Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2007, Article ID 12172, 14 pages
doi:10.1155/2007/12172

Research Article

Blind Identification of FIR Channels in the Presence of Unknown Noise

Xiaojuan He and Kon Max Wong

Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada L8S 4K1

Received 23 December 2005; Revised 20 July 2006; Accepted 29 October 2006

Recommended by Markus Rupp

Blind channel identification techniques based on second-order statistics (SOS) of the received data have been a topic of active research in recent years. Among the most popular is the subspace method (SS) proposed by Moulines et al. (1995). It has good performance when the channel output is corrupted by white noise. However, when the channel noise is correlated and unknown as is often encountered in practice, the performance of the SS method degrades severely. In this paper, we address the problem of estimating FIR channels in the presence of arbitrarily correlated noise whose covariance matrix is unknown. We propose several algorithms according to the different available system resources: (1) when only one receiving antenna is available, by upsampling the output, we develop the maximum a posteriori (MAP) algorithm for which a simple criterion is obtained and an efficient implementation algorithm is developed; (2) when two receiving antennae are available, by upsampling both the outputs and utilizing *canonical correlation decomposition* (CCD) to obtain the subspaces, we present two algorithms (CCD-SS and CCD-ML) to blindly estimate the channels. Our algorithms perform well in unknown noise environment and outperform existing methods proposed for similar scenarios.

1. INTRODUCTION

Channel distortion remains one of the hurdles in high-fidelity data communications because the performance of a digital communication system is invariably affected by the characteristics of the channel over which the signals are transmitted as well as by additive noise. The effects of the channel often manifest themselves as distortions to the transmitted signals in the form of intersymbol interference (ISI), cross-talks, fading, and so forth [2]. Mitigation of such effects is often carried out by filtering, channel equalization, and appropriate signal designs for which a proper knowledge of the channel characteristics is required. Thus, channel estimation is a very important process in digital communications. Traditionally, channel estimation is carried out by observing the received pilot signals sent over the channel and various algorithms for identifying the channel have been developed based on the transmission of pilot signals [3–5]. However, the insertion of pilot signals often means a decrease of bandwidth efficiency, and the resulting limitation of effective data throughput [6] may be a substantial penalty in performance. Thus, blind identification of the channel could be helpful.

Since the pioneering work of Tong et al. [7], a number of blind channel estimation algorithms based on second-order statistics (SOS) have been proposed. A popular method is the subspace (SS) method [1] which performs well in a white noise environment. However, in practice, this method degrades seriously because the "white noise" assumption is seldom satisfied in reality. In addition, cochannel interference often modeled as noise is generally nonwhite and unknown [8]. For practical applications therefore, channel estimation algorithms capable of dealing with arbitrary noise are necessary.

It is proposed in [9] that the noise covariance matrix be iteratively estimated by trying to fit it into an assumed special band-Toeplitz structure, and then be subtracted from the received data covariance so that the SS method can be applied. The estimation of the noise in this way may suffer from being subjective. Thus, algorithms which obviate noise estimation may be more desirable. A modified subspace (MSS) method was proposed in [10] transmitting two adjacent nonoverlapping signal sequences. Due to the channel response, the received signal vectors will overlap. By making use of the fact that the noise in the received signal vectors is uncorrelated, the SS method can then be applied. However, this algorithm depends on the signal property, and severe restrictions on the length of the transmitted signal sequences may have to be imposed for the method to be applicable. More recently, a *semiblind* ML estimator of single-input multiple-output flat-fading channels in spatially correlated noise is proposed in [11]. On the other hand, applying the EM algorithm to evaluate ML, the channel coefficients and the spatial noise covariance can be computed [12], and this estimator is also proposed for estimating space-time fading channels under unknown spatially correlated noise.

In this paper, based on SOS, we consider different system models having unknown correlated noise environments and accordingly develop different algorithms for the estimation of the channel. Natural and man-made noise in wireless communications can occur as both temporally and spatially correlated. These include electromagnetic pickup of radiating signals, switching transients, atmospheric disturbances, extra-terrestrial radiation, internal circuit noise, and so forth. If only one transmitter antenna and one receiver antenna are available in the communication system, we only have to deal with temporally correlated noise, and for this case, we develop the maximum a posteriori (MAP) criterion utilizing Jeffreys' Principle. On the other hand, if two (or more) receiving antennae are available (such as in the case of a base station), we may encounter noise which is both temporally and spatially correlated. However, since spatial correlation of noise is negligible when the two receiving antennae are separated by more than a few wavelengths of the transmission carrier [13], a condition not hard to satisfy in the case of a base station, therefore, we assume in this paper, that the noise vectors from the two antennae are uncorrelated while the temporal correlation of the individual noise vector still maintains. For this case, we employ the *canonical correlation decomposition* (CCD) [14, 15] for identifying the subspaces and forming the corresponding projectors, and develop a subspace-based algorithm (CCD-SS) and a maximum likelihood-based algorithm (CCD-ML) for the estimation of the channel. Computer simulations show that all these methods achieve superior performance to the MSS method under different signal-to-noise ratio (SNR).

2. SYSTEM MODEL AND SUBSPACE CHANNEL ESTIMATION

2.1. System model

The output of a linear time-invariant complex channel can be represented in baseband as

$$r(t) = \sum_{k=0}^{+\infty} s(k)h(t-kT) + \eta(t), \tag{1}$$

where T is the symbol period, $\{s(k)\}$ is the sequence of complex symbols transmitted, $h(t)$ is the complex impulse response of the channel, and $\eta(t)$ is the additive complex noise process independent of $\{s(k)\}$. Since most channels have impulse responses approximately finite in time support, we can assume that $h(t) = 0$ for $t \notin [0, LT]$, where $L > 0$ is an integer, that is, we consider FIR channels with maximum channel order L. Let the received signal $r(t)$ be upsampled by a positive integer M. Then, the upsampled received signal $r(t_0 + mT/M)$ can be divided into M subsequences such that

$$r_m(n) = \sum_{\ell=0}^{L} h_m(\ell)s(n-\ell) + \eta_m(n), \quad m = 1, 2, \ldots, M, \tag{2}$$

where $r_m(n) = r(t_0 + nT + (m-1)T/M)$, $h_m(n) = h(t_0 + nT + (m-1)T/M)$, $\eta_m(n) = \eta(t_0 + nT + (m-1)T/M)$, $m = 1, 2, \ldots, M$. Clearly, these M subsequences can be conveniently viewed as outputs of M discrete FIR channels with a common input sequence $\{s(n)\}$. At time instant n, the upsampled received signal can now be represented in vector form at the symbol rate as

$$\mathbf{r}_o(n) = \sum_{\ell=0}^{L} \mathbf{h}(\ell)s(n-\ell) + \boldsymbol{\eta}_o(n) = \mathbf{H}_o\mathbf{s}_o(n) + \boldsymbol{\eta}_o(n), \tag{3}$$

where

$$\begin{aligned} \mathbf{s}_o(n) &= \left[s(n)\cdots s(n-L)\right]^T, \\ \mathbf{r}_o(n) &= \left[r_1(n)\cdots r_M(n)\right]^T, \\ \boldsymbol{\eta}_o(n) &= \left[\eta_1(n)\cdots \eta_M(n)\right]^T, \end{aligned} \tag{4}$$

$$\mathbf{H}_o = \left[\mathbf{h}(0)\mathbf{h}(1)\cdots\mathbf{h}(L)\right] \tag{5}$$

with $\mathbf{h}(\ell) = [h_1(\ell)\cdots h_M(\ell)]^T$. Assume the channel is invariant during the time period of K symbols, then the received $MK \times 1$ signal vector can be represented as

$$\begin{aligned} \mathbf{r}(n) &= \left[\mathbf{r}_o^T(nK)\mathbf{r}_o^T(nK-1)\cdots\mathbf{r}_o^T(nK-K+1)\right]^T \\ &= \mathbf{H}\mathbf{s}(n) + \boldsymbol{\eta}(n), \end{aligned} \tag{6}$$

where $\mathbf{s}(n) = [s(nK)s(nK-1)\cdots s(nK-K-L+1)]^T$ is the transmitted signal vector, $\boldsymbol{\eta}(n) = [\boldsymbol{\eta}_o^T(nK)\boldsymbol{\eta}_o^T(nK-1)\cdots\boldsymbol{\eta}_o^T(nK-K+1)]^T$ is the noise vector, and $\mathbf{H}$ is the $MK \times (K+L)$ channel matrix which has a block Toeplitz structure such that

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(0) & \cdots & \mathbf{h}(L) & \cdots & \mathbf{0} \\ \vdots & \ddots & & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{h}(0) & \cdots & \mathbf{h}(L) \end{bmatrix} \tag{7}$$

with $\mathbf{0}$ being the M dimensional null vector.

2.2. Subspace channel estimation

Let the covariance matrix of the received signal vector $\mathbf{r}$ of (6) be denoted by $\boldsymbol{\Sigma}_r$, that is,

$$\boldsymbol{\Sigma}_r = \mathrm{E}\{\mathbf{r}(n)\mathbf{r}^H(n)\} = \mathbf{H}\boldsymbol{\Sigma}_s\mathbf{H}^H + \boldsymbol{\Sigma}_\eta, \tag{8}$$

where $\boldsymbol{\Sigma}_s = \mathrm{E}\{\mathbf{s}(n)\mathbf{s}^H(n)\}$ and $\boldsymbol{\Sigma}_\eta = \mathrm{E}\{\boldsymbol{\eta}(n)\boldsymbol{\eta}^H(n)\}$ are the covariance matrices of the transmitted signal and the noise, respectively. The following assumptions are usually made for the channel to be identifiable using SS method:

(1) the channel matrix $\mathbf{H}$ is of full column rank, that is, the subchannels share no common zeros;
(2) the signal covariance matrix $\mathbf{\Sigma}_s$ is of full rank;
(3) the noise process is uncorrelated with the signal;
(4) the channel order L is known or has been correctly estimated;
(5) the noise process is complex and white, that is, $\mathbf{\Sigma}_\eta = \mathrm{E}\{\boldsymbol{\eta}(n_1)\boldsymbol{\eta}^H(n_2)\} = \sigma_\eta^2\mathbf{I}\delta_{n_1n_2}$, where σ_η^2 is the noise variance and $\delta_{n_1n_2}$ is the Kronecker delta.

Applying eigendecomposition (ED) on $\mathbf{\Sigma}_r$, we have

$$\mathbf{\Sigma}_r = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^H + \mathbf{U}_\eta\mathbf{\Lambda}_\eta\mathbf{U}_\eta^H, \tag{9}$$

where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_{MK})$ with $\lambda_1 \geq \cdots \geq \lambda_{K+L} > \lambda_{K+L+1} = \cdots = \lambda_{MK} = \sigma_\eta^2$ being the eigenvalues of $\mathbf{\Sigma}_r$. Since $\mathbf{\Sigma}_r$ is Hermitian and positive definite, its eigenvalues are real and positive and its eigenvectors are orthonormal. The columns of $\mathbf{U}_s$ and $\mathbf{U}_\eta$ are the eigenvectors corresponding to the largest $K+L$ eigenvalues and to the remaining eigenvalues, respectively. Thus, the *noise subspace* spanned by the columns of $\mathbf{U}_\eta$ is orthogonal to the *signal subspace* spanned by the columns of the channel matrix $\mathbf{H}$.

In practice, we can only estimate the covariance matrix $\mathbf{\Sigma}_r$ by observing N received signal vectors of (6) such that $\widehat{\mathbf{\Sigma}}_r = (1/N)\sum_{n=1}^{N}\mathbf{r}(n)\mathbf{r}^H(n)$ so that the *estimated* noise subspace $\widehat{\mathbf{U}}_\eta$ can be obtained by replacing $\mathbf{\Sigma}_r$ by $\widehat{\mathbf{\Sigma}}_r$ in (9). Since $\widehat{\mathbf{\Sigma}}_r$ is still Hermitian and positive definite, its eigenvalues are still real and positive and its eigenvectors are still orthonormal. However, the $MK-(K+L)$ smallest eigenvalues are no longer equal. Also since the noise subspace is estimated, therefore it will not be truly orthogonal to the true signal subspace spanned by the columns of the channel matrix. Hence, we should search for the subspace that is closest to being orthogonal to the estimated noise subspace, that is,

$$\begin{aligned} &\min \|\widehat{\mathbf{U}}_\eta^H\mathbf{H}\|_F^2, \\ &\text{s.t. } \|\mathbf{h}\|_2 = 1, \end{aligned} \tag{10}$$

where

$$\mathbf{h} = [\mathbf{h}^H(0)\mathbf{h}^H(1)\cdots\mathbf{h}^H(L)]^H \tag{11}$$

is the channel coefficient vector to be estimated. For the use of (10) to estimate the channel, we need the following theorem [1].

