Title	Performance of an ℓ1 Regularized Subspace-based MIMO Channel Estimation with Random Sequences
Author(s)	Takano, Yasuhiro; Juntti, Markku; Matsumoto, Tad
Citation	IEEE Wireless Communications Letters, 5(1): 112-115
Issue Date	2015-12-04
Type	Journal Article
Text version	author
URL	http://hdl.handle.net/10119/12999
Rights	This is the author's version of the work. Copyright (C) 2015 IEEE. IEEE Wireless Communications Letters, 5(1), 2015, pp.112-115. DOI:10.1109/LWC.2015.2505727. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Description	

Performance of an $\ell 1$ Regularized Subspace-based MIMO Channel Estimation with Random Sequences

Yasuhiro Takano, *Student Member, IEEE,* Markku Juntti, *Senior Member, IEEE,* and Tad Matsumoto, *Fellow, IEEE*

***Abstract*—The conventional $\ell 2$ multi-burst (MB) channel estimation can achieve the Cramér-Rao bound asymptotically by using the subspace projection. However, the $\ell 2$ MB technique suffers from the noise enhancement problem if the training sequences (TSs) are not ideally uncorrelated. We clarify that the problem is caused by an inaccurate noise whitening process. The $\ell 1$ regularized MB channel estimation can, however, improve the problem by a channel impulse response length constraint. Asymptotic performance analysis shows that the $\ell 1$ MB can improve channel estimation performance significantly over the $\ell 2$ MB technique in a massive multiple-input multiple-output system when the TSs are not long enough and not ideally uncorrelated.**

***Index Terms*—Subspace-based channel estimation, noise whitening, massive MIMO, pilot contamination, compressive sensing.**

I. Introduction

CONVENTIONAL $\ell 2$ multi-burst (MB) channel estimation techniques (e.g., [1]) can achieve the Cramér-Rao bound (CRB) asymptotically under the following two assumptions: 1) channel impulse responses (CIRs) follow the subspace channel model assumption [1]; and 2) the training sequences (TSs) are ideally uncorrelated between transmission (TX) streams. However, finding optimal TS combinations is a non-polynomial (NP) hard problem in a massive multiple-input multiple-output (MIMO) system, since binomial coefficients increase in factorial orders. Moreover, the number of the ideally uncorrelated sequences with a given bandwidth is limited, which can cause the pilot contamination problem [2].

This letter studies, therefore, performance of the MB algorithm where the assumption 2) is not always correct. Specifically, random TS is assumed as a typical *moderately* uncorrelated sequence. This letter shows that the $\ell 2$ MB technique with the random TS can suffer from the noise enhancement problem due to inaccurate noise whitening process. However, the $\ell 1$ MB algorithm [3] can improve the problem by a CIR length constraint. The main objective of this letter is to clarify the reason for the improvement.

This letter is organized as follows. Section II defines the signal model and summarizes the $\ell 1$ MB algorithm. Section III describes analytical MSE performance by taking account of the noise whitening accuracy. Section IV verifies the analytical performance via computer simulations. Section V shows concluding remarks.

<u>*Notations*</u>: The bold lower-case $\mathbf{x}$ and upper-case $\mathbf{X}$ denote a vector and a matrix, respectively. For matrix $\mathbf{X}$, its transpose and transposed conjugate are denoted as $\mathbf{X}^{\mathrm{T}}$ and $\mathbf{X}^{\mathrm{H}}$, respectively. $\mathbf{X}^{-1}$ and $\mathbf{X}^{\dagger}$ denote the matrix inverse and the Moore-Penrose pseudoinverse of $\mathbf{X}$, respectively. The Cholesky decomposition of $\mathbf{X}$ is denoted by $\mathbf{X}^{\mathrm{H}/2}\mathbf{X}^{1/2}$. $\mathbf{X}|_{i:j}$ is a submatrix composed of the the i-th to j-th column vectors in the matrix $\mathbf{X}$. The expectation and covariance matrices of $\mathbf{X}(l)$ are denoted as $\mathbb{E}_l^L[\mathbf{X}(l)] = \frac{1}{L}\sum_{j=l-L+1}^{l}\mathbf{X}(j)$ and $\mathbb{K}_l^L[\mathbf{X}(l)] = \frac{1}{L}\sum_{j=l-L+1}^{l}\mathbf{X}^{\mathrm{H}}(j)\ \mathbf{X}(j)$, respectively. Moreover, $\mathbb{E}[\mathbf{X}(l)]$ and $\mathbb{K}[\mathbf{X}(l)]$ are $\mathbb{E}_l^{\infty}[\mathbf{X}(l)]$ and $\mathbb{K}_l^{\infty}[\mathbf{X}(l)]$, respectively. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix.

