A Spatial-Temporal Subspace-Based Compressive Channel Estimation Technique in Unknown Interference MIMO Channels

Takano, Yasuhiro
Su, Hsuan-Jung
Shiraishi, Yoshiaki
Morii, Masakatu

(Citation)
IEEE Transactions on Signal Processing,68:300-313

(Issue Date)
2020

(Resource Type)
journal article

(Version)
Accepted Manuscript

A Spatial–Temporal Subspace-based Compressive Channel Estimation Technique in Unknown Interference MIMO Channels

Yasuhiro Takano, *Member, IEEE,* Hsuan-Jung Su, *Senior Member, IEEE,* Yoshiaki Shiraishi, *Member, IEEE,* and Masakatu Morii, *Member, IEEE*

***Abstract*—Spatial–temporal (ST) subspace-based channel estimation techniques formulated with $\ell 2$ minimum mean square error (MMSE) criterion alleviate the multi-access interference (MAI) problem when the interested signals exhibit low-rank property. However, the conventional $\ell 2$ ST subspace-based methods suffer from mean squared error (MSE) deterioration in unknown interference channels, due to the difficulty to separate the interested signals from the channel covariance matrices (CCMs) contaminated with unknown interference. As a solution to the problem, we propose a new $\ell 1$ regularized ST channel estimation algorithm by applying the expectation-maximization (EM) algorithm to iteratively examine the signal subspace and the corresponding sparse-supports. The new algorithm updates the CCM independently of the slot-dependent $\ell 1$ regularization, which enables it to correctly perform the sparse-independent component analysis (ICA) with a reasonable complexity order. Simulation results shown in this paper verify that the proposed technique significantly improves MSE performance in unknown interference MIMO channels, and hence, solves the BER floor problems from which the conventional receivers suffer.**

***Index Terms*—Multi-access interference (MAI), unknown interference, subspace-based channel estimation, compressive sensing, principal component analysis (PCA), independent component analysis (ICA).**

I. Introduction

Interference alignment techniques (e.g., [1]) can improve throughput performance in multi-access interference (MAI) channels by exploiting spatial-degrees of freedom (DoF) of multiple-input multiple-output (MIMO) channels if channel state information (CSI) is known accurately. We can utilize $\ell 2$ spatial-temporal (ST) subspace-based channel estimation techniques [2] to obtain accurate enough CSIs. The ST subspace-based techniques alleviate the MAI problem by utilizing a property that the rank of the channel covariance matrix (CCM) [3] of the interested signals is less than the observed dimension. However, in practice, receivers have to estimate channels over unknown interference caused by hidden terminals [4], or artificial noise [5] for secure transmission, etc. The conventional $\ell 2$ techniques can seriously suffer from mean squared error (MSE) deterioration in such scenarios due to the difficulty to separate the interested signals from the CCM contaminated with unknown interference.

Y. Takano, Y. Shiraishi and M. Morii are with Kobe University, 1-1 Rokkodai-cho, Nada, Kobe, 657-8501, Japan (e-mail: {takano@eedept., zenmei@port., mmorii@}kobe-u.ac.jp).

H.-J. Su is with National Taiwan University, Taipei 10617, Taiwan (e-mail: hjs@ntu.edu.tw).

Compressed sensing (CS)-based algorithms can be utilized to improve the MAI problem if the interested parameters exhibit sparse nature. This is because they are executed in an observed sub-domain so that the $\ell 0$ or $\ell 1$ norm of the estimate is minimized, which consequently perform interference suppression as well. For example, spatio-temporal compressive channel estimation techniques [6], [7] are proposed for massive MIMO systems. However, they aim to improve multiple-input single-output (MISO) reception performance in downlink and do not assume the spatial DoF of multiple receive antennas.

Recently, MIMO channel estimation algorithms are extensively studied for millimeter-wave (mmWave) systems (e.g., [8]–[11]) by leveraging that channel parameters in narrowband transmission can be approximated to a sparse matrix of the angular domain representation [12]. As shown in [8], an orthogonal matching pursuit (OMP)-based technique outperforms the ordinary $\ell 2$ least squares (LS) technique in a mmWave MIMO system. However, the greedy OMP algorithm does not always achieve its analytical MSE performance since the MSE convergence performance depends significantly on the stopping criterion and the design of the dictionary including beamforming matrices.

A sparse Bayesian learning (SBL)-based algorithm [9] improves estimation accuracy over the OMP by exploiting the *block sparsity* property [9] commonly observed for the angular domain channel gain vectors in certain L_M measurements. Note that the estimation algorithms [8], [9] do not consider the MAI problem directly. The signal model in both studies is formulated as a collection of received single-input multi-output (SIMO) signals from each transmission (TX) beam. The formulation is reasonable in narrowband transmission, however, it increases pilot overheads in broadband transmission.

For frequency selective fading (FSF) MIMO channels, CS-based channel estimation algorithms are proposed in [10], [11]. Although they consider the estimation problem in MAI channels, the assumption of a long channel coherent time limits application scenarios such as frequency division duplex (FDD)-mmWave systems with a slow mobility and a short transmission time interval (TTI). Moreover, in [10], [11], the MAI problem is not explicitly considered since the uplink TX precoder is supposed to avoid interference between TX beams. However, in a multiuser uplink MIMO scenario, the interference alignment using TX beamforming only cannot perfectly eliminate the MAI [1]. Moreover, not all uplink terminals are capable of TX beamforming. As demonstrated in this paper, the algorithm [11] can exhibit an MSE floor in a transmission scenario of a carrier frequency of 5 GHz when TX beamforming is not assumed.

Given such a background, this paper aims to improve the MAI problem including unknown interference in FSF channels by ameliorating channel estimation accuracy without resorting to TX beamforming strategies. To this end, we study how a subspace-based rank reduction approach should be jointly utilized with a CS-based algorithm in an uplink time division duplex (TDD) MIMO receiver. Specifically, we propose a new $\ell 1$ regularized ST-minimum mean square error (MMSE) channel estimation which exactly performs the expectation-maximization (EM) algorithm in order to iteratively examine the sparse-subspace and the corresponding supports. Simulation results shown in this paper verify that a receiver using the proposed technique improves the MSE performance in unknown interference MIMO channels and, hence, solves a bit error rate (BER) floor problem from which the conventional receivers suffer.

Contributions of this paper are summarized as follows:

- ST subspace-based channel estimation algorithms using the independent component analysis (ICA) [13] are compared with that using the principal component analysis (PCA) [2]. A new $\ell 1$ regularized ICA estimation algorithm outperforms the conventional $\ell 2$ PCA approach in unknown interference MIMO channels.
- We show a novel CCM updating algorithm which can be performed independently of the $\ell 1$ regularization. Note that the sparse-supports, referred to as *active-set* [14], can be changed over slot-timings, which contradicts a requirement that the active-set has to be consistent in a certain duration to correctly perform the sparse-PCA (e.g., [15]) and/or ICA approaches.
- Performance analysis of the ICA-based ST channel estimation is detailed. Based on the analytical performance, we improve the previously-proposed adaptive active-set detection (AAD) technique [16], [17] for the unknown interference MIMO channels.

Note that the proposed $\ell 1$ ICA approach can improve MSE performance of any $\ell 1$ LS estimates. This paper shows, however, such a naive extension is not always the optimum.

This paper is organized as follows. Section II clarifies a MIMO transmission system assumed in this paper. Section IV presents the new channel estimation techniques. Section V shows analytical estimation performance of the proposed techniques. Section VI verifies the effectiveness of the new algorithms via computer simulations. Section VII shows concluding remarks.

Notations: 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 $\mathbf{X}^\mathsf{T}$ and $\mathbf{X}^\mathsf{H}$, 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 $\mathbf{X}^{\mathsf{H}/2}\mathbf{X}^{1/2}$. $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ are the ceiling and floor functions, respectively. Operators used in this paper are summarized in Table I.

TABLE I
OPERATORS

Operator	Definition
$\mathbf{X}\vert_{\mathcal{A}}$	Submatrix composed of the column vectors in $\mathbf{X}$ corresponding to the set $\mathcal{A}$, where $\mathcal{CAL}$-font is used for index sets. A set of consecutive numbers $\{i, \cdots, j\}$ is denoted by $\mathcal{A} = \{i : j\}$, where $\{i : j\} = \emptyset$ if $i > j$.
$\mathbf{J}_{\mathcal{A}}$	Factor matrix $\mathbf{J}_{\mathcal{A}} \stackrel{\mathrm{def}}{=} \mathbf{I}\vert_{\mathcal{A}}$ denotes a compressed/sparse matrix, where $\mathbf{I}$ is an identify matrix. e.g., $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5]$, $\mathcal{A} = \{1, 3\} \Rightarrow \mathbf{X}\mathbf{J}_{\mathcal{A}} \cdot \mathbf{J}_{\mathcal{A}}^\mathsf{T} = [\mathbf{x}_1, \mathbf{x}_3] \cdot \mathbf{J}_{\mathcal{A}}^\mathsf{T} = [\mathbf{x}_1, \mathbf{0}, \mathbf{x}_3, \mathbf{0}, \mathbf{0}]$.
$\mathtt{vec}(\mathbf{X})$	$MN \times 1$ vector composed by stacking the columns of $\mathbf{X} \in \mathbb{C}^{M \times N}$.
$\mathtt{mat}_N(\mathbf{x})$	Inversion of the vectorization: $\mathtt{mat}_N\{\mathtt{vec}(\mathbf{X})\} = \mathbf{X}$.
$\mathtt{diag}(\mathbf{X})$	Vector composed of the diagonal elements of $\mathbf{X}$.
$\mathtt{D_{IAG}}(\mathbf{x})$	Diagonal matrix formed with the vector $\mathbf{x}$.
$\mathtt{tplz}_M\{\mathbf{r}\}$	$M \times N$ Toeplitz matrix whose first row is length N vector $\mathbf{r}$.
$\Vert\mathbf{X}\Vert^2_{\mathbf{A}\times\mathbf{B}}$	Weighted matrix Frobenius norm: $\mathtt{tr}\{\mathbf{B}\mathbf{X}\mathbf{A}\mathbf{X}^\mathsf{H}\}$ for $\mathbf{X} \in \mathbb{C}^{M \times N}$ with positive definite matrices $\mathbf{A}$ and $\mathbf{B}$. Moreover, $\Vert\mathbf{X}\Vert^2_{\mathbf{A}} = \Vert\mathbf{X}\Vert^2_{\mathbf{A}\times\mathbf{I_M}}$ and $\Vert\mathbf{X}\Vert^2 = \Vert\mathbf{X}\Vert^2_{\mathbf{I_N}\times\mathbf{I_M}}$, where $\mathbf{I_M}$ is the $M \times M$ identity matrix.
$\Vert\mathbf{X}\Vert_1$	Matrix $\ell 1$ norm: $\sum_{i=1}^M \sum_{j=1}^N \vert x_{ij}\vert$ where x_{ij} is the (i, j)-th entry of $\mathbf{X}$.
$\mathbb{E}^{L,\Delta}_{j=l}[\mathbf{X}(j)]$	Average of matrix sequence $\mathbf{X}(j)$ sampled with an interval Δ: $\frac{1}{L}\sum_{j=\Delta(l-L)+1}^{l} \mathbf{X}(j)$ for the past L observations from the timing l. For $\Delta = 1$, we denote $\mathbb{E}^{L}_{j=l}[\mathbf{X}(j)]$. Moreover, $\mathbb{E}[\mathbf{X}(j)] = \mathbb{E}^{\infty}_{j=l}[\mathbf{X}(j)]$.
$\mathbb{K}[\mathbf{X}(j)]$	Column-wise covariance matrix: $\mathbb{E}[\mathbf{X}^\mathsf{H}(j)\mathbf{X}(j)]$.
$\mathbb{R}[\mathbf{X}(j)]$	Row-wise covariance matrix: $\mathbb{E}[\mathbf{X}(j)\mathbf{X}^\mathsf{H}(j)] = \mathbb{K}[\mathbf{X}^\mathsf{H}]$.
$\mathbb{P}(\mathbf{X})$	Projection matrix $\mathbf{X}\mathbf{X}^\dagger$ of $\mathbf{X}$.
$\mathbb{I}(\mathbf{X})$	The indicator function of $\mathbf{X}$.

II. System Model

A. Received Signal

Consider channel estimation in a TDD-MIMO system composed of N_T transmit- and N_R receive-antennas. L_t-symbol training sequence (TS) $\mathbf{x}_k, \forall k \in \{1 : N_T\}$, is transmitted over FSF channels whose channel impulse response (CIR) lengths are at most W symbols. $(W-1)$-symbol guard interval (GI) is added at the front and rear of TS in order to avoid inter-block interference in the received TS signals of $\tilde{L}_t = L_t + W - 1$ symbols. Received TS matrix $\mathbf{Y}(l) \in \mathbb{C}^{N_R \times \tilde{L}_t}$ is written as,

$$\mathbf{Y}(l) = \mathbf{H}(l)\mathbf{X} + \mathbf{N}(l) + \mathbf{Z}(l), \tag{1}$$

where the timing index l is a multiple of a slot interval Δ_T. Channel, TS, additive white Gaussian noise (AWGN) and unknown interference matrices are

$$\begin{aligned} \mathbf{H}(l) &= [\mathbf{H}_1(l), \cdots, \mathbf{H}_{N_T}(l)] &\in \mathbb{C}^{N_R \times W N_T}, \\ \mathbf{X} &= [\mathbf{X}_1^\mathsf{T}, \cdots, \mathbf{X}_{N_T}^\mathsf{T}]^\mathsf{T} &\in \mathbb{C}^{W N_T \times \tilde{L}_t}, \\ \mathbf{N}(l) &= [\mathbf{n}_1(l), \cdots, \mathbf{n}_{N_R}(l)]^\mathsf{T} &\in \mathbb{C}^{N_R \times \tilde{L}_t}, \\ \mathbf{Z}(l) &= [\mathbf{z}_1(l), \cdots, \mathbf{z}_{N_R}(l)]^\mathsf{T} &\in \mathbb{C}^{N_R \times \tilde{L}_t}, \end{aligned}$$

respectively. The noise vector at the n-th receive (Rx) antenna $\mathbf{n}_n(l)$ follows the Complex normal distribution $\mathcal{CN}(\mathbf{0}, \sigma^2_\mathbf{n}\mathbf{I}_{\tilde{L}_t})$ and has the spatially uncorrelated property: $\mathbb{E}[\mathbf{n}_i^\mathsf{H}\mathbf{n}_j]/\tilde{L}_t = 0$ for $i \neq j$. However, the unknown interference vector $\mathbf{z}_n(l)$ does not always have the spatially uncorrelated property, since the interference caused by leakage signals from specific hidden terminals is observed via Rx array antennas [18]. Hence, we assume zero-mean and temporally-white properties only: $\mathbb{E}[\mathbf{z}_n(l)] = \mathbf{0}$ and $\frac{1}{N_R}\sum_{n=1}^{N_R}\mathbb{E}[\mathbf{z}_n(l)\mathbf{z}_n^{\mathsf{H}}(l)] = \sigma_{\mathbf{z}}^2\mathbf{I}_{\tilde{L}_t}$, respectively. The variances $\sigma_{\mathbf{n}}^2$ and $\sigma_{\mathbf{z}}^2$ per receive antenna are determined according to signal-to-noise ratio (SNR) and signal-to-interference ratio (SIR), respectively.

The TS submatrix $\mathbf{X}_k$ is defined by $\texttt{tplz}_W\left\{[\mathbf{x}_k^{\mathsf{T}}, \mathbf{0}_{W-1}^{\mathsf{T}}]\right\}$, where the operation $\texttt{tplz}_W\{\mathbf{r}\}$ constructs a $W \times L$ Toeplitz matrix whose first row vector is $\mathbf{r} \in \mathbb{C}^{1\times L}$.

B. Channel Model

Under the wide-sense stationary uncorrelated scattering (WSSUS) assumption, let the channel coherent time be greater than a slot duration $L_{\rm slot}T_{\rm sym}$ [sec], where $L_{\rm slot}$ and $T_{\rm sym}$ are the slot length in symbol and the symbol interval in second, respectively. The (n, j)-th entry of the CIR matrix $\mathbf{H}_k(l)$ is composed of r_k resolvable paths, and it is observed via pulse shaping filters $p[t]$, as $h^{(k,n)}[t] = p[t] \star \left\{\sum_{r=1}^{r_k} b_r^{(k,n)}[t]\,\delta[t-\tau_r^{(k,n)}]\right\}$ at timing $t = (j + lL_{\rm slot})T_{\rm sym}$ [sec], where $b_r^{(k,n)}[t]$ and $\tau_r^{(k,n)}$ are the complex gain and path delay of the r-th path [19]–[21]. The convolution operator and the Dirac delta function are denoted by $\star$ and $\delta[t]$, respectively.