Theorem 1. *Let* $\mathbf{H}^\perp = \mathrm{span}\{\mathbf{U}_\eta\}$ *be the orthogonal complement of the column space of* $\mathbf{H}$*. For any* $\mathbf{h}$ *and its corresponding estimate* $\widehat{\mathbf{h}}$ *satisfying the identifiable condition that the subchannels are coprime,* $\mathbf{H}^\perp = \widehat{\mathbf{H}}^\perp$ *if and only if* $\mathbf{h} = \alpha\widehat{\mathbf{h}}$*, where* $\widehat{\mathbf{H}}^\perp = \mathrm{span}\{\widehat{\mathbf{U}}_\eta\}$ *is the estimated orthogonal complement of the channel matrix* $\mathbf{H}$.

According to Theorem 1, $\mathbf{h}$ can be obtained up to a constant of proportionality. Due to the specific block Toeplitz structure of the channel matrix, we can carry out the channel estimation of (10) using a more convenient objective function such that

$$\widehat{\mathbf{h}} = \arg\min_{\|\mathbf{h}\|_2=1} \mathbf{h}^H\left(\sum_{j=1}^{MK-(K+L)} \widehat{\mathcal{U}}_j\widehat{\mathcal{U}}_j^H\right)\mathbf{h} \tag{12}$$

from which $\widehat{\mathbf{h}}$ can be obtained as the eigenvector corresponding to the smallest eigenvalue of $(\sum_{j=1}^{MK-(K+L)}\widehat{\mathcal{U}}_j\widehat{\mathcal{U}}_j^H)$. Here in (12), $\widehat{\mathcal{U}}_j$ is constructed from the jth column of $\widehat{\mathbf{U}}_\eta$ according to the following lemma [1].

Lemma 1. *Suppose that* $\mathbf{v} = [v_1 \ \cdots \ v_{MK}]^T$ *is in the orthogonal complement subspace of* $\mathbf{H}$*, then*

$$\begin{aligned} &\Big[\underbrace{v_1^*, \dots, v_M^*}_{\mathbf{v}_1^H} \ \vdots \ \cdots \ \vdots \ \underbrace{v_{(K-1)M+1}^*, \dots, v_{MK}^*}_{\mathbf{v}_K^H}\Big] \\ &\quad\times\begin{bmatrix} \mathbf{h}(0) & \cdots & \mathbf{h}(L) & & \\ & \ddots & & \ddots & \\ & & \mathbf{h}(0) & \cdots & \mathbf{h}(L) \end{bmatrix} = 0 \\ &\Longrightarrow \begin{bmatrix}\mathbf{h}^H(0) & \cdots & \mathbf{h}^H(L)\end{bmatrix}\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_K & & \\ & \ddots & & \ddots & \\ & & \mathbf{v}_1 & \cdots & \mathbf{v}_K \end{bmatrix} \\ &= \mathbf{h}^H\mathbf{V}_K = 0, \end{aligned} \tag{13}$$

where $\mathbf{v}_m$ *is the mth subvector of* $\mathbf{v}$ *and* $\mathbf{V}_K$ *is of dimension* $M(L+1)\times(K+L)$.

Lemma 1 can be easily proved by showing that the results of multiplying out the matrices in the two equations are the same. The above channel estimation method employing (12) is referred to as the subspace (SS) method.

3. CHANNEL ESTIMATION IN UNKNOWN NOISE

Assumptions (1) to (3) for the SS method in the previous section are at least approximately valid in practice. For assumption (4) (known channel order), there are various research work that have addressed the issue in different ways [16–19]. Assumption (5) on white noise, however, is often violated in many applications as mentioned in Section 1, resulting in severe deterioration in the performance of the method. An algorithm, designated *modified subspace* (MSS) method which is based on the above SS, has been proposed [10]. However, this MSS algorithm depends on the signal property, and restrictions on the length of the transmitted signal block have to be imposed for the method to be applicable. To address the problem of Assumption (5), in this section, we examine the situation when noise is temporally correlated and unknown, and develop various effective algorithms to estimate the channel. The algorithms are developed under different considerations of the receiver resources. Specifically, they are developed according to the number of receiver antennas available. To facilitate our algorithms so that the channel estimates can be obtained more directly, we will make use of the following results in matrix algebra.

Channel matrix transformation

It has been shown in detail [20] that a highly structured matrix $\mathbf{G}_\eta$, the columns of which *span the orthogonal complement of a special Sylvester channel matrix*, can be obtained using an efficient recursive algorithm. This Sylvester channel matrix, denoted by $\widetilde{\mathbf{H}}$ in turn, has a structure which is the row-permuted form of the block Toeplitz channel matrix $\mathbf{H}$ shown in (7), that is,

$$\widetilde{\mathbf{H}} = \begin{bmatrix} h_1(0) & \cdots & h_1(L) & & \\ & \ddots & & \ddots & \\ & & h_1(0) & \cdots & h_1(L) \\ h_2(0) & \cdots & h_2(L) & & \\ & \ddots & & \ddots & \\ & & h_2(0) & \cdots & h_2(L) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_M(0) & \cdots & h_M(L) & & \\ & \ddots & & \ddots & \\ & & h_M(0) & \cdots & h_M(L) \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{(1)} \\ \mathbf{H}_{(2)} \\ \vdots \\ \mathbf{H}_{(M)} \end{bmatrix} = \mathbf{\Pi}\mathbf{H}, \tag{14}$$

where $\mathbf{\Pi}$ is a proper row-permutation matrix, and

$$\mathbf{H}_{(m)} = \begin{bmatrix} h_m(0) & \cdots & h_m(L) & & \\ & \ddots & & \ddots & \\ & & h_m(0) & \cdots & h_m(L) \end{bmatrix} \tag{15}$$

with $\{h_m(\ell), m = 1, \dots, M\}$ being the elements of the $(\ell + 1)$th column vector of $\mathbf{H}_o$ in (5). $\mathbf{H}_{(m)}$ is of dimension $K \times (K + L)$ for $m = 1, 2, \dots, M$. Delete the last L rows and L columns of $\mathbf{H}_{(m)}$, and denote the truncated matrix by $\overline{\mathbf{H}}_{(m)}$ which has the dimension of $(K - L) \times K$, then we can form the matrix $\mathbf{G}^H_{\eta,m}$ such that [20]

$$\mathbf{G}^H_{\eta,m} = \left[\begin{array}{cccc|c} & \mathbf{G}^H_{\eta,m-1} & & & \mathbf{0} \\ \hline -\overline{\mathbf{H}}_{(m)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \overline{\mathbf{H}}_{(1)} \\ & -\overline{\mathbf{H}}_{(m)} & \mathbf{0} & \mathbf{0} & \overline{\mathbf{H}}_{(2)} \\ & & \ddots & & \vdots \\ & & & -\overline{\mathbf{H}}_{(m)} & \overline{\mathbf{H}}_{(m-1)} \end{array}\right]_{[((m-1)m/2)(K-L)]\times[mK]} \tag{16}$$

with $m = 2, \dots, M$ being the index of the subchannels. (For $m = 2$, we have $\mathbf{G}^H_{\eta,2} = [-\overline{\mathbf{H}}_{(2)} \;\; \overline{\mathbf{H}}_{(1)}]$.) Specifically, for the channel model with M subchannels ($m = M$), we denote $\mathbf{G}_{\eta,M}$ by $\mathbf{G}_\eta$ which has the following desirable properties useful to our channel estimation algorithms.

Properties of $\mathbf{G}_\eta$

(1) We note that $\mathbf{G}_\eta$ is of dimension $MK \times (M(M-1)(K-L)/2)$ and the orthogonal complement of the columns of $\widetilde{\mathbf{H}}$ is of dimension $MK - (K + L)$. Since the columns of $\mathbf{G}_\eta$ spans the orthogonal complement of the columns of $\widetilde{\mathbf{H}}$, then we have

$$\mathbf{G}^H_\eta \widetilde{\mathbf{H}} = \mathbf{G}^H_\eta (\mathbf{\Pi}\mathbf{H}) = (\mathbf{\Pi}^H \mathbf{G}_\eta)^H \mathbf{H} = \mathbf{0}. \tag{17}$$

Since the $(M(M-1)(K-L)/2)$ columns of $\mathbf{G}_\eta$ spans the orthogonal complement of $\widetilde{\mathbf{H}}$, we must have

$$M(M-1)(K-L)/2 \geq MK - (K + L) \quad \text{or} \quad K \geq \frac{(M+1)}{(M-1)} L. \tag{18}$$

(2) For any vector $\mathbf{b} = [\mathbf{b}_1^T \;\; \mathbf{b}_2^T \;\; \cdots \;\; \mathbf{b}_M^T]^T$, where $\mathbf{b}_m = [b_m(1) \;\; b_m(2) \;\; \cdots \;\; b_m(K)]^T$, $m = 1, 2, \dots, M$, the following relation holds:

$$\mathbf{G}^H_\eta \mathbf{b} = \mathbf{B}_M \widetilde{\mathbf{h}}, \tag{19}$$

where

$$\widetilde{\mathbf{h}} = \left[\widetilde{\mathbf{h}}_1^T \;\; \widetilde{\mathbf{h}}_2^T \;\; \cdots \;\; \widetilde{\mathbf{h}}_M^T\right]^T \tag{20}$$

with $\widetilde{\mathbf{h}}_m = [h_m(0) \;\; h_m(1) \;\; \cdots \;\; h_m(L)]^T$, $m = 1, 2, \dots, M$ being the vector comprising of the coefficients of the mth subchannel and $\mathbf{B}_M$ is constructed from $\mathbf{b}$ recursively according to

$$\mathbf{B}_m = \left[\begin{array}{cccc|c} & \mathbf{B}_{m-1} & & & \mathbf{0} \\ \hline \mathbf{B}_{(m)} & & & & -\mathbf{B}_{(1)} \\ & \mathbf{B}_{(m)} & & & -\mathbf{B}_{(2)} \\ & & \ddots & & \vdots \\ & & & \mathbf{B}_{(m)} & -\mathbf{B}_{(m-1)} \end{array}\right] \tag{21}$$

with $\mathbf{B}_2 = [\mathbf{B}_{(2)} \;\; -\mathbf{B}_{(1)}]$ and

$$\mathbf{B}_{(m)} = \begin{bmatrix} b_m(1) & b_m(2) & \cdots & b_m(L+1) \\ \vdots & \vdots & \vdots & \vdots \\ b_m(K-L) & b_m(K-L+1) & \cdots & b_m(K) \end{bmatrix} \quad \text{for } m = 1, 2, \dots, M. \tag{22}$$

We now present our channel estimation algorithms in the following.

