Block-Sparsity Log-Sum-Induced Adaptive Filter for Cluster Sparse System Identification

AIHUA ZHANG¹, (Member, IEEE), PENGCHENG LIU², (Member, IEEE), JUN SUN³, AND BING NING¹, (Member, IEEE)

¹School of Electronic and Information Engineering, Zhongyuan University of Technology, Zhengzhou 450007, China
²Department of Computer Science, University of York, York YO10 5GH, U.K.
³School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China

Corresponding author: Aihua Zhang (zah@zut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61501530 and Grant 61803396, Foundation of Henan Educational Committee under Grant 21A510015.

ABSTRACT In this work, an effective adaptive block-sparsity log-sum least mean square (BSLS-LMS) algorithm is proposed to improve the convergence performance of cluster sparse system identification. The main idea of the proposed scheme is to add a new block-sparsity induced term into the cost function of the LMS algorithm. We utilize the $l_1$ norm of adaptive tap weights and log-sum as a mixed constraint, via optimizing the cost function through the gradient descent method, the proposed adaptive filtering method can iteratively move the identified signals towards the optimal solutions, and finally identify the unknown system accurately. The cluster-sparse system response, with block length and arbitrary average sparsity, is generated by a Markov-Gaussian (M-G) model. For the white Gaussian input data, the theoretical formulas of the steady-state mis-adjustment and convergence behaviors of the BSLS-LMS are derived in a general sparse system and a block sparse system, respectively. Numerical experiments demonstrate that the proposed adaptive BSLS-LMS algorithm achieves much better convergence behavior than conventional sparse signal recovery solutions. The experimental study also verifies the consistency between the simulation results and the theoretical analysis.

INDEX TERMS Block-sparse system, channel estimation, identification, log-sum, LMS, performance analysis.

I. INTRODUCTION

Signal reconstruction technologies have attracted much attention in the fields of channel estimation, image recovery, sparse unknown system identification [1]–[6]. In many cases, sparse unknown systems have only a few nonzero entries, and these limited nonzero or large coefficients response appear independently in different locations over a long pulse. Different from the general sparse signal, another type of sparse signal is called the block-structured sparse signal, and it has a special cluster structure in the form of non-zero coefficients [7]–[10]. Nonzero coefficients can be located randomly in general sparse systems, but for a block sparse system, the impulse response typically consists of one or several clusters in which nonzero coefficients gather. Multiple-input multiple-output (MIMO) communication networks, the satellite-linked path and other practical applications are typical examples of multi-clustering sparse structures.

A number of block-sparsity signal recovery algorithms based on conventional sparse signal recovery methods have been designed. Convex optimization algorithms have been applied to block sparse signal reconstruction and studied [10]–[12]. Mixed $l_{2,1}$ program block-sparse signal recovery was proposed in [10], [11]. The block version of the zero-point attracting projection algorithm employing an approximate $l_{2,0}$ norm as the cost function was proposed in [12]. Convex relaxations equivalent to the original non-convex formulation using the $l_{q,1}$ norm were studied in [13]. Greedy pursuit algorithms have been proposed for block-sparse systems [14], [15]. The block-version of the orthogonal matching pursuit (BOMP) algorithm that identifies block-sparse signals successfully via a mixed-optimization approach was investigated in [11]. Model-based compressive sampling matching pursuit (CoSaMP) was proposed in [15], and Bayesian Compressed Sensing framework-based...
algorithms were developed in [16]–[18]. The Stochastic Taps Normalized Least Mean Square (STNLMS) [19], the Select and Queue with a Constraint (SELQUE) [20] and improved M-SELQUE algorithm [21] were applied to identify an unknown system by estimating the scattered region. The aforementioned algorithms can reconstruct unknown block sparse signals using prior knowledge of the block partition and cluster structure. To recover a block signal accurately, the block structure is fairly important. However, the key prior knowledge of the block partition of sparse coefficients is not always practically available. In fact, the locations and group sizes of the sparse clusters in an unknown system are random and entirely unknown.

Recently, block sparse least mean squares adaptive filtering has received extensive attention [14], [22]–[26]. These algorithms can recover the block-sparse signals of unknown clustering sparse systems without prior knowledge of the block-structured non-zero coefficients. The sparse adaptive algorithms have been shown to outperform other algorithms due to their low computational complexity and robustness against noise. To accelerate the convergence process of coefficients, zero-attracting block-sparsity induced LMS algorithms have been proposed and applied to recover block-sparse impulse responses in previous studies [22]–[24]. Since most of the partitions in block-sparse systems are zero-coefficient groups, zero-attraction algorithms generally converge faster than other algorithms. The main idea of zero-attracting block-sparsity-aware algorithms is to insert the constraint of block sparsity into the cost function of the standard LMS algorithm, such as the \( l_0 \) norm LMS and block-sparse LMS algorithms [25], a mixed \( l_{2,1} \) norm, or an approximate mixed \( l_{2,0} \) norm of the uniformly divided filter tap-weight vector [26].

It is reported that log-sum minimization requires fewer measurements to recover sparse signals than \( l_1 \) minimization [27]. This advantage of using the log-sum penalty function has been proved in the Reweighted Zero-Attracting LMS (RZA-LMS) algorithm [28]. Inspired by the RZA-LMS and new developments in compressive sensing [29], [30], we propose a new log-sum penalty LMS to identify unknown sparse systems and theoretically verify the adoption of the proposed new penalty as an alternative sparsity-aware function. In the proposed novel block sparse LMS algorithm, we insert a mixed \( l_1 \) norm and log-sum penalty for the coefficients of the unknown system into the cost function, so that the attractor selectively promotes the zero taps instead of uniformly promoting the zeros on all the taps. In this paper, the M-G model [25] is applied to generate and describe block-sparse systems. Then, the optimal group size of the sparse-structure is provided based on M-G model and the proposed algorithm. It is demonstrated that the behavior of the BSLS-LMS algorithm is better than that of the block-sparse LMS (BS-LMS) when the group size is properly selected. The convergence behaviors of the algorithm are simulated by using the mean square deviation (MSD). The experimental results show that the proposed BSLS-LMS algorithm performs better than the BS-LMS and other block sparse filtering algorithms. In addition, when the number of blocks in the system increases, the block-sparsity log-sum LMS demonstrates robustness and the performance loss is very small compared with the standard. The main contributions of this paper are summarized as follows:

- We propose a new block-sparse system identification algorithm to reconstruct the system response. We present theoretical expressions of the steady-state mis-adjustment and transient convergence behavior of the proposed algorithm with an appropriate group partition size for white Gaussian input data. Then, we theoretically demonstrate that the BSLS-LMS has much better convergence behavior than previous methods [19]–[21], [25].

