provide a theoretical guarantee that we may recover the block sparse signals using (5) instead of traditional optimization problem (4). Theorem 2 (section 3.2) will guarantee overall that the D-BSS has the ability to converge. Theorem 3 (section 3.3) will show that the support of D-BSS has the ability to converge in limited time. Theorem 4 (section 3.4) will give the convergence rate of D-BSS.

## 3. Theoretical Analysis to D-BSS: Equivalence and Convergence

### *3.1. Equivalence Between Optimal Model and D-BSS*

In this section, we firstly present an important relationship which is important for the theoretical analysis below between the mixed norm $P(\boldsymbol{a})$ and the thresholding function $h(\cdot)$.

**Lemma 1.** *If we have* $P(\boldsymbol{a}) \triangleq \|\boldsymbol{a}\|_{2,1}$ *and* $\boldsymbol{a} = h(\cdot)$ *defined as (6), i.e.,*

$$h(x_n) = \begin{cases} 0, & \|\boldsymbol{x}[s]\|_2 < \lambda \\ g(x_n), & \|\boldsymbol{x}[s]\|_2 \geq \lambda \end{cases}$$

*with* $g(x_n) = \left(1 - \frac{\lambda}{\|\boldsymbol{x}[s]\|_2}\right) x_n$, *then*

$$\frac{\lambda dP}{d\boldsymbol{a}} = \boldsymbol{x} - \boldsymbol{a} \tag{7}$$

*Proof.* $\frac{\lambda dP}{d\boldsymbol{a}} = \frac{\lambda d\|\boldsymbol{a}\|_{2,1}}{d\boldsymbol{a}}$, for $\boldsymbol{a}$, we discuss this issue in two cases, $a_n = 0$ and $a_n \neq 0$. When $a_n \neq 0$, $\frac{\lambda d\|\boldsymbol{a}\|_{2,1}}{da_n} = \frac{\lambda d\|\boldsymbol{a}[s]\|_2}{da_n} = \frac{\lambda a_n}{\|\boldsymbol{a}[s]\|_2}$, and $s$ is the block index that $a_n$ belongs to. Combining the $g(x_n)$ defined in (6), $a_n = h(x_n) = g(x_n) = \left(1 - \frac{\lambda}{\|\boldsymbol{x}[s]\|_2}\right) x_n$, then

$$\frac{\lambda a_n}{\|\boldsymbol{a}[s]\|_2} = \frac{\lambda x_n}{\|\boldsymbol{x}[s]\|_2} = x_n - a_n$$

When $a_n = 0$, because $P(\cdot)$ is continuous and differentiable everywhere, and $g(x(n))$ is continuous, then we have

$$\frac{\lambda d\|\boldsymbol{a}\|_{2,1}}{da_n}|_{a_n=0} = \frac{\lambda d\|\boldsymbol{a}\|_{2,1}}{da_n}|_{a_n\to 0}$$

according to (6), when $a_n \to 0$, $\|\boldsymbol{x}[s]\|_2 \to \lambda$, so

$$\frac{\lambda d\|\boldsymbol{a}\|_{2,1}}{da_n}|_{a_n\to 0} = \frac{\lambda d\|\boldsymbol{a}\|_{2,1}}{da_n}|_{\|\boldsymbol{x}[s]\|_2\to\lambda}$$

Using the result of $a_n \neq 0$ above,

$$\frac{\lambda d\|\boldsymbol{a}\|_{2,1}}{da_n}|_{\|\boldsymbol{x}[s]\|_2\to\lambda} = \lim_{\|\boldsymbol{x}[s]\|_2\to\lambda}(x_n - a_n) = \lim_{\|\boldsymbol{x}[s]\|_2\to\lambda}\frac{\lambda x_n}{\|\boldsymbol{x}[s]\|_2} = x_n$$

since $a_n$ is 0 here, so $x_n = x_n - a_n$

Consequently, $\frac{\lambda dP}{d\boldsymbol{a}} = \frac{\lambda d\|\boldsymbol{a}\|_{2,1}}{d\boldsymbol{a}} = \mathbf{x} - \boldsymbol{a}$, (7) is proved for all $a_n$. □

With Lemma 1, we have the following theorem showing that we can solve the dynamic system (5) instead of the optimization problem (4).