3.1. Maximum a posteriori estimation

In this channel estimation algorithm which is based on the MAP criterion, we assume that there is only one receiver antenna available, and therefore the signal model is the same as that presented in the last section. Over N snapshots, we represent received data as $\mathbf{R}_N = [\mathbf{r}(1) \;\; \mathbf{r}(2) \;\; \cdots \;\; \mathbf{r}(N)]$, where $\mathbf{r}(n)$, $n = 1, 2, \dots, N$, are the N snapshots of the received data vectors defined in (6). If the noise is Gaussian distributed with zero mean and unknown covariance $\mathbf{\Sigma}_\eta$, then the conditional probability density function (PDF) of the received signal over N snapshots is

$$p(\mathbf{R}_N \mid \mathbf{h}, \mathbf{\Sigma}_\eta^{-1}, \mathbf{s}(n)) = \pi^{-MKN}\big(\det \mathbf{\Sigma}_\eta^{-1}\big)^N \times \exp\bigg(-\sum_{n=1}^{N}\big[\mathbf{r}(n)-\mathbf{H}\mathbf{s}(n)\big]^H \mathbf{\Sigma}_\eta^{-1}\big[\mathbf{r}(n)-\mathbf{H}\mathbf{s}(n)\big]\bigg). \quad (23)$$

If we define the estimate of the noise covariance matrix as

$$\hat{\mathbf{\Sigma}}_\eta = \frac{1}{N}\sum_{n=1}^{N}\big[\mathbf{r}(n)-\mathbf{H}\mathbf{s}(n)\big]\big[\mathbf{r}(n)-\mathbf{H}\mathbf{s}(n)\big]^H, \quad (24)$$

then (23) can be rewritten as

$$p(\mathbf{R}_N \mid \mathbf{h}, \mathbf{\Sigma}_\eta^{-1}) = \pi^{-MKN}\big(\det \mathbf{\Sigma}_\eta^{-1}\big)^N \operatorname{etr}\big\{-\mathbf{\Sigma}_\eta^{-1}(N\hat{\mathbf{\Sigma}}_\eta)\big\}, \quad (25)$$

where $\operatorname{etr}(\cdot)$ denotes $\exp[\operatorname{tr}\{\cdot\}]$. Applying Bayes' rule, that is,

$$p(\mathbf{h}, \mathbf{\Sigma}_\eta^{-1} \mid \mathbf{R}_N) = p(\mathbf{R}_N \mid \mathbf{h}, \mathbf{\Sigma}_\eta^{-1})\,p(\mathbf{h}, \mathbf{\Sigma}_\eta^{-1})/p(\mathbf{R}_N) \quad (26)$$

to (25) and noting that $p(\mathbf{R}_N)$ is independent of $\mathbf{h}$ and $\mathbf{\Sigma}_\eta$, we arrive at the a posteriori PDF containing only the channel coefficients by integrating $p(\mathbf{h}, \mathbf{\Sigma}_\eta^{-1} \mid \mathbf{R}_N)$ with respect to $\mathbf{\Sigma}_\eta^{-1}$ to obtain the marginal density function, that is,

$$p(\mathbf{h} \mid \mathbf{R}_N) \propto p(\mathbf{h})\int_{-\infty}^{\infty} p(\mathbf{R}_N \mid \mathbf{h}, \mathbf{\Sigma}_\eta^{-1})\,p(\mathbf{\Sigma}_\eta^{-1} \mid \mathbf{h})\,\mathrm{d}\mathbf{\Sigma}_\eta^{-1} \quad (27a)$$

$$\propto \int_{-\infty}^{\infty} p(\mathbf{R}_N \mid \mathbf{h}, \mathbf{\Sigma}_\eta^{-1})\,p(\mathbf{\Sigma}_\eta^{-1} \mid \mathbf{h})\,\mathrm{d}\mathbf{\Sigma}_\eta^{-1}, \quad (27b)$$

where, to arrive at (27b), we have assumed that all the channel coefficients are equally likely within the range of distribution. To evaluate the integral in (27b), we must obtain an expression for $p(\mathbf{\Sigma}_\eta^{-1} \mid \mathbf{h})$. Now, $\mathbf{\Sigma}_\eta$ is the covariance matrix of the noise and since we assume that we know nothing about the noise, we choose a *noninformative a priori* PDF [21]. Jeffreys [22] derived a general principle to obtain the noninformative a priori PDF such that: *the priori distribution of a set of parameters is taken to be proportional to the square root of the determinant of the information matrix*. Applying Jeffreys' principle, the noninformative a priori PDF of the noise covariance matrix can be written as [23]

$$p(\mathbf{\Sigma}_\eta^{-1} \mid \mathbf{h}) \propto \big\{\det(\mathbf{\Sigma}_\eta^{-1})\big\}^{-MK}. \quad (28)$$

Substituting (28) into (27b), the a posteriori PDF becomes

$$p(\mathbf{h} \mid \mathbf{R}_N) \propto \big\{\det(N\hat{\mathbf{\Sigma}}_\eta)\big\}^{-N}\int_{-\infty}^{\infty}\big\{\det(N\hat{\mathbf{\Sigma}}_\eta)\big\}^{N} \times \big\{\det(\mathbf{\Sigma}_\eta^{-1})\big\}^{N-MK}\operatorname{etr}\big\{-\mathbf{\Sigma}_\eta^{-1}N\hat{\mathbf{\Sigma}}_\eta\big\}\mathrm{d}\mathbf{\Sigma}_\eta^{-1}. \quad (29)$$

The integrand in (29) can be recognized as the complex Wishart distribution [24] with the role of $\mathbf{\Sigma}_\eta^{-1}$ and $N\hat{\mathbf{\Sigma}}_\eta$ reversed, and hence the integral is a constant. Therefore,

$$p(\mathbf{h} \mid \mathbf{R}_N) \propto \big\{\det(\hat{\mathbf{\Sigma}}_\eta)\big\}^{-N}. \quad (30)$$

To arrive at a MAP estimate of the channel using (30), we need to relate $\hat{\mathbf{\Sigma}}_\eta$ to the channel matrix $\mathbf{H}$. We can employ the ML estimate [25] of the transmitted signal $\hat{\mathbf{s}}(n) = (\mathbf{H}^H\mathbf{\Sigma}_\eta^{-1}\mathbf{H})^{-1}\mathbf{H}^H\mathbf{\Sigma}_\eta^{-1}\mathbf{r}(n)$ and after substituting this for $\mathbf{s}(n)$ in (24), we obtain

$$\hat{\mathbf{\Sigma}}_\eta = \frac{1}{N}\sum_{n=1}^{N}\big[\mathcal{P}_{\mathbf{H}}^{\perp}\mathbf{r}(n)\big]\big[\mathcal{P}_{\mathbf{H}}^{\perp}\mathbf{r}(n)\big]^H, \quad (31)$$

where $\mathcal{P}_{\mathbf{H}}^{\perp} = \mathbf{I} - \mathbf{H}(\mathbf{H}^H\mathbf{\Sigma}_\eta^{-1}\mathbf{H})^{-1}\mathbf{H}^H\mathbf{\Sigma}_\eta^{-1}$ is a weighted projection matrix with the idempotent property $(\mathcal{P}_{\mathbf{H}}^{\perp})^2 = \mathcal{P}_{\mathbf{H}}^{\perp}$. Putting this value of $\hat{\mathbf{\Sigma}}_\eta$ into (30) and taking logarithm, the MAP estimate of the channel coefficients can be obtained as

$$\hat{\mathbf{H}} = \max_{\mathbf{H}}\Big\{-\log\det\Big[\mathcal{P}_{\mathbf{H}}^{\perp}\hat{\mathbf{\Sigma}}_r(\mathcal{P}_{\mathbf{H}}^{\perp})^H\Big]\Big\}. \quad (32)$$

We note that $\mathcal{P}_{\mathbf{H}}^{\perp}$ is a (nonorthogonal) projector onto the $[MK-(K+L)]$-dimensional *noise subspace*. Since $\hat{\mathbf{\Sigma}}_r$ is of rank MK, therefore the matrix $\mathcal{P}_{\mathbf{H}}^{\perp}\hat{\mathbf{\Sigma}}_r(\mathcal{P}_{\mathbf{H}}^{\perp})^H$ is only of rank $[MK-(K+L)]$, that is, its determinant equals zero. Therefore, direct maximization of (32) (which is equivalent to minimization of the determinant) becomes meaningless, and we have to look for modification of the criterion. Let us examine the geometric interpretation of the MAP criterion in (32): it is well known [26] that the determinant of a square matrix is equal to the product of its eigenvalues. It is also well known [26] that the determinant of the covariance matrix $\hat{\mathbf{\Sigma}}_r$ represents the square of the volume of the parallelepiped whose edges are formed by the MK data vectors. Now, consider the projected data represented by $(1/\sqrt{N})\mathcal{P}_{\mathbf{H}}^{\perp}\mathbf{R}_N = [\boldsymbol{\eta}_1 \cdots \boldsymbol{\eta}_{MK}]^H$, where $\boldsymbol{\eta}_m^H$ is an N-dimensional projected data row vector. Since $\mathcal{P}_{\mathbf{H}}^{\perp}$ projects the MK-dimensional vector $\mathbf{r}(n)$ onto an $[MK-(K+L)]$-dimensional hyperplane, the vectors $\boldsymbol{\eta}_1 \cdots \boldsymbol{\eta}_{MK}$ are linearly dependent and span the hyperplane. Thus, for the matrix $\mathcal{P}_{\mathbf{H}}^{\perp}\hat{\mathbf{\Sigma}}_r(\mathcal{P}_{\mathbf{H}}^{\perp})^H = [\boldsymbol{\eta}_1 \cdots \boldsymbol{\eta}_{MK}]^H[\boldsymbol{\eta}_1 \cdots \boldsymbol{\eta}_{MK}]$, each of its $[MK-(K+L)]$-dimensional principal minor (formed by deleting $(K+L)$ of the corresponding rows and columns) is equal to the square of the volume of the $[MK-(K+L)]$-dimensional parallelepiped whose edges are the $[MK-(K+L)]$ vectors $\{\boldsymbol{\eta}_m\}$ involved in the principal minor. Now, since the determinant of the rank deficient matrix $\mathcal{P}_{\mathbf{H}}^{\perp}\hat{\mathbf{\Sigma}}_r(\mathcal{P}_{\mathbf{H}}^{\perp})^H$ represents the square of the volume of a collapsed parallelepiped in the $[MK-(K+L)]$-dimensional hyperplane and is always equal to zero, instead of minimizing this vanishing volume, it is reasonable then to minimize the total volume of all the $[MK-(K+L)]$-dimensional parallelepipeds formed by the $[MK-(K+L)]$-dimensional principal minors of the rank deficient matrix, that is, $\min\sum_{MK-(K+L)}(\prod\lambda_i)$ which is the sum of the products of the eigenvalues taken $MK-(K+L)$ at a time. Since there are only $MK-(K+L)$ nonzero eigenvalues, then there is only one nonzero product of eigenvalues taken $MK-(K+L)$ at a time. Thus, instead of maximizing (32) which will lead us to nowhere, we argue from a geometric viewpoint that it is more fruitful to maximize the

Table 1: Computation complexity of MAP algorithm.