- We propose a new cost function by utilizing the \( l_1 \) norm of the adaptive tap weights and the log-sum as a mixed penalty. The proposed adaptive filtering method can iteratively move the identified signals towards the optimal solutions, and finally identify the unknown system accurately. The proposed adaptive algorithm has low computational complexity and has much better convergence behavior than conventional sparse signal recovery solutions.

The rest of the paper is organized as follows. In section 2, log-sum block sparse algorithm is proposed to improve reconstructor performance with analysis of the computational complexity provided. In section 3, the theoretical results of the steady-state performance and convergence of the BSLS-LMS are analyzed. In section 4, several experiments are implemented to verify the theoretical derivation, and the effectiveness of the BSLS-LMS is evaluated by comparing it with existing block sparse algorithms. Finally, Section 5 concludes the proposed algorithm.

II. BLOCK-SPARSITY LOG-SUM LEAST MEAN SQUARES ALGORITHM

The unknown system coefficients vector to be identified is denoted as \( s = [s_1, s_2, \ldots, s_L]^T \), and the input signal at time instant \( n \) is \( x_n = [x_n, x_{n-1}, \ldots, x_{n-L+1}]^T \), where \( L \) is the length of the unknown system, both \( s \) and \( x_n \) are real-valued vectors, and \( (\cdot)^T \) represents the transposition. The output signal of the unknown system at time instant \( n \) is defined as

\[
d_n = x_n^T s + v_n, \tag{1}
\]

where \( v_n \) is the additive white Gaussian noise of the unknown system.

We denote \( \hat{s} \) as the adaptive tap-weights. In other word, \( \hat{s} \) is the reconstructed signal of the unknown coefficients vector \( s \). As a result, the estimated error at time instant \( n \) between the desired output of the unknown system and the output of the adaptive filter is

\[
e_n = d_n - x_n^T \hat{s}_n, \tag{2}
\]

where \( \hat{s}_n = [\hat{s}_{1,n}, \hat{s}_{2,n}, \ldots, \hat{s}_{L,n}]^T \) represents the adaptive tap-weights vector. In many practical scenarios, the nonzero
coefficients of the unknown systems appear in the form of clusters rather than random distribution. To define block sparsity, we consider $s$ as a concatenation of blocks, and throughout this paper we assume that there are $M$ block groups in the unknown system and the group partition size length of every block is $G$ with $G < L$. The block sparse signal is represented as

$$\hat{s} = [\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_G, \hat{s}_{G+1}, \hat{s}_{G+2}, \ldots, \hat{s}_{2G}]$$

$$\vdots$$

$$\hat{s}_{(M-1)G+1}, \hat{s}_{(M-1)G+2}, \ldots, \hat{s}_{LG}]^T, \quad (3)$$

We adopt the mixed $l_1$ norm and log-sum constraint to evaluate the block sparsity of $\hat{s} = [\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_L]^T$ as

$$L \left[ \| \hat{s} \|_1 \right] = L \left[ \begin{bmatrix} w_1^T \\ \vdots \\ w_M^T \end{bmatrix} \right] = \sum_{i=1}^L \log \left( \varepsilon + \frac{|w_i|}{\eta} \right), \quad (4)$$

where $w_j = [\hat{s}_{(i-1)G+1}, \hat{s}_{(i-1)G+2}, \ldots, \hat{s}_{iG}]^T$ denotes the $i$th group of $\hat{s}$. We assume that the signals in vector $\hat{s}$ can always be evenly divided into $M$ groups with a length of $G$ by adding zero taps at the end of $\hat{s}$. Aiming to study the unknown systems by utilizing the block-sparsity, we propose a new cost function that combines the expectation of the estimated error, the mixed $l_1$ norm, the log-sum constraint of the tap-weighted vector and the balance parameters, i.e.,

$$J_n = E \left[ |e_n|^2 \right] + \gamma \sum_{i=1}^M \log \left( \varepsilon + \frac{|w_i|}{\eta} \right), \quad (5)$$

where $\varepsilon (0 < \varepsilon < 1)$ and $\eta (0 < \eta < 1)$ are positive constants that guarantee that the cost function is well-defined and that better sense the sparsity of the block-sparse signal, respectively. The regularization parameter $\gamma$ is a positive factor that regulates the strength between the block-sparsity penalty and the mean squared error, and the parameter should be appropriately chosen to ensure satisfactory performance [24]. It is noted that this idea differs from RZA-LMS, as the constraint function in RZA-LMS is $\gamma \sum_{i=1}^M \log \left( 1 + |\hat{s}(i)| / \eta \right)$, while we define $0 < \varepsilon < 1$ as a positive constant. By using gradient descent, the new recursion formula of the adaptive tap-weights can be defined as

$$\hat{s}_{n+1} = \hat{s}_n + \mu e_n x_n + \rho f(\hat{s}_n), \quad (6)$$

where $\eta' = 1 / \eta, \mu$ is the step-size, and $\rho = \frac{1}{L} \mu \gamma \eta'$ denotes a zero attraction parameter that balances the strength of the block-sparse penalty term for a given step-size and group zero attraction

$$f(\hat{s}) = \left[ f_1(\hat{s}), f_2(\hat{s}), \ldots, f_L(\hat{s}) \right]^T,$$

and

$$f_i(\hat{s}) = - \frac{\text{sgn}(\hat{s}_i)}{\varepsilon + \eta' |w_i|/G}, \quad (7)$$

where $\text{sgn}[\cdot]$ is a sign function.

