Iterative Channel Estimation in MIMO Antenna Selection Systems for Correlated Gauss-Markov Channel

Yousuke NARUSE(a), Student Member and Jun-ichi TAKADA†, Member

SUMMARY We address the issue of MIMO channel estimation with the aid of a priori temporal correlation statistics of the channel as well as the spatial correlation. The temporal correlations are incorporated to the estimation scheme by assuming the Gauss-Markov channel model. Under the MMSE criteria, the Kalman filter performs an iterative optimal estimation. To take advantage of the enhanced estimation capability, we focus on the problem of channel estimation from a partial channel measurement in the MIMO antenna selection system. We discuss the optimal training sequence design, and also the optimal antenna subset selection for channel measurement based on the statistics. In a highly correlated channel, the estimation works even when the measurements from some antenna elements are omitted at each fading block.

key words: MIMO, channel estimation, Gauss-Markov model, Kalman filter, MIMO antenna selection system, optimal training symbol design

1. Introduction

Recently, MIMO transmission is attracting a lot of attention since it achieves much higher transmission capability by utilizing the spatial freedom of the waves in radio propagation [1], [2]. In MIMO systems, precision of the estimate of the channel state information (CSI) dramatically affects the performance of transmission because the estimation error leads to a decrease of SINR due to the degradation of the separation capability of each signal on the spatial freedom. Generally in pilot symbol aided channel estimation with a constant noise power, the more energy we put into the transmission of training symbols, the lower the estimation error can be. Especially for rapidly fading channels where the channel estimation must be frequently repeated in certain intervals to track the temporal changes of the channel, required energy and time for channel estimation consume a lot of resources of the wireless communication system. In order to estimate channel precisely using a limited set of resources, optimal design of training symbol have been discussed by fully exploiting a priori channel statistics about the spatial correlation [3]. A MIMO channel with $N_{Tx}$ transmit and $N_{Rx}$ receive antenna possesses $N_{Tx}N_{Rx}$ independent freedoms to be estimated. In pilot symbol aided estimation, certain powers are allocated and transmitted for each freedom in order to perform their observation. Without the channel statistics, we have no choice but to treat them equally, which is called the orthogonal training sequence. On the other hand, if the channel is spatially correlated, each freedom of the channel has different variance, and then the one with larger variance is considered to be a more important unknown which contributes more to estimation capability. In order to estimate such important unknowns by priority, optimal training symbols are designed such that the power in the training symbols is concentrated on the important unknown of the channel matrix. Physically, this can be interpreted as the energy of the training sequence is focused on the significant scatterer in the propagation environment [3].

In addition to the spatial correlation, by exploiting the nature of temporal correlation of the channel, a more effective estimation can be realized. A stochastic model for time-varying channel is required in order to incorporate the temporal correlation into the estimation scheme, and autoregressive (AR) modelling is widely accepted among these studies [4]–[7]. In such context, MMSE estimate of the channel can be performed iteratively with the Kalman filter [20]. It should be noted that for the estimation of time-variant frequency-selective fading channels, Kalman-based methods are quite common in the literatures [7], [8].

In this paper, we extend the optimal design of training sequence to the Kalman-based estimation methods. The optimal training sequence of this case is determined respectively for each fading block. Here, the allocation of power for each freedom is determined similarly, but the variances referred are determined by accumulating the estimation errors occurred so far, based on the model of temporal channel transition.

Also we investigate the optimal partial measurements for MIMO antenna selection systems which can be discussed in the similar way as the optimal training sequence design. If the temporal and spatial correlations are high enough, we can partially omit the observation of the channel while satisfying the desired precision of the estimation. This approach might be useful especially in situations where measurement of the full channel state costs much.

In MIMO antenna selection systems with $N_{RF}$ RF chains and $M$ antenna elements, since the number of antenna elements is larger than the number of RF chains, it requires at least $\lceil M/N_{RF} \rceil$ times measurements to cover all the antenna elements by selecting connected subsets of elements sequentially, where $\lceil x \rceil$ means the smallest integer not exceeded by $x$. With presence of highly correlated channel, measurements for all the antenna elements can be avoided by interpolating the not measured portion of the full channel matrix with the aid of a priori correlation statistics. This can
reduce the repeated measurements in the antenna selection systems. In this paper, the method is formulated to replace the repeated measurements with a single measurement by utilizing the Kalman filter.

This paper is organized as follows. In Sect. 2, the system model including the MIMO antenna selection is described. Also the stochastic channel model considering spatial and temporal correlation is explained. Section 3 is devoted to explain a channel estimation using the Kalman filter, and Sect. 4 discusses the optimal training symbol design and measurement antenna selection to enhance the estimation capability. The effectiveness of the proposed method is verified by numerical simulation in Sect. 5.

1.1 Mathematical Notations

Throughout this paper we will use bold-faced upper case letters to denote matrices, and bold-faced lower case letters for column vectors, light-faced letters for scalar quantities. The subscripts $\tau$, $H$, * indicate transpose, Hermitian transpose (transpose and complex conjugate), and complex conjugate respectively. $I_N$ denotes the $N \times N$ identity matrix. Also the inverse, Moore-Penrose pseudo inverse, trace, determinant, and Frobenius norm of the matrix $X$ are denoted by $X^{-1}$, $X^\dagger$, tr $X$, det $X$, and $\|X\|_F$, respectively. The $m$-th row and $n$-th column element of the matrix $X$ is denoted by $[X]_{m,n}$. $E_x$ means the expectation with respect to $x$. The bracket $\langle \cdot, \cdot \rangle$ is used for inner product on the column vector space. Also the vector Euclidean norm is expressed as $\| \cdot \|$. In this paper, since we often discuss correlations between each matrix element, it is convenient to treat matrix as one column vector that consists of all its elements. For any $m \times n$ matrix $A = [a_1 \ a_2 \ \cdots \ a_n]$, the vec operator generates a $mn \times 1$ vector defined as

$$\text{vec} \ A \triangleq [a_1^\top \ a_2^\top \ \cdots \ a_n^\top]^\top$$

(1)

where $\triangleq$ means definition. The Kronecker product $\otimes$ is required with the use of the vec operator.