	No. of multiplications
compute $\hat{\mathbf{\Sigma}}_r$	NM^2K^2
compute $\mathbf{G}_\eta^H\mathbf{\Pi}\hat{\mathbf{\Sigma}}_r\mathbf{\Pi}^H\mathbf{G}_\eta$	$\frac{M(M-1)}{2}(K-L)M^2K^2+MK\left[\frac{M(M-1)}{2}(K-L)\right]^2$
compute $(\mathbf{G}_\eta^H\mathbf{\Pi}\hat{\mathbf{\Sigma}}_r\mathbf{\Pi}^H\mathbf{G}_\eta)^\dagger$	$\mathcal{O}\left\{\left[\frac{M(M-1)}{2}(K-L)\right]^3\right\}$
compute $\sum_{i=1}^{MK}\mathcal{V}_i^H(\mathbf{G}_\eta^H\mathbf{\Pi}\hat{\mathbf{\Sigma}}_r\mathbf{\Pi}^H\mathbf{G}_\eta)^\dagger\mathcal{W}_i$	$MK\left[M(L+1)\frac{M^2(M-1)^2}{4}(K-L)^2+M^2(L+1)^2\frac{M(M-1)}{2}(K-L)\right]$
compute SVD ($\sum_{i=1}^{MK}\mathcal{V}_i^H(\mathbf{G}_\eta^H\mathbf{\Pi}\hat{\mathbf{\Sigma}}_r\mathbf{\Pi}^H\mathbf{G}_\eta)^\dagger\mathcal{W}_i$)	$\mathcal{O}\{M^3(L+1)^3\}$
Total	Sum of the rows

following criterion:

$$\hat{\mathbf{H}}=\max_{\mathbf{H}}\left\{-\log\left(\prod_{i=1}^{MK-(K+L)}\hat{\lambda}_i\right)\right\},\qquad \hat{\lambda}_1\geq\cdots\geq\hat{\lambda}_{MK-(K+L)}, \tag{33}$$

with $\hat{\lambda}_i, i=1,2,\ldots,MK-(K+L)$ being the nonzero eigenvalues of $\mathcal{P}_{\mathbf{H}}^{\perp}\hat{\mathbf{\Sigma}}_r(\mathcal{P}_{\mathbf{H}}^{\perp})^H$. (A different approach [23] using an orthonormal basis of $\mathcal{P}_{\mathbf{H}}^{\perp}$ can be taken to develop (33) from (32)). Following the same mathematical manipulation as in [23], (33) can be written as

$$\hat{\mathbf{H}}\approx\max_{\mathbf{H}}\left\{-\operatorname{tr}\left[\mathcal{P}_{\mathbf{H}}^{\perp}\left(\log\hat{\mathbf{\Sigma}}_r\right)\right]\right\}, \tag{34}$$

where the logarithm of a positive definite matrix $\mathbf{A}$ is defined such that if $\mathbf{A}$ can be eigendecomposed as $\mathbf{A}=\mathbf{V}_a\mathbf{\Lambda}_a\mathbf{V}_a^H$, then $\log\mathbf{A}=\mathbf{V}_a(\log\mathbf{\Lambda}_a)\mathbf{V}_a^H$ and the logarithm of a diagonal matrix is the matrix with the diagonal entries being the logarithm of the original entries [23].

Equation (34) is our MAP estimate of the channel coefficients under unknown correlated noise. However, it is not very convenient to use since $\mathcal{P}_{\mathbf{H}}^{\perp}$ is an implicit function of $\mathbf{h}$. We overcome this difficulty by applying the result of channel matrix transformation [20] as summarized in the beginning of this section. By permuting the rows of the channel matrix $\mathbf{H}$ using $\mathbf{\Pi}$, we obtain the Sylvester form $\tilde{\mathbf{H}}$ of the channel matrix from which we recursively generate the matrix $\mathbf{G}_\eta$. Now, from (17), we have

$$\mathbf{I}-\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H=\mathbf{\Pi}^H\mathbf{G}_\eta(\mathbf{G}_\eta^H\mathbf{\Pi}\mathbf{\Pi}^H\mathbf{G}_\eta)^\dagger\mathbf{G}_\eta^H\mathbf{\Pi}, \tag{35}$$

where, because of the relation of (18), the pseudoinverse, denoted by $\dagger$, of the matrix $(\mathbf{G}_\eta^H\mathbf{\Pi}\mathbf{\Pi}^H\mathbf{G}_\eta)$ has to be used. Combining the projection matrix $\mathcal{P}_{\mathbf{H}}^{\perp}$ and (35), we obtain

$$\mathcal{P}_{\mathbf{H}}^{\perp}=\mathbf{\Sigma}_\eta\mathbf{\Pi}^H\mathbf{G}_\eta(\mathbf{G}_\eta^H\mathbf{\Pi}\mathbf{\Sigma}_\eta\mathbf{\Pi}^H\mathbf{G}_\eta)^\dagger\mathbf{G}_\eta^H\mathbf{\Pi}. \tag{36}$$

Thus, the MAP criterion in (34) can now be written as

$$\begin{aligned}\hat{\mathbf{H}}&\approx\max_{\mathbf{H}}\left\{-\operatorname{tr}\left[(\mathbf{G}_\eta^H\mathbf{\Pi}\mathbf{\Sigma}_\eta)^H(\mathbf{G}_\eta^H\mathbf{\Pi}\mathbf{\Sigma}_\eta\mathbf{\Pi}^H\mathbf{G}_\eta)^\dagger\mathbf{G}_\eta^H\mathbf{\Pi}(\log\hat{\mathbf{\Sigma}}_r)\right]\right\}\\&\approx\max_{\mathbf{H}}\left\{-\operatorname{tr}\left[(\mathbf{G}_\eta^H\mathbf{\Pi}\hat{\mathbf{\Sigma}}_r)^H(\mathbf{G}_\eta^H\mathbf{\Pi}\hat{\mathbf{\Sigma}}_r\mathbf{\Pi}^H\mathbf{G}_\eta)^\dagger\mathbf{G}_\eta^H\mathbf{\Pi}(\log\hat{\mathbf{\Sigma}}_r)\right]\right\},\end{aligned} \tag{37}$$

where in the second step, we have used the facts that $(\mathbf{\Pi}^H\mathbf{G}_\eta)^H\mathbf{H}=\mathbf{0}$ and thus $\mathbf{\Sigma}_r$ can be substituted for $\mathbf{\Sigma}_\eta$, and that as N increases, $\hat{\mathbf{\Sigma}}_r\rightarrow\mathbf{\Sigma}_r$. Now, let $\mathbf{v}_i$ denote the ith column of $\mathbf{\Pi}\hat{\mathbf{\Sigma}}_r$ and $\mathbf{w}_i$ denote the ith column of $\mathbf{\Pi}(\log\hat{\mathbf{\Sigma}}_r)$, then using Property (2) of $\mathbf{G}_\eta$ in (19) such that $\mathbf{G}_\eta^H\mathbf{v}_i=\mathcal{V}_i\tilde{\mathbf{h}}$ and $\mathbf{G}_\eta^H\mathbf{w}_i=\mathcal{W}_i\tilde{\mathbf{h}}$ with $\mathcal{V}_i$ and $\mathcal{W}_i$ constructed from $\mathbf{v}_i$ and $\mathbf{w}_i$ respectively as indicated in (21), then the channel coefficients can be estimated as

$$\hat{\mathbf{h}}=\arg\min_{\|\tilde{\mathbf{h}}\|_2=1}\tilde{\mathbf{h}}^H\left(\sum_{i=1}^{MK}\mathcal{V}_i^H(\mathbf{G}_\eta^H\mathbf{\Pi}\hat{\mathbf{\Sigma}}_r\mathbf{\Pi}^H\mathbf{G}_\eta)^\dagger\mathcal{W}_i\right)\tilde{\mathbf{h}}. \tag{38}$$

We can see that the estimated channel vector $\tilde{\mathbf{h}}$ from (38) is a permuted version of the channel vector defined in (11). $(\mathbf{G}_\eta^H\mathbf{\Pi}\hat{\mathbf{\Sigma}}_r\mathbf{\Pi}^H\mathbf{G}_\eta)^\dagger$ in (38) is a weighting matrix which has the unknown channel coefficients. The IQML [27] algorithm can now be applied to solve this optimization problem. The computation complexity for each iteration of the MAP algorithm using IQML is summarized in Table 1. It can be observed that the computation is dominated by the calculation of $\sum_{i=1}^{MK}\mathcal{V}_i^H(\mathbf{G}_\eta^H\mathbf{\Pi}\hat{\mathbf{\Sigma}}_r\mathbf{\Pi}^H\mathbf{G}_\eta)^\dagger\mathcal{W}_i$. When the number of iteration is small (which is the case according to the simulation results), the computation complexity is of the same order as that summarized in Table 1.

3.2. *Channel estimation using canonical correlation decomposition*

For the MAP algorithm, only one set of received data from the transmitted signals is needed. However, if two versions of the same set of transmitted signals can be received at different points in space by applying two sufficiently separated receiver antennae (as may be in the case of a base station), channel estimation algorithms with better performance may be developed. Here, we develop two algorithms based on the CCD of two sets of received data.

Consider a receiver activated by the same transmitted signal having two antennae the outputs of which are upsampled by factors M_1 and M_2, respectively. For mathematical convenience, we assume the order of the two channels linking the transmitter to the two receiver antennae to be the same. Then, similar to (6), the two outputs from the antennae over

K symbols can be represented as

$$\mathbf{r}_1(n) = \mathbf{H}_1\mathbf{s}(n) + \boldsymbol{\eta}_1(n); \qquad \mathbf{r}_2(n) = \mathbf{H}_2\mathbf{s}(n) + \boldsymbol{\eta}_2(n). \quad (39)$$

Let the two antennae be sufficiently separated so that the noise vectors are uncorrelated, that is, $\mathrm{E}\{\boldsymbol{\eta}_1(n)\boldsymbol{\eta}_2^H(n)\} = \mathbf{0}$ and $\mathrm{E}\{\boldsymbol{\eta}_2(n)\boldsymbol{\eta}_1^H(n)\} = \mathbf{0}$, and we allow the covariance matrix of $\boldsymbol{\eta}_1(n)$ and $\boldsymbol{\eta}_2(n)$ to be arbitrary and unknown. We now stack the two received vectors to form vector $\mathbf{r}$, the covariance matrix of which is given by

$$\boldsymbol{\Sigma} = \mathrm{E}\left\{\begin{bmatrix}\mathbf{r}_1\\ \mathbf{r}_2\end{bmatrix}\begin{bmatrix}\mathbf{r}_1^H & \mathbf{r}_2^H\end{bmatrix}\right\} = \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}, \quad (40)$$

where the submatrices $\boldsymbol{\Sigma}_{ij}$ are given by $\boldsymbol{\Sigma}_{ii} = \mathbf{H}_i\mathbf{R}_s\mathbf{H}_i^H + \boldsymbol{\Sigma}_{i\eta}$, $i = 1,2$, and $\boldsymbol{\Sigma}_{12} = \mathbf{H}_1\mathbf{R}_s\mathbf{H}_2^H = \boldsymbol{\Sigma}_{21}^H$. (40) can be employed in different ways to estimate the channel in the presence of correlated noise. The modified subspace method (MSS) [10] mentioned in Section 1, for example, uses received signal vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ in consecutive time slots and employed their cross-correlation matrix to estimate the channel taking advantage of the zero noise correlation term. In doing so, some arbitrarily restrictive assumptions of the signals have to be made. This method generally achieves higher accuracy in the channel estimate than the simple SS method.

3.2.1. CCD-based subspace algorithm

We now introduce the matrix product $(\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1/2})$ on which a singular value decomposition (SVD) [28] can be performed such that

$$\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1/2} = \mathbf{U}_1\boldsymbol{\Gamma}_0\mathbf{U}_2^H, \quad (41)$$

where $\mathbf{U}_1$ and $\mathbf{U}_2$ are of dimension $M_1K \times M_1K$ and $M_2K \times M_2K$, respectively, and $\boldsymbol{\Gamma}_0$ is of dimension $M_1K \times M_2K$, given by

$$\boldsymbol{\Gamma}_0 = \begin{bmatrix}\boldsymbol{\Gamma} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} \quad (42)$$

with $\boldsymbol{\Gamma} = \mathrm{diag}(\gamma_1, \ldots, \gamma_{K+L})$, and γ_k, $k = 1,\ldots,K+L$ are real and positive such that $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_{K+L} > 0$. Equation (41) is referred to as the CCD of the matrix $\boldsymbol{\Sigma}$, and $\{\gamma_1, \ldots, \gamma_{K+L}\}$ are called the *canonical correlation coefficients* [29, 30]. Now, for $i = 1,2$, define the *canonical vector matrices* and the *reciprocal canonical vector matrices* corresponding to the data $\mathbf{r}_i$ as

$$\mathbf{Z}_i \triangleq \boldsymbol{\Sigma}_{ii}^{-1/2}\mathbf{U}_i, \qquad \mathbf{Y}_i \triangleq \boldsymbol{\Sigma}_{ii}^{1/2}\mathbf{U}_i. \quad (43)$$