### Table 1. The proposed BSLS-LMS algorithm.

| Input: $x_n, d_n, n = 0, 1, 2, \ldots, L, G, \mu, \gamma, \rho, \varepsilon, \eta$ |
| Output: $\{\hat{s}_n\}_{n=0, 1, 2, \ldots}$ |
| 1. Initialize $\hat{s}_0 = 0, M = L/G$. |
| 2. for $n = 0, 1, 2, \ldots$, |
| 3. Calculate instantaneous error vector $e_n$ |
| 4. $e_n = d_n - x_n^T \hat{s}_n$ |
| 5. for $i = 1, 2, \ldots, M$, |
| 6. Update step-size and balance factor |
| 7. end for |
| 8. end for |

### Table 2. Computational costs.

| Algorithm | MAC | Mul | Div | Com | SR |
|-----------|-----|-----|-----|-----|----|
| LMS       | $2L$ | –   | –   | –   | –  |
| $l_0$-LMS | $4L$ | $L$ | $2L$ | –   | –  |
| BS-LMS    | $5L$ | $L/G$ | $L/G$ | $L/G$ | – |
| BSLS-LMS  | $4L$ | $L/G$ | –   | –   | –  |

Mul: Multiplication, Div: Division, Com: Comparison, SR: Square Root

The proposed algorithm and comparison of the computational cost are detailed in Table 1 and Table 2, respectively. 

According to the recovery process of the log-sum penalty LMS as summarized in Table 1, the computational complexity of this new algorithm is $O(L)$. The detailed calculations per iteration is listed and compared with those of the $l_0$-LMS and BS-LMS. As shown in Table 2, the computations amount of the BSLS-LMS is less than that of the BS-LMS [17]. In practical applications, since unknown systems vary slowly, partial update can be used for saving calculation.

In fact, literature [27], [28] shows that when $\varepsilon = 0$, the log-sum penalty is virtually the same as the $l_0$-norm. Therefore, it is plausible that when $\varepsilon$ is very small, the behavior of the log-sum penalty function, as shown in equation (6) is similar to that of the $l_0$-norm. Moreover, we define $\eta$ as $0 < \eta < 1$ which guarantees better convergence performance of the log-sum penalty block LMS than the approximate $l_0$-norm LMS.

### III. Performance Analysis of the BSLS-LMS Algorithm in General Sparse Systems

In the proposed BSLS-LMS algorithm, the sparse constraint term adopts a mixed $l_1$ norm and has a modified log-sum penalty on the impulse response coefficients of the unknown system, which is similar to the $l_0$-pseudo-norm of the coefficient vector and forces the solution of the proposed algorithm to be sparse. The basic assumptions of the system are as
follows: (a) the input signal $x_n$ follows i.i.d. zero-mean Gaussian distribution, (b) the input vector $x_n$, tap-weights $\hat{s}_n$, and additive white noise $v_n$ are independent of each other, and (c) the variance of $x_n$ is $\sigma_x^2$, and all the tap-weights $\hat{s}_n$ can be modeled by Gaussian variables. To provide a theoretical basis for this novel induced term as an alternative sparsity-aware function, we analyze the mean and mean-squared performances of the BSLS-LMS based on the above assumptions in this section.

The rationale of making the above assumptions is the following: in the standard LMS, the tap-weights of $\hat{s}_{j,n}$ can converge to their optimal values uniformly under an i.i.d. Gaussian input signal. In the proposed block sparse log-sum LMS, due to the group zero-attracting in (5), uniform convergence exists globally. Therefore, the strength of the temporary tap-weight is very close to that of the unknown coefficient in the group. In fact, numerical experiments have verified that this assumption is valid.

A. MEAN PERFORMANCE BASED ON THE PROPOSED ALGORITHM

The misalignment vector of identification is defined as

$$\begin{align*}
\mathbf{r}_n &= \hat{s}_n - s_n.
\end{align*}$$

Combining (1), (2), and (6), the update equation of the misalignment vector can be formulated by

$$\begin{align*}
\mathbf{r}_{n+1} &= \left( I - \mu \mathbf{x}_n \mathbf{x}_n^T \right) \mathbf{r}_n + \mu \mathbf{v}_n \mathbf{x}_n + \rho f (\hat{s}_n).
\end{align*}$$

Taking the expectation for (9) and using the assumption (c), when $n$ goes to infinity, $E (\mathbf{r}_n)$ and $E [f (\hat{s}_n)]$ converge, as one has

$$\begin{align*}
E [\mathbf{r}_n] &= \lim_{n \to \infty} \frac{\rho}{\mu \sigma_x^2} E [f (\hat{s}_n)].
\end{align*}$$

The upper bound of the derivation is

$$\begin{align*}
E [\mathbf{r}_n^T] &= \left[ \begin{array}{c} \rho \eta \\ \mu \sigma_x^2 \end{array} \right] E \left[ f (\hat{s}_n^T) \right] \\
&\leq \left[ \begin{array}{c} \rho \eta \\ \mu (\epsilon + 1) \sigma_x^2 \end{array} \right] I,
\end{align*}$$

where $I$ is the vector with all one entries. This means that the proposed algorithm has a stability condition for the coefficient misalignment vector convergence.

B. MEAN SQUARED STEADY STATE ANALYSIS OF THE BSLS-LMS ALGORITHM

In this section, we derive the steady-state mean squared deviation (MSD) between the original signal and the estimated signal, and then, we deduce a criterion for zero attraction parameter selection for the proposed algorithm to outperform other block sparse algorithms.

