A Novel Maximum-Likelihood Method for Blind Multichannel Identification

Chengpu Yu†, Cishen Zhang⋆, Lihua Xie†
†EXQUISITUS, Centre for E-City, School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore 639798
⋆Faculty of Engineering and Industrial Sciences, Swinburne University of Technology
Hawthorn VIC 3122, Australia

***Abstract*—Deterministic blind identification algorithms of single-input and multi-output (SIMO) systems can effectively estimate channel functions and the common source signal at high signal-noise-ratio (SNR) and small available data sample scenarios. However, it is difficult for them to identify systems accurately when the noise level is high. To deal with the noise problem, this paper develops an exact Maximum-Likelihood (EML) model which is different from the two-stage Maximum-Likelihood (TSML) method or the semi-blind ML method in the literature. The EML model derived from the cross relation equation of two channels does not contain the source signal but channel functions and output observations, hence the identification performance is barely affected by the unknown source signal. In addition, an iterative optimization approach based on variable splitting technique and alternating direction method of multipliers (ADMM) is derived to minimize the negative log-likelihood function. Simulations are carried out to verify the effectiveness of the proposed method.**

I. Introduction

Blind multi-channel identification has applications in many areas such as wireless communications, underwater acoustic signal processing, speech recognition and enhancement, ultrasound imaging, etc. and has been intensively investigated over the last few decades. The conventional second-order statistical methods do not contain adequate information to identify non-minimum phase single-input single-output (SISO) systems. Instead, by over-sampling the system output, the non-minimum phase system can be identified using its second-order statistics since the over-sampled system output is a cyclostationary signal that contains phase information. The over-sampled system can be represented as a single-input multi-output (SIMO) model [1], [2]. Since then, researches on blind identification of SIMO systems have been intensively carried out and many algorithms had been developed [3]. The well known results on blind identification of SIMO systems in [1], [4] established that unique identification solution for finite impulse response (FIR) channels can be obtained up to a scalar constant under mild conditions.

In general, blind multi-channel identification methods can be classified into two categories: statistical methods and deterministic methods. For statistical methods, a large number of observed data are necessary [2], [5] in order to get accurate statistical information. The main advantage of this kind of methods is that it can effectively deal with the channel noise. Eric et al. [6] used the statistical subspace method to identify FIR channels, Roberto et al. [5] proposed a FIR channel equalization algorithm for systems with colored sources, and Chengpu et al. [7] presented an ARMA system identification algorithm using second-order statistics. On the other hand, the deterministic methods, such as the cross relation (CR) method [1], the subspace method [8] and the least squares smoothing method [9], can perform well using a small set of observed data without knowing any statistical properties of the source signal. These methods are effective only at those scenarios with low noise levels.

In the literature, a two-stage deterministic Maximum-Likelihood method was proposed in [4], which outperforms other deterministic methods at low noise levels. It uses a heuristic approach to resolve the Maximum-Likelihood problem, where the noise affect is neglected when calculating the common source signal from the noisy observed data and the estimated channel functions. In addition, only two iterations are carried in this method, hence the result is not locally (or globally) optimal. On the other hand, an iterative quadratic Maximum-Likelihood approach [10] was used to solve the same optimization problem as in [4], but its convergence can not be guaranteed. The cross relation method [1] has a merit that the source signal does not involved in the optimization formulation, hence the identification performance will not be affected by the source signal. However, it is formulated without concerning the noise effect, hence the identification results may deviate from their true values when the noise level is high. Inspired from the advantages of the above deterministic algorithms, we aim to develop a deterministic method which is robust to noise effect even if the observed data set is small.

To the best of our knowledge, there is not much work in the literature on identifying SIMO systems deterministically by properly dealing with the noise effect. In this paper, an exact Maximum-Likelihood (EML) formulation is derived for the blind SIMO channel identification problem based on the cross relations (CR) among different channels, where the noise

effect is properly dealt with in the EML formulation. Since the derived optimization problem is nonlinear, it is difficult to obtain a global optimal solution. Here, by incorporating the variable splitting technique and the alternating direction method of multipliers (ADMM), a numerical computational approach is derived to minimize the nonlinear negative log-likelihood function. Compared with the TSML method and the CR method, our proposed algorithm can achieve more accurate estimates of FIR channels.