CCD attempts to characterize the correlation structure between two sets of variables $\mathbf{r}_1$ and $\mathbf{r}_2$ by replacing them with two new sets using the transformations $\mathbf{Z}_i$ and $\mathbf{Y}_i$. It has been shown [30] that such transformations render the new sets to attain maximum correlation between corresponding elements while maintaining zero correlations between non-corresponding elements. While such properties separate the signal and noise subspaces, they fully exploit the correlation between the two versions of the transmitted signal. Now, partition $\mathbf{Z}_i$ and $\mathbf{Y}_i$, $i = 1,2$, such that

$$\begin{aligned}\mathbf{Z}_i &= [\mathbf{Z}_{is} \mid \mathbf{Z}_{i\eta}] = [\boldsymbol{\Sigma}_{ii}^{-1/2}\mathbf{U}_{is} \mid \boldsymbol{\Sigma}_{ii}^{-1/2}\mathbf{U}_{i\eta}],\\ \mathbf{Y}_i &= [\mathbf{Y}_{is} \mid \mathbf{Y}_{i\eta}] = [\boldsymbol{\Sigma}_{ii}^{1/2}\mathbf{U}_{is} \mid \boldsymbol{\Sigma}_{ii}^{1/2}\mathbf{U}_{i\eta}],\end{aligned} \quad (44)$$

where $\mathbf{Z}_{is}$ and $\mathbf{Z}_{i\eta}$, $\mathbf{Y}_{is}$ and $\mathbf{Y}_{i\eta}$, $\mathbf{U}_{is}$ and $\mathbf{U}_{i\eta}$ are the first $K+L$ columns and the last $M_iK - (K+L)$ columns of $\mathbf{Z}_i$, $\mathbf{Y}_i$, and $\mathbf{U}_i$, respectively. Then, the following relations hold [29, 30]:

$$\mathrm{span}\{\mathbf{Y}_{is}\} = \mathrm{span}\{\mathbf{H}_i\}, \quad \mathrm{span}\{\mathbf{Z}_{i\eta}\} = \overline{\mathrm{span}}\{\mathbf{H}_i\}, \quad i = 1,2, \quad (45)$$

where $\overline{\mathrm{span}}\{\mathbf{H}_i\}$ denotes the orthogonal complement of $\mathrm{span}\{\mathbf{H}_i\}$. We can see that, in the presence of correlated noise, by applying CCD, the signal and noise subspaces can be partitioned according to the column spaces of $\mathbf{Y}_{is}$ and $\mathbf{Z}_{i\eta}$, respectively.

From (45), we can conclude that

$$\mathbf{Z}_{i\eta}^H\mathbf{H}_i = \mathbf{0}. \quad (46)$$

As usual in practice, we can only estimate the covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{r}$ in (40) such that

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N}\sum_{n=1}^{N}\begin{bmatrix}\mathbf{r}_1(n)\\ \mathbf{r}_2(n)\end{bmatrix}\begin{bmatrix}\mathbf{r}_1^H(n) & \mathbf{r}_2^H(n)\end{bmatrix} = \begin{bmatrix}\hat{\boldsymbol{\Sigma}}_{11} & \hat{\boldsymbol{\Sigma}}_{12}\\ \hat{\boldsymbol{\Sigma}}_{21} & \hat{\boldsymbol{\Sigma}}_{22}\end{bmatrix}, \quad (47)$$

and all the parameter matrices obtained from this are estimates, that is, we apply CCD on $\hat{\boldsymbol{\Sigma}}$ to obtain $\hat{\mathbf{U}}_i$, $\hat{\mathbf{Z}}_i$, and $\hat{\mathbf{Y}}_i$, accordingly. Using the estimate $\hat{\mathbf{Z}}_{i\eta}$, we can employ a technique similar to the SS method in white noise by applying the concept in (46) to obtain the channel coefficient estimates up to a constant of proportionality such that

$$\hat{\mathbf{h}}_i = \arg\min_{\|\mathbf{h}_i\|_2=1}\mathbf{h}_i^H\left(\sum_{j=1}^{M_iK-(K+L)}\hat{\mathcal{Z}}_j\hat{\mathcal{Z}}_j^H\right)\mathbf{h}_i, \quad (48)$$

where $\hat{\mathcal{Z}}_j$ is constructed from the jth column of $\hat{\mathbf{Z}}_{i\eta}$ in a similar way as (13) in Lemma 1. Again, the channel estimate $\hat{\mathbf{h}}_i$ can be obtained from (48) as the eigenvector corresponding to the smallest eigenvalue of $(\sum_{j=1}^{M_iK-(K+L)}\hat{\mathcal{Z}}_j\hat{\mathcal{Z}}_j^H)$. This method is referred to as the "CCD-based subspace" method.

The main computation complexity involved in the CCD-SS method is summarized in Table 2.

3.2.2. CCD-based maximum likelihood algorithm (CCD-ML)

Maximum likelihood (ML) is one of the most powerful methods in parameter estimation. Because of its superior performance, it is also widely used as a criterion in channel estimation when the channel noise can be assumed Gaussian distributed and white. This assumption makes the concentration of the log-likelihood function from the nuisance parameters possible and results in the reduction of the dimension of the parameter space and thus the computational burden. However, when the noise covariance matrix is unknown as is the focus of this paper, the ML estimation cannot

Table 2: Computation complexity of CCD-SS algorithm.

	No. of multiplications
computation of $\hat{\mathbf{\Sigma}}$	$N(M_1+M_2)^2K^2$
computation of $\mathbf{\Sigma}_{11}^{-1/2}$	$M_1^3K^3$
computation of $\mathbf{\Sigma}_{22}^{-1/2}$	$M_2^3K^3$
computation of $\mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1/2}$	$M_1^2M_2K^3+M_1M_2^2K^3$
computation of SVD$(\mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1/2})$	$\mathcal{O}\{\min(M_1^3K^3, M_2^3K^3)\}$
computation of $\mathbf{Z}_{i\eta}$	$M_i^2K^2[M_iK-(K+L)]$
computation of $\sum_{j=1}^{M_iK-(K+L)}\hat{\mathcal{Z}}_j\hat{\mathcal{Z}}_j^H$	$[M_iK-(K+L)]M_i^2(L+1)^2(K+L)$
computation of ED$(\sum_{j=1}^{M_iK-(K+L)}\hat{\mathcal{Z}}_j\hat{\mathcal{Z}}_j^H)$	$\mathcal{O}\{M_i^3(L+1)^3\}$
Total	Sum of the rows

be applied directly. However, we can approach the problem in a different way by examining the asymptotic projection error between the signal subspace and the noise subspace and from the statistical properties of this, we can establish a log-likelihood function from which an ML estimation of the channel can be obtained.

Let us first construct the two eigenprojectors $\mathcal{P}_{is}$ and $\mathcal{P}_{i\eta}$ associated, respectively, with the subspace spanned by $\{\mathbf{z}_{ik}\}$, $k=1,2,\ldots,K+L$, and $\{\mathbf{z}_{ij}\}$, $j=K+L+1,\ldots,M_iK$, which correspondingly are the first $K+L$ and the last $(M_i-1)K-L$ columns of $\mathbf{Z}_i$,

$$\mathcal{P}_{is}=\sum_{k=1}^{K+L}\mathbf{z}_{ik}\mathbf{z}_{ik}^H\mathbf{\Sigma}_{ii}=\mathbf{Z}_{is}\mathbf{Z}_{is}^H\mathbf{\Sigma}_{ii}=\mathbf{Z}_{is}\mathbf{Y}_{is}^H \tag{49a}$$

$$\mathcal{P}_{i\eta}=\sum_{j=K+L+1}^{MK}\mathbf{z}_{ij}\mathbf{z}_{ij}^H\mathbf{\Sigma}_{ii}=\mathbf{Z}_{i\eta}\mathbf{Z}_{i\eta}^H\mathbf{\Sigma}_{ii}=\mathbf{Z}_{i\eta}\mathbf{Y}_{i\eta}^H, \tag{49b}$$

where the last steps of (49a) and (49b) are arrived at directly from the definitions of $\mathbf{Z}_i$ and $\mathbf{Y}_i$ in (43). It can be easily verified that $\mathcal{P}_{is}$ and $\mathcal{P}_{i\eta}$ are both idempotent and are, therefore, valid projectors. Due to the span of the columns of $\mathbf{Y}_{is}$ and $\mathbf{Z}_{i\eta}$, we can see that $\mathcal{P}_{is}^H$ and $\mathcal{P}_{i\eta}$ project onto the signal and the noise subspaces, respectively. Let us now consider the columns of the matrix product $\hat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta}$, where $\hat{\mathbf{Y}}_{is}$ is obtained using the estimate of the covariance matrix $\hat{\mathbf{\Sigma}}$ in (47). Denoting the vector obtained by stacking the column of a matrix by vec$(\cdot)$, we have

$$\operatorname{vec}(\hat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta})\simeq\operatorname{vec}(\mathbf{Y}_{is}^H\hat{\mathbf{Z}}_{is}\hat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta}) \tag{50a}$$
$$=(\mathbf{I}\otimes\mathbf{Y}_{is}^H)\operatorname{vec}(\hat{\mathbf{Z}}_{is}\hat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta}) \tag{50b}$$
$$=(\mathbf{I}\otimes\mathbf{Y}_{is}^H)\operatorname{vec}(\widehat{\mathcal{P}}_{is}\mathbf{Z}_{i\eta}), \tag{50c}$$

where (50a) holds asymptotically as $\hat{\mathbf{Y}}_{is}\to\mathbf{Y}_{is}$, also from vec$(\mathbf{ABC})=(\mathbf{C}^T\otimes\mathbf{A})vec(\mathbf{B})$ with $\mathbf{C}$ being the identity matrix $\mathbf{I}$ of dimension $[M_iK-(K+L)]\times[M_iK-(K+L)]$, (50b) also holds, and finally, (50c) comes directly from the estimated form of the signal space projector $\mathcal{P}_{is}$ in (49a). We now invoke the following important result [29].

Theorem 2. *If $\mathbf{X}_{i\eta}\subseteq\overline{\operatorname{span}}(\mathbf{H}_i)$, then the random vectors vec$(\widehat{\mathcal{P}}_{is}\mathbf{X}_{i\eta})$, $i=1,2$, are asymptotically complex Gaussian with zero mean and covariance matrix*

$$E[\operatorname{vec}(\widehat{\mathcal{P}}_{is}\mathbf{X}_{i\eta})\operatorname{vec}^H(\widehat{\mathcal{P}}_{is}\mathbf{X}_{i\eta})]=\frac{1}{N}[\mathbf{X}_{i\eta}^H\mathbf{\Sigma}_{ii}\mathbf{X}_{i\eta}]^T\otimes[\mathbf{Z}_{is}\mathbf{\Gamma}^{-1}\mathbf{Z}_{\bar{i}s}^H\mathbf{\Sigma}_{\bar{i}\bar{i}}\mathbf{Z}_{\bar{i}s}\mathbf{\Gamma}^{-1}\mathbf{Z}_{is}^H], \tag{51}$$

where the index $\bar{i}$ denotes the complement of i such that $\bar{i}=2$ if $i=1$, and $\bar{i}=1$ if $i=2$.

