ON STRUCTURED TOTAL LEAST-SQUARES FOR BLIND IDENTIFICATION OF MULTICHANNEL FIR FILTERS

Muhammad Z. Ikram

DSP Solutions R&D Center
Texas Instruments, Inc., Dallas, TX 75243
mzi@ti.com

ABSTRACT

Structured total least-squares (STLS) provides a nice framework for approximating a full-rank affinely-structured matrix with a rank-deficient matrix having the same affine structure. In this paper, we investigate the use of STLS method for blind identification of multiple FIR channels driven by an unknown deterministic input. First, we exploit the block-Hankel affine structure of the data matrix, which motivates the use of STLS-based methods. Then, we derive an iterative non-linear solution to the unknown channel parameters by using a generalized form of singular value decomposition. We carry out extensive numerical simulations to compare the performance of the proposed method against the well-known least-squares (LS) method, where the affine structure of the date matrix is overlooked. These results reveal that the STLS based method outperforms the LS method for ill-conditioned as well as well-conditioned channels over a wide range of SNR.

Index Terms— Blind Channel Identification, Structured Total Least Squares, Hankel Matrix

1. INTRODUCTION

Blind channel identification (BCI) refers to the estimation of channel impulse response from its output only. Over the past few years, this area of research has received immense practical interest (see [1], and the references therein). The primary reason for this is the fact that it does not require a training sequence to equalize the channel, thereby saving the channel capacity, which is severely limited in mobile communications. Some earlier work in this direction focused on using higher-order statistics (HOS), which require the use of long data records for reliable estimates. This limits their usage in mobile communications where the channel has to be estimated within a short period of time.

Blind estimation of FIR channels using only second-order statistics (SOS) is first attributed to Tong *et al.* who showed that by sampling the received data at a rate higher than the baud (symbol) rate, SOS can suffice to estimate the channel impulse response up to a constant [2]. Ever since the publication of [2], many different statistical and deterministic approaches for BCI have been proposed, each having its own relative merits and demerits [1].

In this work, we consider BCI *without* assuming any knowledge of the input signal statistics. This is more realistic in mobile communications, where a short data sequence may not yield reliable statistics [8]. A deterministic least-squares (LS) method was proposed by Xu *et al.* [3], which is based on the cross-relation property between the outputs of two channels fed by the same input. A striking feature of the LS method is that for an arbitrary input and in the absence of noise, it yields exact channel estimates. In the presence of noise, the channel parameter estimates are obtained by solving an over-determined set of linear equations which translates to obtaining a reduced rank solution of a block-Hankel structured matrix. Unfortunately, in the presence of noise, the LS method fails to associate this specific structure with the reduced rank matrix. In this contribution, we constrain the LS solution so that the reduced rank data matrix also assumes a block-Hankel structure, as desired. This constrained optimization problem is referred to as structured total least-squares (STLS) [4], whose solution is based on solving a set of non-linear equations. We present a BCI method using STLS where we follow an iterative approach of [4] to estimate the channel parameters. To illustrate the efficacy of our proposed method, we carry out numerical simulations, which compares the performance of the LS and the proposed STLS based blind channel estimation methods.

2. PROBLEM FORMULATION

Consider a P-channel FIR system having impulse responses $h_i(n), i = 1, \ldots, P$, driven by a common input $s(n)$. The outputs of this multichannel system are given by

$$y_i(n) = \sum_{l=0}^{L} h_i(l)s(n-l) + n_i(n), \quad i = 1, \ldots, P, \quad (1)$$

where L is the order of P FIR filters, i.e., $h_i(z) = \sum_{l=0}^{L} h_i(l)z^{-l}$, which is assumed to be known. Note that the P-channel model can be obtained either by using P physical receivers or by oversampling the single channel output at P times the baud rate [1]. Our objective in this paper is to estimate the channels $h_i(n)$ from N output samples $y_i(n)$ of P channels $i = 1, \ldots, P$. We assume that: **A(1)** the P FIR filters are coprime; i.e., they do not share any common zeros; **A(2)** the number of modes in the input sequence is $L+1$, and **A(3)** the additive noise $n_i(n)$ is stationary and white.

