# DISTRIBUTED BLIND SPARSE CHANNEL IDENTIFICATION IN SENSOR NETWORKS

*Ying Liu, Hao Liu*

Department of Information Science
and Electronic Engineering,
Zhejiang University, Hangzhou 310027, P.R. China

## ABSTRACT

Blind channel identification plays an essential role in communications, and various approaches have been proposed in the literature. One of the most important methods for single-input multi-output (SIMO) system identification is the distributed subchannel matching (DSCM) algorithm. As the D-SCM treats each component of the channel coefficient vectors equally, it has no advantage when the channels are sparse. In this paper, we propose a sparse DSCM algorithm to blindly identify some sparse channels. Unlike the common DSCM algorithm, a sparsity-enforcing regularization term based on $\ell_1$-norm or $\ell_0$-norm, is added into the cost function to exploit the sparse structure of channels. Some simulations are then presented to show that the proposed sparse DSCM can improve the performance of estimation in both convergence and accuracy.

***Index Terms***— Blind channel identification, sensor networks, sparse, subchannel matching, distributed estimation

## 1. INTRODUCTION

Blind channel identification is one of the most important topics in digital communication systems. Without the requirement of a training signal, blind estimation can save channel capacity and transmission bandwidth. So, it is more preferable in practical applications [1].

In this paper, the blind identification of multiple channels in a networked system is investigated. In particular, we consider the case that a common source is transmitted to a set of distributed sensors over a geometrical region. Due to the inhomogeneous transmission medium, the channels are distinct at sensors, which correspond to a single-input multi-output (SIMO) convolutional system model. Our objective is to estimate the multiple channels only based on the observed signals.

In the literature, various blind channel identification methods for SIMO systems have been proposed, such as subchannel matching (SCM) method [2] and subspace method [3]. In these algorithms, all the measurements are transmitted to a centralized fusion center for processing, and thus a powerful fusion center is required. Moreover, the centralized processing is also fragile to the failure of the fusion center.

To tackle this problem, in [4,5], a distributed SCM (DSCM) method is proposed, where a set of sensors cooperatively estimate the channels and the source by accessing to the information of each sensor's neighbors in both noiseless and noisy measurements. Such distributed cooperative estimation can properly distribute the overall computational load and reduce the relevant information exchanges.

On the other hand, sparsity (many coefficients of a system/vector are zeros or near-zeros) commonly exists in nature. In many wireless communication systems, the propagation channels involved exhibit a large delay spread, but a sparse impulse response consisting of a small number of dominant echoes. For example, terrestrial transmission of high definition television (HDTV) signals [6], underwater channel models [7] and communication channels [8] are shown to be sparse. Recent advances have shown that exploiting the sparsity of the vector of interest contributes to improving the performance of estimation in both convergence and accuracy [8–10].

Considering this, in this paper, the blind sparse channel identification under noisy measurements of a distributed sensor network is studied. By incorporating a sparsity-enforcing regularization term based on $\ell_1$-norm or $\ell_0$-norm into the cost function, a kind of sparse DSCM algorithm is proposed. Numerical simulations are then performed to show that the proposed sparse DSCM is more effective than the common D-SCM without considering the sparsity of the channels.

The rest of this paper is organized as follows. In Section 2, the problem of distributed blind sparse channel estimation is formulated. In Section 3, a kind of sparse DSCM algorithm using gradient descent is proposed. Some simulation examples are given to validate the proposed algorithms in Section 4, followed by the conclusion in Section 5.

## 2. PROBLEM FORMULATION

In this section, we formulate the problem of blind channel identification in a SIMO networked system model. Consider a network consisting of $L$ sensors spatially distributed over a

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61171153 and 61471320).