The rest of this paper is organized as follows. Section II describes the blind multi-channel identification problem. Section III formulates an exact Maximum-Likelihood model and Section IV develops a numerical approach to solve it. Simulation results are provided in Section V followed by Conclusion in Section VI.

II. Problem Formulation

Consider the following L-channel FIR SIMO model

$$\begin{aligned} y_i(t) &= h_i(t) * s(t) + n_i(t) \\ &= \sum_{k=1}^{M} h_i(k)s(t-k) + n_i(t) \quad i = 1, 2, \cdots, L \end{aligned} \tag{1}$$

where $s(t)$ denotes the common source signal, $h_i(t)$ denotes the i-th FIR channel function, M denotes the maximum order of all channels, $y_i(t)$ denotes i-th output signal, $w_i(t)$ denotes the additive white noise in i-th channel, and $*$ denotes the convolution operator.

The CR method [1] is a classical deterministic method for blind channel identification. When the source signal is informative enough and there is no additive noise, the following CR equations can be obtained

$$y_i(t) * h_j(t) = y_j(t) * h_i(t) \quad 1 \le i \ne j \le L. \tag{2}$$

By combining all possible pairs of L channels, the following matrix-vector formulation can be constructed

$$\mathbf{Yh} = 0 \tag{3}$$

where

$$\mathbf{Y} = \begin{bmatrix} \mathcal{Y}_2 & -\mathcal{Y}_1 & & & \\ \mathcal{Y}_3 & 0 & -\mathcal{Y}_1 & & \\ \vdots & \vdots & & \ddots & \\ \mathcal{Y}_L & 0 & 0 & \cdots & -\mathcal{Y}_1 \\ & \mathcal{Y}_3 & -\mathcal{Y}_2 & & \\ & \vdots & & \ddots & \\ & \mathcal{Y}_L & 0 & \cdots & -\mathcal{Y}_2 \\ & & & \cdots & \\ & & & \mathcal{Y}_L & -\mathcal{Y}_{L-1} \end{bmatrix},$$

$$\mathcal{Y}_i = \begin{bmatrix} y_i(N) & y_i(N-1) & \cdots & y_i(N-M) \\ y_i(N-1) & y_i(N-2) & \cdots & y_i(N-M-1) \\ \vdots & \vdots & \ddots & \vdots \\ y_i(M) & y_i(M-1) & \cdots & y_i(0) \end{bmatrix},$$

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_L \end{bmatrix}^T,$$

$$\mathbf{h}_i = \begin{bmatrix} h_i(0) & h_i(1) & \cdots & h_i(M) \end{bmatrix}^T,$$

and N is the number of available output samples.

The identification of FIR channels by the CR method [1] is based on following assumptions:
1) All FIR channels are coprime, i.e. they do not share any common zeros;
2) The number of observation samples $N \ge 3M + 1$;
3) The modes of input, defined by the rank of its infinite-order Hankel matrix, should be larger than or equal to $2M+1$.

The Assumption 1) is necessary: if all channel functions share some common zeros, they cannot be uniquely identified up to a scalar constant [1]. The Assumption 2) is used to assure that the matrix $\mathbf{Y}$ is a tall matrix, so its rank is determined by the column rank. The Assumption 3) is necessary in that the equality of (2) will not be affected when removing the source signal on both hand-sides.

The concerned problem in this paper is to identify the FIR functions $\{\mathbf{h}_i\}_{i=1}^{L}$ from the corresponding output observations using an exact Maximum-Likelihood estimator. To eliminate the effect of the unknown source signal, the derivation of the exact Maximum-Likelihood estimator is based on the CR equations among different channels.

III. Exact Maximum-Likelihood Method

Taking into account the additive channel noise and under Assumption 3), the CR equations between different channels are given by

$$(y_i(t) - n_i(t)) * h_j(t) = (y_j(t) - n_j(t)) * h_i(t) \quad 1 \le i \ne j \le L. \tag{4}$$

They are bilinear equations with respect to channel functions $h_i(t)$. In noiseless cases, all channels can be uniquely identified up to a scalar constant. In order to eliminate the

scalar ambiguity, we set the value of $h_1(0)$ to $h_1(0) = 1$ (instead of imposing the unit energy constraint on FIR channels). Then other coefficients of FIR channels can be uniquely determined from the CR equations.