**Theorem 1.** *The dynamic system (5) has equilibrium points that are critical points of minimizing (4).*

*Proof.* For the dynamical system (5), any equilibrium point $\hat{\boldsymbol{x}}$ with a corresponding $\hat{\boldsymbol{a}}$ satisfies $\dot{\boldsymbol{x}}(t) = 0$, i.e.,

$$-\hat{\boldsymbol{x}} + h(\hat{\boldsymbol{x}}) - \Phi^T\Phi\hat{\boldsymbol{a}} + \Phi^T\boldsymbol{y} = 0 \tag{8}$$

For the $V_b(\boldsymbol{a})$ in (4), the point $\hat{\boldsymbol{a}}$ is a critical point of $V_b(\cdot)$ if and only if $\frac{dV_b}{d\boldsymbol{a}}|_{\boldsymbol{a}=\hat{\boldsymbol{a}}} = 0$, i.e.,

$$\Phi^T \boldsymbol{y} - \Phi^T \Phi \hat{\boldsymbol{a}} - \frac{\lambda dP}{d\hat{\boldsymbol{a}}} = 0 \tag{9}$$

From (7), the conditions (8) and (9) are the same. □

*3.2. The Convergence Performance of D-BSS*

In this section, we will demonstrate the D-BSS has globally quasi-convergent outputs, i.e., there exists a set $E = \{\boldsymbol{a} : s.t.\ \dot{\boldsymbol{a}} = 0\}$ such that for all $\boldsymbol{x}(0)$, the outputs $\boldsymbol{a}$ with initial state $\boldsymbol{x}(0)$ satisfy $\lim_{t\to\infty} \boldsymbol{a} \in E$. So we have the following Theorem 2:

**Theorem 2.** *The outputs of dynamic system defined in (5) are globally quasi-convergent.*

*Proof.* As the beginning, we define by $\dot{V}_b$ the derivative of $V_b$ with respect to time $t$, and the same as $\dot{\boldsymbol{a}}$, $\dot{\boldsymbol{x}}$. In order to prove the final result, we will firstly prove $\dot{V}_b \leq 0$. Since

$$\begin{aligned} \dot{V}_b &= \frac{dV_b}{d\boldsymbol{a}}\dot{\boldsymbol{a}} \\ &= (-\Phi^T \boldsymbol{y} + \Phi^T \Phi \boldsymbol{a} + \boldsymbol{x} - \boldsymbol{a})^T \dot{\boldsymbol{a}} \\ &= -\dot{\mathbf{x}}^T \dot{\boldsymbol{a}} = -\dot{\boldsymbol{x}}^T H \dot{\boldsymbol{x}} \end{aligned} \tag{10}$$

where $H$ is the Jacobian matrix of the thresholding function (6) and the second equation is due to (7). Then one can easily find that $H$ is with the following form:

$$H = \begin{pmatrix} H_1 & 0 & \cdots & 0 \\ 0 & H_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_l \end{pmatrix} \tag{11}$$

with $H_s \in \mathbb{R}^{m\times m}, s \in \{1,...,l\}$ the sub-matrix of $H$.

Thus (10) can be written as follows:

$$\dot{V}_b = -\dot{\mathbf{x}}^T H \dot{\mathbf{x}} = -\sum_{s\in\{1,...,l\}} \dot{\boldsymbol{x}}^T[s] H_s \dot{\boldsymbol{x}}[s] \tag{12}$$

Then, in order to prove the final result, the only problem is to verify $H_s$ is semi-definite. For the sake of simplicity, define $\rho = \|\boldsymbol{x}[s]\|_2$. The proof is straightforward as follows.

When $\rho < \lambda$, $H_s = \mathbf{0}$. When $\rho \geq \lambda$, the sub-matrix $H_s$ can be written as a sum of a diagonal matrix and a symmetric matrix:

$$H_s = \tilde{H}_s + \bar{H}_s \tag{13}$$

with

$$\tilde{H}_s = \frac{1}{\rho^3} \begin{pmatrix} \rho^3 - \lambda\rho^2 & 0 & \cdots & 0 \\ 0 & \rho^3 - \lambda\rho^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho^3 - \lambda\rho^2 \end{pmatrix}$$

and

$$\bar{H}_s = \frac{1}{\rho^3} \begin{pmatrix} \lambda x_{s,1}^2 & \lambda x_{s,1}x_{s,2} & \cdots & \lambda x_{s,1}x_{s,m} \\ \lambda x_{s,2}x_{s,1} & \lambda x_{s,2}^2 & \cdots & \lambda x_{s,2}x_{s,m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda x_{s,m}x_{s,1} & \lambda x_{s,m}x_{s,2} & \cdots & \lambda x_{s,m}^2 \end{pmatrix}.$$

It is obvious that $\tilde{H}_s$ and $\bar{H}_s$ are both semi-definite, and consequently, $H_s$ is also semi-definite.

Due to the fact that $\dot{V}_b \leq 0$, we can then conclude that $V_b(\cdot)$ is continuous and decreasing. Moreover, because $V_b$ is not negative according to (4), so it will converge to a constant value, i.e., $\lim_{t\to\infty} \dot{V}_b = 0$.