Multiply both sides of (9) by their respective transposes,

$$\begin{align*}
r_{n+1}^T r_{n+1} &= \left( I - \mu \mathbf{x}_n \mathbf{x}_n^T \right) \mathbf{r}_n^T + \mu \mathbf{v}_n \mathbf{x}_n + \rho f (\hat{s}_n)
\end{align*}$$

Using the fact that $D_n = tr [S_n]$ and taking the trace on both sides of (15), one can know

$$\begin{align*}
D_{n+1} &= \left[ 1 - 2 \mu \sigma_x^2 + (L + 2) \mu^2 \sigma_x^4 \right] \cdot D_n
+ 2 (1 - \mu \sigma_x^2) \rho E [r_{n+1} f (\hat{s}_n)]
\end{align*}$$

From (16), we obtain the condition of convergence as follows

$$\begin{align*}
\left| 1 - 2 \mu \sigma_x^2 + (L + 2) \mu^2 \sigma_x^4 \right| < 1.
\end{align*}$$

Using (16) and considering the $k$th diagonal element, we obtain (17), as shown at the bottom of the next page.

To derive $E [r_{\hat{s}_k} f (\hat{s}_k)]$, we classify the index set $\{ k = 0, 1, \cdots, L - 1 \}$ into two groups according to the unknown coefficients. We define $C_{NZ}$ and $C_Z$ as the index sets of the adaptive non-zero tap-weights and zero tap-weights, respectively.

$$\begin{align*}
C_{NZ} &= \{ k : s_k \neq 0, k = 0, 1, \cdots, L - 1 \}, \\
C_Z &= \{ k : s_k = 0, k = 0, 1, \cdots, L - 1 \}.
\end{align*}$$

When $k \in C_{NZ}$, we consider that the tap-weight $\hat{s}_{k,n}$ has the same sign as the corresponding unknown coefficient $s_k$.
From Eq. (8), when \( n \) approaches infinity, \( E(\hat{s}_\infty) \) converges to
\[
\hat{s}_\infty = s + r_\infty,
\]
\[
E(\hat{s}_\infty) = s + E(r_\infty)
\]
\[
= s + \frac{\rho}{\mu \sigma^2_k} \left[ \frac{\text{sgn}(\hat{s}_\infty)}{\epsilon + \eta'} |\hat{s}_\infty| \right].
\]
\[
\rho \text{ has a small enough positive value that for every } k \in C_{NZ},
\]
we have [5]
\[
E\left[\text{sgn}(\hat{s}_\infty)\right] = \text{sgn}(s_k).
\]
Therefore, it can be shown that
\[
E\left[r_{k,\infty} f(\hat{s}_k)\right] = E\left[r_{k,\infty} \frac{\text{sgn}(s_k)}{\epsilon + \eta'} |\hat{s}_k|\right].
\]
When \( k \in C_Z \), we have
\[
E\left[\text{sgn}(\hat{s}_k)\right] = E\left[\text{sgn}(r_{k,\infty})\right].
\]
Thus, we can have
\[
E\left[r_{k,\infty} f(\hat{s}_k)\right] = E\left[\frac{|r_{k,\infty}|}{\epsilon + \eta'} |\hat{s}_k|\right].
\]
According to assumption (c), from the property of the Gaussian distribution, the following results are obtained:
\[
E\left[r_{k,\infty} f(\hat{s}_k)\right] = \begin{cases} \frac{1}{p} E\left[r_{k,\infty} \text{sgn}(s_k)\right], & k \in C_{NZ} \\ \frac{1}{p} \frac{2E[r_{k,\infty}^2]}{\pi}, & k \in C_Z \end{cases}
\]
where \( \frac{1}{p} = E\left[\frac{1}{\epsilon + \eta'} |\hat{s}_k|\right] \).
\[
E\left[f^2(\hat{s}_k)\right] = \begin{cases} \frac{1}{p^2}, & k \in C_{NZ}, \\ \frac{p^2}{E^2[\text{sgn}(r_{k,\infty})]} \frac{2}{\pi}, & k \in C_Z. \end{cases}
\]
For \( k \in C_{NZ} \), substituting (25) and (26) into (17), we obtain
\[
2\mu \sigma^2_k \Delta_0 E\left[r_{k,\infty}^2\right] - \mu^2 \sigma^4_k D_\infty - \mu^2 \sigma^2_k \sigma^2_0
\]
\[
+ \rho^2 \left( 1 - \frac{2}{\mu \sigma^2_k} \right) = 0,
\]
where \( \Delta_0 \neq 1 - \mu \sigma^2_k \).

For \( k \in C_Z \), combining (25), (26) and (17), we obtain
\[
2\mu \sigma^2_k \Delta_0 \omega^2 - \frac{4 \rho \Delta_0}{p \sqrt{2 \pi}} \omega - \mu^2 \sigma^4_0 D_\infty - \mu^2 \sigma^2_k \sigma^2_0 - \rho^2 / p^2 = 0,
\]
\[
E\left[r_{k,\infty}^2\right] = \frac{\mu^2 \sigma^4_k D_\infty + \mu^2 \sigma^2_k \sigma^2_0 + 2 \rho (1 - \mu \sigma^2_k) E\left[r_{k,\infty} f(\hat{s}_k)\right] + \rho^2 E\left[f^2(\hat{s}_k)\right]}{2 \mu \sigma^2_k (1 - \mu \sigma^2_k)}.
\]
between the convergence rate and signal estimation accuracy. The proposed algorithm will obtain the minimum steady-state MSD, when \( \rho \) is taken as,

\[
\rho_{\text{opt}} = \frac{\sqrt{\alpha_3}}{2} \left( \frac{\sqrt{\alpha_1 + \alpha_2}}{\alpha_1 - \alpha_2} - \frac{\sqrt{\alpha_1 - \alpha_2}}{\alpha_1 + \alpha_2} \right). \tag{33}
\]