For a pair of noise-free outputs of any two sensors i and j, we can write

$$x_i(n) = h_i(n) * s(n), \quad x_j(n) = h_j(n) * s(n),$$

which allows us to write [3]

$$h_i(n) * x_j(n) = h_j(n) * x_i(n).$$

In a compact matrix notation

$$\begin{pmatrix} \mathbf{X}_i & -\mathbf{X}_j \end{pmatrix} \begin{pmatrix} \mathbf{h}_j \\ \mathbf{h}_i \end{pmatrix} = \mathbf{0}, \tag{2}$$

where $\mathbf{h}_m = [h_m(L), \ldots, h_m(0)]^T$,

$$\mathbf{X}_m = \begin{pmatrix} x_m(L) & x_m(L+1) & \ldots & x_m(2L) \\ x_m(L+1) & x_m(L+2) & \ldots & x_m(2L+1) \\ \vdots & \vdots & \ddots & \vdots \\ x_m(N-L) & x_m(N-L+1) & \ldots & x_m(N) \end{pmatrix}, \tag{3}$$

and $(\cdot)^T$ denotes the transpose. The cross-relation in (2) is the main idea behind the LS approach of [3] and the STLS approach proposed in this paper, as will be shown in the next two sections.

3. MOTIVATION FOR STLS BASED APPROACH

The cross-relation of (2) can straightforwardly be expressed for each (i, j) pair of channels, which when combined together gives

$$\begin{pmatrix} \mathbf{X}^1 \\ \mathbf{X}^2 \\ \vdots \\ \mathbf{X}^{P-1} \end{pmatrix} \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_P \end{pmatrix} = \mathbf{0} \tag{4}$$

or simply $\mathbf{Xh} = \mathbf{0}$, where $\mathbf{X}$ is of dimension $(N-2L+1)\frac{P(P-1)}{2} \times (L+1)P$ and $\mathbf{h}$ is a $(L+1)P \times 1$ vector. Each constituent block of $\mathbf{X}$ is of dimension $(N-2L+1)(P-i) \times (L+1)P$ and is given by

$$\mathbf{X}^i = \begin{pmatrix} \mathbf{0} & \ldots & \mathbf{0} & \mathbf{X}_{i+1} & -\mathbf{X}_i & \mathbf{0} & \mathbf{0} \\ & \vdots & & \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{X}_P & \mathbf{0} & \ldots & -\mathbf{X}_i \end{pmatrix}. \tag{5}$$

With the identifiability conditions **(A1)–(A3)** satisfied, the matrix $\mathbf{X}$, also sometimes referred to as the data selection matrix, is rank deficient by one, and $\mathbf{h}$ is in its null space. Given a noisy data matrix $\mathbf{Y}$, an estimate of the channel parameters can then be obtained by finding the low-rank approximation $\widehat{\mathbf{Y}}$ of the noisy data selection matrix $\mathbf{Y}$. In other words, we seek a LS solution

$$\min \|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2 \text{ subject to } \begin{cases} \widehat{\mathbf{Y}}\mathbf{h} = 0 \\ \mathbf{h}^T\mathbf{h} = 1 \end{cases}, \tag{6}$$

where $\|\cdot\|$ is the Frobenius norm. The first constraint in (6) guarantees the rank-deficiency of $\widehat{\mathbf{Y}}$ whereas the second constraint avoids the trivial solution $\mathbf{h} = \mathbf{0}$. Using the Eckart-Young-Mirsky theorem, one can easily show that the LS solution can be obtained from the dyadic singular value decomposition (SVD) of $\mathbf{Y}$ [5], i.e.,

$$\widehat{\mathbf{Y}} = \mathbf{Y} - \mathbf{u}\sigma\mathbf{v}^T \text{ and } \mathbf{h} = \mathbf{v},$$

where $\mathbf{u}$ and $\mathbf{v}$ are the left and right singular vectors of $\mathbf{Y}$ corresponding to its smallest singular value σ.