Since $H_s$ is semi-definite, so $\dot{V}_b$ will be zero only if $\|\dot{\boldsymbol{x}}[s]\|_2 = 0$ for any $s$. Then $\lim_{t\to\infty} \|\dot{\boldsymbol{a}}\|_2 = 0$ because of $\dot{\boldsymbol{a}} = H\dot{\boldsymbol{x}}$, i.e., $\boldsymbol{a}$ converges to the set $E = \{\boldsymbol{a} : s.t.\ \dot{\boldsymbol{a}} = 0\}$. So the outputs of D-BSS are globally quasi-convergent. □

### *3.3. Convergence of Support*

In the last subsection, the global convergence of the proposed system (5) has been addressed. In this subsection, we will further analyze the system to show a strengthened result on the support of $\boldsymbol{x}(t)$, denoted by $\Omega(t)$,

$$\Omega(t) \triangleq \{s : \|\boldsymbol{x}[s](t)\|_2 \geq \lambda, s \in \{1, ..., l\}\}$$

Under some mild conditions on the equilibrium point $\hat{\boldsymbol{x}}$ of system (5), it will be shown that this support $\Omega(t)$ will converge after finite switches. Thus we have the following theorem:

**Theorem 3.** *For the equilibrium point $\hat{\boldsymbol{x}}$ of system (5), if there exists a small constant $\varepsilon > 0$ such that, $\forall s \in \{1, ..., l\}$, one of the following two conditions is satisfied,*

$$\|\hat{\boldsymbol{x}}[s]\|_2 \geq \lambda + \varepsilon \tag{14}$$

$$\|\hat{\boldsymbol{x}}[s]\|_2 \leq \lambda - \varepsilon \tag{15}$$

*then there exists a constant $T > 0$, s.t. $\forall t > T$, $\Omega(t) = \Omega(T)$.*

*Proof.* Since $\lim_{t\to\infty} \boldsymbol{x}(t) = \hat{\boldsymbol{x}}$, there exists a real constant $J > 0$, for all $t > J$,

$$|\|\boldsymbol{x}[s](t)\|_2 - \|\hat{\boldsymbol{x}}[s]\|_2| \leq \|\boldsymbol{x}(t) - \hat{\boldsymbol{x}}\|_2 < \varepsilon$$

Below we will show that $\|\boldsymbol{x}[s](t)\|_2 - \lambda$ and $\|\hat{\boldsymbol{x}}[s]\|_2 - \lambda$ has the same sign for all $t > J$, the theorem is then proved and $T = J$.

From (14), we have $\|\hat{\boldsymbol{x}}[s]\|_2 > \lambda$, then

$$\begin{aligned}
\varepsilon &> |\|\boldsymbol{x}[s](t)\|_2 - \|\hat{\boldsymbol{x}}[s]\|_2| \\
&\geq \|\hat{\boldsymbol{x}}[s]\|_2 - \|\boldsymbol{x}[s](t)\|_2 \\
&\geq \lambda + \varepsilon - \|\boldsymbol{x}[s](t)\|_2
\end{aligned}$$

which implies, $\|\boldsymbol{x}[s](t)\|_2 > \lambda$.

The same argument can be applied for (15). More precisely, from (15) we have $\|\hat{\boldsymbol{x}}[s]\|_2 < \lambda$, this yields

$$\begin{aligned}
\|\boldsymbol{x}[s](t)\|_2 - \lambda &\leq \|\boldsymbol{x}[s](t)\|_2 - \varepsilon - \|\hat{\boldsymbol{x}}[s]\|_2 \\
&\leq |\|\boldsymbol{x}[s](t)\|_2 - \|\hat{\boldsymbol{x}}[s]\|_2| - \varepsilon < 0
\end{aligned}$$

So $\|\boldsymbol{x}[s](t)\|_2 < \lambda$. Finally we have proved that the support $\Omega(t)$ will converge after finite time. □

### *3.4. Convergence Rate*

In this part, we firstly introduce a useful function $\Theta(\cdot)$ about D-BSS, then give a theorem about the convergence rate of D-BSS.

For the proof below, we redefine the output and state variables in terms of the distance from the fixed point $\hat{\boldsymbol{x}}$ as

$$\begin{aligned} \boldsymbol{b}(t) &= \boldsymbol{x}(t) - \hat{\boldsymbol{x}} \\ \boldsymbol{e}(t) &= \boldsymbol{a}(t) - \hat{\boldsymbol{a}} = h(\boldsymbol{b}(t) + \hat{\boldsymbol{x}}) - h(\hat{\boldsymbol{x}}) \end{aligned} \tag{16}$$

Representing the relationship between thresholding function (6) and the distance variable (16), the function $\|\boldsymbol{a}[s]\|_2 = \Theta(\cdot)$ with respect to $\|\boldsymbol{x}[s]\|_2$ is deduced:

$$\Theta(\|\boldsymbol{x}[s]\|_2) = \begin{cases} 0, & \|\boldsymbol{x}[s]\|_2 < \lambda \\ \theta(\|\boldsymbol{x}[s]\|_2), & \|\boldsymbol{x}[s]\|_2 \geq \lambda \end{cases} \tag{17}$$

with $\theta(\|\boldsymbol{x}[s]\|_2) = \|\boldsymbol{x}[s]\|_2 - \lambda$, and $\theta'(\|\boldsymbol{x}[s]\|_2) = 1$. Similar with $\Omega(t)$ in section 3.3, we denote $\hat{\Omega}(t)$ referring to $\hat{\boldsymbol{x}}(t)$.