II. Preliminaries

A. System Model

The same MIMO system as that in [3] is used. This letter assumes that, however, channel estimation is performed with TSs only. The received signal corresponding to the transmitted TSs can be described, as $\mathcal{Y}(l) = \mathcal{H}(l)\mathcal{X}(l) + \mathcal{Z}$ at the burst timing l, where

$$\begin{aligned}
\mathcal{Y}(l) &= [\mathbf{y}_1(l), \cdots, \mathbf{y}_{N_R}(l)]^{\mathrm{T}} &\in \mathbb{C}^{N_R \times \tilde{N}_t}, \\
\mathcal{X}(l) &= [\mathbf{X}_1^{\mathrm{T}}(l), \cdots, \mathbf{X}_{N_T}^{\mathrm{T}}(l)]^{\mathrm{T}} &\in \mathbb{C}^{WN_T \times \tilde{N}_t}, \\
\mathcal{H}(l) &= [\mathbf{H}_1(l), \cdots, \mathbf{H}_{N_T}(l)] &\in \mathbb{C}^{N_R \times WN_T}, \\
\mathcal{Z} &= [\mathbf{z}_1, \cdots, \mathbf{z}_{N_R}]^{\mathrm{T}} &\in \mathbb{C}^{N_R \times \tilde{N}_t},
\end{aligned}$$

with $\tilde{N}_t = N_t + W$. N_T and N_R denote the number of transmit (Tx) and receive (Rx) antennas, respectively. Matrix $\mathbf{X}_k(l) \in \mathbb{C}^{W \times \tilde{N}_t}$ is a Toeplitz matrix whose first row vector is $[\mathbf{x}_k^{\mathrm{T}}(l), \mathbf{0}_W^{\mathrm{T}}] \in \mathbb{C}^{1 \times \tilde{N}_t}$, where $\mathbf{x}_k(l)$ denotes a length N_t TS vector. The $N_R \times W$ matrix $\mathbf{H}_k(l)$ describes the CIRs between the k-th Tx and N_R Rx antennas, where the channel lengths are assumed at most W symbols. Noise vector $\mathbf{z}_n$ at the n-th Rx antenna follows $\mathcal{CN}(\mathbf{0}, \sigma_z^2\mathbf{I}_{\tilde{N}_t})$ with the variance σ_z^2 depending on the average signal-to-noise ratio (SNR).

B. $\ell 1$ regularized MB Channel Estimation Algorithm

The $\ell 1$ MB algorithm [3] performs CIR length regularized $\ell 2$ MB channel estimation under an assumption that significant CIR taps are distributed over the first w symbols according to the received SNR. The $\ell 1$ MB estimation performs the subspace projection per a TX stream and it obtains $N_R \times w$

Y. Takano and T. Matsumoto are with Japan Advanced Institute of Science and Technology (JAIST) 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan (e-mail: {yace.takano; matumoto}@jaist.ac.jp).

M. Juntti and T. Matsumoto are with the Centre for Wireless Communications, University of Oulu, FI-90014, Oulu, Finland (e-mail: markku.juntti@ee.oulu.fi).

This research was conducted under the financial support of the double degree program between JAIST and University of Oulu.

channel estimate matrices $\hat{\mathbf{G}}^{MB}_{[w]k}(l)$, $1 \leq k \leq N_T$, for each TX stream. The w-th possible solution corresponding to the length w CIR constraint is, hence, described as $\hat{\mathcal{H}}^{MB}_{[w]}(l) = [\hat{\mathbf{G}}^{MB}_{[w]1}(l), \cdots, \hat{\mathbf{G}}^{MB}_{[w]N_T}(l)]\mathcal{P}^{\mathrm{T}}_{[w]}$, where $\mathcal{P}_{[w]} = \mathbf{I}_{N_T} \otimes \mathbf{P}_{[w]}$ with the $W \times w$ matrix $\mathbf{P}_{[w]} = [\mathbf{I}_w \ \mathbf{O}]^{\mathrm{T}}$. $\otimes$ denotes the Kronecker product. The optimal solution $\hat{\mathcal{H}}^{MB}_{[\hat{w}]}(l)$ may be determined by minimizing the Akaike information criterion (AIC) [4] as $\hat{w} = \arg\min_{1 \leq w \leq W} \mathtt{AIC}(\hat{\mathcal{H}}^{MB}_{[w]}(l))$.