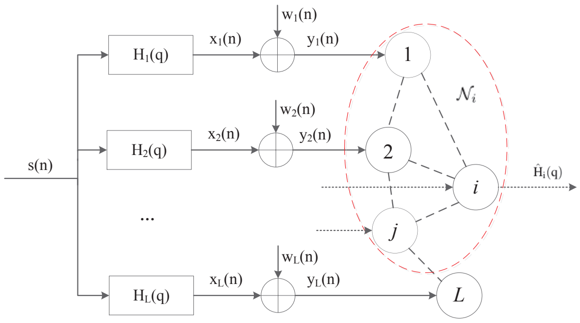


**Fig. 1**. Diagram of blind channel identification in a networked system. Take sensor $i$ as an example.

region. Here, we use an $L \times L$ non-negative weight matrix $\mathbf{C}$ to describe the topology of the network. Assume that the network is connected and $\mathbf{C}$ is a symmetric stochastic matrix with its entities $c_{i,j}$ computed by the following Metropolis rule [4, 9]

$$c_{i,j} = \begin{cases} 1/(1+\max\{d_i, d_j\}) & j \in \mathcal{N}_i, j \neq i, \\ 1 - \sum_{l \in \mathcal{N}_i} c_{i,l} & i = j, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

where $\mathcal{N}_i$ denotes the neighbors of node $i$ and $d_i$ is the degree of node $i$.

All the sensors are interested in the common source $s(n)$ through their own FIR channel with impulse response $\{h_i(n)\}$, resulting to a SIMO system, see Fig. 1. With reference to Fig. 1, the output $y_i(n)$ collected at each sensor $i$, the input data $s(n)$ and the FIR channel $\{h_i(n)\}$ are related by

$$\begin{aligned} & x_i(n) = h_i(n) * s(n) = \sum_{k=0}^{M} h_i(k)s(n-k), \\ & y_i(n) = x_i(n) + w_i(n), \\ & i = 1, 2, \ldots, L, \;\; n = 1, 2, \ldots, N, \end{aligned} \tag{2}$$

where the order of the FIR channel is $M + 1$, $s(n) \in \mathbb{R}$ denotes the source signal, $N$ is the length of source signal, the notation '$*$' denotes convolution operation, $y_i(n) \in \mathbb{R}$ denotes the observation of the $i$-th sensor, $w_i(n) \in \mathbb{R}$ denotes the measurement noise. Here, we are interested in the case that the channels are sparse, that is, a large proportion of channel coefficients $h_i(n)$ are zeros.

Similar to the studies in blind identification [2, 4, 5], some assumptions are assumed throughout the paper.

A1. All transfer functions $\{H_i(q)\}_{i=1}^{L}$ share no common zeros, where $H_i(q) = \sum_{k=0}^{M} h_i(k)q^{-k}$ denotes the transfer function from the source to the $i$-th sensor and $q^{-1}$ stands for the back shift operator in time domain.

A2. The source signal $s(n)$ is stochastic and has linear complexity larger than $2M + 1$.

A3. The noise $w_i(n)$ has bounded amplitude.

Assumptions A1-A2 provide sufficient conditions for the blind identifiability of the SIMO system in (2), see e.g., [2]. The boundedness assumption A3 is required for the estimation using finite observation samples. We also assume that the data transmission between sensors is perfect and each node can only exchange information with its neighbors. Our goal is to identify the multiple sparse channels only based on the observation sequences at each sensor.

## 3. SPARSE DISTRIBUTED SUBCHANNEL MATCHING ALGORITHM FOR BLIND CHANNEL IDENTIFICATION

In this section, we present a sparse DSCM algorithm to tackle the situation that the subchannel coefficient vectors are sparse. Suppose that the noisy observations $\{y_i(n)\}_{i=1,n=0}^{L,N}$ are available. Denote

$$\boldsymbol{\mathcal{Y}}_i \triangleq \begin{bmatrix} y_i(N) & y_i(N-1) & \cdots & y_i(N-M) \\ y_i(N-1) & y_i(N-2) & \cdots & y_i(N-M-1) \\ \vdots & \ddots & \ddots & \vdots \\ y_i(2M) & y_i(2M-1) & \cdots & y_i(M) \end{bmatrix}. \tag{3}$$

According to the network topology, an augment matrix $\boldsymbol{\mathcal{Y}}$ is defined as

$$\boldsymbol{\mathcal{Y}} = \Big[ \underbrace{\mathbf{0} \cdots \mathbf{0}}_{i-1 \text{ block entries}} \;\; -\boldsymbol{\mathcal{Y}}_j \;\; \underbrace{\mathbf{0} \cdots \mathbf{0}}_{j-i-1 \text{ block entries}} \;\; \boldsymbol{\mathcal{Y}}_i \;\; \mathbf{0} \cdots \mathbf{0} \Big], \tag{4}$$

if there is a link between node $i$ and $j$.