Denote by $\delta_k$ the smallest positive constant, with which $\Phi$ satisfies the block RIP [33] for all block sparse signals. With the concept of exponential convergence[18], we have the following theorem.

**Theorem 4.** *The dynamic system (5) has globally exponential convergent rate of* $\frac{1-\delta_k}{\tau}$*, with* $\delta_k < 1$.

The proof of Theorem 4 is similar to that in [34] except the conclusion (Lemma 2) which is necessary. Now we will prove Lemma 2 employing the function $\Theta(\cdot)$.

**Lemma 2.** *Denote by* $\bar{\Omega} = \Omega \cup \hat{\Omega}$*, then* $\|\boldsymbol{e}_{\bar{\Omega}}\|_2 \leq \|\boldsymbol{b}_{\bar{\Omega}}\|_2$

*Proof.* Firstly we choose a block from $\bar{\Omega}$, written as $\boldsymbol{b}[s]$, and we will prove that $\|\boldsymbol{e}[s]\|_2 \leq \|\boldsymbol{b}[s]\|_2$. Because $\bar{\Omega} = \Omega \cup \hat{\Omega}$, there are three cases:

Case 1: $\|\boldsymbol{x}[s]\|_2 \geq \lambda$ and $\|\hat{\boldsymbol{x}}[s]\|_2 < \lambda$. For the function $\theta(\cdot)$, according to the mean value theorem and $\theta(\lambda) = 0$, there exists $c \in (\lambda, \|\boldsymbol{x}[s]\|_2)$ such that

$$\begin{aligned} \|\boldsymbol{e}[s]\|_2 &= \|\boldsymbol{a}[s]\|_2 = \theta(\|\boldsymbol{x}[s]\|_2) \\ &= \theta'(c)(\|\hat{\boldsymbol{x}}[s] + \boldsymbol{b}[s]\|_2 - \lambda) \\ &\leq \|\hat{\boldsymbol{x}}[s]\|_2 + \|\boldsymbol{b}[s]\|_2 - \lambda \leq \|\boldsymbol{b}[s]\|_2 \end{aligned}$$

Case 2: $\|\boldsymbol{x}[s]\|_2 < \lambda$ and $\|\hat{\boldsymbol{x}}[s]\|_2 \geq \lambda$. For the function $\theta(\cdot)$, according to the mean

value theorem, there exists $c \in (\lambda, \|\hat{\boldsymbol{x}}[s]\|_2)$ such that

$$\begin{aligned}\|\boldsymbol{e}[s]\|_2 &= \|\hat{\boldsymbol{a}}[s]\|_2 = \theta(\|\hat{\boldsymbol{x}}[s]\|_2) \\ &= \theta'(c)(\|\boldsymbol{x}[s] - \boldsymbol{b}[s]\|_2 - \lambda) \\ &\leq \|\boldsymbol{x}[s]\|_2 + \|\boldsymbol{b}[s]\|_2 - \lambda \leq \|\boldsymbol{b}[s]\|_2\end{aligned}$$

Case 3: $\|\boldsymbol{x}[s]\|_2 \geq \lambda$ and $\|\hat{\boldsymbol{x}}[s]\|_2 \geq \lambda$. For the function $g(\cdot)$, according to the mean value theorem, there exists $c \in (\hat{\boldsymbol{x}}_n, \boldsymbol{x}_n)$ such that

$$\begin{aligned}|\boldsymbol{e}_n| &= |\boldsymbol{a}_n - \hat{\boldsymbol{a}}_n| = |g(\boldsymbol{x}_n) - g(\hat{\boldsymbol{x}}_n)| \\ &= |g'(c)(\boldsymbol{x}_n - \hat{\boldsymbol{x}}_n)| \\ &= |g'(c)\boldsymbol{b}_n|\end{aligned}$$

For $\forall t \geq 0, \forall n = 1, ..., N$, from (6),

$$g'(\boldsymbol{x}_n) = 1 - \frac{\lambda}{\|\boldsymbol{x}[s]\|_2} + \frac{\lambda \boldsymbol{x}_n^2}{\|\boldsymbol{x}[s]\|_2 \|\boldsymbol{x}[s]\|_2^2} \leq 1$$

so $|\boldsymbol{e}_n| \leq |\boldsymbol{b}_n|$, and then $\|\boldsymbol{e}[s]\|_2 \leq \|\boldsymbol{b}[s]\|_2$.