As discussed in [2], [16], the k-th CIR submatrix $\mathbf{H}_k(l)$ between the k-th transmit- and N_R receive-antennas can be written as

$$\mathbf{H}_k(l) = \mathbf{B}_k(l)\mathbf{E}_k^{\mathsf{H}}, \tag{2}$$

where the matrix $\mathbf{B}_k(l) \in \mathbb{C}^{N_R\times r_k}$ describes slot-dependent complex gains of r_k-resolvable paths. However, the eigenvectors $\mathbf{E}_k \in \mathbb{C}^{W\times r_k}$ are independent of slot-timing since they can be seen as a time-invariant finite impulse response (FIR) filter representing the response of pulse shaping filters and multipath channels. The gain matrix $\mathbf{B}_k(l) = [\mathbf{b}_{k,1}(l), \cdots, \mathbf{b}_{k,r}(l)]$ is generated from a scattering cluster-based model, such as the spatial channel model (SCM) [22]. The i-th path vector $\mathbf{b}_{k,i}(l)$ can, hence, be written in the angular domain representation [12]:

$$\mathbf{b}_{k,i}(l) = \mathbf{U}_{k,i} \cdot \mathbf{a}_{k,i}(l) \cdot \mathrm{u}_{k,i}, \tag{3}$$

where the unitary matrices $\mathbf{U}_{k,i} \in \mathbb{C}^{N_R\times N_R}$ and $\mathrm{u}_{k,i} \in \mathbb{C}^{1\times 1}$ denote characteristics of propagation channels from a reflector to receive-, and to transmit-antennas, respectively. The vector $\mathbf{a}_{k,i}(l)$ follows a Rayleigh distribution. Note that the receiver has no prior-knowledge about the hyper parameters $\{r_k, \mathbf{E}_k, \mathbf{U}_{k,i}, \mathrm{u}_{k,i}\}$, except that the expected variance of $\mathbf{H}_k(l)$ is $\mathbb{E}[\|\mathbf{H}_k(l)\|^2] = \sigma_{\mathbf{H}}^2$ with a constant $\sigma_{\mathbf{H}}^2$.

III. Preliminaries

A. Channel Covariance Matrices

Let us compute CCMs of (2) and (3). By performing the singular value decomposition (SVD), we have

$$\begin{aligned}
\mathbf{V}_k\mathbf{D}_{\mathrm{T},k}\mathbf{V}_k^{\mathsf{H}} &= \mathbb{K}[\mathbf{H}_k(j)], &(4)\\
\mathbf{U}_{\mathrm{P},k}\mathbf{D}_{\mathrm{P},k}\mathbf{U}_{\mathrm{P},k}^{\mathsf{H}} &= \mathbb{R}[\mathbf{H}_k(j)], &(5)\\
\mathbf{U}_{k,i}\mathbf{D}_{\mathrm{I},k,i}\mathbf{U}_{k,i}^{\mathsf{H}} &= \mathbb{R}[\mathbf{b}_{k,i}(j)], &(6)
\end{aligned}$$

where $\mathbf{V}_k \in \mathbb{C}^{W\times W}$, $\mathbf{U}_{\mathrm{P},k} \in \mathbb{C}^{N_R\times N_R}$ and $\mathbf{U}_{k,i} \in \mathbb{C}^{N_R\times N_R}$ are unitary matrices. We denote column-wise and row-wise covariance matrices for a sequence $\mathbf{X}(j)$ by $\mathbb{K}[\mathbf{X}(j)] = \mathbb{E}[\mathbf{X}^{\mathsf{H}}(j)\mathbf{X}(j)]$ and $\mathbb{R}[\mathbf{X}(j)] = \mathbb{E}[\mathbf{X}(j)\mathbf{X}^{\mathsf{H}}(j)]$, respectively. Diagonal matrices $\mathbf{D}_{\mathrm{T},k}$, $\mathbf{D}_{\mathrm{P},k}$ and $\mathbf{D}_{\mathrm{I},k,i}$ are constructed from singular values of the corresponding covariance matrices, respectively. We refer to (4) as *temporal* CCM. The spatial versions (5) and (6) are referred to as *principal (p-)spatial*, and *independent (i-)spatial* CCMs, respectively, by following concept of the ICA [13]. Moreover, joint subspaces reduced from $\{(4), (5)\}$ and $\{(4), (6)\}$ are, respectively, referred to as pST and iST subspaces, hereafter.

B. Channel Rank Properties

The reduced-rank channel estimation techniques (e.g., [2]) utilizes a property that the rank of the interested signals in the noisy CCMs is less than the original dimension. We define *adaptive-ranks* for the CCMs by generalizing that shown in [17], where the following notations are used. For an $N \times M$ parameter matrix $\mathbf{A}$, let $\hat{\mathbf{A}}$ denote a noisy observation: $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{N}_{\mathbf{A}}$, where the matrix $\mathbf{N}_{\mathbf{A}}$ has properties: $\frac{1}{M}\mathbb{K}[\mathbf{N}_{\mathbf{A}}] = \sigma_{\hat{\mathbf{A}}}^2\mathbf{I}_N$ and $\frac{1}{N}\mathbb{R}[\mathbf{N}_{\mathbf{A}}] = \sigma_{\hat{\mathbf{A}}}^2\mathbf{I}_M$ with a constant $\sigma_{\hat{\mathbf{A}}}^2$.

Definition 1 (Adaptive-ranks)**.** *In the temporal CCM of $\hat{\mathbf{A}}$, the dimension of the interested signals above the noise level $\sigma_{\hat{\mathbf{A}}}^2$ can be approximated in a subspace of the dimension given by the temporal adaptive-rank:*

$$\texttt{ar}(\mathbb{K}[\hat{\mathbf{A}}], \sigma_{\hat{\mathbf{A}}}^2) = \textstyle\sum_{\forall n}\mathbb{I}\left\{\lambda_n^2(\mathbb{K}[\hat{\mathbf{A}}]) \geq \gamma\sigma_{\hat{\mathbf{A}}}^2\right\}, \tag{7}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function and $\lambda_n^2(\mathbf{M})$ denotes the n-th singular value of matrix $\mathbf{M}$. Moreover, $\gamma \overset{\text{def}}{=} \texttt{min}\left\{\texttt{ar}(\mathbb{K}[\hat{\mathbf{A}}], \sigma_{\hat{\mathbf{A}}}^2), \texttt{ar}(\mathbb{R}[\hat{\mathbf{A}}], \sigma_{\hat{\mathbf{A}}}^2)\right\}$, where the spatial adaptive-rank is given by

$$\texttt{ar}(\mathbb{R}[\hat{\mathbf{A}}], \sigma_{\hat{\mathbf{A}}}^2) = \texttt{ar}(\mathbb{K}[\hat{\mathbf{A}}^{\mathsf{H}}], \sigma_{\hat{\mathbf{A}}}^2). \tag{8}$$

The parameter γ is, similar to [12, (7.74)], the rank of the signal matrix in $\hat{\mathbf{A}}$ corresponding to the joint spatial-temporal subspace. The adaptive-ranks are determined by iteratively examining (7) and (8), where we initialize $\gamma = \texttt{min}(M, N)$.

Let us see adaptive-ranks for CCMs of CIR estimates. Suppose that the receiver estimates the CIR as $\hat{\mathbf{H}}_k(j) = \mathbf{H}_k(j) + \mathbf{N}_{\mathbf{H}_k}$ using a length L_t TS, after relevant noise whitening transformations for the received signals such that $\texttt{vec}\{\mathbf{N}_{\mathbf{H}_k}\} \sim \mathcal{CN}(\mathbf{0}, \frac{\sigma_{\mathfrak{N}}^2}{L_t}\mathbf{I}_W \otimes \mathbf{I}_{N_R})$. The operator $\otimes$ is Kronecker product. The variance $\sigma_{\mathfrak{N}}^2 \overset{\text{def}}{=} \sigma_{\mathbf{n}}^2 + \sigma_{\mathbf{z}}^2$ is given according to the received signal-to-interference-puls-noise ratio (SINR). We define adaptive-ranks corresponding to (4), (5) and (6), as $r_{\mathrm{T},k} = \texttt{ar}\left(\mathbb{K}[\hat{\mathbf{H}}_k(j)], \frac{\sigma_{\mathfrak{N}}^2}{L_t}\right)$, $r_{\mathrm{P},k} = \texttt{ar}\left(\mathbb{R}[\hat{\mathbf{H}}_k(j)], \frac{\sigma_{\mathfrak{N}}^2}{L_t}\right)$, and $r_{\mathrm{I},k,i} = \texttt{ar}\left(\mathbb{R}[\hat{\mathbf{b}}_{k,i}(j)], \frac{\sigma_{\mathfrak{N}}^2}{L_t}\right)$, respectively.

Property 1 (Bounds of the adaptive-ranks)**.**

$$\begin{aligned}
r_{\mathrm{T},k} &\leq W, &(9)\\
r_{\mathrm{P},k} &\leq N_R &(10)\\
\bar{r}_{\mathrm{I},k} \overset{\text{def}}{=} \tfrac{1}{r_{\mathrm{T}}}\textstyle\sum_{i=1}^{r_{\mathrm{T}}} r_{\mathrm{I},k,i} &\leq r_{\mathrm{P},k}. &(11)
\end{aligned}$$

Proof. We prove (11) only, since (9) and (10) are obvious. Let us omit the indexes k of TX streams and j of slot timings for the sake of simplicity. Moreover, we may assume $r_{\mathrm{P}} \geq r_{\mathrm{T}}$ in spatially dense large-scale MIMO channels. By [23, Corollary 3.4.3], we have $\sum_{n=1}^{N}\sum_{i=1}^{r_{\mathrm{T}}}\lambda_n^2(\mathbb{R}[\hat{\mathbf{b}}_i]) \geq \sum_{n=1}^{N}\lambda_n^2(\sum_{i=1}^{r_{\mathrm{T}}}\mathbb{R}[\hat{\mathbf{b}}_i])$ for $N \leq N_R$. Taking average for r_{T} paths yields $\mathbb{E}\left[\sum_{n=1}^{N}\lambda_n^2(\mathbb{R}[\hat{\mathbf{b}}_i])\right] \geq \frac{1}{r_{\mathrm{T}}}\sum_{n=1}^{N}\lambda_n^2(\mathbb{R}[\hat{\mathbf{H}}])$, since $\mathbb{R}[\hat{\mathbf{H}}] = \mathbb{R}[\hat{\mathbf{B}}] = \sum_{i=1}^{r_{\mathrm{T}}}\mathbb{R}[\hat{\mathbf{b}}_i]$ according to (2) and (3). Because the equality always holds for $N = N_R$, we find that the variance of vector $\{\lambda_n^2(\mathbb{R}[\hat{\mathbf{b}}_i]) \mid \forall n \leq N\}$ is not less than that of $\{\lambda_n^2(\mathbb{R}[\hat{\mathbf{H}}]) \mid \forall n \leq N\}$. Hence, for the threshold $\sigma_{\mathfrak{N}}^2/L_t$, $\bar{r}_{\mathrm{I},i} = \frac{1}{r_{\mathrm{T}}}\sum_{i=1}^{r_{\mathrm{T}}}\sum_{n=1}^{N}\mathbb{I}\left\{\lambda_n^2(\mathbb{R}[\hat{\mathbf{b}}_i]) \geq \frac{\sigma_{\mathfrak{N}}^2}{L_t}\right\} \leq \sum_{n=1}^{N}\mathbb{I}\left\{\lambda_n^2(\frac{1}{r_{\mathrm{T}}}\mathbb{R}[\hat{\mathbf{H}}]) \geq \frac{\sigma_{\mathfrak{N}}^2}{L_t}\right\} = r_{\mathrm{P}}$. □

Proposition 1 (Ranks of joint ST subspaces)**.** *The dimension of the iST subspace is less than that of the pST subspace.*

Proof. $\sum_{i=1}^{r_{\mathrm{T},k}} r_{\mathrm{I},k,i} = r_{\mathrm{T},k}\bar{r}_{\mathrm{I},k} \leq r_{\mathrm{T},k}r_{\mathrm{P},k}$ for $\forall i$. □

C. Examples

Fig. 1 illustrates Proposition 1, where CIR matrices $\mathbf{H}_k$ follow the Pedestrian-B (PB) model [22] assuming SIMO channels with $N_R = 12$ antennas. As observed from Fig. 1(a), the delay profile $\mathtt{diag}\{\mathbb{K}[\mathbf{H}_k]\}$ spreads over $W = 31$ symbols. However, we confirm from Fig. 1(b) that the temporal adaptive-rank is $r_{\mathrm{T},k} = 5 \leq W$ for a noise level $\frac{\sigma_{\mathfrak{N}}^2}{L_t} = 10^{-3}$.

Fig. 1(c) shows spatial profile vectors $\mathtt{diag}\{\mathbb{R}[\mathbf{H}_k]\}$ and $\mathtt{diag}\{\mathbb{R}[\mathbf{b}_{k,i}]\}$. Unlike the delay profile, the spatial profiles exhibit *dense* nature in the observed domain. Indeed, as shown in Fig. 1(d), the p-spatial singular values $\lambda_{\mathrm{PCA}}^2 = \{\lambda_i^2(\mathbb{R}[\mathbf{H}_k]) \mid i \leq N_R\}$ are supported in almost N_R dimensions, which means that, in the SCM-based channels, the conventional spatial-rank reduction techniques (e.g., [2]) cannot always obtain significant improvement in a high SINR regime.

Nevertheless, we aim to estimate CIRs in unknown interference channels. As observed from Fig. 1(d), the number of significant singular values λ_{PCA}^2 is $r_{\mathrm{P},k} = 6$ for the noise level[1] $\frac{\sigma_{\mathfrak{N}}^2}{L_t} = 10^{-3}$. However, in the iST subspace, the singular values $\lambda_{\mathrm{ICA},i}^2 \in \{\lambda_i^2(\mathbb{R}[\mathbf{b}_{k,i}]) \mid \forall i\}$ above the noise level are counted as $\sum_{i=1}^{r_{\mathrm{T},k}} r_{\mathrm{I},k,i} = 22 \leq r_{\mathrm{T},k}r_{\mathrm{P},k} = 30$.

IV. Channel Estimation

A. Problem Formulation

An $\ell 1$ regularized ST-MMSE channel estimation problem for TDD reception with slot-interval Δ_{T} is written, as

$$\hat{\mathbf{H}}_{\mathrm{ST}}^{\ell 1}(l) = \arg\min_{\mathbf{H}(l)} \mathbb{E}_{j=l}^{L_{\mathrm{S}},\Delta_{\mathrm{T}}}\left[\mathcal{L}(j, \mathbf{H}(j)) + \zeta\|\mathbf{H}(j)\|_1\right] \quad (12)$$

with a Lagrange multiplier ζ [24], where $\|\cdot\|_1$ is the matrix $\ell 1$ norm. We denote the expectation operation as $\mathbb{E}_{j=l}^{L,\Delta}[s(j)] = \frac{1}{L}\sum_{j=\Delta(l-L)+1}^{l} s(j)$ for a length-L sequence $s(j)$ sampled with an interval Δ. The log-likelihood function $\mathcal{L}(\cdot)$ is

$$\mathcal{L}(j, \mathbf{H}) = \frac{1}{\sigma_{\mathfrak{N}}^2}\|\mathbf{Y}(j) - \mathbf{H}\mathbf{X}\|_{\mathbf{I}_{\bar{L}_t}\times\mathbf{\Gamma}}^2. \quad (13)$$

[1]The p-spatial subspace has a threshold $r_{\mathrm{T},k}\frac{\sigma_{\mathfrak{N}}^2}{L_t} = 6 \times 10^{-3}$ according to Definition 1. However, it is is set at $1 \cdot \frac{\sigma_{\mathfrak{N}}^2}{L_t}$ in the i-spatial subspace.

The spatial weight matrix $\mathbf{\Gamma}$ is defined as $\mathbf{\Gamma} = \sigma_{\mathfrak{N}}^2\mathbf{R}_{\mathfrak{N}}^{-1}$, where $\mathbf{R}_{\mathfrak{N}} \overset{\mathrm{def}}{=} \frac{1}{\bar{L}_t}\mathbb{R}\left[\mathfrak{N}(j)\right]$ with $\mathfrak{N}(j) = \mathbf{N}(j) + \mathbf{Z}(j)$. We may use

$$\mathbf{R}_{\mathfrak{N}} \approx (1-\alpha)\sigma_{\mathbf{n}}^2\mathbf{I}_{N_R} + \frac{\alpha}{\bar{L}_t}\mathbb{R}_{j=l-1}^{L_{\mathrm{S}},\Delta_{\mathrm{T}}}[\mathbf{Y}(j) - \hat{\mathbf{H}}(j)\mathbf{X}], \quad (14)$$

where the parameter α is given by $\sigma_{\mathbf{z}}^2/\mathtt{tr}\{\frac{1}{\bar{L}_t}\mathbb{R}_{j=l-1}^{L_{\mathrm{S}},\Delta_{\mathrm{T}}}[\mathbf{Y}(j) - \hat{\mathbf{H}}(j)\mathbf{X}]\}$, so that the SINR of $\mathbf{R}_{\mathfrak{N}}$ is consistent to $\sigma_{\mathfrak{N}}^2$.

The problem (12) aims to find a temporally sparse solution $\hat{\mathbf{H}}_{\mathcal{A}}(l)$ supported with column indexes $\mathcal{A}$ referred to as *active-set*. The problem (12) can, hence, be reformulated as an EM problem composed of the following E- and M-steps:

$$\begin{cases} \hat{\mathbf{G}}_{\mathcal{A}^{[n]}}(l) = \\ \quad \arg\min_{\mathbf{G}_{\mathcal{A}^{[n]}}} \mathbb{E}_{j=l}^{L_{\mathrm{S}},\Delta_{\mathrm{T}}}\left[\mathcal{L}\left(j, \mathbf{G}_{\mathcal{A}^{[n]}}(j)\mathbf{J}_{\mathcal{A}^{[n]}}^{\mathsf{T}} \mid \mathcal{A}^{[n]}\right)\right] & (15) \\ \mathcal{A}^{[n+1]} = \\ \quad \arg\min_{\mathcal{A}^{[n+1]}\subseteq\mathcal{A}^{[n]}} \mathbb{E}_{j=l}^{L_{\mathrm{S}},\Delta_{\mathrm{T}}}\left[\|\hat{\mathbf{G}}_{\mathcal{A}^{[n]}}(j)\mathbf{J}_{\mathcal{A}^{[n+1]}}^{\mathsf{T}} - \mathbf{H}(j)\|^2\right] & (16) \end{cases}$$

with $\mathbf{J}_{\mathcal{A}} \overset{\mathrm{def}}{=} \mathbf{I}_{WN_T}|_{\mathcal{A}}$, where $\mathbf{G}_{\mathcal{A}}(j) = \mathbf{H}(j)\mathbf{J}_{\mathcal{A}} \in \mathbb{C}^{N_R\times|\mathcal{A}|}$ is a column-shrunk CIR matrix supported with active-set $\mathcal{A}$. As depicted in Fig. 2, we iteratively perform the pair of sub-problems for predefined constant N_{AAD} times, and determine the optimal solution by using corrected-Akaike information criterion (AICc) [25], where the active-set is initialized at $\mathcal{A}^{[0]} = \{1 : WN_T\}$.