Considering that the channels are sparse and motivated by the idea of compresses sensing [8–10], we incorporate a sparsity-enforcing regularization term, $\ell_p$-norm( $p = 0, 1$) into the cost function of the conventional DSCM algorithm [4]. That is, the channel can be identified by minimizing the following $\ell_p$-norm penalized cost function

$$\arg\min_{\mathbf{h} \neq \mathbf{0}} \|\boldsymbol{\mathcal{Y}}\mathbf{h}\|^2 + \gamma \xi_p(\mathbf{h}), \tag{5}$$

where $\mathbf{h} = [\mathbf{h}_1^T \cdots \mathbf{h}_L^T]^T$ with $\mathbf{h}_i = [h_i(0) \cdots h_i(M)]^T$, $\xi_p(\mathbf{h})$ denotes the $\ell_p$-norm of $\mathbf{h}$, $\gamma$ is the regularization parameter to balance the penalty of the norm constraint and the data fidelity term $\|\boldsymbol{\mathcal{Y}}\mathbf{h}\|^2$. When $\gamma = 0$, (5) equals to the common DSCM algorithm.

To solve the above optimization problem, we use the gradient descent algorithm. However, due to the existence of noise, the estimation of $\mathbf{h}$ may converge to a trivial solution. To avoid this, we use a similar method as [4]. In the following, we give the centralized solution first and then extend it to distributed implementation. Inspired by [4], the channel vector $\mathbf{h}$ can be iteratively estimated by

$$\hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) - \alpha_k \|\hat{\mathbf{h}}(k)\|^4 [\boldsymbol{\mathcal{Y}}^T \boldsymbol{\mathcal{Y}} \hat{\mathbf{h}}(k) + \gamma \partial \xi_p(\hat{\mathbf{h}}(k))] + \alpha_k \hat{\mathbf{h}}(k) \hat{\mathbf{h}}^T(k) \boldsymbol{\mathcal{Y}}^T \boldsymbol{\mathcal{Y}} \hat{\mathbf{h}}(k) + \gamma \alpha_k \hat{\mathbf{h}}(k) \hat{\mathbf{h}}^T(k) \partial \xi_p(\hat{\mathbf{h}}(k)), \tag{6}$$

where $\partial \xi_p(\cdot)$ denotes the derivative of $\xi_p(\cdot)$.

Note that the term $\partial \xi_p(\cdot)$ in (6) depends on the norm constraint in-used. For the $\ell_1$-norm, we have

$$\xi_1(\mathbf{x}) = \|\mathbf{x}\|_1. \tag{7}$$

where $\mathbf{x} = [x(1) \cdots x(n)]^T$, and thus its component-wise derivation with regard to $x(i)$ is

$$\partial \xi_1(x(i)) = \text{sgn}(x(i)) = \begin{cases} \frac{x(i)}{|x(i)|} & \text{if } x(i) \neq 0, \\ 0 & \text{if } x(i) = 0. \end{cases} \tag{8}$$

The $\ell_0$-norm counts the number of nonzero entries in a vector, thus indicating the model's complexity [10]. As it is nonconvex, the $\ell_0$-norm is usually approximated by

$$\xi_0(\mathbf{x}) = \|\mathbf{x}\|_0 \approx \sum_{i=1}^{M} (1 - e^{-a|x(i)|}), \tag{9}$$

where $a$ is a positive constant. Its component-wise derivation with regard to $x(i)$ is

$$\partial \xi_0(x(i)) = a \text{sgn}(x(i)) e^{-a|x(i)|}, \quad \forall\ 1 \leq i \leq M. \tag{10}$$

By approximating $e^{-a|x(i)|}$ with its first order Taylor expansion, (10) can be written as

$$\begin{aligned} \partial \xi_0(x(i)) =& a(\text{sgn}(x(i)) - a x(i)) \\ =& \begin{cases} -(a^2 x(i) + a) & \text{if } -\frac{1}{a} \leq x(i) < 0, \\ -(a^2 x(i) - a) & \text{if } 0 < x(i) \leq \frac{1}{a}, \\ 0 & \text{elsewhere.} \end{cases} \end{aligned} \tag{11}$$