In a word, for any block with index set in $\bar{\Omega}$, $\|\boldsymbol{e}[s]\|_2 \leq \|\boldsymbol{b}[s]\|_2$. Thus $\|\boldsymbol{e}_{\bar{\Omega}}\|_2 \leq \|\boldsymbol{b}_{\bar{\Omega}}\|_2$. □

**Remark 1.** *In addition, the convergence rate in [34] is* $\frac{1-\beta\delta}{\tau}$ *with* $\delta$ *bigger than the parameter* $\delta_k$ *[33], and* $\beta = 1$. *So* $\frac{1-\delta_k}{\tau}$ *is bigger than* $\frac{1-\beta\delta}{\tau}$, *which means that the D-BSS will improve the convergence rate. The experiments about the comparison of convergence rate will be shown below.*

## 4. Simulations

In this section, we use simulations with respect to the support, amplitude and convergence of the recovered signal to verify the previous theoretical results. A sparse signal vector $\boldsymbol{a_0} \in \mathbb{R}^N$ was generated with two nonzero "blocks" based on the model defined before. The amplitude of each nonzero block is drawn from normalized Gaussian distribution. The measurement vector is $\boldsymbol{y} = \Phi \boldsymbol{a_0} + \boldsymbol{\varepsilon}$, where the $\Phi$ is generated

randomly with entries drawn from normalized Gaussian distribution, the value of $M$ and $N$ will change with different purpose. For the dynamic system algorithm LCA and D-BSS, to be fair with LCA, we set the threshold $\lambda = 0.2$ in LCA, and $\lambda = 0.2\sqrt{m}$ in D-BSS, where $m$ is the block length. So the sparse level is $k = 3m$, and it is $m$ that represents the sparse level afterwards in this paper. At last, we set $\tau = 1000$, and the initial condition to $\boldsymbol{x}(0) = \mathbf{0}$. The normalized mean square error (NMSE) and reconstructed error rate of nonzero locations (ERZ) are respectively exploited to measure the algorithms' performance. Denote $\hat{\boldsymbol{a}}$ the reconstruction of original $\boldsymbol{a}_0$, then the NMSE [3] can be computed as

$$\text{NMSE} = 20\log_{10}\frac{\|\hat{\boldsymbol{a}} - \boldsymbol{a}_0\|}{\|\boldsymbol{a}_0\|}$$

On the other hand, define $T(\boldsymbol{a}_0)$ the location set of zero elements of $\boldsymbol{a}_0$, then the reconstructed ERZ [35] is defined as

$$\text{ERZ} = \frac{|T(\boldsymbol{a}_0)\backslash T(\hat{\boldsymbol{a}})|}{n}$$

Consequently, the better algorithm shall have larger SNR but smaller SER.

### *4.1. Convergence*

In this subsection, theoretical properties addressed in this paper will be verified by simulations. At first, setting the block size $m = 3$ and with only $M = 15$ measurements, the reconstruction of $\boldsymbol{a}_0$ via LCA and D-BSS are respectively given in Fig. 1. Apparently, the D-BSS performs better than LCA.

The convergence property of the proposed D-BSS is also verified. In this place, the sparse signal $\boldsymbol{a}_0$ is randomly generated with length $N = 1024$ and 2 nonzero blocks with $m = 10$. Then $M = 200$ measurements are collected after contamination of noises with SNR=25 dB. In order to evaluate the convergence property, four noes of $\boldsymbol{a}_0$ are selected and the corresponding evolution curves with respective to time $t$ are plotted in Fig. 2. One can easily find that with consideration of block structures the convergence speed and the reconstruction accuracy are both improved by D-BSS.

Moreover, the convergence property with respect to different sparsity level and different signal length is evaluated. In this simulation, we first fix the number of measurements $M = 200$, measurement SNR=30 dB and the number of blocks to 3 for sparse

signals $\boldsymbol{a}_0$. Then ranging the sparse level $m$ from 2 to 64, one can obtain the convergence time by evaluating the NMSE. In this place, the convergence is determined if the difference of NMSE with a short period $\Delta t = 1e-3s$ is less than $1e-3$. The convergence time for both LCA and D-BSS with respect to different sparse level is shown in Fig. 3(a). Apparently, the proposed D-BSS converges faster than LCA. On the other hand, the convergence time with respect to different signal length is also given in Fig. 3(b). Similarly, the result conveys that D-BSS converges much faster than LCA.

### *4.2. Comparison to discrete algorithms*

In this subsection, comparisons are made to the typical discrete algorithms, including ISTB [12], BOMP [10] and block-CoSaMP [26]. It is worth mentioning that BOMP and block-CoSaMP are also block sparse recovery algorithm, while ISTB isn't. At first, the robustness to noises is illustrated, where the parameter setting of this simulation is $M = 200, N = 400, m = 25$, and then range the $SNR$ from 15dB to 40dB. Similarly, the NMSE and EZR of the estimation can be obtained by averaging over 50 trials and the results are shown in Fig.4. We can find that the error rate of the zero locations is much less for D-BSS than the other algorithms, as well as the NMSE.

The second simulation evaluates the estimation accuracy with respect to different sparsity levels. In order to observe significant influence of the sparsity levels, we generate a longer sparse signal vector $\boldsymbol{a_0}$ with $N = 400$ and $M = 200$. In this simulation, we still focus on the support and the amplitude respectively. The evaluation standards are the same as the first simulation. We set $SNR = 25$dB, and range the sparsity level $m$ from 5 to 40. In Fig.5(a), the error rate of estimation with respect to sparsity levels is shown. It is apparent that the performance of D-BSS is superior to the others except the block-CoSaMP. As a result, the D-BSS achieves excellent performance.