B. Solution to the Conditional MMSE Problem (15)

For the sake of simplicity, let us omit the slot interval Δ_{T}. The active-set $\mathcal{A}^{[n]}$ at the n-th EM iteration is abbreviated as $\mathcal{A}$ in Sections IV-B and IV-C. The problem (15) is rewritten [26] as, given active-set $\mathcal{A}$,

$$\hat{\mathbf{G}}_{\mathcal{A}}(l) = \arg\min_{\mathbf{G}_{\mathcal{A}}} \mathbb{E}_{j=l}^{L_{\mathrm{S}}}\left[\|\mathbf{G}_{\mathcal{A}}(j) - \hat{\mathbf{G}}_{\mathcal{A}}^{\mathrm{LS}}(j)\|_{\mathbf{R}_{\mathbf{XX}_{\mathcal{A}}}\times\mathbf{\Gamma}}^2\right] \quad (17)$$

under an assumption that CIRs unsupported with $\mathcal{A}$ are minor. The matrix $\hat{\mathbf{G}}_{\mathcal{A}}^{\mathrm{LS}}(j)$ denotes the conditional $\ell 1$ LS channel estimate given active-set $\mathcal{A}$ [17]:

$$\hat{\mathbf{G}}_{\mathcal{A}}^{\mathrm{LS}}(j) = \mathbf{R}_{\mathbf{YX}}(j)\mathbf{J}_{\mathcal{A}}\mathbf{R}_{\mathbf{XX}_{\mathcal{A}}}^{-1}, \quad (18)$$

where $\mathbf{R}_{\mathbf{YX}}(j) = \mathbf{Y}(j)\mathbf{X}^{\mathsf{H}}$ and $\mathbf{R}_{\mathbf{XX}_{\mathcal{A}}} = \mathbf{J}_{\mathcal{A}}^{\mathsf{T}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{J}_{\mathcal{A}}$. According to [27], we can solve the MMSE problem (17) by using the PCA. In order to perform the PCA accurately, first of all, we transform the problem, as

$$\hat{\tilde{\mathbf{G}}}_{\mathcal{A}}(l) = \arg\min_{\tilde{\mathbf{G}}_{\mathcal{A}}} \mathbb{E}_{j=l}^{L_{\mathrm{S}}}\left[\|\tilde{\mathbf{G}}_{\mathcal{A}}(j) - \hat{\tilde{\mathbf{G}}}_{\mathcal{A}}^{\mathrm{LS}}(j)\|^2\right] \quad (19)$$

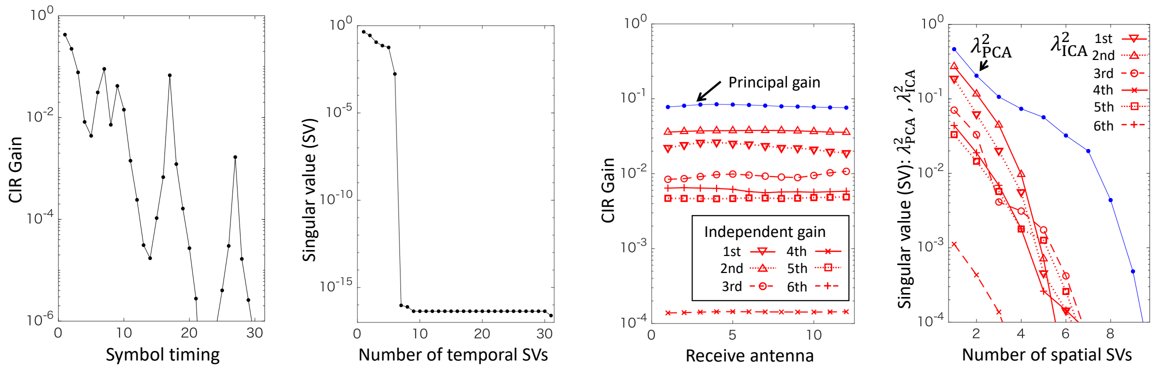
with $\tilde{\mathbf{G}}_{\mathcal{A}}(j) = \mathfrak{W}\{\mathbf{G}_{\mathcal{A}}(j)\}$ and

$$\hat{\tilde{\mathbf{G}}}_{\mathcal{A}}^{\mathrm{LS}}(j) = \mathfrak{W}\{\hat{\mathbf{G}}_{\mathcal{A}}^{\mathrm{LS}}(j)\}, \quad (20)$$

where the noise whitening operation $\mathfrak{W}\{\cdot\}$ is defined by

$$\mathfrak{W}\{\mathbf{M}\} = \mathbf{\Gamma}^{1/2}\left(\mathbf{M}\mathbf{R}_{\mathbf{XX}_{\mathcal{A}}}^{\mathsf{H}/2} - \mathbf{G}_{\mathcal{A}}(j)\nabla\mathbf{R}_{\mathbf{XX}_{\mathcal{A}}}^{\mathsf{H}/2}\right) \quad (21)$$

for matrix $\mathbf{M}$. The upper triangular matrices $\nabla\mathbf{R}_{\mathbf{XX}_{\mathcal{A}}}^{1/2}$ without diagonal blocks is given by $\mathbf{R}_{\mathbf{XX}_{\mathcal{A}}}^{1/2} - \bigoplus_{i=1}^{N_T}\mathbf{Q}_{\mathcal{A},i,i}$, where the



(a) Delay profile. (b) Temporal singular values. (c) Spatial profiles. (d) Spatial singular values.

Fig. 1. CIR analysis for PB realizations in 1×12 SIMO channels.

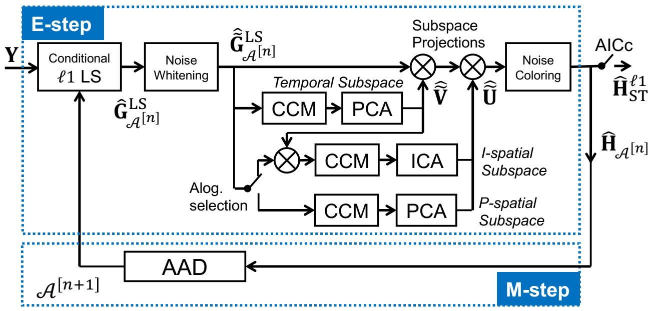


Fig. 2. Block diagram of the $\ell 1$ ST-MMSE channel estimation.

$|\mathcal{A}_i| \times |\mathcal{A}_j|$ matrix $\mathbf{Q}_{\mathcal{A},i,j}$ is the (i,j)-th block submatrix of

$$\mathbf{R}^{1/2}_{\mathbf{XX}_\mathcal{A}} = \begin{bmatrix} \mathbf{Q}_{\mathcal{A},1,1} & \cdots & \mathbf{Q}_{\mathcal{A},1,N_T} \\ & \ddots & \vdots \\ \mathbf{O} & & \mathbf{Q}_{\mathcal{A},N_T,N_T} \end{bmatrix}. \quad (22)$$

The operator $\bigoplus$ denotes the matrix direct sum. The index set $\mathcal{A}_i$ denotes an active-*subset* corresponding to the significant CIR taps in the i-th TX stream.

The solution to the original problem (17) can be obtained by performing inversion of (21) for (19). In the following, we show the iST subspace-based channel estimate after deriving the temporal subspace of (20). The pST solution is, then, illustrated by showing modifications from the iST estimate.

1) Temporal subspace: The $|\mathcal{A}_k| \times |\mathcal{A}_k|$ temporal subspace matrix $\hat{\hat{\mathbf{V}}}_{\mathcal{A}_k}$ is obtained by performing the SVD [27]:

$$\hat{\hat{\mathbf{V}}}_{\mathcal{A}_k} \hat{\hat{\mathbf{D}}}^{\mathcal{A}_k}_{\rm T} \hat{\hat{\mathbf{V}}}^{\rm H}_{\mathcal{A}_k} = \mathbb{K}^{L_{\rm S}}_{j=l} \left[\hat{\hat{\mathbf{G}}}^{\rm LS}_{\mathcal{A}_k}(j) \right], \quad (23)$$

where the submatrix $\hat{\hat{\mathbf{G}}}^{\rm LS}_{\mathcal{A}_k}(l) \overset{\rm def}{=} \hat{\hat{\mathbf{G}}}^{\rm LS}_{\mathcal{A}}(l)\,|_{\mathcal{A}_k}$ of (20) is

$$\mathbf{\Gamma}^{1/2} \left[\hat{\mathbf{G}}^{\rm LS}_{\mathcal{A}_k}(l) \mathbf{Q}^{\rm H}_{\mathcal{A},k,k} + \sum_{i=k+1}^{N_T} \{ \hat{\mathbf{G}}^{\rm LS}_{\mathcal{A}_i}(l) - \mathbf{G}_{\mathcal{A}_i}(l) \} \mathbf{Q}^{\rm H}_{\mathcal{A},k,i} \right]. \quad (24)$$

The notation $\mathbf{M}|_{\mathcal{J}}$ represents a submatrix of $\mathbf{M}$ specified by the column index set $\mathcal{J}$. The temporal subspace $\mathbf{E}_k$ in (2) corresponds to $\hat{\hat{\mathbf{V}}}_{\mathcal{A}_k}|_{1:\hat{r}^{\rm MDL}_{{\rm T},k}}$ in the transformed space by (21), where $\hat{r}^{\rm MDL}_{{\rm T},k}$ is estimated by the minimum description length (MDL) [28] of the vector $\mathtt{diag}\{\hat{\hat{\mathbf{D}}}^{\mathcal{A}_k}_{\rm T}\}$.

2) iST solution: In (2), the slot-independent gain is written as $\mathbf{B}_k(l) = \mathbf{H}_k(l)\mathbf{E}_k$. Similarly, we define a coarse-estimate gain matrix by

$$\hat{\hat{\mathbf{B}}}_{{\rm T},k}(l) = \hat{\hat{\mathbf{G}}}^{\rm LS}_{\mathcal{A}_k}(l)\,(\hat{\hat{\mathbf{V}}}_{\mathcal{A}_k}|_{1:r^{\rm max}_{\rm T}}) \quad (25)$$

for a pre-defined constant[2] $r^{\rm max}_{\rm T}$. We obtain a fine-estimate gain $\hat{\hat{\mathbf{B}}}_k(l)$ by applying i-spatial subspace projections to each column vector in $\hat{\hat{\mathbf{B}}}_{{\rm T},k}$ that corresponds to (3):

$$\mathtt{vec}\{\hat{\hat{\mathbf{B}}}_k(l)\} = \hat{\hat{\mathbf{\Phi}}}_k \mathtt{vec}\{\hat{\hat{\mathbf{B}}}_{{\rm T},k}(l)\}. \quad (26)$$

The $N_R r^{\rm max}_{\rm T} \times N_R r^{\rm max}_{\rm T}$ projection matrix $\hat{\hat{\mathbf{\Phi}}}_k$ is given by

$$\hat{\hat{\mathbf{\Phi}}}_k = \bigoplus\nolimits_{i=1}^{r^{\rm max}_{\rm T}} \hat{\hat{\mathbf{\Phi}}}_{k,i} \quad (27)$$

with $\hat{\hat{\mathbf{\Phi}}}_{k,i} = \mathbb{P}(\hat{\hat{\mathbf{U}}}_{k,i}|_{1:\hat{r}_{{\rm I},k,i}})$, where the unitary matrix $\hat{\hat{\mathbf{U}}}_{k,i}$ is obtained from the SVD:

$$\hat{\hat{\mathbf{U}}}_{k,i} \hat{\hat{\mathbf{D}}}_{{\rm I},k,i} \hat{\hat{\mathbf{U}}}^{\rm H}_{k,i} = \mathbb{R}^{L_{\rm S}}_{j=l} \left[\hat{\hat{\mathbf{b}}}_{k,i}(j) \right] \quad (28)$$

with $\hat{\hat{\mathbf{b}}}_{k,i}(j) = \hat{\hat{\mathbf{B}}}_{{\rm T},k}(j)|_i$. The parameter $\hat{r}_{{\rm I},k,i}$ can be determined by using the MDL of the singular values $\mathtt{diag}\{\hat{\hat{\mathbf{D}}}_{{\rm I},k,i}\}$.

The estimate (17) for the k-th TX stream is written as $\mathbf{\Gamma}^{-1/2}\hat{\hat{\mathbf{G}}}_{\mathcal{A}_k}(l)\mathbf{Q}^{-\rm H}_{\mathcal{A},k,k}$ by performing inversion of (21), where $\hat{\hat{\mathbf{G}}}_{\mathcal{A}_k}(l)$ is obtained by using (26) and (23). Specifically, the iST solution is given by

$$\hat{\mathbf{G}}_{\mathcal{A}_k}(l) = \mathbf{\Gamma}^{-1/2}(\hat{\hat{\mathbf{B}}}_k(l)|_{1:\hat{r}_{{\rm T},k}})\,(\hat{\hat{\mathbf{V}}}_{\mathcal{A}_k}|_{1:\hat{r}_{{\rm T},k}})^{\rm H}\mathbf{Q}^{-\rm H}_{\mathcal{A},k,k}, \quad (29)$$

where $\hat{r}_{{\rm T},k} \overset{\rm def}{=} \hat{r}^{\rm MDL}_{{\rm T},k} + r_\triangle$ with a parameter[3] $r_\triangle$.

3) pST solution: The pST subspace-based channel estimate [2], [27] can be written in the same form as (29). Specifically, we replace $\hat{\hat{\mathbf{B}}}_k(l)$ with $\hat{\hat{\mathbf{\Phi}}}_{{\rm P},\mathcal{A}_k} \cdot \hat{\hat{\mathbf{G}}}^{\rm LS}_{\mathcal{A}_k}(l)\,\hat{\hat{\mathbf{V}}}_{\mathcal{A}_k}$, where the p-spatial projection $\hat{\hat{\mathbf{\Phi}}}_{{\rm P},\mathcal{A}_k}$ is obtained from the PCA of the covariance matrix $\mathbb{R}^{L_{\rm S}}_{j=l}[\hat{\hat{\mathbf{G}}}^{\rm LS}_{\mathcal{A}_k}(j)]$.

Remark: In the right-hand side (RHS) of (23), the active-subset $\mathcal{A}_k$ has to be consistent over the past $L_{\rm S}$ slot durations.

[2] The constant is chosen so that $r^{\rm max}_{\rm T} \gg r_{{\rm T},k}$ for $\forall k$, since the rank estimate $\hat{r}^{\rm MDL}_{{\rm T},k}$ can change over slot-timings. Hence, we need to update all possible i-spatial CCMs corresponding to the first $r^{\rm max}_{\rm T}$-largest paths.

[3] $r_\triangle$ is used for the M-step (16). However, $r_\triangle = 0$ in the final E-step.

That is, calculating the temporal covariance matrix (23) directly from (20) incurs two problems:

P1 non-polynomial (NP)-hard, and
P2 the operation (21) requires the CIR to be estimated.

Note that the active-subset $\mathcal{A}_k$ has to be known *a priori* although it can be changed in the middle of the transmission. We may take a greedy approach that computes the CCM (23) for all possible $\mathcal{A}_k$. However, it causes Problem P1 since the number of all possible active-sets is of binomial order.

C. New Solutions to the E-step (15)

This subsection shows, first of all, a solution to Problems P1 and P2 which enables to specify the active-set *a posteriori*. We proposes, then, the solution to the E-step (15), referred to as $\ell 1\, p$ST technique, by performing the PCA for p-spatial and temporal subspaces. After that, another new conditional $\ell 1\, i$ST channel estimation using the ICA approach is presented by showing modifications from the $\ell 1\, p$ST algorithm.

1) Temporal CCM: We consider Proposition 2 that computes a temporal covariance for all TX streams. For the sake of simplicity, the slot timings l and j are omitted.

Proposition 2 (Temporal covariance matrix of (20))**.**

$$\mathbb{K}\left[\hat{\hat{\mathbf{G}}}^{\mathrm{LS}}_{\mathcal{A}}\right] = \mathbf{C}_{\mathrm{Ka}}(\mathcal{A}) + \mathbf{C}_{\mathrm{Kb}}(\mathcal{A}) - \left\{\mathbf{C}_{\mathrm{Kc}}(\mathcal{A}) + \mathbf{C}^{\mathsf{H}}_{\mathrm{Kc}}(\mathcal{A})\right\},$$

where

$$\begin{aligned}\mathbf{C}_{\mathrm{Ka}}(\mathcal{A}) &= \mathbf{R}^{-\mathsf{H}/2}_{\mathbf{XX}_{\mathcal{A}}}\mathbf{J}^{\mathsf{T}}_{\mathcal{A}}\mathbb{K}[\tilde{\mathbf{R}}_{\mathbf{YX}}]\mathbf{J}_{\mathcal{A}}\mathbf{R}^{-1/2}_{\mathbf{XX}_{\mathcal{A}}},\\ \mathbf{C}_{\mathrm{Kb}}(\mathcal{A}) &= \nabla\mathbf{R}^{1/2}_{\mathbf{XX}_{\mathcal{A}}}\mathbf{J}^{\mathsf{T}}_{\mathcal{A}}\mathbb{K}[\tilde{\mathbf{H}}]\mathbf{J}_{\mathcal{A}}\nabla\mathbf{R}^{\mathsf{H}/2}_{\mathbf{XX}_{\mathcal{A}}},\\ \mathbf{C}_{\mathrm{Kc}}(\mathcal{A}) &= \mathbf{R}^{-\mathsf{H}/2}_{\mathbf{XX}_{\mathcal{A}}}\mathbf{J}^{\mathsf{T}}_{\mathcal{A}}\mathbb{E}[\tilde{\mathbf{R}}^{\mathsf{H}}_{\mathbf{YX}}\tilde{\mathbf{H}}]\mathbf{J}_{\mathcal{A}}\nabla\mathbf{R}^{\mathsf{H}/2}_{\mathbf{XX}_{\mathcal{A}}}\end{aligned}$$

with $\tilde{\mathbf{R}}_{\mathbf{YX}} = \mathbf{\Gamma}^{1/2}\mathbf{R}_{\mathbf{YX}}$ *and* $\tilde{\mathbf{H}} = \mathbf{\Gamma}^{1/2}\mathbf{H}$.