2. System Model

2.1 MIMO Antenna Selection System

For a narrowband MIMO channel with $N_{Tx}$ transmit antennas and $N_{Rx}$ receive antennas, received vector $y \in \mathbb{C}^{N_{Rx}}$ can be expressed as

$$y = \sqrt{\frac{P_r}{N_{Tx}}} H_k x + n$$

(2)

where $P_r$ is average receive signal power at each receive antenna. $n \in \mathbb{C}^{N_{Rx}}$ is the additive noise vector typically assumed to have a white complex normal distribution with average power $\sigma_n^2$. $x \in \mathbb{C}^{N_{Tx}}$ is the normalized transmit vector such that $E_{\text{vec}} x x^H = I_{N_{Tx}}$. $H_k \in \mathbb{C}^{N_{Rx} \times N_{Tx}}$ is the normalized channel matrix for time instant $k$. Since $H_k$ varies for each time instant $k$, we define the normalization as $E_{\text{vec}} \|H_k\|_F^2 = N_{Tx}N_{Rx}$. In this formulation, average signal to noise ratio (SNR) per receive antenna can be expressed as follows:

$$\text{SNR} = \frac{E_{H_k \text{vec}} \left[ \sqrt{\frac{P_r}{N_{Tx}}} H_k x \right]^2}{E_{H_k \text{vec}} \|n\|^2} = \frac{P_r}{\sigma_n^2}$$

(3)

We employ antenna selection system only for receiver side with $N_{RF}$ RF chains satisfying $1 \leq N_{RF} < N_{Rx}$. If we connect the $i$-th RF chains to the $c_i$-th ($1 \leq c_i \leq N_{Rx}$) antenna element, only the $[c_i]_{N_{RF}} \cdot$-th row vectors of $H_k$ work for reception. If we define $\tau_k$ as a vector specifying connection of RF switches for transmission at time instant $k$ by putting together all $[c_i]_{N_{RF}} \cdot$ into one vector as $\tau_k \triangleq [c_1 \ c_2 \ \cdots \ c_{N_{RF}}]^\top$, the matrix employed for transmission can be written as

$$\begin{bmatrix}
[H_k]_{c_1,:} \\
[H_k]_{c_2,:} \\
\vdots \\
[H_k]_{c_{N_{RF}},:}
\end{bmatrix} = A_{\tau_k} H_k, \quad A_{\tau_k} \triangleq \sum_{i=1}^{N_{RF}} e_i e_{\tau_i}^\top$$

(4)

where $[H_k]_{c,:}$ means the $c$-th row vector of $H_k$, and $e_i, f_j$ are the so-called standard basis of $\mathbb{C}^{N_{Rx}}, \mathbb{C}^{N_{RF}}$, respectively. We call $\tau_k$ as the connection vector at $k$ for later discussion. Corresponding $A_{\tau_k}$ works as an extraction and permutation of row vectors interest on the time instant $k$. Similarly for antenna selection systems, channel model becomes $y = \sqrt{P_r/N_{Rx}} A_{\tau_k} H_k + n \ (y, n \in \mathbb{C}^{N_{Rx}})$.

Antenna selection is performed so that the Shannon capacity of the channel matrix $\sqrt{P_r/N_{Tx}} A_{\tau_k} H_k$ becomes larger. Since the true channel matrix is not available directly, we determine the connection $\tau_k$ by referring to the estimated channel matrix $\hat{H}_k$ instead. From the Foschini-Telatar equation [1], [2] for equal transmit power allocation, $\tau_k$ is determined such that

$$\tau_k = \arg \max_{\tau} \log_2 \det \left( I_{N_{RF}} + \frac{\text{SNR}}{N_{Tx}} A_{\tau} \hat{H}_k \hat{H}_k^H A_{\tau}^H \right)$$

subject to $\ \text{rank} \ A_{\tau} = N_{RF}$. (5)

The imposed condition $\text{rank} \ A_{\tau} = N_{RF}$ suggests that the selected antenna elements should not be overlapped. Practical method to get a perfectly optimal solution of the above equation is not yet discovered. In order to obtain a near-optimal solution with lower computational complexity, many antenna selection algorithms have been proposed [9]–[12]. Many of them require full instant channel state $\hat{H}_k$ to get selection for each time instant $k$. In order to keep the optimal selection in the time-varying environment, channel measurements must be frequently repeated to track the temporal changes.

2.2 Spatially Correlated MIMO Channel Model

We consider quasi-static block fading channel where the channel remains constant during one transmit block. Let us denote the channel state in the $k$-th fading block by $H_k$. 

Now we assume the channel is independent between each blocks. Hence we will suppress the time index \( k \) in \( H_k \) on such assumption.