In the case the channel estimation is performed with TSs only, the $N_R \times w$ estimated matrix $\hat{\mathbf{G}}^{MB}_{[w]k}(l)$ is given by

$$\hat{\mathbf{G}}^{MB}_{[w]k}(l) = \hat{\hat{\mathbf{G}}}^{LS}_{[w]k}(l) \cdot \hat{\hat{\mathbf{\Pi}}}_{[w]k} \cdot \bar{\mathbf{Q}}^{-\mathrm{H}}_{[w]kk} \tag{1}$$

for the k-th TX stream, where the $w \times w$ matrix $\bar{\mathbf{Q}}_{[w]ij}$ denotes the (i,j)-th block matrix of $\bar{\mathbf{R}}^{1/2}_{\mathcal{X}\mathcal{X}[w]}$ with $\bar{\mathbf{R}}_{\mathcal{X}\mathcal{X}[w]} = \mathbb{E}^L_l[\mathcal{P}^{\mathrm{T}}_{[w]}\mathbf{R}_{\mathcal{X}\mathcal{X}}(l)\mathcal{P}_{[w]}]$ and $\mathbf{R}_{\mathcal{X}\mathcal{X}}(l) = \mathcal{X}(l)\mathcal{X}^{\mathrm{H}}(l)$. Moreover,

$$\begin{aligned}\hat{\hat{\mathbf{G}}}^{LS}_{[w]k}(l) &\triangleq \hat{\mathbf{G}}^{LS}_{[w]k}(l) \cdot \bar{\mathbf{Q}}^{\mathrm{H}}_{[w]kk} \\ &+ \sum_{i=k+1}^{N_T} \left\{\hat{\mathbf{G}}^{LS}_{[w]i}(l) - \mathfrak{G}_{[w]i}(l)\right\} \bar{\mathbf{Q}}^{\mathrm{H}}_{[w]ki}.\end{aligned} \tag{2}$$

with $\mathfrak{G}_{[w]i}(l) = \hat{\mathbf{G}}^{MB}_{[w]i}(l)$, where $\hat{\mathbf{G}}^{LS}_{[w]k}(l)$ is the LS channel estimate corresponding to an $N_R \times w$ CIR matrix $\mathbf{G}_{[w]k}(l) = \mathbf{H}_k(l)\mathbf{P}_{[w]}$. The projection matrix $\hat{\hat{\mathbf{\Pi}}}_{[w]k}$ denotes $\hat{\hat{\mathbf{V}}}_{[w]k}|_{1:r_k}(\hat{\hat{\mathbf{V}}}_{[w]k}|_{1:r_k})^{\dagger}$, where the unitary matrix $\hat{\hat{\mathbf{V}}}_{[w]k}$ is the singular vectors of the covariance matrix $\mathbb{K}^L_l[\hat{\hat{\mathbf{G}}}^{LS}_{[w]k}(l)]$. The parameters r_k and L denote the number of paths for the k-th TX stream and the sliding window length in the MB algorithm, respectively.

Notice that (2) is performed for the noise whitening. Let us denote $\Delta\hat{\hat{\mathbf{G}}}^{LS}_{[w]k}(l) = \hat{\hat{\mathbf{G}}}^{LS}_{[w]k}(l) - \tilde{\mathbf{G}}_{[w]k}(l)$ with $\tilde{\mathbf{G}}_{[w]k}(l) = \mathbf{G}_{[w]k}(l)\bar{\mathbf{Q}}^{\mathrm{H}}_{[w]kk}$ and concatenate the N_T residual matrices into an $N_R \times wN_T$ matrix as $\Delta\hat{\hat{\mathcal{G}}}^{LS}_{[w]}(l) = [\Delta\hat{\hat{\mathbf{G}}}^{LS}_{[w]1}(l), \cdots, \Delta\hat{\hat{\mathbf{G}}}^{LS}_{[w]N_T}(l)]$. Suppose $\mathfrak{G}_{[w]i}(l) = \mathbf{G}_{[w]i}(l)$ in (2), we observe that

$$\begin{aligned}\mathbb{K}^L_l[\Delta\hat{\hat{\mathcal{G}}}^{LS}_{[w]}(l)] &= \sigma^2_z N_R \bar{\mathbf{R}}^{1/2}_{\mathcal{X}\mathcal{X}[w]}\mathbb{E}^L_l\left[\mathbf{R}^{-1}_{\mathcal{X}\mathcal{X}[w]}(l)\right]\bar{\mathbf{R}}^{\mathrm{H}/2}_{\mathcal{X}\mathcal{X}[w]} \\ &\approx \sigma^2_z N_R \mathbf{I}_{WN_T}\end{aligned} \tag{3}$$

holds when the TSs are fixed to a consistent sequence or the TSs are ideally uncorrelated $\mathbf{R}_{\mathcal{X}\mathcal{X}}(l)/N_t \approx \mathbf{I}_{WN_T}$ for $\forall l$.