It is interesting to note that the matrix $\mathbf{Y}$ has a unique block-Hankel structure associated with it. On the other hand, its low-rank approximant $\widehat{\mathbf{Y}}$, obtained using (6), is *not* guaranteed to have the same structure. In the BCI problem, the matrix $\widehat{\mathbf{Y}}$ is desired to have a block-Hankel structure as in (4). Motivated by this fact, we now investigate and search for an estimate $\widehat{\mathbf{Y}}$ which is guaranteed to have the same structure as $\mathbf{Y}$. In this case, the solution belongs to a set of matrices S, which satisfies the following property

$$S = \widehat{\mathbf{Y}} \text{ given} \begin{cases} \widehat{\mathbf{Y}}\mathbf{h} = \mathbf{0} \\ \widehat{\mathbf{Y}} \text{ has the same block Hankel structure as } \mathbf{Y}. \end{cases}$$

By adding another constraint, the optimization problem can no longer be solved simply using the SVD. In fact, as will be shown in the next section, a non-linear estimation procedure is required to solve this problem. The use of STLS for blind channel identification was indicated by De Moor in [6]. No formal theoretical solution and numerical performance details were, however, given there.

4. PROPOSED STLS BASED BCI

For the sake of simplicity, we consider a 2-channel case, i.e., $P = 2$. The matrix $\mathbf{X}$, therefore, consists of only two row sub-blocks $\mathbf{X}_2$ and $-\mathbf{X}_1$, and is a function of $2(N-L+1)$ data values $x_i(L), \ldots, x_i(N)$ for $i = 1, 2$. In terms of a set of fixed matrices $\mathbf{B}_i, i = 0, \ldots, 2N-2L+1$, this affine matrix can be expressed as

$$\begin{aligned} \mathbf{B}(x) = \mathbf{X} = x_2(L)\mathbf{B}_0 + x_2(L+1)\mathbf{B}_1 + \ldots + x_2(N)\mathbf{B}_{N-L} \\ + x_1(L)\mathbf{B}_{N-L+1} + \ldots + x_1(N)\mathbf{B}_{2N-2L+1}, \end{aligned} \tag{7}$$

where $\mathbf{B}_i$ ($i = 0, \ldots, N-L+1$) has 1 along the $(i+1)$th anti-diagonal of the left $(N-2L+1) \times (L+1)$ block and zeros elsewhere. Similarly, $\mathbf{B}_i$ ($i = N-L+2, \ldots, 2N-2L+1$) has -1 along the $(i+1)$th anti-diagonal of the right $(N-2L+1) \times (L+1)$ block and zeros elsewhere. For example,

$$\mathbf{B}_1 = \left.\left(\begin{array}{c|c} \underbrace{\begin{matrix} 0 & 1 & \ldots & 0 \\ 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{matrix}}_{(L+1)} & \underbrace{\begin{matrix} 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{matrix}}_{(L+1)} \end{array} \right)\right\} (N-2L+1).$$

Based on the above formulation and given the noisy data $y_i(n)$, we define a STLS criteria to estimate the block-Hankel structured low-rank matrix $\mathbf{B}(\widehat{y})$, which consists of minimizing the following cost function

$$\min_{\widehat{y}_1(i),\widehat{y}_2(i),\mathbf{h}} \sum_{i=L}^{N} \left\{ [y_1(i) - \widehat{y}_1(i)]^2 + [y_2(i) - \widehat{y}_2(i)]^2 \right\}$$
$$\text{subject to} \left\{ \begin{array}{l} \mathbf{B}(\widehat{y})\mathbf{h} = \mathbf{0} \\ \mathbf{h}^T\mathbf{h} = 1 \end{array} \right. . \tag{8}$$

Using Lagrange's multipliers, a solution to (8) is obtained by finding the vectors $\mathbf{u}$ and $\mathbf{v}$ corresponding to the smallest scalar σ that satisfies [4]

$$\mathbf{Y}\mathbf{v} = \mathbf{D_v}\mathbf{u}\sigma, \quad \mathbf{u}^T\mathbf{D_v}\mathbf{u} = 1, \tag{9}$$
$$\mathbf{Y}^t\mathbf{u} = \mathbf{D_u}\mathbf{v}\sigma, \quad \mathbf{v}^T\mathbf{D_u}\mathbf{v} = 1 \tag{10}$$

where $\mathbf{D_v}$ and $\mathbf{D_v}$ are symmetric positive definite matrices defined as

$$\sum_{i=1}^{N-L+1} \mathbf{B}_i^T(\mathbf{u}^T\mathbf{B}_i\mathbf{v})\mathbf{u} = \mathbf{D_u}\mathbf{v}, \quad \sum_{i=1}^{N-L+1} \mathbf{B}_i(\mathbf{u}^T\mathbf{B}_i\mathbf{v})\mathbf{v} = \mathbf{D_v}\mathbf{u}. \tag{11}$$