Substituting (32) and (33), the minimum steady-state MSD is yielded as

\[
D^\text{min}_\infty = \frac{\mu \sigma^2 L}{\Delta L} + \frac{\alpha_3}{2} \left( \sqrt{\alpha_1^2 - \alpha_2^2} - \alpha_1 \right). \tag{34}
\]

### IV. PERFORMANCE OF THE LOG-SUM LMS FOR BLOCK SPARSE SYSTEMS

The performance of the log-sum block sparse LMS in block-sparse applications is further studied by using the M-G model. An appropriate assumption \( 2 \ll Q(G) \ll L \) is adopted, which means that the partition size \( G \) is very small relative to the filter length to ensure that the system response is still sparse in the group partition. The principle behind this assumption is that \( G \) is an important predefined parameter that needs to be carefully selected. In addition, it is acceptable that the log-sum LMS penalizes sparsity in group-partition and a too large \( G \) will definitely destroy the sparsity.

In this paper, we study the performance of the BSLS-LMS for an unknown system response generated by the M-G model proposed in [22]. Following the approach in [22], we define \( \theta \) as

\[
\theta = \begin{cases} 
\left( \frac{(G-1)!}{G!} \right)^{\frac{3}{2}} \rho_{\text{opt}}^{-1}, & G \text{ is odd;} \\
\left( \frac{(G-1)!}{G} \right)^{\frac{3}{2}} \rho_{\text{opt}}^{-1}, & G \text{ is even.}
\end{cases}
\]

In block sparse systems, the other parameters in (32) are defined as

\[
\alpha_1 = \left[ Q \Delta_0^2 + (L - Q) \frac{\mu \sigma^2 L}{G} \left( \frac{2 \Delta_0^2 \Delta L}{\pi \Delta_0} \right) \right] / (p^2 \mu \sigma_4^2 \Delta L),
\]

\[
\alpha_2 = 2(L - Q) \theta \sqrt{\frac{\Delta_0^2}{\pi}} / (p^2 \mu \sigma_4^2 \Delta^2),
\]

\[
\alpha_3 = 2 \mu \sigma_4^2 \Delta_0^2 / \alpha,
\]

\[
\alpha = \mu \sigma_4^2 Q \Delta_0^2 / \alpha_1 + \Delta_0^2 \left( \frac{\mu \sigma_4^2 \Delta L}{G} + \frac{\theta^2 \Delta_0 \Delta L}{\pi} \right).
\]

Here we present some approximations that \( \mu \sigma_4^2 \rightarrow 0 \). Utilizing \( \Delta_0 \approx 1 \) and \( \Delta_0^2 = 2 \), we have

\[
\alpha \approx \theta^2 \Delta_0^2 / \pi, \tag{35}
\]

\[
\alpha_1 \approx \frac{2(L - Q) \Delta_0^2 \theta^2}{\pi p^2 \mu \sigma_4^2 \Delta L} + \frac{2Q}{p^2 \mu \sigma_4^2 \Delta L}, \tag{36}
\]

\[
\sqrt{\alpha_1^2 - \alpha_2^2} \approx \frac{2Q \sqrt{(L - Q) \Delta_0^2}}{p^2 \mu \sigma_4^2 \Delta L / \pi}, \tag{37}
\]

\[
\alpha_3 \approx \frac{2 \mu \sigma_4^2 \Delta_0^2 \Delta L}{\Delta_0^2 \theta^2}. \tag{38}
\]

We denote \( \tilde{D}^{\text{min}}_\infty \) as the minimum transient MSD. Utilizing (35)-(38) in (33), we achieve the temporary result that

\[
\tilde{D}^{\text{min}}_\infty = \frac{\mu \sigma^2 L}{\Delta L} \left( \frac{L - 2(L - Q)}{\Delta_0} \right) + \frac{2 \mu \sigma^2 L \sqrt{(L - Q) \Delta_0^2}}{p^2 \theta \Delta_0^2} - \frac{2 \mu \sigma^2 Q}{p^2 \Delta_0^2 \theta^2}. \tag{39}
\]

In a sparse unknown system, we assume that \( 2 \ll Q(G) \ll L \), and then, the above equation can be simplified as

\[
\tilde{D}^{\text{min}}_\infty \approx \frac{\mu \sigma^2 L}{\Delta L} \left(Q + \frac{2 \pi \mu \sigma^2 Q}{p^2 \theta \Delta_0^2} - \frac{2 \mu \sigma^2 Q}{p^2 \Delta_0^2 \theta^2}. \tag{40}
\]

We denote the optimal group partition size as \( G_{\text{opt}} \) which can be found numerically by

\[
G_{\text{opt}} = \arg \min_G \tilde{D}^{\text{min}}_\infty. \tag{41}
\]

The above equation shows the selection criteria of the optimal partition size.

### V. SIMULATION RESULTS

In this section, we design three experiments to verify the effectiveness of the proposed algorithm. The first two experiments verify the performance of the log-sum sparse LMS for a general sparse system and a block sparse system under different signal-to-noise ratio, and the last experiment theoretically analyzes the BSLS-LMS. Finally, all the simulation results prove that the convergence rate of the BSLS-LMS is remarkably superior to those of other classical algorithms when the group partition size \( G \) is close to its optimal value \( G_{\text{opt}} \).

#### A. THE PERFORMANCE OF THE BSLS-LMS ON CHANNEL ESTIMATION FOR COOPERATIVE SYSTEM

In this section, the particular application considered is that of estimating the channel state information (CSI) for a communication system. We apply the proposed algorithm to study the performance of channel estimation in a cooperative communication network. Channel estimation is a system identification problem that seeks to identify the CSI of the unknown system. Suppose the cooperation channel model is \( d_n = x^T_n h + v_n \) as in [32]. \( h \equiv h_1 \ast h_2 \) is the unknown channel impulse response, corresponding to \( s \) in Eq.(1). Denote \( h_1(i) \) \((i = 0, 1, \cdots, L_1 - 1)\) as the baseband channel between the source node and the relay node, and \( h_2(i) \) \((i = 0, 1, \cdots, L_2 - 1)\) as the channel impulse response from the relay node to the destination node. \( h \) (with a length of \( L = L_1 + L_2 - 1 \)) is the cascaded channel that is the convolution between \( h_1 \) and \( h_2 \).