III. MSE Analysis

The burst index l is omitted for the sake of simplicity.

Theorem 1. *Denote the channel estimation error $\hat{\mathbf{H}}^{MB}_{[w]k} - \mathbf{H}_{[w]k}$ by $\Delta\hat{\mathbf{H}}^{MB}_{[w]k}$. The MSE for the $\ell 1$ MB estimate $\hat{\mathbf{H}}^{MB}_{[w]k}$ can be decomposed into the following three terms:*

$$\begin{aligned}\mathbb{E}\left[\|\Delta\hat{\mathbf{H}}^{MB}_{[w]k}\|^2\right] &= \mathbb{E}\left[\|\mathbf{H}^{\perp}_{[w]k}\|^2\right] \\ &+ \mathbb{E}\left[\|\epsilon_{\mathcal{Z},k}(w)\|^2\right] + \mathbb{E}\left[\|\epsilon_{\Pi,k}(w)\|^2\right],\end{aligned} \tag{4}$$

where the discarded part of CIR $\mathbf{H}^{\perp}_{[w]k}$ due to the CIR length constraint, the residual noise $\epsilon_{\mathcal{Z},k}(w)$ and the projection error $\epsilon_{\Pi,k}(w)$ are respectively defined as

$$\mathbf{H}^{\perp}_{[w]k} = \mathbf{H}_k(\mathbf{I}_W - \mathbf{P}_{[w]}\mathbf{P}^{\mathrm{T}}_{[w]}), \tag{5}$$

$$\epsilon_{\mathcal{Z},k}(w) = \Delta\hat{\hat{\mathbf{G}}}^{LS}_{[w]k} \cdot \hat{\hat{\mathbf{\Pi}}}_{[w]k} \cdot \bar{\mathbf{Q}}^{-\mathrm{H}}_{[w]kk}, \tag{6}$$

$$\epsilon_{\Pi,k}(w) = \tilde{\mathbf{G}}_{[w]k} \cdot \Delta\hat{\hat{\mathbf{\Pi}}}_{[w]k} \cdot \bar{\mathbf{Q}}^{-\mathrm{H}}_{[w]kk}. \tag{7}$$

Furthermore, $\Delta\hat{\hat{\mathbf{\Pi}}}_{[w]k} = \hat{\hat{\mathbf{\Pi}}}_{[w]k} - \tilde{\mathbf{\Pi}}_{[w]k}$, where $\tilde{\mathbf{\Pi}}_{[w]k}$ is obtained from the first r_k singular vectors of $\mathbb{K}[\tilde{\mathbf{G}}_{[w]k}]$.

Proof. $\mathbb{E}\left[\|\Delta\hat{\mathbf{H}}^{MB}_{[w]k}\|^2\right] = \mathbb{E}\left[\|\mathbf{H}^{\perp}_{[w]k}\|^2\right] + \mathbb{E}\left[\|\Delta\hat{\mathbf{G}}^{MB}_{[w]k}\|^2\right]$, where $\Delta\hat{\mathbf{G}}^{MB}_{[w]k} = \epsilon_{\mathcal{Z},k}(w) + \epsilon_{\Pi,k}(w)$. Moreover, $\mathrm{tr}\{\mathbb{E}[\epsilon^{\mathrm{H}}_{\Pi,k}(w) \cdot \epsilon_{\mathcal{Z},k}(w)]\} = 0$ since $\mathbb{E}[\tilde{\mathbf{G}}^{\mathrm{H}}_{[w]k} \cdot \Delta\hat{\hat{\mathbf{G}}}^{LS}_{[w]k}] = \mathbf{O}$. $\square$