Applying Theorem 2 to (50), we can conclude that vec$\{\hat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta}\}$ is also asymptotically Gaussian with zero mean and its covariance matrix (after some algebraic simplifications) given by

$$\mathrm{E}[\operatorname{vec}(\hat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta})\operatorname{vec}^H(\hat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta})]=\frac{1}{N}(\mathbf{Z}_{i\eta}^H\mathbf{\Sigma}_{ii}\mathbf{Z}_{i\eta})^T\otimes(\mathbf{\Gamma}^{-1}\mathbf{Z}_{\bar{i}s}^H\mathbf{\Sigma}_{\bar{i}\bar{i}}\mathbf{Z}_{\bar{i}s}\mathbf{\Gamma}^{-1}). \tag{52}$$

With this Gaussian distribution, the log-likelihood function of vec$(\hat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta})$ can be written as

$$\begin{aligned}\mathcal{L}_{\text{ccd}}\propto&-\log\det\{(\mathbf{Z}_{i\eta}^H\mathbf{\Sigma}_{ii}\mathbf{Z}_{i\eta})^T\otimes(\mathbf{\Gamma}^{-1}\mathbf{Z}_{\bar{i}s}^H\mathbf{\Sigma}_{\bar{i}\bar{i}}\mathbf{Z}_{\bar{i}s}\mathbf{\Gamma}^{-1})\}\\&-N\operatorname{tr}\{[(\mathbf{Z}_{i\eta}^H\mathbf{\Sigma}_{ii}\mathbf{Z}_{i\eta})^T\otimes(\mathbf{\Gamma}^{-1}\mathbf{Z}_{\bar{i}s}^H\mathbf{\Sigma}_{\bar{i}\bar{i}}\mathbf{Z}_{\bar{i}s}\mathbf{\Gamma}^{-1})]^{-1}\\&\times\operatorname{vec}(\hat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta})\operatorname{vec}^H(\hat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta})\}.\end{aligned} \tag{53}$$

For large sample size N, the first term of this likelihood function can be omitted and, carrying further simplifications, we have

$$\begin{aligned}\mathcal{L}_{\text{ccd}}\approx&-N\operatorname{tr}\{\operatorname{vec}^H(\hat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta})\\&\times[(\mathbf{Z}_{i\eta}^T\mathbf{\Sigma}_{ii}^T\mathbf{Z}_{i\eta}^*)^{-1}\otimes(\mathbf{\Gamma}^{-1}\mathbf{Z}_{\bar{i}s}^H\mathbf{\Sigma}_{\bar{i}\bar{i}}\mathbf{Z}_{\bar{i}s}\mathbf{\Gamma}^{-1})^{-1}]\\&\times\operatorname{vec}(\hat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta})\}\end{aligned} \tag{54a}$$

$$\begin{aligned}\propto&-\operatorname{tr}\{\operatorname{vec}^H\mathbf{I}\cdot\operatorname{vec}([\hat{\mathbf{Y}}_{is}(\mathbf{\Gamma}^{-1}\mathbf{Z}_{\bar{i}s}^H\mathbf{\Sigma}_{\bar{i}\bar{i}}\mathbf{Z}_{\bar{i}s}\mathbf{\Gamma}^{-1})^{-1}\hat{\mathbf{Y}}_{is}^H]\\&\times[\mathbf{Z}_{i\eta}(\mathbf{Z}_{i\eta}^H\mathbf{\Sigma}_{ii}\mathbf{Z}_{i\eta})^{-1}\mathbf{Z}_{i\eta}^H])\}\end{aligned} \tag{54b}$$

$$=-\operatorname{tr}\{[\mathbf{Z}_{i\eta}(\mathbf{Z}_{i\eta}^H\mathbf{\Sigma}_{ii}\mathbf{Z}_{i\eta})^{-1}\mathbf{Z}_{i\eta}^H][\hat{\mathbf{Y}}_{is}\mathbf{\Gamma}^2\hat{\mathbf{Y}}_{is}^H]\}, \tag{54c}$$

where we have used the identities tr$(\mathbf{AB})$ = tr$(\mathbf{BA})$ and $(\mathbf{A}\otimes\mathbf{B})^{-1}=\mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$ to arrive at (54a), vec$(\mathbf{ABC})=(\mathbf{C}^T\otimes\mathbf{A})vec(\mathbf{B})$ and $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D})=(\mathbf{AC})\otimes(\mathbf{BD})$ to

TABLE 3: Computation complexity of CCD-ML algorithm.

	No. of multiplications
compute $\widehat{\boldsymbol{\Sigma}}$	$N(M_1+M_2)^2K^2$
compute $\boldsymbol{\Sigma}_{11}^{-1/2}$	$M_1^3K^3$
compute $\boldsymbol{\Sigma}_{22}^{-1/2}$	$M_2^3K^3$
compute $\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1/2}$	$M_1^2M_2K^3+M_1M_2^2K^3$
compute SVD$(\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1/2})$	$\mathcal{O}\{\min(M_1^3K^3, M_2^3K^3)\}$
compute $\mathbf{Y}_{is}$	$M_i^2K^2(K+L)$
compute $\mathbf{G}_{i\eta}^H\boldsymbol{\Pi}\widehat{\boldsymbol{\Sigma}}_{ii}\boldsymbol{\Pi}^H\mathbf{G}_{i\eta}$	$M_i^2K^2\times\dfrac{M_i(M_i-1)}{2}(K-L)+M_iK\times\left[\dfrac{M_i(M_i-1)}{2}(K-L)\right]^2$
compute $(\mathbf{G}_{i\eta}^H\boldsymbol{\Pi}\widehat{\boldsymbol{\Sigma}}_{ii}\boldsymbol{\Pi}^H\mathbf{G}_{i\eta})^\dagger$	$\mathcal{O}\left\{\left[\dfrac{M_i(M_i-1)}{2}(K-L)\right]^3\right\}$
compute $\sum_{j=1}^{K+L}\mathcal{F}_{ij}^H(\mathbf{G}_{i\eta}^H\boldsymbol{\Pi}\widehat{\boldsymbol{\Sigma}}_{ii}\boldsymbol{\Pi}^H\mathbf{G}_{i\eta})^\dagger\mathcal{F}_{ij}$	$(K+L)\left\{M_i(L+1)\left[\dfrac{M_i(M_i-1)}{2}(K-L)\right]^2+M_i^2(L+1)^2\dfrac{M_i(M_i-1)}{2}(K-L)\right\}$
compute ED$\{\sum_{j=1}^{K+L}\mathcal{F}_{ij}^H(\mathbf{G}_{i\eta}^H\boldsymbol{\Pi}\widehat{\boldsymbol{\Sigma}}_{ii}\boldsymbol{\Pi}^H\mathbf{G}_{i\eta})^\dagger\mathcal{F}_{ij}\}$	$\mathcal{O}\{M_i^3(L+1)^3\}$
Total	Sum of the rows

arrive at (54b), and the fact that $\mathbf{Z}_{\bar{i}s}^H\boldsymbol{\Sigma}_{\bar{i}\bar{i}}\mathbf{Z}_{\bar{i}s}=\mathbf{I}$ (this relation comes directly from the definition of $\mathbf{Z}_{\bar{i}s}$) together with $\text{tr}\{\text{vec}(\mathbf{A})\,\text{vec}^H(\mathbf{I})\}=\text{tr}\mathbf{A}$ to arrive at (54c). Equation (54c) is the log-likelihood function used in the ML estimation of the channel matrix $\mathbf{H}_i$. Note that in (54c), we did not make use of the relation $\mathbf{Z}_{is}^H\boldsymbol{\Sigma}_{ii}\mathbf{Z}_{is}=\mathbf{I}$ to further simplify the log-likelihood function. This is because we will use this factor to arrive at a form suitable for channel estimation as can be seen in the following.

As it is, (54c) is not convenient to use for the ML channel estimation in unknown noise since $\mathbf{Z}_{i\eta}$ is only an implicit function of the channel. Again, we can apply the channel matrix transformation [20] technique summarized in the beginning of this section. For $i=1,2$, we first obtain the matrix $\mathbf{G}_{i\eta}$ as described in the channel matrix transformation. In a similar way to the development of the MAP estimate, we obtain $\boldsymbol{\Pi}^H\mathbf{G}_{i\eta}$ where $\boldsymbol{\Pi}$ is a permutation matrix. Since the columns of both $\mathbf{Z}_{i\eta}$ and $\boldsymbol{\Pi}^H\mathbf{G}_{i\eta}$ span the orthogonal complement of $\mathbf{H}_i$, then there exists a nonsingular matrix $\mathbf{V}_{i\eta}$, such that $\mathbf{Z}_{i\eta}=\boldsymbol{\Pi}^H\mathbf{G}_{i\eta}\mathbf{V}_{i\eta}$. Substituting this expression of $\mathbf{Z}_{i\eta}$ into (54c), we note that by having retained the term $\mathbf{Z}_{is}^H\boldsymbol{\Sigma}_{ii}\mathbf{Z}_{is}$ in (54c) as mentioned previously, we have

$$\mathcal{L}_{\text{ccd}}\approx-\text{tr}\{(\mathbf{G}_{i\eta}^H\boldsymbol{\Pi}\widehat{\mathbf{Y}}_{is}\widehat{\boldsymbol{\Gamma}})^H(\mathbf{G}_{i\eta}^H\boldsymbol{\Pi}\widehat{\boldsymbol{\Sigma}}_{ii}\boldsymbol{\Pi}^H\mathbf{G}_{i\eta})^\dagger(\mathbf{G}_{i\eta}^H\boldsymbol{\Pi}\widehat{\mathbf{Y}}_{is}\widehat{\boldsymbol{\Gamma}})\}, \tag{55}$$

where we have substituted $\widehat{\boldsymbol{\Gamma}}$ for $\boldsymbol{\Gamma}$ and $\widehat{\boldsymbol{\Sigma}}_{ii}$ for $\boldsymbol{\Sigma}_{ii}$ without affecting the asymptotical property. Now, let $\mathbf{F}_i=\boldsymbol{\Pi}\widehat{\mathbf{Y}}_{is}\widehat{\boldsymbol{\Gamma}}$ and denote $\mathbf{f}_{ij}$ as the jth column of $\mathbf{F}_i$, then

$$\mathbf{G}_{i\eta}^H\mathbf{f}_{ij}=\mathcal{F}_{ij}\mathbf{h}_i, \tag{56}$$

where $\mathcal{F}_{ij}$ can be constructed from $\mathbf{f}_{ij}$ according to (19) of Property (2) of $\mathbf{G}_{i\eta}$. Thus, the ML estimate of $\mathbf{h}_i$, which is in the same form as $\widetilde{\mathbf{h}}$ in (19), can be obtained as

$$\widehat{\mathbf{h}}_i=\arg\min_{\|\mathbf{h}_i\|_2=1}\left\{\mathbf{h}_i^H\left(\sum_{j=1}^{K+L}\mathcal{F}_{ij}^H(\mathbf{G}_{i\eta}^H\boldsymbol{\Pi}\widehat{\boldsymbol{\Sigma}}_{ii}\boldsymbol{\Pi}^H\mathbf{G}_{i\eta})^\dagger\mathcal{F}_{ij}\right)\mathbf{h}_i\right\}. \tag{57}$$

Equation (57) is designated the CCD-ML method of channel estimation. Since the information of $\mathbf{h}_i$ is also embedded in the matrix contained in the parentheses, the IQML [27] algorithm can again be applied to solve this optimization problem with the approximate computation complexity summarized in Table 3.

The computation of the last four lines will be repeated according to the number of iterations. When the number of iteration is small (which is the case according to the simulation results), the complexity of the CCD-ML algorithm will be of the same order as that shown in Table 3.

4. COMPUTER SIMULATION RESULTS

In this section, using computer simulations, we examine the performance of our channel estimation algorithms (MAP, CCD-SS, and CCD-ML) and compare their performance with that of the two subspace methods: the SS [1] and MSS [10] under different SNR. Since the MSS method [10] is developed for channel estimation in unknown correlated noise, it is a main competitor with the algorithms developed in this paper. We, therefore, briefly summarize the MSS algorithm here.

In MSS, we collect two blocks of data $\mathbf{r}(n)$ and $\mathbf{r}(n+1)$, and a cross-correlation is calculated between these two vectors such that $\boldsymbol{\Sigma}_r=\text{E}\{\mathbf{r}(n+1)\mathbf{r}^H(n)\}=\mathbf{H}\cdot\text{E}\{\mathbf{s}(n+1)\mathbf{s}^H(n)\}\mathbf{H}^H=\mathbf{H}\boldsymbol{\Sigma}_s\mathbf{H}^H$ for which the noise correlation term disappears because the noise in the two blocks of data transmitted at different times are *assumed to be uncorrelated* whereas intrablock correlation of the noise is nonzero. Then a new matrix $\boldsymbol{\Sigma}'=\boldsymbol{\Sigma}_r+\boldsymbol{\Sigma}_r^H=\mathbf{H}(\boldsymbol{\Sigma}_s+\boldsymbol{\Sigma}_s^H)\mathbf{H}^H=\mathbf{H}\boldsymbol{\Sigma}_s'\mathbf{H}^H$ is created so that the signal correlation matrix $\boldsymbol{\Sigma}_s'$ is full rank, for which the two transmitted signal blocks need to be either totally correlated or, the block length K has to be equal to the channel order L if the signals are independent. Then the standard SS method is applied to this "noise-cleaned" covariance matrix $\boldsymbol{\Sigma}'$ to obtain the channel coefficients. (Equivalently, this method can also be applied to the model having two versions of the same transmitted signal vector from two different antennas by forming the “noise cleaned” covariance matrix through the cross correlation between the received vectors.)