Next, we use the average consensus method to derive the distributed solution, where a set of sensors collaboratively estimate the multiple channels. As proposed in [4], the two global variables in (6) can be equivalently represented as

$$\begin{aligned} \|\hat{\mathbf{h}}(k)\|^4 &= L^2 \left( \frac{1}{L} \sum_{i=1}^{L} \|\hat{\mathbf{h}}_i(k)\|^2 \right)^2, \\ \|\boldsymbol{\mathcal{Y}} \hat{\mathbf{h}}(k)\|^2 &= \frac{L}{2} \left( \frac{1}{L} \sum_{i=1}^{L} \sum_{j \in \mathcal{N}_i} \|\boldsymbol{\mathcal{Y}}_i \hat{\mathbf{h}}_j(k) - \boldsymbol{\mathcal{Y}}_j \hat{\mathbf{h}}_i(k)\|^2 \right). \end{aligned} \tag{12}$$

These two global variables can be estimated by running the average consensus operation such that each sensor shares a certain computation to approximate the global variables.

Let $\boldsymbol{\phi}(k) = [\phi_1(k) \cdots \phi_L(k)]^T$ and $\boldsymbol{\varphi}(k) = [\varphi_1(k) \cdots \varphi_L(k)]^T$, where $\phi_i(k) = \|\hat{\mathbf{h}}_i(k)\|^2$ and $\varphi_i(k) = \sum_{j \in \mathcal{N}_i} \|\boldsymbol{\mathcal{Y}}_i \hat{\mathbf{h}}_j(k) - \boldsymbol{\mathcal{Y}}_j \hat{\mathbf{h}}_i(k)\|^2$.

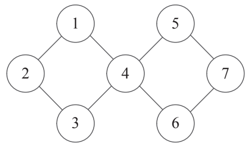


**Fig. 2**. The network topology of 7 sensors.

Let $\overline{\boldsymbol{\phi}}(k) = [\overline{\phi}_1(k) \cdots \overline{\phi}_L(k)]^T$ and $\overline{\boldsymbol{\varphi}}(k) = [\overline{\varphi}_1(k) \cdots \overline{\varphi}_L(k)]^T$, where $\overline{\boldsymbol{\phi}}(k) = \mathbf{C}^R \boldsymbol{\phi}(k)$ and $\overline{\boldsymbol{\varphi}}(k) = \mathbf{C}^R \boldsymbol{\varphi}(k)$. $\mathbf{C}$ denotes the weight matrix as defined in (1). According to the concept of average consensus [4], for each sensor $i$, the estimates of these two global variables become $L^2 \overline{\phi}_i^2(k)$ and $\frac{L}{2} \overline{\varphi}_i(k)$, respectively. Then, the channels of each sensor $i$ can be estimated by the following iteratively

$$\begin{aligned} \hat{\mathbf{h}}_i(k+1) = \hat{\mathbf{h}}_i(k) - \alpha_k L^2 \overline{\phi}_i^2(k) \sum_{j \in \mathcal{N}_i} \big( & \boldsymbol{\mathcal{Y}}_j^T \boldsymbol{\mathcal{Y}}_j \hat{\mathbf{h}}_i(k)) \\ & - \boldsymbol{\mathcal{Y}}_j^T \boldsymbol{\mathcal{Y}}_i \hat{\mathbf{h}}_j(k) \big) + \frac{\alpha_k L}{2} \overline{\varphi}_i(k) \hat{\mathbf{h}}_i(k) \\ & - \alpha_k \gamma L^2 \overline{\phi}_i^2(k) \partial \xi_p(\hat{\mathbf{h}}_i(k)) \\ & + \frac{\alpha_k \gamma L}{2} \hat{\mathbf{h}}_i(k) \hat{\mathbf{h}}_i^T(k) \partial \xi_p(\hat{\mathbf{h}}_i(k)), \end{aligned} \tag{13}$$

where $\partial \xi_p(\cdot)$ is given in (8) and (11) for $\ell_1$ and $\ell_0$-norm based sparse DSCM, denoted as $\ell_1$-DSCM and $\ell_0$-DSCM respectively.