Two cases are designed to prove the estimation performance of the block sparse log-sum constraint LMS algorithm through comparison with several existing classical algorithms, including the conventional LMS, Zero-Attracting LMS (ZA-LMS), the RZA-LMS, the reweighted \( l_1 \) norm...
penalized LMS [29]–[31] and the reweighted \( l_p \) norm constrained LMS [32]. The non-zero coefficients of the channel impulse response are Gaussian variables with a zero mean and unit variance and their positions are randomly selected. The input signal and additive noise are zero mean Gaussian sequences with various SNRs.

In case 1, we assume that the channel vectors of the cooperative relay system have the same length \( L_1 = L_2 = 16 \), so the length of the convolution channel vectors is \( L = L_1 + L_2 - 1 = 31 \). There are four large coefficients of \( h_j \) (\( j = 1, 2 \)) uniformly distributed and the rest are all zeroes, hence the sparsity of the system is \( 4/31 \). In case 2, the concatenated channel has the same length \( L_1 = L_2 = 32 \) and the convolution channel vectors length is \( L = L_1 + L_2 - 1 = 63 \). Eight random tap-weights of \( h_j \) (\( j = 1, 2 \)) are nonzero in case 2, and the sparsity of the cooperative channel is \( 8/63 \).

The parameters of the proposed BSLS-LMS channel estimation algorithm are set as follows: \( \rho = 5 \times 10^{-4} \), \( \eta = 0.02 \) and \( \varepsilon = 0.08 \). For the reference algorithms, all the adjustment parameters are properly chosen to obtain the fastest convergence speed. For the ZA-LMS and RZA-LMS, we set the series zero-attraction parameters as \( \rho_{ZA} = \rho_{RZA} = 5 \times 10^{-4} \) and \( \varepsilon_{RZA} = 10 \).

The MSE results of the estimated channel impulse response in sparsity case 1 are shown in Figures 1(a) ~ (b) and the estimated result of sparsity case 2 are shown in Figure 2 (a) ~ (b). It can be seen from Figures 1 and 2 that the convergence of the sparse-aware parameter estimation algorithms decreases as the channel sparsity increases. By comparing the convergence curves of all algorithms, it is concluded that the block-sparsity log-sum LMS algorithm is generally superior to the other algorithms. However, when the SNR = 10 dB, the performance of the reweighted \( l_1 \) norm constrained sparse filtering algorithm gets more closely to that of the log-sum penalty LMS algorithm. In both cases, the proposed algorithm achieves much better MSE performance when the SNR becomes large.
the convergence performances of the sparse parameter estimation algorithms decrease accordingly. Through analysis of the convergence curves, it is concluded that the performance of the proposed algorithm is superior to those of all the other algorithms in the cases of SNR = 10dB and SNR = 20dB. We also proved the MSE performance against iteration number in the scenario of density channel, as shown in Figure 4, it is clear that the block sparse log-sum algorithm has a better performance than the traditional algorithm and all the other sparse-aware constrained adaptive filter algorithms, which is due to the fact that the upper bound of $E[r_{\infty}^2]$ in Eq. (11).

The MSEs between the coefficients of the BSLS-LMS algorithm and the CSI are shown in Figure 1~Figure 4. The proposed BSLS-LMS algorithm converges faster and has better steady-state performance than all the reference algorithms. When there are very few non-zero coefficients in the impulse response of the unknown system, all the sparse-aware adaptive filters have faster convergence speeds and better steady-state performances than the traditional LMS. In addition, it can be observed that the proposed method achieves a lower MSE than the other algorithms. After 500 iterations, when the number of non-zero coefficients increases the BSLS-LMS still achieves the best convergence among the algorithms for both low and high SNRs.

**FIGURE 3.** The sparsity of the system is 16/63.

**FIGURE 4.** Density channel (SNR = 20dB).

**FIGURE 5.** The sparsity of the system is 16/63.

**FIGURE 6.** Density channel (SNR = 20dB).

B. ON THE PERFORMANCE OF THE BLOCK LOG-SUM LMS

In the second experiment, we test the convergence performance of the BSLS-LMS algorithm by identifying block-sparse systems and comparing it with the reference algorithms, including the BS-LMS [25], $l_0$-LMS [30], STNLMS [19], SELQUE [20], and M-SELQUE [21]. The unknown systems have the same length $L = 800$ and the impulse response of the unknown system is generated by the M-G model. Based on assumption (a), the input signal and the additive noise are independent zero-mean Gaussian sequences, and the variance of the input signal is unit, i.e., $\sigma_x^2 = 1$. In this experiment, the SNRs are set as 40dB and 20dB, the balance parameters $\varepsilon$ and $\eta$ are set as 0.03 and 0.008, respectively. For the BSLS-LMS and BS-LMS, the step-size $\mu$ is always set as half of the maximum value $1/(L\sigma_x^2)$ and the zero attraction parameter $\rho_{opt}$ and the group partition size $G_{opt}$ are set as 8.66e-7 and 4, respectively. For the reference algorithms, all the parameters are properly adjusted to obtain their fastest convergence speed and their best performance. The simulation results are averaged by 10 independent tests for each unknown system, 100 unknown systems are generated and identified, and then the learning curves of identifying these systems are averaged.

The simulation results are plotted in Figure 6 corresponding to the unknown systems whose impulse response is plotted in Figure 5(a). When there are four clusters in the system response, the BS-LMS achieves the fastest steady-state at 20dB and 40dB. The convergence performance of the BSLS-LMS is better than those of other block-sparsity recovery algorithms.