We assume that all element of the \( H \) obeys multivariate complex normal distribution, therefore the second-order statistics provide enough information to characterize the model. We adopt the typical correlation model so-called the Kronecker model [15] which assumes that spatial Tx and Rx correlations are separable. This means that the channel is determined only by the Tx’s and Rx’s surrounding scattering environments, and they are independent each other. Hence the existence of dependence like direct-path waves is not considered. Generally, the Kronecker model is said to be suited to NLOS environments. Applicability of this model assumptions, if a sudden change of the spatial correlation statistics occurs, the correlations must be remeasured.

At equilibrium, it becomes

\[
\text{vec} H_k \sim \mathcal{CN}(0, R_{Tx} \otimes R_{Rx})
\]

regardless of initial channel state if \( \rho \neq 1 \).

### 3. Iterative Channel Estimation for Gauss-Markov Channel

#### 3.1 Channel Observation Model

We formulate the observation model of the channel state for training symbol aided channel estimation with MIMO antenna selection systems employed.

Let \( s_1^{(k)}, s_2^{(k)}, \ldots, s_N^{(k)} \in \mathbb{C}^{N_{RF}} \) be normalized training sequences, each of them being launched from Tx side in number order at fading block \( k \). They are normalized such that \( \| s_i \|^2 = N_t N_r \) where \( s_i = [s_1^{(k)} s_2^{(k)} \cdots s_N^{(k)}] \). Launched sequences are caught at Rx side with its RF switches connected as the connection vector \( \xi_k \) prepared for channel measurement. Then, the received sequences are written as

\[
[y_1^{(k)} y_2^{(k)} \cdots y_N^{(k)}] = \sqrt{\frac{p_{t}}{N_t}} A_{\xi_k} H_k [s_1^{(k)} s_2^{(k)} \cdots s_N^{(k)}] + [n_1 n_2 \cdots n_N]
\]

where \( n_i \in \mathbb{C}^{N_{RF}} \) is the \( i \)-th additive noise vector and \( y_i^{(k)} \) is the received vector corresponding to the transmitted vector \( s_i^{(k)} \). If we do not employ antenna selection systems, \( A_{\xi_k} \) shall be altered by \( I_{N_{RF}} \). As for the design of training sequences, if we have no \textit{a priori} knowledge of the channel, and also the noise is white gaussian, it has been proven that orthogonal training matrix defined as \( S_k S_k^H = N_t I_{N_{RF}} \) provides the best estimation capability [19].

By letting \( Y_k \) and \( N_k \) be \( N_{RF} \times N_t \) matrices each of them consisting of received vectors and noise vectors, respectively, (12) is rewritten as

\[
Y_k = A_{\xi_k} H_k S_k + N_k
\]
the observation noise covariance matrix defined as $R_k \triangleq E_{\mathbf{v}_k} \vec{v}_k \vec{v}_k$. $K_k$ and $Z_k$ are called Kalman gain and innovation term, respectively. $P_{k|k}$ and $P_{k|k-1}$ are error covariance matrices defined as

$$P_{k|l} \triangleq E_{\mathbf{u}_k} \vec{H}_k \vec{H}_k^H (\vec{H}_k - \bar{H}_k \mathbf{u}_k)^H$$

where $l \in \{k, k-1\}$. In general formulation of the Kalman filter, $R_k$ and $Q_k$ can be dependent on $k$. But we assume they remain constant during the period of our interest, hence we will omit the index $k$ from now on. If the noise is white gaussian, letting $\mathbf{R} = \sigma_n^2 I_{N_{\text{RF}}}$ and substituting (18), (19) into (21), and with the aid of the matrix inversion lemma, (21) is rewritten as follows:

$$P_{k|k} = \left( P_{k|k-1}^{-1} + \frac{1}{\sigma_n^2} S_k^T S_k \otimes A_k^H A_k \right)^{-1} \left( \frac{P_{k|k}^{-1}}{\sigma_n^2} + \frac{1}{\sigma_n^2} S_k^T S_k \otimes A_k^H A_k \right)$$

In this paper, we assume model parameters $\rho$ and $Q$ are known in advance whereas they are not available in practical situations. Therefore we must get the estimate of them and utilize them instead of their true values. For the model parameter estimation, we require some sequences of full channel matrix measured in training period provided right before the use of the Kalman filter. A concrete estimation method is described in Appendix A.

### 4. Active Enhancement of Estimation

In this section, we incorporate the concept of active learning into our estimation scheme. Generally in linear inverse problems, estimation capability highly depends on the observation matrix in (14). In this case, since observation matrix is composed of connection of RF switches $\xi_k$ and transmitted training sequences $S_k$, we can enhance estimation quality by optimally designing $\xi_k$ and $S_k$ adaptively to channel environments for each time instant $k$.

#### 4.1 Problem Formulation

Let us define the MSE criteria as $J_k$. From (22), it is same as a trace of error covariance matrix kept in the Kalman filter. For observation matrix $S^T \otimes A_k$, $J_k$ is written as a functional with respect to $\mathbf{z}$ and $\mathbf{S}$:

$$J_k[\mathbf{z}, \mathbf{S}] \triangleq E_{\mathbf{v}_k} \left[ \| H_k - \bar{H}_k \mathbf{u}_k \|_F^2 = \text{tr} \ P_{k|k} \right]$$

In order to obtain the best estimate, $\mathbf{z}_k$ and $S_k$ should be chosen such that

$$(\mathbf{z}_k, S_k) = \arg \min_{\mathbf{z}, S} J_k[\mathbf{z}, \mathbf{S}],$$

subject to: $\text{rank} \ A_k = N_{\text{RF}}$, $\| S \|_F^2 = P_{k|k}$. (25)

Finding the solution to this conditional optimization problem requires much computational complexity. But relying
4.2 Optimal Training Sequence Design without Antenna Selection System

In this section, we consider the case in which antenna selection system is not employed. In this case, all $A_{\ell s}$ in equations shall be replaced by $I_{N_{as}}$.