To summarize, the implementation procedures of sparse D-SCM algorithms are given in Algorithm 1.

**Algorithm 1** Sparse DSCM algorithm.

1. Given the initial condition $\hat{\mathbf{h}}_i(0)$ for $i = 1, \cdots, L$.
2. Using average consensus algorithm (12) to estimate two global variables.
3. For $i = 1 : L$

   Update $\hat{\mathbf{h}}_i(k)$ according to (13), where $\partial \xi_p(\cdot)$ is computed by (8) and (11) for $\ell_1$-DSCM and $\ell_0$-DSCM respectively.

   End for
4. $k \leftarrow k + 1$ and go to step 2.

## 4. SIMULATIONS

In this section, to investigate the performance of the proposed sparse DSCM algorithm for blind sparse channel identification, some numerical simulations are performed.

In the simulation, a network of $L = 7$ sensors is considered and its topology is shown in Fig. 2. For the transmitted symbols, we use BPSK sequences with equiprobable symbols. The length of source signal is $N = 150$ and the highest order of each FIR channel is $M + 1 = 20$. The signal-to-noise ratio (SNR) is set to 30dB. The number of consensus operations $R$ is set to 50. The step size $\alpha_k$ is set to $\alpha_k = k^{-1}$. Since the step size is quite large at first few iterations, an extra operation is carried out to avoid the 'blow-up' of the estimation during the first few iterations [11]. That is,

$$\hat{\mathbf{h}}_i(k) = \hat{\mathbf{h}}_i(k) I_{||\hat{\mathbf{h}}_i(k)||<1} + \hat{\mathbf{h}}_i(0) I_{||\hat{\mathbf{h}}_i(k)||\geq 1}, \qquad (14)$$

where $I_{[\cdot]}$ is an indicator function.

Comparing to the common DSCM, a zero-forcing term with factor $\gamma$ is introduced in the optimization function (5) of the sparse DSCM. So, the performance of sparse DSCM largely depends on the parameter $\gamma$. Moreover, constant $a$ in (11) is also an important parameter to control the performance of $\ell_0$-DSCM. So, in order to obtain good estimation performance, suitable parameters $\gamma$ and $a$ are selected at first.

Fig. 3 depicts the normalized steady-state network error of the channel coefficients against different $\gamma$, where the number of the nonzero components is set as $M_{nz} = 5$ and it is the same for each subchannel. Note that the normalized errors are obtained by averaging the last 1000 instantaneous samples after 15000 iterations. For the $\ell_0$-DSCM, the results of different $a$ are also presented. For the purpose of performance comparison, the results of the common DSCM algorithm (i.e. $\gamma = 0$) are also shown in Fig. 3. From Fig. 3, it is noticed that the sparse DSCM algorithms are largely dependent on parameters $\gamma$. For both $\ell_1$-DSCM and $\ell_0$-DSCM, with the increasing of $\gamma$, the normalized error is decreased first and then increased after a certain value $\gamma$. There exists a corresponding region of $\gamma$ for $\ell_1$-DSCM and $\ell_0$-DSCM respectively, in which they outperform the common DSCM algorithm. Moreover, from Fig. 3, it can be seen that the performance of $\ell_0$-DSCM is also dependent on parameter $a$. From Fig. 3, we can see that with appropriate $a$, for example, $a \in [10, 50]$, $\ell_0$-DSCM outperforms $\ell_1$-DSCM. Moreover, when $\gamma$ is set within $[10^{-2}, 10^{-0.6}]$, sparse DSCM outperforms the common DSCM significantly.

Based on the former simulation results, we will show the performances of sparse DSCM algorithms compared with that of the common DSCM. In this simulation, the number of nonzero taps in each subchannels is set as $M_{nz} = 5$. The regularization parameter $\gamma$ is set to be $10^{-0.8}$ and $10^{-1.6}$ for $\ell_1$-DSCM and $\ell_0$-DSCM respectively. The parameter $a$ is set to 20 for $\ell_0$-DSCM. Initialization conditions of channels are the same for different algorithms. Fig. 4 shows the normalized steady-state network error of channel estimation for different algorithms. From Fig. 4, it is obvious that both the $\ell_1$-DSCM and $\ell_0$-DSCM outperform the common DSCM in estimation accuracy, where the $\ell_0$-DSCM is a little bit better than the $\ell_1$-DSCM from both the viewpoints of accuracy and convergence.