<u>*Remark*</u>: For TSs satisfying $\mathbb{E}^L_l[\mathbf{R}_{\mathcal{X}\mathcal{X}}(l)]/N_t = \mathbf{I}_{WN_T}$, we have $\bar{\mathbf{Q}}^{-1/2}_{[w]kk} = \mathbf{I}_w/\sqrt{N_t}$. Hence,

$$\begin{aligned}\mathbb{E}[\|\epsilon_{\mathcal{Z},k}(w)\|^2] &= \frac{1}{N_t}\mathrm{tr}\left\{\mathbb{K}^L_l\left[\Delta\hat{\hat{\mathbf{G}}}^{LS}_{[w]k}(l)\right]\right\}\frac{r_k}{w} \\ &= \sigma^2_z N_R \frac{\omega(w)}{N_t} r_k,\end{aligned} \tag{8}$$

where we define *whitening ratio* $\omega(w)$ as

$$\begin{aligned}\omega(w) &= \mathrm{tr}\left\{\bar{\mathbf{R}}_{\mathcal{X}\mathcal{X}[w]} \cdot \mathbb{E}^L_l[\mathbf{R}^{-1}_{\mathcal{X}\mathcal{X}[w]}(l)]\right\} \ / \ \mathrm{tr}\{\mathbf{I}_{wN_T}\} \\ &= N_t \cdot \mathrm{tr}\left\{\mathbb{E}^L_l[\mathbf{R}^{-1}_{\mathcal{X}\mathcal{X}[w]}(l)]\right\} \ / \ wN_T.\end{aligned} \tag{9}$$

It should be noted that $\mathbb{E}^L_l[\mathbf{R}^{-1}_{\mathcal{X}\mathcal{X}[w]}(l)] = \mathbf{I}_{wN_T}/N_t$ is not always satisfied although $\bar{\mathbf{R}}_{\mathcal{X}\mathcal{X}[w]} \approx N_t\mathbf{I}_{wN_T}$. This is because $(\mathbf{A}+\mathbf{B})^{-1} = (\mathbf{A}^{-1}+\mathbf{B}^{-1})$ does not hold in general for arbitrary invertible matrices $\mathbf{A}$ and $\mathbf{B}$.

IV. Numerical Examples

A. Simulation Setups

The CIRs are generated with the spatial channel model (SCM) [5]. This letter assumes 4×4 and 16×16 MIMO channels, where the antenna element spacing at the base station and the mobile station are, respectively, set at 4 and 0.5 wavelength. Spatial parameters such as the direction of arrival (DoA) are randomly chosen. Moreover, six path fading channel realizations based on the Vehicular-A model [5] with a 30 km/h (VA30) mobility is assumed. The path positions are set at $\{1\ 3.2\ 6\ 8.6\ 13.1\ 18.6\}$ symbol timings assuming that a TX bandwidth is 7 MHz with a carrier frequency of 2 GHz. However, the CIRs observed at the receiver can be distributed over more than 19 symbol duration due to the effect of the matched filtering. The maximum CIR length W is, hence, set at 31. Moreover, the path number r_k is assumed to be known in order to focus on analysis of the residual error (6).

B. Normalized MSE (NMSE) Performance of the $\ell 1$ MB

Fig. 1(a) shows NMSE performance with random TSs in a 4×4 MIMO system with the parameters $(N_t, L) = (127, 50)$. The NMSE is defined by $\mathbb{E}[\|\hat{\mathcal{H}}^{MB}_{[\hat{w}]} - \mathcal{H}\|^2]/\mathbb{E}[\|\mathcal{H}\|^2]$. The TSs are re-generated every burst timing so that $\bar{\mathbf{R}}_{\mathcal{X}\mathcal{X}}/N_t = \mathbf{I}_{WN_T}$ holds. As shown in Fig. 1(a), the NMSE with the $\ell 2$ MB is 8 dB away from the performance bound, normalized CRB (NCRB), given by $NCRB(\sigma_z^2) = N_R\sigma_z^2 \sum_{k=1}^{N_T} r_k/(N_t\mathbb{E}[\|\mathcal{H}\|^2])$. This is because the whitening ratio with the random TSs becomes $\omega(W) = 6.4 \gg 1$ and, thereby, the $\ell 2$ MB suffers from the noise enhancement in (8). As observed from Fig. 1(a), the NMSE with the $\ell 1$ MB can be improved significantly over that of the $\ell 2$ MB. The reason for the improvement is detailed in Section IV-C.