4.2.1 Absence of Temporal Correlation ($\rho = 0$)

$\rho = 0$ means the estimation is not affected from the previous channel state, and in this case, the situation becomes identical to the MMSE estimation by Wiener filter. Optimal transmit signal design assuming this situation is already discussed in [3]. The optimal training sequence becomes the weighted eigenvectors of the spatial correlation matrix of Tx side whose power allocation is determined by water filling solution [14].

4.2.2 Presence of Temporal Correlation ($0 < \rho < 1$)

This case corresponds to a simple extension of [3] to Gauss-Markov channel model. Let us denote eigenvectors and eigenvalues of $R_{Tx}$ and $R_{Rx}$ as

$$R_{Tx}u_i = \lambda_i^{(Tx)}u_i, \quad R_{Rx}v_j = \lambda_j^{(Rx)}v_j,$$

where $\lambda_i^{(Tx)}$ and $\lambda_j^{(Rx)}$ are sorted by descending order. If we assume a structure of $S_k$ as

$$S_k = \sum_{l=1}^{N_k} \sqrt{\alpha_i^{(k)}} u_i^* \varphi^T,$$

where $\{\varphi_i\}_{i=1}^{N_k}$ is any subset of arbitrary orthonormal basis of $\mathbb{C}^{N_k}$, and $\alpha_i^{(k)}$ is non-negative value specifying power allocation along the $u_i$, then all updating equations of the error covariance matrix can be calculated independently for each component along each eigen vector, and $P_{kj}$ can be always expressed as

$$P_{kj} = \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} [\Gamma_{kj}]_{ij} u_i u_i^\dagger \otimes v_j v_j^\dagger$$

(28)

where $l \in \{k, k-1\}$, and $\Gamma_{kj} \in \mathbb{R}^{N_{as} \times N_{as}}$ contains all eigenvalues characterizing $P_{kj}$. $[\Gamma_{kj}]_{ij}$ is the eigenvalue for the eigenvector $u_i \otimes v_j$. Now, updating equations are reduced to only arithmetic operations on the each element of $\Gamma_{kj}$, and hence much computations are saved. Inversely, if $P_{kj-1}$ is expressed as (28), the structure of the optimal training sequence becomes as a form of (27) (see Appendix C).

In this case, (16), (18), (19), and (21) can be rewritten with much smaller computations as follows:

$$\Gamma_{kj-1} = \rho^2 \Gamma_{k-1,k-1} + (1 - \rho^2) [\lambda_1^{(Tx)} \lambda_2^{(Tx)} \cdots \lambda_{N_k}^{(Tx)} | \lambda_1^{(Rx)} \lambda_2^{(Rx)} \cdots \lambda_{N_k}^{(Rx)}]$$

(29)

$$C_k = \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} (a_i^{(k)} \Gamma_{kj-1} \lambda_i + \alpha_i^2) \varphi_i^{\dagger} u_i \otimes v_j v_j^\dagger$$

(30)

$$K_k = \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} \sqrt{\alpha_i^{(k)}} [\Gamma_{kj-1}]_{ij} \lambda_i u_i \varphi_i^{\dagger} \otimes v_j v_j^\dagger$$

(31)

$$[\Gamma_{kj}]_{ij} = \frac{\alpha_i^2}{\alpha_i^{(k)}} [\Gamma_{kj-1}]_{ij} \lambda_i + \alpha_i^2$$

(32)

The sum power constraint $\|S_k\|^2 = P_s N_s$ is now reduced to $\sum_{i=1}^{N_k} \alpha_i^{(k)} = P_s N_s$. Now, we want to minimize $J_k = \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} [\Gamma_{kj}]_{ij}$, but to simplify the problem, we will consider the minimization only for the first column of $\Gamma_{kj}$. From (29) to (32), we can confirm that if the initial value $\Gamma_{01}$ satisfies $[\Gamma_{01}]_{11} \geq [\Gamma_{01}]_{12} \geq \cdots \geq [\Gamma_{01}]_{1N_k}$, it always holds for all $k$ that $[\Gamma_{kj}]_{11} \geq [\Gamma_{kj}]_{12} \geq \cdots \geq [\Gamma_{kj}]_{1N_k}$. Hence we can say that the first column of $\Gamma_{kj}$ has the most important effect on our criteria, and we can determine the upper bound of $J_k$ which only depends on the first column of $\Gamma_{kj}$ as follows:

![Fig. 2 Schematic block diagram of the whole system model.](image-url)
calculations are rewritten as:

\[ J_k[S_k] = \sum_{i=1}^{N_{\text{tx}}} \sum_{j=1}^{N_{\text{rx}}} [\Gamma_{kj}]_{ij} \leq N_{\text{rx}} \sum_{i=1}^{N_{\text{tx}}} [\Gamma_{kj}]_{ii} = J'_k[S_k] \]  

(33)

Minimizing \( J'_k \) can be solved by using Lagrange multipliers and Kuhn-Tucker conditions. \( \{a_i^{(k)}\}_{i=1}^{N_{\text{tx}}} \) is determined as

\[ a_i^{(k)} = \left[ W - \frac{\sigma_i^2}{[\Gamma_{kj}]_{ii}} \right]^+ \]  

(34)

where \([x]^+ = \max(0, x)\) and \(W\) is so-called the water line calculated by the water filling algorithm.