### *4.3. Application to DOA*

In this section, we apply the proposed algorithm D-BSS to Direction Of Arrival (DOA) estimation. The DOA estimation algorithm based on sparsity is a theory of sparse signal recovery in nature. It converts the estimation of azimuth angle to the estimation of non-zero entries in the sparse signal, so the number of signal source is

not necessary to be known in advance. Take specific algorithms to recover the sparse signal, we can obtain the estimation of azimuth value by the non-zero entries observed from the position of the peak, and obtain the number of signal sources by the number of spikes.

There are many factors that can affect the estimation results. As the specific details of DOA estimation are not the focus of this paper, here we only explain how we did in this simulation. We choose the classical algorithms Capon and MUSIC as the competitors. The parameters in this simulation are set as follows. The fixed signal resource number is 30 distributed in two consecutive blocks, as shown in Fig. 6. On the other hand, the array element number is 40, the snapshot number is 500 and noises are added with SNR=$-10$dB. Then Capon, MUSIC and D-BSS are respectively exploited to estimate the DOA, and the results are shown in Fig. 6. Apparently, we can find that D-BSS performs better than the other two algorithms. In addition, since the D-BSS method is a continuous approach, it is more nature to exploit D-BSS in the context of DOA where the target is moving continuously and real-time processing is required. Consequently, it is very attractive to exploit D-BSS to cope with such kind of problems, for instance, video processing.

## 5. Conclusion

This paper mainly puts forward a signal recovery algorithm D-BSS for block sparse signals using a dynamic system. The most important theoretical result of D-BSS is that it is equivalent to the traditional problem in (4). Furthermore, D-BSS has globally quasi-convergent outputs and has an exponential convergence rate that is faster than LCA, and its support set can be obtained in limited time. The simulation results and analysis show that both the support and the entries of D-BSS for sparse recovery have a better results. The real application on DOA estimation indicates that it is convincing for future engineering applications, for example target tracking and so on.

## References


[1] D. L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006) 1289–1306.

[2] M. Elad, M. A. T. Figueiredo, Y. Ma, On the role of sparse and redundant representations in image processing, Proc. IEEE 98 (6) (2010) 972–982.

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

[4] Y. Luo, L. Tongliang, D. Tao, C. Xu, Decomposition-based transfer distance metric learning for image classification, IEEE Transactions on Image Processing 23 (9) (2014) 3789–3801.

[5] Y. Luo, Y. Wen, D. Tao, Large margin multi-modal multi-task feature extraction for image classification, IEEE Transactions on Image Processing 25 (1) (2016) 414–427.

[6] W. Liu, D. Tao, J. Cheng, Y. Tang, Multiview hessian discriminative sparse coding for image annotation, Computer Vision and Image Understanding 118 (2014) 50–60.

[7] E. Candes, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory 51 (12) (2005) 4203–4215.

[8] T. Blumensath, M. E. Davies, Sampling theorems for signals from the union of finite-dimensional linear subspaces, IEEE Trans. Inf. Theory 4 (55) (2009) 1872–1882.

[9] I. Daubechies, R. DeVore, M. Fornasier, C. S. Güntürk, Iteratively reweighted least squares minimization for sparse recovery, Communications on Pure and Applied Mathematics 63 (1) (2010) 1–38.

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

[11] P. L. Combettes, J.-C. Pesquet, Proximal splitting methods in signal processing, in: Fixed-point algorithms for inverse problems in science and engineering, Springer, 2011, pp. 185–212.

[12] I. Daubechies, M. Defrise, C. D. Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Communications on pure and applied mathematics 57 (11) (2004) 1413–1457.

[13] M. S. Asif, J. Romberg, Sparse recovery of streaming signals using-homotopy, IEEE Transactions on Signal Processing 62 (16) (2014) 4209–4223.

[14] R. Baraniuk, P. Steeghs, Compressive radar imaging, IEEE Radar Conference. IEEE, apr (2007) 128–133.

[15] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE T. Pattern. Anal. 31 (2) (2009) 210–227.

[16] X. Xu, X. Wei, Z. Ye, Doa estimation based on sparse signal recovery utilizing weighted-norm penalty, IEEE Signal Processing Letters 19 (3) (2012) 155–158.

[17] C. J. Rozell, D. H. Johnson, R. G. Baraniuk, B. A. Olshausen, Sparse coding via thresholding and local competition in neural circuits, Neural computation 20 (10) (2008) 2526–2563.

[18] A. Balavoine, J. Romberg, C. Rozell, Convergence and rate analysis of neural networks for sparse approximation, IEEE Transactions on Neural Networks and Learning Systems 23 (9) (2012) 1377–1389.

[19] A. Balavoine, C. J. Rozell, J. Romberg, Discrete and continuous-time soft-thresholding for dynamic signal recovery, IEEE Transactions on Signal Processing 63 (12) (2015) 3165–3176.