Proof. Substituting (20) into $\mathbb{K}\left[\hat{\hat{\mathbf{G}}}^{\mathrm{LS}}_{\mathcal{A}}\right] = \mathbb{E}\left[\{\hat{\hat{\mathbf{G}}}^{\mathrm{LS}}_{\mathcal{A}}\}^{\mathsf{H}}\,\hat{\hat{\mathbf{G}}}^{\mathrm{LS}}_{\mathcal{A}}\right]$ obtains Proposition 2. □

Proposition 2 shows that the temporal CCM of the compressed estimate (20) can be written as a function of active-set $\mathcal{A}$. However, all the auto/cross-correlation $\mathbb{K}[\cdot]$ terms are independent of $\mathcal{A}$. We can, thereby, perform sparse-PCA given active-set $\mathcal{A}$, after updating the correlation matrices. Hence, Proposition 2 solves Problem P1 since it is not necessary to compute $\mathbb{K}\left[\hat{\hat{\mathbf{G}}}^{\mathrm{LS}}_{\mathcal{A}}\right]$ for all possible active-sets $\forall\mathcal{A}$.

Nevertheless, Proposition 2 still has Problem P2. We show the next proposition to be used in a solution of the P2.

Proposition 3. *The temporal covariance matrix of the k-th TX stream can be computed as*

$$\begin{aligned}\mathbb{K}\left[\hat{\hat{\mathbf{G}}}^{\mathrm{LS}}_{\mathcal{A}_k}\right] &= \mathbf{C}_{\mathrm{Ka}}(\mathcal{A},k) + \mathbf{C}_{\mathrm{Kb}}(\mathcal{A},k)\\ &\quad - \left\{\mathbf{C}_{\mathrm{Kc}}(\mathcal{A},k) + \mathbf{C}^{\mathsf{H}}_{\mathrm{Kc}}(\mathcal{A},k)\right\}, \qquad (30)\end{aligned}$$

where

$$\mathbf{C}_{\mathrm{Ka}}(\mathcal{A},k) = \sum_{j=1}^{k}\sum_{i=1}^{k}\mathfrak{Q}^{-\mathsf{H}}_{\mathcal{A},i,k}\mathsf{J}^{\mathsf{T}}_{\mathcal{A}_k}\mathbf{K}^{\mathrm{a}}_{i,j}\mathsf{J}_{\mathcal{A}_k}\mathfrak{Q}^{-1}_{\mathcal{A},j,k}, \qquad (31)$$

$$\mathbf{C}_{\mathrm{Kb}}(\mathcal{A},k) = \sum_{j=k+1}^{N_T}\sum_{i=k+1}^{N_T}\mathbf{Q}_{\mathcal{A},k,j}\mathsf{J}^{\mathsf{T}}_{\mathcal{A}_k}\mathbf{K}^{\mathrm{b}}_{j,i}\mathsf{J}_{\mathcal{A}_k}\mathbf{Q}^{\mathsf{H}}_{\mathcal{A},k,i}, \qquad (32)$$

$$\mathbf{C}_{\mathrm{Kc}}(\mathcal{A},k) = \sum_{j=1}^{k}\sum_{i=k+1}^{N_T}\mathfrak{Q}^{-\mathsf{H}}_{\mathcal{A},i,k}\mathsf{J}^{\mathsf{T}}_{\mathcal{A}_k}\mathbf{K}^{\mathrm{c}}_{i,j}\mathsf{J}_{\mathcal{A}_k}\mathbf{Q}^{\mathsf{H}}_{\mathcal{A},k,j}. \qquad (33)$$

The submatrices $\mathbf{K}^{\mathrm{a}}_{i,j}, \mathbf{K}^{\mathrm{b}}_{i,j}$ and $\mathbf{K}^{\mathrm{c}}_{i,j}$ can be updated independently of the active-set $\mathcal{A}$:

$$\begin{aligned}\mathbf{K}^{\mathrm{a}}_{i,j} &= \mathbb{E}[\tilde{\mathbf{R}}^{\mathsf{H}}_{\mathbf{YX}_i}\tilde{\mathbf{R}}_{\mathbf{YX}_j}], &(34)\\ \mathbf{K}^{\mathrm{b}}_{i,j} &= \mathbb{E}[\tilde{\mathbf{H}}^{\mathsf{H}}_i\tilde{\mathbf{H}}_j], &(35)\\ \mathbf{K}^{\mathrm{c}}_{i,j} &= \mathbb{E}[\tilde{\mathbf{R}}^{\mathsf{H}}_{\mathbf{YX}_i}\tilde{\mathbf{H}}_j], &(36)\end{aligned}$$

where $\mathbf{R}_{\mathbf{YX}_i} = \mathbf{R}_{\mathbf{YX}}|_{\mathcal{D}_i}$. The $|\mathcal{A}_i|\times|\mathcal{A}_j|$ matrix[4] *$\mathfrak{Q}^{-1}_{\mathcal{A},i,j}$ is the (i,j)-th block submatrix of $\mathbf{R}^{-1/2}_{\mathbf{XX}_{\mathcal{A}}}$. Multiplying the $|W|\times|\mathcal{A}_k|$ matrix $\mathsf{J}_{\mathcal{A}_k} = \mathbf{J}^{\mathsf{T}}_{\mathcal{D}_k}\mathbf{J}_{\mathcal{A}_k}$ extracts the entries corresponding to $\mathcal{A}_k$ from the domain $\mathcal{D}_k = \{[1:W] + (k-1)W\}$.*

Proof. We can obtain (30) from Proposition 2, by focusing on the (k,k)-th block matrix. □

We find from (32) and (33) that the CIR submatrices $\mathbf{H}_i$ are required only for $i \geq k+1$ in order to calculate the k-th covariance matrix $\mathbb{K}[\hat{\hat{\mathbf{G}}}_{\mathcal{A}_k}]$. Therefore, Problem P2 can be solved by the back-substitution algorithm [27] which estimates the TX stream-wise CIRs by descending order of $k = N_T, \cdots, 1$ using approximations $\mathbf{H}_{k'} \approx \hat{\mathbf{H}}_{\mathcal{A}}|_{\mathcal{D}_{k'}}, \forall k' > k$. Specifically, for the k-th TX stream, we can partially update covariance submatrices[5] $\mathbf{K}^{\mathrm{b}}_{i,j}$ and $\mathbf{K}^{\mathrm{c}}_{i,j}$, as

$$\begin{aligned}\mathbf{K}^{\mathrm{b}}_{k+1,j} &= \mathbb{E}[\hat{\tilde{\mathbf{H}}}^{\mathsf{H}}_{k+1}\hat{\tilde{\mathbf{H}}}_j] \qquad (\forall j \geq k+1), &(37)\\ \mathbf{K}^{\mathrm{c}}_{i,k+1} &= \mathbb{E}[\tilde{\mathbf{R}}^{\mathsf{H}}_{\mathbf{YX}_i}\hat{\tilde{\mathbf{H}}}_{k+1}] \qquad (\forall i \leq k), &(38)\end{aligned}$$

respectively, by removing terms already calculated for the $k'(\geq k+1)$-th covariance matrices $\mathsf{C}_{\mathrm{Kb}}(\mathcal{A},k')$ and $\mathsf{C}_{\mathrm{Kc}}(\mathcal{A},k')$.

2) P-spatial CCM : Proposition 3 can be extended for the p-spatial covariance matrices (5) of the compressed version using the vectorization operation.

Proposition 4. *The p-spatial covariance matrix of the k-th TX stream can be computed as*

$$\begin{aligned}\mathbb{R}\left[\hat{\hat{\mathbf{G}}}_{\mathcal{A}_k}\right] &= \mathbf{C}_{\mathrm{Ra}}(\mathcal{A},k) + \mathbf{C}_{\mathrm{Rb}}(\mathcal{A},k)\\ &\quad - \left\{\mathbf{C}_{\mathrm{Rc}}(\mathcal{A},k) + \mathbf{C}^{\mathsf{H}}_{\mathrm{Rc}}(\mathcal{A},k)\right\}. \qquad (39)\end{aligned}$$

We obtain the three matrices in the left-hand side (LHS) of (39) *via their vectorized versions, as*

$$\begin{aligned}\mathtt{vec}\{\mathbf{C}_{\mathrm{Ra}}(\mathcal{A},k)\} &= \mathbf{T}_{\mathrm{a}}\,\mathbf{v}_{\mathrm{a}}(\mathcal{A},k),\\ \mathtt{vec}\{\mathbf{C}_{\mathrm{Rb}}(\mathcal{A},k)\} &= \mathbf{T}_{\mathrm{b}}\,\mathbf{v}_{\mathrm{b}}(\mathcal{A},k),\\ \mathtt{vec}\{\mathbf{C}_{\mathrm{Rc}}(\mathcal{A},k)\} &= \mathbf{T}_{\mathrm{c}}\,\mathbf{v}_{\mathrm{c}}(\mathcal{A},k),\end{aligned}$$

[4] In general, $\mathfrak{Q}^{-1}_{\mathcal{A},i,j} \neq \{\mathbf{Q}_{\mathcal{A},i,j}\}^{-1} = \mathbf{Q}^{-1}_{\mathcal{A},i,j}$.

[5] $\mathbf{K}^{\mathrm{b}}_{i,k+1}$, $\forall i \geq k+2$ are also needed in (32). However, $\mathbf{K}^{\mathrm{b}}_{i,k+1} = \{\mathbf{K}^{\mathrm{b}}_{k+1,i}\}^{\mathsf{H}}$, since $\mathbb{K}[\mathbf{H}]$ is an Hermitian matrix. Note that the submatrix $\mathbf{K}^{\mathrm{a}}_{i,j}$ may, however, be obtained by directly calculating $\mathbb{K}[\tilde{\mathbf{R}}_{\mathbf{YX}}]$.

where we denote $\mathbf{T}_{\rm a} = \mathbb{E}[\tilde{\mathbf{R}}^*_{\mathbf{YX}} \otimes \tilde{\mathbf{R}}_{\mathbf{YX}}]$, $\mathbf{T}_{\rm b} = \mathbb{E}[\tilde{\mathbf{H}}^* \otimes \tilde{\mathbf{H}}]$, $\mathbf{T}_{\rm c} = \mathbb{E}[\tilde{\mathbf{H}}^* \otimes \tilde{\mathbf{R}}_{\mathbf{YX}}]$, *and*

$$\begin{aligned}
\mathbf{v}_{\rm a}(\mathcal{A},k) &= \mathtt{vec}\{\mathbf{J}_{\mathcal{A}}\mathbf{R}^{-1/2}_{\mathbf{XX}_{\mathcal{A}}}\mathtt{E}(\mathcal{A},k)\mathbf{R}^{-\mathsf{H}/2}_{\mathbf{XX}_{\mathcal{A}}}\mathbf{J}^{\mathsf{T}}_{\mathcal{A}}\},\\
\mathbf{v}_{\rm b}(\mathcal{A},k) &= \mathtt{vec}\{\mathbf{J}_{\mathcal{A}}\nabla\mathbf{R}^{\mathsf{H}/2}_{\mathbf{XX}_{\mathcal{A}}}\mathtt{E}(\mathcal{A},k)\nabla\mathbf{R}^{1/2}_{\mathbf{XX}_{\mathcal{A}}}\mathbf{J}^{\mathsf{T}}_{\mathcal{A}}\},\\
\mathbf{v}_{\rm c}(\mathcal{A},k) &= \mathtt{vec}\{\mathbf{J}_{\mathcal{A}}\mathbf{R}^{-1/2}_{\mathbf{XX}_{\mathcal{A}}}\mathtt{E}(\mathcal{A},k)\nabla\mathbf{R}^{1/2}_{\mathbf{XX}_{\mathcal{A}}}\mathbf{J}^{\mathsf{T}}_{\mathcal{A}}\}
\end{aligned}$$

with $\mathtt{E}(\mathcal{A},k) = \mathtt{D}_{\mathtt{IAG}}([\{\mathbb{I}(i \in \mathcal{A}_k) \mid \forall i \in \mathcal{A}\}]^{\mathsf{T}}) \in \mathbb{Z}^{|\mathcal{A}|\times|\mathcal{A}|}$.

The back-substitution can be applied[6] to Proposition 4, too.

3) Conditional $\ell 1\, p$ST algorithm: By combining Sections IV-C1 and IV-C2, the conditional $\ell 1\, p$ST channel estimation given active-set $\mathcal{A}$: $f^{\ell 1}_{\rm PCA}(\mathbf{Y}, \mathbf{R}_{\mathbf{YX}}, \theta_{\mathbf{T}}, \theta_{\mathbf{K}} \mid \mathcal{A})$ is summarized in Algorithm 1. The input sets $\theta_{\mathbf{T}}$ and $\theta_{\mathbf{K}}$ are

$$\begin{aligned}
\theta_{\mathbf{T}} &= \{\mathbf{T}_{\rm a}, \mathbf{T}_{\rm b}, \mathbf{T}_{\rm c}\},\\
\theta_{\mathbf{K}} &= \{\mathbf{K}^{\rm a}_{i,j}, \mathbf{K}^{\rm b}_{i,j}, \mathbf{K}^{\rm c}_{i,j} \mid 1 \le i,j \le N_T\}.
\end{aligned}$$

For the sake of conciseness, the known TS matrix $\mathbf{X}$ is omitted from input parameters. The parameters $\mathbf{T}_{\rm a}$ and $\mathbf{K}^{\rm a}_{i,j}$ can be updated before executing Algorithm 1, whereas the others are updated at Steps 4 and 7.

The $\ell 1$ MMSE channel estimate (29) is computed at Step 10, by the descending oder $k = N_T, \cdots, 1$ to perform the back-substitution, where the temporal subspace $\hat{\hat{\mathbf{V}}}_{\mathcal{A}_k}$ and the spatial projection $\hat{\hat{\mathbf{\Phi}}}_{\mathcal{A}_k}$ are obtained at Steps 6 and 9, respectively. It should be emphasized that, at Steps 5 and 8, the temporal and spatial CCMs are reduced via Propositions 3 and 4, but they do not directly require the compressed LS estimates $\hat{\hat{\mathbf{G}}}^{\rm LS}_{\mathcal{A}_k}(l)$.

The output CIR estimate matrix $\hat{\mathbf{H}}_{\mathcal{A}}$ is written, as

$$\hat{\mathbf{H}}_{\mathcal{A}} = [\hat{\mathbf{G}}_{\mathcal{A}_1}, \cdots, \hat{\mathbf{G}}_{\mathcal{A}_{N_T}}]\,\mathbf{J}^{\mathsf{T}}_{\mathcal{A}} \tag{40}$$

by using (29). Set $\theta_{\mathbf{\Pi}}$ is output as

$$\theta_{\mathbf{\Pi}} = \{\hat{\hat{\mathbf{V}}}_{\mathcal{A}_k}, \hat{\mathbf{r}}_{{\rm S},k}, \hat{r}_{{\rm T},k} \mid k = 1, \cdots, N_T\} \tag{41}$$

with $\hat{\mathbf{r}}_{{\rm S},k} = \hat{r}_{{\rm P},k}\mathbf{1}_{\hat{r}_{{\rm T},k}}$, for the AAD algorithm discussed later.

4) Conditional $\ell 1\, i$ST algorithm : We may obtain the i-spatial projection matrix (27) by using the slot-dependent gain estimate (25) when the temporal subspace estimates $\hat{\hat{\mathbf{V}}}_{\mathcal{A}_k}$ are accurate enough. This is because the gain estimate vector $\hat{\hat{\mathbf{b}}}_{k,i}(j)$ is independent of the active-set $\mathcal{A}$ representing the sparse-supports of the temporal-subspace.

Algorithm 2 summarizes the conditional $\ell 1\, i$ST technique given active-set $\mathcal{A}$: $f^{\ell 1}_{\rm ICA}(\mathbf{Y}, \mathbf{R}_{\mathbf{YX}}, \theta_{\mathbf{R}}, \theta_{\mathbf{K}} \mid \mathcal{A})$, where all steps are the same as that in Algorithm 1 except the steps between 7 and 11 that perform the ICA to find the i-spatial projectors. The input parameter $\theta_{\mathbf{R}}$ is defined by

$$\theta_{\mathbf{R}} = \left\{\mathbb{R}^{L_{\rm S}}_{j=l}[\hat{\hat{\mathbf{b}}}_{k,i}(j)] \;\middle|\; 1 \le k \le N_T, 1 \le i \le r^{\max}_{\rm T}\right\}.$$

Moreover, in the output parameter $\theta_{\mathbf{\Pi}}$ (41), we let $\hat{\mathbf{r}}_{{\rm S},k} = [\hat{r}_{{\rm I},k,1}, \cdots, \hat{r}_{{\rm I},k,\hat{r}_{{\rm T},k}}]^{\mathsf{T}}$.