To simplify the notation but without loss of generality, we consider a two-channel SIMO model. The corresponding CR equation is shown below

$$(y_1(t) - n_1(t)) * h_2(t) = (y_2(t) - n_2(t)) * h_1(t).$$

With the fixed $h_1(0) = 1$, an alternative formulation of the above equation is given by

$$\begin{aligned} y_2(t) - n_2(t) = & \left[\begin{array}{cc} -y_2(t-1)* & y_1(t)* \end{array} \right] \left[\begin{array}{c} h_1(t-1) \\ h_2(t) \end{array} \right] \\ & + \left[\begin{array}{cc} -h_2(t)* & h_1(t-1)* \end{array} \right] \left[\begin{array}{c} n_1(t) \\ n_2(t-1) \end{array} \right]. \end{aligned}$$

For computational convenience, this convolution equation is written in the following matrix equation

$$\mathbf{y} = \left[\begin{array}{cc} -\mathcal{Y}_2 & \mathcal{Y}_1 \end{array} \right] \left[\begin{array}{c} \mathbf{h}_1 \\ \mathbf{h}_2 \end{array} \right] + \left[\begin{array}{cc} -\mathbf{H}_2 & \mathbf{H}_1 \end{array} \right] \left[\begin{array}{c} \mathbf{n}_1 \\ \mathbf{n}_2 \end{array} \right] \quad (5)$$

where $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are Hankel matrices defined earlier except that $\mathcal{Y}_2$ only keeps the last M columns, $\mathbf{h}_1$ and $\mathbf{h}_2$ are coefficient vectors defined earlier except that $\mathbf{h}_1$ only keeps the last M coefficients, $\mathbf{y} = \left[\begin{array}{cccc} y_2(N) & y_2(N-1) & \cdots & y_2(M) \end{array} \right]^T$, $\mathbf{n}_1$ and $\mathbf{n}_2$ are noise vectors with the same size as $\mathbf{y}$, $\mathbf{H}_1$ and $\mathbf{H}_2$ are Toeplitz matrices which have the same structure as

$$\mathbf{H}_1 = \left[\begin{array}{cccccc} 1 & h_1(1) & \cdots & h_1(M) & & \\ & \ddots & \ddots & & \ddots & \\ & & 1 & h_1(1) & \cdots & h_1(M) \end{array} \right].$$

In equation (5), $\mathbf{n}_1$ and $\mathbf{n}_2$ are white Gaussian sequences following the distribution $\left[\begin{array}{c} \mathbf{n}_1 \\ \mathbf{n}_2 \end{array} \right] \sim \mathcal{N}\left(0, \sigma^2 I\right)$ where σ^2 denotes the variance. $\mathbf{h}_1$ and $\mathbf{h}_2$ are coefficient vectors to be estimated from the observed data. Following the Maximum-Likelihood estimation framework, the negative log-likelihood function of the blind identification problem can be written as

$$\begin{aligned} J(\mathbf{h}) \quad = \quad & \log |\Sigma_{\mathbf{h}}| + \left(\mathbf{y} - \mathcal{Y}_1\mathbf{h}_2 + \mathcal{Y}_2\mathbf{h}_1\right)^T \Sigma_{\mathbf{h}}^{-1} \cdot \\ & \left(\mathbf{y} - \mathcal{Y}_1\mathbf{h}_2 + \mathcal{Y}_2\mathbf{h}_1\right) \end{aligned} \quad (6)$$

where $\Sigma_{\mathbf{h}} = \sigma^2(\mathbf{H}_1\mathbf{H}_1^T + \mathbf{H}_2\mathbf{H}_2^T)$. It can be found that the covariance matrix $\Sigma_{\mathbf{h}}$ depends on the unknown $\mathbf{h} = [\mathbf{h}_1^T \quad \mathbf{h}_2^T]^T$ and $J(\mathbf{h})$ is a nonlinear function of $\mathbf{h}$, which imposes challenges to achieve a global optimal solution.