In the examples below, 40 (for MAP algorithm) or 40 pairs of (for CCD based algorithms) randomly generated channels are used. Our estimation performance are evaluated by averaging over these 40 or 40 pairs of different channels. Over each channel realization, signals are transmitted. At the receiver, we upsample the received signal by a factor M. While in theory, we can choose any value of $M \geq 2$, in practice, to reduce the computational load, we should keep M as low as possible. Therefore, in our simulations, we focus on the case when oversampling is carried out by a factor of $M = 2$ to minimize the additional computational requirement. At the receiver, for the ith trial of each channel realization, utilizing the received signal and noise, we employ the various methods to obtain the estimate $\widehat{\mathbf{h}}^{(i)}$ of the channel. We then evaluate the error of estimation ($\mathbf{e}_i = \widehat{\mathbf{h}}^{(i)} - \mathbf{h}$). The criterion of performance comparison is the normalized root mean square error (NRMSE) of estimates defined as

$$\overline{\epsilon}_j = \sqrt{\frac{1}{N_T} \sum_{i=1}^{N_T} \frac{\left\|\widehat{\mathbf{h}}^{(i)} - \mathbf{h}\right\|^2}{\|\mathbf{h}\|^2}}, \tag{58}$$

where $\overline{\epsilon}_j$ denote the NRMSE performance for the jth channel realization and N_T is the number of trials for each channel realization. The NRMSE of the channel estimation for each algorithm is averaged over all the channel realizations which can be calculated as

$$\overline{\epsilon} = \frac{1}{J} \sum_{j=1}^{J} \overline{\epsilon}_j. \tag{59}$$

As mentioned above, J, the total number of channel realizations, is 40.

Example 1. In this example, we examine the performance of the algorithms MAP, MSS, and SS which are developed under the condition that only one receiving antenna is available. The transmitted signals are randomly chosen from the 4-QAM constellation and transmitted through the ISI induced FIR channel with order $L = 3$. During the collection of $N = 200$ snapshots of the data blocks, the channel is assumed to be stationary. We choose the additive correlated noise to have the similar model as presented in [10] such that the noise subsamples within one signal sampling period are assumed to have the correlation matrix given by

$$\begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix}^H, \tag{60}$$

whereas the noise subsamples from two different sampling periods are assumed to be uncorrelated. We designate this noise Model 1. The estimation error is averaged over $N_T = 100$ trials for each channel realization.

As mentioned in the beginning of Section 3, the condition that $K \geq ((M+1)/(M-1))L$ has to be satisfied for the MAP algorithm to apply the channel matrix transformation.

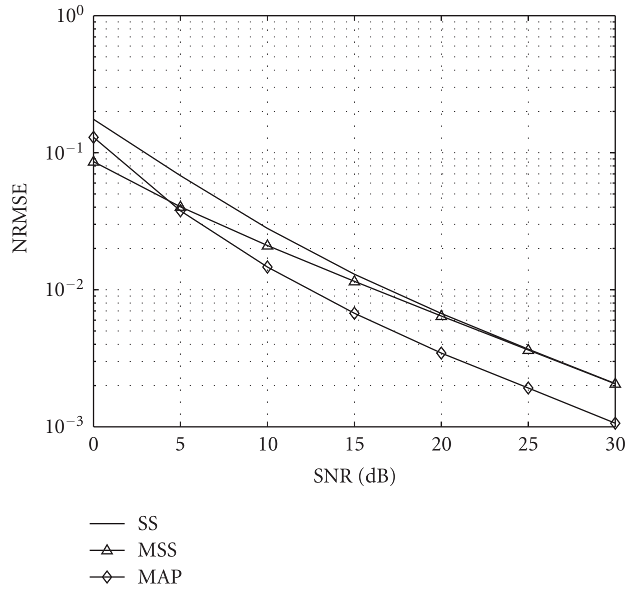

Figure 1: Comparison of NRMSE performance of SS, MSS, and MAP under Noise Model 1.

Here, we choose the block size to be $K = 12$. The weighting matrix $(\mathbf{G}_\eta^H \mathbf{\Pi} \widehat{\mathbf{\Sigma}}_r \mathbf{\Pi}^H \mathbf{G}_\eta)^\dagger$ in (38) is initialized by the estimate from the SS method and the IQML algorithm is then applied iteratively. The stopping criterion is such that the norm of the difference vector between two consecutive iterations is less than 10^{-6} and the average number of iterations for each estimate is taken over 100 trials. Also, as discussed previously in this section, the MSS method can be applied with one receiving antenna if the transmitted signals are fully correlated such that the lag-K correlation matrix of the signals is full rank. Thus, for the MSS method, we transmit the same signal vector $\mathbf{s}(n)$ in two consecutive blocks and obtain the MSS estimates. Now, since the MAP algorithm does not need two correlated signal vectors, the repeated transmission in MSS is redundant for the MAP method. Therefore, for fairness of comparison, the length of the transmitted signal block for MSS is chosen to be half of that for the MAP method.

Figure 1 shows the NRMSE performance of the MAP algorithm in comparison with those of the SS and MSS methods with respect to different SNR. As expected, since the SS method is developed under the assumption of white noise, it does not work well under correlated noise environments and therefore, we can see that under all the SNR considered, both the MSS method and the MAP algorithm are superior in performance to the SS method. Furthermore, the MAP algorithm shows substantially better performance than the MSS algorithm under higher SNR where the performance gain of the MAP algorithm over that of MSS is considerable. The average number of iterations needed in the MAP algorithm to achieve such performance are shown in Table 4. It can be observed that the number of iterations required is small. At high SNR (20 dB and beyond), the performance of SS and MSS become quite close because at high SNR, the effect of the correlation of the noise becomes less dominant.

Table 4: Averaged number of iterations for MAP to acquire the NRMSE performance at different SNR in Figure 1.

SNR (dB)	0 dB	5 dB	10 dB	15 dB	20 dB	25 dB	30 dB
Averaged number of iterations for MAP	4.6761	2.5617	2.0250	1.9750	1.6422	1.1533	1.0078
NRMSE (MAP)	0.1295	0.0377	0.0146	0.0067	0.0034	0.0019	0.0011

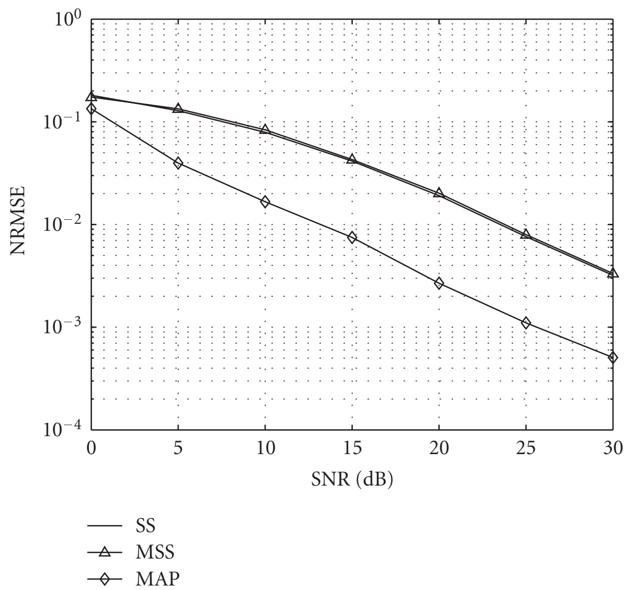

Figure 2: Comparison of NRMSE performance of SS, MSS, and MAP under AR noise Model 2.

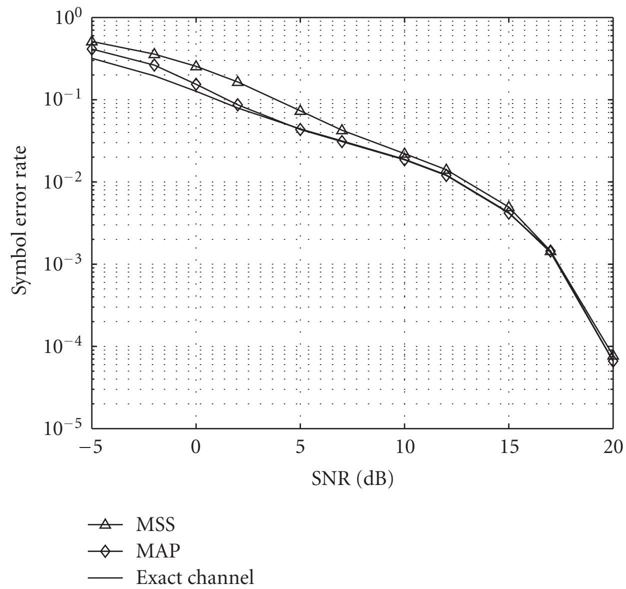

Figure 3: SER comparison by applying channel estimates of MSS and MAP or using the exact channel coefficients in separately received AR noise under different SNR.

Example 2. The noise subsamples in Example 1 are assumed to be correlated within one signal sampling period but are uncorrelated with the noise subsamples outside the sampling periods. This assumption is not easy to satisfy in practical situations. In the present example, we test the performance of the MAP algorithm in comparison with those of the SS and MSS methods under a second-order AR model having coefficients $[1, -1.8, 0.82]$. We designate this noise Model 2. The channel parameters remain the same as in Example 1. The performance of the various algorithms in terms of the NRMSE of estimated channel coefficients are shown in Figure 2.

It is observed that the MAP algorithm still performs just as well, whereas due to the violation of its noise assumption, the MSS method has a performance even slightly inferior to that of the SS algorithm which assumes a white noise environment. On the other hand, the MAP algorithm assumes that the noise is simply *unknown* and therefore is independent of the noise model and robust to change the noise environments.

Using a zero-forcing (ZF) equalizer together with a threshold detector, the corresponding symbol error rates (SER) using the various channel estimation methods are shown in Figure 3. We can see that the SER performance of the MAP algorithm is very close to the SER when the channel is exactly known.

Example 3. In this example, we compare the performance of CCD-SS, CCD-ML, and MSS. The development of CCD-SS and CCD-ML are based on having two versions of received data $\mathbf{r}_1$ and $\mathbf{r}_2$ when the same signal vector is transmitted. The noise in the two received signals are assumed uncorrelated. Such a scenario usually occurs in the case when there are two receiver antennae sufficiently separated, and the uncorrelated noise also fits well with the assumption made in the MSS method. For MSS, the cross-correlation of these two received vectors are calculated so that the effect of the uncorrelated noise in the two separate channels are removed. For CCD-SS and CCD-ML, on the other hand, these two received signal vectors collected by two receiving antennae are stacked up and CCD is applied to the correlation matrix of the stacked vector.

In this example, we assume that the channel order of the two channels are the same, that is, $L_1 = L_2 = 4$. The upsampling factors are $M_1 = M_2 = 2$, the block size is chosen to be $K = 12$ to satisfy $K \geq ((M+1)/(M-1))L$ in the channel matrix transformation. We choose the noise model to be that given by Example 2. For this noise model, since there are two separate transmission channels, the noise will be independent and therefore, the AR model will be used to generate two independent noise sequences in the two channels. This satisfies the assumption made in the MSS algorithm. To facilitate the CCD-ML method, the weighting matrix in

Table 5: Averaged number of iterations for CCD-ML to achieve the NRMSE performance in Figure 4.