In the third experiment, there are more than four clusters in the identifying various unknown block-sparse system whose impulse response are plotted in Figure 5(b). The experimental results are shown in Figure 7, where the BSLS-LMS and BS-LMS show their advantages. It is noted that the BSLS-LMS, similar to BS-LMS, does not need to detect the active regions of the unknown system. The convergence speeds of the SELQUE and STNLMS are obviously reduced because the latency between the two clusters and other active regions
are regarded as a long active region [20]; meanwhile, the other two algorithms cannot detect the active region effectively. Although the M-SELQUE algorithm can detect all the regions, the convergence of the unknown system is still poor when it has multiple clusters.

According to Figure 7, the BSLS-LMS is always the best way to identify various block-sparse systems generated by the M-G model with different parameter sets. Utilizing the blocksparse prior, the BSLS-LMS algorithms has almost the same convergence speed but a lower steady-state performance comparing with $l_0$-LMS; on the other hand, comparing against BS-LMS, BSLS-LMS shows slower convergence speed but a better steady-state performance. Although the M-SELQUE algorithm performs better than other active-region-detection algorithms, its convergence speed is still slower than the proposed algorithm. The main reason is that more iterations will be needed to determine the location of non-zero coefficients when there are more and scattered clusters, hence the convergence rate will reduce correspondingly, which is highly likely to be caused by using the M-G model. From the learning curves, the convergence performances of SELQUE and STNLMS are poor because they are not suitable for the multi-cluster system. It can be seen from Figure 7 and Figure 6 that the convergence performance of the proposed algorithm becomes worse slightly as the number of clusters increases. In summary, we can infer that the BSLS-LMS is more robust than other classical algorithms, especially in multiple distributed cluster sparse systems.

In the fourth experiment, the steady-state MSDs of the BSLS-LMS with various group partition sizes $G$ against $\rho$ are tested. The results of this experiment are plotted in Figure 8 corresponding to the unknown systems shown in Figure 5(c). The step-size is set as $0.66/(L\sigma^2_x)$. $G$ is set as 1, 5, and 15. For each $G$, the variation range of $\rho$ is from $10^{-9}$ to $10^{-5}$. It can be seen from Figure 8 that the theoretical steady-state performance of the log-sum LMS is in good agreement with the simulation results. For each group partition size, the steady-state MSD of the BSLS-LMS decreases when $\rho$ increases from $10^{-9}$, which means that the appropriate group zero-attraction helps to decrease the amplitudes of the coefficients in $C_{Z_2}$. Meanwhile, the stronger group zero attraction enhances the deviation of the coefficients in $C_{NZ}$ as parameter $\rho$ continues to increase. The minimum steady-state MSD and the corresponding optimal $\rho$ are different for different group
TABLE 3. List of abbreviations.

| Abbreviations | Full name |
|---------------|-----------|
| LMS           | Least Mean Square |
| BSLS-LMS      | Block Sparsity Log-Sum LMS |
| M-G           | Markov-Gaussian model |
| MIMO          | Multiple Input Multiple Output |
| BOMP          | Block-version of Orthogonal Matching Pursuit |
| STNLMS        | Stochastic Taps Normalized Least Mean Square |
| SELQUE        | Select and Queue with a Constraint |
| ZA-LMS        | Zero-Attracting LMS |
| RZA-LMS       | Reweighted Zero-Attracting LMS |
| BS-LMS        | Block-Sparse LMS |
| SNR           | Signal-to-Noise Ratio |
| MSD           | Mean Square Deviation |

VI. CONCLUSION

In this paper, we proposed a new block sparse LMS algorithm for unknown system identification. Specifically, we utilized the $l_1$ norm and the log-sum of adaptive tap weights as a mixed constraint in the cost function algorithm. The Markov-Gaussian model was adapted to generate the impulse response of the cluster-sparse unknown system. In addition, based on the expressions of the mean square misalignment, the performance of the block sparse log-sum LMS was theoretically analyzed and the results proved that the proposed algorithm is superior to other algorithms in theory. Finally, several experiments were designed to verify the effectiveness of the theoretical results, and the simulation results demonstrated the superior performance of the proposed algorithm. Moreover, the proposed BSLS-LMS algorithm has low computational complexity which makes the method practical for channel estimation and system identification. Therefore, we can conclude that the theoretical results well predict the trend for MSD when the signal-to-noise ratio is 20dB/40dB, and the simulations results agree well with the analytical values.

VII. ABBREVIATIONS

Please see Table 3.

REFERENCES

[1] Y. Yin, Y. Peng, M. Liu, J. Yang, and G. Gui, “Dynamic user grouping-based NOMA over Rayleigh fading channels,” IEEE Access, vol. 7, pp. 110964–110971, 2019.
[2] Z. Yang, L. Fan, Y. Yang, Z. Yang, and G. Gui, “Generalized singular value thresholding operator based nonconvex low-rank and sparse decomposition for moving object detection,” J. Franklin Inst., vol. 356, no. 16, pp. 10138–10154, Nov. 2019.
[3] Z. Yang, L. Fan, Y. Yang, Z. Yang, and G. Gui, “Generalized nuclear norm and Laplacian scale mixture based low-rank and sparse decomposition for video foreground-background separation,” Signal Process., vol. 172, pp. 1–10, Jul. 2020.
[4] K. Shi and P. Shi, “Convergence analysis of sparse LMS algorithms with $l_1$-norm penalty based on white input signal,” Signal Process., vol. 90, no. 12, pp. 3289–3293, Dec. 2010.
[5] Y. Li, Z. Jiang, O. M. O. Osman, X. Han, and J. Yin, “Mixed norm constrained sparse APA algorithm for satellite and network echo channel estimation,” IEEE Access, vol. 6, pp. 65901–65908, 2018.
[6] W. Shi, Y. Li, and B. Chen, “A separable maximum correntropy adaptive algorithm,” IEEE Trans. Circuits Syst. I, Exp. Briefs, early access, Mar. 3, 2020, doi: 10.1109/TCSII.2020.2977608.