At equilibrium, \( \Gamma_{kj} \) converges to the solution of the algebraic Riccati equation if the system is observable and controllable. In this paper, we will not discuss such steady-state Kalman filter, but it might be possible to calculate the optimal training sequence and corresponding filter gains in advance, prior to the actual iterative observations.

4.3 Optimal Measurement Antenna Selection

Now we consider the case of antenna selection system with its channel matrix being partially observed. A selection of partial measurement \( \bar{\xi}_k \) is optimized for each iteration under the criteria of (25). In this time, \( S_k \) is not jointly optimized due to simplicity.

A naive implementation of (16) to (21) requires calculation of inverse of the \( N_{\text{tx}}N_{\text{rx}} \times N_{\text{tx}}N_{\text{rx}} \) matrix for each step which results in much requirement of computations. Now we show that if we do not optimize on \( S_k \), and also it is given as a form of (27), we can implement the Kalman filter with smaller computations by avoiding the direct calculation of the inverse of the large error covariance matrix.

On this assumption, \( F_{kl} \) is always expressed as

\[ P_{kl} = \sum_{i=1}^{N_{\text{tx}}} u_i u_i^H \otimes F_{kl}^{(i)} \]  

(35)

where \( F_{kl}^{(i)} \) is \( N_{\text{rx}} \times N_{\text{rx}} \) matrix characterizing error covariance of Rx side which corresponds to Tx’s component along \( u_i \). Then, thanks to the relationship of Appendix B, filter calculations are rewritten as:

\[ F_{kl}^{(i)} = \rho^2 F_{kl}^{(i)} + (1 - \rho^2) a_i^{(i)}(\mathbf{r}_x)^\dagger R_{\text{rx}} \]  

(36)

\[ K_k = \sum_{i=1}^{N_{\text{tx}}} u_i \varphi_i^H \otimes \sqrt{a_i^{(k)}} A_{\xi_k}^H \]  

\[ (a_i^{(k)} A_{\xi_k} F_{kl}^{(i)} A_{\xi_k}^H + \sigma_i^2 I_{N_{\text{rx}}})^{-1} \]  

(37)

\[ F_{kl}^{(i)} = \left[ (F_{kl}^{(i)})^\dagger + \frac{a_i^{(k)}}{\sigma_i^2} A_{\xi_k}^H A_{\xi_k} \right]^{-1} \]  

(38)

By comparing (38) with (23), we can see the update of \( F_{kl}^{(i)} \) is done by replacing the observation matrix by \( A_{\xi_k} \) with its noise power being scaled by \( 1/\alpha_i^{(k)} \). From (35), selection criteria is rewritten as follows:

\[ J_k [\xi_k, S_k] = \sum_{i=1}^{N_{\text{tx}}} \text{tr} F_{kl}^{(i)} \]  

(39)

We must compromise the best solution of \( \xi_k \), due to the required computations for optimization. In order to find a near optimal solution, we resort to Greedy algorithm which starts from an empty set of selected antennas, then add one antenna which best contributes to the selection criteria per step, continues until it reaches the desired number of selection. We show the implementation of the Greedy algorithm with much simpler calculations by utilizing a similar approach proposed in fast antenna subset selection [9].

Now we consider the \((t + 1)\)-th step of Greedy algorithm. Until the \( t \)-th \((0 \leq t < N_{\text{rf}}) \) step, \( t \) elements have been already selected, and we will denote \( \omega_t \) as a \( t \times 1 \) vector that contains already selected elements. Also we denote the extracted channel matrix specified by \( \omega_t \) as \( \sqrt{P_t/N_{\text{tx}}}A_{\omega_t}H_k \). \( A_{\omega_t} \) is a \( t \times N_{\text{rx}} \) matrix defined as \( A_{\omega_t} = \sum_{i=1}^{t} e_i e_i^H, \) where \( e_i \) and \( f_i \) are the \( i \)-th standard basis of \( \mathbb{C}^{N_{\text{rx}}} \) and \( \mathbb{C}^t \), respectively. If we select \( x \)-th element at the \((t + 1)\)-th step, \( x \) is chosen so that

\[ J_k [\omega_{t+1}, S_k] = J_k [\omega_{t}^x x^\dagger, S_k] \]  

(40)

is minimized. Substituting the relationship of

\[ A_{\omega_t}^H A_{\omega_{t+1}} = A_{\omega_t}^H A_{\omega_t} + e_t e_t^H \]  

(41)

into (40), and thanks to the matrix inversion lemma, we have

\[ \Phi_{\omega_t}^{(i)} = \left( \Phi_{\omega_t}^{(i)} \right)^{-1} + \frac{\alpha_i}{\sigma_i^2} e_t e_t^H \]  

(42)

where \( \Phi_{\omega_t}^{(i)} = F_{kl}^{(i)}(i-1) \). Finally, the element which best contributes to the criteria (40) is chosen as follows:

\[ x_{\text{opt}} = \arg \min_x J_k [\omega_{t}^x x^\dagger, S_k] \]  

(43)

\[ = \arg \max_x \sum_{i=1}^{N_{\text{tx}}} \left[ \frac{\left\| \Phi_{\omega_t}^{(i)} x_{\text{opt}} \right\|^2}{\sigma_i^2/\alpha_i + \left\| \Phi_{\omega_t}^{(i)} x_{\text{opt}} \right\|^2} \right] \]

After that, the selection result is stored into \( \omega_{t+1} \) as \( \omega_{t+1} \leftarrow [\omega_{t}^x x_{\text{opt}}]^\dagger \). Also \( \Phi_{\omega_{t+1}}^{(i)} \) is generated by (42) and be used for the selection of the next step.