The channel estimates are given by $\widehat{\mathbf{h}} = \mathbf{v}/\|\mathbf{v}\|$. Note that the non-linearity of (9)–(10) stems from the dependence of $\mathbf{D_u}$ and $\mathbf{D_v}$ on the square of the singular vectors $\mathbf{u}$ and $\mathbf{v}$ respectively. This makes the analytical solution to (9)–(10) rather difficult to compute. Independent research efforts due to Abatzoglou and Mendel [7] and De Moor [4] led to iterative solutions to this non-linear optimization problem. The method in [4], however, has shown to posses better convergence properties and will, therefore, be used in our BCI setup. The idea is to treat $\mathbf{D_u}$ and $\mathbf{D_v}$ independent of $\mathbf{u}$ and $\mathbf{v}$ in each iteration, and compute the QR decomposition of $\mathbf{Y}$, i.e.,

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{pmatrix} \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \end{pmatrix}. \tag{12}$$

where $\mathbf{Q}_1$ is $(N-2L+1) \times 2(L+1)$, $\mathbf{Q}_2$ is $(N-2L+1) \times (N-1)$, and $\mathbf{R}$ is a $2(L+1)\times 2(L+1)$ matrix. Equations (9)–(10) can then be manipulated to obtain the following upper triangular form

$$\begin{pmatrix} \mathbf{R}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{Q}_2^T\mathbf{D_v}\mathbf{Q}_1 & \mathbf{Q}_2^T\mathbf{D_v}\mathbf{Q}_2 & \mathbf{0} \\ \mathbf{Q}_1^T\mathbf{D_v}\mathbf{Q}_1\sigma & \mathbf{Q}_1^T\mathbf{D_v}\mathbf{Q}_2\sigma & -\mathbf{R} \end{pmatrix} \begin{pmatrix} \mathbf{z} \\ \mathbf{w} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{D_u}\mathbf{v}\sigma \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \tag{13}$$

where $\mathbf{u} = \mathbf{Q}_1\mathbf{z} + \mathbf{Q}_2\mathbf{w}$ for certain arbitrary vectors $\mathbf{z}$ and $\mathbf{w}$ of lengths $2(L+1)$ and $(N-1)$ respectively. This set of equations allows one to efficiently compute $\mathbf{u}$ and $\mathbf{v}$, which are then used to update $\mathbf{D_u}$ and $\mathbf{D_v}$ for the next iteration.

In summary, the proposed STLS based BCI algorithm can be summarized as follows:

1. Given the noisy data $y_i(n)$, $n = L, \ldots, N$ for $i = 1, 2$, construct the block-Hankel matrix $\mathbf{Y}$ of the form defined in (4).
2. Compute the QR decomposition of $\mathbf{Y}$.

Initialization:

3. Compute the eigenvector corresponding to the smallest eigenvalue of $\mathbf{Y}$ as an initial choice for $\mathbf{u}^{(0)}$ and $\mathbf{v}^{(0)}$. Note that this choice gives the channel estimates in the LS sense.
4. Choose normalized $\mathbf{D_u}^{(0)}$ and $\mathbf{D_v}^{(0)}$, such that
$$\mathbf{v}^{(0)^T}\mathbf{D_u}^{(0)}\mathbf{v}^{(0)} = \mathbf{u}^{(0)^T}\mathbf{D_v}^{(0)}\mathbf{u}^{(0)} = 1.$$

For $k = 1, \ldots$:

5. $\mathbf{z}^{(k)} = \mathbf{R}^{-T}\mathbf{D_u}^{(k-1)}\mathbf{v}^{(k-1)}\sigma^{(k-1)}$
6. $\omega^{(k)} = -\left(\mathbf{Q}_2^T\mathbf{D_v}^{(k-1)}\mathbf{Q}_2\right)^{-1}\left(\mathbf{Q}_2^T\mathbf{D_v}^{(k-1)}\mathbf{Q}_1\right)\mathbf{z}^{(k)}$
7. $\mathbf{u}^{(k)} = \mathbf{Q}_1\mathbf{z}^{(k)} + \mathbf{Q}_2\mathbf{w}^{(k)}$
8. $\mathbf{v}^{(k)} = \mathbf{R}^{-1}\mathbf{Q}_1^T\mathbf{D_v}^{(k-1)}\mathbf{u}^{(k)}$, $\mathbf{v}^{(k)} = \mathbf{v}^{(k)}/\|\mathbf{v}^{(k)}\|$
9. $\gamma^{(k)} = \left(\mathbf{u}^{(k)^T}\mathbf{D_v}^{(k)}\mathbf{u}^{(k)}\right)^{1/4}$
10. $\mathbf{u}^{(k)} = \mathbf{u}^{(k)}/\gamma^{(k)}$, $\mathbf{v}^{(k)} = \mathbf{v}^{(k)}/\gamma^{(k)}$
11. $\mathbf{D_u}^{(k)} = \mathbf{D_u}^{(k)}/(\gamma^{(k)})^2$, $\mathbf{D_v}^{(k)} = \mathbf{D_v}^{(k)}/(\gamma^{(k)})^2$
12. $\sigma^{(k)} = \mathbf{u}^{(k)^T}\mathbf{Y}\mathbf{v}^{(k)}$
13. *Convergence:* If $\|\mathbf{v}^{(k)} - \mathbf{v}^{(k-1)}\| \leq \epsilon$ (a pre-determined scalar), then stop. Otherwise, repeat step 5–13.

At convergence ($k = K$), the structured low rank matrix $\widehat{\mathbf{Y}}$ can be estimated using

$$\mathbf{B}(\widehat{y}) = \sum_{i=0}^{N-L} \mathbf{B}_i\widehat{y}_2^{(K)}(L+i) + \sum_{i=N-L+1}^{2N-2L+1} \mathbf{B}_i\widehat{y}_1^{(K)}(2L-N-1+i), \tag{14}$$

where $\widehat{y}_1^{(K)}(L+i) = y_1(L+i) - \mathbf{u}^{(K)^T}\mathbf{B}_i\mathbf{v}^{(K)}\sigma^{(K)}$ and $\widehat{y}_2^{(K)}(L+i) = y_2(L+i) - \mathbf{u}^{(K)^T}\mathbf{B}_{N-L+1+i}\mathbf{v}^{(K)}\sigma^{(K)}$ for $i = 0, \ldots, N-L$. Finally, the STLS channel estimates $\mathbf{h}$ are obtained from the null space of $\mathbf{B}(\widehat{y})$.

5. SIMULATION RESULTS

In this section, we investigate the performance of the proposed STLS based method and compare it against the LS method. To study the effect of ill-conditioned channels, we consider a two-channel second-order FIR model ($L = 2$, $P = 2$) proposed in [8]

$$\mathbf{h}_i = [1 \quad -2\cos(\theta_i) \quad 1]^T, \quad i = 1, 2,$$

where $\theta_1 = \theta$, $\theta_2 = \theta_1 + \delta$, and δ is the parameter that controls the angular separation between the zeros of the two channels. When δ is small, the channels are considered ill-conditioned. We consider the input $s(n)$ as an i.i.d. sequence of $(+1, -1)$ having unit variance $\sigma_s^2 = 1$. The output data consist of $N = 100$ samples. The outputs of the two channels are corrupted by additive white Gaussian noise of variance σ_n^2. The signal to noise ratio in dB is defined using

$$\text{SNR} = 10 \log_{10} \left(\frac{\sigma_s^2 \|\mathbf{h}\|^2}{M \sigma_n^2} \right)$$

and the performance of STLS and LS methods is evaluated using the mean squared error in dB

$$\text{MSE} = 20 \log_{10} \left(\frac{1}{\|\mathbf{h}\|} \sqrt{\frac{1}{N_r} \sum_{i=1}^{N_r} \|\widehat{\mathbf{h}}^i - \mathbf{h}\|^2} \right),$$

where $\widehat{\mathbf{h}}^i$ is the i-th estimate of channel parameters and $N_r = 100$ is the number of independent Monte-Carlo runs. To alleviate the effect of scaling in the estimates, the first element of $\widehat{\mathbf{h}}^i$ is normalized to be one.