IV. Computational Algorithm

A. Algorithm Description

In view of (6), its first term is a concave function and the second term is highly nonlinear. To minimize $J(\mathbf{h})$ using the standard Newton method may fail to converge to an optimal solution. Instead of applying the Newton method directly, we apply the variable splitting technique [11] on (6), which can be transformed into

$$\begin{aligned} \min \quad & J(\mathbf{h}, S) = \\ & -\log |S| + \left(\mathbf{y} - \mathcal{Y}_1\mathbf{h}_2 + \mathcal{Y}_2\mathbf{h}_1\right)^T S \left(\mathbf{y} - \mathcal{Y}_1\mathbf{h}_2 + \mathcal{Y}_2\mathbf{h}_1\right) \\ s.t. \quad & \Sigma_{\mathbf{h}} S = I. \end{aligned}$$

The above optimization problem can be solved using a fixed-point iterative approach as follows

$$\begin{aligned} & S^{k+1} = \arg\min_S J(\mathbf{h}^k, S) \text{ subject to } \Sigma_{\mathbf{h}^k} S = I, \\ & \mathbf{h}^{k+1} = \arg\min_{\mathbf{h}} J(\mathbf{h}, S^{k+1}) \text{ subject to } \Sigma_{\mathbf{h}} S^{k+1} = I. \end{aligned}$$

By fixing $\mathbf{h}$, the optimal solution of S can be obtained through solving the following optimization problem

$$\min_S \quad J(\mathbf{h}, S), \quad s.t. \quad \Sigma_{\mathbf{h}} S = I. \quad (7)$$

From the definition of the matrix $\Sigma_{\mathbf{h}} = \sigma^2(\mathbf{H}_1\mathbf{H}_1^T + \mathbf{H}_2\mathbf{H}_2^T)$, it can be easily verified that it is symmetric positive definite since both $\mathbf{H}_1$ and $\mathbf{H}_2$ are of full row rank. Thus, the optimal solution of the above optimization problem is just the unique feasible solution, i.e. $S = (\Sigma_{\mathbf{h}})^{-1}$.

Next, by fixing S, the optimal solution of $\mathbf{h}$ can be obtained by solving the following optimization problem

$$\begin{aligned} \min_{\mathbf{h}} \quad & \left(\mathbf{y} - \mathcal{Y}_1\mathbf{h}_2 + \mathcal{Y}_2\mathbf{h}_1\right)^T S \left(\mathbf{y} - \mathcal{Y}_1\mathbf{h}_2 + \mathcal{Y}_2\mathbf{h}_1\right), \\ s.t. \quad & \Sigma_{\mathbf{h}} S = I. \end{aligned} \quad (8)$$

In the above equation, $\Sigma_{\mathbf{h}}$ can be determined from the constraint condition equation, but the vector $\mathbf{h}$ cannot be uniquely determined from the matrix $\Sigma_{\mathbf{h}}$.

To solve this problem, we construct a regularized optimization problem based on the alternating direction method of multipliers (ADMM) [12]:

$$\begin{aligned} L(\mathbf{h}, U) \quad = \quad & \left(\mathbf{y} - \mathcal{Y}_1\mathbf{h}_2 + \mathcal{Y}_2\mathbf{h}_1\right)^T S \left(\mathbf{y} - \mathcal{Y}_1\mathbf{h}_2 + \mathcal{Y}_2\mathbf{h}_1\right) \\ & + (\mu/2)\|\Sigma_{\mathbf{h}} S - I + U\|_F^2, \end{aligned} \quad (9)$$

where $\mu > 0$ is small value representing the dual update step length, and U is a scaled dual variable matrix. The above optimization problem can be solved iteratively as follows:

$$\begin{aligned} & \mathbf{h}^{k+1} = \min_{\mathbf{h}} \quad L(\mathbf{h}, U^k), \\ & U^{k+1} = U^k + \mu(\Sigma_{\mathbf{h}^{k+1}} S - I). \end{aligned} \quad (10)$$

In the first step, the cost function is still nonlinear with respect to $\mathbf{h}$ but is much easier to deal with than that in (6) since no log-determinant and inverse operations are involved. Similar to the inexact Augmented Lagrange Multiplier method [13], we replace the minimization of $L(\mathbf{h}, U^k)$ by a few Newton iterations, thus assuring the decrease of the value of the cost function $L(\mathbf{h}, U^k)$.