It should be noticed that the noise whitening problem can be avoided by using a fixed TS pattern so that $\mathbb{E}_l^L[\mathbf{R}_{\mathcal{X}\mathcal{X}}^{-1}(l)] = \bar{\mathbf{R}}_{\mathcal{X}\mathcal{X}}^{-1}$. However, the NMSE with the $\ell 2$ MB is not improved significantly from that shown in Fig. 1(a) due to $\bar{\mathbf{R}}_{\mathcal{X}\mathcal{X}[W]}/N_t = \mathbf{R}_{\mathcal{X}\mathcal{X}[W]}(l)/N_t \neq \mathbf{I}_{WN_T}$ for a fixed random TS. After all, ideally uncorrelated TSs are needed to essentially solve the noise whitening problem. Fig. 1(b) shows the NMSE performance using the pseudo noise (PN) sequences. As observed from Fig. 1(b), both the $\ell 1$ MB and $\ell 2$ MB channel estimation techniques achieve the NCRB asymptotically, if the PN sequences are selected so that the cross-correlations between TX streams are ideally low.[1]

In a large-scale MIMO system, nevertheless, finding the optimal sequence combinations is an NP hard problem. The Gold sequence [6] is known as one of the most promising solutions to the problem, although it can be inferior to the ideally chosen PN sequence. It is worth noting that, as shown in Fig. 1(c), the NMSE improvement of the $\ell 1$ MB over the $\ell 2$ MB technique becomes significant in a large-scale 16×16 MIMO system, where $N_t = 511$ is assumed.

C. Error Analysis

Figs. 2 show the NMSE performance in the 4×4 MIMO system for possible CIR lengths w, $r_k < w \leq W$, where the random and PN TSs are used in Figs. 2(a) and (b), respectively. The received SNR is set at 15 dB. As observed from Figs. 2, $\bar{\delta}(w) = \bar{\delta}^{\perp}(w) + \bar{\delta}_{\mathcal{Z}}(w) + \bar{\delta}_{\Pi}(w)$ is satisfied, according to Theorem 1, where we define $\bar{\delta}(w) = \sum_{k=1}^{N_T}\mathbb{E}[\|\Delta\hat{\mathbf{H}}_{[w]k}^{MB}\|^2]/\mathbb{E}\left[\|\mathcal{H}\|^2\right]$. $\bar{\delta}^{\perp}(w)$, $\bar{\delta}_{\mathcal{Z}}(w)$ and $\bar{\delta}_{\Pi}(w)$ are defined similarly corresponding to the variances of (5), (6) and (7), respectively.

In the case the random TSs are used, as shown in Fig. 2(a), the $\ell 1$ MB can improve the NMSE of channel estimates significantly by selecting the CIR length as $\arg\min_w\{\bar{\delta}^{\perp}(w) \ll \bar{\delta}(w)\}$. In the case the TSs are generated with the PN sequences, however, the improvement by the CIR length constraint is very slight as shown in Fig. 2(b). This is because the whitening ratio becomes $\omega(w) = 1$ for $\forall w$ when the TSs are ideally uncorrelated sequences.

It should be emphasized that the NMSE of channel estimates is dominated by $\bar{\delta}_{\mathcal{Z}}(w)$ in the CIR length range $\{w \mid \bar{\delta}^{\perp}(w) \ll \bar{\delta}(w)\}$. Furthermore, in that CIR length range, the NMSE $\bar{\delta}_{\mathcal{Z}}(w)$ follows the analytical curve given by (8). In other words, the NMSE performance of the $\ell 1$ MB algorithm can be described via the whitening ratio (9). The next subsection shows, therefore, asymptotic property of the whitening ratio for system setups assuming very long training lengths and/or massive TX streams.

[1]The MB techniques with the PN sequences do not always achieve the NCRB if the sequence combination is not correctly chosen.

D. Asymptotic Property of the Whitening Ratio

Fig. 3(a) illustrates asymptotic property of the whitening ratio for the length N_t of random TSs. The maximum CIR length W and the number of TX streams N_T are fixed at 31 and 4, respectively. As observed from Fig. 3(a), the whitening ratio becomes much grater than 1 for a short training length $N_t = WN_T$. However, the whitening ratio can be decreased significantly by the CIR length constraint. This is because, as discussed in [3], the following holds by [7, Theorem 7.7.8]:

$$\exists w \leq W,\ \operatorname{tr}\{\mathbf{R}_{\mathcal{X}\mathcal{X}[w]}^{-1}(l)\}/w \leq \operatorname{tr}\{\mathbf{R}_{\mathcal{X}\mathcal{X}[W]}^{-1}(l)\}/W. \quad (10)$$

When long TSs are utilized, nevertheless, the $\ell 1$ MB cannot improve NMSE performance over the $\ell 2$ MB algorithm since $\omega(w) \approx 1$ for any CIR length constraint $\forall w$.