[6]Tensor $\mathbf{T}_{\rm b}$ can be updated per sub-column: $\mathbf{T}_{\rm b}|_{\mathcal{J}^{\rm b}_k \setminus \mathcal{J}^{\rm b}_{k+1}}$ is given by $\mathbb{E}\left[\hat{\mathbf{H}}^*_{k+1} \otimes [\hat{\mathbf{H}}_{k+1}, \cdots, \hat{\mathbf{H}}_{N_T}],\ [\hat{\mathbf{H}}_{k+2}, \cdots, \hat{\mathbf{H}}_{N_T}]^* \otimes \hat{\mathbf{H}}_{k+1}\right]$. Similarly, $\mathbf{T}_{\rm c}|_{\mathcal{J}^{\rm c}_k} = \mathbb{E}\left[\hat{\mathbf{H}}^*_{k+1} \otimes [\mathbf{R}_{\mathbf{YX}_1}, \cdots, \mathbf{R}_{\mathbf{YX}_k}]\right]$. The index sets are

$$\begin{aligned}
\mathcal{J}^{\rm b}_k &= \{s(i_1,i_2) \mid 1 + kW \le \forall i_1, i_2 \le WN_T\}\\
\mathcal{J}^{\rm c}_k &= \{s(i_1,i_2) \mid i_1 = W\min(k+1, N_T), 1 \le \forall i_2 \le kN_T\}
\end{aligned}$$

with $s(i_1,i_2) = (i_1 - 1)WN_T + i_2$.

Algorithm 1 $f^{\ell 1}_{\rm PCA}(\mathbf{Y}, \mathbf{R}_{\mathbf{YX}}, \theta_{\mathbf{T}}, \theta_{\mathbf{K}} \mid \mathcal{A})$.

Input: $\mathbf{Y}, \mathbf{R}_{\mathbf{YX}}, \theta_{\mathbf{K}}, \theta_{\mathbf{T}}$ and $\mathcal{A}$.

1: Compute the conditional $\ell 1$ LS estimate $\hat{\mathbf{G}}^{\rm LS}_{\mathcal{A}}$ (18).
2: **for** $k = N_T$ to 1 **do**
3: Compute $\hat{\hat{\mathbf{G}}}^{\rm LS}_{\mathcal{A}_k}$ by (24), where $\mathbf{G}_{\mathcal{A}_i} \approx \hat{\mathbf{G}}_{\mathcal{A}_i}$ for $\forall i > k$.
4: Update $\theta_{\mathbf{K}}$ partially by (37) and (38).
5: Update $\mathbb{K}\left[\hat{\hat{\mathbf{G}}}_{\mathcal{A}_k}\right]$ by using (30).
6: Obtain the temporal subspace $\hat{\hat{\mathbf{V}}}_{\mathcal{A}_k}$ by (23).
7: Update $\theta_{\mathbf{T}}$ partially using (37) and (38).
8: Update $\mathbb{R}\left[\hat{\hat{\mathbf{G}}}_{\mathcal{A}_k}\right]$ by using (39).
9: Obtain the projector $\hat{\hat{\mathbf{\Phi}}}_{{\rm P},\mathcal{A}_k}$ corresponding to (5) .
10: Obtain the k-th channel estimate $\hat{\mathbf{G}}_{\mathcal{A}_k}$ by (29), where $\hat{\hat{\mathbf{B}}}_k = \hat{\hat{\mathbf{\Phi}}}_{{\rm P},\mathcal{A}_k} \cdot \hat{\hat{\mathbf{G}}}^{\rm LS}_{\mathcal{A}_k}\hat{\hat{\mathbf{V}}}_{\mathcal{A}_k}$.
11: **end for**

Output: $\hat{\mathbf{H}}_{\mathcal{A}}$, $\theta_{\mathbf{\Pi}}$, $\theta_{\mathbf{T}}$ and $\theta_{\mathbf{K}}$.

Algorithm 2 $f^{\ell 1}_{\rm ICA}(\mathbf{Y}, \mathbf{R}_{\mathbf{YX}}, \theta_{\mathbf{R}}, \theta_{\mathbf{K}}, r_{\Delta} \mid \mathcal{A})$.

Input: $\mathbf{Y}, \mathbf{R}_{\mathbf{YX}}, \theta_{\mathbf{R}}, \theta_{\mathbf{K}}, r_{\Delta}$, and $\mathcal{A}$.

1: Compute the conditional $\ell 1$ LS estimate $\hat{\mathbf{G}}^{\rm LS}_{\mathcal{A}}$ (18).
2: **for** $k = N_T$ to 1 **do**
3: Compute $\hat{\hat{\mathbf{G}}}^{\rm LS}_{\mathcal{A}_k}$ by (24), where $\mathbf{G}_{\mathcal{A}_i} \approx \hat{\mathbf{G}}_{\mathcal{A}_i}$ for $\forall i > k$.
4: Update $\theta_{\mathbf{K}}$ partially by (37) and (38).
5: Update $\mathbb{K}\left[\hat{\hat{\mathbf{G}}}_{\mathcal{A}_k}\right]$ by using (30).
6: Obtain the temporal-subspace $\hat{\hat{\mathbf{V}}}_{\mathcal{A}_k}$ by (23).
7: Obtain the coarse gain estimate $\hat{\hat{\mathbf{B}}}_{{\rm T},k}$ (25).
8: **for** $i = 1$ to $r^{\max}_{\rm T}$ **do**
9: Update $\mathbb{R}\left[\hat{\hat{\mathbf{b}}}_{{\rm T},k,i}\right]$ in (28).
10: Obtain the projector $\hat{\hat{\mathbf{\Phi}}}_{k,i}$ in (27).
11: **end for**
12: Obtain the k-th channel estimate $\hat{\mathbf{G}}_{\mathcal{A}_k}$ by (29) using r_{Δ}, where $\hat{\hat{\mathbf{B}}}_k(l)$ is computed via (26).
13: **end for**

Output: $\hat{\mathbf{H}}_{\mathcal{A}}$, $\theta_{\mathbf{\Pi}}$, $\theta_{\mathbf{R}}$, and $\theta_{\mathbf{K}}$.

D. Solution to the M-step (16)

We solve the problem (16) by extending the AAD algorithm [17] as to optimize MSE performance of the conditional $\ell 1\, i$ST channel estimation. In the following, let us denote $\hat{r}_{{\rm S},k,i} = \hat{r}_{{\rm I},k,i}$, since the AAD algorithm for the $\ell 1\, p$ST approach is also obtained by assuming $\hat{r}_{{\rm S},k,i} = \hat{r}_{{\rm P},k}$ for $\forall i$.

The AAD updates the active-set recursively by

$$\begin{aligned}
\mathcal{A}^{[n+1]} &= \mathtt{AAD}(\hat{\mathbf{d}}^{[n]}_{\mathbf{H}}, \theta^{[n]}_{\mathbf{\Pi}}, \sigma^2_{\mathfrak{N}} \mid \mathcal{A}^{[n]})\\
&= \left\{ j \;\middle|\; \Delta\hat{d}^{[n]}_j > 0,\ \forall j \in \mathcal{A}^{[n]} \cap \mathcal{T}_W(\Delta\hat{\mathbf{d}}^{[n]}, E) \right\},
\end{aligned} \tag{42}$$

where the superscript $[n]$ denotes n-th iteration. The parameter $\Delta\hat{d}^{[n]}_j$ is the j-th entry of residue vector $\Delta\hat{\mathbf{d}}^{[n]} = \hat{\mathbf{d}}^{[n]}_{\mathbf{H}} - \{\mathfrak{m}^{[n]}(\sigma^2_{\mathfrak{N}}) + \mathfrak{e}^{[n]}(\sigma^2_{\mathfrak{N}})\}$, where the delay profile estimate is

$$\hat{\mathbf{d}}^{[n]}_{\mathbf{H}} \approx \begin{cases} \mathtt{diag}\{\mathbb{K}^{L_{\rm S}}_{j=l}[\hat{\mathbf{H}}_{\ell 2}(j)]\} & (n = 0)\\ \mathtt{diag}\{\hat{\mathbf{H}}^{\mathsf{H}}_{\mathcal{A}^{[n]}}(l)\,\hat{\mathbf{H}}_{\mathcal{A}^{[n]}}(l)\} & (n \ge 1) \end{cases}. \tag{43}$$

The vector[7] $\mathfrak{m}^{[n]}(\sigma^2_{\mathfrak{N}})$ approximately describes the distribution of the squared errors over CIR taps for the channel estimate $\hat{\mathbf{H}}_{\mathcal{A}^{[n]}}$: $[\mathfrak{m}^{[n]}_{\mathcal{A}_1}(\sigma^2_{\mathfrak{N}})^{\mathsf{T}}, \cdots, \mathfrak{m}^{[n]}_{\mathcal{A}_{N_T}}(\sigma^2_{\mathfrak{N}})^{\mathsf{T}}]^{\mathsf{T}}$, where $\mathfrak{m}^{[n]}_{\mathcal{A}_k}(\sigma^2_{\mathfrak{N}}) \overset{\text{def}}{=} \mathtt{J}_{\mathcal{A}^{[n]}_k}\mathtt{diag}\left\{\mathbf{K}_{\Delta\hat{\mathbf{G}}_k}(\sigma^2_{\mathfrak{N}}, \mathcal{A}^{[n]}_k)\right\}$ with

$$\mathbf{K}_{\Delta\hat{\mathbf{G}}_k}(\sigma^2_{\mathfrak{N}}, \mathcal{A}_k) = \sigma^2_{\mathfrak{N}}\,\mathbf{Q}^{-1}_{\mathcal{A}k,k}\hat{\hat{\mathtt{V}}}_{\mathcal{A}_k}\,\mathbf{\Lambda}_{\mathrm{S},k}\,\hat{\hat{\mathtt{V}}}^{\mathsf{H}}_{\mathcal{A}_k}\mathbf{Q}^{-\mathsf{H}}_{\mathcal{A}k,k}. \quad (44)$$

We denote $\hat{\hat{\mathtt{V}}}_{\mathcal{A}_k} = \hat{\hat{\mathbf{V}}}_{\mathcal{A}_k}|_{1:\hat{r}_{\mathrm{T},k}}$ and $\mathbf{\Lambda}_{\mathrm{S},k} = \mathtt{diag}\{\hat{\mathbf{r}}_{\mathrm{S},k}\}$. The error $\mathfrak{e}^{[n]}(\sigma^2_{\mathfrak{N}})$ vector of the estimated delay profile $\hat{\mathbf{d}}^{[n]}_{\mathbf{H}}$ may be approximated by $[\mathfrak{e}^{[n]}_1(\sigma^2_{\mathfrak{N}})^{\mathsf{T}}, \cdots, \mathfrak{e}^{[n]}_{N_T}(\sigma^2_{\mathfrak{N}})^{\mathsf{T}}]^{\mathsf{T}}$, where

$$\mathfrak{e}^{[n]}_k(\sigma^2_{\mathfrak{N}}) = \mu(\sigma^2_{\mathfrak{N}}, \hat{\mathbf{r}}_{\mathrm{S},k}, \hat{r}_{\mathrm{T},k}, \mathcal{A}^{[n]}_k)\,\mathtt{diag}\left\{\mathtt{J}_{\mathcal{A}^{[n]}_k}\mathtt{J}^{\mathsf{T}}_{\mathcal{A}^{[n]}_k}\right\}/|\mathcal{A}^{[n]}_k|. \quad (45)$$

The function $\mu(\cdot)$ represents, as discussed later in Section V, an MSE performance of the compressed estimate $\hat{\mathbf{G}}_{\mathcal{A}^{[n]}_k}$:

$$\begin{aligned}\mu(\sigma^2_{\mathfrak{N}}, \mathbf{r}_{\mathrm{S},k}, r_{\mathrm{T},k}, \mathcal{A}_k) &= \mathtt{tr}\{\mathbf{K}_{\Delta\hat{\mathbf{G}}_k}(\sigma^2_{\mathfrak{N}}, \mathcal{A}_k)\} \\ &= \sigma^2_{\mathfrak{N}}\left(\textstyle\sum_{i=1}^{r_{\mathrm{T},k}} r_{\mathrm{S},k,i}\right)\mathtt{tr}\{\mathbf{R}^{-1}_{\mathbf{XX}_{\mathcal{A}}}\}/|\mathcal{A}|. \quad (46)\end{aligned}$$

The operation $\mathcal{T}_W(\mathbf{x}, E)$ forms an index subset of the *top E* entries in each length-W subvector of vector $\mathbf{x}$ for a predefined constant $E \le W$, which imposes the maximum cardinality regularization onto the active-set.

E. Solutions to (12)

We show a new (*unconditional*) $\ell1\,i$ST channel estimation algorithm. Notice that the (*unconditional*) $\ell1\,p$ST algorithm is straightforwardly obtained by replacing the conditional ICA $f^{\ell1}_{\mathrm{ICA}}(\cdot \mid \mathcal{A})$ with the PCA version $f^{\ell1}_{\mathrm{PCA}}(\cdot \mid \mathcal{A})$.

1) $\ell1\,i$ST: Algorithm 3 summarizes the $\ell1\,i$ST channel estimation obtained by combining the conditional $\ell1\,i$ST and AAD techniques. The EM sub-problems (15) and (16) are iteratively performed at Steps 5 and 7, respectively. The function $f^{\ell1}_{\mathrm{ICA}}(\cdot \mid \mathcal{A})$ is the conditional $\ell1\,i$ST technique shown in Algorithm 2, where it represents the $\ell2\,i$ST channel estimation in the initial iteration since $\mathcal{A}^{[0]} = \{1, \cdots, WN_T\}$. The input parameter $r_{\mathbf{d_H}} \ge 0$ is used to prevent underestimating the delay profile (43).

The optimal solution $\hat{\mathbf{H}}^{[\hat{n}]}$ to the problem (12) is determined from all the possible N_{AAD} candidates, where the index $\hat{n}$ may be chosen according to the minimum of AICc:

$$\mathtt{AICc}(\hat{\mathbf{H}}^{[n]}) = 2\mathcal{L}(l, \hat{\mathbf{H}}^{[n]}) + \tfrac{2K_{\mathrm{free}}(K_{\mathrm{free}}+1)}{N_{\mathrm{in}}-K_{\mathrm{free}}-1}. \quad (47)$$

The parameters K_{free} and N_{in} are given by $\sum_{k=1}^{N_T}\sum_{i=1}^{\hat{r}^{[n]}_{\mathrm{T},k}}\hat{r}^{[n]}_{\mathrm{S},k,i}$ and $N_R\tilde{L}_t$, respectively. The output $\theta^{[\hat{n}]}_{\mathbf{K}}$ is reused as the input parameters $\theta_{\mathbf{K}}$ at the next slot timing. However, we do not update the original $\theta_{\mathbf{K}}$ in the for-loop between Steps 3 and 8.

F. Computational Complexity Order

Table II summarizes complexity orders needed for the proposed algorithms. The $\ell1\,i$ST and $\ell1\,p$ST algorithms increase the complexity order by $\mathcal{O}(\kappa^5 N_{\mathrm{AAD}})$ and $\mathcal{O}(\kappa^6 N_{\mathrm{AAD}})$, respectively, from the order $\mathcal{O}(\kappa^6 + \kappa^4)$ needed for the conventional $\ell2\,p$ST technique, where $\kappa \overset{\text{def}}{=} \mathtt{max}\{W, N_T, N_R\}$.

[7]In the first $\lceil|\mathcal{A}^{[n]}_k|/N_R\rceil$ slots, we may use $\mathfrak{m}^{[n]}_{\mathcal{A}_k}(\sigma^2_{\mathfrak{N}}) \approx \mathfrak{e}^{[n]}_{\mathcal{A}_k}(\sigma^2_{\mathfrak{N}})$, since the estimated subspace can be inaccurate to describe the symbol-wise error.

Algorithm 3 The $\ell1\,i$ST channel estimation.

Input: $\mathbf{Y}$, $\sigma^2_{\mathfrak{N}}$, $\mathbf{\Gamma}$, $\theta_{\mathbf{R}}$, $\theta_{\mathbf{K}}$, and $r_{\mathbf{d_H}}$.
1: Initialize $\mathcal{A}^{[0]} = \{1, \cdots, WN_T\}$.
2: Compute $\mathbf{R_{YX}}$ and update $\mathbf{K}^{\mathrm{a}}_{i,j}$ (34).
3: **for** $n = 0$ to N_{AAD} **do**
4: Set r_Δ at 0 when $n = N_{\mathrm{AAD}}$, otherwise $r_\Delta = r_{\mathbf{d_H}}$.
5: $\{\hat{\mathbf{H}}^{[n]}, \theta^{[n]}_{\mathbf{\Pi}}, \theta^{[n]}_{\mathbf{R}}, \theta^{[n]}_{\mathbf{K}}\} = f^{\ell1}_{\mathrm{ICA}}(\mathbf{Y}, \mathbf{R_{YX}}, \theta_{\mathbf{R}}, \theta_{\mathbf{K}}, r_\Delta \mid \mathcal{A}^{[n]})$.
6: Update $\hat{\mathbf{d}}^{[n]}_{\mathbf{H}}$ (43) by using $\hat{\mathbf{H}}^{[n]}$.
7: $\mathcal{A}^{[n+1]} = \mathtt{AAD}(\hat{\mathbf{d}}^{[n]}_{\mathbf{H}}, \theta^{[n]}_{\mathbf{\Pi}}, \sigma^2_{\mathfrak{N}} \mid \mathcal{A}^{[n]})$ by (42).
8: **end for**
9: $\hat{n} = \arg\min_{n\ge1}\mathtt{AICc}(\hat{\mathbf{H}}^{[n]})$ by using (47).
Output: $\hat{\mathbf{H}}^{[\hat{n}]}$, $\theta^{[\hat{n}]}_{\mathbf{R}}$, and $\theta^{[\hat{n}]}_{\mathbf{K}}$.