5. Simulation

We evaluated the effectiveness of the proposed scheme by the numerical simulations with different parameters for spatial and temporal correlation. As a measure of the estimation capability, we employed the following three criteria.

- Evaluate the Frobenius norm of the estimation error,
defined as (24).

- After the estimation by the Kalman filter, antenna subset used for transmission is chosen according to the information of the estimation result. Then the transmission starts by directly using the estimated channel state information. In this case, the contributing factor for the channel capacity degradation will be the validity of antenna selection and the decrease of SINR, both are caused by the estimation error by the Kalman filter.

- After the antenna selection based on the estimated channel, precise channel estimation for which only the elements used for transmission is repeated with much longer training sequences and without using any channel statistics. In this case, only the validity of antenna selection can be evaluated.

The Shannon capacity of the MIMO channel is evaluated considering the estimation error of the channel under the MIMO eigenmode transmission system. Let singular value decomposition of the estimated channel matrix be $A_r \tilde{H} = USV^H$, where $U$ and $V$ are unitary matrices and $S$ is the diagonal matrix, respectively. Then the diagonal element of the $T = U^H A_r \tau_i H V$ stands for the SNR for each sub channel, and the $[T]_{i,j}$ is a leak from the $i$-th sub channel to the $j$-th sub channel which is caused by the estimation error. The SINR of the $i$-th sub channel is expressed as

$$\text{SINR}_i = \frac{P_r [T]_{i,i} [T]_{i,i}^*}{N_{Tx} \sigma_n^2 + P_r \sum_{j \neq i} [T]_{i,j} [T]_{i,j}^*}. \quad (44)$$

From this, the Shannon capacity of the MIMO channel is derived as $C = \log_2 \prod_{i=1}^{N_s} (1 + \text{SINR}_i)$.

The spatial correlation matrices of the Kronecker model was generated under the assumption that the waves incoming from the different angle is uncorrelated, and also they have the same power. Let $a(\theta)$ be a steering vector of the array antenna for a plane wave coming from the direction of $\theta$. The spatial correlation matrix is generated as

$$R = \frac{1}{\Delta \theta} \int_{\theta_0 - \Delta \theta/2}^{\theta_0 + \Delta \theta/2} a(\theta) a(\theta)^H d\theta \quad (45)$$

where $\theta_0$ is the center of the arrival wave and $\Delta \theta$ is the angular spread. We assume the linear array whose distance between elements is half of the wavelength. $\theta_0$ is chosen randomly for 8 times, and the resultant estimation error was averaged over them. We adopt $\Delta \theta$ as the indicator of the degree of spatial correlation.

For each simulation parameters, an iteration of the Kalman filter is repeated 60 times, and the result is averaged over them. The average SNR is set to 15 dB. The number of the Tx elements $N_{Tx}$ is fixed to 6. Also the length of the training sequence $N_t$ is set to 32.

5.1 Simulation Procedure

The simulation procedure is shown below.

1. Initialize the channel model parameters including the correlation matrices generated as (45).
2. For the first $N_t$ fading blocks, this period is devoted to learn the channel statistics required for the Kalman filter. The estimation method is described in Appendix A.
3. Update the true channel state by (10).
4. From $P_{k-1}$, optimal training sequence $S_k$ is calculated by (27) and (34). In the case of the antenna selection system, $\xi_k$ is chosen by (43).
5. Based on $\xi_k$ and $S_k$, channel is observed by (14).
6. Obtain the estimate of the channel by the Kalman filter. Also $P_{k|k}$ in the filter is updated.
7. Evaluate the estimation error $H_k - \tilde{H}_k$.
8. Based on $\tilde{H}_k$, the connection $\tau_k$ is chosen by the fast antenna subset selection algorithm [9].
9. Evaluate the Shannon capacity of $A_{\tau_k} \tilde{H}_k$ by using (44).
10. In order to evaluate only the antenna selection capability, Evaluate the capacity of $A_{\tau_k} H_k$ as well.
11. Back to step 3 until the required number of samples obtained.

In step 2 of the procedure, parameters of the Gauss-Markov model shall be estimated by using the $N_t$ times measurements of the channel. Since the period that channel parameters remain stationary is limited, we must learn its statistics as fast as possible. Therefore $N_t$ should not be too large as long as the required precision is obtained. Considering them, we choose 10 as $N_t$ for the estimation of the spatial correlation matrices. However, as for the temporal correlation coefficient $\rho$, a precise estimate cannot be obtained until the $N_t$ is large enough. Therefore for the estimation of $\rho$, we selected $N_t$ as 500. In this case, the estimation error becomes approximately less than 0.01. Later, we will show that the channel estimation capability does not depend very much on the precision of the estimate of $\rho$. Therefore, by sacrificing the precision of $\rho$, there might be another way that can reduce the required $N_t$ without losing estimation capability significantly.