[7] X. Zhang, T. Jiang, Y. Li, and Y. Zakhvor, “A novel block sparse reconstruction method for DOA estimation with unknown mutual coupling,” IEEE Commun. Lett., vol. 23, no. 10, pp. 1845–1848, Oct. 2019.

[8] F. Wu, K. Yang, X. Sheng, and F. Huang, “A blocked MCC estimator for group sparse system identification,” AEU-Int. J. Electron. And Commun., vol. 115, pp. 1–6, Feb. 2020.

[9] Y. Li, Z. Jiang, W. Shi, X. Han, and B. Chen, “Blocked maximum correntropy criterion algorithm for cluster-sparse system identifications,” IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 66, no. 11, pp. 1915–1919, Nov. 2019, doi: 10.1109/TCSII.2019.2891654.

[10] Y. C. Eldar and M. Mishali, “Robust recovery of signals from a structured union of subspaces,” IEEE Trans. Inf. Theory, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[11] M. Stojnic, F. Parvaresh, and B. Hassibi, “On the reconstruction of block-sparse signals with an optimal number of measurements,” IEEE Trans. Signal Process., vol. 57, no. 8, pp. 3075–3085, Aug. 2009.

[12] J. Liu, J. Jin, and Y. Gu, “Efficient recovery of block sparse signals via zero-point attracting projection,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Mar. 2012, pp. 3333–3336.

[13] E. Elhamifar and R. Vidal, “Block-sparse recovery via convex optimization,” IEEE Trans. Signal Process., vol. 60, no. 8, pp. 4094–4107, Aug. 2012.

[14] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, “Block-sparse signals: Uncertainty relations and efficient recovery,” IEEE Trans. Signal Process., vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[15] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, “Model-based compressive sensing,” IEEE Trans. Inf. Theory, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[16] Z. Zhang and B. D. Rao, “Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation,” IEEE Trans. Signal Process., vol. 61, no. 8, pp. 2009–2015, Apr. 2013.

[17] L. Wang, L. Zhao, S. Rahardja, and G. Bi, “Alternative to extended block-sparse Bayesian learning and its relation to pattern-coupled sparse Bayesian learning,” IEEE Trans. Signal Process., vol. 66, no. 10, pp. 2759–2771, May 2018.

[18] R. Zhao, Q. Wang, J. Fu, and L. Ren, “Exploiting block-sparsity for hyperspectral kronecker compressive sensing: A tensor-based Bayesian method,” IEEE Trans. Image Process., vol. 29, pp. 1654–1668, 2020.

[19] Y. Gu, Y. Chen, and K. Tang, “Network echo canceller with active taps stochastic localization,” in Proc. IEEE Int. Symp. Commun. Inf. Technol. ISCIT, Oct. 2005, pp. 556–559.

[20] A. Sugiyama, H. Sato, A. Hirano, and S. Ikeda, “A fast convergence algorithm for adaptive FIR filters under computational constraint for adaptive tap-position control,” IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process., vol. 43, no. 9, pp. 629–636, Sep. 1996.

[21] A. Sugiyama, S. Ikeda, and A. Hirano, “A fast convergence algorithm for space-tap adaptive FIR filters identifying an unknown number of dispersive regions,” IEEE Trans. Signal Process., vol. 50, no. 12, pp. 3008–3017, Dec. 2002.

[22] Y. Wei, Y. Zhang, and C. Wang, “Block-sparsity-aware LMS algorithm for network echo cancellation,” Electron. Lett., vol. 54, no. 15, pp. 951–953, Jul. 2018.

[23] C. Ye, G. Gui, L. Xu, and T. Ohtsuki, “Recovery of block-structured sparse signal using block-sparse adaptive algorithms via dynamic grouping,” IEEE Access, vol. 6, pp. 55689–55693, 2018.

[24] U. Subudhi, H. K. Sahoo, and S. K. Mishra, “Harmonics and decaying DC estimation using volterra LMS/algorithm,” IEEE Trans. Ind. Appl., vol. 54, no. 2, pp. 1108–1118, Mar. 2018.

[25] S. Jiang and Y. Gu, “Block-sparse-induced adaptive filter for multi-clustering system identification,” IEEE Trans. Signal Process., vol. 63, no. 20, pp. 5318–5330, Oct. 2015.

[26] Y. Li, Z. Zhang, Z. Jin, Y. Han, and J. Yin, “Cluster-sparse proportionate NLMS algorithm with the hybrid norm constraint,” IEEE Access, vol. 6, pp. 47794–47803, 2018.

[27] Y. Shen, J. Fang, and H. Li, “Exact reconstruction analysis of log-sum minimization for compressed sensing,” IEEE Signal Process. Lett., vol. 20, no. 12, pp. 1223–1226, Dec. 2013.

[28] E. J. Candès, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted l1 minimization,” J. Fourier Anal. Appl., vol. 14, nos. 5–6, pp. 877–905, 2008.
JUN SUN received the bachelor’s and master’s degrees from Zhengzhou University and Huazhong University of Science and Technology, in 2003 and 2008, respectively. He is currently pursuing the Ph.D. degree with Zhengzhou University. He is also a Lecturer with the School of Electronic Information, Zhongyuan University of Technology. His research interests include massive MIMO systems and signal processing in wireless communication systems.

BING NING (Member, IEEE) received the B.E. degree in electronic information engineering from the Nanjing Artillery Academy of PLA, Nanjing, China, in 2009, and the Ph.D. degree in information and communication system from Zhengzhou University, Zhengzhou, China, in 2016. She is currently working in the Zhongyuan University of Technology as a Lecturer. From October 2012 to January 2013, she was a Visiting Researcher with the Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, USA. Her research interests are in the areas of wireless communications, including cognitive radio, non-orthogonal multiple access, and multiple-input multiple-output.

***