B. Computational Issues

The standard Newton minimization of the cost function $L(\mathbf{h})$ in (9) can be carried out iteratively according to

$$\mathbf{h}^{k+1} = \mathbf{h}^k - (\nabla^2 L(\mathbf{h}^k))^{-1} \nabla L(\mathbf{h}^k)$$

where $\nabla L(\mathbf{h}^k)$ and $\nabla^2 L(\mathbf{h}^k)$ denote the gradient and the Hessian matrix of the function $L(\mathbf{h})$ evaluated at $\mathbf{h}^k$, respectively. The stopping criterion of the Newton iteration is defined in terms of the relative error

$$\frac{\|\mathbf{h}^{k+1} - \mathbf{h}^k\|^2}{\|\mathbf{h}^k\|^2} < \epsilon, \tag{11}$$

where ϵ denotes a small constant value.

In the Newton iterations, only if the initial point of $\mathbf{h}^0$ is in a convex region, we can find the local minimal point. Otherwise, it may be trapped into some unexpected local minimal point or divergent. In our implementation, the initial point of $\mathbf{h}$ is set by the solution of the CR method, which is quite close to the global optimum solution when the SNR is high. Compared with the cost function in (6), the cost function in (9) does not involve the log-determinant and inverse operations on a positive definite matrix. Thus, the computational load is much less. In addition, the value of the cost function decreases as the iterations go on. Hence, the convergence of developed algorithm is guaranteed.

Generally, the Crammer-Rao bound (CRB) is the smallest error covariance for any unbiased estimators. The Fisher Information (FI) matrix and the CRB associated with the model in (5) is given by

$$F_{i,j} = \frac{\partial^2 J(\mathbf{h})}{\partial h_i \partial h_j}, \quad \text{CRB}_{\mathbf{h}} = \text{Tr}[F^{-1}],$$

where $F_{i,j}$ is the entry at i-th row and j-th column of the matrix F, and $\text{Tr}[\cdot]$ denotes the trace operator. Being different from the TSML method [4], the Crammer-Rao bound derived here does not depend on the source signal but the true values of FIR channel functions and the output observations. If the FI matrix is positive definite, then it provides the correct CRB. Otherwise, when the FI matrix is rank deficient, the CRB will be infinitely large and channel functions $\mathbf{h}$ cannot be properly estimated.

V. Simulation Results

A series of simulation experiments are carried out to evaluate the performance of our proposed method. For comparison purpose, the CR method [1] and the TSML method [4] are also implemented and simulated. To assess the performance of the proposed method, the following factors are considered: (1) signal noise ratio (SNR) of the observed sequence; (2) the similarity of all channels; and (3) the channel estimation accuracy.

To assess the robustness of the developed algorithm against the noise effect. The SNR of the observed data generated by the model (5) is defined as

$$\text{SNR} = 10 \log_{10} \left(\sum_{t=1}^{N} \|\mathbf{h}(t) * s(t)\|_2^2 \right) / \left(NL\sigma^2 \right),$$

where N is the length of observed data, σ^2 is the variance of the additive noise, and $NL\sigma^2$ in the denominator denotes the noise power of L channels.

As shown in Assumption 1), all channels should be coprime in order to obtain a unique estimate of FIR channels up a scalar constant. In practical applications, if zeros in different channel functions are quite close to each other, the matrix Y in (3) will be ill-conditioned and the estimation of channel functions is sensitive to the computational errors. Here, the degree of closeness among different channels can be measured by the condition number of the associated Sylvester matrix. To demonstrate the robustness of the developed algorithm, two cases with different irreducibilities (closeness of different channels) will be examined below.

The common criterion to describe the accuracy of channel estimation is defined by

$$\text{nMSE} = 10 \log_{10} E \left(\frac{\min_\alpha \|\alpha \hat{\mathbf{h}} - \mathbf{h}\|_2^2}{\|\mathbf{h}\|_2^2} \right)$$

where α denotes the scalar ambiguity for channel estimation, $\hat{\mathbf{h}}$ denotes the estimated channel sequences, and E denotes the mathematical expectation which is computed from 500 Menta Carlo runs. In this paper, the scalar ambiguity is eliminated by fixing the value of $h_1(0)$, thus $\alpha = 1$.