Fig. 5 shows the normalized steady-state network error of sparse DSCM and common DSCM algorithms with respect to different SNRs. In this simulation, the number of non-zero taps in each subchannels is set as $M_{nz} = 5$. The parameter $a$ is set to 20 for $\ell_0$-DSCM. It can be found that the sparse

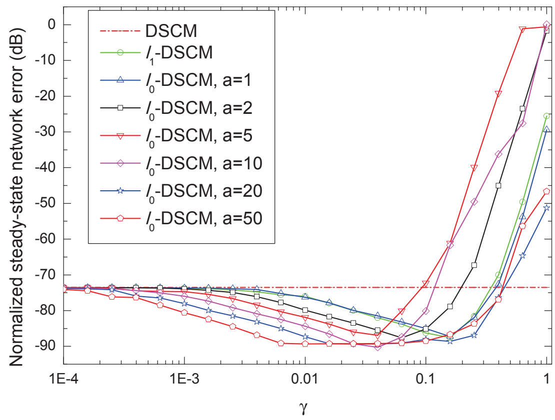


**Fig. 3**. Normalized steady-state network error vs parameter $\gamma$ of the $\ell_1$-DSCM and the $\ell_0$-DSCM.

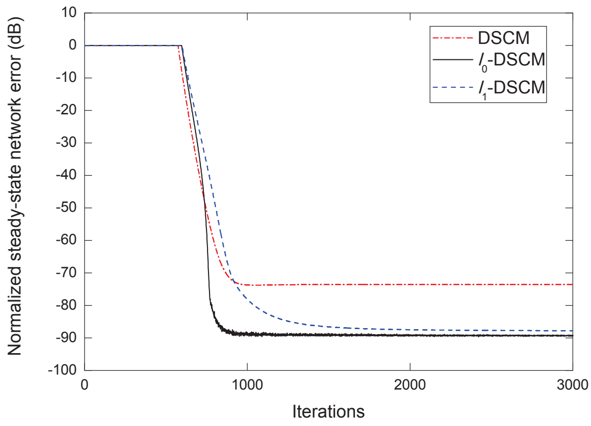


**Fig. 4**. Normalized steady-state network error vs iteration for different algorithms.

DSCM algorithms outperform the common DSCM, and with the increasing of the SNRs, the errors of three algorithms decrease correspondingly.

Next, we investigate the influence of different degree of sparsity on the performance of estimation. In the simulation, the highest order of each FIR channel is $M + 1 = 20$. we set the number of the nonzero components $M_{nz}$ varying from 4 to 20 with an increment of 2 and thus the degree of sparsity changes from high to low. The simulation results are given in Fig. 6. Note that in Fig. 6, the y-coordinate values of each curve present the differences in the error between the common DSCM algorithm (i.e. $\gamma = 0$) and the sparse DSCM algorithms. From Fig. 6, it is noticed that the sparse DSCM algorithms outperform the common DSCM algorithm when the channels are much sparser, and the degree of superiority decreases as the degree of sparsity reduces. Finally, when the sparsity disappears, i.e. $M_{nz} = 20$, the sparse D-