1) $\ell1\,i$ST : Table III details the complexity order required for the $\ell1\,i$ST. The first seven items[8] in Table III are the complexity for Algorithm 2 $f^{\ell1}_{\mathrm{ICA}}(\cdot|\mathcal{A}^{[n]})$ performed at Step 5 in Algorithm 3, whereas the last item describes the complexity for the AAD algorithm performed at Step 7 of Algorithm 3. As observed from Table III, the complexity for the $\ell1\,i$ST is dominated by that needed to update the CCM[9] $\mathbb{K}[\hat{\hat{\mathbf{G}}}^{\mathrm{LS}}_{\mathcal{A}_k}]$ and the projection matrices $\hat{\hat{\mathbf{\Phi}}}_{k,i}$, where $\mathcal{O}(r^{\max}_{\mathrm{T}}) \le W$.

2) $\ell2\,i$ST: The complexity for the $\ell2\,i$ST is also described by Table III with two modifications: 1) the CCM $\mathbb{K}[\hat{\hat{\mathbf{G}}}^{\mathrm{LS}}_{\mathcal{A}_k}]$ can be updated in $\mathcal{O}(N_T \cdot W^3)$ for the fixed active-set $\mathcal{A}_{[0]} = \{1 : WN_T\}$ according to the definition of the operation $\mathbb{K}[\cdot]$. 2) Only the first AAD iteration ($N_{\mathrm{AAD}} = 1$) is performed.

3) Unstructured-iST: As a benchmark, we summarize the complexity of the conventional approach [16] referred to as unstructured-iST (u-iST). It also performs the ST-subspace projection for the $\ell1$ LS channel estimate as in (29). However, as discussed in [16], the projection matrix is approximated by that obtained by the $\ell2\,i$ST. The complexity order needed for the u-iST technique is, hence, the same as that of the $\ell2\,i$ST.

4) $\ell1\,p$ST: Table IV details the complexity order for the $\ell1\,p$ST. The first six items in Table IV are, similar to Table III, the complexity required for the function $f^{\ell1}_{\mathrm{PCA}}(\cdot|\mathcal{A}^{[n]})$. As shown in Table IV, the complexity for the $\ell1\,p$ST is dominated by that to update the spatial CCMs $\mathbb{R}[\hat{\hat{\mathbf{G}}}^{\mathrm{LS}}_{\mathcal{A}_k}]$.

5) $\ell2\,p$ST: The details of the complexity for the $\ell2\,p$ST can be also described by the first six items in Table IV, where we update the covariance matrices $\mathbb{K}[\hat{\hat{\mathbf{G}}}^{\mathrm{LS}}_{\mathcal{A}_k}]$ and $\mathbb{R}[\hat{\hat{\mathbf{G}}}^{\mathrm{LS}}_{\mathcal{A}_k}]$ with the complexity order $\mathcal{O}(N_T \cdot W^2 N_R)$ and $\mathcal{O}(N_T \cdot WN_R^2)$, respectively, only for the initial active-set $\mathcal{A}_{[0]} = \{1 : WN_T\}$.

V. Performance Analysis

MSE performance of the proposed algorithm is shown after detailing the estimate error $\Delta\hat{\mathbf{H}}_{\mathcal{A}} \overset{\text{def}}{=} \hat{\mathbf{H}}_{\mathcal{A}} - \mathbf{H}$. According to [30], the vectorized error $\mathtt{vec}\{\Delta\hat{\mathbf{H}}_{\mathcal{A}}\}$ of the estimate (40) can be decomposed, as

$$(\mathbf{J}_{\mathcal{A}} \otimes \mathbf{I}_{N_R})\{\Delta_{\mathfrak{N}}(\mathcal{A}) + \Delta_{\hat{\mathbf{\Pi}}}(\mathcal{A})\} - \mathtt{vec}\{\mathbf{H}^{\perp}_{\mathcal{A}}\}, \quad (48)$$

[8]The complexity needed to obtain the $\ell1$ LS estimates $\hat{\mathbf{G}}^{\mathrm{LS}}_{\mathcal{A}}$ is dominated by $\mathcal{O}(W^3 N_T^3)$ in N_{AAD} iterations [17], [29].

[9]$\bar{\mathbf{M}}(l) \overset{\text{def}}{=} \mathbb{E}^{L}_{j=l}[\mathbf{M}(j)] \approx \{(L-1)\bar{\mathbf{M}}(l-1) + \mathbf{M}(l)\}/L$, is used to compute the CCMs of the PCA. However, for the ICA, an exact recursion: $\bar{\mathbf{M}}(l) = \{L\bar{\mathbf{M}}(l-1) - \mathbf{M}(l-L) + \mathbf{M}(l)\}/L$, may be utilized.

TABLE II
COMPLEXITY ORDER COMPARISON

Algorithm	Complexity order
$\ell 1\, i$ST	$\mathcal{O}(W^3N_T^3 + N_{\mathrm{AAD}}\{N_T^2W^2(W+N_R) + N_R^3WN_T\})$
$\ell 2\, i$ST	$\mathcal{O}(W^3N_T^3 + N_R^3WN_T)$
u-iST	$\mathcal{O}(W^3N_T^3 + N_R^3WN_T)$
$\ell 1\, p$ST	$\mathcal{O}(W^3N_T^3 + N_{\mathrm{AAD}}W^2N_T^2N_R^2)$
$\ell 2\, p$ST	$\mathcal{O}(W^3N_T^3 + W^2N_TN_R + WN_TN_R^2)$

TABLE III
COMPLEXITY DETAILS OF THE $\ell 1\, i$ST

Symbol	Eqn.	Complexity	Exec. Counts
$\hat{\mathbf{G}}_{\mathcal{A}}^{\mathrm{LS}}$	(18)	$\mathcal{O}(W^3N_T^3)$	1
$\mathbb{K}[\hat{\hat{\mathbf{G}}}_{\mathcal{A}_k}^{\mathrm{LS}}]$	(30)	$\mathcal{O}(N_T^2\{W^3 + W^2N_R\})$	N_{AAD}
$\hat{\mathbf{V}}_{\mathcal{A}_k}$	(23)	$\mathcal{O}(N_T \cdot W^3)$	N_{AAD}
$\hat{\hat{\mathbf{B}}}_{\mathrm{T},k}$	(25)	$\mathcal{O}(N_T \cdot r_{\mathrm{T}}^{\max}WN_R)$	N_{AAD}
$\mathbb{R}[\hat{\hat{\mathbf{b}}}_{\mathrm{T},k,i}]$	(28)	$\mathcal{O}(r_{\mathrm{T}}^{\max}N_T \cdot N_R^2)$	N_{AAD}
$\hat{\hat{\mathbf{\Phi}}}_{k,i}$	(28)	$\mathcal{O}(r_{\mathrm{T}}^{\max}N_T \cdot N_R^3)$	N_{AAD}
$\hat{\mathbf{G}}_{\mathcal{A}}$	(29)	$\mathcal{O}(N_T \cdot r_{\mathrm{T}}^{\max}\{W^2 + N_R^2\})$	N_{AAD}
$\mathcal{A}^{[n+1]}$	(42)	$\mathcal{O}(N_T \cdot W^3)$	N_{AAD}

where $\mathbf{H}_{\mathcal{A}}^{\perp}$ is the CIR unsupported with the active-set $\mathcal{A}$ and

$$\Delta_{\mathfrak{N}}(\mathcal{A}) = \hat{\hat{\mathbf{\Pi}}}(\mathcal{A})\,\mathtt{vec}\{\mathbf{\Delta}\hat{\hat{\mathbf{G}}}_{\mathcal{A}}^{\mathrm{LS}}\}, \tag{49}$$

$$\Delta_{\hat{\hat{\mathbf{\Pi}}}}(\mathcal{A}) = \left(\hat{\hat{\mathbf{\Pi}}}(\mathcal{A}) - \mathbf{I}_{|\mathcal{A}|N_R}\right)\mathtt{vec}\{\mathbf{G}_{\mathcal{A}}\}. \tag{50}$$

The projection matrix $\hat{\hat{\mathbf{\Pi}}}(\mathcal{A})$ is $\bigoplus_{k=1}^{N_T}\hat{\hat{\mathbf{\Pi}}}_k(\mathcal{A})$, where

$$\hat{\hat{\mathbf{\Pi}}}_k(\mathcal{A}) = \left[(\mathbf{Q}_{\mathcal{A}\,k,k}^{-*}\hat{\hat{\mathtt{V}}}_{\mathcal{A}_k}^{*}) \otimes \mathbf{\Gamma}^{-1/2}\right]\hat{\hat{\mathbf{\Phi}}}_k\left[\hat{\hat{\mathtt{V}}}_{\mathcal{A}_k}^{\mathsf{T}} \otimes \mathbf{I}_{N_R}\right].$$

Proposition 5 (Symbol-wise error variance bound)**.**

$$\mathtt{diag}\left\{\mathbb{K}[\Delta\hat{\mathbf{G}}_{\mathcal{A}_k}]\right\} \succeq \mathtt{diag}\left\{\mathbf{K}_{\Delta\hat{\mathbf{G}}_k}(\sigma_{\mathfrak{N}}^2, \mathcal{A}_k)\right\}. \tag{51}$$

Proof. According to (48),

$$\mathbb{K}\left[\Delta\hat{\mathbf{G}}_{\mathcal{A}}\right] \succeq \mathbb{K}\left[\mathtt{mat}_{N_R}\{\Delta_{\mathfrak{N}}(\mathcal{A})\}\right], \tag{52}$$

where $\mathbf{A} \succeq \mathbf{B}$ denotes that the residual $\mathbf{A} - \mathbf{B}$ is a positive semi-definite matrix and the operation $\mathtt{mat}_N\{\cdot\}$ performs inversion of the vectorization: $\mathtt{mat}_N\{\mathtt{vec}(\mathbf{X})\} = \mathbf{X} \in \mathbb{C}^{N\times M}$.

TABLE IV
COMPLEXITY DETAILS OF THE $\ell 1\, p$ST

Symbol	Eqn.	Complexity	Exec. Counts
$\hat{\mathbf{G}}_{\mathcal{A}}^{\mathrm{LS}}$	(18)	$\mathcal{O}(W^3N_T^3)$	1
$\mathbb{K}[\hat{\hat{\mathbf{G}}}_{\mathcal{A}_k}^{\mathrm{LS}}]$	(30)	$\mathcal{O}(N_T^2\{W^3 + W^2N_R\})$	N_{AAD}
$\hat{\mathbf{V}}_{\mathcal{A}_k}$	(23)	$\mathcal{O}(N_T \cdot W^3)$	N_{AAD}
$\mathbb{R}[\hat{\hat{\mathbf{G}}}_{\mathcal{A}_k}^{\mathrm{LS}}]$	(39)	$\mathcal{O}(W^2N_T^2N_R^2)$	N_{AAD}
$\hat{\hat{\mathbf{\Phi}}}_{\mathcal{A}_k}$	(5)	$\mathcal{O}(N_T \cdot N_R^3)$	N_{AAD}
$\hat{\mathbf{G}}_{\mathcal{A}}$	(29)	$\mathcal{O}(N_T \cdot \{W^3 + WN_R^2\})$	N_{AAD}
$\mathcal{A}^{[n+1]}$	(42)	$\mathcal{O}(N_T \cdot W^3)$	N_{AAD}

This is because $\Delta\hat{\mathbf{G}}_{\mathcal{A}}$ is dominated by

$$\mathtt{mat}_{N_R}\{\Delta_{\mathfrak{N}}(\mathcal{A})\} = \Delta\hat{\mathtt{B}}\left[\bigoplus_{k=1}^{N_T}\hat{\hat{\mathtt{V}}}_{\mathcal{A}_k}^{\mathsf{H}}\mathbf{Q}_{\mathcal{A}\,k,k}^{-\mathsf{H}}\right], \tag{53}$$

where we use (29) and $\Delta\hat{\mathtt{B}} = [\Delta\hat{\mathtt{B}}_1, \cdots, \Delta\hat{\mathtt{B}}_{N_T}]$ with

$$\Delta\hat{\mathtt{B}}_k = \mathbf{\Gamma}^{-1/2}\mathtt{mat}_{N_R}\left\{\hat{\hat{\mathbf{\Phi}}}_k\,\mathtt{vec}(\Delta\hat{\hat{\mathbf{G}}}_{\mathcal{A}}^{\mathrm{LS}}\,\hat{\hat{\mathtt{V}}}_{\mathcal{A}_k})\right\}.$$

Moreover, we have

$$\mathbb{K}\left[\Delta\hat{\mathtt{B}}\right] \succeq \bigoplus_{k=1}^{N_T}\mathbb{K}\left[\Delta\hat{\mathtt{B}}_k\right] = \bigoplus_{k=1}^{N_T}\sigma_{\mathfrak{N}}^2\mathbf{\Lambda}_{\mathrm{S},k}. \tag{54}$$

This is because $\mathbb{E}[\Delta\hat{\mathtt{b}}_{k,i}^{\mathsf{H}}\Delta\hat{\mathtt{b}}_{k,j}] = \sigma_{\mathfrak{N}}^2 r_{\mathrm{S},k,i}$ for $i = j$, otherwise 0, where $\Delta\hat{\mathtt{b}}_{k,i}$ is the i-th column vector of $\Delta\hat{\mathtt{B}}_k$. Therefore, (51) is obtained from (52), (53), and (54). □

Proposition 6 (MSE performance of the $\ell 1\, i$ST algorithm)**.**

$$\begin{aligned}\mathrm{MSE}(\hat{\mathbf{H}}_{\mathrm{ST}}^{\ell 1}) &= \mathtt{min}_{\mathcal{A}}\,\mathrm{MSE}(\hat{\mathbf{H}}_{\mathcal{A}} \mid \mathcal{A}) \\ &\geq \textstyle\sum_{k=1}^{N_T}\mu(\sigma_{\mathfrak{N}}^2, \mathbf{r}_{\mathrm{S},k}, r_{\mathrm{T},k}, \mathcal{A}_k^*) + \beta(\mathbf{r}_{\mathrm{S},k}, r_{\mathrm{T},k})\end{aligned} \tag{55}$$

with $\beta(\mathbf{r}_{\mathrm{S},k}, r_{\mathrm{T},k}) = \mathbb{E}[\|\mathbf{H}_k\|^2] - \sum_{i=1}^{r_{\mathrm{T},k}}\sum_{n=1}^{r_{\mathrm{S},k,i}}\lambda_n^2(\mathbb{R}[\mathbf{b}_{k,i}])$, *where the optimal active-set* $\mathcal{A}^*$ *is determined by*

$$\begin{aligned}\mathcal{A}^* &= \arg\min_{\mathcal{A}}\mathbb{E}[\|\mathbf{\Delta}\hat{\mathbf{H}}_{\mathcal{A}}\|^2] \\ &= \left\{j \mid d_j \geq m_j(\sigma_{\mathfrak{N}}^2, \mathcal{A}),\, j \in \mathcal{A}\right\}.\end{aligned} \tag{56}$$

The parameter d_j *is the* j*-th entry of the vector* $\mathtt{diag}\{\mathbb{K}[\mathbf{H}]\}$.

Proof. The MSE (55) is obtained from (48), (51) and (46), where the bias errors $\mathbb{E}[\|\Delta_{\hat{\hat{\mathbf{\Pi}}}}(\mathcal{A})\|^2] + \mathbb{E}[\|\mathbf{H}_{\mathcal{A}}^{\perp}\|^2]$ is lower-bounded by $\beta(\mathbf{r}_{\mathrm{S},k}, r_{\mathrm{T},k})$. We then consider (56). By (48),

$$\begin{aligned}&\mathbb{E}[\|\mathbf{\Delta}\hat{\mathbf{H}}_{\mathcal{A}}\|^2] \geq \mathbb{E}[\|\Delta\hat{\mathbf{G}}_{\mathcal{A}}\|^2] + \mathbb{E}[\|\mathbf{H}_{\mathcal{A}}^{\perp}\|^2] \\ &\quad = \mathbb{E}[\|\mathbf{H}\|^2] + \mathbf{1}_{WN_T}^{\mathsf{T}}\{\mathfrak{m}(\sigma_{\mathfrak{N}}^2, \mathcal{A}) - \mathbf{d}_{\mathbf{H}_{\mathcal{A}}}\},\end{aligned} \tag{57}$$

where $\mathbf{d}_{\mathbf{H}_{\mathcal{A}}} = \mathtt{diag}\{\mathbb{K}[\mathbf{H}_{\mathcal{A}}]\}$. Hence, we have

$$\mathcal{A}^* = \arg\max_{\mathcal{A}}\textstyle\sum_{j\in\mathcal{A}}\{d_j - m_j(\sigma_{\mathfrak{N}}^2, \mathcal{A})\},$$

which is equivalent to (56). □

The AAD (42) is derived from the optimization (56) by substituting the estimate $\hat{\mathbf{d}}_{\mathbf{H}} = \mathbf{d}_{\mathbf{H}} + \mathfrak{e}(\sigma_{\mathfrak{N}}^2)$ for $\mathbf{d}_{\mathbf{H}}$ and considering the error $\mathfrak{e}(\sigma_{\mathfrak{N}}^2)$ in the LHS of the inequality.

VI. NUMERICAL EXAMPLES

A. Simulation Setups

We assume $N_T \times N_R = 3 \times 24$ and 6×12 MIMO transmission scenarios in unknown interference channels. The interference is caused by unknown transmitters in the neighboring service areas, where they also communicate with their base station (BS) using the same MIMO system setup as that of the target user. CIRs are generated according to the SCM [22], where six path fading channel realizations based on the Vehicular-A (VA) model with a 30 km/h mobility and the PB model with a 3 km/h mobility are used in the urban micro cell scenario. The path positions of the VA and PB models are respectively set at $\{1.0, 3.2, 6.0, 8.6, 13.1, 18.6\} + \Delta_{\mathrm{synch}}$ and $\{1.0, 2.4, 6.6, 9.4, 17.1, 26.9\} + \Delta_{\mathrm{synch}}$ symbol timings assuming that the TX bandwidth is 7 MHz with a carrier frequency of 5 GHz. The timing offset Δ_{synch} is set at 0 for the target user, whereas it is chosen randomly from the range

$[0.0, 3.0]$ for the unknown users.[10] The maximum CIR length W is, hence, set at 31. The TTI Δ_{T} is set at 2 slots.