Fig. 3(b) depicts the whitening ratio (9) for massive numbers of the TX streams. The training length is set at $N_t = WN_T$ for the number N_T of TX streams, where the the maximum CIR length W is fixed at 31. As shown in Fig. 3(b), the whitening ratio is deteriorated as N_T increases. Therefore, the $\ell 1$ MB algorithm is expected to improve NMSE performance significantly in a massive MIMO system when ideally uncorrelated TSs are not used. In the case $N_T = 24$ for example, the $\ell 1$ MB has a possibility to achieve up to 14 dB of NMSE gain over the $\ell 2$ MB. However, in a SISO or SIMO system, the NMSE gain becomes at most 3 dB since $\omega(w) \leq 2$ for $\forall w \leq W$.

V. Conclusions

When the ideally uncorrelated TSs are not used, the subspace-based $\ell 2$ MB technique can suffer from the noise enhancement problem due to the inaccurate noise whitening process. The $\ell 1$ MB algorithm can, however, mitigate the problem according to the property (10), if the length w of the effective CIRs above the noise level is shorter than the maximum CIR length W assumed in the system.

Furthermore, this letter has discussed the asymptotic NMSE performance of the $\ell 1$ MB algorithm via the whitening ratio $\omega(w)$. The whitening ratio is deteriorated as the TS length decreases or the number of TX streams increases. The $\ell 1$ MB algorithm can, therefore, improve the NMSE performance over the conventional $\ell 2$ MB technique in a massive MIMO system when the TSs are not long enough and not ideally uncorrelated.

References

[1] M. Nicoli and U. Spagnolini, "Reduced-rank channel estimation for time-slotted mobile communication systems," *IEEE Trans. Signal Process.*, vol. 53, no. 3, pp. 926–944, 2005.

[2] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 8, no. 5, pp. 742–758, Oct 2014.

[3] Y. Takano, M. Juntti, and T. Matsumoto, "$\ell 1$ LS and $\ell 2$ MMSE-based hybrid channel estimation for intermittent wireless connections," *IEEE Trans. Wireless Commun.*, accepted for publication.

[4] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716 – 723, Dec. 1974.

[5] European Telecommunications Standards Institute (ETSI), "Spatial channel model for MIMO simulations (3GPP TR 25.996)," Sep. 2014.

[6] R. Gold, "Optimal binary sequences for spread spectrum multiplexing (corresp.)," *Information Theory, IEEE Transactions on*, vol. 13, no. 4, pp. 619–621, October 1967.

[7] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge university press, 2012.

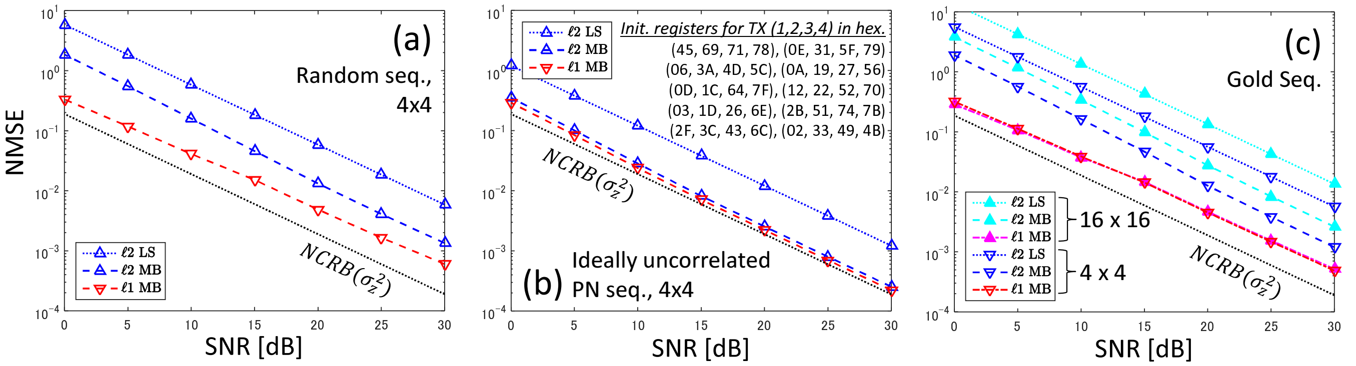

Fig. 1. The NMSE performance in the VA30 scenario by using the random (a), the PN (b) and the Gold (c) sequences. The subfigures (a) and (b) assume the 4×4 MIMO system, whereas the subfigure (c) shows comparison between the 4×4 and 16×16 MIMO systems. TS lengths are $N_t = 127$ and 511 for the 4×4 and 16×16 MIMO systems, respectively. The PN sequences with the generator polynomial $1 + x^6 + x^7$ are obtained by initializing the shift register with the least significant 7 bits of hexadecimal initial values shown in (b) so that the cross-correlations between TX streams are ideally low. The Gold sequences are generated by initializing the two shift registers with the indexes of the frame timing and the TX stream, where the generator polynomials are $\{1 + x^3 + x^7,\ 1 + x + x^2 + x^3 + x^7\}$ and $\{1 + x^4 + x^9,\ 1 + x^3 + x^4 + x^6 + x^9\}$ for $N_t = 127$ and 511, respectively.