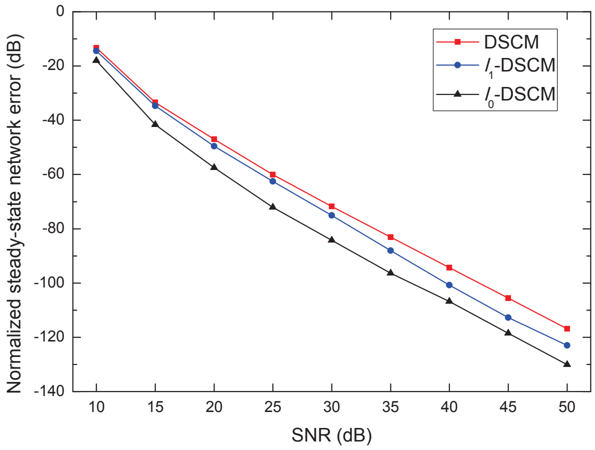


**Fig. 5**. Normalized steady-state network errors of different algorithms vs SNR.

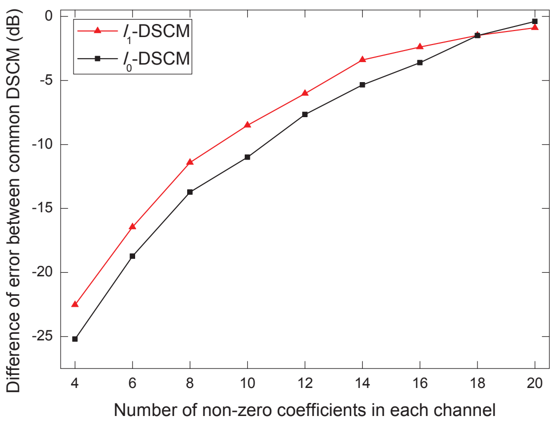


**Fig. 6**. Difference of normalized steady-state network error between the sparse DSCM and the common DSCM with different numbers of non-zero coefficients in the channels.

SCM algorithms converge to the common DSCM algorithm. Moreover from Fig. 6, it is noticed that $\ell_0$-DSCM outperforms $\ell_1$-DSCM when the number of non-zero coefficients in each channel is less than 18. Therefore, $\ell_0$-DSCM is recommended when the channels are much sparser.

## 5. CONCLUSION

In this paper, we have proposed a kind of sparse DSCM algorithm for blind channel identification in sensor networks. In contrast with the common DSCM algorithm, our method takes full advantage of the sparsity of channels by introducing $\ell_1$ and $\ell_0$-norm constraints into the cost function. Numerical simulations are presented to show the effectiveness of the proposed methods. From the simulation results, we find that the sparse DSCM, including $\ell_1$-DSCM and $\ell_0$-DSCM, both outperform the common DSCM with suitable parametric settings when the channels are indeed sparse. Provided that suitable parameters are selected, $\ell_0$-DSCM outperforms $\ell_1$-DSCM in both convergence rate and accuracy. Moreover, we also investigate that the degree of sparsity influences the performance of the sparse DSCM algorithms. The sparse DSCM algorithms are significantly better than the common DSCM when the channels are much sparser, and they converge to the latter when the sparsity disappears.

## REFERENCES

[1] Z. Ding, and L. Ye, *Blind equalization and identification, CRC press*, 2001.

[2] G. H. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification, *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982-2993, Dec. 1995.

[3] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516-525, Feb. 1995.

[4] C. Yu, L. Xie and Y. Soh, "Blind channel and source estimation in networked systems," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4611-4626, Sept. 2014.

[5] C. Yu, L. Xie and Y. Soh, "Distributed blind system identification in sensor networks," *IEEE International Conference on Acoustics, Speech and Signal Processing,* 2014.

[6] A. A. Rontogiannis and K. Berberidis, "Efficient decision feedback equalization for sparse wireless channels," *IEEE Trans. Wireless Commun.*, vol. 2, no. 3, pp. 570-581, May 2003.

[7] M. Kocic, D. Brady, and M. Stojanovic, "Sparse equalization for real time digital underwater acoustic communications," *OCEANS'95,* (San Diego, CA), Oct. 1995, pp. 1417-1422.

[8] J. A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1847-1862, Mar. 2010.

[9] Y. Liu, C. Li, and Z. Zhang, "Diffusion sparse least-mean squares over networks," *IEEE Trans. Signal Process.,* vol. 60, no. 8, pp. 1-6, 2012.

[10] Y. Gu, J. Jin, and S. Mei, "$\ell_0$-norm constraint LMS algorithm for sparse system identification," *IEEE Signal Process. Lett.* vol. 16, no. 9, pp. 774-777, Sept. 2009.

[11] H. Chen, *Stochastic approximation and its applications,* Norwell, MA, USA: Kluwer, 2002.