SNR (dB)	0 dB	5 dB	10 dB	15 dB	20 dB	25 dB	30 dB
averaged number of iterations for CCD-ML	1.9359	1.2691	1.0412	0.9707	0.7579	0.3697	0.1064
NRMSE (CCD-ML)	0.0210	0.0099	0.0053	0.0029	0.0016	0.0009	0.0005

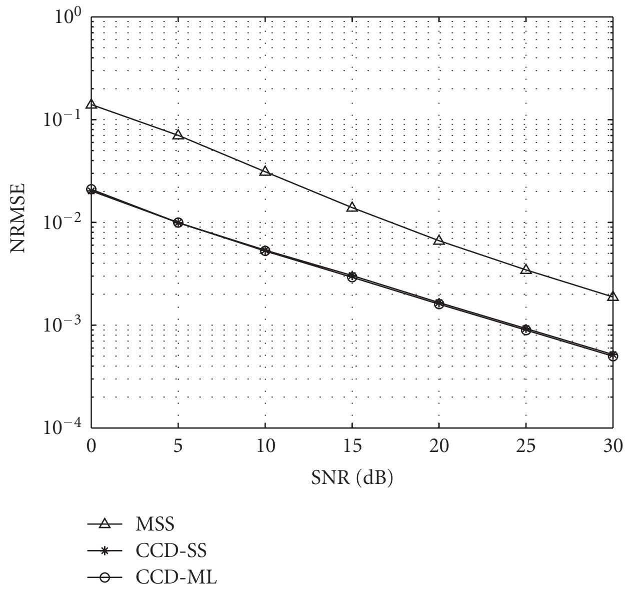

Figure 4: Comparison of NRMSE performance for MSS, CCD-SS, and CCD-ML under separately received AR noise.

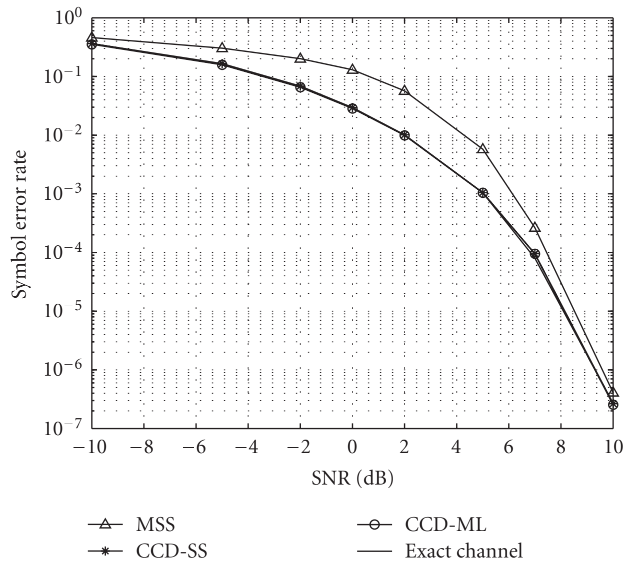

Figure 5: SER comparison by applying channel estimates of MSS, CCD-SS, and CCD-ML or using the exact channel coefficients in separately received AR noise under different SNR.

(57) is initialized using the channel coefficients estimated from CCD-SS and then the channel is estimated iteratively with the updated weighting matrix. We use the same iteration stopping criterion as in Examples 1 and 2. Table 5 shows the average number of iterations needed for each CCD-ML estimate over 100 trials under different SNR, averaged over 40 different channel realizations, and as can be seen, these average numbers of iterations are reasonably small under a wide range of SNR.

Figure 4 shows the performance of the three methods. It can be observed that both the CCD-ML and the CCD-SS methods are far superior in performance to the MSS method under all the SNR considered. Thus, employing CCD definitely provides us with performance advantage. While the CCD-ML method yields the best NRMSE performance, as shown in Figure 4, it is only marginally better than CCD-SS. This may lead us to conclude that the extra computation needed by the CCD-ML algorithm may not be worth the amount of improvement achieved. The corresponding SER are shown in Figure 5, where the performance of both CCD-SS and CCD-ML are observed to be very close to that when the channel is exactly known.

Several tests following similar scenarios as those in the above examples have been carried out under other channel and noise models. Similar observations as presented in the above examples are obtained [31, 32].

5. CONCLUSION

In this paper, we address the important practical problem of FIR channel estimation in unknown correlated noise environments. We examine the effect of additive correlated noise with arbitrary unknown covariance matrix in FIR channels and develop different algorithms according to the different number of antennae available at the receiver. For receivers having only one antenna, we develop an algorithm which maximizes the criterion of a posteriori PDF (MAP) derived by employing the Jeffreys' principle. For receivers having two antennae and therefore, having two copies of the transmitted signal vector infested with independent unknown noise, we employ the *canonical correlation decomposition* (CCD) to separate the signal and noise subspaces arriving at the CCD-SS algorithm. By further examining the asymptotic distribution of the projected noise onto the estimated signal subspace, we formulate the likelihood function for which we could maximize and obtain the CCD-ML algorithm of channel estimation. The advantage of these new methods is that they do not need to assume any noise model, and therefore, their performance are relatively robust. All these algorithms when employed under the conditions for which they have been developed (i.e., either having one antenna or two antennae in the receiver) have been shown to have superior performance to existing (SS and MSS) methods. It is also observed that there is very little difference in performance between the CCD-SS and CCD-ML methods. Hence, if there are two antennae available, the comparatively simpler algorithm of CCD-SS will be favored.

Throughout the paper, we have assumed that the order of the channel is either known or has been accurately estimated using appropriate methods [16–19]. However, the error in the channel order estimation may affect the outcome of the channel estimation. Since the MAP and the CCD methods employ the noise subspace to identify the channel, if the channel order is underestimated, the estimated noise subspace will contain vectors which belong to the signal subspace. The identification of the channel will then be based on an erroneous subspace leading to significant errors. On the other hand, if the channel order is overestimated, the dimension of the estimated noise subspace will be reduced from that of the true noise subspace. If this dimension difference is not substantial, the algorithms will be affected only in their estimation accuracy because in this case, the noise subspace is not fully utilized. Simulations [1, 32] show some of the relative effects on the estimation of the channels due to channel order under- and overestimation.

ACKNOWLEDGMENT

Part of the paper was presented at the International Symposium on Signal Processing and Its Applications, Sydney, Australia, August 2005 [31].

REFERENCES

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

[2] J. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 4th edition, 2001.

[3] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1910–1917, 2003.

[4] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 6, pp. 1615–1624, 2003.

[5] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *IEEE Signal Processing Magazine*, vol. 21, no. 6, pp. 12–25, 2004.

[6] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[7] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 340–349, 1994.

[8] A. Jeremić, T. A. Thomas, and A. Nehorai, "OFDM channel estimation in the presence of interference," *IEEE Transactions on Signal Processing*, vol. 52, no. 12, pp. 3429–3439, 2004.

[9] M. Frikel, W. Utschick, and J. Nossek, "Blind noise and channel estimation," in *Proceedings of the 10th IEEE Signal Processing Workshop on Statistical Signal and Array Processing (SSAP '00)*, pp. 141–145, Pennsylvania, Pa, USA, August 2000.

[10] K. Abed-Meraim, Y. Hua, P. Loubaton, and E. Moulines, "Subspace method for blind identification of multichannel FIR systems in noise field with unknown spatial covariance," *IEEE Signal Processing Letters*, vol. 4, no. 5, pp. 135–137, 1997.

[11] A. Dogandžić, W. Mo, and Z. Wang, "Semi-blind SIMO flat-fading channel estimation in unknown spatially correlated noise using the EM algorithm," *IEEE Transactions on Signal Processing*, vol. 52, no. 6, pp. 1791–1797, 2004.

[12] A. Dogandžić and A. Nehorai, "Space-time fading channel estimation and symbol detection in unknown spatially correlated noise," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 457–474, 2002.

[13] W. B. M. Schwartz and S. Stein, *A Communication Systems and Techniques*, McGraw-Hill, New York, NY, USA, 1966.

[14] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics. Vol. 2*, Charles Griffin, London, UK, 1961.

[15] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, New York, NY, USA, 1982.

[16] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[17] Q. T. Zhang and K. M. Wong, "Information theoretic criteria for the determination of the number of signals in spatially correlated noise," *IEEE Transactions on Signal Processing*, vol. 41, no. 4, pp. 1652–1663, 1993.

[18] H. Gazzah, P. A. Regalia, J.-P. Delmas, and K. Abed-Meraim, "A blind multichannel identification algorithm robust to order overestimation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1449–1458, 2002.

[19] H. Yan, "New autocorrelation decomposition method for robust channel estimation with unknown channel order," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '04)*, vol. 1, pp. 12–17, Atlanta, Ga, USA, March 2004.

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

[21] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, Mass, USA, 1973.

[22] H. Jeffreys, *Theory of Probability*, Oxford University Press, London, UK, 3rd edition, 1961.

[23] K. M. Wong, J. P. Reilly, Q. Wu, and S. Qiao, "Estimation of the directions of arrival of signals in unknown correlated noise—I: the MAP approach and its implementation," *IEEE Transactions on Signal Processing*, vol. 40, no. 8, pp. 2007–2017, 1992.

[24] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1984.

[25] H. L. V. Trees, *Detection, Estimation and Modulation Theory. Part I*, John Wiley & Sons, New York, NY, USA, 1968.

[26] A. Basilevsky, *Applied Matrix Algebra in the Statistical Sciences*, North-Holland, Amsterdam, The Netherlands, 1983.

[27] 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, no. 5, pp. 1081–1089, 1986.

[28] G. Strang, *Introduction to Linear Algebra*, Wellesley-Cambridge Press, Wellesley, Mass, USA, 1993.

[29] K. M. Wong, Q. Wu, and P. Stoica, "Generalized correlation decomposition applied to array processing in unknown noise

environments," in *Advances in Spectrum Analysis and Array Processing Vol. III*, S. Haykin, Ed., pp. 219–323, Prentice-Hall, Upper Saddle River, NJ, USA, 1995.

[30] Q. Wu and K. M. Wong, "UN-MUSIC and UN-CLE: an application of generalized correlation analysis to the estimation of the direction of arrival of signals in unknown correlated noise," *IEEE Transactions on Signal Processing*, vol. 42, no. 9, pp. 2331–2343, 1994.

[31] X. He and K. M. Wong, "Channel estimation in unknown noise: application of canonical correlation decomposition in subspaces," in *Proceedings of the 8th International Symposium on Signal Processing and Its Applications (ISSPA '05)*, vol. 2, pp. 475–478, Sydney, Australia, August-September 2005.

[32] X. He, "Blind FIR channel estimation in the presence of unknown noise," MASc dissertation, McMaster University, Hamilton, Ontario, Canada, 2005.

Xiaojuan He was born in Datong, Shanxi Province, China. She attended the Nanjing University of Science and Technology and was awarded the degree of Bachelor of Engineering (with distinction) in June 2001. She continued her studies in the same institute and obtained the Master of Engineering degree in July 2004. She then attended McMaster University pursuing further studies in the Department of Electrical Engineering and completed the degree of Master of Applied Science (with distinction) in June 2006. Her thesis was awarded the "Outstanding Master's Thesis" prize. After working for a short while as a Research Associate at McMaster University, she is now working as a DSP Engineer at Lyngsoe Systems Ltd. in Canada.

Kon Max Wong received his B.S. (Eng.), DIC, Ph.D., and D.S. (Eng.) degrees, all in electrical engineering, from the University of London, England, in 1969, 1972, 1974, and 1995, respectively. In 1981, he joined McMaster University, Hamilton, Canada, where he has been a Professor since 1985 and served as Chairman of the Department of Electrical and Computer Engineering in 1986–1987, 1988–1994, and 2003–2008. At present, he holds the Canada Research Professorship of Signal Processing and is the Director of the Communication Technology Research Centre. His research interest is in signal processing and communication theory and has published over 200 papers in the area. Professor Wong was the recipient of the IEE Overseas Premium for the best paper in 1989, and is a Fellow of IEEE, a Fellow of the IEE, a Fellow of the Royal Statistical Society, and a Fellow of the Institute of Physics. He was an Associate Editor of the IEEE Transaction on Signal Processing, and served as Chair of the IEEE Sensor Array and Multi-channel Signal Processing Technical Committee in 2002–2004. Professor Wong was awarded a medal by the IBC for his "outstanding contributions to the research and education in signal processing" in May 2000. He is also Coauthor of the paper that received the IEEE Signal Processing Society *Best Young Author* Paper Award, 2006.