Note that, as discussed in [16], the CIR matrix $\mathbf{H}_k(l)$ following (2) can be obtained by re-sampling[11] row-vectors of the complex channel gain $\mathbf{B}_k(l)$. Concretely, the matrix $\mathbf{B}_k(l)$ is generated with the SCM implementation [31] and the raised cosine filter of roll-off 0.3 is used to take account of the pulse shaping and the propagation delay of the multipath models. Moreover, the antenna element spacing at the BS and the mobile station (MS) are set at 0.5 wavelength. The angle θ_{MS} between the BS-MS and the MS broadside [22] is fixed at 125° for the target user. However, the angle θ_{MS} is chosen randomly from $[-180°, 90°]$ for the interference sources.

B. Channel Estimation Performance as SNR varies

The algorithms are verified with normalized MSE (NMSE) performance, where $\mathrm{NMSE}(\hat{\mathbf{H}}) \overset{\mathrm{def}}{=} \mathbb{E}[\|\hat{\mathbf{H}}-\mathbf{H}\|^2]/\mathbb{E}[\|\mathbf{H}\|^2]$ for estimates $\hat{\mathbf{H}}$. In this subsection, SIR is set at a certain value, where we define the SIR and SNR by $\mathbb{E}[\|\mathbf{HX}\|^2]/\mathbb{E}[\|\mathbf{Z}\|^2]$ and $\mathbb{E}[\|\mathbf{HX}\|^2]/\mathbb{E}[\|\mathbf{N}\|^2]$, respectively. The target user transmits signals over an unknown interference user, where the PB and VA models are assumed for the target and interference users, respectively. We refer to this TX scenario as PB/VA. The parameters in the AAD are set at $(N_{\mathrm{AAD}}, E) = (3, \lfloor 0.8W \rfloor)$. The input parameter $r_{\mathbf{d_H}}$ of Algorithm 3 is set at 1.

1) ST vs. Temporal-Only: Fig. 3 shows NMSE performance of the $\ell 2\,p$ST in the 6×12 MIMO system, where a TS of $L_t = 255$ symbols is generated with the Gold sequence. The sliding window length L_{S} in (17) is set at 50. A benchmark referred to as normalized adaptive-CRB (NaCRB) is also presented, where it is defined by normalizing (55) with $\mathbb{E}[\|\mathbf{H}\|]^2$.

As depicted in Fig. 3, the $\ell 2\,p$ST obtains a significant NMSE gain over an approach using the temporal subspace only ($\ell 2$ T-only) [26] at $\mathrm{SIR} = 4\,\mathrm{dB}$. However, the NMSE gain decreases as SNR increases when there is no unknown interference. This observation confirms that, according to the property (10), the joint ST subspace-based estimators obtain the rank reduction gains in a low to moderate SINR regime.

2) iST vs. pST: Fig. 3 shows NMSE of the $\ell 2\,i$ST technique, where the constant $r_{\mathrm{T}}^{\max}$ in (25) is set at $\lfloor 0.5W \rfloor$. The $\ell 2\,i$ST improves NMSE performance significantly over that of the $\ell 2\,p$ST at the moderate $\mathrm{SIR} = 4\,\mathrm{dB}$. This is because the $\ell 2\,p$ST technique has difficulty to separate unknown interferences from the signal of interest, whereas the $\ell 2\,i$ST improves the problem based on Proposition 1. However, the NMSE gain decreases in the negative SNR regime. This is because the CCM (28) of the ICA is computed from $1/W$ times fewer samples than that (39) of the PCA.

3) NMSE details of the iST methods: We verify the NMSE performance of the iST algorithms according to (48). As shown in Fig. 4(a), the $\ell 2\,i$ST estimates the average-rank $\bar{r}_{\mathrm{ST}}$ accurate enough compared with the *true* adaptive-rank based on Definition 1, where $\bar{r}_{\mathrm{ST}} \overset{\mathrm{def}}{=} \sum_{k=1}^{N_T} \sum_{i=1}^{r_{\mathrm{T},k}} r_{\mathrm{I},k,i}/N_T$. For a constant SIR, the interference-to-noise ratio (INR) is proportional to the SNR, which incurs an approximation error of

[10]The interference terminals are not always synchronized to the receiver.

[11]We do not compute the response matrix $\mathbf{E}_k$ in (2) to generate CIRs.

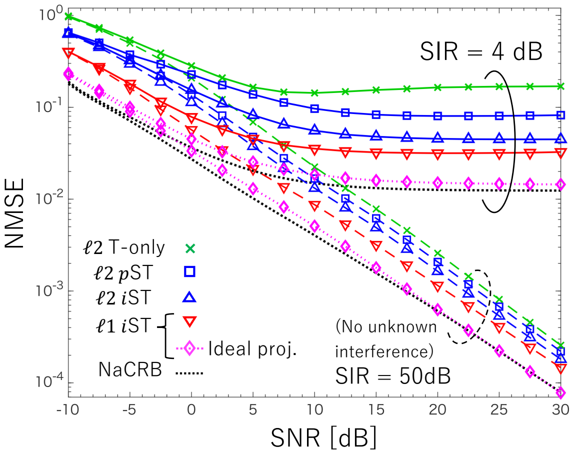
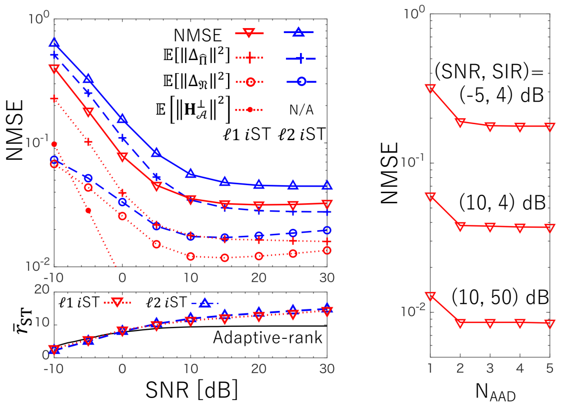


Fig. 3. NMSE performance in the PB/VA scenario.



(a) NMSE details at SIR = 4dB. (b) NMSE vs. N_{AAD}.

Fig. 4. NMSE details (a) and convergence property (b) of the $\ell 1\,i$ST.

(14). Hence, we observe from Fig. 4(a) that the rank estimates and the residual (49) are degraded in a high SNR regime.

However, in a low SNR regime, the $\ell 2\,i$ST suffers from the projection error (50) significantly. This is because, as shown in (50), the projection error can be increased for a redundant active-set $\mathcal{A}$. The $\ell 1\,i$ST improves the projection error by iteratively pruning the active-set. As observed from Fig. 4(b), the AAD (42) gets the NMSE performance converged in three iterations. Of course, the $\ell 1$ regularization does not completely suppress unknown interference in the CCM. As shown in Fig. 3, the $\ell 1\,i$ST asymptotically achieve the NaCRB if the subspace projection is estimated ideally given $\mathbf{H}$.

C. Channel Estimation Performance as SIR varies

In this subsection, channel estimation algorithms are performed in the 3×24 MIMO system using the TS of $L_t = 127$ symbols. $L_{\mathrm{S}} = 100$ is assumed. The CIRs follow the VA and PB models for the target user and two unknown interference users, receptively. We refer to the propagation scenario as VA/$\{\mathrm{PB}, \mathrm{PB}\}$ hereafter.

1) $\ell 1\,p$ST vs. $\ell 1\,i$ST: The $\ell 1\,p$ST is expected to improve the performance under unknown interference by leveraging the $\ell 1$ regularization as well as the $\ell 1\,i$ST. However, as shown in Fig. 5, the $\ell 1\,p$ST does not outperform the $\ell 2\,i$ST algorithm

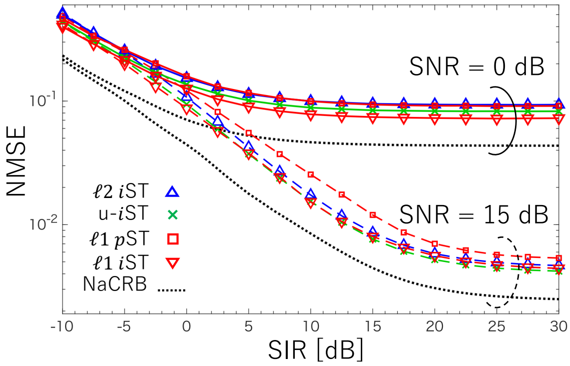


Fig. 5. NMSE performance in the VA/{PB, PB} scenario.

even in a low SINR regime. This is because Proposition 1 holds in the SCM-based channel realizations.

2) u-iST vs. $\ell1\,i$ST: In a high SIR regime of SNR = 15 dB, the conventional u-iST [16] obtains equivalent NMSE performance[12] to that of the $\ell1\,i$ST, since it is also the $\ell1$ ST subspace-based channel estimation (12). However, at SNR = 0 dB, the u-iST technique is inferior to the $\ell1\,i$ST in the entire SIR regime. This is because, in a low to moderate SINR regime, the u-iST inherits the projection accuracy problem, since it approximates the subspace projection with that obtained by the $\ell2\,i$ST.

D. Comparison with Conventional Techniques

1) $\ell1$ LS techniques: Fig. 6 shows NMSE performance of the conventional adaptive structured subspace pursuit (ASSP) technique [7] in the 6×12 MIMO system. The PB/VA scenario is assumed. INR is set at 0 dB. The TTI Δ_{T} is set at 1 slot in this subsection. As shown in Fig. 6, the ASSP outperforms the OMP since it improves accuracy of the active-set detection by using the maximum correlation in the past L_{ASSP} slots, where the optimal L_{ASSP} is chosen for 10. The stopping condition of the ASSP iteration is modified to inspect for each TX stream so that it approximately achieves the $\ell1$ LS NMSE bound.[13] However, the ASSP does not achieve the NaCRB since it does not consider the spatial subspace.

2) $\ell1$ LS + iST-subspace method: The $\ell1\,i$ST may be composed of an arbitrary $\ell1$ LS and the conditional $\ell1$ MMSE channel estimation techniques. However, such a naive extension does not always achieve the optimal performance. Fig. 6 shows the NMSE performance of an iST subspace-based compressive channel estimation constructed with the ASSP and (15). We refer to it as ASSP + iST, hereafter. The ASSP + iST does not outperform the $\ell2\,i$ST, although the ASSP follows the $\ell1$ LS NMSE bound in the MIMO channels.

Fig. 7 details the NMSE performance of the ASSP + iST, where SNR is set at 6 dB. It is observed from Fig. 7 that the NMSE performance of the ASSP + iST diverges from that of the $\ell1\,i$ST as SIR decreases. We find that, in both algorithms, the iST-subspace analysis is performed as expected since the average ranks $\bar{r}_{\mathrm{ST}}$ are estimated relevantly. The residual terms $\mathbb{E}[\|\Delta_{\mathfrak{N}}\|^2]$ (49) and the projection errors $\mathbb{E}[\|\Delta_{\hat{\mathbf{\Pi}}}\|^2]$ (50) are also computed equivalently in the two algorithms. We can find from Fig. 7 that, however, the ASSP + iST increases the bias error $\mathbb{E}[\|\mathbf{H}_{\mathcal{A}}^{\perp}\|^2]$ in the low SIR regime. This is because the ASSP + iST selects the active-set under the $\ell1$ LS criterion, which can deteriorates accuracy as the $\ell1$ MMSE estimates when channels are not exactly sparse.

3) Frequency domain CS-based estimation: The simultaneous weighted OMP (SW-OMP) algorithm [11] can be performed by transforming the received TS matrix after the spatial whitening into the frequency domain:

$$\mathbf{\Gamma}^{\frac{1}{2}}\mathbf{Y}(l)\cdot\mathbf{F}^{\mathsf{T}} \;=\; \mathbf{\Gamma}^{\frac{1}{2}}\mathbf{H}_F(l)\mathbf{\Lambda}_p(l)+\mathbf{\Gamma}^{\frac{1}{2}}\mathfrak{N}(l)\mathbf{F}^{\mathsf{T}},$$

where $\mathbf{F}=\frac{1}{\sqrt{\tilde{L}_t}}[\{\exp(-\frac{2\pi}{\tilde{L}_t}(i-1)(j-1)\sqrt{-1})\}_{i,j}]$ is the $\tilde{L}_t\times\tilde{L}_t$ discrete Fourier transform (DFT) matrix. We denote $[\{a(i,j)\}_{i,j}]$ for a matrix whose (i,j)-th entry is $a(i,j)$. The matrix $\mathbf{H}_F(l)$ is $[\mathbf{H}_1^o(l),\cdots,\mathbf{H}_{N_T}^o(l)](\mathbf{1}_{N_T}\otimes\mathbf{F}^{\mathrm{T}})$ with $\mathbf{H}_k^o(l)=[\mathbf{H}_k(l),\mathbf{O}_{N_R\times(\tilde{L}_t-W)}]$ for $\forall k\in\{1:N_T\}$. The $N_T\tilde{L}_t\times\tilde{L}_t$ block diagonal matrix $\mathbf{\Lambda}_p(l)$ is composed by stacking $\mathtt{D}_{\mathtt{IAG}}\{\mathbf{p}_k(l)\}$ with $\mathbf{p}_k=\mathtt{diag}\{\mathbf{F}\mathbf{X}_{c,k}^{\mathsf{T}}(l)\mathbf{F}^{\mathsf{H}}\}$, where the circulant matrix $\mathbf{X}_{c,k}(l)$ has $[\mathbf{x}_k^{\mathsf{T}}(l),\mathbf{0}_{W-1}^{\mathsf{T}}]^{\mathrm{T}}$ on the first column. In the SW-OMP algorithm, we estimate[14] channel parameters $\mathtt{H}_f(l)=\mathbf{H}_F(l)|_{\{f+\tilde{L}_t(n-1)|n\in\{1:N_T\}\}}$ for the f-th bin. Let the coherent time be longer than L_M slots,

$$\mathtt{H}_f(j)\approx\mathbf{A}_R\mathtt{H}_{f,l}^a\mathbf{A}_T^{\mathsf{H}}\quad(\forall j\in\{l:L_M\}) \tag{58}$$

is used, where $\mathtt{H}_{f,l}^a$ is a $G_R\times G_T$ complex gain matrix in the quantized angular domain representation. The quantized receive antenna response matrix $\mathbf{A}_R$ is given by $\frac{1}{\sqrt{N_R}}\left[\{\exp(-(i-1)\sqrt{-1}\frac{2\pi}{\lambda}d_R\cos\theta_{i,j})\}_{i,j}\right]$ for $i\in\{1:N_R\}$ and $j\in\{1:G_R\}$, where $\cos\theta_{i,j}=\frac{2}{G_R}(j-1)-1$. λ and d_R are the signal wavelength and the antenna element spacing. The quantized transmit antenna response matrix $\mathbf{A}_T\in\mathbb{C}^{N_T\times G_T}$ is defined similarly.

We assume *oracle* stopping criterion given $\mathbf{H}$ so that the MSE of the time domain estimate is minimized. The resolution and coherent time parameters are set at $(G_R,G_T)=(36,12)$ and $L_M=10$ slots. As shown in Fig. 6, the SW-OMP outperforms the ASSP in a low to moderate SINR regime by using the spatial sparse property of the gain matrix $\mathtt{H}_{f,l}^a$ in (58). However, the SW-OMP exhibits an MSE floor in a high SINR regime since the approximation (58) is inaccurate in this scenario. Moreover, the SW-OMP is inferior to the $\ell2\,i$ST in a moderate SINR regime, although the oracle stopping criterion is assumed. According to our experiment, increasing the resolution (G_R,G_T) improves the problem insignificantly.

[12]Technically, the u-iST can achieve slightly better NMSE than the $\ell1\,i$ST in a high SINR regime, since the bias error of the u-iST can be less than that of the $\ell1\,i$ST [16]: $\mathbb{E}[\|\hat{\tilde{\mathbf{\Pi}}}(\mathcal{A}^{[0]})\,\mathtt{vec}\{\mathbf{H}_{\mathcal{A}}^{\perp}\}\|^2]\le\mathbb{E}[\|\mathbf{H}_{\mathcal{A}}^{\perp}\|^2]$ with $\mathcal{A}^{[0]}=\{1:WN_T\}$ holds when the subspace projection is accurate enough.

[13]It can be defined by (57) assuming known active-sets and $\hat{\tilde{\mathbf{\Pi}}}=\mathbf{I}$.