[20] M. Mishali, Y. C. Eldar, Blind multi–band signal reconstruction: Compressed sensing for analog signals, IEEE Trans. Signal Process. 57 (3) (2009) 993–1009.

[21] M. Mishali, Y. C. Eldar, From theory to practice: Sub-nyquist sampling of sparse wideband analog signals, IEEE J. Sel. Topics Signal Process. 4 (2) (2009) 375–391.

[22] H. J. Landau, Necessary density conditions for sampling and interpolation of certain entire functions, Acta Math. 117 (1) (1967) 37–52.

[23] F. Parvaresh, H. Vikalo, S. Misra, B. Hassibi, Recovering sparse signals using sparse measurement matrices in compressed dna microarrays, IEEE J. Sel. Topics Signal Process. 2 (3) (2008) 275–285.

[24] S. F. Cotter, B. D. Rao, Sparse channel estimation via matching pursuit with application to equalization, IEEE Trans. Commun. 50 (3) (2002) 374–377.

[25] D. Malioutov, M. Cetin, A. Willsky, Sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. Signal Process. 53 (8) (2005) 3010–3022.

[26] R. G. Baraniuk, V. Cevher, M. Duarte, C. Hegde, Model-based compressive sensing, IEEE Transactions on Information Theory 56 (4) (2010) 1982–2001.

[27] L. Jacob, G. Obozinski, J. P. Vert, Group lasso with overlap and graph lasso, in: ICML, ACM, 2009, pp. 433–440.

[28] L. Yu, J. Barbot, G. Zheng, H. Sun, Bayesian compressive sensing for cluster structured sparse signals, Signal Processing 92 (1) (2012) 259–269.

[29] L. Yu, H. Sun, G. Zheng, J. Barbot, Model based bayesian compressive sensing via local beta process, Signal Processing 108 (3) (2015) 259–271.

[30] M. F. Duarte, Y. C. Eldar, Structured compressed sensing: From theory to applications, IEEE Transactions On Signal Processing 59 (9) (2011) 4053–4085.

[31] A. Gramfort, M. Kowalski, M. Hämäläinen, Mixed-norm estimates for the m/eeg inverse problem using accelerated gradient methods, Physics in medicine and biology 57 (7) (2012) 1937.

[32] A. Gramfort, D. Strohmeier, J. Haueisen, M. S. Hämäläinen, M. Kowalski, Time-frequency mixed-norm estimates: Sparse m/eeg imaging with non-stationary source activations, NeuroImage 70 (2013) 410–422.

[33] Y. Eldar, M. Mishali, Robust recovery of signals from a structured union of subspaces, IEEE Trans. Inf. Theory 55 (11) (2009) 5302–5316.

[34] A. Balavoine, J. Romberg, C. J. Rozell, Correction to "convergence and rate analysis of neural networks for sparse approximation", IEEE Transactions on Neural Networks and Learning Systems 25 (8) (2014) 1595–1596.

[35] K.-J. Lee, T.-Y. Hsieh, M. Breuer, et al., A novel test methodology based on error-rate to support error-tolerance, in: IEEE International Test Conference, 2005, 2005.

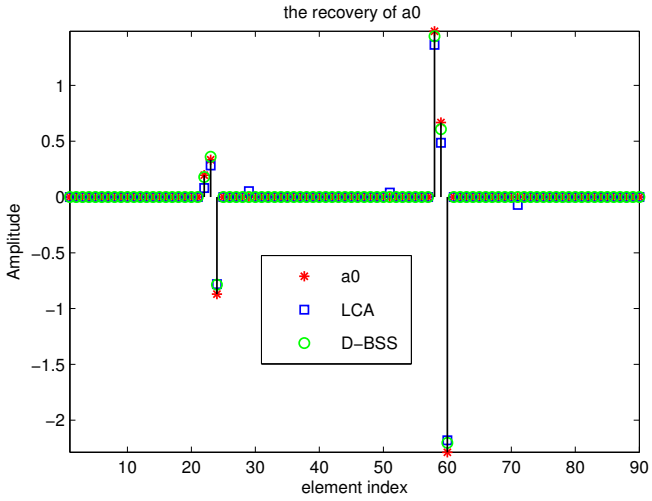
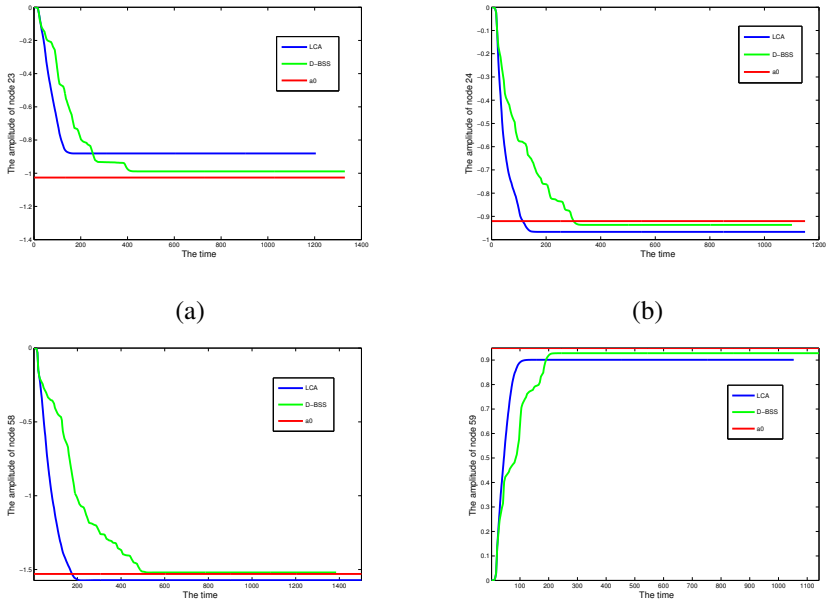