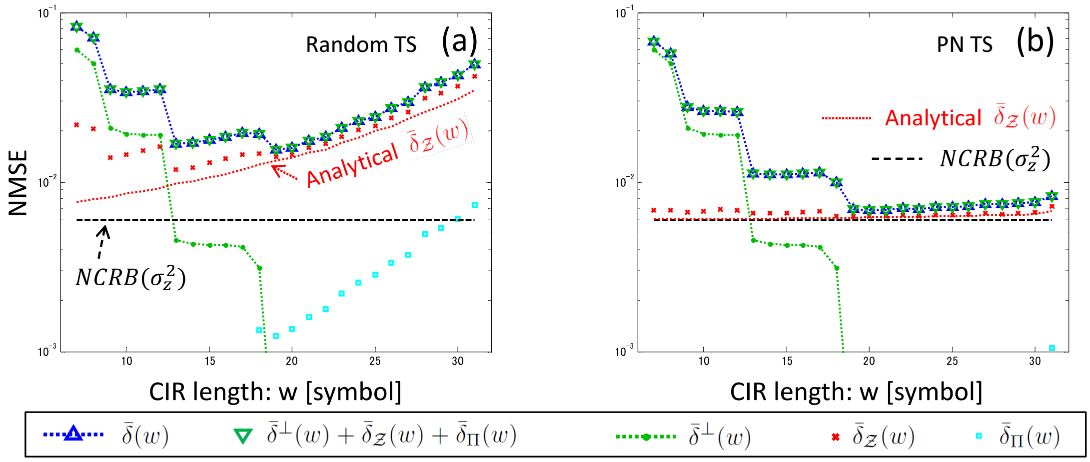

Fig. 2. The NMSE performance for possible CIR lengths w, where the random and ideally uncorrelated PN TSs are respectively used in (a) and (b). The 4×4 MIMO system and the VA30 scenario are assumed. $\bar{\delta}(w)$ denotes the NMSE of the channel estimate $\hat{\mathcal{H}}^{MB}_{[w]}$. $\bar{\delta}^{\perp}(w), \bar{\delta}_{\mathcal{Z}}(w)$ and $\bar{\delta}_{\Pi}(w)$ are normalized variances of (5), (6) and (7), respectively. The red dotted curve *Analytical* $\bar{\delta}_{\mathcal{Z}}(w)$ is the NMSE normalized (8) with $\mathbb{E}[\|\mathcal{H}\|^2]$.

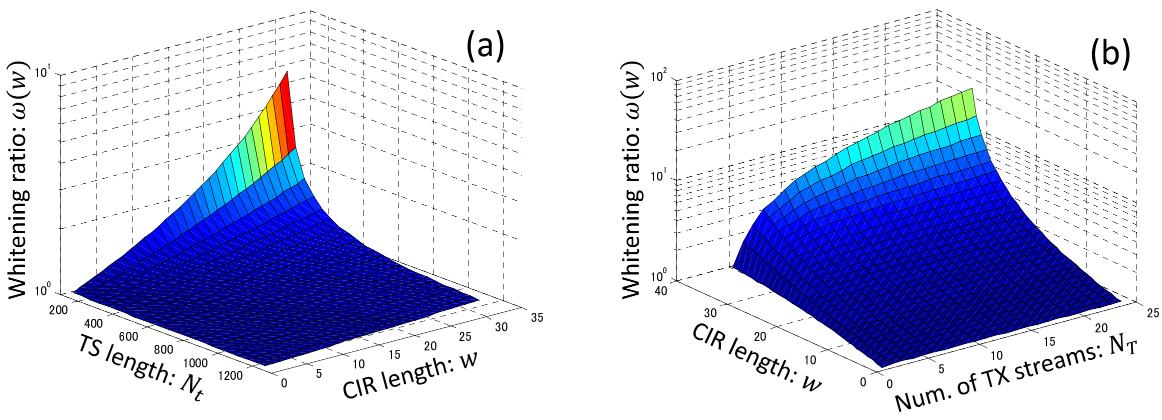

Fig. 3. The whitening ratio $\omega(w)$ (9) for the TS length(a) and the number of the TX streams(b). Random TSs are assumed.