[14]For the f-th bin, $\mathbf{y}_f(l)\overset{\mathrm{def}}{=}(\mathbf{\Gamma}^{\frac{1}{2}}\mathbf{Y}(l)\cdot\mathbf{F}^{\mathsf{T}})|_f=\mathbf{\Gamma}^{\frac{1}{2}}\mathtt{H}_f(l)\mathbf{q}_f(l)+\tilde{\mathfrak{N}}|_f(l)$, where $\mathbf{q}_f(l)=[\mathbf{p}_1(l)|_f,\cdots,\mathbf{p}_{N_T}(l)|_f]^{\mathsf{T}}$. By (58), we collect L_M observations: $\mathtt{Y}_{f,l}=[\mathbf{y}_f(l),\cdots,\mathbf{y}_f(L_M)]=\mathbf{\Gamma}^{\frac{1}{2}}\mathbf{A}_R\mathtt{H}_{f,l}^a\mathbf{A}_T^{\mathsf{H}}\mathbf{Q}_f(l)+\tilde{\mathtt{N}}_f(l)$, where $\mathbf{Q}_f(l)=[\mathbf{q}_f(l),\cdots,\mathbf{q}_f(L_M)]$, and the noise matrix $\tilde{\mathtt{N}}_f(l)$ is defined similarly. We may obtain a compressed estimate of $\mathtt{h}_{f,l}^a=\mathtt{vec}\{\mathtt{H}_{f,l}^a\}$ as $\hat{\mathtt{h}}_{f,l}^a|_{\mathcal{A}}=(\mathbf{\Upsilon}_{\mathcal{A}}\mathbf{\Upsilon}_{\mathcal{A}}^{\mathsf{H}}+\frac{N_R}{G_RG_T}\sigma_{\mathfrak{N}}^2\mathbf{I}_{G_RG_T})^{-1}\mathbf{\Upsilon}_{\mathcal{A}}^{\mathsf{H}}\mathtt{vec}\{\mathtt{Y}_{f,l}\}$ with $\mathbf{\Upsilon}_{\mathcal{A}}=\{(\mathbf{Q}_f^{\mathsf{T}}(l)\otimes\mathbf{\Gamma}^{\frac{1}{2}})(\mathbf{A}_T^*\otimes\mathbf{A}_R))\}|_{\mathcal{A}}$, where $\mathcal{A}$ is determined by the SW-OMP.

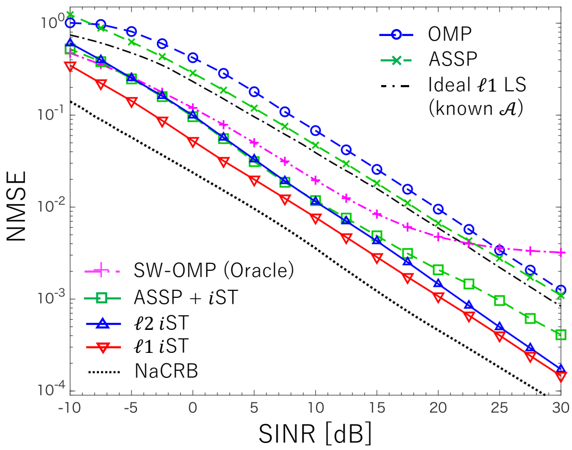


Fig. 6. NMSE performance in the 6×12 MIMO.

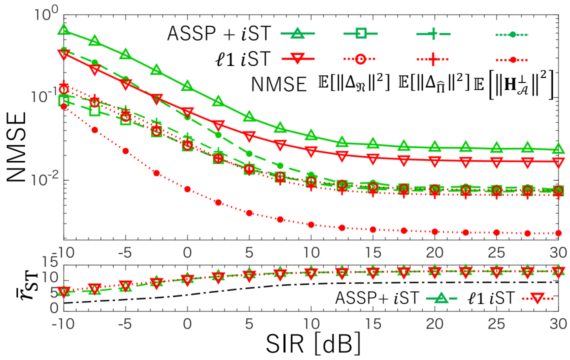


Fig. 7. NMSE details at SNR = 6 dB. The *true* adaptive-rank is depicted with the dash-dot line in the bottom subfigure.

This is because, unlike the iST methods cancel co-streams in (24), the SW-OMP does not solve the MAI problem directly.

E. BER Performance

1) BER in Low SINR: Fig. 8 shows BER performance in the 3×24 MIMO system. An $L_b = 2048$ bit binary data sequence is turbo encoded with rate $R_c = 1/3$ by using the transfer function $[1, g_1/g_0]$ with $(g_0, g_1) = (13, 15)_8$. The channel-encoded sequence is mapped to the binary phase shift keyed (BPSK) symbols and is modulated to $N_T R_c L_b / N_c$ orthogonal frequency division multiplexing (OFDM)-symbols with $N_c = 1024$ subcarriers. We transmit an OFDM-symbol followed by the TS section. The VA/{PB,PB} scenario is assumed at SNR = 0 dB.

As shown in Fig. 8, the receiver achieves $\text{BER} = 10^{-5}$ at SIR = -10 dB if CIR matrix $\mathbf{H}$ is known perfectly. We can find from Fig. 5 that, however, the NMSE of channel estimates is greater than 10^{-1} in the negative SIR regime when SNR is set at 0 dB. Hence, the receiver using actual channel estimates has a large BER performance gap from the ideal receiver. For example, the receiver using the ASSP estimation does not achieve $\text{BER} = 10^{-5}$ in the negative SIR region. The receiver in a large-scale MIMO system is, hence, necessary to jointly utilize the CS and ST subspace-based approaches. As shown in Fig. 8, the receiver using the $\ell 1\, p$ST improves BER more than 3 dB over that of the ASSP.

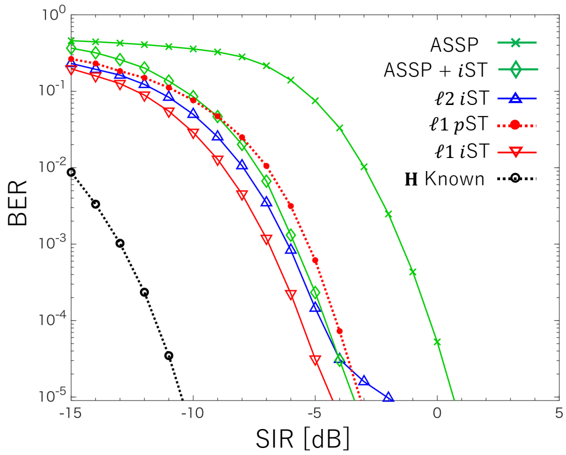


Fig. 8. BER performance in the 3×24 MIMO. SNR is set at 0 dB.

According to Proposition 1, the iST methods are expected to improve receiver performance. However, it is observed from Fig. 8 that the receiver with the $\ell 2\, i$ST suffers from a BER floor due to the subspace projection error. The projection accuracy can be improved by using $\ell 1$ regularization since the CIRs are observed as asymptotically sparse parameters in the low SIR regime. As observed from Fig. 8, the receiver with the ASSP+iST technique solves the BER floor. By using the proposed $\ell 1\, i$ST algorithm, the receiver further improves the BER performance by 1 dB over that of the ASSP+iST. This is because the AAD algorithm optimally selects the active-set by taking the ST-subspace into account.

2) BER with Interference Variation: We verify the proposed algorithms in a practical situation where the unknown interference changes abruptly. The variation of interference is set at $\overline{\text{SIR}} \pm 5$ dB in every 100 slot-interval, where the $\overline{\text{SIR}}$ denotes the average of an SIR configuration. SNR is fixed at 6 dB. The PB/VA scenario is used in 6×12 MIMO channels.

As shown in Fig. 9, the receiver using the $\ell 2\, i$ST exhibits a BER floor even in a moderate SINR regime. This is because the $\ell 2\, i$ST suffers from the projection error due to the abrupt SIR variation. As above-mentioned, CS-based techniques can improve the projection accuracy problem. However, as observed from Fig. 9, the ASSP+iST technique does not outperform the $\ell 2\, i$ST. This is because the CIRs generated from the PB model are not observed as exactly sparse channels in a moderate to high SINR regime.

The AAD algorithm can improve BER performance even in the approximately sparse PB channels, since it is designed to adaptively minimize the MSE performance according to the sparsity of the interested signals. The conventional u-iST performs the EM algorithm {(15), (16)} by using the $\ell 2\, i$ST and the AAD, respectively. Nevertheless, the u-iST technique does not completely solve the BER floor problem since it inherits the projection accuracy problem from the $\ell 2\, i$ST.

Hence, it is necessary to exactly perform the EM algorithm. As shown in Fig. 9, the receiver with the $\ell 1\, p$ST improves receiver performance at $\text{BER} = 10^{-5}$ over that of the $\ell 2\, i$ST, since it performs the E-step (15) correctly. According to Proposition 1, the receiver using the $\ell 1\, i$ST algorithm further

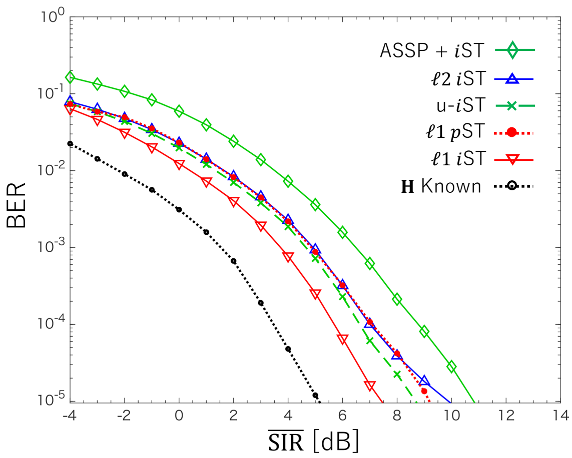


Fig. 9. BER performance in abrupt interference variation. The 6×12 MIMO transmission over the PB/VA scenario is assumed at SNR = 6 dB.

improves BER than that of the $\ell 1\,p$ST, and it achieves a BER gain of 2.5 dB in $\overline{\text{SIR}}$ over that of the $\ell 2\,i$ST.

VII. Conclusions

The ordinary $\ell 2\,i$ST channel estimation technique based on the ICA improves MSE performance over the $\ell 2\,p$ST using the PCA. However, a receiver using the $\ell 2\,i$ST can suffer from a BER floor problem in unknown interference MIMO channels. As a solution to the problem, we proposed a new $\ell 1\,i$ST algorithm which suppresses unknown interference in the CCM by leveraging the temporally sparse property of the observed CIRs. Hence, the $\ell 1\,i$ST accurately performs a spatial-ICA for independent path components.

Simulation results shown in this paper verify that the proposed algorithm achieves a significant MSE gain by using the AAD algorithm that detects the active-sets optimally. Hence, a receiver using the $\ell 1\,i$ST channel estimation solves the BER floor problem in unknown interference MIMO channels.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers JP17K06423, JP19K11963, and JP17K00184. The work of Y. Takano was supported in part by the Telecommunications Advancement Foundation. The work of H.-J. Su was supported in part by the Ministry of Science and Technology (MOST), Taiwan, under grants 106-2923-E-002-015-MY3, 108-2221-E-002-034-MY2 and 108-2218-E-002-060. Authors would like to acknowledge Prof. T. Matsumoto at Japan Advanced Institute of Science and Technology (JAIST) for technical advice, and we would like to appreciate the Associate Editor Prof. B. Clerckx and the anonymous reviewers for their helpful comments and suggestions.

References

[1] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the k-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug 2008.

[2] 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.

[3] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, 2016.

[4] A. Garcia-Rodriguez, G. Geraci, L. G. Giordano, A. Bonfante, M. Ding, and D. Lopez-Perez, "Massive MIMO unlicensed: A new approach to dynamic spectrum access," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 186–192, June 2018.

[5] P. H. Lin, S. H. Lai, S. C. Lin, and H. J. Su, "On secrecy rate of the generalized artificial-noise assisted secure beamforming for wiretap channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1728–1740, September 2013.

[6] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec 2015.

[7] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, "Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 601–617, Feb 2016.

[8] J. Lee, G. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, June 2016.

[9] S. Srivastava, A. Mishra, A. Rajoriya, A. K. Jagannatham, and G. Ascheid, "Quasi-static and time-selective channel estimation for block-sparse millimeter wave hybrid MIMO systems: Sparse bayesian learning (SBL) based approaches," *IEEE Trans. Signal Process.*, vol. 67, no. 5, pp. 1251–1266, March 2019.

[10] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, June 2016.

[11] J. Rodrìguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, May 2018.

[12] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge university press, 2005.

[13] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[14] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[15] J. Banks, C. Moore, R. Vershynin, N. Verzelen, and J. Xu, "Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 4872–4894, July 2018.

[16] Y. Takano, H. Su, M. Juntti, and T. Matsumoto, "A conditional $\ell 1$ regularized MMSE channel estimation technique for IBI channels," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6720–6734, Oct 2018.

[17] 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.*, vol. 15, no. 1, pp. 314–328, Jan 2016.

[18] P. Schniter and E. Byrne, "Adaptive detection of structured signals in low-rank interference," *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3439–3454, July 2019.

[19] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," *IEEE J. Sel. Areas Commun.*, vol. 4, no. 2, pp. 255–271, April 2010.

[20] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, "Estimation of sparse MIMO channels with common support," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3705–3716, Dec 2012.

[21] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse bayesian learning," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3591–3603, Jul 2014.

[22] European Telecommunications Standards Institute (ETSI), "Spatial channel model for MIMO simulations (3GPP TR 25.996)," Jan. 2017.

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

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.

[25] N. Sugiura, "Further analysis of the data by Akaike' s information criterion and the finite corrections," *Communications in Statistics - Theory and Methods*, vol. 7, no. 1, pp. 13–26, 1978.

[26] M. Nicoli, O. Simeone, and U. Spagnolini, "Multislot estimation of frequency-selective fast-varying channels," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1337–1347, Aug 2003.

[27] ——, "Multislot estimation of fast-varying space-time communication channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1184 – 1195, may 2003.

[28] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387 – 392, Apr. 1985.

[29] S. Van Vaerenbergh, I. Santamaria, W. Liu, and J. Principe, "Fixed-budget kernel recursive least-squares," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, March 2010, pp. 1882–1885.

[30] Y. Takano, M. Juntti, and T. Matsumoto, "Performance of an $\ell 1$ regularized subspace-based MIMO channel estimation with random sequences," *IEEE Wireless Commun. Lett.*, vol. 5, no. 1, pp. 112–115, Feb 2016.

[31] J. Salo, G. Del Galdo, J. Salmi, P. Kyosti, M. Milojevic, D. Laselva, and C. Schneider, "MATLAB implementation of the 3GPP spatial channel model (3GPP TR 25.996)," On-line, Jan. 2005, http://www.ist-winner.org/3gpp_scm.html.

Yasuhiro Takano (S'11-M'16) received his Ph.D. and Dr.Sc. (Tech.) degrees, respectively, from Japan Advanced Institute of Science and Technology (JAIST), Japan, and University of Oulu, Finland, in 2016. He is currently with Kobe University, Japan. His research interests include signal processing for wireless communications.

Hsuan-Jung Su (S'90-M'99-SM'12) received the B.S. degree in electronics engineering from National Chiao Tung University, Taiwan, in 1992, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1996 and 1999, respectively. From 1999 to 2000, he was a Post-Doctoral Research Associate with the Institute for Systems Research, University of Maryland. From 2000 to 2003, he was with Bell Laboratories, Lucent Technologies, Holmdel, NJ, USA, where he received the Central Bell Labs Teamwork Award in 2002 and the Bell Labs President's Gold Award in 2003 for his contribution to the 3G wireless network design and standardization. In 2003, he joined the Department of Electrical Engineering, National Taiwan University, where he is currently a Professor and the Director of the Graduate Institute of Communication Engineering. From 2014 to 2015, he was a Visiting Fellow with Princeton University. His research interests cover coding, modulation, signal processing, interference management, resource allocation, and MAC protocols of wireless communication, cognitive, M2M (IoT), and D2D networks. He has also served on the organizing committees and TPCs of many international conferences, including serving as the Finance Chair of the IEEE ICASSP 2009, the Finance Co-Chair and a TPC Track Chair of the IEEE VTC 2010 Spring, a TPC Co-Chair of WPMC 2012, a TPC Co-Chair of IEEE GreenCom 2014, the TPC Chair of WOCC 2015, and a TPC Co-Chair of IEEE GLOBECOM 2020. He was the Chair of the Taipei Chapter of the IEEE Information Theory Society from 2013 to 2015, the Secretary and Treasurer from 2014 to 2015, the Vice Chair of the Technical Affairs Committee from 2016 to 2017, and the Vice Chair of the Membership Development Committee of the IEEE Communications Society Asia-Pacific Board from 2018 to 2019. He is an Area Editor of the *Physical Communication (PHYCOM)* journal (Elsevier). He has guest-edited special issues for journals, such as *IEEE Access*.

Yoshiaki Shiraishi (S'98-M'00) received his B.E. and M.E. degrees from Ehime University, Japan, and the Ph.D. degree from the University of Tokushima, Japan, in 1995, 1997, and 2000, respectively. From 2002 to 2006 he was a lecturer at the Department of Informatics, Kindai University, Japan. From 2006 to 2013 he was an associate professor at the Department of Computer Science and Engineering, Nagoya Institute of Technology, Japan. Since 2013, he has been an associate professor at the Department of Electrical and Electronic Engineering, Kobe University, Japan. His current research interests include information security, cryptography, computer networks, and machine learning-based cyber attack analysis. He received the 20th anniversary award of the Symposium on Cryptography and Information Security (SCIS) and the SCIS paper award from the Information Security (ISEC) group of IEICE in 2003 and 2006, respectively. He received an excellent paper award from the Special Interest Group of Intelligent Transport Systems (SIG-ITS) of the Information Processing Society of Japan (IPSJ) in 2015. He is a member of ACM, and a senior member of IPSJ and IEICE.

Masakatu Morii (M'89) received his B.E. degree in electrical engineering and theM.E. degree in electronics engineering from Saga University, Saga, Japan, and the D.E. degree in communication engineering from Osaka University, Osaka, Japan, in 1983, 1985, and 1989, respectively. From 1989 to 1990 he was an Instructor in the Department of Electronics and Information Science, Kyoto Institute of Technology, Japan. From 1990 to 1995 he was an Associate Professor at the Department of Computer Science, Faculty of Engineering, Ehime University, Japan. From 1995 to 2005 he was a Professor at the Department of Intelligent Systems and Information Science, Faculty of Engineering, the University of Tokushima, Japan. Since 2005, he has been a Professor at the Department of Electrical and Electronic Engineering, Faculty of Engineering, Kobe University, Japan. His research interests are in error correcting codes, cryptography, discrete mathematics, and computer networks and information security. He is a fellow of IEICE.