In the simulation, the source signal is randomly generated such that the Assumption 3) can be assured. The considered SIMO system contains two FIR channels with the same length. Each available output sequence is of length 200, which is much smaller than that of other statistical identification algorithms. The derived algorithm in this paper is applicable for both the real and complex signals, and real cases are examined here for simplicity but without loss of generality.

Two sets of FIR channels for simulations are tabulated in Table I. The first group has strong irreducibility with a condition number 4.13 and the second one has weak irreducibility whose condition number is 52.62. The identification performances on two different sets of FIR channels under different SNRs are displayed in Fig. 1 and 2, respectively. The nMSE curves are obtained by the CR method, the TSML method and our proposed method respectively. It is shown that our method outperforms other methods in terms of nMSE. The worse performance of the CR method is obvious since it does not consider the noise effect properly. The TSML method only involves two iterations so that the corresponding estimation is not locally (or globally)

TABLE I
COEFFICIENTS FOR FIR CHANNELS

	Group 1		Group 2	
	$h_1(n)$	$h_2(n)$	$h_1(n)$	$h_2(n)$
$n=0$	1.0000	1.0000	1.0000	1.0000
$n=1$	-0.3222	1.5292	-1.2910	0.3509
$n=2$	-0.1696	0.2649	1.1954	0.8129
$n=3$	-0.3540	0.2311	-1.2374	1.2722
$n=4$	-0.0968	-0.2259	-0.1352	1.5546
$n=5$	1.8210	0.6573	-0.1729	-0.1847

optimal.

In Fig. 1, the nMSE curves are obtained using the first group of FIR channels with strong irreducibility. Compared with the results in Fig. 2 which are obtained using the second group of FIR channels with weak irreducibility, we can find that the identification performance using the first group of FIR channels is much better when the SNR value is fixed. In addition, from Fig. 2, we can find that the identification results demonstrate more significant advantage of our proposed algorithm at low SNRs.

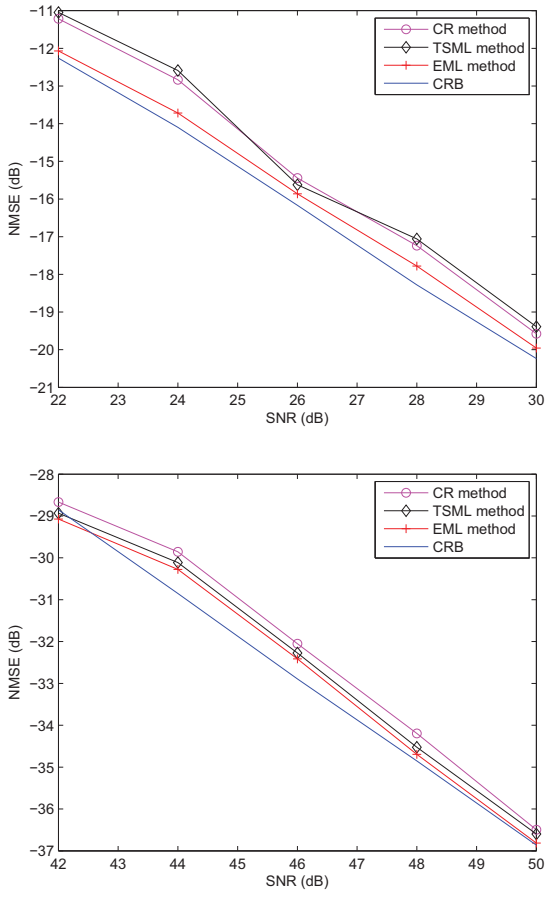

Fig. 1. Performances of the CR, TSML, and EML method versus SNR when all channels have strong irreducibility.