5.2 Effect of the Spatial Correlation

Figure 3 depicts the mean squared error defined as (24) with different spatial correlation for various temporal correlations. The antenna selection system is not employed, hence the equations of (29) to (32) are used. The solid lines and the dotted lines indicate the case of optimal training sequence design by the water-filling solution, and the case of orthogonal training sequence (i.e. $\alpha^{(0)} = P, N_t/N_{Tx}$), respectively. We also compared them with the case of maximum likelihood (ML) channel estimation [17], [18] without any channel statistics. An advantage of the optimal training sequence design was observed especially under a presence of stronger spatial correlations of the Tx side and weaker temporal correlations. The reason of this seems that these conditions tend to yield nonuniform error variances among each fading block which work advantageously for the opti-
mal training sequence design.

5.3 Effect of Correlation Model Mismatch

Although the proposed method assumes the Kronecker model which is applicable to NLOS environments, if a true channel obeys LOS model, the model mismatch might cause a degradation of estimation capability. In order to verify this, we generated a true channel by typical LOS channel model which considers specular and diffuse components. The channel was generated as

$$H = \sqrt{\frac{K}{K + 1}} \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} g_i \cdot a_{Rx}(q_{i}^{Rx}) \left[ a_{Tx}(q_{i}^{Tx}) \right]^T}$$

$$+ \sqrt{\frac{1}{K + 1}} R_{DiffRx}^{1/2} G \left( R_{DiffTx}^{1/2} \right)^T$$

(46)

where $K$ is the Ricean K factor, $N_p$ is a number of specular paths, and $\{g_i\}_{i=1}^{N_p} \sim \mathcal{CN}(0, 1)$. $a_{Tx}$ and $a_{Rx}$ are steering vectors for each side. $R_{DiffTx}$ and $R_{DiffRx}$ are correlation matrices for diffuse component. The DOAs $[\theta_{i}^{Tx}]_{i=1}^{N_p}$ and $[\theta_{i}^{Rx}]_{i=1}^{N_p}$ are generated randomly within the range of the angular spread.

For various K factors, estimation error was evaluated in Fig. 4 using the optimal training sequence design. Since the diffuse component was modeled as the Kronecker model, $K = 0$ matches exactly to our original assumption. Contrary to our expectations, the method works more effectively as the specular component becoming stronger. It is considered that the presence of specular component yield the strong spatial correlation which is advantageous to the proposed method rather than the disadvantage due to the model mismatch.

5.4 Effect of the Temporal Correlation

From now on, we employ the antenna selection system of $N_{Rx} = 7$ and $N_{RF} = 3$ or 6. The channel estimation is done by the equations of (36) to (38). Figure 5 show the degraded capacity with the change of temporal correlation of the actual channel. From the figure, by utilizing the estimated channel by the Kalman filter, almost perfect selection can be achieved, but the slight selection error is observed especially for the lower correlations. If we directly employ the estimated channel for the transmission, the capacity is highly influenced by the temporal correlation. We also show the result obtained by the ML channel estimation without exploiting any channel statistics. In order to compare under a fair condition, we imposed the received signal power of the training sequence being equal among each cases. In the case of $N_{RF} = 6$, estimation by Kalman filter always exceeds the one without channel statistics. As for the case of $N_{RF} = 3$, since over half of the number of variables is not observed directly, the estimation error becomes significant. We can read the large degradations with the correlations below 0.95 mainly caused by the leaks between sub channels.
5.5 Effect of the Measurement Antenna Selection

We investigated the effect of the optimal antenna selection for the channel measurement. Figure 6 shows the mean squared error of the channel matrix for different spatial correlations of the Rx side. For comparison, we also show the case of random selection which means \( \xi_k \) is determined randomly for each iteration.

The estimation quality of the optimal selection is always higher than the other method. Difference in estimation error becomes large when the spatial correlations are small, and also when the number of the not measured elements \( (N_{\text{Rx}} - N_{\text{RF}}) \) is large.

5.6 Robustness of the Estimate of \( \rho \)

Estimation error of the channel statistics causes the degradation of the channel estimation capability. In Fig. 7, the channel is generated with the temporal correlation coefficient being fixed to 0.97, although the estimator assumes its correlation coefficient as a value on the horizontal axis. The capacity with the estimated CSI seems not to be degraded even when the estimation error increases to 0.07. We can observe the robustness of the estimation of channel statistics, but from the discussions so far, it is shown that the estimation is sensitive to the actual temporal correlation itself.

6. Concluding Remarks

We proposed the optimal training sequence design for channel estimation by the Kalman filter. Also in order to reduce the measurement costs for antenna selection systems, partial channel measurement and its optimal control based on the statistics were discussed.

The numerical simulation assuming the Gauss-Markov channel revealed that the optimal training sequence design works effectively in the case of lower temporal correlation and higher spatial correlation of Tx side, although the partial channel measurement has advantage where higher temporal correlation and higher spatial correlation of Rx side. If the partial channel measurement was employed, the estimated channel is precise enough as the coarse estimation which is utilized for only the selection criteria. Meanwhile, if we utilize the estimated channel to the transmission directly, the estimation error causes the degradation of the channel capacity to a large extent. In that case, it was shown that high spatial and temporal correlations are required to keep the accuracy of the method.

As future tasks, this method has a shortage that comes from the definition of the channel model about a frequent temporal change of spatial correlation statistics. This might not be negligible especially in rapidly mobile wireless scenarios. Therefore, now we are investigating a more robust estimation scheme for antenna selection systems which have a less dependence on the spatial correlations.