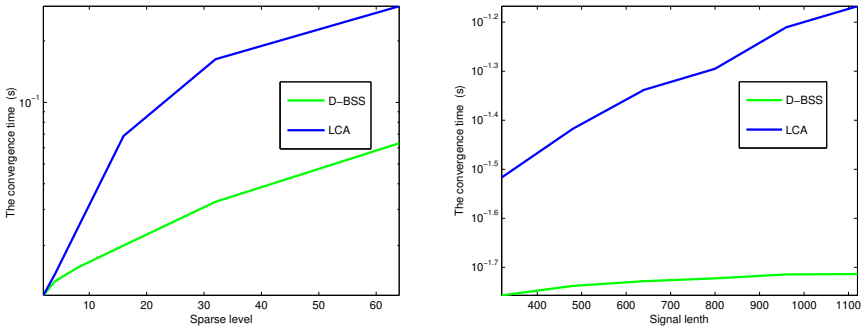

Figure 1: The recovered sparse signal of thte original $\boldsymbol{a_0}$ whose sparsity level m=3 via LCA and our proposed dynamic system D-BSS with measurement SNR= 25dB.




(a) (b)

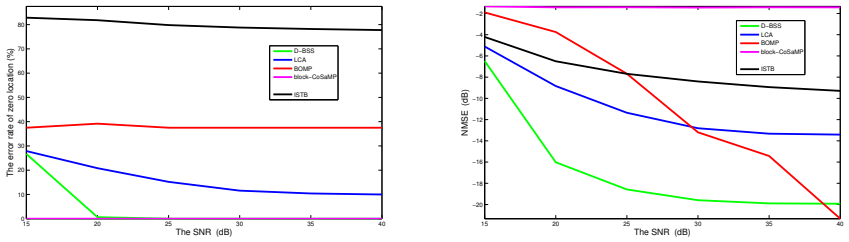
(c) (d)

Figure 2: The evolution process of some single nodes.



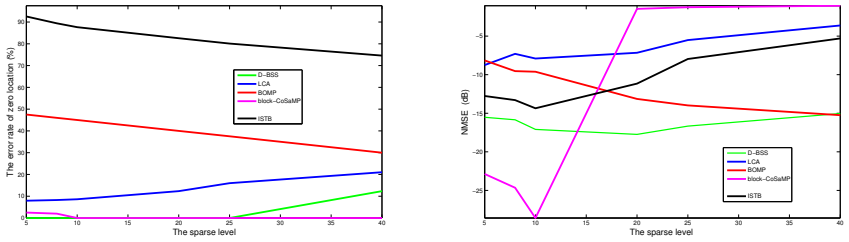

(a) (b)

Figure 3: The converge time with respect to (a) different sparsity levels and (b) different signal length.




(a) (b)

Figure 4: The estimation accuracy of support (a) and amplitude (b) with respect to different SNR levels.




(a) (b)

Figure 5: The estimation accuracy of support (a) and amplitude (b) with respect to different sparsity levels.

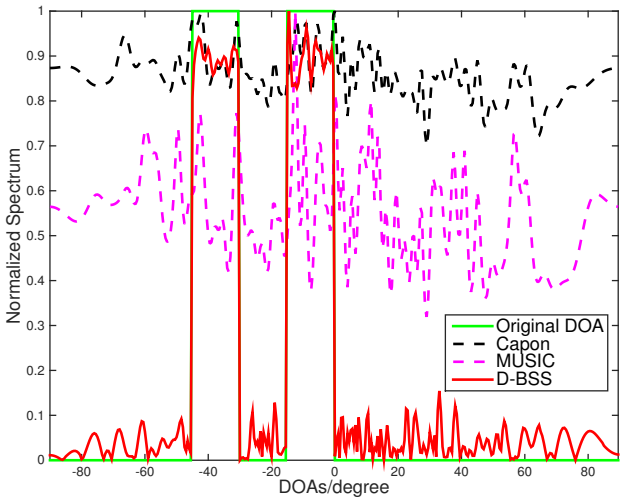



Figure 6: The results of DOA estimation with different methods.