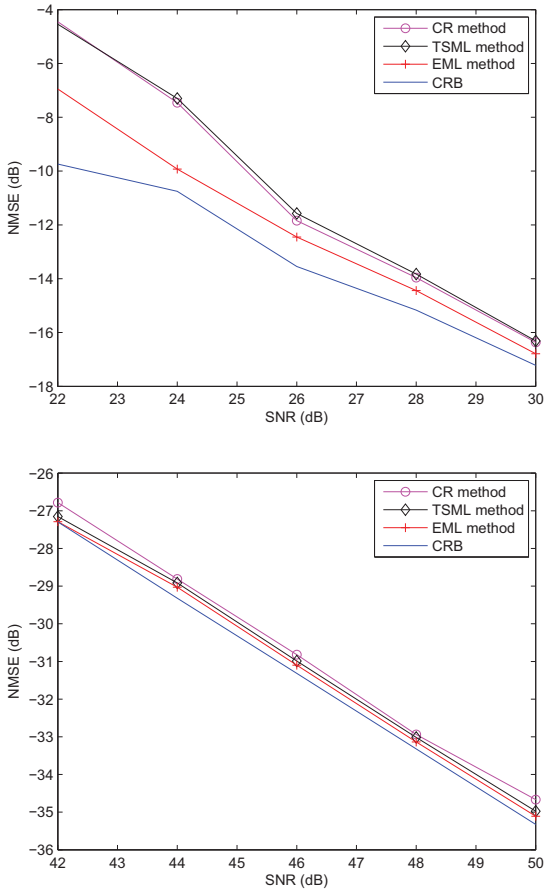

Fig. 2. Performances of the CR, TSML, and EML method versus SNR when all channels have weak irreducibility

VI. CONCLUSION

This paper has proposed an exact Maximum-Likelihood method for blind multi-channel identification. The Maximum-Likelihood problem is formulated based on the cross relations among different FIR channels by normalizing the first coefficient. The major difficulty to minimize the negative log-likelihood function lies in its high nonlinearity with respect to the FIR channels. To ease the numerical computation and improve the algorithm's convergence, an iterative optimization approach based on the variable split technique and the ADMM algorithm is derived. Finally, simulation results show that our proposed method outperforms the CR method and the TSML method in terms of nMSE.

The main contribution of our work is to formulate an exact Maximum-Likelihood estimator and develop an effective approach to resolve the associated nonlinear optimization problem. The proposed identification algorithm can achieve a local optimal solution, which depends on the choice of the initialization point at the beginning of the iterative procedure. However, in order to improve its applicability, more efforts are needed such that a global optimal solution can be achieved.

REFERENCES

[1] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Transactions on Signal Processing*, vol. 43, pp. 2982–2993, Dec 1995.

[2] K. AbedMeraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Transactions on Signal Processing*, vol. 45, pp. 694–705, Mar 1997.

[3] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proceedings of the IEEE*, vol. 86, pp. 1951–1968, Oct 1998.

[4] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Transactions on Signal Processing*, vol. 44, pp. 661–672, Mar 1996.

[5] R. Lopez-Valcarce and S. Dasgupta, "Blind channel equalization with colored sources based on second-order statistics: A linear prediction approach," *IEEE Transactions on Signal Processing*, vol. 49, pp. 2050–2059, Sep 2001.

[6] E. MOULINES, P. DUHAMEL, J. CARDOSO, and S. MAYRARGUE, "Subsapce Methods for the Blind Identification of Multichannel FIR Filters," *IEEE Transactions on Signal Processing*, vol. 43, pp. 516–525, Feb 1995.

[7] C. Yu, C. Zhang, and L. Xie, "Blind Identification of Multi-Channel ARMA Models Based on Second-Order Statistics," *To appear in IEEE Transactions on Signal Processing*, 2012.

[8] W. Qiu, S. K. Saleem, and M. Pham, "Blind identification of multichannel systems driven by impulsive signals," *Digital Signal Processing*, vol. 20, pp. 736–742, May 2010.

[9] L. Tong and Q. Zhao, "Joint order detection and blind channel estimation by least squares smoothing," *IEEE Transactions on Signal Processing*, vol. 47, pp. 2345–2355, Sep 1999.

[10] Y. BRESLER and A. MACOVSKI, "Exact Maximum-Likelihood Parameter-Estimation of Superimposed Exponential Signals in Noise," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 34, pp. 1081–1089, Oct 1986.

[11] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast Image Recovery Using Variable Splitting and Constrained Optimization," *IEEE Transactions on Image Processing*, vol. 19, pp. 2345–2356, Sep 2010.

[12] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An Augmented Lagrangian Approach to the Constrained Optimization Formulation of Imaging Inverse Problems," *IEEE Transactions on Image Processing*, vol. 20, pp. 681–695, Mar 2011.

[13] Z. Lin, M. Chen, and L. Wu, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *Analysis*, vol. math.OC, pp. 2209–2215, 2009.