References

We require \( N_t \) sequences of full channel matrix in order to estimate the parameters in the Kalman filter. Estimates \( \hat{\rho} \) and \( \hat{Q} \) are given as:

\[
\hat{\rho} = \frac{1}{N_{tx}N_{rx}(N_t - 1)} \sum_{i=1}^{N_t} \sum_{j=1}^{N_t-1} \frac{[H_{k+1}]_{i,j} [H_k]_{i,j}}{[H_{k+1}^H]_{i,j} [H_k^H]_{i,j}}
\]

(A-1)

\[
2(1 - \rho) \hat{Q} = 2(1 - \rho)E_k \text{ vec } X_k (\text{vec } X_k)^H
\]

\[
= (1 - \rho)^2 E_k \text{ vec } H_k (\text{vec } H_k)^H
\]

\[
+ (1 - \rho^2)E_k \text{ vec } X_k (\text{vec } X_k)^H
\]

(A-2)

\[
\text{vec } H_k (\text{vec } H_k)^H = E_k \text{ vec } X_k (\text{vec } X_k)^H
\]

In addition, if channel obeys (10), spatial correlation matrices of the Kronecker model \( \tilde{R}_{tx} \) and \( \tilde{R}_{rx} \) can be estimated as follows:

\[
\tilde{R}_{tx} = \sum_{k=1}^{N_t-1} (H_k - H_{k+1})^H (H_k - H_{k+1})
\]

(A-3)

\[
\tilde{R}_{rx} = \sum_{k=1}^{N_t-1} (H_k - H_{k+1}) (H_k - H_{k+1})^H
\]

(A-4)

\[
\tilde{R}_{tx} = N_{tx} \bar{R}_{tx} / (tr \bar{R}_{tx}), \quad \tilde{R}_{rx} = N_{rx} \bar{R}_{rx} / (tr \bar{R}_{rx})
\]

(A-5)

Appendix B

Let \( \{\varphi_i\}_{i=1}^M \) be any subset of arbitrary orthonormal basis of \( \mathbb{C}^N (N \geq M) \), and \( \{X_i\}_{i=1}^M \) be a set of any matrices of the same size, following relationship holds:

\[
\left( \sum_{i=1}^M \varphi_i \varphi_i^H \otimes X_i \right) = \sum_{i=1}^M \varphi_i \varphi_i^H \otimes X_i
\]

(A-6)

(Proof) Let \( \{A_i\}_{i=1}^M \) and \( \{B_i\}_{i=1}^M \) be sets of any matrices of the conforming size, then we have:

\[
\left( \sum_{i=1}^M \varphi_i \varphi_i^H \otimes A_i \right) = \sum_{i=1}^M \varphi_i \varphi_i^H \otimes B_i
\]

(A-7)

From the above relationship, we can confirm that the right side of (A-6) satisfies the four conditions of the Moore-Penrose generalized inverse, i.e. \( XX^*X = X \), \( X'XX' = X' \), \( XX' = (XX')^H \), and \( X'X = (X')^H \).

Appendix C

Let singular value decomposition of \( S_k^T \) be \( \Phi \Sigma \Psi^H \) where \( \Phi \), \( \Psi \) and \( \Sigma \) are \( N_t \times N_t \), \( N_{tx} \times N_{tx} \) unitary matrices and \( N_t \times N_{tx} \) diagonal matrix, respectively. Substituting it into (24) yields,

\[
J_k[S_k] = \text{tr} \left( P_{k-1}^{-1} + \frac{1}{\sigma_n^2} \Psi \Sigma^H \Psi^H \otimes I_{N_{rx}} \right)^{-1}
\]

(A-8)

This implies that \( \Phi \) can be chosen arbitrarily. For any \( N_{rx} \times N_{rx} \) unitary matrix \( V \), \( J_k \) can be transformed into:

\[
J_k = \text{tr} \left( \Psi \otimes V \right)^H P_{k-1}^{-1} (\Psi \otimes V) + \frac{1}{\sigma_n^2} \Sigma \Sigma^H \otimes I_{N_{rx}} \right)^{-1}
\]

(A-9)
It is proven that $J_k$ is minimized if inside of $[\cdot]$ in (A-9) becomes diagonal form ([3] : Theorem 1). If $P_{l,k-1}$ have the form of (28), diagonalization of (A-9) is achieved if and only if $\Psi = [u_1 u_2 \cdots u_{N_{\text{Rx}}}^T]$ and $V = [v_1 v_2 \cdots v_{N_{\text{Rx}}}^T]$. Denoting the $i$-th column vector of $\Phi$ as $\varphi_i$, and $\Sigma^H \Sigma = \text{diag} [\alpha_1 \alpha_2 \cdots \alpha_{N_{\text{Rx}}}]$ yield (27).

Yousuke Naruse was born in 1981. He received the B.E. and M.E. degrees from Tokyo Institute of Technology, Japan, in 2003 and 2006 respectively. He joined FUJIFILM Corporation in 2006, and engages in medical image processing. He is currently working towards the D.E. degree at Tokyo Institute of Technology.

Jun-ichi Takada received B.E. and D.E. degrees from Tokyo Institute of Technology, Japan, in 1987 and 1992, respectively. From 1992 to 1994, he has been a Research Associate at Chiba University, Chiba, Japan. From 1994 to 2006, he has been an Associate Professor at Tokyo Institute of Technology, Tokyo, Japan. Since 2006, he has been a Professor at Tokyo Institute of Technology. His current research interests are wireless propagation and channel modeling, antennas and antenna systems for wireless applications, and cognitive radio technology. He received Achievement Award from IEICE in 2008. He is a member of IEEE, ACES, and ECTI Association Thailand.