We first considered well-conditioned channels for which $\theta = \pi/10$ and $\delta = \pi$. In this case, the performance of the LS and STLS methods against the SNR is shown in Fig. 1.

In the next experiment, we computed the MSE for ill-conditioned channels ($\theta = \pi/10, \delta = \pi/5$) at different SNR levels. The performance of both the methods is shown in Fig. 2. As we see that by preserving the block-Hankel structure of the data matrix in its low rank approximant, better channel estimates are obtained in terms of MSE, thus supporting our theoretical claims.

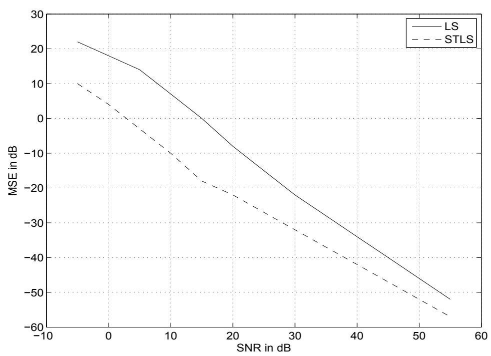

Fig. 1. Performance comparison of LS and STLS methods for well-conditioned channels ($\delta = \pi$) with data size $N = 100$.

For ill-conditioned channels, the STLS method also proves to be more data efficient as shown in Fig. 3, where its performance is compared against LS method for different values of data size N at an SNR of 5 dB.

6. CONCLUSIONS

We proposed an algorithm for blind estimation of multiple FIR channels driven by an arbitrary unknown input. The method uses the cross-relation approach and exploits the specific block-Hankel structure of the data selection matrix. A decomposition of this matrix in terms of an affine set of matrices is defined and the channel estimates are obtained by solving a non-linear constrained optimization problem. Simulation results show the improved performance of the proposed method over the LS method.

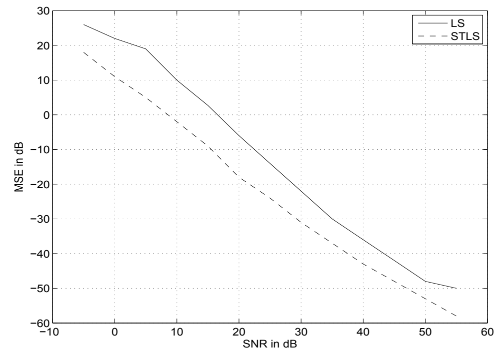

Fig. 2. Performance comparison of LS and STLS methods for ill-conditioned channels ($\delta = \pi/5$) with data size $N = 100$.

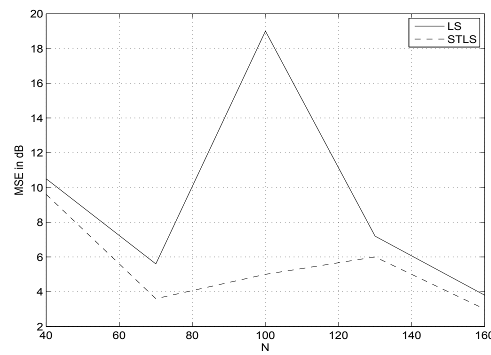

Fig. 3. Performance comparison of LS and STLS methods for ill-conditioned channels ($\delta = \pi/5$) and different data sizes N at an SNR of 5 dB.

7. REFERENCES

[1] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.

[2] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Proc. 25th Asilomar Conf. Signals, Syst. Comp.*, 1991, pp. 856–860.

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

[4] B. De Moor, "Total least squares for affinely structured matrices and noisy realization problems," *IEEE Trans. Signal Processing*, vol. 42, no. 11, pp. 3104–3113, Nov. 1994.

[5] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 1996.

[6] B. De Moor, "Linear system identification, structured total least squares and the Riemannian SVD," in *Proc. 2nd Int. Workshop on Total Least Squares and Errors-in-Variables Modeling*, pp. 225–238, 1996.

[7] T. J. Abatzoglou, J. M. Mendel, and G. A. Harada, "The constrained total least squares and its application to harmonic superresolution," *IEEE Trans. Signal Processing*, vol. 39, no. 5, pp. 1070–1087, May 1991.

[8] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Processing*, vol. 44, no. 3, pp. 661–672, Mar. 1996.